HAL
open science

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

Hossein Aghamiry, A. Gholami, Stéphane Operto

To cite this version:

Hossein Aghamiry, A. Gholami, Stéphane Operto. ADMM-based multiparameter wavefield reconstruction inversion in VTI acoustic media with TV regularization. Geophysical Journal International, 2019, 219 (2), pp.1316-1333. 10.1093/gji/ggz369 . hal-02326376

HAL Id: hal-02326376

https://hal.science/hal-02326376v1

Submitted on 9 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Geophys. J. Int. (XXX) **XXX**, 1–**??**

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

H. S. Aghamiry.[1,2], A. Gholami[1], S. Operto[2]
[1] *University of Tehran, Institute of Geophysics, Tehran, Iran.*
E-mail: h.aghamiry@ut.ac.ir, agholami@ut.ac.ir
[2]*Geoazur, Université Côte d'Azur, CNRS, IRD, OCA, Valbonne, France.*
E-mail: aghamiry@geoazur.unice.fr, operto@geoazur.unice.fr

Received XXX; in original form XXX

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).

Key words: Waveform inversion, Inverse theory, Seismic anisotropy, Controlled source seismology.

1 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 & 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 & Herrmann 2013) rather than by variable projection (van Leeuwen & 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 & 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. 2018b, 2019c). Later, Aghamiry et al. (2018c, 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, 2019a) 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 subproblems: a least-squares quadratic subproblem and a proximity subproblem (Goldstein & 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 band-pass 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.

2 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.

2.1 Bilinearity of the wave equation: preliminaries

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

$$\mathbf{A}(\mathbf{m})\mathbf{u} = \mathbf{s}, \tag{2.1}$$

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

$$\mathbf{L}(\mathbf{u})\mathbf{m} = \mathbf{y}(\mathbf{u}), \tag{2.2}$$

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:

$$\mathbf{A}(\mathbf{m})\mathbf{u} = \mathbf{B}\,\mathbf{M}(\mathbf{m})\,\mathbf{C}\mathbf{u} + \mathbf{D}\mathbf{u}, \tag{2.3}$$

where $\mathbf{M}(\mathbf{m})$ is a block matrix, whose blocks of dimension $n \times n$ are either 0 or of the form $\mathrm{diag}(\mathbf{m}_i)$, and matrices $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ don't depend on $\mathbf{m}$. The operator $\mathrm{diag}(\bullet)$ denotes a diagonal matrix of coefficients $\bullet$ and $\mathbf{m}_i$ is the subsurface parameter vector of class i. By noting that $\mathrm{diag}(\mathbf{x})\mathbf{z} = \mathrm{diag}(\mathbf{z})\mathbf{x}$, 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 2.3 and the above permutation between $\mathbf{u}$ and $\mathbf{m}$ that the wave equation can be re-written as

$$\mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{s} = \mathbf{B}\,\mathbf{C}'(\mathbf{u})\mathbf{m} + \mathbf{D}\mathbf{u} - \mathbf{s} = \mathbf{L}(\mathbf{u})\mathbf{m} - \mathbf{y}(\mathbf{u}). \tag{2.4}$$

Moreover, in the framework of multi-parameter analysis, it is worth noting that

$$\frac{\partial \mathbf{A}(\mathbf{m})}{\partial m_k}\mathbf{u} = \mathbf{L}(\mathbf{u})\mathbf{e}_k, \tag{2.5}$$

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.

2.2 Bilinearity of the acoustic VTI wave equation

2.3 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 & Bakker 2011; Operto et al. 2014)

$$\begin{aligned}
-\hat{i}\omega\mathbf{v}_{x,l} &= \text{diag}(\mathbf{b})\nabla_x\mathbf{u}_{x,l},\\
-\hat{i}\omega\mathbf{v}_{z,l} &= \text{diag}(\mathbf{b})\nabla_z\mathbf{u}_{z,l},\\
-\hat{i}\omega\mathbf{u}_{x,l} &= \text{diag}(\mathbf{c}_{11})\nabla_x\mathbf{v}_{x,l} + \text{diag}(\mathbf{c}_{13})\nabla_z\mathbf{v}_{z,l} - \hat{i}\omega\mathbf{s}_l,\\
-\hat{i}\omega\mathbf{u}_{z,l} &= \text{diag}(\mathbf{c}_{13})\nabla_x\mathbf{v}_{x,l} + \text{diag}(\mathbf{c}_{33})\nabla_z\mathbf{v}_{z,l} - \hat{i}\omega\mathbf{s}_l,
\end{aligned} \tag{2.6}$$

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 & 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 2.6 for all sources leads to the following matrix equation:

$$\left(\begin{bmatrix}\mathbf{M}_b & \mathbf{0}\\ \mathbf{0} & \mathbf{M}_c\end{bmatrix}\begin{bmatrix}\mathbf{0} & \nabla\\ \nabla & \mathbf{0}\end{bmatrix} + \hat{i}\omega\mathbf{I}\right)\begin{bmatrix}\mathbf{V}\\ \mathbf{U}\end{bmatrix} = \hat{i}\omega\begin{bmatrix}\mathbf{0}\\ \mathbf{S}\end{bmatrix}, \tag{2.7}$$

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

$$\nabla = \begin{bmatrix}\nabla_x & \mathbf{0}\\ \mathbf{0} & \nabla_z\end{bmatrix},$$

$$\mathbf{U} = \begin{bmatrix}\mathbf{U}_h\\ \mathbf{U}_v\end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix}\mathbf{V}_x\\ \mathbf{V}_z\end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix}\mathbf{S}_x\\ \mathbf{S}_z\end{bmatrix}, \tag{2.8}$$

with

$$\mathbf{V}_x = \begin{bmatrix}\mathbf{v}_{x,1} & \mathbf{v}_{x,2} & \dots & \mathbf{v}_{x,n_s}\end{bmatrix} \in \mathbb{C}^{n\times n_s},$$

and analogously for $\mathbf{V}_z$, $\mathbf{U}_h$, $\mathbf{U}_v$, and

$$\mathbf{S}_x = \mathbf{S}_z = \begin{bmatrix}\mathbf{s}_1 & \mathbf{s}_2 & \dots & \mathbf{s}_{n_s}\end{bmatrix} \in \mathbb{C}^{n\times n_s}.$$

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 2.7 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

$$\begin{bmatrix}\mathbf{L}_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{L}_{21} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{L}_{32} & \mathbf{L}_{33} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{L}_{43} & \mathbf{L}_{44}\end{bmatrix}\begin{bmatrix}\mathbf{b}\\ \mathbf{c}_{11}\\ \mathbf{c}_{13}\\ \mathbf{c}_{33}\end{bmatrix} = \hat{i}\omega\begin{bmatrix}\mathbf{0} - \mathbf{v}_{x,l}\\ \mathbf{0} - \mathbf{v}_{z,l}\\ \mathbf{s}_l - \mathbf{u}_{x,l}\\ \mathbf{s}_l - \mathbf{u}_{z,l}\end{bmatrix}, \tag{2.9}$$

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 2.7 and 2.9 are equivalent forms of the original equation 2.6. 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.

2.4 Second-order wave equation

Following a parsimonious approach (e.g., Operto et al. 2007), we eliminate $\mathbf{v}_{x,l}$ and $\mathbf{v}_{z,l}$ from equation 2.6 to derive a system of two second-order partial differential equations as

$$\begin{aligned}
-\omega^2\mathbf{u}_{x,l} &= \text{diag}(\mathbf{c}_{11})\nabla_x\text{diag}(\mathbf{b})\nabla_x\mathbf{u}_{x,l}\\
&\quad + \text{diag}(\mathbf{c}_{13})\nabla_z\text{diag}(\mathbf{b})\nabla_z\mathbf{u}_{z,l} - \omega^2\mathbf{s}_l,\\
-\omega^2\mathbf{u}_{z,l} &= \text{diag}(\mathbf{c}_{13})\nabla_x\text{diag}(\mathbf{b})\nabla_x\mathbf{u}_{x,l}\\
&\quad + \text{diag}(\mathbf{c}_{33})\nabla_z\text{diag}(\mathbf{b})\nabla_z\mathbf{u}_{z,l} - \omega^2\mathbf{s}_l.
\end{aligned} \tag{2.10}$$

Equation 2.10 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 and Vigh et al. 2014; Yang et al. 2016). Alternatively,

the parameter estimation can be performed with the bilinear first-order wave equation, equation 2.9, while wavefield reconstruction is performed with the second-order wave equation, equation 2.10, or the fourth-order wave equation reviewed in Appendix C 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 2.10 as

$$\begin{aligned} -\omega^2 \text{diag}\left(1/\mathbf{v}_0^2\right)\mathbf{u}_{x,l} &= \text{diag}(1+2\epsilon)\nabla_{xx}\mathbf{u}_{x,l} \\ &\quad + \text{diag}(\sqrt{1+2\delta})\nabla_{zz}\mathbf{u}_{z,l} - \text{diag}(1/\mathbf{v}_0^2)\mathbf{s}_l, \\ -\omega^2 \text{diag}\left(1/\mathbf{v}_0^2\right)\mathbf{u}_{z,l} &= \text{diag}(\sqrt{1+2\delta})\nabla_{xx}\mathbf{u}_{x,l} \\ &\quad + \nabla_{zz}\mathbf{u}_{z,l} - \text{diag}\left(1/\mathbf{v}_0^2\right)\mathbf{s}_l, \end{aligned} \tag{2.11}$$

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

$$\mathbf{A}(\mathbf{m})\mathbf{U} = \mathbf{S}, \tag{2.12}$$

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 2.8, 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 2.12 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.

$$\begin{bmatrix}\mathbf{L}_1 \\ \vdots \\ \mathbf{L}_l \\ \vdots \\ \mathbf{L}_{n_s}\end{bmatrix}\mathbf{m} = \begin{bmatrix}\mathbf{y}_1 \\ \vdots \\ \mathbf{y}_l \\ \vdots \\ \mathbf{y}_{n_s}\end{bmatrix}, \tag{2.13}$$

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 2.13, will change. Table 1 presents this system for the different possible configurations.

