



HAL
open science

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

Hossein Aghamiry, A. Gholami, Stéphane Operto

► **To cite this version:**

Hossein Aghamiry, A. Gholami, Stéphane Operto. 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*, 2019, 218 (2), pp.855-872. 10.1093/gji/ggz189 . hal-02166848

HAL Id: hal-02166848

<https://hal.science/hal-02166848>

Submitted on 31 Aug 2021

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.



Distributed under a Creative Commons Attribution 4.0 International License

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

H. S. Aghamiry^{1,2}, A. Gholami¹ and S. Operto²

¹*Institute of Geophysics, University of Tehran, Tehran, Iran. E-mail: agholami@ut.ac.ir*

²*Université Côte d'Azur, CNRS, Observatoire de la Côte d'Azur, IRD, Géozur, F-06560, Valbonne, France*

Accepted 2019 April 19. Received 2018 September 6; in original form 2019 April 12

SUMMARY

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

Key words: Inverse theory; Numerical modelling; Waveform inversion; Controlled source seismology.

1 INTRODUCTION

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

the optimization parameters are embedded in the coefficients of the PDE. Due to the computational burden of multiple source modelling and the size of the data and parameter spaces, this PDE-constrained optimization problem is solved with iterative local (linearized) optimization techniques, namely gradient-based methods (Nocedal & Wright 2006). Moreover, it is often solved with a reduced-space formulation, which means that the full search space that encompasses the unknown wavefield and the subsurface parameters is first projected onto the parameter space by computing exactly the incident wavefields in the current subsurface model before updating this later (Haber *et al.* 2000; Askan *et al.* 2007; Epanomeritakis *et al.* 2008).

It is well acknowledged that the oscillating nature of seismic signals makes the reduced-space formulation highly non-linear as the modelled seismograms computed in the current subsurface model may be too far away from the recorded ones to satisfy the cycle-skipping criterion, that is the modelled seismograms should predict the recorded traveltimes with an error lower than half a period (e.g. Virieux & Operto 2009).

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

In this context, the objective of this study is to present a novel implementation of TV regularization and bound constraints in frequency-domain FWI based upon wavefield-reconstruction inversion (WRI). WRI has been originally proposed by van Leeuwen & Herrmann (2013) to extend the search space and mitigate the risk of cycle skipping accordingly. WRI recasts the PDE-constrained optimization problem underlying FWI into an unconstrained quadratic penalty method, where the penalty term is the ℓ_2 norm of the source residuals (namely, the PDE-constraint violation) that is weighted by a positive penalty parameter λ . The penalty method relaxes the wave-equation constraint at the benefit of the data fitting during early iterations, hence mitigating the risk of cycle skipping. To make WRI computationally tractable, van Leeuwen & Herrmann (2013) perform the wavefield reconstruction and the subsurface parameter estimation in an alternating way: first, keeping the current subsurface model fixed, the wavefields, which best jointly fit the observations and satisfy the wave equation in a least-squares sense, are reconstructed for each source; Second, keeping the previously reconstructed wavefields fixed, the subsurface parameters are estimated by least-squares minimization of the source residuals the wave-equation relaxation generated. This cycle being iterated until

convergence. A nice property of the alternating-direction strategy is to linearize the parameter-estimation subproblem around the reconstructed wavefield because the wave-equation constraint is bilinear. However, a significant pitfall of WRI resides in the tuning of the penalty parameter λ . Ideally, increasing values should be used during iterations to progressively enforce the wave-equation constraint and, hence satisfy the first-order optimality conditions of the original constrained problem with acceptable precision at the minimizer. A significant issue is that this continuation approach is tedious to implement and the Hessian is ill conditioned for large λ . Therefore, van Leeuwen & Herrmann (2013) implement WRI with a small preset value of λ , which leads to slow convergence and a subsurface model of limited accuracy.

Later, van Leeuwen & Herrmann (2016) reformulated WRI as a reduced penalty method implemented with a variable projection approach: the closed-form expression of the extended-domain reconstructed wavefield is injected as a function of the subsurface parameters in the penalty function instead of using this wavefield as a passive variable (i.e. independent to the subsurface parameters). Unlike the alternating-direction approach, this variable projection leaves the parameter-estimation subproblem non-linear. van Leeuwen & Herrmann (2016) assess their method with a Gauss–Newton method (by opposition to the full Newton counterpart) to mitigate the computational burden. Moreover, using a sparse approximation of the Gauss–Newton Hessian makes the descent direction of the reduced approach identical to that of the alternating-direction WRI of van Leeuwen & Herrmann (2013). Also, Aravkin *et al.* (2017) analysed the convergence properties of the reduced penalty method when the full Hessian is taken into account and concluded that the variable projection penalty method is insensitive to the penalty parameter. However, this convergence property still needs to be verified against realistic numerical experiments.

To make the alternating-direction WRI of van Leeuwen & Herrmann (2013) (referred to as WRI in the following for sake of brevity) more independent to the penalty parameter, Aghamiry *et al.* (2019b, 2018b) have replaced the penalty method by an augmented Lagrangian (AL) method (Nocedal & Wright 2006), leading to the so-called iteratively-refined (IR)-WRI method. As in WRI, IR-WRI performs the primal wavefield and parameter updates in an alternating mode, while the Lagrange multipliers (i.e. the dual variables) are updated with a gradient ascent method. As above mentioned, this alternating direction strategy makes the parameter-estimation subproblem linear due to the bilinearity of the wave-equation constraint. It follows from this linearization that the alternating direction strategy combined with the AL method is equivalent to an extension of the alternating direction method of multiplier (ADMM) to biconvex problem (Boyd *et al.* 2010). Also, using a scaled form of the AL, we recast IR-WRI as a penalty method where the right-hand sides (the data and the sources) in the objective functions are iteratively updated with the running sum of the data and source residuals (the dual gradient steps). This reformulation of IR-WRI as a penalty method with right-hand side updating clearly draws some similarities and differences with WRI. The right-hand side updating makes IR-WRI largely insensitive to the penalty parameter for a wide range of preset values (Aghamiry *et al.* 2019b, their figs 2 and 3). Using a moderate value of the penalty parameter allows for significant wave-equation error and improved data fitting during early iterations for search space extension, without preventing the fulfillment of the wave-equation constraint with small error at the minimizer (Nocedal & Wright 2006, chapter 17, theorem 17.6). This adaptivity makes IR-WRI resilient to cycle skipping as WRI with however a much faster convergence toward a more accurate

where $\lambda = \lambda_1/\lambda_0$. They solve this biconvex minimization problem with an alternating-direction approach to breakdown the full problem into a sequence of two linear subproblems: A cycle of the algorithm first reconstructs, for each source, the wavefield \mathbf{u} that best fits the data and satisfies the wave equation in a least-squares sense for the current subsurface model. Then, the subsurface model is updated by minimization of the source residuals with a Gauss–Newton algorithm keeping the reconstructed wavefields fixed. A difficulty with the penalty method given by eq. (7) resides in the tuning of the penalty parameter λ during iterations. An increasing values of λ should be used during iterations, known as the penalty algorithm, to progressively enforce the wave-equation constraint in iterations and hence satisfy the Karush–Kuhn–Tucker optimality conditions (Nocedal & Wright 2006) associated with the original constrained problem with acceptable precision. The main problem is that this continuation strategy is tedious to implement and a large λ makes the problem severely ill-conditioned.

2.1.2 IR-WRI augmented Lagrangian method

To bypass this difficulty, IR-WRI implements the original constrained problem, eq. (4), with the AL method (Hestenes 1969; Nocedal & Wright 2006; Boyd *et al.* 2010; Bertsekas 2016).

$$\min_{\mathbf{u}, \mathbf{m}} \max_{\mathbf{v}} \mathbf{0} + \mathbf{v}^T [\mathbf{F}(\mathbf{m})\mathbf{u} - \mathbf{s}] + \frac{1}{2} \|\mathbf{F}(\mathbf{m})\mathbf{u} - \mathbf{s}\|_{\Lambda}^2, \quad (8)$$

where Λ is defined as in eq. (6) and $\mathbf{v} \in \mathbb{C}^{(M+N) \times 1}$ is the Lagrangian multiplier (known as dual variable). Comparing the penalty function, eq. (5), and the AL function, eq. (8), clearly shows that the AL combines a penalty method with a Lagrangian method. The first advantage of the AL method relative to the penalty method is to prevent ill-conditioning by introducing explicit estimate of the Lagrange multiplier in the optimization (Nocedal & Wright 2006, chap. 17). Moreover, the Lagrange multiplier gives the AL method one more way of improving the accuracy of the minimizer in addition to the penalty parameter, hence allows fixed value to be used for this latter (Nocedal & Wright 2006, theorem 17.6). Applying the alternating direction strategy of WRI on the AL method leads to an adaptation of the ADMM to biconvex problem (Boyd *et al.* 2010, section 9.2). One ADMM iteration first minimizes the AL function with respect to the primal variables \mathbf{u} and \mathbf{m} via a single Gauss–Seidel-like iteration (namely, fix one variable and solve for the other) and then update the Lagrangian multiplier via a gradient ascent method.

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

2.2 BTV-regularized ADMM-based IR-WRI

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

$$\min_{\mathbf{u}, \mathbf{m} \in \mathcal{C}} \|\mathbf{m}\|_{\text{TV}} \quad \text{subject to} \quad \mathbf{F}(\mathbf{m})\mathbf{u} = \mathbf{s}, \quad (9)$$

where \mathbf{F} and \mathbf{s} are defined as in eq. (2), $\|\mathbf{m}\|_{\text{TV}} = \sum \sqrt{|\nabla_1 \mathbf{m}|^2 + |\nabla_2 \mathbf{m}|^2}$ is the blockiness-promoting isotropic TV

norm (Rudin *et al.* 1992) and ∇_1 and ∇_2 are first-order finite-difference operators in the horizontal and vertical directions, respectively. With notation abuse, the absolute sign, square power and the square root operations are done componentwise, and the sum runs over all elements (the domain of parameters). Also $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^{N \times 1} \mid \mathbf{m}_l \leq \mathbf{x} \leq \mathbf{m}_u\}$ is the set of all feasible models bounded by the lower bound \mathbf{m}_l and the upper bound \mathbf{m}_u .

Compared to the FWI definition given in the previous section, eq. (4), we have replaced the identically zero objective function by the TV norm of the model and restricted the space of feasible models to \mathcal{C} . Accordingly, the AL function for the problem defined by eq. (9) is

$$\mathcal{L}_A(\mathbf{m}, \mathbf{u}, \mathbf{v}) = \|\mathbf{m}\|_{\text{TV}} + \mathbf{v}^T [\mathbf{F}(\mathbf{m})\mathbf{u} - \mathbf{s}] + \frac{1}{2} \|\mathbf{F}(\mathbf{m})\mathbf{u} - \mathbf{s}\|_{\Lambda}^2, \quad (10)$$

where the same notations as those of eq. (8) are used. Eq. (10) can also be written in a more compact form as

$$\mathcal{L}_A(\mathbf{m}, \mathbf{u}, \bar{\mathbf{s}}) = \|\mathbf{m}\|_{\text{TV}} + \frac{1}{2} \|\mathbf{F}(\mathbf{m})\mathbf{u} - \mathbf{s} - \bar{\mathbf{s}}\|_{\Lambda}^2 - \frac{1}{2} \|\bar{\mathbf{s}}\|_{\Lambda}^2, \quad (11)$$

where $\bar{\mathbf{s}} = -\Lambda^{-1} \mathbf{v}$ is the scaled dual variable and eq. (11) is the scaled form of the AL (Boyd *et al.* 2010, p. 15 and appendix A). The scaled form recasts the AL function as a quadratic penalty function where the right-hand sides are updated with the scaled dual variables. This highlights similarities and differences between WRI and IR-WRI, since this right-hand side updating is lacking in the former.

