Multi-parameter wavefield reconstruction inversion with the alternating direction method of multipliers (ADMM) and compound regularization

Seyed Hossein Seyed Aghamiry

UNIVERSITÉ CÔTE D'AZUR

ÉCOLE DOCTORALE
SCIENCES FONDAMENTALES ET APPLIQUÉES

University of TEHRAN

THÈSE DE DOCTORAT

Imagerie sismique multi-paramètre par reconstruction de champs d'ondes: apport de la méthode des multiplicateurs de Lagrange avec directions alternées (ADMM) et des régularisations hybrides

Multi-parameter Wavefield Reconstruction Inversion with the Alternating Direction Method of Multipliers (ADMM) and Compound Regularization

Hossein S. Aghamiry

Institut de géophysique/UMR Géoazur

Présentée en vue de l'obtention
du grade de docteur en Sismologie/Sciences de la Planète et de l'Univers
de l'université de Téhéran
et de l'université Côte d'Azur
Dirigée par : Ali Gholami / Stéphane Operto
Soutenue le : 18 décembre 2019

Devant le jury, composé de :
Josselin Garnier, Pr., Ecole Polytechnique, Rapporteur
Mohammad-Reza Eslahchi, Pr., Modares University, Rapporteur
Tristan van Leeuwen, Pr., Utrecht University, Rapporteur
Hamid-Reza Siahkoohi, Pr., University of Tehran, Examinateur
Didier Auroux, Pr., Université Côte d'Azur, Examinateur
Arash Amini, Pr., Sharif University of Technology, Examinateur
Mehrdad Pakzad, Pr., University of Tehran, Examinateur

SEISCOPE

Imagerie sismique multi-paramètre par reconstruction de champs d'ondes: apport de la méthode des multiplicateurs de Lagrange avec directions alternées (ADMM) et des régularisations hybrides

Multi-parameter Wavefield Reconstruction Inversion with the Alternating Direction Method of Multipliers (ADMM) and Compound Regularization

Jury:

Président du jury

Hamid-Reza Siahkoohi, Professeur, University of Tehran

Rapporteurs

Josselin Garnier, Professeur, Ecole Polytechnique
Mohammad-Reza Eslahchi, Professeur, Modares University
Tristan van Leeuwen, Professeur, Utrecht University

Examinateurs

Didier Auroux, Professeur, Université Côte d'Azur
Arash Amini, Professeur, Sharif University of Technology
Mehrdad Pakzad, Professeur, University of Tehran

Acknowledgement

First of all, I would like to express my sincere gratitude to my supervisors Ali Gholami and Stéphane Operto for the continuous support of my Ph.D study and related research, for their patience, motivation, and immense knowledge. Their guidance helped me in all the time of research and writing of my papers and this thesis. I could not have imagined having a better advisors and mentors for my Ph.D. study. Many thanks also to Hamid Reza Siahkoohi, who was my counselor and I have opportunity for many fruitful discussions with him during my first part of Ph.D. study in university of Tehran.

Josselin Garnier, Mohammad-Reza Eslahchi, and Tristan van Leeuwen reviewed my thesis and Didier Auroux, Arash Amini, and Mehrdad Pakzad accepted to be a part of my defense jury. I would like to thank them for their time, insightful comments and suggestions which helped to improve this manuscript.

I would like to thank the SEISCOPE consortium (*http://seiscope2.osug.fr*) to fund two years of my stay in France. Also, I thank the IDEX UCA JEDI with the WIMAG project to fund the last three months of my Ph.D.

I would like to thank Alessandra Ribodetti for her support, her kindness, and her helpful advice about attenuation and anisotropy. I thank Jean Virieux, Romain Brossier, Ludovic Métevier and all the SEISCOPE members in Grenoble for the fruitful discussions we had during the meetings and for their continuous support. I would like to thank Rene-Edouard Plessix, Jacques Blum and Didier Auroux for all the useful discussions about data-assimilation and time-domain extension of our method.

I would like to thank all the members of Geoazur laboratory for providing such a comfortable environment to work. I specially thank Valérie Mercier-Valero, Véronique Large, Jean-Xavier Dessa, Lionel Maurino, David Chapeau and Caroline Ramel for all their help and considerations. I also thank all the members of Institute of Geophysics of University of Tehran who helped me during my Ph.D. study. I specially thank Aghavali Zabanbor. Always his motivation was an inspiration to me. Also, I would like to thank Mr. Bistoni, Mrs. Taheri and Mrs. Keshtkar.

I would like to acknowledge my friends in Institute of Geophysics and Geoazur for their kindness and pure friendship. I would like to thank Borhan Tavakoli for his help and for introducing me to Stéphane . I would like to express my appreciation to Serge Sambolian for all the fun and all the discussions we had together.

My greatest appreciations go for my father, mother, Samira, Reza, Najmeh, Narjes and brothers in law, Mehdi and Ehsan. Words can not express what they have done for me and how much I love them. I thank my family for their long distance but limitless support and encouragement that yield a constant and great motivation during all of my long-term educations, especially these last years. Samira never left me alone in the tough moments and she was always my inspiration. She has been extremely supportive of me throughout this entire process and has done her best to help me get to this point. Lastly, it is my lovely and lively nieces Mahan, Iman, Hania and Hana who have been the source of my liveliness.

I sincerely apologize if I forgot some names, but there were many others who helped me during my Ph.D.

Dedicated to the memory of my grandmother.

Abstract

Full Waveform Inversion (FWI) is a PDE-constrained optimization which reconstructs subsurface parameters from sparse measurements of seismic wavefields. FWI generally relies on local optimization techniques and a reduced-space approach where the wavefields are eliminated from the variables. In this setting, two bottlenecks of FWI are nonlinearity and ill-posedness. One source of nonlinearity is cycle skipping, which drives the inversion to spurious minima when the starting subsurface model is not kinematically accurate enough. Ill-posedness can result from incomplete subsurface illumination, noise and parameter cross-talks. This thesis aims to mitigate these pathologies with new optimization and regularization strategies. I first improve the wavefield reconstruction method (WRI). WRI extends the FWI search space by computing wavefields with a relaxation of the wave equation to match the data from inaccurate parameters. Then, the parameters are updated by minimizing wave equation errors with either alternating optimization or variable projection. In the former case, WRI breaks down FWI into two linear subproblems thanks to wave equation bilinearity. WRI was initially implemented with a penalty method, which requires a tedious adaptation of the penalty parameter in iterations. Here, I replace the penalty method by the alternating-direction method of multipliers (ADMM). I show with numerical examples how ADMM conciliates the search space extension and the accuracy of the solution at the convergence point with fixed penalty parameters thanks to the dual ascent update of the Lagrange multipliers. The second contribution is the implementation of bound constraints and non-smooth Total Variation (TV) regularization in ADMM-based WRI. Following the Split Bregman method, suitable auxiliary variables allow for the decoupling of the $\ell 1$ and $\ell 2$ subproblems, the former being solved efficiently with proximity operators. Then, I combine Tikhonov and TV regularizations by infimal convolution to account for the different statistical properties of the subsurface (smoothness and blockiness). At the next step, I show the ability of sparse promoting regularization in reconstructing the model when ultra-long offset sparse fixed-spread acquisition such as those carried out with OBN is used. This thesis continues with the extension of the ADMM-based WRI to multiparameter reconstruction in vertical transversely isotropic (VTI) acoustic media. I first show that the bilinearity of the wave equation is satisfied for the elastodynamic equations. I discuss the joint reconstruction of the vertical wavespeed and epsilon in VTI media. Second, I develop ADMM-based WRI for

attenuation imaging, where I update wavefield, squared-slowness, and attenuation in an alternating mode since the viscoacoustic wave equation can be approximated, with a high degree of accuracy, as a multilinear equation. This alternating solving provides the necessary flexibility to Taylor the regularization to each parameter class and invert large data sets. Then, I overcome some limitations of ADMM-based WRI when a crude initial model is used. In this case, the reconstructed wavefields are accurate only near the receivers. The inaccuracy of the phase of wavefields may be the leading factor that drives the inversion towards spurious minimizers. To mitigate the role of the phase during the early iterations, I update the parameters with phase retrieval, a process that reconstructs a signal from the magnitude of linear measurements. This approach, combined with efficient regularizations, leads to a more accurate reconstruction of the shallow structure, which is decisive to drive ADMM-based WRI toward good solutions at higher frequencies. The last part of this Ph.D. is devoted to time-domain WRI, where a challenge is to perform accurate wavefield reconstruction with an acceptable computational cost. In the frequency domain, wavefield reconstruction consists of solving in a least-squares sense an augmented wave equation system gathering the weighted wave equation and the observation equation. In the time domain, I show how to reconstruct wavefields with explicit time stepping through an iterative quasi-Newton method, whose convergence is discussed with the 2004 BP salt model.

Résumé

La FWI (Full Waveform Inversion) est un problème d'optimisation sous contraintes dédié à l'estimation des paramètres constitutifs du sous-sol à partir de mesures parcimonieuses des champs d'ondes sismiques. La FWI est fondée sur des approches locales d'optimisation et sur un espace de recherche réduit obtenu par projection de variables. La non linéarité et le caractère mal posé de la FWI sont deux difficultés majeures. Une source de non linéarité est liée au repliement de la phase, qui conduit à un minimum local dès que le modèle initial n'est pas suffisamment précis. Le caractère mal posé résulte de l'éclairage incomplet du sous-sol depuis la surface, le bruit et les couplages inter-paramètres. L'objectif de cette thèse est de réduire ces deux pathologies par de nouvelles approches d'optimisation et de régularisation. J'améliore tout d'abord la méthode d'inversion par reconstruction des champs d'onde (WRI: Wavefield Reconstruction Inversion). WRI étend l'espace de recherche en calculant les champs d'onde avec une relaxation de l'équation d'onde afin d'ajuster les données avec des modèles imprécis avant d'estimer les paramètres en minimisant les erreurs générées par cette relaxation. Quand ces deux estimations sont effectuées de manière alternée, WRI décompose l'inversion non linéaire en deux sous-problèmes linéaires en vertue de la bilinéarité de l'équation d'onde. WRI a été implémentée avec une méthode de pénalité, nécessitant une adaptation du paramètre de pénalité lors des itérations. Je remédie à cela avec ADMM (Alternating-Direction Method of Multipliers), qui concilie l'extension de l'espace de recherche et la précision de la solution au point de convergence avec un paramètre de pénalité fixe grâce à la mise à jour itérative des multiplicateurs de Lagrange. Une seconde contribution est l'implémentation de contraintes de bornes et de régularisation par variation totale (TV) dans WRI. Suivant la méthode de Split Bregman, des variables auxiliaires permettent de découpler les termes impliquant des normes $\ell 2$ et $\ell 1$ et de traiter les seconds efficacement avec des opérateurs de proximité. Ensuite, j'ai combiné une régularisation de Tikhonov et de TV par convolution infimale pour prendre en compte les différentes propriétés statistiques du milieu (constantes par morceau et lisses). Ma thèse aborde ensuite des reconstructions multi-paramètres. Je montre dans un premier temps que la bilinéarité de l'équation d'onde est vérifiée pour les équations de l'elastodynamique. Ensuite, je traite le cas de milieux acoustique VTI où je reconstruis conjointement la vitesse verticale et epsilon pour un modèle synthétique représentatif d'un champ pétrolier en mer du Nord. Je m'intéresse

ensuite à l'imagerie de l'atténuation qui est introduite en domaine harmonique sous forme d'une vitesse complexe. J'étends WRI à la reconstruction de paramètres complexes tout en développant une régularisation adaptable à la vitesse réelle et au facteur de qualité. Durant les premières itérations, les champs d'onde reconstruits sont précis uniquement au voisinnage des récepteurs. Les imprécisions de la phase pourraient avoir un rôle préjudiciable sur la solution de l'inversion. Afin de réduire cette empreinte, j'estime les paramètres par "phase retrieval", un processus qui vise la reconstruction d'un signal complexe à partir de l'amplitude de sa mesure linéaire. Une fois un premier modèle obtenu, je ré-injecte l'information de la phase pour converger vers la solution finale. Je montre la pertinence de cette stratégie lorsque le modèle initial est homogène. WRI a été initialement développée dans le domaine fréquentiel car la reconstruction des champs d'onde y est raisonablement aisée avec des méthodes d'algèbre linéaire. En domaine temporel, une approche fondée sur un schéma explicite d'intégration temporelle a été proposée mais repose sur une linéarisation autour des sources supposées connues. Je montre les imprécisions du champ d'onde reconstruit avec cette approximation que je corrige par une approche itérative, dont je discute la convergence avec un exemple numérique.

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

Chapter overview: In this introductory chapter, I first define seismic imaging as an inverse problem for the estimation of the underground parameters. Then, I review the principles of the so-called full-waveform inversion (FWI) from the historical and technical perspectives. I discuss its potential and limits related to nonlinearity and ill-posedness. I review recent works which attempted to extend the linear regime of FWI. This review allows me to introduce the main objectives of my thesis, whose content is outlined chapter by chapter.

1.1 Seismic imaging as an inverse problem

An inverse problem is the process of extracting causal factors from indirect observations (Aster et al., 2004). Seismic imaging, as an inverse problem, is the process of extracting the physical properties of the earth from seismic data. Generally, imaging methods are very important in many scientific fields because they provide a complete picture of an object that is invisible to the human eye. Seismic data are collected during seismic exploration when an active seismic source such as dynamite, vibroseis (for land) or air-gun (for marine) generates wave propagation into the earth. The seismic waves, that can be elastic or acoustic, travel into the earth and are affected and filtered by subsurface heterogeneities, and finally, some of them return to the surface. These seismic data are recorded by seismic detectors, which can be geophones (particle velocity measurements) or hydrophones (pressure perturbation measurements in marine environments) (Sheriff and Geldart, 1995).
After the data acquisition, the seismic data must be inverted to build up images of subsurface structures. Basically, seismic data inversion is a two-step procedure (Claerbout, 1985). First, a smoothly-varying background model of the earth is reconstructed such that it describes the kinematic information of the data, which contains the traveltimes of recorded waves. Then, the dynamic information is used to image the short-wavelength information by solving a linear least-squares problem which is known as migration. The migration step minimizes the difference between the recorded and modeled reflection data when the modeled data are generated

by solving the linearized wave equation with the single-scattering Born approximation (Gray et al., 2001; Etgen et al., 2009). In contrast, Full waveform inversion (FWI) (Tarantola, 1984; Pratt et al., 1998; Virieux and Operto, 2009) refers to a nonlinear optimization problem which estimates a subsurface model that jointly fits the kinematic and dynamic information. The word "Full" is used since FWI uses both the amplitude and the phase of the recorded waveforms simultaneously, unlike tomography and migration methods. Moreover, FWI can use any kind of waves without any distinction (diving waves, reflected waves, diffractions). As such, FWI is a seismic imaging technique which can take advantage of acquisition devices allowing for a rich illumination of the subsurface through the recording a wide variety of wave types.

1.2 Full waveform inversion

1.2.1 Problem statement and state of the art

Limits of the classical FWI

FWI is a leading-edge seismic imaging method that has gained renewed interest in oil and gas exploration for one decade due to the emergence of supercomputers, long-offset wide-azimuth acquisition, and broadband seismic sources (Virieux and Operto, 2009; Virieux et al., 2017). FWI seeks to estimate subsurface parameters (P and S wavespeeds, density, attenuation, anisotropic parameters) with a resolution close to the seismic wavelength by fitting seismic records with modeled seismograms that are computed with two-way modeling engines to account for the full complexity of the wavefields (Tarantola, 1984; Pratt et al., 1998).

Besides the computational cost resulting from two-way full-waveform modeling for large-scale 3D surveys, it is well acknowledged that one of the main challenges raised by FWI is related to the high nonlinearity and ill-posedness of the underlying optimization problem. One of the main sources of nonlinearity is related to the cycle-skipping pathology, which occurs when the initial subsurface model does not predict recorded traveltimes with an error lower than half the period and a classical least-squares difference-based misfit function is used (e.g. Virieux and Operto, 2009, their Figure 7). Ill-posedness may result from the deficit of subsurface illumination provided by surface acquisitions, noise, and parameter cross-talk in the framework of multi-parameter reconstruction (Operto et al., 2013).

From the numerical optimization viewpoint, the data-fitting/parameter-estimation problem underlying FWI is a partial differential equation (PDE)-constrained nonlinear optimization problem where the equality constraint is the wave equation, and the optimization parameters are embedded in the coefficients of the PDE. One of the most popular methods to solve such nonlinear constrained optimization problem is the method of Lagrange multiplier (Haber et al., 2000) where the constraints are implemented through a Lagrangian function. A local minimizer satisfies the so-called KKT (Karush-Kuhn-Tucker) first-order optimality conditions, which state that the variations of the Lagrangian with respect to the wavefield (state variables), the Lagrange multipliers (adjoint-state variables) and the parameters are zero at the optimum. The so-called

full-space approach consists of solving the nonlinear KKT system with an iterative Newton method to jointly update the three classes of unknowns. However, the dimension of the system to be solved for FWI application makes this full space approach intractable from the computational viewpoint in particular for time-domain formulation (Askan et al., 2007; Epanomeritakis et al., 2008). Instead, the FWI is classically solved with a reduced-space approach where the optimization problem is recast into an unconstrained problem after explicitly eliminating the wavefield from the unknowns by solving exactly the wave equation for the current guess of the subsurface parameters. Therefore, the reduced approach embeds at each nonlinear iteration the following steps: [1] compute the state variables by solving the wave equation, [2] from the state variable estimation, compute exactly the adjoint-state variables by solving the adjoint-state equation, [3] update the subsurface parameters with a gradient-based optimization method, where the gradient of the objective with respect to the subsurface parameters can be easily expressed in a closed form as a function of the state and adjoint-state variables (Plessix, 2006). In virtue of the spatial reciprocity of wave propagation, this gradient can be either viewed as the zero-lag correlation between the partial derivative wavefields sampled at the receiver positions with the data residuals or as the zero-lag correlation between the virtual sources of the partial derivative wavefields and the adjoint-state variables. For a classical least-squares difference-based misfit function, a partial derivative wavefield represents the wavefield single-scattered by a parameter perturbation in the framework of the Born approximation (Pratt et al., 1998; Virieux and Operto, 2009).

Although this reduced-space approach strictly enforces the constraints and is computationally tractable, it is prone to cycle skipping, as mentioned above. Indeed, if the initial model is not accurate enough to predict the recorded traveltimes with an error lower than half the dominant period, the zero-lag correlations underlying the gradient building will smear in an inconsistent way the offset-dependent difference-based residuals at wrong positions in the subsurface (i.e., along the wrong isochrones of the FWI sensitivity kernel), leading to unfocused images (Virieux and Operto, 2009, Their Figure 5). The cycle skipping criterion is increasingly difficult to satisfy when the number of propagated wavelengths increases as might be expected with the development of long-offset wide-azimuth acquisitions (Pratt et al., 1996; Pratt, 2008). This prompts the oil industry to develop broadband sources with a richer low-frequency content to balance the increasing propagation distance and mitigate the cycle skipping issue accordingly (Plessix et al., 2012; Baeten et al., 2013). In parallel with this, many heuristic data-driven approaches based upon frequency, traveltime, and offset continuations have been designed to drive the optimization problem toward a multiscale subsurface update from the long wavelengths to the shorter ones (e.g. Górszczyk et al., 2017).

Extending the linear regime of FWI

More automatic approaches, to reduce human intervention, have focused on the design of more convex distances between recorded and modeled data such as those based upon optimal transport theory (Engquist and Froese, 2014; Métivier et al., 2016b; Yang et al., 2018; Métivier et al., 2018, 2019), correlation (van Leeuwen and Mulder, 2010), deconvolution and matching filter (Luo and Sava, 2011; Warner and Guasch, 2016; Zhu and Fomel, 2016), instantaneous phase and envelope (Bozdağ et al., 2011; Luo and Wu, 2015) and dynamic warping (Ma and

Hale, 2013). Some of these distances improve the convexity of the function at the expense of the spatial resolution of the imaging. Therefore, these approaches should be viewed as tools to build initial models for classical FWI.

Another category of approaches seeks to extend the linear regime of FWI by enlarging the parameter search space. A pioneering implementation of this idea is presented by Shaw (1986) and Shaw (1988), who jointly estimates the velocity model and traveltime statics in the framework of 1D waveform inversion. Huang et al. (2017) proposes the source-receiver extension method, which interlace two optimization problems: first, the data is perfectly matched through a trace-per-trace source signature estimation before updating the subsurface model by minimizing the deviation of each artificial source signature from a reference one through an annihilator. This approach shares a connection with the adaptive waveform inversion of Warner and Guasch (2016) in the sense of the trace-per-trace source estimation problem can be viewed as a matching filter process. Biondi and Almomin (2014) propose to extend the subsurface model along the time-lag axis to account for large traveltime residuals and augment the misfit function with a penalty term which penalizes large time lags through an annihilator. The misfit function mixes a nonlinear relation with respect to the background model and a linear relationship with respect to the perturbation model, whose expression relies on the extended Born approximation. According to the Born approximation, the observables are the difference between the recorded data and the modeled data from the previous iteration. This approach, which jointly updates the background model and the reflectivity performs a joint tomography-like and a migration-like update of the subsurface, by opposition to reflection (Xu et al., 2012a; Brossier et al., 2015) or joint reflection + diving wave waveform inversion (Zhou et al., 2015) where a scale separation is explicitly enforced in the formalism through an alternating optimization scheme. The computational cost of this approach remains a major bottle-neck for 3D applications.

The Wavefield Reconstruction Inversion (WRI) method

van Leeuwen and Herrmann (2013, 2016) proposed another extended search space method in the frequency domain by recasting the PDE-constrained optimization problem into a quadratic penalty method, where the least-squares norm of the observation equation (the data misfit function) is penalized by the least-squares norm of the wave equation error (the violation of the physical constraint). The original motivation of this approach is that the subsurface model could be retrieved exactly if one monochromatic wavefield could be recorded everywhere in the subsurface. This statement results from the bilinearity of the wave equation in wavefields and subsurface parameters (A function of two variables is bilinear if it is linear with respect to each of its variables). Since this recording is not possible, the method of van Leeuwen and Herrmann (2013) and van Leeuwen and Herrmann (2016) seeks to approximate the true wavefields by those that best fit the sparse observations and satisfy the wave equation in a least-squares sense. In order to prevent the cycle skipping issue (i.e., fit the observations), this overdetermined least-squares problem is tuned such that a dominant weight is assigned to the observation equation during the early iterations of the waveform inversion (when the velocity model is inaccurate), hence generating significant relaxation of the wave equation constraint. Once the wavefields have been reconstructed for each source, the subsurface parameters are updated by minimizing

in the least-squares sense the wave-equation constraint violations. Since the governing idea of the method is to reconstruct wavefields, which are as close as possible to the true ones, van Leeuwen and Herrmann (2013) and van Leeuwen and Herrmann (2016) called their method Wavefield Reconstruction Inversion (WRI). Similar approaches for parameter estimation problems are referred to as Error in Constitutive Equations (ECE) method by other communities (Banerjee et al., 2013).
This penalty problem is solved for the wavefield and subsurface parameters in an alternating mode (van Leeuwen and Herrmann, 2013) or through variable projection (van Leeuwen and Herrmann, 2016). In the former case, WRI is recast as a sequence of two linear subproblems in virtue of the wave equation bilinearity. A well-known difficulty associated with the penalty method is related to the necessary adaptation of the penalty parameter in iterations, such that the constraint is relaxed during the early iterations and fulfilled at the convergence point (Fu and Symes, 2017). Compared to the above-mentioned full-space approach, the method of van Leeuwen and Herrmann (2016) is more computationally attractive as each source is processed independently during wavefield reconstruction and the update of the subsurface parameters does not require solving an adjoint-state equation. This approach has been adapted in the time domain by Wang et al. (2016, 2017b); Huang et al. (2018b) where the penalty term seeks to penalize the misfit between the true and the reconstructed sources instead of the true and the reconstructed wavefields. This approach has been referred to as the reconstructed wavefield method with extended sources.
Another frequency-domain approach which interlace the wavefield update and the subsurface parameters is the contrast-source method (Abubakar et al., 2011), which consists in the alternated update of the contrast (i.e., a term depending on the subsurface parameters) and the contrast source (i.e., a second-order scattering source formed by the product of the contrast with the incident wavefield). However, the contrast-source method may not have been optimally assessed as a search space expansion technique in the sense that the two functionals associated with the observation and the wave equations were linearly combined with a preset weight, the aim of which is to make the two functionals dimensionless, rather than with an adaptive weight as in the penalty method of van Leeuwen and Herrmann (2013). In fact, the motivation behind the contrast-source method was mainly to mitigate the computational burden of iterative frequency-domain seismic modeling based upon Gauss elimination techniques since the contrast-source method requires only one LU decomposition per frequency, this decomposition being computed in the starting model.

1.2.2 Objectives and content of the thesis

Improving WRI with ADMM

In the **Chapter 2** of this thesis, I will first focus on an easy-to-implement improvement of the WRI, which strongly reduces the sensitivity of the method to the choice of the penalty parameter, resulting in a faster convergence and more accurate subsurface and wavefield reconstruction. To achieve this goal, I replace the penalty method by an augmented Lagrangian method, which is implemented with a classical primal descent - dual ascent approach. In constrained optimization, the augmented Lagrangian method can be viewed as a combination of a

classical Lagrangian method and a penalty method, where the Lagrangian multiplier steers the inversion toward the constraint, while the penalization term regularizes the inversion (Nocedal and Wright, 2006, Chapter 17). To make the computational problem tractable, as in the penalty method of van Leeuwen and Herrmann (2013, 2016), I perform wavefield reconstruction and parameter estimation (primal subproblems) in an alternating mode, while the Lagrange multipliers are updated with a classical dual ascent approach after the two primal updates according to the popular alternating direction method of multiplier (ADMM) (Boyd et al., 2010) or after each of the two primal updates according to the Peaceman-Rachford splitting method (PRSM) (Peaceman and Rachford, 1955; He et al., 2014). This operator splitting draws some connection between my approach and ADMM with the difference that the wavefield reconstruction and the parameter estimation are not separable problems. This non-separability is managed by solving the two subproblems in sequence rather than in parallel, the solution of one subproblem being passed as a passive variable to the next problem. This sequential splitting strategy linearizes the parameter estimation problem around the reconstructed wavefield, which means that the original nonlinear FWI has been recast as a sequence of two linear subproblems.
By using a scaled form of the Lagrangian (Boyd et al., 2010, Section 3.1.1), I will also show that the augmented Lagrangian method is similar to the WRI penalty method, except that the original right-hand sides in the two objective functions of the penalty function (namely, those related to the data and the source misfits) are updated at each iteration with the running sum of the data and source residuals of previous iterations. This iterative right-hand side updating performs a self-adaptive weighting of the competing data-fitting and wave-equation objectives of the penalty function, leading to more efficient wavefield reconstruction and parameter estimation across iterations.

We can have a different insight into the ADMM-based WRI. As in the original WRI, the nonlinear constrained problem for subsurface parameter estimation is recast as a sequence of two unconstrained subproblems via a quadratic penalty method, where the two subproblems seek to update the reconstructed wavefield and the subsurface parameters in an alternating way within a cycle workflow. The key differences are that I re-inject the errors in the data misfit function and the penalty term (the wave equation residual) back to their right-hand sides (the data and the source) at each iteration. This iterative improvement of the solution by re-injecting the error back to the right-hand side of the system, which is known as defect correction or iterative refinement, is adapted here to a waveform-inversion problem. Similar iterative improvement of the solutions of linear inverse problems is found in the framework of asymptotic least-squares waveform inversion where the iterations seek to account for the approximations resulting from the linearization of the wave equation with the Born approximation and approximate Hessian (Lambaré et al., 1992; Jin et al., 1992; Ribodetti et al., 2011). The iterative refinement of the solution presented in this study is also closely related to the so-called Bregman distance (Bregman, 1967). Goldstein and Osher (2009) proposed to use Bregman distance to solve constrained optimization problems of the form $\min_{\mathbf{u}} E(\mathbf{u})$ such that $\mathbf{Au} = \mathbf{b}$ where E denotes a convex functional (in the framework of this study, E would represent the FWI misfit function and $\mathbf{Au} = \mathbf{b}$ is the Helmholtz equation). Yin et al. (2008) and Goldstein and Osher (2009) have shown that minimization of the Bregman distance is equivalent to recast the constrained problem as a quadratic penalty problem in a way comparable to the approach of van

Leeuwen and Herrmann (2013) and van Leeuwen and Herrmann (2016) approach. However, in a way that is comparable to the iterative refinement mentioned above, the right-hand side of the constraint is updated at each iteration $k+1$ with the error $\mathbf{b}-\mathbf{A}\mathbf{u}^k$ from the previous iteration (see Yin et al. (2008, Theorem 3.1). Goldstein and Osher (2009, Theorem 2.2) have shown that this iterative update of the right-hand side strictly enforces the constraint as long as E is convex. Accordingly, the value of the penalty parameter (λ in van Leeuwen and Herrmann (2013, their equation 5) can remain constant over iterations, while it should progressively tend to ∞ in classical penalty method to satisfy the constraint. In the Bregman iterations, this value can be chosen by trial and error such that the convergence rate is optimized. In this thesis, I adapt the iterative refinement strategy embedded in the Bregman method to frequency-domain FWI. I apply iterative right-hand side update to the two subproblems for wavefield reconstruction and subsurface parameter updates. Moreover, I perform the right-hand side update not only in the constraint (the wave equation) but also in the data misfit function during the wavefield reconstruction. Accordingly, I will refer to my method as IR-WRI (Iteratively-Refined WRI).
The **Chapter 2** of this thesis is mostly extracted from a paper published in *Geophysics* (Aghamiry et al., 2019c).

Interfacing leading-edge regularizations with IR-WRI

Ill-posedness is another significant pathology of FWI. Overcoming this issue requires robust and versatile regularized optimization approaches, which should, at the same time, preserve the intrinsic resolution power of FWI. To achieve this goal, I show in the first part of **Chapter 3** how to implement bound constraints and total variation (TV) (Rudin et al., 1992) regularization in IR-WRI with the split Bregman method, a particular instance of ADMM that has been proposed by Goldstein and Osher (2009) to solve $\ell 1$-regularized convex problems. The decomposability of ADMM provides the necessary recipe, after the introduction of suitable auxiliary variables, to split the non-differentiable TV regularization problem into two easy-to-solve subproblems, namely a least-square subproblem and a $\ell 1$ subproblem which is solved efficiently with proximity operators. This study has been published in the *Geophysical Journal International* (Aghamiry et al., 2019b).

A proper regularization should be driven by the shape and statistical characteristics of the medium to be imaged. For example, in geophysical imaging, the subsurface can be conceptualized as a piece-wise smooth medium, which is a model that contains smoothly varying components (for example, sedimentary cover) and blocky components (salt bodies, basalt flows). The widespread Tikhonov regularizations (Tikhonov et al., 2013) rely on the smoothness assumption and hence fail to recover sharp interfaces of such media. Conversely, TV regularizations are based on blockiness assumption and hence are more suitable to image large contrasts. However, they generate undesirable staircase imprints in smooth regions.
Regions characterized by smoothly-varying properties and those containing sharp contrasts have different statistical properties. The former are characterized by the normal prior, while the latter by a heavy-tailed prior (Polson and Sokolov, 2019). Consequently, the simultaneous recovery of both properties is difficult when one type of regularization is used (Tikhonov, TV, etc.). To overcome this issue, I propose to use a combination of different regularizations (Gholami and

Siahkoohi, 2010; Benning and Burger, 2018). A naive approach consists of the simple additive coupling or convex combinations (CC) of regularizations. Alternatively, Gholami and Hosseini (2013) proposed to explicitly decompose the model into several components of different statistical properties and use an appropriate regularization to reconstruct each component. Using this strategy, they combined Tikhonov and TV regularizations (referred to as TT regularization) to reconstruct piece-wise smooth media. The smooth components are captured by the Tikhonov regularization, while the TV counterpart determines the blocky ones. In many applications, it has been shown that a compound regularization based upon infimal convolution (IC) outperforms the one based upon additive coupling (Bergmann et al., 2018). In the second part of **Chapter 3**, I develop a general framework to combine a couple of regularization terms in IR-WRI through CC and IC and compare the performances of each implementation with the large contrast BP salt model.
The analysis of compound regularizations and their application in IR-WRI has been published in the *IEEE Transactions on Geoscience and Remote Sensing* journal (Aghamiry et al., 2020b).

Extending IR-WRI to multi-parameter reconstruction

So far, IR-WRI was mainly assessed for wavespeed estimation from the scalar Helmholtz equation. In the **Chapter 4** of this thesis, I go one step further and develop and assess the extension of IR-WRI to multi-parameter reconstruction. As I mentioned before, IR-WRI exploits the wave-equation bilinearity to recast FWI as a sequence of two linear subproblems. Therefore, the first step is to discuss which forms of the wave equation satisfy the bilinearity property. To achieve this goal, I started with the first order velocity-stress elastodynamic equations in VTI media and show that it is a bilinear system in wavefield and model parameters. Then, I build a second-order wave equation through a parsimonious variable-elimination approach since second-order equations are often more computationally-efficient in frequency-domain wave modeling (Operto et al., 2007a), and I discuss the bilinearity of this second-order equation. Then, I review different optimization scenarios depending on which of the constitutive parameters are processed as passive variables. In a second step, I will review in detail how to implement the IR-WRI method with bound constraints and total variation (TV) regularization for multi-parameter reconstruction in VTI acoustic media, where the different parameter classes can be either updated jointly or sequentially.
The analysis of multi-parameter IR-WRI in VTI acoustic media is the topic of a paper published in the *Geophysical Journal International* (Aghamiry et al., 2019a).
The second multi-parameter problem I will discuss deals with attenuation reconstruction. It is well acknowledged that attenuation effects are easily implemented in frequency-domain seismic modeling by using complex-valued wavespeeds. The imaginary part of the wavespeed depends on the attenuation factor $\boldsymbol{\alpha}$ (or quality factor $\mathbf{Q}$, which is $1/\boldsymbol{\alpha}$) and describes absorption effects, while the real part contains a frequency-dependent term which describes dispersion effects and guarantees causality (Toksöz and Johnston, 1981). In the frequency band, the quality factor is generally considered as frequency independent. During the inverse problem for complex parameters, it is common to process the physical quantity of the complex number (wavespeed and $\boldsymbol{\alpha}$) as two independent real-valued parameters. In frequency-domain FWI, the real part of the wavespeed for a reference frequency and the quality factor are generally processed as two in-

dependent real variables (Liao and McMechan, 1995; Hicks and Pratt, 2001; Malinowski et al., 2011; Hak and Mulder, 2011; Kamei and Pratt, 2013; Operto and Miniussi, 2018). Based on this, I break the complex wavespeed to squared slowness and attenuation factor, and I develop an accurate multilinear viscoacoustic wave equation in the wavefield, squared slowness and attenuation factor and extend IR-WRI for inversion of viscoacoustic wavefields corresponding to attenuative media. The proposed viscoacoustic IR-WRI treats the nonlinear viscoacoustic waveform inversion as a multiconvex optimization problem by replacing the original nonlinear problem with the recursive solution of three linear subproblems corresponding to the reconstruction of the wavefield, squared slowness, and attenuation factor. This new formulation also provides the necessary flexibility to separate and tailor the regularization to the squared slowness and attenuation factor, while providing the necessary computational efficiency to tackle large data sets. The analysis of viscoacoustic IR-WRI with two real-valued quantities is the topic of a paper published in the *Geophysics* journal (Aghamiry et al., 2020d).

Phase retrieval in IR-WRI

When wavefields are estimated from inaccurate velocity models and sparse observations, the reconstructed wavefields match the true wavefields in the vicinity of the receivers but may be quite inaccurate elsewhere. These inaccuracies are mapped into the velocity models during the parameter estimation subproblem, and it remains unclear from the theoretical viewpoint up to what point these inaccuracies can be progressively canceled out in iterations from the surface to depth. Furthermore, it is well acknowledged in image processing that the phase has a dominant imprint in a signal relative to amplitude (Oppenheim and Lim, 1981). This suggests that the phase inaccuracy of the wavefields that are reconstructed during the early iterations may be the main factor that drives IR-WRI toward spurious minima. To check this conjecture, I recast in **Chapter 5** the linear parameter-estimation subproblem of IR-WRI as a phase retrieval problem through a majorization-minimization (MM) technique to mitigate the imprint of the phase inaccuracies in IR-WRI. Phase retrieval refers to any process which aims to reconstruct a complex-valued signal from the amplitudes of linear measurements (Fienup, 1982; Gholami, 2014). I apply phase retrieval during the early stages of IR-WRI at low frequencies. Then, I switch to classical IR-WRI to re-inject the phase information in the waveform inversion. After illustrating the information carried out by amplitudes in the parameter-estimation subproblem, I show that phase retrieval allows me to reconstruct a large contrast medium starting the inversion from a homogeneous model, unlike the classical IR-WRI.
This study has been published in the *Geophysical Journal International* (Aghamiry et al., 2020e)

Time-domain formulation of IR-WRI

WRI has been originally developed in the frequency domain (van Leeuwen and Herrmann, 2013, 2016). The main reason is that the wavefield reconstruction with the feedback term to the observations can be easily implemented in the frequency domain by solving an overdetermined

linear system per source and frequency with any suitable linear algebra techniques (direct or iterative solvers). Solving such a linear system in the time domain with an implicit scheme would be intractable due to the extra temporal dimension. To address this issue, Wang et al. (2016, 2017b) recast the wavefield reconstruction subproblem as a source reconstruction one. Wavefield reconstruction and source reconstruction are theoretically equivalent in virtue of the linearity between the wavefields and the sources. However, Wang et al. (2016, 2017b) exploits the fact that the spatial distribution and the temporal signature of the true sources are generally reasonably well known to linearize the source reconstruction around the true source. This linearization allows them to implement the source estimation problem with an explicit time-stepping procedure, which includes one state and one adjoint-state simulation. I will show in the **Chapter 6** that the approximation proposed by Wang et al. (2016, 2017b) leads to significant inaccuracies in the reconstructed wavefields relative to those estimated in the frequency domain. These inaccuracies have a significant impact on the convergence speed of IR-WRI and the accuracy of the minimizer. Then, I will propose a recipe to refine the accuracy of the reconstructed wavefields iteratively. I will prove mathematically the convergence of my iterative scheme and I will assess the method numerically against a realistic synthetic example.
This study has been published as an express letter in *Geophysics* (Aghamiry et al., 2020a).

Chapter 7 will conclude my Ph.D. thesis, where I will recap the main results and review some perspectives of my work in terms of new methodological developments and applications.

Chapter 2

Wavefield reconstruction inversion

Chapter overview: Full waveform inversion (FWI) is a bi-linear waveform matching procedure which can provide the subsurface model with a wavelength-scale resolution. However, this high resolution makes FWI prone to cycle skipping, which drives inversion to a local minimum when the initial model is not accurate enough. To mitigate cycle skipping, wavefield reconstruction inversion (WRI) extends the inversion search space with a penalty method to relax the wave-equation constraint and fit the data. Then, the subsurface parameters are updated by minimizing the source residuals generated by the wave-equation relaxation, either through alternating optimization or variable projection. In the former case, WRI decomposes FWI into two linear sub-problems, capitalizing on the wave-equation bilinearity.
In this chapter, I am going to review FWI and WRI from a qualitative and quantitative viewpoint. I start with a quick review of different strategies to tackle constrained optimization problems. Then I describe FWI and WRI and illustrate, with some simple examples, how WRI extends the search space relative to FWI. WRI is sensitive to the penalty parameter and requires a tedious dynamic control of this parameter. To overcome this difficulty, I propose the iteratively-refined WRI (IR-WRI), which replaces the penalty method that is used in WRI by the alternating direction method of multipliers (ADMM). ADMM relies on the augmented Lagrangian method, a combination of Lagrangian and penalty methods. Compared to penalty methods, augmented Lagrangian methods refine the solution more efficiently and accurately when a fixed penalty parameter is used, thanks to the iterative update of the Lagrange multipliers. This section includes our *Geophysics* paper, *Improving full-waveform inversion by wavefield reconstruction with the alternating direction method of multipliers* (Aghamiry et al., 2019c). Finally, I will finish the chapter with a local convergence proof of IR-WRI.

2.1 Non-linear constrained optimization problem

Finding the best of all possible solutions for a problem is referred to optimization. More precisely, find a solution in the feasible region which has the minimum (or maximum) value of

the objective function. If there is not any restriction on the feasible region, it is called an unconstrained optimization problem. In contrast, it is called a constrained optimization problem when there are some priors about the feasible region (Nocedal and Wright, 2006). The constrained optimization problem that will be discussed in this thesis has the form

$$\min_{\mathbf{u},\mathbf{m}} \quad \mathcal{P}(\mathbf{m},\mathbf{u}) \tag{2.1a}$$

$$\text{subject to} \quad \mathcal{C}(\mathbf{m},\mathbf{u}) = 0. \tag{2.1b}$$

where $\mathbf{u} \in \mathbb{C}^{N\times 1}$ is a field, N is the number of discrete grid points, and $\mathbf{m} \in \mathbb{C}^{N\times 1}$ is the model parameters. The constraint $\mathcal{C}(\mathbf{m},\mathbf{u}) = 0$ is a partial differential equation (PDE) and $\mathcal{P}(\mathbf{m},\mathbf{u})$ is the distance between observed data and modeled data to be minimized (it is called data-fitting or fidelity term).
In FWI, we seek to find $\mathbf{m}$ and $\mathbf{u}$ when sparse measurements of $\mathbf{u}$ as well as a low resolution approximation of $\mathbf{m}$ are available. Also, reconstructing $\mathbf{u}$ for a given $\mathbf{m}$ is called forward problem or data modeling.

2.1.1 The method of Lagrange multipliers

In mathematics, the method of Lagrange multipliers is a method to find the local stationary point (minimum or maximum) of a function when some equality constraints are satisfied (Nocedal and Wright, 2006). The basic idea is to convert a constrained problem into a form such that the derivative test of an unconstrained problem can still be applied. To find the stationary points of eq. 2.1, it is required to compute the stationary point of the Lagrangian which reads as

$$\mathcal{L}(\mathbf{m},\mathbf{u},\mathbf{v}) = \mathcal{P}(\mathbf{m},\mathbf{u}) + \mathbf{v}^T\mathcal{C}(\mathbf{m},\mathbf{u}), \tag{2.2}$$

where $\mathbf{v} \in \mathbb{C}^{N\times 1}$ is the Lagrange multiplier.
Lagrangian is a saddle point function and its stationary point occurs at a point which minimizes the Lagrangian w.r.t $\mathbf{m}$ and $\mathbf{u}$ and maximizes it w.r.t $\mathbf{v}$ as

$$\min_{\mathbf{u},\mathbf{m}} \max_{\mathbf{v}} \quad \mathcal{L}(\mathbf{m},\mathbf{u},\mathbf{v}). \tag{2.3}$$

The necessary condition for $(\mathbf{m},\mathbf{u})$ to be a local minimizer of the problem 2.1, referred to as the Kush-Kun-Tucker (KKT) conditions, is that $(\mathbf{m},\mathbf{u},\mathbf{v})$ is a stationary point of the Lagrangian

$$\nabla\mathcal{L}(\mathbf{m},\mathbf{u},\mathbf{v}) = 0 \rightarrow \begin{cases} \frac{\partial\mathcal{P}(\mathbf{m},\mathbf{u})}{\partial\mathbf{u}} + \frac{\partial\mathcal{C}(\mathbf{m},\mathbf{u})^T}{\partial\mathbf{u}}\mathbf{v} = 0 \\ \frac{\partial\mathcal{P}(\mathbf{m},\mathbf{u})}{\partial\mathbf{m}} + \frac{\partial\mathcal{C}(\mathbf{m},\mathbf{u})^T}{\partial\mathbf{m}}\mathbf{v} = 0 \\ \mathcal{C}(\mathbf{m},\mathbf{u}) = 0 \end{cases}. \tag{2.4}$$

This is a nonlinear system for $\mathbf{m}$, $\mathbf{u}$ and $\mathbf{v}$, which can be solved with Newton family methods (Haber et al., 2000).

2.1.2 Reduced approach

Because of the computational burden and storage problem, it is impossible to use the former method for large scale problems. The most common approach to solve such problems is called reduced approach. In reduced approach, the KKT conditions are satisfied separately, which means that the full search space that encompasses the unknown field and the model parameters is first projected onto the parameter space by computing exactly the incident fields in the current model before updating this later (Haber et al., 2000; Askan et al., 2007; Epanomeritakis et al., 2008). The objective function can be written as an unconstrained optimization problem as

$$\min_{\mathbf{m}} \qquad \mathcal{P}(\mathbf{m}, \mathbf{u}(\mathbf{m})), \tag{2.5}$$

where $\mathbf{u}(\mathbf{m})$ is the solution of $\mathcal{C}(\mathbf{m}, \mathbf{u}) = 0$. It is well acknowledged that the oscillating nature of seismic signals makes the reduced approach highly nonlinear when the starting $\mathbf{m}$ is far from the solution (e.g. Virieux and Operto, 2009).

2.1.3 Penalty methods

A penalty method replaces a constrained optimization problem by a series of unconstrained problems whose solutions ideally converge to the solution of the original constrained problem (Nocedal and Wright, 2006). The unconstrained problems are formed by adding a penalty function to the objective function that consists of a penalty parameter multiplied by a measure of the violation of the constraints. The measure of the constraint violation is nonzero when the constraints are violated and is zero in the region where constraints are not violated. The penalty equivalent of equation 2.1 reads as

$$\min_{\mathbf{u},\mathbf{m}} \qquad \mathcal{P}(\mathbf{m}, \mathbf{u}) + \lambda \mathcal{G}(\mathcal{C}(\mathbf{m}, \mathbf{u})), \tag{2.6}$$

where $\mathcal{G}(\bullet)$ is a measure of the constraint violation and λ is the penalty parameter. The unconstrained problem 2.6 is solved iteratively, while penalty parameter λ is progressively increased to satisfy the constraint $\mathcal{C}(\mathbf{m}, \mathbf{u}) = 0$ at the convergence point. The advantage of the penalty method is that one does not exactly satisfy the constraint $\mathcal{C}(\mathbf{m}, \mathbf{u}) = 0$ when we are far from the solution, but tuning the penalty parameter in a correct way is a big issue in penalty family methods (Nocedal and Wright, 2006).

2.1.4 Augmented Lagrangian and method of multipliers

Augmented Lagrangian method is another method to solve constrained optimization problems (Nocedal and Wright, 2006; Bertsekas, 2016). It combines Lagrangian and penalty methods (a penalty term is added to the ordinary Lagrangian term). It is close to the penalty method since it replaces a constrained optimization problem by a series of unconstrained problems. It solves in an alternating mode the minimization subproblem for the primal variables and the

maximization subproblem for the dual variables (Lagrange multipliers). The augmented Lagrangian of the problem 2.1 reads

$$\mathcal{L}_A(\mathbf{m},\mathbf{u},\mathbf{v}) = \mathcal{P}(\mathbf{m},\mathbf{u}) + \mathbf{v}^T\mathcal{C}(\mathbf{m},\mathbf{u}) + \lambda\mathcal{G}(\mathcal{C}(\mathbf{m},\mathbf{u})), \tag{2.7}$$

where the second and third terms of the right-hand side are the Lagrangian term and the penalty term (the augmentation), respectively.
Augmented Lagrangian is a saddle point problem, like Lagrangian, which can be solved in alternating mode with the method of multipliers as

$$(\mathbf{u}^{k+1},\mathbf{m}^{k+1}) = \arg\min_{\mathbf{u},\mathbf{m}} \ \mathcal{L}_A(\mathbf{m},\mathbf{u},\mathbf{v}^k), \tag{2.8}$$

$$\mathbf{v}^{k+1} = \mathbf{v}^k + \mathcal{C}(\mathbf{m^{k+1}},\mathbf{u^{k+1}}), \tag{2.9}$$

where k denotes the iteration number. Here the augmented Lagrangian is minimized jointly with respect to the two primal variables, $\mathbf{u}$ and $\mathbf{m}$.

2.1.5 Alternating direction method of multipliers (ADMM)

The alternating direction method of multipliers (ADMM) is an algorithm that solves the first (primal) subproblem of method of multipliers by breaking it into smaller pieces, each of which are then easier to handle. ADMM consists of the iterations

$$\mathbf{u}^{k+1} = \arg\min_{\mathbf{u}} \ \mathcal{L}_A(\mathbf{m}^k,\mathbf{u},\mathbf{v}^k), \tag{2.10}$$

$$\mathbf{m}^{k+1} = \arg\min_{\mathbf{m}} \ \mathcal{L}_A(\mathbf{m},\mathbf{u}^{k+1},\mathbf{v}^k), \tag{2.11}$$

$$\mathbf{v}^{k+1} = \mathbf{v}^k + \mathcal{C}(\mathbf{m^{k+1}},\mathbf{u^{k+1}}). \tag{2.12}$$

In ADMM, $\mathbf{u}$ and $\mathbf{m}$ are updated in an alternating mode, which accounts for the term alternating direction. ADMM can be viewed as a version of the method of multipliers where the primal variables $\mathbf{u}$ and $\mathbf{m}$ are updated with a single Gauss-Seidel iteration rather than jointly. Originally, ADMM was proposed for linear problems. However, the convergence for bi-linear problems has been demonstrated by Xu et al. (2012b) for a specific application. Also, recently, a nonlinear version of ADMM has been proposed by Benning et al. (2015). I will prove the local convergence property of ADMM for FWI problem in section 2.4.

2.2 Full waveform versus wavefield reconstruction inversions

Full Waveform Inversion (FWI) is a leading-edge seismic imaging method that has gained renewed interest in oil and gas exploration since one decade due to the emergence of supercomputers, long-offset wide-azimuth acquisition, and broadband seismic sources (Virieux and Operto, 2009).
FWI can be formulated in the frequency domain as the following PDE constrained optimization

problem:

$$\min_{\mathbf{m},\mathbf{u}} \quad \|\mathbf{P}\mathbf{u}-\mathbf{d}\|_2^2, \quad \text{subject to} \quad \mathbf{A}(\mathbf{m})\mathbf{u}=\mathbf{b}, \tag{2.13}$$

where $\mathbf{m} \in \mathbb{R}^{N\times 1}$ is the discrete subsurface model parameters, $\mathbf{b} \in \mathbb{C}^{N\times 1}$ is the sparse monochromatic source term, $\mathbf{u} \in \mathbb{C}^{N\times 1}$ is the monochromatic modeled wavefield, $\mathbf{d} \in \mathbb{C}^{M\times 1}$ is the recorded seismic data and $\mathbf{P} \in \mathbb{R}^{M\times N}$ (which $M << N$) is a linear observation operator that samples $\mathbf{u}$ at the receiver positions. Also, $\mathbf{A}(\mathbf{m})$ is the discretized wave-equation operator. Here, I assume a single-source problem. However, the extension to multiple sources is straightforward. In its simplest form, it is the Helmholtz operator given by

$$\mathbf{A}(\mathbf{m}) = \Delta + \omega^2 \text{diag}(\mathbf{m}), \tag{2.14}$$

where Δ denotes the Laplace operator, ω the angular frequency, and $\mathbf{m}$ the squares slownesses. The nonlinear constrained optimization problem, equation 2.13, can be solved with the method of Lagrange multipliers and Newton-type algorithm (Haber et al., 2000).

$$\min_{\mathbf{m},\mathbf{u}} \max_{\mathbf{v}} \mathcal{L}(\mathbf{m},\mathbf{u},\mathbf{v}) = \min_{\mathbf{m},\mathbf{u}} \max_{\mathbf{v}} \|\mathbf{P}\mathbf{u}-\mathbf{d}\|_2^2 + \mathbf{v}^T\left[\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{b}\right], \tag{2.15}$$

where $\mathbf{v} \in \mathbb{C}^{N\times 1}$ denotes the Lagrange multiplier (or, adjoint-state wavefield in the FWI terminology) and ^{T} the complex-conjugate (Hermitian) transpose. A Newton approach would require to solve the large-scale KKT system (Haber et al., 2000; Akçelik, 2002; Askan et al., 2007), and also to store previous values of ($\mathbf{u}$, $\mathbf{v}$, $\mathbf{m}$) in memory, which makes it intractable.
Instead, a reduced approach strictly enforces the PDE constraint ($\mathbf{u} = \mathbf{A}(\mathbf{m})^{-1}\mathbf{b}$) at each iteration, hence leading to the monovariate misfit function (Pratt et al., 1998; Plessix, 2006)

$$\min_{\mathbf{m}} J_r(\mathbf{m}) = \min_{\mathbf{m}} \|\mathbf{P}\mathbf{A}^{-1}(\mathbf{m})\mathbf{b}-\mathbf{d}\|_2^2. \tag{2.16}$$

For a long time, the reduced approach has been the most common approach to solve the FWI optimization problem. The gradient of the reduced objective function at iteration $k+1$ is the zero-lag correlation between the so-called virtual scattering source, namely the incident wavefield, $\mathbf{u}^{k+1} = \mathbf{A}^{-1}(\mathbf{m}^k)\mathbf{b}$, weighted by the radiation pattern matrix $\partial\mathbf{A}(\mathbf{m}^k)/\partial\mathbf{m}$ (Pratt et al., 1998), and the adjoint field $\mathbf{v}^{k+1}$

$$\nabla_{\mathbf{m}} J_r = \Re\left(\omega^2 \text{diag}(\mathbf{u}^{k+1})^T \mathbf{v}^{k+1}\right), \tag{2.17}$$

where the radiation pattern of the squared slowness simply samples $\mathbf{u}^{k+1}$ at the position of the model parameter with a weight ω^2 in the case of the Helmholtz equation 2.14. Also, $\Re(\bullet)$ is the real part of $\bullet$. The Lagrange multiplier $\mathbf{v}^{k+1}$ satisfies the so-called adjoint-state equation

$$\mathbf{A}^T(\mathbf{m}^k)\mathbf{v}^{k+1} = \mathbf{P}^T\left[\mathbf{P}\mathbf{u}^{k+1}-\mathbf{d}\right], \tag{2.18}$$

whose right-hand side is the prolongation of the data residuals in the full computational domain. The adjoint of $\mathbf{A}$ in equation 2.18 means that the residuals in the right-hand side are back-propagated in time (reverse time propagation) in the subsurface model. The zero-lag correlation between the virtual scattering source and the adjoint wavefield in the gradient of J_r

extracts, from all of the residuals, the piece of information that should be mapped at a given position of the subsurface.

Although this reduced approach requires to solve exactly the state and adjoint-state equations at each nonlinear iteration, it is more computationally tractable than the full-space approach. Moreover, after elimination of the state and adjoint-state variables, one needs to update only one class of unknowns, namely $\mathbf{m}$, at each iteration. Due to the highly-oscillating nature of the inverse PDE operator $\mathbf{A}^{-1}$ (the Green's function), the reduced minimization problem, equation 2.16, is prone to convergence to inaccurate minimizer when the initial $\mathbf{m}$ is not accurate enough (Symes, 2008; Virieux and Operto, 2009).

To mitigate this nonlinearity issue, van Leeuwen and Herrmann (2013) recast the original constrained optimization problem, equation 2.13, as a multi-variate unconstrained quadratic penalty problem for $\mathbf{u}$ and $\mathbf{m}$ given by

$$\min_{\mathbf{m},\mathbf{u}} J_p(\mathbf{m},\mathbf{u}) = \min_{\mathbf{m},\mathbf{u}} \|\mathbf{P}\mathbf{u}-\mathbf{d}\|_2^2 + \lambda\|\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{b}\|_2^2, \tag{2.19}$$

where the scalar $\lambda > 0$ is the so-called penalty parameter. The hard constraint in the original constrained problem is replaced by a soft constraint as a quadratic penalty term, which represents the square of the constraint violation. FWI based upon penalty method was called Wavefield Reconstruction Inversion (WRI) by van Leeuwen and Herrmann (2013). In the next section, I will describe in detail how WRI method works and what are the pros and cons of this technology. Hereafter, I first illustrate the extending search space property of WRI compared to the classical reduced approach, and I illustrate in a heuristic way how the method works with some simple examples.

WRI has a lot of similarity with wavefield inversion or Error in Constitutive Equations (ECE) methods (Banerjee et al., 2013). The wavefield inversion relies on the fact that, with known wavefield, the model parameters can be extracted straightforwardly. Accordingly, the final goal of wavefield inversion is to reconstruct wavefields as accurately as possible and estimate the model parameters from these wavefields. In WRI, the inversion is broken down into two subproblems for wavefields and model parameters. In the first one, we try to find the wavefields that are as close as possible to the true wavefields for each source by solving the wave equation and the observation equation in a least-squares sense. This step can also be interpreted as a wavefield extrapolation when we try to reconstruct the wavefields at each point of the subsurface model from sparse measurements of the wavefields at receiver positions. Also, this wavefield reconstruction step can be seen as a data assimilation problem when it tries to combine physics (wave-equation) with observations (Auroux and Blum, 2008). At the next step, it reconstructs the model parameters in a least-squares sense from the reconstructed wavefields. More precisely, the wave-equation relaxation generates source residuals, which are minimized with respect to the subsurface parameters during the second step of WRI to push back the reconstructed wavefields toward the wave equation constraint.

The gradient associated with this minimization problem is formed by the zero-lag correlation between the scattering virtual source associated with the reconstructed wavefield and the spatially-distributed source residuals, sometimes called extended sources (Wang et al., 2017a; Huang et al., 2018b). In fact, the source residuals replace the back-propagated adjoint wavefields of the reduced approach (eq. 2.17).

The alternating solving in WRI find its roots in the bilinearity of the wave-equation. From the definition of $\mathbf{A}$ in (2.14), we get that

$$\mathbf{A}(\mathbf{m})\mathbf{u} \quad = \quad \Delta\mathbf{u} + \mathbf{L}\mathbf{m}, \tag{2.20}$$

where

$$\mathbf{L} = \frac{\partial \mathbf{A}(\mathbf{m})}{\partial \mathbf{m}}\mathbf{u} = \omega^2 \text{diag}(\mathbf{u}). \tag{2.21}$$

Bilinearity of the wave-equation tells us that, if we record the wavefield generated by a monochromatic source everywhere in the subsurface, then the subsurface model can be reconstructed exactly in a single iteration by solving the linear system

$$\mathbf{L}\mathbf{m} = \mathbf{y}, \tag{2.22}$$

where $\mathbf{y} = \mathbf{b} - \Delta\mathbf{u}$. The bi-linearity of the wave equation on model parameters and wavefield is illustrated in Figure 2.1. A part of BP 2004 is shown in Figure 2.1a and the monochromatic wavefield computed in this model for one source located at the surface in the middle of the model is shown in Figure 2.1b. In Figure 2.1c, the subsurface model reconstructed from this monochromatic wavefield by a direct inversion matches perfectly the true velocity model shown in Figure 2.1a.

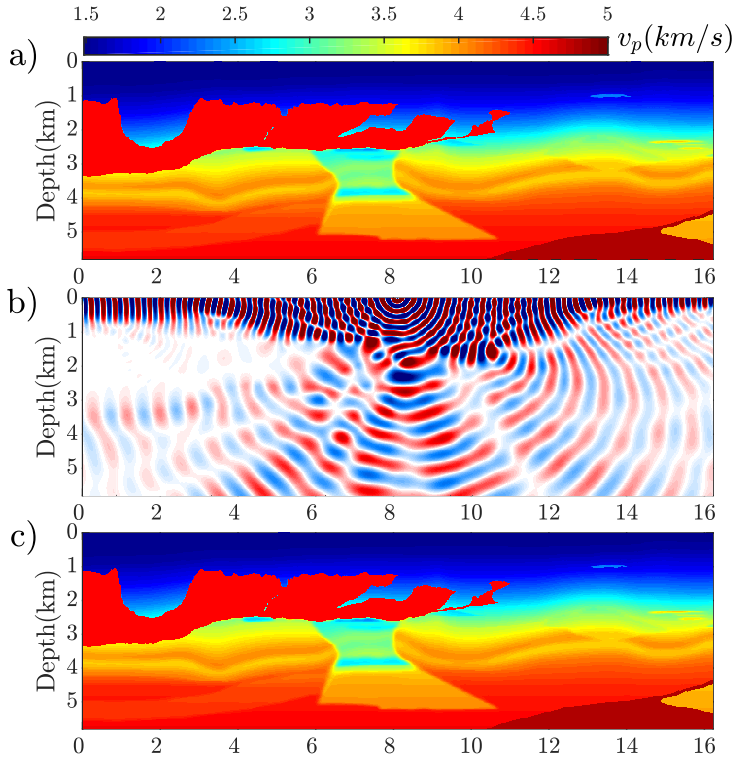

Figure 2.1 – Bilinearity of the wave equation. (a) The left target of the 2004 BP salt model. (b) Monochromatic $5\ Hz$ wavefield for a source in the middle of the model. (c) The reconstructed model using the monochromatic wavefield in (b).

2.2.1 Illustration with a homogeneous, gradient and two-layer velocity models

I now illustrate the mean features of WRI with some simple model examples.

Homogeneous velocity model

First, I now illustrate the mean features of WRI with a simple homogeneous model example. A single-ended long spread surface acquisition is used for a homogeneous velocity model of dimensions $8\ km \times 3\ km$ (distance $\times$ depth). A single source located at $x = 200m$ is used, and receivers are deployed all along the surface with a $50\ m$ spacing. The velocity model is homogeneous and is parametrized by a single wavespeed v_0, $v(x, z) = v_0$. The data are generated with $v_0 = 2000m/s$. The cost function of the reduced approach as well as that of the WRI method for several values of λ are evaluated for $v_0 \in [1750, 2250]$, while the discrimination step is $10\ m/s$ (Figure 2.2). The figure shows that the WRI cost function for a sufficiently-small value of λ is convex, unlike that of the reduced approach.

To further emphasize the role of the WRI method in reducing the cycle skipping and improve

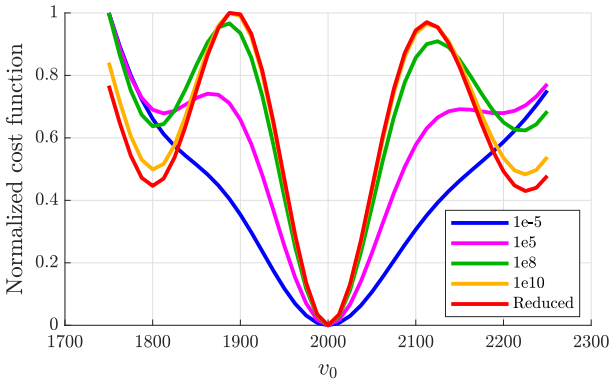

Figure 2.2 – Cost function of the reduced approach and WRI for several values of λ. For WRI, the cost function depends on the velocity (v_0) and λ, while it is only a function of v_0 for the reduced approach. The background velocity model is homogeneous (see text for details).

the convexity of the cost function, I show the monochromatic 5 Hz wavefield computed with the reduced approach and WRI. The true wavefield is shown in Figure 2.3a, while the wavefield computed with the reduced approach for a velocity of $v_0 = 2200m/s$ is shown in Figure 2.3b. The 200 m/s mismatch between the true and background wavespeeds is sufficient to generate cycle skipping at large offsets, as shown by the comparison between the frequency domain data computed in the true and background models (Figure 2.3d) and by the time-domain residuals (Figure 2.3e). The wavefield reconstructed by WRI with a small value of the penalty parameter, such that data fitting is fostered at the expense of the wave-equation constraint, is shown in Figure 2.3c. Comparing the WRI wavefield and the wavefield computed in the true model clearly shows how the wave-equation relaxation pushes the reconstructed wavefield toward the true one. Indeed, at the first WRI iteration, the wavefield reconstruction is accurate only in the areas which are illuminated by arrivals that have a significant signature in the data, which is

the shallow part in this test (direct wave propagation). If the true model would have contained deep reflectors, or the acquisition would have involved longer offsets or lower frequency, the reconstructed wavefield accuracy would have extended downward according to the extended spatial support of the propagated wavefield. The comparison between the frequency domain data computed in the true and initial models with WRI (Figure 2.3d) and by the time-domain residuals (Figure 2.3f), clearly shows the ability of the method to overcome cycle skipping.

The data fit in WRI wavefield has been obtained at the expense of the accuracy with which the

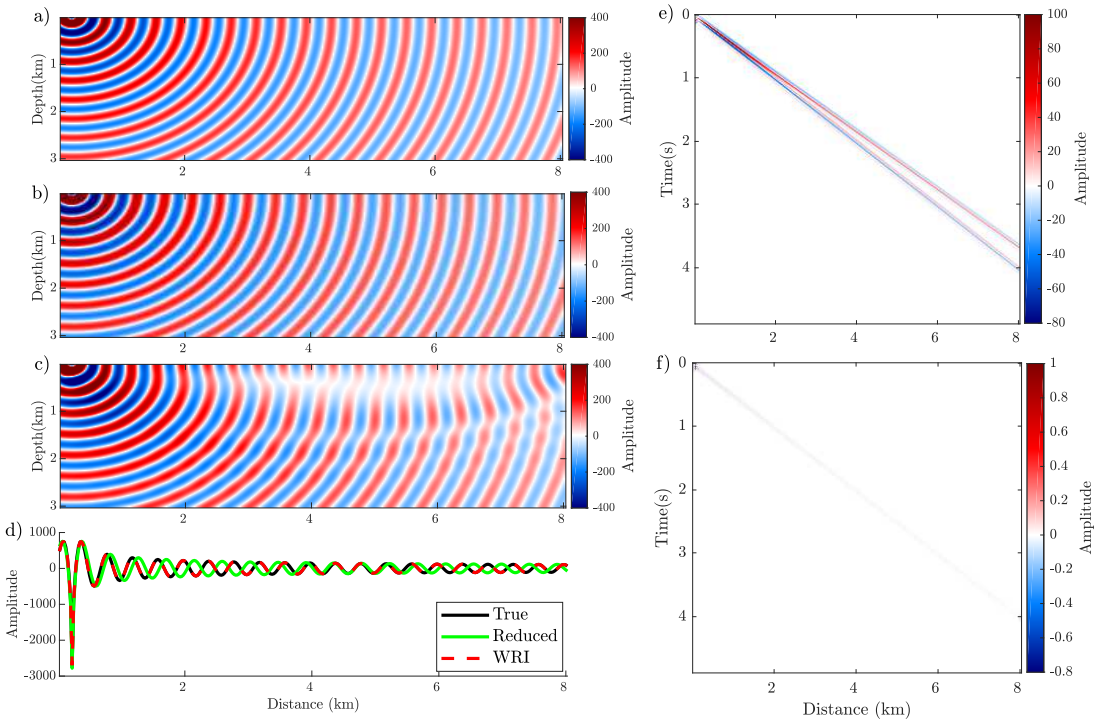

Figure 2.3 – (a) Monochromatic wavefield computed in the true model $v_0 = 2000\ m/s$ for the 5 Hz frequency. (b) Reduced approach wavefield ($v_0 = 2200\ m/s$). (c) WRI wavefield ($v_0 = 2200\ m/s$). (d) Direct comparison between the real part of the true (black), reduced-approach (green) and WRI (red) wavefields at receiver positions. (e-f) Time-domain residual seismograms between true and reduced-approach (e) and WRI (f).

reconstructed wavefields satisfy the wave equation. This wave equation relaxation generates source residuals or extended sources (Huang et al., 2018b) which can be minimized in a least-squares sense to update the subsurface parameters and push back the reconstructed wavefields toward the wave equation. This alternating direction two-step cycle can be iterated until convergence. I used a line of sources, which are deployed all along the surface with 200 m spacing, for inversion and I inverted the 5 Hz data up to 100 iterations. I used ADMM based WRI and apply a smoothing filter with edge preserving property on the updated model at each iteration. The ADMM based WRI will be discussed in detail in the next section. Figures 2.4a-g show iterations 1, 2, 4, 10, 20, 50 and 100 of this process, respectively, where the updated model are in the left column and the source residuals are shown in the right column of the figure. At the first iteration (Figure 2.4a) the velocity updates are quite shallow, hence generating some artificial boundaries between the shallow and the deeper parts. These boundaries feed the wavefield

reconstruction with a new set of data residuals, which contribute to progressively downward continue the model updates until the convergence point is reached following some kind of layer stripping approach.

Gradient velocity model

In the next test similar to that designed by Mulder and Plessix (2008), I consider a more complex vertical velocity gradient model given by $v(x,z) = v_0 + \alpha z$, where v_0 is the velocity at the surface and α is the vertical velocity gradient. The size of the velocity model is $8\ km \times 3\ km$. The acquisition consists of a single source located at $x = 200m$, and a line of receivers spaced $50\ m$ apart all along the surface. The data are generated with a velocity $v_0 = 2000\ m/s$ and a gradient $\alpha = 0.65\ 1/s$ (Figure 2.5a). The cost function of the reduced approach as well as that of WRI with $\lambda = 1e-2$ are evaluated on a grid of 41×41 testing points when $v_0 \in [1750, 2250]$ and $\alpha \in [0.4, 0.9]$ (the discrimination step is $12.5\ m/s$ for v_0 and $0.0125\ 1/s$ for α). For each point of this 2D grid, the cost functions of the reduced approach and WRI are shown in Figures 2.6a-b, respectively.

To facilitate the comparison between the cost functions of the reduced approach and WRI methods and emphasize the role of λ, I select two profiles of Figure 2.6 along the axis $v_0 = 2000\ m/s$ and $\alpha = 0.65\ 1/s$ (the true values) and calculate the cost function for different values of λ. The results are shown in Figure 2.7a for fixed α and Figure 2.7b for fixed v_0. Again, Figures 2.6 and 2.7 highlight the convexity of the WRI cost function when a sufficient relaxation of the wave equation is applied.

Like homogeneous test, I show the mono-chromatic 5 Hz wavefield computed with the reduced approach and WRI. I use a homogeneous initial velocity with $v_0 = 3000m/s$ and $\alpha = 0$ (Figure 2.5b). The true, reduced approach and WRI wavefields are shown in Figures 2.8a-c, respectively. Also, a comparison between the frequency domain data at receiver positions computed in the true and background models are shown in Figure 2.8d while the time-domain residuals of these wavefields are shown in Figures 2.8e-f, for reduced approach and WRI, respectively. The mismatch between the true and background wavespeeds is sufficient to generate cycle skipping. Comparing the WRI wavefield and the wavefield computed in the true model clearly shows how the wave-equation relaxation pushes the reconstructed wavefield toward the true one. The comparison between the frequency domain data computed in the true and initial models with WRI (Figure 2.8d) and by the time-domain residuals (Figure 2.8f), clearly shows the ability of the method to overcome cycle skipping.

I used a line of sources, which are deployed all along the surface with $200\ m$ spacing, for inversion and I inverted the $5\ Hz$ data up to 100 iterations with ADMM based WRI when a smoothing filter on the updated model at each iteration is used. Figures 2.9a-g shows iterations 1, 2, 4, 10, 20, 50 and 100 of this process, respectively, where the updated model are in the left column and the source residuals are shown in the right column of the figure. Like the homogeneous test, the velocity updates are quite shallow at early iterations and they generate some artificial boundaries between the shallow and the deeper parts. These boundaries feed

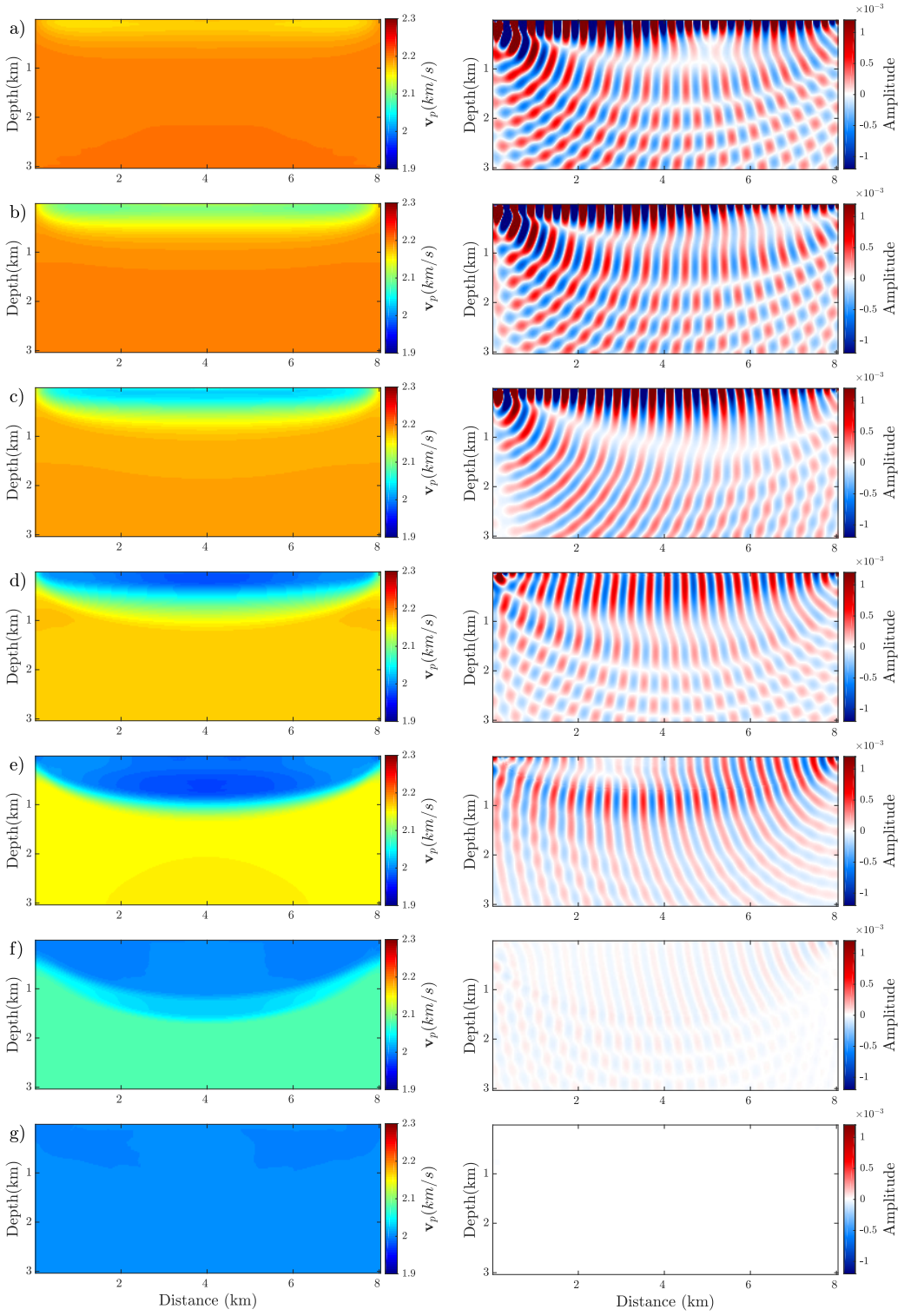

Figure 2.4 – Homogeneous test. The left column are updated models and the right column are source residuals at iteration (a) 1, (b) 2, (c) 4, (d) 10, (e) 20, (f) 50 and (f) 100. The initial and true velocity model are $2200m/s$ and $2000m/s$, respectively.

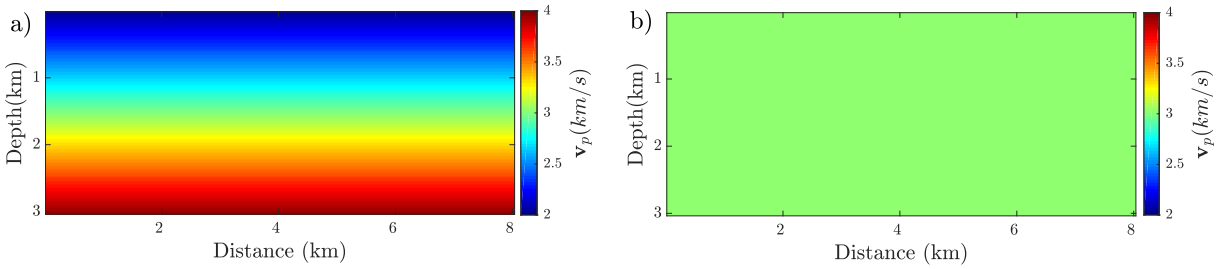

Figure 2.5 – Velocity gradient test. (a) True model. (b) Initial model.

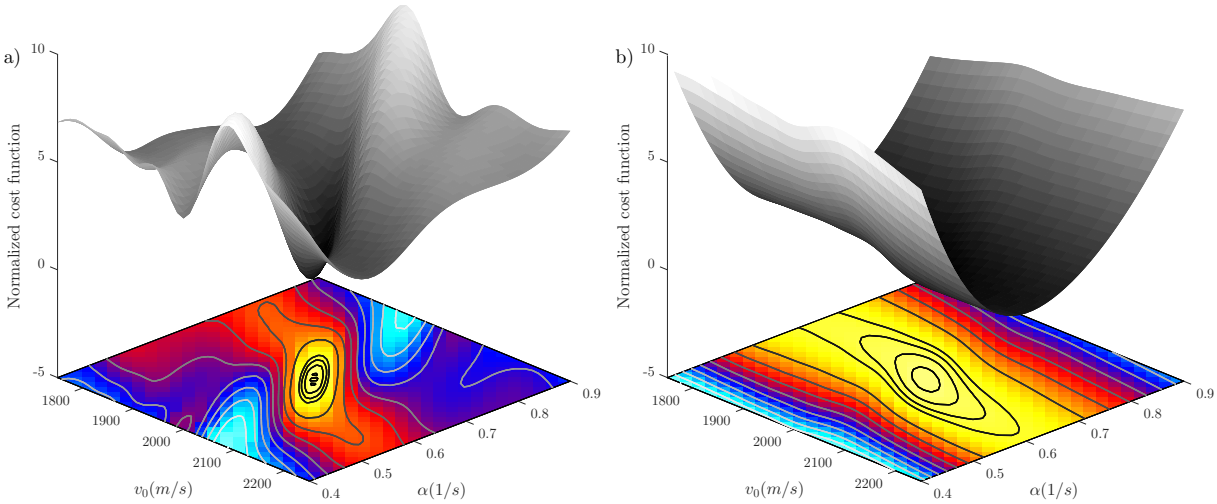

Figure 2.6 – Velocity gradient test. Cost function for reduced approach (a) and WRI(b). Each cost function depends on the velocity at the surface (v_0) and the velocity gradient (α).

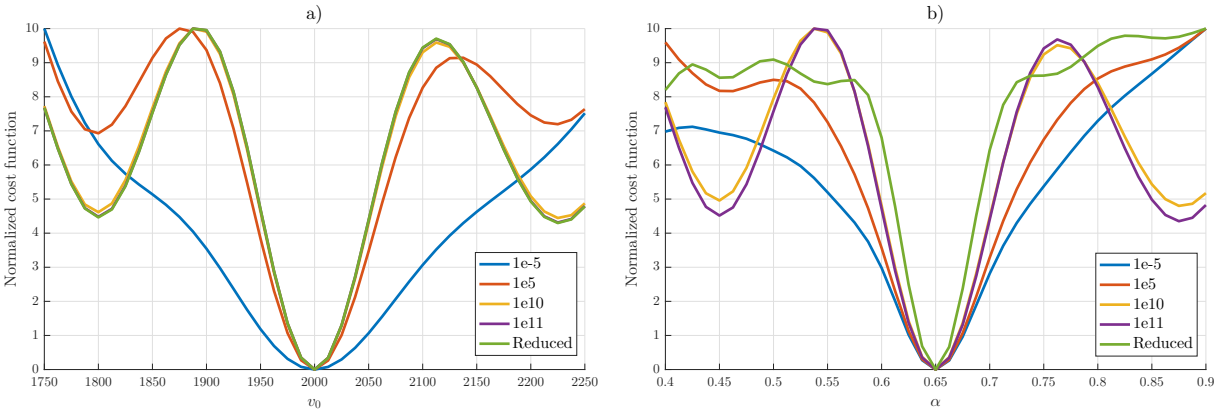

Figure 2.7 – Cost function of gradient velocity model for reduced approach and WRI. (a) When α is kept fixed equal to the true α. (b) When v_0 is kept fixed equal to the true v_0. Both of them are shown for WRI with different values of λ as well as reduced approach.

the wavefield reconstruction with a new set of data residuals, which contribute to progressively downward continue the model updates until the convergence point is reached.

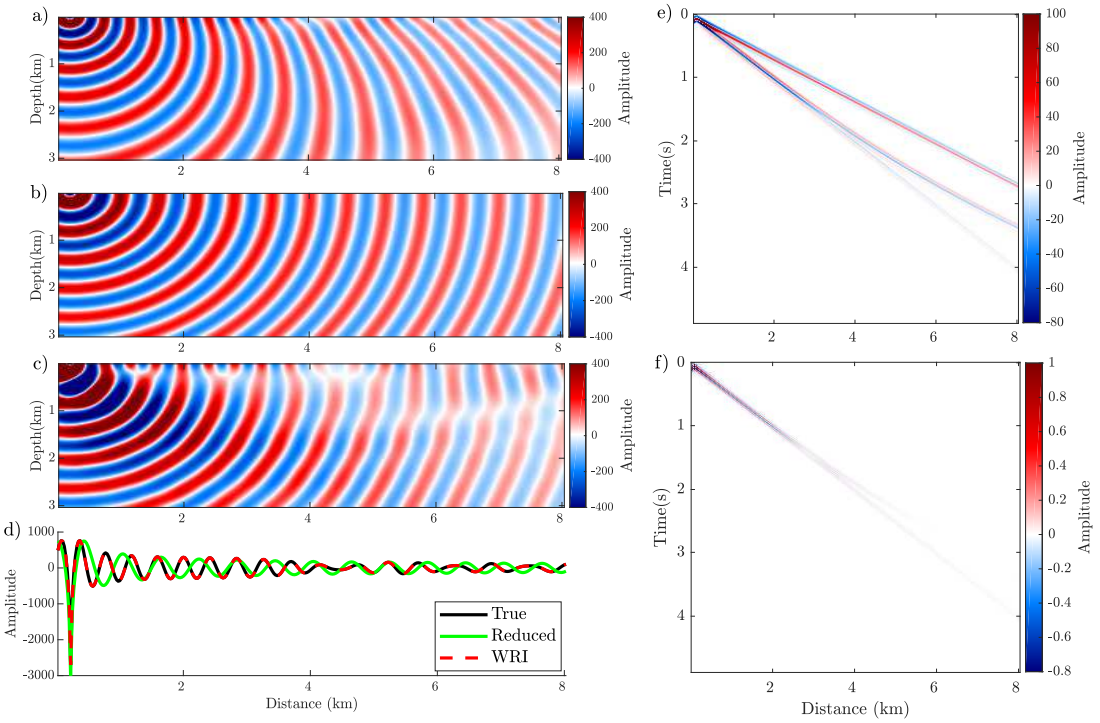

Figure 2.8 – Velocity gradient test. (a) Monochromatic wavefield computed in the true gradient model $v_0 = 2000\ m/s$ and $\alpha = 0.651/s$ (Figure 2.5a) for the 5 Hz frequency. (b) Reduced approach wavefield ($v_0 = 3000\ m/s$ and $\alpha = 0$). (c) WRI wavefield ($v_0 = 3000\ m/s$ and $\alpha = 0$). (d) Direct comparison between the real part of the true (black), reduced-approach (green) and WRI (red) wavefields at receiver positions. (e-f) Time-domain residual seismograms between true and reduced-approach (e) and WRI (f).

Two-layer velocity model

I continue with a two-layer model test. A single-ended spread surface acquisition is used for a two-layers velocity model which has $8\ km$ distance and $3\ km$ depth. A line of sources which are deployed at the surface with source interval $200m$ will be recorded by some receivers which are deployed on the top of the model with $50\ m$ spacing. The velocity model contains two layers with thickness 1 and $2\ km$ with constant velocity 1.9 and $2\ km/s$, respectively (Figure 2.10a). I show the mono-chromatic 5 Hz wavefield computed with the reduced approach and WRI when the homogeneous initial velocity $v_0 = 2200m/s$ (Figure 2.10b) is used as the initial model. The true, reduced approach and WRI wavefields are shown in Figures 2.11a-c, respectively. Also, a comparison between the frequency domain data at receiver positions computed in the true and background models are shown in Figure 2.11d while the time-domain residuals of these wavefields are shown in Figure 2.11e-f, for reduced approach and WRI, respectively. The mismatch between the true and background wavespeeds is sufficient to generate cycle skipping. Comparing the WRI wavefield and the wavefield computed in the true model (Figures 2.11a-

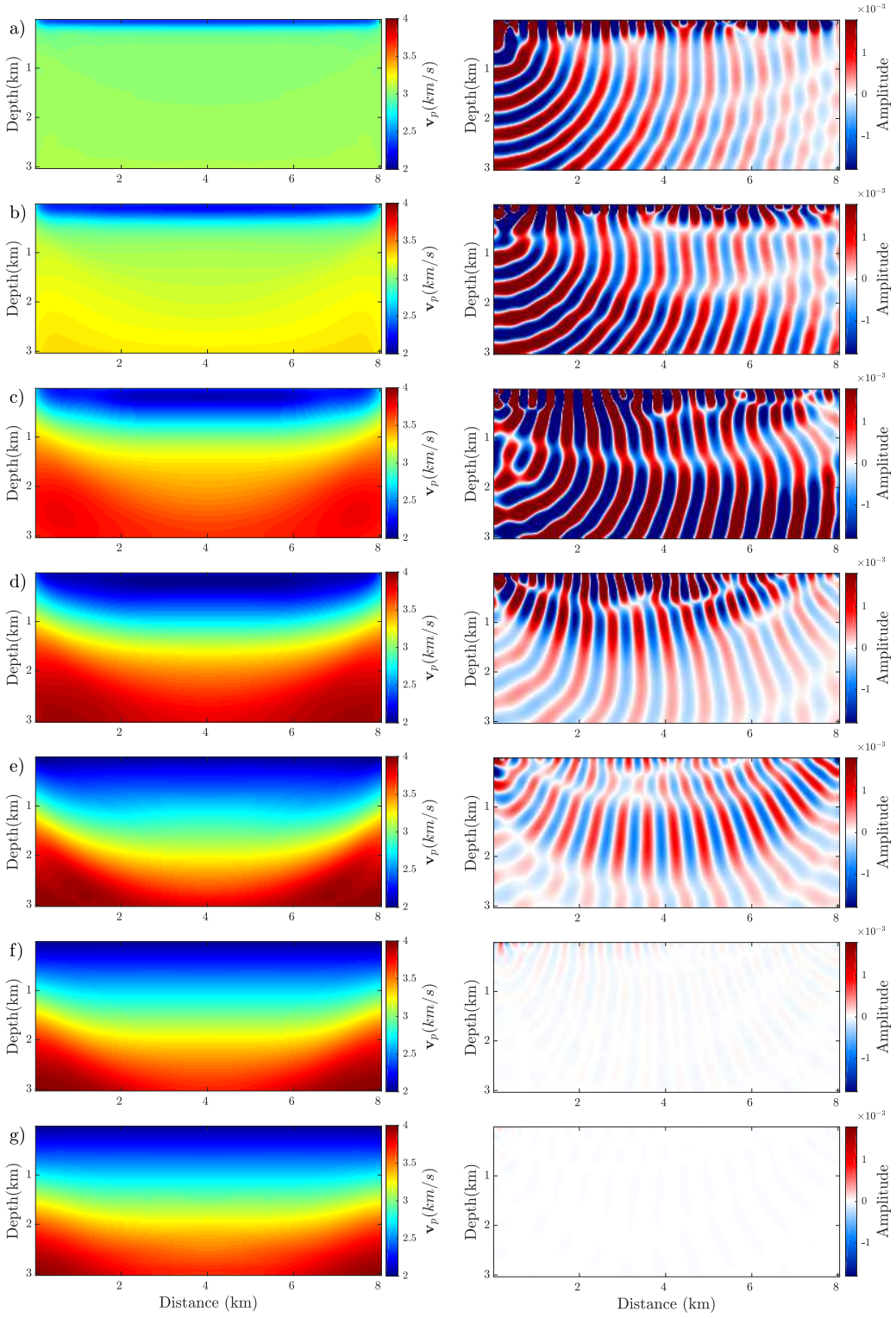

Figure 2.9 – Velocity gradient test. The left column are updated models and the right column are source residuals at iteration (a) 1, (b) 2, (c) 3, (d) 10, (e) 20, (f) 50 and (f) 100. The true and initial velocity model are shown in Figure 2.5a-b, respectively.

c) clearly shows how the wavefield accuracy would have extended downward because of the reflector in the true model. I invert the 5 Hz data up to 100 iterations with ADMM based

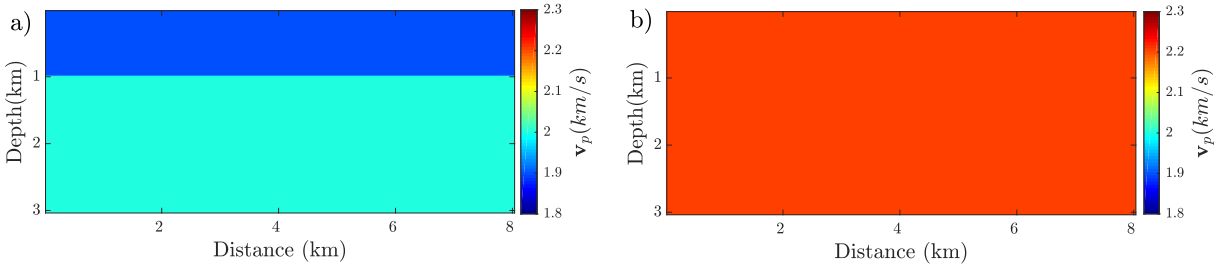

Figure 2.10 – Two-layer test. (a) True model. (b) Initial model.

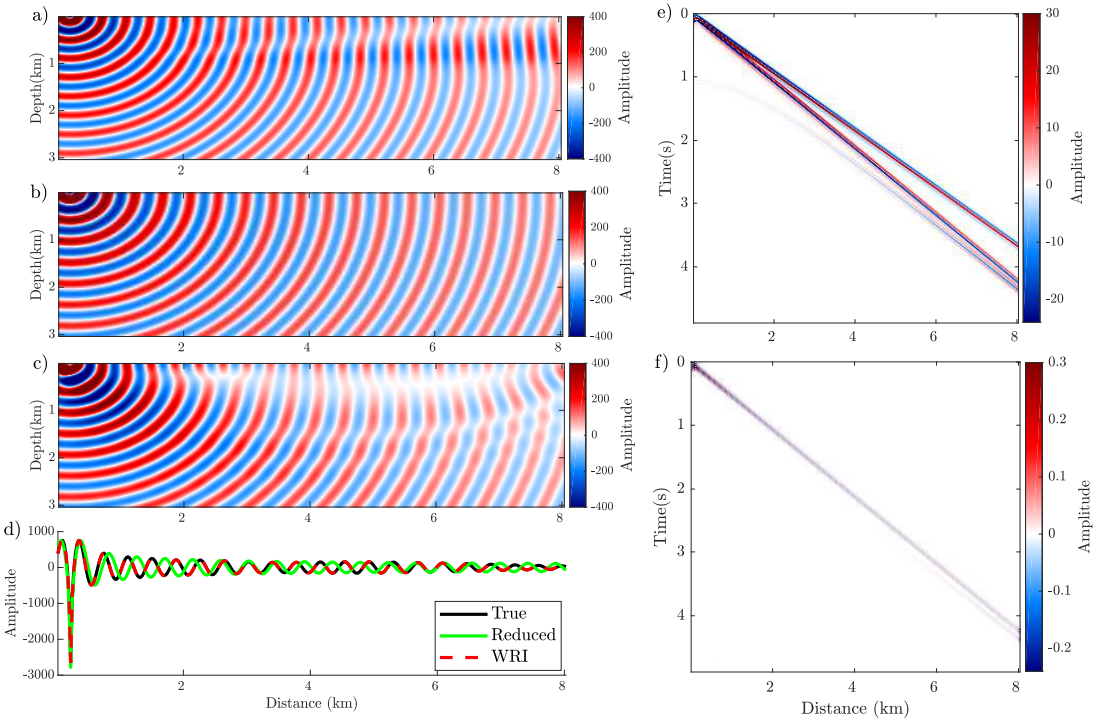

Figure 2.11 – Two-layer test. (a) Monochromatic wavefield computed in the true model, Figure 2.10a, for the 5 Hz frequency. (b) Reduced approach wavefield ($v_0 = 2200\ m/s$). (c) WRI wavefield ($v_0 = 2200\ m/s$). (d) Direct comparison between the real part of the true (black), reduced-approach (green) and WRI (red) wavefields at receiver positions. (e-f) Time-domain residual seismograms between true and reduced-approach (e) and WRI (f).

WRI when a smoothing filter with edge preserving property on the updated model at each iteration is applied. Figures 2.12a-g shows iterations 1, 2, 4, 10, 20, 50 and 100 of this process, respectively, where the updated model are in the left column and the source residuals are shown in the right column of the figure. The velocity updates are relatively shallow at early iterations. However, we already guess the top of the reflector at 1km depth after the first velocity update (Figure 2.12a). This highlights how the footprint in the surface measurements of the wavefields

reflected from the top of the second layer and refracted along the reflector helped to improve the wavefield reconstruction at greater depths compared to the two former tests. The next iterations contribute to refine the velocity in the upper layer (Figure 2.12f), before refining the velocity in the bottom layer (Figure 2.12g).

Two-layer velocity model: on the benefit of long-offset acquisitions

I continue with an example which illustrates when IR-WRI can fail due to the velocity-depth ambiguity and how this failure can be avoided by increasing the offsets. I keep the configuration of the two-layer test and only increase the depth of the reflector to $2.5\ km$ (Figure 2.13a). By doing so the recorded wavefield only contains the direct wave which propagates on the top part of the first layer and short-spread reflection from the reflector. Like previous test, the $5\ Hz$ data are inverted up to 100 iterations with ADMM based WRI when a smoothing filter with edge preserving property is applied on the updated model at each iteration and a homogenous initial model with a velocity of $2.2km/s$ is used as initial model (Figure 2.13b). The velocity models obtained at iterations 1, 2, 4, 10, 20, 50 and 100 are shown in the left column of Figures 2.14(a-f), while the related source residuals are shown in the right column. It can be seen that the inversion remains stuck in a local minimum because the maximum source-receiver offset is not long enough to allow for the recording of the post-critical reflections and refracted waves from the reflector. Moreover, the half width of the first Fresnel zone associated with the direct wave propagation in the upper layer is not enough (850m) for the 8 km maximum offset to illuminate the deep part of the upper layer. This incomplete illumination of the subsurface leads to an ambiguity between the velocity in the upper layer and the depth of the reflector which cannot be resolved by WRI.

To validate this claim, I repeat this test but when the maximum offset is extended to $16\ km$ (Figure 2.15a). The estimated models as well as source residuals for same iterations are shown in Figures 2.16(a-f). The final velocity model shows that IR-WRI was able to position the reflector and estimate the top and bottom wavespeeds accurately.
This series of two tests highlight how WRI takes advantage of long-offset acquisition not only to perform broadband imaging but also to converge more rapidly as long offsets will increase the maximum penetration depth of waves and hence, the sensitivity of the data recorded at the surface to the deep parts of the subsurface. This improved illumination contributes to perform more accurate wavefield reconstruction, which is translated to more accurate velocity estimation by source residual minimization. Indeed, long offsets exacerbate cycle skipping, which is however mitigated by forcing the data fit through the wave equation relaxation.

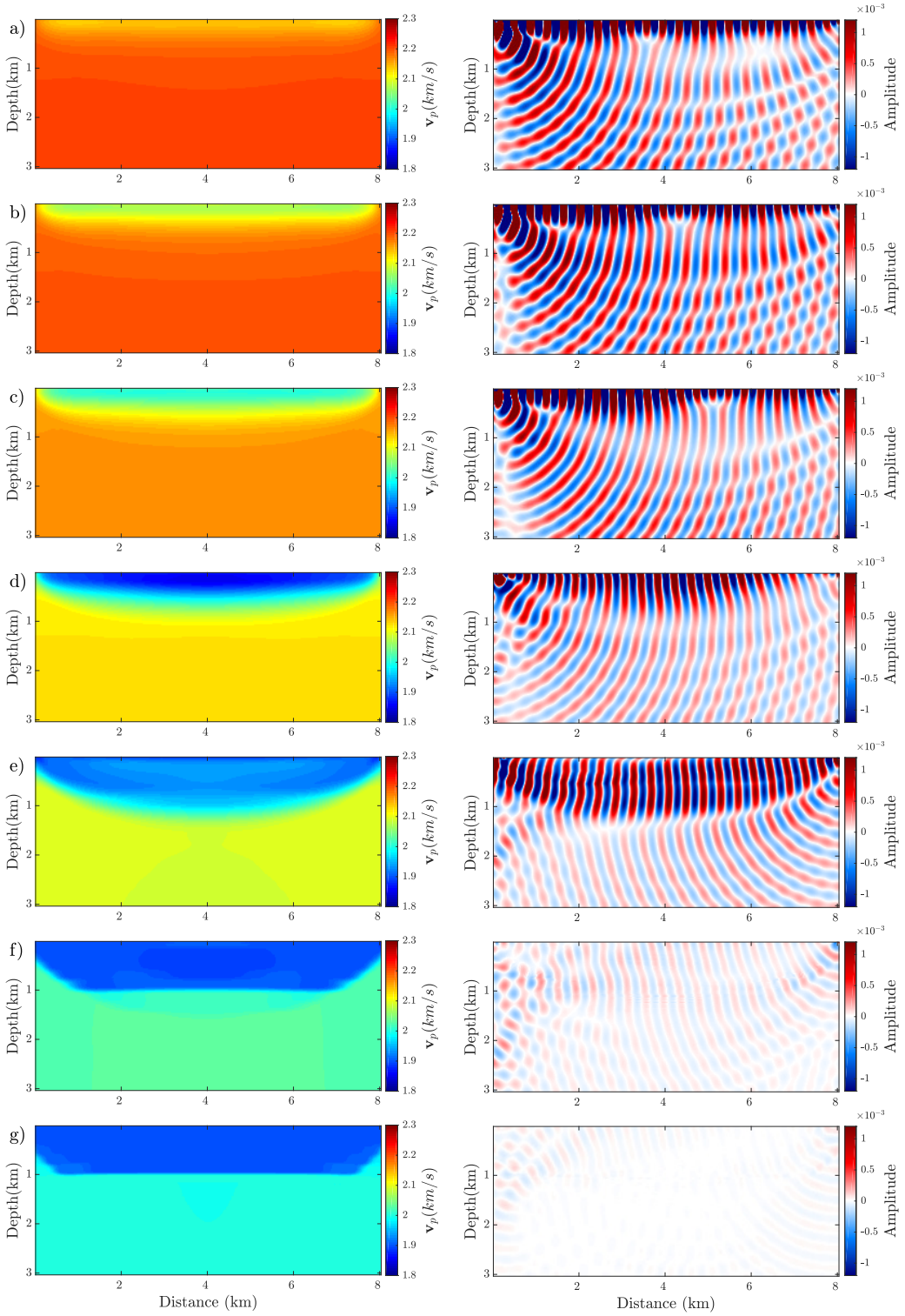

Figure 2.12 – Two-layer test. The left column are updated models and the right column are source residuals at iteration (a) 1, (b) 2, (c) 3, (d) 10, (e) 20, (f) 50 and (f) 100. The initial and true velocity model are shown in Figures 2.10a-b, respectively.

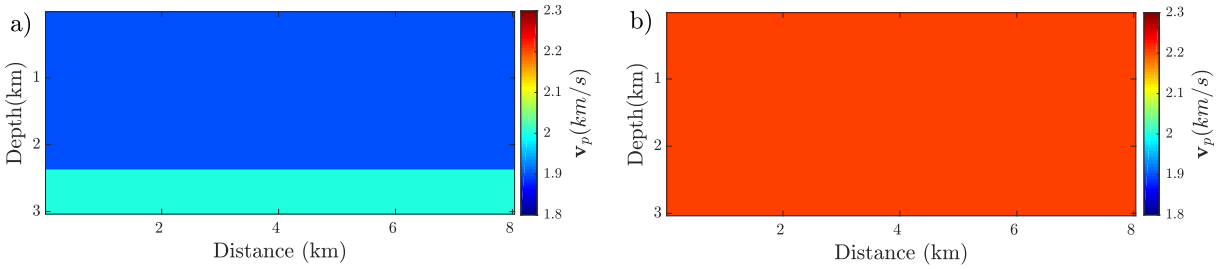

Figure 2.13 – Two-layer test with a deep reflector. (a) True model. (b) Initial model.

Two-layer velocity model: contribution of reflections

I show another two-layer test which is similar to the previous one except that the velocity of the homogeneous starting model is equal to the velocity of the top layer of the true model ($v_0 = 1800m/s$) (Figure 2.17b). With this experimental setup, I want to illustrate how WRI can reconstruct the subsurface from reflected and refracted waves only via a migration-like (positioning of the reflector) and a tomography-like reconstruction (estimation of the bottom velocity). The two-layer velocity model of dimension $8\ km \times 3\ km$ contains two layers of thickness $1.5\ km$ with constant velocities of 1.8 and $2\ km/s$, respectively (Figure 2.17a). The acquisition setup for this test is the same as the previous ones. The monochromatic 5 Hz wavefields computed (exactly) in the true model and the homogeneous initial model are shown in Figures 2.18a-b, while the wavefield computed in the initial model with a relaxation of the wave equation is shown in Figure 2.18c. Also, a comparison between the frequency domain data at receiver positions computed in the true and initial models are shown in Figure 2.18d while the time-domain residuals of these wavefields are shown in Figure 2.18e-f, when wave equation relaxation is used or not. The time-domain residuals computed without relaxation clearly show significant residuals of the reflection and refraction from the reflector, while the wavefield computed with relaxation match these events thanks to the feedback term to the data (the residual is near zero in this case). Since I use here a maximum offset, which is long enough, the $5\ Hz$ inversion with ADMM based WRI (with a smoothing filter with edge-preserving property on the updated model) manages to reconstruct the subsurface model as shown in Figure 2.19.

2.3 ADMM based wavefield reconstruction inversion

WRI has shown good convexity properties when small values of penalty parameter are used. However, one well known issue of penalty method resides in the tuning of penalty parameter which requires tedious continuation strategies proceeding from small values during early iterations to relax the constraint and fit the data from inaccurate model to large value near the convergence point to guarantee that the constraint is satisfied. We overcome this convergence issue by replacing the penalty method by the augmented Lagrangian method and ADMM to solve the constrained optimization problem 2.13. Compared to penalty methods, augmented

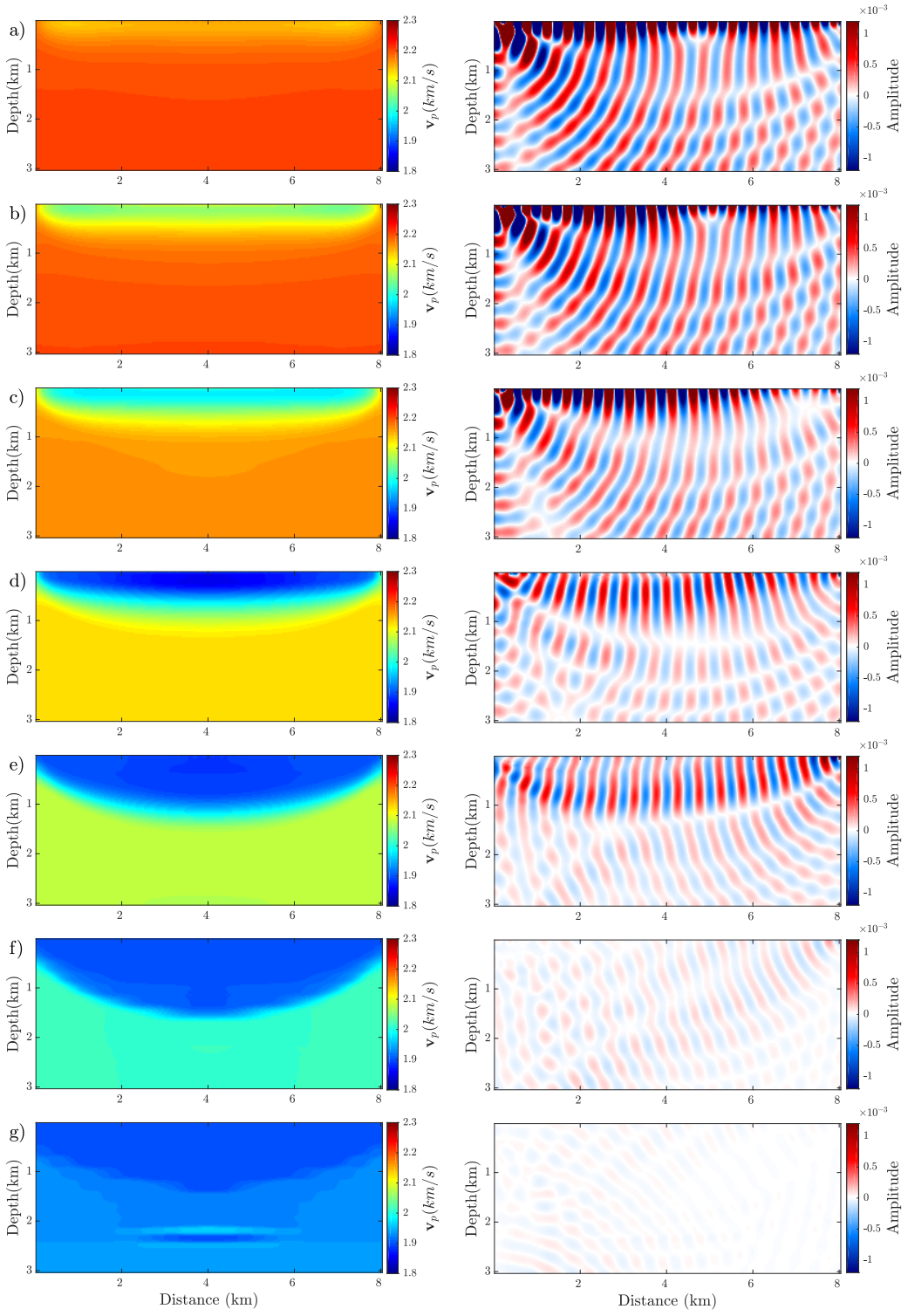

Figure 2.14 – Two-layer test with a deep reflector. The left column are updated models and the right column are source residuals at iteration (a) 1, (b) 2, (c) 3, (d) 10, (e) 20, (f) 50 and (f) 100. The initial and true velocity model are shown in Figures 2.10a-b, respectively.

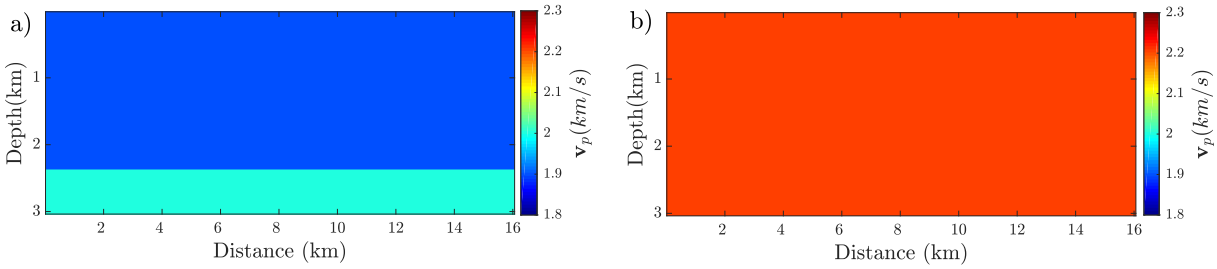

Figure 2.15 – Extended two-layer test with a deep reflector. (a) True model. (b) Initial model.

Lagrangian methods refine the solution more efficiently and accurately when a fixed penalty parameter is used. The following section is a paper published in *Geophysics* (Aghamiry et al., 2019c), which reviews in details the theory, the algorithm and the numerical assessment (accuracy of the solution, convergence history) with synthetic examples of ADMM-based WRI.

Improving full waveform inversion by wavefield reconstruction with the alternating direction method of multipliers
Hossein S. Aghamiry, Ali Gholami and Stéphane Operto
Geophysics, 2019, **84(1)**, pages R139-R162
DOI: 10.1190/geo2018-0093.1

2.3.1 Summary

Full waveform inversion (FWI) is an iterative nonlinear waveform matching procedure subject to wave-equation constraint. FWI is highly nonlinear when the wave-equation constraint is enforced at each iteration. To mitigate nonlinearity, wavefield-reconstruction inversion (WRI) expands the search space by relaxing the wave-equation constraint with a penalty method. The pitfall of this approach resides in the tuning of the penalty parameter because increasing values should be used to foster data fitting during early iterations while progressively enforcing the wave-equation constraint during late iterations. However, large values of penalty parameter lead to ill-conditioned problems. Here, this tuning issue is solved by replacing the penalty method by an augmented Lagrangian method equipped with operator splitting (IR-WRI as iteratively-refined WRI). It is shown that IR-WRI is similar to a penalty method in which data and sources are updated at each iteration by the running sum of the data and source residuals of previous iterations. Moreover, the alternating direction strategy exploits the bilinearity of the wave equation constraint to linearize the subsurface model estimation around the reconstructed wavefield. Accordingly, the original nonlinear FWI is decomposed into a sequence of two linear subproblems, the optimization variable of one subproblem being passed as a passive variable for the next subproblem. The convergence of WRI and IR-WRI are first compared with a simple transmission experiment, which lies in the linear regime of FWI. Under the same conditions, IR-WRI converges to a more accurate minimizer with a smaller number of iterations than WRI.

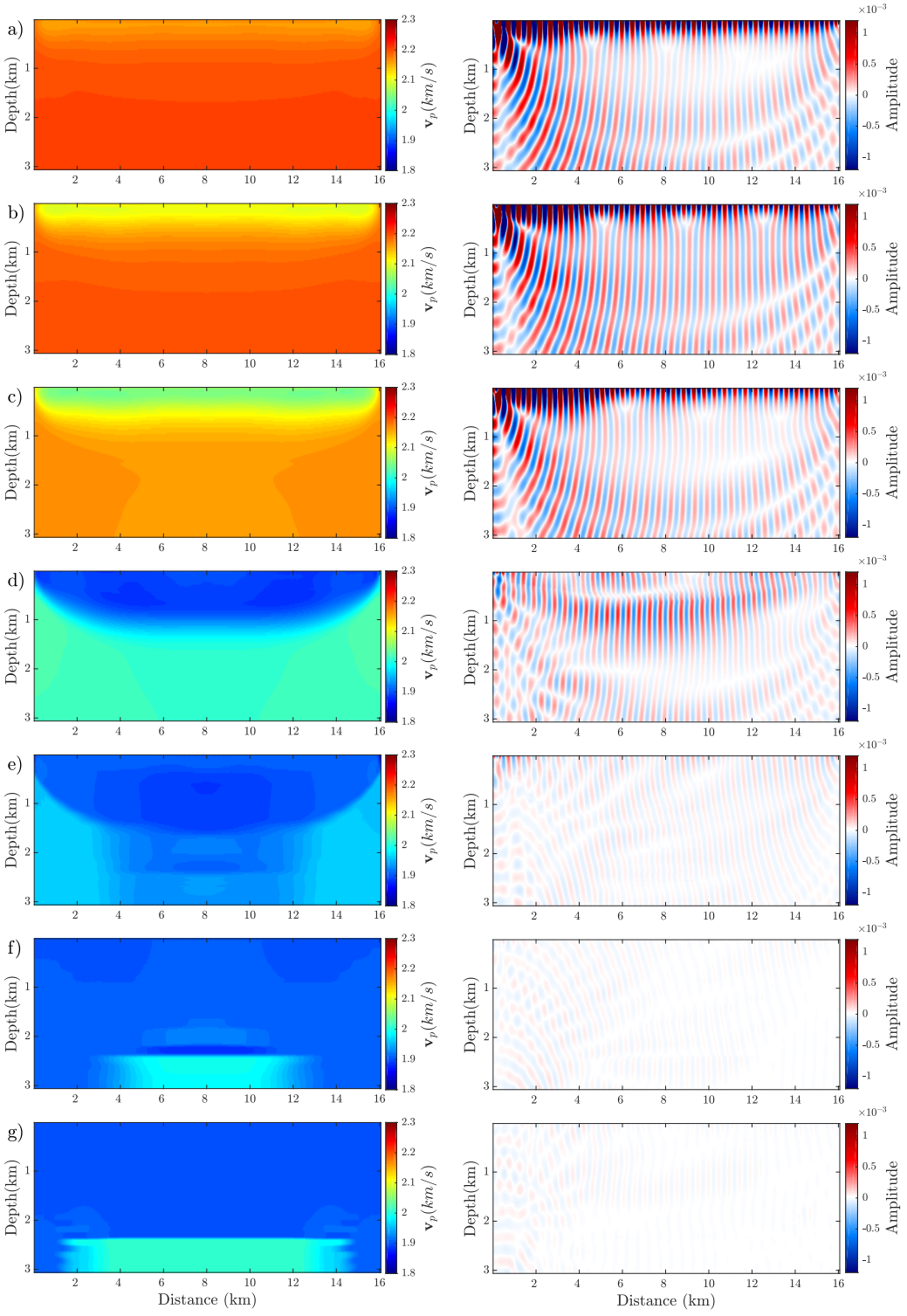

Figure 2.16 – The extended two-layer test with a deep reflector. The left column are updated models and the right column are source residuals at iteration (a) 1, (b) 2, (c) 3, (d) 10, (e) 20, (f) 50 and (f) 100. The initial and true velocity model are shown in Figures 2.10a-b, respectively.

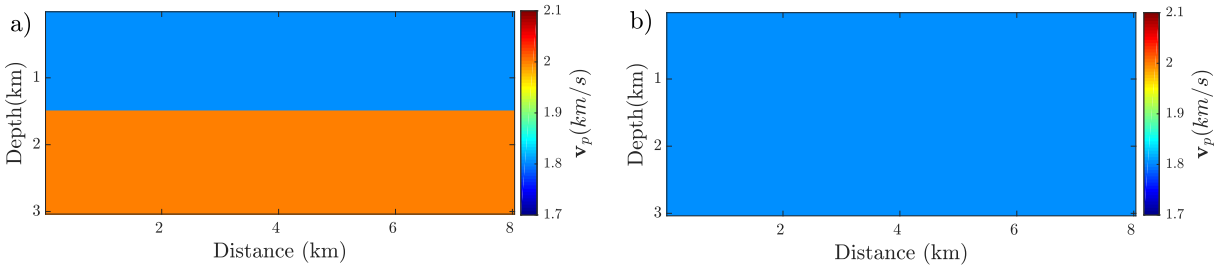

Figure 2.17 – Two-layer test. (a) True model. (b) Initial model.

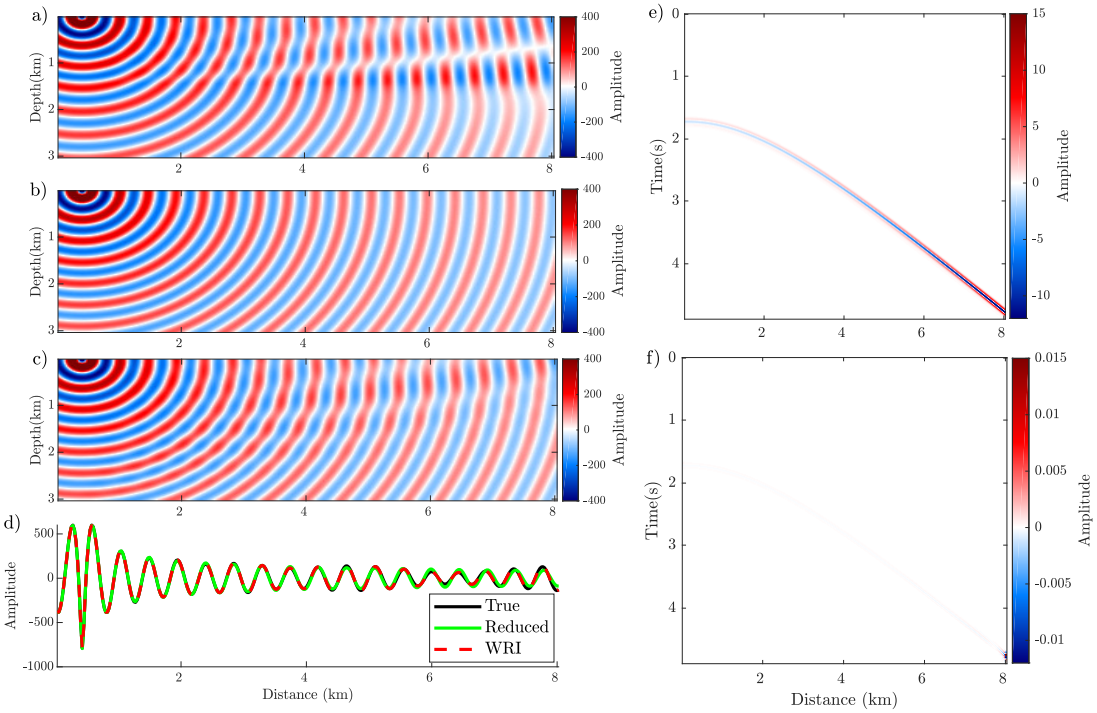

Figure 2.18 – Two-layer test. (a) Monochromatic wavefield computed in the true model, Figure 2.17a, for the 5 Hz frequency. (b) Reduced approach wavefield ($v_0 = 1800\ m/s$). (c) WRI wavefield ($v_0 = 1800\ m/s$). (d) Direct comparison between the real part of the true (black), reduced-approach (green) and WRI (red) wavefields at receiver positions. (e-f) Time-domain residual seismograms between true and reduced-approach (e) and WRI (f).

More realistic case studies performed with the Marmousi II and the BP salt models show the resilience of IR-WRI to cycle skipping and noise, as well as its ability to reconstruct with high fidelity large-contrast salt bodies and sub-salt structures starting the inversion from crude initial models and a 3-Hz starting frequency.

2.3.2 Introduction

Full Waveform Inversion (FWI) is a leading-edge seismic imaging method that has gained renewed interest in oil and gas exploration since one decade due to the emergence of super-

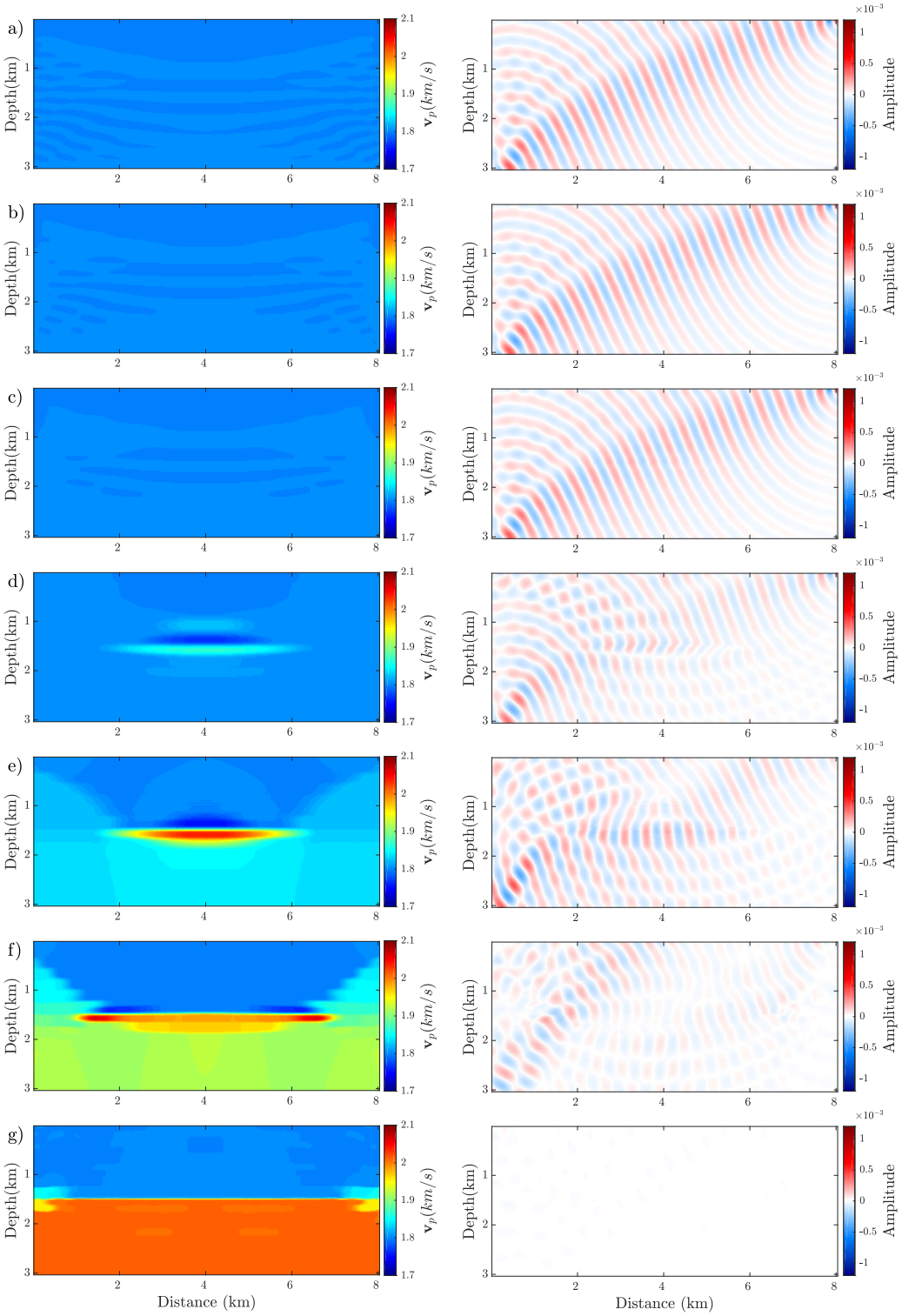

Figure 2.19 – Two-layer test. The left column are updated models and the right column are source residuals at iteration (a) 1, (b) 2, (c) 3, (d) 10, (e) 20, (f) 50 and (f) 100. The initial and true velocity model are shown in Figures 2.17a-b, respectively.

computers, long-offset wide-azimuth acquisition, and broadband seismic sources (Virieux and Operto, 2009; Virieux et al., 2017). FWI seeks to estimate subsurface parameters with a resolution close to the seismic wavelength by fitting seismic records with modeled seismograms that are computed with two-way modeling engines to account for the full complexity of the wavefields (Tarantola, 1984; Pratt et al., 1998).

From the numerical optimization viewpoint, FWI can be defined as a partial differential equation (PDE)-constrained nonlinear optimization problem, where the equality constraint is the wave equation. A popular method to solve such nonlinear constrained optimization problem is the method of Lagrange multiplier, where the constraints are implemented through a Lagrangian function (e.g. Haber et al., 2000). A local minimizer satisfies the so-called KKT (Karush-Kuhn-Tucker) first-order optimality conditions at the stationary point of the Lagrangian. In the frame of local optimization methods, the so-called full-space approach satisfies the nonlinear KKT conditions with an iterative Newton method which jointly updates three classes of variables (the state variables, the adjoint-state variables and the parameters). However, the dimension of the KKT system makes this full space approach computationally intractable for FWI application. Instead, FWI is classically solved with a reduced approach, namely a variational projection approach where the full-space optimization problem is recast as an unconstrained problem after successive elimination of the state and adjoint-state variables from the KKT system (Haber et al., 2000; Akçelik, 2002; Askan et al., 2007; Epanomeritakis et al., 2008). Although this reduced-approach strictly enforces the constraints (because the wave equation is solved exactly with the current guess of the subsurface parameters) and is computationally tractable, it is highly nonlinear and ill-posed due to the oscillating nature of seismic signals (e.g. Symes, 2008). One of the main source of nonlinearity is cycle skipping, which happens as soon as the current subsurface model is not accurate enough to predict recorded traveltimes with an error lower than half the dominant period (e.g. Virieux and Operto, 2009, their Figure 7). In contrast, full-space approaches are more resilient to cycle skipping because the joint updating of wavefields, adjoint wavefields and subsurface model makes them more versatile to fit the data with inaccurate subsurface models. The cycle skipping criterion is difficult to satisfy when the number of propagated wavelengths increases as might be expected with the development of long-offset wide-azimuth acquisitions (Pratt, 2008; Virieux and Operto, 2009). This prompts the oil industry to develop broadband sources with a richer low frequency content to balance increasing propagation distances and mitigate cycle skipping accordingly (Plessix et al., 2012; Baeten et al., 2013). In parallel with this, many heuristic data-driven approaches based upon frequency, traveltime and offset continuations have been designed to drive the optimization problem toward multiscale imaging from the long wavelengths to the shorter ones (e.g. Górszczyk et al., 2017). To reduce human intervention, more automatic approaches rely on more convex distances between recorded and modeled data such as those based upon instantaneous phase and envelope (Shin and Min, 2006; Bozdağ et al., 2011; Luo and Wu, 2015) or correlation (van Leeuwen and Mulder, 2010). However, these approaches should be rather viewed as tools to build initial models for FWI rather than FWI methods per se because they often achieve an improved convexity at the expense of spatial resolution.

More ambitious approaches extend the linear regime of FWI by enlarging the search space. Extended-domain approaches have been intensively developed for velocity macromodel building in the framework of migration-based velocity analysis (Symes (2008) for a review). These

approaches extends the model space with additional degree of freedom as subsurface offset or time lag, which makes them quite computationally intensive. Then, the inverse problem seeks a velocity model that produces a physical reflectivity through an annihilator applied to the extended migration operator. More recently, van Leeuwen and Herrmann (2013) extended the search space in frequency-domain FWI by replacing the wave-equation constraint by a quadratic penalty term. This penalty method, called wavefield-reconstruction inversion (WRI), relaxes the requirement to satisfy exactly the wave equation at each iteration for the benefit of an improved data fitting. This is achieved by reconstructing the wavefield that best jointly fits in the least-squares sense the observations and satisfy the wave equation. To make WRI computationally tractable, van Leeuwen and Herrmann (2013) perform the wavefield reconstruction and the subsurface parameter estimation in an alternating way: the wavefield is first reconstructed using the available subsurface model as fixed background model, before updating the subsurface parameters by minimizing the source residuals generated by the wave-equation relaxation using the previously-reconstructed wavefield as a fixed background wavefield, this cycle being iterated until convergence. This alternating-direction strategy linearizes the parameter-estimation subproblem around the reconstructed wavefield. Accordingly, parameter updating can be performed with Gauss-Newton iterations, the Hessian being diagonal when the forward problem equation is the Helmholtz equation (van Leeuwen and Herrmann, 2013, their equation 8). Later, van Leeuwen and Herrmann (2016) reformulated WRI as a reduced penalty method implemented with a variable projection approach: the closed-form expression of the extended-domain reconstructed wavefield is injected as a function of the subsurface parameters in the penalty function instead of using this wavefield as a passive variable (i.e., independent to the subsurface parameters). Although this variable elimination makes the parameter-estimation subproblem non linear, van Leeuwen and Herrmann (2016) assess their method with a Gauss-Newton method (by opposition to the full Newton counterpart) to mitigate the computational burden. Moreover, using a sparse approximation of the Gauss-Newton Hessian makes the descent direction of the reduced approach identical to that of the alternating-direction WRI of van Leeuwen and Herrmann (2013). In the following, for sake of concise notations, we will refer WRI to as the alternating-direction penalty method of van Leeuwen and Herrmann (2013), because it is more closely related to the approach developed in this study. Compared to full-space approaches, WRI is less computationally expensive as each source is processed independently and the parameter estimation does not require solving an adjoint-state equation because the Lagrange multiplier has been eliminated from the optimization problem at the benefit of a penalty parameter.