2.5 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 2.12:

$$\begin{aligned} &\min_{\mathbf{U},\mathbf{m}\in\mathcal{C}} \quad \sum \sqrt{|\partial_x\mathbf{m}|^2 + |\partial_z\mathbf{m}|^2}, \\ &\text{subject to} \quad \begin{cases}\mathbf{PU} = \mathbf{D}, \\ \mathbf{A}(\mathbf{m})\mathbf{U} = \mathbf{S},\end{cases} \end{aligned} \tag{2.14}$$

where ∂_x and ∂_z are, respectively, first-order finite-difference operators in the horizontal and vertical directions with appropriate boundary conditions, $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^{3n\times 1} \mid \mathbf{m}_l \le \mathbf{x} \le \mathbf{m}_u\}$ 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

$$\begin{aligned} &\min_{\mathbf{U},\mathbf{m}\in\mathcal{C}} \max_{\bar{\mathbf{D}},\bar{\mathbf{S}}} \quad \sum \sqrt{|\partial_x\mathbf{m}|^2 + |\partial_z\mathbf{m}|^2} \\ &+ \left\langle \bar{\mathbf{D}}, \mathbf{PU} - \mathbf{D}\right\rangle + \lambda_0\|\mathbf{PU} - \mathbf{D}\|_F^2 \\ &+ \left\langle \bar{\mathbf{S}}, \mathbf{A}(\mathbf{m})\mathbf{U} - \mathbf{S}\right\rangle + \lambda_1\|\mathbf{A}(\mathbf{m})\mathbf{U} - \mathbf{S}\|_F^2, \end{aligned} \tag{2.15}$$

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 2.15 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)

$$\begin{aligned} &\min_{\mathbf{U},\mathbf{m}\in\mathcal{C}} \max_{\bar{\mathbf{D}},\bar{\mathbf{S}}} \quad \sum \sqrt{|\partial_x\mathbf{m}|^2 + |\partial_z\mathbf{m}|^2} \\ &+ \lambda_0\|\mathbf{PU} - \mathbf{D} - \tilde{\mathbf{D}}\|_F^2 - \|\tilde{\mathbf{D}}\|_F^2 \\ &+ \lambda_1\|\mathbf{A}(\mathbf{m})\mathbf{U} - \mathbf{S} - \tilde{\mathbf{S}}\|_F^2 - \|\tilde{\mathbf{S}}\|_F^2, \end{aligned} \tag{2.16}$$

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 2.16, 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 2.16 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 2.16 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

Table 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.	$\underbrace{\begin{bmatrix}\omega^2\text{diag}(\mathbf{u}_{x,l}) & \text{diag}(\nabla_{zz}\mathbf{u}_{z,l})\\ \omega^2\text{diag}(\mathbf{u}_{z,l}) & \text{diag}(\nabla_{xx}\mathbf{u}_{x,l})\end{bmatrix}}_{\mathbf{L}_l\in\mathbb{C}^{2n\times 2n}}\underbrace{\begin{bmatrix}\mathbf{m}_{v_0}\\ \mathbf{m}_\delta\end{bmatrix}}_{\mathbf{m}\in\mathbb{R}^{2n\times 1}}=\underbrace{\begin{bmatrix}\omega^2\mathbf{s}_l-\text{diag}(\mathbf{m}_\epsilon)\nabla_{xx}\mathbf{u}_{x,l}\\ \omega^2\mathbf{s}_l-\nabla_{zz}\mathbf{u}_{z,l}\end{bmatrix}}_{\mathbf{y}_l\in\mathbb{C}^{2n\times 1}}$

bi-variate primal problem, equation 2.14, 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 for the complete development):

2.5.1 *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:

$$\begin{aligned}&\big[\lambda_0\mathbf{P}^T\mathbf{P}+\lambda_1\mathbf{A}(\mathbf{m}^k)^T\mathbf{A}(\mathbf{m}^k)\big]\mathbf{U}=\\&\big[\lambda_0\mathbf{P}^T[\mathbf{D}+\tilde{\mathbf{D}}^k]+\lambda_1\mathbf{A}(\mathbf{m}^k)^T[\mathbf{S}+\tilde{\mathbf{S}}^k]\big].\end{aligned}\tag{2.17}$$

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 2.17 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 2.17 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 2.11. The elimination procedure allowing to transform the system of two second-order wave equations for $\mathbf{u}_x$ and $\mathbf{u}_z$, equation 2.11, 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.

2.5.2 *Step 2: The primal model estimation.*

The reconstructed wavefields, equation 2.17, are injected in the linearized equation 2.13 to update the subsurface parameters by solving the following system of linear equations:

$$\left[\lambda_1 \sum_{l=1}^{n_s} \mathbf{L}_l^T \mathbf{L}_l + \overline{\nabla}^T \Gamma \overline{\nabla} + Z\mathbf{I}\right] \mathbf{m} = \lambda_1 \sum_{l=1}^{n_s} \mathbf{L}_l^T (\mathbf{y}_l^k + \tilde{\mathbf{s}}_l^k) + \overline{\nabla}^T \Gamma (\mathbf{p}^k + \tilde{\mathbf{p}}^k) + Z(\mathbf{q}^k + \tilde{\mathbf{q}}^k), \tag{2.18}$$

where

$$\overline{\nabla} = \begin{bmatrix} \partial_x \\ \partial_z \end{bmatrix}. \tag{2.19}$$

In equation 2.18, $\mathbf{p}$ 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). 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

$$Z = \begin{bmatrix} \zeta_{v_0}\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \zeta_\epsilon \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \zeta_\delta \mathbf{I} \end{bmatrix} \in \mathbb{R}_+^{3n\times 3n}, \tag{2.20}$$

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 2.18.

In the same way, $\mathbf{\Gamma}$ is a diagonal matrix defined as

$$\mathbf{\Gamma} = \begin{bmatrix} \mathbf{\Gamma}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}_{22} \end{bmatrix}, \tag{2.21}$$

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 2.23. 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.

2.5.3 *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

$$\mathbf{p}^{k+1} \leftarrow \text{prox}_{\Gamma^{-1}}(\mathbf{z}) = \begin{bmatrix} \xi \circ \mathbf{z}_x \\ \xi \circ \mathbf{z}_z \end{bmatrix}, \tag{2.22}$$

where

$$\xi = \max(1 - \frac{1}{\mathbf{\Gamma}\sqrt{\mathbf{z}_x^2 + \mathbf{z}_z^2}}, 0) \tag{2.23}$$

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. $\mathbf{x}^2 = \mathbf{x} \circ \mathbf{x}$.

2.5.4 *Step 4: The bounding constraint primal update.*

Update the primal variable $\mathbf{q}$ via a projection operator, which has the following component-wise form

$$\mathbf{q}^{k+1} \leftarrow \text{proj}_{\mathcal{C}}(\mathbf{m}^{k+1} - \tilde{\mathbf{q}}^k), \tag{2.24}$$

where the projection operator is given by

$$\text{proj}_{\mathcal{C}}(\mathbf{x}) = \min(\max(\mathbf{x}, \mathbf{m}_l), \mathbf{m}_u).$$

2.5.5 *Step 5: Dual updates.*

Update the scaled dual variables with gradient ascent steps

$$\begin{cases} \tilde{\mathbf{S}}^{k+1} & \leftarrow \tilde{\mathbf{S}}^k + \mathbf{S} - \mathbf{A}(\mathbf{m}^{k+1})\mathbf{U}^{k+1}, \\ \tilde{\mathbf{D}}^{k+1} & \leftarrow \tilde{\mathbf{D}}^k + \mathbf{D} - \mathbf{P}\mathbf{U}^{k+1}, \\ \tilde{\mathbf{p}}^{k+1} & \leftarrow \tilde{\mathbf{p}}^k + \mathbf{p}^{k+1} - \overline{\nabla}\mathbf{m}^{k+1}, \\ \tilde{\mathbf{q}}^{k+1} & \leftarrow \tilde{\mathbf{q}}^k + \mathbf{q}^{k+1} - \mathbf{m}^{k+1}, \end{cases} \tag{2.25}$$

2.5.6 *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

2.6 Hyperparameter tuning

We tune the different penalty parameters by extending the procedure reviewed by Aghamiry et al. (2019b,a) 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 2.22. 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 2.18. 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 2.17, according to the criterion proposed by van Leeuwen & 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 $\|\mathbf{P}\mathbf{U} - \mathbf{D}\|_2^2$ 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 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.

3 NUMERICAL EXAMPLES

3.1 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 2.18) 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. 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. 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

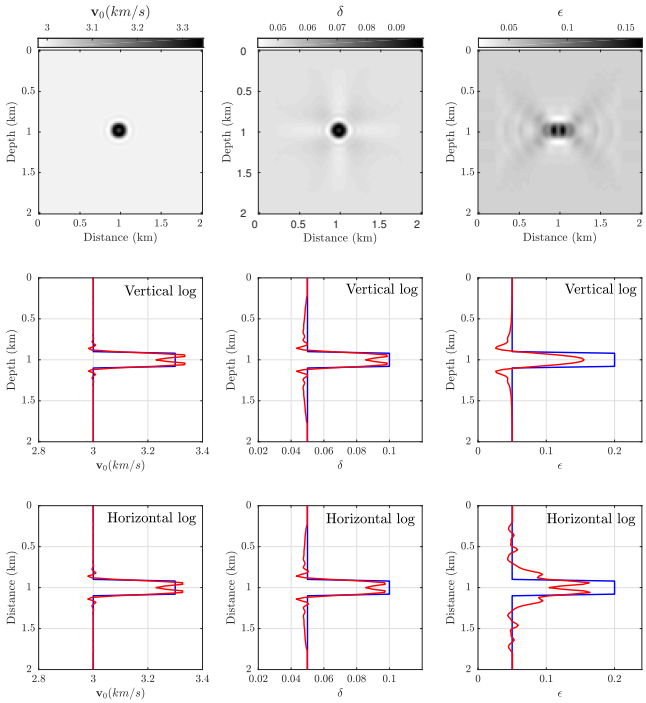

Figure 1. Inclusion test: Mono-parameter IR-WRI results. For $\mathbf{v}_0$ (first column), δ (second column), and ϵ (third column). The initial models are the true homogeneous background models. The vertical and horizontal profiles in the true model (blue) and estimated model (red) are extracted across the center of the inclusions.

(Fig. 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. 1 and 2).

Then, we complement DMP regularization with TV regularization during bound-constrained IR-WRI for the above mono-parameter and multi-parameter experiments (Figs. 3 and 4). The results show how TV regularization contributes to remove the wavenumber filtering performed by radiation patterns (compare for example Figs. 1 and 3) as well as the cross-talk artifacts during the multi-parameter inversion (compare Figs. 2 and 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.