The method of multipliers seeks to find the saddle point of the scaled AL (11) through a primal descent–dual ascent updating resulting in the following iteration:

$$\mathbf{m}^{k+1}, \mathbf{u}^{k+1} = \arg \min_{\mathbf{u}, \mathbf{m} \in \mathcal{C}} \|\mathbf{m}\|_{\text{TV}} + \frac{1}{2} \|\mathbf{F}(\mathbf{m})\mathbf{u} - \mathbf{s} - \bar{\mathbf{s}}^k\|_{\Lambda}^2, \quad (12a)$$

$$\bar{\mathbf{s}}^{k+1} = \bar{\mathbf{s}}^k + \mathbf{s} - \mathbf{F}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}, \quad (12b)$$

for $k = 0, 1, \dots$ beginning with a prior estimate $\bar{\mathbf{s}}^0 = \mathbf{0}$. The iteration (12) can be viewed as follows: we begin with a prior estimate of the dual $\bar{\mathbf{s}}^0$, and minimize the objective function with respect to the primal variables \mathbf{m} and \mathbf{u} , eq. (12a). Subsequently, we maximize the objective function with respect to the dual variable $\bar{\mathbf{s}}$ with a gradient ascent method when \mathbf{m} and \mathbf{u} are kept fixed, eq. (12b). The steepest-ascent step (eq. 12b) shows that the scaled dual variable $\bar{\mathbf{s}}$ is updated with the residual constraint violation of the current iteration. Remembering that the constraint combines the observation equation $\mathbf{P}\mathbf{u} = \mathbf{d}$ and the wave equation $A(\mathbf{m})\mathbf{u} = \mathbf{b}$, eq. (2), the scaled dual variable $\bar{\mathbf{s}}$ updates the right-hand sides of the quadratic penalty function, eq. (12a), with the running sum of the data and source residuals in iterations. This right-hand side updating describes the well-known iterative solution refinement procedure for ill-posed linear inverse problems as reviewed by Aghamiry *et al.* (2019b, their appendix B). This process is iterated until convergence, i.e. when $\mathbf{F}(\mathbf{m}^{k+1})\mathbf{u}^{k+1} = \mathbf{s}$. In the following, we remove the bar of $\bar{\mathbf{s}}^k$ for the sake of simplicity.

Solving the subproblem (12a) jointly for the primal variables (\mathbf{m} , \mathbf{u}) is computationally too intensive. A splitting method is useful here to breakdown the joint optimization over \mathbf{m} and \mathbf{u} into two subproblems (the readers can refer to Glowinski *et al.* 2017 for an overview of splitting methods). The ADMM (Boyd *et al.* 2010) provides a simple framework to achieve this goal via a single Gauss–Seidel-like iteration leading to the following iteration:

$$\mathbf{u}^{k+1} = \arg \min_{\mathbf{u}} \frac{1}{2} \|\mathbf{F}(\mathbf{m}^k)\mathbf{u} - \mathbf{s} - \mathbf{s}^k\|_{\Lambda}^2, \quad (13a)$$

$$\mathbf{m}^{k+1} = \arg \min_{\mathbf{m} \in \mathcal{C}} \|\mathbf{m}\|_{\text{TV}} + \frac{1}{2} \|\mathbf{F}(\mathbf{m})\mathbf{u}^{k+1} - \mathbf{s} - \mathbf{s}^k\|_{\Lambda}^2, \quad (13b)$$

$$\mathbf{s}^{k+1} = \mathbf{s}^k + \mathbf{s} - \mathbf{F}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}. \quad (13c)$$

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

A second modification related to ADMM resides in the updating of the dual variable. In ADMM, the dual variables are updated only once per iteration after the primal-variable updates, eq. (13c). A variant of ADMM, referred to as the Peaceman–Rachford splitting method (PRSM, Peaceman & Rachford 1955), consists of updating the Lagrange multipliers several times, once after the update of each primal variable (see He *et al.* 2014, compare eqs 1.3 and 1.4). One issue with PRSM relative to ADMM is that PRSM requires more restrictive assumptions to ensure its convergence, while it is always faster than ADMM whenever it is convergent (He *et al.* 2014). This issue prompted He *et al.* (2014) to implement a relaxation factor (or, step length) $\alpha \in (0, 1)$ to guarantee the strict contraction of the PRSM iterative sequence. Applying the strictly contractive PRSM algorithm to our minimization problem gives

$$\mathbf{u}^{k+1} = \arg \min_{\mathbf{u}} \frac{1}{2} \|\mathbf{F}(\mathbf{m}^k)\mathbf{u} - \mathbf{s} - \mathbf{s}^k\|_{\Lambda}^2, \quad (14a)$$

$$\mathbf{s}^{k+\frac{1}{2}} = \mathbf{s}^k + \alpha[\mathbf{s} - \mathbf{F}(\mathbf{m}^k)\mathbf{u}^{k+1}], \quad (14b)$$

$$\mathbf{m}^{k+1} = \arg \min_{\mathbf{m} \in \mathcal{C}} \|\mathbf{m}\|_{\text{TV}} + \frac{1}{2} \|\mathbf{F}(\mathbf{m})\mathbf{u}^{k+1} - \mathbf{s} - \mathbf{s}^{k+\frac{1}{2}}\|_{\Lambda}^2, \quad (14c)$$

$$\mathbf{s}^{k+1} = \mathbf{s}^{k+\frac{1}{2}} + \alpha[\mathbf{s} - \mathbf{F}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}]. \quad (14d)$$

The reader is referred to Aghamiry *et al.* (2019b, their fig. 4) for a comparative numerical analysis of the convergence speed of ADMM and PRSM in IR-WRI. In our numerical tests, we found that $\alpha = 0.5$ can serve as a suitable value. We now provide the closed-form solution of the two primal subproblems associated with \mathbf{u} and \mathbf{m} .

2.2.1 Solving for \mathbf{u}

The primal subproblem associated with \mathbf{u} (eq. 14a) is a linear optimization problem whose solution satisfies in a least-squares sense the following system of linear eqs:

$$\begin{bmatrix} \lambda_0^{\frac{1}{2}} \mathbf{P} \\ \lambda_1^{\frac{1}{2}} \mathbf{A}(\mathbf{m}^k) \end{bmatrix} \mathbf{u}^{k+1} = \begin{bmatrix} \lambda_0^{\frac{1}{2}} [\mathbf{d} + \mathbf{d}^k] \\ \lambda_1^{\frac{1}{2}} [\mathbf{b} + \mathbf{b}^k] \end{bmatrix}, \quad (15)$$

where \mathbf{d}^k and \mathbf{b}^k are the components of the dual variable \mathbf{s}^k associated with the observation- and wave-equation constraints, and are formed by the running sum of the data and source residuals in iteration (see

eq. 12). The closed-form expression of the reconstructed wavefield is given by

$$\mathbf{u}^{k+1} = [\lambda_0 \mathbf{P}^T \mathbf{P} + \lambda_1 \mathbf{A}(\mathbf{m}^k)^T \mathbf{A}(\mathbf{m}^k)]^{-1} [\lambda_0 \mathbf{P}^T [\mathbf{d} + \mathbf{d}^k] + \lambda_1 \mathbf{A}(\mathbf{m}^k)^T [\mathbf{b} + \mathbf{b}^k]]. \quad (16)$$

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

2.2.2 Solving for \mathbf{m}

In order to solve the bound-constrained TV-regularized non-linear problem described by eq. (14c), we first tackle the non-linearity issue by considering the special structure of the Helmholtz operator \mathbf{A} given in eq. . Using the following approximation

$$\begin{aligned} \mathbf{A}(\mathbf{m})\mathbf{u}^{k+1} &= \Delta \mathbf{u}^{k+1} + \omega^2 \mathbf{C}(\mathbf{m}) \text{diag}(\mathbf{m}) \mathbf{B} \mathbf{u}^{k+1} \\ &\approx \Delta \mathbf{u}^{k+1} + \underbrace{\omega^2 \mathbf{C}(\mathbf{m}^k) \text{diag}(\mathbf{B} \mathbf{u}^{k+1})}_{\mathbf{L}(\mathbf{u}^{k+1})} \mathbf{m}, \end{aligned} \quad (17)$$

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

$$\mathbf{m}^{k+1} = \arg \min_{\mathbf{m} \in \mathcal{C}} \|\mathbf{m}\|_{\text{TV}} + \frac{\lambda_1}{2} \|\mathbf{L}(\mathbf{u}^{k+1})\mathbf{m} - \mathbf{y}^k\|_2^2, \quad (18)$$

where

$$\mathbf{y}^k = \mathbf{b} + \mathbf{b}^k - \Delta \mathbf{u}^{k+1}. \quad (19)$$

We solve the constrained optimization problem described by eq. (18) with the split-Bregman method (Goldstein & Osher 2009) to decouple the TV minimization subproblem (first term in eq. 18) from the ℓ_2 subproblem (second term in eq. 18) and force the box constraint. We introduce the auxiliary variables \mathbf{p}_0 , \mathbf{p}_1 and \mathbf{p}_2 to perform this splitting, in which these variables being related to \mathbf{m} by means of a simple equality constraint. The auxiliary variable \mathbf{p}_0 is used to enforce the box constraint, while \mathbf{p}_1 and \mathbf{p}_2 are the variables of the TV minimization problem. Let us define $(\mathbf{p}_1 \mathbf{p}_2)$ as a two-column matrix, $\|(\mathbf{p}_1 \mathbf{p}_2)\| = \sqrt{\mathbf{p}_1^2 + \mathbf{p}_2^2}$ be a vector which contains the ℓ_2 norm of each row of $(\mathbf{p}_1 \mathbf{p}_2)$, and $\sum \|(\mathbf{p}_1 \mathbf{p}_2)\|$ be the mixed $\ell_{2,1}$ norm (ℓ_1 norm of $\|(\mathbf{p}_1 \mathbf{p}_2)\|$), which promotes sparsity. Then by defining the objective function J as

$$J(\mathbf{m}, \mathbf{p}) = \sum \|(\mathbf{p}_1 \mathbf{p}_2)\| + \frac{\lambda_1}{2} \|\mathbf{L}(\mathbf{u}^{k+1})\mathbf{m} - \mathbf{y}^k\|_2^2, \quad (20)$$

the unconstrained optimization problem described by eq. (18) can be written in a split and constrained form as

$$\arg \min_{\mathbf{m}, \mathbf{p}, \mathbf{p}_0 \in \mathcal{C}} J(\mathbf{m}, \mathbf{p}) \quad \text{subject to} \quad \mathbf{p} = \nabla \mathbf{m}, \quad (21)$$

where

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

with \mathbf{I} denoting the identity matrix.

The scaled AL function (Appendix A) for the problem defined by eq. (21) is

$$\mathcal{L}_A(\mathbf{m}, \mathbf{p}, \bar{\mathbf{q}}) = J(\mathbf{m}, \mathbf{p}) + \frac{1}{2} \|\nabla \mathbf{m} - \mathbf{p} - \bar{\mathbf{q}}^k\|_{\Gamma}^2 - \frac{1}{2} \|\bar{\mathbf{q}}^k\|_{\Gamma}^2, \quad (22)$$

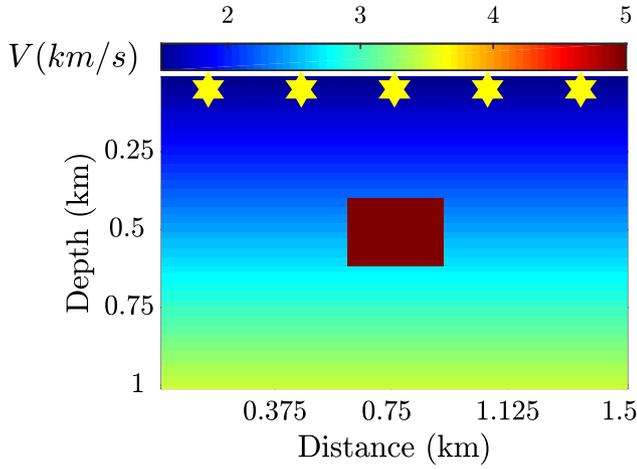


Figure 1. The true velocity of the box-shaped anomaly example. The yellow stars show the source positions.