One potential difficulty with WRI resides in the tuning of the penalty parameter λ. Ideally, increasing values should be used during iterations to progressively enforce the wave-equation constraint and, hence satisfy the KKT optimality conditions of the original constrained problem with acceptable precision at the minimizer. A significant issue with this continuation approach is that the Hessian is ill conditioned for large λ. However, van Leeuwen and Herrmann (2016) suggested that a fixed λ equal to a fraction of the largest eigenvalue of the normal operator associated with the augmented wave-equation system may allow for convergence toward solutions with acceptable error bounds.

Several other extended-search methods have been proposed. Luo and Sava (2011), Warner and Guasch (2016), Zhu and Fomel (2016) estimate the filter that allows for the matching be-

tween the recorded and modeled data before minimizing the deviation of this matching filter from a delta function. Closely related to this approach, the source-receiver extension method first matches the data through a trace-per-trace source signature estimation before updating the subsurface model by minimizing the deviation of each non-physical source signature from a reference one through an annihilator (Huang et al., 2017). In line with the approach promoted by Symes (2008), Biondi and Almomin (2014) extend the velocity model along the time-lag axis to account for large traveltime residuals and augment the misfit function with a penalty term which penalizes large time lags through an annihilator. With the same objective of managing large time shifts, Engquist and Froese (2014), Métivier et al. (2016a), Métivier et al. (2016b), Yang et al. (2018) compute the optimal transport between the observation and the modeled data before updating the subsurface parameters by minimizing the optimal-transport distance. Ma and Hale (2013) build traveltime residual maps between recorded and modeled shot gathers by dynamic warping and minimize in a second step these traveltime residuals.

In this framework, this study focuses on an easy-to-implement improvement of WRI, which strongly reduces the sensitivity of the method to the penalty parameter, resulting in a faster convergence and more accurate subsurface and wavefield reconstruction. To achieve this goal, we replace the penalty method by an augmented Lagrangian method which is implemented with a classical primal descent - dual ascent approach. In constrained optimization, the augmented Lagrangian method can be viewed as a combination of a classical Lagrangian method and a penalty method, where the Lagrangian multiplier steers the inversion toward the constraint, while the penalization term regularizes the inversion (Nocedal and Wright, 2006, Chapter 17). As in WRI, we perform wavefield reconstruction and parameter estimation in an alternating mode to make the computational problem tractable, while the Lagrange multipliers are updated with a classical dual ascent approach. This operator splitting draw some connection between our approach and the popular alternating direction method of multiplier (ADMM) (Boyd and Vandenberghe, 2004) with the difference that the wavefield reconstruction and the parameter estimation subproblems are not separable. This non-separability is managed by solving the two subproblems in sequence rather than in parallel, the solution of one subproblem being passed as a passive variable to the next problem. Taking advantage of the bilinearity of the Helmholtz equation, this sequential splitting strategy linearizes the parameter estimation problem around the reconstructed wavefield, which means that the original non linear FWI has been recast as a sequence of two linear subproblems.
A scaled form of the augmented Lagrangian function clearly highlights the main difference between WRI and our method: In our method, the right-hand sides in the two objectives of the penalty function (namely, the data and the seismic sources) are updated at each iteration with the running sum of the data and source residuals of previous iterations, namely, a scaled form of the Lagrange multipliers. This right-hand side updating decreases more efficiently the data and source residuals in iterations and performs a self-adaptive weighting of the competing data-fitting and wave-equation objectives of the penalty function, leading to more accurate minimizers and faster convergence when a fixed penalty parameter is used. Accordingly, we refer our method to as IR-WRI (Iteratively-Refined WRI).
We assess IR-WRI against three synthetic examples. First, a toy example highlights how the convergence history of IR-WRI relative to WRI is improved by data and source residual updating. Then, we apply IR-WRI to two challenging synthetic models (the Marmousi II and the

2004 BP salt models) starting from crude initial models and realistic frequency (3 Hz). As WRI, IR-WRI shows a good resilience to cycle skipping. However, IR-WRI converges to subsurface models of much improved resolution and, in addition, it is more resilient to noise.

This paper is mainly organized in a method, numerical example and discussion sections. In the method section, we first briefly review the principles of full-space and reduced-space FWI before introducing WRI. Then, we introduce the basic principles of IR-WRI, highlight the main differences between WRI and IR-WRI and review the IR-WRI algorithm. We review in more details the mathematical concepts behind IR-WRI in Appendix A (section 2.3.7): we first introduce a scaled form of the augmented Lagrangian method to recast IR-WRI as a penalty method with right-hand side updating before describing the operator splitting procedure for alternated wavefield reconstruction and parameter estimation with dual updating of the Lagrange multipliers. In the second part of the paper, we present three numerical examples where we compare the results of reduced-space FWI, WRI and IR-WRI. We first illustrate the impact of the right-hand side updating (data and source) on the convergence history of IR-WRI with a toy example. Then, we illustrate the ability of IR-WRI to image complex and large-contrast media with the Marmousi II and BP salt models. In the final discussion section, the results of the numerical examples are interpreted in light of the IR-WRI methodological ingredients, some analogies with other optimization method such as the Bregman method are discussed and some perspective works are introduced.

2.3.3 Method

From FWI to WRI: a short review

The FWI can be formulated in the frequency domain as the following nonlinear multivariate PDE-constrained optimization problem (e.g. van Leeuwen and Herrmann, 2013, 2016):

$$\min_{\mathbf{m},\mathbf{u}} \quad \|\mathbf{P}\mathbf{u}-\mathbf{d}\|_2^2, \quad \text{subject to} \quad \mathbf{A}(\mathbf{m})\mathbf{u}=\mathbf{b}, \tag{2.23}$$

where $\|\cdot\|_2^2$ denotes the Euclidean norm, $\mathbf{m}\in\mathbb{R}^{N\times 1}$ the discrete subsurface model parameters, $\mathbf{b}\in\mathbb{C}^{N\times 1}$ the source term, $\mathbf{u}\in\mathbb{C}^{N\times 1}$ the modeled wavefield, $\mathbf{d}\in\mathbb{C}^{M\times 1}$ the recorded seismic data and $\mathbf{P}\in\mathbb{R}^{M\times N}$ a linear observation operator that samples $\mathbf{u}$ at the receiver positions. We consider one frequency and one source in equation 2.23 for sake of notation compactness. Extension to multiple frequencies and sources is simply implemented by summation over sources and frequencies in the objective function and by adding multiple right-hand sides in the constraint, equation 2.23. The matrix $\mathbf{A}(\mathbf{m})\in\mathbb{C}^{N\times N}$, whose coefficients depend on $\mathbf{m}$, represents the discretized PDE (Pratt et al., 1998; Plessix, 2007). In this study, we limit ourselves to the Helmholtz equation assuming a constant density equal to 1,

$$\mathbf{A}(\mathbf{m}) \;=\; \mathbf{\Delta}+\omega^2\mathbf{C}(\mathbf{m})\text{diag}(\mathbf{m})\mathbf{B}, \tag{2.24}$$

where ω is the angular frequency, $\mathbf{\Delta}$ a discretized Laplace operator and the subsurface parameters $\mathbf{m}$ are parametrized by the squared slowness. The operator $\mathbf{C}$ enclose boundary

conditions, which can be a function of $\mathbf{m}$ (e.g., Robin paraxial conditions (Engquist and Majda, 1977)) or independent from $\mathbf{m}$ (e.g., sponge-like absorbing boundary conditions such as perfectly-matched layers (Bérenger, 1994)). Also, the linear operator $\mathbf{B}$ spreads the "mass" term $\omega^2\mathbf{C}(\mathbf{m})\text{diag}(\mathbf{m})$ over all the coefficients of the stencil to improve its accuracy following an anti-lumped mass strategy (Marfurt, 1984; Jo et al., 1996; Hustedt et al., 2004). The nonlinear constrained optimization problem, equation 2.23, can be solved with the method of Lagrange multipliers and Newton-type algorithm (Haber et al., 2000).

$$\min_{\mathbf{m},\mathbf{u}} \max_{\mathbf{v}} \mathcal{L}(\mathbf{m},\mathbf{u},\mathbf{v}) = \min_{\mathbf{m},\mathbf{u}} \max_{\mathbf{v}} \|\mathbf{P}\mathbf{u}-\mathbf{d}\|_2^2 + \mathbf{v}^T\left[\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{b}\right], \tag{2.25}$$

where $\mathbf{v} \in \mathbb{C}^{N\times 1}$ denotes the Lagrange multiplier (or, adjoint-state wavefield in the FWI terminology) and ^{T} the complex-conjugate (Hermitian) transpose. A full-space approach jointly updates the three classes of variables ($\mathbf{u}$, $\mathbf{v}$, $\mathbf{m}$). A Newton approach would require to solve the large-scale KKT system, whose matrix operator is the multivariate Hessian, the unknowns are gathered in the multivariate descent direction and the right-hand side is the multivariate gradient of the Lagrangian (Haber et al., 2000; Akçelik, 2002; Askan et al., 2007). Instead, a reduced approach (Pratt et al., 1998; Plessix, 2006), which strictly enforces the PDE constraint at each iteration by projection of the full multivariate search space onto the parameter search space, is more commonly used for sake of computational efficiency leading to the monovariate misfit function

$$\min_{\mathbf{m}} J_r(\mathbf{m}) = \min_{\mathbf{m}} \|\mathbf{P}\mathbf{A}^{-1}(\mathbf{m})\mathbf{b}-\mathbf{d}\|_2^2. \tag{2.26}$$

Due to the highly-oscillating nature of the inverse PDE operator $\mathbf{A}^{-1}$ (the Green functions), the reduced minimization problem, equation 2.26, is prone to convergence to inaccurate minimizer when the initial $\mathbf{m}$ is not accurate enough (Symes, 2008; Virieux and Operto, 2009).

To mitigate this nonlinearity issue, van Leeuwen and Herrmann (2013) recast the original constrained optimization problem, equation 2.23, as a multi-variate unconstrained quadratic penalty problem for $\mathbf{u}$ and $\mathbf{m}$ given by

$$\min_{\mathbf{m},\mathbf{u}} J_p(\mathbf{m},\mathbf{u}) = \min_{\mathbf{m},\mathbf{u}} \|\mathbf{P}\mathbf{u}-\mathbf{d}\|_2^2 + \lambda\|\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{b}\|_2^2, \tag{2.27}$$

where the scalar $\lambda > 0$ is the so-called penalty parameter. The hard constraint in the original constrained problem is replaced by a quadratic penalty term, which represents the squares of the constraint violation. Also, the penalty problem, equation 2.27, can be viewed as a compromise between the full-space and the reduced approach in the sense that the full search space $(\mathbf{m},\mathbf{u},\mathbf{v})$ has been projected onto the $(\mathbf{m},\mathbf{u})$ space by enforcing $\mathbf{v} = \lambda[\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{b}]$ in the Lagrangian function, equation 2.25 (van Leeuwen and Herrmann, 2016). By allowing for the PDE constraint violation, van Leeuwen and Herrmann (2013) enlarge the search space and potentially mitigate the inversion nonlinearity. However, the joint update of $\mathbf{u}$ and $\mathbf{m}$ remains challenging from the computational viewpoint and van Leeuwen and Herrmann (2013) resort to an alternating-direction minimization strategy to solve the penalty problem.

A cycle of the algorithm first reconstructs $\mathbf{u}$ by minimizing the objective function, equation 2.27, keeping $\mathbf{m}$ fixed. Then, $\mathbf{m}$ is updated by minimizing the source residuals generated by the wave-equation relaxation, while keeping the reconstructed wavefield fixed. This

minimization pushes the reconstructed wavefield toward the wave equation constraint. This alternating-direction strategy leads to the following sequence of two minimization problems

$$\mathbf{u}^{k+1} = \arg\min_{\mathbf{u}} \|\mathbf{P}\mathbf{u} - \mathbf{d}\|_2^2 + \lambda\|\mathbf{A}(\mathbf{m}^k)\mathbf{u} - \mathbf{b}\|_2^2, \tag{2.28a}$$

$$\mathbf{m}^{k+1} = \arg\min_{\mathbf{m}} \|\mathbf{A}(\mathbf{m})\mathbf{u}^{k+1} - \mathbf{b}\|_2^2, \tag{2.28b}$$

where the solution of one subproblem is passed to the second subproblem as a passive variable.

A well known limitation of penalty methods is the tedious tuning of the penalty parameter through iterations. Ideally, a sequence of increasing λ_k should be used such that the constraint is satisfied with a prescribed tolerance at the last iteration of the minimization problem. However, using large values of λ_k generally leads to ill-conditioned and unstable problems (Nocedal and Wright, 2006, page 505). In the next section, we review our approach based on augmented Lagrangian method with alternating directions to overcome this issue.

IR-WRI: WRI with alternating-direction method of multiplier

A combination of Lagrangian and penalty functionals, equations 2.25 and 2.27, leads to the following augmented Lagrangian function (Nocedal and Wright, 2006; Bertsekas, 2016)

$$\min_{\mathbf{m},\mathbf{u}} \max_{\mathbf{v}} J_a(\mathbf{m},\mathbf{u},\mathbf{v}) = \min_{\mathbf{m},\mathbf{u}} \max_{\mathbf{v}} \|\mathbf{P}\mathbf{u} - \mathbf{d}\|_2^2 + \mathbf{v}^T\left[\mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{b}\right] + \lambda\|\mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{b}\|_2^2. \tag{2.29}$$

The penalty term (third term in the right-hand side of equation 2.29) allows for the relaxation of the wave equation constraint as in the WRI method, while the Lagrange multiplier $\mathbf{v}$ pushes the inversion toward the wave equation constraint. It has been proven mathematically that the augmented Lagrangian method converges to more accurate minimizer than penalty method when a fixed penalty parameter is used provided that the estimation of the Lagrange multiplier is accurate enough (Nocedal and Wright, 2006, Theorems 17.5 and 17.6). Therefore, WRI implemented with augmented Lagrangian method is expected to benefit from the same search space extension than the classical WRI thanks to the penalty term, while converging toward more accurate solution when a fixed penalty parameter is used thanks to the additional leverage provided by the Lagrange multiplier. The remaining issue is related to the computational burden associated with the update of $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{m}$. We review in Appendix A (section 2.3.7) how we overcome this issue: First, we update the primal ($\mathbf{u}$ and $\mathbf{m}$) and dual ($\mathbf{v}$) variables with the so-called method of multiplier, namely a classical primal descent/dual ascent gradient method (e.g. Nocedal and Wright, 2006, Chapter 17). Second, following the governing idea of WRI, we update the two primal variables $\mathbf{u}$ and $\mathbf{m}$ in an alternating mode. This directs us to the alternating-direction method of multiplier (ADMM) (Boyd et al., 2010) applied on biconvex optimization problem, namely, an optimization problem in which two variables (here, $\mathbf{u}$ and $\mathbf{m}$) can be partitioned into sets over which the problem is convex when the other variable is fixed. The reader is referred to Boyd et al. (2010, page 76) for a short discussion on the applicability of ADMM to biconvex problems. Using these strategies, we show in Appendix A (section 2.3.7) that the wavefield reconstruction and the subsurface parameter estimation problem can be recast

as the following minimization subproblems

$$\begin{aligned}
\mathbf{u}^{k+1} &= \arg\min_{\mathbf{u}} \ \|\mathbf{P}\mathbf{u}-\mathbf{d}-\mathbf{d}^k\|_2^2+\lambda\|\mathbf{A}(\mathbf{m}^k)\mathbf{u}-\mathbf{b}-\mathbf{b}^k\|_2^2, && \text{(2.30a)}\\
\mathbf{d}^{k+1} &= \mathbf{d}^k+\mathbf{d}-\mathbf{P}\mathbf{u}^{k+1}, && \text{(2.30b)}\\
\mathbf{b}^{k+\frac{1}{2}} &= \mathbf{b}^k+\alpha[\mathbf{b}-\mathbf{A}(\mathbf{m}^k)\mathbf{u}^{k+1}], && \text{(2.30c)}\\
\mathbf{m}^{k+1} &= \arg\min_{\mathbf{m}} \ \|\mathbf{A}(\mathbf{m})\mathbf{u}^{k+1}-\mathbf{b}-\mathbf{b}^{k+\frac{1}{2}}\|_2^2, && \text{(2.30d)}\\
\mathbf{b}^{k+1} &= \mathbf{b}^{k+\frac{1}{2}}+\alpha[\mathbf{b}-\mathbf{A}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}], && \text{(2.30e)}
\end{aligned}$$

beginning from $\mathbf{d}^0=\mathbf{0}$ and $\mathbf{b}^0=\mathbf{0}$ where $\alpha\in(0,1)$. Equations 2.30a and 2.30d shows that we recast the augmented Lagrangian function as a penalty function in which the right-hand sides (the data and the seismic sources) in the objectives are updated with the running sum of the data and source residuals at previous iterations, namely, $\mathbf{d}^k$ and $\mathbf{b}^k$. These running sums of the residuals represent a scaled form of the Lagrange multipliers, which are updated with gradient ascent steps in equations 2.30b, 2.30c and 2.30e. This right-hand side updating in equations 2.30a and 2.30d is the only difference with the WRI penalty function, equation 2.28. We now provide the closed-form solution of the two primal subproblems for $\mathbf{u}$ and $\mathbf{m}$ and discuss the linearization of the subproblem for $\mathbf{m}$ resulting from the operator splitting.

Optimization over $\mathbf{u}$

The primal subproblem associated with $\mathbf{u}$, equation 2.30a, is a quadratic optimization problem whose closed-form solution is given by

$$\mathbf{u}^{k+1} = \Big[\mathbf{P}^T\mathbf{P}+\lambda\mathbf{A}(\mathbf{m}^k)^T\mathbf{A}(\mathbf{m}^k)\Big]^{-1}\Big[\mathbf{P}^T[\mathbf{d}+\mathbf{d}^k]+\lambda\mathbf{A}(\mathbf{m}^k)^T[\mathbf{b}+\mathbf{b}^k]\Big]. \qquad (2.31)$$

The only difference with the expression given in van Leeuwen and Herrmann (2016, page 8) is the running sum of the data and source residuals $\mathbf{d}^k$ and $\mathbf{b}^k$ in the gradient term (right-most term).

Optimization over $\mathbf{m}$

The primal subproblem for $\mathbf{m}$, equation 2.30d, is more challenging due to the nonlinearity of the forward operator $\mathbf{A}$ in $\mathbf{m}$. However, exploiting the special structure of $\mathbf{A}$, equation 2.24, it is straightforward to show that the operator splitting has linearized the optimization subproblem for $\mathbf{m}$ around $\mathbf{u}$.
Indeed, after rewriting $\mathbf{A}(\mathbf{m})\mathbf{u}^{k+1}$ as

$$\begin{aligned}
\mathbf{A}(\mathbf{m})\mathbf{u}^{k+1} &= \boldsymbol{\Delta}\mathbf{u}^{k+1}+\omega^2\mathbf{C}(\mathbf{m})\mathrm{diag}(\mathbf{m})\mathbf{B}\mathbf{u}^{k+1},\\
&\approx \boldsymbol{\Delta}\mathbf{u}^{k+1}+\mathbf{L}(\mathbf{u}^{k+1})\mathbf{m}, && \text{(2.32)}
\end{aligned}$$

where

$$\mathbf{L}(\mathbf{u}^{k+1}) \;=\; \omega^2\mathbf{C}(\mathbf{m}^k)\mathrm{diag}(\mathbf{B}\mathbf{u}^{k+1}), \tag{2.33}$$

we can recast the subproblem 2.30d as the following linear problem

$$\mathbf{m}^{k+1} \;=\; \arg\min_{\mathbf{m}} \|\mathbf{L}(\mathbf{u}^{k+1})\mathbf{m} - \mathbf{y}^k\|_2^2, \tag{2.34}$$

where

$$\mathbf{y}^k = \mathbf{b} + \mathbf{b}^{k+\frac{1}{2}} - \mathbf{\Delta}\mathbf{u}^{k+1}. \tag{2.35}$$

Note that we have also linearized the operator $\mathbf{A}$ with respect to $\mathbf{m}$ by building the matrix $\mathbf{C}$ from $\mathbf{m}^k$ in equation 2.33 to manage potential nonlinear boundary conditions. However, in case of PML absorbing boundary conditions, this extra linearization is not used. The linear problem, equation 2.34, has a closed-form solution given by

$$\mathbf{m}^{k+1} \;=\; \big[\mathbf{L}(\mathbf{u}^{k+1})^T\mathbf{L}(\mathbf{u}^{k+1})\big]^{-1}\big[\mathbf{L}(\mathbf{u}^{k+1})^T\mathbf{y}^k\big]. \tag{2.36}$$

This expression is the same as van Leeuwen and Herrmann (2013, their equation 8) except that we iteratively update the source $\mathbf{b}$ with the running sum of the wave equation errors in iterations, $\mathbf{b}^{k+\frac{1}{2}}$ in equation 2.35. Note that the Gauss-Newton Hessian ($\mathbf{L}^T\mathbf{L}$) in equation 2.36 is always diagonal.

Now that we have linearized the minimization problem for $\mathbf{m}$ around $\mathbf{u}$ through the splitting procedure, the two subproblems embedded in IR-WRI, equation 2.30, can be discussed in the framework of linear inverse problem theory. The principle of iterative refinement in linear inverse problem theory is reviewed in Appendix B (section 2.3.8). The right-hand side correction terms $\mathbf{b}^k$ and $\mathbf{d}^k$ in the objective function, equation 2.30, gather the running sum of the data and source residuals of previous iterations. For linear problems, these accumulated residuals are linearly related to the running sum of the solution refinements (model perturbations in the classical FWI terminology) performed at previous iterations. Therefore, these right-hand side correction terms gather in the objective function the imprint of the solution refinements performed at previous iterations such that the solution refinement performed at the current iteration relies only on the residual errors (see equation 2.55 in Appendix B, section 2.3.8). This iterative defect correction leads to the error forgetting property discussed by Yin and Osher (2013) in the frame of Bregman iterations, which means that the error correction performed at the current iteration is made independent to the error corrections performed at previous iterations. Here, this iterative solution refinement by right-hand side updating is necessary to correct two sources of errors: the first results from the fact that each primal subproblem is solved keeping fixed the other primal variable and the second from the fact that we solve a constrained problem with a penalty method keeping the penalty parameter fixed. Other applications of iterative error correction in geophysics are for example least-squares migration (e.g. Lambaré et al., 1992; Jin et al., 1992; Ribodetti et al., 2011), which can be viewed as the linear counterpart of the iterative data residual minimization performed by nonlinear FWI. Another application is presented by Gholami et al. (2018) who perform nonlinear amplitude versus offset (AVO) inversion, using a linearized Zoeppritz equation to formulate the iterative linearized inverse problem, while computing the

errors with the nonlinear equation.
Based on the above, the proposed IR-WRI algorithm is summarized in Algorithm 1. Note that, in the outer loop (lines 3-12), the algorithm performs one iteration of the wavefield reconstruction (lines 4-6) and one iteration of parameter estimation (lines 9-10), while several inner iterations of each subproblem can be viewed (these inner iterations are omitted in Algorithm 1 for sake of compactness). The usefulness of these inner iterations is discussed in the next section. Also, note that, in classical ADMM, the dual variables are updated only once per iteration after the residuals generated by each primal subproblem have been gathered. In contrast, we use a variant of ADMM referred to as the Peaceman-Rachford splitting method (PRSM) (Peaceman and Rachford, 1955) to update the dual variables as soon as we update one primal variable (lines 5 and 10 in Algorithm 1). For a review, see Boyd et al. (2010, page 23) and Esser (2009, page 3). The accelerate convergence achieved by PRSM relative to ADMM is illustrated in the next section. In the next section, we will also consider that the results obtained with IR-WRI when right-hand side updating is not activated are representative of those that would be obtained with WRI as shown by the comparison between equations 2.28 and 2.30.

Algorithm 1: FWI algorithm based on the Peaceman-Rachford splitting algorithm.

Initialize: set $k = 0$, $\mathbf{d}^k = \mathbf{0}$, $\mathbf{b}^k = \mathbf{0}$.

Input: $\mathbf{m}^0$

while *convergence criteria not satisfied* **do**

- $\mathbf{u}^{k+1} \leftarrow$ update according to equation 2.31
- $\mathbf{d}^{k+1} \leftarrow \mathbf{d}^k + \mathbf{d} - \mathbf{P}\mathbf{u}^{k+1}$
- $\mathbf{b}^{k+\frac{1}{2}} \leftarrow \mathbf{b}^k + \alpha\left[\mathbf{b} - \mathbf{A}(\mathbf{m}^k)\mathbf{u}^{k+1}\right]$
- $\mathbf{y}^k \leftarrow \mathbf{b} + \mathbf{b}^{k+\frac{1}{2}} - \boldsymbol{\Delta}\mathbf{u}^{k+1}$
- $\mathbf{L}(\mathbf{u}^{k+1}) \leftarrow \omega^2\mathbf{C}(\mathbf{m}^k)\text{diag}(\mathbf{B}\mathbf{u}^{k+1})$
- $\mathbf{m}^{k+1} \leftarrow$ update according to equation 2.36
- $\mathbf{b}^{k+1} \leftarrow \mathbf{b}^{k+\frac{1}{2}} + \alpha\left[\mathbf{b} - \mathbf{A}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}\right]$
- $k \leftarrow k + 1$

2.3.4 Numerical examples

We compare the performance of WRI and IR-WRI against three 2D mono-parameter synthetic examples: a toy example built with a simple inclusion model to compare the convergence history of the two methods and two complex synthetic experiments where the synthetics computed in the starting model are strongly cycle skipped relative to the observables. For each example, no regularization is used. We only apply bound constraints using the true values of the minimum and maximum velocities as bounds for the velocity-model update. Because the interfacing of regularization and additional constraints in our method is not the scope of this study, we have not reviewed the bound-constraint implementation in the method section for sake of clarity. We implement bound constraints through an auxiliary variable and add an extra constraint to the problem following the split Bregman method (Goldstein and Osher, 2009).

The interested reader is also referred to Maharramov and Levin (2015), where the split Bregman method (Goldstein and Osher, 2009) is used to imply bound constraints and total-variation regularization in optimization problems.
For all numerical examples the forward modeling is performed with a nine-point stencil finite-difference method implemented with anti-lumped mass and PML absorbing boundary conditions to solve the Helmholtz equation, where the stencil coefficients are optimized to the frequency (Chen et al., 2013). In this setting the diagonal matrix $\mathbf{C}$ contains the damping PML coefficients and does not depend on $\mathbf{m}$. With this setting, the update of $\mathbf{m}$ can be recast as a linear inverse problem, without additional linearization to that induced by the operator splitting (see equation 2.36).

Inclusion model

The subsurface model contains a box-shape anomaly of side 100 m embedded in a homogeneous background (Figure 2.20a). The dimensions of the model are 1000 m in distance and 700 m in depth and the grid spacing is 10 m. The cross-hole acquisition consists of one source at (x,z)=(0 m,350 m) and 18 receivers deployed vertically on the opposite side of the grid (Figure 2.20a). The source signature is a Ricker wavelet with a 5 Hz dominant frequency. We start the inversion from the homogeneous background model ($v_p = 1800$ m/s) and invert simultaneously three frequency components (2.5, 5 and 7 Hz) with noiseless data. The stopping criterion of iteration is related to the level of the wave-equation error.
First, we show the final reconstructed wavefields (Figure 2.20e,f) and velocity models (Figure 2.20b,c) as well as the final data misfit (Figure 2.20g) obtained with WRI and IR-WRI when the value of the penalty parameter λ, equation 2.31, has been chosen to be a small fraction (1e-4) of the largest eigenvalue of $\mathbf{A}^{-T}\mathbf{P}^T\mathbf{P}\mathbf{A}^{-1}$ (μ_1) according to the scaling proposed by van Leeuwen and Herrmann (2016). WRI fails to build a velocity model as well resolved as the one found by IR-WRI (Figure 2.20b,c). Here, the WRI failure does not obviously result from cycle skipping as this case study lies in the linear regime of classical FWI but instead highlights the inability of the penalty method to retrieve accurate minimizer with a fixed λ. Instead, the iterative updating of the data residuals and wave equation errors performed by the augmented Lagrangian method provides an efficient and automatic substitute to the adaptive tuning of the penalty parameter to jointly satisfy the two competing data-fitting and wave-equation objectives at the final iteration. The true scattered wavefield can be compared with the final WRI and IR-WRI scattered wavefields in Figure 2.20d-f. Due to the transmission acquisition geometry, only the forward component of the scattered wavefield can be reconstructed, leading to a smeared reconstruction of the inclusion. The final scattered wavefield and data residuals confirm that IR-WRI outperforms WRI to reconstruct the wavefield and fit the data (Figure 2.20g-i).

Sensitivity of WRI and IR-WRI to λ and convergence analysis

A more comprehensive convergence analysis can be performed by applying WRI and IR-WRI with several values of λ ranging between 1e-9 μ_1 and 1e3 μ_1 (Figure 2.21). The value of

μ_1 (=1e7) is indicated by the vertical dash line in Figure 2.21a,b,e,f. We stop iteration when a preset value of the wave-equation misfit $\|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b}\|_2$ is reached (Figure 2.21c,g, horizontal dash line). We compare the relative model error $\|\mathbf{m}^k - \mathbf{m}^*\|_2/\|\mathbf{m}^*\|_2$ and the relative wavefield error $\|\mathbf{u}^k - \mathbf{u}^*\|_2/\|\mathbf{u}^*\|_2$ achieved by WRI and IR-WRI as functions of the iteration count k and λ in Figure 2.21a,b,e,f. In parallel with this, we compare the convergence history of the wave-equation objective function $\|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b}\|_2$ and the data misfit objective function $\|\mathbf{P}\mathbf{u}^k - \mathbf{d}\|_2$ achieved by WRI and IR-WRI as functions of k for several λ in Figure 2.21c,d,g,h. The true model and the wavefield computed in the true model (namely, the global minimizers) are denoted by $\mathbf{m}^*$ and $\mathbf{u}^*$, respectively.

For small values of λ ranging between 1e-9 μ_1 and 1e-4 μ_1, IR-WRI reaches the stopping criterion of iteration after a much smaller number of iterations (typically, one order of magnitude smaller) than WRI (Figure 2.21c,g). The most-accurate minimizers (wavefield and subsurface model) are obtained for these small values of λ, whatever WRI or IR-WRI is used, because small values of λ foster data fitting during early iterations before minimizing source residuals during late iterations (due to the small weight assigned to the wave equation objective function) and fulfill the stopping criterion of iteration accordingly (Figure 2.21a,b,e,f). However, the minimizers estimated by IR-WRI are significantly more accurate than those of WRI for a fixed value of λ (compare Figure 2.21a and 2.21e, Figure 2.21b and 2.21f). This highlights that IR-WRI has decreased data residuals not only faster but also more significantly than WRI before reaching the stopping criterion of iteration.

This more efficient data residual minimization performed by IR-WRI likely results from the joint updating of the data and source residuals in iterations, equation 2.30b,2.30c and 2.30e. However, one may wonder whether this joint residual updating performed by IR-WRI presents the risk that the wave-equation based stopping criterion of iteration is prematurely satisfied before achieving a sufficient data fit. The numerical results suggest that this does not occur because each time the wave equation error is decreased, this pushes back the optimization against the data fitting constraint, hence re-balancing the relative weight of the two objectives. This interpretation highlights the self-adaptivity of the augmented Lagrangian method to process competing constraints by iteratively updating the Lagrange multipliers (see Nocedal and Wright (2006, Section 12.8) for the intuitive significance of Lagrange multipliers). As such, the augmented Lagrangian method written in a scaled form, equation 2.45, can be viewed as a self-adaptive penalty method, where the iterative updating of the Lagrange multipliers provides a systematic and easy-to-implement alternative to an ad hoc penalty parameter continuation strategy.

When λ is decreased, the data fit achieved by WRI and IR-WRI becomes closer and closer, since the small weight assigned to the wave-equation objective function leaves some time for WRI and IR-WRI to get close to a local minimum where the objective function is nearly flat before satisfying the stopping criterion of iteration. At the same time, the number of iterations performed by the two approaches becomes increasingly different because the poorly-conditioned WRI needs a prohibitively-high number of iterations before the wave-equation based stopping criterion of iteration is fulfilled.

When higher values of λ are used up to a value of 1e3 μ_1, WRI and IR-WRI satisfy the stopping criterion after a smaller number of iterations (down to around 60 iterations for λ = 10 μ_1) (Figure 2.21c,g) before an acceptable data fit has been achieved (Figure 2.21d,h). This degraded data fit translates into minimizers of more limited accuracy (Figure 2.21a,b,e,f). However, this

degradation of the data fit and minimizer quality is less significant in IR-WRI than in WRI due to the above-mentioned self-adaptive management of the two objectives performed by IR-WRI. It is also instructive to look at the joint evolution in iterations of the data misfit and the wave-equation error on the one hand (Figure 2.22a,b), and the wavefield and subsurface model errors on the other hand (Figure 2.22c,d). Figure 2.22a,b shows how small values of λ foster data fitting with inaccurate $\mathbf{m}$ by allowing for significant wave-equation error during early iterations, which is consistent with the governing idea of expanding the search space. The more complex zigzag path followed by IR-WRI relative to WRI in the $(\|\mathbf{P}\mathbf{u}^k - \mathbf{d}\|_2 - \|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b}\|_2)$ plane for intermediate values of λ (this trend is also illustrated in Figure 2.21g and 2.21h by the non monotonicity of the IR-WRI convergence curves) highlights how the joint updating of the data misfit and wave-equation error dynamically balances the weight of the two objectives in iterations. This zigzag convergence trend translates also into more complex path in the $(\|\mathbf{u}^k - \mathbf{u}^*\|_2/\|\mathbf{u}^*\|_2 - \|\mathbf{m}^k - \mathbf{m}^*\|_2/\|\mathbf{m}^*\|_2)$ plane, which highlights how the solution refinement is pushed toward the wavefield reconstruction or the velocity model estimation according to the self-adapting weighting of the two objectives (Figure 2.22c,d). Figure 2.22c,d also shows that the accuracy gap between the WRI and IR-WRI minimizers becomes increasingly significant as λ is decreased, while the data fit achieved by the two methods becomes closer (Figure 2.22a,b). This highlights the ability of IR-WRI to keep on refining the wavefield and the velocity model when the optimization becomes close to a local minimum where the objective function is increasingly flat.

ADMM versus Peaceman-Rashford dual updating

In IR-WRI algorithm based on the Peaceman-Rachford splitting (a variant of ADMM reviewed in Appendix A, section 2.3.7), the dual variable $\mathbf{b}^k$ is updated twice with a step length equal to α, equation 2.30c and 2.30e, while this updating is performed only once with the ADMM method, equation 2.47. To show the relevance of this double dual updating, we perform IR-WRI for several weights α, equation 2.30, and plot the model and wavefield errors as well as the wave-equation and data misfit as functions of k in Figure 2.23 and compare these results with those obtained with ADMM. The PRSM results obtained with α=1 can lead to an improved minimizers (Figure 2.23a,b), although this setting leads potentially to instabilities. These instabilities are illustrated by a poorer data fit and higher wave equation error when α=1 in Figure 2.23c,d). Overall, PRSM builds improved minimizers when α becomes closer to 1. Also, the PRSM minimizers obtained with $\alpha \geq 0.5$ are more accurate than those obtained with ADMM. Accordingly, we perform the following Marmousi II and BP salt experiments with PRSM using α=0.5.

Waveform inversion linearization and iteration strategy

The IR-WRI algorithm potentially embeds two levels of nested iterations, Algorithm 1. The outer loop over k manages the iterations of the nonlinear multi-variate optimization. Inside one cycle, we can refine the wavefield several times through the iterative updating of $\mathbf{d}$ and $\mathbf{b}$ for the

current $\mathbf{m}^k$, lines 4-6 in Algorithm 1, before refining the parameters several times through the iterative updating of $\mathbf{b}$ for the current $\mathbf{u}^{k+1}$, lines 9-10 in Algorithm 1. For sake of compactness, these two successive inner loops are not explicitly written in Algorithm 1, where one iteration of the wavefield reconstruction and parameter estimation are performed per cycle. Let's denote by n the number of inner iterations for the wavefield reconstruction and the parameter update. We perform IR-WRI for several values of n and show the wavefield and model errors as well as the wave-equation misfit and data misfit as functions of the number of PDE resolution (each inner iteration pair requires one additional PDE resolution) in Figure 2.24. The results clearly show that the best choice is n=1. This can be intuitively understood by reminding that the original nonlinear multivariate optimization for $\mathbf{u}$ and $\mathbf{m}$ has been recast as two subproblems for $\mathbf{u}$ and $\mathbf{m}$ that are solved in sequence. Moreover, this operator splitting allows for the linearization of the model update around the reconstructed wavefield, equation 2.36. This operator splitting strategy implies that, during the resolution of each subproblem, the optimization variable is updated from an inaccurate passive variable. Therefore, the resulting residuals cannot be decreased efficiently by inner iterations due to the inaccuracy of the passive variable which is kept fixed. A more efficient procedure is therefore to pass the residuals and the optimization variable of one subproblem to the next one as soon as they have been updated by one inner iteration. This conclusion is similar to the one reached by Goldstein and Osher (2009, section 3.2) in the different framework of $l1$-regularized constrained optimization problem solved with the split Bregman method, namely an operator splitting method similar to ADMM.

Marmousi model

We consider now the more complex Marmousi II model, which covers a $13750\ m \times 3800\ m$ spatial domain (Figure 2.25a). We perform WRI and IR-WRI in the 3 Hz - 15 Hz frequency band and we re-sample the original model on a 25 m grid accordingly. The fixed-spread surface acquisition consists of 137 sources spaced 100 m apart and 548 receivers spaced 25 m apart at the surface.

A PML absorbing boundary condition is implemented on top of the grid (i.e., free-surface multiples are not involved in the inversion) and the source signature is a Ricker wavelet with a 10 Hz dominant frequency. We design the inversion with a classical multiscale frequency continuation proceeding from the low frequencies to the higher ones. One frequency is processed at a time between 3 Hz and 15 Hz with a 0.5 Hz interval, leading to 25 successive mono-frequency inversions. This frequency continuation strategy does not guarantee that, when the subsurface model is updated at a given frequency, the lower-frequency data are still fitted. To overcome this issue, we perform three cycles of multiscale inversion using the final model of the previous cycle as the initial model of the current cycle and starting the second and third cycles at 5 Hz and 7.5 Hz, respectively. For each frequency, the stopping criterion of iteration is

$$k_{max} = 10 \qquad \text{or} \qquad (\|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b}\|_F \leq \delta \qquad \text{and} \qquad \|\mathbf{P}\mathbf{u}^k - \mathbf{d}\|_F \leq \epsilon_n), \quad (2.37)$$

with δ=1e-3 and ϵ_n=1e-5 and k_{max} is the maximum number of iterations. Here, F refers to the Frobenius norm.
The initial velocity model is a crude laterally homogeneous velocity model in which the veloc-

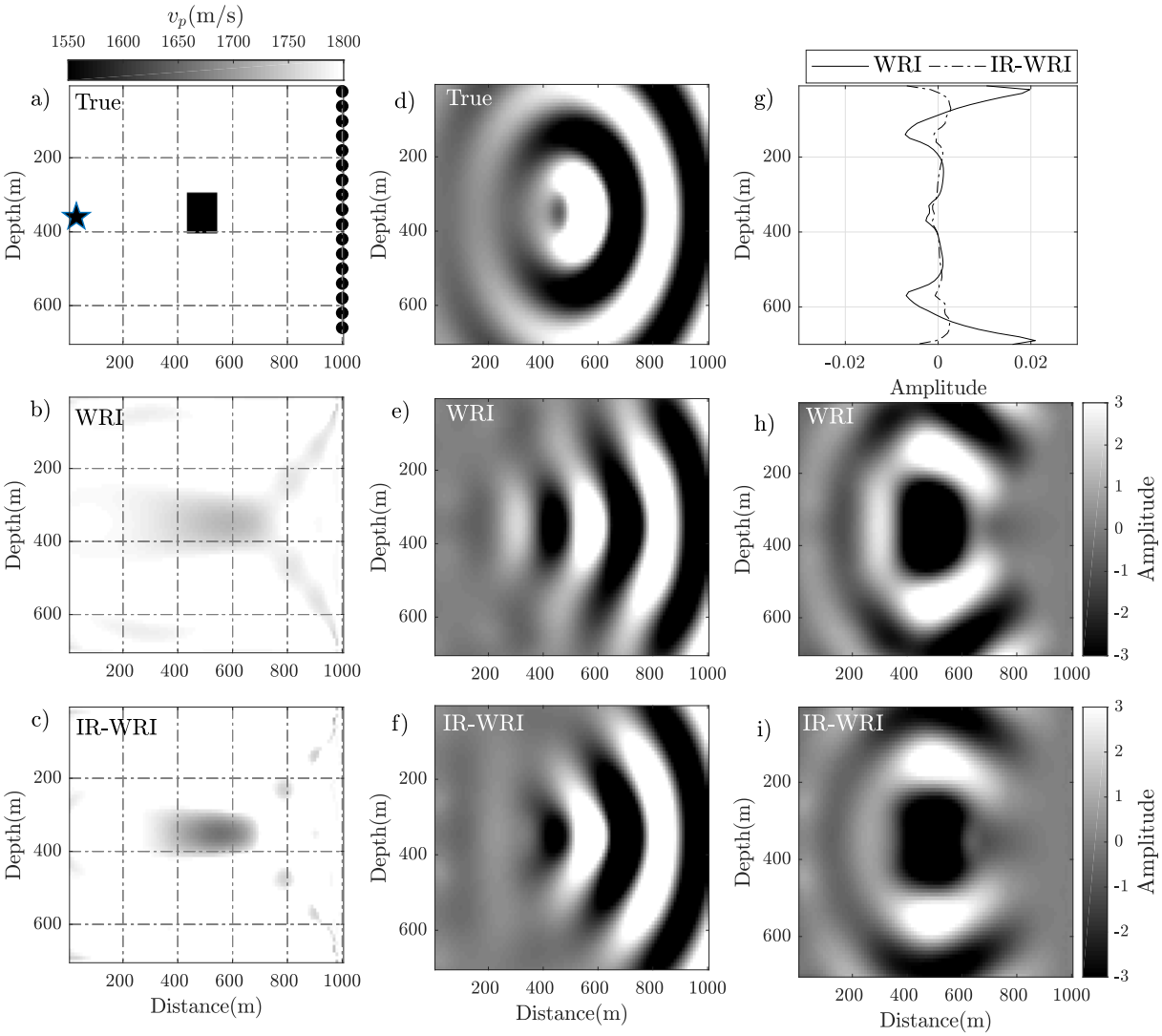

Figure 2.20 – Box-shape anomaly example. (a) True model. The star points the source position, while circles denote receivers. (b-c) WRI (b) and IR-WRI (c) final models for λ=1000 (see Figure 2.21a,e). (d) Wavefield (real part) scattered by the true inclusion for the 7 Hz frequency. (e-f) Wavefield (real part) scattered by the inclusion reconstructed by (e) WRI and (f) IR-WRI for the 7 Hz frequency. (g) Final 7 Hz data residuals $\mathbf{P}\mathbf{u}^k - \mathbf{d}$ (real part) at the receiver positions for WRI (solid line) and IR-WRI (dashed line). (h) Difference between scattered wavefields shown in (d) and (e). (i) Same as (h) for scattered wavefields shown in (d) and (f).

ity linearly increases with depth (Figure 2.25b).

The comparison between a common-shot gather computed in the true and initial models highlights the kinematic inaccuracy of the initial velocity model (Figure 2.26). Considering noiseless data, this inaccuracy makes the reduced-space frequency-domain FWI to be stuck in a local minimum during the 3-Hz inversion (Figure 2.25c). Here, we perform FWI with the

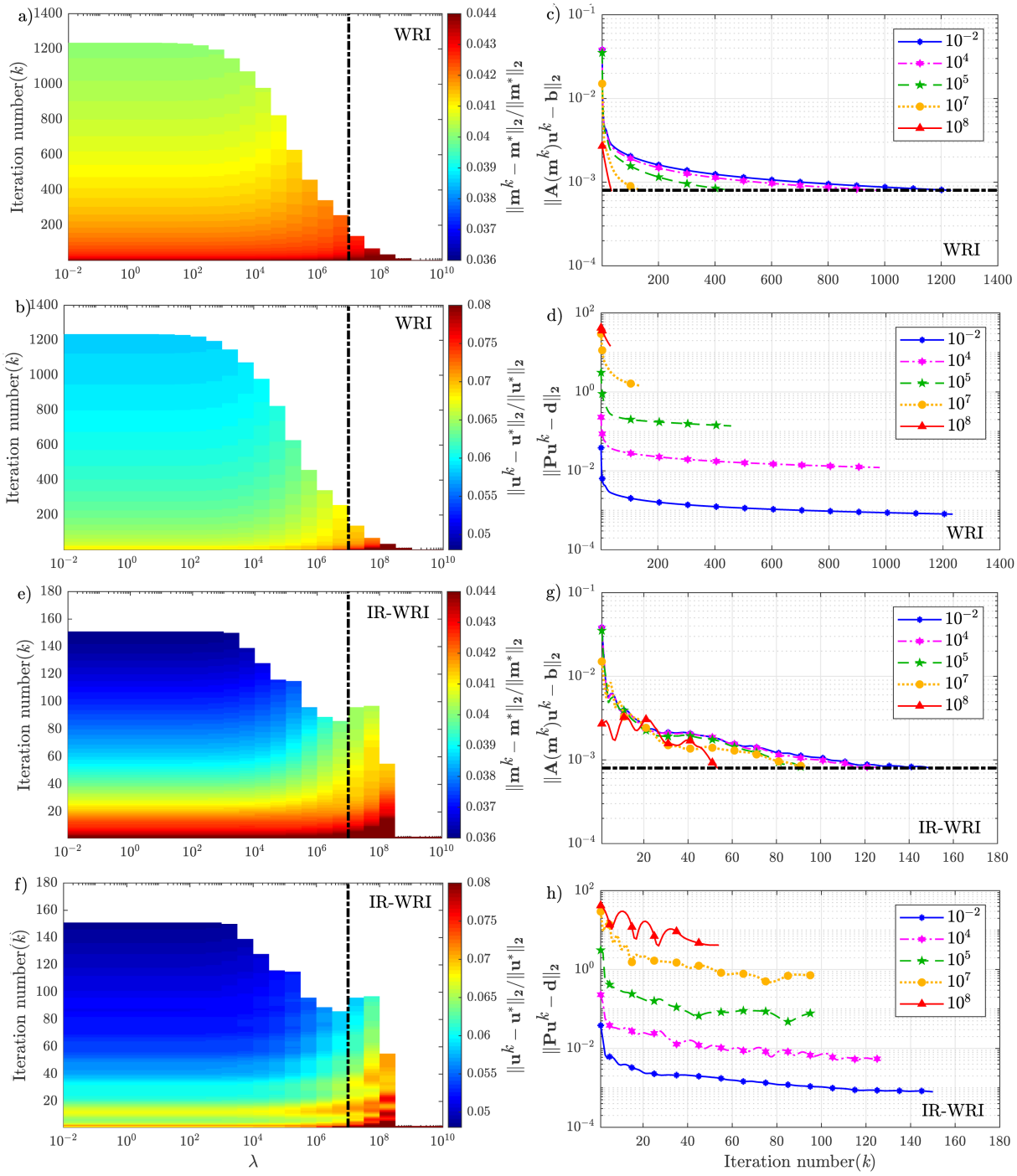

Figure 2.21 – Box-shape anomaly example. Sensitivity of WRI and IR-WRI to λ. (a-d) WRI results. (a-b) Convergence history of (a) model error function $\|\mathbf{m}^k-\mathbf{m}^*\|_2/\|\mathbf{m}^*\|_2$ and (b) wavefield error function $\|\mathbf{u}^k-\mathbf{u}^*\|_2/\|\mathbf{u}^*\|_2$ in the (λ-k) plane. (c-d) Convergence history of (c) wave-equation misfit function $\|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k-\mathbf{b}\|_2$ and (d) data misfit function $\|\mathbf{P}\mathbf{u}^k-\mathbf{d}\|_2$ as function of k for several λ. (e-h) Same as (a-d) for IR-WRI. The vertical dashed line in (a), (b), (e) and (f) points the value of the highest eigenvalue of matrix $\mathbf{A}^{-T}\mathbf{P}^T\mathbf{P}\mathbf{A}^{-1}$ (namely, 1e7) and the horizontal dashed line in (c) and (g) shows the preset stopping criteria based upon the wave-equation misfit function $\|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k-\mathbf{b}\|_2$.

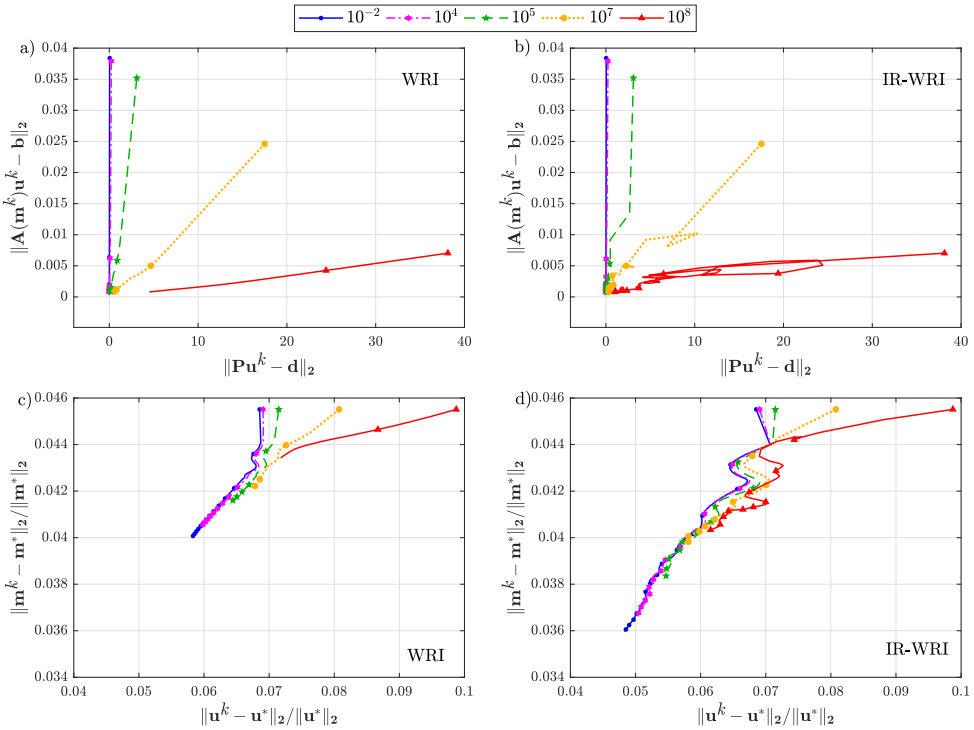

Figure 2.22 – Box-shape anomaly example. Convergence history of WRI (a,c) and IR-WRI (b,d) (see also Figure 2.21). (a-b) Convergence path in the $(\|\mathbf{P}\mathbf{u}^k - \mathbf{d}\|_2 - \|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b}\|_2)$ plane for several values of λ. (c-d) Convergence path in the $(\|\mathbf{u}^k - \mathbf{u}^*\|_2/\|\mathbf{u}^*\|_2 - \|\mathbf{m}^k - \mathbf{m}^*\|_2/\|\mathbf{m}^*\|_2)$ plane for several values of λ.

limited-memory Broyden-Fletcher-Goldfarb-Shanno (l-BFGS) quasi-Newton optimization and a line search procedure for step length estimation that satisfy the Wolfe conditions. The final velocity models obtained with WRI and IR-WRI are shown in Figure 2.25d,e. For WRI and IR-WRI, we use a constant penalty parameter λ = 1e-2 μ_1 during iterations. WRI and IR-WRI performed 615 and 533 iterations, respectively. A direct comparison between the true model, the initial model and the final WRI and IR-WRI models along three vertical logs at horizontal distances of 4.5 km, 7 km and 9.5 km are shown in Figure 2.27a. These results show that IR-WRI successfully converges toward a satisfying model. WRI leads to an acceptable velocity model in the shallow part down to roughly 2.5 km depth, with however a much poorer resolution and less accurate positioning of sharp reflectors at depth than the IR-WRI model.

To emphasize the key role of wavefield reconstruction to make the inversion resilient to cycle skipping, we compare, for the first shot (x=0km), the recorded and the modeled data computed by FWI and WRI/IR-WRI at the first iteration (before model update) and for the 3-Hz frequency (Figure 2.28a). Note that the wavefields reconstructed by WRI and IR-WRI are identical at the first iteration. We show that WRI/IR-WRI allow for a quite good data fit from the first iteration when a small value of λ is used, unlike FWI. After the first wavefield reconstruction, WRI/IR-WRI updates the initial velocity model by least-squares minimization of the source residuals

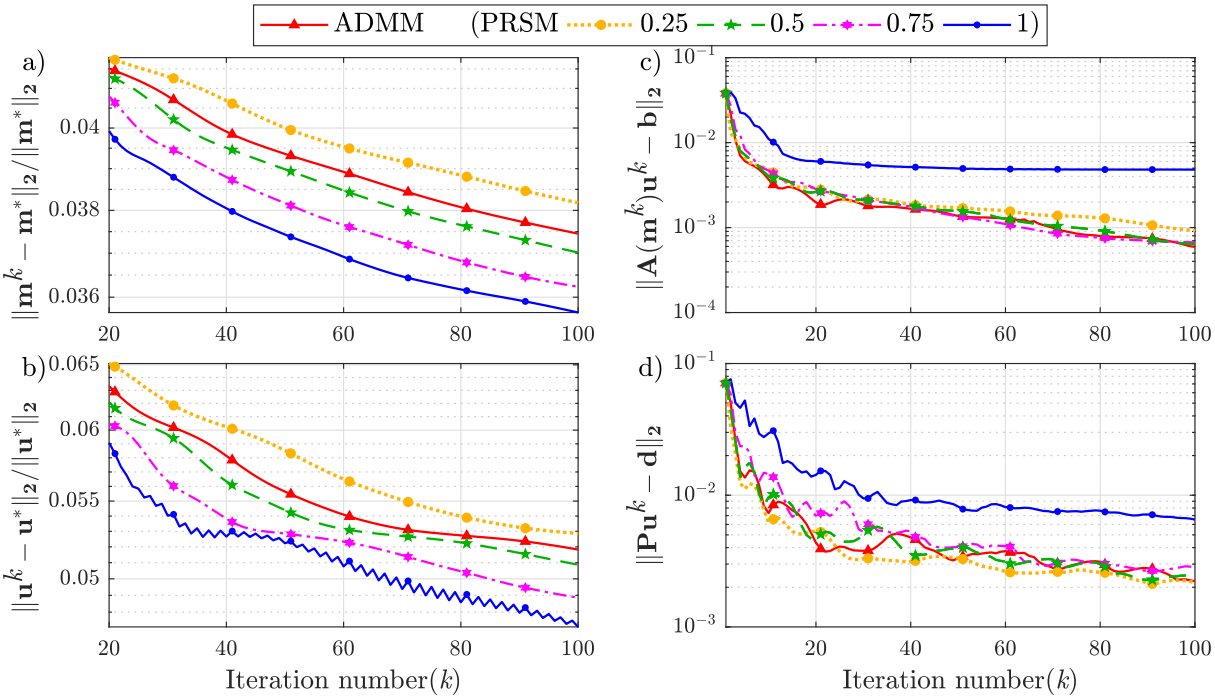

Figure 2.23 – Box-shape anomaly example. Comparison between ADMM and PRSM and sensitivity of PRSM to the step length α applied to the $\mathbf{b}$ updates. (a-d) Convergence history of (a) the model error function $\|\mathbf{m}^k - \mathbf{m}^*\|_2/\|\mathbf{m}^*\|_2$, (b) the wavefield error function $\|\mathbf{u}^k - \mathbf{u}^*\|_2/\|\mathbf{u}^*\|_2$, (c) the wave-equation misfit function $\|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b}\|_2$, and (d) the data misfit function $\|\mathbf{P}\mathbf{u}^k - \mathbf{d}\|_2$.

to push back the reconstructed wavefields toward the wave-equation constraint. The source residual for the first shot after the first wavefield reconstruction is shown in Figure 2.28i. The wavefield scattered by this source residual using the initial model as background is shown in Figure 2.28c. To gain some qualitative insight on the extent to which the model updating will be driven toward the true model perturbation, it is therefore instructive to check how much this scattered reconstructed wavefield is close to the true scattered wavefield (the difference between the wavefields computed in the true and initial models) (Figure 2.28b). Comparison between Figure 2.28b and 2.28c shows that the true and reconstructed wavefields are in phase in most parts of the subsurface, the differences being more related to amplitude mismatches. This phase agreement between the true and reconstructed wavefields might explain why IR-WRI converges toward an accurate velocity model in this case. Indeed, the wavefield reconstruction procedure fails in some parts of the subsurface, which are poorly illuminated by the surface acquisition or because the wavefields that have propagated through these parts of the subsurface are recorded at the receivers with weak amplitudes (for example, deep short-spread reflections). However, these wavefield misfit in the deep part progressively vanish with iteration according to a depth continuation updating.

To emphasize the role of the iterative right hand sides updating on IR-WRI, we now compare the WRI and IR-WRI scattered wavefields with the true scattered wavefield (Figure 2.28b,d,e) as

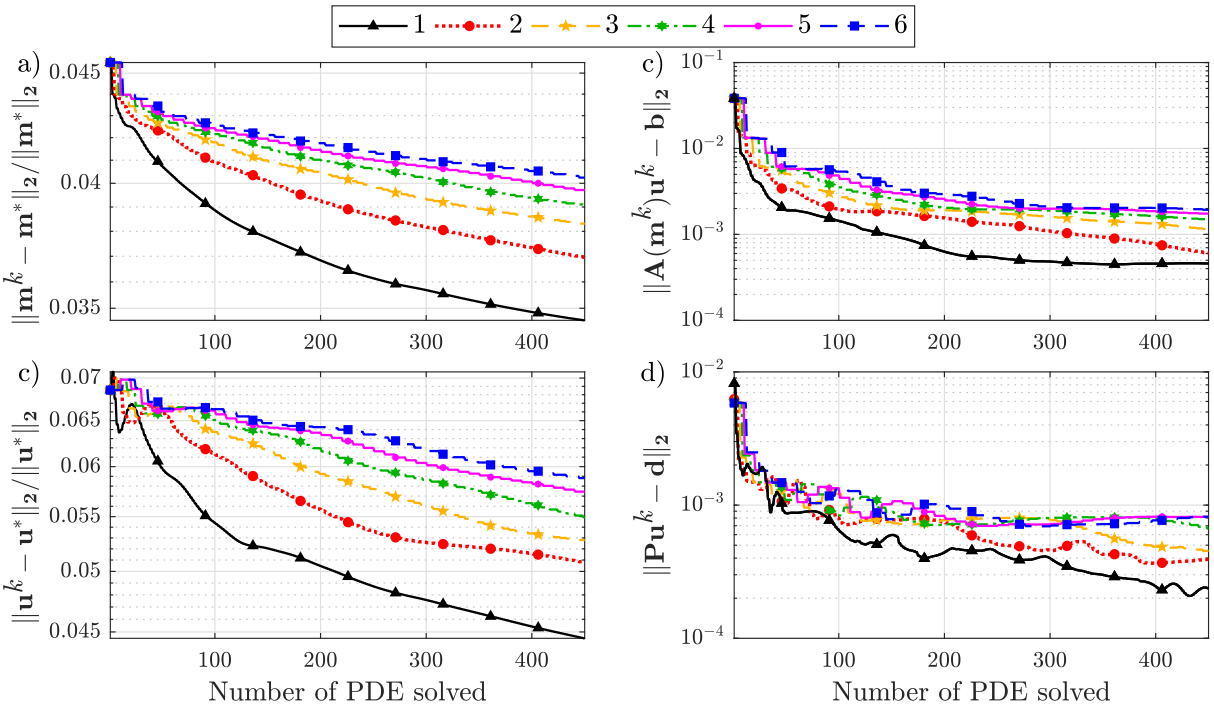

Figure 2.24 – Box-shape anomaly example. Sensitivity of the IR-WRI convergence to the number n of inner iterations (see text for details). (a-b) Convergence history of (a) the model error function $\|\mathbf{m}^k - \mathbf{m}^*\|_2/\|\mathbf{m}^*\|_2$, (b) the wavefield error function $\|\mathbf{u}^k - \mathbf{u}^*\|_2/\|\mathbf{u}^*\|_2$, (c) the wave-equation misfit function $\|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b}\|_2$, and (d) the data misfit function $\|\mathbf{P}\mathbf{u}^k - \mathbf{d}\|_2$. The horizontal axis is labeled with the number of PDE solve. In the figure, n ranges from 1 to 6 and the same n is used during the update of $\mathbf{u}$ and $\mathbf{m}$.

well as the WRI and IR-WRI data and source residuals (Figure 2.28g,h,j,k) at the final iteration. First, IR-WRI manages to cancel out the data and source residuals at the final iteration, unlike WRI (Figure 2.28g,h,j,k). This highlights how the augmented Lagrangian method embedded in IR-WRI manages to satisfy the data fitting and wave equation constraints at the convergence point with a fixed penalty parameter thanks to the Lagrange multiplier updating, unlike WRI. Moreover, although the WRI scattered wavefield is in phase with the true scattered wavefield in most part of the subsurface, the IR-WRI scattered wavefield matches better subtle wavefield variations and amplitudes in particular in the deep part of the model. The imprint of this final improved wavefield reconstruction achieved by IR-WRI relative to WRI on the reconstructed velocity models can be assessed by comparing Figure 2.25d and 2.25e.

As a final quality control of the inversion results, we compute synthetic seismograms in the final WRI and IR-WRI models to assess how the errors in the velocity reconstruction translate into phase and amplitude data fit. We first superimpose in transparency the seismograms computed in the WRI and IR-WRI models on those computed in the true model to assess whether WRI and IR-WRI were affected by cycle skipping (Figure 2.29a,b). This direct comparison confirms that none of the inversion was significantly hampered by cycle skipping, although we show small traveltime mismatches for deep reflections in the case of WRI (Figure 2.29a,b, black

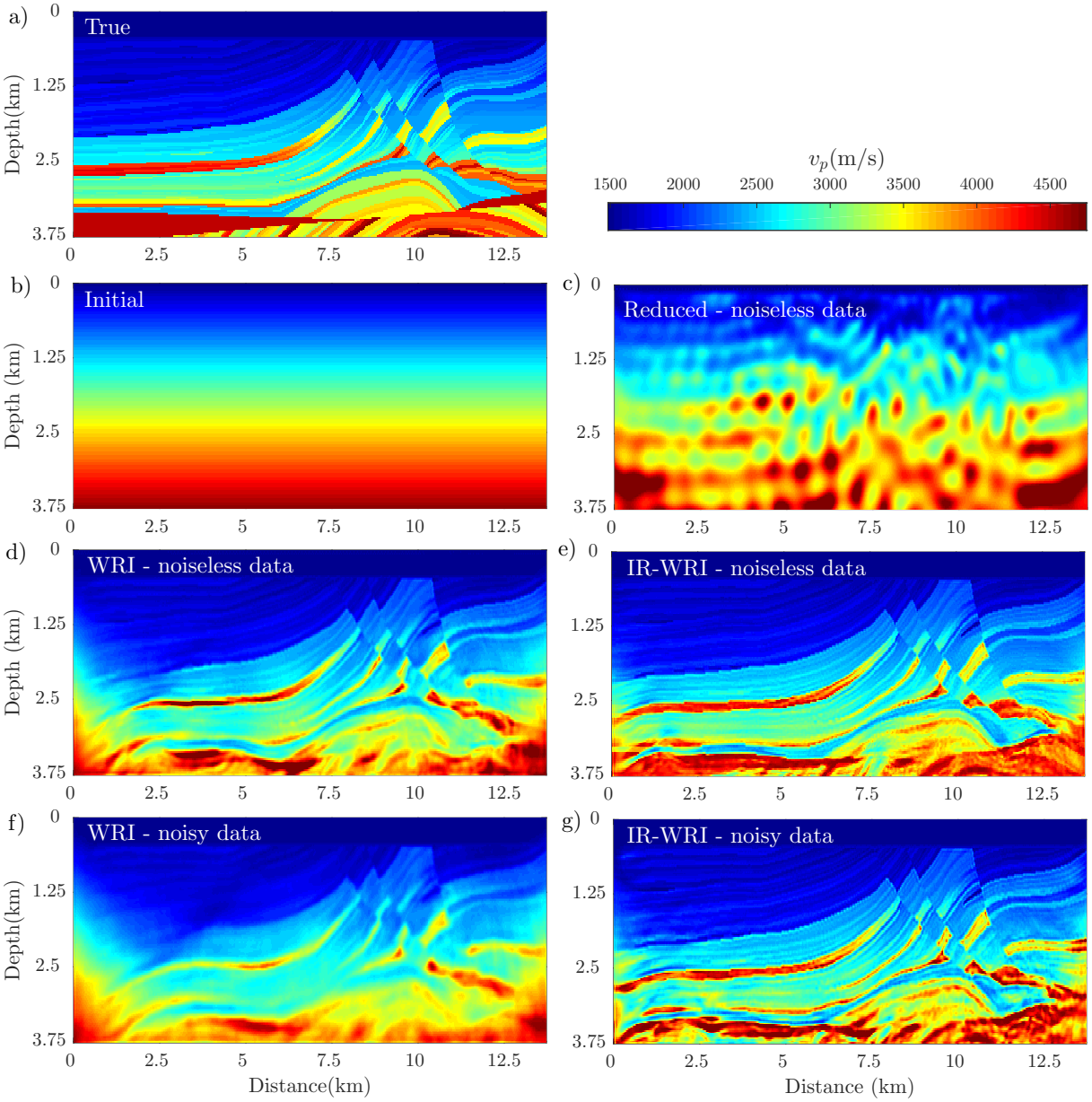

Figure 2.25 – Marmousi II example. (a) True Marmousi II model. (b) Initial velocity model. (c) Reduced approach. (d-e) For noiseless data, velocity models built by (d) WRI and (e) IR-WRI. (f-g) Same as (d-e) for noisy data with a SNR of 10 db.

arrow). These traveltime mismatches can be related to the poor reconstruction of the deep reflectors above mentioned. Second, we show the residuals between the seismograms computed in the true model and the final WRI/IR-WRI models (Figure 2.30a,b). These residuals clearly show that the iterative data and source residual updating allows for a better amplitude fit of both diving waves and reflections. This improved amplitude fit reflects the improved reconstruction of small-scale structures and velocity contrasts in the subsurface models when this updating is performed (Figures 2.25d,e and 2.27a).

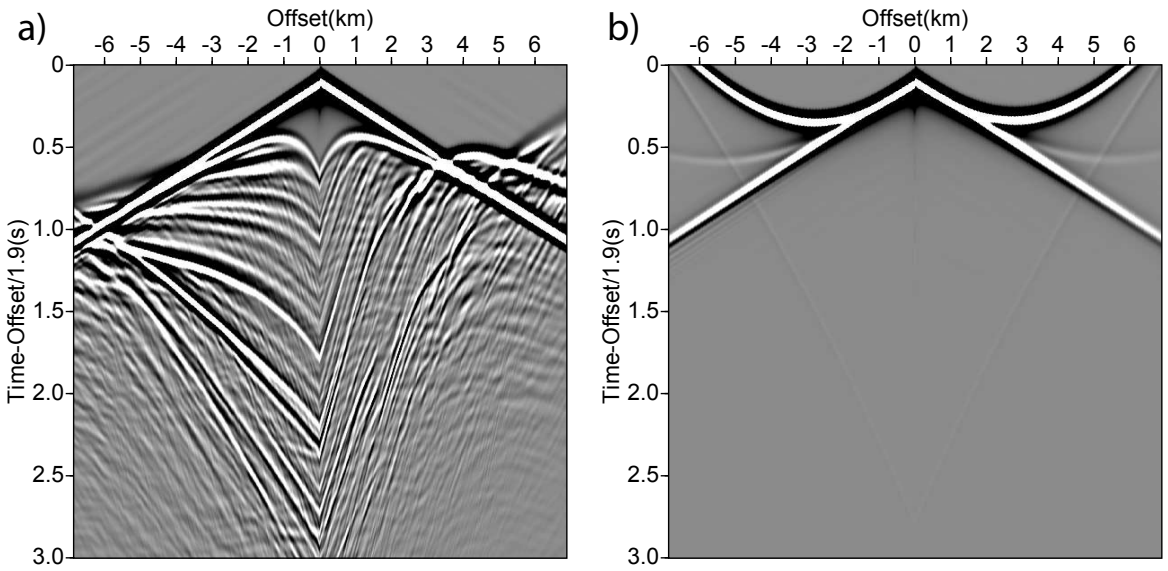

Figure 2.26 – Marmousi II example. Synthetic common-shot gather computed in the true velocity model (a) and the initial velocity model (b). The seismograms are plotted with a reduction velocity of 1900 m/s (origin time of each seismograms is $|\text{offset}|/1900$ s) to compress the time scale.

To reproduce more realistic conditions, we add random noise to the recorded data with a signal-to-noise ratio (SNR) of 10 db. We added the noise in the frequency domain which means that each frequency involved in the inversion has the same SNR. We used the same experimental setup as above performing three inversion paths through the frequency batches. We used δ=1e-3 and we set ϵ_n equal to the noise level of each frequency. The final WRI and IR-WRI models are obtained after 461 and 396 iterations, respectively, and are shown in Figure 2.25f,g). A direct comparison between the true model, the initial model and the final WRI/IR-WRI models along the vertical logs are shown in Figure 2.27b. While the noise weekly impacts upon the IR-WRI results (compare Figures 2.25e and 2.25g and Figures 2.27a and 2.27b, red curves), the resolution of the model inferred from WRI is significantly degraded relative to that inferred from noiseless data, leading to a quite smooth reconstruction (compare Figure 2.25d and 2.25f and Figures 2.27a, bleue curves and 2.27b, blue curves). This highlights the higher resilience to noise of IR-WRI.

The data fit assessment for noisy data is shown in Figures 2.29c,d and 2.30c,d). The direct comparison between the seismograms computed in the true and WRI/IR-WRI models still reveal a good traveltime match without obvious evidence of cycle skipping (Figures 2.29c,d). However, as for the noiseless case, we notice small traveltime mismatches in the case of WRI due to the poor focusing and mispositioning of deep reflectors shown in Figure 2.27b, red curves. More significantly, this direct comparison shows that the WRI fails to match the amplitudes of several deep short-spread reflections due to poorly resolved model reconstruction (Figures 2.29c, white arrows), unlike IR-WRI (Figure 2.29d). This is further confirmed by the

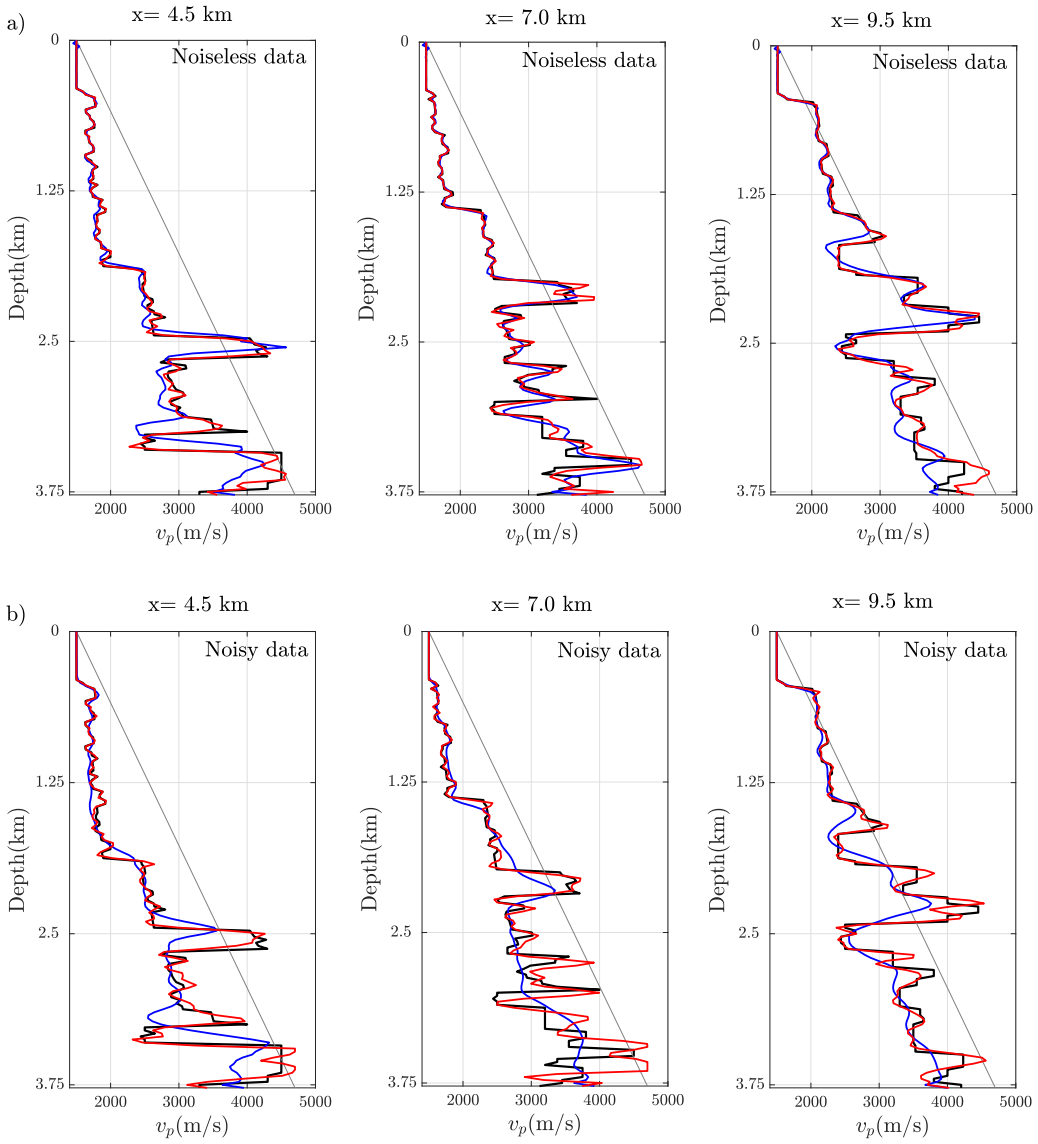

Figure 2.27 – Marmousi II example. Direct comparison between true (black), initial (gray), WRI (blue) and IR-WRI (red) velocity models along three logs at X=4.5km, 7km, 9.5km from left to right. (a) Noiseless data case (Figure 2.25d,e). (b) Noisy data case (Figure 2.25f,g). Note the deficit of resolution and mispositioning of deep reflectors in WRI results. These defaults increase significantly with noise, while IR-WRI reconstructions are more resilient to noise.

data residuals shown in Figure 2.30c,d. For both inversion, the amplitudes of the residuals as well as their low frequency content (due to the increase smoothness of the velocity reconstruction) have increased compared to those inferred from noiseless data. However, this increase is much more moderate in the case of IR-WRI.

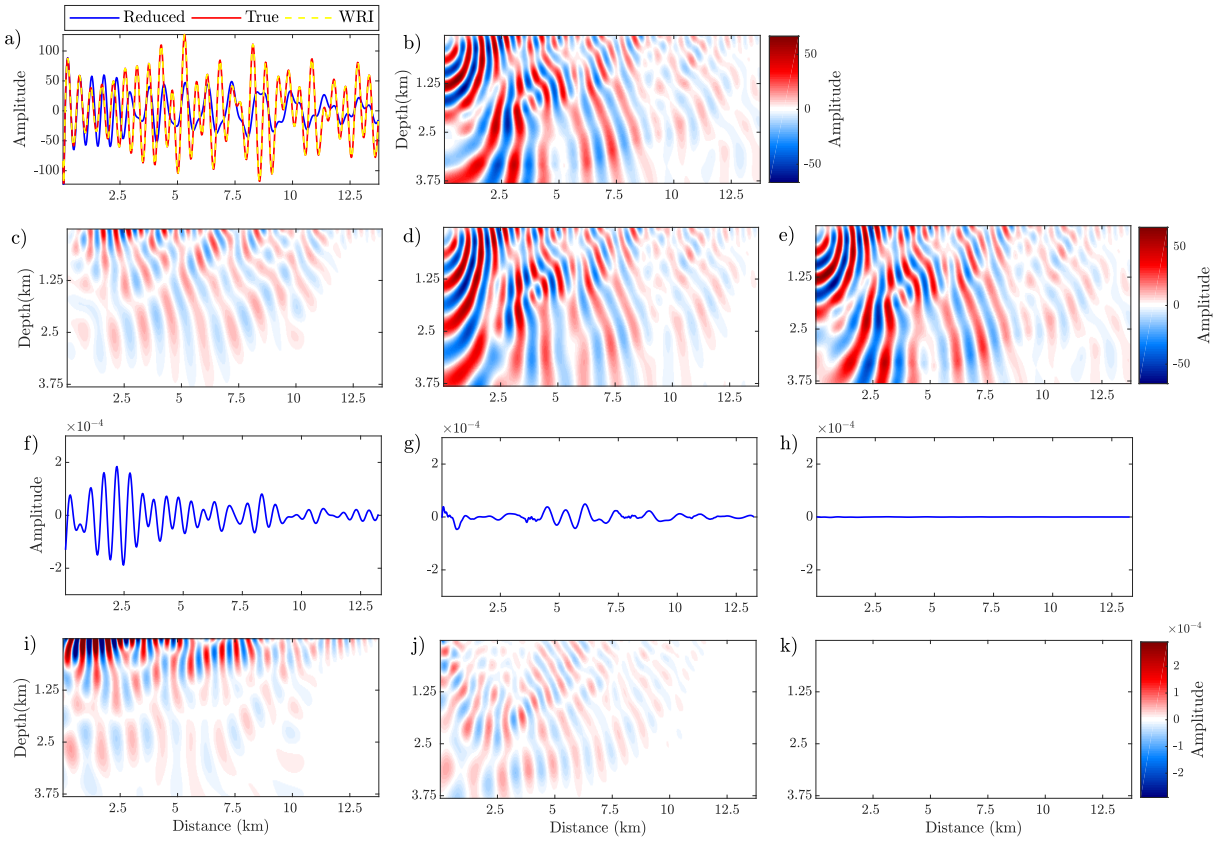

Figure 2.28 – Marmousi II example. For a source at x=0 m and the 3-Hz frequency: (a) Direct comparison between data (real part) computed in the initial (blue) and true (black) models and by WRI wavefield reconstruction at first iteration (orange). Note the almost perfect agreement between true and WRI data. (b) True scattered wavefield (difference between the wavefield computed in the true and initial velocity models). (c) Scattered wavefield reconstructed by WRI at first iteration (difference between the wavefield reconstructed by WRI and the wavefield computed in the initial model). (d-e) Scattered wavefields reconstructed by WRI (d) and IR-WRI (e) at final iteration. (f-h) Data residual ($\mathbf{Pu} - \mathbf{d}$) at first WRI iteration (f), at final WRI iteration (g) and at final IR-WRI iteration (h). (i-k) Wave-equation residual at first iteration of wavefield reconstruction, namely ($\mathbf{A}(\mathbf{m}^0)\mathbf{u}^1 - \mathbf{b}$) (i), final iteration of wavefield reconstruction for WRI (j) and final iteration of wavefield reconstruction for IR-WRI (k) ($\mathbf{A}(\mathbf{m}^{k-1})\mathbf{u}^k - \mathbf{b}$)

2004 BP salt model

Finally, we assess WRI and IR-WRI against a target of the challenging 2004 BP salt model (Figure 2.31a) when a crude stating model is used (Figure 2.31b). The selected target corresponds to the left part of the 2004 BP salt model and was previously used in Métivier et al. (2016a) for an application of FWI based upon an optimal-transport distance. It is representative of the geology of the deep offshore Gulf of Mexico and mainly consists of a simple background with a complex rugose multi-valued salt body, sub-salt slow velocity anomalies related to over-pressure zones and a fast velocity anomaly to the right of the salt body (Billette

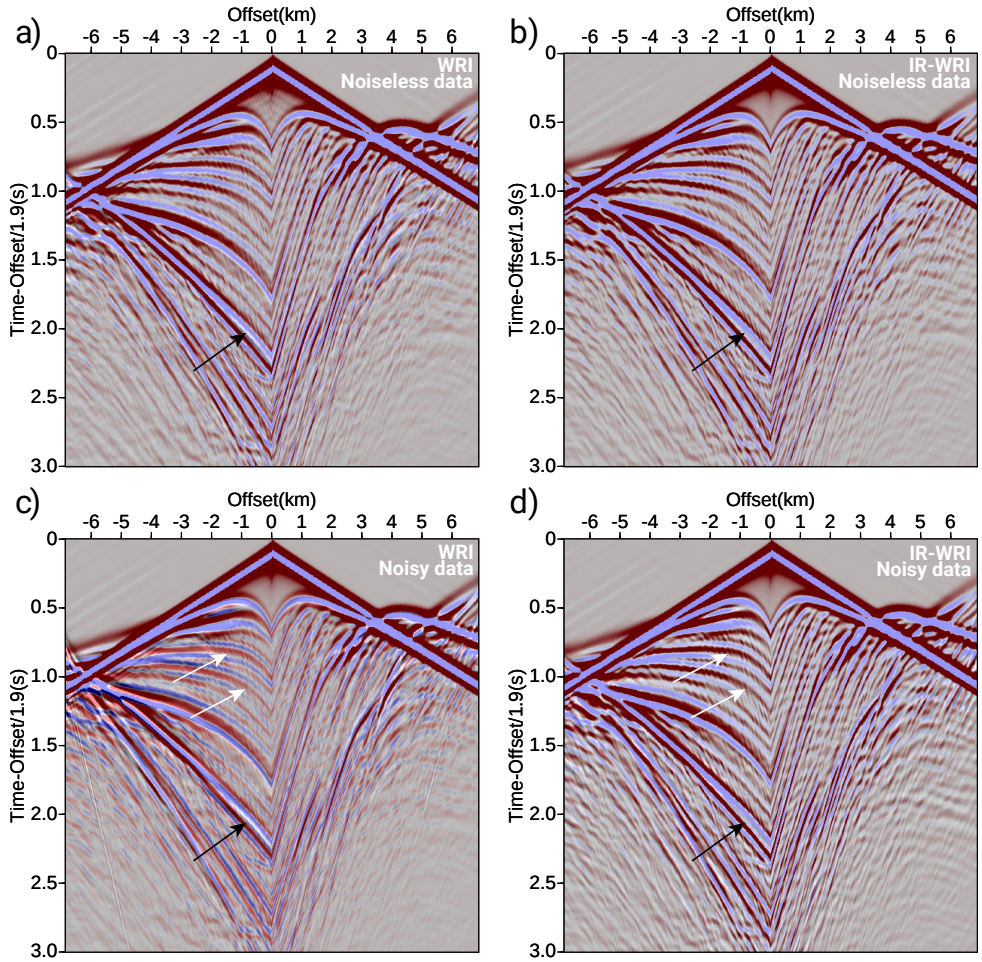

Figure 2.29 – Marmousi II example. Direct comparison between synthetic seismograms computed in true (red/white/blue scale scale) and WRI/IR-WRI (black/gray/white scale) models. Seismograms computed in the WRI/IR-WRI models are superimposed in transparency on seismograms computed in the true model. The two sets of seismograms are in phase if black/white wiggles of seismograms computed in the WRI/IR-WRI models match red/blue wiggles of seismograms computed in the true model. Seismograms are plotted with a reduction velocity of 1900 m/s. (a-b) For noiseless data, black/white seismograms are computed in (a) WRI and (b) IR-WRI models shown in Figure 2.25d,e. (c-d) Same as (a-b) for noisy data (WRI/IR-WRI models are shown in Figures 2.25f,g). Arrows point the most obvious differences in the data fit achieved by WRI and IR-WRI. In the WRI results, data misfit take the form of significant amplitude underestimation (white arrows) and small time mismatches resulting from poorly resolved model reconstruction (black arrows).

and Brandsberg-Dahl, 2004). As for the Marmousi II model, we perform WRI and IR-WRI for noiseless and noisy data. The subsurface model is 16250 m wide and 5825 m deep and is discretized with a 25 m grid interval. We used 162 sources spaced 100 m apart on the top side of the model. The source signature is a Ricker wavelet with a 10 Hz dominant frequency. A line of receivers with a 25 m spacing are deployed at the surface leading to a stationary-receiver

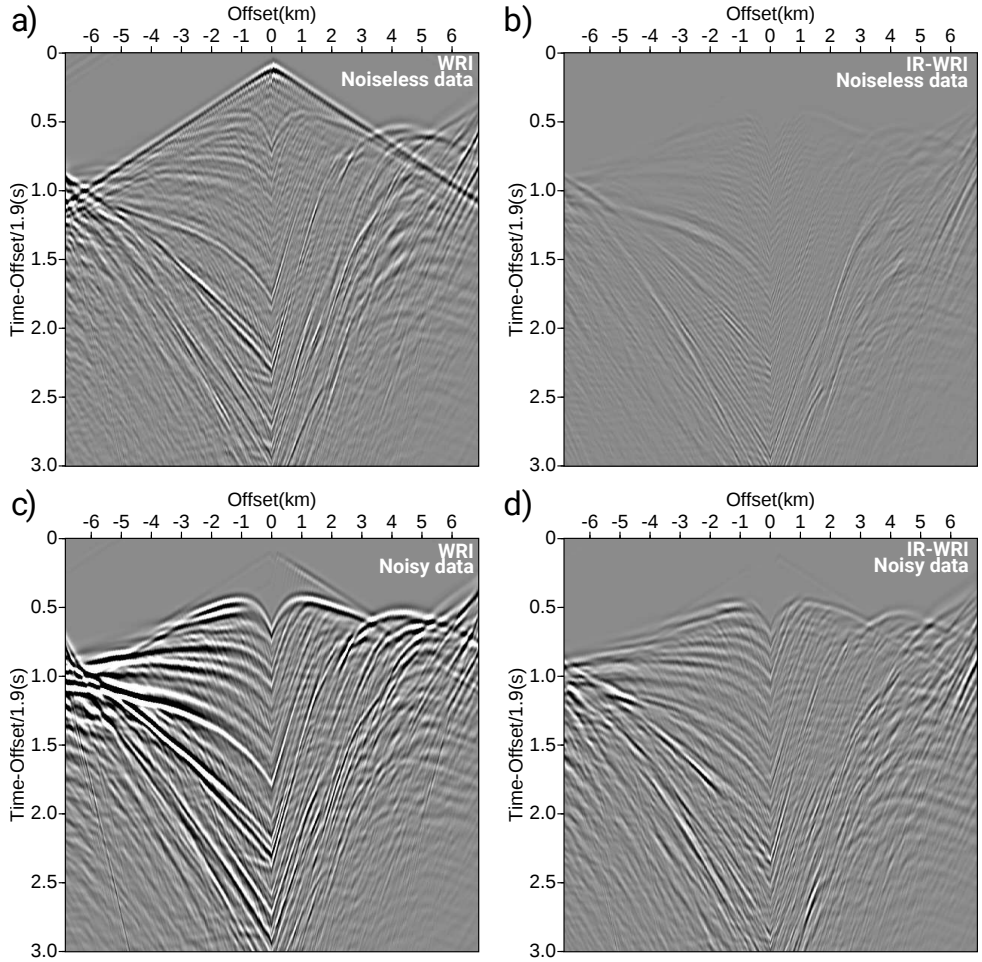

Figure 2.30 – Marmousi II example. Residuals between seismograms computed in the true and WRI/IR-WRI models. Each panel is plotted with the amplitude scale used in Figure 2.26. (a-b) For noiseless data, residuals are computed from (a) WRI and (b) IR-WRI models. See Figure 2.29a,b for a direct comparison of the two sets of seismograms. (c-d) Same as (a-b) for noisy data. See Figure 2.29c,d for a direct comparison of the two sets of seismograms.

acquisition.

We used a smoothed version of the true velocity model as initial model (Figure 2.31b). Following Métivier et al. (2016a), we design the correlation length of the smoother such that the imprint of the salt body is canceled out.

Comparison between a common shot gather computed in the true and initial models show significant traveltime mismatches (Figure 2.32) that drive reduced-space FWI to a local minimum when the starting frequency is 3Hz (Figure 2.31c).

To perform WRI and IR-WRI, we used small batches of three frequencies with one frequency overlap between two consecutive batches, moving from the low frequencies to the higher ones according to a classical frequency continuation strategy. The starting and final frequencies are 3 Hz and 15 Hz and the sampling interval in one batch is 0.5 Hz. For each batch, the stop-

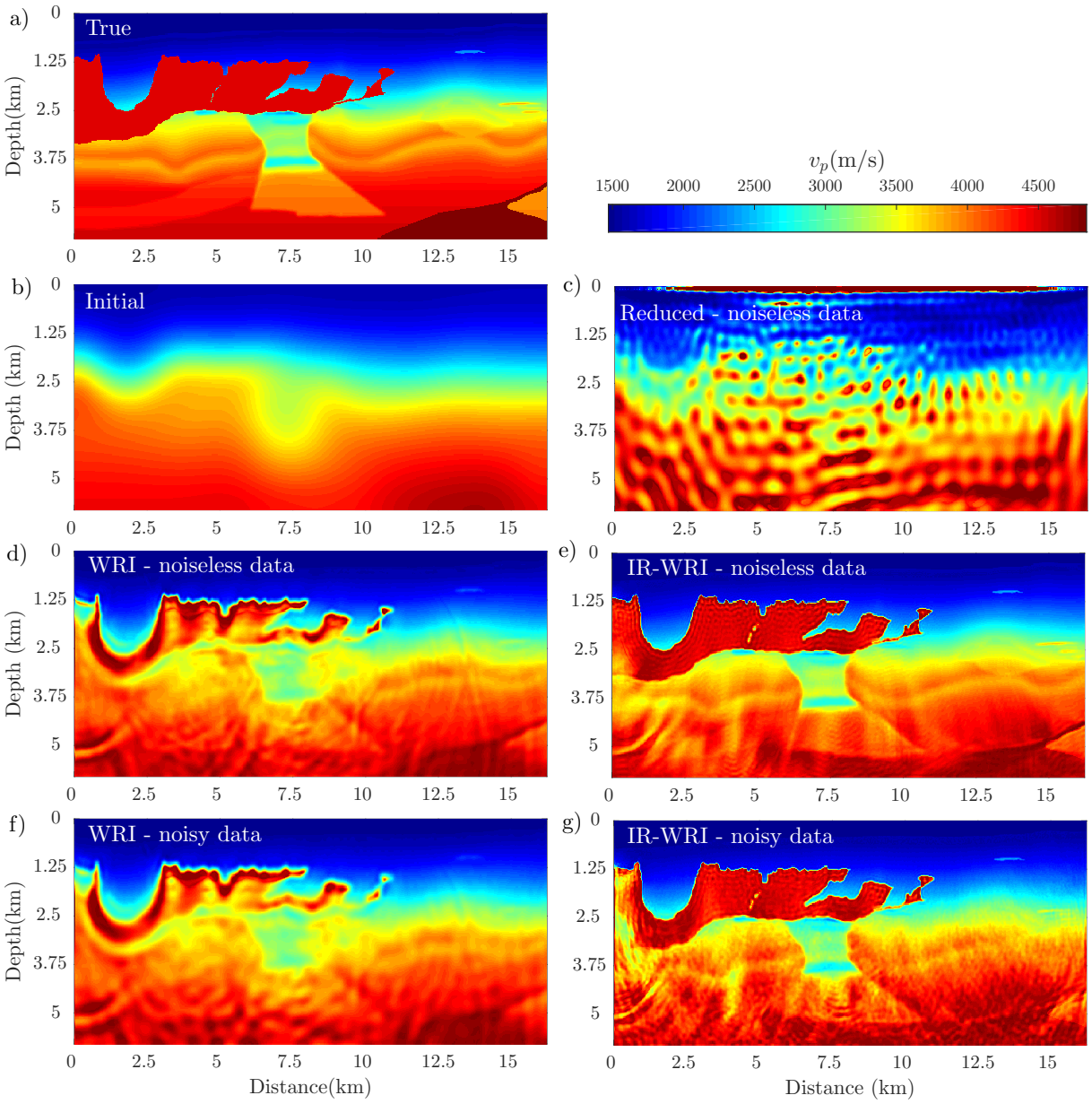

Figure 2.31 – 2004 BP salt case study. (a) True 2004 BP model. (b) Initial velocity model. (c) Reduced approach. (d-e) Final WRI (d) and IR-WRI (e) velocity models for noiseless data. (f-g) Same as (d-e) for noisy data for a SNR of 10 db.

ping criterion of iteration is given by equation 2.37 with k_{max}=20. We use δ=1e-3 for noiseless and noisy data, while ϵ_n is set to 1e-5 and the noise level of the batch for noiseless and noisy data, respectively. For the penalty parameter, we use λ=1e-3 μ_1 and λ=2e-2 μ_1 for noiseless and noisy data, respectively. We perform three paths through the frequency batches to improve the inversion results (the starting frequency of the second and third path is 5 Hz and 7.5 Hz, respectively). The WRI and IR-WRI models inferred from noiseless data are shown in Figures 2.31d,e. WRI and IR-WRI perform 561 and 448 iterations, respectively. Direct comparison

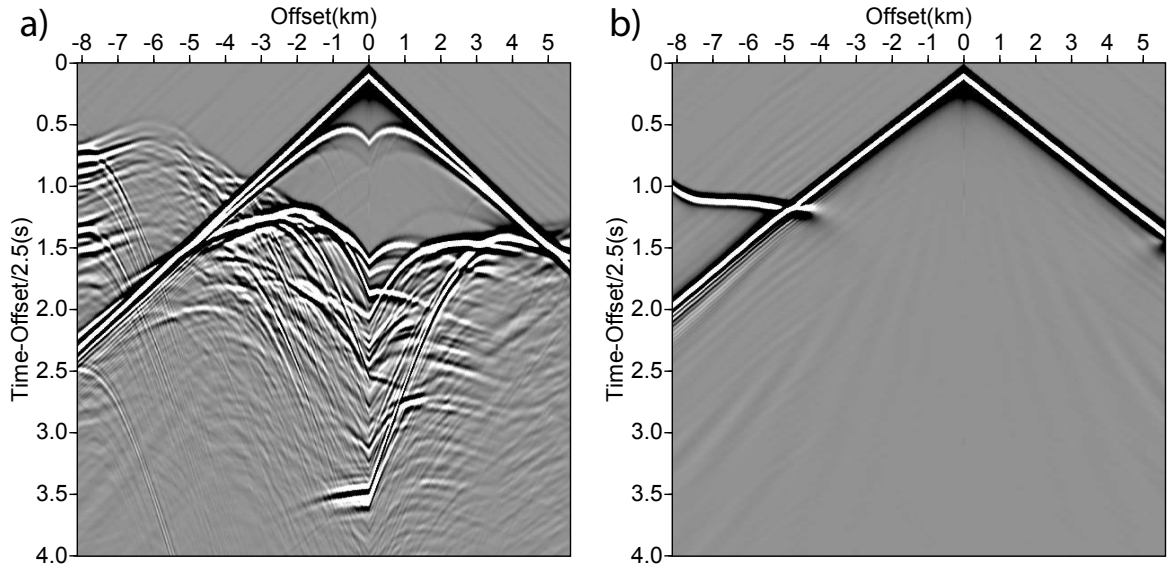

Figure 2.32 – 2004 BP salt case study. Synthetic common-shot gather computed in the (a) true and (b) initial velocity models. Seismograms are plotted with a reduction velocity of 2500 m/s (the origin time of each seismograms is $|\text{offset}|/2500$ s) to compress the time scale.

between the true model, the starting model and the two WRI models along three vertical logs cross-cutting the salt body at 5 km, 7.5 km and 10 km distance is shown in Figure 2.33a. We show that IR-WRI recovers with a quite high resolution the precise geometry of the rugose salt body, the sub-salt low-velocity anomalies as well as the shallow fast anomaly to the right (Figure 2.31e). In contrast, WRI successfully recovers the upper part of the salt but fails to retrieve the bottom part of the salt body. Moreover, the sub-salt structures are not recovered and the image of the fast anomaly to the right is blurred (Figure 2.31f).

We compare the 3-Hz recorded and modeled data computed by FWI and WRI/IR-WRI at the first iteration in Figure 2.34a. Again, WRI/IR-WRI allow for a good data fit from the first iteration, while the smooth initial model generates strong amplitude and phase mismatches between the recorded and the FWI modeled data. IR-WRI has canceled out more efficiently the data and source residuals at the convergence point than WRI (Figure 2.34g,h,j,k. The scattered wavefield reconstructed by IR-WRI is also more accurate than the one reconstructed by WRI at the convergence point even if the differences look like quite subtle in terms of phase and amplitude (Figure 2.34d,e). These subtle differences in the wavefield reconstruction performed by WRI and IR-WRI translate however in quite different velocity reconstruction as shown in Figure 2.31d,e.
As for the Marmousi example, the direct comparison between seismograms computed in the true and WRI/IR-WRI models does not show any evidence of cycle skipping (Figure 2.35a,b). However, the residuals have much lower amplitudes in case of IR-WRI (Figure 2.36a,b) due to the improved reconstruction of the velocity contrasts and high wavenumbers highlighted in Figures 2.31e and 2.33a. When noisy data are used (Figures 2.31f,g and 2.33b), the shallow

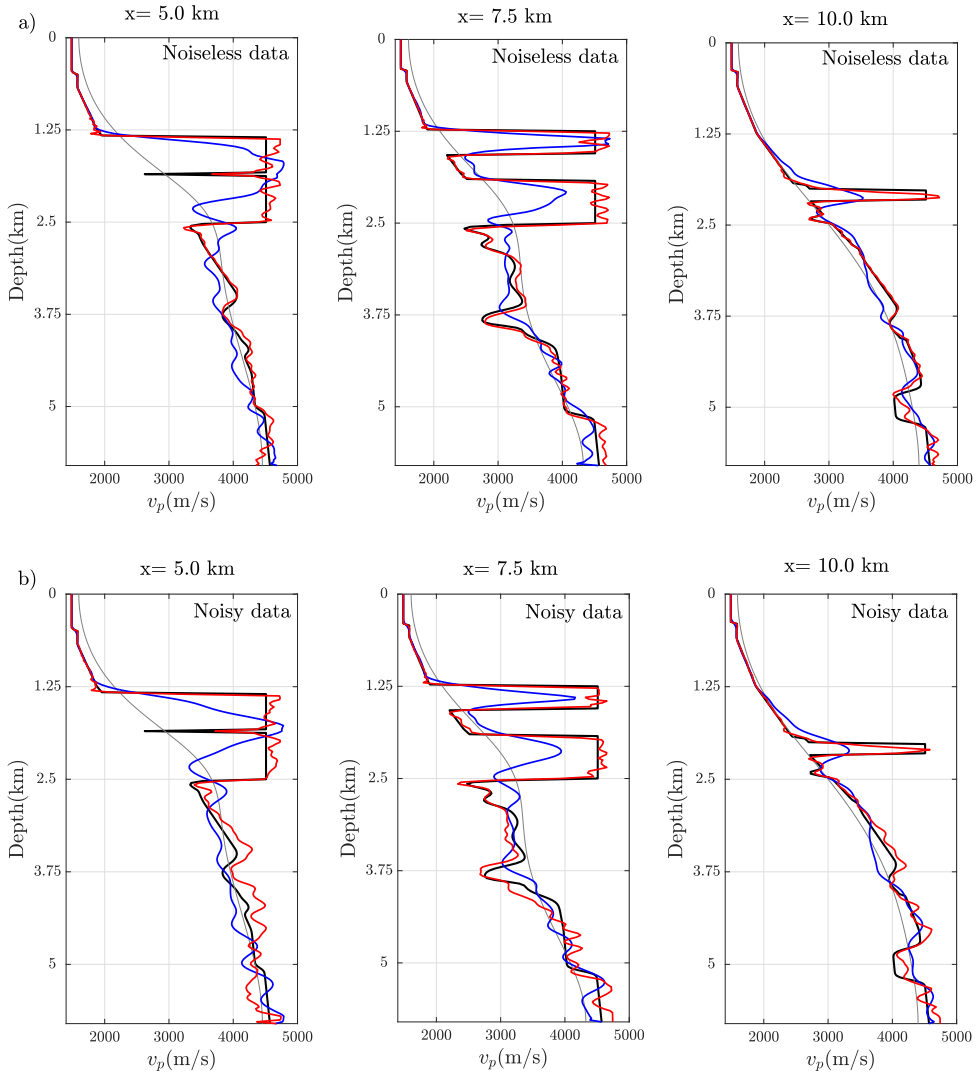

Figure 2.33 – 2004 BP salt case study. Direct comparison between true (black), initial (gray), WRI (blue) and IR-WRI (red) velocity models along three logs at X=5km, 7.5km, 10km from left to right. (a) Noiseless data. (b) Noisy data.

imaging performed by the IR-WRI is weakly impacted upon by the noise. WRI and IR-WRI perform 461 and 396 iterations, respectively, and confirms again that the iterative residual updating not only improves the quality of the minimizer but fastens also the convergence. The sub-salt imaging is more affected by noise. However, the slow velocity anomalies below the salt body is still fairly well identified in the IR-WRI model. In the WRI case, we show the same trend as the one revealed by the Marmousi example. The noise further degrades the results by hampering the reconstruction of short-scale features below the top of the salt but also in the overburden as revealed by the poor reconstruction of the fast anomaly on the right. The comparison of the seismograms computed in the true model and the WRI/IR-WRI models and the data residuals are shown in Figures 2.35c,d and 2.36c,d. A similar trend than for the Marmousi

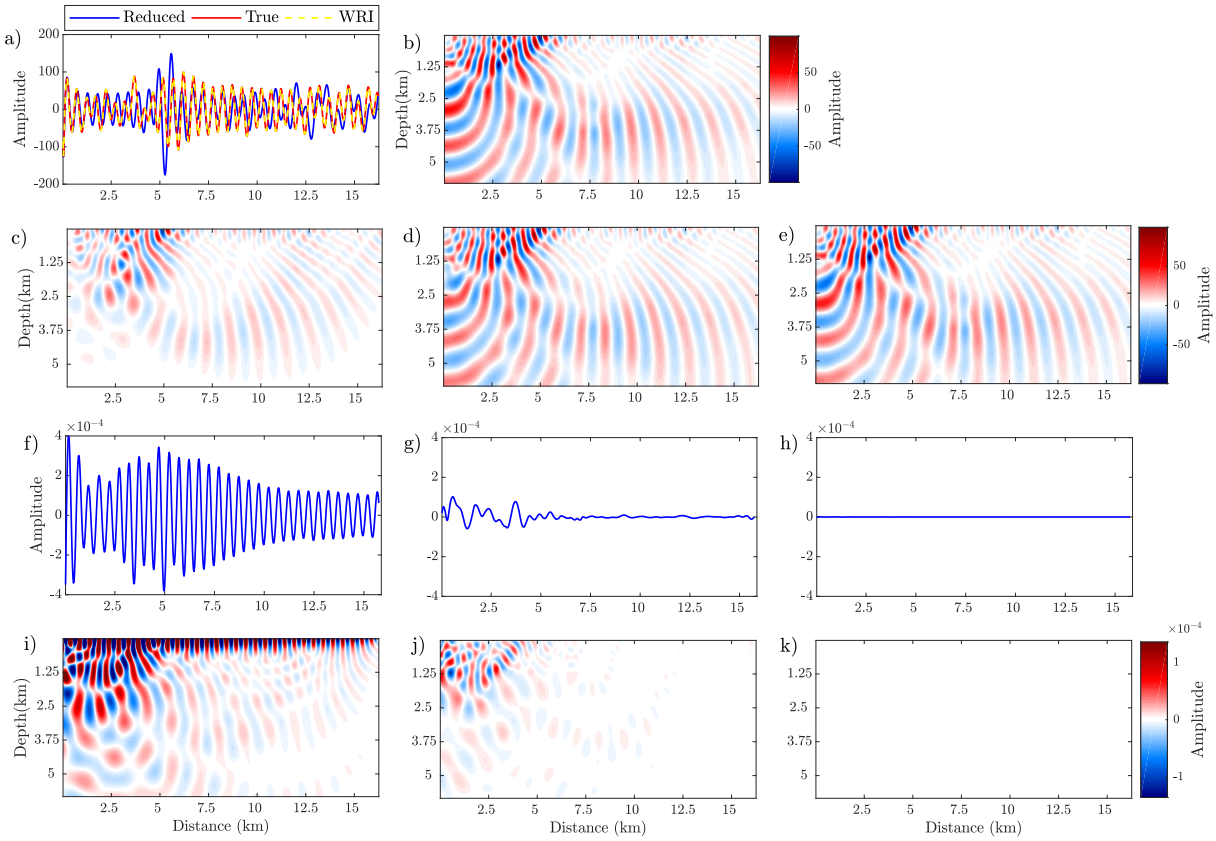

Figure 2.34 – Same as Figure 2.28 for the 2004 BP salt case study.

model is shown: When noise is added to the data, the amplitudes of the data residuals increase in much larger proportion in the WRI case.

2.3.5 Discussion

We have presented a significant yet easy-to-implement improvement of the WRI method of van Leeuwen and Herrmann (2013). In Appendix A (section 2.3.7), we have first recast the original FWI problem as a feasibility problem where both data and source misfit are processed as constraints. This feasibility problem is equivalent to constrained optimization problem where the objective function is identically zero. We solve this constrained problem with an augmented Lagrangian method (method of multiplier). A scaled form of the augmented Lagrangian function shows that it is equivalent to a quadratic penalty function formed by the squares of the constraint violations (data and source misfit), to which are added the running sum of the data and source residuals of previous iterations. In the objectives of the penalty function, the data and the source represent the right-hand side of the constraints, and the running sum of the data and source residuals correspond to the scaled Lagrange multipliers. Following van Leeuwen and Herrmann (2013), we perform the refinement of the two primal variables (the wavefield and

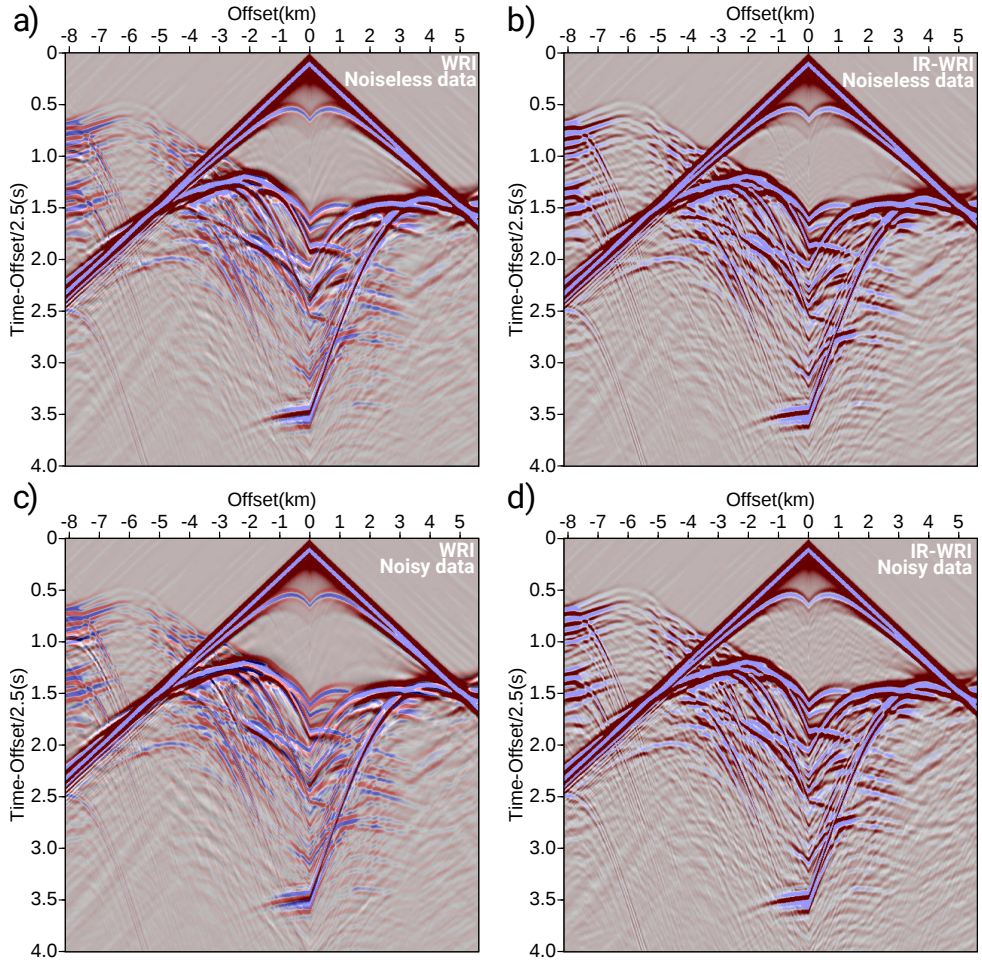

Figure 2.35 – 2004 BP salt case study. Direct comparison between synthetic seismograms computed in true model (red/white/blue scale) and WRI/IR-WRI models (black/gray/white scale). Same showing as in Figure 2.29 is used to compare the two sets of seismograms. (a-b) For noiseless data, black/white seismograms are computed in (a) WRI and (b) IR-WRI models shown in Figure 2.31d,e. (c-d) Same as (a-b) for noisy data (WRI and IR-WRI models are shown in Figure 2.31f,g). For WRI, data misfit take the form of significant amplitude underestimation and small time mismatch resulting from poorly resolved model reconstruction. Seismograms are plotted with a reduction velocity of 2500 m/s as in Figure 2.32.

the subsurface parameters) in alternating mode, using the solution of one primal subproblem as a passive variable for the next subproblem. Compared to a classical application of ADMM where separable convex subproblems are solved in parallel, this sequential resolution of the two primal subproblems allows us, on the one hand to linearize the parameter-estimation problem around the reconstructed wavefield and, on the other hand to manage the non-separability of the wavefield reconstruction and parameter estimation. This sequential solving of the two primal subproblems also prompts us to update the dual variable associated with the source residuals two times, once after each primal subproblem.

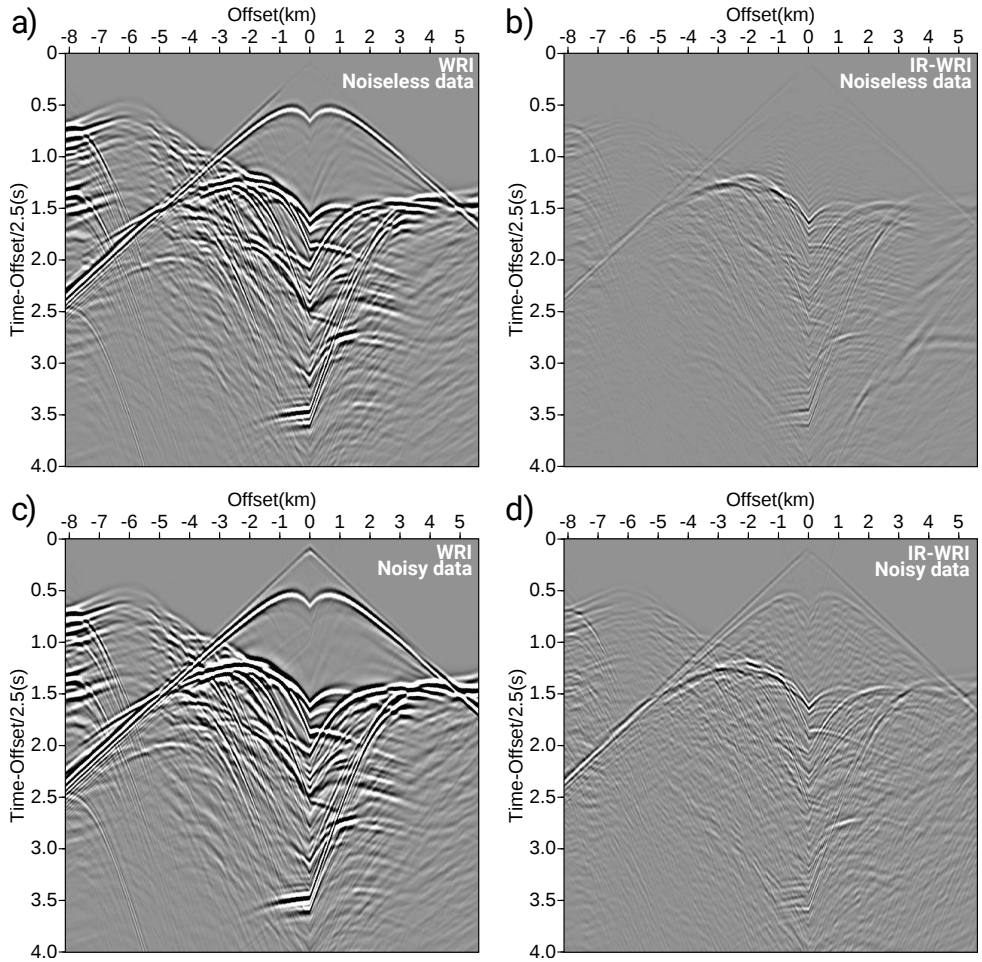

Figure 2.36 – 2004 BP salt case study. Residuals between seismograms computed in true model and WRI/IR-WRI models. Each panel is plotted with the amplitude scale used in Figure 2.32. (a-b) For noiseless data, residuals are computed from (a) WRI and (b) IR-WRI models. See Figure 2.35a,b for a direct comparison of the two sets of seismograms. (c-d) Same as (a-b) for noisy data. See Figure 2.35c,d for a direct comparison of the two sets of seismograms.

The convergence properties and the error bounds on the solution of penalty and augmented Lagrangian methods are discussed in the framework of nonlinear programming by (Nocedal and Wright, 2006, Chapter 17). As mentioned in the above reference (page 519), the augmented Lagrangian method gives us two ways of improving the accuracy of the minimizer: improve the accuracy of the Lagrange multiplier (namely, decrease the data and source residuals enough) or increase the penalty parameter, whereas the quadratic penalty approach gives us only one option: increase the penalty parameter. Accordingly and considering that we perform all of the numerical experiments with a fixed penalty parameter, the augmented Lagrangian approach implemented in IR-WRI has logically converged to more accurate minimizers with a smaller number of iterations than the penalty method implemented in WRI. The two leverages that control the accuracy of the IR-WRI minimizer also allow us to use moderate values of the penalty parameter without impacting prohibitively convergence speed. This moderate values

allows the method to fit the data during the early iterations with large wave-equation error, which provides the most suitable framework to enlarge the search space and account for large time shifts, while satisfying the wave-equation constraint with small error at the convergence point. According to the analogy between WRI and IR-WRI highlighted by the scaled-form Lagrangian, IR-WRI can be viewed as a self-adaptive penalty method, where a tedious and potentially unstable continuous increasing of the penalty parameter is replaced by a stable dual steepest-ascent updating of the data and source residuals. This iterative residual updating in a quadratic misfit function is a well known procedure to refine solution of a wide class of linear inverse problems with gradient or Gauss-Newton steps as reminded in Appendix B (section 2.3.8).

The second key ingredient in IR-WRI is operator splitting implemented with an alternating direction strategy. IR-WRI relies on a generalization of the alternating-direction method of multiplier (ADMM) to biconvex problems (Boyd et al., 2010, section 9.2). A biconvex problem is one in which two variables can be partitioned into sets over which the problem is convex when the other variable is fixed (Shen et al., 2017). IR-WRI is bi-convex because the wave-equation constraint is bilinear. This property leads to the linearization of the subproblem for $\mathbf{m}$ around $\mathbf{u}$ in our alternating-direction algorithm. Accordingly, the standard form of the ADMM can be readily implemented in IR-WRI, without however guarantee that IR-WRI benefits from the convergence properties of classical ADMM for convex problems. The interested reader is referred to Brás et al. (2012) for additional discussion on the applicability of ADMM on biconvex problems in the framework of elliptic PDE constraint.

As a variant of WRI based upon the above-mention alternating direction strategy, van Leeuwen and Herrmann (2016) recast WRI as a reduced variable projection penalty method which leaves the parameter estimation subproblem nonlinear. Aravkin et al. (2017) analyze the convergence properties of the reduced penalty method when the full Hessian is taken into account and conclude (their Theorem 2.5) that the variable projection penalty method is insensitive to the penalty parameter. We can view to develop a reduced formulation of our augmented Lagrangian method by enforcing the closed-form expression of the reconstructed wavefield in the subsurface parameter estimation problem. More theoretical and numerical investigations are necessary to assess whether it is worth paying the price to solve a more complex nonlinear subproblem for $\mathbf{m}$ in the variable projection method compared to the ADMM-based IR-WRI which alternates two linear problems.

Although we have shown that the accuracy of the minimizer found by IR-WRI is weakly sensitive to the penalty parameter for a wide range of moderate values, the value which maximizes the convergence speed may be found accordingly to mitigate the computational cost. It is also worth reminding that updating the right-hand sides of the objectives in the IR-WRI penalty function does not introduce significant computational overhead in one iteration of IR-WRI compared to the WRI counterpart.

The reader familiar with the Bregman optimization method for constrained convex problem may have noticed that IR-WRI borrows some ingredients from this approach. The original Bregman method (Bregman, 1967) has been recast by Yin et al. (2008); Goldstein and Osher (2009) as a penalty method where the right-hand side of the constraint is iteratively updated with the errors of previous iterations. Analogy between Bregman method and augmented Lagrangian

method is recognized in Yin et al. (2008), while Esser (2009) reviewed the connection between ADMM and the split Bregman method. The split Bregman method has been developed to solve l1-regularization problems for image denoising applications. The strategies used in the split Bregman method to decouple l1 and l2 objectives via auxiliary variables have been used by Aghamiry et al. (2018b) to interface total-variation regularization with IR-WRI, where the identically-zero objective function in equation 2.39 is replaced by the total variation norm of the subsurface model. A short discussion on the choice of the penalty parameter in the framework of the Bregman optimization method is provided in Goldstein and Osher (2009, section 2.2) and is consistent with the results obtained with IR-WRI.

The original motivation of WRI was to make waveform inversion more resilient to cycle skipping. We apply WRI and IR-WRI on two complex synthetic examples to assess their ability to account for large time shifts. The results show that both methods remain immune to cycle skipping for these case studies. However, only IR-WRI shows accurate reconstruction of deep targets such as the base of salt bodies and subsalt structures due to improved convergence history. This improved convergence manifests in the seismograms by a better amplitude fit of weak phases such as deep short-spread reflections. The improved convergence also makes IR-WRI more resilient to noise. Indeed, IR-WRI still needs to be assessed against more realistic applications. The subsurface parameters are updated by minimizing the source residuals generated by the previously-reconstructed wavefield. The wavefield is reconstructed by fitting the sparse seismic observables through the relaxation of the wave equation constraint. The accuracy with which the wavefield should be reconstructed to make IR-WRI converge toward an accurate velocity model and the relative role of the acquisition geometry and initial velocity model on the accuracy of this wavefield reconstruction still need to be better understood.

Perspective work aims to extend IR-WRI to 3D configuration and multiparameter reconstruction and assess the method on real stationary-receiver sea-bottom case studies as those tackled by Operto et al. (2015); Operto and Miniussi (2018) where sharp contrasts generated by gas and chalky reservoir provide a suitable environment to assess the ability of IR-WRI to image large-contrast media and account for significant amplitude attenuation effects in realistic setting.

2.3.6 Conclusion

We have presented a new formulation of frequency-domain FWI based on wavefield reconstruction, which converges faster and improves the ability of the method to reconstruct short-scale structures, while preserving the resilience of the original formulation to cycle skipping. The improvement relies on an augmented Lagrangian method, which makes the method far less sensitive to the penalty parameter, when the value of this later is kept fixed in iterations. Accordingly, moderate values of the penalty parameter can be used such that the search space is significantly enlarged during the early iterations by relaxing the wave equation constraint without preventing to honor it with good accuracy at the final iteration. As the original penalty method, the augmented Lagrangian method is implemented with a splitting strategy, during which the wavefield reconstruction and the parameter estimation are performed in an alternating way and the solution of one subproblem is passed as a passive variable for the next problem. This splitting

strategy linearizes the parameter estimation around the reconstructed wavefield and manages the non-separability of the wavefield reconstruction and parameter estimation subproblems. A scaled form of the augmented Lagrangian also shows that the new method is equivalent to a self-adaptive penalty method, where the dual updating of the Lagrange multipliers is recast as the iterative updating of the data and source right-hand sides in the objectives of the penalty function. Therefore, the new method does not generate significant computational overhead relative to the original penalty method. Numerical examples suggest that the iterative updating of both the data and source residuals dynamically manage the two competing objectives of the penalty function leading to improved convergence history. Application on the BP salt model reveals the ability of the method to reconstruct large-contrast structures including salt bodies and sub-salt structures, starting from crude initial models and a 3-Hz starting frequency. Synthetic seismogram modeling confirms that the inversion didn't suffer from cycle skipping for this case study and manages to reasonably fit amplitudes of weak phases such as deep short-spread reflections. Perspectives deal with regularization, multiparameter reconstruction, extension to 3D and application to real data.

Acknowledgments

We would like to thank the editor A. Guitton, G. Barnier, M. Maharramov and three anonymous reviewers for their insightful comments which greatly help improving the manuscript. This study was partially funded by the SEISCOPE consortium (*http://seiscope2.osug.fr*), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXON-MOBIL, JGI, PETROBRAS, SCHLUMBERGER, SHELL, SINOPEC and TOTAL. This study was granted access to the HPC resources of SIGAMM infrastructure (http://crimson.oca.eu), hosted by Observatoire de la Côte d'Azur and which is supported by the Provence-Alpes Côte d'Azur region, and the HPC resources of CINES/IDRIS/TGCC under the allocation 0596 made by GENCI."

2.3.7 Appendix A: WRI based upon augmented Lagrangian

In this appendix, we develop WRI in the framework of augmented Lagrangian implemented with the method of multiplier and alternated directions. In particular, we show how we end up with two quadratic penalty subproblems for $\mathbf{u}$ and $\mathbf{m}$ similar to those of WRI, except that the objectives of the penalty functions embed now an error correction term in the form of right-hand side updating (equation 2.30). Let us first recast FWI as the following feasibility problem

$$\text{find} \quad \mathbf{m} \text{ and } \mathbf{u} \quad \text{subject to} \quad \mathbf{A}(\mathbf{m})\mathbf{u} = \mathbf{b} \quad \text{and} \quad \mathbf{P}\mathbf{u} = \mathbf{d}, \tag{2.38}$$

where the term $\mathbf{P}\mathbf{u} = \mathbf{d}$ is now processed as an equality constraint in case of noiseless data. In presence of noise, this equality constraint becomes an inequality constraint, $\|\mathbf{P}\mathbf{u} - \mathbf{d}\|_2^2 < \epsilon_n$, where ϵ_n represents the noise level. We assume that the feasible set is nonempty, namely the constraints are consistent or there are vectors $\mathbf{m}$ and $\mathbf{u}$ satisfying the constraints. The feasibility problem can be considered as a special case of a constrained optimization problem where the objective function is identically zero (Boyd and Vandenberghe, 2004, pages 128-129).