3.2 Synthetic North Sea case study

3.2.1 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. 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

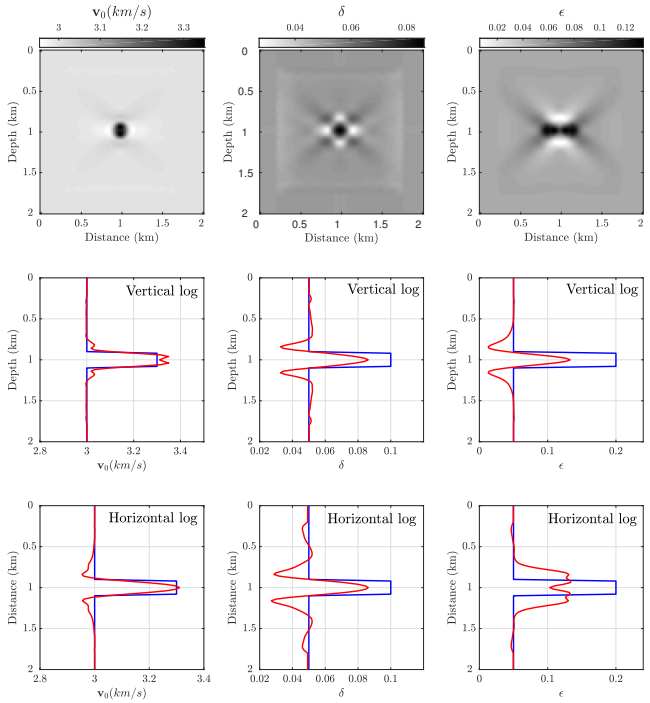

Figure 2. Inclusion test: Multiparameter IR-WRI results for joint update of $\mathbf{v}$ (first column), δ (second column), and ϵ (third column). The initial models are the homogeneous background models.

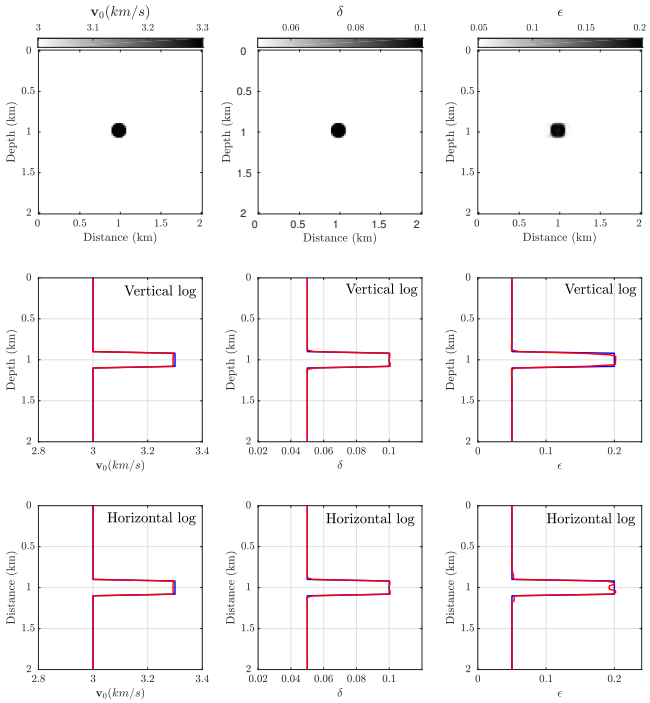

Figure 3. Inclusion test: Same as Fig. 1, but with TV regularization.

is indicated by a sharp positive velocity contrast at around 2.5 km depth, and a flat reflector at 5 km depth (Fig. 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. 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. 5a,c,e) with the forward modelling engine described in Appendix C. 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. 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. 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 & 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 B.4a, in order to force the updates of ϵ to be smooth and close to ϵ^0.

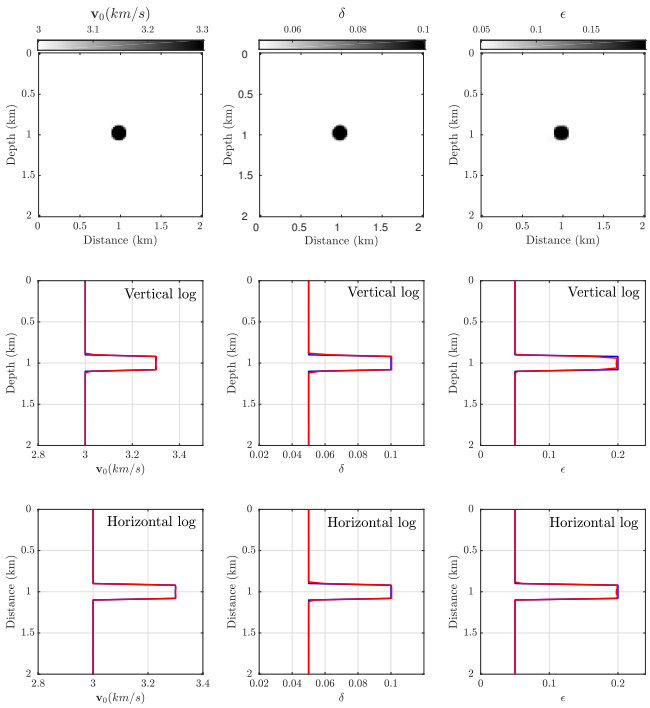

Figure 4. Inclusion test: Same as Fig. 2, but with TV regularization.

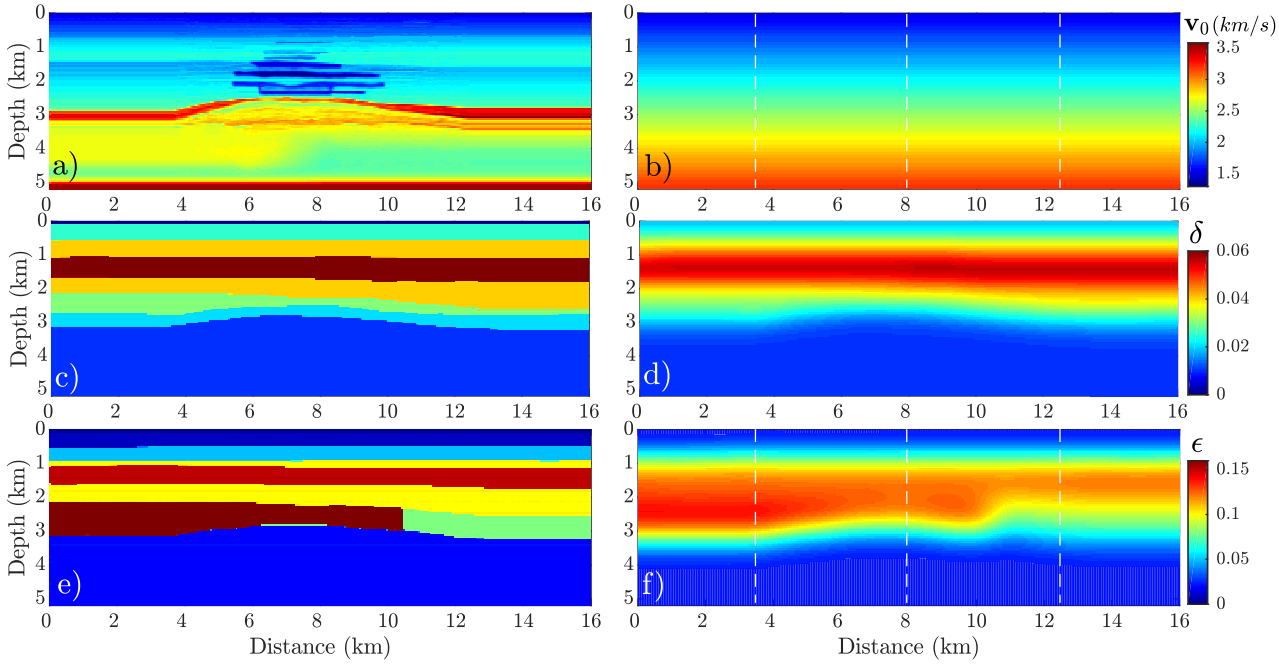

Figure 5. North Sea case study. (a) True $\mathbf{v}_0$ model. (b) Initial $\mathbf{v}_0$ model. (c) True δ model. (d) Smoothed δ model, which is used as passive background model during inversion tests. (e) True ϵ model. (f) Smoothed ϵ model, which is used as a passive background model during the mono-parameter 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.

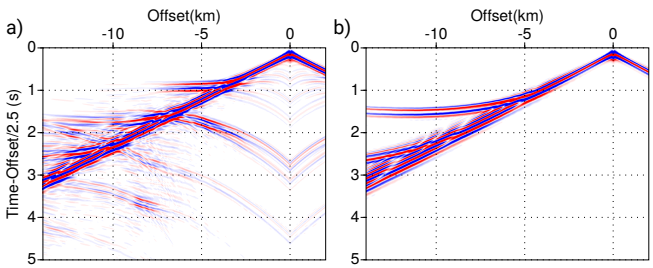

Figure 6. North Sea case study. Time domain seismograms computed in (a) the true model (Fig. 5a,c,e) and in (b) the initial model (Fig. 5b,d,f). The seismograms are plotted with a reduction velocity of 2.5 km/s.

The background models used for the passive anisotropic parameters (either ϵ and δ or δ alone) are the smooth models (Fig. 5d,f), which means that the IR-WRI results will be impacted upon by the smoothness of the passive 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

$$\|\mathbf{A}(\mathbf{m}^{k+1})\mathbf{U}^{k+1} - \mathbf{S}\|_F \leq \varepsilon_b \;\&\; \|\mathbf{P}\mathbf{U}^{k+1} - \mathbf{D}\|_F \leq \varepsilon_d, \quad (3.1)$$

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 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.

3.2.2 *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 a series of $\mathbf{v}_0$ and ϵ models that are generated according to

$$\mathbf{v}_0(\alpha) = \mathbf{v}_{0_{true}} + |\alpha|[\mathbf{v}_{0_{init}} - \mathbf{v}_{0_{true}}], \quad (3.2a)$$
$$\epsilon(\beta) = \epsilon_{true} + |\beta|[\epsilon_{init} - \epsilon_{true}], \quad (3.2b)$$

where $\mathbf{v}_{0_{true}}$ and $\mathbf{v}_{0_{init}}$ denote the true and the initial $\mathbf{v}_0$ models, respectively (Fig. 5a,b) and $-1 \leq \alpha \leq 1$. Similarly, ϵ_{true} and ϵ_{init} are the true and the initial ϵ models, respectively (Fig. 5e,f) and $-1 \leq \beta \leq 1$. Finally, we use the true δ model (Fig. 5c) to generate the recorded data and the smoothed version (Fig. 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. 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.

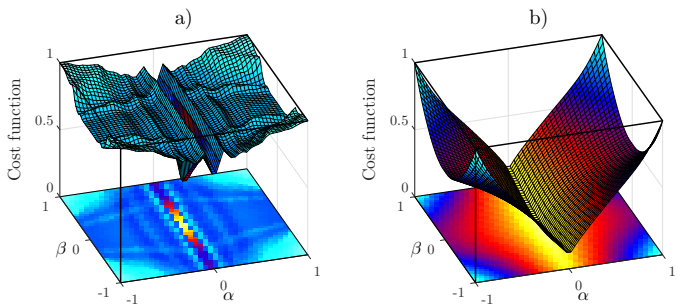

Figure 7. North Sea case study. 3-Hz misfit function for the first iteration when $\mathbf{v}_0$ and ϵ are active parameters and δ is kept fixed to the smooth background model (Fig. 5f). (a) Classical reduced approach, (b) WRI. The variables α and β parametrize the $\mathbf{v}_0$ and ϵ models, respectively, for which the misfit function is computed (see text for details).