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

$$\Gamma = \begin{bmatrix} \gamma_0 & 0 & & & \\ & \ddots & & & \\ 0 & & \gamma_0 & & \\ \hline & & & \gamma_1 & 0 \\ & & & & \ddots \\ 0 & & & 0 & \gamma_1 \\ \hline & & & & & \gamma_2 & 0 \\ & & & & & & \ddots \\ 0 & & & & & & 0 & \gamma_2 \end{bmatrix} \begin{matrix} N \\ N \\ N \end{matrix}, \quad (23)$$

with the penalty parameters $\gamma_0, \gamma_1, \gamma_2 > 0$. Again, applying the method of multipliers to find the saddle point of the problem (22) gives

$$\mathbf{m}^{k+1}, \mathbf{p}^{k+1} = \arg \min_{\mathbf{m}, \mathbf{p}, \mathbf{p}_0 \in \mathcal{C}} J(\mathbf{m}, \mathbf{p}) + \frac{1}{2} \|\nabla \mathbf{m} - \mathbf{p} - \mathbf{q}^k\|_{\Gamma}^2, \quad (24a)$$

$$\bar{\mathbf{q}}^{k+1} = \bar{\mathbf{q}}^k + \mathbf{p}^{k+1} - \nabla \mathbf{m}^{k+1}, \quad (24b)$$

where the bar of $\bar{\mathbf{q}}^k$ is removed for simplicity.

Substituting the explicit expression of J , eq. (20), into eq. (24a) leads to the following PRSM iteration:

$$\mathbf{m}^{k+1} = \arg \min_{\mathbf{m}} \frac{\lambda_1}{2} \|\mathbf{L}(\mathbf{u}^{k+1})\mathbf{m} - \mathbf{y}^k\|_2^2 + \frac{1}{2} \|\nabla \mathbf{m} - \mathbf{p}^k - \bar{\mathbf{q}}^k\|_{\Gamma}^2, \quad (25a)$$

$$\mathbf{q}^{k+\frac{1}{2}} = \bar{\mathbf{q}}^k + \frac{1}{2} [\mathbf{p}^k - \nabla \mathbf{m}^{k+1}], \quad (25b)$$

$$\mathbf{p}^{k+1} = \arg \min_{\mathbf{p}_0 \in \mathcal{C}, (\mathbf{p}_1, \mathbf{p}_2)} \sum \|(\mathbf{p}_1 \ \mathbf{p}_2)\| + \frac{1}{2} \|\nabla \mathbf{m}^{k+1} - \mathbf{p} - \mathbf{q}^{k+\frac{1}{2}}\|_2^2, \quad (25c)$$

$$\bar{\mathbf{q}}^{k+1} = \bar{\mathbf{q}}^{k+\frac{1}{2}} + \frac{1}{2} [\mathbf{p}^{k+1} - \nabla \mathbf{m}^{k+1}]. \quad (25d)$$

Note that the weight $\frac{1}{2}$ in eqs (25b) and (25d) has a similar role as α in eqs (14b) and (14d). Now we come up with a linear inverse subproblem for \mathbf{m} , eq. (25a). Accordingly, the update \mathbf{m}^{k+1} is obtained by solving the following system of linear equations in a least-squares sense:

$$\begin{bmatrix} \lambda_1^{\frac{1}{2}} \mathbf{L}(\mathbf{u}^{k+1}) \\ \Gamma^{\frac{1}{2}} \nabla \end{bmatrix} \mathbf{m}^{k+1} = \begin{bmatrix} \lambda_1^{\frac{1}{2}} [\mathbf{b} + \mathbf{b}^k - \Delta \mathbf{u}^{k+1}] \\ \Gamma^{\frac{1}{2}} [\mathbf{p}^k + \bar{\mathbf{q}}^k] \end{bmatrix}, \quad (26)$$

where we have substituted \mathbf{y}^k by its explicit expression, eq. (19). In eq. (26), the first line describes the information carried out by the reconstructed wavefield to update \mathbf{m} via the wave-equation rewriting, while the second line describes the action of the TV regularization and bound constraints on \mathbf{m} via its linear relation with the auxiliary variable \mathbf{p} .

The closed-form expression of \mathbf{m} is given by

$$\mathbf{m}^{k+1} = [\lambda_1 \mathbf{L}(\mathbf{u}^{k+1})^T \mathbf{L}(\mathbf{u}^{k+1}) + \nabla^T \Gamma \nabla]^{-1} [\lambda_1 \mathbf{L}(\mathbf{u}^{k+1})^T [\mathbf{b} + \mathbf{b}^k - \Delta \mathbf{u}^{k+1}] + \nabla^T \Gamma [\mathbf{p}^k + \bar{\mathbf{q}}^k]]. \quad (27)$$

As for the linear system (16), \mathbf{m} can be computed numerically with any suitable sparse linear algebra method.

The subproblem for \mathbf{p} (eq. 25c) is straightforward to solve. The objective function is separable with respect to the variable \mathbf{p}_0 and the variables \mathbf{p}_1 and \mathbf{p}_2 (i.e. the optimization can be performed for \mathbf{p}_0 and $\mathbf{p}_1, \mathbf{p}_2$ separately). The variable \mathbf{p}_0 is solution of the following linear inverse problem

$$\mathbf{p}_0^{k+1} = \arg \min_{\mathbf{p}_0 \in \mathcal{C}} \frac{\gamma_0}{2} \|\mathbf{m}^{k+1} - \mathbf{p}_0 - \mathbf{q}_0^k\|_2^2. \quad (28)$$

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

$$\mathbf{p}_0^{k+1} = \text{proj}_{\mathcal{C}}(\mathbf{m}^{k+1} - \mathbf{q}_0^k), \quad (29)$$

where the projection operator is $\text{proj}_{\mathcal{C}}(\bullet) = \min(\max(\bullet, \mathbf{m}_l), \mathbf{m}_u)$.

The variables \mathbf{p}_1 and \mathbf{p}_2 are updated via the following proximity operator:

$$\mathbf{p}_1^{k+1}, \mathbf{p}_2^{k+1} = \arg \min_{(\mathbf{p}_1 \ \mathbf{p}_2)} \sum \|(\mathbf{p}_1 \ \mathbf{p}_2)\| + \frac{\gamma}{2} \|(\mathbf{p}_1 \ \mathbf{p}_2) - (\mathbf{z}_1 \ \mathbf{z}_2)\|_2^2, \quad (30)$$

where $\gamma = \gamma_1 = \gamma_2$, and $\mathbf{z}_i = \nabla_i \mathbf{m}^{k+1} - \mathbf{q}_i^k$ for $i = 1, 2$. Proximity operators are generalization of projection operators (Combettes & Pesquet 2011). Eq. (30) describes a separable optimization problem with respect to \mathbf{p}_1 and \mathbf{p}_2 . Furthermore, \mathbf{p}_1 and \mathbf{p}_2 have closed-form expressions (Goldstein & Osher 2009)

$$\mathbf{p}_i^{k+1} = \text{prox}_{\gamma}(\mathbf{z}_i) \quad \text{for } i = 1 \text{ and } 2, \quad (31)$$

where

$$\text{prox}_{\gamma}(\mathbf{z}_i) = \frac{\mathbf{z}_i}{\|(\mathbf{z}_1 \ \mathbf{z}_2)\|} \max(\|(\mathbf{z}_1 \ \mathbf{z}_2)\| - \gamma, 0). \quad (32)$$

It can be seen that, for a single vector, the proximity operator defined in eq. (32) reduces to soft thresholding.

Considering all the above-mentioned processes, a pseudo-code for the BTV-regularized IR-WRI algorithm is summarized in Algorithm 1. Note that the lines 6–11 of the algorithm correspond to one ADMM iteration of the model-parameter updating. These operations could be iterated in an inner loop to update the BTV-regularized model several times after the wavefield reconstruction at each outer iteration. However, we observed numerically that only a single iteration of the inner loop guarantees the most efficient convergence of the full algorithm. This property has been noticed by Goldstein & Osher (2009) and is discussed more extensively in the framework of IR-WRI by Aghamiry *et al.* (2019b). The inefficiency of the inner iterations can be understood by the fact that the original non-linear problem is solved with an alternating-direction strategy (managed by the outer loop). This implies that each subproblem is solved from potentially inaccurate passive variables, this inaccuracy preventing an efficient minimization of the objective during inner iterations. Furthermore, in order to drive the algorithm, we assumed that the constraint is feasible. However, it has been shown that in the case of infeasible linear constraints the ADMM iteration can still produce approximate solutions that are stable (Frick *et al.* 2011; Jiao *et al.* 2016).

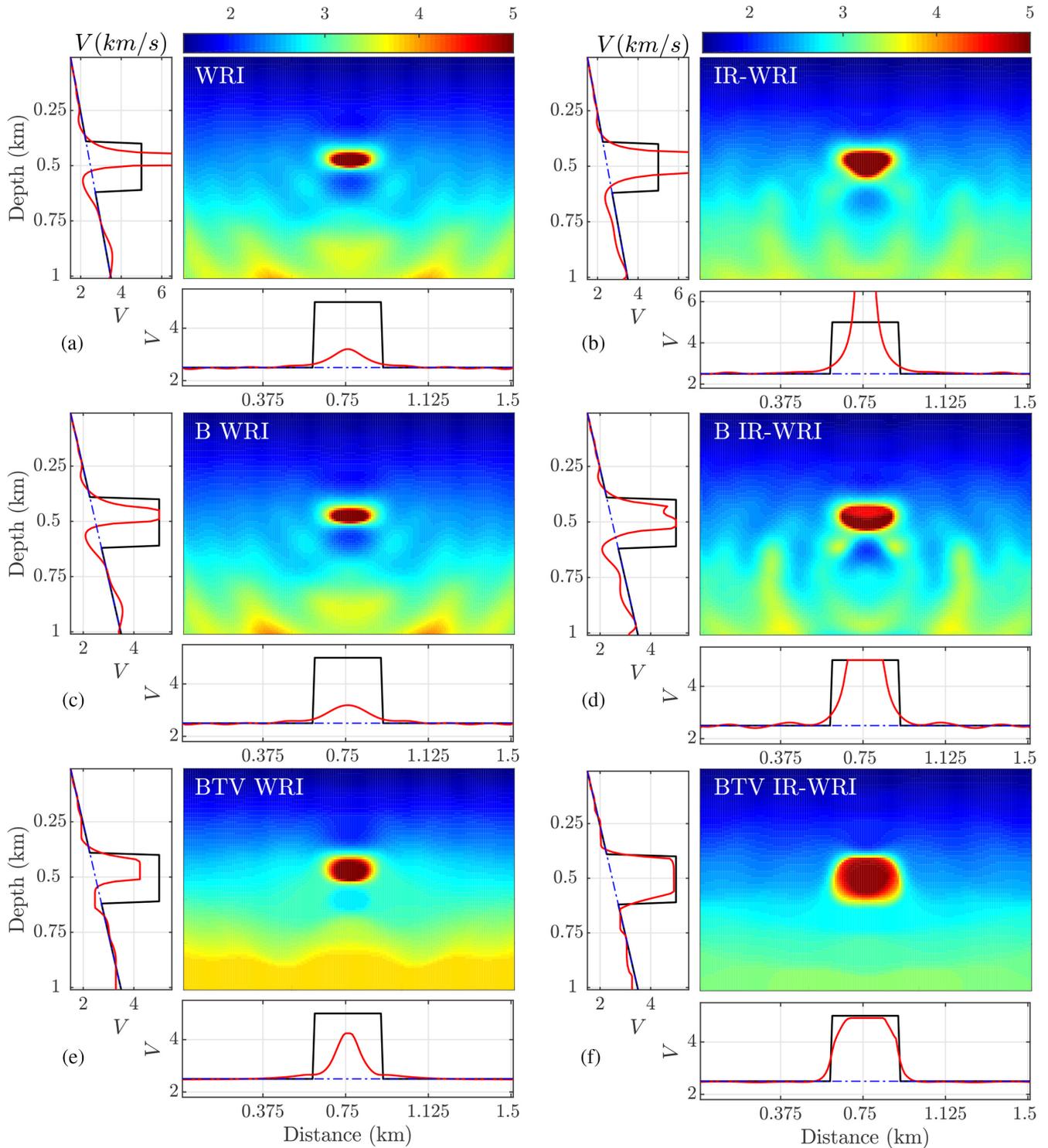


Figure 2. A box-shaped anomaly example with the true background velocity gradient as the initial model. Velocity models reconstructed by (a) WRI. (b) IR-WRI. (c and d) Bound-constrained WRI (c) and IR-WRI (d). (e and f) BTV-regularized WRI (e) and IR-WRI (f). Horizontal and vertical profiles across the centre of the inclusion from the true (black), initial (dash blue) and reconstructed (red) models are shown below and on the left-hand side of the models.