Accordingly, the feasibility problem, equation 2.38, is recast as the constrained problem

$$\min_{\mathbf{u},\mathbf{m}} \ \mathbf{0} \quad \text{subject to} \quad \mathbf{A}(\mathbf{m})\mathbf{u} = \mathbf{b} \quad \text{and} \quad \mathbf{P}\mathbf{u} = \mathbf{d}. \tag{2.39}$$

where the first and second constraints are bilinear and linear, respectively. The non-linear feasibility problem 2.39 can be written in more compact form as

$$\min_{\mathbf{u},\mathbf{m}} \ \mathbf{0} \quad \text{subject to} \quad \mathbf{F}(\mathbf{m})\mathbf{u} = \mathbf{s}, \tag{2.40}$$

where

$$\mathbf{s} = \begin{bmatrix} \mathbf{d} \\ \mathbf{b} \end{bmatrix} \in \mathbb{C}^{(M+N)\times 1} \quad \text{and} \quad \mathbf{F}(\mathbf{m}) = \begin{bmatrix} \mathbf{P} \\ \mathbf{A}(\mathbf{m}) \end{bmatrix} \in \mathbb{C}^{(M+N)\times N}.$$

The augmented Lagrangian function, Boyd et al. (2010, their equation 2.6) and Nocedal and Wright (2006, their equation 17.36), for the problem defined by equation 2.40 is

$$\mathcal{L}(\mathbf{m}, \mathbf{u}, \mathbf{v}) = \mathbf{v}^T[\mathbf{F}(\mathbf{m})\mathbf{u} - \mathbf{s}] + \|\mathbf{F}(\mathbf{m})\mathbf{u} - \mathbf{s}\|^2_{\mathbf{\Gamma}}, \tag{2.41}$$

where $\|\mathbf{x}\|^2_{\mathbf{Q}} = \mathbf{x}^T\mathbf{Q}\mathbf{x}$ and

$$\mathbf{\Gamma} = \left(\begin{array}{cc|c} \gamma_0 & & 0 \\ & \ddots & \\ 0 & & \gamma_0 \\ \hline & 0 & \end{array}\right)\begin{matrix} \\ \\ \\ M \\ \\ \end{matrix} \quad \left(\begin{array}{cc} 0 & \\ & \gamma_1 \\ & \ddots \\ & 0 \end{array}\right)\begin{matrix} \\ \\ \\ \\ \gamma_1 \end{matrix} \begin{matrix} \\ \\ \\ \\ N \end{matrix}$$

Applying the method of multipliers, Boyd et al. (2010, their equations 2.7-2.8) and Nocedal and Wright (2006, their equation 17.39), we get

$$(\mathbf{m}^{k+1}, \mathbf{u}^{k+1}) = \arg\min_{\mathbf{m},\mathbf{u}} \ \mathcal{L}(\mathbf{m}, \mathbf{u}, \mathbf{v}^k), \tag{2.42a}$$

$$\mathbf{v}^{k+1} = \mathbf{v}^k + \mathbf{\Gamma}[\mathbf{F}(\mathbf{m}^{k+1})\mathbf{u}^{k+1} - \mathbf{s}], \tag{2.42b}$$

for $k = 0, 1, ...$ beginning from $\mathbf{v}^0 = \mathbf{0}$. This iteration can be interpreted as follows: we begin with a prior estimate $\mathbf{v}^0$ of the dual variable $\mathbf{v}$ and minimize the objective function with respect to the primal variables $\mathbf{m}$ and $\mathbf{u}$, equation 2.42a. Then, we update the dual variable $\mathbf{v}$ via a steepest ascent method by keeping the primal variables $\mathbf{m}$ and $\mathbf{u}$ fixed (equation 2.42b). This process is iterated until convergence, i.e., when $\mathbf{F}(\mathbf{m}^{k+1})\mathbf{u}^{k+1} = \mathbf{s}$. To highlight more explicitly the key role of iterative right-hand side updating in our method and the relationship with the penalty method of van Leeuwen and Herrmann (2013), we rewrite the Lagrangian, equation 2.41, in a scaled form (Boyd et al., 2010, Section 3.1.1). Introducing the scaled dual variable $\mathbf{s}^k = -\mathbf{\Gamma}^{-1}\mathbf{v}^k$ in the Lagrangian and adding the constant terms $\frac{1}{2}\|\mathbf{s}^k\|^2_{\mathbf{\Gamma}}$ to it, we find

$$\mathcal{L}(\mathbf{m}, \mathbf{u}, \mathbf{s}^k) = \|\mathbf{F}(\mathbf{m})\mathbf{u} - \mathbf{s} - \mathbf{s}^k\|^2_{\mathbf{\Gamma}}. \tag{2.43}$$

Without modifying its minimum, we can scale the augmented Lagrangian function 2.43 by $1/\gamma_0$ to eliminate one of the penalty parameters and get

$$\mathcal{L}(\mathbf{m},\mathbf{u},\mathbf{s}^k) = \|\mathbf{F}(\mathbf{m})\mathbf{u}-\mathbf{s}-\mathbf{s}^k\|^2_{\mathbf{\Lambda}}, \tag{2.44}$$

where

$$\mathbf{\Lambda} = \begin{pmatrix} \begin{matrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{matrix} & \mathbf{0} \\ \mathbf{0} & \begin{matrix} \lambda & & 0 \\ & \ddots & \\ 0 & & \lambda \end{matrix} \end{pmatrix} \begin{matrix} M \\ , \\ N \end{matrix}$$

and $\lambda = \gamma_1/\gamma_0$. Method of multipliers for the scaled-form Lagrangian reads as

$$(\mathbf{m}^{k+1},\mathbf{u}^{k+1}) = \underset{\mathbf{m},\mathbf{u}}{\arg\min}\ \|\mathbf{F}(\mathbf{m})\mathbf{u}-\mathbf{s}-\mathbf{s}^k\|^2_{\mathbf{\Lambda}}, \tag{2.45a}$$

$$\mathbf{s}^{k+1} = \mathbf{s}^k + \mathbf{s} - \mathbf{F}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}, \tag{2.45b}$$

with $\mathbf{s}^0 = 0$.
The scaled-form Lagrangian, equation 2.44, is similar to the penalty function of van Leeuwen and Herrmann (2016), equation 2.27, except that the original right-hand side $\mathbf{s}$ (which gathers $\mathbf{d}$ and $\mathbf{b}$ according to our compact notation) is updated at each iteration with the scaled dual variable $\mathbf{s}^k$. This scaled dual variable gathers the running sum of the constraint violations of previous iterations, namely, the data and source residuals. This right-hand side (or residual) updating can be viewed as the counterpart of the residual updating performed at each iteration of the reduced FWI, this residual updating being limited to the data misfit term in this later case (since the wave equation error is 0 in reduced FWI).

Alternating-direction method of multipliers

The joint update of the primal variables ($\mathbf{m}$,$\mathbf{u}$) in equation 2.45a is computationally too intensive. A splitting method can be useful here to decouple the optimization tasks with respect to $\mathbf{m}$ and $\mathbf{u}$ (the readers can refer to Glowinski et al., 2017, for an overview of splitting methods). Instead of joint updating, the alternating-direction method of multipliers (ADMM) (Boyd et al., 2010) updates the variables $\mathbf{m}$ and $\mathbf{u}$ with Gauss-Seidel like iterations, i.e., fixing one variable and solving for the other. Using ADMM, we end up with the following iteration:

$$\mathbf{u}^{k+1} = \underset{\mathbf{u}}{\arg\min}\ \|\mathbf{F}(\mathbf{m}^k)\mathbf{u}-\mathbf{s}-\mathbf{s}^k\|^2_{\mathbf{\Lambda}}, \tag{2.46a}$$

$$\mathbf{m}^{k+1} = \underset{\mathbf{m}}{\arg\min}\ \|\mathbf{F}(\mathbf{m})\mathbf{u}^{k+1}-\mathbf{s}-\mathbf{s}^k\|^2_{\mathbf{\Lambda}}, \tag{2.46b}$$

$$\mathbf{s}^{k+1} = \mathbf{s}^k + \mathbf{s} - \mathbf{F}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}. \tag{2.46c}$$

ADMM algorithm has decoupled the full problem into two subproblems associated with primal variables $\mathbf{u}$ and $\mathbf{m}$. Compared to classical ADMM for separable convex problems and linear

constraints, the non-separability between wavefield reconstruction and parameter estimation is managed by passing the solution of one primal subproblem as a passive variable for the next subproblem. We will show soon that this alternating-direction approach linearizes the primal subproblem for $\mathbf{m}$ around the reconstructed wavefield $\mathbf{u}$. As such, our approach can also be seen as a linearly constrained augmented Lagrangian method, the linearization simply resulting from the splitting procedure (Nocedal and Wright, 2006, pages 522-523). The fact that the primal update of one subproblem is passed to the next subproblem implies obviously that the two subproblems are solved in sequence rather than in parallel as in classical ADMM.
For reasons that will become apparent below, we reintroduce explicitly the two components $\mathbf{d}$ and $\mathbf{b}$ of the scaled dual vector $\mathbf{s}$ in the notations

$$\begin{aligned}
\mathbf{u}^{k+1} &= \arg\min_{\mathbf{u}} \ \|\mathbf{P}\mathbf{u}-\mathbf{d}-\mathbf{d}^k\|_2^2+\lambda\|\mathbf{A}(\mathbf{m}^k)\mathbf{u}-\mathbf{b}-\mathbf{b}^k\|_2^2, && \text{(2.47a)}\\
\mathbf{m}^{k+1} &= \arg\min_{\mathbf{m}} \ \|\mathbf{A}(\mathbf{m})\mathbf{u}^{k+1}-\mathbf{b}-\mathbf{b}^k\|_2^2, && \text{(2.47b)}\\
\mathbf{d}^{k+1} &= \mathbf{d}^k+\mathbf{d}-\mathbf{P}\mathbf{u}^{k+1}, && \text{(2.47c)}\\
\mathbf{b}^{k+1} &= \mathbf{b}^k+\mathbf{b}-\mathbf{A}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}. && \text{(2.47d)}
\end{aligned}$$

A second modification related to classical ADMM resides in the updating of the dual variables. In classical ADMM, the dual variables are updated only once per iteration after the residuals generated by each primal subproblem have been gathered, equations 2.47c-2.47d. As reviewed by Boyd et al. (2010, page 23) and Esser (2009, page 3), a variant of ADMM, referred to as the Peaceman-Rachford splitting method (PRSM) (Peaceman and Rachford, 1955), consists of updating the Lagrange multipliers several times, once after the update of each primal variable (see He et al. (2014, equations 1.3 versus 1.4). One issue with PRSM relative to ADMM is to require more restrictive assumptions to ensure its convergence, while it is always faster than ADMM whenever it is convergent (He et al., 2014). This issue prompts He et al. (2014) to implement relaxation factors (or, step lengths) $\alpha \in (0,1)$ to guarantee the strict contraction of the PRSM iterative sequence. Accordingly, applying the strictly contractive PRSM algorithm to our minimization problem gives

$$\begin{aligned}
\mathbf{u}^{k+1} &= \arg\min_{\mathbf{u}} \ \|\mathbf{P}\mathbf{u}-\mathbf{d}-\mathbf{d}^k\|_2^2+\lambda\|\mathbf{A}(\mathbf{m}^k)\mathbf{u}-\mathbf{b}-\mathbf{b}^k\|_2^2, && \text{(2.48a)}\\
\mathbf{d}^{k+1} &= \mathbf{d}^k+\mathbf{d}-\mathbf{P}\mathbf{u}^{k+1}, && \text{(2.48b)}\\
\mathbf{b}^{k+\frac{1}{2}} &= \mathbf{b}^k+\alpha[\mathbf{b}-\mathbf{A}(\mathbf{m}^k)\mathbf{u}^{k+1}], && \text{(2.48c)}\\
\mathbf{m}^{k+1} &= \arg\min_{\mathbf{m}} \ \|\mathbf{A}(\mathbf{m})\mathbf{u}^{k+1}-\mathbf{b}-\mathbf{b}^{k+\frac{1}{2}}\|_2^2, && \text{(2.48d)}\\
\mathbf{b}^{k+1} &= \mathbf{b}^{k+\frac{1}{2}}+\alpha[\mathbf{b}-\mathbf{A}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}], && \text{(2.48e)}
\end{aligned}$$

where $\alpha \in (0,1)$. The dual variable $\mathbf{b}^k$ is updated twice with a step α, once after the update of $\mathbf{u}$ and once after the update of $\mathbf{m}$, since it depends upon both primal variables, while $\mathbf{d}^k$ is updated only once after the update of $\mathbf{u}$. In the numerical section, we use the PRSM formulation, equation 2.48 at the expense of ADMM, equation 2.47, as we found that it was leading to a faster convergence provided that $\alpha < 1$ to guarantee stability.

2.3.8 Appendix B: Iterative solution refinement for linear inverse problems

When one solves a system of equations, the computed solution may deviate from the desired one. In linear algebra, this might result from round-off errors in large-scale problems (e.g. Press et al., 2007, page 61). In convex and nonlinear optimization, this might arise when regularization or penalty methods are used to solve an ill-conditioned problem (Gholami and Gheymasi, 2016). Another example deals with approximate solution of nonlinear inverse problem through linearization, as for example least-squares migration/inversion, where the Born approximation, sometimes combined with ray theory, is used to linearize the wave equation around a background velocity model (Lambaré et al., 1992; Jin et al., 1992) or linearized AVO inversion (Gholami et al., 2018). The residual errors in the computed solution may require to solve repeatedly the problem to iteratively refine the solution accuracy. In this appendix, we review this iterative solution refinement procedure which amounts to update at a given iteration the right-hand side of the linear problem with the running sum of the residuals of previous iterations.

Let's solve iteratively

$$\mathbf{A}\mathbf{x} = \mathbf{b}. \tag{2.49}$$

At iteration 1, we find an approximate solution $\mathbf{x}_1$ by solving the quadratic program in a least-squares sense

$$\mathbf{x}_1 = \arg\min_x \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 = \mathbf{A}^{-g}\mathbf{b}, \tag{2.50}$$

where $\mathbf{A}^{-g}$ denotes the generalized or pseudoinverse of $\mathbf{A}$, $\mathbf{A}^{-g} = \left[\mathbf{A}^T\mathbf{A} + \beta\mathbf{I}\right]^{-1}\mathbf{A}^T$ and β is a stabilizing parameter (Menke, 2012).
The solution $\mathbf{b}_1$ of the forward problem is given by

$$\mathbf{b}_1 = \mathbf{A}\mathbf{x}_1, \tag{2.51}$$

and fits $\mathbf{b}$ with an error $\delta\mathbf{b}_1 = \mathbf{b} - \mathbf{b}_1$. We search the perturbation $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_1$ that needs to be applied on $\mathbf{x}_1$ to correct the residual $\delta\mathbf{b}_1$:

$$\delta\mathbf{x} = \arg\min_x \|\mathbf{A}\delta\mathbf{x} - \delta\mathbf{b}_1\|_2^2 = \mathbf{A}^{-g}\delta\mathbf{b}_1 = \mathbf{A}^{-g}\left[\mathbf{b} - \mathbf{A}\mathbf{x}_1\right]. \tag{2.52}$$

The refined solution is given by

$$\mathbf{x}_2 = \mathbf{x}_1 + \mathbf{A}^{-g}\delta\mathbf{b}_1 = \mathbf{A}^{-g}\mathbf{b} + \mathbf{A}^{-g}\left[\mathbf{b} - \mathbf{A}\mathbf{x}_1\right]. \tag{2.53}$$

Proceeding with the next iteration leads to the following refined solution

$$\mathbf{x}_3 = \mathbf{x}_2 + \mathbf{A}^{-g}[\mathbf{b} - \mathbf{A}\mathbf{x}_2] = \mathbf{A}^{-g}\mathbf{b} + \mathbf{A}^{-g}\left[\mathbf{b} - \mathbf{A}\mathbf{x}_1\right] + \mathbf{A}^{-g}[\mathbf{b} - \mathbf{A}\mathbf{x}_2]. \tag{2.54}$$

Repeating this procedure k times, we obtain

$$\mathbf{x}_k = \mathbf{A}^{-g}\mathbf{b} + \sum_{i=1}^{k-1} \mathbf{A}^{-g}\left[\mathbf{b} - \mathbf{A}\mathbf{x}_i\right] = \mathbf{A}^{-g}\left[\mathbf{b} + \sum_{i=1}^{k-1}[\mathbf{b} - \mathbf{A}\mathbf{x}_i]\right]. \tag{2.55}$$

Analogy between equations 2.50 and 2.55 allows us to write $\mathbf{x}_k$ as the solution of the quadratic program

$$\mathbf{x}_k = \arg\min_x \|\mathbf{A}\mathbf{x} - \mathbf{b} - \mathbf{b}^{k-1}\|_2^2, \tag{2.56}$$

where the original right-hand side $\mathbf{b}$ has been updated with the running sum of the residuals of the first $k-1$ iterations, $\mathbf{b}^{k-1} = \sum_{i=1}^{k-1}[\mathbf{b} - \mathbf{A}\mathbf{x}_i]$. This means that, instead of looking for a solution perturbation and adding this perturbation to the current solution to refine it, the imprint of the current solution is directly injected in the objective function via its linear relationship with the residuals of previous iterations. This iterative updating of the right-hand sides of a linear problem describes the approach that is used in this study to refine the solutions of the wavefield reconstruction and subsurface model estimation subproblems when the initial Lagrangian method is recast as a penalty method after scaling the dual variables by the penalty parameter, equation 2.44.

2.4 Local convergence proof

To the best of our knowledge, there is no global convergence result in general for non-convex problems, although the global convergence of bi-convex problems is an open problem. As in Xu et al. (2012b), we show the necessary condition for local convergence. We provide a convergence property under some assumptions.
We can claim that algorithm 1 has a local convergence property if the solution at k-iteration satisfies Karush-Kuhn-Tucker (KKT) condition for augmented Lagrangian

$$\mathcal{L}_A(\mathbf{m}, \mathbf{u}, \bar{\mathbf{b}}, \bar{\mathbf{d}}) = \lambda\|\mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{b} + \bar{\mathbf{b}}\|_2^2 + \|\mathbf{P}\mathbf{u} - \mathbf{d} + \bar{\mathbf{d}}\|_2^2 - \lambda\|\bar{\mathbf{b}}\|_2^2 - \|\bar{\mathbf{d}}\|_2^2. \tag{2.57}$$

From mathematical point of view, the primals and duals at k-iteration $(\mathbf{u}^k, \mathbf{m}^k, \bar{\mathbf{b}}^k, \bar{\mathbf{d}}^k)$ satisfy KKT condition for augmented Lagrangian (equation 2.57) if

$$\frac{\partial \mathcal{L}_A}{\partial \mathbf{m}} = 0 \rightarrow \mathbf{L}(\mathbf{u}^k)^T[\mathbf{L}(\mathbf{u}^k)\mathbf{m}^k + \mathbf{b} - \bar{\mathbf{b}}^k - \boldsymbol{\Delta}\mathbf{u}^k] = 0 \tag{2.58a}$$

$$\frac{\partial \mathcal{L}_A}{\partial \mathbf{u}} = 0 \rightarrow \lambda\mathbf{A}(\mathbf{m}^k)^T[\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b} - \bar{\mathbf{b}}^k] + \mathbf{P}^T[\mathbf{P}\mathbf{u}^k - \mathbf{d} - \bar{\mathbf{d}}^k] = 0 \tag{2.58b}$$

$$\frac{\partial \mathcal{L}_A}{\partial \bar{\mathbf{b}}} = 0 \rightarrow \mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b} = 0 \tag{2.58c}$$

$$\frac{\partial \mathcal{L}_A}{\partial \bar{\mathbf{d}}} = 0 \rightarrow \mathbf{P}\mathbf{u}^k - \mathbf{d} = 0. \tag{2.58d}$$

Theorem 1. *Let* $(\mathbf{u}^k, \mathbf{m}^k, \bar{\mathbf{b}}^k, \bar{\mathbf{d}}^k)$ *be the primals and duals which are generated at the k-th*

iteration of ADMM (Algorithm 1). If the duals $(\bar{\mathbf{b}}^k, \bar{\mathbf{d}}^k)$ *are bounded and satisfy*

$$\sum_{k=0}^{\infty} \left[\|\bar{\mathbf{b}}^{k+1} - \bar{\mathbf{b}}^k\|_2^2 + \|\bar{\mathbf{d}}^{k+1} - \bar{\mathbf{d}}^k\|_2^2 \right] < \infty, \tag{2.59}$$

then any accumulation point of $\mathbf{m}^k$ *and* $\mathbf{u}^k$ *satisfies the KKT conditions 2.58*

Proof. We claim that

$$\mathbf{m}^{k+1} - \mathbf{m}^k \to 0 \tag{2.60a}$$
$$\mathbf{u}^{k+1} - \mathbf{u}^k \to 0 \tag{2.60b}$$
$$\bar{\mathbf{b}}^{k+1} - \bar{\mathbf{b}}^k \to 0 \tag{2.60c}$$
$$\bar{\mathbf{d}}^{k+1} - \bar{\mathbf{d}}^k \to 0. \tag{2.60d}$$

The third and the forth terms are directly extracted from 2.59 as $k \to \infty$. We begin the proof of the first and the second one by the bi-convexity of augmented Lagrangian (equation 2.43). Because of the bi-convexity, $\mathcal{L}_A$ is strongly convex w.r.t each primal variable $(\mathbf{u}, \mathbf{m})$, and so the following inequality hold (Nishihara et al., 2015)

$$\mathcal{L}_A(\mathbf{m}^k, \mathbf{u}, \bar{\mathbf{b}}, \bar{\mathbf{d}}) - \mathcal{L}_A(\mathbf{m}^{k+1}, \mathbf{u}, \bar{\mathbf{b}}, \bar{\mathbf{d}}) \geq \partial_{\mathbf{m}} \mathcal{L}_A(\mathbf{m}^{k+1}, \mathbf{u}, \bar{\mathbf{b}}, \bar{\mathbf{d}})^T [\mathbf{m}^k - \mathbf{m}^{k+1}] + \|\mathbf{m}^k - \mathbf{m}^{k+1}\|_2^2. \tag{2.61}$$

In addition, if $\mathbf{m}^{k+1}$ be a minimizer of $\mathcal{L}_A$ w.r.t $\mathbf{m}$, we have

$$\partial_{\mathbf{m}} \mathcal{L}_A(\mathbf{m}^{k+1}, \mathbf{u}, \bar{\mathbf{b}}, \bar{\mathbf{d}})^T [\mathbf{m}^k - \mathbf{m}^{k+1}] \geq 0 \tag{2.62}$$

and so

$$\mathcal{L}_A(\mathbf{m}^k, \mathbf{u}, \bar{\mathbf{b}}, \bar{\mathbf{d}}) - \mathcal{L}_A(\mathbf{m}^{k+1}, \mathbf{u}, \bar{\mathbf{b}}, \bar{\mathbf{d}}) \geq \|\mathbf{m}^k - \mathbf{m}^{k+1}\|_2^2. \tag{2.63}$$

In the same way, we can have the following inequality for $\mathbf{u}$

$$\mathcal{L}_A(\mathbf{m}, \mathbf{u}^k, \bar{\mathbf{b}}, \bar{\mathbf{d}}) - \mathcal{L}_A(\mathbf{m}, \mathbf{u}^{k+1}, \bar{\mathbf{b}}, \bar{\mathbf{d}}) \geq \|\mathbf{u}^k - \mathbf{u}^{k+1}\|_2^2. \tag{2.64}$$

Using equations 2.63 and 2.64, the change in augmented Lagrangian value at iteration k and $k+1$ will be

$$\begin{aligned}
&\mathcal{L}_A(\mathbf{m}^k, \mathbf{u}^k, \bar{\mathbf{b}}^k, \bar{\mathbf{d}}^k) - \mathcal{L}_A(\mathbf{m}^{k+1}, \mathbf{u}^{k+1}, \bar{\mathbf{b}}^{k+1}, \bar{\mathbf{d}}^{k+1}) \\
&= \mathcal{L}_A(\mathbf{m}^k, \mathbf{u}^k, \bar{\mathbf{b}}^k, \bar{\mathbf{d}}^k) - \mathcal{L}_A(\mathbf{m}^{k+1}, \mathbf{u}^k, \bar{\mathbf{b}}^k, \bar{\mathbf{d}}^k) \\
&+ \mathcal{L}_A(\mathbf{m}^{k+1}, \mathbf{u}^k, \bar{\mathbf{b}}^k, \bar{\mathbf{d}}^k) - \mathcal{L}_A(\mathbf{m}^{k+1}, \mathbf{u}^{k+1}, \bar{\mathbf{b}}^k, \bar{\mathbf{d}}^k) \\
&+ \mathcal{L}_A(\mathbf{m}^{k+1}, \mathbf{u}^{k+1}, \bar{\mathbf{b}}^k, \bar{\mathbf{d}}^k) - \mathcal{L}_A(\mathbf{m}^{k+1}, \mathbf{u}^{k+1}, \bar{\mathbf{b}}^{k+1}, \bar{\mathbf{d}}^k) \\
&+ \mathcal{L}_A(\mathbf{m}^{k+1}, \mathbf{u}^{k+1}, \bar{\mathbf{b}}^{k+1}, \bar{\mathbf{d}}^k) - \mathcal{L}_A(\mathbf{m}^{k+1}, \mathbf{u}^{k+1}, \bar{\mathbf{b}}^{k+1}, \bar{\mathbf{d}}^{k+1}) \\
&\geq \|\mathbf{m}^k - \mathbf{m}^{k+1}\|_2^2 + \|\mathbf{u}^k - \mathbf{u}^{k+1}\|_2^2 + \|\bar{\mathbf{b}}^k - \bar{\mathbf{b}}^{k+1}\|_2^2 + \|\bar{\mathbf{d}}^k - \bar{\mathbf{d}}^{k+1}\|_2^2,
\end{aligned} \tag{2.65}$$

where

$$\mathcal{L}_A(\mathbf{m}^{k+1},\mathbf{u}^{k+1},\bar{\mathbf{b}}^{k},\bar{\mathbf{d}}^{k})-\mathcal{L}_A(\mathbf{m}^{k+1},\mathbf{u}^{k+1},\bar{\mathbf{b}}^{k+1},\bar{\mathbf{d}}^{k})\geq\|\bar{\mathbf{b}}^{k}-\bar{\mathbf{b}}^{k+1}\|_2^2,$$
$$\mathcal{L}_A(\mathbf{m}^{k+1},\mathbf{u}^{k+1},\bar{\mathbf{b}}^{k+1},\bar{\mathbf{d}}^{k})-\mathcal{L}_A(\mathbf{m}^{k+1},\mathbf{u}^{k+1},\bar{\mathbf{b}}^{k+1},\bar{\mathbf{d}}^{k+1})\geq\|\bar{\mathbf{d}}^{k}-\bar{\mathbf{d}}^{k+1}\|_2^2,$$

are easily extracted using the triangle inequality when the values are substituted in the augmented Lagrangian (equation 2.57).
Taking summation of the inequality equation 2.65 from 1 to ∞, and consider that augmented Lagrangian is below bounded (which observed from the boundness of $\bar{\mathbf{b}}$ and $\bar{\mathbf{d}}$ and the positivity of other terms in augmented Lagrangian [equation 2.57]), then we have

$$\sum_{k=0}^{\infty}\left[\|\mathbf{m}^{k}-\mathbf{m}^{k+1}\|_2^2+\|\mathbf{u}^{k}-\mathbf{u}^{k+1}\|_2^2+\|\bar{\mathbf{b}}^{k}-\bar{\mathbf{b}}^{k+1}\|_2^2+\|\bar{\mathbf{d}}^{k}-\bar{\mathbf{d}}^{k+1}\|_2^2\right] \tag{2.66a}$$
$$\leq\sum_{k=0}^{\infty}\left[\mathcal{L}_A(\mathbf{m}^{k},\mathbf{u}^{k},\bar{\mathbf{b}}^{k},\bar{\mathbf{d}}^{k})-\mathcal{L}_A(\mathbf{m}^{k+1},\mathbf{u}^{k+1},\bar{\mathbf{b}}^{k+1},\bar{\mathbf{d}}^{k+1})\right]<\infty \tag{2.66b}$$

Since the third and forth term of equation 2.66a are bounded (equation 2.59), we have

$$\sum_{k=0}^{\infty}\left[\|\mathbf{m}^{k}-\mathbf{m}^{k+1}\|_2^2+\|\mathbf{u}^{k}-\mathbf{u}^{k+1}\|_2^2\right]<\infty, \tag{2.67}$$

from which we immediately have $\mathbf{m}^{k+1}-\mathbf{m}^{k}\to 0$ and $\mathbf{u}^{k+1}-\mathbf{u}^{k}\to 0$ as $k\to\infty$.

□

We can get the following corollary from the proof of Theorem 1 as

Corollary 1 *Let* $(\mathbf{u}^k,\mathbf{m}^k,\bar{\mathbf{b}}^k,\bar{\mathbf{d}}^k)$ *be the primals and duals at k-iteration of Algorithm 1.* $(\mathbf{u}^k,\mathbf{m}^k,\bar{\mathbf{b}}^k,\bar{\mathbf{d}}^k)$ *satisfies Karush-Kuhn-Tucker (KKT) condition for problem 2.57.*

Proof. First we subtract $\mathbf{u}^k$ and $\mathbf{m}^k$ from both side of equation 2.31 and 2.36, respectively, and we rearrange them and dual variables update equations as

$$\mathbf{L}(\mathbf{u}^{k+1})^T\mathbf{L}(\mathbf{u}^{k+1})\left[\mathbf{m}^{k+1}-\mathbf{m}^{k}\right]$$
$$=\mathbf{L}(\mathbf{u}^{k+1})^T[\mathbf{b}-\bar{\mathbf{b}}^{k}-\boldsymbol{\Delta}\mathbf{u}^{k+1}]-\mathbf{L}(\mathbf{u}^{k+1})^T\mathbf{L}(\mathbf{u}^{k+1})\mathbf{m}^{k}, \tag{2.68a}$$
$$\left[\lambda\mathbf{A}(\mathbf{m}^{k})^T\mathbf{A}(\mathbf{m}^{k})+\mathbf{P}^T\mathbf{P}\right]\left[\mathbf{u}^{k+1}-\mathbf{u}^{k}\right]$$
$$=\lambda\mathbf{A}(\mathbf{m}^{k})^T[\mathbf{b}-\bar{\mathbf{b}}^{k}]+\mathbf{P}^T[\mathbf{d}-\bar{\mathbf{d}}^{k}]-\left[\lambda\mathbf{A}(\mathbf{m}^{k})^T\mathbf{A}(\mathbf{m}^{k})+\mathbf{P}^T\mathbf{P}\right]\mathbf{u}^{k}, \tag{2.68b}$$
$$\bar{\mathbf{b}}^{k+1}-\bar{\mathbf{b}}^{k}=\mathbf{A}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}-\mathbf{b}, \tag{2.68c}$$
$$\bar{\mathbf{d}}^{k+1}-\bar{\mathbf{d}}^{k}=\mathbf{P}\mathbf{u}^{k+1}-\mathbf{d}. \tag{2.68d}$$

Theorem 1 implies that the left side of equation 2.68 goes to zero as soon as $k\to\infty$ and the right side of equation 2.68 is exactly KKT condition of the augmented Lagrangian (equation 2.58). □

Chapter 3

Regularization and bound constraints in ADMM-WRI

Chapter overview: The focus of this chapter is the implementation of suitable regularization and constraint for subsurface imaging in IR-WRI as well as its applications. The failure that we saw in the inversion results of IR-WRI has not resulted from cycle skipping but from a poor convergence of the optimization, which generates subsurface model of limited resolution in areas that are located far away from the receivers. Seismic surveys perform only a sparse recording of the wavefields at the receiver positions and so the inverse problem has a huge null space, which means that there are a lot of models that can fit the recorded data. To tackle this ill-conditioned problem, we need to inject prior information via regularization to drive the inversion toward a good model everywhere. Selecting a suitable regularization function is directly related to the statistical properties of the model that we are interested in. The two most popular regularizations in seismic imaging are Tikhonov regularization, which drives FWI toward smooth reconstruction by minimization of the roughness of the estimated model, and total-variation (TV) regularization, which drives the inversion toward piecewise constant or blocky models by fostering the sparsity of the first differences of the estimated model. According to our geological knowledge of the earth's interior, the earth usually can be viewed as a piecewise smooth medium, that is a combination of blocky and smooth media. Therefore, neither Tikhonov nor TV are fully representative of such models. To reconstruct a better model, we can use more complicated regularization functions or combine simple regularizations to jointly catch the smooth and blocky trends of the model. Independently from the choice of a suitable regularization function, sometimes applying regularization is difficult because of the non-differentiability of the associated function, as is the case of sparse regularizations.
In this chapter, I begin by reviewing the basic concepts and definitions of regularization. Then, I apply TV regularization and bound constraints to the IR-WRI method using splitting methods and ADMM (split- Bregman) to overcome the non-differentiability of TV regularization. This section contains our article which has been published in the *Geophysical journal international* (Aghamiry et al., 2019b). As I mentioned, because of the piecewise smooth nature of the earth, TV is not the best choice as only the blocky components of the model are properly reconstructed with such regularization. I continue by combining simple regularization func-

tions via convex combination and infimal convolution to find a better regularization function for earth imaging. This section contains the article which has been published in *IEEE Transactions on Geoscience and Remote Sensing journal* (Aghamiry et al., 2020b). Finally, I finish this chapter with the application of sparse regularization in a compressive sensing process to assess the robustness of IR-WRI against sparse long-offset stationary-recording acquisitions.

3.1 Introduction to regularization

Regularization is the process of adding some prior and statistical information to an ill-posed inverse problem to prevent over-fitting and obtain a unique and stable solution among all the possible solutions.
In order to obtain a unique and stable solution, it is possible to reformulate the problem as an optimization problem with a regularization term, so that, out of all the models that fit the data to some extent we keep the one that matches the regularization priors. In regularization, the original ill-posed problem will be replaced with a well-posed problem, which has a solution close to the correct solution.
The regularization term can be added to the optimization problem via a penalty term or a constraint. With penalty term, it reads

$$\min_{\mathbf{u},\mathbf{m}} \quad \mathcal{P}(\mathbf{m},\mathbf{u}) + \gamma\mathcal{R}(\mathbf{m}) \tag{3.1a}$$

$$\text{subject to} \quad \mathcal{C}(\mathbf{m},\mathbf{u}) = 0, \tag{3.1b}$$

where $\gamma > 0$ is the regularization weight and $\mathcal{R}(\mathbf{m})$ is the regularization function.
Also, it can take the form of an extra constraint as

$$\min_{\mathbf{u},\mathbf{m}} \quad \mathcal{P}(\mathbf{m},\mathbf{u}) \tag{3.2a}$$

$$\text{subject to} \quad \begin{cases} \mathcal{C}(\mathbf{m},\mathbf{u}) = 0, \\ \mathcal{R}(\mathbf{m}) < \tau, \end{cases} \tag{3.2b}$$

where $\tau > 0$ is the reconstructed model regularization value.
It can be shown that these two optimization problems can reach the same minimizer using a specific relation between γ and τ (Hansen, 2005).
The regularization term $\mathcal{R}(\mathbf{m})$ contains a measure of the model coefficients in a specific domain. Most of the time this measure is a p-norm function. The selected norm for regularization is directly related to the statistical distribution of the model coefficient in the regularization domain. For simple regularization functions it will be

$$\mathcal{R}(\mathbf{m}) = \|\mathbf{X}\mathbf{m}\|_p, \tag{3.3}$$

where $\mathbf{X}$ is the regularization domain and $\|\bullet\|_p$ is the p-norm of $\bullet$. If $\bullet$ has a normal distribution, 2-norm can be a good choice, while 1-norm is a good choice for long-tail distributions.
In Tikhonov family regularizations, p is equal to 2 and $\mathbf{X}$ is identity matrix or derivative operators. In total variation regularization family, the 1-norm is applied on the first or higher order derivative of the model coefficients. The derivative is not the only choice for regularization domain. It can be every kind of transform like wavelet, Fourier, curvelet etc. For example, sparsity in the Fourier domain can be achieved by $\mathcal{R}(\mathbf{m}) = \|\mathbf{F}\mathbf{m}\|_1$, where $\mathbf{F}$ denotes the Fourier transform.
When the regularization function is non-differentiable, finding the minimizer is more challenging. Among plenty of methods that were proposed to tackle this kind of non-differentiable optimization (Glowinski et al., 2017, chapter 7), I am going to use splitting methods and specifically the split-Bregman method (Goldstein and Osher, 2009) in this thesis. According to the split-Bregman method, I recast the unconstrained problem as a constrained problem by defining some auxiliary variables that are equal to the non-differentiable terms and build the augmented Lagrangian and find its stationary point using ADMM. The splitting techniques allow for breaking down the original problems to a few easy to solve convex subproblems. I describe the split-Bregman algorithm completely in the following sections.
The other priors that I am going to add to the optimization problem are bound constraints or box constraints, which bound the estimated model values between a preset bounds as

$$\min_{\mathbf{u},\mathbf{m}} \quad \mathcal{P}(\mathbf{m},\mathbf{u}) \tag{3.4a}$$

$$\text{subject to} \quad \begin{cases} \mathcal{C}(\mathbf{m},\mathbf{u}) = 0 \\ \mathbf{m}_l \leq \mathbf{m} \leq \mathbf{m}_u \end{cases}, \tag{3.4b}$$

where $\mathbf{m}_l$ and $\mathbf{m}_u$ are the lower and upper bound, respectively. Split-Bregman is also a straightforward method to apply such constraints. It will be discussed precisely in the second section.

3.2 Implementing bound constraints and total-variation regularization in extended full waveform inversion with the alternating direction method of multiplier: application to large contrast media

This section includes our paper about applying TV regularization and bound constraints on IR-WRI using splitting techniques and ADMM (Aghamiry et al., 2019b).

Implementing bound constraints and total-variation regularization in extended full waveform inversion with the alternating direction method of multiplier: application to large contrast media

Hossein S. Aghamiry, Ali Gholami and Stéphane Operto

Geophysical Journal International, 2019, **218**(2), pages 855-872
DOI:10.1093/gji/ggz189

3.2.1 Summary

Full waveform inversion (FWI) is a waveform matching procedure, which can provide a subsurface model with a wavelength-scale resolution. However, this high resolution makes FWI prone to cycle skipping, which drives the inversion to a local minimum when the initial model is not accurate enough. Other sources of nonlinearities and ill-posedness are noise, uneven illumination, approximate wave physics and parameter cross-talks. All these sources of error require robust and versatile regularized optimization approaches to mitigate their imprint on FWI while preserving its intrinsic resolution power. To achieve this goal, we implement bound constraints and total variation (TV) regularization in the so-called frequency-domain wavefield-reconstruction inversion (WRI) with the alternating direction method of multipliers (ADMM). In the ADMM framework, WRI relies on an augmented Lagrangian function, a combination of penalty and Lagrangian functions, to extend the FWI search space by relaxing the wave-equation constraint during early iterations. Moreover, ADMM breaks down the joint wavefield reconstruction plus parameter estimation problem into a sequence of two linear subproblems, whose solutions are coordinated to provide the solution of the global problem. The decomposability of ADMM is further exploited to interface in a straightforward way bound constraints and TV regularization with WRI via variable splitting and proximal operators. The resilience of our regularized WRI formulation to cycle skipping and noise as well as its resolution power are illustrated with two targets of the large-contrast BP salt model. Starting from a 3Hz frequency and a crude initial model, the extended search space allows for the reconstruction of the salt and subsalt structures with a high fidelity. The TV regularization filters out the imprint of ambient noise and artifacts associated with multi-scattering and Gibbs effects, while fostering large-contrast reconstruction. Compared to other TV-regularized WRI implementations, the proposed method is easy to tune due to its moderate sensitivity to penalty parameters and does not require a prior guess of the TV-norm ball.

3.2.2 Introduction

During the last decade, full waveform inversion (FWI) has been used to estimate subsurface parameters (P and S wavespeeds, density, attenuation, anisotropic parameters) with a resolution close to the seismic wavelength by matching recorded and synthetic seismograms (Tarantola, 1984; Pratt et al., 1998; Virieux and Operto, 2009). From the numerical optimization viewpoint, the data-fitting/parameter-estimation problem underlying FWI is a nonlinear partial differential equation (PDE)-constrained optimization problem, where the equality constraint is the wave

equation and the optimization parameters are embedded in the coefficients of the PDE. Due to the computational burden of multiple source modelling and the size of the data and parameter spaces, this PDE-constrained optimization problem is solved with iterative local (linearized) optimization techniques, namely gradient-based methods (Nocedal and Wright, 2006). Moreover, it is often solved with a reduced-space formulation, which means that the full search space that encompasses the unknown wavefield and the subsurface parameters is first projected onto the parameter space by computing exactly the incident wavefields in the current subsurface model before updating this later (Haber et al., 2000; Askan et al., 2007; Epanomeritakis et al., 2008). It is well acknowledged that the oscillating nature of seismic signals makes the reduced-space formulation highly nonlinear as the modelled seismograms computed in the current subsurface model may be too far away from the recorded ones to satisfy the cycle-skipping criterion, that is the modelled seismograms should predict the recorded traveltimes with an error lower than half a period (e.g. Virieux and Operto, 2009).

Beyond cycle skipping, other sources of nonlinearity and ill-posedness such as noise, uneven subsurface illumination, approximate wave physics and parameter cross-talks in multi-parameter reconstruction require the use of stabilizing or regularization techniques that drive the inversion towards subsurface models that satisfy some a priori assumptions. Among the penalization techniques, Tikhonov regularization is probably the most popular one and seeks to penalize the roughness of the subsurface model to force smooth reconstruction (see Benning and Burger (2018) for a review). As the subsurface may be better represented by piecewise smooth media with potentially sharp contrasts as in presence of salt, edge-preserving techniques such as total-variation (TV) regularization have been proposed to steer the inversion to the space of blocky structured models. TV regularization has been applied on several geophysical applications such as FWI (Askan et al., 2007; Anagaw and Sacchi, 2011; Guitton, 2012; Maharramov and Biondi, 2015; Peters and Herrmann, 2017; Brandsberg-Dahl et al., 2017; Esser et al., 2018), seismic tomography (Gholami and Siahkoohi, 2010; Loris and Verhoeven, 2012), impedance inversion (Gholami, 2015, 2016), amplitude versus offset (AVO) inversion (Gholami et al., 2018), and seismic deconvolution (Gholami and Sacchi, 2013). For FWI applications, TV regularization can be implemented as a penalty function (Askan et al., 2007; Anagaw and Sacchi, 2011; Brandsberg-Dahl et al., 2017) or as a constraint (Peters and Herrmann, 2017; Esser et al., 2018). Choosing the most suitable implementation strategy may depend on the prior information on the TV norm of the model and on the optimization method that is used to minimize the objective function (Alkhalifah et al., 2018). If the information about the value of TV norm of the model is available, TV regularization can be implemented as a constraint in the FWI objective function. Otherwise, one may resort to a penalty method with the difficulty to design an adaptive penalty parameter, which optimally balances over iterations the relative weight of the data misfit and the total variation of the model in the objective.

In this context, the objective of this study is to present a novel implementation of TV regularization and bound-constraints in frequency-domain FWI based upon wavefield reconstruction inversion (WRI). WRI has been originally proposed by van Leeuwen and Herrmann (2013) to extend the search space and mitigate the risk of cycle skipping accordingly. WRI recasts the PDE-constrained optimization problem underlying FWI into an unconstrained quadratic penalty method, where the penalty term is the ℓ_2 norm of the source residuals (namely, the PDE-

constraint violation) that is weighted by a positive penalty parameter λ. The penalty method relaxes the wave-equation constraint at the benefit of the data fitting during early iterations, hence mitigating the risk of cycle skipping. To make WRI computationally tractable, van Leeuwen and Herrmann (2013) perform the wavefield reconstruction and the subsurface parameter estimation in an alternating way: first, keeping the current subsurface model fixed, the wavefields, which best jointly fit the observations and satisfy the wave equation in a least-squares sense, are reconstructed for each source; Second, keeping the previously-reconstructed wavefields fixed, the subsurface parameters are estimated by least-squares minimization of the source residuals the wave-equation relaxation generated. This cycle being iterated until convergence. A nice property of the alternating-direction strategy is to linearise the parameter-estimation subproblem around the reconstructed wavefield because the wave equation constraint is bilinear. However, a significant pitfall of WRI resides in the tuning of the penalty parameter λ. Ideally, increasing values should be used during iterations to progressively enforce the wave-equation constraint and, hence satisfy the first-order optimality conditions of the original constrained problem with acceptable precision at the minimizer. A significant issue is that this continuation approach is tedious to implement and the Hessian is ill conditioned for large λ. Therefore, van Leeuwen and Herrmann (2013) implement WRI with a small preset value of λ, which leads to slow convergence and a subsurface model of limited accuracy.
Later, van Leeuwen and Herrmann (2016) reformulated WRI as a reduced penalty method implemented with a variable projection approach: the closed-form expression of the extended-domain reconstructed wavefield is injected as a function of the subsurface parameters in the penalty function instead of using this wavefield as a passive variable (i.e., independent to the subsurface parameters). Unlike the alternating-direction approach, this variable projection leaves the parameter-estimation subproblem non linear. van Leeuwen and Herrmann (2016) assess their method with a Gauss-Newton method (by opposition to the full Newton counterpart) to mitigate the computational burden. Moreover, using a sparse approximation of the Gauss-Newton Hessian makes the descent direction of the reduced approach identical to that of the alternating-direction WRI of van Leeuwen and Herrmann (2013). Also, Aravkin et al. (2017) analysed the convergence properties of the reduced penalty method when the full Hessian is taken into account and concluded that the variable projection penalty method is insensitive to the penalty parameter. However, this convergence property still needs to be verified against realistic numerical experiments.

To make the alternating-direction WRI of van Leeuwen and Herrmann (2013) (referred to as WRI in the following for sake of brevity) more independent to the penalty parameter, Aghamiry et al. (2019c, 2018c) have replaced the penalty method by an augmented Lagrangian method (Nocedal and Wright, 2006), leading to the so-called iteratively-refined (IR)-WRI method. As in WRI, IR-WRI performs the primal wavefield and parameter updates in an alternating mode, while the Lagrange multipliers (i.e., the dual variables) are updated with a gradient ascent method. As above mentioned, this alternating direction strategy makes the parameter estimation subproblem linear due to the bilinearity of the wave equation constraint. It follows from this linearization that the alternating direction strategy combined with the augmented Lagrangian method is equivalent to an extension of the alternating direction method of multiplier (ADMM) to biconvex problem (Boyd et al., 2010). Also, using a scaled form of the augmented La-

grangian, we recast IR-WRI as a penalty method where the right-hand sides (the data and the sources) in the objective functions are iteratively updated with the running sum of the data and source residuals (the dual gradient steps). This reformulation of IR-WRI as a penalty method with right-hand side updating clearly draws some similarities and differences with WRI. The right-hand side updating makes IR-WRI largely insensitive to the penalty parameter for a wide range of preset values (Aghamiry et al., 2019c, Their Figures 2 and 3). Using a moderate value of the penalty parameter allows for significant wave equation error and improved data fitting during early iterations for search space extension, without preventing the fulfilment of the wave equation constraint with small error at the minimizer (Nocedal and Wright, 2006, Chapter 17, Theorem 17.6). This adaptivity makes IR-WRI resilient to cycle skipping as WRI with however a much faster convergence toward a more accurate minimizer. The reader is referred to Aghamiry et al. (2019c) for a thorough comparative convergence and accuracy analysis of WRI and IR-WRI based upon toy and complex large-contrast synthetic examples.

The objective of this study is to show how to interface TV regularization and bound constraints (hereafter, we refer to it as BTV regularization) with IR-WRI by taking advantage of the alternating-direction strategy implemented in ADMM and the split-Bregman variable splitting scheme developed by Goldstein and Osher (2009). More precisely, we recast the BTV regularized IR-WRI as a TV minimization problem subject to constraints, that are the modelled wavefield fit the observables and satisfy the wave equation with prescribed errors, and the model parameters preset bounds. As in WRI and IR-WRI, we solve the wavefield and subsurface parameter subproblems in an alternating mode. However, the later one involves now a combination of ℓ_1 and ℓ_2 norms related to the TV minimization and wave-equation error minimization, respectively, with additional bound constraints. This is managed by the split-Bregman variable splitting scheme, which de-couples the ℓ_1 and ℓ_2 components and bound constraints of the functional through the introduction of auxiliary variables and solves each related subproblem in sequence (Goldstein and Osher, 2009).

We first apply our method on a toy example corresponding to a high-velocity box-shape anomaly embedded in a background model where the velocity increases with depth. Then, we consider two more realistic examples corresponding to the left and central parts of the large-contrast 2004 BP salt model (Billette and Brandsberg-Dahl, 2004). We show that the BTV regularized IR-WRI converges to accurate minimizers when we start from a crude initial model and a realistic 3 Hz frequency. We also compare the results of WRI and IR-WRI without any priors, and when performed only with bound constraints and with BTV regularization to highlight the impact of each ingredient upon the quality of the results and the computational burden. We also assess the resilience of the method to noise by comparing the results that are obtained with noiseless and noisy data.

This paper is organized as follows. In the first section, we review the principles of WRI and IR-WRI. We first recast FWI as a feasibility problem and review different approaches that are suitable to solve PDE-constrained optimization problems such as penalty and augmented Lagrangian methods. Then, we review how we can easily interface some stabilizing terms with the feasibility problem through variable splitting and ADMM. In the second part, we present the results of the synthetic examples involving the inclusion model with two different starting

models and the two targets of the BP2004 model with noiseless and noisy data. The results confirm that the combined use of TV regularization and bound constraint in the ADMM-based IR-WRI method defines a suitable framework to make high-resolution FWI immune to cycle skipping in large-contrast media.

3.2.3 Method

In the following, we first recast frequency-domain FWI as a bi-convex feasibility problem, which can be formulated as a constrained optimization problem with identically-zero objective function (Aghamiry et al., 2019c, their appendix A). Then, we review the penalty and augmented Lagrangian methods as optimization techniques to solve this constrained optimization problem with an extended search-space, leading to WRI and IR-WRI, respectively. Finally, we interface bound constraints and the isotropic TV regularization with IR-WRI by replacing the identically-zero objective function by the TV norm of the subsurface model, and we show how to solve efficiently the regularized IR-WRI with ADMM and split Bregman iterations.

WRI and IR-WRI principles

FWI can be formulated in the frequency domain as the following bi-convex feasibility problem (Aghamiry et al., 2019c):

$$\text{Find} \qquad \mathbf{m} \text{ and } \mathbf{u} \tag{3.5a}$$

$$\text{subject to} \quad \mathbf{F}(\mathbf{m})\mathbf{u} = \mathbf{s} \tag{3.5b}$$

with

$$\mathbf{s} = \begin{bmatrix} \mathbf{d} \\ \mathbf{b} \end{bmatrix}, \qquad \mathbf{F}(\mathbf{m}) = \begin{bmatrix} \mathbf{P} \\ \mathbf{A}(\mathbf{m}) \end{bmatrix}, \tag{3.6}$$

where $\mathbf{m} \in \mathbb{R}^{N\times 1}$ denotes the vector of discrete model parameters (here, the squared slowness), $\mathbf{u} \in \mathbb{C}^{N\times 1}$ the wavefield, $\mathbf{b} \in \mathbb{C}^{N\times 1}$ the source term, $\mathbf{d} \in \mathbb{C}^{M\times 1}$ the recorded wavefield (data) at receiver locations, and $\mathbf{P} \in \mathbb{R}^{M\times N}$ is a linear observation operator that samples the modelled wavefield at the receiver positions. The matrix $\mathbf{A}(\mathbf{m}) \in \mathbb{C}^{N\times N}$ represents the discretized PDE Helmholtz operator (Pratt et al., 1998; Plessix, 2007; Chen et al., 2013).

$$\mathbf{A}(\mathbf{m}) \;=\; \boldsymbol{\Delta} + \omega^2 \mathbf{C}(\mathbf{m})\text{diag}(\mathbf{m})\mathbf{B}, \tag{3.7}$$

where ω is the angular frequency and $\boldsymbol{\Delta}$ is the discretized Laplace operator. The operator $\mathbf{C}$ encloses boundary conditions, which can be a function of $\mathbf{m}$ (e.g., Robin paraxial conditions, Engquist and Majda, 1977) or independent from $\mathbf{m}$ (e.g., sponge-like absorbing boundary conditions such as perfectly-matched layers, Bérenger, 1994). Also, the linear operator $\mathbf{B}$ can be used to spread the "mass" term $\omega^2\mathbf{C}(\mathbf{m})\text{diag}(\mathbf{m})$ over all the coefficients of the stencil to improve its accuracy following an anti-lumped mass strategy (Marfurt, 1984; Jo et al., 1996; Hustedt et al., 2004).

In the feasibility problem 3.5, we just want to find $\mathbf{m}$ and $\mathbf{u}$ that satisfy the constraint (we assume that the feasible set is non-empty, namely the constraint is consistent). The feasibility problem can be formulated as the following constrained optimization problem with identically-zero objective function (Boyd and Vandenberghe, 2004, Page 128)

$$\min_{\mathbf{u},\mathbf{m}} \quad \mathbf{0} \tag{3.8a}$$

$$\text{subject to} \quad \mathbf{F}(\mathbf{m})\mathbf{u} = \mathbf{s}. \tag{3.8b}$$

WRI penalty method

WRI implements this constrained optimization problem with a penalty method.

$$\min_{\mathbf{u},\mathbf{m}} \; \mathbf{0} + \|\mathbf{F}(\mathbf{m})\mathbf{u} - \mathbf{s}\|^2_{\mathbf{\Lambda}}, \tag{3.9}$$

where $\|\mathbf{x}\|^2_{\mathbf{Q}} := \mathbf{x}^T\mathbf{Q}\mathbf{x}$ for a vector $\mathbf{x}$ and a square matrix $\mathbf{Q}$ with superscript T denoting matrix transposition. In equation 3.9, $\mathbf{\Lambda}$ is a diagonal matrix which includes the penalty parameters λ_0, $\lambda_1 > 0$ on its main diagonal,

$$\mathbf{\Lambda} = \left[\begin{array}{ccc|ccc} \lambda_0 & & 0 & & & \\ & \ddots & & & \mathbf{0} & \\ 0 & & \lambda_0 & & & \end{array}\right] \begin{matrix} \\ \\ M \end{matrix} \\ \left[\begin{array}{ccc|ccc} & & & \lambda_1 & & 0 \\ & \mathbf{0} & & & \ddots & \\ & & & 0 & & \lambda_1 \end{array}\right] \begin{matrix} \\ \\ N \end{matrix} \tag{3.10}$$

The objective function in equation 3.9 is a compact written version of the penalty function of van Leeuwen and Herrmann (2013)

$$\min_{\mathbf{u},\mathbf{m}} \|\mathbf{P}\mathbf{u} - \mathbf{d}\|_2^2 + \lambda\|\mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{b}\|_2^2, \tag{3.11}$$

where $\lambda = \lambda_1/\lambda_0$. They solve this biconvex minimization problem with an alternating-direction approach to break down the full problem into a sequence of two linear sub-problems: A cycle of the algorithm first reconstructs, for each source, the wavefield $\mathbf{u}$ that best fits the data and satisfies the wave equation in a least-squares sense for the current subsurface model. Then, the subsurface model is updated by minimization of the source residuals with a Gauss-Newton algorithm keeping the reconstructed wavefields fixed. A difficulty with the penalty method given by equation 3.11 resides in the tuning of the penalty parameter λ during iterations. An increasing values of λ should be used during iterations, known as the penalty algorithm, to progressively enforce the wave-equation constraint in iterations and hence satisfy the Karush-Kuhn-Tucker (KKT) optimality conditions (Nocedal and Wright, 2006) associated with the original constrained problem with acceptable precision. The main problem is that this continuation strategy is tedious to implement and a large λ makes the problem severely ill-conditioned.

IR-WRI augmented Lagrangian method

To bypass this difficulty, IR-WRI implements the original constrained problem, equation 3.8, with the augmented Lagrangian (AL) method (Hestenes, 1969; Nocedal and Wright, 2006; Boyd et al., 2010; Bertsekas, 2016).

$$\min_{\mathbf{u},\mathbf{m}} \max_{\mathbf{v}} \quad \mathbf{0} + \mathbf{v}^T[\mathbf{F}(\mathbf{m})\mathbf{u} - \mathbf{s}] + \frac{1}{2}\|\mathbf{F}(\mathbf{m})\mathbf{u} - \mathbf{s}\|^2_{\mathbf{\Lambda}}, \tag{3.12}$$

where $\mathbf{\Lambda}$ is defined as in equation 3.10 and $\mathbf{v} \in \mathbb{C}^{(M+N)\times 1}$ is the Lagrangian multiplier (known as dual variable). Comparing the penalty function, equation 3.9, and the augmented Lagrangian function, equation 3.44, clearly shows that the augmented Lagrangian combines a penalty method with a Lagrangian method. A first advantage of the augmented Lagrangian method relative to the penalty method is to prevent ill-conditioning by introducing explicit estimate of the Lagrange multiplier in the optimization (Nocedal and Wright, 2006, chapter 17). Moreover, the Lagrange multiplier gives the augmented Lagrangian method one more way of improving the accuracy of the minimizer in addition to the penalty parameter, hence allows fixed value to be used for this latter (Nocedal and Wright, 2006, Theorem 17.6). Applying the alternating direction strategy of WRI on the augmented Lagrangian method leads to an adaptation of the alternating-direction method of multiplier (ADMM) to biconvex problem (Boyd et al., 2010, Section 9.2). One ADMM iteration first minimizes the augmented Lagrangian function with respect to the primal variables $\mathbf{u}$ and $\mathbf{m}$ via a single Gauss-Seidel like iteration (namely, fix one variable and solve for the other) and then update the Lagrangian multiplier via a gradient ascent method.

In the following section, we review each step of the ADMM-based IR-WRI algorithm when equipped with TV regularization and bound constraints. The reader is also referred to Aghamiry et al. (2019c) for the detailed IR-WRI algorithm when no regularization is used.

BTV-regularized ADMM-based IR-WRI

To implement TV regularization and bound constraints in IR-WRI, we first recast FWI as a constrained TV minimization problem given by

$$\min_{\mathbf{u},\mathbf{m}\in\mathcal{C}} \quad \|\mathbf{m}\|_{\text{TV}} \tag{3.13a}$$

$$\text{subject to} \quad \mathbf{F}(\mathbf{m})\mathbf{u} = \mathbf{s}, \tag{3.13b}$$

where $\mathbf{F}$ and $\mathbf{s}$ are defined as in equation 3.6, $\|\mathbf{m}\|_{\text{TV}} = \sum \sqrt{|\nabla_1\mathbf{m}|^2 + |\nabla_2\mathbf{m}|^2}$ is the blockiness-promoting isotropic TV norm (Rudin et al., 1992), and ∇_1 and ∇_2 are first-order finite-difference operators in the horizontal and vertical directions, respectively. With notation abuse, the absolute sign, square power, and the square root operations are done component-wise, and the sum runs over all elements (the domain of parameters). Also $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^{N\times 1} \mid \mathbf{m}_l \le \mathbf{x} \le \mathbf{m}_u\}$ is the set of all feasible models bounded by the lower bound $\mathbf{m}_l$ and the upper bound $\mathbf{m}_u$.
Compared to the FWI definition given in the previous section, equation 3.8, we have replaced the identically-zero objective function by the TV norm of the model and restricted the space of

feasible models to $\mathcal{C}$. Accordingly, the augmented Lagrangian function for the problem defined by equation 3.13 is

$$\mathcal{L}_A(\mathbf{m},\mathbf{u},\mathbf{v}) = \|\mathbf{m}\|_{\text{TV}} + \mathbf{v}^T[\mathbf{F}(\mathbf{m})\mathbf{u}-\mathbf{s}] + \frac{1}{2}\|\mathbf{F}(\mathbf{m})\mathbf{u}-\mathbf{s}\|^2_{\mathbf{\Lambda}}, \tag{3.14}$$

where the same notations as those of equation 3.44 are used.
Equation 3.14 can also be written in a more compact form as

$$\mathcal{L}_A(\mathbf{m},\mathbf{u},\bar{\mathbf{s}}) = \|\mathbf{m}\|_{\text{TV}} + \frac{1}{2}\|\mathbf{F}(\mathbf{m})\mathbf{u}-\mathbf{s}-\bar{\mathbf{s}}\|^2_{\mathbf{\Lambda}} - \frac{1}{2}\|\bar{\mathbf{s}}\|^2_{\Lambda}, \tag{3.15}$$

where $\bar{\mathbf{s}} = -\mathbf{\Lambda}^{-1}\mathbf{v}$ is the scaled dual variable and equation 3.15 is the scaled form of the augmented Lagrangian (Boyd et al. (2010, Page 15) and Appendix A, section 3.2.7). The scaled form recasts the augmented Lagrangian function as a quadratic penalty function where the right-hand sides are updated with the scaled dual variables. This highlights similarities and differences between WRI and IR-WRI, since this right-hand side updating is lacking in the former.
The method of multipliers seeks to find the saddle point of the scaled augmented Lagrangian 3.15 through a primal descent - dual ascent updating resulting in the following iteration:

$$\mathbf{m}^{k+1},\mathbf{u}^{k+1} = \underset{\mathbf{u},\mathbf{m}\in\mathcal{C}}{\arg\min}\ \|\mathbf{m}\|_{\text{TV}} + \frac{1}{2}\|\mathbf{F}(\mathbf{m})\mathbf{u}-\mathbf{s}-\bar{\mathbf{s}}^k\|^2_{\mathbf{\Lambda}} \tag{3.16a}$$

$$\bar{\mathbf{s}}^{k+1} = \bar{\mathbf{s}}^k + \mathbf{s} - \mathbf{F}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}, \tag{3.16b}$$

for $k = 0, 1, ...$ beginning with a prior estimate $\bar{\mathbf{s}}^0 = \mathbf{0}$. The iteration 3.16 can be viewed as follows: we begin with a prior estimate of the dual $\bar{\mathbf{s}}^0$, and minimize the objective function with respect to the primal variables $\mathbf{m}$ and $\mathbf{u}$, equation 3.16a. Subsequently, we maximize the objective function with respect to the dual variable $\bar{\mathbf{s}}$ with a gradient ascent method when $\mathbf{m}$ and $\mathbf{u}$ are kept fixed, equation 3.16b. The steepest-ascent step, equation 3.16b, shows that the scaled dual variable $\bar{\mathbf{s}}$ is updated with the residual constraint violation of the current iteration. Remembering that the constraint combines the observation equation $\mathbf{Pu} = \mathbf{d}$ and the wave equation $A(\mathbf{m})\mathbf{u} = \mathbf{b}$, equation 3.6, the scaled dual variable $\bar{\mathbf{s}}$ updates the right-hand sides of the quadratic penalty function, equation 3.16a, with the running sum of the data and source residuals in iterations. This right-hand side updating describes the well-known iterative solution refinement procedure for ill-posed linear inverse problems as reviewed by Aghamiry et al. (2019c, Their appendix B). This process is iterated until convergence, i.e., when $\mathbf{F}(\mathbf{m}^{k+1})\mathbf{u}^{k+1} = \mathbf{s}$. In the following, we remove the bar of $\bar{\mathbf{s}}^k$ for the sake of simplicity.
Solving the subproblem 3.16a jointly for the primal variables ($\mathbf{m}$, $\mathbf{u}$) is computationally too intensive. A splitting method is useful here to break down the joint optimization over $\mathbf{m}$ and $\mathbf{u}$ into two subproblems (the readers can refer to Glowinski et al. (2017) for an overview of splitting methods). The alternating-direction method of multipliers (ADMM) (Boyd et al., 2010) provides a simple framework to achieve this goal via a single Gauss-Seidel like iteration leading to

the following iteration:

$$\mathbf{u}^{k+1} = \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{F}(\mathbf{m}^k)\mathbf{u} - \mathbf{s} - \mathbf{s}^k\|^2_{\mathbf{\Lambda}} \tag{3.17a}$$

$$\mathbf{m}^{k+1} = \arg\min_{\mathbf{m}\in\mathcal{C}} \|\mathbf{m}\|_{\text{TV}} + \frac{1}{2}\|\mathbf{F}(\mathbf{m})\mathbf{u}^{k+1} - \mathbf{s} - \mathbf{s}^k\|^2_{\mathbf{\Lambda}} \tag{3.17b}$$

$$\mathbf{s}^{k+1} = \mathbf{s}^k + \mathbf{s} - \mathbf{F}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}. \tag{3.17c}$$

The ADMM iteration has decomposed the full problem into two subproblems associated with primal variables $\mathbf{u}$ and $\mathbf{m}$ by passing the primal update of one subproblem as a passive variable for the next subproblem. We show below that, taking advantage of the bi-convexity of the problem, this alternating-direction approach has linearized the primal subproblem for $\mathbf{m}$ around the reconstructed wavefield $\mathbf{u}$. The fact that the primal update of one subproblem is passed to the next subproblem implies obviously that the two subproblems are solved in sequence rather than in parallel as in ADMM for linear separable problems.

A second modification related to ADMM resides in the updating of the dual variable. In ADMM, the dual variables are updated only once per iteration after the primal-variable updates, equation 3.17c. A variant of ADMM, referred to as the Peaceman-Rachford splitting method (PRSM) (Peaceman and Rachford, 1955), consists of updating the Lagrange multipliers several times, once after the update of each primal variable (see He et al. (2014, Compare equations 1.3 and 1.4). One issue with PRSM relative to ADMM is that PRSM requires more restrictive assumptions to ensure its convergence, while it is always faster than ADMM whenever it is convergent (He et al., 2014) . This issue prompted He et al. (2014) to implement a relaxation factor (or, step length) $\alpha \in (0, 1)$ to guarantee the strict contraction of the PRSM iterative sequence. Applying the strictly contractive PRSM algorithm to our minimization problem gives

$$\mathbf{u}^{k+1} = \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{F}(\mathbf{m}^k)\mathbf{u} - \mathbf{s} - \mathbf{s}^k\|^2_{\mathbf{\Lambda}} \tag{3.18a}$$

$$\mathbf{s}^{k+\frac{1}{2}} = \mathbf{s}^k + \alpha[\mathbf{s} - \mathbf{F}(\mathbf{m}^k)\mathbf{u}^{k+1}] \tag{3.18b}$$

$$\mathbf{m}^{k+1} = \arg\min_{\mathbf{m}\in\mathcal{C}} \|\mathbf{m}\|_{\text{TV}} + \frac{1}{2}\|\mathbf{F}(\mathbf{m})\mathbf{u}^{k+1} - \mathbf{s} - \mathbf{s}^{k+\frac{1}{2}}\|^2_{\mathbf{\Lambda}} \tag{3.18c}$$

$$\mathbf{s}^{k+1} = \mathbf{s}^{k+\frac{1}{2}} + \alpha[\mathbf{s} - \mathbf{F}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}]. \tag{3.18d}$$

The reader is referred to Aghamiry et al. (2019c, Their Figure 4) for a comparative numerical analysis of the convergence speed of ADMM and PRSM in IR-WRI. In our numerical tests we found that $\alpha = 0.5$ can serve as a suitable value.

We now provide the closed-form solution of the two primal subproblems associated with $\mathbf{u}$ and $\mathbf{m}$.

Solving for $\mathbf{u}$

The primal subproblem associated with $\mathbf{u}$, equation 3.18a, is a linear optimization problem whose solution satisfies in a least-squares sense the following system of linear equations:

$$\begin{bmatrix} \lambda_0^{\frac{1}{2}}\mathbf{P} \\ \lambda_1^{\frac{1}{2}}\mathbf{A}(\mathbf{m}^k) \end{bmatrix} \mathbf{u}^{k+1} = \begin{bmatrix} \lambda_0^{\frac{1}{2}}[\mathbf{d}+\mathbf{d}^k] \\ \lambda_1^{\frac{1}{2}}[\mathbf{b}+\mathbf{b}^k] \end{bmatrix}, \tag{3.19}$$

where $\mathbf{d}^k$ and $\mathbf{b}^k$ are the components of the dual variable $\mathbf{s}^k$ associated with the observation-equation and wave-equation constraints, and are formed by the running sum of the data and source residuals in iteration (see equation 3.16). The closed-form expression of the reconstructed wavefield is given by

$$\mathbf{u}^{k+1} = \left[\lambda_0\mathbf{P}^T\mathbf{P} + \lambda_1\mathbf{A}(\mathbf{m}^k)^T\mathbf{A}(\mathbf{m}^k)\right]^{-1}\left[\lambda_0\mathbf{P}^T[\mathbf{d}+\mathbf{d}^k] + \lambda_1\mathbf{A}(\mathbf{m}^k)^T[\mathbf{b}+\mathbf{b}^k]\right]. \tag{3.20}$$

The reconstructed wavefield can be computed numerically with linear algebra methods (direct or iterative methods) suitable for sparse matrices as reviewed by van Leeuwen and Herrmann (2016).

Solving for $\mathbf{m}$

In order to solve the bound-constrained TV regularized nonlinear problem described by equation 3.18c, we first tackle the nonlinearity issue by considering the special structure of the Helmholtz operator $\mathbf{A}$ given in equation 3.7. Using the following approximation

$$\begin{aligned} \mathbf{A}(\mathbf{m})\mathbf{u}^{k+1} &= \boldsymbol{\Delta}\mathbf{u}^{k+1} + \omega^2\mathbf{C}(\mathbf{m})\mathrm{diag}(\mathbf{m})\mathbf{B}\mathbf{u}^{k+1}, \\ &\approx \boldsymbol{\Delta}\mathbf{u}^{k+1} + \underbrace{\omega^2\mathbf{C}(\mathbf{m}^k)\mathrm{diag}(\mathbf{B}\mathbf{u}^{k+1})}_{\mathbf{L}(\mathbf{u}^{k+1})}\mathbf{m}, \end{aligned} \tag{3.21}$$

we linearise the operator $\mathbf{A}$ with respect to $\mathbf{m}$ by building the matrix $\mathbf{C}$ from $\mathbf{m}^k$ to manage potential nonlinear boundary conditions. Note that this linearization step is not necessary when the PML absorbing conditions are used. We also exploit the bilinearity of the wave equation to permute $\mathbf{B}\mathbf{u}$ and $\mathbf{m}$ in the operator $\mathbf{L}(\mathbf{u}^{k+1})$. With these manipulations, the sub-problem associated with $\mathbf{m}$ is recast as

$$\mathbf{m}^{k+1} = \underset{\mathbf{m}\in\mathcal{C}}{\arg\min}\ \|\mathbf{m}\|_{\mathrm{TV}} + \frac{\lambda_1}{2}\|\mathbf{L}(\mathbf{u}^{k+1})\mathbf{m} - \mathbf{y}^k\|_2^2. \tag{3.22}$$

where

$$\mathbf{y}^k = \mathbf{b} + \mathbf{b}^k - \boldsymbol{\Delta}\mathbf{u}^{k+1}. \tag{3.23}$$

We solve the constrained optimization problem described by equation 3.22 with the split Bregman method (Goldstein and Osher, 2009) to decouple the TV minimization subproblem (first term in equation 3.22) from the ℓ_2 subproblem (second term in equation 3.22) and force the box constraint. We introduce the auxiliary variables $\mathbf{p}_0$, $\mathbf{p}_1$ and $\mathbf{p}_2$ to perform this splitting, in

which these variables being related to $\mathbf{m}$ by means of a simple equality constraint. The auxiliary variable $\mathbf{p}_0$ is used to enforce the box constraint, while $\mathbf{p}_1$ and $\mathbf{p}_2$ are the variables of the TV minimization problem. Let us define $(\mathbf{p}_1\ \mathbf{p}_2)$ as a two-column matrix, $\|(\mathbf{p}_1\ \mathbf{p}_2)\| = \sqrt{\mathbf{p}_1^2 + \mathbf{p}_2^2}$ be a vector which contains the ℓ_2 norm of each row of $(\mathbf{p}_1\ \mathbf{p}_2)$, and $\sum \|(\mathbf{p}_1\ \mathbf{p}_2)\|$ be the mixed $\ell_{2,1}$ norm (ℓ_1 norm of $\|(\mathbf{p}_1\ \mathbf{p}_2)\|$), which promotes sparsity. Then by defining the objective function J as

$$J(\mathbf{m}, \mathbf{p}) = \sum \|(\mathbf{p}_1\ \mathbf{p}_2)\| + \frac{\lambda_1}{2}\|\mathbf{L}(\mathbf{u}^{k+1})\mathbf{m} - \mathbf{y}^k\|_2^2, \tag{3.24}$$

the unconstrained optimization problem described by equation 3.22 can be written in a split and constrained form as

$$\begin{aligned} &\underset{\mathbf{m},\mathbf{p},\mathbf{p}_0 \in \mathcal{C}}{\arg\min} && J(\mathbf{m}, \mathbf{p}) && \text{(3.25a)} \\ &\text{subject to} && \mathbf{p} = \nabla\mathbf{m}, && \text{(3.25b)} \end{aligned}$$

where

$$\nabla = \begin{bmatrix} \mathbf{I} \\ \nabla_1 \\ \nabla_2 \end{bmatrix} \in \mathbb{R}^{3N \times N}, \qquad \mathbf{p} = \begin{bmatrix} \mathbf{p}_0 \\ \mathbf{p}_1 \\ \mathbf{p}_2 \end{bmatrix} \in \mathbb{R}^{3N \times 1},$$

and $\mathbf{I}$ is the identity matrix.

The scaled augmented Lagrangian function (Appendix A, section 3.2.7) for the problem defined by equation 3.25 is

$$\mathcal{L}_A(\mathbf{m}, \mathbf{p}, \bar{\mathbf{q}}) = J(\mathbf{m}, \mathbf{p}) + \frac{1}{2}\|\nabla\mathbf{m} - \mathbf{p} - \bar{\mathbf{q}}^k\|_{\mathbf{\Gamma}}^2 - \frac{1}{2}\|\bar{\mathbf{q}}^k\|_{\mathbf{\Gamma}}^2, \tag{3.26}$$

where $\bar{\mathbf{q}}$ is the scaled Lagrangian multipliers and

$$\mathbf{\Gamma} = \left[\begin{array}{c|c|c} \begin{matrix} \gamma_0 & & 0 \\ & \ddots & \\ 0 & & \gamma_0 \end{matrix} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \begin{matrix} \gamma_1 & & 0 \\ & \ddots & \\ 0 & & \gamma_1 \end{matrix} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \begin{matrix} \gamma_2 & & 0 \\ & \ddots & \\ 0 & & \gamma_2 \end{matrix} \end{array}\right] \begin{matrix} N \\ N, \\ N \end{matrix} \tag{3.27}$$

with the penalty parameters $\gamma_0, \gamma_1, \gamma_2 > 0$. Again, applying the method of multipliers to find the saddle point of the problem 3.26 gives

$$\begin{aligned} \mathbf{m}^{k+1}, \mathbf{p}^{k+1} &= \underset{\mathbf{m},\mathbf{p},\mathbf{p}_0 \in \mathcal{C}}{\arg\min}\ J(\mathbf{m}, \mathbf{p}) + \frac{1}{2}\|\nabla\mathbf{m} - \mathbf{p} - \mathbf{q}^k\|_{\mathbf{\Gamma}}^2 && \text{(3.28a)} \\ \mathbf{q}^{k+1} &= \mathbf{q}^k + \mathbf{p}^{k+1} - \nabla\mathbf{m}^{k+1}. && \text{(3.28b)} \end{aligned}$$

where the bar of $\bar{\mathbf{q}}^k$ is removed for simplicity.
Substituting the explicit expression of J, equation 3.24, into equation 3.28a leads to the follow-

ing PRSM iteration:

$$\mathbf{m}^{k+1} = \arg\min_{\mathbf{m}} \frac{\lambda_1}{2}\|\mathbf{L}(\mathbf{u}^{k+1})\mathbf{m} - \mathbf{y}^k\|_2^2 + \frac{1}{2}\|\nabla\mathbf{m} - \mathbf{p}^k - \mathbf{q}^k\|_{\mathbf{\Gamma}}^2 \tag{3.29a}$$

$$\mathbf{q}^{k+\frac{1}{2}} = \mathbf{q}^k + \frac{1}{2}[\mathbf{p}^k - \nabla\mathbf{m}^{k+1}]. \tag{3.29b}$$

$$\mathbf{p}^{k+1} = \arg\min_{\mathbf{p}_0\in\mathcal{C},(\mathbf{p}_1,\mathbf{p}_2)} \sum \|(\mathbf{p}_1\ \mathbf{p}_2)\| + \frac{1}{2}\|\nabla\mathbf{m}^{k+1} - \mathbf{p} - \mathbf{q}^{k+\frac{1}{2}}\|_{\mathbf{\Gamma}}^2 \tag{3.29c}$$

$$\mathbf{q}^{k+1} = \mathbf{q}^{k+\frac{1}{2}} + \frac{1}{2}[\mathbf{p}^{k+1} - \nabla\mathbf{m}^{k+1}]. \tag{3.29d}$$

Note that the weight $\frac{1}{2}$ in equations 3.29b and 3.29d has a similar role as α in equations 3.18b and 3.18d. Now we come up with a linear inverse subproblem for $\mathbf{m}$, equation 3.29a. Accordingly, the update $\mathbf{m}^{k+1}$ is obtained by solving the following system of linear equations in a least-squares sense:

$$\begin{bmatrix} \lambda_1^{\frac{1}{2}}\mathbf{L}(\mathbf{u}^{k+1}) \\ \mathbf{\Gamma}^{\frac{1}{2}}\nabla \end{bmatrix} \mathbf{m}^{k+1} = \begin{bmatrix} \lambda_1^{\frac{1}{2}}[\mathbf{b} + \mathbf{b}^k - \mathbf{\Delta}\mathbf{u}^{k+1}] \\ \mathbf{\Gamma}^{\frac{1}{2}}[\mathbf{p}^k + \mathbf{q}^k] \end{bmatrix}, \tag{3.30}$$

where we have substituted $\mathbf{y}^k$ by its explicit expression, equation 3.23. In equation 3.30, the first line describes the information carried out by the reconstructed wavefield to update $\mathbf{m}$ via the wave-equation rewriting, while the second line describes the action of the TV regularization and bound constraints on $\mathbf{m}$ via its linear relation with the auxiliary variable $\mathbf{p}$.
The closed-form expression of $\mathbf{m}$ is given by

$$\mathbf{m}^{k+1} = \left[\lambda_1\mathbf{L}(\mathbf{u}^{k+1})^T\mathbf{L}(\mathbf{u}^{k+1}) + \nabla^T\mathbf{\Gamma}\nabla\right]^{-1}\left[\lambda_1\mathbf{L}(\mathbf{u}^{k+1})^T[\mathbf{b} + \mathbf{b}^k - \mathbf{\Delta}\mathbf{u}^{k+1}] + \nabla^T\mathbf{\Gamma}[\mathbf{p}^k + \mathbf{q}^k]\right]. \tag{3.31}$$

As for the linear system 3.20, $\mathbf{m}$ can be computed numerically with any suitable sparse linear algebra method.
The sub-problem for $\mathbf{p}$, equation 3.29c, is straightforward to solve. The objective function is separable with respect to the variable $\mathbf{p}_0$ and the variables $\mathbf{p}_1$ and $\mathbf{p}_2$ (i.e., the optimization can be performed for $\mathbf{p}_0$ and $\mathbf{p}_1$, $\mathbf{p}_2$ separately). The variable $\mathbf{p}_0$ is solution of the following linear inverse problem

$$\mathbf{p}_0^{k+1} = \arg\min_{\mathbf{p}_0\in\mathcal{C}} \frac{\gamma_0}{2}\|\mathbf{m}^{k+1} - \mathbf{p}_0 - \mathbf{q}_0^k\|_2^2. \tag{3.32}$$

The ith element of the solution, $\mathbf{p}_0^{k+1}(i)$, is the closest element of $\mathbf{m}^{k+1}(i) - \mathbf{q}_0^k(i)$ to the desired set $[\mathbf{m}_l(i), \mathbf{m}_u(i)]$. Therefore,

$$\mathbf{p}_0^{k+1} = \text{proj}_{\mathcal{C}}(\mathbf{m}^{k+1} - \mathbf{q}_0^k), \tag{3.33}$$

where the projection operator is $\text{proj}_{\mathcal{C}}(\bullet) = \min(\max(\bullet, \mathbf{m}_l), \mathbf{m}_u)$.
The variables $\mathbf{p}_1$ and $\mathbf{p}_2$ are updated via the following proximity operator:

$$\mathbf{p}_1^{k+1}, \mathbf{p}_2^{k+1} = \arg\min_{(\mathbf{p}_1\ \mathbf{p}_2)} \sum \|(\mathbf{p}_1\ \mathbf{p}_2)\| + \frac{\gamma}{2}\|(\mathbf{p}_1\ \mathbf{p}_2) - (\mathbf{z}_1\ \mathbf{z}_2)\|_2^2, \tag{3.34}$$

where $\gamma = \gamma_1 = \gamma_2$, and $\mathbf{z}_i = \nabla_i \mathbf{m}^{k+1} - \mathbf{q}_i^k \; i = 1, 2$. Proximity operators are generalization of projection operators (Combettes and Pesquet, 2011). Equation 3.34 describes a separable optimization problem with respect to $\mathbf{p}_1$ and $\mathbf{p}_2$. Furthermore, $\mathbf{p}_1$ and $\mathbf{p}_2$ have closed-form expressions (Goldstein and Osher, 2009)

$$\mathbf{p}_i^{k+1} = \text{prox}_\gamma(\mathbf{z}_i), \quad i = 1, 2, \tag{3.35}$$

where

$$\text{prox}_\gamma(\mathbf{z}_i) = \frac{\mathbf{z}_i}{\|(\mathbf{z}_1 \; \mathbf{z}_2)\|} \max(\|(\mathbf{z}_1 \; \mathbf{z}_2)\| - \gamma, 0). \tag{3.36}$$

It can be seen that, for a single vector, the proximity operator defined in equation 3.36 reduces to soft thresholding.

Considering all the above-mentioned processes, a pseudocode for the BTV-regularized IR-WRI algorithm is summarized in Algorithm 2. Note that the lines 6-11 of the algorithm correspond to one ADMM iteration of the model-parameter updating. These operations could be iterated in an inner loop to update the BTV-regularized model several times after the wavefield reconstruction at each outer iteration. However, we observed numerically that only a single iteration of the inner loop guarantees the most efficient convergence of the full algorithm. This property has been noticed by Goldstein and Osher (2009) and is discussed more extensively in the framework of IR-WRI by Aghamiry et al. (2019c). The inefficiency of the inner iterations can be understood by the fact that the original nonlinear problem is solved with an alternating-direction strategy (managed by the outer loop). This implies that each subproblem is solved from potentially inaccurate passive variables, this inaccuracy preventing an efficient minimization of the objective during inner iterations. Furthermore, in order to drive the algorithm, we assumed that the constraint is feasible. However, it has been shown that in the case of infeasible linear constraints the ADMM iteration can still produce approximate solutions that are stable (Frick et al., 2011; Jiao et al., 2016).

Algorithm 2: BTV regularized IR-WRI algorithm based on the PRS algorithm.

Initialize: set the RHS errors $k = 0, \mathbf{s}^0 = \mathbf{0}, \mathbf{q}^0 = \mathbf{0}$
Input: $\mathbf{m}^0$ (initial model parameters)
while *convergence criteria not satisfied* **do**
- $\mathbf{u}^{k+1} \leftarrow$ update according to 3.20
- $\mathbf{s}^{k+\frac{1}{2}} \leftarrow \mathbf{s}^k + 0.5[\mathbf{s} - \mathbf{F}(\mathbf{m}^k)\mathbf{u}^{k+1}]$
- $\mathbf{m}^{k+1} \leftarrow$ update according to 3.30
- $\mathbf{q}^{k+\frac{1}{2}} \leftarrow \mathbf{q}^k + 0.5[\mathbf{p}^k - \nabla \mathbf{m}^{k+1}]$
- $\mathbf{p}_0^{k+1} \leftarrow \text{proj}_{\mathcal{C}}(\mathbf{m}^{k+1} - \mathbf{q}_0^{k+\frac{1}{2}})$
- $\mathbf{p}_1^{k+1} \leftarrow \text{prox}_\gamma(\nabla_1 \mathbf{m}^{k+1} - \mathbf{q}_1^{k+\frac{1}{2}})$
- $\mathbf{p}_2^{k+1} \leftarrow \text{prox}_\gamma(\nabla_2 \mathbf{m}^{k+1} - \mathbf{q}_2^{k+\frac{1}{2}})$
- $\mathbf{q}^{k+1} \leftarrow \mathbf{q}^{k+\frac{1}{2}} + 0.5[\mathbf{p}^{k+1} - \nabla \mathbf{m}^{k+1}]$
- $\mathbf{s}^{k+1} \leftarrow \mathbf{s}^{k+\frac{1}{2}} + 0.5[\mathbf{s} - \mathbf{F}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}]$
- $k \leftarrow k + 1$

3.2.4 Numerical examples

Experimental setup and parameter tuning

We assess the performance of our BTV regularized IR-WRI against 2D mono-parameter synthetic examples. We start with a toy example built with a high-velocity inclusion model that is embedded in a background medium where velocity linearly increases with depth . To tackle more realistic applications, we proceed with two scaled targets of the challenging 2004 BP salt model (Billette and Brandsberg-Dahl, 2004). With the BP salt case study, we seek to illustrate the potential of IR-WRI equipped with our BTV regularization to image salt bodies and sub-salt structures starting from crude initial models and realistic frequencies.
For all the numerical examples, forward modelling is performed with a 9-point stencil implemented with anti-lumped mass and PML absorbing boundary conditions (Chen et al., 2013). In this setting the diagonal matrix $\mathbf{C}$ contains the damping PML coefficients and does not depend on $\mathbf{m}$. With this setting, equation 3.21 does not require any approximation for linearization.
We will compare the results of WRI and IR-WRI to highlight the improved convergence history of IR-WRI resulting from the iterative updating of the right-hand sides in the penalty function associated with the scaled-form augmented Lagrangian, equations 3.15-3.18. We assume that our IR-WRI algorithm, when this right-hand side updating is not activated, is representative of the WRI penalty method (van Leeuwen and Herrmann, 2013). For a fair comparison, we will use the same experimental setup (penalty parameters and stopping criterion of iteration) for the two methods. We also compare the WRI and IR-WRI results when they are obtained without any priors ($\gamma_0 = \gamma = 0$), with only bound constraints ($\gamma = 0$) and with BTV regularization ($\gamma_0 \neq 0$ and $\gamma \neq 0$), where it is reminded that γ_0 and γ are the penalty parameters that control the weight of the bound constraints and TV regularization, respectively, in the objective function (see equation 3.27).
We tune the different penalty parameters according to the following guideline. We start from the last subproblem of the splitting procedure and set the parameter γ, which controls the soft thresholding performed by the TV regularization, equation 3.35-3.36. In this study, we find that $\gamma = 2\% \max \|(\mathbf{z}_1\ \mathbf{z}_2)\|$ was a good pragmatical value. This tuning can be refined according to prior knowledge of the geological structure, coming from well logs for example. In this study, we use the same weight for the bound constraints and the TV regularization: $\gamma_0 = \gamma$. Once we set γ, we define λ_1 such that γ/λ_1 is a percentage of the mean absolute value of the diagonal coefficients of $\mathbf{L}^T\mathbf{L}$ during the parameter estimation subproblem, equation 3.31. This percentage is set according to the weight that we want to assign to the TV regularization and the bound constraints relative to the wave equation constraint during the parameter estimation. Parameter λ_1 may be increased during iterations to reduce the weight of TV regularization and bound constraints near the convergence point. We found this adaptation useful when we start from very crude initial models. Finally, we set λ_0 such that $\lambda = \lambda_1/\lambda_0$ is a small fraction of the highest eigenvalue ξ of the normal operator $\mathbf{A}(\mathbf{m})^{-T}\mathbf{P}^T\mathbf{P}\mathbf{A}(\mathbf{m})^{-1}$ during the wavefield reconstruction subproblem, equation 3.20, according to the criterion proposed by van Leeuwen and Herrmann (2016). In all the numerical tests, we use $\lambda =$ 1e-5ξ and $\lambda =$ 1e-3ξ for noiseless and noisy data, respectively. This tuning of λ is indeed important because it controls the extension of the search space. A too high value of λ reduces the weight of $\|\mathbf{P}\mathbf{u} - \mathbf{d}\|_2^2$ during the wavefield reconstruction and makes IR-WRI behave like a reduced approach. Conversely, using a small

value for λ fosters data fitting and expends the search space accordingly. However, a too small value can lead to a prohibitively high number of iterations of the augmented Lagrangian method before the wave equation constraint is fulfilled with sufficient accuracy. Moreover, when data are contaminated by noise, a too small value for λ will make the wavefield reconstruction overfit the data and drive WRI to be a poor minimizer. We always use λ as a fixed percentage of ξ in iterations for both WRI and IR-WRI. This does not prevent IR-WRI to converge towards accurate minimizers thanks to the iterative error correction performed by the Lagrange multiplier updating. The reader is referred to Aghamiry et al. (2019c) for a more thorough sensitivity analysis of IR-WRI to the penalty parameter λ.

Inclusion model

The subsurface model contains a sharp box-shape anomaly of sides 0.2 $\times$ 0.3 km with a velocity (V_P) of 5 km/s. It is embedded in a smooth background model where V_P increases linearly with depth from 1.5 to 3.5 km/s (Figure 3.1). The model is 1.5 km long and 1 km deep, and is discretized with a 10 m grid interval. The regular surface acquisition consists of five sources (as depicted with yellow stars in Figure 3.1) and 65 receivers deployed on the surface. The source signature is a Ricker wavelet with a 5 Hz dominant frequency. We start the inversion from the true background model and invert simultaneously three frequency components (2.5, 5 and 7 Hz) with noiseless data. This frequency bandwidth has been selected to cover a significant band of vertical wavenumbers in the waveform-inversion sensitivity kernels, considering the limited aperture illumination provided by the surface acquisition. Moreover, a realistic starting frequency of 2.5Hz allows us to assess the resilience to cycle skipping of IR-WRI. A maximum number of iterations set to 70 is used as a stopping criterion for all of the tests shown in Figure 3.2. When bound constraints are used, the bounds $\mathbf{m}_l$ and $\mathbf{m}_u$, equation 3.33, are set to the true minimum and maximum square slownesses, respectively.

We first compare WRI and IR-WRI results when bound constraints and TV regularization are not activated, i.e. $\gamma_0 = \gamma = 0$ (Figure 3.2(a-b)). WRI and IR-WRI reconstruct only the top of the anomaly with strongly overestimated velocities. Then, we add bound constraints in WRI and IR-WRI, using $\gamma_0/\lambda_1 = 0.01\zeta$, where ζ is the mean absolute value of the diagonal coefficients of $\mathbf{L}^T\mathbf{L}$. The bound-constrained WRI and IR-WRI only reconstruct the top of the anomaly as in Figure 3.2(a-b). However, the inclusion velocities are now well controlled by the bound constraints (Figure 3.2(c-d)). These first two tests show that IR-WRI reconstructs better the shape of the anomaly than WRI with however more significant artifacts on both sides of the anomaly. These artifacts may have resulted from the deficit in horizontal-wavenumber illumination provided by the sparse limited-offset surface acquisition and by multi-scattering pollutions.
Then, we apply BTV regularization with $\gamma/\lambda_1 = \gamma_0/\lambda_1 = 0.01\zeta$ (Figure 3.2(e-f)). Since the initial model matches the true velocity-gradient background model, we use a small value of γ/λ_1 (i.e., a high value of λ_1) and keep it constant in iterations to preserve the smooth components of the subsurface model. We show that the BTV regularized WRI still fails to reconstruct the full anomaly (Figure 3.2e). In contrast, BTV regularized IR-WRI keeps on improving the reconstruction of the anomaly in depth, while efficiently mitigating the oscillating artifacts (Figure 3.2f).

A vertical profile across the reconstructed anomaly also highlights some limitations of the BTV regularization (Figure 3.2f): below the anomaly, the BTV regularization superimposes staircase artifacts on the velocity gradient, consistently with the piecewise constant assumption underlying TV regularization.

To emphasize the resilience to cycle skipping of the BTV regularized WRI and IR-WRI, we repeat this toy example using a 2.2 km/s homogeneous velocity model as initial model (Figure 3.3). We perform a first test without any prior. Compared to the previous test, we just stabilize the inversion by adding a small damping term (=0.01ζ) to $\mathbf{L}^T\mathbf{L}$. Compared to Figure 3.2(a-b), the artifacts have a much stronger imprint due to the inaccuracy of the starting model (Figure 3.3(a-b)). Then, we move to bound-constrained and TV regularized tests (Figure 3.3(c-f)). To decrease the above-mentioned artifacts, we assign a high initial weight to the BTV regularization ($\gamma/\lambda_1 = \gamma_0/\lambda_1 = \zeta$) and decrease it by a factor 2 every 10 iterations until it reach a minimal value set to 0.01ζ (i.e., the constant value previously used). In accordance with the former test, bound constraints alone are not sufficient to reconstruct the bottom part of the anomaly and cancel out the oscillating artifacts (Figure 3.3(c-d)). In contrast, BTV IR-WRI achieves these two goals, although it leaves a significant staircase footprint below the anomaly (Figure 3.3f). Compared to Figure 3.2f, the edges of the anomaly are better reconstructed at the expense of the background velocity-gradient model. This results, because of the more aggressive TV regularization was used during the early iterations of this test, allowing for a better reconstruction of the blocky components of the medium, while injecting undesired staircase footprint on its smooth components. As for the former test, IR-WRI clearly outperforms WRI due to the more efficient solution refinement procedure resulting from the right-hand side updating.

We also show the joint evolution in iterations of the observation-equation and wave-equation errors (Figure 3.4) and the wavefield and subsurface model errors (Figure 3.5), when the initial model is the true velocity-gradient background model and the homogeneous model. IR-WRI fits the data and wave equation better than WRI after 70 iterations with both initial models because the right-hand side updating embedded in IR-WRI cancels out more efficiently the data and source residuals in iterations and, hence better refines the solution accordingly. Moreover, BTV regularization in IR-WRI further improves the data and wave equation fit for both initial models because it reduces more efficiently the oscillating artifacts in the reconstructed velocity model

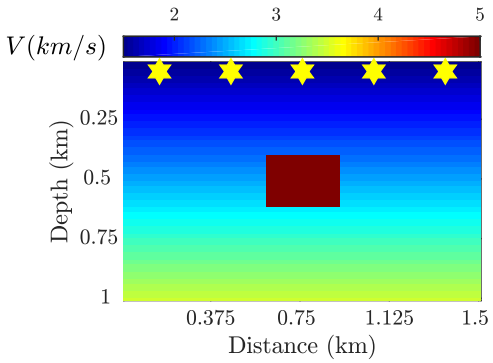

Figure 3.1 – The true velocity of the box-shaped anomaly example. The yellow stars show the source positions.

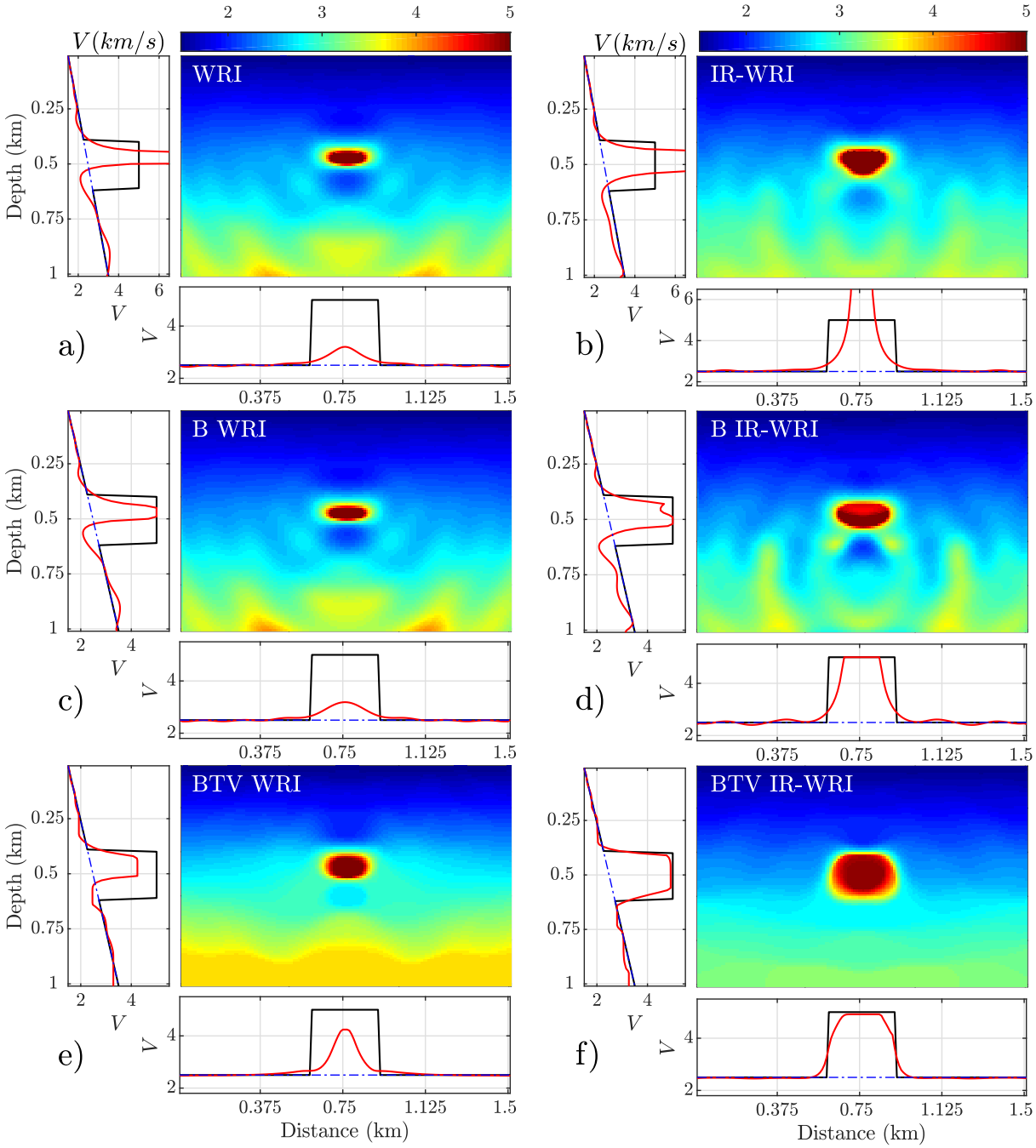

Figure 3.2 – A box-shaped anomaly example with the true background velocity gradient as the initial model. Velocity models reconstructed by (a) WRI. (b) IR-WRI. (c-d) Bound constrained WRI (c) and IR-WRI (d). (e-f) BTV regularized WRI (e) and IR-WRI (f). Horizontal and vertical profiles across the center of the inclusion from the true (black), initial (dash blue) and reconstructed (red) models are shown below and on the left-hand side of the models.

and better reconstructs the edges of the anomaly (Figure 3.2f). As above mentioned, the oscillating artifacts may result from the deficit of horizontal wavenumber illumination generated by the

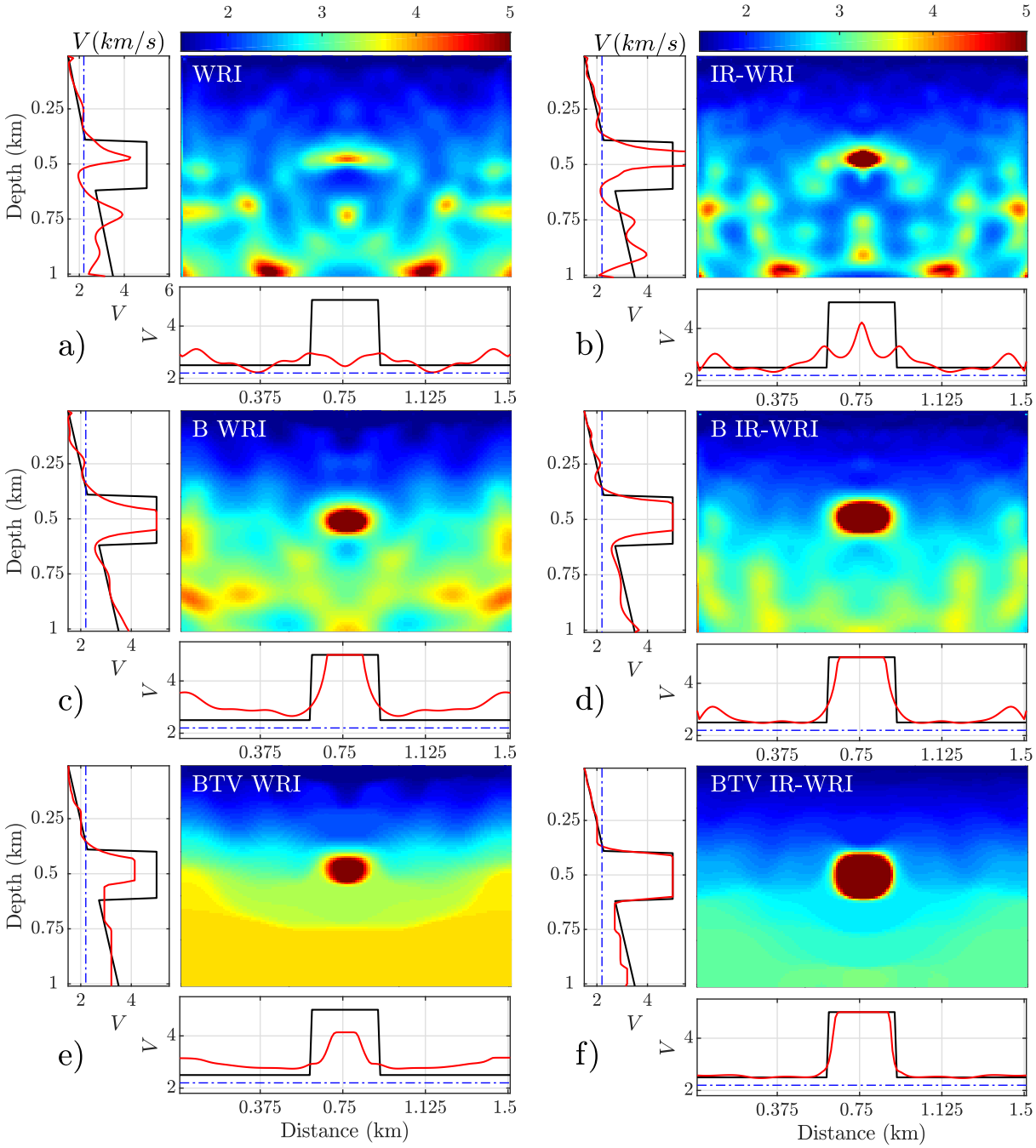

Figure 3.3 – The box-shape anomaly example. Same as Figure 3.2 for a homogeneous initial velocity model (V_P=2.2km/s).

limited-offset surface acquisition. In this framework, the prior contained in the BTV regularization efficiently narrows the null space of the inversion. The more complex zigzag path followed by IR-WRI relative to WRI in the $(\|\mathbf{P}\mathbf{u}^k - \mathbf{d}\|_2 - \|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b}\|_2)$ plane highlights how the joint updating of the data and source by their associated residuals dynamically balances the weight of the two objective functions in iterations. This more complex convergence history of

IR-WRI, which has been already noticed in Aghamiry et al. (2019c), suggests that the right-hand side updating perform a self-adaptive weighting of the two competing objective functions driven by the relative reduction of the data and source residuals in iterations. This zigzag convergence trend translates also into more complex path in the $(\|\mathbf{u}^k-\mathbf{u}^*\|_2/\|\mathbf{u}^*\|_2-\|\mathbf{m}^k-\mathbf{m}^*\|_2/\|\mathbf{m}^*\|_2)$ plane ($\mathbf{m}^*$ and $\mathbf{u}^*$ denote the true model and wavefields, respectively), which illustrates how the solution refinement is pushed toward the wavefield reconstruction or the velocity model estimation according to the self-adaptive weighting of the observation-equation and wave-equation objective functions (Figure 3.5). Note that the relative model errors increase in the case of the initial velocity-gradient model (Figure 3.5a,b). This results from the fact that the smooth background model is degraded by the oscillating artifacts and the staircase footprint during the sharp inclusion reconstruction. Indeed, this degradation of the smooth components has a much higher weight in the ℓ_2 misfit function than the more accurate reconstruction of the blocky components. However, this increase is much more moderate in IR-WRI (Figure 3.5b) than in WRI (Figure 3.5a). Finally, we plot the TV norm of the reconstructed models in iterations for the different tests (Figure 3.6). As expected, the models reconstructed with bound constraints and TV regularization match better the TV of the true model. When the homogeneous initial model is used, BTV IR-WRI smoothly converges to the TV of the true model after iteration 20 (when bound constraints are activated), while the weight of the TV regularization is progressively decreased (Figure 3.6b, blue curve). The BTV WRI model matches slightly better the TV of the true model than the BTV IR-WRI one when the initial model is the true background model. Indeed, this does not reflect that WRI better reconstructs the anomaly than IR-WRI. Instead, it reflects the slower convergence of WRI relative to IR-WRI which contributes to keep the background model smooth (namely, which a TV close to that of the true background model).

2004 BP salt model - central target

We now consider a more realistic application with a first target of the challenging 2004 BP salt model. The 2004 BP salt model is representative of the geology of the deep water offshore Gulf of Mexico and mainly consists of a simple background with a complex rugose multi-valued salt body, sub-salt slow velocity anomalies related to over-pressure zones and a fast velocity anomaly to the right of the salt body (Billette and Brandsberg-Dahl, 2004). The first selected target corresponds to the central part of the 2004 BP salt model characterized by a deeply rooted salt body (Figure 3.7a). Note that we rescale the spatial dimensions and the sampling of the original target for sake of computational efficiency.
Accordingly, our subsurface model is 8.8 km wide and 2.9 km deep, and is discretized with a 25 m grid interval. We used 50 sources spaced 175 m apart on the top side of the model. The source signature is a Ricker wavelet with a 10 Hz dominant frequency. A line of receivers with a 50 m spacing are deployed at the surface leading to a stationary-receiver acquisition.
We start the inversion from a smoothed version of true velocity model where the imprint of the salt body and other structures were cancelled out (Figure 3.7b) and invert the 3-Hz frequency with noiseless data.
We compare the results of WRI and IR-WRI with bound constraints and BTV regularization. To highlight the specific role of bound constraints, we activate them after 21 iterations. Since we start from a rough initial model, we set $\gamma_0/\lambda_1=\gamma/\lambda_1=\zeta$ and decrease them during iterations

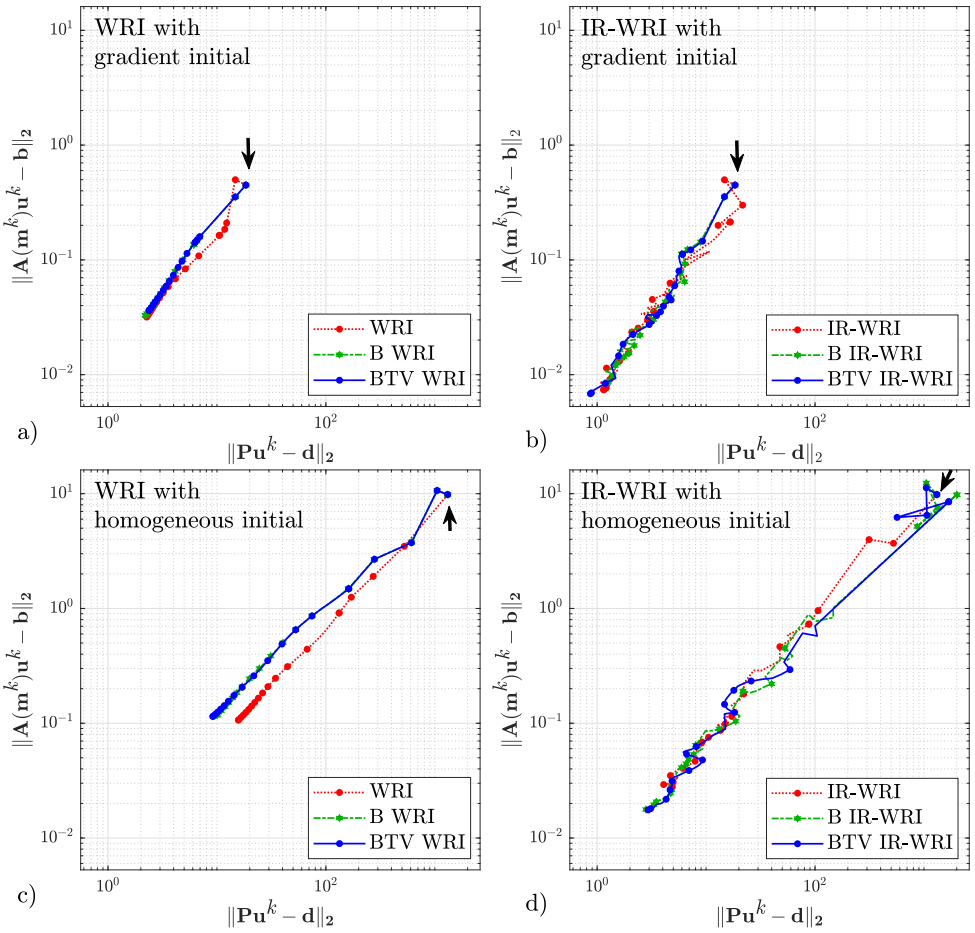

Figure 3.4 – The box-shape anomaly example. Convergence history in the ($\|\mathbf{P}\mathbf{u}^k - \mathbf{d}\|_2 - \|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b}\|_2$) plane of WRI and IR-WRI without priors, with bound constraints and with BTV regularizations. (a,c) WRI. (b,d) IR-WRI. Initial model is (a-b) the true velocity-gradient background model, and (c-d) the homogeneous velocity model. All the panels are plotted with the same horizontal and vertical logarithmic scale. The black arrow points the starting point.

in a manner similar to the box-shape anomaly test with a homogeneous starting model. Also, we add a damping term to $\mathbf{L}^T\mathbf{L}$ with a weight equal to 0.01ζ to further stabilize the inversion. We stop inversion after 70 iterations. The estimated models are shown in Figure 3.7 together with horizontal profiles at 1.65 km depth and vertical profiles at 4.15 km distance extracted from the true, initial and reconstructed models. As for the inclusion test, WRI fails to reconstruct the salt body and the subsalt structure because the data and source residuals are not re-injected in the right-hand sides of the penalty function at each iteration as in equation 3.17, leading to a stagnant convergence of the inversion (Figure 3.7c,e). When IR-WRI is applied with bound constraints alone, the reconstructed model is affected by noise with a periodic horizontal pattern. This noise likely results from the monochromatic nature of the inversion, multi-scattering within the salt body and limited illumination of the horizontal wavenumbers of the salt body leading to wraparound (Figure 3.7d). The BTV regularized IR-WRI mitigates efficiently this noise without degrading the resolution of the salt body and the sub-salt structures (Figure 3.7f).

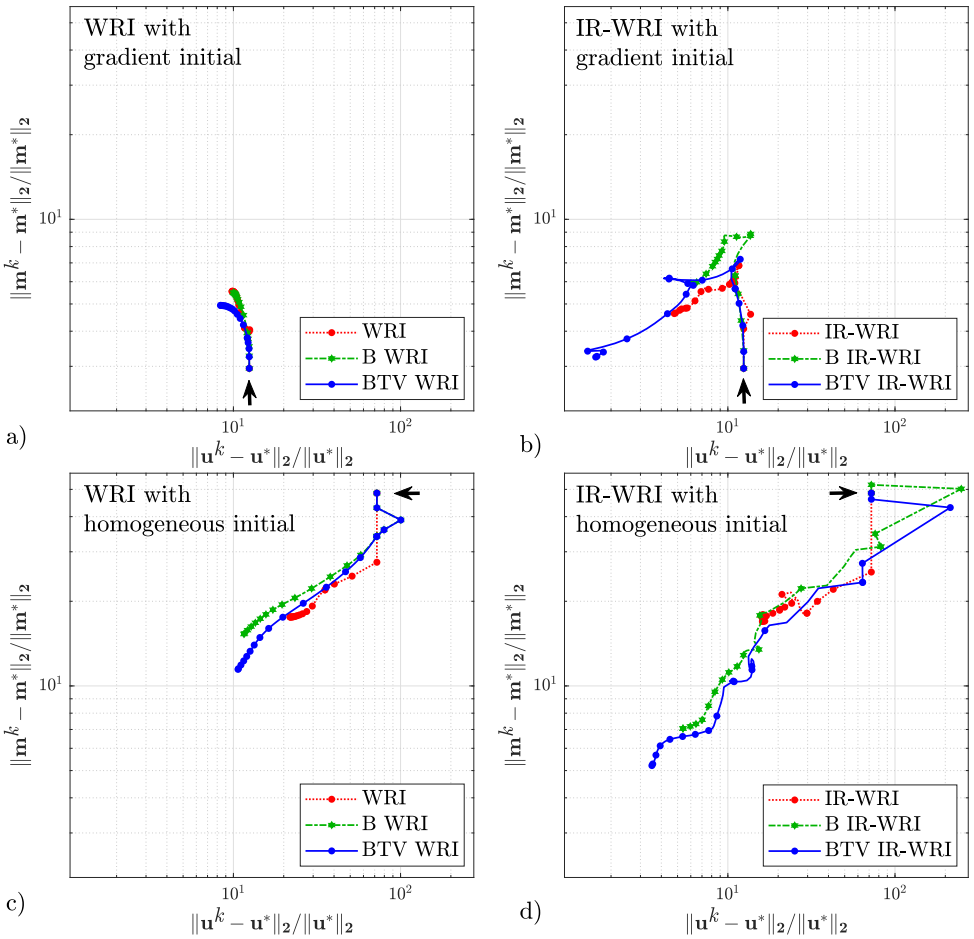

Figure 3.5 – The box-shaped anomaly example. Convergence history in the ($\|\mathbf{u}^k - \mathbf{u}^*\|_2/\|\mathbf{u}^*\|_2 - \|\mathbf{m}^k - \mathbf{m}^*\|_2/\|\mathbf{m}^*\|_2$) plane of WRI and IR-WRI without priors, with bound constraints and with BTV regularization and for the two initial models. The wavefield misfit function $\|\mathbf{u}^k - \mathbf{u}^*\|_2/\|\mathbf{u}^*\|_2$ is computed by summation over all the shots. (a,c) WRI. (b,d) IR-WRI. The initial model is (a-b) the true velocity-gradient background model, and (c-d) the homogeneous velocity model. Note the increase of $\|\mathbf{m}^k - \mathbf{m}^*\|_2/\|\mathbf{m}^*\|_2$ over iterations in (a-b) (see text for explanations). All the panels are plotted with the same horizontal and vertical logarithmic scale. The black arrow indicates the starting point.

Figure 3.8(a) shows the joint evolution in iterations of the data misfit and the wave-equation error. As for the inclusion test, note the zigzag path followed by the IR-WRI objective functions over iterations. Also, the joint evolution of wavefield and subsurface model errors and the evolution of TV norm over iteration are shown in Figure 3.8(b,c). The TV norm evolution emphasizes how the bound constraints fasten the convergence of TV-regularized IR-WRI after iteration 20.

We continue the inversion at higher frequencies using the final models of the 3 Hz inversion as initial models (Figure 3.7c-f). We used small batches of two frequencies with one frequency overlapping between two consecutive batches, moving from the low frequencies to the higher ones according to a classical frequency continuation strategy. We set $\gamma_0/\lambda_1 = \gamma/\lambda_1 = 0.01\zeta$

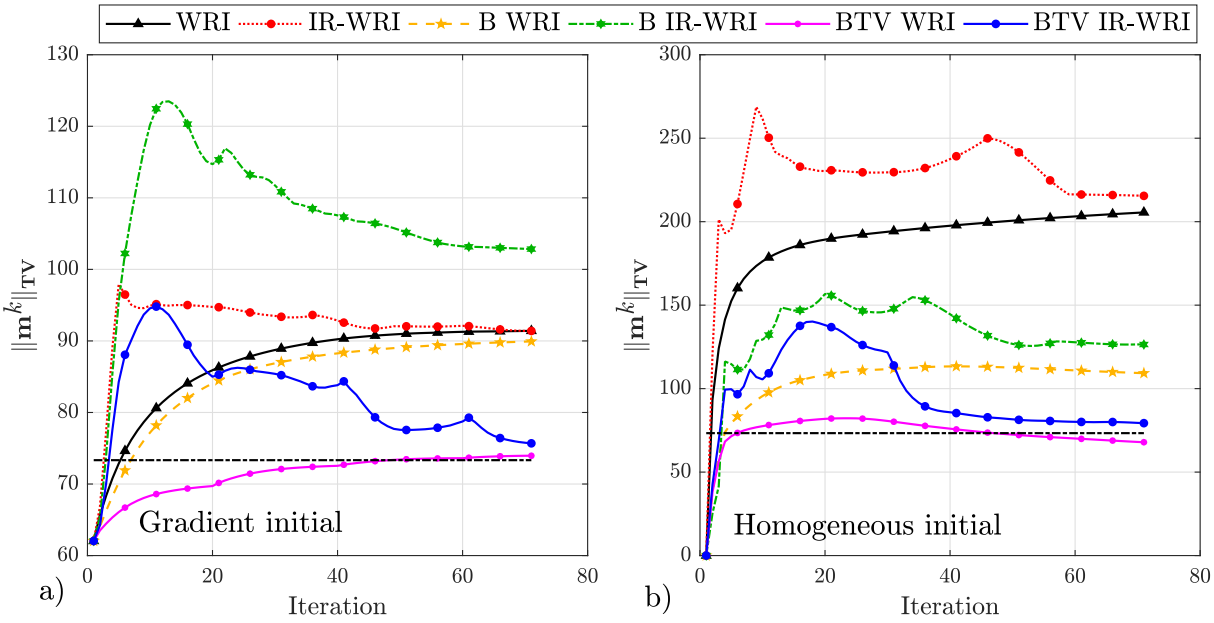

Figure 3.6 – The box-shape anomaly example. TV norm history over iterations of bound constrained, BTV regularized and ordinary WRI and IR-WRI for (a) velocity-gradient and (b) homogeneous initial model. The TV norm of true model is plotted with dashed-black line.

and remove the damping of $\mathbf{L}^T\mathbf{L}$. The starting and final frequencies are 3.5 Hz and 12 Hz and the sampling interval in one batch is 0.5 Hz. The algorithm performs at most 15 iterations per frequency batch and the number of iterations that have been performed is 170. The inversion results are shown in Figure 3.9. WRI with bound constraints and BTV (Figure 3.9(a,c)) fails to converge toward satisfactory results, while BTV-regularized IR-WRI converges to accurate velocity model, although a significant imprint of the TV regularization is shown (Figure 3.9d). When IR-WRI is performed with only bound constraints, the oscillating artifacts are not cancelled out (Figure 3.9b). This highlights the role of TV regularization in reconstructing blocky structures and removing wraparound artifacts.

2004 BP salt model - Left target

We now consider a second target of the 2004 BP salt model located on the left side of the model (Figure 3.10a). This target was previously used by among others Métivier et al. (2016a), Brandsberg-Dahl et al. (2017) and Esser et al. (2018) for FWI applications. After rescaling of the original model with the ratio used by Métivier et al. (2016a), our subsurface target is 16250 m wide and 5825 m deep, and is discretized with a 25 m grid interval. We used 108 sources spaced 150 m apart on the top side of the model. The source signature is a Ricker wavelet with a 10 Hz dominant frequency. A line of receivers with a 25 m spacing are deployed at the surface leading to a stationary-receiver acquisition. We perform IR-WRI with bound constraints alone and with BTV regularization, for noiseless and noisy data.

We used a crude laterally-homogeneous velocity-gradient model as initial model (Figure 3.10b). We used small batches of two frequencies with one frequency overlap between two

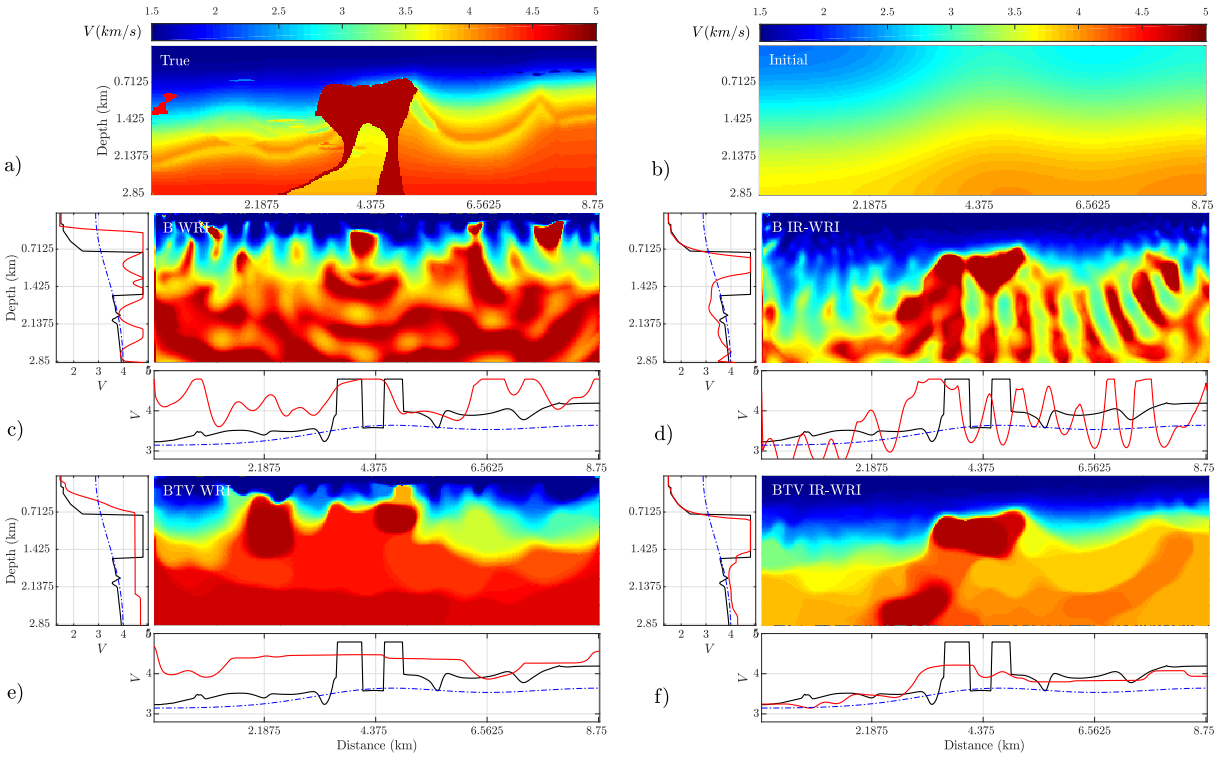

Figure 3.7 – 2004 BP salt model - Central target. 3-Hz frequency. (a) True model. (b) Starting model. (c-d) Bound constrained WRI(c) and IR-WRI(d) models. (e-f) BTV regularized WRI(e) and IR-WRI(f) models. Horizontal and vertical profiles at 4.15 km distance and 1.65 km depth from the true (black), initial (dash blue) and reconstructed (red) models are shown below and on the left-hand side of the models.

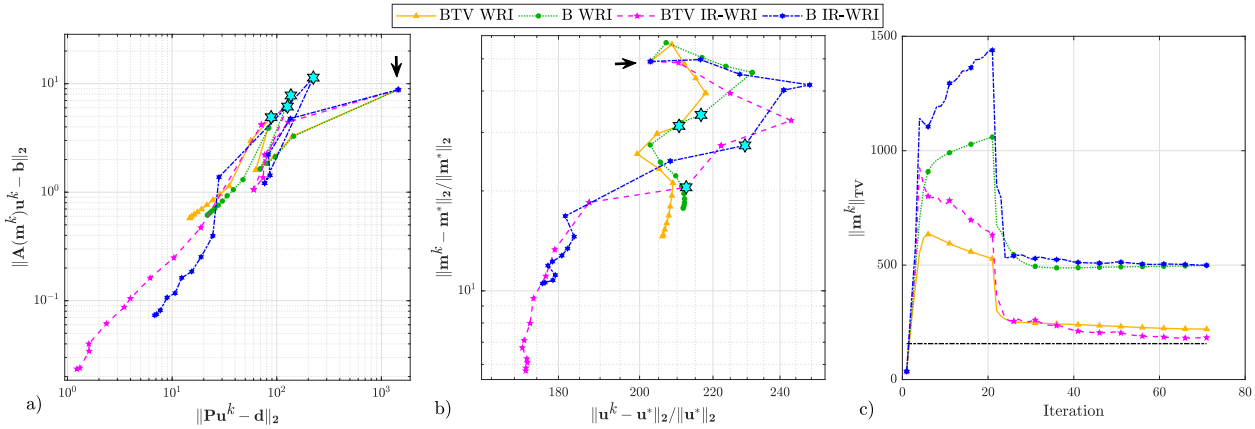

Figure 3.8 – 2004 BP salt model - Central target. 3-Hz frequency. Convergence history (a) in the $(\|\mathbf{P}\mathbf{u}^k-\mathbf{d}\|_2-\|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k-\mathbf{b}\|_2)$ plane (b) in the $(\|\mathbf{u}^k-\mathbf{u}^*\|_2/\|\mathbf{u}^*\|_2-\|\mathbf{m}^k-\mathbf{m}^*\|_2/\|\mathbf{m}^*\|_2)$ plane. (c) Evaluation of TV norm $\|\mathbf{m}_T V\|$ over iterations. Note again the more complex convergence history of IR-WRI compared to WRI due to the self-adaptive weighting of the data-fitting and wave-equation objective functions performed by right-hand side updating. The iteration 22 are located with cyan stars in (a) and (b).