3.2.3 *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. 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. 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. 9).

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. 10). The direct comparison between the true model, the initial model and the IR-WRI models of Fig. 10 at distances $x = 3.5\ km$, $x = 8.0\ km$ and $x = 12.5\ km$ are shown in Fig. 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. 11). The TV regularization significantly reduces these amplitude and mis-positioning errors (Figs. 10b and 11) and hence produces $\mathbf{v}_0$ models which are much more consistent with those obtained for noiseless data (Figs. 8b and 9).

To assess how the differences between the velocity models shown in Fig. 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. 12 and 13). The time-domain seismograms and the residuals shown in Fig. 12 give an overall vision of the achieved data fit, while the direct comparison between the recorded and modelled seismograms shown in Fig. 13 allow for a more detailed assessment of the waveform match for a specific arrival.

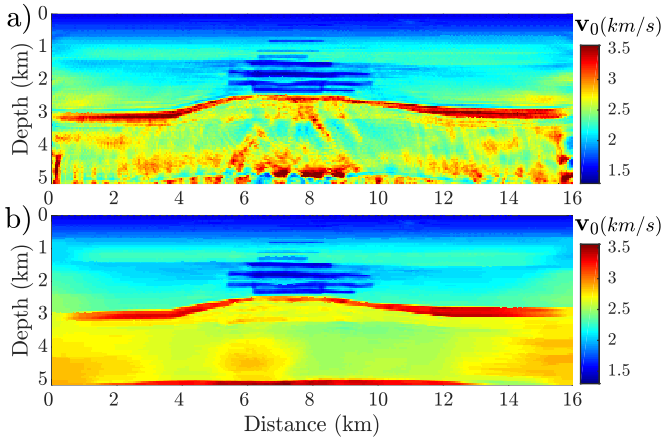

Figure 8. North Sea case study with noiseless data. Mono-parameter IR-WRI. Vertical wavespeed $\mathbf{v}_0$ models inferred from IR-WRI with (a) DMP regularization, (b) TV regularization.

A first conclusion is that, for all of the models shown in Fig. 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 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

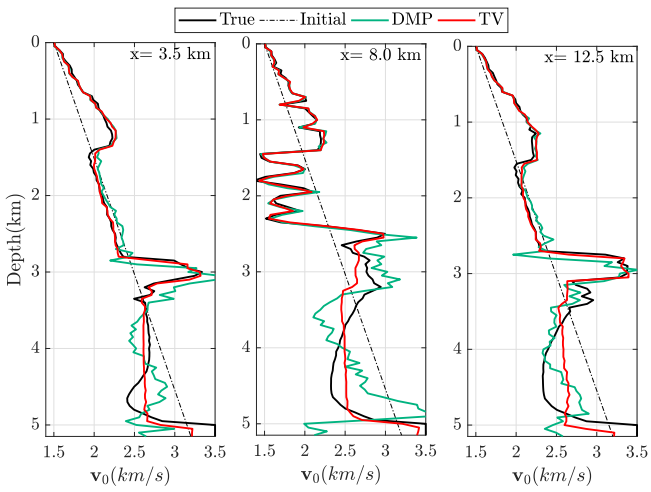

Figure 9. North Sea case study with noiseless data. Direct comparison along the logs at $x = 3.5$ (left), $x = 8.0$ (center) and $x = 12.5\ km$ (right) between the true velocity model (black), the initial model (dashed line) and the bound constrained IR-WRI models shown in Fig. 8. The logs of the IR-WRI models obtained with DMP regularization and DMP+TV regularization are the green lines and the red lines, respectively.

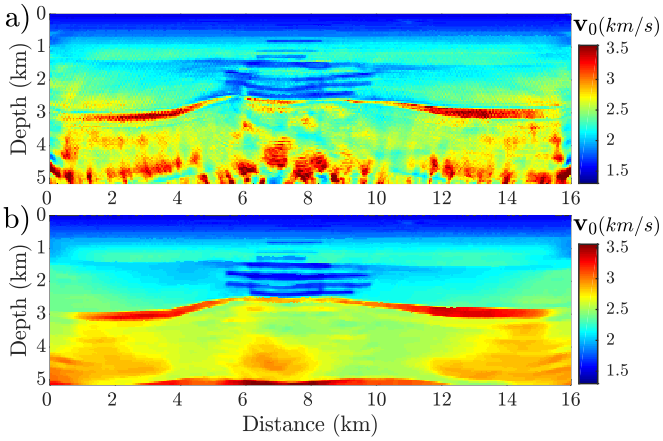

Figure 10. North Sea case study with noisy data (SNR=10 db). Mono-parameter bound constrained IR-WRI with (a) DMP (b) TV regularization.

pre- and post-critical incidences relative to the DMP regularization alone (Fig. 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. 13a and 13c for DMP regularization, and Figs. 13b and 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. 8a and 10a, and Figs. 9 and 11, green curves), compared to its more moderate impact on the data fit (Compare Figs. 13a and 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. 13b and 13b).

3.2.4 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

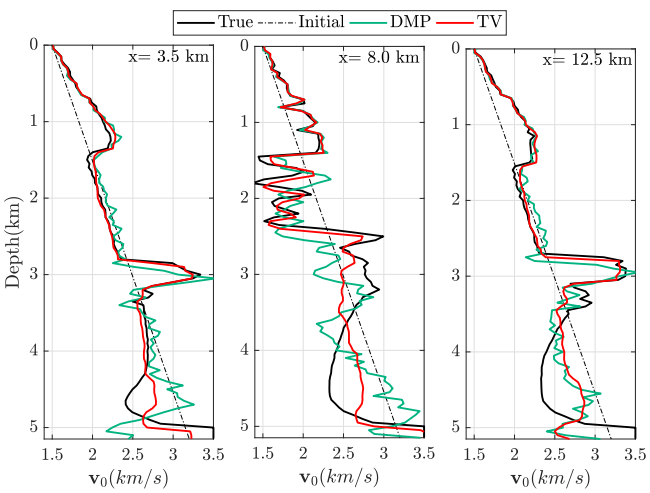

Figure 11. North Sea case study with noisy data (SNR=10 db). Direct comparison along the logs at $x = 3.5$ (left), $x = 8.0$ (center) and $x = 12.5\ km$ (right) between the true velocity model (black), the initial model (dashed line) and the bound constrained IR-WRI estimated models with DMP, Fig. 8a, (green) and TV regularization, Fig. 8b, (red).

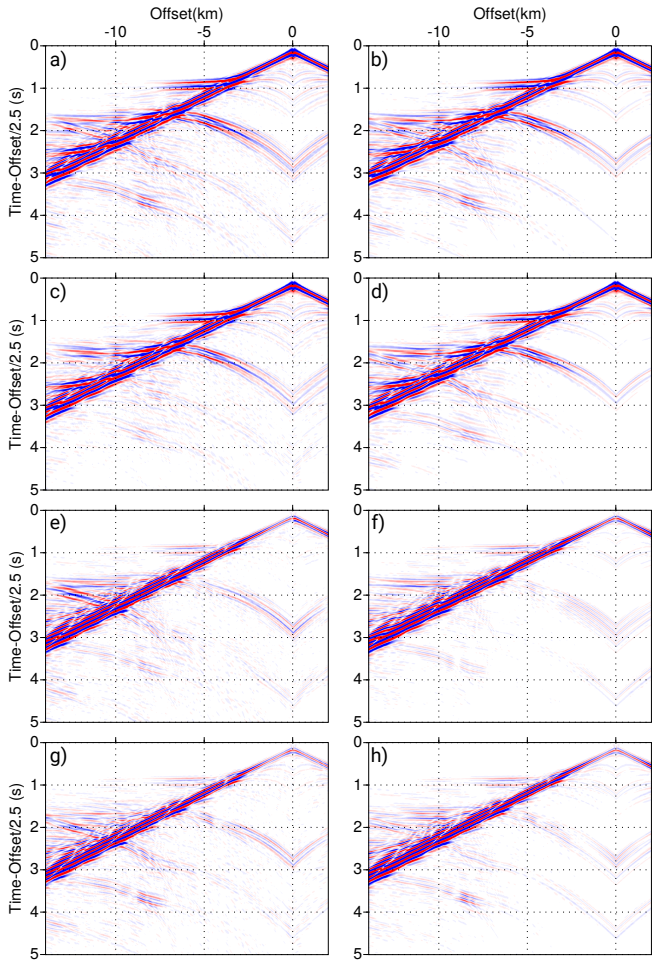

Figure 12. North Sea case study. Time-domain seismograms computed in the subsurface models obtained by ($\mathbf{v}_0$) mono-parameter IR-WRI. (a-b) Noiseless data: (a) DMP regularization, (b) DMP+TV regularization. (c-d) Same as (a-b) for noisy data. (e-h) Residuals between the seismograms computed in the true model (Fig. 6a) and those shown in (a-d). The seismograms are plotted with a reduction velocity of 2.5 km/s.