3 NUMERICAL EXAMPLES

3.1 Experimental setup and parameter tuning

We assess the performance of our BTV-regularized IR-WRI against 2-D monoparameter synthetic examples. We start with a toy example built with a high-velocity inclusion model that is embedded

in a background medium where velocity linearly increases with depth. To tackle more realistic applications, we proceed with two scaled targets of the challenging 2004 BP salt model (Billette & Brandsberg-Dahl 2004). With the BP salt case study, we seek to illustrate the potential of IR-WRI equipped with our BTV regularization to image salt bodies and subsalt structures starting from crude initial models and realistic frequencies.

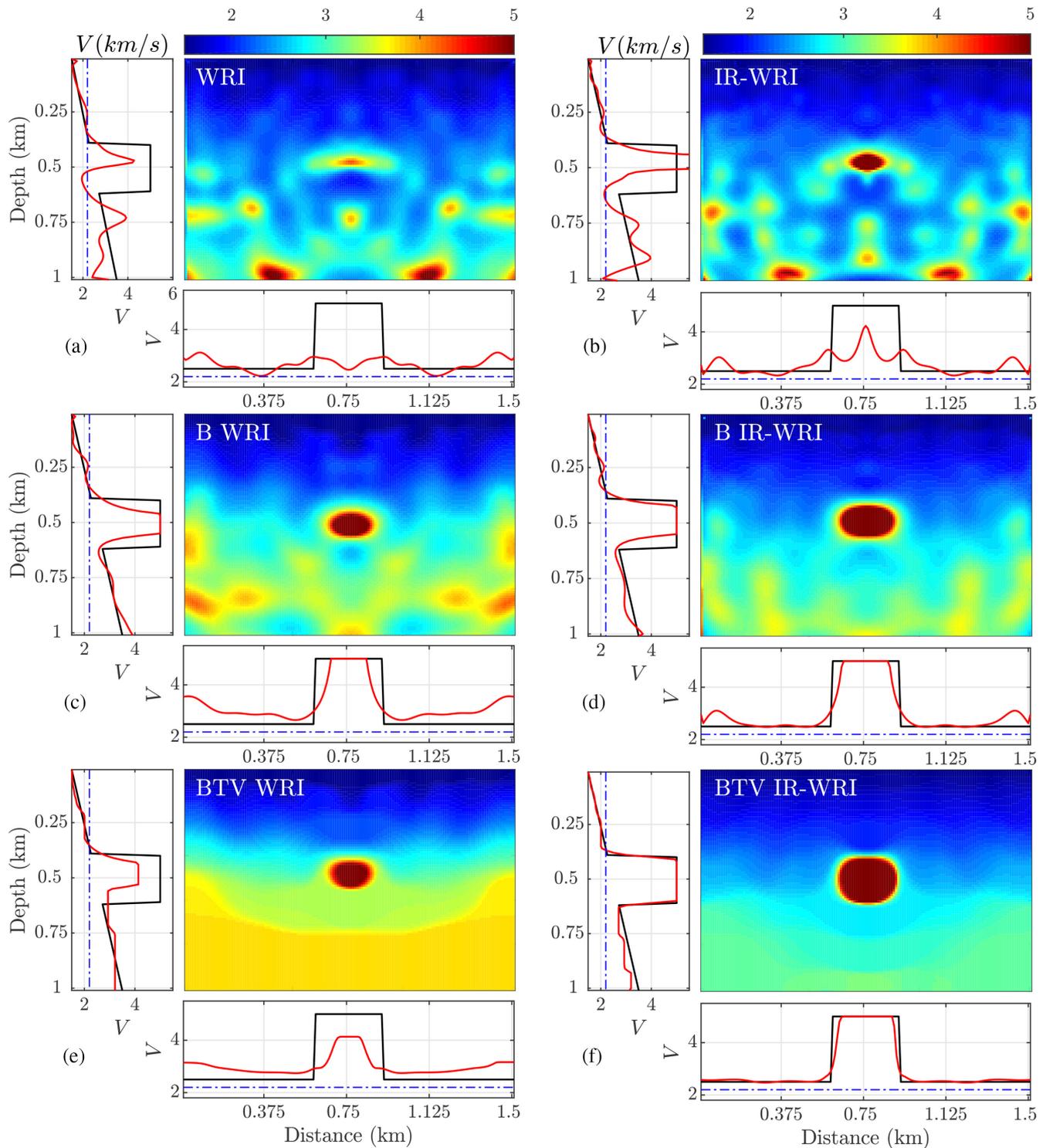


Figure 3. The box-shaped anomaly example. Same as Fig. 2 for a homogeneous initial velocity model ($V = 2.2 \text{ km s}^{-1}$).

For all the numerical examples, forward modelling is performed with a nine-point stencil implemented with antilumped mass and PML absorbing boundary conditions (Chen *et al.* 2013). In this setting, the diagonal matrix C contains the damping PML coefficients and does not depend on \mathbf{m} . With this setting, eq. (17) does not require any approximation for linearization.

We will compare the results of WRI and IR-WRI to highlight the improved convergence history of IR-WRI resulting from the

iterative updating of the right-hand sides in the penalty function associated with the scaled-form AL (eqs 11–14). We assume that our IR-WRI algorithm, when this right-hand side updating is not activated, is representative of the WRI penalty method (van Leeuwen & Herrmann 2013). For a fair comparison, we will use the same experimental setup (penalty parameters and stopping criterion of iteration) for the two methods. We also compare the WRI and IR-WRI results when they are obtained without any priors ($\gamma_0 = \gamma =$

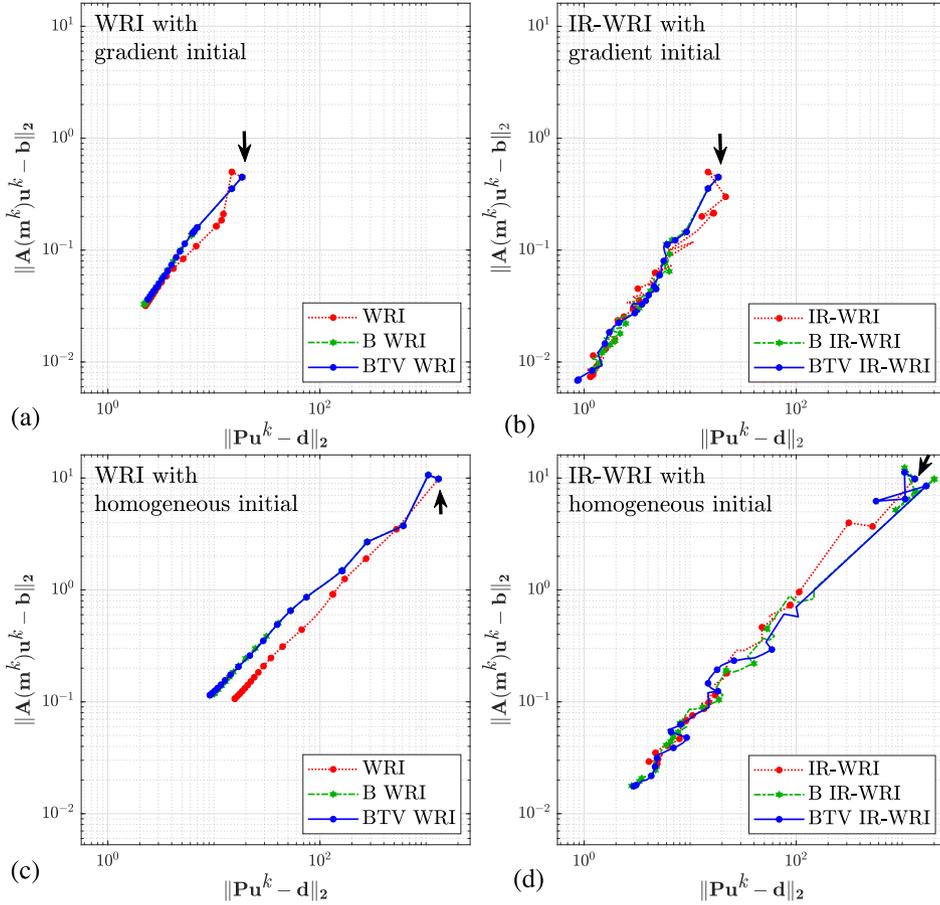


Figure 4. The box-shaped anomaly example. Convergence history in the $(\|\mathbf{P}\mathbf{u}^k - \mathbf{d}\|_2 - \|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b}\|_2)$ plane of WRI and IR-WRI without priors, with bound constraints and with BTV regularizations. (a) and (c) WRI. (b) and (d) IR-WRI. Initial model is (a) and (b) the true velocity-gradient background model, and (c) and (d) the homogeneous velocity model. All the panels are plotted with the same horizontal and vertical logarithmic scale. The black arrow points the starting point.

Algorithm 1 BTV regularized IR-WRI algorithm based on the PRS algorithm.

- 1: Initialize: set the RHS errors $k = 0$, $\mathbf{s}^0 = \mathbf{0}$, $\mathbf{q}^0 = \mathbf{0}$
- 2: Input: \mathbf{m}^0 (initial model parameters)
- 3: **while** convergence criteria not satisfied **do**
- 4: $\mathbf{u}^{k+1} \leftarrow$ update according to 2.16
- 5: $\mathbf{s}^{k+\frac{1}{2}} \leftarrow \mathbf{s}^k + 0.5[\mathbf{s} - \mathbf{F}(\mathbf{m}^k)\mathbf{u}^{k+1}]$
- 6: $\mathbf{m}^{k+1} \leftarrow$ update according to 2.26
- 7: $\mathbf{q}^{k+\frac{1}{2}} \leftarrow \mathbf{q}^k + 0.5[\mathbf{p}^k - \nabla \mathbf{m}^{k+1}]$
- 8: $\mathbf{p}_0^{k+1} \leftarrow \text{proj}_C(\mathbf{m}^{k+1} - \mathbf{q}_0^{k+\frac{1}{2}})$
- 9: $\mathbf{p}_1^{k+1} \leftarrow \text{prox}_\gamma(\nabla_1 \mathbf{m}^{k+1} - \mathbf{q}_1^{k+\frac{1}{2}})$
- 10: $\mathbf{p}_2^{k+1} \leftarrow \text{prox}_\gamma(\nabla_2 \mathbf{m}^{k+1} - \mathbf{q}_2^{k+\frac{1}{2}})$
- 11: $\mathbf{q}^{k+1} \leftarrow \mathbf{q}^{k+\frac{1}{2}} + 0.5[\mathbf{p}^{k+1} - \nabla \mathbf{m}^{k+1}]$
- 12: $\mathbf{s}^{k+1} \leftarrow \mathbf{s}^{k+\frac{1}{2}} + 0.5[\mathbf{s} - \mathbf{F}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}]$
- 13: $k \leftarrow k + 1$
- 14: **end while**

0), with only bound constraints ($\gamma = 0$) and with BTV regularization ($\gamma_0 \neq 0$ and $\gamma \neq 0$), where it is reminded that γ_0 and γ are the penalty parameters that control the weight of the bound constraints and TV regularization, respectively, in the objective function (see eq. 23).