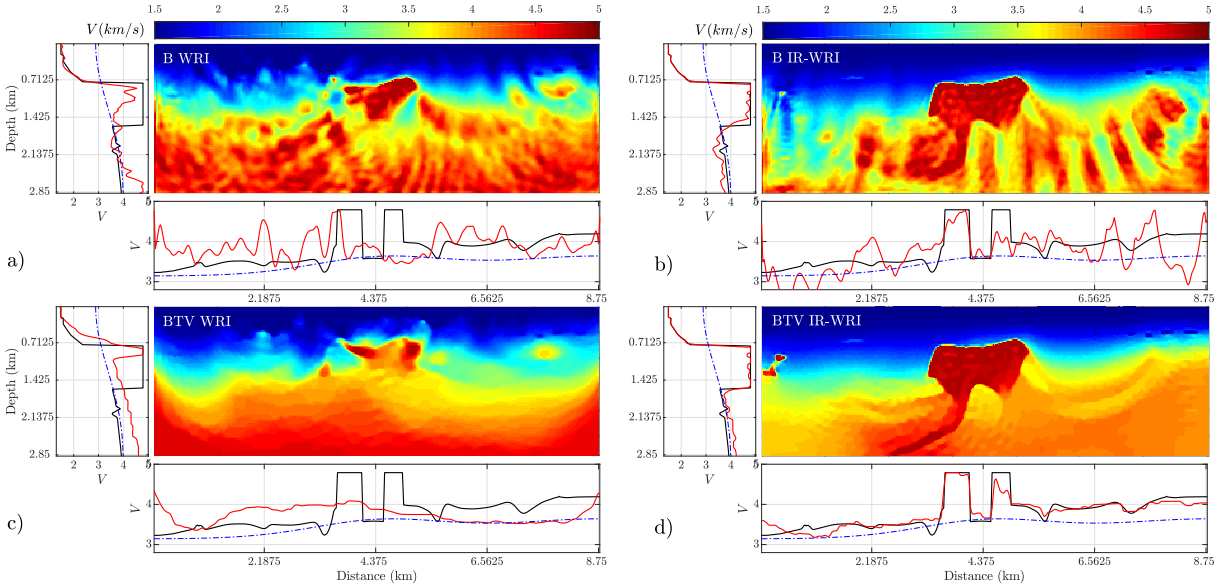

Figure 3.9 – 2004 BP salt case study. Central target. (a) Final bound constrained WRI with Figure 3.7(c) as initial model.(b) Final bound constrained IR-WRI with Figure 3.7(d) as initial model. (c) Final BTV regularized WRI with Figure 3.7(e) as initial model. (d) Final BTV regularized IR-WRI with Figure 3.7(f) as initial model.

consecutive batches, moving from the low frequencies to the higher ones according to a classical frequency continuation strategy. The starting and final frequencies are 3 Hz and 13 Hz and the sampling interval in one batch is 0.5 Hz. The stopping criterion of iteration for each batch is given by

$$k_{max} = 15 \quad \text{or} \quad (\|\mathbf{A}(\mathbf{m}^{k+1})\mathbf{u}^{k+1} - \mathbf{b}\|_2 \leq \varepsilon_b \qquad \text{and} \qquad \|\mathbf{P}\mathbf{u}^{k+1} - \mathbf{d}\|_2 \leq \varepsilon_d), \quad (3.37)$$

where k_{max} denotes the maximum iteration count, ε_b=1e-3, and ε_d=1e-5 for noiseless data and ε_b=1e-3 , ε_d= noise level of batch for noisy data. We perform three paths through the frequency batches to improve the IR-WRI results, using the final model of one path as the initial model of the next one (these cycles can be viewed as outer iterations of IR-WRI). The starting frequency of the second and third path is 6 Hz and 8.5 Hz, respectively. The IR-WRI models inferred from noiseless data with bound constraints and with BTV regularization are shown in Figure 3.10(c-d). The number of iterations that have been performed with bound constraints and with BTV regularization are 441 and 340, respectively. Direct comparison between the true model, the starting model and the IR-WRI models along three vertical logs cross-cutting the salt body at 5 km, 7.5 km and 10 km distance (vertical dashed lines in Figure 3.10a) are shown in Figure 3.11a. The results show the resilience of IR-WRI to cycle skipping with a pretty accurate reconstruction of the salt body and sub-salt structures (Figure 3.10(c-d)). However, the model obtained with bound constraints alone shows high-frequency noise in the salt body and below (Figure 3.10c). This noise can result from Gibbs phenomenon caused by the frequency decimation and artifacts resulting from multi scattering (Alkhalifah et al., 2018). The BTV regularization efficiently removes these artifacts except those resulting from truncation

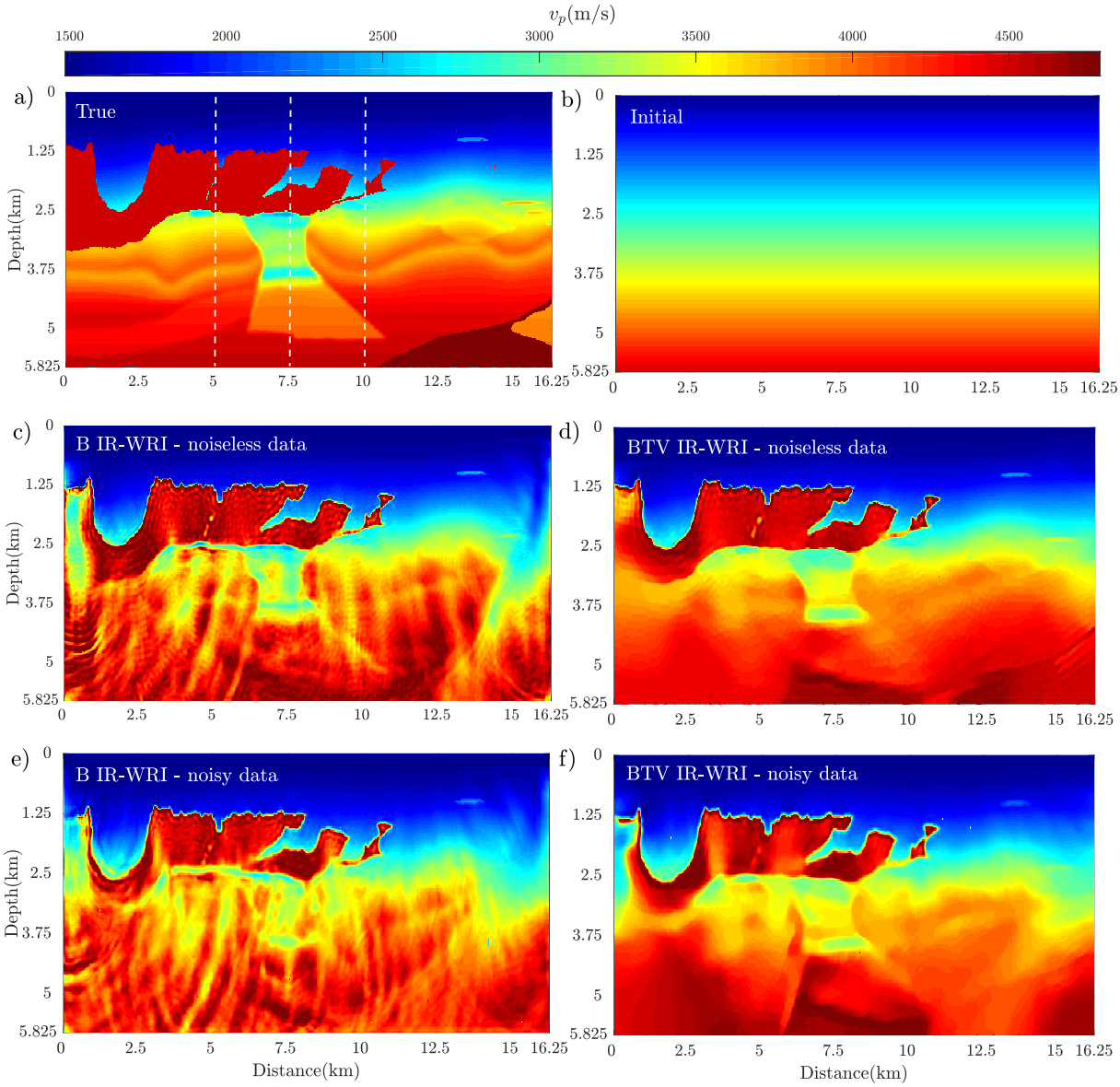

Figure 3.10 – The 2004 BP salt case study. Left target. (a) True BP model. The vertical dashed lines indicate the location of vertical logs of Figure 3.11. (b) Initial velocity model. (c-d) Final IR-WRI velocity models obtained with bound constraints (c) and with BTV regularization (d) for noiseless data. (e-f) Same as (c-d) for noisy data for a SNR of 10 db.

of the acquisition near the left end of the model at 5-km depth (Figure 3.10d). The inversion captures reasonably well the subsalt structures, including the low-velocity over-pressure zone at (x,z)=(7.5km,4km) as well as the smooth velocity variations.
When noisy data are used (Figure 3.10(e-f) and 3.11b), the number of iterations that have been performed with bound constraints alone and with BTV regularization is 263 and 254, respectively. As for the noiseless case, a direct comparison between the true model, the starting model and the two IR-WRI models along three vertical logs cross-cutting the salt body at 5 km, 7.5 km

and 10 km distance is shown in Figure 3.11b. The artifacts have now a more significant imprint when only bound constraints are used (Figure 3.10e). Accordingly, we apply a more aggressive TV regularization to obtain the results shown in Figure 3.10f. The artifacts have been efficiently removed with however a more obvious imprint of the piecewise-constant approximation underlying BTV regularization. This blocky pattern is clearly visible in deep part of the vertical profiles of Figure 3.11b where velocity gradients have been replaced by stack of constant-velocity layers.

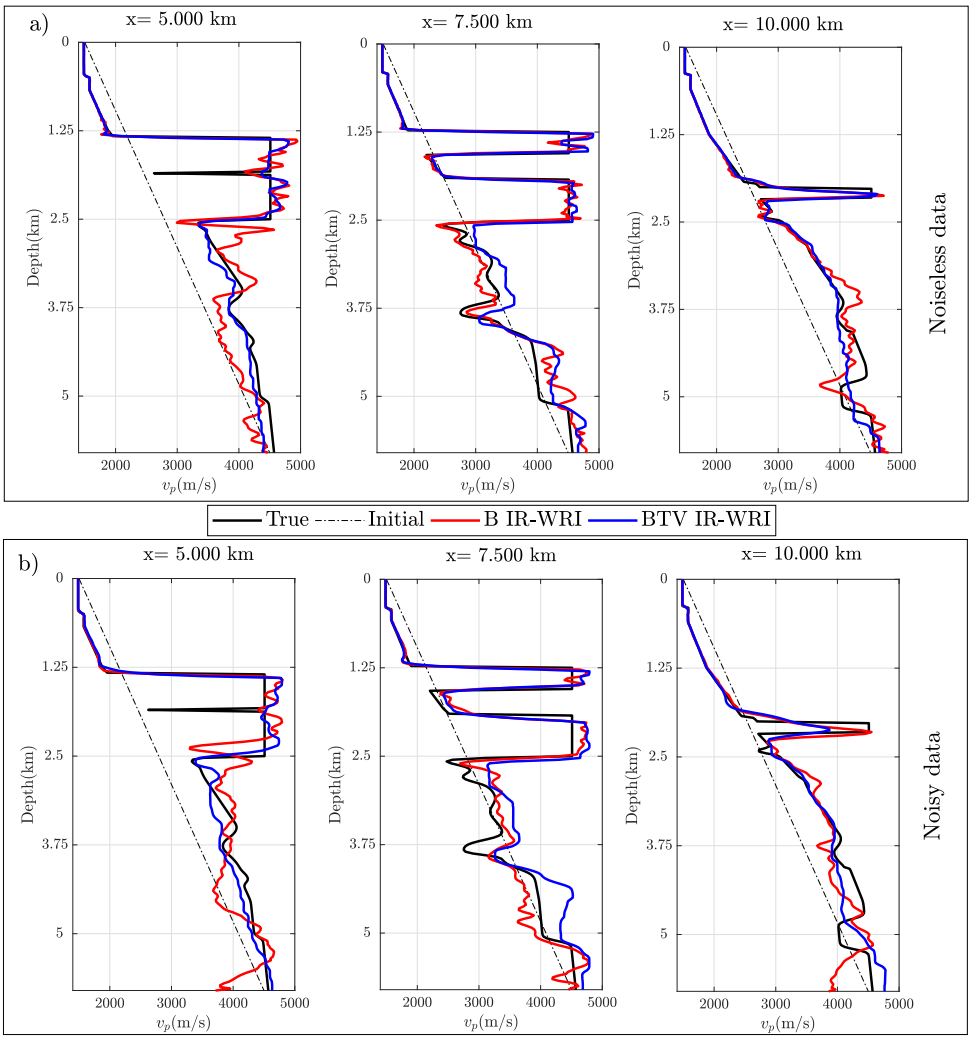

Figure 3.11 – 2004 BP salt case study - Left target. Direct comparison between the true velocity model (black), the initial model (dashed line) and the IR-WRI models obtained with bound constraints (red) and with BTV regularization (blue) along three logs at x=4.5 km, 7 km, 10 km (vertical dashed lines in Figure 3.10a) from left to right. (a) Noiseless data. (b) Noisy data.

3.2.5 Discussion

We have implemented bound constraints and TV regularization in the wavefield reconstruction inversion (WRI) method (van Leeuwen and Herrmann, 2013) which has been recently improved by Aghamiry et al. (2019c) in the framework of the alternating-direction method of multiplier (ADMM), leading to the iteratively-refined WRI (IR-WRI).

There are some key differences between our implementation of BTV-regularized WRI and previous ones based upon projected-gradient method (Peters and Herrmann, 2017) and primal dual hybrid gradient (PDHG) methods (Esser et al., 2018; Yong et al., 2018) that we review below. In the projected-gradient method, Peters and Herrmann (2017) first update model parameters with FWI or WRI and then project the updated model into the intersection of the TV and bound constraint with Dykstra projection algorithm (Boyle and Dykstra, 1986), an alternating projection onto some constraints until satisfying all of them. Therefore, their workflow is subdivided in two different parts, that are the model update followed by the projection onto intersection of all constraints (Peters and Herrmann, 2017, their equation 7). Unlike Peters and Herrmann (2017) method, which relies on independent update and projection steps, all the ingredients of BTV-regularized IR-WRI (i.e., wavefield reconstruction, parameter estimation, TV regularization and bound constraints) are consistently integrated in the theoretical framework of ADMM optimization (the readers can also refer to Maharramov and Levin, 2015).
Esser et al. (2018); Yong et al. (2018) implemented TV regularization and bound constraints in the reduced variable projection WRI of van Leeuwen and Herrmann (2016) with PDHG. PDHG is a method to solve constrained optimization problems by alternating gradient descent (for primal variable) and gradient ascent (for dual variable), which can be interpreted as linearized ADMM (Goldstein et al., 2015). PDHG can be helpful if the least squares minimizations embedded in ADMM are difficult to solve efficiently (Goldstein et al., 2015). This is not really the case in WRI, which mainly requires to solve two sparse linear systems for $\mathbf{u}$ and $\mathbf{m}$. Meanwhile, the selection of step size in PDHG that guarantees fast convergence, or even convergence at all, is not intuitive at all and can make PDHG impractical. This issue prompted Goldstein et al. (2015) to develop step size tuning rules which contribute to make PDHG self-adaptive as illustrated recently by Yong et al. (2018) in the frame of WRI. Beyond PDHG method, Esser et al. (2018); Yong et al. (2018) minimize data and source residuals with a penalty method, which lacks the convergence property of the augmented Lagrangian method promoted in IR-WRI. Moreover, they implement TV regularization as a hard constraint in the parameter-estimation subproblem through a Lagrangian function, whose saddle point is estimated with PDHG. In contrast, we implement the TV regularization as a soft constraint after introducing the auxiliary variable $\mathbf{p}$ and solve the regularized sub-problem for $\mathbf{m}$ with the split Bregman method (or equivalently ADMM), equation 3.26. This gives us the necessary flexibility to implement aggressive regularization during early iterations and relax it progressively, when very crude initial models are used. With more accurate initial models, the augmented Lagrangian embedded in the split Bregman method, equation 3.26, allows for constant penalty parameter to be used. These penalty parameters are used as step lengths in the augmented Lagrangian method, hence leading to a self-adaptive TV regularization implementation rid of tedious TV-norm ball continuation strategies and/or adaptive step lengths (Esser et al., 2018; Yong et al., 2018). Third, while Esser et al. (2018, equation 18) and Yong et al. (2018, equation 41) implement bound constraints

as hard constraints by projection at each iteration of the TV-regularized perturbation model onto the feasible set defined by the bound constraints, we implement bound constraints consistently with TV regularization in the framework of the method of multiplier, equations 3.30-3.34. Finally, our approach does not require prior guess of TV-norm ball because we implement TV regularization as a minimization problem rather than as a constraint. That being said, Esser et al. (2018) show promising results with the τ continuation approach and asymmetric TV norm on the BP salt model starting from a very crude initial model. It will be interesting in future work to assess the benefit of asymmetric TV norm in BTV IR-WRI.

Indeed, implementation of BTV-regularized IR-WRI requires to setup different parameters, which have been discussed at the beginning of the section *Experimental setup and parameter tuning*. The penalty parameter which controls the relative weight between the wave equation objective and the data fitting objective during the wavefield-reconstruction subproblem has been discussed in length in Aghamiry et al. (2019c). They have concluded that a fixed penalty parameter can be used during iterations because the accuracy of the minimizer in the method of multiplier is controlled both by the penalty parameter and the accuracy of the updated multiplier (Nocedal and Wright, 2006, Theorem 17.6). In this study, we suggest good pragmatical values of this penalty parameter as a percentage of the maximum eigenvalue of the augmented wave-equation normal operator for noiseless and noisy data. Optimal values may be refined by trial and error to prevent noisy data over-fitting and keep the iteration count within reasonable limits. The TV and bound parameters should be easily determined from well logs or a priori geological knowledge. Finally, the relative weight between the TV regularization and the wave-equation constraint, which is controlled by λ_1, needs also to be estimated during the parameter-estimation subproblem. As above mentioned, this penalty parameter is kept fixed during iterations or is progressively decreased to relax the TV regularization and bound constraints near the convergence point, depending of the accuracy of the initial model. The reader is also referred to Goldstein and Osher (2009, section 2.2) who discuss the sensitivity of the split Bregman method (an optimization method similar to ADMM (Esser, 2009)) to the penalty parameter for ℓ_1-regularized problems.

One drawback of the BTV regularization is related to the piecewise constant approximation underlying TV regularization, which tends to superimpose some blocky patterns on the smooth part of the subsurface. To overcome this issue, ongoing work seeks to optimally combine Tikhonov and TV regularization in IR-WRI (Gholami and Hosseini, 2013; Aghamiry et al., 2018a, 2020b).

Other perspective developments involve extension to multiparameter reconstruction, 3D geometries and application to real data to further assess the potential and limits of IR-WRI. For 3D applications, the wavefield reconstruction in the frequency domain requires to solve a large-scale linear algebra problem, equation 3.20. Operto et al. (2015), Amestoy et al. (2016), and Operto and Miniussi (2018) have shown the computational efficiency of 3D frequency-domain FWI based on sparse direct solvers in the 3.5-10 Hz frequency band for dense stationary-recording ocean-bottom cable acquisitions. Mary (2017) showed that block low-rank multifrontal solver allows one to tackle numerical problems involving up to 100 million unknowns. The symmetry of the normal operator, equation 3.20, should balance the computational overhead resulting from the higher number of non-zero coefficients relative to the impedance matrix $\mathbf{A}$. Note that $\mathbf{A}$ was processed as an unsymmetric matrix in the above-mentioned references due to dis-

cretization issues, although more suitable discretisations which preserve the symmetry of the impedance matrix may be considered in the future (Pratt and Smithyman, 2018). Alternatively, domain decomposition methods suitable for Helmholtz problems can be interfaced with hybrid direct-iterative solvers to perform wavefield reconstruction in bigger computational domains (Dolean et al., 2015). These approaches may be suitable for stationary-receiver acquisition involving a more limited number of reciprocal sources such as ocean bottom seismometer acquisitions.

3.2.6 Conclusion

We have presented a new method to implement TV regularization and bound constraints in frequency-domain FWI based on wavefield reconstruction (WRI). In a previous study, we have reformulated WRI in the framework of the alternating-direction method of multiplier (ADMM), leading to the iteratively-refined wavefield-reconstruction inversion method (IR-WRI). We have shown how the augmented Lagrangian embedded in ADMM makes IR-WRI weakly sensitive to a wide range of penalty parameter thanks to the Lagrange multiplier updating. Using a small value of this penalty parameter efficiently extends the search space during early iterations to foster data fitting without preventing the wave-equation constraint to be fulfilled at the convergence point with a preset prescribed error. IR-WRI performs a first ADMM step to alternate wavefield reconstruction and subsurface parameter estimation as in the original WRI method. When BTV regularization is used, we perform a second ADMM step to decompose the BTV-regularized parameter estimation sub-problem into a sequence of two simpler sub-problems through the introduction of auxiliary variables. This variable splitting allows for the de-coupling between the ℓ_1 and the ℓ_2 components of the penalty function according to the so-called split Bregman method. An interesting property exploited by WRI methods is the bi-linearity of the wave equation constraint with respect to the wavefield and the parameter, which makes the ℓ_2 wave-equation objective of the second sub-problem quadratic. Our implementation of BTV-regularization in IR-WRI with ADMM (or, equivalently split Bregman) provides a versatile framework to cascade constraints and regularization of different nature and is reasonably easy-to-tune due to the limited sensitivity of the augmented Lagrangian method to the choice of the penalty parameters. For challenging subsurface targets with contrasted structures such as salt bodies, we have shown that our BTV-regularized WRI shows a high resilience to cycle skipping and noise and efficiently mitigates high-frequency artifacts associated with incomplete illumination and multi scattering without detriment to the resolution of the imaging.

Acknowledgements

We would like to thank the editor H. Chauris, T. Alkhalifah and one anonymous reviewer for their comments. This study was partially funded by the SEISCOPE consortium (*http://seiscope2.osug.fr*), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXON-MOBIL, JGI, PETROBRAS, SCHLUMBERGER, SHELL, SINOPEC and TOTAL. This study was granted access to the HPC resources of SIGAMM infrastructure (http://crimson.oca.eu), hosted by Observatoire de

la Côte d'Azur and which is supported by the Provence-Alpes Côte d'Azur region, and the HPC resources of CINES/IDRIS/TGCC under the allocation 0596 made by GENCI.

3.2.7 Appendix A: Scaled form of augmented Lagrangian

Let's start with the following constrained problem

$$\min_{\mathbf{x}} \quad \|P(\mathbf{x})\|_2^2 \quad \text{subject to} \quad Q(\mathbf{x}) = \mathbf{0}. \tag{3.38a}$$

The augmented Lagrangian function for the problem is (Nocedal and Wright, 2006)

$$\mathcal{L}_A(\mathbf{x}, \mathbf{v}) = \|P(\mathbf{x})\|_2^2 + \mathbf{v}^T Q(\mathbf{x}) + \frac{\lambda}{2}\|Q(\mathbf{x})\|_2^2. \tag{3.39}$$

The problem 3.39 can be written in a more compact form by introducing the scaled dual variable $\bar{\mathbf{q}} = -\frac{\mathbf{v}}{\lambda}$ and adding and subtracting the term $\frac{\lambda}{2}\|\bar{\mathbf{q}}\|_2^2$ to the augmented Lagrangian 3.39. In this case, we arrive at the following scaled-form of the method of multipliers:

$$\begin{aligned} \mathcal{L}_A(\mathbf{x}, \bar{\mathbf{q}}) &= \|P(\mathbf{x})\|_2^2 - \lambda \bar{\mathbf{q}}^T Q(\mathbf{x}) + \frac{\lambda}{2}\|Q(\mathbf{x})\|_2^2 + \frac{\lambda}{2}\|\bar{\mathbf{q}}\|_2^2 - \frac{\lambda}{2}\|\bar{\mathbf{q}}\|_2^2 && \text{(3.40a)} \\ &= \|P(\mathbf{x})\|_2^2 + \frac{\lambda}{2}\|Q(\mathbf{x}) - \bar{\mathbf{q}}\|_2^2 - \frac{\lambda}{2}\|\bar{\mathbf{q}}\|_2^2. && \text{(3.40b)} \end{aligned}$$

3.3 Compound Regularization of Full-waveform Inversion for Imaging Piecewise Media

This section includes our paper about applying compound regularization, combination of simple regularizations, and bound constraints on IR-WRI using splitting techniques and ADMM (Aghamiry et al., 2020b).

Compound Regularization of Full-waveform Inversion for Imaging Piecewise Media

Hossein S. Aghamiry, Ali Gholami and Stéphane Operto

*IEEE Transactions on Geoscience and Remote Sensing, 2020, **58**(2), pages 1192-1204, DOI:10.1109/TGRS.2019.2944464*

3.3.1 Summary

Full waveform inversion (FWI) is an iterative nonlinear waveform matching procedure, which seeks to reconstruct unknown model parameters from partial waveform measurements. The nonlinear and ill-posed nature of FWI requires sophisticated regularization techniques to solve it. In most applications, the model parameters may be described by physical properties (e.g., wave speeds, density, attenuation, anisotropy) which are piecewise functions of space. Compound regularizations are thus beneficial to capture these different functions by FWI. We consider different implementations of compound regularizations in the wavefield reconstruction inversion (WRI) method, a formulation of FWI that extends its search space and prevents the so-called cycle skipping pathology. Our hybrid regularizations rely on Tikhonov and total variation (TV) functionals, from which we build two classes of hybrid regularizers: the first class is simply obtained by a convex combination (CC) of the two functionals, while the second relies on their infimal convolution (IC). In the former class, the model parameters are required to simultaneously satisfy different priors, while in the latter the model is broken into its basic components, each satisfying a distinct prior (e.g. smooth, piecewise constant, piecewise linear). We implement these compound regularizations in WRI using the alternating direction method of multipliers (ADMM). Then, we assess our regularized WRI for seismic imaging applications. Using a wide range of subsurface models, we conclude that the compound regularizer based on IC leads to the lowest error in the parameter reconstruction compared to that obtained with the CC counterpart and the Tikhonov and TV regularizers when used independently.

3.3.2 Introduction

Full waveform inversion (FWI) seeks to estimate constitutive parameters by nonlinear minimization of a distance between recorded and simulated wavefield measurements. This technology was originally developed in geophysical imaging (Tarantola, 1984), and has spread more recently into other fields of imaging sciences such as medical imaging (Sandhu et al., 2015) and oceanography (Wood et al., 2008). This partial-differential equation (PDE)-constrained nonlinear inverse problem is classically solved with local reduced-space optimization methods (Haber et al., 2000). In this linearized framework, a challenging source of non linearity is the so-called cycle skipping pathology which occurs when the initial model does not allow to match the data with a kinematic error smaller than half a period (Virieux and Operto, 2009; Virieux et al., 2017). Other sources of error are noise, approximate wave physics and ill-posedness resulting from parameter cross-talk, coarse acquisition sampling and uneven illumination of the targeted

structure. Designing regularization techniques that mitigate these source of errors is therefore a key challenge for the success of FWI applications.

A proper regularization should be driven by the shape and statistical characteristics of the medium to be imaged. For example, in geophysical imaging, the subsurface can be represented by a piece-wise smooth medium, that is a model which contains smoothly varying and blocky components. The widespread Tikhonov regularizations (Tikhonov et al., 2013) rely on the smoothness assumption and hence fail to recover sharp interfaces of such media. Conversely, first order total variation (TV) regularizations (Rudin et al., 1992) are based on a blockiness assumption and hence are more suitable to image large contrasts. However, they generate undesirable staircase imprints in smooth regions (Gholami and Hosseini, 2013). Hereafter we refer TV to as first order TV regularization, unless we explicitly mention the order of the TV regularization. Regions characterized by smoothly-varying properties and those containing sharp contrasts have different statistical properties. The former are characterized by the normal prior, while the latter by a heavy tailed prior (Polson and Sokolov, 2019). Consequently, simultaneous recovery of both properties is difficult when one type of regularization is used (Tikhonov, TV, etc). To overcome this issue, a combination of different regularizations can be used (Gholami and Siahkoohi, 2010; Benning and Burger, 2018; Bioucas-Dias and Figueiredo, 2008). A naive approach consists of the simple additive coupling or convex combinations (CC) of regularizations. Alternatively, Gholami and Hosseini (2013) proposed to explicitly decompose the model into several components of different statistical properties and use an appropriate regularization to reconstruct each component. Using this strategy, they combined Tikhonov and TV regularizations (referred to as TT regularization) to reconstruct piece-wise smooth media. The smooth components are captured by the Tikhonov regularization, while the blocky ones are determined by the TV counterpart. In many applications, it has been shown that a compound regularization based upon infimal convolution (IC) outperforms the one based upon additive coupling (Bergmann et al., 2018).

TT regularization based upon IC has been successfully applied to FWI for seismic subsurface imaging in the framework of iteratively-refined wavefield reconstruction inversion (IR-WRI) (Aghamiry et al., 2018a). IR-WRI extends the search space of FWI and decreases cycle skipping through a relaxation of the wave-equation constraint (van Leeuwen and Herrmann, 2013; Aghamiry et al., 2019c, 2018c). Taking advantage of the bilinearity of the wave equation, IR-WRI breaks down FWI into two linear subproblems which are solved in an alternating mode: wavefield reconstruction driven by the observables and model-parameter estimation by minimization of the source residuals the relaxation generated. Furthermore, Aghamiry et al. (2019a) extended the method to acoustic multiparameter inversion. The linearity of the parameter-estimation subproblem provides a suitable framework to implement sophisticated nonsmooth regularizations.

In this study, following Gholami and Hosseini (2013) and Aghamiry et al. (2018a), we develop a general framework to combine a couple of regularization terms in IR-WRI through IC. Then, we specifically develop this framework for Tikhonov and TV regularizations, which are suitable for seismic subsurface imaging applications. Compared to Aghamiry et al. (2018a), we jointly update the blocky and the smooth components through a variable projection process rather than in an alternating mode. We first show that our new IC-based TT regularization out-

performs the CC-based counterpart with several well-documented numerical benchmarks in the field of seismic imaging. We also compare the results obtained with these two TT regularizations with those obtained with total generalized variation (TGV) regularization, a combination of first and second order TV, and those obtained with Tikhonov and TV regularizations when used independently.

3.3.3 Notation

The mathematical symbols adopted in this paper are as follows. We use italics for scalar quantities, boldface lowercase letters for vectors, and boldface capital letters for matrices and tensors. We use the superscript T to denote the adjoint of an operator. The ith component of the column vector $\mathbf{x}$ is shown by $\mathbf{x}_i$ and its absolute value is returned by $|\mathbf{x}_i|$. For the real-valued n-length column vectors $\mathbf{x}$ and $\mathbf{y}$ the dot product is defined by $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{y}$ and their Hadamard product, denoted by $\mathbf{x} \circ \mathbf{y}$, is another vector made up of their componentwise products, i.e. $(\mathbf{x} \circ \mathbf{y})_i = \mathbf{x}_i\mathbf{y}_i$. The ℓ_2- and ℓ_1-norms of $\mathbf{x}$ are, respectively, defined by $\|\mathbf{x}\|_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_{i=1}^{n} |\mathbf{x}_i|^2}$ and $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |\mathbf{x}_i|$.

3.3.4 Method

In this section, we briefly review the frequency-domain FWI as a bi-convex feasibility problem and describe the extended forms of FWI. We show how the problem can be solved with the alternating direction method of multipliers (ADMM) (Boyd et al., 2010) for a general regularization function.

Full-waveform inversion

Frequency-domain FWI with a general regularization term and bounding constraints can be formulated as (Aghamiry et al., 2019c,b)

$$\begin{aligned} \min_{\mathbf{u},\mathbf{m}\in\mathcal{C}} \quad & \Phi(\mathbf{m}) \\ \text{subject to} \quad & \mathbf{A}(\mathbf{m})\mathbf{u} = \mathbf{b}, \\ & \mathbf{P}\mathbf{u} = \mathbf{d}, \end{aligned} \tag{3.41}$$

where $\mathbf{m} \in \mathbb{R}^{n\times 1}$ gathers unknown squared slowness, n is the number of discrete grid points, $\Phi(\mathbf{m})$ is an appropriate regularization term which we assume to be convex, $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^{n\times 1} \mid \mathbf{m}_l \leq \mathbf{x} \leq \mathbf{m}_u\}$ is the set of all feasible models bounded by the lower bound $\mathbf{m}_l$ and the upper bound $\mathbf{m}_u$.

The first constraint in (3.41), $\mathbf{A}(\mathbf{m})\mathbf{u} = \mathbf{b}$, is a partial-differential equation (PDE) wherein $\mathbf{u} \in \mathbb{C}^{n\times 1}$ is the wavefield and $\mathbf{b} \in \mathbb{C}^{n\times 1}$ is the source term. In this study, $\mathbf{A}(\mathbf{m}) \in \mathbb{C}^{n\times n}$ is the discretized PDE Helmholtz operator (Pratt et al., 1998; Chen et al., 2013) given by

$$\mathbf{A}(\mathbf{m}) \;=\; \boldsymbol{\Delta} + \omega^2\mathbf{C}(\mathbf{m})\text{diag}(\mathbf{m})\mathbf{B}, \tag{3.42}$$

with ω the angular frequency and $\boldsymbol{\Delta}$ the discretized Laplace operator. The diagonal matrix $\mathbf{C}$ embeds boundary conditions and can be dependent or independent on $\mathbf{m}$ depending on the kinds of absorbing boundary conditions (radiation versus sponge) (Aghamiry et al., 2019c). Also, $\mathbf{B}$ is used to spread the *mass* term $\omega^2\mathbf{C}(\mathbf{m})\text{diag}(\mathbf{m})$ over all the coefficients of the stencil to improve its accuracy following an anti-lumped mass strategy (Marfurt, 1984; Jo et al., 1996; Hustedt et al., 2004).

The second constraint in (3.41), $\mathbf{Pu} = \mathbf{d}$, is the observation equation, in which $\mathbf{d} \in \mathbb{C}^{m\times 1}$ is the recorded seismic data, m is the number of recorded data and $\mathbf{P} \in \mathbb{R}^{m\times n}$ is a linear operator that samples the wavefield at the receiver positions.

WRI approach to solving (3.41)

The extended approach, known as wavefield reconstruction inversion (WRI) (van Leeuwen and Herrmann, 2013), recasts the constrained optimization problem, equation (3.41), as an unconstrained problem where both constraints are implemented with quadratic penalty functions.

$$\min_{\mathbf{u},\mathbf{m}\in\mathcal{C}} \Phi(\mathbf{m}) + \frac{\lambda_0}{2}\|\mathbf{Pu}-\mathbf{d}\|_2^2 + \frac{\lambda_1}{2}\|\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{b}\|_2^2, \tag{3.43}$$

where $\lambda_0, \lambda_1 > 0$ are the penalty parameters. For the unregularized case where $\Phi(\mathbf{m}) = 0$ and also without the bounding constraint, van Leeuwen and Herrmann (2013) solved this biconvex minimization problem with an alternating-direction algorithm, whereby the joint minimization over $\mathbf{u}$ and $\mathbf{m}$ is replaced by an alternating minimization over each variable separately. The main property of the penalty formulation given by equation (3.43) is that the PDE constraint in the original problem is replaced by a quadratic penalty term, which enlarges the search space and mitigates the inversion nonlinearity accordingly (van Leeuwen and Herrmann, 2013). Its main drawback, however, is the difficulty related to the adaptive tuning of the penalty parameter, which is common to all penalty methods (Nocedal and Wright, 2006).

IR-WRI approach to solving (3.41)

To overcome the above limitation, the iteratively-refined WRI (IR-WRI) implements the original constrained problem (3.41) with the augmented Lagrangian (AL) method (Nocedal and Wright, 2006; Hestenes, 1969).

$$\begin{aligned}\min_{\mathbf{u},\mathbf{m}\in\mathcal{C}} \max_{\mathbf{v}_0,\mathbf{v}_1} \Phi(\mathbf{m}) + \frac{\lambda_0}{2}\|\mathbf{Pu}-\mathbf{d}\|_2^2 + \frac{\lambda_1}{2}\|\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{b}\|_2^2 \\ + \mathbf{v}_0^T[\mathbf{Pu}-\mathbf{d}] + \mathbf{v}_1^T[\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{b}],\end{aligned} \tag{3.44}$$

where $\mathbf{v}_0 \in \mathbb{C}^{m\times 1}$ and $\mathbf{v}_1 \in \mathbb{C}^{n\times 1}$ are the dual variables (the Lagrangian multipliers). The min-max problem (3.44) can also be written in a more compact form (the scaled form AL) as

$$\begin{aligned}\min_{\mathbf{u},\mathbf{m}\in\mathcal{C}} \max_{\mathbf{v}_0,\mathbf{v}_1} \Phi(\mathbf{m}) &+ \frac{\lambda_0}{2}\|\mathbf{P}\mathbf{u}-\mathbf{d}+\frac{1}{\lambda_0}\mathbf{v}_0\|_2^2 - \frac{\lambda_0}{2}\|\frac{\mathbf{v}_0}{\lambda_0}\|_2^2 \\ &+ \frac{\lambda_1}{2}\|\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{b}+\frac{1}{\lambda_1}\mathbf{v}_1\|_2^2 - \frac{\lambda_1}{2}\|\frac{\mathbf{v}_1}{\lambda_1}\|_2^2.\end{aligned} \tag{3.45}$$

Applying a gradient ascent to (3.45) with respect to the duals, after a simple change of variables $\mathbf{d}^k = -\mathbf{v}_0^k/\lambda_0$ and $\mathbf{b}^k = -\mathbf{v}_1^k/\lambda_1$, gives the following iteration:

$$\begin{aligned}&\min_{\mathbf{u},\mathbf{m}\in\mathcal{C}} \ \Phi(\mathbf{m}) + \frac{\lambda_0}{2}\|\mathbf{P}\mathbf{u}-\mathbf{d}-\mathbf{d}^k\|_2^2 \\ &\qquad\qquad + \frac{\lambda_1}{2}\|\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{b}-\mathbf{b}^k\|_2^2, \\ &\mathbf{d}^{k+1} = \mathbf{d}^k + \mathbf{d} - \mathbf{P}\mathbf{u}, \\ &\mathbf{b}^{k+1} = \mathbf{b}^k + \mathbf{b} - \mathbf{A}(\mathbf{m})\mathbf{u},\end{aligned} \tag{3.46}$$

beginning with $\mathbf{d}^0 = 0$ and $\mathbf{b}^0 = 0$. Capitalizing on the bilinearity of the wave equation in $\mathbf{m}$ and $\mathbf{u}$, ADMM Boyd et al. (2010) is a powerful method to solve this kind of multivariate optimization problem. ADMM updates $\mathbf{m}$ and $\mathbf{u}$ separately through a Gauss-Seidel like iteration, i.e., fixing $\mathbf{m}$ and solving for $\mathbf{u}$ and vice versa. Accordingly, beginning with an initial guess $\mathbf{m}^0$, we end up with the following iteration to solve (3.46) (Aghamiry et al., 2019c,b):

$$\mathbf{u}^{k+1} = \arg\min_{\mathbf{u}} \ \left\| \begin{bmatrix} \sqrt{\frac{\lambda_0}{\lambda_1}}\mathbf{P} \\ \mathbf{A}(\mathbf{m}^k) \end{bmatrix} \mathbf{u} - \begin{bmatrix} \sqrt{\frac{\lambda_0}{\lambda_1}}(\mathbf{d}+\mathbf{d}^k) \\ \mathbf{b}+\mathbf{b}^k \end{bmatrix} \right\|_2^2 \tag{3.47a}$$

$$\mathbf{m}^{k+1} = \arg\min_{\mathbf{m}\in\mathcal{C}} \ \Phi(\mathbf{m}) + \frac{\lambda_1}{2}\|\mathbf{A}(\mathbf{m})\mathbf{u}^{k+1}-\mathbf{b}-\mathbf{b}^k\|_2^2, \tag{3.47b}$$

$$\mathbf{d}^{k+1} = \mathbf{d}^k + \mathbf{d} - \mathbf{P}\mathbf{u}^{k+1}, \tag{3.47c}$$

$$\mathbf{b}^{k+1} = \mathbf{b}^k + \mathbf{b} - \mathbf{A}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}. \tag{3.47d}$$

The subproblem (3.47a) associated with the wavefield reconstruction is quadratic and admits a closed-form solution. It relaxes the requirement to satisfy exactly the wave equation ($\mathbf{A}(\mathbf{m}^k)\mathbf{u} = \mathbf{b}$) for the benefit of improved data fitting ($\mathbf{P}\mathbf{u} = \mathbf{d}$). This is achieved by reconstructing the wavefields that best jointly fit the observations and satisfy the wave equation in a least-squares sense. While wavefields generated by the reduced approach (Appendix A, section 3.3.8) satisfy exactly the wave equation, $\mathbf{u}_r = \mathbf{A}(\mathbf{m}^k)^{-1}\mathbf{b}$, this makes classical FWI highly non-convex.

Equation (3.46) shows that the duals are updated with the running sum of the data and source residuals in iterations and are used to update the right-hand sides in the penalty functions of the scaled AL. These error correction terms in the AL method are the key ingredients that allow for a constant penalty parameter to be used in iterations, while guaranteeing convergence to accurate minimizer (Nocedal and Wright, 2006). In the next section we focus on the solution of the model subproblem (3.47b) when compound regularizations are used as the regularization

term.

3.3.5 Model Subproblem

This section presents compound regularization functionals and the details of our approach to solve the model subproblem (3.47b) with these functionals.

Compound regularizers

Simple regularizers, Appendix B (section 3.3.9), are effective for recovering models which can be characterized by a single prior and structure e.g. smooth, blocky, piecewise linear, etc. The compound regularizers are more effective for recovering complicated models that are represented by more than one prior. They are constructed by combining two or more separate simple regularizers. This can be done by either a *convex combination* (CC) or an *infimal convolution* (IC).

Convex combination of simple regularizers

A CC of r simple regularizer functionals $\Phi_1, ..., \Phi_r$ is a compound regularizer functional of the form

$$\Phi_\alpha(\mathbf{x}) = \alpha_1\Phi_1(\mathbf{x}) + ... + \alpha_r\Phi_r(\mathbf{x}), \tag{3.48}$$

where weights α_i satisfy $\alpha_i \geq 0$ and

$$\alpha_1 + \alpha_2, ..., +\alpha_r = 1. \tag{3.49}$$

Definitely, if all of the functions $\Phi_1, ..., \Phi_r$ are convex then Φ is so. In CC models, the regularized solution is forced to satisfy the individual priors simultaneously. As an example, a compound regularizer functional constructed by a CC of ℓ_1- and squared ℓ_2-norms ($\ell_1 + \ell_2$), known as an elastic net (Zou and Hastie, 2005; Gholami, 2013), is

$$\Phi_\alpha(\mathbf{x}) = (1-\alpha)\|\mathbf{x}\|_2^2 + \alpha\|\mathbf{x}\|_1, \tag{3.50}$$

with $0 \leq \alpha \leq 1$. The convexity of ℓ_1- and ℓ_2-norms implies that $\Phi_\alpha(\mathbf{x})$ in (3.50) is convex. One may also construct a compound regularizer functional by a CC of two ℓ_1-norms which are applied in different domains, such as those spanned by two different wavelet transforms, or those spanned by a wavelet transform and the gradient operator (Gholami and Siahkoohi, 2010).

Infimal convolution of simple regularizers

In IC models, the solution is decomposed into simple components and then each component is regularized by an appropriate prior. Accordingly, the IC of r simple regularizer functionals

$\Phi_1, ..., \Phi_r$ is a compound functional of the form

$$\Phi_\alpha(\mathbf{x}) = \min_{\mathbf{x}=\mathbf{x}_1+...+\mathbf{x}_r} \{\alpha_1\Phi_1(\mathbf{x}_1) + ... + \alpha_r\Phi_r(\mathbf{x}_r)\}. \tag{3.51}$$

In the case of two functionals, Φ_α in (3.51) takes the form

$$\Phi_\alpha(\mathbf{x}) = \min_{\mathbf{z}}\{(1-\alpha)\Phi_1(\mathbf{x}-\mathbf{z}) + \alpha\Phi_2(\mathbf{z})\}, \tag{3.52}$$

which is similar to the classical formula of convolution, and hence the term *infimal convolution*.

The IC of ℓ_1- and (squared) ℓ_2-norms ($\ell_1\square\ell_2$) is

$$\Phi_\alpha(\mathbf{x}) = \min_{\mathbf{z}}\{(1-\alpha)\|\mathbf{x}-\mathbf{z}\|_2^2 + \alpha\|\mathbf{z}\|_1\}, \tag{3.53}$$

which is a denoising problem whose solution is unique and is given by the well-known soft-threshold function (Donoho, 1995):

$$\mathbf{z} = \max\left(1 - \frac{\alpha}{2(1-\alpha)|\mathbf{x}|}, 0\right) \circ \mathbf{x}, \tag{3.54}$$

Putting $\mathbf{z}$ from (3.54) into (3.53) gives that

$$\Phi_\alpha(\mathbf{x}) = \begin{cases} (1-\alpha)|\mathbf{x}|^2 & \text{if } |\mathbf{x}| \le \frac{\alpha}{2(1-\alpha)} \\ \alpha|\mathbf{x}| - \frac{\alpha^2}{4(1-\alpha)} & \text{if } |\mathbf{x}| > \frac{\alpha}{2(1-\alpha)} \end{cases}, \tag{3.55}$$

which is nothing other than the Huber function (Huber, 1973). As seen, this function has a hybrid behavior: it has a quadratic behavior for small values of $|\mathbf{x}|$ and linear behavior for large values. The parameter $\frac{\alpha}{2(1-\alpha)}$ determines where the transition from quadratic to linear behavior takes place.

Geometrical illustrations of the ℓ_1-norm, ℓ_2-norm, $(\ell_1+\ell_2)$-norm, and $(\ell_1\square\ell_2)$-norm for $\alpha = 0.7$ are shown in Fig. 3.12. This figure shows that the ℓ_1- and ℓ_2-norms have a uniform behavior for all values, while the CC norm (the $(\ell_1 + \ell_2)$-norm) has a hybrid behavior: it approaches the ℓ_1-norm near zero, where it behaves as a linear function, but approaches the ℓ_2-norm for large values, where it behaves as a quadratic function. Unlike $\ell_1 + \ell_2$, the IC function $\ell_1\square\ell_2$ approaches the ℓ_2-norm near zero but is linear and approaches the ℓ_1-norm for large values.

In this paper, we consider (3.52) in the following settings, though other configurations are possible:

$$\Phi_\alpha^{\text{TT}}(\mathbf{x}) = \min_{\mathbf{x}=\mathbf{x}_1+\mathbf{x}_2} (1-\alpha)\|\nabla^2\mathbf{x}_2\|_2^2 + \alpha\|\nabla\mathbf{x}_1\|_1, \tag{3.56}$$

and

$$\Phi_\alpha^{\text{TGV}}(\mathbf{x}) = \min_{\mathbf{x}=\mathbf{x}_1+\mathbf{x}_2} (1-\alpha)\|\nabla^2\mathbf{x}_2\|_1 + \alpha\|\nabla\mathbf{x}_1\|_1, \tag{3.57}$$

where, in both (3.56) and (3.57), the norms are applied on the absolute valued components of $\nabla^2\mathbf{x}_2$ ((3.79) and (3.81)) and $\nabla\mathbf{x}_1$ (3.77). The compound regularizer Φ_α^{TT} is a combination of the second order Tikhonov and TV (TT) regularizations (Gholami and Hosseini, 2013) and Φ_α^{TGV} is a combination of the first and second order TV regularizations, called total generalized

variation (TGV) (Bredies et al., 2010; Setzer et al., 2011). The former is suitable for recovering piecewise-smooth models, while the latter is better suited for piecewise linear models. The next section gives a solution procedure to solve (3.47b) with these regularizers.

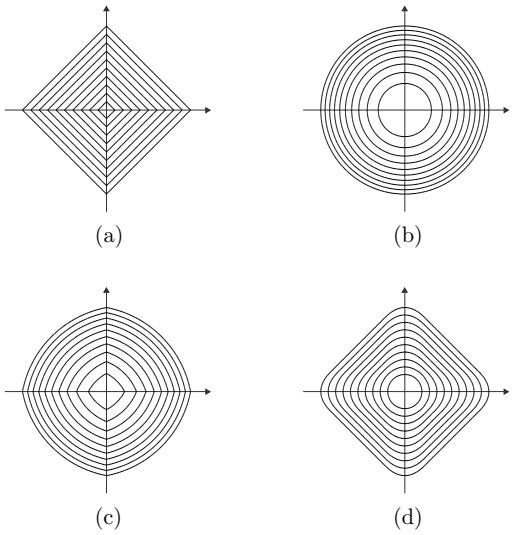

Figure 3.12 – Geometrical illustration of different regularizers. (a) the ℓ_1-norm, (b) the ℓ_2-norm, (c) the $(\ell_1 + \ell_2)$-norm, and (d) the $(\ell_1 \square \ell_2)$-norm.

Solving the subproblem (3.47b)

In this section we present how to solve the subproblem (3.47b) with TT regularization. The solution procedure for the TGV regularizer follows easily. From the definition of $\mathbf{A}$ in (3.42), we get that

$$\mathbf{A}(\mathbf{m})\mathbf{u} \quad = \quad \boldsymbol{\Delta}\mathbf{u} + \mathbf{L}\mathbf{m}, \tag{3.58}$$

where

$$\mathbf{L} = \frac{\partial \mathbf{A}(\mathbf{m})}{\partial \mathbf{m}}\mathbf{u} = \omega^2 \mathbf{C}\mathrm{diag}(\mathbf{B}\mathbf{u}), \tag{3.59}$$

and we assume that $\mathbf{C}$ does not depend on $\mathbf{m}$ (this is the case for perfectly-matched absorbing boundary conditions (Aghamiry et al., 2019c)). From the explicit decomposition $\mathbf{m} = \mathbf{m}_1 + \mathbf{m}_2$ and (3.58), the solution of the optimization problem (3.47b) can be expressed as

$$\underset{\substack{\mathbf{m}=\mathbf{m}_1+\mathbf{m}_2 \\ \mathbf{m}\in\mathcal{C}}}{\arg\min} \quad \Phi_\alpha^{\mathrm{TT}}(\mathbf{m}_1, \mathbf{m}_2) + \frac{\lambda_1}{2}\|\mathbf{L}[\mathbf{m}_1 + \mathbf{m}_2] - \mathbf{y}\|_2^2, \tag{3.60}$$

where $\mathbf{y} = \mathbf{b} + \mathbf{b}^k - \boldsymbol{\Delta}\mathbf{u}^{k+1}$. Defining auxiliary variables $\mathbf{p} = \nabla\mathbf{m}_1 \in \mathbb{R}^{2n\times 1}$ and $\mathbf{q} = \mathbf{m}_1 + \mathbf{m}_2 \in \mathbb{R}^{n\times 1}$ and recasting (3.60) as a constrained problem and then applying ADMM

leads to the following iteration (Aghamiry et al., 2019b, Section 2.2.2)

$$\begin{bmatrix}\mathbf{m}_1^{k+1}\\ \mathbf{m}_2^{k+1}\end{bmatrix} = \underset{\mathbf{m}_1,\mathbf{m}_2}{\arg\min}\, C(\mathbf{m}_1,\mathbf{m}_2,\mathbf{q}^k,\tilde{\mathbf{q}}^k,\mathbf{p}^k,\tilde{\mathbf{p}}^k), \tag{3.61a}$$

$$\mathbf{p}^{k+1} = \underset{\mathbf{p}}{\arg\min}\, \alpha\|\mathbf{p}\|_1 + \frac{\gamma_1}{2}\|\nabla\mathbf{m}_1^{k+1} - \mathbf{p} - \tilde{\mathbf{p}}^k\|_2^2, \tag{3.61b}$$

$$\mathbf{q}^{k+1} = \underset{\mathbf{q}\in\mathcal{C}}{\arg\min}\, \frac{\gamma_0}{2}\|\mathbf{m}_1^{k+1} + \mathbf{m}_2^{k+1} - \mathbf{q} - \tilde{\mathbf{q}}^k\|_2^2, \tag{3.61c}$$

where $\gamma_0, \gamma_1 > 0$ and

$$\begin{aligned} C(\mathbf{m}_1,\mathbf{m}_2,\mathbf{q}^k,\tilde{\mathbf{q}}^k,\mathbf{p}^k,\tilde{\mathbf{p}}^k) = \frac{\lambda_1}{2}\|\mathbf{L}[\mathbf{m}_1+\mathbf{m}_2]-\mathbf{y}\|_2^2 \\ + (1-\alpha)\|\nabla^2\mathbf{m}_2\|_2^2 + \frac{\gamma_1}{2}\|\nabla\mathbf{m}_1 - \mathbf{p}^k - \tilde{\mathbf{p}}^k\|_2^2 \\ + \frac{\gamma_0}{2}\|\mathbf{m}_1+\mathbf{m}_2-\mathbf{q}^k-\tilde{\mathbf{q}}^k\|_2^2. \end{aligned}$$

The auxiliary primal variables $\mathbf{p}$ and $\mathbf{q}$ are introduced to decouple the ℓ_1 and the ℓ_2 minimization problems and solve the former ones with proximal algorithms following operator splitting methods. The dual variables $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{q}}$ are updated through a gradient ascent step according to the method of multipliers (Nocedal and Wright, 2006)

$$\tilde{\mathbf{p}}^{k+1} = \tilde{\mathbf{p}}^k + \mathbf{p}^{k+1} - \nabla\mathbf{m}_1^{k+1}, \tag{3.62a}$$

$$\tilde{\mathbf{q}}^{k+1} = \tilde{\mathbf{q}}^k + \mathbf{q}^{k+1} - [\mathbf{m}_1^{k+1} + \mathbf{m}_2^{k+1}], \tag{3.62b}$$

We now discuss how to solve the subsubproblems given in (3.61).

The subsubproblem (3.61a)

A solution of subsubproblem (3.61a) occurs at the point where the derivatives of the objective function C with respect to $\mathbf{m}_1$ and $\mathbf{m}_2$ vanish simultaneously. Accordingly, we end up with the following linear system of equations:

$$\begin{bmatrix}\mathbf{G}_{11} & \mathbf{G}_{12}\\ \mathbf{G}_{21} & \mathbf{G}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{m}_1\\ \mathbf{m}_2\end{bmatrix} = \begin{bmatrix}\mathbf{h}_1\\ \mathbf{h}_2\end{bmatrix}, \tag{3.63}$$

with

$$\begin{cases}\mathbf{G}_{11} = \lambda_1\mathbf{L}^T\mathbf{L} + \gamma_1\nabla^T\nabla + \gamma_0\mathbf{I},\\ \mathbf{G}_{12} = \mathbf{G}_{21} = \lambda_1\mathbf{L}^T\mathbf{L} + \gamma_0\mathbf{I},\\ \mathbf{G}_{22} = \lambda_1\mathbf{L}^T\mathbf{L} + (1-\alpha)(\nabla^2)^T\nabla^2 + \gamma_0\mathbf{I},\end{cases}$$

and

$$\begin{cases}\mathbf{h}_1 = \lambda_1\mathbf{L}^T\mathbf{y} + \gamma_1\nabla^T[\mathbf{p}^k + \tilde{\mathbf{p}}^k] + \gamma_0[\mathbf{q}^k + \tilde{\mathbf{q}}^k],\\ \mathbf{h}_2 = \lambda_1\mathbf{L}^T\mathbf{y} + \gamma_0[\mathbf{q}^k + \tilde{\mathbf{q}}^k],\end{cases}$$

where $\mathbf{I}$ is the identity matrix.

Aghamiry et al. (2018a) broke down the $2n \times 2n$ problem (3.63) into two smaller $n \times n$ systems and updates $\mathbf{m}_1$ and $\mathbf{m}_2$ in alternating mode at the expense of convergence speed (Aghamiry et al., 2018a, their eqs. 10 and 11). Instead, we solve here the original system exactly using a variable projection scheme, thus leading to faster convergence and more accurate results. From the first equation of (3.63), we find that

$$\mathbf{m}_2 = \mathbf{G}_{12}^{-1}[\mathbf{h}_1 - \mathbf{G}_{11}\mathbf{m}_1] \tag{3.64}$$

and plugging this into the second equation of (3.63) we get the following:

$$\mathbf{m}_1 = [\mathbf{G}_{11} - \mathbf{G}_{22}\mathbf{G}_{12}^{-1}\mathbf{G}_{11}]^{-1}[\mathbf{h}_2 - \mathbf{G}_{22}\mathbf{G}_{12}^{-1}\mathbf{h}_1]. \tag{3.65}$$

Interestingly, $\mathbf{L}$ is diagonal, implying that $\mathbf{G}_{12}$ is also diagonal. Thus we only need to solve an $n \times n$ system to estimate $\mathbf{m}_1$, from which $\mathbf{m}_2$ easily follows.

The subsubproblem (3.61b)

The sub-problem for $\mathbf{p}$, equation (3.61b), is a denoising problem and is straightforward to solve. Note that $\mathbf{p}$ has two components associated with the gradient in each direction:

$$\mathbf{p} = \begin{bmatrix} \mathbf{p}_x \\ \mathbf{p}_z \end{bmatrix}. \tag{3.66}$$

Equation (3.61b) is solved with a generalized proximity operator (Combettes and Pesquet, 2011) leading to

$$\mathbf{p}^{k+1} = \text{prox}_{\alpha/\gamma_1}(\mathbf{z}) = \begin{bmatrix} \xi \circ \mathbf{z}_x \\ \xi \circ \mathbf{z}_z \end{bmatrix}, \tag{3.67}$$

where

$$\mathbf{z} = \nabla \mathbf{m}_1^{k+1} - \tilde{\mathbf{p}}^k = \begin{bmatrix} \mathbf{z}_x \\ \mathbf{z}_z \end{bmatrix}, \tag{3.68}$$

and

$$\xi = \max(1 - \frac{\alpha}{\gamma_1 \sqrt{\mathbf{z}_x^2 + \mathbf{z}_z^2}}, 0). \tag{3.69}$$

The subsubproblem (3.61c)

The optimization problem (3.61c) also has an entrywise solution given by

$$\mathbf{q}^{k+1} = \text{proj}_{\mathcal{C}}(\mathbf{m}_1^{k+1} + \mathbf{m}_2^{k+1} - \tilde{\mathbf{q}}^k), \tag{3.70}$$

where the projection operator projects its argument onto the desired box $[\mathbf{m}_l, \mathbf{m}_u]$ according to $\text{proj}_{\mathcal{C}}(\bullet) = \min(\max(\bullet, \mathbf{m}_l), \mathbf{m}_u)$. Based on the above, the proposed ADMM based TT regularized IR-WRI algorithm is summarized in Algorithm 3. It should be noted that the total algorithm consists of two levels of iterations: an outer iteration (lines 4-13) given in (3.47) and an inner iteration (lines 5-11) given in (3.61) corresponding to the model subproblem (3.47b).

Numerical results, however, show that only one inner iteration suffices for convergence of the algorithm, hence significantly reducing the total computational cost (Goldstein and Osher, 2009; Aghamiry et al., 2019c). The main computational cost of the algorithm is on lines 4 and 5 where we need to solve the augmented PDE and the $n \times n$ sparse system given in equation 3.65. The other steps don't introduce significant computational overheads. The computational overhead introduced by compound regularizer compared to single regularizer results from the larger numerical bandwidth of the matrix $[\mathbf{G}_{11} - \mathbf{G}_{22}\mathbf{G}_{12}^{-1}\mathbf{G}_{11}]$ in equation 3.65 compared to that associated with a single regularizer, either $\mathbf{G}_{11}$ or $\mathbf{G}_{22}$.

Algorithm 3: ADMM based TT regularized IR-WRI algorithm

1: Initialize: set the dual variables $\mathbf{b}^0$, $\mathbf{d}^0$, $\tilde{\mathbf{p}}^0$ and $\tilde{\mathbf{q}}^0$ equal to $\mathbf{0}$
2: Input: $\mathbf{m}^0$ (initial model parameters)
3: **while** convergence criteria not satisfied **do**
4: $\quad \mathbf{u}^{k+1} \leftarrow \left[\frac{\lambda_0}{\lambda_1}\mathbf{P}^T\mathbf{P} + \mathbf{A}(\mathbf{m}^k)^T\mathbf{A}(\mathbf{m}^k)\right]^{-1}\left[\frac{\lambda_0}{\lambda_1}\mathbf{P}^T[\mathbf{d}+\mathbf{d}^k] + \mathbf{A}(\mathbf{m}^k)^T[\mathbf{b}+\mathbf{b}^k]\right]$
5: $\quad \mathbf{m}_1^{k+1} \leftarrow$ update according to eq. (3.65)
6: $\quad \mathbf{m}_2^{k+1} \leftarrow$ update according to eq. (3.64)
7: $\quad \mathbf{m}^{k+1} \leftarrow \mathbf{m}_1^{k+1} + \mathbf{m}_2^{k+1}$
8: $\quad \mathbf{p}^{k+1} \leftarrow \text{prox}_{\alpha/\gamma_1}(\nabla\mathbf{m}_1^{k+1} - \tilde{\mathbf{p}}^k)$
9: $\quad \mathbf{q}^{k+1} \leftarrow \text{proj}_{\mathcal{C}}(\mathbf{m}^{k+1} - \tilde{\mathbf{q}}^k)$
10: $\quad \tilde{\mathbf{p}}^{k+1} \leftarrow \tilde{\mathbf{p}}^k + \mathbf{p}^{k+1} - \nabla\mathbf{m}_1^{k+1}$
11: $\quad \tilde{\mathbf{q}}^{k+1} \leftarrow \tilde{\mathbf{q}}^k + \mathbf{q}^{k+1} - \mathbf{m}^{k+1}$
12: $\quad \mathbf{d}^{k+1} \leftarrow \mathbf{d}^k + \mathbf{d} - \mathbf{P}\mathbf{u}^{k+1}$
13: $\quad \mathbf{b}^{k+1} \leftarrow \mathbf{b}^k + \mathbf{b} - \mathbf{A}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}$
14: **end while**

3.3.6 Numerical examples

We assess the performance of our algorithm against 1D and 2D mono-parameter synthetic examples. In Table 3.1 we give different regularization functions which are applied for stabilizing the FWI solution. We start with zero-offset Vertical-Seismic-Profiling (VSP) examples (1D IR-WRI) where the targeted wave speed profiles are selected from well-documented 2D benchmark subsurface velocity models in exploration seismic. To tackle more realistic applications, we proceed with a target of the 2D challenging 2004 BP salt model (Billette and Brandsberg-Dahl, 2004) with noiseless and noisy data when a crude initial model and realistic frequencies are used as starting points.

Performance comparison using 1D test on benchmark models

First, we assess the performance of our regularized IR-WRI against 1D mono-parameter synthetic examples when the true models are 100 vertical profiles selected from the 2004 BP salt (Billette and Brandsberg-Dahl, 2004), Marmousi II (Martin et al., 2006), SEG/EAGE overthrust (Aminzadeh et al., 1997), SEG/EAGE salt (Aminzadeh et al., 1997) and synthetic Valhall

(Prieux et al., 2011) benchmark velocity models (we extracted 20 profiles from each benchmark model). For all of the experiments, a single source is used at the surface and the receivers are evenly deployed along the entire profile. A single frequency, whose value is set so that the reduced-space inversion is prone to cycle skipping, is considered for inversion. The model dimension, the inverted frequency and the receiver spacing are outlined for each model in Table 3.2. We perform forward modeling with a 3-point $\mathcal{O}(\Delta x^2)$ staggered-grid finite-difference stencil and PML absorbing boundary conditions at the two ends of the model. The starting model for IR-WRI is a homogeneous velocity model in which the velocity is the mean value of each profile. We set the penalty parameters according to the guideline given in Appendix C (section 3.3.10). Moreover, for a fair comparison of the compound regularizers (JTT, TT and TGV), we select for each of them the optimum value of α among a range of preset values that minimizes the error in the models estimated by the IR-WRI. Also, we set the parameter bounds $\mathbf{m}_l$ and $\mathbf{m}_u$ equal to 50% and 150% of the minimum and maximum velocities of the true model, respectively. The monochromatic inversion is performed with noiseless data when a maximum number of iterations, equal to 100, is used as a stopping criterion. The average error of the estimated velocity profiles for the five benchmark models and the different regularizations are plotted in Fig. 3.13. In this paper, the model error is defined as the energy of the difference between the true model and the estimated one compared to the energy of the true model. The errors in each model for different regularizations are normalized to 1 for sake of clarity (the error of DMP regularizer is not shown because of its worse performance). Fig. 3.13 clearly shows that the compound regularizations based upon infimal convolution (TT and TGV) always behave better than the CC regularization and the single regularization functionals (TV and Tikhonov). To emphasize the effects of the different regularization functions, we plot some close-ups of the reconstructed profiles in Fig. 3.14. These results show that TT provides the most accurate reconstruction for the 2004 BP salt (Fig. 3.14a) and Overthrust (Fig. 3.14c) models. This is consistent with the fact that the velocity trends of these two models match well the piecewise smooth prior. In contrast, TGV behaves slightly better than TT for the Valhall model, whose velocity trend is the closest one to the piecewise linear prior (Fig. 3.14d). For Marmousi II (Fig. 3.14b), TT and TGV give similar results.

Table 3.1 – Different regularization functions.

Abbreviation	Expression of $\Phi(\mathbf{m})$
DMP	$\|\mathbf{m}\|_2^2$
Tikhonov	$\|\nabla^2\mathbf{m}\|_2^2$
TV	$\|\nabla\mathbf{m}\|_1$
JTT	$(1-\alpha)\|\nabla^2\mathbf{m}\|_2^2 + \alpha\|\nabla\mathbf{m}\|_1$
TT	$\min_{\mathbf{m}=\mathbf{m}_1+\mathbf{m}_2}\{(1-\alpha)\|\nabla^2\mathbf{m}_2\|_2^2 + \alpha\|\nabla\mathbf{m}_1\|_1\}$
TGV	$\min_{\mathbf{m}=\mathbf{m}_1+\mathbf{m}_2}\{(1-\alpha)\|\nabla^2\mathbf{m}_2\|_1 + \alpha\|\nabla\mathbf{m}_1\|_1\}$

Table 3.2 – Experimental setup of 1D model tests

	Length (km)	Inverted frequency (Hz)	Grid interval (m)	Receiver interval (m)
2004 BP salt	11.46	5	6	180
Marmousi II	3.75	12	5	85
Overthrust	4.6	12	20	120
SEG/EAGE salt model	4.2	10	20	120
Synthetic valhall	5.22	5	25	175

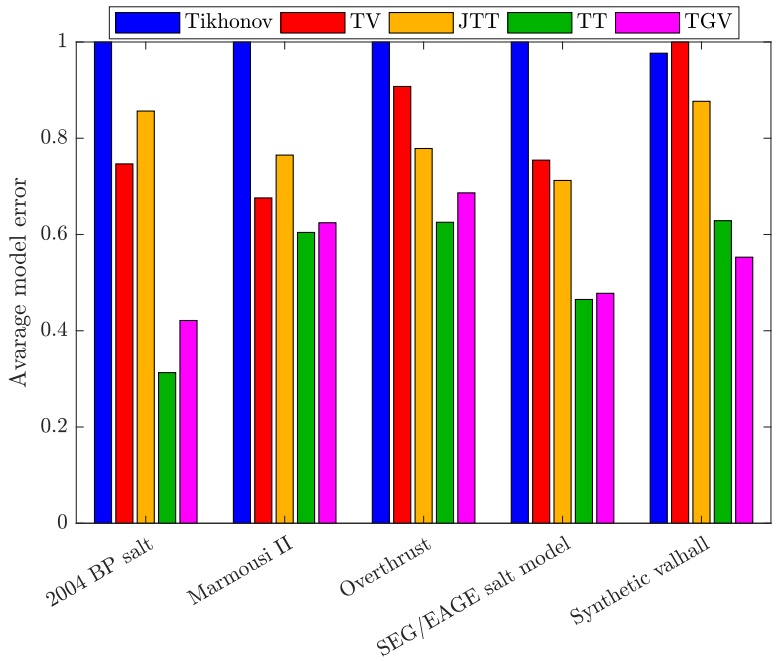

Figure 3.13 – Zero offset VSP test. Average model error in estimated 1D profiles for different velocity models and different regularization functions.

2004 BP salt model

We now consider a more realistic application with a target of the challenging 2004 BP salt model (Billette and Brandsberg-Dahl, 2004). The 2004 BP salt model is representative of the geology of the deep offshore Gulf of Mexico and mainly consists of a simple background with a complex rugose multi-valued salt body, sub-salt slow velocity anomalies related to over-pressure zones and a fast velocity anomaly to the right of the salt body. The selected subsurface model is 16250 m wide and 5825 m deep, and is discretized with a 25 m grid interval (Fig. 3.15a). We used 108 sources spaced 150 m apart on the top side of the model. We perform forward modeling with a staggered-grid 9-point finite-difference method (Chen et al., 2013)

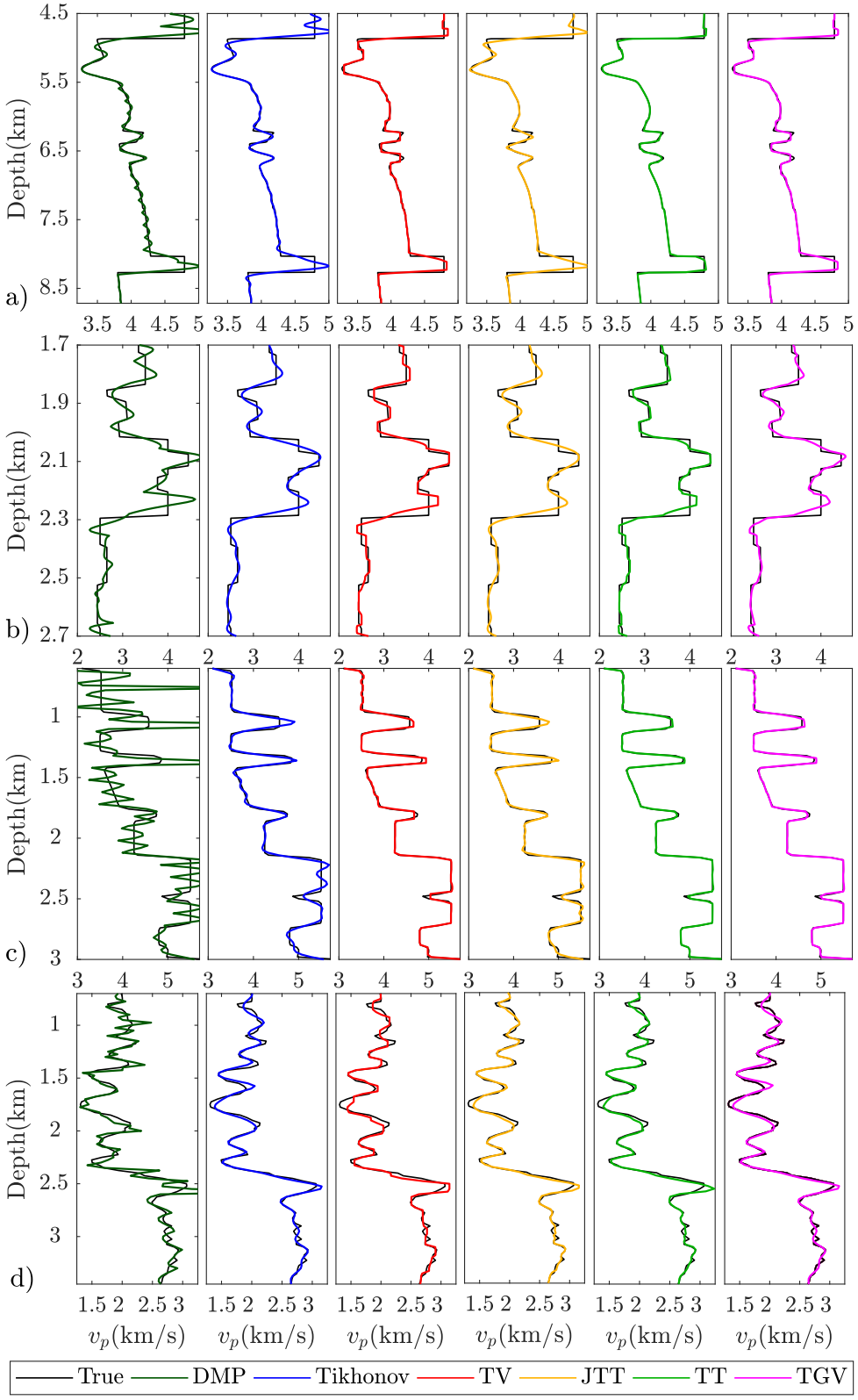

Figure 3.14 – Zero offset VSP test. Some part of (a) 2004 BP salt, (b) Marmousi II, (c) Overthrust, and (d) synthetic Valhall models estimated with different regularizations.

with PML boundary conditions along the four edges of the model. The source signature is a 10 Hz Ricker wavelet. A line of receivers with a 25 m spacing are deployed all along the surface leading to a long-offset stationary-receiver acquisition. We used small batches of two frequencies with one frequency overlap between two consecutive batches, moving from the low frequencies to the higher ones according to a classical frequency continuation strategy. We use $\alpha = 0.7$ and set the rest of hyperparameters according to the guidelines reviewed in appendix C (section 3.3.10). We set the parameter bounds $\mathbf{m}_l$ and $\mathbf{m}_u$ equal to the minimum and maximum velocity of the true model, respectively. The starting and final frequencies are 3 Hz and 13 Hz and the sampling interval in one batch is 0.5 Hz. The initial velocity model is a crude laterally-homogeneous velocity-gradient model with velocities ranging between 1.5 to 4.5 km/s (Fig. 3.15b). We start with inverting the first batch of frequencies ({3, 3.5} Hz) with noiseless data using a maximum number of iterations equal to 45 as a stopping criterion. To highlight the specific role of bound constraints, we activate them after 20 iterations. To emphasize the effect of regularization, the result of bound constrained IR-WRI with a simple DMP regularization is shown in Fig. 3.16a, while the bound-constrained IR-WRI results with Tikhonov and TV regularizations are shown in Figs. 3.16b and 3.16c, respectively. Although the TV reconstruction is better than the Tikhonov one, it provides a velocity model which is far from the optimal one. A direct comparison between the true model, the starting model and the estimated models is shown in Fig. 3.17a along three vertical logs at 2.5, 9.0 and 15.0 km distance (as depicted with dashed white lines in Fig. 3.15a). We continue with compound regularization results which are shown in Fig. 3.16d-f and Fig. 3.17b. Clearly, the TT regularizer better captures the long wavelengths of the salt body and the smooth subsalt background model. The joint evolution in iterations of the observation-equation ($\|\mathbf{Pu} - \mathbf{d}\|_2$) and wave-equation errors ($\|\mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{b}\|_2$), Fig. 3.18a-b, and the relative model errors in iterations, Fig. 3.18c-d, further confirm the relative performance of each regularizer during the inversion of the first frequency batch. Note the complex zigzag path followed by the inversion to jointly minimize the data residuals and the wave equation error in Fig. 3.18a-b. As already highlighted by Aghamiry et al. (2019c), this results from the dynamic balancing in iterations of the observation-equation and wave-equation constraints performed by the dual updates with the data and source residuals.

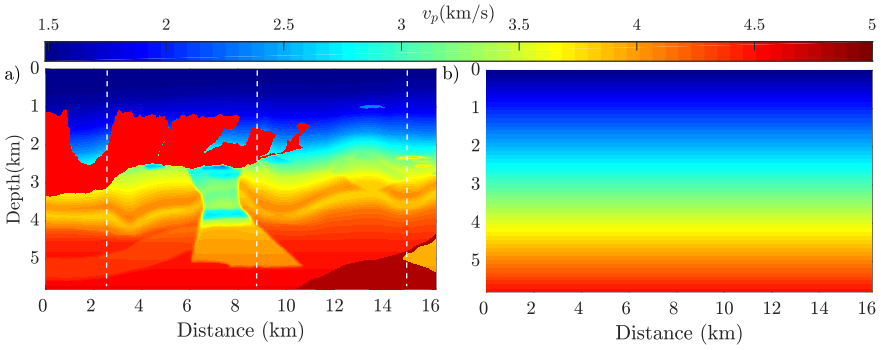

Figure 3.15 – 2004 BP salt case study. (a) True velocity model. The vertical dashed lines indicate the location of vertical logs of Figs. 3.17, 3.20 and 3.26. (b) The velocity-gradient initial model.

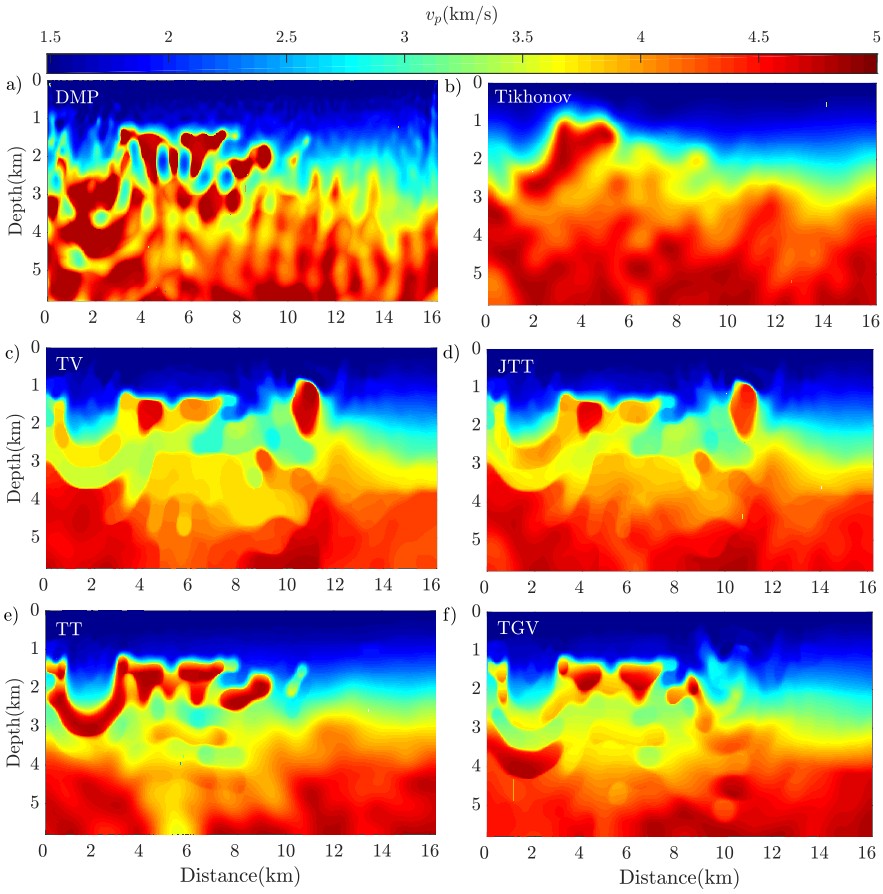

Figure 3.16 – Noiseless 2004 BP salt case study. Velocity models obtained after the {3, 3.5} Hz inversion when the velocity-gradient model (Fig. 3.15b) is used as initial model. (a-f) Bound constrained IR-WRI with (a) DMP, (b) Tikhonov, (c) TV, (d) JTT, (e) TT and (f) TGV regularization.

We continue the inversion at higher frequencies using the final models of the {3, 3.5} Hz inversion, Fig. 3.16a-f, as initial models when the stopping criteria is either $k_{max} = 15$ or

$$\|\mathbf{A}(\mathbf{m}^{k+1})\mathbf{u}^{k+1} - \mathbf{b}\|_2 \leq \varepsilon_b, \quad \|\mathbf{P}\mathbf{u}^{k+1} - \mathbf{d}\|_2 \leq \varepsilon_d, \tag{3.71}$$

where k_{max} denotes the maximum iteration count, ε_b=1e-3, and ε_d=1e-5. We perform three paths through the frequency batches to improve the IR-WRI results, using the final model of one path as the initial model of the next one (these cycles can be viewed as outer iterations of IR-WRI). The starting and finishing frequencies of the paths are [3.5, 6], [4, 8.5], [6, 13] Hz respectively, where the first element of each pair shows the starting frequency and the second one is the finishing frequency. The bound-constrained IR-WRI models obtained from noiseless data are shown in Fig. 3.19. As for the inversion of the first batch, direct comparison between the true model, the starting model and the estimated models are shown in Fig. 3.20 along three vertical logs at 2.5 km, 9.0 km and 15 km distance (vertical dashed lines in Fig. 3.15a). The TT and TGV regularizers lead to high-quality velocity models, that capture both the fine-scale structure

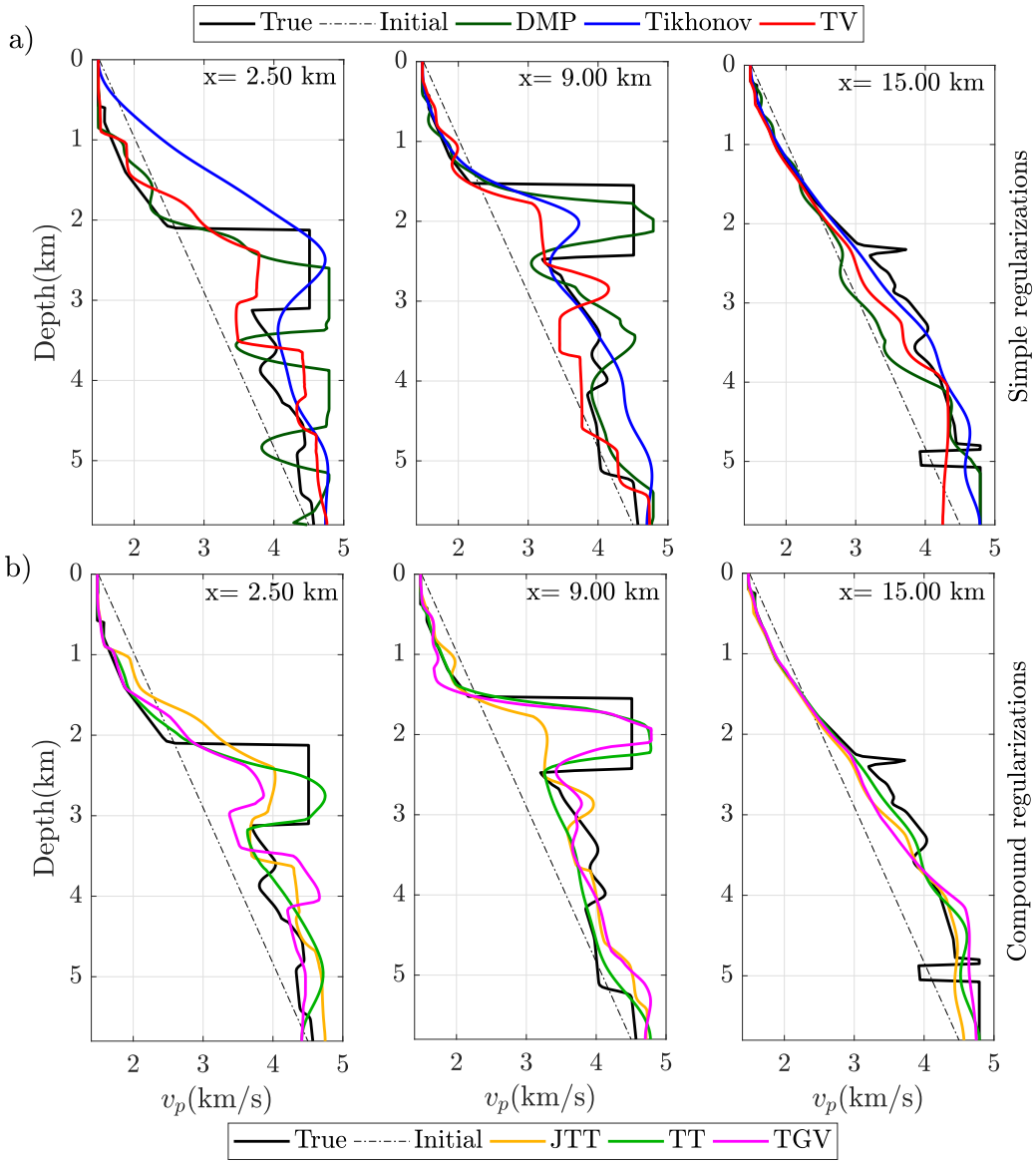

Figure 3.17 – Noiseless 2004 BP salt case study. Direct comparison along the logs shown in Fig. 3.15a between the true velocity model (black), the initial model (dashed line) and the estimated models obtained after the {3, 3.5} Hz inversion (Fig. 3.16) with (a) simple regularizations (DMP in olive-green, Tikhonov in blue and TV in red) and (b) compound regularizations (JTT in orange, TT in green and TGV in pink).

of the rugose large-contrast salt body and the high-velocity shallow anomaly on the right, as well as the smoother sub-salt background model including the low-velocity over-pressure structure. It is also worth noting the significant differences between the JTT and TT IR-WRI models in particular in the deep part of the model. Moreover, the number of iterations performed by IR-WRI for each regularization shows that TT has the best convergence speed (Table 3.3).

As a final quality control of the different IR-WRI models, it is instructive to check the

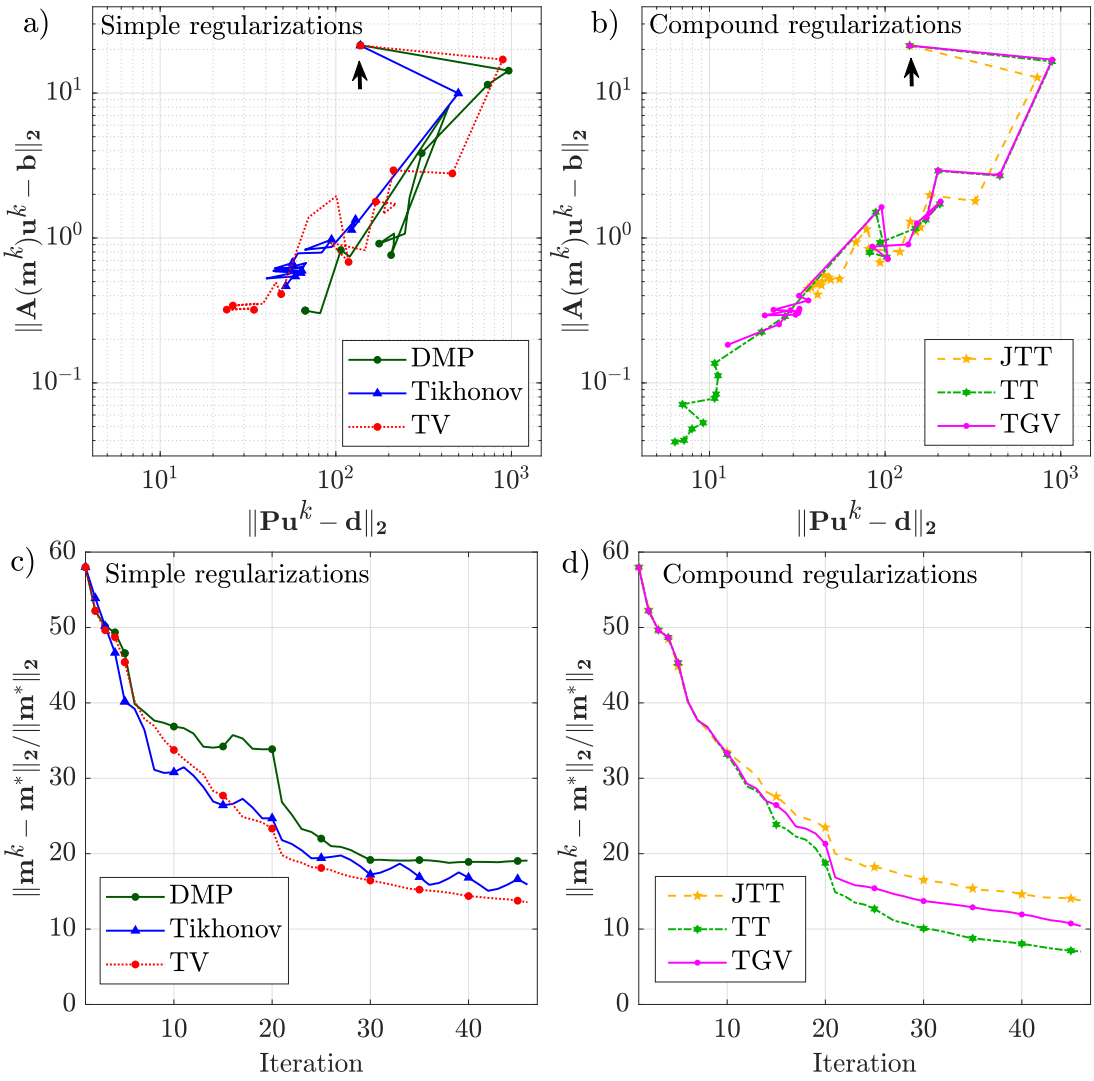

Figure 3.18 – Noiseless 2004 BP salt case study. {3, 3.5} Hz inversion with the velocity-gradient initial model. (a-b) convergence history of the algorithm in the ($\|\mathbf{P}\mathbf{u}^k - \mathbf{d}\|_2 - \|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b}\|_2$) plane for (a) simple regularizations and (b) compound regularizations. The black arrow points the starting point. (c-d) evaluation of $\|\mathbf{m}^k - \mathbf{m}^*\|_2/\|\mathbf{m}^*\|_2$ during the iteration where $\mathbf{m}^*$ is the true model. The panels (a) and (b) as well as (c) and (d) are plotted with the same horizontal and vertical scale.

wave-equation and data residuals left by the different regularization methods for the starting 3-Hz frequency (Fig. 3.21 and 3.22). The real part of wave-equation error (Fig. 3.21) and data residuals (Fig. 3.22) are plotted at the first and final iterations of the inversion. Both of the final data and source residuals suggest that the TT regularizer slightly outperforms the TGV counterpart at low frequencies. To further illustrate the ability of compound regularizations to manage the blocky and smooth components of the subsurface, we show separately the two model components of IC based regularizations ($\mathbf{m}_1$ and $\mathbf{m}_2$) estimated with TT (Fig. 3.23a-b) and TGV regularizations (Fig. 3.23c-d) as well as histograms of $\nabla\mathbf{m}_1$ and $\nabla^2\mathbf{m}_2$ (Fig. 3.24).

The trend of the reconstructed $\mathbf{m}_1$ and $\mathbf{m}_2$ shows that the IC based compound regularizations mostly succeeded in de-coupling the reconstruction of the blocky structure from that of the smooth background for TT regularization (Fig. 3.23a-b) and de-coupling the reconstruction of the blocky structure from that of the piecewise linear background for TGV regularization (Fig. 3.23c-d). This statement is further supported by the long-tail shape and the Gaussian shape of the histograms of the $\nabla\mathbf{m}_1$ and $\nabla^2\mathbf{m}_2$ components of TT regularization (Fig. 3.24a-b) and the long-tail shapes of the $\nabla\mathbf{m}_1$ and $\nabla^2\mathbf{m}_2$ components of TGV regularization (Fig. 3.24c-d).

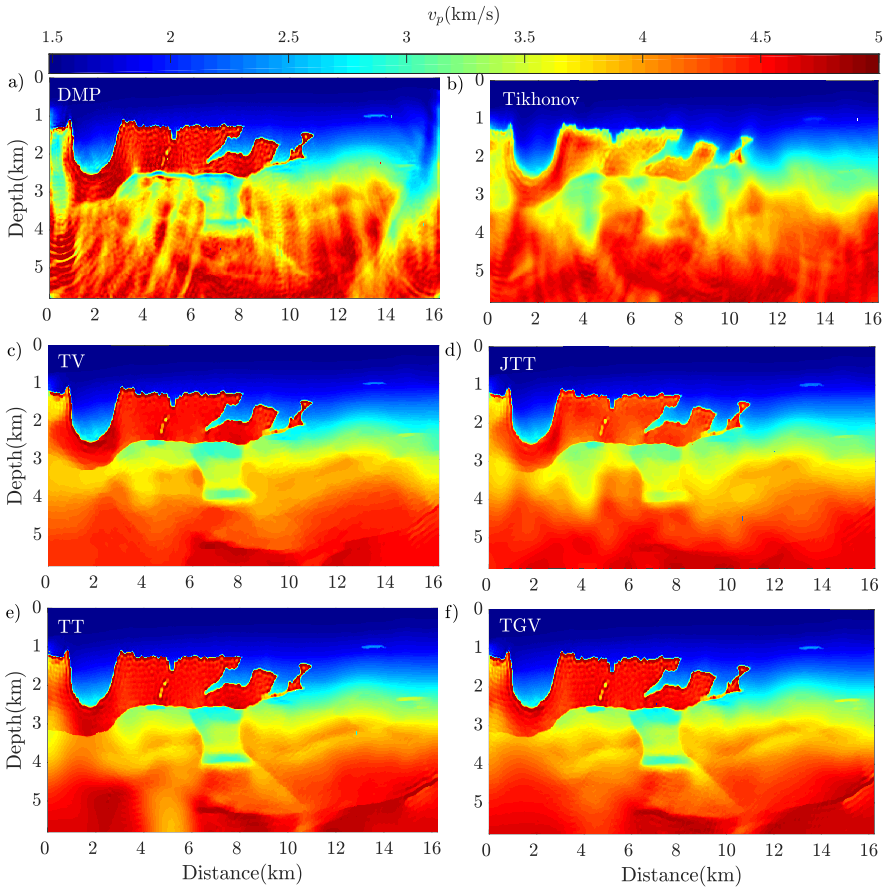

Figure 3.19 – Noiseless 2004 BP salt case study. Final inversion results with the velocity models of Fig. 3.16a-f as initial models. The display of the panels is the same as that of Fig. 3.16.

Table 3.3 – Total number of IR-WRI iterations for each regularization.

	DMP	Tikhonov	TV	JTT	TT	TGV
Noiseless data	426	448	399	415	361	394
Noisy data	285	293	264	271	270	274

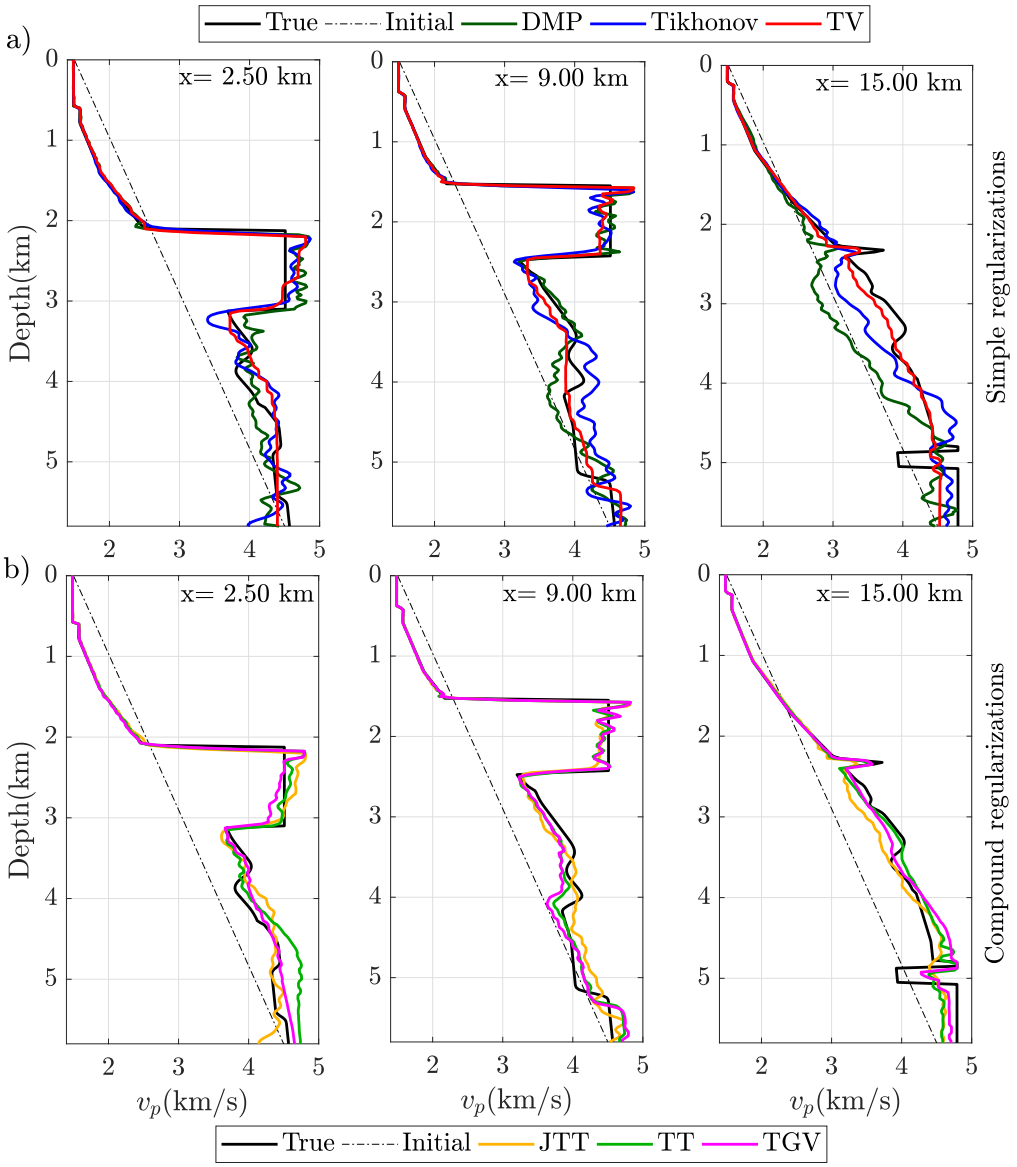

Figure 3.20 – Noiseless 2004 BP salt case study. Direct comparison of final inversion results with Fig. 3.16a-f as initial models. The panels are same as Fig. 3.17 for final results of Fig. 3.19.

We continue by assessing the resilience of the different regularization strategies to noise when data are contaminated with a Gaussian random noise with a SNR=10 db, where SNR is defined as

$$\text{SNR} = 20 \log \left(\frac{A_{signal}}{A_{noise}} \right), \tag{3.72}$$

in which A denotes root mean square (RMS) amplitude. We use the same setup and the same initial velocity model (Fig. 3.15b) as those used for the noiseless case. The stopping criterion is defined by (3.71), where ε_d is now set to the noise level. The final models of bound-constrained IR-WRI obtained from noisy data are shown in Fig. 3.25. The number of iterations performed

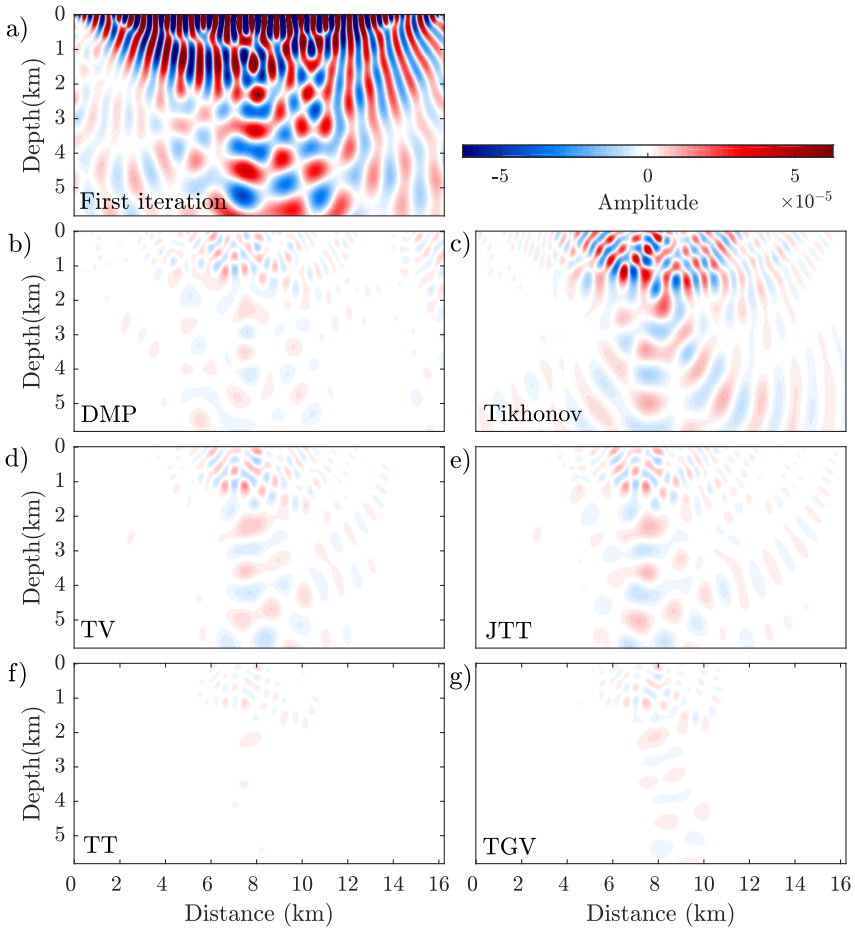

Figure 3.21 – Noiseless 2004 BP salt case study. For a source at x=8.12 km and the 3-Hz frequency: (a) Real part of wave-equation residual at first iteration of wavefield reconstruction, namely $(\mathbf{A}(\mathbf{m}^0)\mathbf{u}^1 - \mathbf{b})$. (b-g) Real part of wave-equation residual at the final iteration achieved respectively with DMP, Tikhonov, TV, JTT, TT, and TGV regularization.

by IR-WRI with the different regularizations are outlined in Table 3.3. As for the inversion of the first batch, direct comparison between the true model, the starting model and the estimated models are shown in Fig. 3.26 along three vertical logs at 2.5 km, 9.0 km and 15 km distance. The results further confirm that TT regularization provides the most reliable results and illustrate the resilience of this compound regularization to noise.

3.3.7 Conclusions

In this study, we first show how to efficiently implement different kinds of regularization and bound constraints in the wavefield reconstruction inversion method with the alternating direction method of multipliers (ADMM). Then, we show the capability of IR-WRI when equipped with compound Tikhonov and TV regularizations to accurately reconstruct large-contrast subsurface media when starting from a very crude initial model. This compound regularization is suitable for seismic imaging of the subsurface as it can often be represented by piecewise

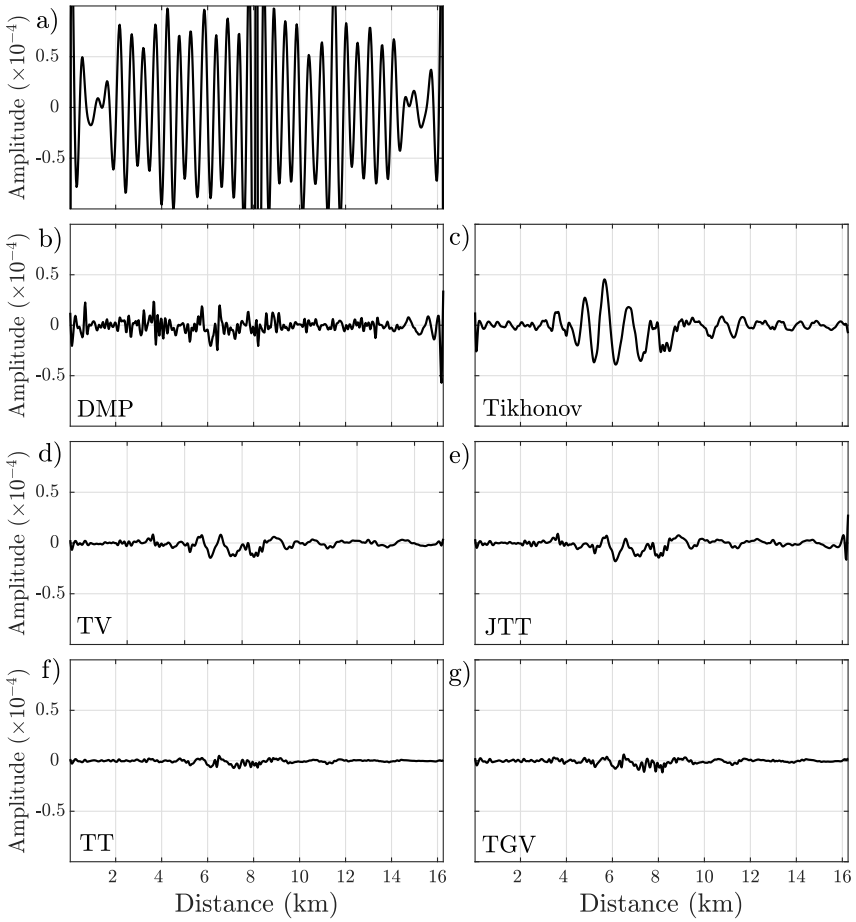

Figure 3.22 – Noiseless 2004 BP salt case study. For a source at x=8.12 km and the 3-Hz frequency: (a) Real part of data residual ($\mathbf{Pu} - \mathbf{d}$) at first iteration. (b-g) Real part of data residual at the final iteration achieved respectively with DMP, Tikhonov, TV, JTT, TT, and TGV regularization.

smooth media. We show that the infimal convolution (IC) of the Tikhonov and TV regularizers captures much more accurately the blocky and smooth components of the subsurface than the convex combination of the two regularizers. It also outperforms the Tikhonov and TV regularizers when used alone. We also show how the infimal-convolution regularizer can be efficiently implemented by jointly updating the smooth and blocky subsurface components through variable projection. Alternatively, TGV regularized IR-WRI can be a suitable tool to reconstruct piecewise linear media and provides similar results than TT IR-WRI. We conclude that such hybrid regularizations in the extended search-space IR-WRI potentially provide a suitable framework to reconstruct, without cycle skipping, large-contrast subsurface media from ultra-long offset seismic data. It should also find applications in other fields of imaging sciences such as medical imaging.

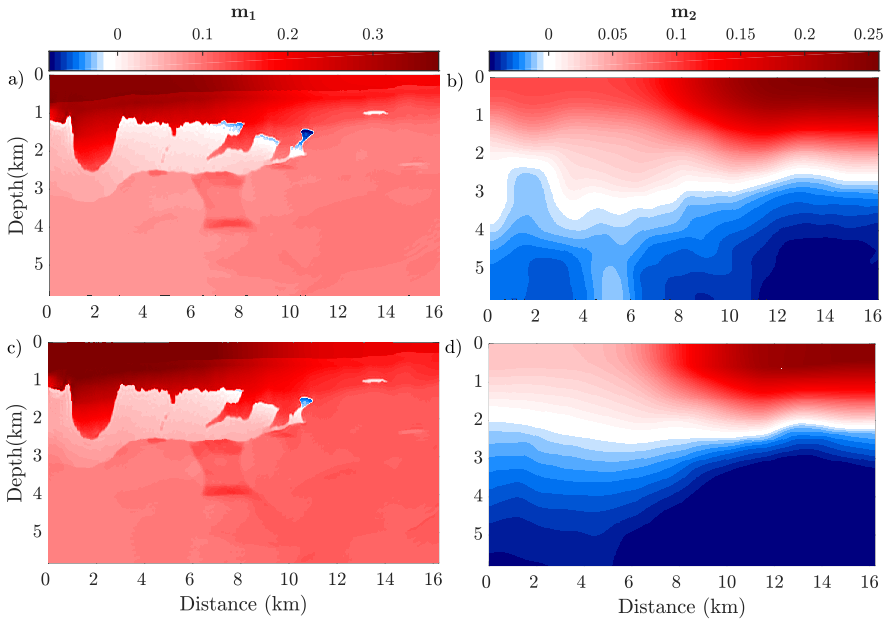

Figure 3.23 – Noiseless 2004 BP salt case study. Final blocky and smooth components, $\mathbf{m}_1$ and $\mathbf{m}_2$, reconstructed by IR-WRI with (a-b) TT and (c-d) TGV regularizations. The corresponding velocity models are shown in Figs. 3.19e and 3.19f. (a,c) $\mathbf{m}_1$. (b,d) $\mathbf{m}_2$. Note that $\mathbf{m}_1$ and $\mathbf{m}_2$ are parametrized with squared slownesses.

Acknowledgments

We would like to thank the Editor Mauricio D. Sacchi, Gian Matharu and two anonymous reviewers for their comments which help improving the manuscript. This study was partially funded by the SEISCOPE consortium (*http://seiscope2.osug.fr*), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXON-MOBIL, JGI, PETROBRAS, SCHLUMBERGER, SHELL, SINOPEC and TOTAL. This study was granted access to the HPC resources of SIGAMM infrastructure (http://crimson.oca.eu), hosted by Observatoire de la Côte d'Azur and which is supported by the Provence-Alpes Côte d'Azur region, and the HPC resources of CINES/IDRIS/TGCC under the allocation A0050410596 made by GENCI."

3.3.8 Appendix A: Reduced approach to solving (3.41)

The reduced approach, which is more commonly used for sake of computational efficiency, strictly enforces the PDE constraint at each iteration by projection of the full space onto the parameter search space, leading to the following unconstrained optimization problem (Virieux and Operto, 2009; Pratt et al., 1998; Plessix, 2006)

$$\min_{\mathbf{m}\in\mathcal{C}} \; \Phi(\mathbf{m}) + \frac{\lambda_0}{2}||\mathbf{P}\mathbf{A}^{-1}(\mathbf{m})\mathbf{b} - \mathbf{d}||_2^2, \tag{3.73}$$

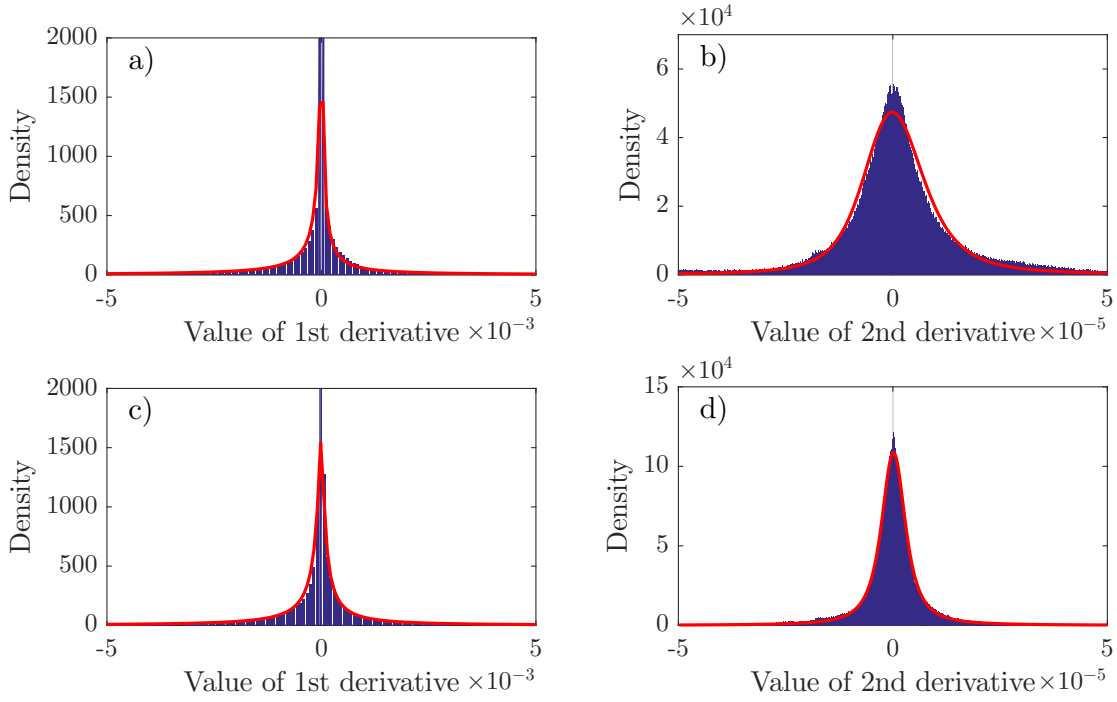

Figure 3.24 – Noiseless 2004 BP salt case study. Histograms of the model components shown in Fig. 3.23. The red lines show the probability density functions fitted to the histograms.

where $\lambda_0 > 0$ is the penalty parameter. A number of methods have been proposed to solve the optimization problems of the form (3.73), either for unregularized form $\Phi(\mathbf{m}) = 0$ (Pratt et al., 1998), or the regularized form (Asnaashari et al., 2013; Esser et al., 2018). Although this reduced approach is more computationally tractable than the full-space approach, the highly-oscillating nature of the inverse PDE operator $\mathbf{A}^{-1}$ makes the inverse problem highly nonlinear, and hence prone to convergence to a spurious local minima when the initial $\mathbf{m}$ is not accurate enough (Virieux and Operto, 2009; Symes, 2008). The extended approach described in this paper (section 3.3.4) is an alternative way which is more immune to local minima.

3.3.9 Appendix B: Simple regularizers

The two most widely used regularizers rely on the (squared) ℓ_2 and ℓ_1-norms. The squared ℓ_2-norm, defined as

$$\|\mathbf{x}\|_2^2 = \sum_{i=1}^{n} |\mathbf{x}_i|^2, \tag{3.74}$$

promotes smooth reconstruction, since the minimization of the squared value of components will penalize large components more severely than small ones.

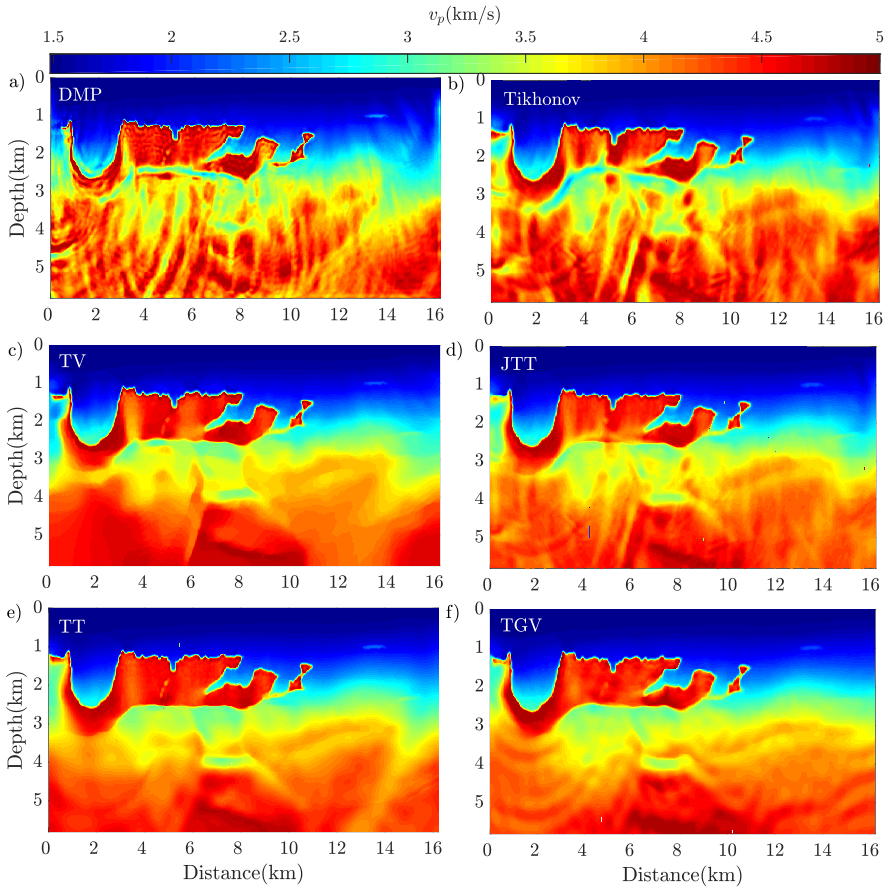

Figure 3.25 – Noisy 2004 BP salt case study. Final inversion results with Fig. 3.15b as initial model. (a-f) bound constrained IR-WRI with (a) DMP, (b) Tikhonov, (c) TV, (d) JTT, (e) TT and (f) TGV regularization.

In contrast, the ℓ_1-norm, defined as

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |\mathbf{x}_i|, \tag{3.75}$$

promotes sparse reconstruction (with many zero components), since the minimization of the absolute value of components will penalize small components more severely than the large counterparts.

The priors can be defined under a suitable transformation. For example, one may minimize the ℓ_1- or ℓ_2-norms of the first and/or second order differences of the model. The first order forward differences for discrete scalar field f in x- and z-direction are denoted by $\nabla_x f$ and $\nabla_z f$, with

$$\begin{cases} (\nabla_x f)_{i,j} = f_{i,j} - f_{i,j-1}, \\ (\nabla_z f)_{i,j} = f_{i,j} - f_{i-1,j}, \end{cases} \tag{3.76}$$

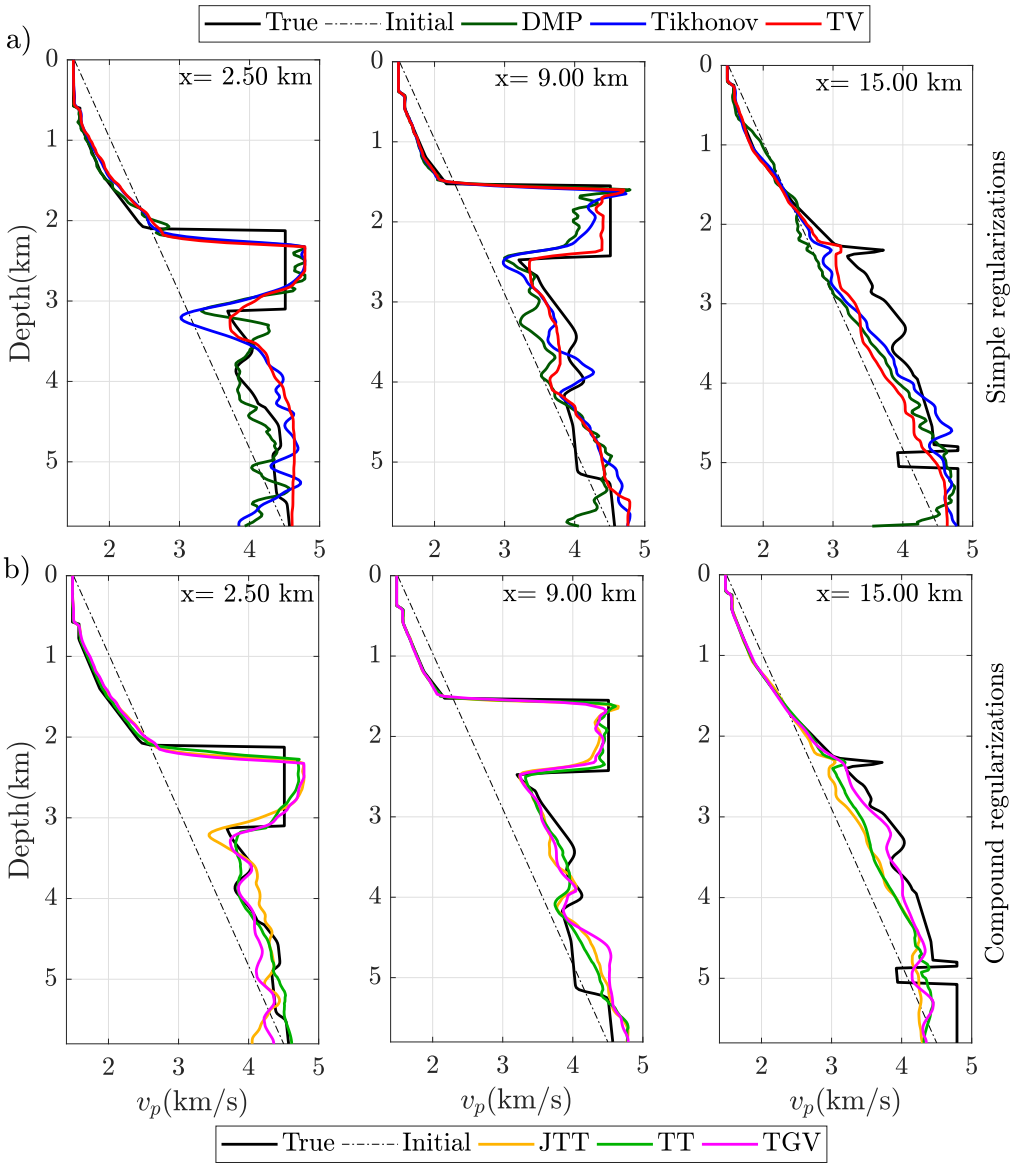

Figure 3.26 – Noisy 2004 BP salt case study. Direct comparison (along the logs shown in Fig. 3.15a) between the true velocity model (black), the initial model (dashed line) and the estimated models with (a) simple regularizations (DMP in olive-green, Tikhonov in blue and TV in red) and (b) compound regularizations (JTT in orange, TT in green and TGV in pink).

with appropriate boundary conditions, where i and j run over the domain of the model parameters. Accordingly, the discrete first order operator in 2D is defined as $\nabla = \begin{bmatrix}\nabla_x^T & \nabla_z^T\end{bmatrix}^T$ with

$$(|\nabla f|)_{i,j} = \sqrt{(\nabla_x f)_{i,j}^2 + (\nabla_z f)_{i,j}^2}. \tag{3.77}$$

The squared ℓ_2-norm of $|\nabla f|$ gives the first order Tikhonov regularization (Tikhonov et al., 2013), which returns a flat regularized model (with a small gradient), while its ℓ_1-norm gives the total variation regularization (Rudin et al., 1992), which returns a piecewise constant model

(with a sparse gradient).

Analogously, the second order forward differences are denoted by $\nabla_{xx} f$ and $\nabla_{zz} f$, with

$$\begin{cases} (\nabla_{xx} f)_{i,j} = f_{i,j-1} - 2f_{i,j} + f_{i,j+1}, \\ (\nabla_{zz} f)_{i,j} = f_{i-1,j} - 2f_{i,j} + f_{i+1,j}, \end{cases} \tag{3.78}$$

with appropriate boundary conditions, where again i and j run over the domain of the model parameters. Accordingly, the discrete second order operator is defined as $\nabla^2 = \begin{bmatrix} \nabla_{xx}^T & \nabla_{zz}^T \end{bmatrix}^T$ with

$$(|\nabla^2 f|)_{i,j} = \sqrt{(\nabla_{xx} f)_{i,j}^2 + (\nabla_{zz} f)_{i,j}^2}. \tag{3.79}$$

The squared ℓ_2-norm of $|\nabla^2 f|$ gives the second-order Tikhonov regularization, which returns a smooth regularized model (with a small Laplacian), while its ℓ_1-norm gives the second order TV regularization, which returns a piecewise linear model (with a sparse Laplacian).

Mixed second-order differences can also be constructed as $\nabla_{xz} f \equiv \nabla_z \nabla_x f$ with

$$(\nabla_{xz} f)_{i,j} = f_{i,j} - f_{i,j-1} - f_{i-1,j} + f_{i-1,j-1}. \tag{3.80}$$

A discrete second-order operator, which includes mixed differences is defined as $\nabla^2 = \begin{bmatrix} \nabla_{xx}^T & \sqrt{2}\nabla_{xz}^T & \nabla_{zz}^T \end{bmatrix}^T$ with

$$(|\nabla^2 f|)_{i,j} = \sqrt{(\nabla_{xx} f)_{i,j}^2 + 2(\nabla_{xy} f)_{i,j}^2 + (\nabla_{yy} f)_{i,j}^2}, \tag{3.81}$$

which equals the Frobenius norm of the Hessian matrix (Lefkimmiatis et al., 2012; Gholami and Sacchi, 2013; Gholami and Naeini, 2019).

3.3.10 Appendix C: Parameter tuning

Here, we provide some guidelines to tune the different hyperparameters in the regularized IR-WRI method. The reader is also referred to Aghamiry et al. (2019b) for more details. We start with γ_1 (step 8 of Algorithm 3), which controls the soft thresholding performed by the TV regularization, and set it equal to $0.02 \times \max |\nabla \mathbf{m}_1^k - \tilde{\mathbf{p}}^k|$. This tuning can be refined according to prior knowledge of the geological structure, coming from well logs for example. Also, we use the same weight for the bound constraints and the TV regularization: $\gamma_0 = \gamma_1$. Once we set γ_1, we define λ_1 such that γ_1/λ_1 is a percentage of mean absolute value of the diagonal coefficients of $\mathbf{L}^T\mathbf{L}$. Parameter λ_1 may be increased during iterations to reduce the weight of TV regularization and bound constraints near the convergence point. Finally, we set λ_0 such that $\lambda = \lambda_1/\lambda_0$ is a small fraction of the highest eigenvalue ξ of the normal operator $\mathbf{A}(\mathbf{m})^{-T}\mathbf{P}^T\mathbf{P}\mathbf{A}(\mathbf{m})^{-1}$ during the wavefield reconstruction subproblem according to the criterion proposed by van Leeuwen and Herrmann (2016). In all the numerical tests, we use $\lambda =$ 1e-5ξ and $\lambda =$ 1e-3ξ for noiseless and noisy data, respectively.

3.4 Application of sparse regularization to compressive sensing for sparse acquisition

Ultra-long offset stationary-recording acquisitions as those carried out with Ocean Bottom Nodes (OBN) are emerging in the industry for deep water exploration (Beaudoin and Ross, 2007; Shen et al., 2018; Huang et al., 2019). Indeed, ultra-long offsets are obtained with a limited pool of autonomous instruments at the expense of the receiver spacing (acquisition sampling). Moreover, even the shot dimension may be downsampled in the cross direction to maintain reasonable acquisition time. This subsampling of the acquisition parameters can translate in a subsampling of the spectral components of the subsurface model below the Nyquist rate, which translates itself to wraparound of subsurface heterogeneities in the spatial domain (Plessix and Mulder, 2004). In this section, I discuss sparsity-promoting regularization in IR-WRI as a tool to mitigate the footprint of sparse acquisition in FWI.
Ultra-long offsets are beneficial for FWI for at least two reasons which are resolution and redundancy. Long offsets provide a broad illumination of scattering, dip and azimuth angles (Figure 3.27), which translates into an extension of the wavenumber spectrum towards long wavelengths and steep dips (Figure 3.28a). Furthermore, as maximum offset increases, frequency and scattering angles sample wavenumbers more and more redundantly (Figure 3.28b).
To illustrate further these points, we remind that the wavenumber vector that is mapped at a given position in a 2D subsurface model is related to the frequency f, the scattering angle θ and the dip angle ϕ by (Thierry et al., 1999):

$$\mathbf{k} = \mathbf{k}\left[\cos(\phi), \sin(\phi)\right] \quad \text{with} \quad \mathbf{k} = \frac{2f}{c}\cos(\theta/2), \tag{3.82}$$

where c is the local wavespeed (Figure 3.27). Equation 3.82 shows that the sampling of the

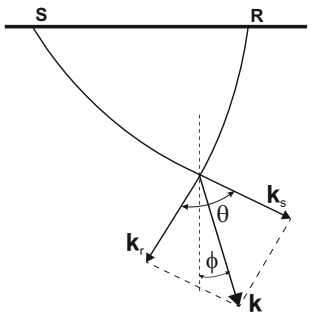

Figure 3.27 – Resolution analysis (diffraction tomography). Two rays starting from the source S and the receiver R connects a diffractor point in the subsurface with a scattering angle θ. The local wavenumber vector at the diffractor position shown by $\mathbf{k}$ while ϕ defines the orientation of $\mathbf{k}$ relative to source and receiver wavenumber vectors $\mathbf{k}_s$ and $\mathbf{k}_r$.

wavenumber modulus k is controlled by the frequency and θ samplings, the samplings of the angle θ being mostly controlled by the sampling of the finer acquisition dimension, while the sampling of ϕ is controlled by the source and receiver spacings. Also, it shows that frequencies

and scattering angles can sample the wavenumber spectrum of the subsurface in a redundant way, namely, several pairs of frequency and scattering angle sample the same wavenumber. I show the wavenumber components that would be injected in a smooth background model by FWI of 7 Hz monochromatic data sets with different maximum offsets in figure 3.28a. The wavenumber spectrum is extended toward the low wavenumber as the result of increasing the maximum offset and, hence improving the FWI resolution. Also, the wavenumber components of three different monochromatic data sets with same maximum offsets are shown in figure 3.28b. The figure shows that three frequencies are sufficient to continuously sample the domain of interest of the wavenumber space (the one spanned by the wavenumber bandwidth of the subsurface). This redundancy can be reduced by subsampling frequencies to design

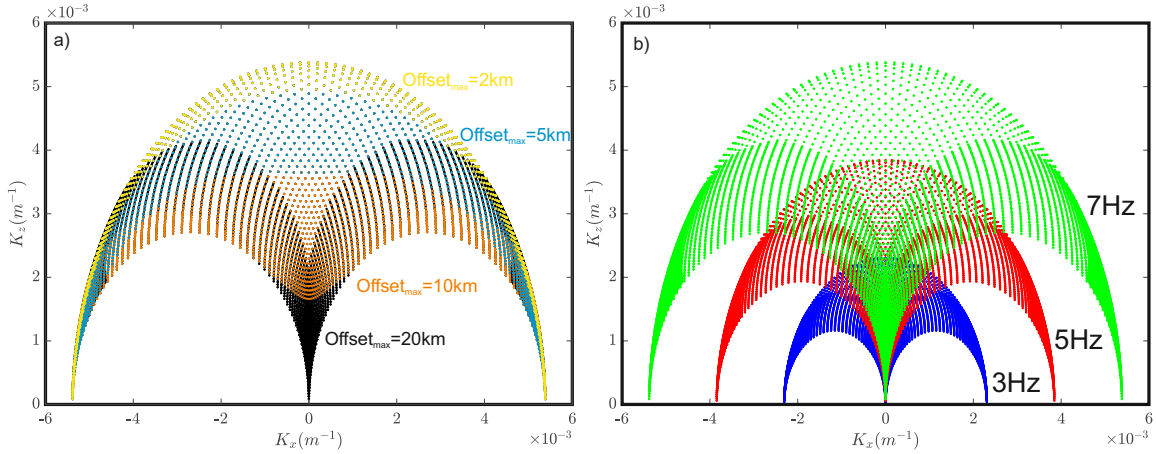

Figure 3.28 – Local wavenumber spectrum at the diffractor position in 3 km depth when a fixed-spread acquisition is used. (a) for 7 Hz monochromatic data set spanned by four different maximum offsets 2 (yellow), 5 (blue), 10 (orange) and 20 km (black). (b) for three discrete frequencies 3 (blue), 5 (red) and 7 Hz (green) and with 20 km maximum offset.

computationally-efficient frequency-domain FWI (Pratt et al., 1998; Sirgue and Pratt, 2004). Alternatively, this redundancy can be preserved to sample each wavenumber with a wide range of scattering angles, hence mitigating parameter cross-talks during multi-parameter reconstruction.