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. 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. 15a, and the same comparisons for ϵ are depicted in Fig. 15b. Compared to the 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. 8 and 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. 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. 18. The direct comparison between the recorded and modelled seismograms is shown in Fig. 19a,b. Clearly, using both $\mathbf{v}_0$ and ϵ as

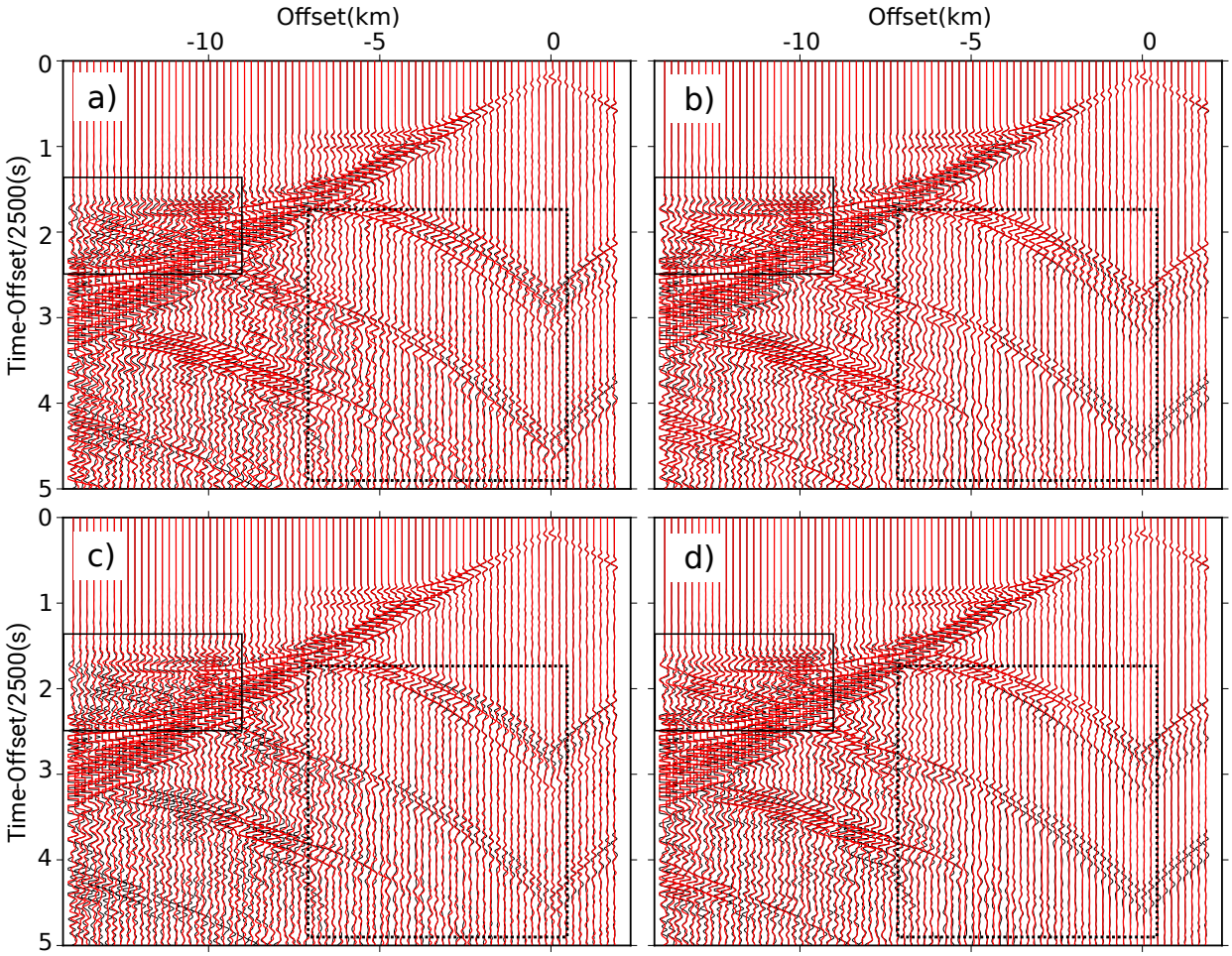

Figure 13. North Sea case study. Mono-parameter bound-constrained IR-WRI results. Direct comparison between seismograms computed in the true model (black) and in the IR-WRI models (red). (a-b) Noiseless data (IR-WRI models of Fig. 8). (a) DMP regularization. (b) DMP+TV regularization. (c-d) Same as (a-b) for noisy data (IR-WRI models of Fig. 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.

optimization parameters allows us to better conciliate the fit of the pre- and post-critical reflections (Compare Figs. 13a and 19a for DMP regularization, and Figs. 13b and 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. 19a and 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. 16. Direct comparisons along logs 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. 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. 14 and 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. 17b).

The time-domain seismograms are shown in Figs. 18 and 19c,d. The data fit obtained with DMP regularization has been significantly degraded compared to that obtained with noiseless data (compare Figs. 19a and 19c), while the data fit obtained with DMP+TV regularization is more consistent with noiseless and noisy data (compare Figs. 19b and 19d). This is consistent with the previous conclusions drawn from the mono-parameter inversion.

4 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 buoy-

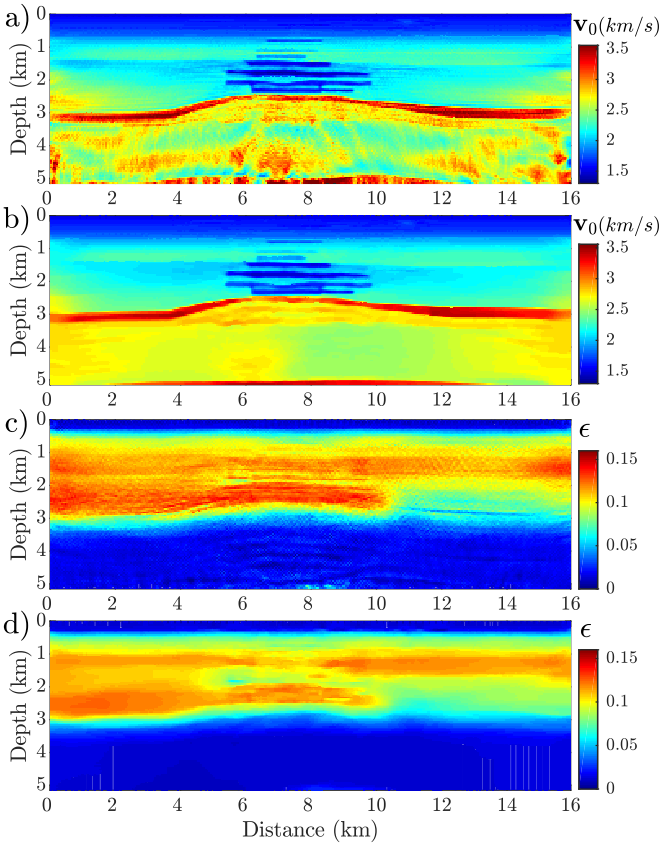

Figure 14. North Sea case study with noiseless data. Multi-parameter IR-WRI results. (a-b) $\mathbf{v}_0$ models obtained with (a) DMP (a) and (b) DMP+TV regularizations. (c-d) Same as (a-b) for ϵ.

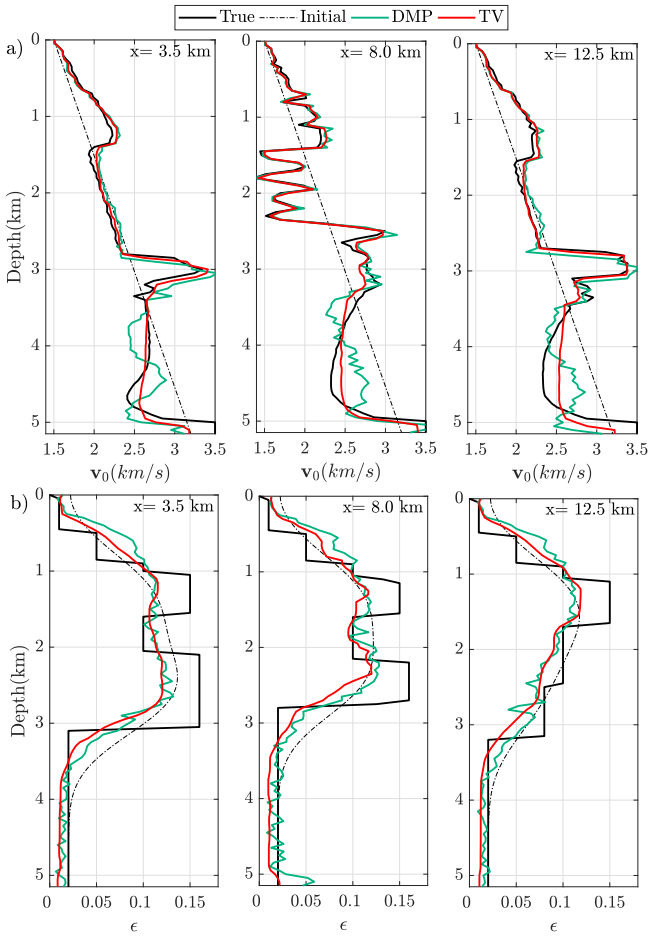

Figure 15. North Sea case study with noiseless data. Multi-parameter bound-constrained IR-WRI results. Direct comparison along logs at $x = 3.5$ (left), $x = 8.0$ (center) and $x = 12.5$ km (right) between the true model (black), the initial model (dashed line) and the models inferred from IR-WRI with DMP (green) and DMP+TV regularization (red) for (a) $\mathbf{v}_0$ and (b) ϵ (Fig. 14).

ancy) 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 (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 wave equation, hence keeping the parameter estimation subproblem, equation B.4a, nonlinear. Accordingly, equation B.4a 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 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. 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. 8-10 and Figs. 14-16; for the data fit, compare Fig. 13 and Fig. 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

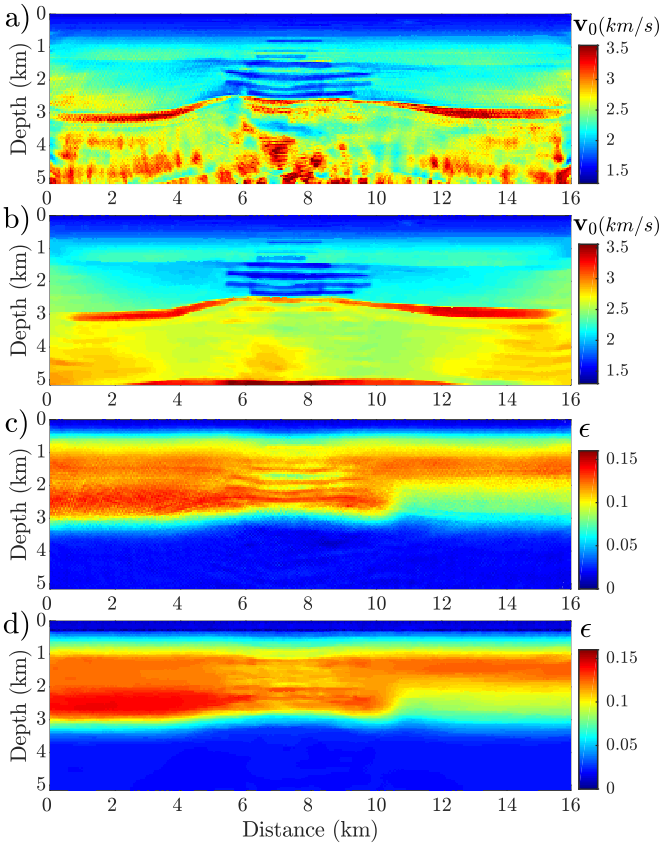

Figure 16. North Sea case study with noisy data (SNR=10 db). Multi-parameter bound-constrained IR-WRI results. (a-b) $\mathbf{v}_0$ models obtained with (a) DMP (a) and (b) DMP+TV regularizations. (c-d) Same as (a-b) for ϵ.