We tune the different penalty parameters according to the following guideline. We start from the last subproblem of the splitting

procedure and set the parameter γ , which controls the soft thresholding performed by the TV regularization (eqs 31 and 32). In this study, we find that $\gamma = 0.02 \max \|\mathbf{z}_1 \ \mathbf{z}_2\|$ was a good pragmatical value. This tuning can be refined according to prior knowledge of the geological structure, coming from well logs for example. In this study, we use the same weight for the bound constraints and the TV regularization: $\gamma_0 = \gamma$. Once we set γ , we define λ_1 such that γ/λ_1 is a percentage of the mean absolute value of the diagonal coefficients of $\mathbf{L}^T \mathbf{L}$ during the parameter-estimation subproblem (eq. 27). 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, eq. (16), according to the criterion proposed by van Leeuwen & Herrmann (2016). In all the numerical tests, we use $\lambda = 1e-5\xi$ and $1e-3\xi$ 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 expands the search space accordingly. However, a too small value can lead to a prohibitively high number of iterations of the AL method before the wave-equation constraint is fulfilled with sufficient accuracy. Moreover, when data are contaminated by noise, a too small value for λ will make the wavefield reconstruction overfit the data and drive WRI to be a poor minimizer. We always use λ as a fixed percentage of ξ in iterations for both WRI and IR-WRI. This does not prevent IR-WRI to converge towards accurate minimizers thanks to the iterative error correction performed by the Lagrange multiplier updating. The reader is referred to Aghamiry *et al.* (2019b) for a more thorough sensitivity analysis of IR-WRI to the penalty parameter λ .

3.2 Inclusion model

The subsurface model contains a sharp box-shape anomaly of sides 0.2×0.3 km with a velocity (V) of 5 km s^{-1} . It is embedded in a smooth background model where V increases linearly with depth from 1.5 to 3.5 km s^{-1} (Fig. 1). The model is 1.5 km long and 1 km deep, and is discretized with a 10 m grid interval. The regular surface acquisition consists of five sources (as depicted with yellow stars in Fig. 1) and 65 receivers deployed on the surface. The source signature is a Ricker wavelet with a 5 Hz dominant frequency. We start the inversion from the true background model and invert simultaneously three frequency components (2.5 , 5 and 7 Hz) with noiseless data. This frequency bandwidth has been selected to cover a significant band of vertical wavenumbers in the waveform-inversion sensitivity kernels, considering the limited aperture illumination provided by the surface acquisition. Moreover, a realistic starting frequency of 2.5 Hz allows us to assess the resilience to cycle skipping of IR-WRI. A maximum number of iterations set to 70 is used as a stopping criterion for all of the tests shown in Fig. 2. When bound constraints are used, the bounds \mathbf{m}_l and \mathbf{m}_u , eq. (29), are set to the true minimum and maximum square slownesses, respectively.

We first compare WRI and IR-WRI results when bound constraints and TV regularization are not activated, i.e. $\gamma_0 = \gamma = 0$ (Figs 2a and b). WRI and IR-WRI reconstruct only the top of the anomaly with strongly overestimated velocities. Then, we add bound constraints in WRI and IR-WRI, using $\gamma_0/\lambda_1 = 0.01\zeta$, where ζ is the mean absolute value of the diagonal coefficients of $\mathbf{L}^T\mathbf{L}$. The bound-constrained WRI and IR-WRI only reconstruct the top of the anomaly as in Figs 2(a) and (b). However, the inclusion velocities are now well controlled by the bound constraints (Figs 2c and d). These first two tests show that IR-WRI reconstructs better the shape of the anomaly than WRI with however more significant artefacts on both sides of the anomaly. These artefacts may have resulted from the deficit in horizontal-wavenumber illumination provided by the sparse limited-offset surface acquisition and by multiscattering pollutions.

Then, we apply BTV regularization with $\gamma/\lambda_1 = \gamma_0/\lambda_1 = 0.01\zeta$ (Figs 2e and f). Since the initial model matches the true velocity-gradient background model, we use a small value of γ/λ_1 (i.e. a high value of λ_1) and keep it constant in iterations to preserve the smooth components of the subsurface model. We show that the BTV-regularized WRI still fails to reconstruct the full anomaly (Fig. 2e). In contrast, BTV-regularized IR-WRI keeps on improving the reconstruction of the anomaly in depth, while efficiently mitigating the oscillating artefacts (Fig. 2f).

A vertical profile across the reconstructed anomaly also highlights some limitations of the BTV regularization (Fig. 2f): below

the anomaly, the BTV regularization superimposes staircase artefacts on the velocity gradient, consistently with the piecewise constant assumption underlying TV regularization.

To emphasize the resilience to cycle skipping of the BTV-regularized WRI and IR-WRI, we repeat this toy example using a 2.2 km s^{-1} homogeneous velocity model as initial model (Fig. 3). We perform a first test without any prior. Compared to the previous test, we just stabilize the inversion by adding a small damping term ($=0.01\zeta$) to $\mathbf{L}^T\mathbf{L}$. Compared to Figs 2(a) and (b), the artefacts have a much stronger imprint due to the inaccuracy of the starting model (Figs 3a and b). Then, we move to bound-constrained and TV-regularized tests (Figs 3c–f). To decrease the above-mentioned artefacts, we assign a high initial weight to the BTV regularization ($\gamma/\lambda_1 = \gamma_0/\lambda_1 = \zeta$) and decrease it by a factor 2 every 10 iterations until it reaches a minimal value set to 0.01ζ (i.e. the constant value previously used). In accordance with the former test, bound constraints alone are not sufficient to reconstruct the bottom part of the anomaly and cancel out the oscillating artefacts (Figs 3c and d). In contrast, BTV IR-WRI achieves these two goals, although it leaves a significant staircase footprint below the anomaly (Fig. 3f). Compared to Fig. 2(f), the edges of the anomaly are better reconstructed at the expense of the background velocity-gradient model. This results, because of the more aggressive TV regularization used during the early iterations of this test, allowing for a better reconstruction of the blocky components of the medium, while injecting undesired staircase footprint on its smooth components. As for the former test, IR-WRI clearly outperforms WRI due to the more efficient solution refinement procedure resulting from the right-hand side updating.

We also show the joint evolution in iterations of the observation- and wave-equation errors (Fig. 4) and the wavefield and subsurface model errors (Fig. 5), when the initial model is the true velocity-gradient background model and the homogeneous model. IR-WRI fits the data and wave equation better than WRI after 70 iterations with both initial models because the right-hand side updating embedded in IR-WRI cancels out more efficiently the data and source residuals in iterations and, hence better refines the solution accordingly. Moreover, BTV regularization in IR-WRI further improves the data and wave-equation fit for both initial models because it reduces more efficiently the oscillating artefacts in the reconstructed velocity model and better reconstructs the edges of the anomaly (Fig. 2f). As above mentioned, the oscillating artefacts may result from the deficit of horizontal wavenumber illumination generated by the limited-offset surface acquisition. In this framework, the prior contained in the BTV regularization efficiently narrows the null space of the inversion. The more complex zigzag path followed by IR-WRI relative to WRI in the $(\|\mathbf{P}\mathbf{u}^k - \mathbf{d}\|_2 - \|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b}\|_2)$ plane highlights how the joint updating of the data and source by their associated residuals dynamically balances the weight of the two objective functions in iterations. This more complex convergence history of IR-WRI, which has been already noticed in Aghamiry *et al.* (2019b), suggests that the right-hand side updating perform a self-adaptive weighting of the two competing objective functions driven by the relative reduction of the data and source residuals in iterations. This zigzag convergence trend translates also into more complex path in the $(\|\mathbf{u}^k - \mathbf{u}^*\|_2/\|\mathbf{u}^*\|_2 - \|\mathbf{m}^k - \mathbf{m}^*\|_2/\|\mathbf{m}^*\|_2)$ plane (\mathbf{m}^* and \mathbf{u}^* denote the true model and wavefields, respectively), which illustrates how the solution refinement is pushed toward the wavefield reconstruction or the velocity model estimation according to the self-adaptive weighting of the observation- and wave-equation objective functions (Fig. 5). Note that the relative model errors increase in the

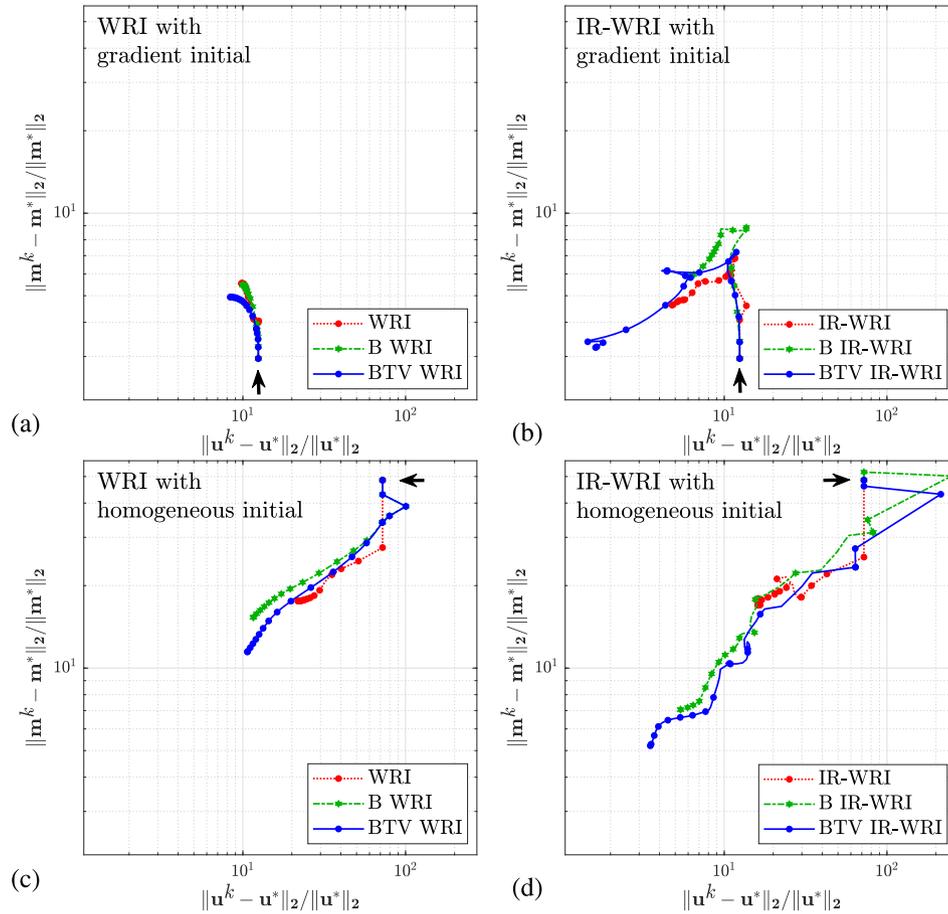


Figure 5. The box-shaped anomaly example. Convergence history in the $(\|\mathbf{u}^k - \mathbf{u}^*\|_2 / \|\mathbf{u}^*\|_2 - \|\mathbf{m}^k - \mathbf{m}^*\|_2 / \|\mathbf{m}^*\|_2)$ plane of WRI and IR-WRI without priors, with bound constraints and with BTV regularization and for the two initial models. The wavefield misfit function $\|\mathbf{u}^k - \mathbf{u}^*\|_2 / \|\mathbf{u}^*\|_2$ is computed by summation over all the shots. (a) and (c) WRI. (b) and (d) IR-WRI. The initial model is (a) and (b) the true velocity-gradient background model, and (c) and (d) the homogeneous velocity model. Note the increase of $\|\mathbf{m}^k - \mathbf{m}^*\|_2 / \|\mathbf{m}^*\|_2$ over iterations in (a) and (b) (see the text for explanations). All the panels are plotted with the same horizontal and vertical logarithmic scales. The black arrow indicates the starting point.