In a classical reduced-space formulation, the full search space is projected onto the parameter space by enforcing the closed-form expression of the wavefields as a function of the parameters in the data-misfit objective function. This variable projection makes FWI highly sensitive to cycle skipping. Many data-driven continuation approaches have been proposed to mitigate this non-linearity, the most classical one proceeding from low frequencies to higher ones. Furthermore, efficient frequency-domain FWI have been designed by limiting the inversion to a few discrete frequencies when long-offset stationary-receiver acquisitions provide a broad aperture illumination of the subsurface. This frequency decimation is tuned such that the redundancy of the wavenumber coverage is decreased, while preserving a wavenumber sampling that fulfills the Nyquist criterion and prevent aliasing accordingly (Sirgue and Pratt, 2004).

Also, equation 3.82 shows that a combination of coarse frequency, source and receiver samplings can easily subsample the two polar coordinates of $\mathbf{k}$, leading to wraparound artifacts in

space. I show the wavenumber components that would be injected in a smooth background model by FWI of 7 Hz monochromatic data set with maximum offset 15 km in Figure 3.29 when two different source and receiver spacings are used. Figure 3.29a shows the wavenumber spectrum for 1 km source and receiver spacings, while Figure 3.29b is for 100 m source and receiver spacings. This figure clearly shows for a surface acquisition how the wavenumber spectrum is sparsified when the source and receiver spacings increase from 100 m to 1 km.
Ultra-long offset acquisition implemented with multi-component Ocean Bottom Nodes (OBN)

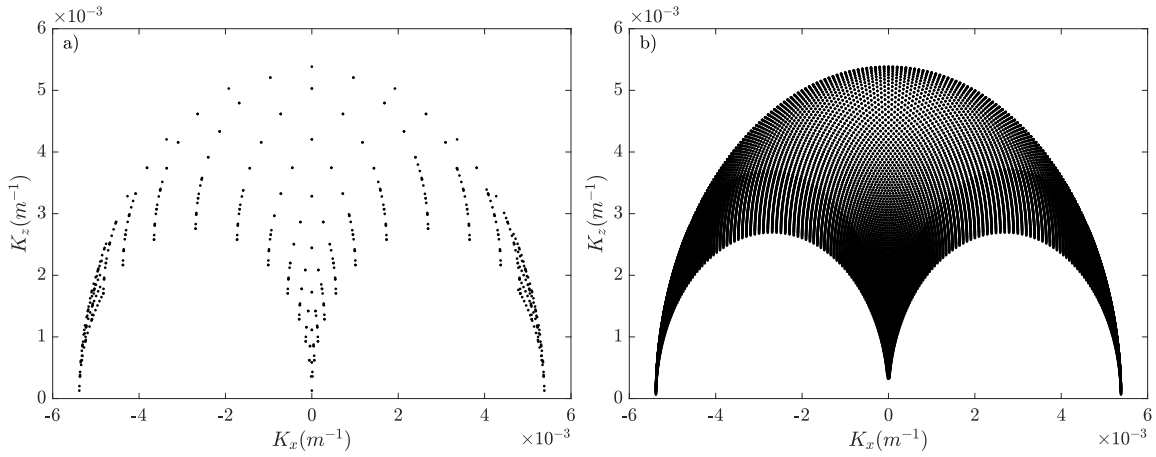

Figure 3.29 – Local wavenumber spectrum at the diffractor position in 3 km depth for 7 Hz monochromatic data set when a fixed-spread acquisition maximum offset 15 km is used. (a-b) source and receiver interval are equal to 1 km (a) and 100 m (b).

is a suitable technology for deep offshore subsalt imaging by FWI because of undershooting the deepest targeted structures with diving waves. Ultra-long offsets can be only designed via sparse stationary-recording acquisitions such as OBN (Vermeer, 2012). These acquisition designs raise however two major issues, cycle skipping and undersampling of the wavenumber spectrum. The cycle skipping is exacerbated by long propagation distances as shown by the condition that should be satisfied by the relative traveltime error. This condition reads as

$$\frac{\Delta t}{L} < \frac{1}{N_\lambda}, \tag{3.83}$$

where Δt is traveltime error, L propagation time, N_λ number of propagated wavelengths (Pratt, 2008). The subsampled acquisition parameters can potentially translate into subsampling of the FWI kernel resulting in spectral leakage in the model domain.
In this section, I want to design a FWI technology which is resilient to cycle skipping as well as subsampling of the acquisition parameters. Regardless of cycle skipping issue that can be solved using WRI, the subsampled wavenumber issue must be addressed. There are mainly three strategies to manage these sampling issues. The first is to recover finely-sampled data by means of compressive sensing (Donoho, 2006a,b; Candès et al., 2006; Candès and Wakin, 2008) and then perform FWI on the interpolated dataset (Herrmann, 2010). Indeed, this approach is demanding in terms of computational cost and storage. The second is to perform FWI of the

sparse data followed by denoising of the FWI model. This projection approach is however not efficient. The third one that we propose is to interface sparsity-promoting regularization in the optimization algorithm to manage subsampling effects on the fly during FWI (Li et al., 2016). The ability of BTT-regularized IR-WRI (TT regularized and bound constrained IR-WRI) for imaging large-contrast media from stationary-recording acquisition was shown in Aghamiry et al. (2020b). The resilience to cycle skipping makes IR-WRI amenable to ultra-long offset stationary-recording acquisitions. In this section we are going to show how the blockiness and sparsity-promoting $\ell 1$ TV regularization embedded in BTT-regularized IR-WRI efficiently removes aliasing artifacts in the subsurface model, and narrow the FWI null space resulting from incomplete illumination when complex large-contrast media are reconstructed from sparse long-offset acquisitions.

3.4.1 Numerical results

Removing aliasing artifacts: illustration with a spike test

To illustrate the potential of sparsity-promoting TV regularization to filter out aliasing artifacts in the subsurface model (i.e., wavenumber subsampling along some spatial directions) generated by sparse acquisitions, we image a point diffractor in a $2 \times 2\ km^2$ homogeneous background model (Fig. 3.30). The velocities at the point diffractor and in the homogeneous background are 5 km/s and 1.75 km/s, respectively. For all numerical examples, the Helmholtz equation is solved with a 9-point stencil finite-difference method implemented with anti-lumped mass and PML absorbing boundary conditions. The source signature is a Ricker wavelet with a 15 Hz dominant frequency. We start the inversion from the homogeneous background model and invert simultaneously frequencies between 2 to 34 Hz with 2 Hz interval. We consider noiseless data and perform 20 iterations. We first design a full-illumination albeit sparse acquisition centered on the scatterer with 8 sources and 180 receivers regularly spaced along a circle of radius 1 km. The estimated model with IR-WRI without any regularization is shown in Fig. 3.30a, while the result of this test with TV regularization is shown in Fig. 3.30b. The results show how TV regularization removes the moderate aliasing and focus the perturbation at the diffractor position. To strengthen aliasing, we repeat the previous test with only 8 receivers regularly deployed along the circle. The estimated models without and with TV regularization are shown in Fig. 3.30c-3.30d, respectively. Again, the TV regularization efficiently removes wraparound artifacts. Finally, we test a surface acquisition using one source in the middle of the model and a line of receivers spaced 20 m apart. The results of IR-WRI without regularization shows the smearing of the diffractor point resulting from the acquisition truncation and the spatial wraparound of the diffractor point with a period of around 1 km ($\approx 1/\Delta k_{z_{max}} = v_{p_{diff}}/(2\Delta f) = 5000/4 = 1.25\ km$) resulting from the coarse frequency sampling and the narrow scattering angle illumination. The TV regularization cancels out the undesired truncation and wraparound effects to focus the perturbation along a small horizontal reflector segment related to the Fresnel zone. Also, the $\ell 1$ regularization fosters the reconstruction of intermediate vertical wavenumbers, which belongs to the null space of the un-regularized IR-WRI (inset in Fig. 3.30f).

2004 BP salt model

For a more realistic test, we consider a target of the 2004 BP salt model located on the left side of the model (Figure 3.15a). The 2004 BP salt model consists of a simple background

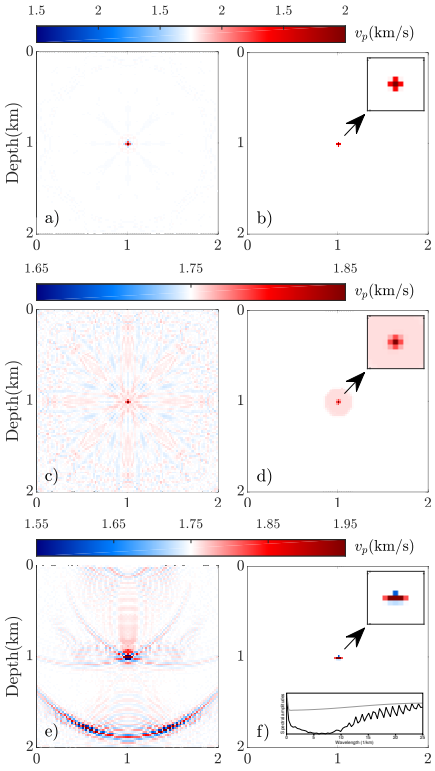

Figure 3.30 – Spike test. IR-WRI results (a,c,e) without regularization (b,d,f) with TV regularization. (a-b) Circular acquisition with 8 sources and 180 receivers. (c-d) Same as (a-b) for 8 receivers. (e-f) Single-source surface acquisition. In (f), the inset shows the spectral amplitudes of the vertical profile cross-cutting the reconstructed spikes shown in (e) and (f) (black and gray curves, respectively). Note the reconstruction of the intermediate wavenumbers fostered by $\ell 1$ regularization in the gray curve and aliasing in the black curve.

with a complex rugose multi-valued salt body, sub-salt slow velocity anomalies related to over-pressure zones and a fast velocity anomaly to the right of the salt body. The selected subsurface model is 16250 m wide and 5825 m deep and is discretized with a 25 m grid interval. We perform forward modeling with a staggered-grid 9-point finite-difference method with PML boundary conditions using a 10 Hz Ricker wavelet as source signature. We used small batches of two frequencies with one frequency overlap between two consecutive batches, moving from the low frequencies to the higher ones according to a classical frequency continuation strategy. The starting and final frequencies are 3 Hz and 13 Hz, and the sampling interval in one batch is 0.5 Hz. We perform three paths through the frequency batches to improve the IR-WRI results, using the final model of one path as the initial model of the next one (these cycles can be viewed as outer iterations of IR-WRI). The starting and finishing frequencies of the paths are [3,6], [4,10], [6,13] Hz respectively, where the first element of each pair shows the starting frequency and the second one shows the finishing frequency.

We consider five regular source devices on the top side of the model with 66, 34, 16, 12, and 8 sources, respectively. First, we used 131 receivers regularly spaced 125 m apart on the top side of the model. We perform two IR-WRIs for noiseless data: without TT regularization but with bound constraints (Fig. 3.31, left column) and with BTT regularization (Fig. 3.31, right column). The hyper parameters in BTT IR-WRI are set following the guidelines provided in Aghamiry et al. (2019b) and 10 iterations are performed per frequency batch. Therefore, all the results are obtained with the same number of iterations. Using the same setup, we repeat these tests with only 66 receivers regularly spaced 250 m apart (Fig. 3.32). We proceed by repeating

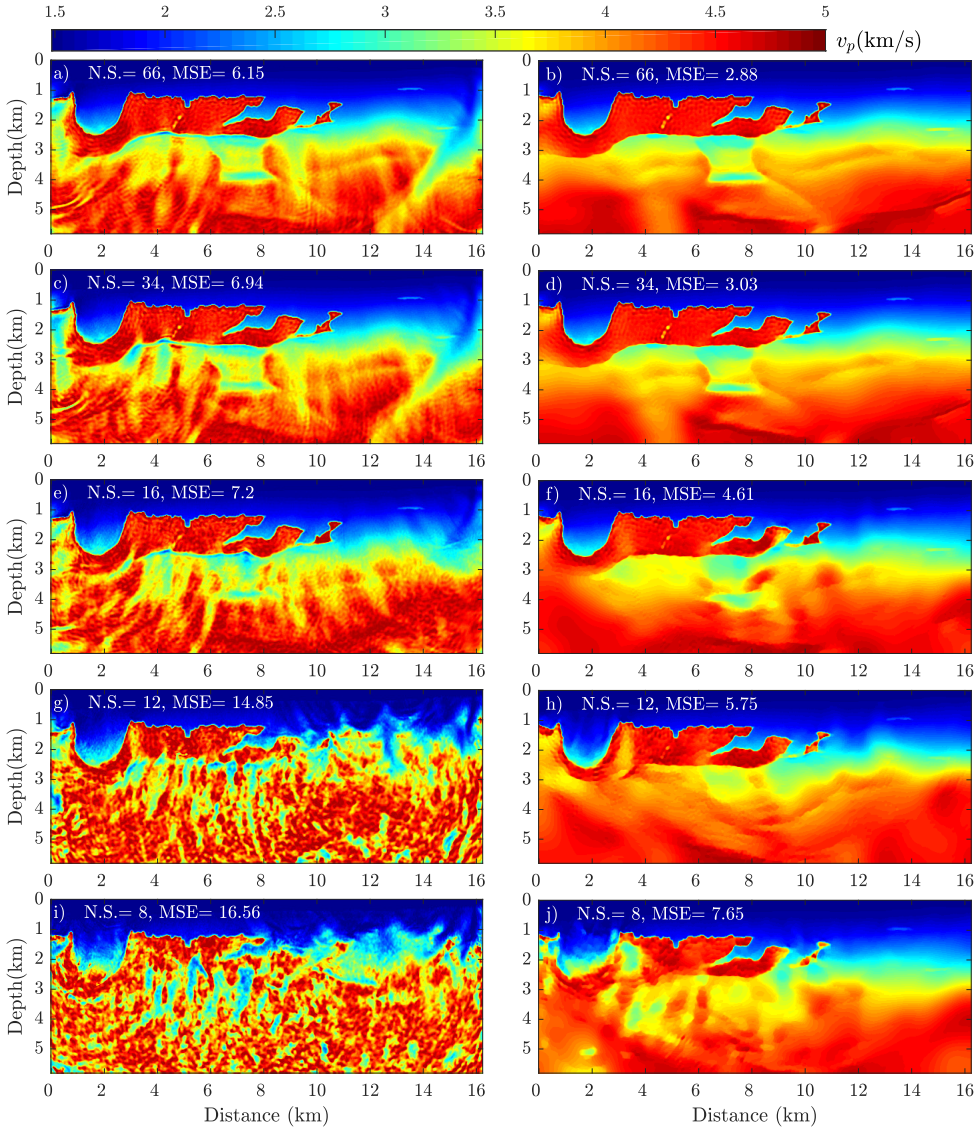

Figure 3.31 – 2004 BP test. IR-WRI results for different number of sources. The recording system involves 131 receivers regularly spaced 125 m apart. (a,c,e,g,i) Bound constrained IR-WRI with 66(a), 34(c), 16(e), 12(g) and 8(i) sources. (b,d,f,h,j) Same as (a,c,e,g,i) with BTT regularization. The number of sources and estimated model error are provided in each panel.

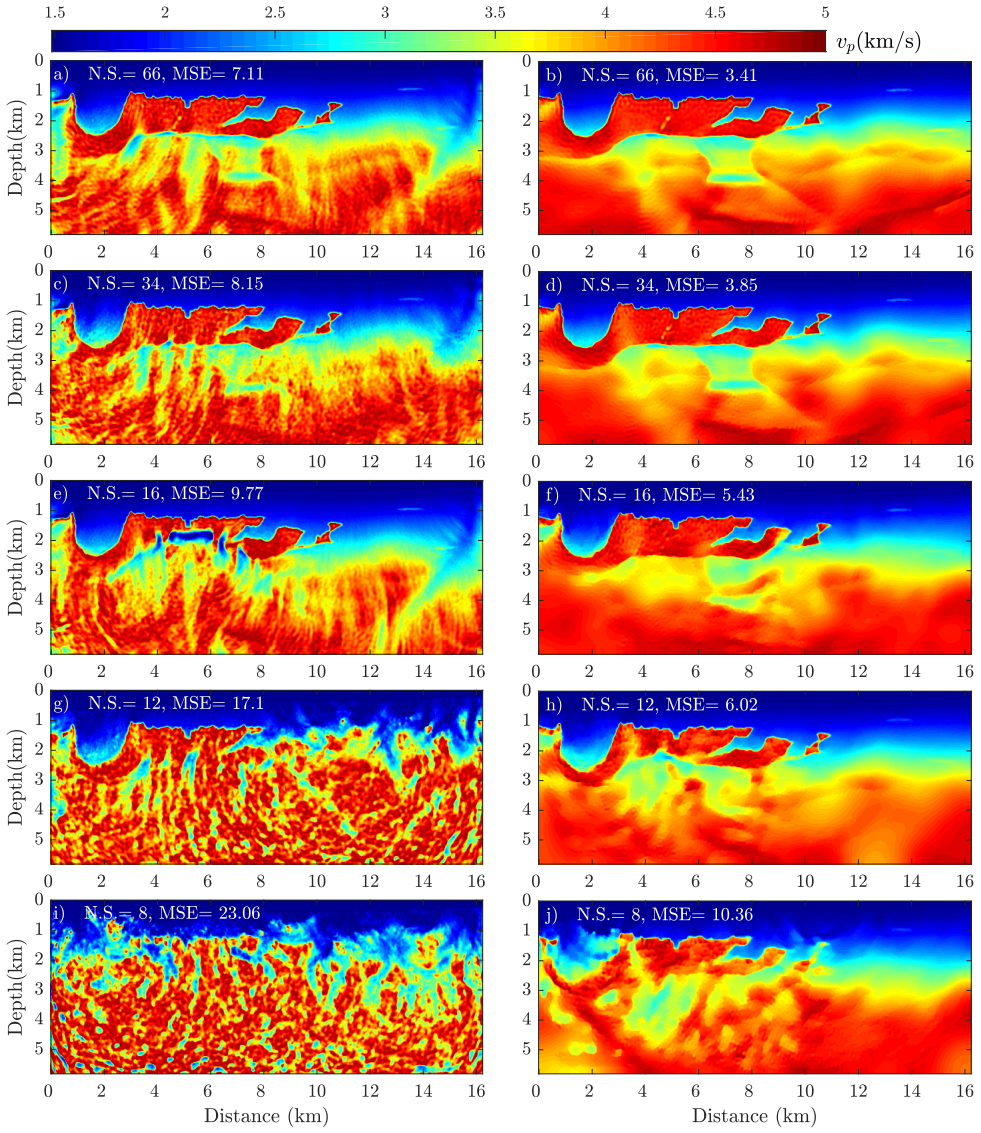

Figure 3.32 – 2004 BP test. Same as Fig. 3.31 for 66 receivers regularly spaced 250 m apart.

the tests shown in Fig. 3.31 and 3.32 for a random distribution of source locations. The source positions are generated such that the source interval ranges between 0.5 and 1.5 of the source interval of the regular acquisition. The results of BTT regularized IR-WRI with 131 and 66 receivers are shown in the left and right columns of Fig. 3.33, respectively.

Finally, we sparsify the frequency interval when 131 receivers and 66 sources are used. We set the frequency interval to 0.5 Hz, 1 Hz and 2 Hz for the first, second and third batch of frequencies, respectively. The IR-WRI results without and with TT regularization are shown in Fig. 3.34.

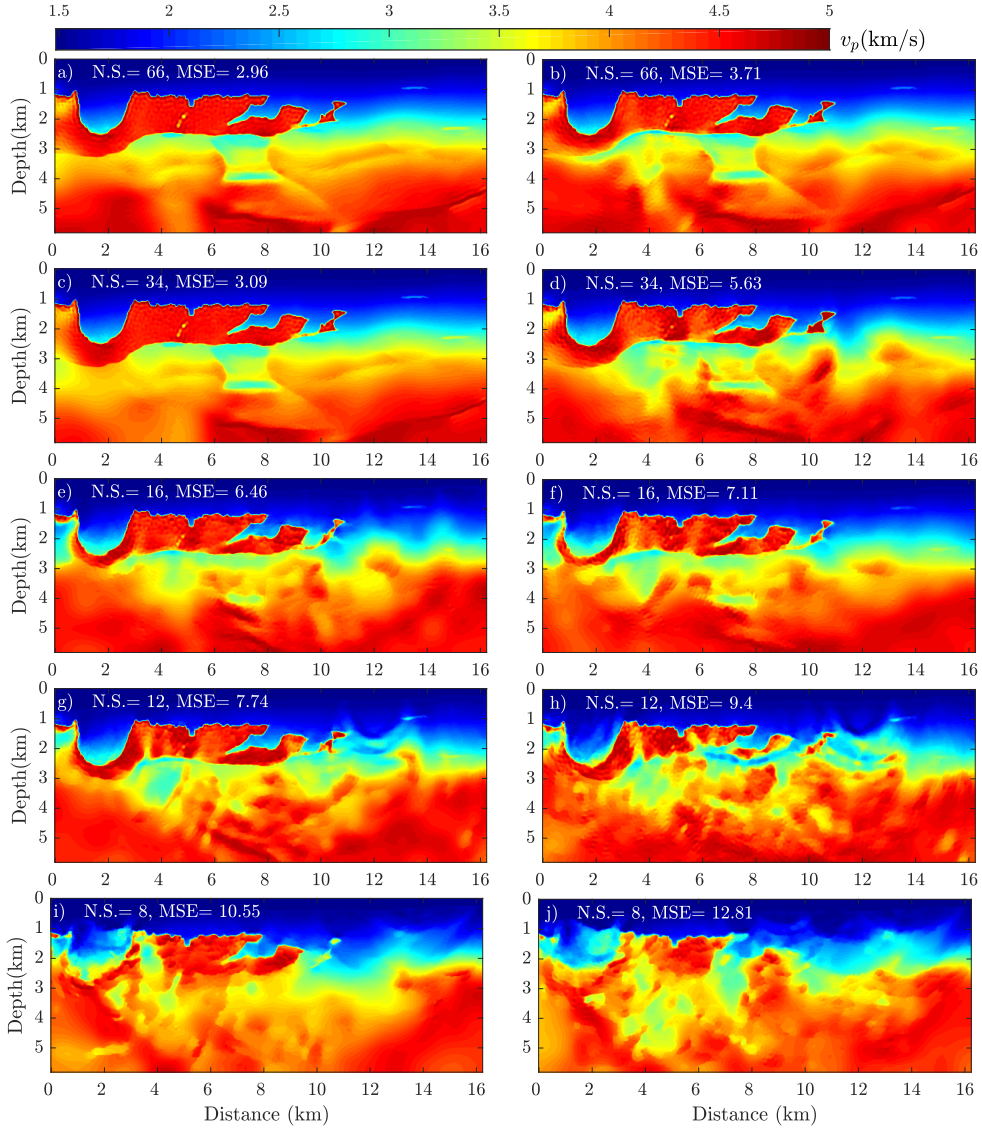

Figure 3.33 – 2004 BP test. BTT regularized IR-WRI with 131 receivers spaced 125 m apart (left column), and 66 receivers regularly spaced 250 m apart (right column). Sources are randomly distributed. The number of sources are (a-b) 66, (c-d) 34, (e-f) 16, (g-h) 12 and (i-j) 8 sources. The number of sources and estimated model error are written in each panel.

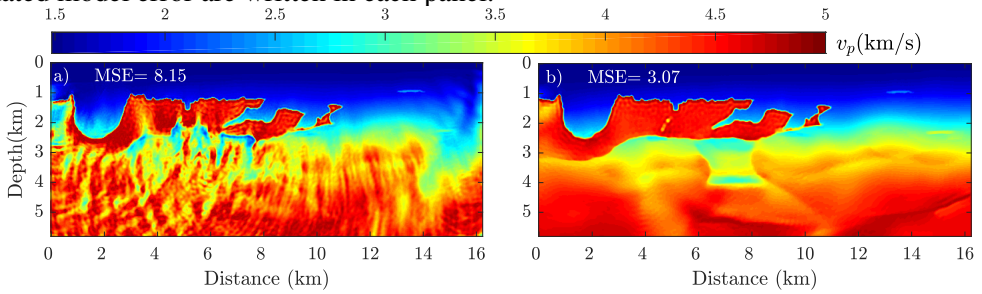

Figure 3.34 – 2004 BP test. IR-WRI results with coarse frequency interval. The acquisition involves 66 sources and 131 receivers regularly spaced 250 m and 125 m apart, respectively. (a) Bound constrained IR-WRI. (b) BTT regularized IR-WRI. The estimated model error is provided in each panel. Reliable reconstruction is shown in (b).

Chapter 4

Extension of ADMM-based WRI to multi-parameter

Chapter overview: The focus of this chapter is to extend ADMM-based FWI or IR-WRI to multi-parameter reconstruction. Since bi-linearity of the wave-equation in model parameters and wavefields has a vital role in IR-WRI, I start this chapter by showing the bi-linearity of the elastodynamic equations for the most general triclinic anisotropic media. I will consider both stiffness and compliance notations for the first and second order elastodynamic equations. By showing this, I proof that IR-WRI can be extended to every kind of physics to perform mono/multi-parameter imaging in visco-acoustic/elastic media.
In the second section of this chapter, I perform a proof of concept study by extending IR-WRI to multiparameter reconstruction for VTI acoustic media. By investigating the structure of the IR-WRI gradient, I find that the radiation patterns of the virtual sources generate similar wavenumber filtering and parameter cross-talks as in classical FWI. I develop VTI IR-WRI more specifically for parameterization involving vertical wavespeed and Thomsen's parameters δ and ϵ, although other parameterization can be used as reviewed in the "Discussion" of this section. I investigate different situations when some of the parameters are used as active parameters during inversion, while others are kept fixed as passive parameters. I equipped VTI IR-WRI with bound constraints and TV regularization to remove undesired effects. Of course, compound regularizations (Aghamiry et al., 2020b) are also a relevant alternative for the regularization of each parameter class reconstruction. This section contains the article which has been published in the *Geophysical journal international* (Aghamiry et al., 2019a).
Finally, for exploiting the full potential of FWI in attenuating media, viscous effects are taken into account by processing the complex-valued wavespeed as two real variables, squared-slowness and attenuation. I update these two classes of parameter in alternating mode which means that IR-WRI is recast as a multi-convex optimization on wavefield, squared-slowness and attenuation. This description of the complex-valued wavespeed as two real variables also allows me to taylor more easily the regularization to the squared-slowness and attenuation. I implement bound constraints and TV regularization on visco-acoustic IR-WRI and show the key role of TV regularization to decrease the ill-posedness of the problem. This section contains the article which has been published in the *Geophysics* journal (Aghamiry et al., 2020d).

4.1 Bi-linearity of the wave-equation in triclinic media

The wave equation is a non-linear function of the wavefield (state variable) and earth parameters (optimization variable). However, when the wavefield is known, the wave equation becomes a linear function of the earth parameters and vice versa (Aghamiry et al., 2019c). This property, which is called bi-linearity, makes FWI be a bi-convex optimization problem which can be solved efficiently with the alternating direction method of multipliers (ADMM) (Boyd et al., 2010). In this section, we show that the bi-linearity of the wave-equation is valid for all types of wave-equation which are used in acoustic and elastic multiparameter FWI. First, we show the bi-linearity of elastodynamic in triclinic media, which is the most general case, with stiffness and compliance notations. Then we move to the acoustic wave equation when vertical transversely isotropy (VTI) anisotropy is considered. Finally, we show the bi-linearity of the acoustic equation for isotropic media. For all the above-mentioned cases, we describe first and second order equations and investigate their bi-linearity.

4.1.1 First-order equations with stiffness notation

The governing equations of FWI in elastic anisotropic media of size $n = n_x \times n_y \times n_z$ are the equation of motion and the constitutive Hooke's law (e.g., Carcione, 2015) as:

$$\begin{aligned} \text{diag}(\rho(\mathbf{x}))\partial_t \mathbf{v}_i(\mathbf{x},t) &= \partial_j \sigma_{ij}(\mathbf{x},t) + \mathbf{f}_i(\mathbf{x},t) && (4.1a)\\ \sigma_{ij}(\mathbf{x},t) &= \text{diag}(\mathbf{c}_{ijkl}(\mathbf{x}))\varepsilon_{kl}(\mathbf{x},t) + \mathbf{s}_i(\mathbf{x},t), && (4.1b) \end{aligned}$$

where the independent variables of each component are shown inside parenthesis, $\mathbf{x} = (x, y, z)$ denote the spatial coordinates in 3D, t denotes the time coordinate and $\text{diag}(\bullet)$ denotes a diagonal matrix. Also, $\mathbf{f}_i \in \mathbb{R}^{n\times 1}$ denotes i'th component of external forces, $\mathbf{s}_i \in \mathbb{R}^{n\times 1}$ denotes i'th component of explosive sources or fracture, $\rho \in \mathbb{R}^{n\times 1}$ density, $\mathbf{v}_i \in \mathbb{R}^{n\times 1}$ particle velocities, $\sigma_{ij} \in \mathbb{R}^{n\times 1}$ the ij'th component of stress tensor, $\mathbf{c}_{ijkl} \in \mathbb{R}^{n\times 1}$ the $ijkl$'th component of stiffness tensor, $\varepsilon_{kl} \in \mathbb{R}^{n\times 1}$ the kl'th strain tensor component with $i, j, k, l \in [x, y, z]$ and $\partial_i \in \mathbb{R}^{n\times n}$ discrete derivative operator in i'th direction. We introduce perfectly-matched layer (PML) (Bérenger, 1994) absorbing conditions in equations 4.1 through a change of coordinates in the complex space such that $\partial_{\tilde{x}} = \frac{1}{\xi_x}\partial_x$, $\partial_{\tilde{y}} = \frac{1}{\xi_y}\partial_y$, and $\partial_{\tilde{z}} = \frac{1}{\xi_z}\partial_z$. We have $\xi_x = 1 + \hat{\imath}\frac{\gamma_x}{\omega}$, $\xi_y = 1 + \hat{\imath}\frac{\gamma_y}{\omega}$ and $\xi_z = 1 + \hat{\imath}\frac{\gamma_z}{\omega}$, where $\hat{\imath} = \sqrt{-1}$, ω is angular frequency and the functions γ_x, γ_y and γ_z control the damping of the wavefield in the PMLs (Operto et al., 2007b). For sake of simplicity, we remove tilde from partial derivative in the rest of this section and the effects of PML boundary condition is considered for all the following partial derivative operators.
It is recalled that strains are related to displacement $\mathbf{u}_i \in \mathbb{R}^{n\times 1}$ by

$$\varepsilon_{ij}(\mathbf{x},t) = \frac{1}{2}[\partial_i \mathbf{u}_j(\mathbf{x},t) + \partial_j \mathbf{u}_i(\mathbf{x},t)], \quad i,j \in \{x,y,z\}. \tag{4.2}$$

Next, according to the Voigt's notation convention (e.g., Helbig, 2013), we change the double-index notation for stresses to be single-indexes as

$$\begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} = \begin{bmatrix} \sigma_1 & \sigma_6 & \sigma_5 \\ \sigma_6 & \sigma_2 & \sigma_4 \\ \sigma_5 & \sigma_4 & \sigma_3 \end{bmatrix}. \tag{4.3}$$

For strain, aiding by the Voigt's index rule we have

$$\begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{bmatrix} = \begin{bmatrix} \varepsilon_1 & \varepsilon_6 & \varepsilon_5 \\ \varepsilon_6 & \varepsilon_2 & \varepsilon_4 \\ \varepsilon_5 & \varepsilon_4 & \varepsilon_3 \end{bmatrix} = \begin{bmatrix} \partial_x \mathbf{u}_x & \frac{1}{2}[\partial_x \mathbf{u}_y + \partial_y \mathbf{u}_x] & \frac{1}{2}[\partial_x \mathbf{u}_z + \partial_z \mathbf{u}_x] \\ \frac{1}{2}[\partial_y \mathbf{u}_x + \partial_x \mathbf{u}_y] & \partial_y \mathbf{u}_y & \frac{1}{2}[\partial_y \mathbf{u}_z + \partial_z \mathbf{u}_y] \\ \frac{1}{2}[\partial_z \mathbf{u}_x + \partial_x \mathbf{u}_z] & \frac{1}{2}[\partial_z \mathbf{u}_y + \partial_y \mathbf{u}_x] & \partial_z \mathbf{u}_z \end{bmatrix}, \tag{4.4}$$

and it can be used to rewrite elastic parameters c_{ijkl} $i, j, k, l \in [x, y, z]$ as c_{IJ} $I, J \in [1, 2, 3, 4, 5, 6]$. We now introduce Auld differential matrix (Auld, 1990) as

$$\nabla^T = \underbrace{\begin{bmatrix} \partial_x & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \partial_y & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \partial_z \\ \mathbf{0} & \partial_z & \partial_y \\ \partial_z & \mathbf{0} & \partial_x \\ \partial_y & \partial_x & \mathbf{0} \end{bmatrix}}_{3n} \Bigg\} 6n. \tag{4.5}$$

The matrix form of strain-displacement relation, equation 4.2, by using ∇^T, equation 4.5, can be written as

$$\varepsilon = \nabla^T \mathbf{u}, \tag{4.6}$$

where

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix} \in \mathbb{R}^{6n \times 1} \quad \text{and} \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_x \\ \mathbf{u}_y \\ \mathbf{u}_z \end{bmatrix} \in \mathbb{R}^{3n \times 1}, \tag{4.7}$$

are vectors that contain components of the strain and wavefield, respectively.
With the Voigt's convention the stress-strain relation, equation 4.1b, in matrix form becomes

$$\sigma = \mathbf{M}_c \varepsilon + \mathbf{s}. \tag{4.8}$$

where

$$\mathbf{M}_c = \underbrace{\left.\begin{bmatrix} \mathrm{diag}(\mathbf{c}_{11}) & \mathrm{diag}(\mathbf{c}_{12}) & \mathrm{diag}(\mathbf{c}_{13}) & \mathrm{diag}(\mathbf{c}_{14}) & \mathrm{diag}(\mathbf{c}_{15}) & \mathrm{diag}(\mathbf{c}_{16}) \\ \mathrm{diag}(\mathbf{c}_{21}) & \mathrm{diag}(\mathbf{c}_{22}) & \mathrm{diag}(\mathbf{c}_{23}) & \mathrm{diag}(\mathbf{c}_{24}) & \mathrm{diag}(\mathbf{c}_{25}) & \mathrm{diag}(\mathbf{c}_{26}) \\ \mathrm{diag}(\mathbf{c}_{31}) & \mathrm{diag}(\mathbf{c}_{32}) & \mathrm{diag}(\mathbf{c}_{33}) & \mathrm{diag}(\mathbf{c}_{34}) & \mathrm{diag}(\mathbf{c}_{35}) & \mathrm{diag}(\mathbf{c}_{36}) \\ \mathrm{diag}(\mathbf{c}_{41}) & \mathrm{diag}(\mathbf{c}_{42}) & \mathrm{diag}(\mathbf{c}_{43}) & \mathrm{diag}(\mathbf{c}_{44}) & \mathrm{diag}(\mathbf{c}_{45}) & \mathrm{diag}(\mathbf{c}_{46}) \\ \mathrm{diag}(\mathbf{c}_{51}) & \mathrm{diag}(\mathbf{c}_{52}) & \mathrm{diag}(\mathbf{c}_{53}) & \mathrm{diag}(\mathbf{c}_{54}) & \mathrm{diag}(\mathbf{c}_{55}) & \mathrm{diag}(\mathbf{c}_{56}) \\ \mathrm{diag}(\mathbf{c}_{61}) & \mathrm{diag}(\mathbf{c}_{62}) & \mathrm{diag}(\mathbf{c}_{63}) & \mathrm{diag}(\mathbf{c}_{64}) & \mathrm{diag}(\mathbf{c}_{65}) & \mathrm{diag}(\mathbf{c}_{66}) \end{bmatrix}\right\} 6n,}_{6n} \tag{4.9}$$

$$\mathbf{s} = \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \mathbf{s}_3 \\ \mathbf{s}_4 \\ \mathbf{s}_5 \\ \mathbf{s}_6 \end{bmatrix} \in \mathbb{R}^{6n\times 1}, \tag{4.10}$$

and

$$\sigma = \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{bmatrix} \in \mathbb{R}^{6n\times 1}. \tag{4.11}$$

Plugging the expression of ε, equation 4.6, in equation 4.8 gives the strain-displacement relation in matrix form as

$$\sigma = \mathbf{M}_c \nabla^T \mathbf{u} + \mathbf{s}. \tag{4.12}$$

In the other hand, the equation of motion,4.1a, can be written in matrix form as

$$\partial_t \mathbf{v} = \mathbf{M}_b \nabla \sigma + \mathbf{f}, \tag{4.13}$$

where

$$\mathbf{M}_b = \underbrace{\left.\begin{bmatrix} \mathrm{diag}(\mathbf{b}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}(\mathbf{b}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathrm{diag}(\mathbf{b}) \end{bmatrix}\right\} 3n,}_{3n} \tag{4.14}$$

$\mathbf{b} = \frac{1}{\rho} \in \mathbb{R}^{n\times 1}$ is the buoyancy vector,

$$\mathbf{v} = \begin{bmatrix} \mathbf{v}_x \\ \mathbf{v}_y \\ \mathbf{v}_z \end{bmatrix} \in \mathbb{R}^{3n\times 1}, \tag{4.15}$$

is the particle velocities vector and

$$\mathbf{f} = \begin{bmatrix} \mathbf{f}_x \\ \mathbf{f}_y \\ \mathbf{f}_z \end{bmatrix} \in \mathbb{R}^{3n\times 1}, \tag{4.16}$$

is the external forces vector.
Because in this Ph.D. thesis, we work on the frequency domain FWI, we continue with frequency domain form of equations. By taking Fourier transform of problem 4.13 with respect to time, the derivative ∂_t changes to $-\hat{i}\omega$ and also the independent variable t changes to ω.
By multiplying equation 4.12 by $-\hat{i}\omega$, we have

$$-\hat{i}\omega\sigma = \mathbf{M}_c\nabla^T\mathbf{v} - \hat{i}\omega\mathbf{s}. \tag{4.17}$$

It is possible to combine equations 4.17 and the frequency form of equation 4.13 which is

$$-\hat{i}\omega\mathbf{v} = \mathbf{M}_b\nabla\sigma + \mathbf{f}, \tag{4.18}$$

to write the velocity-stress elastodynamic equation in triclinic media as

$$\mathbf{MDw} + \hat{i}\omega\mathbf{w} = \mathbf{S}, \tag{4.19}$$

where

$$\mathbf{w} = \begin{bmatrix} \mathbf{v} \\ \sigma \end{bmatrix} \in \mathbb{C}^{9n\times 1} \quad \text{and} \quad \mathbf{S} = \begin{bmatrix} -\mathbf{f} \\ \hat{i}\omega\mathbf{s} \end{bmatrix} \in \mathbb{C}^{9n\times 1}, \tag{4.20}$$

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_c \end{bmatrix} \in \mathbb{C}^{9n\times 9n} \quad \text{and} \quad \mathbf{D} = \begin{bmatrix} \mathbf{0} & \nabla \\ \nabla^T & \mathbf{0} \end{bmatrix} \in \mathbb{C}^{9n\times 9n}. \tag{4.21}$$

Equation 4.19 is a bi-linear function, which is linear in $\mathbf{w}$ with fixed $\mathbf{M}$ and vice versa. The $9n \times 9n$ linear system with fixed subsurface properties ($\mathbf{M}$) on velocity and stress wavefields ($\mathbf{w}$) reads as

$$[\mathbf{MD} + \hat{i}\omega\mathbf{I}]\mathbf{w} = \mathbf{S}, \tag{4.22}$$

where $\mathbf{I}$ is the identity matrix.
The $9n \times 37n$ linear system with fixed $\mathbf{w}$ with respect to subsurface parameters reads

$$\begin{bmatrix} \text{diag}(\partial_x\sigma_1 + \partial_y\sigma_6 + \partial_z\sigma_5) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \text{diag}(\partial_x\sigma_6 + \partial_y\sigma_2 + \partial_z\sigma_4) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \text{diag}(\partial_x\sigma_5 + \partial_y\sigma_4 + \partial_z\sigma_3) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D_v} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D_v} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D_v} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D_v} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D_v} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D_v} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{c}_3 \\ \mathbf{c}_4 \\ \mathbf{c}_5 \\ \mathbf{c}_6 \end{bmatrix} = \mathbf{S} - \hat{i}\omega\mathbf{w} \tag{4.23}$$

where

$$\mathbf{D_v} = \big[\text{diag}(\partial_x \mathbf{v}_x)\ \ \text{diag}(\partial_y \mathbf{v}_y)\ \ \text{diag}(\partial_z \mathbf{v}_z)\ \ \text{diag}(\partial_z \mathbf{v}_y + \partial_y \mathbf{v}_z)\ \ ... \\ \text{diag}(\partial_z \mathbf{v}_x + \partial_x \mathbf{v}_z)\ \ \text{diag}(\partial_y \mathbf{v}_x + \partial_x \mathbf{v}_y)\big] \in \mathbb{C}^{n \times 6n} \tag{4.24}$$

and

$$\mathbf{c}_i = \begin{bmatrix} \mathbf{c}_{i1} \\ \mathbf{c}_{i2} \\ \mathbf{c}_{i3} \\ \mathbf{c}_{i4} \\ \mathbf{c}_{i5} \\ \mathbf{c}_{i6} \end{bmatrix} \in \mathbb{C}^{6n \times 1}. \tag{4.25}$$

4.1.2 Second-order equations with stiffness notation

Alternatively, we can recast the elastodynamic equations as a system of second-order PDEs by elimination of the auxiliary stress wavefields (inject equation 4.12 in equation 4.18) as

$$\left[\omega^2 \mathbf{M}_\rho + \nabla \mathbf{M}_c \nabla^T\right] \mathbf{u} = \underbrace{-\mathbf{M}_\rho \mathbf{f} - \nabla \mathbf{s}}_{\tilde{\mathbf{f}}}. \tag{4.26}$$

where

$$\mathbf{M}_\rho = \underbrace{\begin{bmatrix} \text{diag}(\rho) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \text{diag}(\rho) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \text{diag}(\rho) \end{bmatrix}}_{3n} \Bigg\} 3n, \tag{4.27}$$

$\mathbf{M}_b \mathbf{M}_\rho = \mathbf{I}$ and $\mathbf{v} = -\hat{i}\omega \mathbf{u}$.
Equation 4.26 is linear in $\mathbf{u}$ with fixed $\mathbf{M}_c$ and $\mathbf{M}_\rho$. Also, the $3n \times 37n$ linear system with subsurface properties as unknown and fixed wavefields ($\mathbf{u}$) reads

$$\begin{bmatrix} \omega^2 \text{diag}(\mathbf{u}_x) & \mathbf{D}^x_{\mathbf{u}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}^z_{\mathbf{u}} & \mathbf{D}^y_{\mathbf{u}} \\ \omega^2 \text{diag}(\mathbf{u}_y) & \mathbf{0} & \mathbf{D}^y_{\mathbf{u}} & \mathbf{0} & \mathbf{D}^z_{\mathbf{u}} & \mathbf{0} & \mathbf{D}^x_{\mathbf{u}} \\ \omega^2 \text{diag}(\mathbf{u}_z) & \mathbf{0} & \mathbf{0} & \mathbf{D}^z_{\mathbf{u}} & \mathbf{D}^y_{\mathbf{u}} & \mathbf{D}^x_{\mathbf{u}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \rho \\ \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{c}_3 \\ \mathbf{c}_4 \\ \mathbf{c}_5 \\ \mathbf{c}_6 \end{bmatrix} = \tilde{\mathbf{f}}, \tag{4.28}$$

where

$$\begin{aligned}\mathbf{D}_{\mathbf{u}}^{x} &= \big[\text{diag}(\partial_{xx}\mathbf{u}_x)\ \ \text{diag}(\partial_{xy}\mathbf{u}_y)\ \ \text{diag}(\partial_{xz}\mathbf{u}_z)\ \ \text{diag}(\partial_{xz}\mathbf{u}_y+\partial_{xy}\mathbf{u}_z)\ \ ...\\ &\qquad \text{diag}(\partial_{xz}\mathbf{u}_x+\partial_{xx}\mathbf{u}_z)\ \ \text{diag}(\partial_{xy}\mathbf{u}_x+\partial_{xx}\mathbf{u}_y)\big] \in \mathbb{C}^{n\times 6n},\\ \mathbf{D}_{\mathbf{u}}^{y} &= \big[\text{diag}(\partial_{yx}\mathbf{u}_x)\ \ \text{diag}(\partial_{yy}\mathbf{u}_y)\ \ \text{diag}(\partial_{yz}\mathbf{u}_z)\ \ \text{diag}(\partial_{yz}\mathbf{u}_y+\partial_{yy}\mathbf{u}_z)\ \ ...\\ &\qquad \text{diag}(\partial_{yz}\mathbf{u}_x+\partial_{yx}\mathbf{u}_z)\ \ \text{diag}(\partial_{yy}\mathbf{u}_x+\partial_{yx}\mathbf{u}_y)\big] \in \mathbb{C}^{n\times 6n},\\ \mathbf{D}_{\mathbf{u}}^{z} &= \big[\text{diag}(\partial_{zx}\mathbf{u}_x)\ \ \text{diag}(\partial_{zy}\mathbf{u}_y)\ \ \text{diag}(\partial_{zz}\mathbf{u}_z)\ \ \text{diag}(\partial_{zz}\mathbf{u}_y+\partial_{zy}\mathbf{u}_z)\ \ ...\\ &\qquad \text{diag}(\partial_{zz}\mathbf{u}_x+\partial_{zx}\mathbf{u}_z)\ \ \text{diag}(\partial_{zy}\mathbf{u}_x+\partial_{zx}\mathbf{u}_y)\big] \in \mathbb{C}^{n\times 6n},\end{aligned} \tag{4.29}$$

$\partial_{ij}\ \ i,j \in \{x,y,z\}$ is the second order derivative and its subscript shows the direction and $\mathbf{c}_i$ is defined in equation 4.25.

4.1.3 First-order equations with compliance notation

We can also introduce the compliance matrix $\mathbf{M}_{\int} = \mathbf{M}_{\mathbf{c}}^{-1}$ in order to manipulate symmetric (self-adjoint) operator in the right-hand side of equation 4.19 and to decouple the subsurface properties from the differential operators. By multiplying both sides of equation 4.8 by $\mathbf{M}_{\int}$ we have

$$\mathbf{M}_{\int}\sigma = \varepsilon + \underbrace{\mathbf{M}_{\int}\mathbf{s}}_{\tilde{\mathbf{s}}}, \tag{4.30}$$

where

$$\mathbf{M}_{\int} = \underbrace{\begin{bmatrix} \text{diag}(\int_{11}) & \text{diag}(\int_{12}) & \text{diag}(\int_{13}) & \text{diag}(\int_{14}) & \text{diag}(\int_{15}) & \text{diag}(\int_{16}) \\ \text{diag}(\int_{21}) & \text{diag}(\int_{22}) & \text{diag}(\int_{23}) & \text{diag}(\int_{24}) & \text{diag}(\int_{25}) & \text{diag}(\int_{26}) \\ \text{diag}(\int_{31}) & \text{diag}(\int_{32}) & \text{diag}(\int_{33}) & \text{diag}(\int_{34}) & \text{diag}(\int_{35}) & \text{diag}(\int_{36}) \\ \text{diag}(\int_{41}) & \text{diag}(\int_{42}) & \text{diag}(\int_{43}) & \text{diag}(\int_{44}) & \text{diag}(\int_{45}) & \text{diag}(\int_{46}) \\ \text{diag}(\int_{51}) & \text{diag}(\int_{52}) & \text{diag}(\int_{53}) & \text{diag}(\int_{54}) & \text{diag}(\int_{55}) & \text{diag}(\int_{56}) \\ \text{diag}(\int_{61}) & \text{diag}(\int_{62}) & \text{diag}(\int_{63}) & \text{diag}(\int_{64}) & \text{diag}(\int_{65}) & \text{diag}(\int_{66}) \end{bmatrix}}_{6n} \Bigg\} 6n, \tag{4.31}$$

and $\int_{ij} \in \mathbb{C}^{n\times 1}$.

By putting $\varepsilon = \nabla^T\mathbf{u}$,equation 4.6, into equation 4.30 and multiplying by $-\hat{i}\omega$ we have

$$-\hat{i}\omega\mathbf{M}_{\int}\sigma = \nabla^T\mathbf{v} - \hat{i}\omega\tilde{\mathbf{s}}, \tag{4.32}$$

then by combining equations 4.18, when both sides are multiplied by $\mathbf{M}_{\rho}$, and 4.32 we have the first-order velocity-stress elastodynamic equation in triclinic media as

$$\hat{i}\omega\tilde{\mathbf{M}}\mathbf{w} + \mathbf{D}\mathbf{w} = \mathbf{S}, \tag{4.33}$$

where

$$\tilde{\mathbf{M}} = \begin{bmatrix} \mathbf{M}_{\rho} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{\int} \end{bmatrix} \in \mathbb{C}^{9n\times 9n}, \tag{4.34}$$

$$\mathbf{S} = \begin{bmatrix} -\mathbf{M}_\rho \mathbf{f} \\ \hat{i}\omega\tilde{\mathbf{s}} \end{bmatrix} \in \mathbb{C}^{9n\times 1}, \tag{4.35}$$

and definition of $\mathbf{D}$ and $\mathbf{w}$ are same as before.
Again, the first-order velocity-stress equation, equation 4.33, defines a bilinear function with respect to the subsurface parameters ($\tilde{\mathbf{M}}$) and the wavefield ($\mathbf{w}$). This $9n \times 9n$ linear system with fixed $\tilde{\mathbf{M}}$ and $\mathbf{w}$ as unknown reads

$$[\hat{i}\omega\tilde{\mathbf{M}} + \mathbf{D}]\mathbf{w} = \mathbf{S}, \tag{4.36}$$

and with fixed $\mathbf{w}$, the $9n \times 37n$ linear system on subsurface parameters reads

$$\begin{bmatrix} \hat{i}\omega\text{diag}(\mathbf{v}_x) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hat{i}\omega\text{diag}(\mathbf{v}_y) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hat{i}\omega\text{diag}(\mathbf{v}_z) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_\sigma & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_\sigma & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_\sigma & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_\sigma & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_\sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_\sigma \end{bmatrix} \begin{bmatrix} \rho \\ \int_1 \\ \int_2 \\ \int_3 \\ \int_4 \\ \int_5 \\ \int_6 \end{bmatrix} = \mathbf{S} - \mathbf{D}\mathbf{w}, \tag{4.37}$$

where

$$\mathbf{D}_\sigma = \hat{i}\omega \begin{bmatrix} \text{diag}(\sigma_1) & \text{diag}(\sigma_2) & \text{diag}(\sigma_3) & \text{diag}(\sigma_4) & \text{diag}(\sigma_5) & \text{diag}(\sigma_6) \end{bmatrix} \in \mathbb{C}^{n\times 6n}, \tag{4.38}$$

and

$$\int_i = \begin{bmatrix} \int_{i1} \\ \int_{i2} \\ \int_{i3} \\ \int_{i4} \\ \int_{i5} \\ \int_{i6} \end{bmatrix} \in \mathbb{C}^{6n\times 1}. \tag{4.39}$$

4.1.4 Second-order equations with compliance notation

It is possible to recast the elastodynamic equations with compliance notation as a system of second-order PDEs by elimination of the particle velocity (inject equation 4.18 in equation 4.32) as

$$[\nabla^T \mathbf{M}_b \nabla + \omega^2 \mathbf{M}_f]\sigma = \underbrace{\omega^2\tilde{\mathbf{s}} - \nabla^T\mathbf{f}}_{\tilde{\mathbf{s}}}, \tag{4.40}$$

Equation 4.40 is linear in σ with fixed $\mathbf{M}_f$ and $\mathbf{M}_b$. Also, the $6n \times 37n$ linear system with subsurface properties as unknown and fixed stress wavefields (σ) reads

$$\begin{bmatrix} \partial_x \text{diag}(\partial_x\sigma_1 + \partial_y\sigma_6 + \partial_z\sigma_5) & \mathbf{D}_\sigma & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \partial_y \text{diag}(\partial_x\sigma_6 + \partial_y\sigma_2 + \partial_z\sigma_4) & \mathbf{0} & \mathbf{D}_\sigma & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \partial_z \text{diag}(\partial_x\sigma_5 + \partial_y\sigma_4 + \partial_z\sigma_3) & \mathbf{0} & \mathbf{0} & \mathbf{D}_\sigma & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_\sigma & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_\sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_\sigma \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \int_1 \\ \int_2 \\ \int_3 \\ \int_4 \\ \int_5 \\ \int_6 \end{bmatrix} = \tilde{\mathbf{s}}, \tag{4.41}$$

where

$$\mathbf{D}_\sigma = \omega^2 \begin{bmatrix} \text{diag}(\sigma_1) & \text{diag}(\sigma_2) & \text{diag}(\sigma_3) & \text{diag}(\sigma_4) & \text{diag}(\sigma_5) & \text{diag}(\sigma_6) \end{bmatrix} \in \mathbb{C}^{n \times 6n}. \tag{4.42}$$

4.2 Application of ADMM-based WRI to anisotropy

This section presents a proof of concept study of multi-parameter IR-WRI in acoustic media where transverse isotropy with a vertical symmetry axis (VTI) is implemented. I apply TV regularization and bound constraint in IR-WRI using splitting techniques and ADMM as reviewed in the previous chapters to stabilize the multi-parameter reconstruction. Although VTI acoustic media are non physical (e.g., Grechka et al., 2004), the VTI acoustic approximation has been intensively used by the oil industry for FWI since the proof of concept study of Operto et al. (2009) who have shown its validity in weak anisotropic media from both the kinematic and dynamic viewpoints.

The source is a pressure explosive source which is applied on the normal stress components. In this section, I denote the normal stress components by $\mathbf{u}_x$ and $\mathbf{u}_z$ instead of σ_1 and σ_3 as used in section 1 for sake of notation consistency with previous chapters where wavefields are denoted by $\mathbf{u}$. Also, the notations used in this section introduce explicitly multiple sources with matrix notations in contrast to the vector notations used in previous chapters.

ADMM-based multi-parameter wavefield reconstruction inversion in VTI acoustic media with TV regularization

Hossein S. Aghamiry, Ali Gholami and Stéphane Operto

Geophysical Journal International, 2019, **219**(2), pages 1316-1333
DOI:10.1093/gji/ggz369

4.2.1 Summary

Full waveform inversion (FWI) is a nonlinear waveform matching procedure, which suffers from cycle skipping when the initial model is not kinematically-accurate enough. To mitigate cycle skipping, wavefield reconstruction inversion (WRI) extends the inversion search space by computing wavefields with a relaxation of the wave equation in order to fit the data from the first iteration. Then, the subsurface parameters are updated by minimizing the source residuals the relaxation generated. Capitalizing on the wave-equation bilinearity, performing wavefield reconstruction and parameter estimation in alternating mode decomposes WRI into two linear subproblems, which can solved efficiently with the alternating-direction method of multiplier (ADMM), leading to the so-called iteratively refined wavefield reconstruction inversion (IR-WRI). Moreover, ADMM provides a suitable framework to implement bound constraints and different types of regularizations and their mixture in IR-WRI. Here, IR-WRI is extended to multiparameter reconstruction for VTI acoustic media. To achieve this goal, we first propose different forms of bilinear VTI acoustic wave equation. We develop more specifically IR-WRI for the one that relies on a parametrisation involving vertical wavespeed and Thomsen's parameters δ and ϵ. With a toy numerical example, we first show that the radiation patterns of the virtual sources generate similar wavenumber filtering and parameter cross-talks in classical FWI and IR-WRI. Bound constraints and TV regularization in IR-WRI fully remove these undesired effects for an idealized piecewise constant target. We show with a more realistic long-offset case study representative of the North Sea that anisotropic IR-WRI successfully reconstruct the vertical wavespeed starting from a laterally homogeneous model and update the long-wavelengths of the starting ϵ model, while a smooth δ model is used as a passive background model. VTI acoustic IR-WRI can be alternatively performed with subsurface parametrisations involving stiffness or compliance coefficients or normal moveout velocities and η parameter (or horizontal velocity).

4.2.2 Introduction

Full waveform inversion (FWI) is a waveform matching procedure which provides subsurface model with a wavelength resolution. However, it suffers from cycle skipping when the initial model is not accurate enough according to the lowest available frequency. To mitigate cycle skipping, the search space of frequency-domain FWI can be extended by wavefield reconstruction inversion (WRI) (van Leeuwen and Herrmann, 2013, 2016). In WRI, the search space is extended with a penalty method to relax the wave-equation constraint at the benefit of the observation equation (i.e., the data fit) during wavefield reconstruction. Then, the subsurface

parameters are estimated from the reconstructed wavefields by minimizing the source residuals the relaxation generated. If these two subproblems (wavefield reconstruction and parameter estimation) are solved in alternating mode (van Leeuwen and Herrmann, 2013) rather than by variable projection (van Leeuwen and Herrmann, 2016), WRI can be recast as a sequence of two linear subproblems capitalizing on the bilinearity of the scalar Helmholtz equation with respect to the wavefield and the squared slownesses (the Helmholtz equation is linear with respect to the wavefield for a given squared slowness model and is linear with respect to the squared slownesses for a given wavefield).
Aghamiry et al. (2019c) improved WRI by replacing the penalty method with the augmented Lagrangian method (Nocedal and Wright, 2006) and solve it using the alternating-direction method of multipliers (ADMM) (Boyd et al., 2010), leading to iteratively-refined WRI (IR-WRI). Although ADMM was originally developed for separable convex problems, the bilinearity of the wave equation makes IR-WRI biconvex, which allows for the use of ADMM as is (Boyd et al., 2010, Section 3.1.1). Moreover, a scaled form of ADMM draws clear connection between WRI and IR-WRI in the sense that it shows that IR-WRI reduces to a penalty method in which the right-hand sides (RHSs) in the quadratic objective functions associated with the observation equation and the wave equation are iteratively updated with the running sum of the data and source residuals in iteration (namely, the scaled Lagrange multipliers). This RHS updating, the lacking feature in WRI, efficiently refines the solution of the two linear subproblems when a fixed penalty parameter is used (Aghamiry et al., 2018c, 2019c). Later, Aghamiry et al. (2018b, 2019b) interfaced bound constraints and edge-preserving regularizations with ADMM to manage large-contrast media. Also, to preserve the smooth components of the subsurface when edge preserving regularizations are used, Aghamiry et al. (2018a, 2020b) combine blocky and smooth promoting regularization in the framework of IR-WRI by using infimal convolution of Tikhonov and Total Variation (TV) regularization functions.

IR-WRI was mainly assessed for wavespeed estimation from the scalar Helmholtz equation. The objective of this study is to develop and assess the extension of IR-WRI to multi-parameter reconstruction in VTI acoustic media. To achieve this goal, we first need to review different formulations of the VTI acoustic wave equation and different subsurface parametrisations for which bilinearity of the wave equation is fulfilled, in order to keep the parameter estimation subproblem linear. Since the wavefield reconstruction requires to solve a large and sparse system of linear equations, second-order or fourth-order wave equation for pressure will be favored at the expense of first-order velocity-stress formulations to mitigate the number of unknowns during wavefield reconstruction. However, we stress that different forms of the wave equation can be used to perform wavefield reconstruction and parameter estimation, provided they provide consistent solutions (Gholami et al., 2013b, Their Appendix A and B).
Also, we will favor subsurface parametrisation involving the vertical wavespeed $\mathbf{v}_0$ and the Thomsen's parameter ϵ and δ at the expense of that involving stiffness coefficients according to the trade-off analysis of Gholami et al. (2013b, Their Appendix A and B).

When bilinearity of the wave equation is fulfilled, IR-WRI can be extended to multi-parameter estimation following the procedure promoted by Aghamiry et al. (2019b), where TV regularization and bound constraints are efficiently implemented in the parameter-estimation subproblem using variable splitting schemes (Glowinski et al., 2017). The splitting procedure allows us to break down the non-differentiable TV regularization problem into two easy-to-solve subprob-

lems: a least-squares quadratic subproblem and a proximity subproblem (Goldstein and Osher, 2009).

In this study, we perform a first assessment of multi-parameter IR-WRI for $\mathbf{v}_0$ and ϵ using a toy inclusion example and the more realistic synthetic North Sea case study tackled by Gholami et al. (2013b) and Gholami et al. (2013a). With the toy example, we first show that the radiation patterns and the parameter cross-talks have the same effects as in classical FWI when TV regularization is not applied. Then, we show how TV regularization fully removes bandpass filtering and cross-talk effects generated by these radiation patterns for this idealized piecewise constant target. With the North Sea example, we first show the resilience of IR-WRI against cycle skipping when using a crude initial $\mathbf{v}_0$ model. The reconstruction of the $\mathbf{v}_0$ model is accurate except in the deep smooth part of the subsurface which suffers from a deficit of wide-angle illumination, while the reconstruction of ϵ is more challenging and requires additional regularization to keep its update smooth and close to the starting model. However, we manage to update significantly the long-to-intermediate wavelengths of ϵ by IR-WRI, unlike Gholami et al. (2013a). Also, comparison between mono-parameter IR-WRI for $\mathbf{v}_0$ and multi-parameter IR-WRI for $\mathbf{v}_0$ and ϵ allows us to gain qualitative insights on the sensitivity of the IR-WRI to ϵ in terms of subsurface model quality and data fit.

This paper is organized as follow. We first discuss the bilinearity of the acoustic VTI wave equation as well as its implication on the gradient and the Gauss-Newton Hessian of the parameter-estimation subproblem. From the selected bilinear formulation of the wave equation, we develop bound-constrained TV-regularized IR-WRI for VTI acoustic media parametrized by $1/\mathbf{v}_0^2$, $1+2\epsilon$ and $\sqrt{1+2\delta}$. Third, we assess IR-WRI for VTI acoustic media against the inclusion and North Sea case studies. Finally, we discuss the perspectives of this work.

4.2.3 Theory

In this section, we first show that the VTI acoustic wave equations is bilinear with respect to the wavefield and model parameters. Then, we rely on this bilinearity to formulate multi-parameter IR-WRI in VTI acoustic media with bounding constraints and TV regularization.

Bilinearity of the wave equation: preliminaries

If we write the wave equation in a generic matrix form as

$$\mathbf{A}(\mathbf{m})\mathbf{u} = \mathbf{s}, \tag{4.43}$$

where $\mathbf{A} \in \mathbb{C}^{(n\times n_c)\times(n\times n_c)}$ is the wave-equation operator, $\mathbf{u} \in \mathbb{C}^{(n\times n_c)\times 1}$ is the wavefield vector, $\mathbf{s} \in \mathbb{C}^{(n\times n_c)\times 1}$ is the source vector, $\mathbf{m} \in \mathbb{R}^{(n\times n_m)\times 1}$ is the subsurface parameter vector, n is the number of degrees of freedom in the spatial computational mesh, n_c is the number of wavefield components, n_m is the number of parameter classes, then the wave equation is bilinear if there exists a linear operator $\mathbf{L}(\mathbf{u}) \in \mathbb{C}^{(n\times n_m)\times(n\times n_m)}$ such that

$$\mathbf{L}(\mathbf{u})\mathbf{m} = \mathbf{y}(\mathbf{u}), \tag{4.44}$$

where $\mathbf{y}(\mathbf{u}) \in \mathbb{C}^{(n\times n_m)\times 1}$. Bilinearity is verified when the left-hand side of the wave equation can be decomposed as:

$$\mathbf{A}(\mathbf{m})\mathbf{u} = \mathbf{B}\,\mathbf{M}(\mathbf{m})\,\mathbf{C}\mathbf{u} + \mathbf{D}\mathbf{u}, \tag{4.45}$$

where $\mathbf{M}(\mathbf{m})$ is a block matrix, whose blocks of dimension $n \times n$ are either 0 or of the form $\text{diag}(\mathbf{m}_i)$, and matrices $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ don't depend on $\mathbf{m}$. The operator $\text{diag}(\bullet)$ denotes a diagonal matrix of coefficients $\bullet$ and $\mathbf{m}_i$ is the subsurface parameter vector of class i. By noting that $\text{diag}(\mathbf{x})\mathbf{z} = \text{diag}(\mathbf{z})\mathbf{x}$, the block diagonal structure of $\mathbf{M}$ allows one to rewrite the term $\mathbf{B}\,\mathbf{M}(\mathbf{m})\,\mathbf{C}\mathbf{u}$ as $\mathbf{B}\,\mathbf{C}'(\mathbf{u})\mathbf{m}$, where $\mathbf{C}'(\mathbf{u})$ is a block matrix, whose blocks of dimension $n \times n$ are either 0 or diagonal with coefficients depending on $\mathbf{u}$.

It follows from equation 4.45 and the above permutation between $\mathbf{u}$ and $\mathbf{m}$ that the wave equation can be re-written as

$$\mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{s} = \mathbf{B}\,\mathbf{C}'(\mathbf{u})\mathbf{m} + \mathbf{D}\mathbf{u} - \mathbf{s} = \mathbf{L}(\mathbf{u})\mathbf{m} - \mathbf{y}(\mathbf{u}). \tag{4.46}$$

Moreover, in the framework of multi-parameter analysis, it is worth noting that

$$\frac{\partial \mathbf{A}(\mathbf{m})}{\partial m_k}\mathbf{u} = \mathbf{L}(\mathbf{u})\mathbf{e}_k, \tag{4.47}$$

where the left-hand side is the so-called virtual source associated with m_k (Pratt et al., 1998) and $\mathbf{e}_k \in \mathbb{C}^{(n\times n_m)\times 1}$ denotes a column vector whose kth component is one while all the others are zeros.

Accordingly, the normal operator $\mathbf{L}^T\mathbf{L}$, i.e., the Gauss-Newton Hessian of the parameter estimation subproblem in IR-WRI, is formed by the zero-lag correlation of the virtual sources and, hence is extremely sparse. We also point that $\mathbf{L}$ and its associated normal operator $\mathbf{L}^T\mathbf{L}$ are block diagonals if $\mathbf{B}$ is diagonal. This means that, if the model parameters are first sorted according to their position in the mesh (fast index) and second according to the parameter class they belong to (slow index), then the diagonal coefficients of the off-diagonal blocks describe the inter-parameter coupling. In the following section, we show the bilinearity of the acoustic VTI wave equation based upon the above matrix manipulations.

Bilinearity of the acoustic VTI wave equation

First-order velocity-stress wave equation

We first consider the frequency-domain first-order velocity-stress wave equation in 2D VTI acoustic media (Duveneck et al., 2008; Duveneck and Bakker, 2011; Operto et al., 2014)

$$\begin{aligned}
-\hat{i}\omega\mathbf{v}_{x,l} &= \text{diag}(\mathbf{b})\nabla_x\mathbf{u}_{x,l},\\
-\hat{i}\omega\mathbf{v}_{z,l} &= \text{diag}(\mathbf{b})\nabla_z\mathbf{u}_{z,l},\\
-\hat{i}\omega\mathbf{u}_{x,l} &= \text{diag}(\mathbf{c}_{11})\nabla_x\mathbf{v}_{x,l} + \text{diag}(\mathbf{c}_{13})\nabla_z\mathbf{v}_{z,l} - \hat{i}\omega\mathbf{s}_l,\\
-\hat{i}\omega\mathbf{u}_{z,l} &= \text{diag}(\mathbf{c}_{13})\nabla_x\mathbf{v}_{x,l} + \text{diag}(\mathbf{c}_{33})\nabla_z\mathbf{v}_{z,l} - \hat{i}\omega\mathbf{s}_l,
\end{aligned} \tag{4.48}$$

where $\hat{i} = \sqrt{-1}$, ω is the angular frequency, $\mathbf{v}_{x,l} \in \mathbb{C}^{n\times 1}$ and $\mathbf{v}_{z,l} \in \mathbb{C}^{n\times 1}$ are the horizontal and vertical particle velocity wavefields, $\mathbf{s}_l \in \mathbb{C}^{n\times 1}$ denote the pressure sources, and $\mathbf{u}_{x,l} \in \mathbb{C}^{n\times 1}$ and $\mathbf{u}_{z,l} \in \mathbb{C}^{n\times 1}$ are the so-called horizontal and vertical pressure wavefields (Plessix and Cao, 2011). The subscript $l \in \{1, 2, ..., n_s\}$ is the source index, where n_s denotes the number of sources. The subsurface properties are parametrized by the buoyancy $\mathbf{b} \in \mathbb{R}^{n\times 1}$ (inverse of density) and the stiffness coefficients $\mathbf{c}_{ij} \in \mathbb{R}^{n\times 1}$. Operators ∇_x and ∇_z are finite difference approximation of first order derivative operators with absorbing perfectly matched layer (PML) coefficients (Bérenger, 1994). Gathering equation 4.48 for all sources leads to the following matrix equation:

$$\left(\begin{bmatrix}\mathbf{M}_b & \mathbf{0}\\ \mathbf{0} & \mathbf{M}_c\end{bmatrix}\begin{bmatrix}\mathbf{0} & \nabla\\ \nabla & \mathbf{0}\end{bmatrix} + \hat{i}\omega\mathbf{I}\right)\begin{bmatrix}\mathbf{V}\\ \mathbf{U}\end{bmatrix} = \hat{i}\omega\begin{bmatrix}\mathbf{0}\\ \mathbf{S}\end{bmatrix}, \tag{4.49}$$

where $\mathbf{I}$ is the identity matrix,

$$\nabla = \begin{bmatrix}\nabla_x & \mathbf{0}\\ \mathbf{0} & \nabla_z\end{bmatrix},$$

$$\mathbf{U} = \begin{bmatrix}\mathbf{U}_h\\ \mathbf{U}_v\end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix}\mathbf{V}_x\\ \mathbf{V}_z\end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix}\mathbf{S}_x\\ \mathbf{S}_z\end{bmatrix}, \tag{4.50}$$

with

$$\mathbf{V}_x = \begin{bmatrix}\mathbf{v}_{x,1} & \mathbf{v}_{x,2} & ... & \mathbf{v}_{x,n_s}\end{bmatrix} \in \mathbb{C}^{n\times n_s},$$

and analogously for $\mathbf{V}_z$, $\mathbf{U}_h$, $\mathbf{U}_v$, and

$$\mathbf{S}_x = \mathbf{S}_z = \begin{bmatrix}\mathbf{s}_1 & \mathbf{s}_2 & ... & \mathbf{s}_{n_s}\end{bmatrix} \in \mathbb{C}^{n\times n_s}.$$

Furthermore,

$$\mathbf{M}_b = \begin{bmatrix}\text{diag}(\mathbf{b}) & \mathbf{0}\\ \mathbf{0} & \text{diag}(\mathbf{b})\end{bmatrix}, \mathbf{M}_c = \begin{bmatrix}\text{diag}(\mathbf{c}_{11}) & \text{diag}(\mathbf{c}_{13})\\ \text{diag}(\mathbf{c}_{13}) & \text{diag}(\mathbf{c}_{33})\end{bmatrix}.$$

Note that, according to the decomposition introduced in the previous section, $\mathbf{M} = \begin{bmatrix}\mathbf{M}_b & \mathbf{0}\\ \mathbf{0} & \mathbf{M}_c\end{bmatrix}$, $\mathbf{B} = \mathbf{I}$, $\mathbf{C} = \nabla$ and $\mathbf{D} = \hat{i}\omega\mathbf{I}$. Equation 4.49 is linear in $\mathbf{U}$ and $\mathbf{V}$ when the model parameters embedded in $\mathbf{M}_b$ and $\mathbf{M}_c$ are known. When $\mathbf{U}$ and $\mathbf{V}$ are known, this system can be also recast as a new linear system in which the unknowns are the model parameters, thus highlighting the bilinearity of the wave equation. For the lth source, the new equations become

$$\begin{bmatrix}\mathbf{L}_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{L}_{21} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{L}_{32} & \mathbf{L}_{33} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{L}_{43} & \mathbf{L}_{44}\end{bmatrix}\begin{bmatrix}\mathbf{b}\\ \mathbf{c}_{11}\\ \mathbf{c}_{13}\\ \mathbf{c}_{33}\end{bmatrix} = \hat{i}\omega\begin{bmatrix}\mathbf{0} - \mathbf{v}_{x,l}\\ \mathbf{0} - \mathbf{v}_{z,l}\\ \mathbf{s}_l - \mathbf{u}_{x,l}\\ \mathbf{s}_l - \mathbf{u}_{z,l}\end{bmatrix}, \tag{4.51}$$

where

$$\begin{cases}\mathbf{L}_{11} = \text{diag}(\nabla_x\mathbf{u}_{x,l}),\\ \mathbf{L}_{21} = \text{diag}(\nabla_z\mathbf{u}_{z,l}),\\ \mathbf{L}_{32} = \mathbf{L}_{43} = \text{diag}(\nabla_x\mathbf{v}_{x,l}),\\ \mathbf{L}_{33} = \mathbf{L}_{44} = \text{diag}(\nabla_z\mathbf{v}_{z,l}).\end{cases}$$

Equations 4.49 and 4.51 are equivalent forms of the original equation 4.48. The former expresses the discretized wavefields as the unknowns of a linear system, whose coefficients depend on the known subsurface parameters, while the latter expresses the model parameters as the unknowns of an another linear system, whose coefficients depend on the known wavefields. In the framework of WRI, this bilinearity allows one to recast the waveform inversion problem as two linear subproblems for wavefield reconstruction and parameter estimation, which can be solved efficiently in alternating mode with ADMM (Aghamiry et al., 2019c,b). In the next section, we show the bilinearity of the second-order frequency-domain wave equation, which may be more convenient to solve with linear algebra methods than the first-order counterpart, since it involves fewer unknowns for a computational domain of given size.

Second-order wave equation

Following a parsimonious approach (e.g., Operto et al., 2007a), we eliminate $\mathbf{v}_{x,l}$ and $\mathbf{v}_{z,l}$ from equation 4.48 to derive a system of two second-order partial differential equations as

$$\begin{aligned} -\omega^2\mathbf{u}_{x,l} &= \text{diag}(\mathbf{c}_{11})\nabla_x\text{diag}(\mathbf{b})\nabla_x\mathbf{u}_{x,l} \\ &\quad + \text{diag}(\mathbf{c}_{13})\nabla_z\text{diag}(\mathbf{b})\nabla_z\mathbf{u}_{z,l} - \omega^2\mathbf{s}_l, \\ -\omega^2\mathbf{u}_{z,l} &= \text{diag}(\mathbf{c}_{13})\nabla_x\text{diag}(\mathbf{b})\nabla_x\mathbf{u}_{x,l} \\ &\quad + \text{diag}(\mathbf{c}_{33})\nabla_z\text{diag}(\mathbf{b})\nabla_z\mathbf{u}_{z,l} - \omega^2\mathbf{s}_l. \end{aligned} \tag{4.52}$$

Equation 4.52 defines a tri-linear equation with respect to buoyancy, stiffness parameters and pressure wavefields. A tri-linear function is a function of three variables which is linear in one variable when the other two variables are fixed. In this study, we will assume that density is constant and equal to 1 to focus on the estimation of the anisotropic parameters. If heterogeneous density needs to be considered, the second-order wave equation can be recast as a bilinear system if the first-order wave equation is parametrized with compliance coefficients instead of stiffness coefficients (see Appendix A, section 4.2.7, and Vigh et al., 2014; Yang et al., 2016b). Alternatively, the parameter estimation can be performed with the bilinear first-order wave equation, equation 4.51, while wavefield reconstruction is performed with the second-order wave equation, equation 4.52, or the fourth-order wave equation reviewed in Appendix C (section 4.2.9) for sake of computational efficiency.

We continue by assuming that the density is constant and equal to 1 and parametrize the VTI equation in terms of vertical wavespeed $\mathbf{v}_0$ and Thomsen's parameters ϵ and δ (Thomsen, 1986). Accordingly, we rewrite equation 4.52 as

$$\begin{aligned} -\omega^2\text{diag}\left(1/\mathbf{v}_0^2\right)\mathbf{u}_{x,l} &= \text{diag}(1+2\epsilon)\nabla_{xx}\mathbf{u}_{x,l} \\ &\quad + \text{diag}(\sqrt{1+2\delta})\nabla_{zz}\mathbf{u}_{z,l} - \text{diag}(1/\mathbf{v}_0^2)\mathbf{s}_l, \\ -\omega^2\text{diag}\left(1/\mathbf{v}_0^2\right)\mathbf{u}_{z,l} &= \text{diag}(\sqrt{1+2\delta})\nabla_{xx}\mathbf{u}_{x,l} \\ &\quad + \nabla_{zz}\mathbf{u}_{z,l} - \text{diag}\left(1/\mathbf{v}_0^2\right)\mathbf{s}_l, \end{aligned} \tag{4.53}$$

where $\nabla_{xx} = \nabla_x\nabla_x$ and $\nabla_{zz} = \nabla_z\nabla_z$. We write this $2n \times 2n$ linear system in a more compact

form as

$$\mathbf{A}(\mathbf{m})\mathbf{U} = \mathbf{S}, \tag{4.54}$$

where $\mathbf{S}^T = \omega^2 \begin{bmatrix}\mathbf{S}_x^T & \mathbf{S}_z^T\end{bmatrix}$, $\mathbf{U}$ is defined as in equation 4.50, and the matrix $\mathbf{A}$ is given by

$$\begin{bmatrix}\omega^2\text{diag}(\mathbf{m}_{v_0}) + \text{diag}(\mathbf{m}_\epsilon)\nabla_{xx} & \text{diag}(\mathbf{m}_\delta)\nabla_{zz} \\ \text{diag}(\mathbf{m}_\delta)\nabla_{xx} & \omega^2\text{diag}(\mathbf{m}_{v_0}) + \nabla_{zz}\end{bmatrix},$$

where the model parameters are

$$\mathbf{m} = \begin{bmatrix}\mathbf{m}_{v_0} \\ \mathbf{m}_\epsilon \\ \mathbf{m}_\delta\end{bmatrix} = \begin{bmatrix}1/\mathbf{v}_0^2 \\ 1+2\epsilon \\ \sqrt{1+2\delta}\end{bmatrix}.$$

Equation 4.54 is linear in wavefields $\mathbf{U}$ when the model parameters $\mathbf{m}$ are known. When $\mathbf{U}$ is known, this system can be recast as a linear system in which the unknowns are the model parameters.

$$\begin{bmatrix}\mathbf{L}_1 \\ \vdots \\ \mathbf{L}_l \\ \vdots \\ \mathbf{L}_{n_s}\end{bmatrix} \mathbf{m} = \begin{bmatrix}\mathbf{y}_1 \\ \vdots \\ \mathbf{y}_l \\ \vdots \\ \mathbf{y}_{n_s}\end{bmatrix}, \tag{4.55}$$

where $\mathbf{L}_l$ is given by

$$\begin{bmatrix}\omega^2\text{diag}(\mathbf{u}_{x,l}) & \text{diag}(\nabla_{xx}\mathbf{u}_{x,l}) & \text{diag}(\nabla_{zz}\mathbf{u}_{z,l}) \\ \omega^2\text{diag}(\mathbf{u}_{z,l}) & \mathbf{0} & \text{diag}(\nabla_{xx}\mathbf{u}_{x,l})\end{bmatrix}$$

and

$$\mathbf{y}_l = \begin{bmatrix}\omega^2\mathbf{s}_l \\ \omega^2\mathbf{s}_l - \nabla_{zz}\mathbf{u}_{z,l}\end{bmatrix}.$$

Note that each block of $\mathbf{L}_l$ is diagonal. In the next section, we develop multi-parameter acoustic VTI IR-WRI with bound constraints and TV regularization. We give the most general formulation in which all the tree parameter classes $\mathbf{m}_{v_0}$, $\mathbf{m}_\epsilon$, $\mathbf{m}_\delta$ are optimization parameters (updated by the inversion). However, one may process some of them as passive parameters or update the parameter classes in sequence rather than jointly. In this case, the linear system associated with the parameter estimation subproblem, equation 4.55, will change. Table 4.1 presents this system for the different possible configurations.

Table 4.1 – The linear systems corresponding to the update of $\mathbf{m}_{v_0}$, $\mathbf{m}_\epsilon$ and $\mathbf{m}_\delta$ when they are active or passive during the inversion. In the first three columns, « a. » denotes an active parameter and « p. » denotes a passive parameter.

$\mathbf{m}_{v_0}$	$\mathbf{m}_\epsilon$	$\mathbf{m}_\delta$	The corresponding linear system for updating the model
a.	p.	p.	$\underbrace{\begin{bmatrix}\omega^2\text{diag}(\mathbf{u}_{x,l})\\ \omega^2\text{diag}(\mathbf{u}_{z,l})\end{bmatrix}}_{\mathbf{L}_l\in\mathbb{C}^{2n\times n}}\underbrace{\begin{bmatrix}\mathbf{m}_{v_0}\end{bmatrix}}_{\mathbf{m}\in\mathbb{R}^{n\times 1}}=\underbrace{\begin{bmatrix}\omega^2\mathbf{s}_l-\text{diag}(\mathbf{m}_\epsilon)\nabla_{xx}\mathbf{u}_{x,l}-\text{diag}(\mathbf{m}_\delta)\nabla_{zz}\mathbf{u}_{z,l}\\ \omega^2\mathbf{s}_l-\nabla_{zz}\mathbf{u}_{z,l}-\text{diag}(\mathbf{m}_\delta)\nabla_{xx}\mathbf{u}_{x,l}\end{bmatrix}}_{\mathbf{y}_l\in\mathbb{C}^{2n\times 1}}$
p.	a.	p.	$\underbrace{\begin{bmatrix}\text{diag}(\nabla_{xx}\mathbf{u}_{x,l})\end{bmatrix}}_{\mathbf{L}_l\in\mathbb{C}^{n\times n}}\underbrace{\begin{bmatrix}\mathbf{m}_\epsilon\end{bmatrix}}_{\mathbf{m}\in\mathbb{R}^{n\times 1}}=\underbrace{\begin{bmatrix}\omega^2\mathbf{s}_l-\omega^2\text{diag}(\mathbf{m}_{v_0})\mathbf{u}_{x,l}-\text{diag}(\mathbf{m}_\delta)\nabla_{zz}\mathbf{u}_{z,l}\end{bmatrix}}_{\mathbf{y}_l\in\mathbb{C}^{n\times 1}}$
p.	p.	a.	$\underbrace{\begin{bmatrix}\text{diag}(\nabla_{zz}\mathbf{u}_{z,l})\\ \text{diag}(\nabla_{xx}\mathbf{u}_{x,l})\end{bmatrix}}_{\mathbf{L}_l\in\mathbb{C}^{2n\times n}}\underbrace{\begin{bmatrix}\mathbf{m}_\delta\end{bmatrix}}_{\mathbf{m}\in\mathbb{R}^{n\times 1}}=\underbrace{\begin{bmatrix}\omega^2\mathbf{s}_l-\omega^2\text{diag}(\mathbf{m}_{v_0})\mathbf{u}_{x,l}-\text{diag}(\mathbf{m}_\epsilon)\nabla_{xx}\mathbf{u}_{x,l}\\ \omega^2\mathbf{s}_l-\nabla_{zz}\mathbf{u}_{z,l}-\omega^2\text{diag}(\mathbf{m}_{v_0})\mathbf{u}_{z,l}\end{bmatrix}}_{\mathbf{y}_l\in\mathbb{C}^{2n\times 1}}$
p.	a.	a.	$\underbrace{\begin{bmatrix}\text{diag}(\nabla_{xx}\mathbf{u}_{x,l}) & \text{diag}(\nabla_{zz}\mathbf{u}_{z,l})\\ \mathbf{0} & \text{diag}(\nabla_{xx}\mathbf{u}_{x,l})\end{bmatrix}}_{\mathbf{L}_l\in\mathbb{C}^{2n\times 2n}}\underbrace{\begin{bmatrix}\mathbf{m}_\epsilon\\ \mathbf{m}_\delta\end{bmatrix}}_{\mathbf{m}\in\mathbb{R}^{2n\times 1}}=\underbrace{\begin{bmatrix}\omega^2\mathbf{s}_l-\omega^2\text{diag}(\mathbf{m}_{v_0})\mathbf{u}_{x,l}\\ \omega^2\mathbf{s}_l-\nabla_{zz}\mathbf{u}_{z,l}-\omega^2\text{diag}(\mathbf{m}_{v_0})\mathbf{u}_{z,l}\end{bmatrix}}_{\mathbf{y}_l\in\mathbb{C}^{2n\times 1}}$
a.	a.	p.	$\underbrace{\begin{bmatrix}\omega^2\text{diag}(\mathbf{u}_{x,l}) & \text{diag}(\nabla_{xx}\mathbf{u}_{x,l})\\ \omega^2\text{diag}(\mathbf{u}_{z,l}) & \mathbf{0}\end{bmatrix}}_{\mathbf{L}_l\in\mathbb{C}^{2n\times 2n}}\underbrace{\begin{bmatrix}\mathbf{m}_{v_0}\\ \mathbf{m}_\epsilon\end{bmatrix}}_{\mathbf{m}\in\mathbb{R}^{2n\times 1}}=\underbrace{\begin{bmatrix}\omega^2\mathbf{s}_l-\text{diag}(\mathbf{m}_\delta)\nabla_{zz}\mathbf{u}_{z,l}\\ \omega^2\mathbf{s}_l-\nabla_{zz}\mathbf{u}_{z,l}-\text{diag}(\mathbf{m}_\delta)\nabla_{xx}\mathbf{u}_{x,l}\end{bmatrix}}_{\mathbf{y}_l\in\mathbb{C}^{2n\times 1}}$
a.	p.	a.	$\underbrace{\begin{bmatrix}\omega^2\text{diag}(\mathbf{u}_{x,l}) & \text{diag}(\nabla_{zz}\mathbf{u}_{z,l})\\ \omega^2\text{diag}(\mathbf{u}_{z,l}) & \text{diag}(\nabla_{xx}\mathbf{u}_{x,l})\end{bmatrix}}_{\mathbf{L}_l\in\mathbb{C}^{2n\times 2n}}\underbrace{\begin{bmatrix}\mathbf{m}_{v_0}\\ \mathbf{m}_\delta\end{bmatrix}}_{\mathbf{m}\in\mathbb{R}^{2n\times 1}}=\underbrace{\begin{bmatrix}\omega^2\mathbf{s}_l-\text{diag}(\mathbf{m}_\epsilon)\nabla_{xx}\mathbf{u}_{x,l}\\ \omega^2\mathbf{s}_l-\nabla_{zz}\mathbf{u}_{z,l}\end{bmatrix}}_{\mathbf{y}_l\in\mathbb{C}^{2n\times 1}}$

ADMM-based acoustic VTI wavefield reconstruction inversion

We consider the following bound-constrained TV-regularized multivariate optimization problem associated with the wave equation described by equation 4.54:

$$\begin{aligned}&\min_{\mathbf{U},\mathbf{m}\in\mathcal{C}}\quad\sum\sqrt{|\partial_x\mathbf{m}|^2+|\partial_z\mathbf{m}|^2},\\ &\text{subject to}\quad\begin{cases}\mathbf{PU}=\mathbf{D},\\ \mathbf{A}(\mathbf{m})\mathbf{U}=\mathbf{S},\end{cases}\end{aligned}\tag{4.56}$$

where ∂_x and ∂_z are, respectively, first-order finite-difference operators in the horizontal and vertical directions with appropriate boundary conditions, $\mathcal{C}=\{\mathbf{x}\in\mathbb{R}^{3n\times 1}\mid\mathbf{m}_l\leq\mathbf{x}\leq\mathbf{m}_u\}$ is the set of all feasible models bounded by a priori lower bound $\mathbf{m}_l$ and upper bound $\mathbf{m}_u$,

$$\mathbf{D}=\begin{bmatrix}\mathbf{d}_1 & \mathbf{d}_2 & \dots & \mathbf{d}_{n_s}\end{bmatrix}\in\mathbb{C}^{n_r\times n_s}$$

with $\mathbf{d}_l$ denoting the recorded data for the lth source, each including n_r samples (the number of receivers), $\mathbf{P} \in \mathbb{R}^{n_r \times 2n}$ is a linear observation operator which samples the wavefields at the receiver positions. Here, we assume that the sampling operator is identical across all sources (stationary-recording acquisitions). However, one may used a specific operator for each source.

We solve this constrained optimization problem with ADMM (Boyd et al., 2010; Aghamiry et al., 2019b), an augmented Lagrangian method with operator splitting, leading to the following saddle point problem

$$\begin{aligned} &\min_{\mathbf{U},\mathbf{m}\in\mathcal{C}} \max_{\bar{\mathbf{D}},\bar{\mathbf{S}}} \quad \sum \sqrt{|\partial_x \mathbf{m}|^2 + |\partial_z \mathbf{m}|^2} \\ &+ \left\langle \bar{\mathbf{D}}, \mathbf{P}\mathbf{U} - \mathbf{D} \right\rangle + \lambda_0 \|\mathbf{P}\mathbf{U} - \mathbf{D}\|_F^2 \\ &+ \left\langle \bar{\mathbf{S}}, \mathbf{A}(\mathbf{m})\mathbf{U} - \mathbf{S} \right\rangle + \lambda_1 \|\mathbf{A}(\mathbf{m})\mathbf{U} - \mathbf{S}\|_F^2, \end{aligned} \tag{4.57}$$

where $\| \bullet \|_F^2$ denotes the Frobenius norm of $\bullet$, $\lambda_0, \lambda_1 > 0$ are penalty parameters and $\bar{\mathbf{D}} \in \mathbb{C}^{n_r \times n_s}$ and $\bar{\mathbf{S}} \in \mathbb{C}^{n \times n_s}$ are the Lagrange multipliers (dual variables).

The last two lines of Equation 4.57 shows that the augmented Lagrangian function combines a Lagrangian function (left terms) and a penalty function (right terms). Also, scaling the Lagrange multipliers by the penalty parameters, $\tilde{\mathbf{D}} = -\bar{\mathbf{D}}/\lambda_0$ and $\tilde{\mathbf{S}} = -\bar{\mathbf{S}}/\lambda_1$, allows us to recast the augmented Lagrangian function in a more convenient form (Boyd et al., 2010, Section 3.1.1)

$$\begin{aligned} &\min_{\mathbf{U},\mathbf{m}\in\mathcal{C}} \max_{\tilde{\mathbf{D}},\tilde{\mathbf{S}}} \quad \sum \sqrt{|\partial_x \mathbf{m}|^2 + |\partial_z \mathbf{m}|^2} \\ &+ \lambda_0 \|\mathbf{P}\mathbf{U} - \mathbf{D} - \tilde{\mathbf{D}}\|_F^2 - \|\tilde{\mathbf{D}}\|_F^2 \\ &+ \lambda_1 \|\mathbf{A}(\mathbf{m})\mathbf{U} - \mathbf{S} - \tilde{\mathbf{S}}\|_F^2 - \|\tilde{\mathbf{S}}\|_F^2, \end{aligned} \tag{4.58}$$

where the scaled dual variables have been injected in the penalty functions.

In the WRI framework, the augmented Lagrangian method provides an efficient and easy-to-tune optimization scheme that extends the parameter search space by introducing a significant relaxation of the wave equation at the benefit of the observation equation during the early iterations, while satisfying the two equations at the convergence point. We solve the saddle point problem, equation 4.58, with the method of multiplier, in which the primal variables, $\mathbf{U}$ and $\mathbf{m}$, and the dual variables, $\tilde{\mathbf{D}}$ and $\tilde{\mathbf{S}}$, are updated in alternating mode. The dual problem is iteratively solved with basic gradient ascent steps. Accordingly, we immediately deduce from equation 4.58 that the scaled dual variables $\tilde{\mathbf{D}}$ and $\tilde{\mathbf{S}}$ are formed by the running sum of the constraint violations (the data and source residuals) in iterations. They update the RHSs (the data and the sources) of the quadratic penalty functions in equation 4.58 to refine the primal variables $\mathbf{U}$ and $\mathbf{m}$ at a given iteration from the residual source and data errors (this RHS updating is a well known procedure to iteratively refine solutions of ill-posed linear inverse problems). The bi-variate primal problem, equation 4.56, is biconvex due to the bilinearity of the wave equation highlighted in the previous section. Therefore, it can be broken down into two linear subproblems for $\mathbf{U}$ and $\mathbf{m}$, which can be solved efficiently in alternating mode with ADMM after noting that the TV regularizer is convex (Aghamiry et al., 2019b).

As pointed out by Aghamiry et al. (2019c), a key advantage of augmented Lagrangian methods compared to penalty methods is that fixed penalty parameters λ_0 and λ_1 can be used in

iterations, because the Lagrange multipliers progressively correct for the constraint violations generated by the penalty terms through the above mentioned RHS updating.

Starting from an initial model $\mathbf{m}$ and zero-valued dual variables, the kth ADMM iteration embeds the following steps (see Appendix B, section 4.2.8, for the complete development):

Step 1: The primal wavefield reconstructions.

Build regularized wavefields by solving the following multi-RHS system of linear equations with direct or iterative methods suitable for sparse matrices:

$$\begin{aligned}&\left[\lambda_0 \mathbf{P}^T\mathbf{P} + \lambda_1 \mathbf{A}(\mathbf{m}^k)^T\mathbf{A}(\mathbf{m}^k)\right]\mathbf{U} = \\ &\left[\lambda_0 \mathbf{P}^T[\mathbf{D} + \tilde{\mathbf{D}}^k] + \lambda_1 \mathbf{A}(\mathbf{m}^k)^T[\mathbf{S} + \tilde{\mathbf{S}}^k]\right].\end{aligned} \tag{4.59}$$

By choosing a small value of λ_0/λ_1, the reconstructed wavefields closely fit the observations during the early iterations, while only weakly satisfying the wave equation. Problem 4.59 can be also interpreted as an extrapolation problem to reconstruct $\mathbf{U}$, when the observation equation (i.e. $\mathbf{PU} = \mathbf{D}$) is augmented with the wave equation.

To mitigate the computational burden of the wavefield reconstruction, we solve equation 4.59 with a fourth-order wave equation operator following the parsimonious approach of Operto et al. (2014), while the subsequent model estimation subproblem relies on the bilinear wave equation provided in equation 4.53. The elimination procedure allowing to transform the system of two second-order wave equations for $\mathbf{u}_x$ and $\mathbf{u}_z$, equation 4.53, into a fourth-order wave equation for $\mathbf{u}_x$ coupled with the closed-form expression of $\mathbf{u}_z$ as a function of $\mathbf{u}_x$ is reviewed in Appendix C (section 4.2.9).

Step 2: The primal model estimation.

The reconstructed wavefields, equation 4.59, are injected in the linearized equation 4.55 to update the subsurface parameters by solving the following system of linear equations:

$$\begin{aligned}\left[\lambda_1 \sum_{l=1}^{n_s} \mathbf{L}_l^T\mathbf{L}_l + \overline{\nabla}^T\Gamma\overline{\nabla} + Z\mathbf{I}\right]\mathbf{m} = \lambda_1 \sum_{l=1}^{n_s} \mathbf{L}_l^T(\mathbf{y}_l^k + \tilde{\mathbf{s}}_l^k) \\ + \overline{\nabla}^T\Gamma(\mathbf{p}^k + \tilde{\mathbf{p}}^k) + Z(\mathbf{q}^k + \tilde{\mathbf{q}}^k),\end{aligned} \tag{4.60}$$

where

$$\overline{\nabla} = \begin{bmatrix}\partial_x \\ \partial_z\end{bmatrix}. \tag{4.61}$$

In equation 4.60, $\mathbf{p}$ and $\mathbf{q}$ are auxiliary primal variables, which have been introduced to solve the bound-constrained TV-regularized parameter estimation subproblem with the split Bregman method (Appendix B, section 4.2.8). The vectors $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{q}}$ are the corresponding dual variables. These auxiliary primal and dual variables are initialized to $\mathbf{0}$ during the first ADMM iteration.

The operator Z is a diagonal weighting matrix defined as

$$Z = \begin{bmatrix} \zeta_{v_0}\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \zeta_{\epsilon}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \zeta_{\delta}\mathbf{I} \end{bmatrix} \in \mathbb{R}_{+}^{3n\times 3n}, \tag{4.62}$$

where $\zeta_{v_0}, \zeta_{\epsilon}, \zeta_{\delta} > 0$ control the relative weights assigned to the bound constraints applied on the three parameter classes. Note that the bound constraints introduce also a damping (DMP) or zero-order Tikhonov regularization in the Hessian of equation 4.60.

In the same way, $\mathbf{\Gamma}$ is a diagonal matrix defined as

$$\mathbf{\Gamma} = \begin{bmatrix} \mathbf{\Gamma}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}_{22} \end{bmatrix}, \tag{4.63}$$

where

$$\mathbf{\Gamma}_{11} = \mathbf{\Gamma}_{22} = \begin{bmatrix} \gamma_{v_0}\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \gamma_{\epsilon}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \gamma_{\delta}\mathbf{I} \end{bmatrix} \in \mathbb{R}_{+}^{3n\times 3n},$$

and $\gamma_{v_0}, \gamma_{\epsilon}, \gamma_{\delta} > 0$ control the soft thresholding that is performed by the TV regularizer, equation 4.65. We remind that augmented Lagrangian methods seek to strictly satisfy the constraints at the convergence point only, not at each iteration. Therefore, the relative values of these penalty parameters have a significant impact upon the path followed by the inversion to converge toward this convergence point.

Step 3: The TV primal update.

Update the TV primal variable $\mathbf{p}$ via a TV proximity operator. Set

$$\mathbf{z} \leftarrow \overline{\nabla}\mathbf{m}^{k+1} - \tilde{\mathbf{p}}^{k} = \begin{bmatrix} \mathbf{z}_x \\ \mathbf{z}_z \end{bmatrix},$$

then

$$\mathbf{p}^{k+1} \leftarrow \text{prox}_{\Gamma^{-1}}(\mathbf{z}) = \begin{bmatrix} \xi \circ \mathbf{z}_x \\ \xi \circ \mathbf{z}_z \end{bmatrix}, \tag{4.64}$$

where

$$\xi = \max(1 - \frac{1}{\mathbf{\Gamma}\sqrt{\mathbf{z}_x^2 + \mathbf{z}_z^2}}, 0) \tag{4.65}$$

and $\mathbf{x} \circ \mathbf{y}$ denotes the Hadamard (component-wise) product of $\mathbf{x}$ and $\mathbf{y}$. Also, the power of 2 indicates the Hadamard product of $\mathbf{x}$ with itself, i.e. $\mathbf{x}^2 = \mathbf{x} \circ \mathbf{x}$.

Step 4: The bounding constraint primal update.

Update the primal variable $\mathbf{q}$ via a projection operator, which has the following component-wise form

$$\mathbf{q}^{k+1} \leftarrow \text{proj}_{\mathcal{C}}(\mathbf{m}^{k+1} - \tilde{\mathbf{q}}^{k}), \tag{4.66}$$

where the projection operator is given by

$$\text{proj}_{\mathcal{C}}(\mathbf{x}) = \min(\max(\mathbf{x}, \mathbf{m}_l), \mathbf{m}_u).$$

Step 5: Dual updates.

Update the scaled dual variables with gradient ascent steps

$$\begin{cases} \tilde{\mathbf{S}}^{k+1} & \leftarrow \tilde{\mathbf{S}}^{k} + \mathbf{S} - \mathbf{A}(\mathbf{m}^{k+1})\mathbf{U}^{k+1}, \\ \tilde{\mathbf{D}}^{k+1} & \leftarrow \tilde{\mathbf{D}}^{k} + \mathbf{D} - \mathbf{P}\mathbf{U}^{k+1}, \\ \tilde{\mathbf{p}}^{k+1} & \leftarrow \tilde{\mathbf{p}}^{k} + \mathbf{p}^{k+1} - \overline{\nabla}\mathbf{m}^{k+1}, \\ \tilde{\mathbf{q}}^{k+1} & \leftarrow \tilde{\mathbf{q}}^{k} + \mathbf{q}^{k+1} - \mathbf{m}^{k+1}, \end{cases} \tag{4.67}$$

Step 6: Check the stopping condition.

Exit if the preset stopping conditions are satisfied else go to step 1. We will describe the stopping criteria of iterations in the following "Numerical example" section for each numerical example

Hyperparameter tuning

We tune the different penalty parameters by extending the procedure reviewed by Aghamiry et al. (2019b, 2020b) to multiparameter reconstruction.

We start from the last subproblem of the splitting procedure and set the penalty parameters contained in $\mathbf{\Gamma}$. These hyperparameters control the soft thresholding performed by the TV regularization, equation 4.64. We set $\gamma_i = 2\% \max \sqrt{\mathbf{z}_{i_x}^2 + \mathbf{z}_{i_z}^2}$, where the subscript $i \in \{v_0, \epsilon, \delta\}$ denotes the parameter class ($\mathbf{m}_{v_0}$, $\mathbf{m}_{\epsilon}$, $\mathbf{m}_{\delta}$). This tuning can be refined by using a different thresholding percentage for each parameter class adaptively during iterations or according to prior knowledge of the geological structure, coming from well logs for example. Also, we use the same weight for the damping regularization associated with the bound constraints and the TV regularization: $\zeta_i = \gamma_i$.
Then, we select λ_1 as a percentage of the mean absolute value of the diagonal coefficients of $\sum_{l=1}^{n_s} \mathbf{L}_l^T \mathbf{L}_l$ during the parameter estimation subproblem, equation 4.60. This percentage is set according to the weight that we want to assign to the TV regularization and the bound constraints relative to the wave equation constraint during the parameter estimation. Parameter λ_1 may be increased during iterations to reduce the weight of TV regularization and bound constraints near the convergence point. We found this adaptation useful when we start from very

crude initial models. Finally, we set λ_0 such that $\lambda = \lambda_1/\lambda_0$ is a small fraction of the highest eigenvalue ξ of the normal operator $\mathbf{A}(\mathbf{m})^{-T}\mathbf{P}^T\mathbf{P}\mathbf{A}(\mathbf{m})^{-1}$ during the wavefield reconstruction subproblem, equation 4.59, according to the criterion proposed by van Leeuwen and Herrmann (2016). In all the numerical tests, we use $\lambda =$ 1e-2ξ and $\lambda =$ 1e-0ξ for noiseless and noisy data, respectively. This tuning of λ is indeed important because it controls the extension of the search space. A too high value of λ reduces the weight of $\|\mathbf{PU} - \mathbf{D}\|_2^2$ during the wavefield reconstruction and makes IR-WRI behave like a reduced approach. Conversely, using a small value for λ fosters data fitting and expends the search space accordingly. However, a too small value can lead to a prohibitively high number of iterations of the augmented Lagrangian method before the wave equation constraint is fulfilled with sufficient accuracy. Moreover, when data are contaminated by noise, a too small value for λ will make the wavefield reconstruction over-fit the data and drive the algorithm to be a poor minimizer. We always use λ as a fixed percentage of ξ in iterations.

4.2.4 Numerical examples

Inclusion test

We first assess multi-parameter IR-WRI with a simple inclusion example for noiseless data. The experimental setup in terms of model, acquisition geometry and frequency selection is identical to that used by Gholami et al. (2013b). The vertical velocity $\mathbf{v}_0$ and the Thomsen's parameters δ and ϵ are 3 km/s, 0.05 and 0.05, respectively, in the homogeneous background model and 3.3 km/s, 0.1 and 0.2, respectively, in the inclusion of radius 100 m. Nine frequencies between 4.8 Hz and 19.5 Hz are processed simultaneously during IR-WRI and a maximum of 25 iterations is used as stopping criterion for iterations. An ideal fixed-spread acquisition is used, where 64 sources and 320 receivers surround the inclusion, hence providing a complete angular illumination of the target.

Although we use $(1/\mathbf{v}_0^2, \sqrt{1+2\delta}, 1+2\epsilon)$ as optimization parameters during our inversions, we show the reconstructed models under the form of $\mathbf{v}_0$, δ and ϵ for comparison with the results of Gholami et al. (2013b). Note that the radiation patterns of the $(1/\mathbf{v}_0^2, \sqrt{1+2\delta}, 1+2\epsilon)$ parametrisation are scaled versions of those of the $(\mathbf{v}_0, \delta, \epsilon)$ parametrisation: namely, they exhibit the same amplitude variation with scattering angle, with however different amplitudes. This means that, for an equivalent regularization and parameter scaling, we expect similar resolution and trade-off effects with these two parametrisations. Note that $\mathbf{v}_0$ is provided in km/s in our inversion such that the order of magnitude of $1/\mathbf{v}_0^2$ is of the order of δ and ϵ.

We start with bound-constrained IR-WRI with damping (DMP) regularization only ($\gamma_i = 0, i \in \{v_0, \epsilon, \delta\}$ in equation 4.60) and perform three independent mono-parameter reconstructions for $\mathbf{m}_{v_0}$, $\mathbf{m}_\delta$ and $\mathbf{m}_\epsilon$, respectively. For each mono-parameter inversion, the true model associated with the optimization parameter contains the inclusion, while the true models associated with the two passive parameters are homogeneous (Fig. 4.1). For all three inversions, the starting models are the true homogeneous background models. In other words, the data residuals contain only the footprint of the mono-parameter inclusion to be reconstructed. This

test is used to assess the intrinsic resolution of IR-WRI for each parameter reconstruction, independently from the cross-talk issue (Gholami et al., 2013b). It is reminded that this intrinsic resolution is controlled by the frequency bandwidth, the angular illumination provided by the acquisition geometry and the radiation pattern of the optimization parameter in the chosen subsurface parametrisation. As this test fits the linear regime of classical FWI, we obtain results (Fig. 4.1) very similar to those obtained by Gholami et al. (2013b, Their Fig. 9), where the shape of the reconstructed inclusions is controlled by the radiation pattern of the associated parameter. The reader is referred to Gholami et al. (2013b) for a detailed analysis of these radiation patterns.
Then, we perform the joint reconstruction of $\mathbf{m}_{v_0}$, $\mathbf{m}_\delta$ and $\mathbf{m}_\epsilon$, when the true model contains an inclusion for each parameter class (Fig. 4.2). With our parameter scaling and subsurface parametrisation, $\mathbf{v}_0$, ϵ and δ are reconstructed with well balanced amplitudes compared to the results of Gholami et al. (2013b, Their Fig. 10), where $\mathbf{v}_0$ has a dominant imprint in the inversion. Comparing the models reconstructed by the mono-parameter and multi-parameter inversions highlights however the wavenumber leakage generated by parameter cross-talks (Figs. 4.1 and 4.2).

Then, we complement DMP regularization with TV regularization during bound-constrained IR-WRI for the above mono-parameter and multi-parameter experiments (Figs. 4.3 and 4.4). The results show how TV regularization contributes to remove the wavenumber filtering performed by radiation patterns (compare for example Figs. 4.1 and 4.3) as well as the cross-talk artifacts during the multi-parameter inversion (compare Figs. 4.2 and 4.4). Although this toy example has been designed with a piecewise constant model which is well suited for TV regularization, yet it highlights the potential role of TV regularization to mitigate the ill-posedness of FWI resulting from incomplete wavenumber coverage and parameter trade-off.

Synthetic North Sea case study

Experimental setup

We consider now a more realistic $16\ km \times 5.2\ km$ shallow-water model representative of the North Sea (Munns, 1985). The reader is also referred to Gholami et al. (2013a) for an application of acoustic VTI FWI on this model. The true model and the initial (starting) models for $\mathbf{v}_0$, δ and ϵ are shown in Fig. 4.5. The subsurface model is formed by soft sediments in the upper part, a pile of low-velocity gas layers above a chalk reservoir, the top of which is indicated by a sharp positive velocity contrast at around 2.5 km depth, and a flat reflector at 5 km depth (Fig. 4.5a). The initial $\mathbf{v}_0$ model is laterally homogeneous with velocity linearly increasing with depth between 1.5 to 3.2 km/s (Fig. 4.5b), while the δ and ϵ initial models are Gaussian filtered version of the true models. Note that our initial/background $\mathbf{v}_0$, δ and ϵ models are cruder than those used by Gholami et al. (2013a). The fixed-spread surface acquisition consists of 80 (reciprocal) explosive sources spaced 200 m apart at 75 m depth on the sea bottom and 320 (reciprocal) hydrophone receivers spaced 50 m apart at 25 m depth. Accordingly, the pressure wavefield is considered for the acoustic VTI inversion. A free-surface boundary condition is used on top of the grid and the source signature is a Ricker wavelet with a 10 Hz dominant frequency. We compute the recorded data in the true models (Figs. 4.5a,c,e) with the

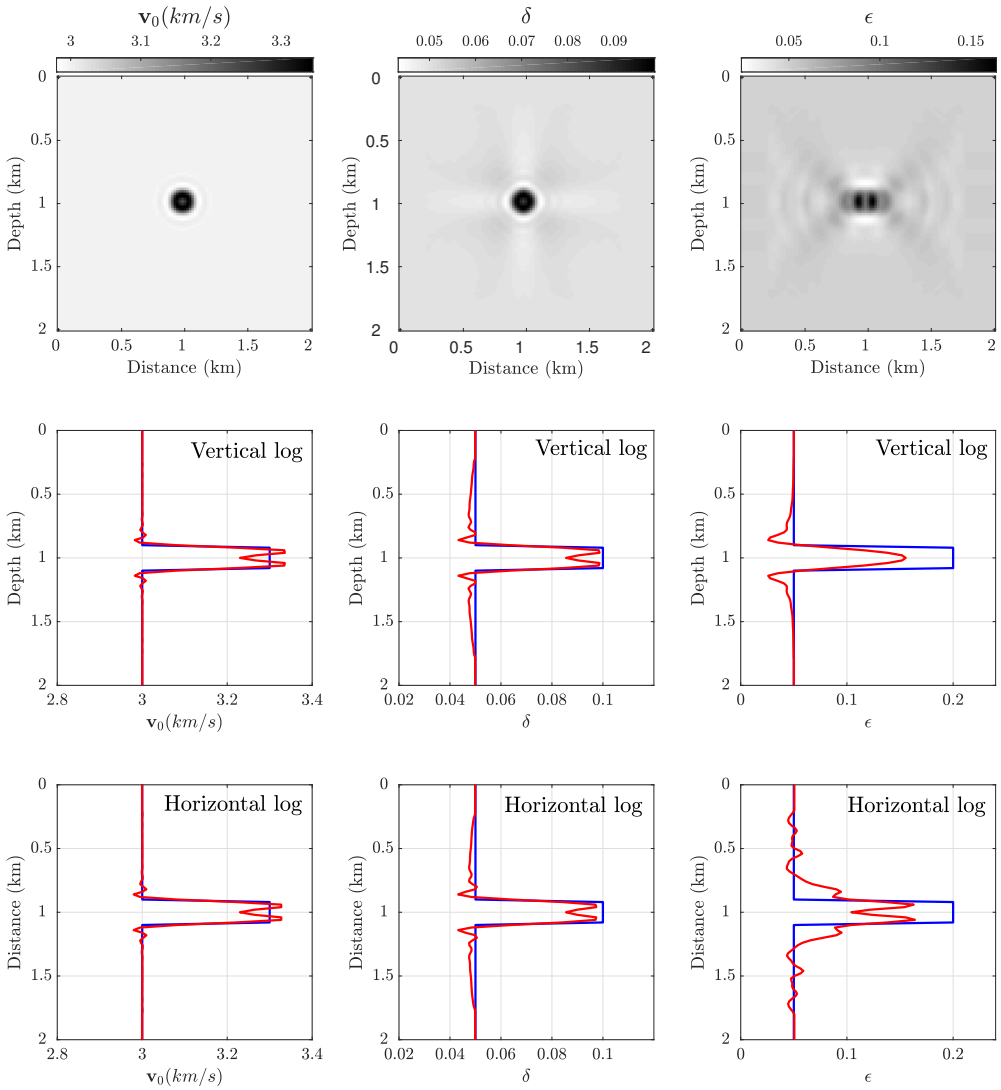

Figure 4.1 – Inclusion test: Mono-parameter IR-WRI results. For $\mathbf{v}_0$ (first column), δ (second column), and ϵ (third column). The initial models are the true homogeneous background models. The vertical and horizontal profiles in the true model (blue) and estimated model (red) are extracted across the center of the inclusions.

forward modelling engine described in Appendix C (section 4.2.9). During IR-WRI, we use the same forward engine to compute the modelled data according to an inverse crime procedure. Common-shot gathers computed in the true model and in the initial model for a shot located at 14 km are shown in Fig. 4.6. The seismograms computed in the true model are dominated by the direct wave, the diving waves from the sedimentary overburden, complex packages of pre- and post-critical reflections from the gas layers, the top of the reservoir and the deep reflector. The refracted wave from the deep reflector is recorded at a secondary arrival between -14km and -10km offset in Fig. 4.6a. Also, energetic reverberating P-wave reflections are generated by the wave guide formed by the shallow water layer and the weathering layer. They take the form of leaking modes with phase velocities higher than the water wave speed (Operto and

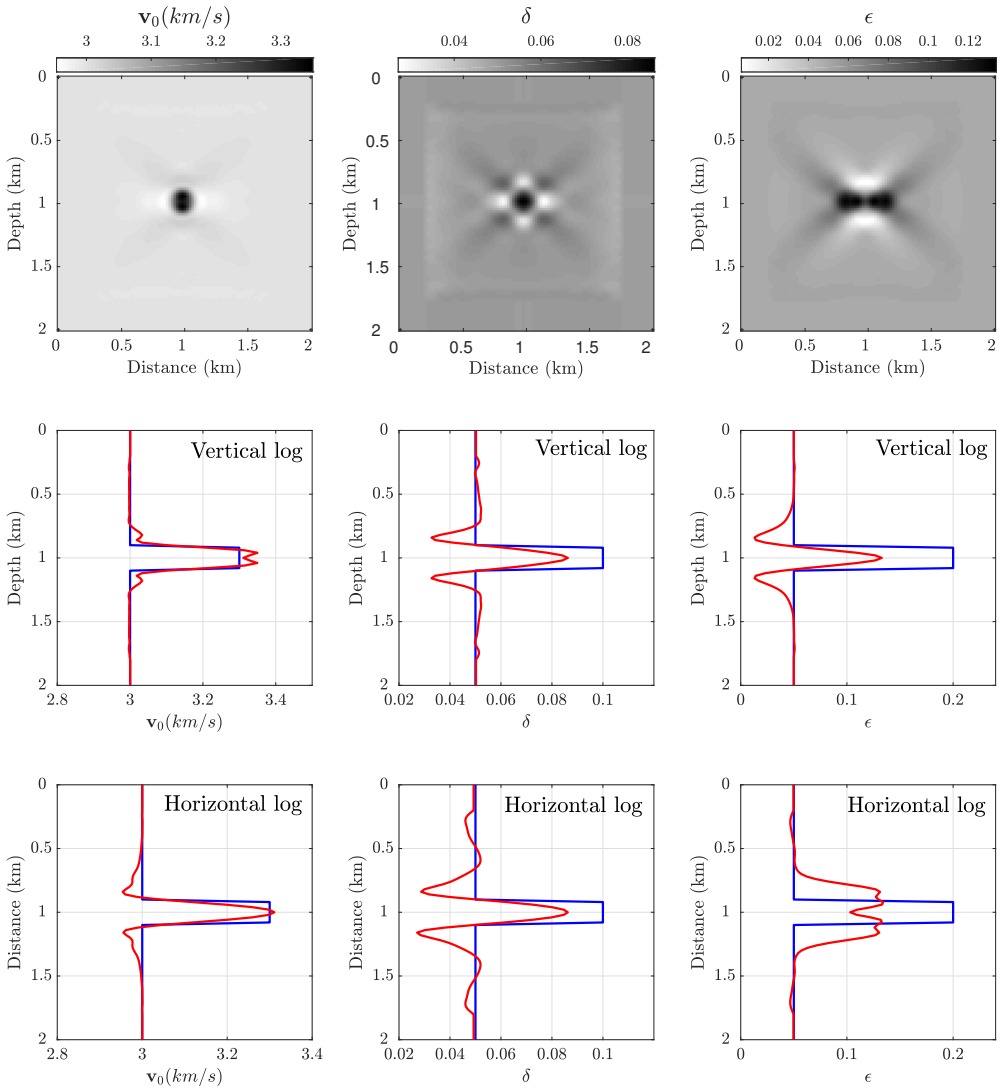

Figure 4.2 – Inclusion test: Multiparameter IR-WRI results for joint update of $\mathbf{v}$ (first column), δ (second column), and ϵ (third column). The initial models are the homogeneous background models.

Miniussi, 2018). The seismograms computed in the starting model mainly show the direct wave and the diving waves, these latter being highly cycle skipped relative to those computed in the true model.

We perform both mono-parameter IR-WRI for $\mathbf{m}_{v_0}$ and multi-parameter IR-WRI for $\mathbf{m}_{v_0}$ and $\mathbf{m}_\epsilon$. In both cases, we compare the results that are obtained when bound-constrained IR-WRI is performed with DMP regularization only ($\gamma_i = 0$) and with DMP + TV regularization. When $\mathbf{m}_\epsilon$ is involved as an optimization parameter, we had to introduce an additional regularization term $\|\mathbf{m} - \mathbf{m}_\epsilon^0\|$ in the parameter-estimation subproblem, equation 4.82a, in order to force the updates of ϵ to be smooth and close to ϵ^0. The background models used for the passive anisotropic parameters (either ϵ and δ or δ alone) are the smooth models (Fig. 4.5d,f), which means that the IR-WRI results will be impacted upon by the smoothness of the passive

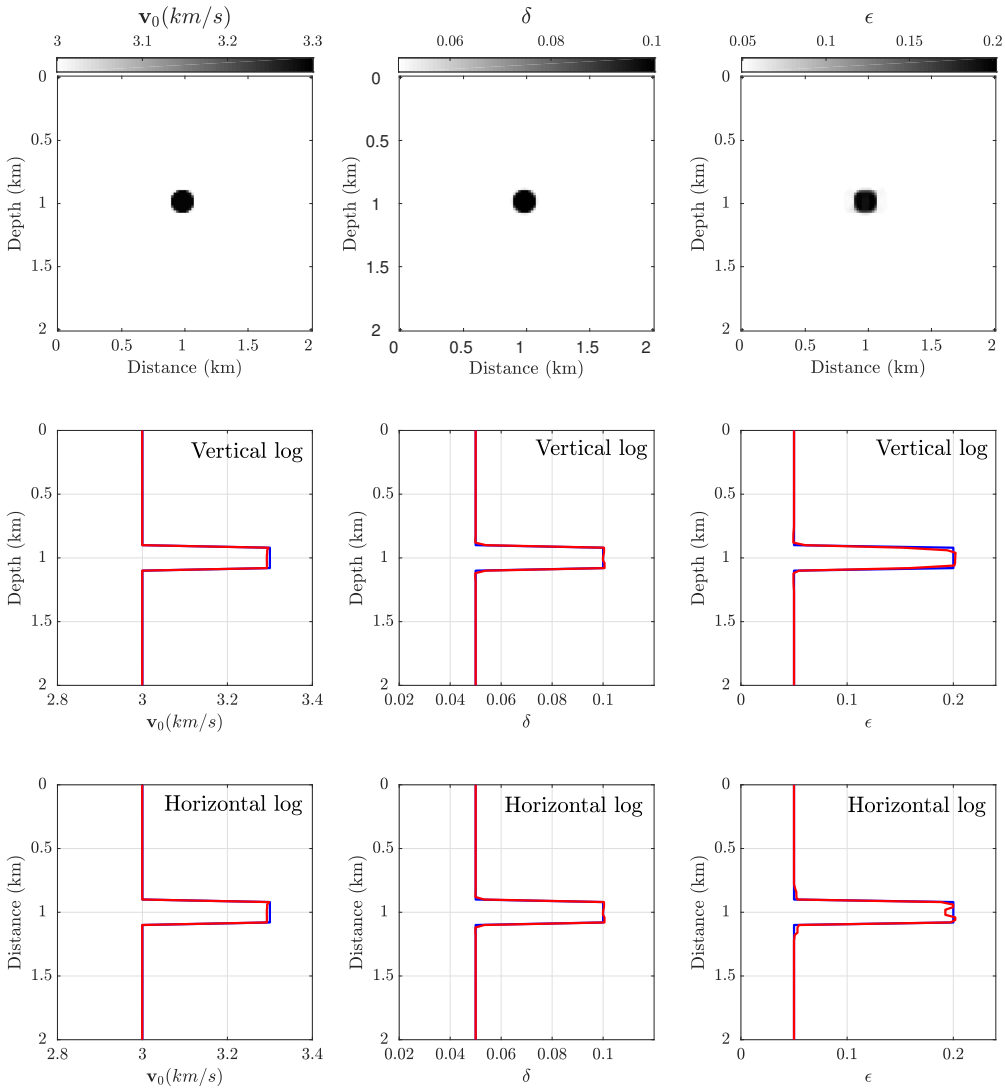

Figure 4.3 – Inclusion test: Same as Fig. 4.1, but with TV regularization.

parameters.

We perform IR-WRI with small batches of two frequencies with one frequency overlap between two consecutive batches, moving from the low frequencies to the higher ones according to a classical frequency continuation strategy. The starting and final frequencies are 3 Hz and 15 Hz and the sampling interval in one batch is 0.5 Hz. The stopping criterion for iterations and for each batch is given by $k_{max} = 15$ or

$$\|\mathbf{A}(\mathbf{m}^{k+1})\mathbf{U}^{k+1} - \mathbf{S}\|_F \leq \varepsilon_b \;\&\; \|\mathbf{P}\mathbf{U}^{k+1} - \mathbf{D}\|_F \leq \varepsilon_d, \tag{4.68}$$

where k_{max} denotes the maximum iteration count. For noiseless data, ε_b and ε_d are set to 1e-3 and 1e-5, respectively. For noisy data, they are set to 1e-3 and the noise level of the batch, respectively.

We perform three paths through the frequency batches to improve the IR-WRI results, using

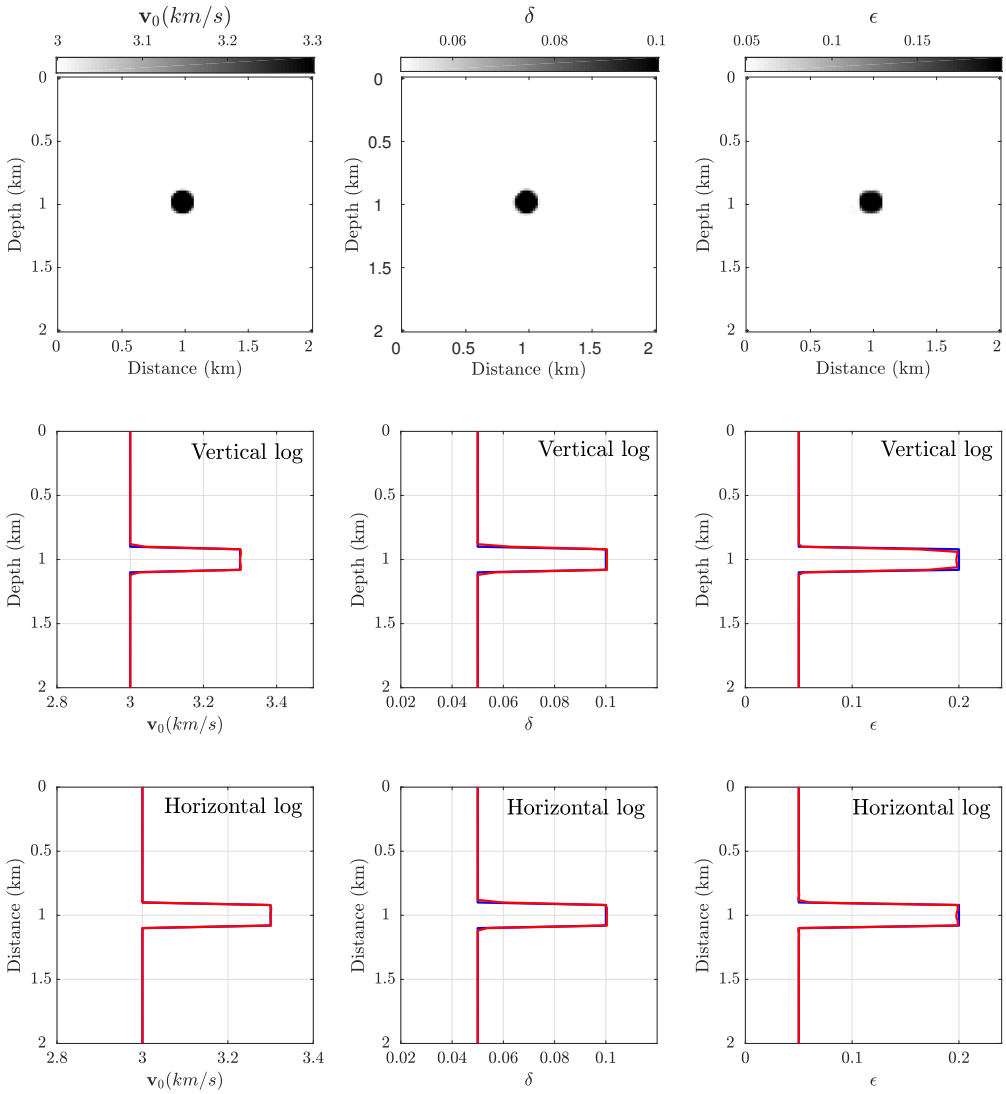

Figure 4.4 – Inclusion test: Same as Fig. 4.2, but with TV regularization.

the final model of one path as the initial model of the next one (these cycles can be viewed as outer iterations of IR-WRI). The starting and finishing frequencies of the paths are [3, 6], [4, 8.5], [6, 15] Hz respectively, where the first element of each pair shows the starting frequency and the second one is the finishing frequency.