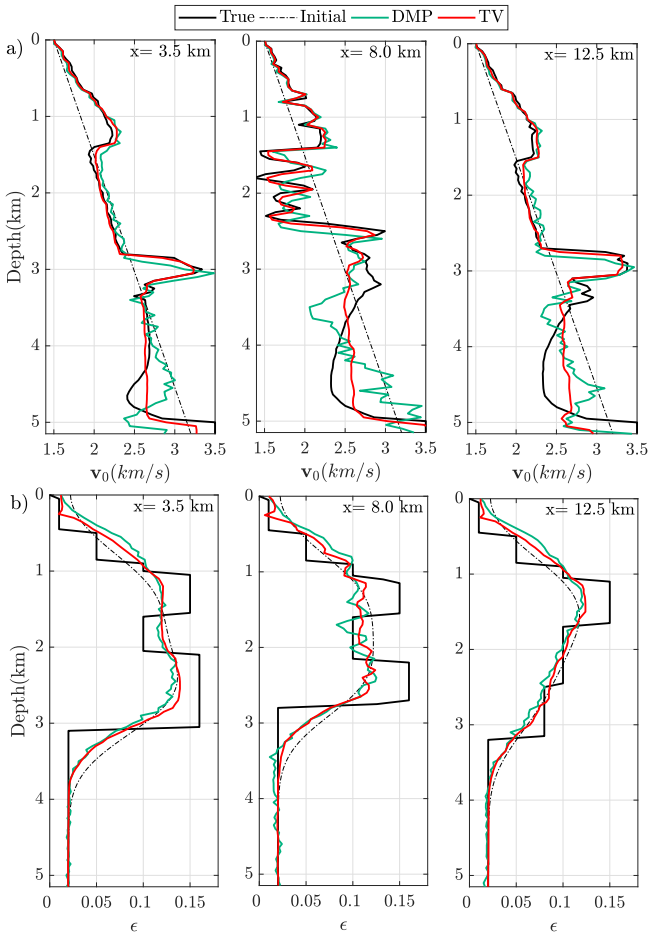

Figure 17. North Sea case study with noisy data. Multi-parameter bound-constrained IR-WRI results. Direct comparison along logs at $x = 3.5$ (left), $x = 8.0$ (center) and $x = 12.5\ km$ (right) between the true model (black), the initial model (dashed line) and the models inferred from IR-WRI with DMP (green) and DMP+TV regularization (red) for (a) $\mathbf{v}_0$ and (b) ϵ (Fig. 16).

estimate a crude ϵ background model with a coarse parametrisation by global optimization.

Among the alternative parametrisations proposed for VTI acoustic FWI, Plessix & 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 & 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 & Cao (2011), the VTI equation developed by Zhou H. (2006) given by

$$\begin{aligned}
\omega^2 \mathrm{diag}\left(\frac{1}{\mathbf{v}_n^2}\right)\mathbf{u}_q + 2\mathrm{diag}(\eta)\nabla_{xx}(\mathbf{u}_p+\mathbf{u}_q) &= \mathbf{s}_q, \\
\omega^2 \mathrm{diag}\left(\frac{1}{\mathbf{v}_n^2}\right)\mathbf{u}_p + \nabla_{xx}(\mathbf{u}_p+\mathbf{u}_q) &+ \\
\mathrm{diag}\left(\frac{1}{\sqrt{1+2\delta}}\right)\nabla_{zz}\mathrm{diag}\left(\frac{1}{\sqrt{1+2\delta}}\right)\mathbf{u}_p &= \mathbf{s}_p, \qquad (4.1)
\end{aligned}$$

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

$$\begin{aligned}
\omega^2 \mathrm{diag}\left(\frac{1}{\mathbf{v}_n^2}\right)\mathbf{u}_q + 2\mathrm{diag}(\eta)\nabla_{xx}(\mathbf{u}_p+\mathbf{u}_q) &= \mathbf{s}_q, \\
\omega^2 \mathrm{diag}\left(\frac{1}{\mathbf{v}_n^2}\right)\mathbf{u}_p + \nabla_{xx}(\mathbf{u}_p+\mathbf{u}_q) &+ \\
\mathrm{diag}\left(\frac{1}{1+2\delta}\right)\nabla_{zz}\mathbf{u}_p &= \mathbf{s}_p \qquad (4.2)
\end{aligned}$$

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 & Plessix 2014), according to Zhou H. (2006) the VTI equations with smooth δ can be written

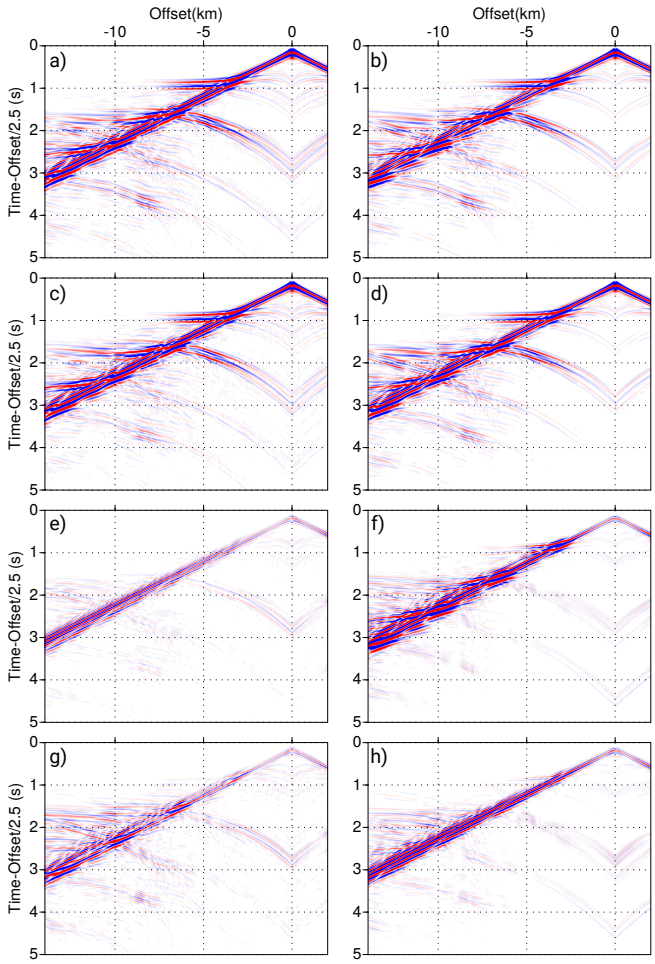

Figure 18. North Sea case study. Same as Fig. 12 for (v_0,ϵ) multi-parameter IR-WRI.

as

$$\begin{aligned}
\omega^2 \text{diag}\left(\frac{1}{\mathbf{v}_h^2}\right)\mathbf{u}_q + \nabla_{xx}(\mathbf{u}_p+\mathbf{u}_q) - \text{diag}\left(\frac{1}{1+2\eta}\right) & \\
\nabla_{xx}(\mathbf{u}_p+\mathbf{u}_q) &= \mathbf{s}_q, \\
\omega^2 \text{diag}\left(\frac{1}{\mathbf{v}_h^2}\right)\mathbf{u}_p + \text{diag}\left(\frac{1}{1+2\eta}\right)\nabla_{xx}(\mathbf{u}_p+\mathbf{u}_q) &+ \\
\text{diag}\left(\frac{1}{1+2\epsilon}\right)\nabla_{zz}\mathbf{u}_p &= \mathbf{s}_p,
\end{aligned} \tag{4.3}$$

which is bilinear in wavefields and parameters $(1/\mathbf{v}_h^2, 1/(1+2\eta), 1/(1+2\epsilon))$.

5 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.

6 ACKNOWLEDGMENTS

We would like to thank the Editor Huajian Yao, Xin Cheng and Daniel 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.

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

$$\begin{aligned}
-\hat{i}\omega\mathbf{v}_{x,l} &= \text{diag}(\mathbf{b})\nabla_x\mathbf{u}_{x,l}, \\
-\hat{i}\omega\mathbf{v}_{z,l} &= \text{diag}(\mathbf{b})\nabla_z\mathbf{u}_{z,l}, \\
-\hat{i}\omega[\text{diag}(\int_{11})\mathbf{u}_{x,l} + \text{diag}(\int_{13})\mathbf{u}_{z,l}] &= \nabla_x\mathbf{v}_{x,l} - \hat{i}\omega\mathbf{s}_l, \\
-\hat{i}\omega[\text{diag}(\int_{13})\mathbf{u}_{x,l} + \text{diag}(\int_{33})\mathbf{u}_{z,l}] &= \nabla_z\mathbf{v}_{z,l} - \hat{i}\omega\mathbf{s}_l,
\end{aligned} \tag{A.1}$$

where $\int_{ij} \in \mathbb{R}^{n\times 1}$ are the compliance coefficients, and the other notations are defined after equation 2.6. Gathering equation A.1 for all sources results in the following matrix equation:

$$\begin{bmatrix} \hat{i}\omega\mathbf{I} & \mathbf{B}\nabla \\ \nabla & \hat{i}\omega\mathcal{S} \end{bmatrix} \begin{bmatrix} \mathbf{V} \\ \mathbf{U} \end{bmatrix} = \hat{i}\omega \begin{bmatrix} \mathbf{0} \\ \mathbf{S} \end{bmatrix}, \tag{A.2}$$

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 2.7. Equation A.2 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