case of the initial velocity-gradient model (Figs 5a and b). This results from the fact that the smooth background model is degraded by the oscillating artefacts and the staircase footprint during the sharp inclusion reconstruction. Indeed, this degradation of the smooth components has a much higher weight in the ℓ_2 misfit function than the more accurate reconstruction of the blocky components. However, this increase is much more moderate in IR-WRI (Fig. 5b) than in WRI (Fig. 5a). Finally, we plot the TV norm of the reconstructed models in iterations for the different tests (Fig. 6). As expected, the models reconstructed with bound constraints and TV regularization match better the TV of the true model. When the homogeneous initial model is used, BTV IR-WRI smoothly converges to the TV of the true model after around iteration 15, while the weight of the TV regularization is progressively decreased (Fig. 6b, blue curve). The BTV WRI model matches slightly better the TV of the true model than the BTV IR-WRI one when the initial model is the true background model. Indeed, this does not reflect that WRI better reconstructs the anomaly than IR-WRI. Instead, it reflects the slower convergence of WRI relative to IR-WRI which contributes to keep the background model smooth (namely, which a TV close to that of the true background model).

3.3 2004 BP salt model—central target

We now consider a more realistic application with a first target of the challenging 2004 BP salt model. The 2004 BP salt model is representative of the geology of the deep water offshore Gulf of Mexico and mainly consists of a simple background with a complex rugose multivalued salt body, subsalt slow velocity anomalies related to overpressure zones and a fast velocity anomaly to the right of the salt body (Billette & Brandsberg-Dahl 2004). The first selected target corresponds to the central part of the 2004 BP salt model characterized by a deeply rooted salt body (Fig. 7a). Note that we rescale the spatial dimensions and the sampling of the original target for sake of computational efficiency.

Accordingly, our subsurface model is 8.8 km wide and 2.9 km deep, and is discretized with a 25 m grid interval. We used 50 sources spaced 175 m apart on the top side of the model. The source signature is a Ricker wavelet with a 10 Hz dominant frequency. A line of receivers with a 50 m spacing are deployed at the surface leading to a stationary-receiver acquisition.

We start the inversion from a smoothed version of true velocity model where the imprint of the salt body and other structures were cancelled out (Fig. 7b) and invert the 3-Hz frequency with noiseless data.

We compare the results of WRI and IR-WRI with bound constraints and BTV regularization. To highlight the specific role of

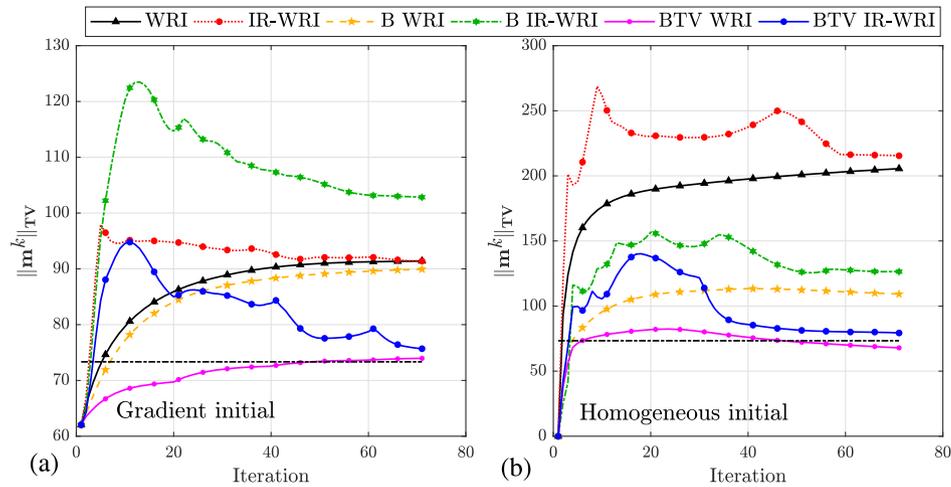


Figure 6. The box-shaped anomaly example. TV norm history over iterations of bound-constrained, BTV-regularized and ordinary WRI and IR-WRI for (a) velocity-gradient and (b) homogeneous initial model. The TV norm of true model is plotted with dashed-black line.

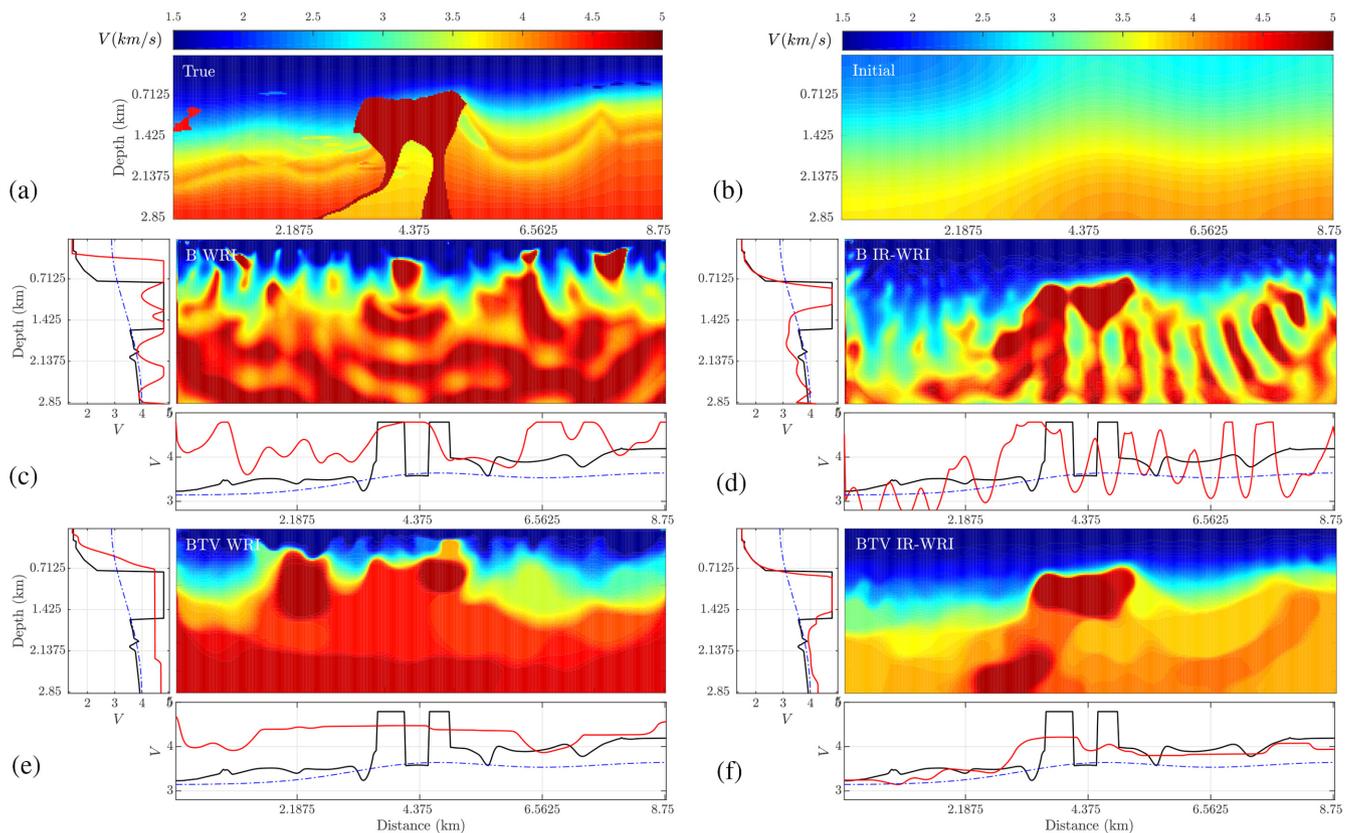


Figure 7. 2004 BP salt model—central target. 3-Hz frequency. (a) True model. (b) Starting model. (c) and (d) Bound-constrained WRI (c) and IR-WRI (d) models. (e) and (f) BTV-regularized WRI (e) and IR-WRI (f) models. Horizontal and vertical profiles at 4.15 km distance and 1.65 km depth from the true (black), initial (dash blue) and reconstructed (red) models are shown below and on the left-hand side of the models.

bound constraints, we activate them after 21 iterations. Since we start from a rough initial model, we set $\gamma_0/\lambda_1 = \gamma/\lambda_1 = \zeta$ and decrease them during iterations in a manner similar to the box-shape anomaly test with a homogeneous starting model. Also, we add a damping term to $\mathbf{L}^T\mathbf{L}$ with a weight equal to 0.01ζ to further stabilize the inversion. We stop inversion after 70 iterations. The estimated models are shown in Fig. 7 together with horizontal profiles at 1.65 km depth and vertical profiles at 4.15 km distance extracted from the true, initial and reconstructed models.

As for the inclusion test, WRI fails to reconstruct the salt body and the subsalt structure because the data and source residuals are not re-injected in the right-hand sides of the penalty function at each iteration as in eq. (13), leading to a stagnant convergence of the inversion (Figs 7c and e). When IR-WRI is applied with bound constraints alone, the reconstructed model is affected by noise with a periodic horizontal pattern. This noise likely results from the monochromatic nature of the inversion, multiscattering within the salt body and limited illumination of the horizontal

wavenumbers of the salt body leading to wraparound (Fig. 7d). The BTV-regularized IR-WRI mitigates efficiently this noise without degrading the resolution of the salt body and the subsalt structures (Fig. 7f).

Fig. 8(a) shows the joint evolution in iterations of the data misfit and the wave-equation error. As for the inclusion test, note the zigzag path followed by the IR-WRI objective functions over iterations. Also, the joint evolution of wavefield and subsurface model errors and the evolution of TV norm over iteration are shown in Figs 8(b) and (c). The TV norm evolution emphasizes how the bound constraints fasten the convergence of TV-regularized IR-WRI after iteration 20.

We continue the inversion at higher frequencies using the final models of the 3 Hz inversion as initial models (Figs 7c–f). We used small batches of two frequencies with one frequency overlapping between two consecutive batches, moving from the low frequencies to the higher ones according to a classical frequency continuation strategy. We set $\gamma_0/\lambda_1 = \gamma/\lambda_1 = 0.01\zeta$ and remove the damping of $\mathbf{L}^T\mathbf{L}$. The starting and final frequencies are 3.5 and 12 Hz and the sampling interval in one batch is 0.5 Hz. The algorithm performs at most 15 iterations per frequency batch and the number of iterations that have been performed is 170. The inversion results are shown in Fig. 9. WRI with bound constraints and BTV (Figs 9a and c) fails to converge toward satisfactory results, while BTV-regularized IR-WRI converges to accurate velocity model, although a significant imprint of the TV regularization is shown (Fig. 9d). When IR-WRI is performed with only bound constraints, the oscillating artefacts are not cancelled out (Fig. 9b). This highlights the role of TV regularization in reconstructing blocky structures and removing wraparound artefacts.