Convexity and sensitivity analysis

Before discussing the IR-WRI results, we illustrate how WRI extends the search space of FWI for the North Sea case study. For this purpose, we compare the shape of the FWI misfit function with that of the parameter-estimation WRI subproblem for the 3 Hz frequency and for

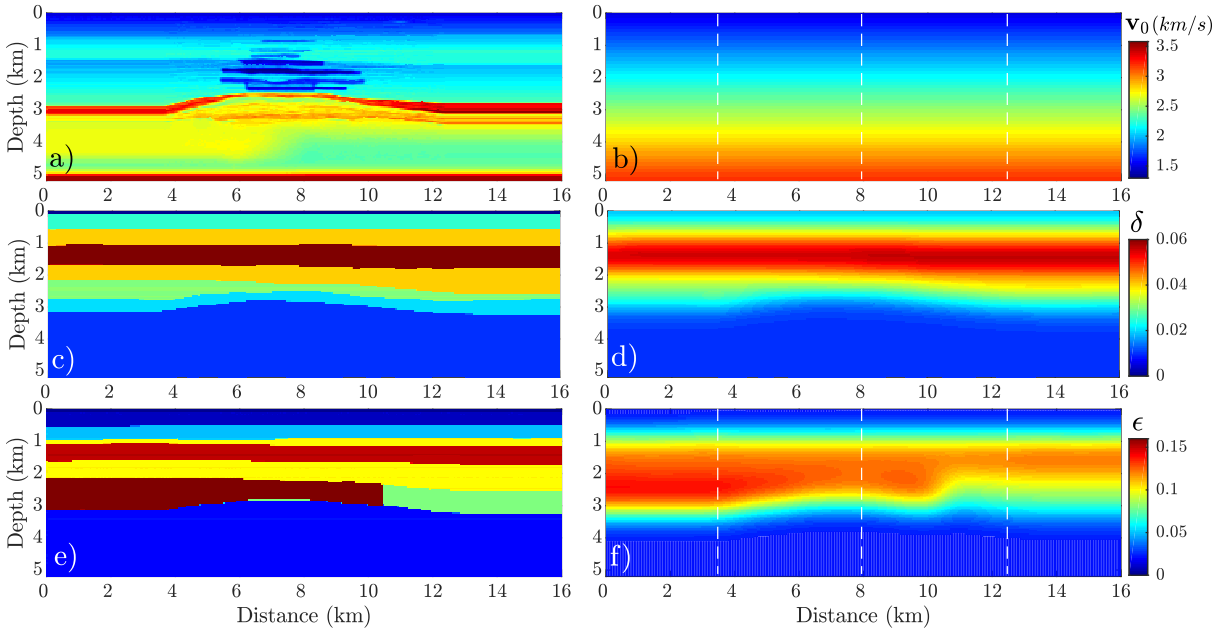

Figure 4.5 – North Sea case study. (a) True $\mathbf{v}_0$ model. (b) Initial $\mathbf{v}_0$ model. (c) True δ model. (d) Smoothed δ model, which is used as passive background model during inversion tests. (e) True ϵ model. (f) Smoothed ϵ model, which is used as a passive background model during the mono-parameter inversion and as an initial model during the joint reconstruction of $\mathbf{v}_0$ and ϵ. The vertical dashed lines in (b) and (f) indicate the location of vertical logs.

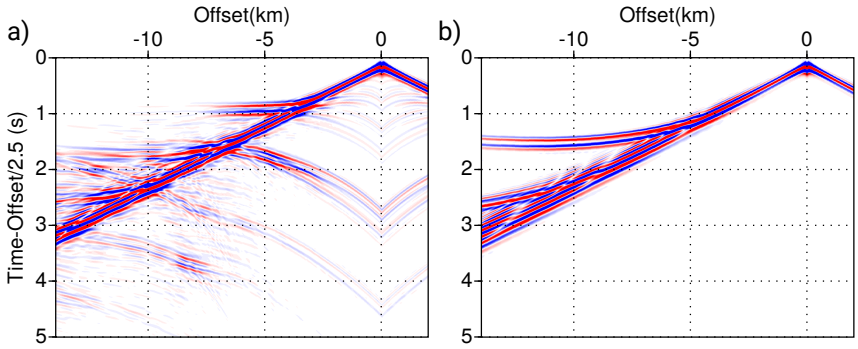

Figure 4.6 – North Sea case study. Time domain seismograms computed in (a) the true model (Fig. 4.5a,c,e) and in (b) the initial model (Fig. 4.5b,d,f). The seismograms are plotted with a reduction velocity of 2.5 km/s.

a series of $\mathbf{v}_0$ and ϵ models that are generated according to

$$\mathbf{v}_0(\alpha) = \mathbf{v}_{0_{true}} + |\alpha|[\mathbf{v}_{0_{init}} - \mathbf{v}_{0_{true}}], \tag{4.69a}$$

$$\epsilon(\beta) = \epsilon_{true} + |\beta|[\epsilon_{init} - \epsilon_{true}], \tag{4.69b}$$

where $\mathbf{v}_{0_{true}}$ and $\mathbf{v}_{0_{init}}$ denote the true and the initial $\mathbf{v}_0$ models, respectively (Fig. 4.5a,b) and $-1 \leq \alpha \leq 1$. Similarly, ϵ_{true} and ϵ_{init} are the true and the initial ϵ models, respectively (Fig. 4.5e,f) and $-1 \leq \beta \leq 1$. Finally, we use the true δ model (Fig. 4.5c) to generate the recorded

data and the smoothed version (Fig. 4.5d) as a passive parameter to evaluate the misfit function. The misfit functions of the classical reduced-approach FWI as well as that of WRI are shown in Fig. 4.7. The WRI misfit function is convex, while that of FWI exhibits spurious local minima along both the α and β dimensions. Also, the sensitivity of the misfit function to $\mathbf{v}_0$ is much higher than that of ϵ for the considered range of models as already pointed out by Gholami et al. (2013b), Gholami et al. (2013a) and Cheng et al. (2016, Their Fig. 2). The weaker sensitivity of the misfit function to ϵ makes the joint update of $\mathbf{v}_0$ and ϵ challenging.

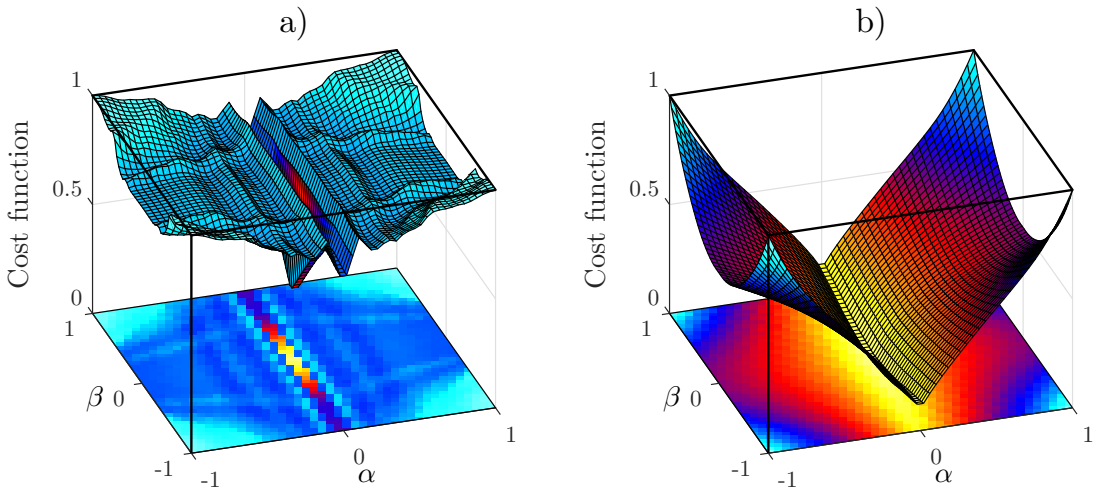

Figure 4.7 – North Sea case study. 3-Hz misfit function for the first iteration when $\mathbf{v}_0$ and ϵ are active parameters and δ is kept fixed to the smooth background model (Fig. 4.5f). (a) Classical reduced approach, (b) WRI. The variables α and β parametrize the $\mathbf{v}_0$ and ϵ models, respectively, for which the misfit function is computed (see text for details).

Mono-parameter IR-WRI

We start with mono-parameter bound-constrained IR-WRI when $\mathbf{v}_0$ is the optimization parameter and δ and ϵ are passive parameters. We first consider noiseless data.
The final $\mathbf{v}_0$ models inferred from bound-constrained IR-WRI with DMP regularization only and DMP+TV regularizations are shown in Fig. 4.8. Also, direct comparisons between the logs extracted from the true model, the initial model, and the IR-WRI models at $x = 3.5\ km$, $x = 8.0\ km$ and $x = 12.5\ km$ are shown in Fig. 4.9. Although the crude initial $\mathbf{v}_0$ model and the smooth δ and ϵ passive models, the shallow sedimentary part and the gas layers are fairly well reconstructed with the two regularization settings. The main differences are shown at the reservoir level and below. Without TV regularization, the reconstruction at the reservoir level is quite noisy and the inversion fails to reconstruct the smoothly-decreasing velocity below the reservoir due to the lack of diving wave illumination at these depths. This in turn prevents the focusing of the deep reflector at 5 km depth by migration of the associated short-spread reflections. When TV regularization is used, IR-WRI provides a more accurate and cleaner image of the reservoir and better reconstructs the sharp contrast on top of it. It also reconstructs the deep

reflector at the correct depth in the central part of the model, while the TV regularization has replaced the smoothly-decreasing velocities below the reservoir by a piecewise constant layer with a mean velocity (Fig. 4.9).

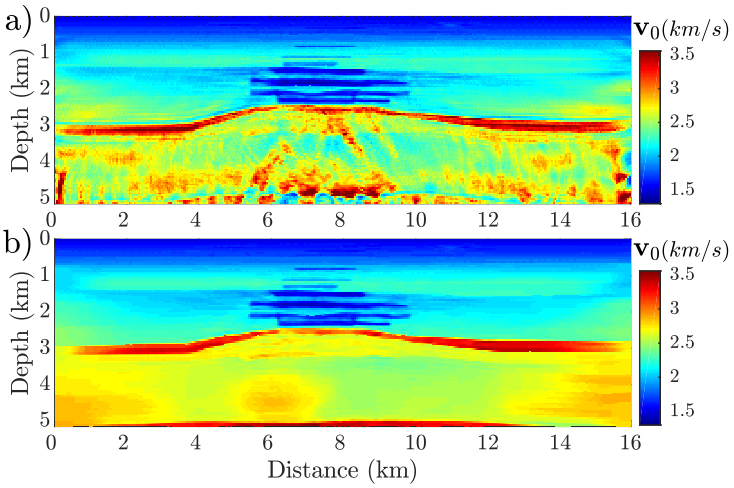

Figure 4.8 – North Sea case study with noiseless data. Mono-parameter IR-WRI. Vertical wavespeed $\mathbf{v}_0$ models inferred from IR-WRI with (a) DMP regularization, (b) TV regularization.

To generate more realistic test, we add random noise with Gaussian distribution to the data with SNR=10 db and re-run the mono-parameter bound-constrained IR-WRI with DMP and TV regularizations (Fig. 4.10). The direct comparison between the true model, the initial model and the IR-WRI models of Fig. 4.10 at distances $x = 3.5\ km$, $x = 8.0\ km$ and $x = 12.5\ km$ are shown in Fig. 4.11. The noise degrades the reconstruction of the gas layers both in terms of velocity amplitudes and positioning in depth when only DMP regularization is used. Also, the sharp reflector on top of the reservoir is now unfocused and mis-positioned in depth accordingly (Fig. 4.11). The TV regularization significantly reduces these amplitude and mis-positioning errors (Figs. 4.10b and 4.11) and hence produces $\mathbf{v}_0$ models which are much more consistent with those obtained for noiseless data (Figs. 4.8b and 4.9).

To assess how the differences between the velocity models shown in Fig. 4.8 impact waveform match, we compute time-domain seismograms in these models as well as the differences with those computed in the true model (Figs. 4.12 and 4.13). The time-domain seismograms and the residuals shown in Fig. 4.12 give an overall vision of the achieved data fit, while the direct comparison between the recorded and modelled seismograms shown in Fig. 4.13 allow for a more detailed assessment of the waveform match for a specific arrival.

A first conclusion is that, for all of the models shown in Fig. 4.8, the main arrivals, namely those which have a leading role in the reconstruction of the subsurface model (diving waves, pre- and post-critical reflections), are not cycle skipped relative to those computed in the true model. We note however more significant residuals for the reverberating guided waves when TV regularization is used. These mismatches were generated by small wavespeed errors generated in the shallow part of the model by the TV regularization. These artifacts can be probably

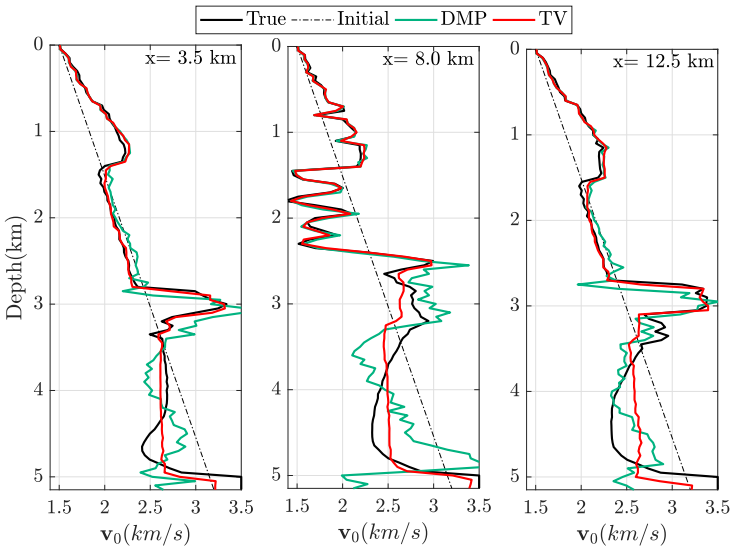

Figure 4.9 – North Sea case study with noiseless data. Direct comparison along the logs at $x = 3.5$ (left), $x = 8.0$ (center) and $x = 12.5$ km (right) between the true velocity model (black), the initial model (dashed line) and the bound constrained IR-WRI models shown in Fig. 4.8. The logs of the IR-WRI models obtained with DMP regularization and DMP+TV regularization are the green lines and the red lines, respectively.

corrected by deactivating or decreasing the weight of the TV regularization locally. For noiseless data, the direct comparison between the seismograms computed in the true model and in the reconstructed ones show how the DMP+TV regularization improves the waveform match both at pre- and post-critical incidences relative to the DMP regularization alone (Fig. 4.13a,b). For noisy data, the data fit is slightly degraded by noise when DMP regularization is used, while the TV regularization produces a data fit which is more consistent with that obtained with noiseless data (Compare Figs. 4.13a and 4.13c for DMP regularization, and Figs. 4.13b and 4.13d for DMP+TV regularization). It is striking to see the strong impact of noise on the quality of the $\mathbf{v}_0$ reconstruction when DMP regularization is used (compare Figs. 4.8a and 4.10a, and Figs. 4.9 and 4.11, green curves), compared to its more moderate impact on the data fit (Compare Figs. 4.13a and 4.13c). This highlights the ill-posedness of the FWI, which is nicely mitigated by the prior injected by the TV regularization as illustrated by the consistency of the $\mathbf{v}_0$ models inferred from noiseless and noisy data (Compare Figs. 4.13b and 4.13b).

Multi-parameter IR-WRI

We continue with multi-parameter bound-constrained IR-WRI when $\mathbf{v}_0$ and ϵ are optimization parameters and δ is a passive parameter. As for the mono-parameter inversion, we start with noiseless data. The final $\mathbf{v}_0$ and ϵ models inferred from bound-constrained IR-WRI with DMP and DMP+TV regularizations are shown in Fig. 4.14. The direct comparisons between the logs extracted from the true $\mathbf{v}_0$ model, the initial model, the bound constrained IR-WRI models with DMP and DMP+TV regularization at $x = 3.5$ km, $x = 8.0$ km and $x = 12.5$ km are shown in Fig. 4.15a, and the same comparisons for ϵ are depicted in Fig. 4.15b. Compared to the

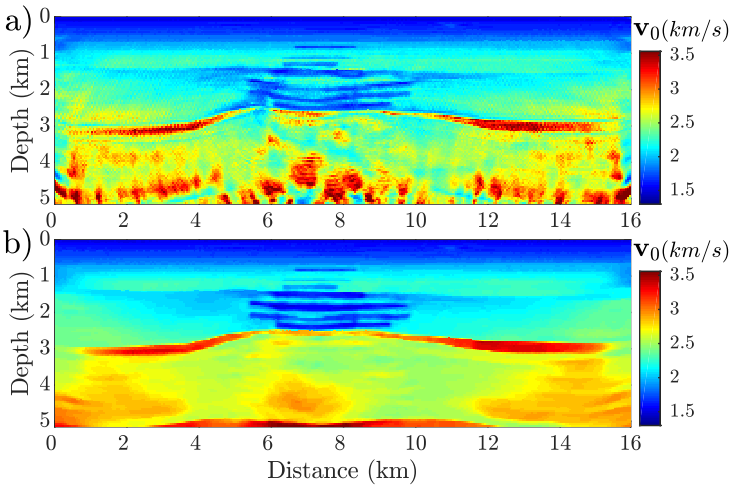

Figure 4.10 – North Sea case study with noisy data (SNR=10 db). Mono-parameter bound constrained IR-WRI with (a) DMP (b) TV regularization.

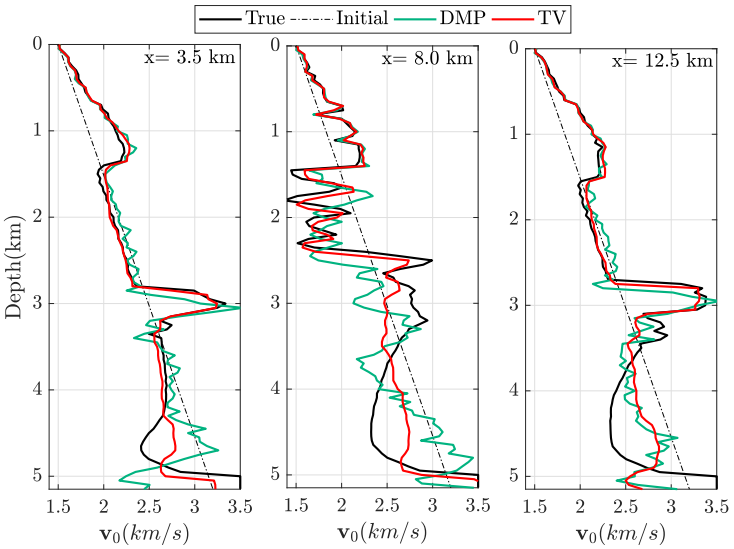

Figure 4.11 – North Sea case study with noisy data (SNR=10 db). Direct comparison along the logs at $x = 3.5$ (left), $x = 8.0$ (center) and $x = 12.5$ km (right) between the true velocity model (black), the initial model (dashed line) and the bound constrained IR-WRI estimated models with DMP, Fig. 4.8a, (green) and TV regularization, Fig. 4.8b, (red).

mono-parameter inversion results, involving ϵ as an optimization parameter clearly improves the reconstruction at the reservoir level down to around 3 km depth (compare Figs. 4.8 and 4.14a,b). The long to intermediate wavelengths of the ϵ model are primarily updated according to the radiation pattern of this parameter in the $(\mathbf{v}_0, \epsilon, \delta)$ parametrisation. The main effect of TV regularization relative to DMP regularization is to remove high-frequency noise from the ϵ model (Fig. 4.15b).

The time-domain seismograms computed in the multi-parameter models inferred from bound-constrained IR-WRI when DMP and DMP+TV regularizations are used are shown in Fig.

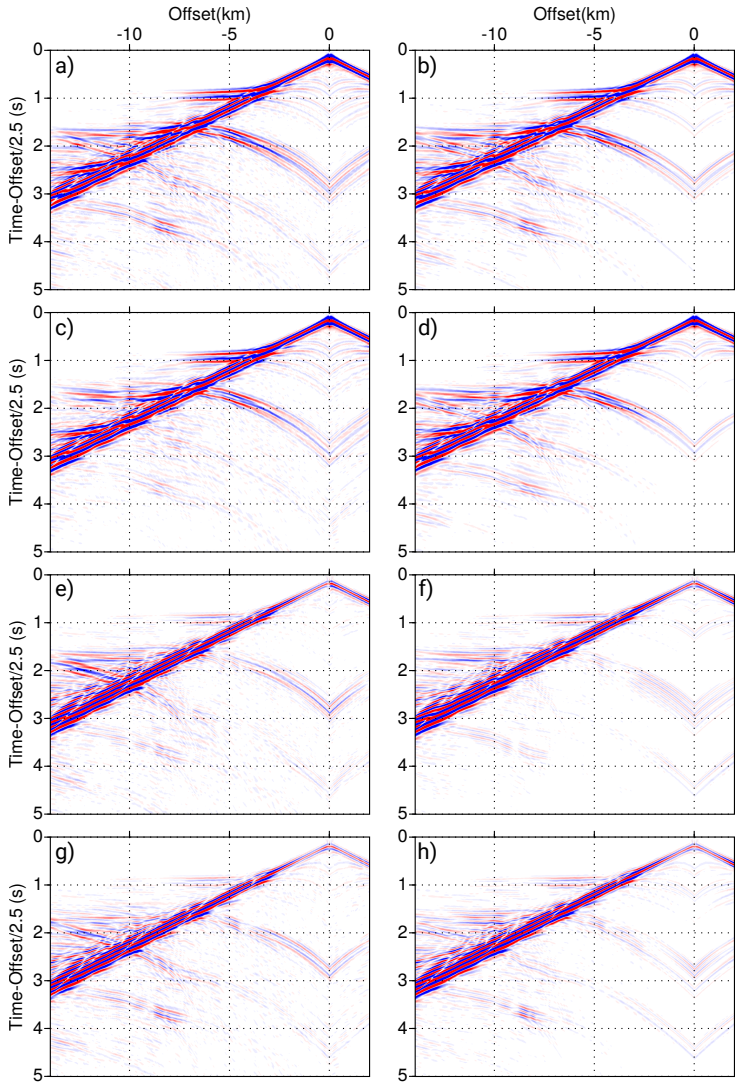

Figure 4.12 – North Sea case study. Time-domain seismograms computed in the subsurface models obtained by ($\mathbf{v}_0$) mono-parameter IR-WRI. (a-b) Noiseless data: (a) DMP regularization, (b) DMP+TV regularization. (c-d) Same as (a-b) for noisy data. (e-h) Residuals between the seismograms computed in the true model (Fig. 4.6a) and those shown in (a-d). The seismograms are plotted with a reduction velocity of 2.5 km/s.

4.18. The direct comparison between the recorded and modelled seismograms is shown in Fig. 4.19a,b. Clearly, using both $\mathbf{v}_0$ and ϵ as optimization parameters allows us to better conciliate the fit of the pre- and post-critical reflections (Compare Figs. 4.13a and 4.19a for DMP regularization, and Figs. 4.13b and 4.19b for DMP+TV regularization). With noiseless data, DMP regularization is enough to achieve a high-fidelity data fit, which looks better than that obtained with DMP+TV regularization (Compare Figs. 4.19a and 4.19b).

We repeat now the joint inversion when data are contaminated by Gaussian random noise with a SNR=10 db. The final $\mathbf{v}_0$ and ϵ models inferred from bound-constrained IR-WRI with DMP and DMP+TV regularization are shown in Fig. 4.16. Direct comparisons along logs

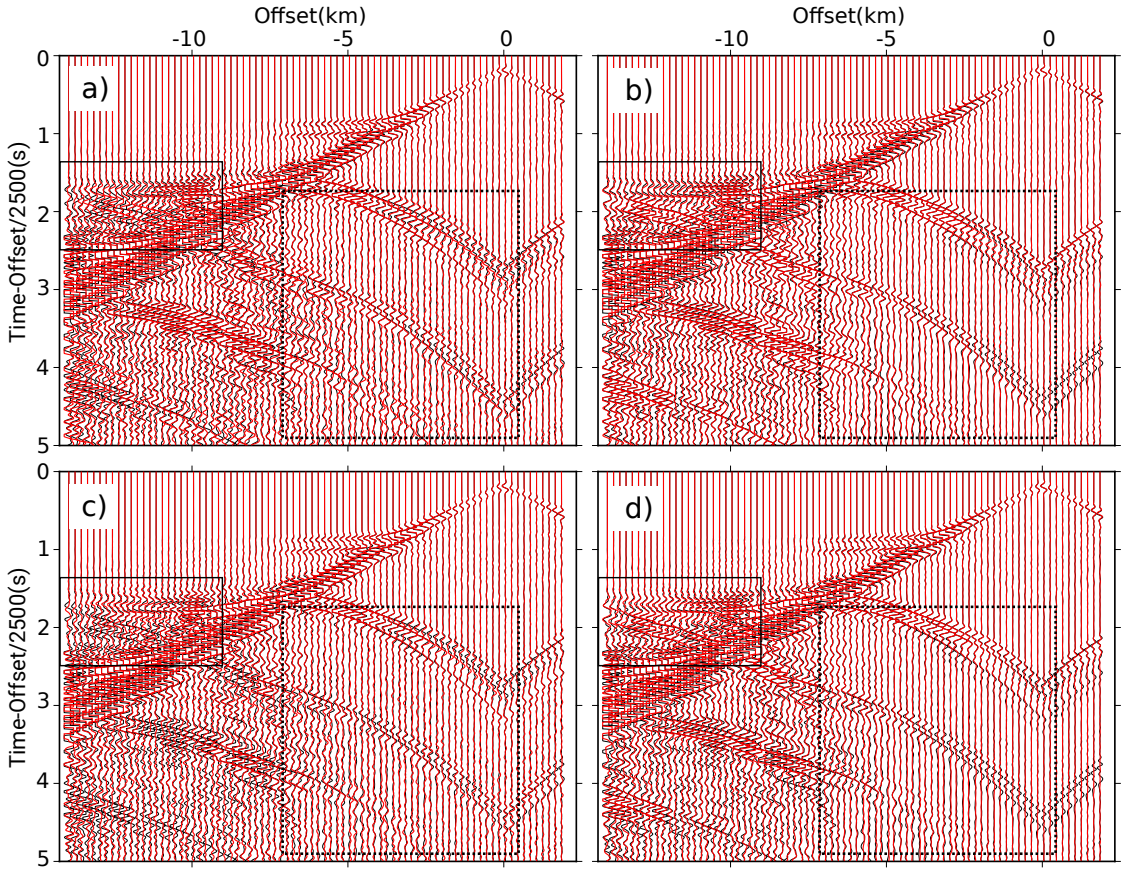

Figure 4.13 – North Sea case study. Mono-parameter bound-constrained IR-WRI results. Direct comparison between seismograms computed in the true model (black) and in the IR-WRI models (red). (a-b) Noiseless data (IR-WRI models of Fig. 4.8). (a) DMP regularization. (b) DMP+TV regularization. (c-d) Same as (a-b) for noisy data (IR-WRI models of Fig. 4.10). The seismograms are plotted with a reduction velocity of 2.5 km/s. True amplitudes are shown after a gain with offset and time for amplitude balancing. The solid box delineates post-critical reflections from the reservoir and the refracted wave from the deep interface, while the dot boxes delineate pre-critical reflections from the reservoir and the deep reflector. The amplitudes of the reverberating guided waves at long offsets are clipped for sake of clarity.

extracted from the true models, the initial models, and the models inferred from the bound constrained multi-parameter IR-WRI with DMP and DMP+TV regularization at $x = 3.5\ km$, $x = 8.0\ km$ and $x = 12.5\ km$ are shown in Fig. 4.17. For the $\mathbf{v}_0$ reconstruction, a trend similar to that shown for the mono-parameter inversion is shown, with a more significant impact of the noise on the velocity model reconstructed with DMP regularization compared to the one reconstructed with DMP+TV regularization (Compare Figs. 4.14 and 4.16). As expected, the impact of noise is more significant on the second-order ϵ model, even when TV regularization is used, in the sense that the estimated perturbations of the initial model have smaller amplitudes compared to the noiseless case (Fig. 4.17b).

The time-domain seismograms are shown in Figs. 4.18 and 4.19c,d. The data fit obtained with DMP regularization has been significantly degraded compared to that obtained with noise-

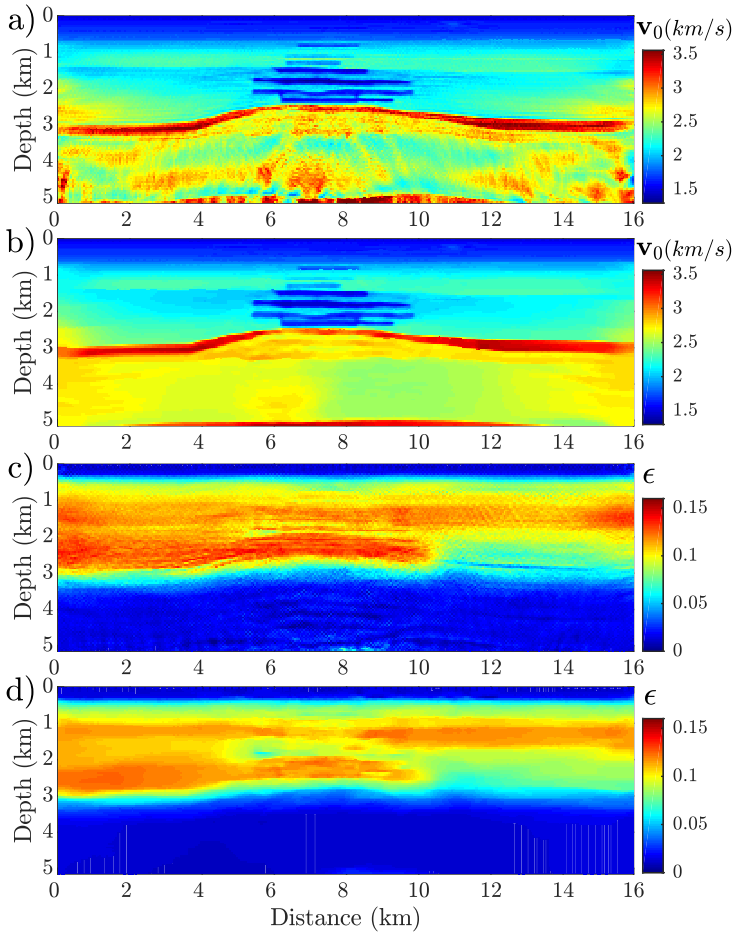

Figure 4.14 – North Sea case study with noiseless data. Multi-parameter IR-WRI results. (a-b) $\mathbf{v}_0$ models obtained with (a) DMP (a) and (b) DMP+TV regularizations. (c-d) Same as (a-b) for ϵ.

less data (compare Figs. 4.19a and 4.19c), while the data fit obtained with DMP+TV regularization is more consistent with noiseless and noisy data (compare Figs. 4.19b and 4.19d). This is consistent with the previous conclusions drawn from the mono-parameter inversion.

4.2.5 Discussion

We have extended the ADMM-based wavefield reconstruction inversion method (IR-WRI), originally developed for mono-parameter wavespeed reconstruction (Aghamiry et al., 2019c,b), to multi-parameter inversion in VTI acoustic media. We have first discussed which formulations of the VTI acoustic wave equation are bilinear in wavefield and subsurface parameters. First-order velocity stress form is often more convenient than the second-order counterpart to fulfill bilinearity, in particular if density (or buoyancy) is an optimization parameter. However, it may be not the most convenient one for frequency-domain wavefield reconstruction as the size of the linear system to be solved scales to the number of wavefield components. To bypass this

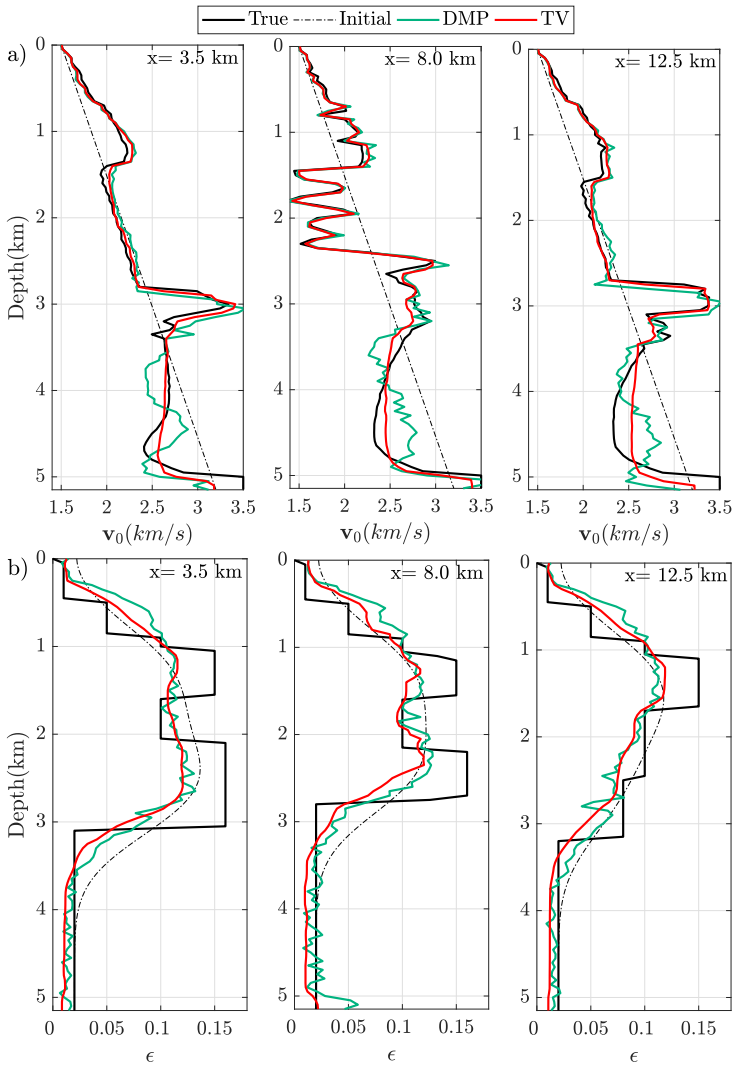

Figure 4.15 – North Sea case study with noiseless data. Multi-parameter bound-constrained IR-WRI results. Direct comparison along logs at $x = 3.5$ (left), $x = 8.0$ (center) and $x = 12.5$ km (right) between the true model (black), the initial model (dashed line) and the models inferred from IR-WRI with DMP (green) and DMP+TV regularization (red) for (a) $\mathbf{v}_0$ and (b) ϵ (Fig. 4.14).

issue, wavefield reconstruction and parameter estimation can be performed with different wave equations, provided they give consistent solutions (Gholami et al. (2013a) and this study).

Bilinearity allows us to recast the parameter estimation subproblem as a linear subproblem and hence the waveform inversion as a biconvex problem. Although ADMM has been originally developed to solve distributed convex problems, it can be used as is to solve biconvex problem (Boyd et al., 2010). From the mathematical viewpoint, biconvex problems should have superior convergence properties compared to fully nonconvex problem (this discussion is out of the scope of this study but we refer the interested reader to Benning et al., 2015). Alternatively, ADMM-like optimization can be used to perform IR-WRI heuristically without bilinear

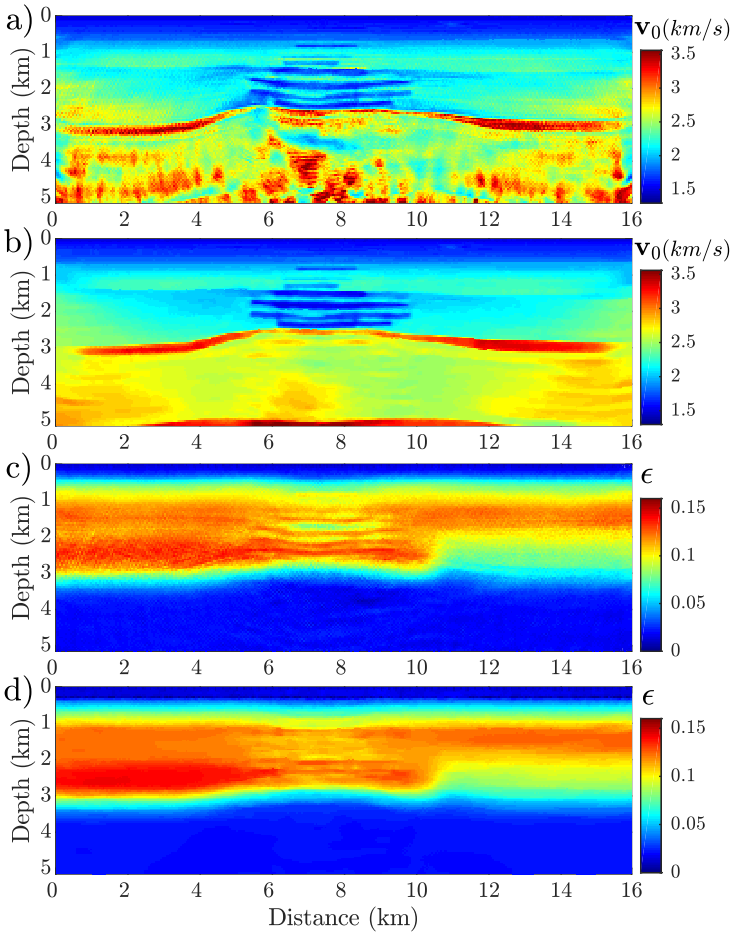

Figure 4.16 – North Sea case study with noisy data (SNR=10 db). Multi-parameter bound-constrained IR-WRI results. (a-b) $\mathbf{v}_0$ models obtained with (a) DMP (a) and (b) DMP+TV regularizations. (c-d) Same as (a-b) for ϵ.

wave equation, hence keeping the parameter estimation subproblem, equation 4.82a, nonlinear. Accordingly, equation 4.82a would be solved with a Newton algorithm rather than with a Gauss-Newton one. This nonlinear updating of the parameters may however require several inner Newton iterations per IR-WRI cycle, while Aghamiry et al. (2019c) showed that one inner Gauss-Newton iteration without any line search was providing the most efficient convergence of IR-WRI when bilinearity is fulfilled.

Indeed, the bilinearity specification limits the choice of subsurface parametrisation for parameter estimation. In the general case of triclinic elastodynamic equations, a subsurface parametrisation involving buoyancy and stiffness or compliance coefficients will be the most natural ones, as they correspond to the coefficients of the equation of motion and the Hooke's law. In the particular case of the VTI acoustic wave equation, we have developed a bilinear wave equation whose coefficients depend on the vertical wavespeed $\mathbf{v}_0$ and the Thomsen's parameters δ and ϵ. Although $\mathbf{v}_0$ and ϵ are coupled at wide scattering angles, the $(\mathbf{v}_0, \epsilon, \delta)$ parametrisation was promoted by Gholami et al. (2013b) and Gholami et al. (2013a) because the dominant parameter $\mathbf{v}_0$ has a radiation pattern which doesn't depend on the scattering angle, and hence can

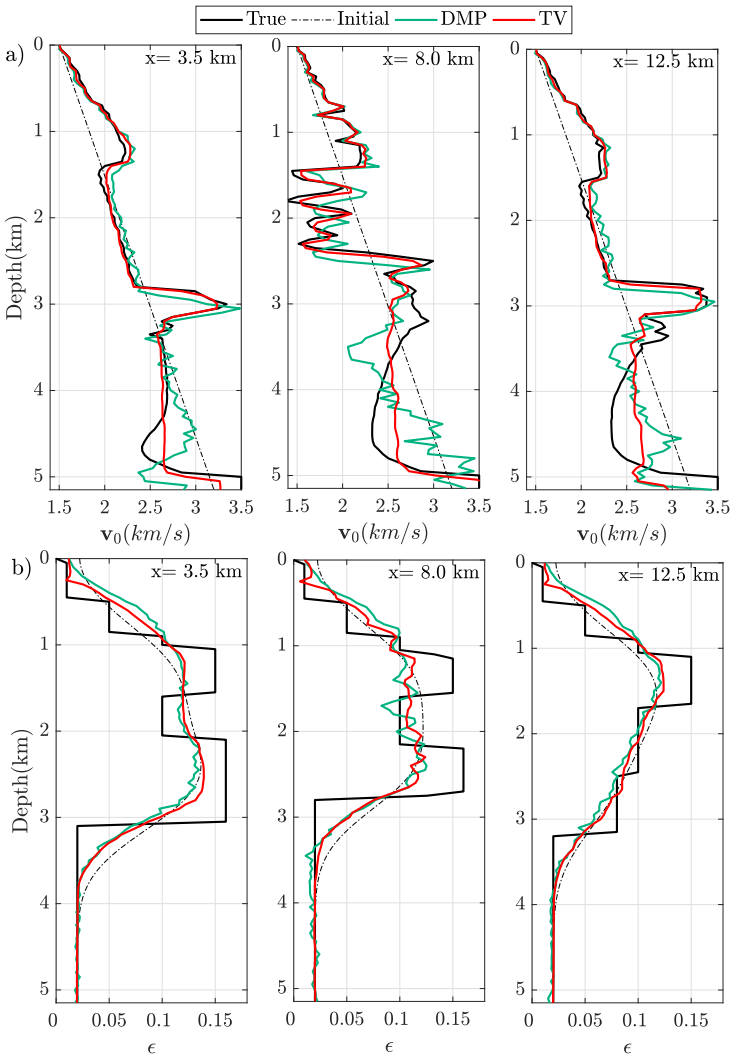

Figure 4.17 – North Sea case study with noisy data. Multi-parameter bound-constrained IR-WRI results. Direct comparison along logs at $x = 3.5$ (left), $x = 8.0$ (center) and $x = 12.5$ km (right) between the true model (black), the initial model (dashed line) and the models inferred from IR-WRI with DMP (green) and DMP+TV regularization (red) for (a) $\mathbf{v}_0$ and (b) ϵ (Fig. 4.16).

be reconstructed with a high resolution from wide-azimuth long-offset data. The counterpart is that updating the long wavelengths of the secondary parameter ϵ is challenging and requires so far a crude initial guess of its long wavelengths (Fig. 4.5e,f), which can be used as prior to regularize the ϵ update. Comparing the results of IR-WRI when ϵ is used as a passive parameter and as an optimization parameter shows that the sensitivity of the inversion to ϵ remains small provided that a reasonable guess of its long wavelengths are provided in the starting model (for the models, compare Figs. 4.8-4.10 and Figs. 4.14-4.16; for the data fit, compare Fig. 4.13 and Fig. 4.19). This limited sensitivity of FWI to the short-to-intermediate wavelengths of ϵ in the $(\mathbf{v}_0, \epsilon, \delta)$ prompted for example Debens et al. (2015) to estimate a crude ϵ background model with a coarse parametrisation by global optimization.

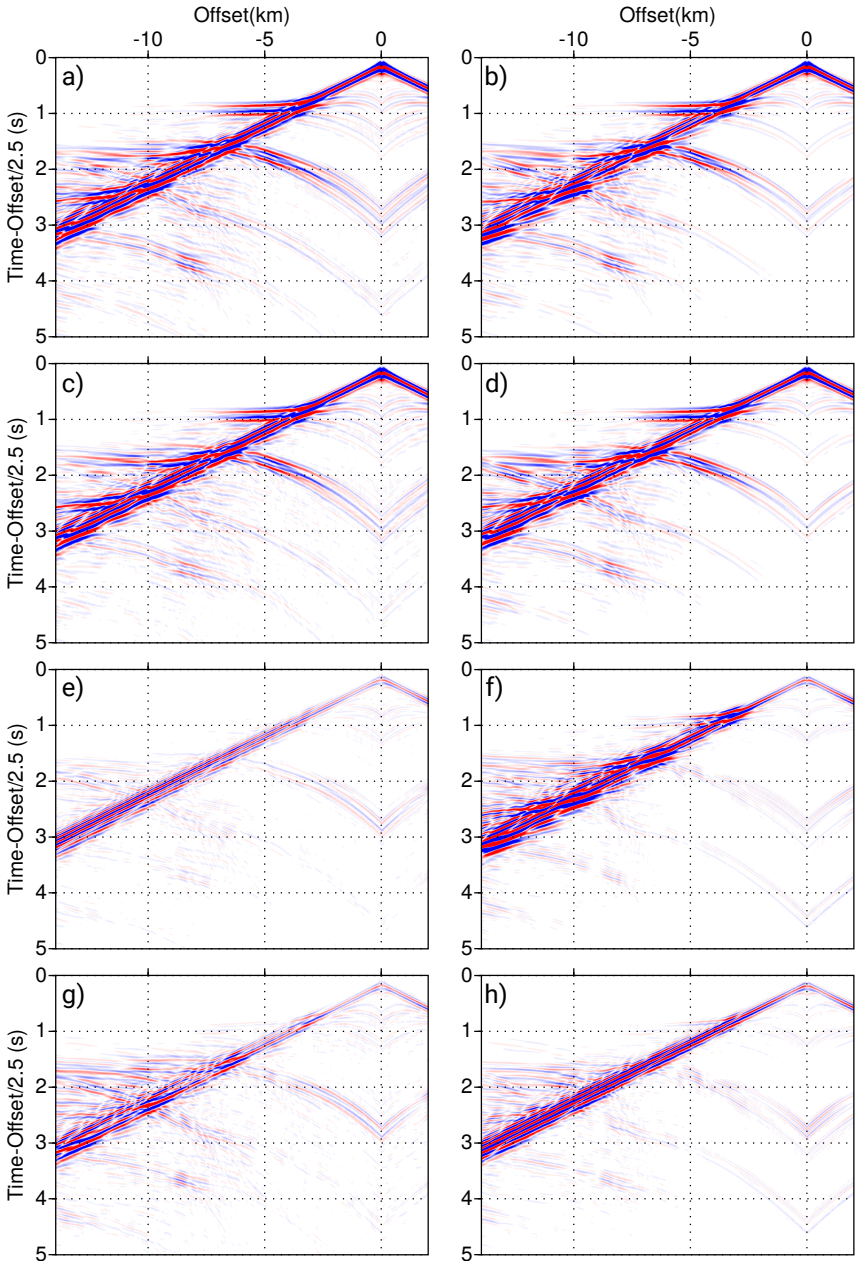

Figure 4.18 – North Sea case study. Same as Fig. 4.12 for (v_0,ϵ) multi-parameter IR-WRI.

Among the alternative parametrisations proposed for VTI acoustic FWI, Plessix and Cao (2011) proposed the $(\mathbf{v}_n, \mathbf{v}_h, \delta)$ or the $(\mathbf{v}_n, \eta, \delta)$ parametrisations for long-offset acquisition, while Alkhalifah and Plessix (2014) promoted the $(\mathbf{v}_h, \eta, \epsilon)$ parametrisation, where $\mathbf{v}_n = \mathbf{v}_0\sqrt{1+2\delta}$ is the so-called NMO velocity, $\mathbf{v}_h = \mathbf{v}_0\sqrt{1+2\epsilon}$ is the horizontal velocity and $\eta = (\epsilon-\delta)/(1+2\delta)$ represents the anellipticity of the anisotropy. For the parametrisation promoted by Plessix

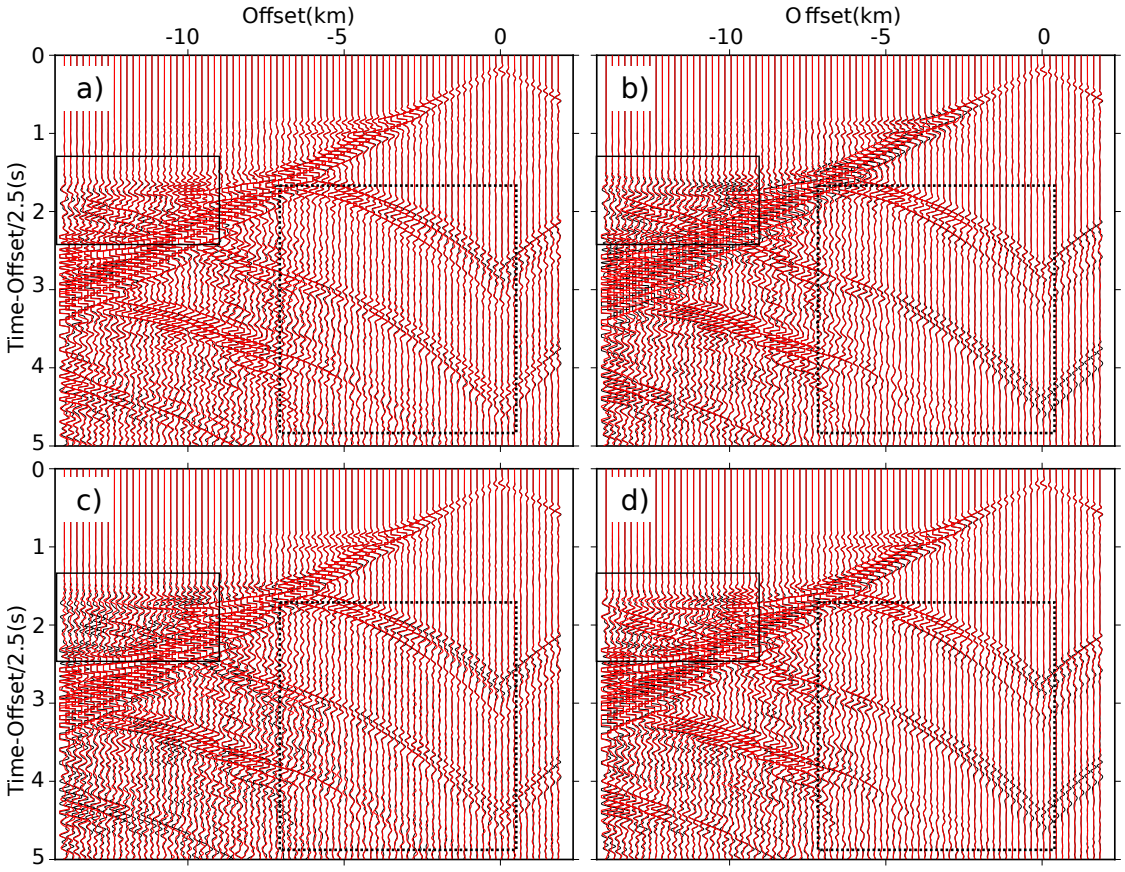

Figure 4.19 – North Sea case study. Same as Fig. 4.13 for (v_0,ϵ) multi-parameter IR-WRI.

and Cao (2011), the VTI equation developed by Zhou H. (2006) given by

$$\begin{aligned}
\omega^2 \text{diag}\left(\frac{1}{\mathbf{v}_n^2}\right)\mathbf{u}_q + 2\text{diag}(\eta)\nabla_{xx}(\mathbf{u}_p+\mathbf{u}_q) &= \mathbf{s}_q,\\
\omega^2 \text{diag}\left(\frac{1}{\mathbf{v}_n^2}\right)\mathbf{u}_p + \nabla_{xx}(\mathbf{u}_p+\mathbf{u}_q) &+\\
\text{diag}\left(\frac{1}{\sqrt{1+2\delta}}\right)\nabla_{zz}\text{diag}\left(\frac{1}{\sqrt{1+2\delta}}\right)\mathbf{u}_p &= \mathbf{s}_p,
\end{aligned} \tag{4.70}$$

is bilinear in wavefields and parameters $(1/\mathbf{v}_n^2, \eta)$, where $\mathbf{u}_p = \sqrt{1+2\delta}\mathbf{u}_z$ and $\mathbf{u}_q = \mathbf{u}_x - \sqrt{1+2\delta}\mathbf{u}_z$. Note that if δ is assumed to be smooth, the above equation can be approximated as

$$\begin{aligned}
\omega^2 \text{diag}\left(\frac{1}{\mathbf{v}_n^2}\right)\mathbf{u}_q + 2\text{diag}(\eta)\nabla_{xx}(\mathbf{u}_p+\mathbf{u}_q) &= \mathbf{s}_q,\\
\omega^2 \text{diag}\left(\frac{1}{\mathbf{v}_n^2}\right)\mathbf{u}_p + \nabla_{xx}(\mathbf{u}_p+\mathbf{u}_q) &+\\
\text{diag}\left(\frac{1}{1+2\delta}\right)\nabla_{zz}\mathbf{u}_p &= \mathbf{s}_p
\end{aligned} \tag{4.71}$$

which is bilinear in wavefields and parameters $(1/\mathbf{v}_n^2, \eta, 1/(1+2\delta)$. This implies that δ can be involved as an optimization parameter if necessary. Note that, if this smoothness approximation is used to update the parameters during the primal problem, the modelled data and the source residuals can be solved with the exact equation to update the dual variables with more accuracy. For the $(\mathbf{v}_h, \eta, \epsilon)$ parametrisations (Alkhalifah and Plessix, 2014), according to Zhou H. (2006) the VTI equations with smooth δ can be written as

$$\begin{aligned} \omega^2 \text{diag}\left(\frac{1}{\mathbf{v}_h^2}\right)\mathbf{u}_q + \nabla_{xx}(\mathbf{u}_p+\mathbf{u}_q) - \text{diag}\left(\frac{1}{1+2\eta}\right) & \\ \nabla_{xx}(\mathbf{u}_p+\mathbf{u}_q) &= \mathbf{s}_q, \\ \omega^2 \text{diag}\left(\frac{1}{\mathbf{v}_h^2}\right)\mathbf{u}_p + \text{diag}\left(\frac{1}{1+2\eta}\right)\nabla_{xx}(\mathbf{u}_p+\mathbf{u}_q) &+ \\ \text{diag}\left(\frac{1}{1+2\epsilon}\right)\nabla_{zz}\mathbf{u}_p &= \mathbf{s}_p, \end{aligned} \tag{4.72}$$

which is bilinear in wavefields and parameters $(1/\mathbf{v}_h^2, 1/(1+2\eta), 1/(1+2\epsilon))$.

4.2.6 Conclusion

We have shown that ADMM-based IR-WRI can be extended to multi-parameter reconstruction for VTI acoustic media. The gradient of the misfit function for the parameter estimation subproblem involves the so-called virtual sources, which carry out the effect of the parameter-dependent radiation patterns. This suggests that, although IR-WRI extends the search-space of FWI to mitigate cycle skipping, it is impacted by ill-posedness associated with parameter cross-talk and incomplete angular illumination as classical FWI. We have verified this statement with a toy numerical example for which we have reproduced the same pathologies in terms of resolution and parameter cross-talks as those produced by classical FWI during a former study. We have illustrated how equipping IR-WRI with bound constraints and TV regularization fully remove the ill-posedness effects for this idealized numerical example. We have provided some guidelines to design bilinear wave equation of different order for different subsurface parametrisations. Although bilinearity puts some limitations on the choice of the subsurface parametrisation, it recasts the parameter estimation subproblem as a quadratic optimization problem, which can be solved efficiently with Gauss-Newton algorithm. Application on the long-offset synthetic case study representative of the North Sea has shown how IR-WRI can be started from a crude laterally-homogeneous vertical velocity model without impacting the inversion with cycle skipping, when a smooth δ parameter is used as a passive background model. However, a smooth initial ϵ model, albeit quite crude, is necessary to guarantee the convergence of the method to a good solution, either when ϵ is used as a passive or as an optimization parameter. For this case study where the low velocity gas layers and the smooth medium below the reservoir suffer from a deficit illumination of diving waves, the TV regularization plays a key role to mitigate the ill-posedness.

Acknowledgments

We would like to thank the editor H. Yao, X. Cheng and D. Köhn for their comments. This study was partially funded by the SEISCOPE consortium (*http://seiscope2.osug.fr*), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXON-MOBIL, JGI, PETROBRAS, SCHLUMBERGER, SHELL, SINOPEC and TOTAL. This study was granted access to the HPC resources of SIGAMM infrastructure (http://crimson.oca.eu), hosted by Observatoire de la Côte d'Azur and which is supported by the Provence-Alpes Côte d'Azur region, and the HPC resources of CINES/IDRIS/TGCC under the allocation 0596 made by GENCI.

4.2.7 Appendix A: First and second-order wave equation with compliance notation

We consider the frequency-domain first-order velocity-stress equation in 2D VTI acoustic media with compliance notation as

$$\begin{aligned}
-\hat{i}\omega\mathbf{v}_{x,l} &= \text{diag}(\mathbf{b})\nabla_x\mathbf{u}_{x,l},\\
-\hat{i}\omega\mathbf{v}_{z,l} &= \text{diag}(\mathbf{b})\nabla_z\mathbf{u}_{z,l},\\
-\hat{i}\omega[\text{diag}(\int_{11})\mathbf{u}_{x,l} + \text{diag}(\int_{13})\mathbf{u}_{z,l}] &= \nabla_x\mathbf{v}_{x,l} - \hat{i}\omega\mathbf{s}_l,\\
-\hat{i}\omega[\text{diag}(\int_{13})\mathbf{u}_{x,l} + \text{diag}(\int_{33})\mathbf{u}_{z,l}] &= \nabla_z\mathbf{v}_{z,l} - \hat{i}\omega\mathbf{s}_l,
\end{aligned} \tag{4.73}$$

where $\int_{ij} \in \mathbb{R}^{n\times 1}$ are the compliance coefficients, and the other notations are defined after equation 4.48. Gathering equation 4.73 for all sources results in the following matrix equation:

$$\begin{bmatrix} \hat{i}\omega\mathbf{I} & \mathbf{B}\nabla \\ \nabla & \hat{i}\omega\mathcal{S} \end{bmatrix} \begin{bmatrix} \mathbf{V} \\ \mathbf{U} \end{bmatrix} = \hat{i}\omega \begin{bmatrix} \mathbf{0} \\ \mathbf{S} \end{bmatrix}, \tag{4.74}$$

where

$$\mathcal{S} = \begin{bmatrix} \text{diag}(\int_{11}) & \text{diag}(\int_{13}) \\ \text{diag}(\int_{13}) & \text{diag}(\int_{33}) \end{bmatrix},$$

and the other notations are defined in equation 4.49. Equation 4.74 is linear in $\mathbf{U}$ and $\mathbf{V}$ when the model parameters embedded in $\mathbf{B}$ and $\mathcal{S}$ are known. When $\mathbf{U}$ and $\mathbf{V}$ are known, this system can be recast as a new linear system in which the unknowns are the model parameters, and hence the bilinearity of the wave equation. For the lth source, the new equation becomes

$$\begin{bmatrix} \mathbf{L}_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{L}_{21} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{32} & \mathbf{L}_{33} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{L}_{43} & \mathbf{L}_{44} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \int_{11} \\ \int_{13} \\ \int_{33} \end{bmatrix} = \begin{bmatrix} \mathbf{0} - \hat{i}\omega\mathbf{v}_{x,l} \\ \mathbf{0} - \hat{i}\omega\mathbf{v}_{z,l} \\ \hat{i}\omega\mathbf{s}_l - \nabla_x\mathbf{v}_{x,l} \\ \hat{i}\omega\mathbf{s}_l - \nabla_z\mathbf{v}_{z,l} \end{bmatrix}, \tag{4.75}$$

where

$$\begin{cases} \mathbf{L}_{11} = \text{diag}(\nabla_x \mathbf{u}_{x,l}), \\ \mathbf{L}_{21} = \text{diag}(\nabla_z \mathbf{u}_{z,l}), \\ \mathbf{L}_{32} = \mathbf{L}_{43} = \hat{i}\omega \text{diag}(\mathbf{u}_{x,l}), \\ \mathbf{L}_{33} = \mathbf{L}_{44} = \hat{i}\omega \text{diag}(\mathbf{u}_{z,l}). \end{cases}$$

To develop the second-order wave equation, we eliminate $\mathbf{v}_{x,l}$ and $\mathbf{v}_{z,l}$ from equation 4.73. We obtain the following equation

$$\begin{aligned} &\frac{1}{\omega^2}\nabla_x \text{diag}(\mathbf{b})\nabla_x \mathbf{u}_{x,l} + \text{diag}(\textstyle\int_{11})\mathbf{u}_{x,l} + \text{diag}(\textstyle\int_{13})\mathbf{u}_{z,l} = \mathbf{s}_l, \\ &\frac{1}{\omega^2}\nabla_z \text{diag}(\mathbf{b})\nabla_z \mathbf{u}_{z,l} + \text{diag}(\textstyle\int_{13})\mathbf{u}_{x,l} + \text{diag}(\textstyle\int_{33})\mathbf{u}_{z,l} = \mathbf{s}_l, \end{aligned} \tag{4.76}$$

which is bilinear with respect to buoyancy, compliance parameters and pressure wavefields. With known buoyancy and compliance parameters, we get the following $2n \times 2n$ linear system to estimate wavefields for all sources

$$\mathbf{A}(\mathbf{m})\mathbf{U} = \mathbf{S}, \tag{4.77}$$

where $\mathbf{A}$ is given by

$$\begin{bmatrix} \frac{1}{\omega^2}\nabla_x \text{diag}(\mathbf{b})\nabla_x + \text{diag}(\textstyle\int_{11}) & \text{diag}(\textstyle\int_{13}) \\ \text{diag}(\textstyle\int_{13}) & \frac{1}{\omega^2}\nabla_z \text{diag}(\mathbf{b})\nabla_z + \text{diag}(\textstyle\int_{33}) \end{bmatrix}.$$

and

$$\mathbf{m} = \begin{bmatrix} \mathbf{b} \\ \textstyle\int_{11} \\ \textstyle\int_{13} \\ \textstyle\int_{33} \end{bmatrix},$$

and $\mathbf{S}$ is defined in equation 4.50. When $\mathbf{U}$ is known, this system can also be recast as a new linear system in which the unknowns are the model parameters as

$$\begin{bmatrix} \mathbf{L}_1 \\ \vdots \\ \mathbf{L}_l \\ \vdots \\ \mathbf{L}_{n_s} \end{bmatrix} \mathbf{m} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_l \\ \vdots \\ \mathbf{y}_{n_s} \end{bmatrix}, \tag{4.78}$$

where $\mathbf{L}_l$ is given by

$$\begin{bmatrix} \frac{1}{\omega^2}\nabla_x \text{diag}(\nabla_x \mathbf{u}_{x,l}) & \text{diag}(\mathbf{u}_{x,l}) & \text{diag}(\mathbf{u}_{z,l}) & \mathbf{0} \\ \frac{1}{\omega^2}\nabla_z \text{diag}(\nabla_z \mathbf{u}_{z,l}) & \mathbf{0} & \text{diag}(\mathbf{u}_{x,l}) & \text{diag}(\mathbf{u}_{z,l}) \end{bmatrix}$$

and

$$\mathbf{y}_l = \begin{bmatrix} \mathbf{s}_l \\ \mathbf{s}_l \end{bmatrix}.$$

4.2.8 Appendix B: Solving the optimization problem, equation 4.58, with ADMM

Starting from an initial model $\mathbf{m}^0$, and setting the dual variables $\tilde{\mathbf{D}}$ and $\tilde{\mathbf{S}}$ equal to zero, the kth ADMM iteration for solving equation 4.58 reads as (see Aghamiry et al., 2019c,b,e, for more details)

$$\mathbf{U}^{k+1} \leftarrow \arg\min_{\mathbf{U}} \; C_{\mathbf{m}^k,\tilde{\mathbf{D}}^k,\tilde{\mathbf{S}}^k}(\mathbf{U}), \tag{4.79a}$$

$$\mathbf{m}^{k+1} \leftarrow \arg\min_{\mathbf{m}\in\mathcal{C}} \; C_{\mathbf{U}^{k+1},\tilde{\mathbf{S}}^k}(\mathbf{m}), \tag{4.79b}$$

$$\tilde{\mathbf{S}}^{k+1} \leftarrow \tilde{\mathbf{S}}^k + \mathbf{S} - \mathbf{A}(\mathbf{m}^{k+1})\mathbf{U}^{k+1}, \tag{4.79c}$$

$$\tilde{\mathbf{D}}^{k+1} \leftarrow \tilde{\mathbf{D}}^k + \mathbf{D} - \mathbf{P}\mathbf{U}^{k+1}, \tag{4.79d}$$

where

$$C_{\mathbf{m}^k,\tilde{\mathbf{D}}^k,\tilde{\mathbf{S}}^k}(\mathbf{U}) = \lambda_0\|\mathbf{P}\mathbf{U} - \mathbf{D} - \tilde{\mathbf{D}}^k\|_F^2 + \lambda_1\|\mathbf{A}(\mathbf{m}^k)\mathbf{U} - \mathbf{S} - \tilde{\mathbf{S}}^k\|_F^2 \tag{4.80}$$

and

$$C_{\mathbf{U}^{k+1},\tilde{\mathbf{S}}^k}(\mathbf{m}) = \sum\sqrt{|\partial_x\mathbf{m}|^2 + |\partial_z\mathbf{m}|^2} + \lambda_1\|\mathbf{A}(\mathbf{m})\mathbf{U}^{k+1} - \mathbf{S} - \tilde{\mathbf{S}}^k\|_F^2. \tag{4.81}$$

The regularized wavefields $\mathbf{U}$ are the minimizers of the quadratic cost function $C_{\mathbf{m}^k,\tilde{\mathbf{D}}^k,\tilde{\mathbf{S}}^k}(\mathbf{U})$, equation 4.80, where $\mathbf{m}^k$, $\tilde{\mathbf{D}}^k$ and $\tilde{\mathbf{S}}^k$ are kept fixed. Zeroing the derivative of $C_{\mathbf{m}^k,\tilde{\mathbf{D}}^k,\tilde{\mathbf{S}}^k}(\mathbf{U})$ gives the wavefields as the solution of a linear system of equations defined by equation 4.59 (step 1 of the algorithm). The regularized wavefields are then introduced as passive quantities in the cost function $C_{\mathbf{U}^{k+1},\tilde{\mathbf{S}}^k}(\mathbf{m})$, equation 4.81, which is minimized to estimate $\mathbf{m}$ over the desired set $\mathcal{C}$. We solve this minimization subproblem with the splitting techniques. Accordingly, we introduce the auxiliary primal variables $\mathbf{p}$ and $\mathbf{q}$ to decouple the $\ell 1$ and $\ell 2$ terms and split the parameter estimation subproblem into three sub-steps for $\mathbf{m}$, $\mathbf{p}$ and $\mathbf{q}$ (Goldstein and Osher, 2009; Aghamiry et al., 2019b):

$$\mathbf{m}^{k+1} \leftarrow \arg\min_{\mathbf{m}} \; \lambda_1\|\mathbf{A}(\mathbf{m})\mathbf{U}^{k+1} - \mathbf{S} - \tilde{\mathbf{S}}^k\|_F^2, + \|\overline{\nabla}\mathbf{m} - \mathbf{p}^k - \tilde{\mathbf{p}}^k\|_{\mathbf{\Gamma}}^2 + \|\mathbf{m} - \mathbf{q}^k - \tilde{\mathbf{q}}^k\|_Z^2, \tag{4.82a}$$

$$\mathbf{p}^{k+1} \leftarrow \arg\min_{\mathbf{p}} \; \sum\sqrt{|\mathbf{p}_x|^2 + |\mathbf{p}_z|^2} + \|\overline{\nabla}\mathbf{m}^{k+1} - \mathbf{p} - \tilde{\mathbf{p}}^k\|_{\mathbf{\Gamma}}^2, \tag{4.82b}$$

$$\mathbf{q}^{k+1} \leftarrow \arg\min_{\mathbf{q}\in\mathcal{C}} \; \|\mathbf{m}^{k+1} - \mathbf{q} - \tilde{\mathbf{q}}^k\|_Z^2, \tag{4.82c}$$

where $\overline{\nabla}$ is defined in equation 4.61,

$$\mathbf{p} = \begin{bmatrix} \mathbf{p}_x \\ \mathbf{p}_z \end{bmatrix}, \tag{4.83}$$

$\|\mathbf{x}\|^2_{\bullet} = \mathbf{x}^T \bullet \mathbf{x}$, Z and $\mathbf{\Gamma}$ are diagonal matrices defined in equation 4.62 and 4.63, respectively.

From the linearized equation 4.55, the subproblem for $\mathbf{m}$, equation 4.82a, can be written as

$$\mathbf{m}^{k+1} \leftarrow \arg\min_{\mathbf{m}} \left\| \begin{bmatrix} \mathbf{L}_1 \\ \vdots \\ \mathbf{L}_{n_s} \\ \overline{\nabla} \\ \mathbf{I} \end{bmatrix} \mathbf{m} - \begin{bmatrix} \mathbf{y}_1^k + \tilde{\mathbf{s}}_1^k \\ \vdots \\ \mathbf{y}_{n_s}^k + \tilde{\mathbf{s}}_{n_s}^k \\ \mathbf{p}^k + \tilde{\mathbf{p}}^k \\ \mathbf{q}^k + \tilde{\mathbf{q}}^k \end{bmatrix} \right\|^2_{\mathbf{\Xi}}, \tag{4.84}$$

where the diagonal weighting matrix $\mathbf{\Xi}$ is defined as

$$\mathbf{\Xi} = \begin{bmatrix} \lambda_1 \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Gamma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & Z \end{bmatrix} \in \mathbb{R}^{[3+n_s]n \times [3+n_s]n}.$$

Equation 4.84 is now quadratic and admits a closed form solution as given in equation 4.60 (step 2 of the algorithm).

The only remaining tasks consist in determining the auxiliary primal variables $(\mathbf{p}, \mathbf{q})$, equations 4.82b,c, and the auxiliary dual variables $(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})$. They are initialized to 0 and are updated as follows: The primal variables $\mathbf{p}$ is updated through a TV proximity operator, which admits a closed form solution given by equation 4.64 (see Combettes and Pesquet, 2011) (step 3 of the algorithm). The primal variable $\mathbf{q}$ is updated by a projection operator, which also admits a closed form solution given by equation 4.66 (step 4 of the algorithm). Finally, the duals are updated according to gradient ascend steps (step 5 of the algorithm)

$$\begin{cases} \tilde{\mathbf{p}}^{k+1} &= \tilde{\mathbf{p}}^k + \mathbf{p}^{k+1} - \overline{\nabla}\mathbf{m}^{k+1} \\ \tilde{\mathbf{q}}^{k+1} &= \tilde{\mathbf{q}}^k + \mathbf{q}^{k+1} - \mathbf{m}^{k+1}. \end{cases} \tag{4.85}$$

4.2.9 Appendix C: Using fourth-order equation for wavefield reconstruction (Step 1 of the algorithm)

The wavefield reconstruction subproblem, equation 4.59, can be written as the following over-determined system

$$\begin{bmatrix} \omega^2 \text{diag}(\mathbf{m}_{v_0}) + \text{diag}(\mathbf{m}_\epsilon)\nabla_{xx} & \text{diag}(\mathbf{m}_\delta)\nabla_{zz} \\ \text{diag}(\mathbf{m}_\delta)\nabla_{xx} & \omega^2 \text{diag}(\mathbf{m}_{v_0}) + \nabla_{zz} \\ \frac{1}{2}\tilde{\mathbf{P}} & \frac{1}{2}\tilde{\mathbf{P}} \end{bmatrix} \begin{bmatrix} \mathbf{U}_x \\ \mathbf{U}_z \end{bmatrix} = \begin{bmatrix} \mathbf{S}_x \\ \mathbf{S}_z \\ \mathbf{D} \end{bmatrix}, \tag{4.86}$$

where $\tilde{\mathbf{P}}$ is the sampling operator of each component of the wavefield at receiver positions and $\mathbf{P} = [\frac{1}{2}\tilde{\mathbf{P}} \;\; \frac{1}{2}\tilde{\mathbf{P}}]$ (the coefficient $\frac{1}{2}$ results because the (isotropic) pressure wavefield $\mathbf{u}$ recorded

in the water is given by $\mathbf{u} = \frac{1}{2}(\mathbf{u}_x + \mathbf{u}_z)$). We can eliminate $\mathbf{U}_z$ from the equation 4.86 to develop a fourth-order partial-differential equation for $\mathbf{U}_x$ and then update $\mathbf{U}_z$ from its explicit expression as a function of $\mathbf{U}_x$ without any computational burden.
By multiplying the second row of equation 4.86 by $\text{diag}(\mathbf{m}_\delta)$ and taking the difference with the first row, we find

$$\begin{bmatrix} \omega^2\text{diag}(\mathbf{m}_{v_0}) + \text{diag}(\mathbf{m}_\epsilon)\nabla_{xx} & \text{diag}(\mathbf{m}_\delta)\nabla_{zz} \\ \omega^2\text{diag}(\mathbf{m}_{v_0}) + \text{diag}(\mathbf{m}_\epsilon - \mathbf{m}_\delta^2)\nabla_{xx} & -\omega^2\text{diag}(\mathbf{m}_\delta \circ \mathbf{m}_{v_0}) \\ \frac{1}{2}\tilde{\mathbf{P}} & \frac{1}{2}\tilde{\mathbf{P}} \end{bmatrix} \begin{bmatrix} \mathbf{U}_x \\ \mathbf{U}_z \end{bmatrix} = \begin{bmatrix} \mathbf{S}_x \\ \mathbf{S}_x - \text{diag}(\mathbf{m}_\delta)\mathbf{S}_z \\ \mathbf{D} \end{bmatrix}. \tag{4.87}$$

where $\mathbf{m}_\delta^2 = \mathbf{m}_\delta \circ \mathbf{m}_\delta$. The second equation of 4.87 provides us the closed-form expression of $\mathbf{U}_z$ as a function of $\mathbf{U}_x$

$$\mathbf{U}_z = \mathbf{A}_z\mathbf{U}_x + \mathbf{B}_z, \tag{4.88}$$

where

$$\begin{cases} \mathbf{A}_z = \frac{\text{diag}(\mathbf{m}_\epsilon - \mathbf{m}_\delta^2)\nabla_{xx} + \omega^2\text{diag}(\mathbf{m}_{v_0})}{\omega^2\text{diag}(\mathbf{m}_{v_0} \circ \mathbf{m}_\delta)}, \\ \mathbf{B}_z = \frac{\text{diag}(\mathbf{m}_\delta)\mathbf{S}_z - \mathbf{S}_x}{\omega^2\text{diag}(\mathbf{m}_{v_0} \circ \mathbf{m}_\delta)}. \end{cases} \tag{4.89}$$

Injecting equation 4.89 into the first equation of 4.87 leads to the over-determined system satisfied by $\mathbf{U}_x$

$$\begin{bmatrix} \omega^2\text{diag}(\mathbf{m}_{v_0}) + \text{diag}(\mathbf{m}_\epsilon)\nabla_{xx} + \text{diag}(\mathbf{m}_\delta)\nabla_{zz}\mathbf{A}_z \\ \frac{1}{2}\tilde{\mathbf{P}} + \frac{1}{2}\tilde{\mathbf{P}}\mathbf{A}_z \end{bmatrix} \mathbf{U}_x = \begin{bmatrix} \mathbf{S}_x - \text{diag}(\mathbf{m}_\delta)\nabla_{zz}\mathbf{B}_z \\ \mathbf{D} - \frac{1}{2}\tilde{\mathbf{P}}\mathbf{B}_z \end{bmatrix}. \tag{4.90}$$

So, instead of solving the $(n_r + 2n) \times 2n$ linear system 4.86 to update $\mathbf{U}$, $\mathbf{U}_x$ is updated by solving the $(n_r + n) \times n$ linear system 4.90 and then $\mathbf{U}_z$ is updated using 4.89 without significant computational overhead. The wave equation operator in equation 4.89 has been broken down into an elliptic wave equation operator and anelliptic correction term (the term of $\mathbf{A}_z$ related to $\text{diag}(\mathbf{m}_\epsilon - \mathbf{m}_\delta^2)$, equation 4.89). The former can be accurately discretized with the 9-point finite-difference method of Chen et al. (2013), while the anelliptic term can be discretized with a basic second-order accurate 5-point stencil without generating significant inaccuracies in the modelling (Operto et al., 2014).

4.3 Application of ADMM-based WRI to attenuation

This section presents our paper dealing with the extension of ADMM-based WRI to viscoacoustic media. Viscoacoustic effects are easily included in the time-harmonic wave equation with frequency-dependent complex-valued velocities as function of phase velocity and attenuation factor (the inverse of quality factor) which are both real-valued parameters (e.g., Toksöz and Johnston, 1981; Carcione, 2015). Classical approaches process the real and imaginary parts of the complex-valued wavespeed as two independent real-valued parameters during the inversion. In this case, the Hessian of viscoacoustic IR-WRI can be singular (Ribodetti et al., 2000; Mulder and Hak, 2009; Hak and Mulder, 2011). Alternatively, it is possible to process the velocity as a complex parameter in frequency-domain visco-acoustic inversion to recast the optimization as a mono-parameter problem. In that case, one can show that the mono-parameter Hessian of viscoacoustic IR-WRI is not singular. This complex-valued inverse problem is an ongoing work in our team (Aghamiry et al., 2019d). In this thesis, the phase velocity and the attenuation parameters are updated in an alternating mode with ADMM-based WRI, taking advantage that the wave equation can be recast as a multi-linear function of the three parameter classes (wavefield, squared-slowness and attenuation factor). As attenuation model, I use a first-order approximation of the complex-valued Kolsky-Futterman model. I regularize the multi-variate inversion with the bound constraints and TV regularization, which are taylored to the squared slowness and attenuation factor.

Multi-parameter wavefield reconstruction inversion for wavespeed and attenuation with bound constraints and total variation regularization
Hossein S. Aghamiry, Ali Gholami and Stéphane Operto
Geophysics, 2020, *DOI: 10.1190/geo2019-0596.1*

4.3.1 Summary

Wavefield reconstruction inversion (WRI) extends the search space of Full Waveform Inversion (FWI) by allowing for wave equation errors during wavefield reconstruction to match the data from the first iteration. Then, the wavespeeds are updated from the wavefields by minimizing the source residuals. Performing these two tasks in alternating mode breaks down the nonlinear FWI as a sequence of two linear subproblems, relaying on the bilinearity of the wave equation. We solve this biconvex optimization with the alternating-direction method of multipliers (ADMM) to cancel out efficiently the data and source residuals in iterations and stabilize the parameter estimation with appropriate regularizations. Here, we extend WRI to viscoacoustic media for attenuation imaging. Attenuation reconstruction is challenging because of the small imprint of attenuation in the data and the cross-talks with velocities. To address these issues, we recast the multivariate viscoacoustic WRI as a triconvex optimization and update wavefields, squared slowness, and attenuation factor in alternating mode at each WRI iteration. This requires to linearize the attenuation-estimation subproblem via an approximated trilinear viscoacoustic wave equation. The iterative defect correction embedded in ADMM corrects the

errors generated by this linearization, while the operator splitting allows us to tailor $\ell 1$ regularization to each parameter class. A toy numerical example shows that these strategies mitigate cross-talk artifacts and noise from the attenuation reconstruction. A more realistic synthetic example representative of the North Sea validates the method.

4.3.2 Introduction

Full waveform inversion (FWI) is a high-resolution nonlinear imaging technology which can provide an accurate subsurface model by matching observed and calculated waveforms (Tarantola, 1984; Pratt et al., 1998; Virieux and Operto, 2009). However, it is well acknowledged that it suffers from two main pathologies. The first one is the nonlinearity associated with cycle skipping: when the distance between the observed and calculated data is the least-squares norm of their differences, FWI remains stuck into spurious local minima when the initial velocity model does not allow to match traveltimes with an error lower than half a period. To mitigate cycle skipping, many variants of FWI have been proposed with more convex distances such as those based on matching filters (Warner and Guasch, 2016; Guasch et al., 2019) or optimal transport (Métivier et al., 2018) among others. The second pathology is ill-posedness resulting from uneven subsurface illumination provided by limited-aperture surface acquisitions (e.g., Tang, 2009) and parameter crosstalks during multiparameter reconstruction (see Operto et al., 2013, for a tutorial). Mitigating this ill-posedness requires accounting for the Hessian in local optimization methods (e.g. Métivier et al., 2017) and regularizing the inversion with prior information such as physical bound constraints (e.g. Asnaashari et al., 2013; Duan and Sava, 2016). Among the methods proposed to mitigate cycle skipping, wavefield reconstruction inversion (WRI) (van Leeuwen and Herrmann, 2013, 2016) extends the parameter search space of frequency-domain FWI by processing the wave-equation as a soft constraint with a penalty method. The resulting wave equation relaxation allows for data fitting with inaccurate velocity models through the reconstruction of data-assimilated wavefields, namely wavefields satisfying the observation equation relating the wavefields to the observations (Aghamiry et al., 2020a). The algorithm then updates the model parameters by least-squares minimization of the wave equation errors (or source residuals) so that the assimilated wavefields explain both the wave equation and the data as well as possible. Performing wavefield reconstruction and parameter estimation in an alternating mode (van Leeuwen and Herrmann, 2013) rather than by variable projection (van Leeuwen and Herrmann, 2016) recasts WRI as a sequence of two linear subproblems as a result of the bilinearity of the wave equation in wavefield and squared slowness. The reader is also referred to Aghamiry et al. (2019a) for a more general discussion on the bilinearity of the elastic anisotropic wave equation. Aghamiry et al. (2019c) solved this biconvex problem with the alternating direction method of multipliers (ADMM) (Boyd et al., 2010). ADMM is an augmented Lagrangian method which makes use of operator splitting and alternating directions to solve convex separable multivariate constrained problems. The augmented Lagrangian function combines a penalty function and a Lagrangian function (Nocedal and Wright, 2006, Chapter 17). The penalty function relaxes the constraints during early iterations as in WRI, while the Lagrangian function progressively corrects the constraint violations via the action of the Lagrange multipliers. The leverage provided by the Lagrange multipliers guarantees to satisfy the constraints at the convergence point with constant penalty parameters (Aghamiry et al.,

2019c). Accordingly, Aghamiry et al. (2019c) called their approach iteratively refined WRI (IR-WRI). Alternatives to satisfy the constraints at the convergence point with penalty methods rely on multiplicative (da Silva and Yao, 2017) or discrepancy-based (Fu and Symes, 2017) approaches. Aghamiry et al. (2019b) implemented bounding constraints and total variation (TV) regularization (Rudin et al., 1992) in IR-WRI with the split Bregman method (Goldstein and Osher, 2009) to improve the imaging of large-contrast media, with however undesirable staircase imprints in smooth regions. To overcome this issue and capture both the blocky and smooth components of the subsurface, Aghamiry et al. (2020b) combine in IR-WRI Tikhonov and TV regularizations by infimal convolution.

The objective of this paper is to extend frequency-domain IR-WRI to viscoacoustic media for attenuation imaging. Attenuation reconstruction by FWI raises two potential issues. The first is related to the crosstalks between wavespeed and attenuation. The ambiguity between velocity and attenuation perturbation in least-squares migration has been emphasized by Mulder and Hak (2009). Many combinations of velocity and attenuation perturbations can fit equally well reflection amplitudes since they are basically related by a Hilbert transform. This ambiguity can be simply illustrated by the radiation pattern of velocity and attenuation perturbations, which have the same amplitude versus angle behavior and a 90° phase shift (Malinowski et al., 2011; da Siva et al., 2019). The conclusions of Mulder and Hak (2009) substantiate those of Ribodetti et al. (2000) who show that the Hessian of ray+Born least-squares migration of single-offset reflection data is singular if the reflector is not illuminated from above and beneath. On the other hand, Hak and Mulder (2011) show that wavespeed and attenuation can be decoupled during nonlinear waveform inversion of multi-offset/multi-frequency data provided that the causality term is properly implemented in the attenuation model. This conclusion has been further supported by several realistic synthetic experiments and real data case studies in marine and land environments, which manage to reconstruct trustworthy attenuation models (Hicks and Pratt, 2001; Askan et al., 2007; Kamei and Pratt, 2008; Malinowski et al., 2011; Takougang and Calvert, 2012; Prieux et al., 2013; Stopin et al., 2016; Operto and Miniussi, 2018; Lacasse et al., 2019). This decoupling between velocity and attenuation can be further argued on the basis of physical considerations. In the transmission regime of wave propagation, wavespeeds control the kinematic of wave propagation. This implies that FWI is dominantly driven toward wavespeed updating to match the traveltimes of the wide-aperture data (diving waves, post-critical reflections) and update the long wavelengths of the subsurface accordingly, while attenuation has a secondary role in matching amplitude and dispersion effects (e.g., see Operto and Miniussi, 2018, for an illustration). This weak imprint of the attenuation in the seismic response was illustrated by the sensitivity analysis carried out by Kurzmann et al. (2013) who concluded that a crude homogeneous background attenuation model might be enough to perform reliable FWI, while da Siva et al. (2019) proposed to reconstruct an under-parametrized attenuation model by semi-global FWI. When a high-resolution attenuation model is sought, the ill-posedness of the attenuation reconstruction may be managed with different recipes including data and model-driven inversions (joint versus sequential updates of the velocity and attenuation of selected subdatasets), parameter scaling, bound constraints and regularizations (e.g. Prieux et al., 2013; Operto et al., 2013).

In this context, the contribution of this study is two fold: first, we show how to implement velocity and attenuation reconstruction in frequency-domain viscoacoustic IR-WRI when equipped

with bound constraints and nonsmooth regularizations. Second, we discuss with numerical examples whether the alternating-direction algorithm driven by the need to expand the search space is suitable to manage ill-conditioned multi-parameter reconstruction. It is well acknowledged that viscous effects are easily included in the time-harmonic wave equation with frequency-dependent complex-valued velocities as a function of phase velocity and attenuation factor (the inverse of quality factor), which are both real-valued parameters (Toksöz and Johnston, 1981). Accordingly, the objective function of viscoacoustic IR-WRI requires to be minimized over a set of three-parameter classes (wavefield, squared slowness, attenuation factor). In this study, we consider the Kolsky-Futterman model as attenuation model (Kolsky, 1956; Futterman, 1962). With this model, the viscoacoustic wave equation is bilinear in wavefield and squared slowness, while it is nonlinear in attenuation factor. This prompts us to introduce a first-order approximation of the viscoacoustic function to form a trilinear viscoacoustic wave equation. This equation allows us to recast the multivariate viscoacoustic IR-WRI as a sequence of three linear subproblems for wavefields, squared slowness and attenuation factor estimation, which are solved in alternating mode following the block relaxation strategy of ADMM. Then, the errors generated by the approximated wave equation during attenuation estimation are corrected by the action of the Lagrange multipliers (dual variables), which are formed by the source residuals computed with the exact wave equation. Another application of the augmented Lagrangian method in AVO inversion is presented in Gholami et al. (2018), where the linearized Zoeppritz equations are used to simplify the primal problem, while the dual problem compensates the linearization-related errors by computing the residuals with the exact Zoeppritz equations. Also, the decomposition of the viscoacoustic IR-WRI into three linear subproblems provides a suitable framework to tailor $\ell 1$ regularizations to each parameter-estimation subproblem (Aghamiry et al., 2019b, 2020b).

The alternating update of the squared slowness and attenuation factor at each IR-WRI iteration is probably non-neutral on how the inversion manages the parameter crosstalks and the contrasted sensitivity of the data to each parameter class, as the multi-parameter inversion is broken down as a sequence of two mono-parameter inversions. This approach differs from those commonly used in the multi-parameter inversion. The most brute-force approach consists in the joint updating of the multiple parameter classes, with the issue of managing multi-parameter Hessian with suitable parameter scaling (e.g., Stopin et al., 2014; Métivier et al., 2015; Yang et al., 2016a). Other approaches rely on ad-hoc hierarchical data and model-driven inversion where the dominant parameter is updated during a first mono-parameter inversion, before involving the secondary parameter in a subsequent multi-parameter inversion (e.g., Prieux et al., 2013; Cheng et al., 2016). Another possible model-driven strategy consists of performing the joint updating of the multiple parameter classes during a first inversion, then reset the secondary parameters to their initial values and restart a multi-parameter inversion involving all the parameter classes (Yang et al., 2014).

In this study, we assess our approach against two synthetic experiments, a toy example and a more realistic well-documented synthetic example representative of the North Sea (Prieux et al., 2013). A comparison of our approach with those reviewed above remains however beyond the scope of this paper. One reason is that all the above approaches have been implemented in conventional FWI, which would remain stuck in a local minimum when starting from the crude initial model used in this study. This is to remind that IR-WRI provides a practical framework

to conciliate the search space expansion to mitigate cycle skipping and easy-to-design multi-parameter reconstruction via the alternating update of the multiple parameter classes.
This paper is organized as follows. In the method section, we first review the forward problem equation before going into the details of viscoacoustic IR-WRI: We first formulate the constrained optimization problem to be solved and recast it as a saddle point problem with an augmented Lagrangian function (Appendix A, section 4.3.7). Then, we review the solution of the three primal subproblems for wavefields, squared slowness and attenuation factor in the framework of ADMM, as well as the expression of the dual variables or Lagrange multipliers that capture the history of the solution refinement in iterations. The Appendix B (section 4.3.8) reviews in a general setting the split Bregman method to solve $\ell 1$-regularized convex problem. This recipe can be easily applied to the squared slowness and attenuation reconstruction subproblems. The final section presents two synthetic examples, which are performed without and with bound constraints and TV regularization in order to discriminate the role of the augmented-Lagrangian optimization from that of the priors. A toy example allows us to illustrate in a simple setting how well IR-WRI manages the parameter crosstalk and the ill-posedness of the attenuation reconstruction and how bound constraints and TV regularization remove the corresponding artifacts. A second synthetic example representative of the North Sea environment allows one to assess the method in a more realistic setting.

4.3.3 Notation

As we implement IR-WRI in the frequency domain, we use discrete matrix notations. Italics refer to scalar quantities, boldface lowercase letters refer to vectors, and boldface capital letters refer to matrices and tensors. By default, vectors are considered as column vectors. We use the superscript T to denote the adjoint of an operator. The ith element of $\mathbf{x}$ is denoted by $[\mathbf{x}]_i$ and its absolute value is returned by $|\,[\mathbf{x}]_i\,|$. For real-valued vectors $\mathbf{x}$ and $\mathbf{y}$ of length n, the dot product is defined by $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{y} = \sum_{i=1}^{n} [\mathbf{x}]_i\,[\mathbf{y}]_i$ and their Hadamard product, denoted by $\mathbf{x} \circ \mathbf{y}$, is another vector made up of their component-wise products, i.e. $[\mathbf{x} \circ \mathbf{y}]_i = [\mathbf{x}]_i\,[\mathbf{y}]_i$. By an abuse of notation, other arithmetic operations on vectors are performed component wise. For example, $[\mathbf{x}^2]_i = [\mathbf{x}]_i^2$ and $[\mathbf{x}/\mathbf{y}]_i = [\mathbf{x}]_i\,/\,[\mathbf{y}]_i$. Also, let $\rho(\sigma)$ be a function of σ. In discrete form, $[\boldsymbol{\rho}(\boldsymbol{\sigma})]_i = \rho([\boldsymbol{\sigma}]_i)$. The ℓ_1- and ℓ_2-norms of $\mathbf{x}$ are, respectively, defined by $\|\mathbf{x}\|_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_{i=1}^{n} [\mathbf{x}]_i^2}$ and $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |\,[\mathbf{x}]_i\,|$.

4.3.4 Theory

In this section, we first review the viscoacoustic wave equation in the frequency-space domain. Then, we use this wave equation to formulate the iteratively-refined wavefield reconstruction inversion (IR-WRI) for velocity and attenuation.

Forward problem

The viscoacoustic wave equation in the frequency-space domain is given by

$$\left(\Delta + \frac{\omega^2}{c(\mathbf{x})^2}\right) u(\mathbf{x},\omega) = b(\mathbf{x},\omega), \tag{4.91}$$

where Δ is the Laplacian operator, ω is the angular frequency, $\mathbf{x} = (x, z)$ denotes the position in the subsurface model and $b(\mathbf{x},\omega)$ and $u(\mathbf{x},\omega)$ are respectively the source term and the wavefield for frequency ω. In this paper we limit ourselves to the wave equation with a constant density (equal to 1), but the proposed method can be extended to variable density case. Viscoacoustic (attenuative) media can be described by complex-valued velocity $c(\mathbf{x})$. The velocity associated with the Kolsky-Futterman model is given by (Kolsky, 1956; Futterman, 1962)

$$\frac{1}{c(\mathbf{x})} = \frac{1}{v(\mathbf{x})}\left[1 - \frac{1}{\pi Q(\mathbf{x})}\log|\frac{\omega}{\omega_r}| + \iota\frac{\text{sign}(\omega)}{2Q(\mathbf{x})}\right], \tag{4.92}$$

where $v(\mathbf{x})$ denotes the phase velocity, $Q(\mathbf{x})$ the frequency-independent quality factor, both real-valued, and $\iota = \sqrt{-1}$. Also, sign$(\bullet)$ is the sign function that extracts the sign of a real number $\bullet$. The logarithmic term with reference frequency $\omega_r = 2\pi f_r$ implies causality (Aki and Richards, 2002; Hak and Mulder, 2011). In this study, same as Toverud and Ursin (2005), f_r is chosen to be 50 Hz .

Inverse problem

From now on, we use the aforementioned discrete matrix notations. We discretize the 2D partial-differential equation (PDE), equation 4.91, with a $N = N_x \times N_z$ grid points, where N_x and N_z are the number of points in the horizontal and vertical directions, respectively. We parametrize the inversion by squared slowness (inverse of squared velocity) and attenuation factor (inverse of quality factor) that are respectively denoted by column vectors $\mathbf{m}, \boldsymbol{\alpha} \in \mathbb{R}^N$. Accordingly, equation 4.92 in discrete form reads as

$$\frac{1}{\mathbf{c}^2} = \mathbf{m} \circ \boldsymbol{\rho}(\boldsymbol{\alpha}). \tag{4.93}$$

In equation 4.93, $\boldsymbol{\rho}$ is given by

$$\boldsymbol{\rho}(\boldsymbol{\alpha}) = (1 + \beta(\omega)\boldsymbol{\alpha})^2, \quad \beta(\omega) = \iota\frac{\text{sign}(\omega)}{2} - \frac{1}{\pi}\log|\frac{\omega}{\omega_r}|. \tag{4.94}$$

The model parameters $\mathbf{m}$ and $\boldsymbol{\alpha}$ are defined as solution of the following nonlinear PDE-constrained optimization problem (Aghamiry et al., 2019c)

$$\begin{aligned} &\min_{\mathbf{u},\mathbf{m}\in\mathcal{M},\boldsymbol{\alpha}\in\mathcal{A}} \quad \mu E(\mathbf{m}) + \nu F(\boldsymbol{\alpha}) \\ &\text{subject to} \quad \begin{cases} \mathbf{A}(\mathbf{m},\boldsymbol{\alpha})\mathbf{u} = \mathbf{b} \\ \mathbf{P}\mathbf{u} = \mathbf{d}, \end{cases} \end{aligned} \tag{4.95}$$

where $\mathbf{u} \in \mathbb{C}^{N\times 1}$ is the wavefield, $\mathbf{b} \in \mathbb{C}^{N\times 1}$ is the source term, and $\mathbf{A}(\mathbf{m},\boldsymbol{\alpha}) \in \mathbb{C}^{N\times N}$ is the matrix representation of the discretized Helmholtz PDE, in equation 4.91. The observation operator $\mathbf{P} \in \mathbb{R}^{M\times N}$ samples the reconstructed wavefields at the M receiver positions for comparison with the recorded data $\mathbf{d} \in \mathbb{C}^{M\times 1}$ (we assume a single source experiment for the sake of compact notation; however, the extension to multiple sources is straightforward). The functions E and F are appropriate regularization functions for $\mathbf{m}$ and $\boldsymbol{\alpha}$, respectively, which are weighted by the penalty parameters μ and $\nu > 0$, respectively. $\mathcal{M}$ and $\mathcal{A}$ are convex sets defined according to our prior knowledge of $\mathbf{m}$ and $\boldsymbol{\alpha}$. For example, if we know the lower and upper bounds on $\mathbf{m}$ and $\boldsymbol{\alpha}$ then

$$\mathcal{M} = \{\mathbf{m} | \mathbf{m}_{min} \leq \mathbf{m} \leq \mathbf{m}_{max}\}, \tag{4.96}$$

and

$$\mathcal{A} = \{\boldsymbol{\alpha} | \boldsymbol{\alpha}_{min} \leq \boldsymbol{\alpha} \leq \boldsymbol{\alpha}_{max}\}. \tag{4.97}$$

The PDE constraint $\mathbf{A}(\mathbf{m},\boldsymbol{\alpha})\mathbf{u} = \mathbf{b}$ in equation 4.95 is nonlinear in $\mathbf{m}$ and $\boldsymbol{\alpha}$ and indefinite (Dolean et al., 2015), while the data constraint $\mathbf{P}\mathbf{u} = \mathbf{d}$ is linear but the operator $\mathbf{P}$ is rank-deficient with a huge null space because $M \ll N$. Therefore, determination of the optimum multivariate solution $(\mathbf{u}^*, \mathbf{m}^*, \boldsymbol{\alpha}^*)$ satisfying both constraints (the wave equation and the observation equation) simultaneously is extremely difficult, and requires sophisticated regularizations. In this paper, we use the first-order isotropic TV regularization (Rudin et al., 1992) for both $\mathbf{m}$ and $\boldsymbol{\alpha}$, i.e. $E(\mathbf{m}) = \|\mathbf{m}\|_{\text{TV}}$ and $F(\boldsymbol{\alpha}) = \|\boldsymbol{\alpha}\|_{\text{TV}}$. However, other regularizations such as compound regularizations (Aghamiry et al., 2020b) can be used in a similar way. The isotropic TV norm of a 2D image $\mathbf{w} \in \mathbb{R}^N$ is defined as

$$\|\mathbf{w}\|_{\text{TV}} = \sum \sqrt{(\nabla_x \mathbf{w})^2 + (\nabla_z \mathbf{w})^2}, \tag{4.98}$$

where ∇_x and ∇_z are respectively first-order difference operators in the horizontal and vertical directions with appropriate boundary conditions (Gholami and Naeini, 2019).

Beginning with an initial model $\mathbf{m}^0$ and $\boldsymbol{\alpha}^0 = \mathbf{0}$, $\mathbf{b}^0 = \mathbf{0}$, $\mathbf{d}^0 = \mathbf{0}$, ADMM solves iteratively the multivariate optimization problem, equation 4.95, with alternating directions as (see Boyd

et al., 2010; Benning et al., 2015; Aghamiry et al., 2019c,b, for more details)

$$\begin{cases} \mathbf{u}^{k+1} = \arg\min_{\mathbf{u}} \Psi(\mathbf{u}, \mathbf{m}^k, \boldsymbol{\alpha}^k, \mathbf{b}^k, \mathbf{d}^k) & (4.99a) \\ \mathbf{m}^{k+1} = \arg\min_{\mathbf{m}\in\mathcal{M}} \Psi(\mathbf{u}^{k+1}, \mathbf{m}, \boldsymbol{\alpha}^k, \mathbf{b}^k, \mathbf{d}^k) & (4.99b) \\ \boldsymbol{\alpha}^{k+1} = \arg\min_{\boldsymbol{\alpha}\in\mathcal{A}} \Psi(\mathbf{u}^{k+1}, \mathbf{m}^{k+1}, \boldsymbol{\alpha}, \mathbf{b}^k, \mathbf{d}^k) & (4.99c) \\ \mathbf{b}^{k+1} = \mathbf{b}^k + \mathbf{b} - \mathbf{A}(\mathbf{m}^{k+1}, \boldsymbol{\alpha}^{k+1})\mathbf{u}^{k+1} & (4.99d) \\ \mathbf{d}^{k+1} = \mathbf{d}^k + \mathbf{d} - \mathbf{P}\mathbf{u}^{k+1}, & (4.99e) \end{cases}$$

where

$$\begin{aligned} \Psi(\mathbf{u}, \mathbf{m}, \boldsymbol{\alpha}, \mathbf{b}^k, \mathbf{d}^k) = \mu\|\mathbf{m}\|_{\text{TV}} + \nu\|\boldsymbol{\alpha}\|_{\text{TV}} \\ + \lambda\|\mathbf{b}^k + \mathbf{b} - \mathbf{A}(\mathbf{m}, \boldsymbol{\alpha})\mathbf{u}\|_2^2 + \gamma\|\mathbf{d}^k + \mathbf{d} - \mathbf{P}\mathbf{u}\|_2^2, \end{aligned} \tag{4.100}$$

is the augmented Lagrangian function written in scaled form (Appendix A, section 4.3.7), $\bullet^k$ is the value of $\bullet$ at iteration k, the scalars $\lambda, \gamma > 0$ are the penalty parameters assigned to the wave equation and observation equation constraints, respectively, and $\mathbf{b}^k$, $\mathbf{d}^k$ are the scaled Lagrange multipliers, which are updated through a dual ascent scheme by the running sum of the constraint violations (source and data residuals) as shown by equations 4.99d-4.99e. The penalty parameters $\lambda, \gamma > 0$ can be tuned in equation 4.100 such that a dominant weight γ is given to the observation equation at the expense of the wave equation during the early iterations to guarantee the data fit, while the iterative update of the Lagrange multipliers progressively correct the errors introduced by these penalizations such that both of the observation equation and the wave equation are satisfied at the convergence point with acceptable accuracies. In the next three subsections, we show how to solve each optimization subproblem 4.99a-4.99c.

Update wavefield (subproblem 4.99a)

The objective function Ψ is quadratic in $\mathbf{u}$ and its minimization gives the following closed-form expression of $\mathbf{u}$

$$\left(\lambda\mathbf{A}^T\mathbf{A} + \gamma\mathbf{P}^T\mathbf{P}\right)\mathbf{u}^{k+1} = \lambda\mathbf{A}^T(\mathbf{b}^k + \mathbf{b}) + \gamma\mathbf{P}^T(\mathbf{d}^k + \mathbf{d}), \tag{4.101}$$

where $\mathbf{A} \equiv \mathbf{A}(\mathbf{m}^k, \boldsymbol{\alpha}^k)$ and $\mathbf{A}^T$ denotes the Hermitian transpose of $\mathbf{A}$.

Update squared slowness (subproblem 4.99b)

The PDE operator

$$\mathbf{A}(\mathbf{m}, \boldsymbol{\alpha}) = \boldsymbol{\Delta} + \omega^2\mathbf{C}\text{diag}(\mathbf{m} \circ \boldsymbol{\rho}(\boldsymbol{\alpha}))\mathbf{B} \tag{4.102}$$

is discretized with the finite-difference method of Chen et al. (2013) where $\boldsymbol{\Delta}$ is the discretized Laplace operator, $\mathbf{C}$ introduces boundary conditions such as perfectly matched layers (Bérenger, 1994), $\mathbf{B}$ is the mass matrix (Marfurt, 1984) which spreads the mass term

$\omega^2\mathbf{C}\text{diag}(\mathbf{m}\circ\boldsymbol{\rho}(\boldsymbol{\alpha}))$ over all the coefficients of the stencil to improve its accuracy following an anti-lumped mass strategy, and diag($\bullet$) denotes a diagonal matrix. From equation 4.102 we get that

$$\mathbf{A}(\mathbf{m},\boldsymbol{\alpha})\mathbf{u} = \mathbf{A}(\mathbf{0},\boldsymbol{\alpha})\mathbf{u} + \omega^2\mathbf{C}\text{diag}(\mathbf{B}\mathbf{u}\circ\boldsymbol{\rho}(\boldsymbol{\alpha}))\mathbf{m}, \tag{4.103}$$

where $\mathbf{A}(\mathbf{0},\boldsymbol{\alpha}) \equiv \boldsymbol{\Delta}$. Therefore, subproblem 4.99b can be written as

$$\mathbf{m}^{k+1} = \underset{\mathbf{m}\in\mathcal{M}}{\arg\min}\ \mu\|\mathbf{m}\|_{\text{TV}} + \lambda\|\mathbf{L}\mathbf{m} - \mathbf{y}^k\|_2^2, \tag{4.104}$$

where

$$\mathbf{L} = \omega^2\mathbf{C}\text{diag}(\mathbf{B}\mathbf{u}^{k+1}\circ\boldsymbol{\rho}(\boldsymbol{\alpha}^k)), \tag{4.105}$$

and

$$\mathbf{y}^k = \mathbf{b}^k + \mathbf{b} - \boldsymbol{\Delta}\mathbf{u}^{k+1}. \tag{4.106}$$

Equation 4.104 describes the box-constrained TV-regularized subproblem for $\mathbf{m}$ which is convex but non-smooth. This box-constrained TV-regularized problem can be solved efficiently with ADMM and splitting methods, also referred to as the split Bregman method (Goldstein and Osher, 2009). Using splitting methods, the unconstrained subproblem 4.104 is recast as a multivariate constrained problem, through the introduction of auxiliary variables. These auxiliary variables are introduced to decouple the $\ell 2$ subproblem from the $\ell 1$ subproblem such that they can be solved in alternating mode with ADMM. Moreover, a closed-form expression of the auxiliary variables is easily obtained by solving the $\ell 1$ subproblem with proximity operators (Combettes and Pesquet, 2011; Parikh and Boyd, 2013). We refer the reader to Appendix B (section 4.3.8) for a more detailed review of this method.

Update attenuation factor (subproblem 4.99c)

Subproblem 4.99c is nonlinear due to the nonlinearity of the PDE with respect to $\boldsymbol{\alpha}$. We linearize this subproblem by using a first-order approximation of $\rho(x)$, equation 4.94, as

$$\rho(x) \approx 1 + 2\beta(\omega)x, \tag{4.107}$$

which is accurate for $|x| \ll 1$ (Hak and Mulder, 2011), and gives

$$\mathbf{A}(\mathbf{m},\boldsymbol{\alpha})\mathbf{u} \approx \mathbf{A}(\mathbf{m},\mathbf{0})\mathbf{u} + 2\omega^2\beta(\omega)\mathbf{C}\text{diag}(\mathbf{B}\mathbf{u}\circ\mathbf{m})\boldsymbol{\alpha}. \tag{4.108}$$

Accordingly, subproblem 4.99c can be written as the following linear problem

$$\boldsymbol{\alpha}^{k+1} \approx \underset{\boldsymbol{\alpha}\in\mathcal{A}}{\arg\min}\ \nu\|\boldsymbol{\alpha}\|_{\text{TV}} + \lambda\|\mathbf{H}\boldsymbol{\alpha} - \mathbf{h}^k\|_2^2, \tag{4.109}$$

where

$$\mathbf{H} = 2\omega^2\beta(\omega)\mathbf{C}\text{diag}(\mathbf{B}\mathbf{u}^{k+1}\circ\mathbf{m}^{k+1}), \tag{4.110}$$

and

$$\mathbf{h}^k = \mathbf{b}^k + \mathbf{b} - \mathbf{A}(\mathbf{m}^{k+1},\mathbf{0})\mathbf{u}^{k+1}. \tag{4.111}$$

Equation 4.109 is also a box-constrained TV-regularized convex problem, which can be solved with ADMM (Appendix A, section 4.3.7) in a manner similar to the previous subproblem for squared slowness. It is important to stress that the errors generated by the first-order approximation of $\rho(x)$ during the update of $\boldsymbol{\alpha}$ are iteratively compensated by the action of the scaled Lagrange multiplier $\mathbf{b}^k$. These Lagrange multipliers are formed by the running sum of the wave equation errors, which are computed with the exact wave-equation operator (namely, without linearization of $\rho(x)$). The overall workflow described above is summarized in Algorithm 4.

Practical implementation

The ADMM optimization that is used to solve equations 4.104 and 4.109 is reviewed in Appendix B (section 4.3.8). We also refer the reader to Goldstein and Osher (2009), Boyd et al. (2010) and Aghamiry et al. (2019b) as complements. A key property of the ADMM algorithm, equation 4.99, is that, at iteration k, we don't need to solve each optimization subproblems 4.99a-4.99c exactly via inner iterations. The intuitive reason is that the updating of the primal variable performed by one subproblem is hampered by the errors of the other primal variables that are kept fixed. In this framework, the errors at each iteration k are more efficiently compensated by the gradient-ascent update of the Lagrange multipliers (dual variable). This statement was corroborated by numerical experiments which showed that one (inner) iteration of each subproblem per ADMM cycle k generates solutions which are accurate enough to guarantee the fastest convergence of the ADMM algorithm (Goldstein and Osher, 2009; Boyd et al., 2010; Gholami et al., 2018; Aghamiry et al., 2019a,b). Moreover, this error compensation is more efficient when the dual variables are updated after each primal subproblem 4.99a-4.99c rather than at the end of an iteration k as indicated in Algorithm 4 for sake of compactness. This variant of ADMM is referred to as the Peaceman-Rachford splitting method (Peaceman and Rachford, 1955; He et al., 2014) and will be used in the following numerical experiments. The reader is referred to Aghamiry et al. (2019c) for more details about the improved convergence of the Peaceman-Rachford splitting method compared to ADMM in the framework of IR-WRI.

We properly scale the different parameters to reduce the ill-conditioning of the multi-parameter optimization resulting from the different order of magnitude of $\mathbf{m}$ and $\boldsymbol{\alpha}$ (Operto et al., 2013). In this paper, we scale $\mathbf{m}$ by 1e6 to make it dimensionless as $\boldsymbol{\alpha}$. Also, for tuning the penalty parameters, we follow the guideline presented in Aghamiry et al. (2019b, section 3.1). The overall procedure is as follow: we first set μ=0.6 and ν=0.4 to tune the relative weight of the regularizations of the squared slowness ($\mathbf{m}$) and attenuation factor ($\boldsymbol{\alpha}$), equation 4.100. Then, we set the ratios μ/ξ_m and ν/ξ_α, lines 7, 9, 10, 12 in Algorithm 4, to tune the soft thresholding performed by the TV regularization of $\mathbf{m}$ and $\boldsymbol{\alpha}$ (subproblems 4.99b and 4.99c). We refer the reader to Appendix B (section 4.3.8) for the role of the penalty parameters ξ_m and ξ_α in TV regularization (denoted generically by ξ in equation 4.120). We set μ/ξ_m and ν/ξ_α equal to $0.02\times\max(\mathbf{r})$, in equations 4.136. These values can be refined according to our prior knowledge of the subsurface medium. Then, we set a constant λ to balance the relative weight of the regularization and the wave equation misfit function during the parameter-estimation subproblems, equations 4.104 and 4.109, and lines 5 and 6 in Algorithm 4. If necessary, λ can be increased during iterations to mitigate the imprint of the regularization near the convergence

Algorithm 4: Viscoacoustic Wavefield Inversion with Bound Constrained TV Regularization. Lines 4 to 6 are the primal subproblems for wavefield reconstruction and parameter estimation. Lines 7 to 12 are primal subproblems for auxiliary variables introduced to implement nonsmooth regularizations and bound constraints (Appendix B, section 4.3.8). Lines 13 to 20 are the dual subproblems solved with gradient ascent steps.

1: Begin with $k = 0$, an initial squared slowness $\mathbf{m}^0$, and attenuation $\boldsymbol{\alpha}^0$,

2: Set to zero the values of $\mathbf{d}^0, \mathbf{b}^0, \mathbf{p}^0_{x,m}, \mathbf{p}^0_{y,m}, \mathbf{p}^0_{z,m}, \mathbf{p}^0_{x,\alpha}, \mathbf{p}^0_{y,\alpha}, \mathbf{p}^0_{z,\alpha}, \mathbf{q}^0_{x,m}, \mathbf{q}^0_{y,m}, \mathbf{q}^0_{z,m}, \mathbf{q}^0_{x,\alpha}, \mathbf{q}^0_{y,\alpha}, \mathbf{q}^0_{z,\alpha},$

3: **while** convergence criteria not satisfied **do**

4: $\mathbf{u}^{k+1} = \left[\lambda \mathbf{A}^T\mathbf{A} + \gamma \mathbf{P}^T\mathbf{P}\right]^{-1}\left[\lambda \mathbf{A}^T[\mathbf{b}^k + \mathbf{b}] + \gamma \mathbf{P}^T[\mathbf{d}^k + \mathbf{d}]\right]$

5: $\mathbf{m}^{k+1} = \left[\lambda \mathbf{L}^T\mathbf{L} + \xi_m\nabla_x^T\nabla_x + \xi_m\mathbf{I} + \xi_m\nabla_z^T\nabla_z\right]^{-1}\Big[\lambda \mathbf{L}^T\mathbf{y}^k + \xi_m\nabla_x^T[\mathbf{p}^k_{x,m} + \mathbf{q}^k_{x,m}] + \xi_m[\mathbf{p}^k_{y,m} + \mathbf{q}^k_{y,m}] + \xi_m\nabla_z^T[\mathbf{p}^k_{z,m} + \mathbf{q}^k_{z,m}]\Big]$

6: $\boldsymbol{\alpha}^{k+1} = \left[\lambda \mathbf{H}^T\mathbf{H} + \xi_\alpha\nabla_x^T\nabla_x + \xi_\alpha\mathbf{I} + \xi_\alpha\nabla_z^T\nabla_z\right]^{-1}\Big[\lambda \mathbf{H}^T\mathbf{h}^k + \xi_\alpha\nabla_x^T[\mathbf{p}^k_{x,\alpha} + \mathbf{q}^k_{x,\alpha}] + \xi_\alpha[\mathbf{p}^k_{y,\alpha} + \mathbf{q}^k_{y,\alpha}] + \xi_\alpha\nabla_z^T[\mathbf{p}^k_{z,\alpha} + \mathbf{q}^k_{z,\alpha}]\Big]$

7: $\mathbf{p}^{k+1}_{x,m} = \max(1 - \frac{\mu/\xi_m}{\sqrt{|\nabla_x\mathbf{m}^{k+1} - \mathbf{q}^k_{x,m}|^2 + |\nabla_z\mathbf{m}^{k+1} - \mathbf{q}^k_{z,m}|^2}}, 0) \circ (\nabla_x\mathbf{m}^{k+1} - \mathbf{q}^k_{x,m})$

8: $\mathbf{p}^{k+1}_{y,m} = \text{proj}_{\mathcal{M}}(\mathbf{m}^{k+1} - \mathbf{q}^k_{y,m})$

9: $\mathbf{p}^{k+1}_{z,m} = \max(1 - \frac{\mu/\xi_m}{\sqrt{|\nabla_x\mathbf{m}^{k+1} - \mathbf{q}^k_{x,m}|^2 + |\nabla_z\mathbf{m}^{k+1} - \mathbf{q}^k_{z,m}|^2}}, 0) \circ (\nabla_z\mathbf{m}^{k+1} - \mathbf{q}^k_{z,m})$

10: $\mathbf{p}^{k+1}_{x,\alpha} = \max(1 - \frac{\nu/\xi_\alpha}{\sqrt{|\nabla_x\boldsymbol{\alpha}^{k+1} - \mathbf{q}^k_{x,\alpha}|^2 + |\nabla_z\boldsymbol{\alpha}^{k+1} - \mathbf{q}^k_{z,\alpha}|^2}}, 0) \circ (\nabla_x\boldsymbol{\alpha}^{k+1} - \mathbf{q}^k_{x,\alpha})$

11: $\mathbf{p}^{k+1}_{y,\alpha} = \text{proj}_{\mathcal{A}}(\boldsymbol{\alpha}^{k+1} - \mathbf{q}^k_{y,\alpha})$

12: $\mathbf{p}^{k+1}_{z,\alpha} = \max(1 - \frac{\nu/\xi_\alpha}{\sqrt{|\nabla_x\boldsymbol{\alpha}^{k+1} - \mathbf{q}^k_{x,\alpha}|^2 + |\nabla_z\boldsymbol{\alpha}^{k+1} - \mathbf{q}^k_{z,\alpha}|^2}}, 0) \circ (\nabla_z\boldsymbol{\alpha}^{k+1} - \mathbf{q}^k_{z,\alpha})$

13: $\mathbf{q}^{k+1}_{x,m} = \mathbf{q}^k_{x,m} + \mathbf{p}^{k+1}_{x,m} - \nabla_x\mathbf{m}^{k+1}$

14: $\mathbf{q}^{k+1}_{y,m} = \mathbf{q}^k_{y,m} + \mathbf{p}^{k+1}_{y,m} - \mathbf{m}^{k+1}$

15: $\mathbf{q}^{k+1}_{z,m} = \mathbf{q}^k_{z,m} + \mathbf{p}^{k+1}_{z,m} - \nabla_z\mathbf{m}^{k+1}$

16: $\mathbf{q}^{k+1}_{x,\alpha} = \mathbf{q}^k_{x,\alpha} + \mathbf{p}^{k+1}_{x,\alpha} - \nabla_x\boldsymbol{\alpha}^{k+1}$

17: $\mathbf{q}^{k+1}_{y,\alpha} = \mathbf{q}^k_{y,\alpha} + \mathbf{p}^{k+1}_{y,\alpha} - \boldsymbol{\alpha}^{k+1}$

18: $\mathbf{q}^{k+1}_{z,\alpha} = \mathbf{q}^k_{z,\alpha} + \mathbf{p}^{k+1}_{z,\alpha} - \nabla_z\boldsymbol{\alpha}^{k+1}$

19: $\mathbf{b}^{k+1} = \mathbf{b}^k + \mathbf{b} - \mathbf{A}(\mathbf{m}^{k+1}, \boldsymbol{\alpha}^{k+1})\mathbf{u}^{k+1}$

20: $\mathbf{d}^{k+1} = \mathbf{d}^k + \mathbf{d} - \mathbf{P}\mathbf{u}^{k+1}$

21: $k = k + 1$

22: **end while**

point. Finally, we set γ for wavefield reconstruction such that λ/γ is a small fraction of the highest eigenvalue of the regularized normal operator, equation 4.101 and line 1 in Algorithm 4 (van Leeuwen and Herrmann, 2016). This parameter γ can be kept constant during iterations. The reader is referred to Aghamiry et al. (2019c) for an analysis of the sensitivity of IR-WRI to the choice of the weight balancing the role of the observation equation and the wave equation during wavefield reconstruction.

4.3.5 Numerical examples

Simple inclusions test

We first consider a simple 2D example to validate viscoacoustic IR-WRI without and with TV regularization. The true velocity model is a homogeneous background model with a wavespeed of 1.5 km/s, which contains two inclusions: a 250-m-diameter circular inclusion at position (1 km,1.6 km) with a wavespeed of 1.8 km/s and a 0.2 $\times$ 0.8 km rectangular inclusion at the center of the model with a wavespeed of 1.3 km/s (Figure 4.20a). Also, the true $\boldsymbol{\alpha}$ model is a homogeneous background model with attenuation 0.01, which contains two inclusions of values 0.1. The first is a 250-m-diameter circular inclusion at position (1 km,0.4 km) and the second is a 0.2 $\times$ 0.8 km rectangle inclusion at the center of the model (Figure 4.20b). The wavespeed and $\boldsymbol{\alpha}$ rectangular inclusions share the same size and position, while the position of the wavespeed and $\boldsymbol{\alpha}$ circular inclusions are different to test different parameter trade-off scenarios. A vertical and horizontal logs which cross the center of $\mathbf{v}$ and $\boldsymbol{\alpha}$ models are plotted in the left and bottom side of the models, respectively, for all the figures of this test. An ideal acquisition is used with eight sources and 200 receivers along the four edges of the model, and three frequency components (2.5, 5.0, 7.0 Hz) are jointly inverted using a maximum of 30 iterations as stopping criterion of iteration.

We performed viscoacoustic IR-WRI without and with TV regularization starting from homogeneous velocity model 1.5 km/s and zero attenuation. The final $(\mathbf{v}, \boldsymbol{\alpha})$ models estimated by IR-WRI without and with TV regularization are shown in Figures 4.20c and 4.20d and 4.20e and 4.20f, respectively. The IR-WRI results without TV regularization show an acceptable velocity model and a quite noisy attenuation reconstruction (Figure 4.20c and 4.20d). The velocity reconstruction is, however, hampered by significant limited bandwidth effects. We also show underestimated wavespeeds in the rectangular inclusion along with the horizontal log. These underestimated wavespeeds are clearly correlated with underestimated $\boldsymbol{\alpha}$ values, hence highlighting some trade-off between $\mathbf{v}$ and $\boldsymbol{\alpha}$. Other moderate trade-off artifacts are shown in the vertical log of the $\boldsymbol{\alpha}$ model, which shows undesired high-frequency perturbations at the position of the circular velocity inclusion. The TV regularization removes all of these pathologies very efficiently: it extends the wavenumber bandwidth of the models and removes to a large extent the parameter crosstalk as the subsurface medium perfectly matches the piecewise-constant prior associated with TV regularization (Figure 4.20e and 4.20f). Note though that $\boldsymbol{\alpha}$ remains slightly underestimated in the circular inclusion (Figure 4.20f). This correlates with a barely-visible velocity underestimation at this location in Figure 4.20e. The relative magnitude of these errors gives some insight on the relative sensitivity of the data to $\mathbf{m}$ and $\boldsymbol{\alpha}$.

Synthetic North Sea case study

Experimental setup

We consider a more realistic 16.0 km $\times$ 5.2 km shallow-water synthetic model representative of the North Sea (Munns, 1985). The true $\mathbf{v}$ and $\boldsymbol{\alpha}$ models are shown in Figures 4.21a and 4.21b, respectively. The velocity model is formed by soft sediments in the upper part, a pile of

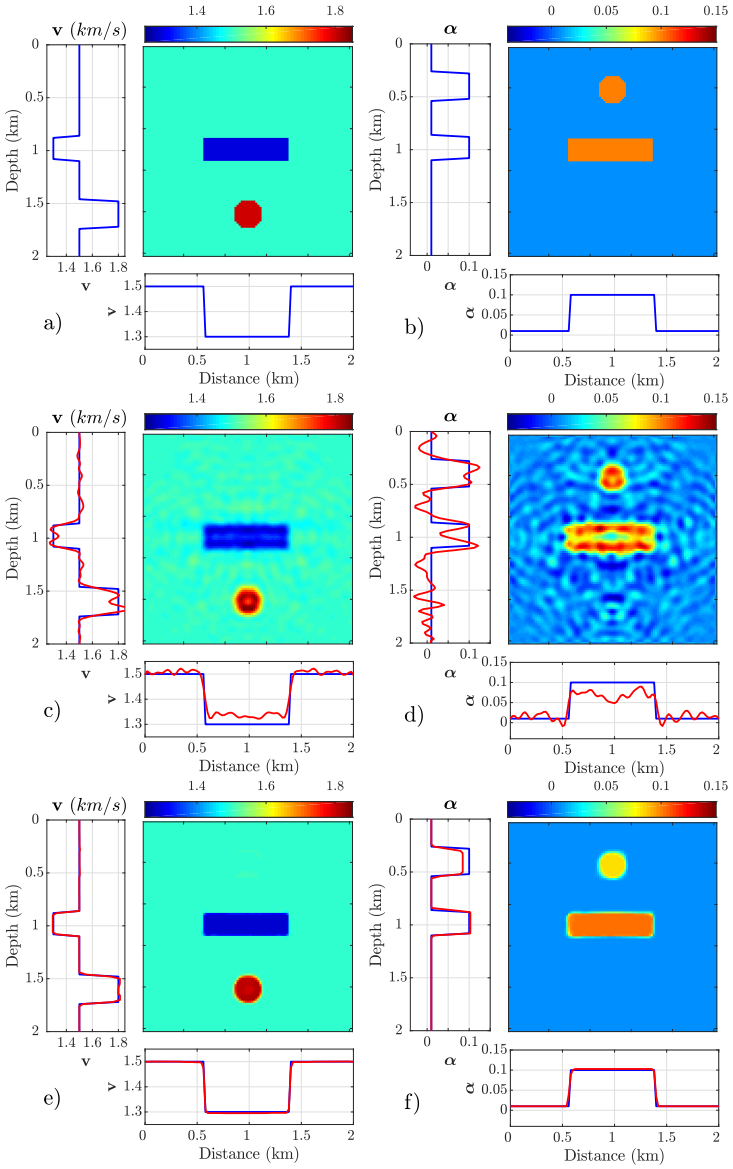

Figure 4.20 – (a) True velocity model. (b) True attenuation model. (c-d) Reconstructed velocity (c) and attenuation (d) from viscoacoustic IR-WRI without TV regularization. (e-f) Same as (a-b) when TV regularization is applied. Profiles of the true (blue) and reconstructed (red) models running across the center of the models are shown on the left and bottom of the reconstructed models.

low-velocity gas layers above a chalk reservoir, the top of which is indicated by a sharp positive velocity contrast at around 2.5 km depth, and a flat reflector at 5.0 km depth (Figure 4.21a). The $\boldsymbol{\alpha}$ model has two highly attenuative zones in the upper soft sediments and gas layers, and the $\boldsymbol{\alpha}$ value is relatively low elsewhere (Figure 4.21b).

The fixed-spread surface acquisition consists of 320 explosive sources spaced 50 m apart at 25 m depth, and 80 hydrophone receivers spaced 200 m apart on the seafloor at 75 m depth.

For the sake of computational efficiency, we use the spatial reciprocity of Green's functions to process sources as receivers and vice versa. A free-surface boundary condition is used on top of the grid, and the source signature is a Ricker wavelet with a 10 Hz dominant frequency. We perform forward modeling with a nine-point stencil finite-difference method implemented with anti-lumped mass and PML absorbing boundary conditions, where the stencil coefficients are optimized to the frequency (Chen et al., 2013). We solve the normal-equation system for wavefield reconstruction, equation 4.101, with a sparse direct solver (Duff et al., 1986). The

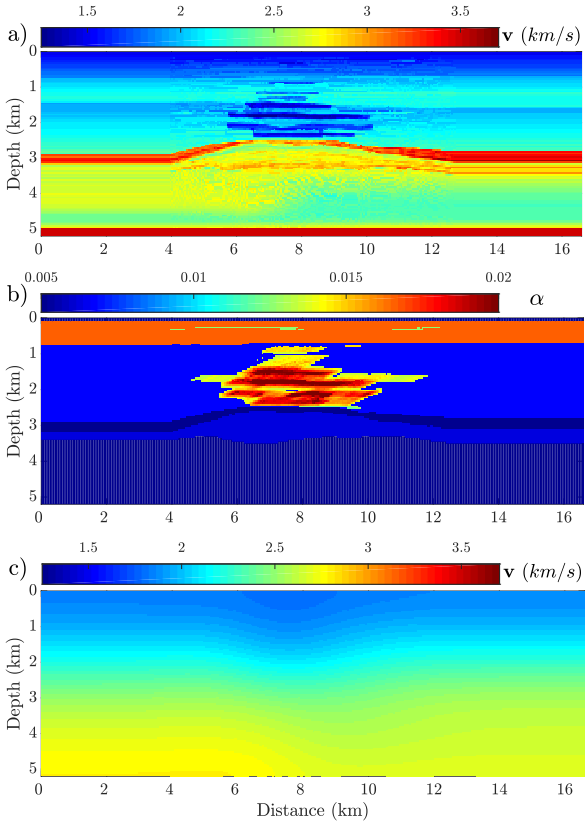

Figure 4.21 – North Sea case study. (a) True $\mathbf{v}$ model. (b) True $\boldsymbol{\alpha}$ model. (c) Initial $\mathbf{v}$ model.

initial model for $\mathbf{v}$ is a highly Gaussian filtered version of the true model (Figure 4.21c), while the starting $\boldsymbol{\alpha}$ model is homogeneous without attenuation. The common-shot gathers computed in the true and initial models are compared in Figure 4.22 for a shot located at 16.0 km. The latter mainly show the direct wave and the diving waves, which are highly cycle skipped relative to those computed in the true model.

We perform the inversion with small batches of three frequencies with one frequency overlap between two consecutive batches, moving from the low frequencies to the higher ones according to a classical frequency continuation strategy. The starting and final frequencies are 3.0 Hz and 15.0 Hz, and the sampling interval in each batch is 0.5 Hz. We perform three paths through the frequency batches to improve the results, using the final model of one path as the initial model of the next one (these paths can be viewed as outer iterations of the algorithm). The starting and finishing frequencies of the paths are [3, 6] Hz, [4, 10] Hz, [6, 15] Hz, respectively, where the first and second elements of each pair show the starting and finishing frequencies, respectively.