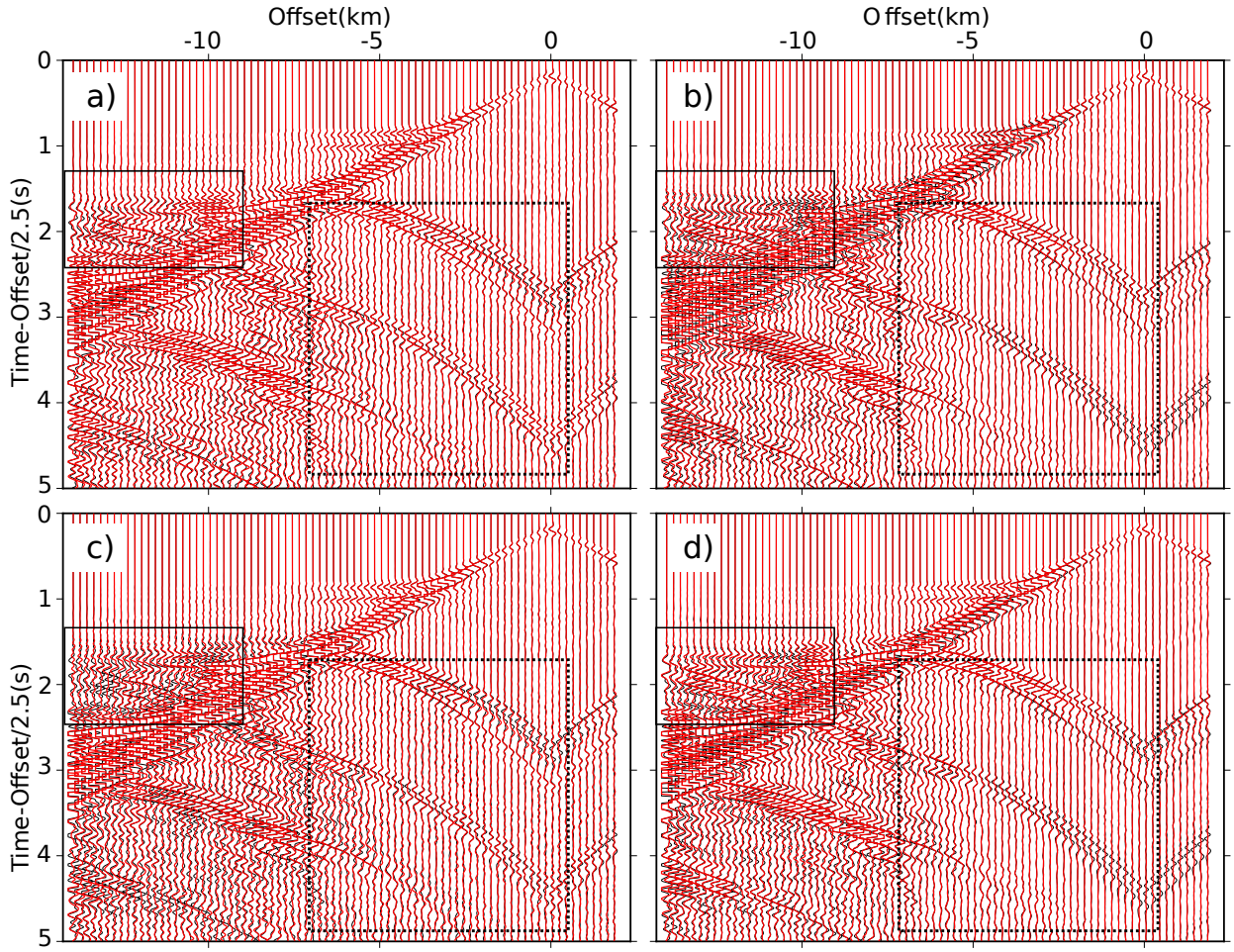

Figure 19. North Sea case study. Same as Fig. 13 for (v_0,ϵ) multi-parameter IR-WRI.

parameters, and hence the bilinearity of the wave equation. For the lth source, the new equation becomes

$$\begin{bmatrix} \mathbf{L}_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{L}_{21} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{32} & \mathbf{L}_{33} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{L}_{43} & \mathbf{L}_{44} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \int_{11} \\ \int_{13} \\ \int_{33} \end{bmatrix} = \begin{bmatrix} \mathbf{0} - \hat{\imath}\omega \mathbf{v}_{x,l} \\ \mathbf{0} - \hat{\imath}\omega \mathbf{v}_{z,l} \\ \hat{\imath}\omega \mathbf{s}_l - \nabla_x \mathbf{v}_{x,l} \\ \hat{\imath}\omega \mathbf{s}_l - \nabla_z \mathbf{v}_{z,l} \end{bmatrix}, \quad \text{(A.3)}$$

where

$$\begin{cases} \mathbf{L}_{11} = \text{diag}(\nabla_x \mathbf{u}_{x,l}), \\ \mathbf{L}_{21} = \text{diag}(\nabla_z \mathbf{u}_{z,l}), \\ \mathbf{L}_{32} = \mathbf{L}_{43} = \hat{\imath}\omega \text{diag}(\mathbf{u}_{x,l}), \\ \mathbf{L}_{33} = \mathbf{L}_{44} = \hat{\imath}\omega \text{diag}(\mathbf{u}_{z,l}). \end{cases}$$

To develop the second-order wave equation, we eliminate $\mathbf{v}_{x,l}$ and $\mathbf{v}_{z,l}$ from equation A.1. We obtain the following equation

$$\begin{aligned} &\frac{1}{\omega^2} \nabla_x \text{diag}(\mathbf{b}) \nabla_x \mathbf{u}_{x,l} + \text{diag}(\int_{11}) \mathbf{u}_{x,l} + \text{diag}(\int_{13}) \mathbf{u}_{z,l} = \mathbf{s}_l, \\ &\frac{1}{\omega^2} \nabla_z \text{diag}(\mathbf{b}) \nabla_z \mathbf{u}_{z,l} + \text{diag}(\int_{13}) \mathbf{u}_{x,l} + \text{diag}(\int_{33}) \mathbf{u}_{z,l} = \mathbf{s}_l, \end{aligned} \quad \text{(A.4)}$$

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

$$\mathbf{A}(\mathbf{m})\mathbf{U} = \mathbf{S}, \quad \text{(A.5)}$$

where $\mathbf{A}$ is given by

$$\begin{bmatrix} \frac{1}{\omega^2} \nabla_x \text{diag}(\mathbf{b}) \nabla_x + \text{diag}(\int_{11}) & \text{diag}(\int_{13}) \\ \text{diag}(\int_{13}) & \frac{1}{\omega^2} \nabla_z \text{diag}(\mathbf{b}) \nabla_z + \text{diag}(\int_{33}) \end{bmatrix}.$$

and

$$\mathbf{m} = \begin{bmatrix} \mathbf{b} \\ \int_{11} \\ \int_{13} \\ \int_{33} \end{bmatrix},$$

and $\mathbf{S}$ is defined in equation 2.8. 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

$$\begin{bmatrix} \mathbf{L}_1 \\ \vdots \\ \mathbf{L}_l \\ \vdots \\ \mathbf{L}_{n_s} \end{bmatrix} \mathbf{m} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_l \\ \vdots \\ \mathbf{y}_{n_s} \end{bmatrix}, \quad \text{(A.6)}$$

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

$$\mathbf{y}_l = \begin{bmatrix} \mathbf{s}_l \\ \mathbf{s}_l \end{bmatrix}.$$

APPENDIX B: SOLVING THE OPTIMIZATION PROBLEM, EQUATION 2.16, 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 2.16 reads as (see Aghamiry et al. 2019c,b,d, for more details)

$$\mathbf{U}^{k+1} \leftarrow \arg\min_{\mathbf{U}} \; C_{\mathbf{m}^k,\tilde{\mathbf{D}}^k,\tilde{\mathbf{S}}^k}(\mathbf{U}), \tag{B.1a}$$

$$\mathbf{m}^{k+1} \leftarrow \arg\min_{\mathbf{m}\in\mathcal{C}} \; C_{\mathbf{U}^{k+1},\tilde{\mathbf{S}}^k}(\mathbf{m}), \tag{B.1b}$$

$$\tilde{\mathbf{S}}^{k+1} \leftarrow \tilde{\mathbf{S}}^k + \mathbf{S} - \mathbf{A}(\mathbf{m}^{k+1})\mathbf{U}^{k+1}, \tag{B.1c}$$

$$\tilde{\mathbf{D}}^{k+1} \leftarrow \tilde{\mathbf{D}}^k + \mathbf{D} - \mathbf{P}\mathbf{U}^{k+1}, \tag{B.1d}$$

where

$$\begin{aligned} C_{\mathbf{m}^k,\tilde{\mathbf{D}}^k,\tilde{\mathbf{S}}^k}(\mathbf{U}) = \lambda_0 \|\mathbf{P}\mathbf{U} - \mathbf{D} - \tilde{\mathbf{D}}^k\|_F^2 \\ + \lambda_1 \|\mathbf{A}(\mathbf{m}^k)\mathbf{U} - \mathbf{S} - \tilde{\mathbf{S}}^k\|_F^2 \end{aligned} \tag{B.2}$$

and

$$\begin{aligned} C_{\mathbf{U}^{k+1},\tilde{\mathbf{S}}^k}(\mathbf{m}) = \sum \sqrt{|\partial_x \mathbf{m}|^2 + |\partial_z \mathbf{m}|^2} \\ + \lambda_1 \|\mathbf{A}(\mathbf{m})\mathbf{U}^{k+1} - \mathbf{S} - \tilde{\mathbf{S}}^k\|_F^2. \end{aligned} \tag{B.3}$$

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 B.2, 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 2.17 (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 B.3, 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 & Osher 2009; Aghamiry et al. 2019b):

$$\begin{aligned} \mathbf{m}^{k+1} \leftarrow \arg\min_{\mathbf{m}} \; \lambda_1 \|\mathbf{A}(\mathbf{m})\mathbf{U}^{k+1} - \mathbf{S} - \tilde{\mathbf{S}}^k\|_F^2, \\ + \|\overline{\nabla}\mathbf{m} - \mathbf{p}^k - \tilde{\mathbf{p}}^k\|_{\mathbf{\Gamma}}^2 + \|\mathbf{m} - \mathbf{q}^k - \tilde{\mathbf{q}}^k\|_Z^2, \end{aligned} \tag{B.4a}$$

$$\begin{aligned} \mathbf{p}^{k+1} \leftarrow \arg\min_{\mathbf{p}} \; \sum \sqrt{|\mathbf{p}_x|^2 + |\mathbf{p}_z|^2} \\ + \|\overline{\nabla}\mathbf{m}^{k+1} - \mathbf{p} - \tilde{\mathbf{p}}^k\|_{\mathbf{\Gamma}}^2, \end{aligned} \tag{B.4b}$$

$$\mathbf{q}^{k+1} \leftarrow \arg\min_{\mathbf{q}\in\mathcal{C}} \; \|\mathbf{m}^{k+1} - \mathbf{q} - \tilde{\mathbf{q}}^k\|_Z^2, \tag{B.4c}$$

where $\overline{\nabla}$ is defined in equation 2.19,

$$\mathbf{p} = \begin{bmatrix} \mathbf{p}_x \\ \mathbf{p}_z \end{bmatrix}, \tag{B.5}$$

$\|\mathbf{x}\|_\bullet^2 = \mathbf{x}^T \bullet \mathbf{x}$, Z and $\mathbf{\Gamma}$ are diagonal matrices defined in equation 2.20 and 2.21, respectively.

From the linearized equation 2.13, the subproblem for $\mathbf{m}$, equation B.4a, can be written as

$$\mathbf{m}^{k+1} \leftarrow \arg\min_{\mathbf{m}} \left\| \begin{bmatrix} \mathbf{L}_1 \\ \vdots \\ \mathbf{L}_{n_s} \\ \overline{\nabla} \\ \mathbf{I} \end{bmatrix} \mathbf{m} - \begin{bmatrix} \mathbf{y}_1^k + \tilde{\mathbf{s}}_1^k \\ \vdots \\ \mathbf{y}_{n_s}^k + \tilde{\mathbf{s}}_{n_s}^k \\ \mathbf{p}^k + \tilde{\mathbf{p}}^k \\ \mathbf{q}^k + \tilde{\mathbf{q}}^k \end{bmatrix} \right\|_{\mathbf{\Xi}}^2, \tag{B.6}$$

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 B.6 is now quadratic and admits a closed form solution as given in equation 2.18 (step 2 of the algorithm).