3.4 2004 BP salt model—left target

We now consider a second target of the 2004 BP salt model located on the left-hand side of the model (Fig. 10a). This target was previously used by among others Métivier *et al.* (2016), Brandsberg-Dahl *et al.* (2017) and Esser *et al.* (2018) for FWI applications. After rescaling of the original model with the ratio used by Métivier *et al.* (2016), our subsurface target is 16.25 km wide and 5.825 km deep, and is discretized with a 25 m grid interval. We used 108 sources spaced 150 m apart on the top side of the model. The source signature is a Ricker wavelet with a 10 Hz dominant frequency. A line of receivers with a 25 m spacing are deployed at the surface leading to a stationary-receiver acquisition. We perform IR-WRI with bound constraints alone and with BTV regularization, for noiseless and noisy data.

We used a crude laterally homogeneous velocity-gradient model as initial model (Fig. 10b). We used small batches of two frequencies with one frequency overlap between two consecutive batches, moving from the low frequencies to the higher ones according to a classical frequency continuation strategy. The starting and final frequencies are 3 and 13 Hz and the sampling interval in one batch is 0.5 Hz. The stopping criterion of iteration for each batch is given by $k_{max} = 15$ or

$$\|\mathbf{A}(\mathbf{m}^{k+1})\mathbf{u}^{k+1} - \mathbf{b}\|_2 \leq \varepsilon_b \quad \text{and} \quad \|\mathbf{P}\mathbf{u}^{k+1} - \mathbf{d}\|_2 \leq \varepsilon_d, \quad (33)$$

where k_{max} denotes the maximum iteration count, $\varepsilon_b = 1e-3$, and $\varepsilon_d = 1e-5$ for noiseless data and $\varepsilon_b = 1e-3$, $\varepsilon_d =$ noise level of batch for noisy data. We perform three paths through the frequency batches to improve the IR-WRI results, using the final model of one path as the initial model of the next one (these cycles can be

viewed as outer iterations of IR-WRI). The starting frequency of the second and third paths is 6 and 8.5 Hz, respectively. The IR-WRI models inferred from noiseless data with bound constraints and with BTV regularization are shown in Figs 10(c) and (d). The number of iterations that have been performed with bound constraints and with BTV regularization are 441 and 340, respectively. Direct comparison between the true model, the starting model and the IR-WRI models along three vertical logs cross-cutting the salt body at 5, 7.5 and 10 km distance (vertical dashed lines in Fig. 10a) are shown in Fig. 11(a). The results show the resilience of IR-WRI to cycle skipping with a pretty accurate reconstruction of the salt body and subsalt structures (Figs 10c and d). However, the model obtained with bound constraints alone shows high-frequency noise in the salt body and below (Fig. 10c). This noise can result from Gibbs phenomenon caused by the frequency decimation and artefacts resulting from multiscattering (Alkhalifah *et al.* 2018). The BTV regularization efficiently removes these artefacts except those resulting from truncation of the acquisition near the left end of the model at 5-km depth (Fig. 10d). The inversion captures reasonably well the subsalt structures, including the low-velocity overpressure zone at $(x,z) = (7.5 \text{ km}, 4 \text{ km})$ as well as the smooth velocity variations.

When noisy data are used (Figs 10e and f, and 11b), the number of iterations that have been performed with bound constraints alone and with BTV regularization is 263 and 254, respectively. As for the noiseless case, a direct comparison between the true model, the starting model and the two IR-WRI models along three vertical logs cross-cutting the salt body at 5, 7.5 and 10 km distance is shown in Fig. 11(b). The artefacts have now a more significant imprint when only bound constraints are used (Fig. 10e). Accordingly, we apply a more aggressive TV regularization to obtain the results shown in Fig. 10(f). The artefacts have been efficiently removed with however a more obvious imprint of the piecewise-constant approximation underlying BTV regularization. This blocky pattern is clearly visible in deep part of the vertical profiles of Fig. 11(b) where velocity gradients have been replaced by stack of constant-velocity layers.

4 DISCUSSION

We have implemented bound constraints and TV regularization in the WRI method (van Leeuwen & Herrmann 2013) which has been recently improved by Aghamiry *et al.* (2019b) in the framework of the ADMM, leading to the IR-WRI.

There are some key differences between our implementation of BTV-regularized WRI and previous ones based upon projected-gradient method (Peters & Herrmann 2017) and primal dual hybrid gradient (PDHG) methods (Esser *et al.* 2018; Yong *et al.* 2018) that we review below. In the projected-gradient method, Peters & Herrmann (2017) first update model parameters with FWI or WRI and then project the updated model into the intersection of the TV and bound constraint with Dykstra projection algorithm (Boyle & Dykstra 1986), an alternating projection onto some constraints until satisfying all of them. Therefore, their workflow is subdivided in two different parts, that are the model update followed by the projection onto intersection of all constraints (Peters & Herrmann 2017, their eq. 7). Unlike Peters & Herrmann (2017) method, which relies on independent update and projection steps, all the ingredients of BTV-regularized IR-WRI (i.e. wavefield reconstruction, parameter estimation, TV regularization and bound constraints) are

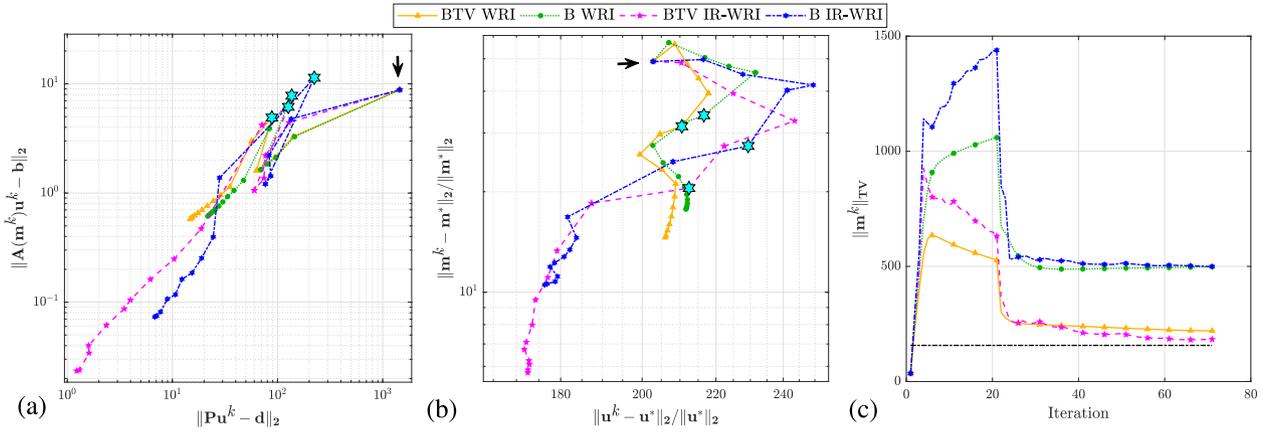


Figure 8. 2004 BP salt model—central target. 3-Hz frequency. Convergence history (a) in the $(\|Pu^k - d\|_2 - \|A(m^k)u^k - b\|_2)$ plane (b) in the $(\|u^k - u^*\|_2 / \|u^*\|_2 - \|m^k - m^*\|_2 / \|m^*\|_2)$ plane. (c) Evaluation of TV norm $\|m^k\|_{TV}$ over iterations. Note again the more complex convergence history of IR-WRI compared to WRI due to the self-adaptive weighting of the data-fitting and wave-equation objective functions performed by right-hand side updating. The iteration 22 are located with cyan stars in (a) and (b).

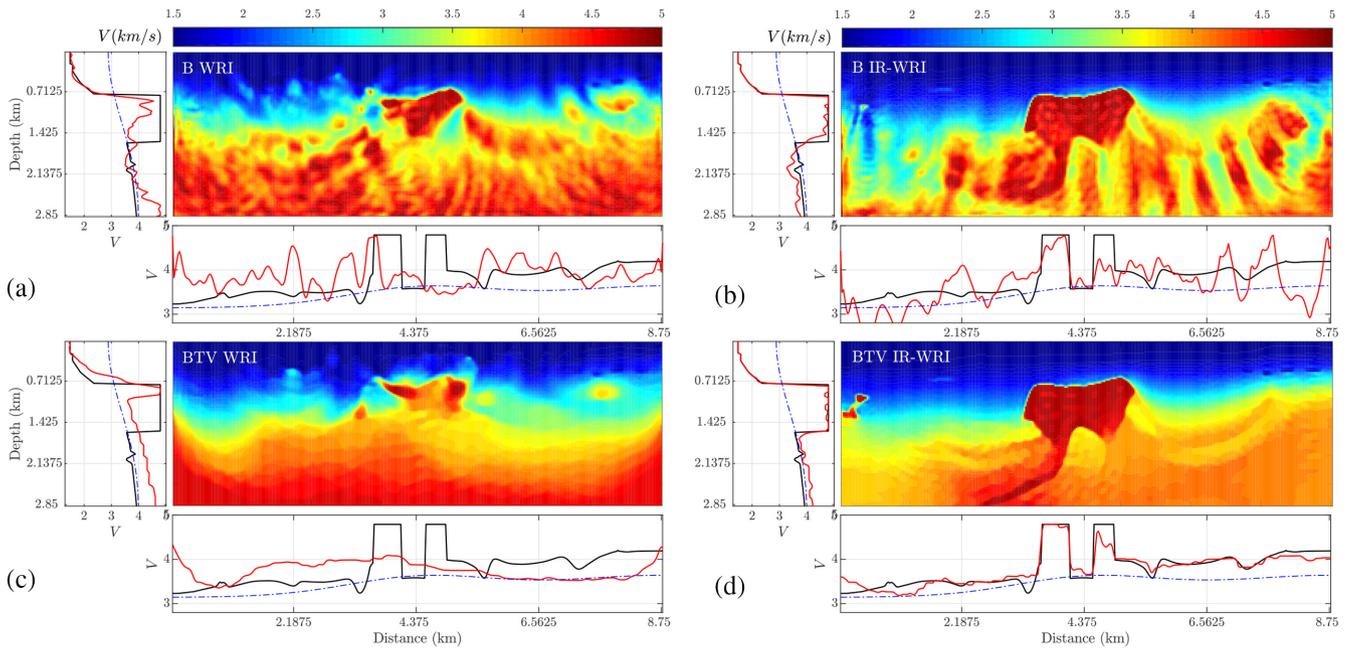


Figure 9. 2004 BP salt case study. Central target. (a) Final bound-constrained WRI with Fig. 7(c) as initial model. (b) Final bound-constrained IR-WRI with Fig. 7(d) as initial model. (c) Final BTV-regularized WRI with Fig. 7(e) as initial model. (d) Final BTV-regularized IR-WRI with Fig. 7(f) as initial model.

consistently integrated in the theoretical framework of ADMM optimization (the readers can also refer to Maharramov & Levin 2015).