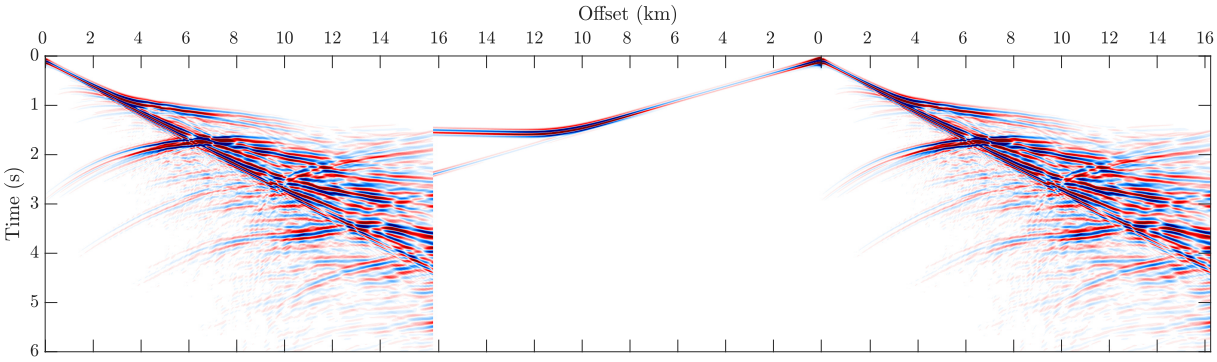

Figure 4.22 – Time domain seismograms computed in the true and initial models. The true seismograms are shown in the left and right panels, while those computed in the initial model are shown in the middle panel with a mirror representation such that the two sets of seismograms can be compared at long and short offsets. The seismograms are plotted with a reduction velocity of 2.5 km/s for sake of time axis compression.

Moreover, as we proceed over the three paths, we increase the number of frequencies par batch from two to five, and the frequency overlap from one to three. The motivation behind this frequency management is to keep the bandwidth of each patch narrow during the first path to mitigate nonlinearities through a progressive frequency continuation, before broadening this bandwidth during the second and third paths to strengthen the imprint of dispersion in the inversion and decouple velocity and attenuation more efficiently.

Comparison of FWI and WRI objective Functions

Before showing the inversion results, it is worth illustrating how WRI extends the search space of FWI for this North Sea case study. For this purpose, we compare the shape of the classical FWI misfit function based upon the $\ell 2$ norm of the data residuals (e.g., Pratt et al., 1998) with that of the parameter-estimation WRI subproblem for the 3 Hz frequency and for a series of $\mathbf{v}$ and $\boldsymbol{\alpha}$ models that are generated according to

$$\mathbf{v}_a = \mathbf{v}_{true} + a^2(\mathbf{v}_{init} - \mathbf{v}_{true}), \tag{4.112a}$$

$$\boldsymbol{\alpha}_b = \boldsymbol{\alpha}_{true} + b^2(\boldsymbol{\alpha}_{init} - \boldsymbol{\alpha}_{true}), \tag{4.112b}$$

where $-1 \leq a, b \leq 1$. In this case, $(\mathbf{v}_a, \boldsymbol{\alpha}_b)$ lies on the line-segment joining the initial point $(\mathbf{v}_{init}, \boldsymbol{\alpha}_{init})$ and final point $(\mathbf{v}_{true}, \boldsymbol{\alpha}_{true})$. We set $\mathbf{v}_{true}$ and $\boldsymbol{\alpha}_{true}$ as those shown in Figures 4.21a and 4.21b, $\mathbf{v}_{init}$ as the initial $\mathbf{v}$ model shown in Figure 4.21c, and $\boldsymbol{\alpha}_{init}$ as a homogeneous model of attenuation 0.004. The misfit functions of FWI and WRI are shown in Figures 4.23a and 4.23b, respectively. The FWI objective function exhibits spurious local minima with respect to both velocity and attenuation (a and b dimensions), while only one minimum is seen in the WRI objective function. Indeed, this highlights the search space expansion generated by the wave equation relaxation during WRI. This search space expansion is displayed through a wider and flatter attraction basin compared to that of FWI. This wide attraction basin exacer-

bates the contrast between the sensitivities of the objective function to velocity and attenuation, with a quite weak sensitivity to the latter. These contrasted sensitivities would likely make the parameter estimation subproblem poorly scaled if the velocity and attenuation were jointly updated during WRI based on variable projection (van Leeuwen and Herrmann, 2016). In this context, the alternating-direction strategy can be viewed as an heuristic to overcome this scaling issue. Indeed, the lack of sensitivity to attenuation during the early WRI iterations requires aggressive regularization and bound constraints to stabilize the attenuation estimation. As WRI proceeds over iterations and the wave equation constraint is satisfied more accurately, the inversion should recover a significant sensitivity to attenuation as that highlighted in Figure 4.23a allowing for a relaxation of the regularization. These statements highlight the need to reconcile search space expansion to manage nonlinearity and regularization plus bound constraints to restrict the range of feasible solutions for $\boldsymbol{\alpha}$.

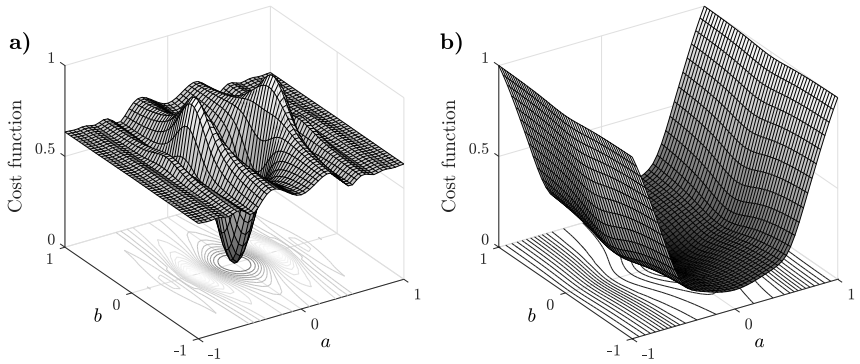

Figure 4.23 – The objective function for the 3 Hz wavefield as a function of $\mathbf{v}_a$ and $\boldsymbol{\alpha}_b$ generated using equation 4.112. (a) Classical reduced-space FWI. (b) WRI.

Viscoacoustic FWI results

To show the need of search space expansion and regularization, we first perform a classical viscoacoustic FWI for noiseless data (e.g., Kamei and Pratt, 2013; Operto and Miniussi, 2018). We use squared-slowness and attenuation as optimization parameters and update them simultaneously with the L-BFGS quasi-Newton optimization and a line search procedure for step length estimation (that satisfies the Wolfe conditions). Neither TV regularization nor bound constraints are applied. Owing the limited kinematic accuracy of the initial models highlighted by the seismograms mismatches in Figure 4.22, the reconstruction of the velocity model remains stuck in a local minimum during the first frequency batch inversion (Figure 4.24a), while the estimated attenuation model shows unrealistic values due to the lack of bound constraints (Figure 4.24b). This failure prompts us to stop the inversion at this stage.

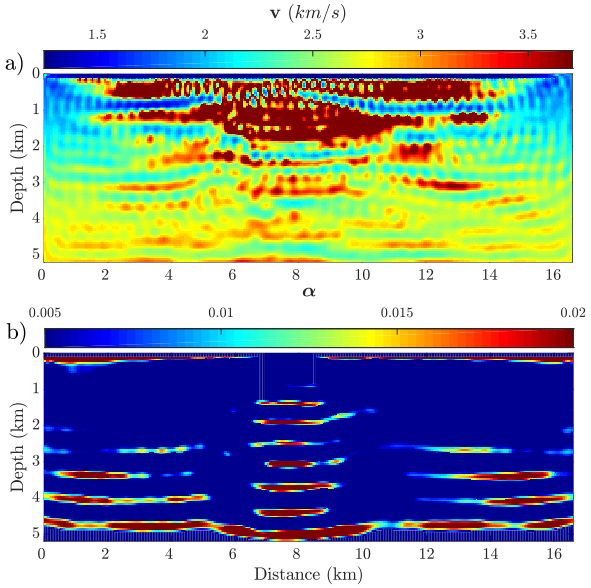

Figure 4.24 – Viscoacoustic FWI results after inverting the first frequency batch. (a) Reconstructed velocity model. (b) Reconstructed attenuation model. TV regularization and bound constraints are not applied.

Viscoacoustic IR-WRI results

We now update $\mathbf{v}$ and $\boldsymbol{\alpha}$ with IR-WRI according to the optimization workflow described in Algorithm 4. IR-WRI is performed without and with bound constraints + TV regularization (referred to as BTV regularization in the following). The lower and upper bounds are 1.2 km/s and 4.0 km/s for velocities, and 0.001 and 0.025 for $\boldsymbol{\alpha}$. For each case, the stopping criterion for each batch is given to be either reaching a maximum iteration count of 20 or

$$\begin{aligned} &\sum \|\mathbf{A}(\mathbf{m}^{k+1}, \boldsymbol{\alpha}^{k+1})\mathbf{u}^{k+1} - \mathbf{b}\|_2^2 \leq \epsilon_b \quad \text{and} \\ &\sum \|\mathbf{P}\mathbf{u}^{k+1} - \mathbf{d}\|_2^2 \leq \epsilon_d, \end{aligned} \tag{4.113}$$

where the sums run over the frequencies of the current batch, ϵ_b=1e-3 and ϵ_d=1e-5. We start with inversion of noiseless data. The final $\mathbf{v}$ and $\boldsymbol{\alpha}$ models, estimated by IR-WRI without and with BTV regularization, are shown in Figure 4.25a-4.25d after 360 and 321 iterations, respectively. A direct comparisons between the logs extracted from the true models, the initial model and the IR-WRI velocity models reconstructed without/with BTV regularization at $x = 3.5$ km, $x = 8.0$ km and $x = 12.0$ km are shown in Figure 4.26a. Although a crude initial velocity model was used, the velocities in the shallow sedimentary cover and the gas layers are fairly well reconstructed in both cases (Figure 4.25a and 4.25c. Also, IR-WRI without BTV regularization manages to reconstruct an acceptable velocity model, unlike classical FWI (Figure 4.24). The main differences between the IR-WRI velocity models built with and without BTV regularization are shown at the reservoir level and below. Without BTV regularization, the top of the reservoir is mispositioned (Figure 4.26a, $x = 8.0$ km, green versus red

curves), and the inversion fails to reconstruct the smoothly decreasing velocity below it due to the lack of diving wave illumination at these depths (Figure 4.26a). This, in turn, prevents the focusing of the deep reflector at 5 km depth by the migration of the associated short-spread reflections (Figure 4.25a). When BTV regularization is used, viscoacoustic IR-WRI provides more accurate and cleaner reservoir images and better reconstructs the sharp contrast on top of it (Figure 4.25c). As expected, the TV regularization replaces the smoothly-varying velocities below the reservoir (between 3 km to 5 km depth) by a piecewise-constant layer due to the lack of wave illumination in this part of the model (Figure 4.26a, green curves). However, this does not prevent a fairly accurate reconstruction of the deep reflector at 5 km depth.

A direct comparisons between the logs extracted from the true and the IR-WRI attenuation models at $x = 3.5$ km, $x = 8.0$ km and $x = 12.0$ km are shown in Figure 4.26b. The reconstruction of $\boldsymbol{\alpha}$ without BTV regularization is quite unstable, with an oscillating trend and overestimated values (Figures 4.25b and red curves in Figure 4.26b). This highlights fairly well the ill-posedness of the attenuation reconstruction. In contrast, the $\boldsymbol{\alpha}$ model reconstructed with BTV regularization captures the large-scale attenuation trend in the shallow sedimentary cover, in the gas layers and below the reservoir (Figures 4.25d and green curves in Figure 4.26b). We note, however, that the attenuation is underestimated on top of the gas layers between 1.0 km and 1.4 km depth (Figure 4.26b, $x = 8.0$ km, green curve). This error in the attenuation reconstruction might be correlated with subtle underestimation of velocities at these depths (Figure 4.26a, $x = 8$ km, green curve). This might indicate, on the one hand, some mild amplitude-related crosstalk effects between velocities and attenuation and, on the other hand, the higher sensitivity of the data to velocities compared to attenuation (in the sense that a small error in the velocity contrasts can compensate more significant attenuation errors). Similar crosstalk artifacts have been previously discussed during the toy inclusion test (Figure 4.20e and 4.20f).

We continue by assessing the resilience of the proposed viscoacoustic IR-WRI to noise when data are contaminated with a Gaussian random noise with a SNR=10 db. Here, SNR is defined based on the root mean square (RMS) amplitude of the signal and that of noise as

$$\text{SNR} = 20 \log \left(\frac{Signal\ RMS\ Amplitude}{Noise\ RMS\ Amplitude} \right). \tag{4.114}$$

We use the same setup and the same initial models as those used for the noiseless case. The stopping criterion is defined by equation 4.113, where ε_d is now set to the noise level. The final models of IR-WRI obtained without and with BTV regularization are shown in Figure 4.25e-4.25h. The total number of IR-WRI iterations are 196 and 185, respectively, for these results. In a similar manner to the noiseless case, a direct comparisons between the logs extracted from the true models, the initial model and the IR-WRI velocity models at $x = 3.5$ km, $x = 8.0$ km and $x = 12.0$ km are shown in Figure 4.27. Overall, a similar trend as for the noiseless case is shown. However, the presence of noise in the data leads to a mispositioning of the reservoir at 8 km distance in the BTV IR-WRI velocity model, which was not observed in the noiseless case (compare Figures 4.26a and 4.27a, x= 8.0 km, green curves). This mispositioning of the reservoir may be correlated with a poorer reconstruction of the attenuating gas layers between 1.0 km and 2.5 km depth (compare Figures 4.26b and 4.27b, x= 8.0 km, green curves). Note also that the velocity model built from noisy data without TV regularization (Figure 4.25e) looks

smoother than that built from noiseless data (Figure 4.25a). This results because the stopping criterion was reached earlier in the noisy case compared to the noiseless case (196 iterations in the noisy case against 360 iterations in the noiseless case).

We also compute a common-shot gather for a shot located at 16.0 km in the IR-WRI models

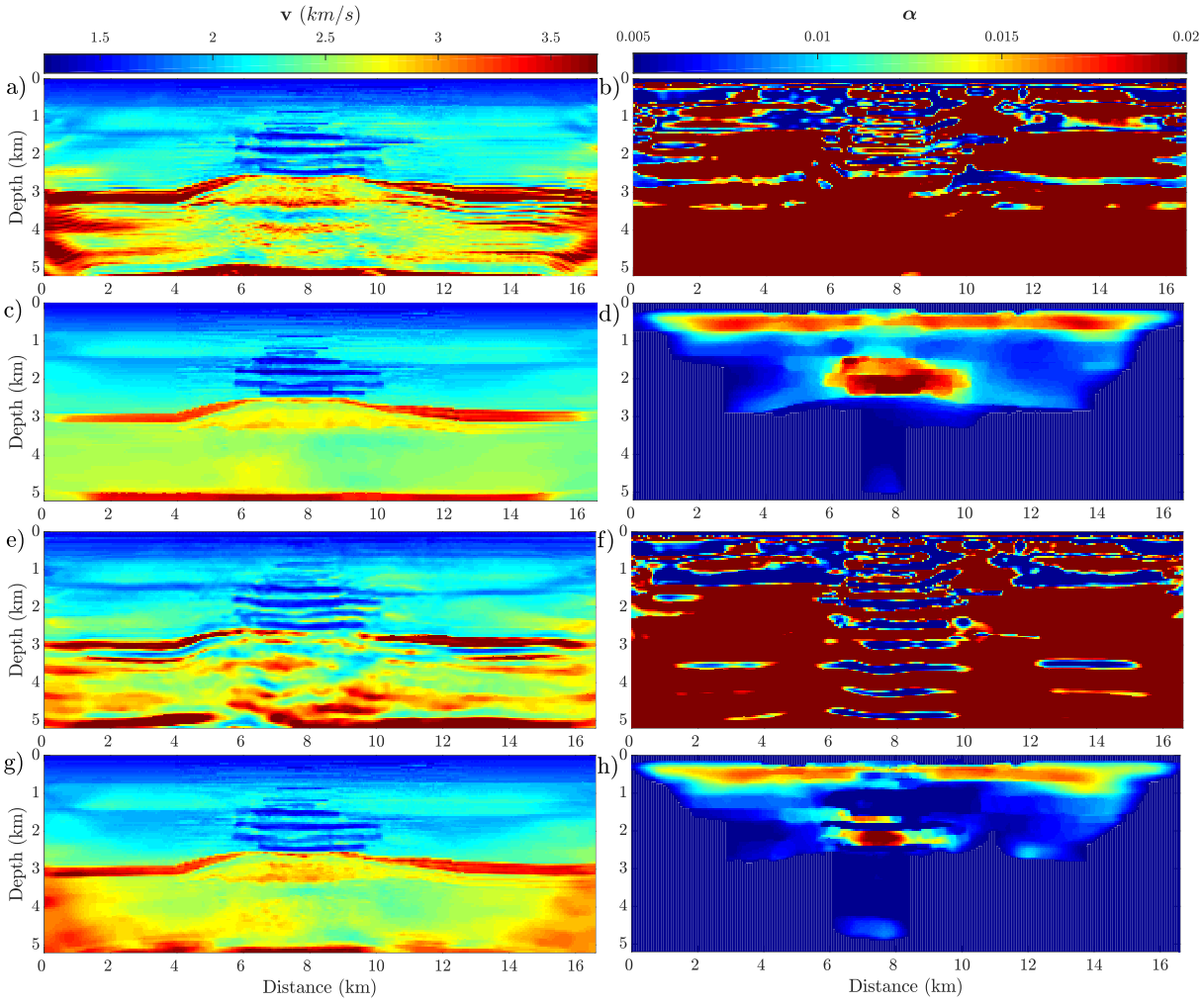

Figure 4.25 – Viscoacoustic IR-WRI results. (a-d) Noiseless data. (a-b) Velocity (a) and attenuation (b) models reconstructed without BTV regularization. (c-d) Same as (a-b) when BTV regularization is applied. (e-h) Same as (a-d) for noisy data (SNR=10db).

inferred from noiseless/noisy data with/without BTV regularization (Figure 4.28). The time-domain seismograms computed in the IR-WRI models obtained without regularization (Figure 4.28a and 4.28c) show underestimated amplitudes and do not match late dispersive arrivals due to the overestimated and oscillating values of $\boldsymbol{\alpha}$ (Figures 4.25b and 4.26b for noiseless data and Figures 4.25f and 4.27b for noisy data).

In the case of noiseless data, the bound constraints and the TV regularization allow for a high-quality data fit (Figure 4.28c), consistently with the accuracy of the models shown in Figure 4.25c and 4.25d. In the case of noisy data, the bound constraints and the TV regularization improve significantly the data match (compare Figure 4.28c and Figure 4.28d). However, the

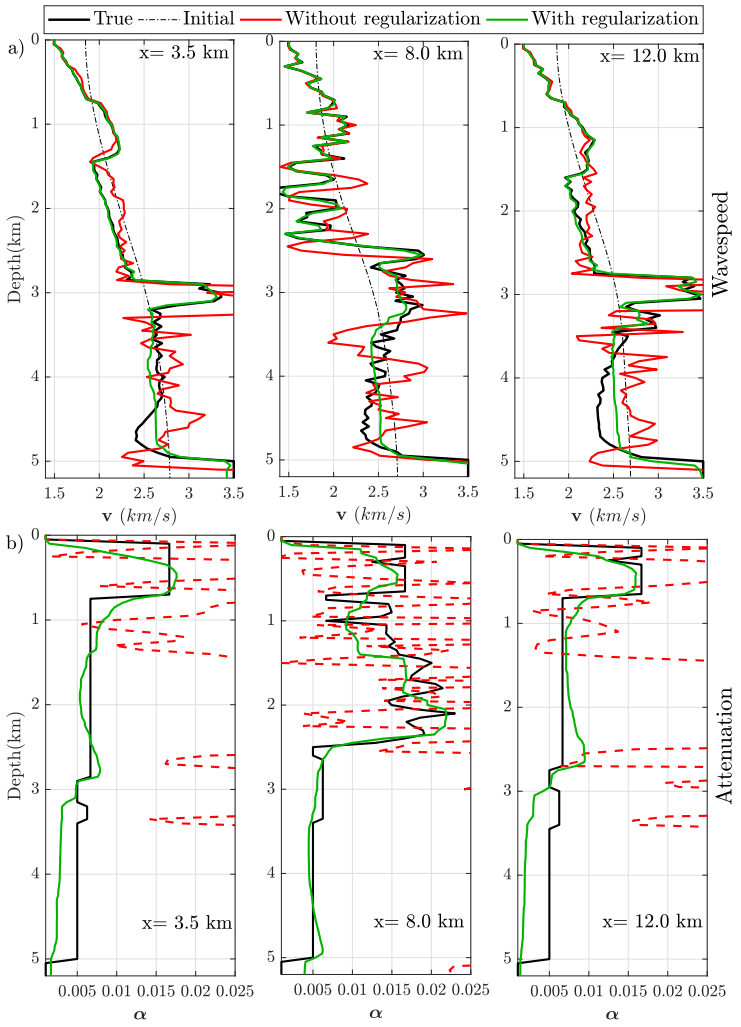

Figure 4.26 – For noiseless data, direct comparison along the logs at $x = 3.5$ km (left), $x = 8.0$ km (center) and $x = 12.0$ km (right) between the true model (black), the initial model (dashed line) and the IR-WRI models without (red) and with BTV (green) regularization (Figure 4.25a-4.25d). (a) Estimated $\mathbf{v}$, (b) estimated $\boldsymbol{\alpha}$.

imprint of the crosstalk artifacts mentioned above are clearly seen at long offsets with a degraded fit of deeply-propagating waves (for example, the refracted wave from the deep reflector at around 1.5s traveltime and the late dispersive waves at around 4 s traveltime) relative to the noiseless-data results (compare Figure 4.28b and 4.28d).

4.3.6 Conclusions

We extended the recently proposed ADMM-based iteratively-refined wavefield reconstruction inversion (IR-WRI) for attenuation imaging by inversion of viscoacoustic wavefields. The proposed viscoacoustic IR-WRI treats the nonlinear viscoacoustic waveform inversion as a multiconvex optimization problem. To achieve this goal, the original nonlinear multi-parameter

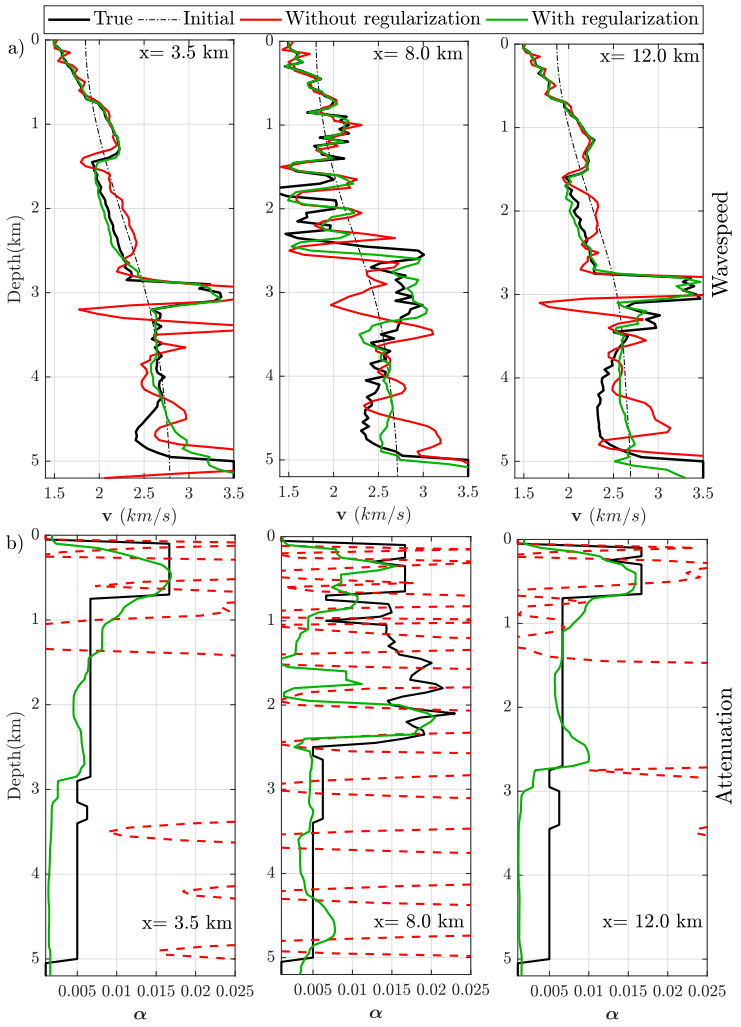

Figure 4.27 – Same as Fig. 4.26 but for noisy data with SNR=10db. The corresponding IR-WRI models are shown in Figure 4.25e-4.25h.

problem for squared slowness and attenuation factor is replaced by three recursive linear mono-parameter subproblems for wavefield, squared slowness and attenuation factor that are solved in alternating mode at each IR-WRI iteration. The attenuation-reconstruction subproblem requires to introduce an approximate multilinear viscoacoustic wave equation in wavefield, squared slowness, and attenuation factor. However, the errors generated by this approximate viscoacoustic wave equation during the attenuation reconstruction are efficiently compensated by the Lagrange multipliers (namely, the running sum of the wave equation errors) that are computed with the exact viscoacoustic equation.

This new formulation first has the flexibility to tailor the regularization to the squared slowness and attenuation factor. Moreover, it simplifies the multi-parameter optimization workflow and mitigates its computational cost since the original poorly-scaled multi-parameter inversion is recast as two interlaced mono-parameter inversions. A realistic synthetic example suggests that the search space extension embedded in IR-WRI efficiently mitigates cycle skipping when a crude initial velocity model is used, while the TV-regularized alternating-direction optimization

reasonably manages the crosstalks between squared slowness and attenuation as well as the limited sensitivity of the data to the attenuation.

Acknowledgments

We would like to thank the editors H. Zhu, J. Blanch, and J. Shragge as well as four anonymous reviewers for their insightful comments which greatly help improving the manuscript. This study was partially funded by the SEISCOPE consortium (*http://seiscope2.osug.fr*), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXON-MOBIL, JGI, PETROBRAS, SCHLUMBERGER, SHELL, SINOPEC, and TOTAL. This study was granted access to the HPC resources of SIGAMM infrastructure (*https://www.oca.eu/fr/mesocentre-sigamm*) and CINES/IDRIS/TGCC under the allocation A0050410596 made by GENCI.

4.3.7 Appendix A: Scaled form of augmented Lagrangian

In this appendix, we briefly review how augmented Lagrangian (AL) function, as the one shown in equation 4.100, is used to solve constrained problems with the method of multiplier (Nocedal and Wright, 2006, Chapter 17). Let's start with the following constrained problem

$$\min_{\mathbf{x}} \ \|\mathcal{P}(\mathbf{x})\|_2^2 \quad \text{subject to} \quad \mathcal{Q}(\mathbf{x}) = \mathbf{0}. \tag{4.115}$$

The AL function associated with the problem 4.115 combines a Lagrangian function and a penalty function as

$$\mathcal{L}_A(\mathbf{x}, \mathbf{v}) = \underbrace{\|\mathcal{P}(\mathbf{x})\|_2^2 + \langle \mathbf{v}, \mathcal{Q}(\mathbf{x})\rangle}_{\text{Lagrangian}} + \underbrace{\frac{\xi}{2}\|\mathcal{Q}(\mathbf{x})\|_2^2}_{\text{Augmentation}}, \tag{4.116}$$

where $\langle \cdot, \cdot \rangle$ denotes inner product and $\mathbf{v}$ and ξ denotes the Lagrange multiplier (dual variable) and the penalty parameter, respectively.
This AL function can be written in a compact form by introducing the scaled dual variable $\mathbf{q} = -\mathbf{v}/\xi$ and adding and subtracting the term $\frac{\xi}{2}\|\mathbf{q}\|_2^2$ to the function 4.116 (See Boyd et al., 2010, page 15 for more details):

$$\begin{aligned}\mathcal{L}_A(\mathbf{x}, \mathbf{q}) &= \|\mathcal{P}(\mathbf{x})\|_2^2 - \xi\langle \mathbf{q}, \mathcal{Q}(\mathbf{x})\rangle + \frac{\xi}{2}\|\mathcal{Q}(\mathbf{x})\|_2^2 + \frac{\xi}{2}\|\mathbf{q}\|_2^2 - \frac{\xi}{2}\|\mathbf{q}\|_2^2 \\ &= \|\mathcal{P}(\mathbf{x})\|_2^2 + \frac{\xi}{2}\|\mathcal{Q}(\mathbf{x}) - \mathbf{q}\|_2^2 - \frac{\xi}{2}\|\mathbf{q}\|_2^2. \end{aligned} \tag{4.117}$$

Equation 4.117 shows the augmented Lagrangian method can be seen as a penalty method with an error correction term in the penalty function, $\frac{\xi}{2}\|\mathcal{Q}(\mathbf{x}) - \mathbf{q}\|_2^2$, corresponding to the scaled Lagrange multipliers. This correction term controls how well the constraint is satisfied at the convergence point. In the framework of the method of multipliers, the AL function is minimized with respect to the primal variable $\mathbf{x}$ and maximized with respect to the scaled dual variable $\mathbf{q}$ in an alternating mode. Expression 4.117 shows that the dual variable is simply updated with

the constraint violation when a gradient ascent method is used. This recipe has been used to derive equation 4.99 with ADMM (AL method with an alternating update of multiple classes of primal variable). The reader is referred to Aghamiry et al. (2019c) for the detailed development.

4.3.8 Appendix B: Bound constrained TV-regularization using ADMM

In this appendix, we review step by step how to solve a bound-constrained TV-regularized convex problem (such as those in equations 4.104 and 4.109) using variable splitting and ADMM (Boyd et al., 2010). Let us consider a general bound-constrained TV-regularized convex problem of the form

$$\min_{\mathbf{x}\in\mathcal{X}} \sum_{i=1}^{N} \sqrt{|\,[\nabla_x \mathbf{x}]_i\,|^2 + |\,[\nabla_z \mathbf{x}]_i\,|^2} + \frac{\lambda}{2}\|\mathbf{G}\mathbf{x}-\mathbf{y}\|_2^2, \tag{4.118}$$

for some column vector $\mathbf{y}$ and matrix $\mathbf{G}$. The model $\mathbf{x}$ is an N-length column vector, ∇_x and ∇_z are square first-order difference matrices, and $\mathcal{X}$ is the desired convex set. The penalty parameter $\lambda > 0$ balances the relative weight of the TV regularizer and the misfit term. Following Aghamiry et al. (2019b, section 2.2.2), equation 4.118 can be solved via the following three easy approaches.

1) *Variable splitting.* Since the variable $\mathbf{x}$ appears simultaneously in the TV, misfit, and bounding terms, it "couples" these terms and makes it difficult to solve the problem. To decouple them, new auxiliary variables $\mathbf{p}_x = \nabla_x\mathbf{x}$, $\mathbf{p}_y \in \mathcal{X}$, and $\mathbf{p}_z = \nabla_z\mathbf{x}$ are substituted in the TV term and the bound constraint, respectively, and their expression as a function of the original variable $\mathbf{x}$ are introduced as new equality constraints. This recasts equation 4.118 as the following constrained problem

$$\min_{\mathbf{x},\mathbf{p}_x,\mathbf{p}_y\in\mathcal{X},\mathbf{p}_z} \sum_{i=1}^{N} \sqrt{|\,[\mathbf{p}_x]_i\,|^2 + |\,[\mathbf{p}_z]_i\,|^2} + \frac{\lambda}{2}\|\mathbf{G}\mathbf{x}-\mathbf{y}\|_2^2 \tag{4.119}$$

$$\text{subject to} \begin{cases} \mathbf{p}_x = \nabla_x\mathbf{x}, \\ \mathbf{p}_y = \mathbf{x}, \\ \mathbf{p}_z = \nabla_z\mathbf{x}. \end{cases}$$

2) *Augmented Lagrangian.* The second trick is to relax these new linking constraints with an augmented Lagrangian function. This recasts 4.119 as the following min-max optimization

problem

$$\begin{aligned}\min_{\mathbf{x},\mathbf{p}_x,\mathbf{p}_y\in\mathcal{X},\mathbf{p}_z}\max_{\mathbf{q}_x,\mathbf{q}_y,\mathbf{q}_z}&\sum_{i=1}^{N}\sqrt{|\,[\mathbf{p}_x]_i\,|^2+|\,[\mathbf{p}_z]_i\,|^2}+\frac{\lambda}{2}\|\mathbf{G}\mathbf{x}-\mathbf{y}\|_2^2\\&+\langle\mathbf{q}_x,\mathbf{p}_x-\nabla_x\mathbf{x}\rangle+\frac{\xi}{2}\|\mathbf{p}_x-\nabla_x\mathbf{x}\|_2^2\\&+\langle\mathbf{q}_y,\mathbf{p}_y-\mathbf{x}\rangle+\frac{\xi}{2}\|\mathbf{p}_y-\mathbf{x}\|_2^2\\&+\langle\mathbf{q}_z,\mathbf{p}_z-\nabla_z\mathbf{x}\rangle+\frac{\xi}{2}\|\mathbf{p}_z-\nabla_z\mathbf{x}\|_2^2,\end{aligned}\tag{4.120}$$

where $\mathbf{q}_x,\mathbf{q}_y,\mathbf{q}_z$ are Lagrange multipliers (dual variables), and $\xi>0$ is the penalty parameter. Note that, here we used the same penalty parameter for all three constraints but one may use different parameter for each of them. The augmented Lagrangian method (a.k.a. method of multipliers, Hestenes, 1969) maximizes the objective in equation 4.120 with respect to the dual variables iteratively by using a simple steepest ascent algorithm (with step length ξ)

$$\mathbf{q}_x^{k+1}=\mathbf{q}_x^k+\xi(\mathbf{p}_x^{k+1}-\nabla_x\mathbf{x}^{k+1})\tag{4.121}$$
$$\mathbf{q}_y^{k+1}=\mathbf{q}_y^k+\xi(\mathbf{p}_y^{k+1}-\mathbf{x}^{k+1})\tag{4.122}$$
$$\mathbf{q}_z^{k+1}=\mathbf{q}_z^k+\xi(\mathbf{p}_z^{k+1}-\nabla_z\mathbf{x}^{k+1})\tag{4.123}$$

where $\mathbf{x}^{k+1}$ and $\mathbf{p}_x^{k+1},\mathbf{p}_y^{k+1},\mathbf{p}_z^{k+1}$ are obtained by solving

$$\begin{aligned}\arg\min_{\mathbf{x},\mathbf{p}_x,\mathbf{p}_y\in\mathcal{X},\mathbf{p}_z}&\sum_{i=1}^{N}\sqrt{|\,[\mathbf{p}_x]_i\,|^2+|\,[\mathbf{p}_z]_i\,|^2}+\frac{\lambda}{2}\|\mathbf{G}\mathbf{x}-\mathbf{y}\|_2^2\\&+\langle\mathbf{q}_x^k,\mathbf{p}_x-\nabla_x\mathbf{x}\rangle+\frac{\xi}{2}\|\mathbf{p}_x-\nabla_x\mathbf{x}\|_2^2\\&+\langle\mathbf{q}_y^k,\mathbf{p}_y-\mathbf{x}\rangle+\frac{\xi}{2}\|\mathbf{p}_y-\mathbf{x}\|_2^2\\&+\langle\mathbf{q}_z^k,\mathbf{p}_z-\nabla_z\mathbf{x}\rangle+\frac{\xi}{2}\|\mathbf{p}_z-\nabla_z\mathbf{x}\|_2^2,\end{aligned}\tag{4.124}$$

beginning with $\mathbf{p}_x^0=\mathbf{p}_y^0=\mathbf{p}_z^0=\mathbf{0}$.

Equations 4.121-4.124 can be simplified by using a change of variables ($\mathbf{q}_x\leftarrow\frac{1}{\xi}\mathbf{q}_x,\mathbf{q}_y\leftarrow\frac{1}{\xi}\mathbf{q}_y,\mathbf{q}_z\leftarrow\frac{1}{\xi}\mathbf{q}_z$) using the fact that for two real vectors $\mathbf{a}$ and $\mathbf{b}$ the following holds:

$$\langle\mathbf{a},\mathbf{b}\rangle+\frac{\xi}{2}\|\mathbf{b}\|_2^2=\frac{\xi}{2}\|\mathbf{b}+\frac{1}{\xi}\mathbf{a}\|_2^2-\frac{\xi}{2}\|\frac{1}{\xi}\mathbf{a}\|_2^2.\tag{4.125}$$

Accordingly,

$$\mathbf{q}_x^{k+1} = \mathbf{q}_x^k + \mathbf{p}_x^{k+1} - \nabla_x \mathbf{x}^{k+1} \tag{4.126}$$
$$\mathbf{q}_y^{k+1} = \mathbf{q}_y^k + \mathbf{p}_y^{k+1} - \mathbf{x}^{k+1} \tag{4.127}$$
$$\mathbf{q}_z^{k+1} = \mathbf{q}_z^k + \mathbf{p}_z^{k+1} - \nabla_z \mathbf{x}^{k+1} \tag{4.128}$$

and

$$\begin{aligned}\underset{\mathbf{x},\mathbf{p}_x,\mathbf{p}_y\in\mathcal{X},\mathbf{p}_z}{\arg\min}\ & \sum_{i=1}^{N}\sqrt{|\mathbf{p}_x|_i^2+|\mathbf{p}_z|_i^2}+\frac{\lambda}{2}\|\mathbf{G}\mathbf{x}-\mathbf{y}\|_2^2 \\ &+\frac{\xi}{2}\|\mathbf{p}_x-\nabla_x\mathbf{x}+\mathbf{q}_x^k\|_2^2-\frac{\xi}{2}\|\mathbf{q}_x^k\|_2^2 \\ &+\frac{\xi}{2}\|\mathbf{p}_y-\mathbf{x}+\mathbf{q}_y^k\|_2^2-\frac{\xi}{2}\|\mathbf{q}_y^k\|_2^2 \\ &+\frac{\xi}{2}\|\mathbf{p}_z-\nabla_z\mathbf{x}+\mathbf{q}_z^k\|_2^2-\frac{\xi}{2}\|\mathbf{q}_z^k\|_2^2. \end{aligned}\tag{4.129}$$

3) *Alternating minimization.* The basic augmented Lagrangian method minimizes the objective function in equation 4.129 (augmented Lagrangian function) jointly over $\mathbf{x}$, $\mathbf{p}_x$, $\mathbf{p}_y$, and $\mathbf{p}_z$, the third trick is to perform this minimization by alternating minimizing with respect to each variable separately (Goldstein and Osher, 2009; Boyd et al., 2010) to arrive at the so-called ADMM.

$$\begin{aligned}\mathbf{x}^{k+1} = \underset{\mathbf{x}}{\arg\min}\ & \frac{\lambda}{2}\|\mathbf{G}\mathbf{x}-\mathbf{y}\|_2^2+\frac{\xi}{2}\|\mathbf{p}_x^k-\nabla_x\mathbf{x}+\mathbf{q}_x^k\|_2^2 \\ &+\frac{\xi}{2}\|\mathbf{p}_y^k-\mathbf{x}+\mathbf{q}_y^k\|_2^2+\frac{\xi}{2}\|\mathbf{p}_z^k-\nabla_z\mathbf{x}+\mathbf{q}_z^k\|_2^2. \end{aligned}\tag{4.130}$$

$$\begin{aligned}(\mathbf{p}_x^{k+1},\mathbf{p}_z^{k+1}) = \underset{\mathbf{p}_x,\mathbf{p}_z}{\arg\min}\ & \sum_{i=1}^{N}\sqrt{|\,[\mathbf{p}_x]_i\,|^2+|\,[\mathbf{p}_z]_i\,|^2} \\ &+\frac{\xi}{2}\|\mathbf{p}_x-\nabla_x\mathbf{x}^{k+1}+\mathbf{q}_x^k\|_2^2 \\ &+\frac{\xi}{2}\|\mathbf{p}_z-\nabla_z\mathbf{x}^{k+1}+\mathbf{q}_z^k\|_2^2. \end{aligned}\tag{4.131}$$

$$\mathbf{p}_y^{k+1} = \underset{\mathbf{p}_y\in\mathcal{X}}{\arg\min}\ \frac{\xi}{2}\|\mathbf{p}_y-\mathbf{x}^{k+1}+\mathbf{q}_y^k\|_2^2. \tag{4.132}$$

The subproblem expressed in equation 4.130 is an easy-to-solve least-squares problem, which has a closed-form solution obtained by setting the derivative of the objective function with

respect to $\mathbf{x}$ equal to zero.

$$\mathbf{x}^{k+1} = \left[\lambda \mathbf{G}^T\mathbf{G} + \xi\nabla_x^T\nabla_x + \xi\mathbf{I} + \xi\nabla_z^T\nabla_z\right]^{-1} \\ \left[\lambda \mathbf{G}^T\mathbf{y} + \xi\nabla_x^T[\mathbf{p}_x^k + \mathbf{q}_x^k] + \xi[\mathbf{p}_y^k + \mathbf{q}_y^k] + \xi\nabla_z^T[\mathbf{p}_z^k + \mathbf{q}_z^k]\right]. \tag{4.133}$$

The subproblem expressed in equation 4.131 also has a closed form solution, given by the generalized soft thresholding function (Goldstein and Osher, 2009),

$$\mathbf{p}_x^{k+1} = \max(1 - \frac{1/\xi}{\mathbf{r}}, 0) \circ (\nabla_x\mathbf{x}^{k+1} - \mathbf{q}_x^k), \tag{4.134}$$

and

$$\mathbf{p}_z^{k+1} = \max(1 - \frac{1/\xi}{\mathbf{r}}, 0) \circ (\nabla_z\mathbf{x}^{k+1} - \mathbf{q}_z^k), \tag{4.135}$$

where

$$\mathbf{r} = \sqrt{|\nabla_x\mathbf{x}^{k+1} - \mathbf{q}_x^k|^2 + |\nabla_z\mathbf{x}^{k+1} - \mathbf{q}_z^k|^2}. \tag{4.136}$$

The subproblem expressed in equation 4.132 is a projection operator given by

$$\mathbf{q}^{k+1} = \text{proj}_{\mathcal{X}}(\mathbf{x}^{k+1} - \mathbf{q}_y^k). \tag{4.137}$$

In the case that $\mathcal{X}$ is a box set of form

$$\mathcal{X} = \{\mathbf{x} | \mathbf{x}_{min} \leq \mathbf{x} \leq \mathbf{x}_{max}\}, \tag{4.138}$$

then the projection operator admits a closed form solution

$$\text{proj}_{\mathcal{X}}(\mathbf{x}^{k+1} - \mathbf{q}_y^k) = \min(\max(\mathbf{x}^{k+1} - \mathbf{q}_y^k, \mathbf{x}_{min}), \mathbf{x}_{max}),$$

where $\mathbf{x}_{min}$ and $\mathbf{x}_{max}$ are lower and upper bounds of $\mathbf{x}$, respectively.

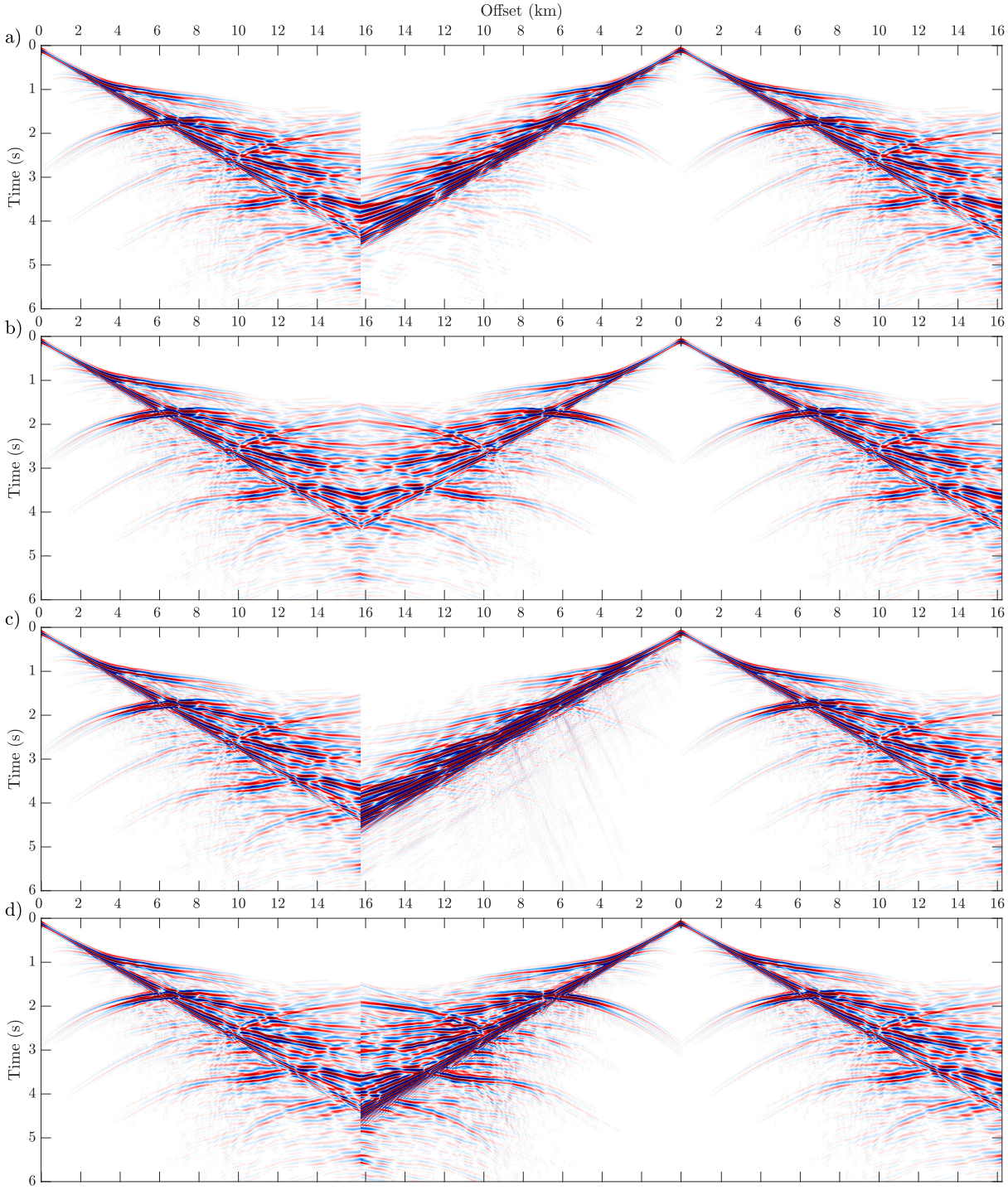

Figure 4.28 – Time domain seismograms computed in (a) IR-WRI without regularization and (b) BTV regularized IR-WRI for noiseless data (Figure 4.25a-4.25d). (c-d) Same as (a-b), but for noisy data (Figure 4.25e-4.25h). The true seismograms are shown in the first and the last panel of the above mentioned seismograms (folded) to have a comparison at short and long offset with true seismograms. The seismograms are plotted with a reduction velocity of 2.5 km/s to compress the time axis.

Chapter 5

Extending the linear regime of ADMM-based WRI with phase retrieval

Chapter overview: The focus of this chapter is to extend the search space of ADMM-based WRI using phase retrieval (e.g., Fienup, 1982). Phase retrieval refers to any process which seeks to reconstruct a complex-valued signal from the magnitude of linear measurement (e.g., Fourier transform). To introduce my motivation behind the use of phase retrieval in ADMM-based WRI, I would like to start with some limitations of ADMM-based WRI in complex media when the initial model is very far away from the true model. In this case, the reconstructed wavefields are accurate only near the receivers (where the measurements are performed). Moreover, it is well acknowledged that the phase has a stronger imprint in a signal relative to amplitude (Oppenheim and Lim, 1981). Accordingly, it sounds reasonable to believe that the inaccuracies of the phase of the reconstructed wavefields may me the leading factor that can drive the inversion towards spurious minimizers. To overcome this issue, I propose to update the parameters with phase retrieval during the early stages of the inversion (i.e., at low frequencies) to mitigate the role of the phase when the background subsurface model is inaccurate. Then, I switch back to the classical form of ADMM-based WRI introduced in the previous chapters to process the higher frequencies with the full-information content of the data. Using a large contrast model with salt bodies as benchmark, I show that phase retrieval combined with efficient regularizations leads to a more accurate reconstruction of the shallow structure and more accurate positioning of the contrasts at low frequencies. The final model obtained after the inversion of the full frequency band shows improved subsalt reconstruction when phase retrieval is used during the early stages of the inversion. This reflects the fastest convergence of classical ADMM-based WRI when it starts from a more accurate low-resolution velocity model built with phase retrieval. This chapter includes our paper, entitled Robust Wavefield Inversion via Phase Retrieval , which has been published in *Geophysical journal international* (Aghamiry et al., 2020e).

Robust Wavefield Inversion via Phase Retrieval
Hossein S. Aghamiry, Ali Gholami and Stéphane Operto
Geophysical Journal International, 2020, **221**(2), pages 1327 - 1340
DOI: 10.1093/gji/ggaa035

5.1 Summary

Extended formulation of Full Waveform Inversion (FWI), called Wavefield Reconstruction Inversion (WRI), offers potential benefits of decreasing the nonlinearity of the inverse problem by replacing the explicit inverse of the ill-conditioned wave-equation operator of classical FWI (the oscillating Green functions) with a suitably defined data-driven regularized inverse. This regularization relaxes the wave-equation constraint to reconstruct wavefields that match the data, hence mitigating the risk of cycle skipping. The subsurface model parameters are then updated in a direction that reduces these constraint violations. However, in the case of a rough initial model, the phase errors in the reconstructed wavefields may trap the waveform inversion in a local minimum leading to inaccurate subsurface models. In this paper, in order to avoid matching such incorrect phase information during the early WRI iterations, we design a new cost function based upon phase retrieval, namely a process which seeks to reconstruct a signal from the amplitude of linear measurements. This new formulation, called Wavefield Inversion with Phase Retrieval (WIPR), further improves the robustness of the parameter estimation subproblem by a suitable phase correction. We implement the resulting WIPR problem with an alternating-direction approach, which combines the Majorization-Minimization (MM) algorithm to linearise the phase-retrieval term and a variable splitting technique based upon the alternating direction method of multipliers (ADMM). This new workflow equipped with Tikhonov-total variation (TT) regularization, which is the combination of second-order Tikhonov and total variation regularizations and bound constraints, successfully reconstructs the 2004 BP salt model from a sparse fixed-spread acquisition using a 3 Hz starting frequency and a homogeneous initial velocity model.

5.2 Introduction

Full waveform inversion (FWI) is a waveform matching procedure which provides a subsurface model with a resolution reaching to the half of the smallest propagated wavelength (Virieux and Operto, 2009). Ultra-long offset wide-azimuth stationary-recording acquisitions provide a varied angular illumination of the subsurface, which is amenable to the development of the broadband velocity model (Sirgue and Pratt, 2004). However, the large number of propagated wavelengths generated by these acquisitions makes FWI prone to cycle skipping or phase wrapping (Shah, 2014; Choi and Alkhalifah, 2015). Various continuation strategies in frequencies, travel-times and offsets can be used to mitigate this pathology (e.g., Shipp and Singh, 2002; Górszczyk et al., 2017). However, these multi-scale approaches can be tedious to implement and still require quite accurate initial velocity models and low frequencies to prevent cycle

skipping at long offsets.

To increase the resilience of FWI to cycle skipping, frequency-domain wavefield reconstruction inversion (WRI) has been proposed by van Leeuwen and Herrmann (2013, 2016). The governing idea of the method relies on the following statement: if we would be able to record a monochromatic wavefield generated by a single source everywhere in the subsurface, then it would be straightforward to reconstruct the subsurface model exactly in virtue of the wave equation bilinearity (Fig. 5.1). Since we cannot record wavefields everywhere, the best we can do to approach the true wavefields is to find those that best match the sparse observations and satisfy the wave equation in a least-squares sense. This least-squares problem can be implemented by solving an overdetermined linear system gathering the weighted wave equation and the observation equation (the equation relating the data to the wavefield through a sparse sampling operator). The observation equation generates a wave equation relaxation whose strength depends on the accuracy of the available subsurface model and the weight assigned to the wave equation. Then, the subsurface parameters are updated from the reconstructed wavefields by the least-squares minimization of the wave equation errors (source residuals), and this two-step cycle is iterated until both the observation equation and the wave equation are satisfied with a prescribed accuracy. Typically, a small weight is assigned to the wave equation during early iterations to foster data fitting and prevent cycle skipping accordingly. This weight is progressively increased to guarantee that the wave equation is fulfilled near the convergence point. The parameter-estimation subproblem can be solved through a variable projection approach (van Leeuwen and Herrmann, 2016) or through alternating optimization (van Leeuwen and Herrmann, 2013). In this latter case, the parameter-estimation subproblem is linearized around the reconstructed wavefield according to the wave-equation bilinearity.

WRI has been originally implemented with a penalty method, which requires a tedious dynamic control of the penalty parameter as above mentioned (Fu and Symes, 2017). To overcome this issue, Aghamiry et al. (2019c) proposed to replace the penalty method with an augmented Lagrangian method equipped with operator splitting, namely the alternating-direction method of multiplier (ADMM) (Boyd et al., 2010). Contrary to penalty methods, augmented Lagrangian methods converge to an accurate solution with a fixed penalty parameter by iteratively updating the Lagrange multipliers with the running sum of the constraint violations (Nocedal and Wright, 2006, Chapter 17). This defines what is sometimes referred to as an iterative refinement or defect correction algorithm. Accordingly, Aghamiry et al. (2019c) called their method iteratively-refined (IR-) WRI.

ADMM also provides a suitable framework to implement bound constraints and $\ell 1$-based nonsmooth regularizations (such as total variation (TV) regularization) in IR-WRI (Aghamiry et al., 2019b). IR-WRI with TV regularization was further improved by using more versatile Tikhonov-TV (TT) regularization combining second-order Tikhonov and TV regularizers through infimal convolution (Aghamiry et al., 2020b). Infimal convolution means that the TT regularization explicitly decomposes the model into two components of different statistical properties (a smooth one and a blocky one) such that a suitable regularization can be tailored to each component (Tikhonov and TV) (Gholami and Hosseini, 2013).

Although IR-WRI, when equipped with basic frequency continuation strategies and advanced regularization techniques, mitigates cycle skipping and helps to overcome nonlinearity

issues broadly speaking, still it can be trapped in a local minimum when the initial model is very far from the true model. In this case, the phase and amplitude of the reconstructed wavefields match those of the true wavefields only in the vicinity of the receivers, while they can be quite inaccurate elsewhere. In IR-WRI, the inaccuracies of the reconstructed wavefields are directly mapped into the subsurface model through a linear deconvolution-like processing. This implies that only the shallow part of the subsurface model is expected to be reliably updated during the early iterations of IR-WRI for surface acquisitions. It remains unclear from a theoretical viewpoint how these inaccuracies are progressively canceled out by IR-WRI in iterations through a depth-continuation process.

It is well acknowledged that the phase often has a more dominant role than amplitude in image processing and waveform inversion methods (Oppenheim and Lim, 1981; Shechtman et al., 2015). Therefore, the inaccuracies in the phase of the reconstructed wavefields may be those which can more likely drive the inversion toward spurious local minima. Accordingly, the objective of this study is to assess whether a reformulation of the parameter-estimation subproblem that mitigates the role of the phase during the early stages of the IR-WRI at the benefit of the more robust amplitude counterpart contributes to stabilizing the inversion.

To achieve this goal, we recast the parameter-estimation subproblem of IR-WRI as a phase retrieval problem (Fienup, 1982). Phase retrieval considers the fundamental problem of how to reconstruct a signal from the magnitude of linear measurements (e.g., the magnitude of its Fourier transform). The name phase retrieval arose because if the phase of the linear measurements can be retrieved, then it is easy to reconstruct the signal by solving a linear problem (Waldspurger et al., 2015). It has a long history of applications in science and engineering when the phase of the linear measurements cannot be recorded or is inaccurate., e.g., optics (Walther, 1963), X-ray crystallography (Millane, 1990; Harrison, 1993), astronomy (Fienup, 1982). Also, it has recently attracted renewed interest in many imaging problems (Fogel et al., 2016), such as X-ray medical imaging (Pfeiffer et al., 2006; Burvall et al., 2011), optical imaging (Shechtman et al., 2015) and seismic processing (Gholami, 2014).

In IR-WRI, the linear operator of the parameter estimation subproblem is formed by the so-called virtual sources (Pratt et al., 1998), while the right-hand sides (the linear measurements) depend on the source and the Laplacian of the wavefields (considering the Helmholtz equation as wave equation). According to the above definition of phase retrieval, we update the subsurface parameters from the magnitude of the right-hand sides rather than from their phase and amplitude. A potential downside of the phase-retrieval algorithm is ill-posedness generated by the rapidly-decreasing sensitivity of the inversion with depth, in particular in large contrast or attenuating media when a small amount of the seismic energy is transmitted in the deep subsurface. This makes the use of efficient regularization necessary, as we will show. As in IR-WRI, we implement the iterative refinement procedure in the phase retrieval algorithm to design an adaptive control on the regularization parameters and achieve a faster convergence to the desired model. Once the velocity model estimation has been stabilized by phase retrieval during a first low-frequency band inversion, we switch to classical IR-WRI to process the higher frequencies with the phase information.

We assess our phase-retrieval based waveform inversion, referred to as WIPR, against two complex synthetic case studies, the Marmousi II and the large contrast 2004 BP salt models

with a sparse long-offset fixed-spread acquisition. We show how WIPR helps to reconstruct more accurately the phase of the reconstructed wavefields, which is translated into more accurate velocity estimation. Then, we show the impact of this more accurate early-stage velocity reconstruction on the convergence of IR-WRI at higher frequencies.

This paper is organized in a method, numerical, and discussion sections. In the method section, we first introduce phase retrieval in a general context. Second, we review the principles of WRI with iterative refinement and show how we implement phase retrieval in the underlying parameter-estimation subproblem. Then, we briefly review how compound regularizations can be implemented in WIPR. In the numerical experiment section, we show the application on the Marmousi and 2004 BP salt models. In the discussion section, we further discuss the resolution power of amplitudes in WIPR and show its impact on the phase reconstruction.

5.3 Notation

The mathematical symbols adopted in this paper are as follows. We use italics for scalar quantities, boldface lower-case letters for vectors, and boldface capital letters for matrices and tensors. We use the superscript T to denote the adjoint of an operator and $*$ to show the complex conjugate. We use diag($\bullet$) to show a square diagonal matrix which includes vector $\bullet$ on its main diagonal. The ith component of the column vector $\mathbf{x}$ is shown by x_i. For a complex number $x = ae^{jb}$ with $j = \sqrt{-1}$, $|x| = a$ denotes the magnitude of x and $\angle x = b$ denotes its phase. For the n-length column vectors $\mathbf{x}$ and $\mathbf{y}$, the dot product is defined by $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{y} = \sum_{i=1}^{n} x_i^* y_i$ and their Hadamard product, denoted by $\mathbf{x} \circ \mathbf{y}$, is another vector made up of their componentwise products, i.e. $(\mathbf{x} \circ \mathbf{y})_i = x_i y_i$. The ℓ_1- and ℓ_2-norms of $\mathbf{x}$ are, respectively, defined by $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$ and $\|\mathbf{x}\|_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_{i=1}^{n} |x_i|^2}$.

5.4 Method

In this section, we briefly review the phase retrieval problem (Gerchberg, 1972) and a simple majorization-minimization (MM) (Lange, 2016) algorithm to solve this non-convex problem. Then we review the frequency-domain FWI (Pratt et al., 1998), its WRI alternative (van Leeuwen and Herrmann, 2016), and its improved version called iteratively refined WRI (IR-WRI) (Aghamiry et al., 2019c). Finally, we introduce phase retrieval in the parameter-estimation subproblem of IR-WRI. In the following of this study, we refer to IR-WRI with phase retrieval as WIPR. WIPR is solved efficiently with the alternating direction method of multipliers (ADMM) (Boyd et al., 2010) and MM algorithms, which provide a suitable framework to implement bound constraints and TT regularization in the parameter estimation subproblem and tackle large data sets.

5.4.1 Phase retrieval

Phase retrieval refers to the reconstruction of a complex signal from the amplitude of linear measurements. Consider the complex-valued linear problem $\mathbf{Lx} = \mathbf{y}$ with $\mathbf{L} \in \mathbb{C}^{m\times n}$, $\mathbf{x} \in \mathbb{C}^{n\times 1}$, and $\mathbf{y} \in \mathbb{C}^{m\times 1}$ and assume that only the amplitude of the right hand side, $|\mathbf{y}|$, can be measured or is reliable. The corresponding phase-retrieval problem can be written as (Gholami, 2014)

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) = \min_{\mathbf{x}} \quad \frac{1}{2}\||\mathbf{Lx}| - |\mathbf{y}|\|_2^2. \tag{5.1}$$

Problem (5.1) is non-convex where this non-convexity finds its root in removing the phase of the right hand side ($|\mathbf{Lx}| = |\mathbf{y}|$ is non-unique and there are many $\mathbf{x}$ that can fit the magnitude of $\mathbf{y}$).
Several algorithms have been proposed to solve the problem ranging from the Gerchberg-Saxton algorithm (Gerchberg, 1972) as a first algorithm to more modern optimization techniques (Candes et al., 2013; Netrapalli et al., 2013; Eldar et al., 2016). In this paper, we present a simple algorithm based on the MM technique (Appendix A, section 5.8) to find a local minimum of $f(\mathbf{x})$ iteratively. Based on the proposed algorithm, the minimizer of the convex surrogate function

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \quad \frac{1}{2}\|\mathbf{Lx} - |\mathbf{y}|e^{j\angle \mathbf{Lx}^k}\|_2^2. \tag{5.2}$$

converges to the minimizer of $f(\mathbf{x})$ as iteration number, k, tends to infinity. The quadratic problem 5.2 is the same as the least-squares norm of the original system $\mathbf{Lx} = \mathbf{y}$, when the phase of $\mathbf{y}$ is replaced with the phase of $\mathbf{Lx}^k$, which is extracted from the solution of k'th iteration.

5.4.2 FWI versus wavefield inversion

The reduced formulation of frequency-domain FWI can be written as (Pratt et al., 1998; Plessix, 2006)

$$\min_{\mathbf{m}} \quad \|\mathbf{Pu}(\mathbf{m}) - \mathbf{d}\|_2^2, \tag{5.3}$$

where $\mathbf{d}$ denotes the recorded seismic data, $\mathbf{P}$ is the linear observation operator that samples the wavefield $\mathbf{u}(\mathbf{m})$ at the receiver positions, and $\mathbf{m}$ denotes the subsurface parameters. The wavefield $\mathbf{u}(\mathbf{m})$ is the solution of the wave equation

$$\mathbf{u}(\mathbf{m}) = \mathbf{A}^{-1}(\mathbf{m})\mathbf{b}, \tag{5.4}$$

where $\mathbf{b}$ is the source term and $\mathbf{A}(\mathbf{m}) \in \mathbb{C}^{n\times n}$ is the discretized wave-equation operator. In this study, the wave equation is the Helmholtz equation, whose operator $\mathbf{A}(\mathbf{m})$ is given by

$$\mathbf{A}(\mathbf{m}) = \mathbf{\Delta} + \omega^2 \mathbf{C}\text{diag}(\mathbf{m})\mathbf{B}, \tag{5.5}$$

where ω is the angular frequency, $\mathbf{m}$ contains the squared slownesses, $\mathbf{\Delta}$ is the discretized Laplace operator, $\mathbf{C}$ introduces boundary conditions (e.g., sponge-like absorbing boundary con-

ditions such as perfectly-matched layers (Bérenger, 1994)) and $\mathbf{B}$ spreads the « mass » term $\omega^2\mathbf{C}\text{diag}(\mathbf{m})$ over all the coefficients of the stencil to improve its accuracy following an anti-lumped mass strategy (Chen et al., 2013).

A main drawback of the reduced formulation in equation (5.3) is that the objective function depends on the model parameters $\mathbf{m}$ through the oscillating inverse operator $\mathbf{A}^{-1}(\mathbf{m})$. This makes the inverse problem highly nonlinear, and hence prone to convergence to inaccurate minimizer when the initial $\mathbf{m}$ is not accurate enough.

An alternative objective function as proposed by van Leeuwen and Herrmann (2016) is defined as

$$\min_{\mathbf{m}} \ \|\mathbf{A}(\mathbf{m})\mathbf{u}(\mathbf{m}) - \mathbf{b}\|_2^2 \tag{5.6}$$

in which the wavefield is given by

$$\mathbf{u}(\mathbf{m}) = \left(\lambda\mathbf{A}(\mathbf{m})^T\mathbf{A}(\mathbf{m}) + \mathbf{P}^T\mathbf{P}\right)^{-1}\left(\lambda\mathbf{A}(\mathbf{m})^T\mathbf{b} + \mathbf{P}^T\mathbf{d}\right). \tag{5.7}$$

The wavefield $\mathbf{u}(\mathbf{m})$ is the solution of an augmented wave equation system, which gathers the wave equation $\mathbf{A}(\mathbf{m})\mathbf{u}(\mathbf{m}) = \mathbf{b}$ weighted by the scalar penalty parameter $\lambda > 0$ and the observation equation $\mathbf{P}\mathbf{u}(\mathbf{m}) = \mathbf{d}$ (van Leeuwen and Herrmann, 2013, their equation 6).

A main advantage of equation (5.6) over equation (5.3) is that its objective function depends on the model parameters $\mathbf{m}$ through a regularized inverse of the wave-equation operator $\left(\lambda\mathbf{A}(\mathbf{m})^T\mathbf{A}(\mathbf{m}) + \mathbf{P}^T\mathbf{P}\right)^{-1}$, where the regularizer injects the prior information on the true wavefields, namely their sparse measurements $\mathbf{d}$, through the sampling operator $\mathbf{P}$. This wavefield reconstruction driven by the observations introduces a relaxation of the wave equation, namely source residuals, which are minimized to update the model parameters, equation 5.6, such that the wavefields, equation 5.7, are pushed back toward the wave equation.

The minimization problem, equation (5.6), can be solved with a variable projection approach by enforcing the closed-form expression of $\mathbf{u}$, equation 5.7, in the objective function, equation 5.6 (van Leeuwen and Herrmann, 2016), or with an alternating-direction strategy for $\mathbf{u}$ and $\mathbf{m}$ through a Gauss-Seidel like iteration (van Leeuwen and Herrmann, 2013). We will follow the latter option, where the model estimation problem, equation (5.6), reduces to a quadratic optimization problem

$$\mathbf{m}^{k+1} = \arg\min_{\mathbf{m}} \ \|\mathbf{A}(\mathbf{m})\mathbf{u}^k - \mathbf{b}\|_2^2, \tag{5.8}$$

which, in virtue of the wave-equation bilinearity (van Leeuwen and Herrmann, 2013; Aghamiry et al., 2019c), can be recast as

$$\mathbf{m}^{k+1} = \min_{\mathbf{m}} \ \|\mathbf{L}(\mathbf{u}^k)\mathbf{m} - \mathbf{y}(\mathbf{u}^k)\|_2^2, \tag{5.9}$$

with

$$\begin{cases} \mathbf{L}(\mathbf{u}^k) = \omega^2\mathbf{C}\text{diag}(\mathbf{B}\mathbf{u}^k), \\ \mathbf{y}(\mathbf{u}^k) = \mathbf{b} - \Delta\mathbf{u}^k. \end{cases} \tag{5.10}$$

and $\mathbf{u}^k \equiv \mathbf{u}(\mathbf{m}^k)$.

During wavefield reconstruction, equation 5.7, a small value of λ is used during the early iterations to foster the data fit, hence mitigating cycle skipping, at the expense of the fidelity with which the wave equation is satisfied. Then, λ is progressively increased in iterations such that the wave equation is satisfied at the convergence point (Fu and Symes, 2017). To avoid the tedious dynamic control of penalty parameters and converge in an automatic way toward accurate solution with fixed λ, Aghamiry et al. (2019c) introduced iteratively-refined WRI (IR-WRI). IR-WRI relies on an iterative refinement procedure, which finds its root in augmented Lagrangian method (Nocedal and Wright, 2006, Chapter 17). In the end, it simply consists of updating at the end of each iteration the right-hand sides of the wave equation and the observation equation, namely the data $\mathbf{d}$ and the source $\mathbf{b}$, with the running sum of the data and source residuals in iteration

$$\mathbf{b}^{k+1} = \mathbf{b}^{k} + \mathbf{b} - \mathbf{A}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}, \tag{5.11a}$$

$$\mathbf{d}^{k+1} = \mathbf{d}^{k} + \mathbf{d} - \mathbf{P}\mathbf{u}^{k+1}. \tag{5.11b}$$

In the framework of linear inverse problems, $\mathbf{d}^{k+1}$ and $\mathbf{b}^{k+1}$ record the wavefield and model refinements performed at previous iterations to refine the wavefields and the model parameters at the current iteration from the residual errors only. This iterative refinement or defect correction (Böhmer and Stetter, 1984) procedure, which is lacking in WRI, is the key ingredient which controls the accuracy of the wavefield and model solutions at the convergence point when fixed λ is used. The reader is also referred to Gholami et al. (2018); Gholami and Naeini (2019) for other recent applications of iterative refinement in the field of geophysics.

IR-WRI has shown promising results even for complicated velocity models when equipped with frequency continuation strategies and appropriate regularization (Aghamiry et al., 2019b, 2020b). However, when the initial model is far from the true model, the phase of the reconstructed wavefields becomes inaccurate in areas that are located far away from the receivers. These phase inaccuracies can lead to an inaccurate model reconstruction of these regions, hence trapping the inversion into a local minimum. In order to mitigate this issue, we show in the next section how to solve the quadratic optimization problem, equation 5.9, with phase retrieval.

5.4.3 Wavefield inversion with phase retrieval (WIPR)

To estimate the model parameters without involving inaccurate phase information, we propose to replace the IR-WRI parameter-estimation problem given by equation 5.9 by the following phase retrieval problem

$$\mathbf{m}^{k+1} = \arg\min_{\mathbf{m}} \; \| |\mathbf{L}(\mathbf{u}^{k})\mathbf{m}| - |\mathbf{y}(\mathbf{u}^{k})| \|_2^2. \tag{5.12}$$

In order to solve (5.12), we replace it by a surrogate majorizing function (as described in section 5.4.1 and Appendix A (section 5.8)

$$\mathbf{m}^{k+1} = \arg\min_{\mathbf{m}} \; \| \mathbf{L}(\mathbf{u}^{k})\mathbf{m} - |\mathbf{y}(\mathbf{u}^{k})| e^{j\angle \mathbf{L}(\mathbf{u}^{k})\mathbf{m}^{k}} \|_2^2. \tag{5.13}$$

This objective function is now quadratic and admits a closed form solution.

In order to better understand why phase retrieval helps to improve the solution, we give a tentative interpretation hereafter. From equation (5.7) it is seen that, at each iteration, the wavefield approximately satisfies the wave equation

$$\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k \approx \mathbf{b} \tag{5.14}$$

or

$$\mathbf{L}(\mathbf{u}^k)\mathbf{m}^k \approx \mathbf{y}(\mathbf{u}^k), \tag{5.15}$$

where the approximation level is controlled by the parameter λ and the approximation becomes an equality as $\lambda \to \infty$. The next iterate $\mathbf{m}^{k+1}$ is then found such that it minimizes the wave-equation errors generated by the wave-equation relaxation, equation (5.8). In this way, simple minimization via equation (5.8) forces the model to match both the amplitude and phase information of $\mathbf{y}(\mathbf{u}^k)$. In order to reduce the imprint of the phase errors, a phase correction or phase alignment step can be applied to equation (5.15) before updating the model parameters:

$$\min_{\phi} \|\mathbf{L}(\mathbf{u}^k)\mathbf{m}^k - |\mathbf{y}(\mathbf{u}^k)|e^{j\phi}\|_2^2. \tag{5.16}$$

The optimal minimizer of equation (5.16) is given by $\phi = \angle\mathbf{L}(\mathbf{u}^k)\mathbf{m}^k$. Using the optimal ϕ, the next iterate is found by minimizing the corrected quadratic problem, as presented in equation (5.13). In this way, the solution of the inverse problem is less affected by the phase errors in $\mathbf{u}$ and thus hopefully less prone to convergence to a local minimum. The reader is referred to Jiang et al. (2017); Qian et al. (2017) for other applications of phase retrieval with alternating optimization.

Furthermore, we apply the iterative refinement procedure to WIPR in the same way as for IR-WRI by updating $\mathbf{b}$ and $\mathbf{d}$ according to equations 5.11a and 5.11b as outlined in the Algorithm 5. This implies that, although we recast the parameter-estimation subproblem as a phase retrieval problem (Algorithm 5, Line 5), we use both the phase and amplitude of the source and data residuals to update the right-hand sides of the wavefield-reconstruction and parameter-estimation subproblems (Algorithm 5, Lines 8 and 9).

5.4.4 Tikhonov-Total variation (TT) regularized WIPR with bounding constraints

In this section, we implement TT regularization and bounding constraints in WIPR to stabilize the updates of the model parameters. By adding a convex regularization term $\|\mathbf{m}\|_{\text{TT}}$ to the cost function of WIPR, we get the following optimization for the $\mathbf{m}$-subproblem

$$\mathbf{m}^{k+1} = \underset{\mathbf{m}\in\mathcal{C}}{\arg\min}\ \|\mathbf{m}\|_{\text{TT}} + \lambda\|\mathbf{L}(\mathbf{u}^k)\mathbf{m} - |\mathbf{y}(\mathbf{u}^k)|e^{j\angle\mathbf{L}(\mathbf{u}^k)\mathbf{m}^k}\|_2^2, \tag{5.17}$$

where $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^{n\times 1} \mid \mathbf{m}_l \leq \mathbf{x} \leq \mathbf{m}_u\}$ is the set of all feasible models bounded by the lower bound $\mathbf{m}_l$ and the upper bound $\mathbf{m}_u$ and $\|\mathbf{m}\|_{\text{TT}}$ is the TT regularization functional as defined by

Algorithm 5: WIPR algorithm

Input: $\mathbf{m}^0$
initialize: $\mathbf{A} \leftarrow \boldsymbol{\Delta} + \omega^2 \mathbf{C}\text{diag}(\mathbf{m}^0)\mathbf{B}$, $\mathbf{b}^0 \leftarrow \mathbf{0}$, $\mathbf{d}^0 \leftarrow \mathbf{0}$
while convergence conditions not satisfied **do**
  $\mathbf{u}^{k+1} \leftarrow \left[\mathbf{P}^T\mathbf{P} + \lambda\mathbf{A}^T\mathbf{A}\right]^{-1}\left[\mathbf{P}^T(\mathbf{d}+\mathbf{d}^k) + \lambda\mathbf{A}^T(\mathbf{b}+\mathbf{b}^k)\right]$
  $\mathbf{L} \leftarrow \omega^2\mathbf{C}\text{diag}(\mathbf{B}\mathbf{u}^{k+1})$
  $\tilde{\mathbf{y}} \leftarrow |\mathbf{b}+\mathbf{b}^k - \boldsymbol{\Delta}\mathbf{u}^{k+1}|e^{j\angle\mathbf{L}\mathbf{m}^k}$
  $\mathbf{m}^{k+1} \leftarrow \left[\mathbf{L}^T\mathbf{L}\right]^{-1}\mathbf{L}^T\tilde{\mathbf{y}}$
  $\mathbf{A} \leftarrow \boldsymbol{\Delta} + \omega^2\mathbf{C}\text{diag}(\mathbf{m}^{k+1})\mathbf{B}$
  $\mathbf{b}^{k+1} \leftarrow \mathbf{b}^k + \mathbf{b} - \mathbf{A}\mathbf{u}^{k+1}$
  $\mathbf{d}^{k+1} \leftarrow \mathbf{d}^k + \mathbf{d} - \mathbf{P}\mathbf{u}^{k+1}$
end while

Aghamiry et al. (2019b, 2020b)

$$\|\mathbf{m}\|_{\text{TT}} = \min_{\mathbf{m}=\mathbf{m}_1+\mathbf{m}_2} \|\mathbf{m}_1\|_{\text{TV}} + \alpha\|\mathbf{m}_2\|_{\text{Tikh}}, \tag{5.18}$$

with

$$\begin{cases} \|\mathbf{m}\|_{\text{TV}} = \sum \sqrt{|\nabla_x\mathbf{m}|^2 + |\nabla_z\mathbf{m}|^2}, \\ \|\mathbf{m}\|_{\text{Tikh}} = \sum \left(|\nabla_{xx}\mathbf{m}|^2 + 2|\nabla_{xz}\mathbf{m}|^2 + |\nabla_{zz}\mathbf{m}|^2\right), \end{cases} \tag{5.19}$$

in which the sum runs over all element, ∇_i are first order difference operators in direction i and ∇_{ij} are second-order differential operators in directions i and j.

The TT regularization is an infimal convolution-based combination of the second-order Tikhonov and TV regularizations and is suitable for recovering piecewise-smooth models (Gholami and Hosseini, 2013; Aghamiry et al., 2020b). It decomposes the model $\mathbf{m}$ explicitly into two components, $\mathbf{m}_1$ and $\mathbf{m}_2$, which have different statistical properties. Here, the blocky component $\mathbf{m}_1$ and the smooth component $\mathbf{m}_2$ are captured by the TV and the Tikhonov regularizations, respectively.

Subproblem (5.17) is solved with ADMM and the split-Bregman scheme, which de-couples the ℓ_1 and ℓ_2 components and bound constraints of the function through the introduction of auxiliary variables and solves each related subproblem in sequence (Goldstein and Osher, 2009). By doing this, we come up with a least-squares problem with a closed-form expression to update $\mathbf{m}_1$ and $\mathbf{m}_2$, a proximity subproblem to update the auxiliary variables and a gradient ascent step to update dual variables. The reader is referred to Aghamiry et al. (2020b) and Aghamiry et al. (2019b, Section 2.2.2) for the detailed derivation of the algorithm.

5.5 Numerical examples

We assess our workflow against the Marmousi II model (Martin et al., 2006) and the large-contrast 2004 BP salt model (Billette and Brandsberg-Dahl, 2004) when the inversion is started

from a homogeneous velocity model and a 3 Hz frequency such that a considerable phase error is generated. In both cases, we consider a long-offset fixed-spread acquisition and perform the frequency-domain inversion in the 3 Hz - 13 Hz frequency band with a frequency interval of 0.5 Hz. Small batches of two frequencies with one frequency overlapping between two consecutive batches are successively inverted following a classical frequency continuation strategy. We apply the phase retrieval algorithm only during the first frequency batch processing to stabilize the inversion. Then, we proceed with regular IR-WRI at higher frequencies. The IR-WRI results obtained from an early WIPR stage will be referred to as IR-WRI$_{pr}$ to prevent confusions with results obtained when IR-WRI is applied from the first frequency-batch inversion. During IR-WRI/IR-WRI$_{pr}$, we perform three paths through the frequency batches to improve the inversion results, using the final model of one path as the initial model of the next one (these cycles can be viewed as outer iterations of the algorithms). The starting and finishing frequencies of the paths are [3.5, 6], [4, 8.5], [6, 13] Hz, respectively, where the first element of each pair shows the starting frequency and the second one is the finishing frequency. We will compare the results of IR-WRI$_{pr}$ and IR-WRI to assess the improvements resulting from the initial WIPR step. We will also compare the results of IR-WRI and IR-WRI$_{pr}$ when regularization is applied or not such that the relative role of phase retrieval and regularization can be discriminated. Namely, we will illustrate that the role of phase retrieval is to extend the linear regime of the waveform inversion further when an inaccurate starting model is used, while regularization injects suitable prior to manage non-uniqueness of the solution in the poorly illuminated area. When regularization is applied, the same tuning is used during WIPR and IR-WRI, and during IR-WRI and IR-WRI$_{pr}$. For all the numerical tests, we use a 9-point finite-difference staggered-grid stencil with PML boundary condition along the four edges of the model (no free-surface boundary condition is used) and anti-lumped mass to solve the Helmholtz equation, where the stencil coefficients are optimized to the frequency (Chen et al., 2013).

5.5.1 Marmousi II model

The Marmousi II model covers a $11500\ m \times 4200\ m$ spatial domain (Fig. 5.1a). The fixed-spread surface acquisition consists of 56 sources spaced 200 m apart, and 230 receivers spaced 50 m apart at the surface. The source signature is a 10 Hz Ricker wavelet, and the wavespeed in the homogeneous starting model is 3 km/s.

We first compare the results of WIPR and IR-WRI after 45 iterations of the [3,3.5] Hz frequency batch inversion when bound constraints are applied without any additional regularization (Fig. 5.2a-b). Bound constraints are used for all of the tests of this section from the first iteration. WIPR manages to reconstruct an accurate long-wavelength approximation of the Marmousi model (Fig. 5.2b), while IR-WRI clear fails to capture the kinematic trend of the model (Fig. 5.2a). This statement is clearly illustrated by the direct comparison between the true model, the initial model, and the reconstructed IR-WRI and WIPR models along with three vertical logs at horizontal distances of 4, 8.2, and 10 km (Fig. 5.3a). We also apply TT regularization during WIPR and IR-WRI (Figures 5.2c-d and 5.3b). The TT regularization improves the WIPR and IR-WRI results in the poorly-illuminated area, although it does not remove high-velocity artefacts near the left-bottom end of the IR-WRI velocity model (Fig. 5.3b, x=4 km).

Since the model parameters are updated from the reconstructed wavefields, the accuracy of

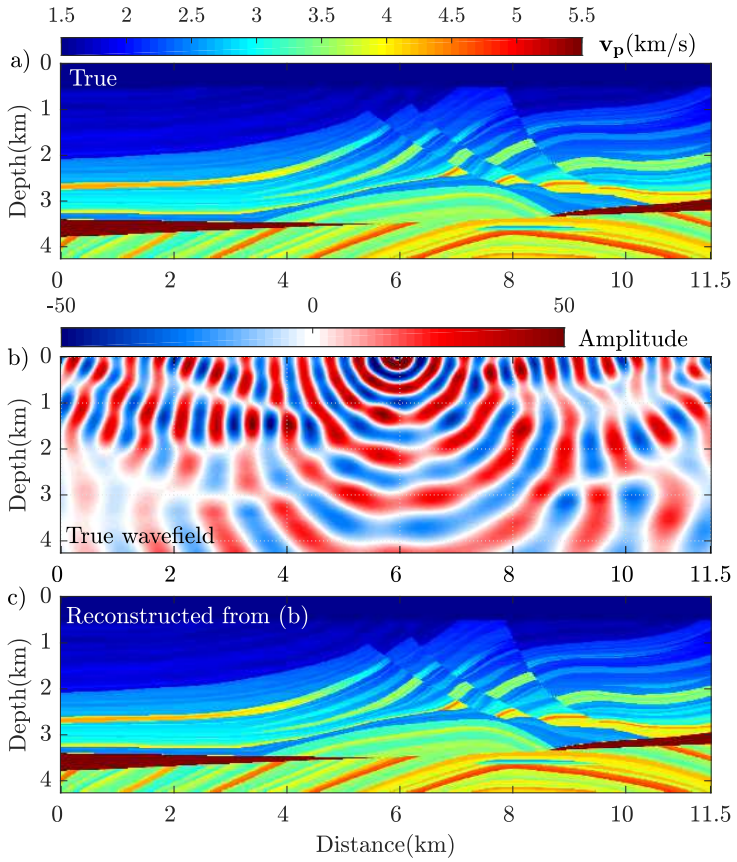

Figure 5.1 – Illustration of the wave-equation bilinearity. (a) True Marmousi II velocity model. (b) Three-Hz wavefield (real part) for the source located at the surface and a distance of 6 km. (c) Velocity model inferred from the amplitude and phase of the monochromatic wavefield, equation 5.24. The same model can be inferred from the sole amplitude of the monochromatic wavefield, equation 5.25. The reconstructed velocity model exactly matches the true velocity model.

these wavefields is a visible indicator of the reliability of the estimated model parameters. We compare the real part of the 3 Hz wavefields reconstructed by IR-WRI and WIPR without and with TT regularization (Fig. 5.4) with the true wavefield (Fig. 5.1b) for a source located at the horizontal distance of 6 km. Both IR-WRI and WIPR wavefields are in phase with the true wavefield along the receiver line since we tune the penalty parameter such that the data are matched from the first iteration. However, the WIPR wavefield matches much better the true wavefield than the IR-WRI counterpart in depth. When regularization is used, the accuracy of the wavefield reconstruction is mostly improved in the poorly-illuminated zones, mainly near the bottom ends of the model.

We continue with IR-WRI at higher frequencies using the final IR-WRI and WIPR models of the [3, 3.5] Hz inversion (Fig. 5.2) as initial models. The stopping criterion of iterations is a

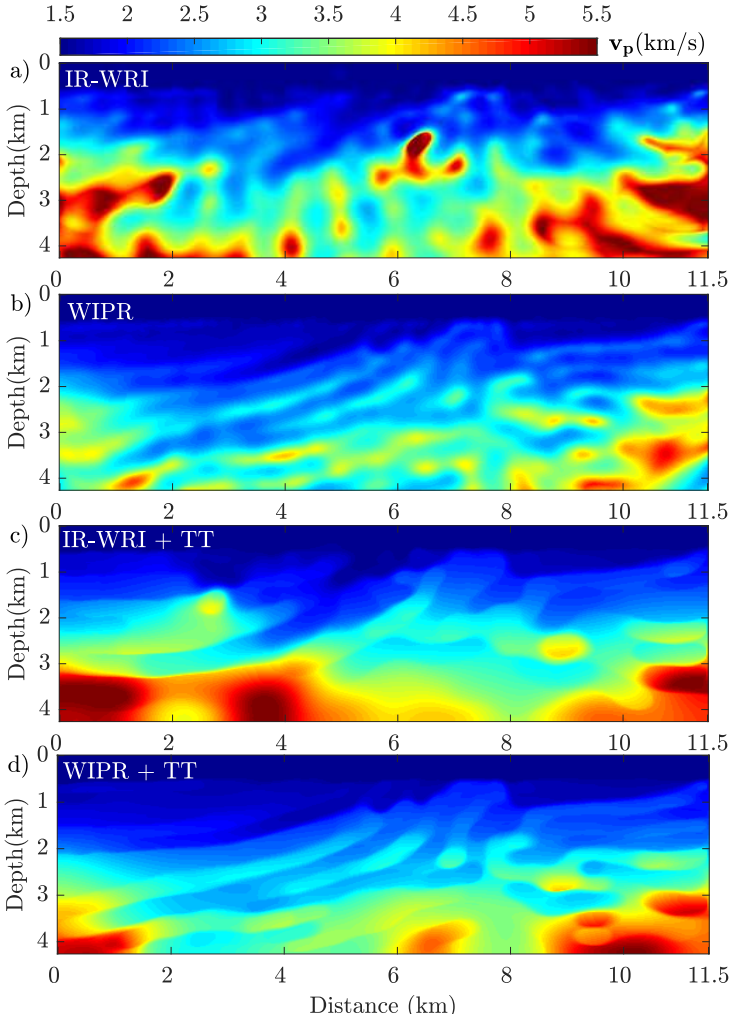

Figure 5.2 – Marmousi II example. Velocity models after the {3,3.5} Hz inversion. (a) IR-WRI, (b) WIPR, (c) TT regularized IR-WRI, and (d) TT regularized WIPR. Bound constraints are used for all of the tests from the first iteration.

maximum of 30 iterations per batch or

$$(\|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b}\|_F \leq 10^{-3} \quad \text{and} \quad \|\mathbf{P}\mathbf{u}^k - \mathbf{d}\|_F \leq 10^{-5}), \tag{5.20}$$

where F refers to the Frobenius norm. The final IR-WRI and IR-WRI$_{pr}$ velocity models are shown in Fig. 5.5, while a direct comparison between them and the true model are plotted in Fig. 5.6 at distances 4, 8.2 and 10 km. IR-WRI performed a total of 382 and 369 iterations without and with TT regularization, respectively (Fig. 5.5a,c), while IR-WRI$_{pr}$ performed a total of 201 and 234 iterations (Fig. 5.5b,d). The smaller number of iteration required to satisfy the stopping criterion of iterations performed by IR-WRI$_{pr}$ compared to IR-WRI highlights the role of the initial model built by WIPR during the [3, 3.5] Hz inversion (Fig. 5.2b,d). Without regularization, IR-WRI$_{pr}$ outperforms IR-WRI in the deep part, namely near the reservoir at around 8 km distance and near the salt layers near the ends of the models. Regularization

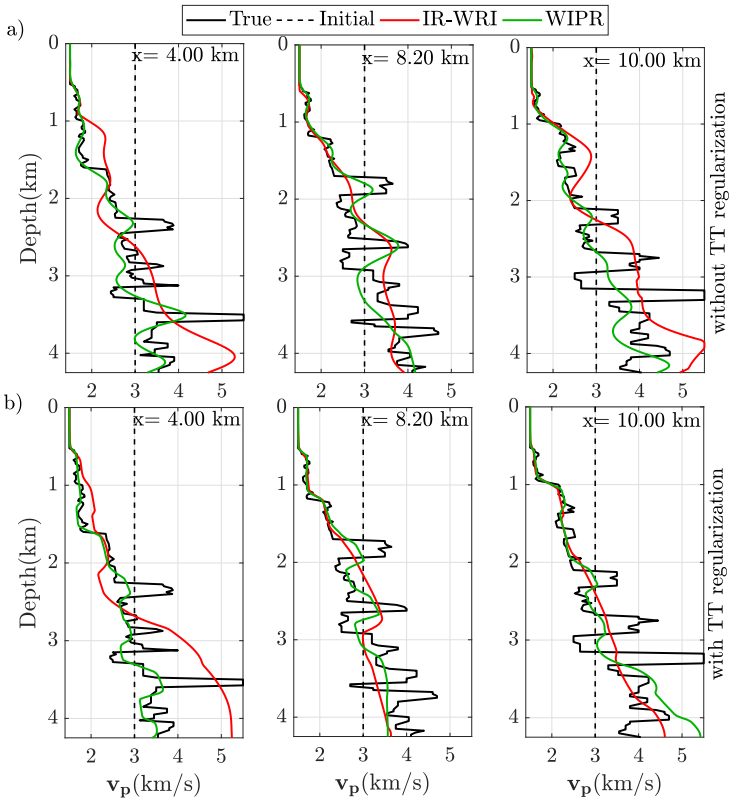

Figure 5.3 – Marmousi II example. Direct comparison along with three vertical logs at $x = 4, 8.2$, and $10\ km$ between the true velocity model (black), the initial model (dashed black), and the IR-WRI (red) and WIPR (green) models shown in Fig. 5.2. (a) Without TT regularization. (b) With TT regularization.

improves significantly both IR-WRI and IR-WRI$_{pr}$ results in these areas. However, IR-WRI$_{pr}$ still provides superior results in the deep part.

5.5.2 The large contrast 2004 BP salt model

We consider now a target of the large contrast 2004 BP salt model located on the left side of the model (Fig. 5.7a). The 2004 BP salt model mainly consists of a simple sediment background with a complex rugose multi-valued salt body, sub-salt slow velocity anomalies related to over-pressure zones, and a fast velocity anomaly to the right of the salt body. The selected subsurface model is 16250 m wide and 5825 m deep and is discretized with a 25 m grid interval. The surface fixed-spread acquisition consists of 66 sources spaced 250 m apart, and 131 receivers spaced 125 m apart. We perform forward modelling using a 10 Hz Ricker wavelet as a source signature. The starting velocity model for inversion is homogeneous with a velocity of 3 km/s.

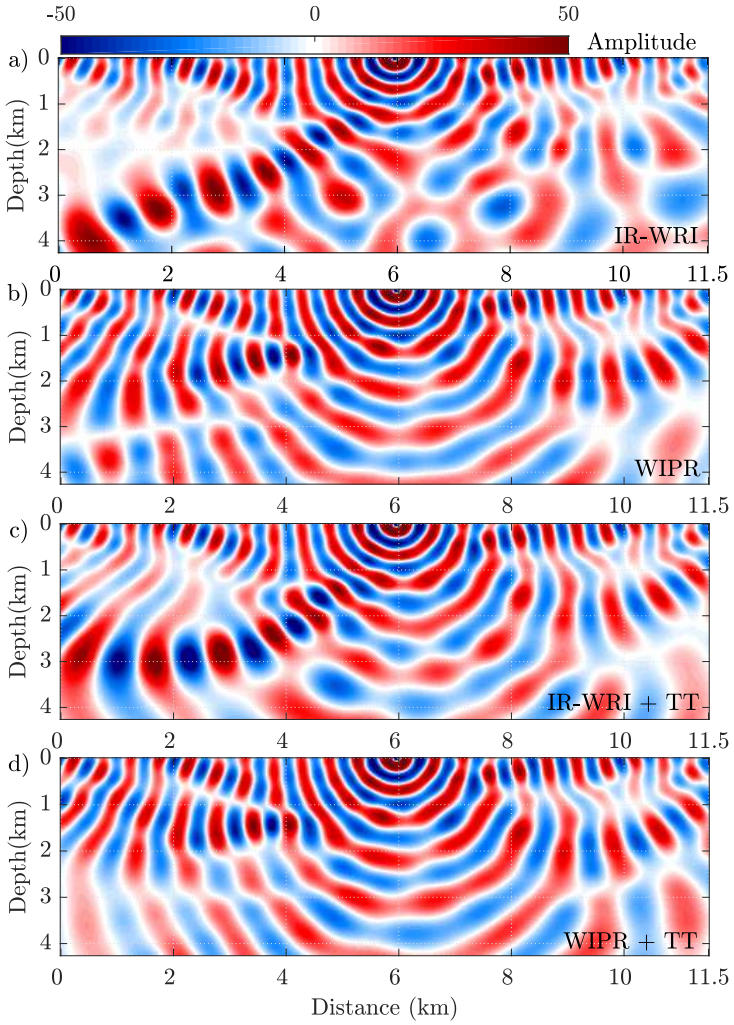

Figure 5.4 – Marmousi II example. Three-Hz wavefields (real part) reconstructed after the first frequency batch inversion for the source located at 6 km. (a) IR-WRI, (b) WIPR, (c) TT regularized IR-WRI, and (d) TT regularized WIPR. The corresponding velocity models are shown in Fig. 5.2. These wavefields can be compared with the true one shown in Fig. 5.1b.

IR-WRI versus WIPR at low frequencies

We first perform 45 iterations of bound-constrained WIPR and IR-WRI for a starting [3,3.5] Hz frequency batch. Note that we turn on the bound constraints after 20 iterations to discriminate their role from those of the TT regularization and optimization scheme (see later discussion in the text). The velocity models reconstructed by bound-constrained IR-WRI and WIPR with and without TT regularization are shown in Fig. 5.8, while the direct comparison between the true model, the initial model, and the models reconstructed by IR-WRI and WIPR are shown in Fig. 5.9 along three vertical logs at horizontal distances of 3.5 km, 7.7 km, and 12 km. The most striking feature is that IR-WRI overestimates velocities above the salt and mispositions the top of the salt layers accordingly, while WIPR captures the top salt much more accurately (Fig. 5.9a, green versus red curves, x=3.5 km and 7.7 km). This overestimation of the shallow

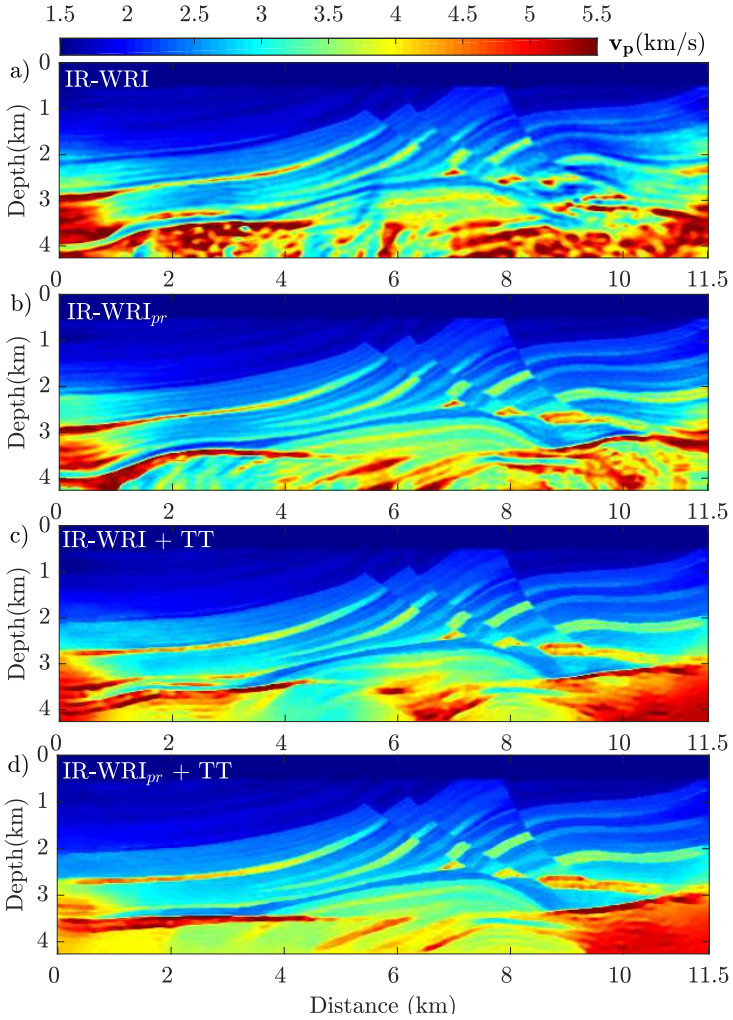

Figure 5.5 – Marmousi II example. Final velocity models obtained by IR-WRI (a,c) and IR-WRI$_{pr}$ (b,d). (a,b) Without TT regularization. (c,d) With TT regularization.

velocities by IR-WRI is also shown in the right part of the model away from the salt (Fig. 5.9a, red curve, x=12 km). On the other hand, WIPR fails to reconstruct the bottom part of the thick salt body and overprints unrealistic variations on the subsalt area when TT regularization is not used (Fig. 5.8b and Fig. 5.9a, green curve, x=3.5 km). These artefacts are likely generated by the increased ill-posedness of WIPR with depth, which itself results from the lack of phase information. This ill-posedness is however, efficiently mitigated by the TT regularization (Fig. 5.8d). The TV component of the TT regularization helps to sharpen the top and bottom of the salt layers (Fig. 5.9b, green curve, x=7.7km), recovers better the geometry of the salt layer (Fig. 5.9b, green curve, x=3.5km), while the Tikhonov component stabilizes the inversion in-depth and drives it toward smooth reconstruction of the poorly-illuminated subsalt structures (Fig. 5.9b, green curve, x=12km). During IR-WRI, TT regularization fails to achieve these goals (Fig. 5.8c) because the mispositioning in depth of the velocity structures generated by the increased nonlinearity of IR-WRI are too heavy to be corrected by the regularization (Fig. 5.9b, red curve).

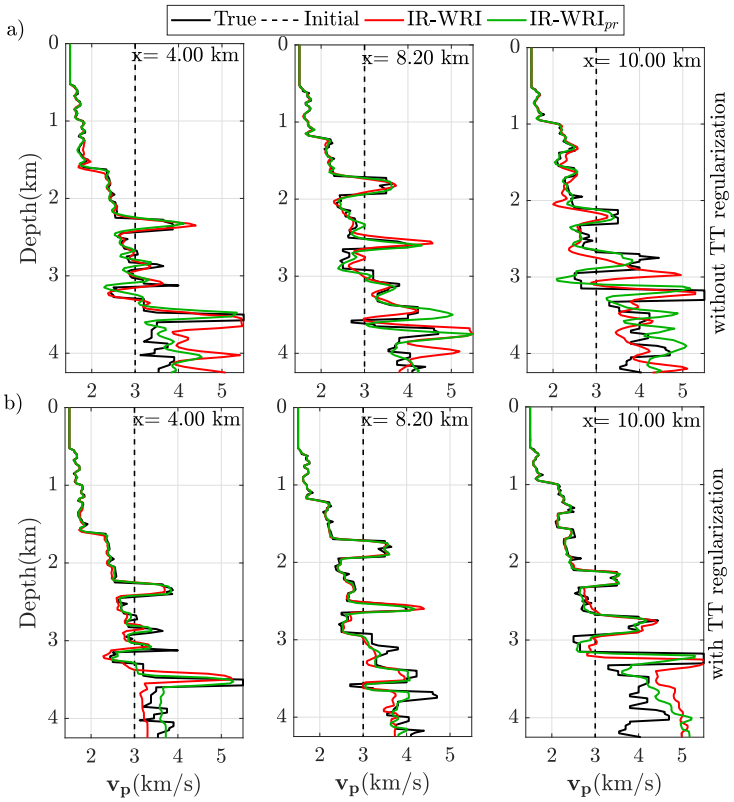

Figure 5.6 – Marmousi II example. Direct comparison at $x = 4, 8.2$, and 10 km between the true velocity model (black), the initial model (dashed line), and the models estimated by IR-WRI (red) and IR-WRI$_{pr}$ (green) (Fig. 5.5). (a) Without TT regularization. (b) With TT regularization.

We compare the true 3-Hz wavefield shown in Fig. 5.7b with the wavefields reconstructed by IR-WRI and WIPR without and with TT regularization (Fig. 5.10). Like the Marmousi test, both IR-WRI and WIPR wavefields are in phase with the true wavefield along the receiver line since we tune the penalty parameter such that the data are matched from the first iteration. However, the WIPR wavefields match much better the true wavefield than the IR-WRI counterpart in the first two kilometers of the model. This is pointed by the arrows in Fig. 5.10, which show that the IR-WRI wavefield transmitted into the salt body is out-of-phase with respect to the true wavefield, unlike the WIPR wavefield.

We also computed time-domain seismograms in the TT-regularized IR-WRI and WIPR models (Fig. 5.8c,d) with a 10-Hz Ricker wavelet for a source located at 15.8 km (Fig. 5.11). We show quite significant travel-time mismatches at long offsets between the true seismograms and those computed in the IR-WRI model (Fig. 5.11a), which result from the overestimated velocities above the salt shown in Fig. 5.8c, while the match between the true seismograms and those computed in the WIPR model is far better. This poor data fit achieved by IR-WRI highlights both the inaccuracies of the reconstructed velocity model and the limited accuracy with which the wave equation has been satisfied at the convergence point.

This is further supported by Fig. 5.12, which shows the relative model error (Fig. 5.12a) and the joint evolution of the data misfit ($\|\mathbf{P}\mathbf{u}^k - \mathbf{d}\|_2$) and wave-equation error ($\|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b}\|_2$)

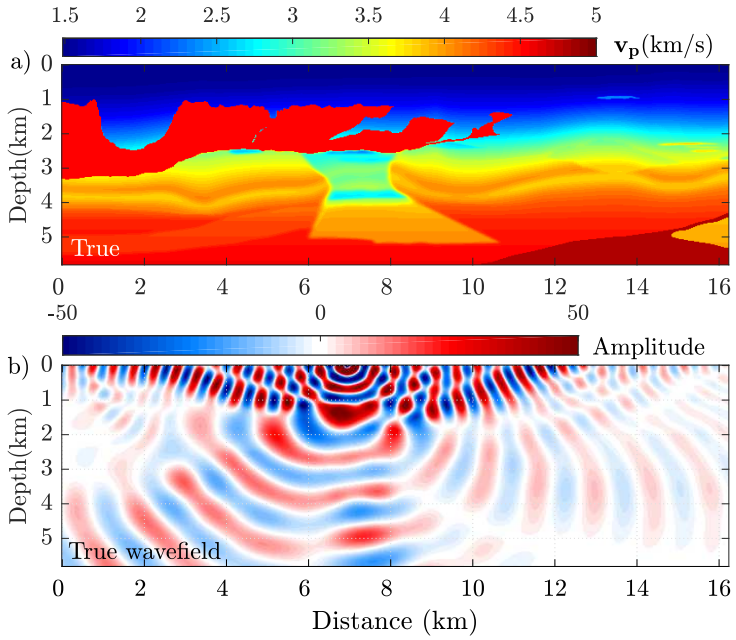

Figure 5.7 – (a) True 2004 BP salt velocity model. (b) Real part of the 3-Hz wavefield computed in (a) for the source located at 7 km.

in iterations (Fig. 5.12b). Also, the evolution of the data misfit and wave-equation error in iterations are shown separately in Figs. 5.12c and 5.12d, respectively. First, these error curves show that WIPR converges toward a more accurate velocity model than IR-WRI (Fig. 5.12a). Accordingly, it shows that WIPR satisfies more accurately the wave-equation constraint than IR-WRI (Fig. 5.12d), although the parameter estimation has been recast as a phase retrieval problem rather than as a wave-equation error minimization. Fig. 5.12a also shows that bound constraints have the least impact on the TT regularized WIPR (blue curve) since the descent direction is rather continuous around iteration 20 (when bound constraints are activated) . This highlights the more stable descent direction followed by TT regularized WIPR relative to IR-WRI.

IR-WRI versus IR-WRI$_{pr}$

We continue with IR-WRI at higher frequencies using the final IR-WRI and WIPR models of the [3, 3.5] Hz inversion (Fig. 5.8) as initial models. The stopping criterion of iterations is 10 iterations per batch. The final velocity models built by IR-WRI and IR-WRI$_{pr}$ are shown in Fig. 5.13, while a direct comparison between them and the true model are plotted in Fig. 5.14 at distances 3.5 km, 7.7 km and 12 km.

Let's first focus on the results obtained without TT regularization (Fig. 5.13a,b). IR-WRI apparently reconstructs the salt body with good accuracy (Fig. 5.13a), although the shallow kinematic inaccuracies of the starting IR-WRI model shown in Figs. 5.8a and 5.9, red curves. This highlights the potential of the search-space expansion implemented in IR-WRI to manage cycle skipping. The final IR-WRI$_{pr}$ velocity model shows low-velocity artefacts in the salt

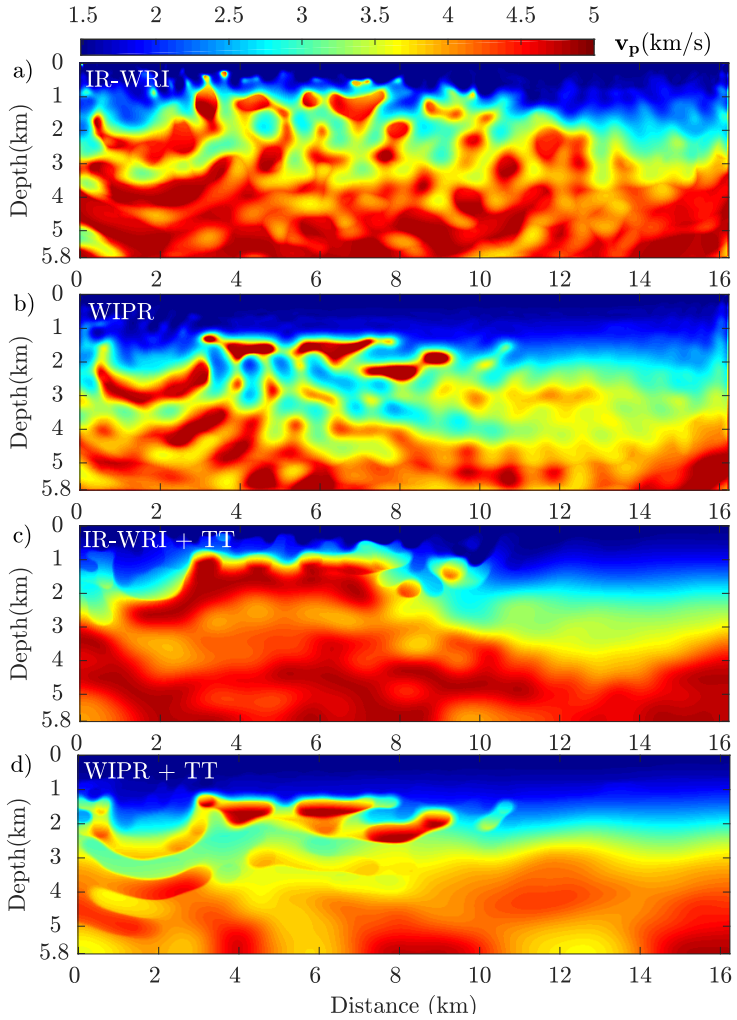

Figure 5.8 – BP salt example. Velocity models inferred from the inversion of the {3,3.5} Hz frequencies. (a) IR-WRI, (b) WIPR, (c) IR-WRI with TT regularization, and (d) WIPR with TT regularization.

body at around 3 km distance (Fig. 5.13b), which are probably inherited from the underestimated velocities below the top of the salt shown in Figs. 5.8b and 5.9a, green curve, x=3.5 km. For both inversions, the reconstruction of the subsalt structures are quite noisy. To gain more quantitative insights on the relative accuracy of the IR-WRI and IR-WRI$_{pr}$ models, we show the difference between these velocity models and the true model in Fig. 5.15(a-b) and outline the model error (ME) for the velocity models as well as the least-squares data and source misfit at the convergence point in Table 5.1. Figure 5.15(a-b) clearly shows that the IR-WRI$_{pr}$ model is as a whole more accurate than the IR-WRI counterpart, although the above-mentioned local artefacts. The TT regularization removes the local artefacts shown in the salt body of the IR-WRI$_{pr}$ model and improves the subsalt imaging (Fig. 5.13d) more significantly than for the IR-WRI model (Fig. 5.13c). The difference between the true model and the IR-WRI/IR-WRI$_{pr}$ models (Fig. 5.15(c-d)) and the model error (ME) provided in Table 5.1 outline the superior accuracy of the IR-WRI$_{pr}$ model relative to the IR-WRI counterpart. These relative model ac-

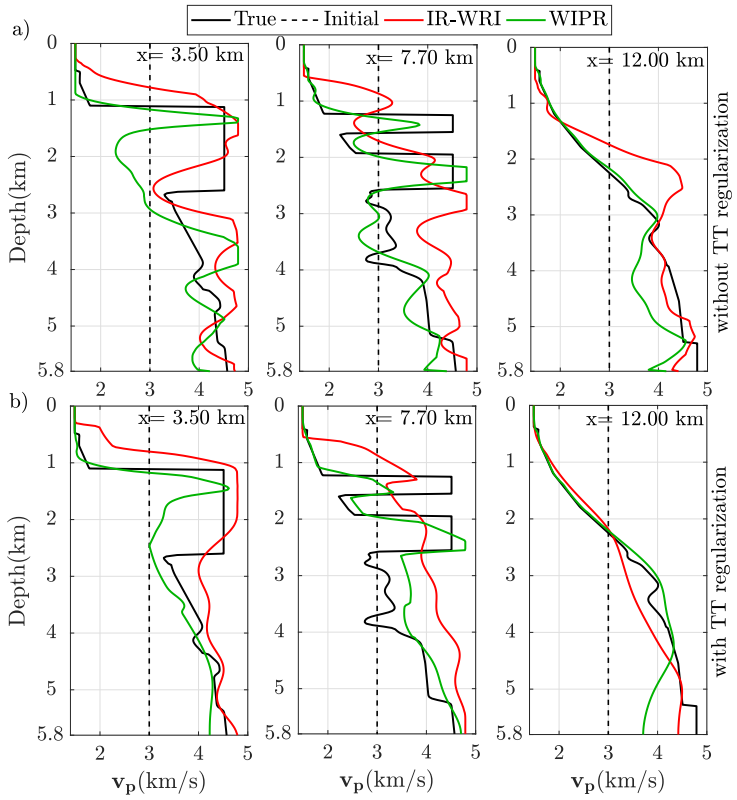

Figure 5.9 – BP salt example. Direct comparison along logs located at $x = 3.5, 7.7$ and 12 km between the true velocity model (black), the initial model (dashed line) and the models estimated by IR-WRI (red) and WIPR (green) shown in Fig. 5.8. (a) Without TT regularization. (b) With TT regularization.

curacies are consistent with the data and source misfit achieved by IR-WRI$_{pr}$ and IR-WRI at the convergence point (Table 5.1).

5.6 Discussion

With the large contrast BP salt model, we have shown that the phase inaccuracies of the reconstructed wavefields prevent accurate estimation of the shallow sediments and accurate positioning in depth of the top of the salt after the first frequency batch inversion. In contrast, WIPR reconstructs more accurately the velocity gradient in the sedimentary cover as well as the sharp velocity contrast on top of the salt. This accurate positioning of the top salt is indeed important to speed up the convergence and improve the solution at higher frequencies.
To further illustrate the sensitivity of IR-WRI to wavefield amplitudes and phases in large contrast media, we show in Figs. 5.16 and 5.17 the amplitude and phase of $\Delta\mathbf{u}$ for the 3 Hz frequency and a source located at 7 km of distance. In these figures, $\Delta\mathbf{u}$ have been computed in the true velocity model, and in the TT regularized IR-WRI and WIPR velocity models obtained after the first frequency batch inversion (Figs. 5.8c,d). We remind that $\Delta\mathbf{u}$ is one of the terms

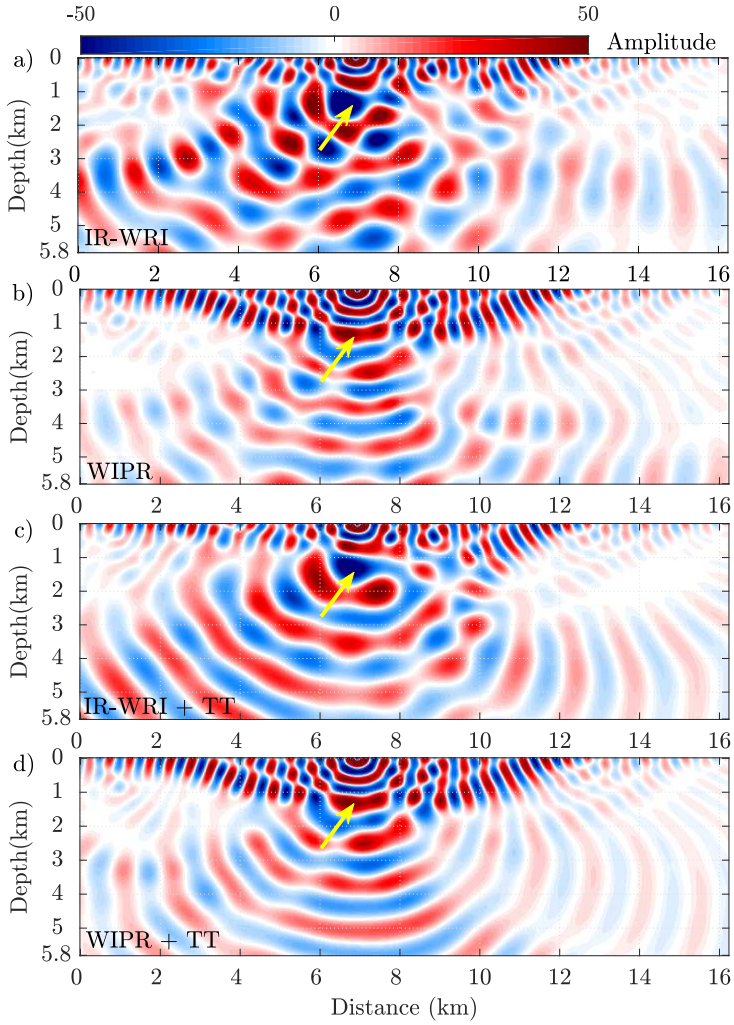

Figure 5.10 – BP salt example. Three-Hz wavefields (real part) reconstructed after the first frequency batch inversion. (a) IR-WRI, (b) WIPR, (c) TT regularized IR-WRI, and (d) TT regularized WIPR. The corresponding velocity models are shown in Fig. 5.8. The source is located at 7 km distance. These wavefields can be compared with the true wavefield shown in Fig. 5.7b. The arrows point where the incident wavefields interact with the top salt.

forming the right-hand side $\tilde{\mathbf{y}}$ of the quadratic parameter-estimation subproblem together with the updated source $\mathbf{b} + \mathbf{b}^k$, Algorithm 5, Line 5. As such, $\Delta\mathbf{u}$ can be viewed as the observable of the parameter-estimation subproblem.

The amplitudes of $\Delta\mathbf{u}$ computed in the true velocity model show the dominant imprint of the reflection and the refraction from the top of the salt (Fig. 5.16a, shallow black zone). Also, the geometry of the seismically-transparent salt body is almost perfectly delineated. This shows that the amplitudes of $\Delta\mathbf{u}$ intrinsically embeds high-resolution information on the subsurface. This high-resolution potential can be simply illustrated with the wave-equation bilinearity as discussed below. Let us remind the wave equation for a single source and a single frequency

$$\boldsymbol{\Delta}\mathbf{u} + \omega^2 \text{diag}(\mathbf{m})\mathbf{u} = \mathbf{b} \tag{5.21}$$

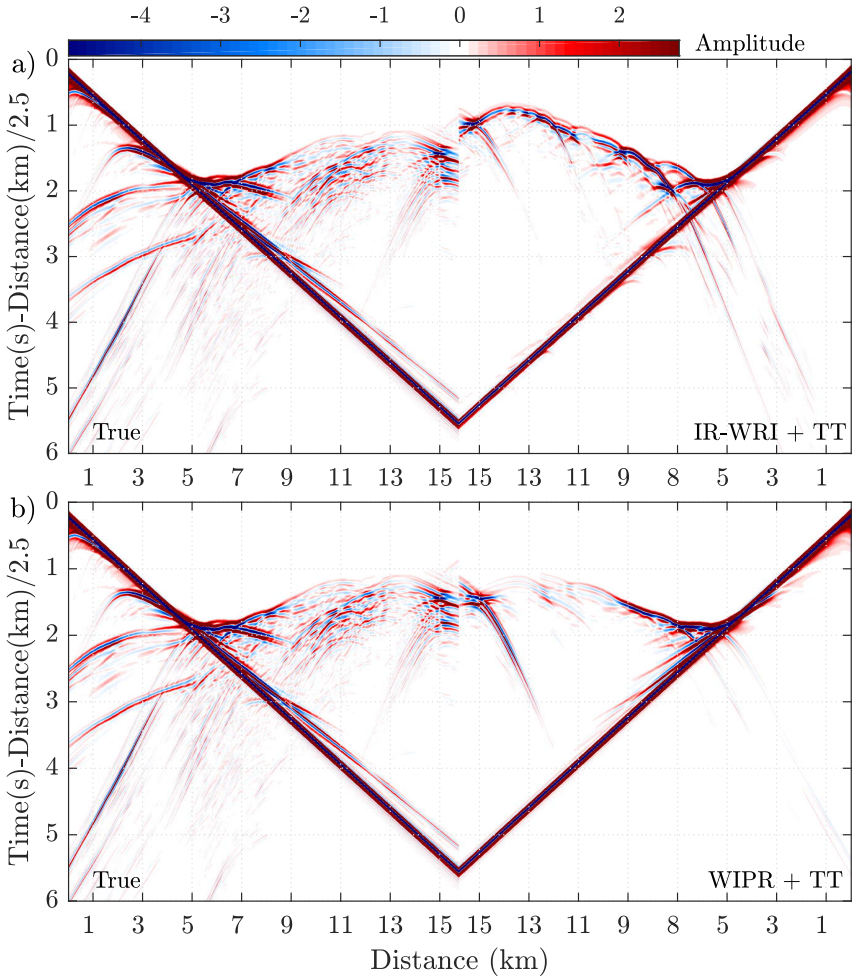

Figure 5.11 – BP salt example. Comparison between time-domain seismograms computed in the true velocity model (left panel) and those computed in the velocity models obtained by (a) IR-WRI (Fig. 5.8c) and (b) WIPR (Fig. 5.8d) (right panel). The seismograms are plotted with a reduction velocity of 2.5 km/s. Note the significant travel-time mismatch in (a) at long offset.

which leads to the following linear equation of the parameter-estimation subproblem:

$$\underbrace{\omega^2 \text{diag}(\mathbf{m})\mathbf{u}}_{\mathbf{L(u)m}} = \underbrace{\mathbf{b} - \mathbf{\Delta u}}_{\mathbf{y(u)}}. \tag{5.22}$$

Compared to equation 5.5, we drop the matrix $\mathbf{B}$ and $\mathbf{C}$ which are not important for the discussion. The equality implies that $|\mathbf{L(u)m}| = |\mathbf{y(u)}|$ and $\angle\mathbf{L(u)m} = \angle\mathbf{y(u)}$. If $\mathbf{u}$ is the true wavefield (if it would be recorded everywhere) then we would be able to recover the true model exactly by minimizing

$$\begin{aligned} \|\mathbf{L(u)m} - \mathbf{y(u)}\|_2^2 &= \|(|\mathbf{L(u)m}| - |\mathbf{y(u)}|)e^{j\angle\mathbf{L(u)m}}\|_2^2 \\ &= \|\mathbf{L(u)m} - |\mathbf{y(u)}|e^{j\angle\mathbf{L(u)m}}\|_2^2. \end{aligned} \tag{5.23}$$

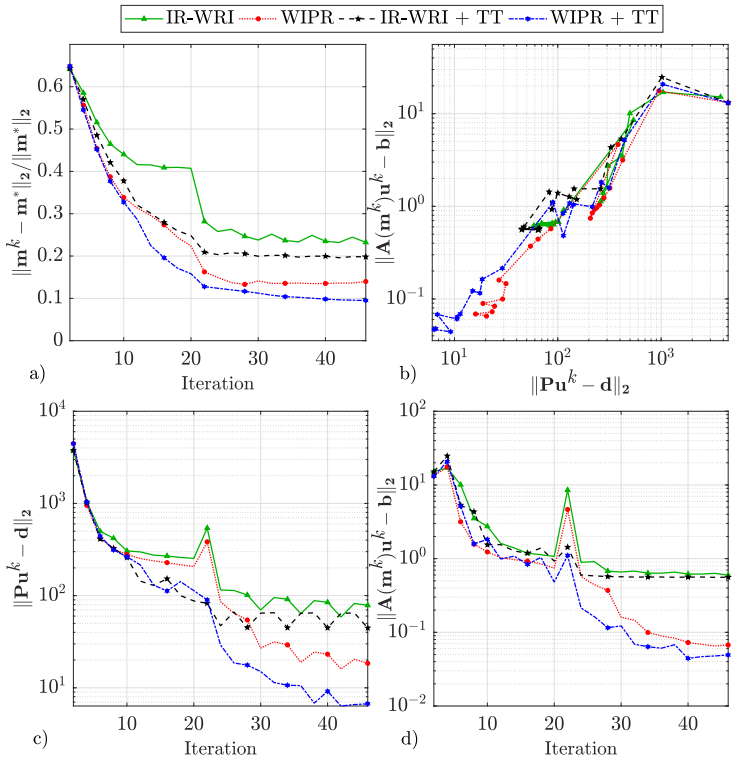

Figure 5.12 – BP salt example. Convergence path of {3,3.5} Hz inversion. (a) Evolution of $\|\mathbf{m}^k-\mathbf{m}^*\|_2/\|\mathbf{m}^*\|_2$ during iterations ($\mathbf{m}^*$ denotes the true model). (b) Convergence history of the algorithm in the ($\|\mathbf{P}\mathbf{u}^k-\mathbf{d}\|_2-\|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k-\mathbf{b}\|_2$) plane. (c-d) Evolution of (c) $\|\mathbf{P}\mathbf{u}^k-\mathbf{d}\|_2$, (d) $\|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k-\mathbf{b}\|_2$ during iterations for IR-WRI (green), WIPR (red), TT regularized IR-WRI (black), and TT regularized WIPR (blue).

It is seen that, basically, the true model is the solution of a weighted phase retrieval problem (as indicated by the middle term) where the complex exponential of the phase serves as the weight. Indeed, for an approximate wavefield which is obtained from eq. (5.7) with a rough velocity model or a large penalty parameter the equality in eq. (5.23) is not necessarily satisfied because, in this case, $\angle\mathbf{L}(\mathbf{u})\mathbf{m} \neq \angle\mathbf{y}(\mathbf{u})$. The proposed WIPR algorithm simply forces this equality to be satisfied at each iteration by aligning the phase of $\mathbf{y}$ with that of $\mathbf{L}(\mathbf{u})\mathbf{m}$ (right term in equation 5.23).

The ability to reconstruct the true model from amplitude information only can be illustrated more explicitly if we assume that $\mathbf{m}$ is real in equation 5.21 (non attenuating medium). The matrix $\mathbf{L}(\mathbf{u})$ is diagonal and contains the so-called virtual source (equation 5.22).

$$\mathbf{m} = \frac{1}{\omega^2}\left(\frac{1}{\mathbf{u}}\right) \circ \mathbf{y}(\mathbf{u}), \tag{5.24}$$

where $\circ$ denotes element wise multiplication. Indeed, we can equally reconstruct the true $\mathbf{m}$ from the magnitude of $\mathbf{y}$ as

$$\mathbf{m} = \frac{1}{\omega^2}\left(\frac{1}{|\mathbf{u}|}\right) \circ |\mathbf{y}(\mathbf{u})|, \tag{5.25}$$

		ME	$\|\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{b}\|_2$	$\|\mathbf{P}\mathbf{u}-\mathbf{d}\|_2$
Iteration No.45	IR-WRI	23.23	0.5954	7.8550
	WIPR	13.97	0.0672	1.8450
	IR-WRI+TT	19.81	0.5640	4.4630
	WIPR+TT	9.49	0.04926	0.6706
Final iteration	IR-WRI	7.57	0.2046	2.8911
	IR-WRI$_{pr}$	6.50	0.0647	1.6009
	IR-WRI+TT	4.89	0.1272	1.6470
	IR-WRI$_{pr}$+TT	3.16	0.0359	0.6350

Table 5.1 – Model error (ME) for the velocity models and least-squares data and source misfit at the end of the first frequency batch inversion and at the convergence point of the four inversion tests (with or without phase retrieval, with or without TT regularization). Here, ME is defined as $100\frac{\|\mathbf{m}_k-\mathbf{m}_*\|_1}{\|\mathbf{m}_*\|_1}$, where $\mathbf{m}_*$ denotes the true model.

where $1/\mathbf{u}$ and $1/|\mathbf{u}|$ are element-wise reciprocal of $\mathbf{u}$ and $|\mathbf{u}|$, respectively. The Marmousi II model computed with equations 5.24 and 5.25 by using the true wavefield (Fig. 5.1b) is shown in Fig. 5.1c and matches exactly the true one (Fig. 5.1a).