The only remaining tasks consist in determining the auxiliary primal variables $(\mathbf{p}, \mathbf{q})$, equations B.4b,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 2.22 (see Combettes & 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 2.24 (step 4 of the algorithm). Finally, the duals are updated according to gradient ascend steps (step 5 of the algorithm)

$$\begin{cases} \tilde{\mathbf{p}}^{k+1} &= \tilde{\mathbf{p}}^k + \mathbf{p}^{k+1} - \overline{\nabla}\mathbf{m}^{k+1} \\ \tilde{\mathbf{q}}^{k+1} &= \tilde{\mathbf{q}}^k + \mathbf{q}^{k+1} - \mathbf{m}^{k+1}. \end{cases} \tag{B.7}$$

APPENDIX C: USING FOURTH-ORDER EQUATION FOR WAVEFIELD RECONSTRUCTION (STEP 1 OF THE ALGORITHM)

The wavefield reconstruction subproblem, equation 2.17, can be written as the following over-determined system

$$\begin{bmatrix} \omega^2 \mathrm{diag}(\mathbf{m}_{v_0}) + \mathrm{diag}(\mathbf{m}_\epsilon)\nabla_{xx} & \mathrm{diag}(\mathbf{m}_\delta)\nabla_{zz} \\ \mathrm{diag}(\mathbf{m}_\delta)\nabla_{xx} & \omega^2 \mathrm{diag}(\mathbf{m}_{v_0}) + \nabla_{zz} \\ \frac{1}{2}\tilde{\mathbf{P}} & \frac{1}{2}\tilde{\mathbf{P}} \end{bmatrix} \begin{bmatrix} \mathbf{U}_x \\ \mathbf{U}_z \end{bmatrix} = \begin{bmatrix} \mathbf{S}_x \\ \mathbf{S}_z \\ \mathbf{D} \end{bmatrix}, \tag{C.1}$$

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 C.1 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 C.1 by $\mathrm{diag}(\mathbf{m}_\delta)$ and taking the difference with the first row, we find

$$\begin{bmatrix} \omega^2 \mathrm{diag}(\mathbf{m}_{v_0}) + \mathrm{diag}(\mathbf{m}_\epsilon)\nabla_{xx} & \mathrm{diag}(\mathbf{m}_\delta)\nabla_{zz} \\ \omega^2 \mathrm{diag}(\mathbf{m}_{v_0}) + \mathrm{diag}(\mathbf{m}_\epsilon - \mathbf{m}_\delta^2)\nabla_{xx} & -\omega^2 \mathrm{diag}(\mathbf{m}_\delta \circ \mathbf{m}_{v_0}) \\ \frac{1}{2}\tilde{\mathbf{P}} & \frac{1}{2}\tilde{\mathbf{P}} \end{bmatrix} \begin{bmatrix} \mathbf{U}_x \\ \mathbf{U}_z \end{bmatrix} = \begin{bmatrix} \mathbf{S}_x \\ \mathbf{S}_x - \mathrm{diag}(\mathbf{m}_\delta)\mathbf{S}_z \\ \mathbf{D} \end{bmatrix}. \tag{C.2}$$

where $\mathbf{m}_\delta^2 = \mathbf{m}_\delta \circ \mathbf{m}_\delta$. The second equation of C.2 provides us the closed-form expression of $\mathbf{U}_z$ as a function of $\mathbf{U}_x$

$$\mathbf{U}_z = \mathbf{A}_z \mathbf{U}_x + \mathbf{B}_z, \tag{C.3}$$

where

$$\begin{cases} \mathbf{A}_z = \frac{\mathrm{diag}(\mathbf{m}_\epsilon - \mathbf{m}_\delta^2)\nabla_{xx} + \omega^2 \mathrm{diag}(\mathbf{m}_{v_0})}{\omega^2 \mathrm{diag}(\mathbf{m}_{v_0} \circ \mathbf{m}_\delta)}, \\ \mathbf{B}_z = \frac{\mathrm{diag}(\mathbf{m}_\delta)\mathbf{S}_z - \mathbf{S}_x}{\omega^2 \mathrm{diag}(\mathbf{m}_{v_0} \circ \mathbf{m}_\delta)}. \end{cases} \tag{C.4}$$

Injecting equation C.4 into the first equation of C.2 leads to the over-determined system satisfied by $\mathbf{U}_x$

$$\begin{bmatrix} \omega^2 \text{diag}(\mathbf{m}_{v_0}) + \text{diag}(\mathbf{m}_\epsilon)\nabla_{xx} + \text{diag}(\mathbf{m}_\delta)\nabla_{zz}\mathbf{A}_z \\ \frac{1}{2}\tilde{\mathbf{P}} + \frac{1}{2}\tilde{\mathbf{P}}\mathbf{A}_z \end{bmatrix} \mathbf{U}_x = \begin{bmatrix} \mathbf{S}_x - \text{diag}(\mathbf{m}_\delta)\nabla_{zz}\mathbf{B}_z \\ \mathbf{D} - \frac{1}{2}\tilde{\mathbf{P}}\mathbf{B}_z \end{bmatrix}. \quad \text{(C.5)}$$

So, instead of solving the $(n_r + 2n) \times 2n$ linear system C.1 to update $\mathbf{U}$, $\mathbf{U}_x$ is updated by solving the $(n_r + n) \times n$ linear system C.5 and then $\mathbf{U}_z$ is updated using C.4 without significant computational overhead. The wave equation operator in equation C.4 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 C.4). 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).

REFERENCES

Aghamiry, H., Gholami, A., & Operto, S., 2018a. Hybrid Tikhonov + total-variation regularization for imaging large-contrast media by full-waveform inversion, in *Expanded Abstracts, 88th Annual SEG Meeting (Anaheim)*, pp. 1253–1257.

Aghamiry, H., Gholami, A., & Operto, S., 2018b. Improving full-waveform inversion based on wavefield reconstruction via Bregman iterations, in *Expanded Abstracts, 80th Annual EAGE Meeting (Copenhagen)*.

Aghamiry, H., Gholami, A., & Operto, S., 2018c. Imaging contrasted media with total variation constrained full waveform inversion and split Bregman iterations, in *Expanded Abstracts, 80th Annual EAGE Meeting (Copenhagen)*.

Aghamiry, H., Gholami, A., & Operto, S., 2019a. Compound Regularization of Full-Waveform Inversion for Imaging Piecewise Media, *arXiv:1903.04405*.

Aghamiry, H., Gholami, A., & 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., & 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., & Operto, S., 2019d. Multi-parameter ADMM-based wavefield reconstruction inversion in VTI acoustic media, in *81th Annual EAGE Meeting (London)*.

Alkhalifah, T. & Plessix, R., 2014. A recipe for practical full-waveform inversion in anisotropic media: An analytical parameter resolution study, *Geophysics*, **79**(3), R91–R101.

Benning, M., Knoll, F., Schönlieb, C.-B., & Valkonen, T., 2015. Preconditioned admm with nonlinear operator constraint, in *IFIP Conference on System Modeling and Optimization*, pp. 117–126, Springer.

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

Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J., 2010. Distributed optimization and statistical learning via the alternating direction of multipliers, *Foundations and trends in machine learning*, **3**(1), 1–122.

Chen, Z., Cheng, D., Feng, W., & 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., & Vigh, D., 2016. Multiparameter estimation with acoustic vertical transverse isotropic full-waveform inversion of surface seismic data, *Interpretation*, **4(4)**, SU1–SU16.

Combettes, P. L. & Pesquet, J.-C., 2011. Proximal splitting methods in signal processing, in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, vol. 49 of **Springer Optimization and Its Applications**, pp. 185–212, eds Bauschke, H. H., Burachik, R. S., Combettes, P. L., Elser, V., Luke, D. R., & Wolkowicz, H., Springer New York.

Debens, H. A., Warner, M., Umpleby, A., & da Silva, N. V., 2015. Global anisotropic 3d fwi, in *SEG Technical Program Expanded Abstracts*.

Duveneck, E. & 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., & Perkins, C., 2008. Acoustic VTI wave equations and their application for anisotropic reverse-time migration, *SEG Technical Program Expanded Abstracts*, **27**(1), 2186–2190.

Gholami, Y., Brossier, R., Operto, S., Prieux, V., Ribodetti, A., & 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., & 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., & Yin, W., 2017. *Splitting Methods in Communication, Imaging, Science, and Engineering*, Springer.

Goldstein, T. & Osher, S., 2009. The split Bregman method for L1-regularized problems, *SIAM Journal on Imaging Sciences*, **2**(2), 323–343.

Munns, J. W., 1985. The Valhall field: a geological overview, *Marine and Petroleum Geology*, **2**, 23–43.

Nocedal, J. & Wright, S. J., 2006. *Numerical Optimization*, Springer, 2nd edn.

Operto, S. & 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., Virieux, J., Amestoy, P., L'Éxcellent, J.-Y., Giraud, L., & Ben Hadj Ali, H., 2007. 3D finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: A feasibility study, *Geophysics*, **72**(5), SM195–SM211.

Operto, S., Brossier, R., Combe, L., Métivier, L., Ribodetti, A., & 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.

Plessix, R. E. & 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.

Pratt, R. G., Shin, C., & Hicks, G. J., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophysical Journal International*, **133**, 341–362.

Thomsen, L. A., 1986. Weak elastic anisotropy, *Geophysics*, **51**, 1954–1966.

van Leeuwen, T. & Herrmann, F., 2016. A penalty method for PDE-constrained optimization in inverse problems, *Inverse Problems*, **32(1)**, 1–26.

van Leeuwen, T. & Herrmann, F. J., 2013. Mitigating local minima in full-waveform inversion by expanding the search space, *Geophysical Journal International*, **195(1)**, 661–667.

Vigh, D., Jiao, K., Watts, D., & Sun, D., 2014. Elastic full-waveform inversion application using multicomponent measurements of seismic data collection, *Geophysics*, **79**(2), R63–R77.

Yang, P., Brossier, R., Métivier, L., & Virieux, J., 2016. A review on the systematic formulation of 3D multiparameter full waveform inversion in viscoelastic medium, *Geophysical Journal International*, **207**(1), 129–149.

Zhou H., Zhang G., B. R., 2006. An anisotropic acoustic wave equation for VTI media, in *Expanded Abstracts, 68th Annual meeting*, EAGE.