The Fig. 5.16c, as well as the vertical profile extracted at 6.5 km of distance, confirm that WIPR manages to match the amplitudes of the true $\Delta\mathbf{u}$ down to the top of the salt at 1.25 km depth, while IR-WRI (Fig. 5.16b) allows for an amplitude fit down to a maximum depth of 0.5 km only, consistently with the accuracy of the velocity fields reconstructed by IR-WRI and WIPR (Fig. 5.8c,d).

It is also instructive to look at the phase of $\Delta\mathbf{u}$ (Fig. 5.17). To facilitate the comparison of the phases computed in the true model (Fig. 5.17a) and in the reconstructed models (Fig. 5.17b,c), we superimpose in Fig. 5.17b,c the contours of the phase computed in the true model at the spatial positions where the phase wraparounds, namely when it jumps from $\pm\pi$ to $\mp\pi$ (Fig. 5.17b,c). We clearly show that WIPR reconstructs the phase of $\Delta\mathbf{u}$ much more accurately than IR-WRI around the salt body. This is further illustrated by the difference between the phases computed in the true model and the reconstructed models along the vertical profile at 6.5 km distance (Fig. 5.17d). The phase difference associated with IR-WRI (solid red curve) reaches a value of $-\pi$ at the top of the salt (1.25 km depth) showing that at these depths the two wavefields are cycle skipped, while the phase differences associated with WIPR (solid green curve) remains far below this limit down to around 2.7 km depth.

Conversely, it is worth noting that IR-WRI matches better the phase of the true wavefield than WIPR in the smooth bottom-right part of the model (Compare Figs. 5.17b and 5.17c). This highlights that the sensitivity of the inversion to amplitudes decreases rapidly with depth due to lack of illumination, geometrical spreading effects and energy partitioning at interfaces. We also want to stress that, although we update the wavespeeds from the amplitudes of $\mathbf{y}$, we update $\mathbf{y}$ at each iteration with the phase and amplitude of the wave equation error (the source residuals) in the framework of the ADMM optimization (Algorithm 5). This right-hand side updating re-injects at each iteration the phase error as a defect correction term in the optimization.

For sake of completeness, we show also the amplitude and phase of $\Delta\mathbf{u}$ for the Marmousi case

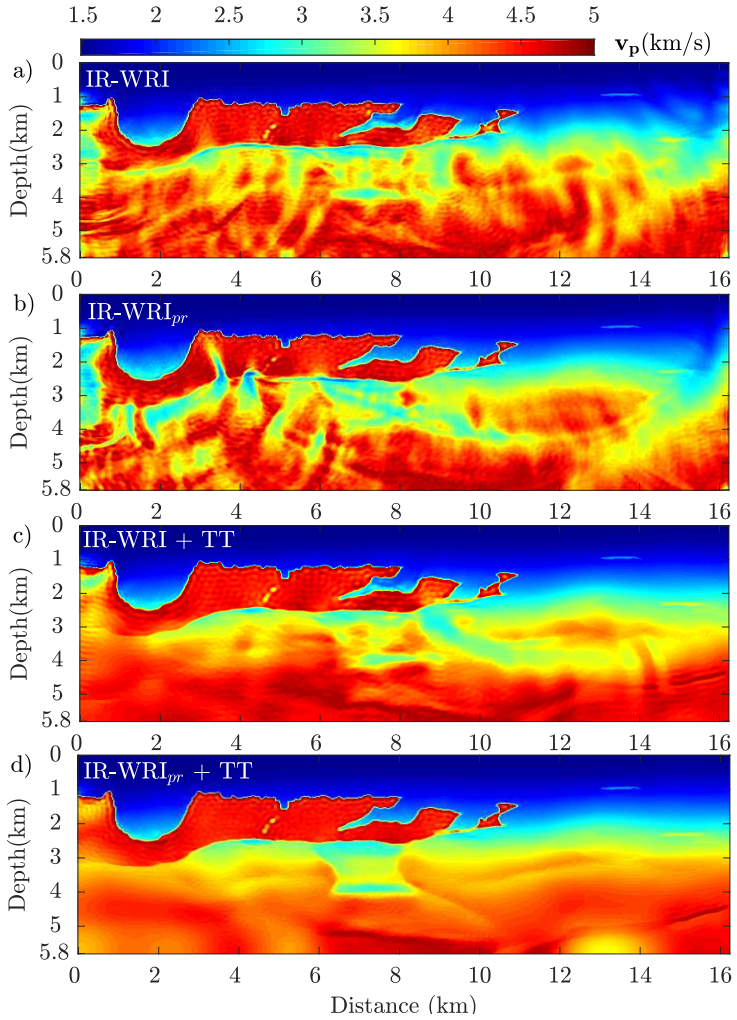

Figure 5.13 – BP salt example. Final velocity models estimated by IR-WRI (a,c) and IR-WRI$_{pr}$ (b,d). (a-b) Without TT regularization. (c-d) With TT regularization.

study in Figs. 5.18 and 5.19. Fig. 5.18 shows how WIPR reproduces more accurately the wide-angle reflection from the dipping layer on the left side of the source than IR-WRI. Fig. 5.19 shows that WIPR better reconstructs the true phase of $\Delta\mathbf{u}$ than IR-WRI.
Although the application of the phase retrieval algorithm has been limited to the scalar Helmholtz equation in this study, the method has been formulated such that more complex physics and multi-parameter reconstruction can be viewed (the reader is referred to Aghamiry et al. (2019d) and Aghamiry et al. (2019a) for application of IR-WRI in visco-acoustic VTI media). The key feature allowing for these extensions is the linearization of the non-convex phase retrieval problem performed by the MM approach, as reviewed in Appendix A (section 5.8).

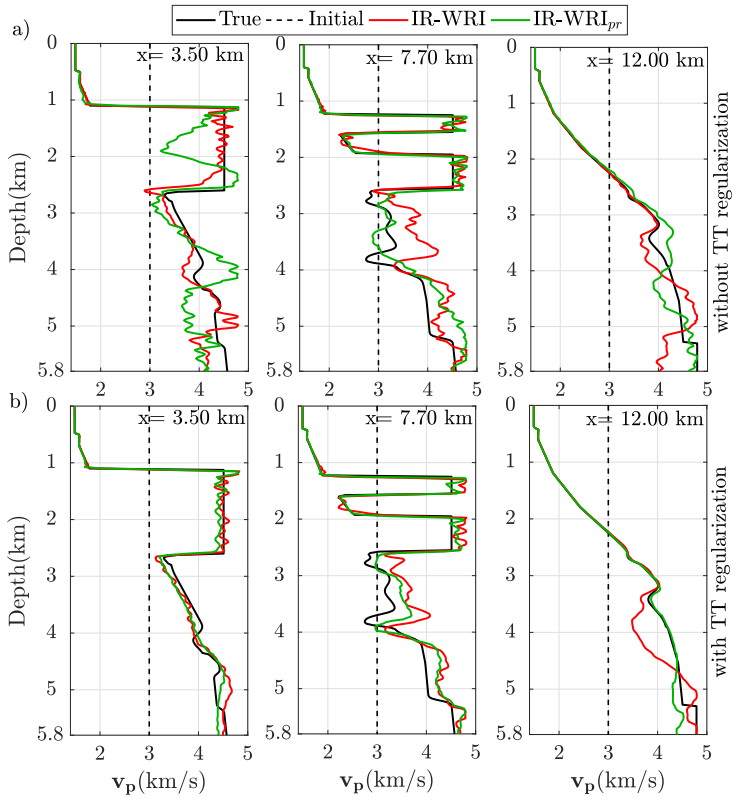

Figure 5.14 – BP salt example. Direct comparison along logs located at $x = 3.5, 7.7$ and $12\ km$ between the true velocity model (black), the initial model (dashed line) and the models estimated by IR-WRI (red) and IR-WRI$_{pr}$ (green) shown in Fig. 5.13. (a) Without TT. (b) With TT regularization.

5.7 Conclusions

We present a preliminary application of bound-constrained and TT regularized WIPR. Using the large contrast BP salt model, we show that, when the inversion is started from scratch (homogeneous initial velocity model), WIPR improves significantly the reconstruction of the sedimentary background and the top of the salt during the early iterations of the inversion (that is, when the wavefields are not yet accurately reconstructed far away from the receivers). Sparsity-promoting regularization is a necessary ingredient to stabilize the phase retrieval inversion at greater depths where the amplitude information becomes more challenging to extract. Sparsity-promoting regularization combined with phase retrieval maybe also two complementary tools to relax the need for finely-sampled stationary-recording acquisitions. Phase retrieval can also be implemented in classical FWI. Also, WIPR can be theoretically extended to multi-parameter reconstruction in attenuating media. Further theoretical and numerical works are, however, necessary to flesh out the potential and limits of phase retrieval in the FWI technology.

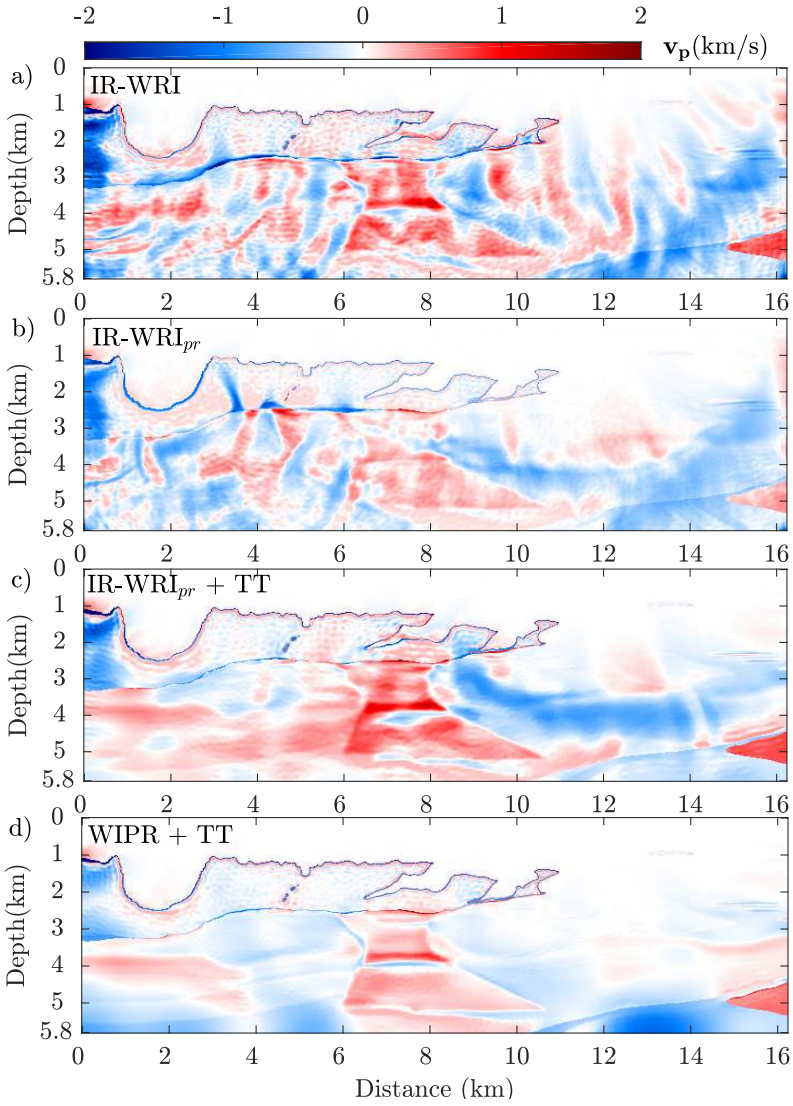

Figure 5.15 – BP salt example. Difference between the true velocity model and the velocity models shown in Fig. 5.13. (a,c) IR-WRI. (b,d) IR-WRI$_{pr}$. (a-b) Without TT regularization. (c-d) With TT regularization.

Acknowledgments

We would like to thank editors H. Yao, L. Alexander, and reviewers H. Zhu and an anonymous for their comments which help improving the manuscript. This study was partially funded by the SEISCOPE consortium (*http://seiscope2.osug.fr*), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXON-MOBIL, JGI, PETROBRAS, SCHLUMBERGER, SHELL, SINOPEC, and TOTAL. This study was granted access to the HPC resources of SIGAMM infrastructure (*https://www.oca.eu/fr/ mesocentre-sigamm*) and CINES/IDRIS/TGCC under the allocation A0050410596 made by GENCI.

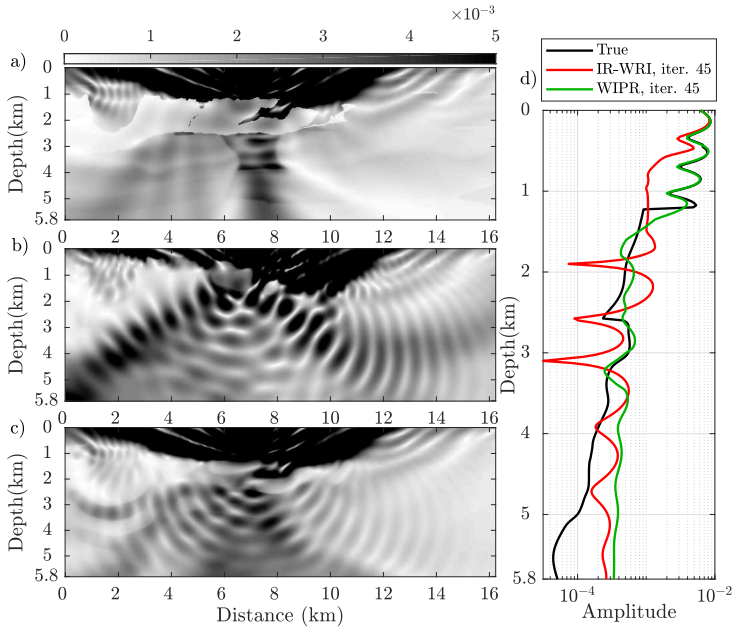

Figure 5.16 – Amplitude of $\Delta\mathbf{u}$ for 2004 BP model. (a) The wavefield $\mathbf{u}$ is computed in the true velocity model for the 3 Hz frequency. (b-c) The wavefields are computed in the velocity models inferred from TT-regularized (b) IR-WRI (Fig. 5.8c) and (c) WIPR (Fig. 5.8d) for the first frequency batch. (d) Comparison between vertical profiles extracted from (a) (black line), (b) (red line), (c) (green line) at a distance of 6.5 km.

5.8 Appendix A: Majorization-minimization to solve phase retrieval problem

Here, we present a simple algorithm to solve phase retrieval problem based on the majorization-minimization (MM) (Lange, 2016) technique. The governing idea of MM is to find the minimum of non-convex/convex function $f(\mathbf{x})$ via the iterative minimization of a simpler convex surrogate function $g(\mathbf{x}, \mathbf{x}^k)$ that majorizes $f(\mathbf{x})$ at step k (i.e. $g(\mathbf{x}, \mathbf{x}^k) \geq f(\mathbf{x})$). Figure 5.20 shows a schematic of the MM process. The non-convex function $f(\mathbf{x})$ is shown in blue while a few surrogate functions $g(\mathbf{x}, \mathbf{x}^k)$ for points $\mathbf{x}^k$, $k \in \{0, 1, 2, 3\}$, are shown in orange. This figure shows how the iterative MM algorithm approaches a local minimum of $f(\mathbf{x})$ through the minimum of easy to minimize surrogate functions $g(\mathbf{x}, \mathbf{x}^k)$.

The phase-retrieval problem for linear system $\mathbf{Lx} = \mathbf{y}$ can be written as (Gholami, 2014)

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) = \min_{\mathbf{x}} \quad \frac{1}{2}\||\mathbf{Lx}| - |\mathbf{y}|\|_2^2, \tag{5.26}$$

where $\mathbf{L} \in \mathbb{C}^{m\times n}$, $\mathbf{x} \in \mathbb{C}^{n\times 1}$, and $\mathbf{y} \in \mathbb{C}^{m\times 1}$. Problem (5.26) is non-convex where this non-convexity finds its root in removing the phase of the right hand side.

Let us introduce the auxiliary variable $\mathbf{z} = \mathbf{Lx}$, then the objective function in (5.26) can be written as

$$\tilde{f}(\mathbf{z}) = \frac{1}{2}\|\sqrt{\mathbf{zz}^*} - |\mathbf{y}|\|_2^2. \tag{5.27}$$

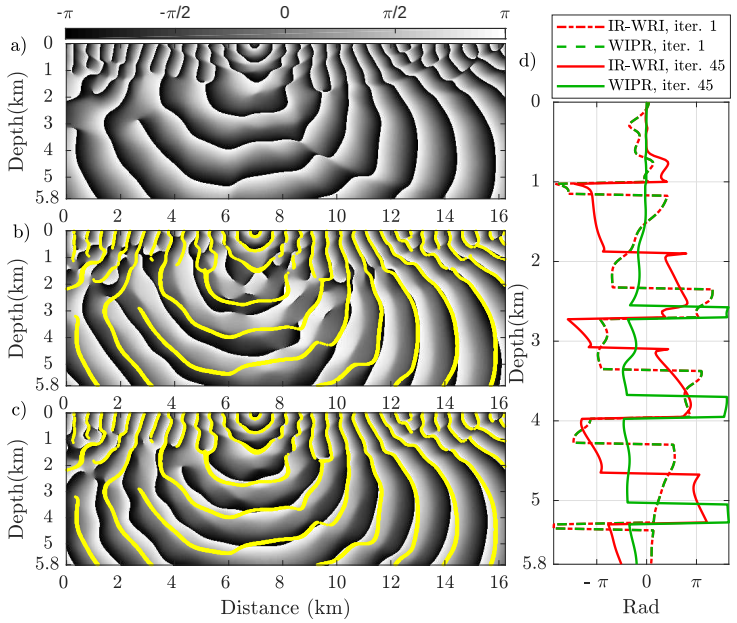

Figure 5.17 – Phase of $\Delta\mathbf{u}$ for 2004 BP model. Same as Fig. 5.16 for the phase of $\Delta\mathbf{u}$. (d) Difference between the phases computed in the true model and the reconstructed models at the first iteration (dashed red and green lines), after the first frequency batch inversion for IR-WRI (red line) and WIPR (green line) at a distance of 6.5 km. Note the phase mismatch between the phase computed in the true model and in the IR-WRI model below 0.5 km depth.

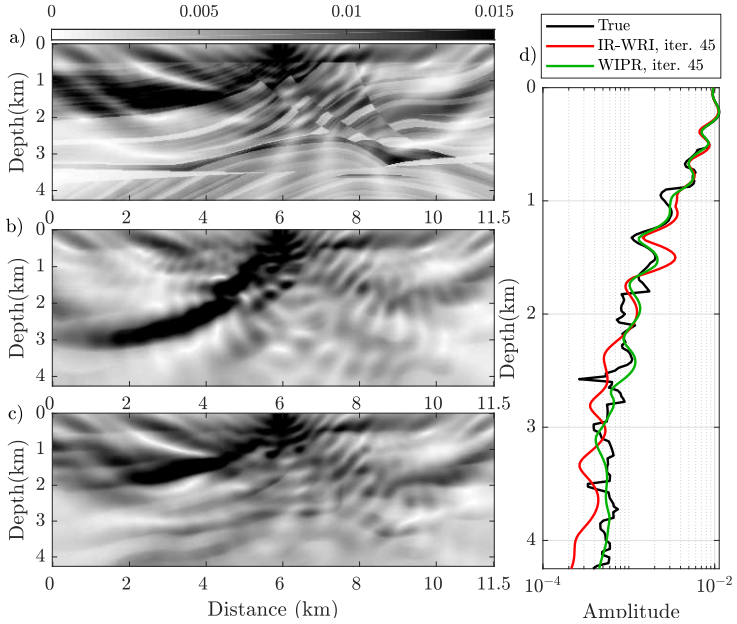

Figure 5.18 – Amplitude of $\Delta\mathbf{u}$ for Marmousi II model. (a) The wavefield $\mathbf{u}$ is computed in the true velocity model for the 3 Hz frequency. (b-c) The wavefields are computed in the velocity models inferred from TT-regularized (b) IR-WRI (Fig. 5.2c) and (c) WIPR (Fig. 5.2d) for the first frequency batch. (d) Comparison between vertical profiles extracted from (a) (black line), (b) (red line), (c) (green line) at a distance of 6.0 km.

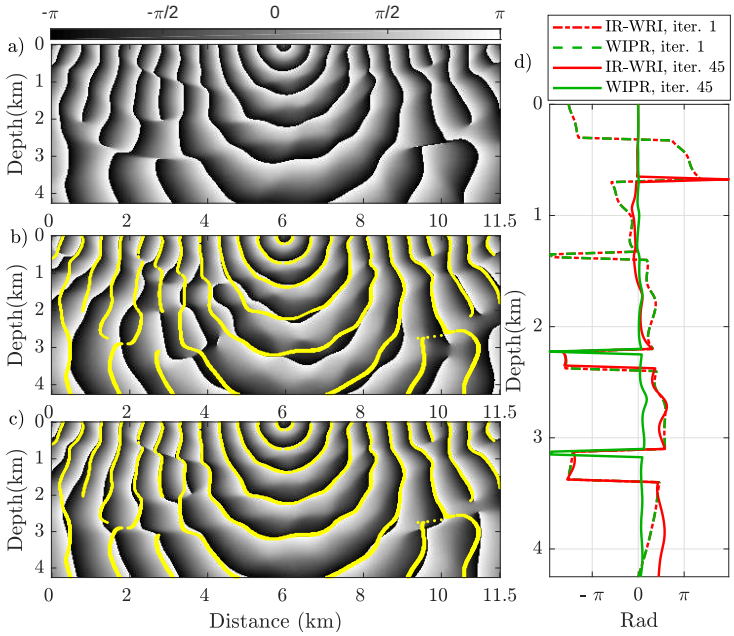

Figure 5.19 – Phase of $\Delta\mathbf{u}$ for Marmousi II model. Same as Fig. 5.18 for the phase of $\Delta\mathbf{u}$. (d) Difference between the phases computed in the true model and the reconstructed models at the first iteration (dashed red and green lines), after the first frequency batch inversion for IR-WRI (red line) and WIPR (green line) at a distance of 6.0 km.

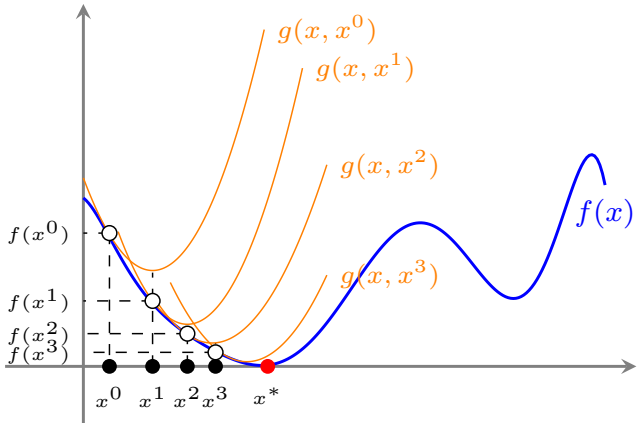

Figure 5.20 – Sketch of the iterative MM algorithm to solve optimization problem 5.26. The function $f(\mathbf{x})$ is shown in blue while a few surrogate functions $g(\mathbf{x}, \mathbf{x}^k), k \in \{0, 1, 2, 3\}$ are shown in orange. The MM algorithm seeks to find a local minimizer of $f(\mathbf{x})$, red point, by iteratively minimizing easy to minimize functions $g(\mathbf{x}, \mathbf{x}^k)$. The surrogate function $g(\mathbf{x}, \mathbf{x}^k)$ is greater than or equal to $f(\mathbf{x})$, and the equality holds at the current optimal point $\mathbf{x}^k$. Minimization of $g(\mathbf{x}, \mathbf{x}^k)$ gives a new optimal point and it reaches to $\mathbf{x}^*$ when k tends to infinity.

Using the Wirtinger calculus (Kreutz-Delgado, 2009), the quadratic approximation of $f(\mathbf{x})$ around an initial guess $\mathbf{x}^k$ is

$$g(\mathbf{x}, \mathbf{x}^k) = f(\mathbf{x}^k) + \nabla f(\mathbf{x}^k)^T \Delta\mathbf{x} + \Delta\mathbf{x}^T \mathbf{H}(\mathbf{x}^k) \Delta\mathbf{x}, \tag{5.28}$$

where $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}^k$, ∇f is the gradient of $f(\mathbf{x})$ and $\mathbf{H}$ denotes the Hessian, both evaluated at $\mathbf{x}^k$. Simple algebra shows that the gradient of $f(\mathbf{x})$ is given by

$$\begin{aligned}\nabla f(\mathbf{x}) &= 2\mathbf{L}^T \frac{\partial \tilde{f}(\mathbf{z})}{\partial \mathbf{z}^*} && (5.29)\\ &= \mathbf{L}^T \text{diag}(\mathbf{1} - \frac{|\mathbf{y}|}{|\mathbf{L}\mathbf{x}|})\mathbf{L}\mathbf{x} && (5.30)\\ &= \mathbf{L}^T(\mathbf{L}\mathbf{x} - |\mathbf{y}|e^{j\angle \mathbf{L}\mathbf{x}}), && (5.31)\end{aligned}$$

where $\mathbf{1}$ is an all-ones vector, and the Hessian operator is given by

$$\begin{aligned}\mathbf{H}(\mathbf{x}) &= 2\mathbf{L}^T \left(\frac{\partial^2 \tilde{f}(\mathbf{z})}{\partial \mathbf{z} \partial \mathbf{z}^*} + \frac{\partial^2 \tilde{f}(\mathbf{z})}{\partial \mathbf{z}^* \partial \mathbf{z}^*} \mathbf{\Omega} \right) \mathbf{L} && (5.32)\\ &= \mathbf{L}^T \left(\text{diag}(\mathbf{1} - \frac{|\mathbf{y}|}{2|\mathbf{L}\mathbf{x}|}) + \text{diag}(\frac{|\mathbf{y}|e^{j2\angle \mathbf{L}\mathbf{x}}}{2|\mathbf{L}\mathbf{x}|})\mathbf{\Omega} \right) \mathbf{L},\end{aligned}$$

where $\mathbf{\Omega}(\mathbf{z}) = \mathbf{z}^*$ denotes the complex conjugate operator.
Ignoring the nonlinear terms of Hessian leads to the following Gauss-Newton approximation

$$g(\mathbf{x}, \mathbf{x}^k) = f(\mathbf{x}^k) + \nabla f(\mathbf{x}^k)^T \Delta\mathbf{x} + \Delta\mathbf{x}^T \mathbf{L}^T \mathbf{L} \Delta\mathbf{x}, \tag{5.33}$$

or equivalently

$$g(\mathbf{x}, \mathbf{x}^k) = \frac{1}{2} \| \mathbf{L}\mathbf{x} - |\mathbf{y}| e^{j\angle \mathbf{L}\mathbf{x}^k} \|_2^2. \tag{5.34}$$

The quadratic function $g(\mathbf{x}, \mathbf{x}^k)$ majorizes $f(\mathbf{x})$ at the point $\mathbf{x}^k$ provided that the following MM conditions are satisfied:

$$\begin{cases} g(\mathbf{x}^k, \mathbf{x}^k) = f(\mathbf{x}^k) & (5.35a)\\ g(\mathbf{x}, \mathbf{x}^k) \geq f(\mathbf{x}) \qquad \text{for all } \mathbf{x}. & (5.35b)\end{cases}$$

Equality (5.35a) can easily be confirmed by substituting $\mathbf{x}$ by $\mathbf{x}^k$ in $f(\mathbf{x})$ and $g(\mathbf{x}, \mathbf{x}^k)$. Furthermore, let $z_i = (\mathbf{L}\mathbf{x})_i$ and $\theta_i = \angle(\mathbf{L}\mathbf{x}^k)_i$ then according to the backward triangle inequality (Meyer, 2000)

$$\begin{aligned}\frac{1}{2}|z_i - |y_i| e^{j\theta_i}|^2 &\geq \frac{1}{2}(|z_i| - ||y_i| e^{j\theta_i}|)^2 && (5.36)\\ &= \frac{1}{2}(|z_i| - |y_i||e^{j\theta_i}|)^2\\ &= \frac{1}{2}(|z_i| - |y_i|)^2.\end{aligned}$$

Hence,

$$\frac{1}{2} \sum_i |z_i - |y_i| e^{j\theta_i}|^2 \geq \frac{1}{2} \sum_i (|z_i| - |y_i|)^2, \tag{5.37}$$

which confirms that the inequality (5.35b) is also satisfied and thus $g(\mathbf{x}, \mathbf{x}^k)$ majorizes $f(\mathbf{x})$. These conditions guarantee that iterative minimization of $g(\mathbf{x}, \mathbf{x}^k)$ converge to a local minimum of $f(\mathbf{x})$ as k tends to infinity (Wu et al., 1983).

Chapter 6

Toward time-domain ADMM-based WRI: efficient wavefield reconstruction

Chapter overview: This chapter is a first attempt toward the implementation of ADMM-based WRI in the time domain. Wavefield reconstruction inversion (WRI) extends the search space of Full Waveform Inversion (FWI) by computing wavefields with a wave-equation relaxation such that the observation equation is satisfied (i.e., the data are fitted) with inaccurate velocity models. This wavefield reconstruction can be easily implemented in the frequency domain by solving in a least-squares senses an overdetermined linear system gathering the weighted wave equation and the observation equation. Wavefield reconstruction in the time domain with explicit time stepping methods is more challenging because I will show that the right-hand side of the wave equation to be solved depends on a back-propagated wavefield, whose excitation term is formed by the residuals between the data and the sought (unknown) wavefield. To bypass this issue, I start with the residuals between the data and the exact solution of the wave equation and I refine the reconstructed wavefield iteratively with a majorize-minimization algorithm (Lange, 2016). Although this iterative refinement procedure generates computational overheads during one WRI iteration, it allows to preserve the capability of WRI to extend the search space and hence allows for the converge toward a more accurate velocity models. This chapter includes our paper, entitled *Accurate and efficient data-assimilated wavefield reconstruction in the time domain* which is published in *Geophysics* journal as a geophysical letter (Aghamiry et al., 2020a).

Accurate and efficient data-assimilated wavefield reconstruction in the time domain

Hossein S. Aghamiry, Ali Gholami and Stéphane Operto

Geophysics, 2019, **85(2)**, pages A7-A12

DOI: 10.1190/GEO2019-0535.1

6.1 Summary

Wavefield reconstruction inversion (WRI) mitigates cycle skipping in Full Waveform Inversion by computing wavefields that do not exactly satisfy the wave-equation to match data with inaccurate velocity models. We refer these wavefields to as "data assimilated wavefield" because they are obtained by combining the physics of wave propagation and the observations. Then, the velocity model is updated by minimizing the wave equation errors, namely the source residuals. Computing these data-assimilated wavefields in the time domain with explicit time stepping is challenging. This is because the right-hand side of the wave equation to be solved depends on the back-propagated residuals between the data and the unknown wavefields. To bypass this issue, a previously proposed approximation replaces these residuals by those between the data and the exact solution of the wave equation. This approximation is questionable during the early WRI iterations when the wavefields computed with and without data-assimilation differ significantly. We propose a simple *backward-forward* time-stepping recursion to refine the accuracy of the data-assimilated wavefields. Each iteration requires us to solve one backward and one forward problem, the former being used to update the right-hand side of the latter. An application to the BP salt model shows that a few iterations are enough to reconstruct data-assimilated wavefields accurately with a crude velocity model. Although this *backward-forward* recursion leads to increased computational overheads during one WRI iteration, it preserves its capability to extend the search space.

6.2 Introduction

Wavefield reconstruction inversion (WRI) (van Leeuwen and Herrmann, 2013) decreases the nonlinearity of full waveform inversion (FWI) (Virieux and Operto, 2009) by extending the parameter search space. WRI first computes wavefields such that the modeled and recorded data are matched to each other (hence, mitigating cycle skipping) despite inaccurate velocity models by relaxing the constraint that the modeled data satisfy the wave equation exactly. This data fit is achieved during wavefield reconstruction by jointly solving in a least-squares sense the weighted wave equation and the observation equation, namely the equation relating the wavefield to the data via a sparse sampling operator. This describes a data-assimilation procedure since the wavefield reconstruction is driven not only by the physics of wave propagation but also by the observations. Accordingly, we refer these wavefields to as data-assimilated wavefields in the following study. This wave-equation relaxation generates source residuals, namely extended sources (Huang et al., 2018a), that are minimized in a second step to update the velocity model. When the starting velocity model is inaccurate, wavefields are reconstructed with aggressive relaxation of wave-equation to guarantee the data fit. In that case, the data-assimilated wavefields can differ significantly from the wavefields that satisfy the wave equation exactly.

Data-assimilated wavefields are reconstructed in the frequency domain by solving with direct or iterative methods an overdetermined sparse linear system gathering the weighted time-harmonic wave equation and the observation equation. In contrast, the wavefield reconstruction in the time domain with explicit time stepping is more challenging because the right-hand side

of the wave equation to solve depends on the final state of the unknown assimilated wavefield. The approaches that have been proposed so far rely on an approximation, which prevents aggressive wave equation relaxation and hence inaccurate starting velocity models to be used (Wang et al., 2016). A first aim of this letter is to propose an efficient *backward-forward* time-stepping recursion for data-assimilated wavefield reconstruction. Then we show that the approximation used by Wang et al. (2016) is obtained at the first iteration of the proposed algorithm. We assess the accuracy and the convergence speed of our method on the 2004 BP salt model. We show that it reconstructs wavefields which are close to those obtained in the frequency domain after a few iterations when the background model is far away from the true model. Then, we illustrate the significant impact of the accuracy of the wavefield reconstruction on the velocity model built by iteratively-refined (IR-)WRI (Aghamiry et al., 2019c).

6.3 Method

6.3.1 Problem statement

FWI seeks the subsurface parameters $\mathbf{m} \in \mathbb{R}^{N\times 1}$ by solving the following partial-differential equation (PDE) constrained optimization problem (van Leeuwen and Herrmann, 2013; Aghamiry et al., 2019c):

$$\min_{\mathbf{m},\mathbf{u}} \quad ||\mathbf{P}\mathbf{u}-\mathbf{d}||_2^2 \quad \text{subject to} \quad \mathbf{A}(\mathbf{m})\mathbf{u}=\mathbf{b}, \tag{6.1}$$

where $\|\cdot\|_2$ denotes the Euclidean norm. In the frequency domain, the source $\mathbf{b} \in \mathbb{C}^{N\times 1}$, the wavefield $\mathbf{u} \in \mathbb{C}^{N\times 1}$, the data $\mathbf{d} \in \mathbb{C}^{M\times 1}$ and the PDE operator $\mathbf{A}(\mathbf{m}) \in \mathbb{C}^{N\times N}$ are complex valued. The linear observation operator $\mathbf{P} \in \mathbb{R}^{M\times N}$ samples $\mathbf{u}$ at the receiver positions. The number of degrees of freedom which discretize the parameter and the data spaces are denoted by N and M, respectively. In the time domain, $\mathbf{b} \in \mathbb{R}^{(N\times N_t)\times 1}$, $\mathbf{u} \in \mathbb{R}^{(N\times N_t)\times 1}$, $\mathbf{d} \in \mathbb{R}^{(M\times N_t)\times 1}$ and $\mathbf{A}(\mathbf{m}) \in \mathbb{R}^{(N\times N_t)\times(N\times N_t)}$ are real valued, where N_t is the number of time steps.

Eliminating $\mathbf{u}$ from equation 6.1, we recast FWI (Pratt et al., 1998) as a reduced-space unconstrained problem

$$\min_{\mathbf{m}} \ \|\mathbf{P}\mathbf{A}(\mathbf{m})^{-1}\mathbf{b}-\mathbf{d}\|_2^2, \tag{6.2}$$

where the misfit function depends on $\mathbf{m}$ through the highly nonlinear inverse operator $\mathbf{A}(\mathbf{m})^{-1}$.

The WRI method (van Leeuwen and Herrmann, 2013) solves problem 6.1 with a penalty method to mitigate the nonlinearity of the reduced-space FWI, problem 6.3, via a relaxation of the constraint

$$\min_{\mathbf{u},\mathbf{m}} \|\mathbf{P}\mathbf{u}-\mathbf{d}\|_2^2+\lambda\|\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{b}\|_2^2, \tag{6.3}$$

where λ is the penalty parameter. A small value of λ is typically used to foster the data fit within the first iterations at the expense of the fidelity with which the wave equation is satisfied. Alternatively, IR-WRI (Aghamiry et al., 2019c) uses an augmented Lagrangian method, which can be recast as a penalty method where the right-hand sides $\mathbf{b}$ of the wave equation are iteratively updated with the wave-equation errors. This right-hand side updating progressively corrects the

errors generated by the penalization and allows us to satisfy the wave-equation constraint at the convergence point when a constant λ is used.

The bivariate problem 6.3 can be solved efficiently in alternating mode, namely, estimate $\mathbf{u}$ keeping $\mathbf{m}$ fixed and *vice versa*. The wavefield reconstruction requires us to solve

$$\left(\lambda \mathbf{A}(\mathbf{m})^T \mathbf{A}(\mathbf{m}) + \mathbf{P}^T \mathbf{P}\right) \mathbf{u} = \lambda \mathbf{A}(\mathbf{m})^T \mathbf{b} + \mathbf{P}^T \mathbf{d}, \tag{6.4}$$

where the left-hand side operator is the Gauss-Newton Hessian, the right-hand side is the steepest descent direction and ^{T} denotes the transpose operator. We refer to the solution of equation 6.4 as the data-assimilated wavefield as it jointly satisfies the weighted wave-equation and the observation equation (i.e. fits the data) in a least-squares sense. Then, $\mathbf{m}$ is updated by minimizing the wave equation residual $\|\mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{b}\|_2^2$ keeping $\mathbf{u}$ fixed.

We focus now on the solution of the first subproblem. In the frequency domain, the system 6.4 can be solved accurately for a limited number of discrete frequencies with either direct or iterative methods. In the time domain, the size of the system would be prohibitively large due to the extra temporal dimension.

6.3.2 Reconstruction of data-assimilated wavefield by backward-forward recursion

In this section, we show how to solve the augmented wave equation system, equation 6.4, with explicit (matrix free) time stepping via an iterative majorize-minimization (MM) algorithm (Lange, 2016). Let us write equation 6.4 in a compact matrix form $\mathbf{B}\mathbf{u} = \mathbf{y}$, where $\mathbf{B} = \lambda \mathbf{A}^T \mathbf{A} + \mathbf{P}^T \mathbf{P}$ is a symmetric positive definite matrix and $\mathbf{y} = \lambda \mathbf{A}^T \mathbf{b} + \mathbf{P}^T \mathbf{d}$ (for compactness, we ignore the argument $\mathbf{m}$ of $\mathbf{A}$, while recalling that $\mathbf{m}$ is kept fixed during the wavefield reconstruction). Solving $\mathbf{B}\mathbf{u} = \mathbf{y}$ is equivalent to minimizing the quadratic function

$$f(\mathbf{u}) = \frac{1}{2} \mathbf{u}^T \mathbf{B} \mathbf{u} - \mathbf{y}^T \mathbf{u}. \tag{6.5}$$

The governing idea of MM is to find the minimum of f via the iterative minimization of a simpler *surrogate* function $\tilde{f}_k$ that majorizes f at step k (i.e. $\tilde{f}_k \geq f$). For reasons which will become clear later, we define $\tilde{f}_k$ as

$$\tilde{f}_k(\mathbf{u}) = f(\mathbf{u}) + \frac{c}{2} \|\mathbf{A}\mathbf{u} - \mathbf{A}\mathbf{u}_k\|_2^2 - \frac{1}{2} \|\mathbf{P}\mathbf{u} - \mathbf{P}\mathbf{u}_k\|_2^2, \tag{6.6}$$

where the scalar c is determined such that the Hessian of the added terms is positive definite

$$c\mathbf{A}^T \mathbf{A} - \mathbf{P}^T \mathbf{P} \succ 0. \tag{6.7}$$

This condition makes $\tilde{f}_k$ strictly convex and guarantees that the following surrogate conditions are satisfied (Lange, 2016)

$$\begin{cases} \tilde{f}_k(\mathbf{u}_k) = f(\mathbf{u}_k), \\ \tilde{f}_k(\mathbf{u}) \geq f(\mathbf{u}), \quad \forall \mathbf{u}, k. \end{cases} \tag{6.8}$$

Figure 6.1 shows a schematic of the MM process. The convex function f is shown in blue while a few surrogate functions $\tilde{f}_k$ for points $\mathbf{u}_k$, $k \in \{0,1,2,3\}$, are shown in orange. This figure shows how the iterative MM algorithm approaches the minimum of f through the minimum of easy to minimize surrogate functions $\tilde{f}_k$.

It follows from equation 6.8 that iterative minimization of $\tilde{f}_k$ will converge to the global minimum of f since

$$f(\mathbf{u}_{k+1}) \leq \tilde{f}_k(\mathbf{u}_{k+1}) \leq \tilde{f}_k(\mathbf{u}_k) = f(\mathbf{u}_k). \tag{6.9}$$

Zeroing the gradient of $\tilde{f}_k$, equation 6.6, gives the wavefield at iteration $k+1$

$$\mathbf{u}_{k+1} = \mathbf{u}_k + \alpha \tilde{\mathbf{B}}^{-1} \left(\mathbf{y} - \mathbf{B}\mathbf{u}_k\right), \tag{6.10}$$

where $\tilde{\mathbf{B}}^{-1} = (\mathbf{A}^T\mathbf{A})^{-1}$ is an approximate inverse of $\mathbf{B}$ and $\alpha = \frac{1}{\lambda + c}$ serves as the step length. Accordingly, our method can be viewed as a quasi-Newton inversion. Note that, since $\mathbf{A}$ does not change during the wavefield-reconstruction iterations, a suitable constant c can be found to satisfy the positive definite condition, equation 6.7, at each step k. However, α can be updated at each iteration by a line search to optimize the convergence speed of the wavefield reconstruction. Starting from an initial guess $\mathbf{u}_0$, a recursive procedure (Aghamiry et al., 2019c, their Appendix B) gives immediately

$$\mathbf{u}_{k+1} = \mathbf{u}_0 + \alpha \sum_{i=0}^{k} \tilde{\mathbf{B}}^{-1} \left(\mathbf{y} - \mathbf{B}\mathbf{u}_i\right). \tag{6.11}$$

If we substitute $\tilde{\mathbf{B}}^{-1}$, $\mathbf{B}$ and $\mathbf{y}$ by their expression ($\mathbf{B} = \lambda\mathbf{A}^T\mathbf{A} + \mathbf{P}^T\mathbf{P}$ and $\mathbf{y} = \lambda\mathbf{A}^T\mathbf{b} + \mathbf{P}^T\mathbf{d}$) and multiply the left- and right-hand sides by $\mathbf{A}$, we obtain

$$\mathbf{A}\mathbf{u}_{k+1} = \mathbf{A}\mathbf{u}_0 + \alpha \sum_{i=0}^{k} \left(\bar{\mathbf{u}}_i + \lambda \Delta\mathbf{b}_i\right), \tag{6.12}$$

where $\Delta\mathbf{b}_i = \mathbf{b} - \mathbf{A}\mathbf{u}_i$ denotes the source residuals at iteration i. The backward wavefield $\bar{\mathbf{u}}_i$ satisfies the so-called adjoint equation

$$\mathbf{A}^T\bar{\mathbf{u}}_i = \mathbf{P}^T\left(\mathbf{d} - \mathbf{P}\mathbf{u}_i\right) = \mathbf{P}^T\Delta\mathbf{d}, \tag{6.13}$$

which propagates the data residuals $\Delta\mathbf{d}$ backward in time (e.g. Pratt et al., 1998).

The most natural choice for $\mathbf{u}_0$ is the (exact) solution of the wave equation

$$\mathbf{u}_0 = \mathbf{A}^{-1}\mathbf{b}. \tag{6.14}$$

Accordingly, equation 6.12 simplifies to

$$\mathbf{A}\mathbf{u}_{k+1} = \mathbf{b} + \alpha \sum_{i=0}^{k} \left(\bar{\mathbf{u}}_i + \lambda \Delta\mathbf{b}_i\right) = \mathbf{b} + \bar{\mathbf{b}}_k, \tag{6.15}$$

where $\bar{\mathbf{b}}_k$ denotes the extended source. Noting from the above equation that $\Delta\mathbf{b}_{k+1} = -\bar{\mathbf{b}}_k$, the

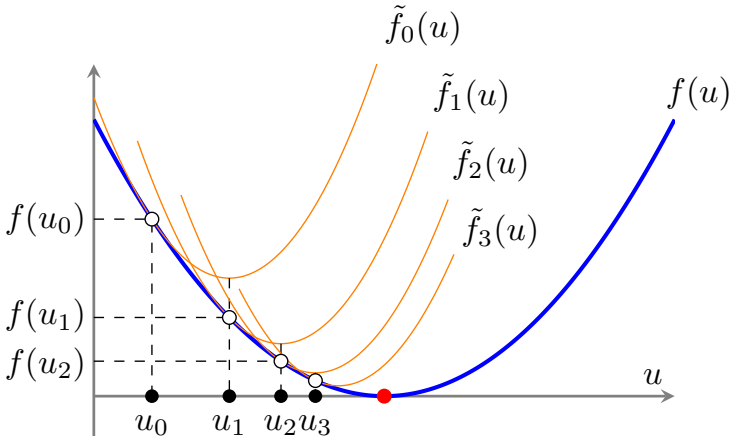

Figure 6.1 – Sketch of the iterative MM algorithm. The function f is shown in blue while a few surrogate functions $\tilde{f}_k, k \in \{0,1,2,3\}$ are shown in orange. The MM algorithm seeks to find the minimizer of f, red point, by iteratively minimizing easy to minimize functions $\tilde{f}_k$. The surrogate function $\tilde{f}_k$ is greater than or equal to f, and the equality holds at the current optimal point u_k. Minimization of $\tilde{f}_k$ gives a new optimal point.

extended source reduces to the following series of backward wavefields

$$\begin{aligned}\bar{\mathbf{b}}_k &= \alpha \sum_{i=0}^{k} \left(\bar{\mathbf{u}}_i - \lambda \bar{\mathbf{b}}_{i-1}\right) = (1-\alpha\lambda)\,\bar{\mathbf{b}}_{k-1} + \alpha \bar{\mathbf{u}}_k \\ &= \alpha \sum_{i=0}^{k} (1-\alpha\lambda)^{k-i}\,\bar{\mathbf{u}}_i. \end{aligned} \tag{6.16}$$

We summarize now the key results of our approach. We have designed the surrogate function $\tilde{f}_k$ of the MM algorithm, equation 6.6, such that it allows us to recast the normal-equation system, equation 6.4, in a form which is amenable to explicit time stepping, equation 6.15. This is indeed shown by the left hand side of equation 6.15 which reduces now to the wave equation operator $\mathbf{A}$. The MM method also provides us with the mathematical framework to prove the convergence of our algorithm. The iterative refinement of the wavefields performed by the right-hand side updating in equation 6.15 corrects the errors generated by the use of the approximate inverse $\tilde{\mathbf{B}}^{-1}$ of $\mathbf{B}$, equation 6.10, associated with $\tilde{f}_k$. In fact, the series of backward wavefields in equation 6.16 converges to the backward wavefield that would have been generated by the back-propagated residuals between the data and the data-assimilated wavefield sought. This is shown by the re-writing of equation 6.4 as

$$\mathbf{A}\mathbf{u} = \mathbf{b} + \frac{1}{\lambda}\mathbf{A}^{-T}\mathbf{P}^T\left(\mathbf{d} - \mathbf{P}\mathbf{u}\right), \tag{6.17}$$

where the unknown data-assimilated wavefield $\mathbf{u}$ appears both in the right- and left-hand sides.

The approximate wavefield reconstructed by Wang et al. (2016, their equations 7 and 8) is

the one obtained at the first iteration of our approach (k=0 and c=0 in equation 6.15).

$$\mathbf{A}(\mathbf{m})\mathbf{u}_1 = \mathbf{b} + \bar{\mathbf{u}}_0/\lambda, \tag{6.18}$$

In this case, the backward wavefield $\bar{\mathbf{u}}_0$ on the right-hand side is computed from the data residuals generated by the forward wavefield $\mathbf{u}_0$, which exactly satisfies the wave equation 6.14. Instead, the data residuals should be generated from the data-assimilated wavefield, i.e. the right-hand side of equation 6.17. This is a gross approximation when a significant wave-equation relaxation is necessary to match the data during the early WRI iterations, which is why further improvements to the method are necessary.

The wavefield reconstruction requires one backward and one forward solve at each iteration (Algorithm 1), hence we call it a *backward-forward* time stepping recursion. The storage requirement scales to the full time history of the sum of the backward wavefields (i.e. $N \times N_t$). The backward-wavefields can be recomputed on the fly forward in time during the wavefield reconstruction if one wants to avoid its storage on disk.

Algorithm 6: Data-assimilated wavefield reconstruction by backward-forward recursion.

Initialize: k = 0; $\mathbf{u}_0 = \mathbf{A}(\mathbf{m})^{-1}\mathbf{b}, \bar{\mathbf{b}}_{-1} = \mathbf{0}$.

Step 1: solve the adjoint equation for $\bar{\mathbf{u}}_k$ (equation 6.13) and update the extended source (equation 6.16).

Step 2: solve the forward equation for $\mathbf{u}_{k+1}$ (equation 6.15).

Step 3: if $\|\mathbf{u}_{k+1} - \mathbf{u}_k\| \leq \epsilon$ exit, else set $k = k + 1$ go to Step 1.

6.4 Numerical examples

Although this study is dedicated to the time-domain implementation, we assess the accuracy and the convergence of the algorithm in the frequency domain because we can easily generate an accurate data-assimilated wavefield by solving equation 6.4 with a direct solver. We compute the assimilated wavefield with a constant value of c that satisfies the positive definite condition, equation 6.7, and a small value of λ to generate a significant wave equation relaxation and fit the data accordingly. We refer to the wavefield computed by the *backward-forward* method at iteration k as BFI_k. We assess the method against a part of the 2004 BP salt model (Figure 6.2a). The selected subsurface model is 16.25 km wide and 5.825 km deep, and is discretized with a 25 m grid interval. We compute the wavefields with a finite-difference method (Chen et al., 2013), using absorbing boundary conditions along the bottom, right and left sides of the model and a free-surface boundary condition at the surface. We consider a long-offset fixed-spread acquisition designed with a line of 80 sources on the sea floor spaced 200 m apart and a line of 215 receivers spaced 75 m apart at 25 m depth.

We compute the data-assimilated wavefields in a laterally homogeneous velocity model in which the velocity linearly increases with depth from 1.5 km/s to 4.5 km/s. The wavefield computed in the true model and the data-assimilated wavefield computed with a direct solver

(the data-assimilated wavefield targeted by our recursive approach) are shown in Figures 6.2b and 6.2c for the 3 Hz frequency and a source located at 450-m distance. A direct comparison between the two wavefields at the receiver positions shows that the chosen λ allows the data to be matched (Figure 6.2d). We show the BFI_1 wavefield, namely the one used by Wang et al. (2016), and the difference with the targeted wavefield (Figure 6.2c) in Figures 6.3a and 6.3d, respectively. The significant mismatch between the two wavefields, highlighted by their direct comparison at three depths ($z = 0, 2.0, 4.0$ km) in Figure 6.3g, illustrates the need for more iterations. The BFI_3 and BFI_5 wavefields show that three to five iterations are enough to achieve the desired accuracy (Figures 6.3b,c,e,f and Figures 6.3h,i, respectively).

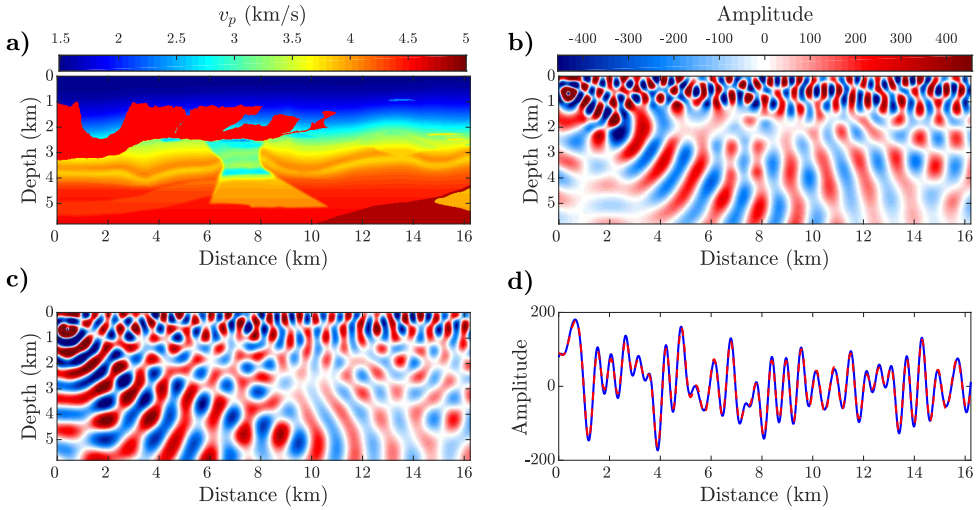

Figure 6.2 – (a) 2004 BP salt velocity model. (b) 3 Hz wavefield computed in (a). (c) Data-assimilated wavefield computed in the laterally-homogeneous velocity model described in the text. (d) Direct comparison between (b, blue) and (c, dashed red) at receiver positions.

To emphasize the impact of the wavefield accuracy on the velocity estimation, we compare the two velocity models inferred from IR-WRI when the BFI_1 and $\text{BFI}_{k,k\geq 1}$ wavefields are used at each (outer) IR-WRI iteration (Figures 6.4a and 6.4b). In the latter case ($k \geq 1$), we stop the inner wavefield-reconstruction iterations when $\|\mathbf{u}_{k+1} - \mathbf{u}_k\| \leq 1e^{-3}\|\mathbf{u}_k\|$. We perform IR-WRI with hybrid Tikhonov + TV (TT) regularization (Aghamiry et al., 2020b). The inversion setup and the hyperparameter tuning for IR-WRI are the same as those reviewed in the above reference. We use a classical continuation frequency strategy in the [3-13]Hz frequency band by proceeding over small batches of two frequencies with a frequency interval of 0.5Hz. We also perform three paths through the batches, where the starting and finishing frequencies of the paths are [3, 9.5], [3.5, 11.5], [5, 13] Hz respectively. A total of 45 frequency batches are processed. A direct comparison between the true velocity model, the initial model and the final IR-WRI velocity models at $x = 5$ km, $x = 9$ km and $x = 15$ km distances are shown in Figure 6.4c. As expected, the results show that the accuracy of the wavefield reconstruction has a significant impact on the quality of the estimated velocity model.

Indeed, the number of *backward-forward* iterations that is required to satisfy the stopping criterion is directly related to the accuracy of the velocity model. To gain some insights on the real computational overhead generated by our iterative wavefield reconstruction during a

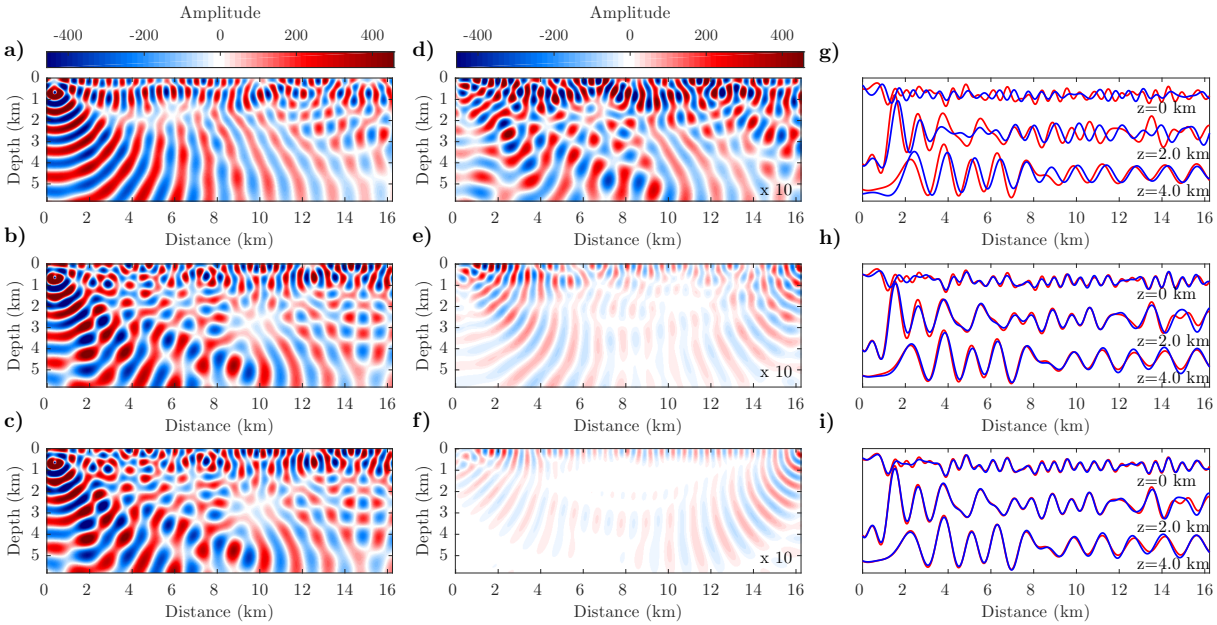

Figure 6.3 – Reconstructed wavefields BFI_k at iterations (a) 1, (b) 3, (c) 5 and (d-f) their difference with the target wavefield (Figure 6.2c). (g-i) Direct comparison between (g) BFI_1, (h) BFI_3 (i) BFI_5 (blue) and target (red) wavefields at different depths.

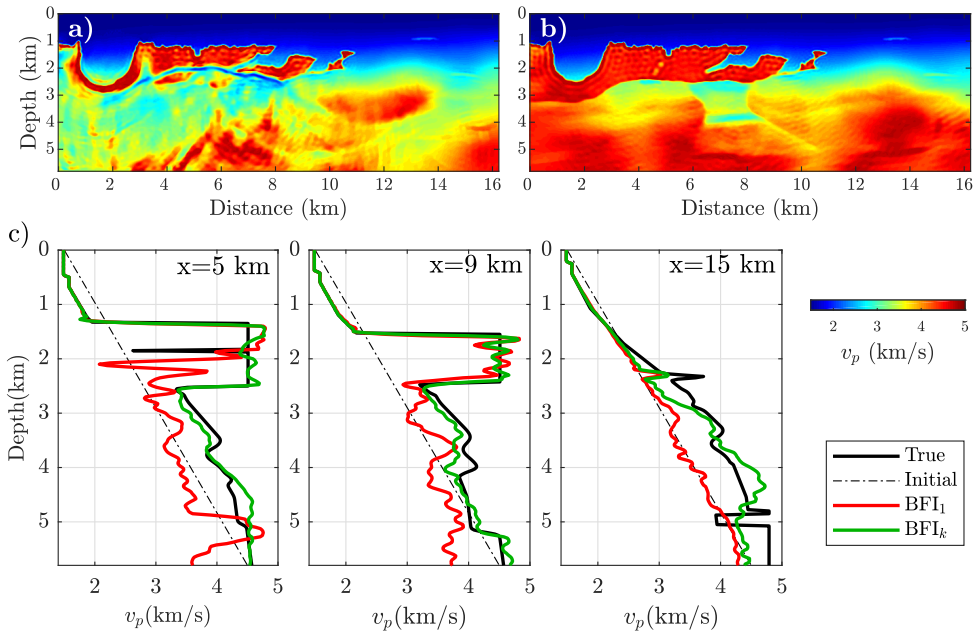

Figure 6.4 – Velocity models inferred from IR-WRI. (a) BFI_1 wavefields are used at each IR-WRI iteration. (b) BFI_k wavefields are used at each IR-WRI iteration in which k is determined with the stopping criterion described in the text. (c) A direct comparison at distances 5, 9 and 15 km between the true velocity model (solid black), the initial model (dash black), and the estimated velocity models shown in (a) (red) and (b) (green).

full run of IR-WRI, we show in Figure 6.5 the average number of *backward-forward* iterations per frequency batch. Although, these forward-backward recursions generate computational overhead, they keep the capability of the method to extend the search space. In addition, this number decreases as the IR-WRI proceeds over the frequency batches and it reduces to one at around the mid-point of the IR-WRI inversion.

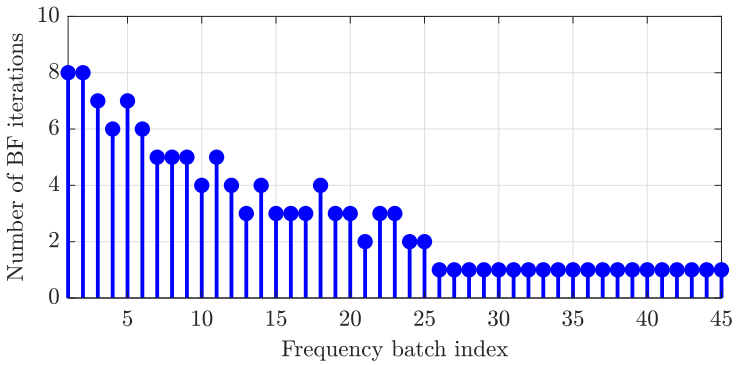

Figure 6.5 – The average number of *backward-forward* iterations required for wavefield reconstruction per frequency batch.

6.5 Conclusions

We have proposed a simple *backward-forward* time-stepping recursion to perform accurate data-assimilated wavefield reconstruction with explicit time stepping methods in the framework of WRI. Although this recursion introduces a computational overhead in the wavefield reconstruction during the early stages of the waveform inversion, they preserve the ability of WRI to extend the search space when crude initial velocity models are used. Therefore, the computational overhead introduced in the wavefield reconstruction should be easily compensated by the improved convergence speed of the WRI and the improved accuracy of the estimated velocity model.

Acknowledgments

We would like to thank the editors A. Malcolm, J. Blanch, and J. Shragge and reviewers R. Plessix and an anonymous for their insightful comments that greatly helped improve the paper. This study was partially funded by the SEISCOPE consortium (*http://seiscope2.osug.fr*), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXON-MOBIL, JGI, PETROBRAS, SCHLUMBERGER, SHELL, SINOPEC, and TOTAL. This study was granted access to the HPC resources of SIGAMM infrastructure (*https://www.oca.eu/fr/mesocentre-sigamm*) and CINES/IDRIS/TGCC under the allocation A0050410596 made by GENCI.

Chapter 7

Conclusion and perspectives

7.1 General conclusion

7.1.1 A new FWI framework: ADMM-based WRI

I developed the ADMM based wavefield reconstruction approach as a new framework for full waveform inversion. First, I recast the original FWI problem as a feasibility problem where both data and source misfit are processed as constraints. This feasibility problem is equivalent to constrained optimization problem where the objective function is identically zero. I solved this constrained problem with an augmented Lagrangian method (method of multiplier). A scaled form of the augmented Lagrangian shows that it is equivalent to a quadratic penalty function formed by the square of the constraint violations (data and source misfit), to which are added the running sum of the data and source residuals of previous iterations. In the objectives of the penalty function, the data and the source represent the right-hand side of the constraints, and the running sum of the data and source residuals correspond to the scaled Lagrange multipliers. Following van Leeuwen and Herrmann (2013), I performed the refinement of the two primal variables (the wavefield and the subsurface parameters) in alternating mode, using the solution of one primal subproblem as a passive variable for the next subproblem. Compared to a classical application of ADMM where separable convex subproblems are solved in parallel, this sequential resolution of the two primal subproblems allows me, on the one hand to linearize the parameter-estimation problem around the reconstructed wavefield and, on the other hand to manage the non-separability of the wavefield reconstruction and parameter estimation. This sequential solving of the two primal subproblems also prompts me to update the dual variable associated with the source residuals two times, once after each primal subproblem.
The augmented Lagrangian method gives me two ways of improving the accuracy of the minimizer: improve the accuracy of the Lagrange multiplier (namely, decrease the data and source residuals enough) or increase the penalty parameter, whereas the quadratic penalty approach gives me only one option: increase the penalty parameter. Accordingly the augmented-Lagrangian approach implemented in IR-WRI has logically converged to more accurate minimizers with a smaller number of iterations than the penalty method implemented in WRI when a fixed penalty parameter is used. The two leverages that control the accuracy of the IR-WRI minimizer also

allow me to use moderate values of the penalty parameter without impacting prohibitively convergence speed. This moderate values allows the method to fit the data during the early iterations with large wave-equation error, which provides the most suitable framework to enlarge the search space and account for large time shifts, while satisfying the wave-equation constraint with small error at the convergence point. According to the analogy between WRI and IR-WRI highlighted by the scaled-form Lagrangian, IR-WRI can be viewed as a self-adaptive penalty method, where a tedious and potentially unstable continuous increasing of the penalty parameter is replaced by a stable dual steepest-ascent updating of the data and source residuals. This iterative residual updating in a quadratic misfit function is a well known procedure to refine solution of a wide class of linear inverse problems.

7.1.2 Implementing regularization and bound constraints in WRI

At the next step to decrease the non-uniqueness of the possible minimizers and decrease the ill-posedness of the problem, I applied bound constraints and total-variation (TV) regularization on IR-WRI. To do this, I formulated IR-WRI as a TV minimization problem subject to data-fitting, wave-equation and bound constraints. At each iteration of the workflow, I performed a first ADMM step to break down the wavefield reconstruction and the parameter estimation into a sequence of two subproblems. Once one iteration of the wavefield reconstruction has been performed, I tackled the parameter estimation subproblem which involves the mixed $\ell_{1,2}$ TV norm of the model, the ℓ_2 wave-equation objective and the bound constraints. I applied a second ADMM step to decompose this multi-objective optimization problem into a sequence of simpler subproblems. Thanks to the introduction of auxiliary variables, I decoupled ℓ_1 and the ℓ_2 components of the penalty function following the split Bregman method proposed by Goldstein and Osher (2009). After this de-coupling, the subsurface parameter are first updated by minimizing the source residuals with one Gauss-Newton iteration involving the resolution of a sparse bound constrained and TV regularized system, before updating the auxiliary variables with proximal operators. This cycle is iterated until convergence.
My implementation of bound constrained and TV regularized IR-WRI with ADMM (or, equivalently split Bregman) provides a versatile framework to cascade constraints and regularization of different nature and is reasonably easy-to-tune due to the limited sensitivity of the augmented Lagrangian method to the choice of the penalty parameters. For challenging subsurface target with large-contrast structures such as salt bodies, my bound constrained and TV regularized IR-WRI has shown a high resilience to cycle skipping and noise and has efficiently mitigated high-frequency artifacts associated with Gibbs effects and multi scattering without detriment to the resolution of the imaging. However, one drawback of the TV regularization was related to the piecewise constant approximation underlying TV regularization, which tends to superimpose some blocky patterns on the smooth part of the subsurface. To overcome this issue, I used more complicated regularization function with the ability to reconstruct more complicated structures. A proper regularization should be driven by the shape and statistical characteristics of the medium to be imaged. For example, in geophysical imaging, the subsurface can be represented by a piece-wise smooth medium, that is a model which contains smoothly varying and blocky components. The widespread Tikhonov regularizations (Tikhonov et al., 2013)

rely on the smoothness assumption and hence fail to recover sharp interfaces of such media. Conversely, TV regularizations are based on blockiness assumption and hence are more suitable to image large contrasts. However, they generate undesirable staircase imprints in smooth regions. Regions characterized by smoothly-varying properties and those containing sharp contrasts have different statistical properties. The former are characterized by the normal prior, while the latter by a heavy tailed prior (Polson and Sokolov, 2019). Consequently, simultaneous recovery of both properties is difficult when one type of regularization is used (Tikhonov, TV, etc). To overcome this issue, a combination of different regularizations can be used (Gholami and Siahkoohi, 2010; Benning and Burger, 2018). A naive approach consists of the simple additive coupling or convex combinations (CC) of regularizations. Alternatively, Gholami and Hosseini (2013) proposed to explicitly decompose the model into several components of different statistical properties and use an appropriate regularization to reconstruct each component. Using this strategy, they combined Tikhonov and TV regularizations (TT regularization) to reconstruct piece-wise smooth media. The smooth components are captured by the Tikhonov regularization, while the blocky ones are determined by the TV counterpart. I showed that a compound regularization based upon infimal convolution outperforms the one based upon additive coupling (Bergmann et al., 2018). I also showed how the infimal convolution regularizer can be efficiently implemented by jointly updating the smooth and blocky subsurface components through variable projection. Alternatively, TGV regularized IR-WRI can be a suitable tool to reconstruct piecewise linear media and provides similar results than TT IR-WRI. I showed that such hybrid regularizations in the extended search-space IR-WRI potentially provide a suitable framework to reconstruct, without cycle skipping, large-contrast subsurface media from ultra-long offset seismic data.

7.1.3 Multi-parameter imaging: anisotropy and attenuation

At the next level, I extended IR-WRI to multi-parameter inversion in VTI acoustic media. I discussed different formulations of the VTI acoustic wave equation which fulfill the bilinearity w.r.t wavefield and subsurface parameters. First-order velocity stress form is often more convenient than the second-order counterpart to fulfill bilinearity, in particular if density (or buoyancy) is an optimization parameter. However, it may be not the most convenient one for frequency-domain wavefield reconstruction as the size of the linear system to be solved scales to the number of wavefield components. To bypass this issue, wavefield reconstruction and parameter estimation can be performed with different wave equations, provided they give consistent solutions. Bilinearity allows me to recast the parameter estimation subproblem as a linear subproblem and hence the waveform inversion as a biconvex problem. Also, ADMM-like optimization can be used to perform IR-WRI heuristically without bilinear wave equation, hence keeping the parameter estimation subproblem nonlinear. Accordingly, the nonlinear subproblem can be solved with a Newton algorithm rather than with a Gauss-Newton one. This nonlinear approach would amount to solve the multivariate nonlinear programming problem with the Augmented Lagrangian method when the two primal subproblems are solved in alternating mode. This nonlinear updating of the parameters may however require several inner Newton iterations per IR-WRI cycle, while Aghamiry et al. (2019c) showed that one inner

Gauss-Newton iteration without any line search was providing the most efficient convergence of IR-WRI when bilinearity is fulfilled. Indeed, the bilinearity specification limits the choice of subsurface parametrisation for parameter estimation. In the general case of triclinic elastodynamic equations, a subsurface parametrisation involving buoyancy and stiffness or compliance coefficients will be the most natural ones, as they correspond to the coefficients of the equation of motion and the Hooke's law. In the particular case of the VTI acoustic wave equation, I have developed a bilinear wave equation whose coefficients depend on the vertical wavespeed $\mathbf{v}_0$ and the Thomsen's parameters δ and ϵ. Although $\mathbf{v}_0$ and ϵ are coupled at wide scattering angles, the $(\mathbf{v}_0, \epsilon, \delta)$ parametrisation was promoted by Gholami et al. (2013b) and Gholami et al. (2013a) because the dominant parameter $\mathbf{v}_0$ has a radiation pattern which doesn't depend on the scattering angle, and hence can be reconstructed with a high resolution from wide-azimuth long-offset data. The counterpart is that updating the long wavelengths of the secondary parameter ϵ is challenging and requires so far a crude initial guess of its long wavelengths which can be used as prior to regularize the ϵ update. Comparing the results of IR-WRI when ϵ is used as a passive parameter and as an optimization parameter showed that the sensitivity of the inversion to ϵ remains small provided that a reasonable guess of its long wavelengths are provided in the starting model. Among the alternative parametrisations proposed for VTI acoustic FWI, Plessix and Cao (2011) proposed the $(\mathbf{v}_n, \mathbf{v}_h, \delta)$ or the $(\mathbf{v}_n, \eta, \delta)$ parametrisations for long-offset acquisition, while Alkhalifah and Plessix (2014) promoted the $(\mathbf{v}_h, \eta, \epsilon)$ parametrisation, where $\mathbf{v}_n = \mathbf{v}_0\sqrt{1+2\delta}$ is the so-called NMO velocity, $\mathbf{v}_h = \mathbf{v}_0\sqrt{1+2\epsilon}$ is the horizontal velocity and $\eta = (\epsilon - \delta)/(1+2\delta)$ represents the anellipticity of the anisotropy. I showed that the bilinearity of second order wave-equation is still valid for all of these parameterizations.

7.1.4 On the contribution of phase retrieval in WRI

FWI is a biconvex optimization problem for the wavefields and the subsurface parameters. Although the alternating optimization linearizes the two subproblems around one passive variable, their solutions are indeed impacted upon by the inaccuracies of these passive variables. In particular, it remains unclear how the phase inaccuracies of the reconstructed wavefields away from the receivers impact upon the parameter estimation during IR-WRI. To gain insights on this issue, I recast the parameter-estimation subproblem of IR-WRI as a phase retrieval problem (WIPR) in order to mitigate the role of the phase of the reconstructed wavefields during parameter estimation. By doing so, I indeed transform the linearized parameter-estimation subproblem into a nonconvex ill-posed problem. To overcome this issue, I further linearize this nonconvex problem around the subsurface model of the previous iteration with a majorization-minimization technique and manage the ill-posedness with suitable sparsity-promoting regularization. With the large contrast BP salt model, I first showed that the phase inaccuracies of the reconstructed wavefields prevent accurate estimation of the shallow sediments and accurate positioning in depth of the top of the salt after the first frequency batch inversion when I start the inversion from scratch (homogeneous initial velocity model). Then, I showed how WIPR improved significantly the reconstruction of the sedimentary background and the top of the salt during the early iterations of the inversion (that is, when the wavefields are not yet accurately reconstructed far away from the receivers). Sparsity-promoting regularization is however a nec-

essary ingredient to stabilize the phase retrieval inversion at greater depths where the amplitude information becomes more challenging to extract. This accurate positioning of the top salt is indeed important to speed up the convergence and improve the solution at higher frequencies when the phase information is re-injected in the inversion.

7.1.5 Toward effective time-domain WRI

The reconstructed wavefield has a crucial role in the IR-WRI technology. It is easy to reconstruct data-assimilated wavefields in the frequency domain by solving in the least-squares sense an overdetermined linear system gathering the weighted time-harmonic wave equation and the observation equation for each frequency, separately. Wavefield reconstruction in the time domain needs to be implemented with explicit (namely, matrix free) time stepping methods to be tractable. However, this explicit implementation is tricky because I have shown that it would require solving the wave equation with a right-hand side which depends on the adjoint-state wavefield associated with the unknown state wavefield. A previously proposed approximation to tackle this issue (Wang et al., 2016) replaces this adjoint wavefield by the one generated with the state wavefield that has been computed without data assimilation. This approximation can be questioned during the early WRI iterations when the wavefields computed with and without data assimilation differ significantly. I propose to refine the wavefield reconstruction with a simple iterative procedure. Each iteration of the wavefield refinement requires to solve one adjoint-state and one state problem with data-related and source-related correction terms in the right-hand sides. A numerical test performed with the BP salt model shows that a few iterations are enough to achieve accurate wavefield reconstruction with a crude velocity background model. Although this iterative refinement procedure generates computational overheads during one WRI iteration, it allows to preserve the capability of WRI to extend the search space. Also the computational overhead introduced in the wavefield reconstruction should be easily compensated by the improved convergence speed of the WRI and the improved accuracy of the estimated velocity model.

7.2 Perspectives

Here I review the main perspectives for the future extensions of the wavefield inversion technology which are under investigation. Therefore, I only address the key points of these extensions. The applications and validations of the proposed extensions are among our future work-plans and publications.

7.2.1 Assessing the limits of WRI with ultra long offset real data

All the proposed methods must be applied on real-data case studies to asses the actual limits of IR-WRI. I target two main long-offset case studies that were recently tackled at Geoazur. The first one involves a 3D Ocean Bottom Cable (OBC) data set collected above the shallow

water Valhall oil field in the North Sea. This oil field involves a large contrast chalky reservoir overlain by a gas cloud (Sirgue et al., 2010). So far, relevant results have been obtained by classical FWI starting from an accurate velocity model developed by reflection traveltime tomography (Operto et al., 2015). The goal will be to assess whether IR-WRI manages large contrasts generated by the chalk and the gas when the inversion starts from a crude initial velocity models and a 3 Hz starting frequency. Also, the waveguide generated by the water layer and the soft sediments as well as the gas cloud generate significant absorption and dispersion effects (Operto and Miniussi, 2018). It will be interesting to check that the viscous extension of IR-WRI manages to reconstruct attenuation by matching these wave phenomena.
The second case study is an ultra-long offset Ocean Bottom Seismometer (OBS) data set collected across the eastern Nankai subduction zone, offshore Japan (Dessa et al., 2004). The maximum offset reaches a value of 140 km. Indeed, the long propagation distances induced by these ultra long offsets make FWI highly sensitive to cycle skipping. Moreover, the structure of the subduction zone is extremely heterogeneous due to the subduction of several topography highs. Relevant results were also obtained for this case study by classical FWI (Górszczyk et al., 2017). However, a kinematically-accurate starting velocity model was developed by first-arrival traveltime tomography with a very careful quality control of the traveltime picking and residuals. Moreover, a Laplace-Fourier FWI with aggressive time damping allowed Górszczyk et al. (2017) to start the inversion with a frequency as small as 1.5 Hz. Finally, a quite tedious hierarchical workflow was designed where increasing frequencies, traveltimes and offsets are progressively injected in the inversion. In this context, reproducing these results with IR-WRI starting from a crude initial velocity model, a 3 Hz starting frequency and a basic frequency continuation strategy would be a major breaktrough. At this exploration scale, a key open question is related to the ability of IR-WRI to progressively image the crust from the surface, where observations are collected, down to Moho depths (15 to 30 km).

7.2.2 Efficient compressive sensing strategies for sparse OBN acquisitions

The resilience to cycle skipping makes IR-WRI amenable to ultra-long offset stationary-recording acquisitions such as those carried out with OBS (or Ocean Bottom Node (OBN) following the terminology used by the industry). These acquisitions are extensively used for deep crustal imaging by the academic community and are emerging in the oil industry for deep water exploration (Beaudoin and Ross, 2007; Shen et al., 2018). These acquisitions record a wide variety of wave types, including diving waves that undershoot the deepest targeted structures, and hence are amenable to broadband subsurface imaging provided that FWI manages properly nonlinearities. These wide-azimuth areal acquisitions are necessarily carried out at the expense of the receiver sampling, in particular in 3D. This coarse sampling of the acquisition parameters can in turn downsample the wavenumber spectrum of the FWI gradient below the Nyquist criterion, hence leading to spatial aliasing artifacts (Vermeer, 2012). A strong challenge is therefore to design optimal sparse acquisitions and efficient compressive sensing strategies allowing for the reconstruction of complex structures without aliasing.
I perform a first attempt in this direction by showing the ability of IR-WRI equipped with sparsity-promoting regularization in the parameter space for imaging large-contrast media from 2D stationary-recording acquisition (Aghamiry et al., 2020b). This preliminary study must

continue with the extension to 3D acquisition and more efficient regularization in the wavefield domain.

7.2.3 Regularization of the reconstructed wavefield

As I explained in this thesis, the wavefield reconstruction subproblem in IR-WRI can be viewed as the extrapolation of the wavefield in depth from its parcimonious measurements at the surface. This underdetermined problem is regularized by a physical constraint provided by the wave equation. Wavefields have a sparse representation in some domains. Therefore, implementing a sparsity-promoting regularization on the wavefield will complement efficiently the wave-equation constraint to extrapolate more accurately the wavefield in depth during the wavefield-reconstruction subproblem. This is indeed a key issue since the accuracy of the estimated model parameters critically depends on the accuracy of the reconstructed wavefields. This regularization will also complement the sparsity-promoting regularization in the model-domain to design efficient compressive sensing strategies in the framework of sparse acquisitions as discussed in the previous section.

7.2.4 Efficient wavefield reconstrucion for 3D IR-WRI

Which solver for 3D wavefield reconstruction?

Frequency-domain wavefield reconstruction is the most expensive task of the WRI cycle. This will become even more critical in 3D. Wavefield reconstruction requires the solution of a large and sparse system of linear equations with multiple right-hand sides. These solutions can be computed with direct or iterative solvers depending on the size of the computational domain and the number of right-hand sides. Continuous advances in the development of efficient sparse direct solvers with efficient low-rank compression strategies allow to tackle problems of increasing size, while the sparsity of the seismic sources can be exploited during the solution step (Amestoy et al., 2015, 2016; Mary, 2017; Amestoy et al., 2018). Alternatively, iterative solvers equipped with suitable preconditioners for Helmholtz problems can be used when the number of unknowns exceeds 100 millions (Dolean et al., 2015). The choice of the optimal solver and the discretization of the wave equation will be a key issue to tackle efficiently realistic 3D IR-WRI applications.
When direct solvers can be used, the computational effort can be significantly reduced by limiting the number of LU factorization during the IR-WRI iterations. One possible strategy to achieve this goal is to recast the contrast-source inversion method in the framework of IR-WRI. In contrast-source inversion, the unknowns are the contrast (which can be viewed as model parameter perturbations) and the contrast source (which represents the interaction of the wavefield with the contrasts) (Abubakar et al., 2009, 2011). This re-parametrization of the waveform inversion allows one to perform a single LU decomposition in iterations of one frequency inversion (instead of one LU decomposition per iteration). As in WRI, the contrast source method linearly combines two objective functions related to the observation equation and the wave equation. These two functionals are combined with fixed weights such that the two objective

functions are made dimensionless. This technology should be readily recast in the framework of IR-WRI by estimating these weights through the scaled Lagrangian multipliers such that a versatile search space extension can be applied.
Alternatively, we can view to avoid performing a new LU factorization for wavefield reconstruction when the parameter updates of the previous iteration were small (for example, during the late iterations of one frequency inversion). Instead, we can reconstruct the wavefields more efficiently from the LU factors of the previous iteration by iterative refinement.

Sparse storage of the dual variables

The dual variables or Lagrange multipliers must be stored during the inversion. Since keeping dual variables is memory demanding, it will be an obstacle to use IR-WRI for 3D and large scale applications. Because of sparse representation of dual variables in some specific domains, it will be possible to store their sparse coefficients or most of their energy by keeping only a few coefficients and overcome the memory issue.

7.2.5 Extending IR-WRI to elastic FWI

I showed in this thesis the bilinearity of the wave-equation in triclinic media. So, it will be conceptually possible to extend IR-WRI to more complicated physics even if the 3D elastic wavefield reconstruction will be a significant computational challenge.

7.2.6 Extension of ADMM-based WRI to RWI

RWI (Reflection Waveform Inversion) is an adaptation of FWI for reflection data (Xu et al., 2012a; Brossier et al., 2015; Zhou et al., 2015, 2018). For narrow-azimuth/short-angle acquisitions, it is well acknowledged that intermediate wavelengths belong to the null space of the seismic imaging problem (Claerbout, 1985; Jannane et al., 1989; Neves and Singh, 1996). Accordingly, unlike FWI, RWI relies on an explicit scale separation between a long-wavelength background velocity model and a short-wavelength reflectivity model. In RWI, the smooth background velocity model is reconstructed by a tomography-like waveform inversion, while the reflectivity is built either by a linear or nonlinear migration-like waveform inversion. These two sub-problems are generally solved in alternating mode, the reflectivity being used as a passive variable when the velocity model is updated and *vice versa*. Two difficulties related to the velocity model building step is the nonlinearity of the problem as in FWI and the potential inaccuracies of the passive reflectivity, which inject undesired amplitude-related residuals during the velocity update (these amplitude-related residuals can update the background velocities even if these velocities are kinematically accurate). It will be interesting to recast RWI in the IR-WRI framework to mitigate the impact of cycle skipping during the velocity model estimation and better account for the errors in the reflectivity. Some attempts have been proposed to mitigate the non linearity of the velocity model estimation by adding an extra dimension in the model space through time lags (Biondi and Almomin, 2014). However, these approaches

are very computationally intensive. Therefore, a clear assessment of the computational cost of ADMM-based RWI compared to these concurrent approaches will have to be done.

7.2.7 IR-WRI with adaptive regularization on model parameters

Since model reconstruction subprobem is linear in IR-WRI, it is easy to apply different kind of regularization. Finding an optimum regularization for IR-WRI is still an open problem. For example improving regularization with prior on structural dips can be one possibility to enhance the reconstructed model.
Another possibility can be using data-driven regularizations. The base of a good regularization is the prior expressed by the regularizer. The TT regularization (Aghamiry et al., 2020b) goes one step further and inject normal as well as long-tail distribution to the inversion, simultaneously. Nevertheless, tailoring a suitable and easy to implement prior for describing geophysical models is a nontrivial task. We are working on a general black-box regularization algorithm for solving linear/nonlinear inverse problems which admits data-driven or empirical priors estimated by sophisticated denoising algorithms. The solution is gradually fed by the prior provided by the denoiser without asking for any information about its functional form, thus treating the denoiser as a black-box regularizer. Beyond efficiency of such methods, they show a local behavior and they can handle models that have different structures with different priors in different parts of the models. Aghamiry et al. (2020c) shows our preliminary efforts for adaptive regularization on FWI/IR-WRI to update velocity model.

7.2.8 Improving the convergence speed of ADMM-based WRI: adaptive ADMM

ADMM has a linear convergence rate when one of the objective function terms is strongly convex. Using new versions of ADMM such as adaptive ADMM (Goldstein et al., 2015) for example) to speed up the convergence rate of IR-WRI is another avenue for research which deserves further investigations.

7.2.9 Introducing other priors in ADMM-based WRI: empirical parameter relationships and well logs

FWI is an ill-posed problem with a huge null space. I have shown that ADMM provides a versatile framework to inject various priors in IR-WRI. Two obvious priors which can be injected in waveform inversion problems are well logs (Asnaashari et al., 2013; Baumstein, 2013, 2014, 2015) and empirical relationships between parameters of different nature (Duan and Sava, 2016).

7.2.10 Computationnally efficient time-domain ADMM-based WRI

I showed our preliminary efforts for time-domain IR-WRI. There are some issues that must be solved for time-domain IR-WRI. One of them is the storage of source residuals and the adjoint wavefields. Another possible compression strategy is to perform wavefield reconstruction in the time domain and parameter estimation in the coarsified frequency domain following the original idea of Nihei and Li (2007) and Sirgue et al. (2008).

7.2.11 Application of IR-WRI to medical imaging

Multi-parameter frequency-domain IR-WRI for wavespeed, attenuation and density equipped with bound constraints and sparsity-promoting regularizations should find obvious applications in medical imaging (e.g. Sandhu et al., 2015).

7.2.12 Complex-valued optimization to joint estimation of velocity and attenuation

Attenuation is classically implemented in frequency-domain wave simulation with complex velocity (Toksöz and Johnston, 1981). Classical frequency-domain FWI and also the method which is proposed in this thesis for velocity and attenuation reconstruction is generally performed by processing the complex velocity as two independent real variables. In complex-valued optimization, we jointly reconstruct the velocity and attenuation with regularized IR-WRI by using the complex velocity as the optimization parameter. Aghamiry et al. (2019d) shows our preliminary efforts for complex-valued optimization to joint estimation of velocity and attenuation.

Bibliography

Abubakar, A., Hu, W., Habashy, T. M., and van den Berg, P. M. (2009). Application of the finite-difference contrast-source inversion algorithm to seismic full-waveform data. *Geophysics*, 74(6):WCC47–WCC58.

Abubakar, A., Pan, G., Li, M., Zhang, L., Habashy, T. M., and van den Berg, P. (2011). Three-dimensional seismic full-waveform inversion using the finite-difference contrast source inversion method. *Geophysical prospecting*, 59:874–888.

Aghamiry, H., Gholami, A., and Operto, S. (2018a). Hybrid tikhonov + total-variation regularization for imaging large-contrast media by full-waveform inversion. In *Expanded Abstracts, 88^{th} Annual SEG Meeting (Anaheim)*, pages 1253–1257.

Aghamiry, H., Gholami, A., and Operto, S. (2018b). Imaging contrasted media with total variation constrained full waveform inversion and split Bregman iterations. In *Expanded Abstracts, 80^{th} Annual EAGE Meeting (Copenhagen)*.

Aghamiry, H., Gholami, A., and Operto, S. (2018c). Improving full-waveform inversion based on wavefield reconstruction via Bregman iterations. In *Expanded Abstracts, 80^{th} Annual EAGE Meeting (Copenhagen)*.

Aghamiry, H., Gholami, A., and Operto, S. (2019a). Admm-based multi-parameter wavefield reconstruction inversion in VTI acoustic media with TV regularization. *Geophysical Journal International*, 219(2):1316–1333.

Aghamiry, H., Gholami, A., and Operto, S. (2019b). Implementing bound constraints and total-variation regularization in extended full waveform inversion with the alternating direction method of multiplier: application to large contrast media. *Geophysical Journal International*, 218(2):855–872.

Aghamiry, H., Gholami, A., and Operto, S. (2019c). Improving full-waveform inversion by wavefield reconstruction with alternating direction method of multipliers. *Geophysics*, 84(1):R139–R162.

Aghamiry, H., Gholami, A., and Operto, S. (2019d). Joint estimation of velocity and attenuation by frequency-domain TV-regularized wavefield reconstruction inversion. In *81^{th} Annual EAGE Meeting (London) - WS01: Attenuation: Challenges in Modelling and Imaging at the Exploration Scale*.

Aghamiry, H., Gholami, A., and Operto, S. (2019e). Multi-parameter ADMM-based wavefield reconstruction inversion in VTI acoustic media. In *81^{th} Annual EAGE Meeting (London)*.

Aghamiry, H., Gholami, A., and Operto, S. (2020a). Accurate and efficient wavefield reconstruction in the time domain. *Geophysics*, 85(2):A7–A12.

Aghamiry, H., Gholami, A., and Operto, S. (2020b). Compound regularization of Full-Waveform Inversion for imaging piecewise media. *IEEE Transactions on Geoscience and Remote Sensing*, 58(2):1192–1204.

Aghamiry, H., Gholami, A., and Operto, S. (2020c). Full waveform inversion with adaptive regularization. *arXiv:2001.09846*.

Aghamiry, H., Gholami, A., and Operto, S. (2020d). Multi-parameter wavefield reconstruction inversion for wavespeed and attenuation with bound constraints and total variation regularization. *Geophysics*, doi: 10.1190/geo2019-0596.1.

Aghamiry, H., Gholami, A., and Operto, S. (2020e). Robust wavefield inversion with phase retrieval. *Geophysical Journal International*, 221(2):1327–1340.

Akçelik, V. (2002). *Multiscale Newton-Krylov methods for inverse acoustic wave propagation*. PhD thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania.

Aki, K. and Richards, P. G. (2002). *Quantitative seismology, theory and methods, second edition*. University Science Books, Sausalito,California.

Alkhalifah, T. and Plessix, R. (2014). A recipe for practical full-waveform inversion in anisotropic media: An analytical parameter resolution study. *Geophysics*, 79(3):R91–R101.

Alkhalifah, T., Sun, B. B., and Wu, Z. (2018). Full model wavenumber inversion: Identifying sources of information for the elusive middle model wavenumbers. *Geophysics*, 83(6):R597–R610.

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

Amestoy, P., Buttari, A., L'Excellent, J.-Y., and Mary, T. (2018). On the complexity of the block low-rank multifrontal factorization. *SIAM Journal on Scientific Computing*, 49(4):A1710–A1740.

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

Aminzadeh, F., Brac, J., and Kunz, T. (1997). *3-D Salt and Overthrust models*. SEG/EAGE 3-D Modeling Series No.1.

Anagaw, A. and Sacchi, M. (2011). Regularized 2D acoustic full waveform inversion. In *Expanded Abstracts*. EAGE.

Aravkin, A. Y., Drusvyatskiy, D., and van Leeuwen, T. (2017). Efficient quadratic penalization through the partial minimization technique. *IEEE Transactions on Automatic Control*, 63:2131–2138.

Askan, A., Akcelik, V., Bielak, J., and Ghattas, O. (2007). Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures. *Bulletin of the Seismological Society of America*, 97(6):1990–2008.

Asnaashari, A., Brossier, R., Garambois, S., Audebert, F., Thore, P., and Virieux, J. (2013). Regularized seismic full waveform inversion with prior model information. *Geophysics*, 78(2):R25–R36.

Aster, R. C., Borchers, B., and Thurber, C. H. (2004). *Parameter Estimation and Inverse Problems*. Academic Press.

Auld, B. A. (1990). *Acoustic fields and waves in solids, volume 1 and 2*. Krieger Publishing Co.

Auroux, D. and Blum, J. (2008). A nudging-based data assimilation method for oceanographic problems: the back and forth nudging (bfn) algorithm. *Nonlinear Processes in Geophysics*, 15:305–319.

Baeten, G., de Maag, J. W., Plessix, R.-E., Klaassen, R., Qureshi, T., Kleemeyer, M., ten Kroode, F., and Rujie, Z. (2013). The use of low frequencies in a full-waveform inversion and impedance inversion land seismic case study. *Geophysical Prospecting*, 61(4):701–711.

Banerjee, B., Walsh, T. F., Aquino, W., and Bonnet, M. (2013). Large scale parameter estimation problems in frequency-domain elastodynamics using an error in constitutive equation functional. *Computational Methods in Applied Mechanics Engineering*, 253:60–72.

Baumstein, A. (2013). POCS-based geophysical constraints in multi-parameter Full Wavefield Inversion. In *EAGE, 75th EAGE Conference and Exhibition incorporating SPE EUROPEC 2013*.

Baumstein, A. (2014). Extended subspace method for attenuation of crosstalk in multi-parameter full wavefield inversion. In *SEG Technical Program Expanded Abstracts 2014*, pages 1121–1125.

Baumstein, A. (2015). Borehole-constrained multi-parameter Full Waveform Inversion. In *Expanded Abstracts, 77th Annual EAGE Meeting (Madrid)*.

Beaudoin, G. and Ross, A. (2007). Field design and operation of a novel deepwater, wide-azimuth node seismic survey. *The Leading Edge*, 26:494–503.

Benning, M. and Burger, M. (2018). Modern regularization methods for inverse problems. *Acta Numerica*, 27:1–111.

Benning, M., Knoll, F., Schönlieb, C.-B., and Valkonen, T. (2015). Preconditioned admm with nonlinear operator constraint. In *IFIP Conference on System Modeling and Optimization*, pages 117–126. Springer.

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

Bergmann, R., Fitschen, J. H., Persch, J., and Steidl, G. (2018). Priors with coupled first and second order differences for manyfold-valued image processing. *Journal of mathematical imaging and vision*, 60:1459–1481.

Bertsekas, D. P. (2016). *Nonlinear programming, Third edition*. Athena Scientific.

Billette, F. J. and Brandsberg-Dahl, S. (2004). The 2004 BP velocity benchmark. In *Extended Abstracts, 67th Annual EAGE Conference & Exhibition, Madrid, Spain*, page B035.

Biondi, B. and Almomin, A. (2014). Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion. *Geophysics*, 79(3):WA129–WA140.

Bioucas-Dias, J. M. and Figueiredo, M. A. (2008). An iterative algorithm for linear inverse problems with compound regularizers. In *2008 15th IEEE International Conference on Image Processing*, pages 685–688. IEEE.

Böhmer, K. and Stetter, H. (1984). Defect correction methods. *Theory and applications*, 5.

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2010). Distributed optimization and statistical learning via the alternating direction of multipliers. *Foundations and trends in machine learning*, 3(1):1–122.

Boyd, S. P. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Boyle, J. P. and Dykstra, R. L. (1986). A method for finding projections onto the intersection of convex sets in Hilbert spaces. In Dykstra, R., Robertson, T., and Wright, F. T., editors, *Advances in Order Restricted Statistical Inference: Proceedings of the Symposium on Order Restricted Statistical Inference held in Iowa City, Iowa, September 11–13, 1985*, pages 28–47. Springer New York, New York, NY.

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

Brandsberg-Dahl, S., Chemingui, N., Valenciano, A., Ramos-Martinez, J., and Qiu, L. (2017). FWI for model updates in large-contrast media. *The Leading Edge*, Special section: Full Waveform Inversion Part II(1):81–87.

Brás, N. B., Bioucas-Dias, J., Martins, R. C., and Serra, A. C. (2012). An Alternating Direction algorithm for Total Variation reconstruction of distributed parameters. *IEEE transactions on image processing*, 21(6):3004–3016.

Bredies, K., Kunisch, K., and Pock, T. (2010). Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526.

Bregman, L. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*.

Brossier, R., Operto, S., and Virieux, J. (2015). Velocity model building from seismic reflection data by full waveform inversion. *Geophysical Prospecting*, 63:354–367.

Burvall, A., Lundstrom, U., Takman, P. A. C., Larsson, D. H., and Hertz, H. M. (2011). Phase retrieval in X-ray phase-contrast imaging suitable for tomography. *Optics Express*, 19(11):10359–10376.

Candès, E. J., Romberg, J. K., and Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Pure and Applied Mathematics*, 59(8):1207–1223.

Candes, E. J., Strohmer, T., and Voroninski, V. (2013). Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274.

Candès, E. J. and Wakin, M. B. (2008). An introduction to compressive sensing. *IEEE SIGNAL PROCESSING MAGAZINE*, March:21–30.

Carcione, J. M. (2015). *Wave Fields in Real Media, Wave Propagation in Anisotropic, Anelastic, Porous and Electromagnetic Media*. Elsevier, third edition edition.

Chen, Z., Cheng, D., Feng, W., and Wu, T. (2013). An optimal 9-point finite difference scheme for the Helmholtz equation with PML. *International Journal of Numerical Analysis & Modeling*, 10(2).

Cheng, X., Jiao, K., Sun, D., and Vigh, D. (2016). Multiparameter estimation with acoustic vertical transverse isotropic full-waveform inversion of surface seismic data. *Interpretation*, 4(4):SU1–SU16.

Choi, Y. and Alkhalifah, T. (2015). Unwrapped phase inversion with an exponential damping. *Geophysics*, 80(5):251–264.

Claerbout, J. (1985). *Imaging the Earth's interior*. Blackwell Scientific Publication.

Combettes, P. L. and Pesquet, J.-C. (2011). Proximal splitting methods in signal processing. In Bauschke, H. H., Burachik, R. S., Combettes, P. L., Elser, V., Luke, D. R., and Wolkowicz, H., editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, volume 49 of *Springer Optimization and Its Applications*, pages 185–212. Springer New York.

da Silva, N. and Yao, G. (2017). Wavefield reconstruction inversion with a multiplicative cost function. *Inverse problems*, 34(1):015004.

da Siva, N. V., Yao, G., and Warner, M. (2019). Semiglobal viscoacoustic full-waveform inversion. *Geophysics*, 84(2):R271–R293.

Debens, H. A., Warner, M., Umpleby, A., and da Silva, N. V. (2015). Global anisotropic 3d fwi. In *SEG Technical Program Expanded Abstracts*.

Dessa, J. X., Operto, S., Kodaira, S., Nakanishi, A., Pascal, G., Uhira, K., and Kaneda, Y. (2004). Deep seismic imaging of the eastern Nankai trough (Japan) from multifold ocean bottom seismometer data by combined traveltime tomography and prestack depth migration. *Journal of Geophysical Research*, 109(B02111):doi:10.1029/2003JB002689.

Dolean, V., Jolivet, P., and Nataf, F. (2015). *An introduction to domain decomposition methods - Algorithms, theory, and parallel implementation*. SIAM.

Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE transactions on information theory*, 41(3):613–627.

Donoho, D. L. (2006a). Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306.

Donoho, D. L. (2006b). For most large underdetermined systems of linear equations the minimal $l1$-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, LIX:0797–0829.

Duan, Y. and Sava, P. (2016). Elastic wavefield tomography with physical model constraints. *Geophysics*, 81(6):R447–R456.

Duff, I. S., Erisman, A. M., and Reid, J. K. (1986). *Direct methods for sparse matrices, second edition*. Oxford Science Publications, Oxford, U. K.

Duveneck, E. and Bakker, P. M. (2011). Stable P-wave modeling for reverse-time migration in tilted TI media. *Geophysics*, 76(2):S65–S75.

Duveneck, E., Milcik, P., Bakker, P. M., and Perkins, C. (2008). Acoustic VTI wave equations and their application for anisotropic reverse-time migration. *SEG Technical Program Expanded Abstracts*, 27(1):2186–2190.

Eldar, Y. C., Hammen, N., and Mixon, D. G. (2016). Recent advances in phase retrieval. *IEEE signal processing magazine, Lecture Notes*, September:158–162.

Engquist, B. and Froese, B. D. (2014). Application of the Wasserstein metric to seismic signals. *Communications in Mathematical Science*, 12(5):979–988.

Engquist, B. and Majda, A. (1977). Absorbing boundary conditions of the numerical simulation of waves. *Mathematics of Computation*, 31:629–651.

Epanomeritakis, I., Akçelik, V., Ghattas, O., and Bielak, J. (2008). A Newton-CG method for large-scale three-dimensional elastic full waveform seismic inversion. *Inverse Problems*, 24:1–26.

Esser, E. (2009). Applications of lagrangian-based alternating direction methods and connections to split Bregman. Technical report, UCLA.

Esser, E., Guasch, L., van Leeuwen, T., Aravkin, A. Y., and Herrmann, F. J. (2018). Total variation regularization strategies in Full-Waveform Inversion. *SIAM Journal Imaging Sciences*, 11(1):376–406.

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

Fienup, J. R. (1982). Phase retrieval algorithms: a comparison. *Applied optics*, 21(15):2758–2769.

Fogel, F., Waldspurger, I., and d'Aspremont, A. (2016). Phase retrieval for imaging problems. *Mathematical programming computation*, 8(3):311–335.

Frick, K., Lorenz, D. A., and Resmerita, E. (2011). Morozov's principle for the augmented lagrangian method applied to linear inverse problems. *Multiscale Modeling & Simulation*, 9(4):1528–1548.

Fu, L. and Symes, W. W. (2017). A discrepancy-based penalty method for extended waveform inversion. *Geophysics*, R282-R298:78–82.

Futterman, W. (1962). Dispersive body waves. *Journal Geophysical Research*, 67:5279–5291.

Gerchberg, R. W. (1972). A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237–246.

Gholami, A. (2013). Sparse time–frequency decomposition and some applications. *IEEE Transactions on Geoscience and Remote Sensing*, 51(6):3598–3604.

Gholami, A. (2014). Phase retrieval through regularization for seismic problems. *Geophysics*, 79(5):V153–V164.

Gholami, A. (2015). Nonlinear multichannel impedance inversion by total-variation regularization. *Geophysics*, 80(5):R217–R224.

Gholami, A. (2016). A fast automatic multichannel blind seismic inversion for high-resolution impedance recovery. *Geophysics*.

Gholami, A., Aghamiry, H., and Abbasi, M. (2018). Constrained nonlinear AVO inversion using Zoeppritz equations. *Geophysics*, 83(3):R245–R255.

Gholami, A. and Gheymasi, H. M. (2016). Regularization of geophysical ill-posed problems by iteratively re-weighted and refined least-squares. *Computer Geosciences*, 20:19–33.

Gholami, A. and Hosseini, S. M. (2013). A balanced combination of tikhonov and total variation regularizations for reconstruction of piecewise-smooth signals. *Signal Processing*, 93:1945–1960.

Gholami, A. and Naeini, E. Z. (2019). 3d dix inversion using bound-constrained tv regularization. *Geophysics*, 84(3):1–43.

Gholami, A. and Sacchi, M. D. (2013). Fast 3D blind seismic deconvolution via constrained total variation and gcv. *SIAM Journal on Imaging Sciences*, 6(4):2350–2369.

Gholami, A. and Siahkoohi, H. (2010). Regularization of linear and non-linear geophysical ill-posed problems with joint sparsity constraints. *Geophysical Journal International*, 180(2):871–882.

Gholami, Y., Brossier, R., Operto, S., Prieux, V., Ribodetti, A., and Virieux, J. (2013a). Which parametrization is suitable for acoustic VTI full waveform inversion? - Part 2: application to Valhall. *Geophysics*, 78(2):R107–R124.

Gholami, Y., Brossier, R., Operto, S., Ribodetti, A., and Virieux, J. (2013b). Which parametrization is suitable for acoustic VTI full waveform inversion? - Part 1: sensitivity and trade-off analysis. *Geophysics*, 78(2):R81–R105.

Glowinski, R., Osher, S. J., and Yin, W. (2017). *Splitting Methods in Communication, Imaging, Science, and Engineering*. Springer.

Goldstein, T., Li, M., and Yuan, X. (2015). Adaptive primal-dual splitting methods for statistical learning and image processing. In *Advances in Neural Information Processing Systems*, pages 2089–2097.

Goldstein, T. and Osher, S. (2009). The split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343.

Górszczyk, A., Operto, S., and Malinowski, M. (2017). Toward a robust workflow for deep crustal imaging by FWI of OBS data: The eastern nankai trough revisited. *Journal of Geophysical Research: Solid Earth*, 122(6):4601–4630.

Gray, S. H., Etgen, J., Dellinger, J., and Whitmore, D. (2001). Seismic migration problems and solutions. *Geophysics*, 66(5):1622–1640.

Grechka, V., Zhang, L., and Rector III, J. W. (2004). Shear waves in acoustic anisotropic media. *Geophysics*, 69:576–582.

Guasch, L., Warner, M., and Ravaut, C. (2019). Adaptive waveform inversion: Practice. *Geophysics*, 84(3):R447–R461.

Guitton, A. (2012). Blocky regularization schemes for full waveform inversion. *Geophysical Prospecting*, 60(5):870–884.

Haber, E., Ascher, U. M., and Oldenburg, D. (2000). On optimization techniques for solving nonlinear inverse problems. *Inverse problems*, 16(5):1263.

Hak, B. and Mulder, W. A. (2011). Seismic attenuation imaging with causality. *Geophysical Journal International*, 184(1):439–451.

Hansen, P. C. (2005). *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*, volume 4. Siam.

Harrison, R. W. (1993). Phase problem in crystallography. *JOSA a*, 10(5):1046–1055.

He, B., Liu, H., Wang, Z., and Yuan, X. (2014). A strictly contractive peaceman–rachford splitting method for convex programming. *SIAM Journal on Optimization*, 24(3):1011–1040.

Helbig, K. (2013). What Kelvin might have written about Elasticity. *Geophysical prospecting*, 61:O21–O31.

Herrmann, F. J. (2010). Randomized sampling and sparsity: Getting more information from fewer samples. *Geophysics*, 75(6):WB173–WB187.

Hestenes, M. R. (1969). Multiplier and gradient methods. *Journal of optimization theory and applications*, 4(5):303–320.

Hicks, G. J. and Pratt, R. G. (2001). Reflection waveform inversion using local descent methods: estimating attenuation and velocity over a gas-sand deposit. *Geophysics*, 66(2):598–612.

Huang, G., Nammour, R., and Symes, W. (2017). Full-waveform inversion with source-receiver extension. *Geophysics*, 82(3):R153–R171.

Huang, G., Nammour, R., and Symes, W. W. (2018a). Source-independent extended waveform inversion based on space-time source extension: Frequency-domain implementation. *Geophysics*, 83(5):R449–R461.

Huang, G., Nammour, R., and Symes, W. W. (2018b). Volume source-based extended waveform inversion. *Geophysics*, 83(5):R369–387.

Huang, Y., Maio, J., and Sheng, J. (2019). FWI salt model update with sparse nodes: a feasibility test on OGO FAN survey. In *81^{th} Annual EAGE Meeting (London)*.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures, and Monte Carlo. *The Annals of Statistics*, 1(5):799–821.

Hustedt, B., Operto, S., and Virieux, J. (2004). Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling. *Geophysical Journal International*, 157:1269–1296.

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

Jiang, X., So, H. C., and Liu, X. (2017). Robust phase retrieval via ADMM with outliers. *arXiv:1702.06157v1*.

Jiao, Y., Jin, Q., Lu, X., and Wang, W. (2016). Alternating direction method of multipliers for linear inverse problems. *SIAM Journal on Numerical Analysis*, 54(4):2114–2137.

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

Jo, C. H., Shin, C., and Suh, J. H. (1996). An optimal 9-point, finite-difference, frequency-space 2D scalar extrapolator. *Geophysics*, 61:529–537.

Kamei, R. and Pratt, R. G. (2008). Waveform tomography strategies for imaging attenuation structure for cross-hole data. In *Extended Abstracts*, page F019.

Kamei, R. and Pratt, R. G. (2013). Inversion strategies for visco-acoustic waveform inversion. *Geophysical Journal International*, 194:859–894.

Kolsky, H. (1956). The propagation of stress pulses in viscoelastic solids. *Philosophical Magazine*, 1:693–710.

Kreutz-Delgado, K. (2009). The complex gradient operator and the cr-calculus. *arXiv preprint arXiv:0906.4835*.

Kurzmann, A., Przebindowska, A., Kohn, D., and Bohlen, T. (2013). Acoustic full waveform tomography in the presence of attenuation: a sensitivity analysis. *Geophysical Journal International*, 195(2):985–1000.

Lacasse, M., Denli, H., White, L., Gudipati, V., Lee, S., and Tan, S. (2019). Accounting for heterogeneous attenuation in full-wavefield inversion. In *81th Annual EAGE Meeting (London) - WS01: Attenuation: Challenges in Modelling and Imaging at the Exploration Scale*.

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

Lange, K. (2016). *MM optimization algorithms*, volume 147. SIAM.

Lefkimmiatis, S., Bourquard, A., and Unser, M. (2012). Hessian-based norm regularization for image restoration with biomedical applications. *IEEE Transactions on Image Processing*, 21(3):983–995.

Li, X., Esser, E., and Herrmann, F. J. (2016). Modified gauss-newton full-waveform inversion explained - why sparsity-promoting updates do matter. *Geophysics*, 81(3):R99–R112.

Liao, O. and McMechan, G. A. (1995). 2.5D full-wavefield viscoacoustic inversion. *Geophysical Prospecting*, 43:1043–1059.

Loris, I. and Verhoeven, C. (2012). Iterative algorithms for total variation-like reconstructions in seismic tomography. *GEM-International Journal on Geomathematics*, 3(2):179–208.

Luo, J. and Wu, R.-S. (2015). Seismic envelope inversion: reduction of local minima and noise resistance. *Geophysical Prospecting*, 63(3):597–614.

Luo, S. and Sava, P. (2011). A deconvolution-based objective function for wave-equation inversion. *SEG Technical Program Expanded Abstracts*, 30(1):2788–2792.

Ma, Y. and Hale, D. (2013). Wave-equation reflection traveltime inversion with dynamic warping and full waveform inversion. *Geophysics*, 78(6):R223–R233.

Maharramov, M. and Biondi, B. (2015). Robust simultaneous time-lapse full-waveform inversion with total-variation regularization of model difference. In *77th EAGE Conference & Exhibition, Expanded Abstracts, We P3 09*.

Maharramov, M. and Levin, S. A. (2015). Total-variation minimization with bound constraints. *arXiv preprint arXiv:1505.05694*.

Malinowski, M., Operto, S., and Ribodetti, A. (2011). High-resolution seismic attenuation imaging from wide-aperture onshore data by visco-acoustic frequency-domain full waveform inversion. *Geophysical Journal International*, 186(3):1179–1204.

Marfurt, K. (1984). Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations. *Geophysics*, 49:533–549.

Martin, G. S., Wiley, R., and Marfurt, K. J. (2006). Marmousi2: An elastic upgrade for Marmousi. *The Leading Edge*, 25(2):156–166.

Mary, T. (2017). *Block Low-Rank multifrontal solvers: complexity, performance and scalability*. PhD thesis, PhD thesis, Université de Toulouse.

Menke, W. (2012). *Geophysical data analysis: Discrete inverse theory*, volume 45. Academic press.

Métivier, L., Allain, A., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J. (2018). Optimal transport for mitigating cycle skipping in full waveform inversion: a graph space transform approach. *Geophysics*, 83(5):R515–R540.

Métivier, L., Brossier, R., Mérigot, Q., and Oudet, E. (2019). A graph space optimal transport distance as a generalization of L^p distances: application to a seismic imaging inverse problem. *Inverse Problems*, 35(8):085001.

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

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

Métivier, L., Brossier, R., Operto, S., and J., V. (2017). Full waveform inversion and the truncated Newton method. *SIAM Review*, 59(1):153–195.

Métivier, L., Brossier, R., Operto, S., and Virieux, J. (2015). Acoustic multi-parameter FWI for the reconstruction of P-wave velocity, density and attenuation: preconditioned truncated Newton approach. In *SEG Technical Program Expanded Abstracts*, pages 1198–1203. SEG.

Meyer, C. D. (2000). *Matrix analysis and applied linear algebra*, volume 71. Siam.

Millane, R. P. (1990). Phase retrieval in crystallography and optics. *JOSA A*, 7(3):394–411.

Mulder, W. and Plessix, R. E. (2008). Exploring some issues in acoustic full waveform inversion. *Geophysical Prospecting*, 56(6):827–841.

Mulder, W. A. and Hak, B. (2009). An ambiguity in attenuation scattering imaging. *Geophysical Journal International*, 178:1614–1624.

Munns, J. W. (1985). The Valhall field: a geological overview. *Marine and Petroleum Geology*, 2:23–43.

Netrapalli, P., Jain, P., and Sanghavi, S. (2013). Phase retrieval using alternating minimization. In *Advances in Neural Information Processing Systems*, pages 2796–2804.

Neves, F. A. and Singh, S. C. (1996). Sensitivity study of seismic reflection/refraction data. *Geophysical Journal International*, 126(2):470–476.

Nihei, K. T. and Li, X. (2007). Frequency response modelling of seismic waves using finite difference time domain with phase sensitive detection (TD-PSD). *Geophysical Journal International*, 169:1069–1078.

Nishihara, R., Lessard, L., Recht, B., Packard, A., and Jordan, M. I. (2015). A general analysis of the convergence of ADMM. *arXiv preprint arXiv:1502.02009*.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, 2nd edition.

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

Operto, S., Brossier, R., Gholami, Y., Métivier, L., Prieux, V., Ribodetti, A., and Virieux, J. (2013). A guided tour of multiparameter full waveform inversion for multicomponent data: from theory to practice. *The Leading Edge*, Special section Full Waveform Inversion(September):1040–1054.

Operto, S. and Miniussi, A. (2018). On the role of density and attenuation in 3D multi-parameter visco-acoustic VTI frequency-domain FWI: an OBC case study from the North Sea. *Geophysical Journal International*, 213:2037–2059.

Operto, S., Miniussi, A., Brossier, R., Combe, L., Métivier, L., Monteiller, V., Ribodetti, A., and Virieux, J. (2015). Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: application to Valhall in the visco-acoustic vertical transverse isotropic approximation. *Geophysical Journal International*, 202(2):1362–1391.

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

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

Operto, S., Virieux, J., Ribodetti, A., and Anderson, J. E. (2009). Finite-difference frequency-domain modeling of visco-acoustic wave propagation in two-dimensional TTI media. *Geophysics*, 74 (5):T75–T95.

Oppenheim, A. V. and Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541.

Parikh, N. and Boyd, S. (2013). Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):123–231.

Peaceman, D. W. and Rachford, Jr, H. H. (1955). The numerical solution of parabolic and elliptic differential equations. *Journal of the Society for industrial and Applied Mathematics*, 3(1):28–41.

Peters, B. and Herrmann, F. J. (2017). Constraints versus penalties for edge-preserving full-waveform inversion. *The Leading Edge*, 36(1):94–100.

Pfeiffer, F., Weitkamp, T., Bunk, O., and David, C. (2006). Phase retrieval and differential phase-contrast imaging with low-brillance X-ray sources. *Nature Physics*, 2:258–261.

Plessix, R. E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503.

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

Plessix, R.-E., Baeten, G., de Maag, J. W., and ten Kroode, F. (2012). Full waveform inversion and distance separated simultaneous sweeping: a study with a land seismic data set. *Geophysical Prospecting*, 60:733 – 747.

Plessix, R. E. and Cao, Q. (2011). A parametrization study for surface seismic full waveform inversion in an acoustic vertical transversely isotropic medium. *Geophysical Journal International*, 185:539–556.

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

Polson, N. G. and Sokolov, V. (2019). Bayesian regularization: From tikhonov to horseshoe. *Wiley Interdisciplinary Reviews: Computational Statistics*, page e1463.

Pratt, R. G. (2008). Waveform tomography - successes, cautionary tales, and future directions. In *Presented at the 70th Annual EAGE Conference & Exhibition, Roma*, pages WO11 – Full–Waveform Inversion: current status and perspectives.

Pratt, R. G., Shin, C., and Hicks, G. J. (1998). Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International*, 133:341–362.

Pratt, R. G. and Smithyman, B. (2018). Full waveform inversion for density: Reciprocity, divergence, parameter resolution, and other issues. In *Expanded Abstracts, 80th Annual EAGE Meeting (Copenhagen)*.

Pratt, R. G., Song, Z. M., Williamson, P. R., and Warner, M. (1996). Two-dimensional velocity models from wide-angle seismic data by wavefield inversion. *Geophysical Journal International*, 124:323–340.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 3 edition.

Prieux, V., Brossier, R., Gholami, Y., Operto, S., Virieux, J., Barkved, O., and Kommedal, J. (2011). On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study. *Geophysical Journal International*, 187:1495–1515.

Prieux, V., Brossier, R., Operto, S., and Virieux, J. (2013). Multiparameter full waveform inversion of multicomponent OBC data from Valhall. Part 1: imaging compressional wavespeed, density and attenuation. *Geophysical Journal International*, 194(3):1640–1664.

Qian, C., Fu, X., Sidiropoulos, N. D., Huang, L., and Xie, J. (2017). Inexact alternating optimization for phase retrieval in the presence of outliers. *IEEE Transactions on Signal Processing*, 65(22):6069–6082.

Ribodetti, A., Operto, S., Agudelo, W., Collot, J.-Y., and Virieux, J. (2011). Joint ray+born least-squares migration and simulated annealing optimization for high-resolution target-oriented quantitative seismic imaging. *Geophysics*, 76(2):R23.

Ribodetti, A., Operto, S., Virieux, J., Lambaré, G., Valéro, H.-P., and Gibert, D. (2000). Asymptotic viscoacoustic diffraction tomography of ultrasonic laboratory data : a tool for rock properties analysis. *Geophysical Journal International*, 140:324–340.

Rudin, L., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268.

Sandhu, G. Y., Li, C., Roy, O., Schmidt, S., and Duric, N. (2015). Frequency domain ultrasound waveform tomography: breast imaging using a ring transducer. *Physics in Medecine and Biology*, 60:5381–5398.

Setzer, S., Steidl, G., and Teuber, T. (2011). Infimal convolution regularizations with discrete ℓ1-type functionals. *Communications in Mathematical Sciences*, 9(3):797–827.

Shah, N. K. (2014). *Seismic Full Waveform Inversion for wrapped and unwrapped phase*. PhD thesis, Imperial College London.

Shaw, P. R. (1986). Extending the linear regime of the seismic waveform inversion problem. *Geophysical Research Letters*, 13(11):1157–1160.

Shaw, P. R. (1988). Waveform inversion of time-corrected refraction data. *Geophysical Journal*, 93:75–90.

Shechtman, Y., Eldar, Y. C., Cohen, O., Chapman, H. N., Miao, J., and Segev, M. (2015). Phase retrieval with application to optical imaging: a contemporary overview. *IEEE signal processing magazine*, 32(3):87–109.

Shen, L., Ahmed, I., Brenders, A., Dellinger, J., Etgen, J., and Michell, S. (2018). Full-waveform inversion: the next leap forward in subsalt imaging. *The Leading Edge*, January:76b1–67b6.

Shen, X., Diamond, S., Udell, M., Gu, Y., and Boyd, S. (2017). Disciplined multi-convex programming. In *29th Chinese Control and Decision Conference (CCDC), IEEE control systems society, https://doi.org/10.1016/j.jcp.2016.09.036*, volume 327, pages 895–900.

Sheriff, R. E. and Geldart, L. P. (1995). *Exploration seismology, second edition*. Cambridge University Press.

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

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

Sirgue, L., Barkved, O. I., Dellinger, J., Etgen, J., Albertin, U., and Kommedal, J. H. (2010). Full waveform inversion: the next leap forward in imaging at Valhall. *First Break*, 28:65–70.

Sirgue, L., Etgen, J. T., and Albertin, U. (2008). 3D Frequency Domain Waveform Inversion using Time Domain Finite Difference Methods. In *Proceedings 70th EAGE, Conference and Exhibition, Roma, Italy*, page F022.

Sirgue, L. and Pratt, R. G. (2004). Efficient waveform inversion and imaging : a strategy for selecting temporal frequencies. *Geophysics*, 69(1):231–248.

Stopin, A., Plessix, R.-E., and Al Abri, S. (2014). Multiparameter waveform inversion of a large wide-azimuth low-frequency land data set in Oman. *Geophysics*, 79(3):WA69–WA77.

Stopin, A., Plessix, R.-E., Kuehl, H., Goh, V., and Overgaag, K. (2016). Application of visco-acoustic full waveform inversion for gas cloud imaging and velocity model building. In *Expanded Abstracts*. EAGE.

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

Takougang, E. M. T. and Calvert, A. J. (2012). Seismic velocity and attenuation structures of the Queen Charlotte Basin from full-waveform tomography of seismic reflection data. *Geophysics*, 77(3):B107–B124.

Tang, Y. (2009). Target-oriented wave-equation least-squares migration/inversion with phase-encoded hessian. *Geophysics*, 74(6):WCA95–WCA107.

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266.

Thierry, P., Operto, S., and Lambaré, G. (1999). Fast 2D ray-Born inversion/migration in complex media. *Geophysics*, 64(1):162–181.

Thomsen, L. A. (1986). Weak elastic anisotropy. *Geophysics*, 51:1954–1966.

Tikhonov, A., Goncharsky, A., Stepanov, V., and Yagola, A. (2013). *Numerical methods for the solution of ill-posed problems*. Springer Science & Business Media.

Toksöz, M. N. and Johnston, D. H. (1981). *Geophysics reprint series, No. 2: Seismic wave attenuation*. Society of exploration geophysicists, Tulsa, OK.

Toverud, T. and Ursin, B. (2005). Comparison of seismic attenuation models using zero-offset vertical seismic profiling (vsp) data. *Geophysics*, 70:F17–F25.

van Leeuwen, T. and Herrmann, F. (2016). A penalty method for PDE-constrained optimization in inverse problems. *Inverse Problems*, 32(1):1–26.

van Leeuwen, T. and Herrmann, F. J. (2013). Mitigating local minima in full-waveform inversion by expanding the search space. *Geophysical Journal International*, 195(1):661–667.

van Leeuwen, T. and Mulder, W. A. (2010). A correlation-based misfit criterion for wave-equation traveltime tomography. *Geophysical Journal International*, 182(3):1383–1394.

Vermeer, G. J. O. (2012). *3D Seismic Survey Design*. Society of Exploration Geophysics.

Vigh, D., Jiao, K., Watts, D., and Sun, D. (2014). Elastic full-waveform inversion application using multicomponent measurements of seismic data collection. *Geophysics*, 79(2):R63–R77.

Virieux, J., Asnaashari, A., Brossier, R., Métivier, L., Ribodetti, A., and Zhou, W. (2017). An introduction to Full Waveform Inversion. In Grechka, V. and Wapenaar, K., editors, *Encyclopedia of Exploration Geophysics*, pages R1–1–R1–40. Society of Exploration Geophysics.

Virieux, J. and Operto, S. (2009). An overview of full waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26.

Waldspurger, I., d'Aspremont, A., and Mallat, S. (2015). Phase recovery, maxcut and complex semidefinite programming. *Mathematical Programming*, 149(1-2):47–81.

Walther, A. (1963). The question of phase retrieval in optics. *Optica Acta: International Journal of Optics*, 10(1):41–49.

Wang, C., Yingst, D., Farmer, P., Jones, I., Martin, G., and Leveille, J. (2017a). *Reconstructed full-waveform inversion with the extended source*, pages 1449–1453.

Wang, C., Yingst, D., Farmer, P., Jones, I., Martin, G., and Leveille, J. (2017b). Reflection-based full waveform inversion (rfwi) in the frequency domain. In *SEG Technical Program Expanded Abstracts*, pages 4656–4660. SEG.

Wang, C., Yingst, D., Farmer, P., and Leveille, J. (2016). *Full-waveform inversion with the reconstructed wavefield method*, pages 1237–1241.

Warner, M. and Guasch, L. (2016). Adaptive waveform inversion: Theory. *Geophysics*, 81(6):R429–R445.

Wood, W. T., Holbrook, W. S., Sen, M. K., and Stoffa, P. L. (2008). Full waveform inversion of reflection seismic data for ocean temperature profiles. *Geophysical Research Letters*, 35:L04608.

Wu, C. J. et al. (1983). On the convergence properties of the em algorithm. *The Annals of statistics*, 11(1):95–103.

Xu, S., Wang, D., Chen, F., Lambaré, G., and Zhang, Y. (2012a). Inversion on reflected seismic wave. *SEG Technical Program Expanded Abstracts 2012*, pages 1–7.

Xu, Y., Yin, W., Wen, Z., and Zhang, Y. (2012b). An alternating direction algorithm for matrix completion with nonnegative factors. *Frontiers of Mathematics in China*, 7(2):365–384.

Yang, J., Liu, Y., and Dong, L. (2016a). Simultaneous estimation of velocity and density in acoustic multiparameter full-waveform inversion using an improved scattering-integral approach. *Geophysics*, 81(6):R399–R415.

Yang, J.-Z., Liu, Y.-Z., and Dong, L.-G. (2014). A multi-parameter full waveform strategy for acoustic media with variable density. *Chinese Journal of Geophysics*, 57(2):628–643.

Yang, P., Brossier, R., Métivier, L., and Virieux, J. (2016b). A review on the systematic formulation of 3D multiparameter full waveform inversion in viscoelastic medium. *Geophysical Journal International*, 207(1):129–149.

Yang, Y., Engquist, B., Sun, J., and Hamfeldt, B. F. (2018). Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion. *Geophysics*, 83(1):R43–R62.

Yin, W. and Osher, S. (2013). Error forgetting of Bregman iteration. *Journal of Scientific Computing*, 54:684–695.

Yin, W., Osher, S., Goldfarb, D., and Darbon, J. (2008). Bregman iterative algorithms for l1-minimization with applications to compressed sensing. *SIAM Journal Imaging Sciences*, 1(1):143–168.

Yong, P., Liao, W., Huang, J., and Li, Z. (2018). Total variation regularization for seismic waveform inversion using an adaptive primal dual hybrid gradient method. *Inverse Problems*, 34(4):045006.

Zhou, W., Brossier, R., Operto, S., and Virieux, J. (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.

Zhou, W., Brossier, R., Operto, S., Virieux, J., and Yang, P. (2018). Velocity model building by waveform inversion of early arrivals and reflections: a 2d ocean-bottom-cable study with gas cloud effects. *Geophysics*, 83(2):R141–R157.

Zhou H., Zhang G., B. R. (2006). An anisotropic acoustic wave equation for VTI media. In *Expanded Abstracts, 68^{th} Annual meeting*. EAGE.

Zhu, H. and Fomel, S. (2016). Building good starting models for full-waveform inversion using adaptive matching filtering misfit. *Geophysics*, 81(5):U61–U72.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320.