Esser *et al.* (2018) and Yong *et al.* (2018) implemented TV regularization and bound constraints in the reduced variable projection WRI of van Leeuwen & Herrmann (2016) with PDHG. PDHG is a method to solve constrained optimization problems by alternating gradient descent (for primal variable) and gradient ascent (for dual variable), which can be interpreted as linearized ADMM (Goldstein *et al.* 2015). PDHG can be helpful if the least-squares minimizations embedded in ADMM are difficult to solve efficiently (Goldstein *et al.* 2015). This is not really the case in WRI, which mainly requires to solve two sparse linear systems for u and m . Meanwhile, the selection of step size in PDHG that guarantees fast convergence, or even convergence at all, is not intuitive at all and can make PDHG impractical. This issue prompted Goldstein *et al.* (2015) to develop step size tuning rules which contribute to make

PDHG self-adaptive as illustrated recently by Yong *et al.* (2018) in the frame of WRI. Beyond PDHG method, Esser *et al.* (2018) and Yong *et al.* (2018) minimize data and source residuals with a penalty method, which lacks the convergence property of the AL method promoted in IR-WRI. Moreover, they implement TV regularization as a hard constraint in the parameter-estimation subproblem through a Lagrangian function, whose saddle point is estimated with PDHG. In contrast, we implement the TV regularization as a soft constraint after introducing the auxiliary variable p and solve the regularized subproblem for m with the split-Bregman method (or equivalently ADMM), eq. (22). This gives us the necessary flexibility to implement aggressive regularization during early iterations and relax it progressively, when very crude initial models are used. With more accurate initial models, the AL embedded in the split-Bregman method (eq. 22), allows for constant penalty parameter to be used. These penalty parameters are used as step lengths in the AL method,

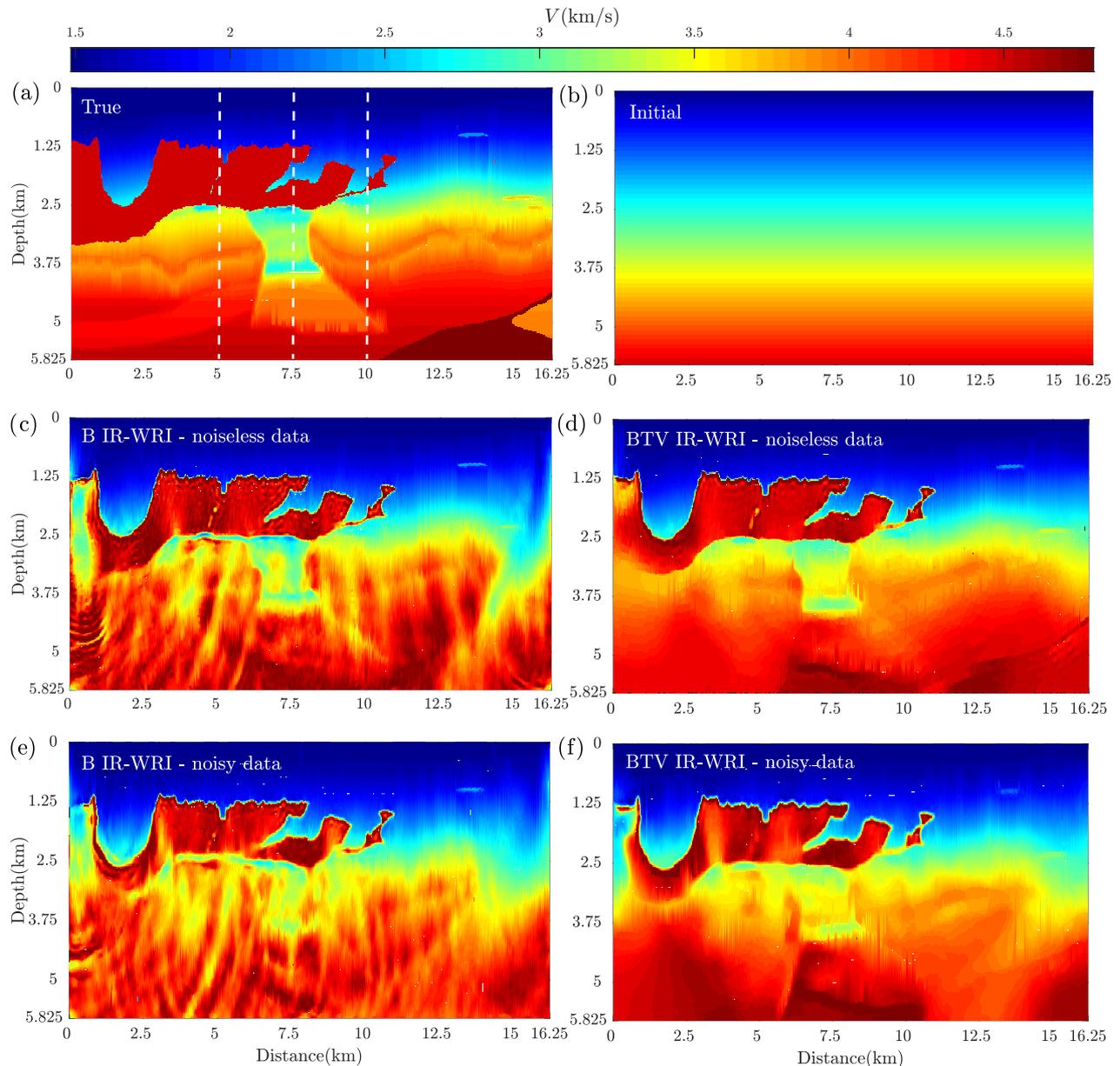


Figure 10. The 2004 BP salt case study. Left target. (a) True BP model. The vertical dashed lines indicate the location of vertical logs of Fig. 11. (b) Initial velocity model. (c) and (d) Final IR-WRI velocity models obtained with bound constraints (c) and with BTV regularization (d) for noiseless data. (e) and (f) Same as (c) and (d) for noisy data for an S/N of 10 db.

hence leading to a self-adaptive TV regularization implementation rid of tedious TV-norm ball continuation strategies and/or adaptive step lengths (Esser *et al.* 2018; Yong *et al.* 2018). Third, while Esser *et al.* (2018, eq. 18) and Yong *et al.* (2018, eq. 41) implement bound constraints as hard constraints by projection at each iteration of the TV-regularized perturbation model onto the feasible set defined by the bound constraints, we implement bound constraints consistently with TV regularization in the framework of the method of multiplier (eqs 26–30). Finally, our approach does not require prior guess of TV-norm ball because we implement TV regularization as a minimization problem rather than as a constraint. That being said, Esser *et al.* (2018) show promising results with the τ continuation approach and asymmetric TV norm on the BP salt model starting from a very crude initial model. It will be interesting in

future work to assess the benefit of asymmetric TV norm in BTV IR-WRI.

Indeed, implementation of BTV-regularized IR-WRI requires to setup different parameters, which have been discussed at the beginning of the section *Experimental setup and parameter tuning*. The penalty parameter which controls the relative weight between the wave-equation objective and the data fitting objective during the wavefield-reconstruction subproblem has been discussed in length in Aghamiry *et al.* (2019b). They have concluded that a fixed penalty parameter can be used during iterations because the accuracy of the minimizer in the method of multiplier is controlled both by the penalty parameter and the accuracy of the updated multiplier (Nocedal & Wright 2006, theorem 17.6). In this study, we suggest good

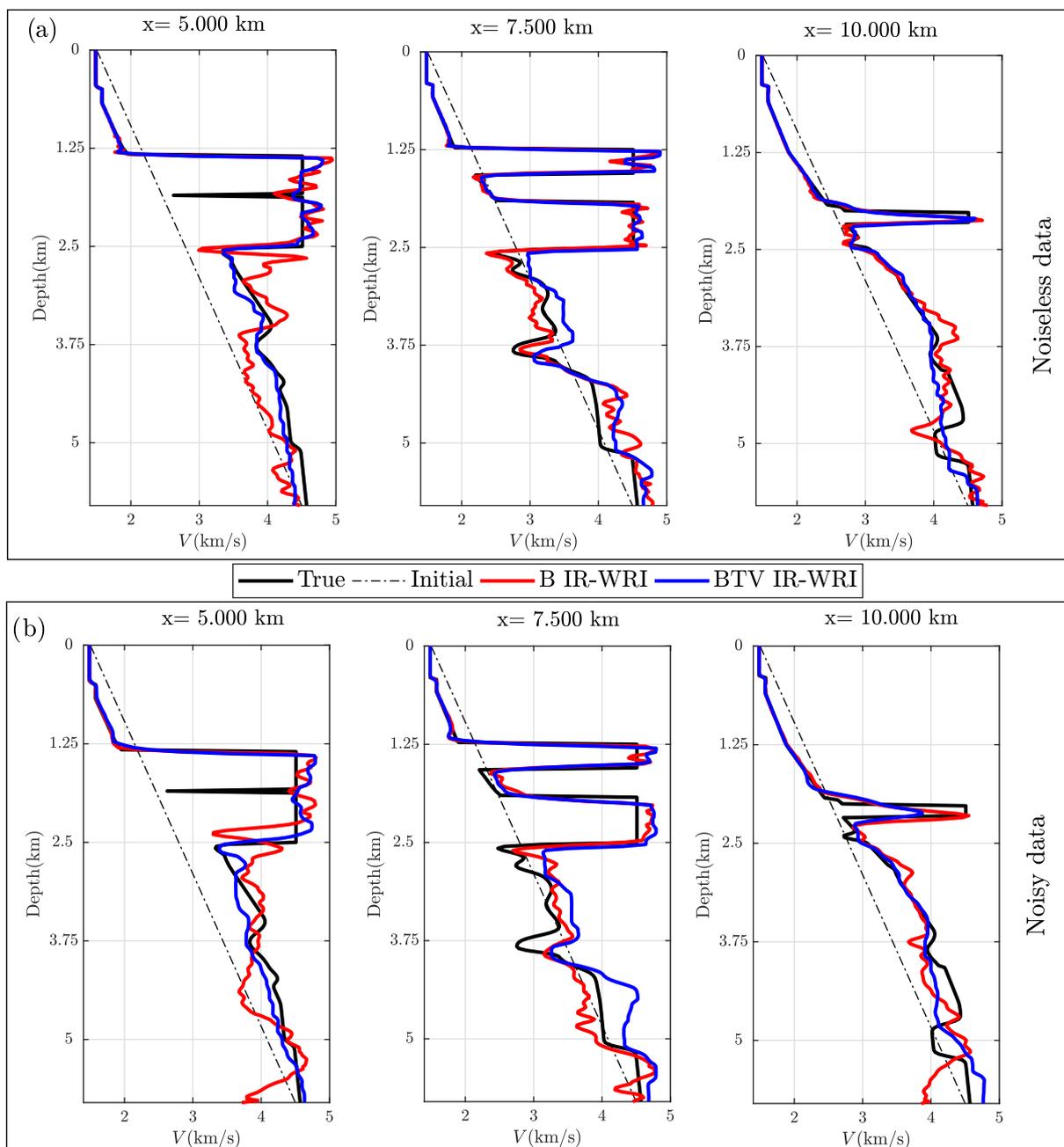


Figure 11. 2004 BP salt case study—left target. Direct comparison between the true velocity model (black), the initial model (dashed line) and the IR-WRI models obtained with bound constraints (red) and with BTV regularization (blue) along three logs at $x = 4.5, 7$ and 10 km (vertical dashed lines in Fig. 10a) from left to right. (a) Noiseless data. (b) Noisy data.

pragmatical values of this penalty parameter as a percentage of the maximum eigenvalue of the augmented wave-equation normal operator for noiseless and noisy data. Optimal values may be refined by trial and error to prevent noisy data overfitting and keep the iteration count within reasonable limits. The TV and bound parameters should be easily determined from well logs or *a priori* geological knowledge. Finally, the relative weight between the TV regularization and the wave-equation constraint, which is controlled by λ_1 , needs also to be estimated during the parameter-estimation subproblem. As above mentioned, this penalty parameter is kept fixed during

iterations or is progressively decreased to relax the TV regularization and bound constraints near the convergence point, depending of the accuracy of the initial model. The reader is also referred to Goldstein & Osher (2009, section 2.2) who discuss the sensitivity of the split-Bregman method (an optimization method similar to ADMM, Esser 2009) to the penalty parameter for ℓ_1 -regularized problems.

One drawback of the BTV regularization is related to the piecewise constant approximation underlying TV regularization, which tends to superimpose some blocky patterns on the smooth part of the

subsurface. To overcome this issue, ongoing work seeks to optimally combine Tikhonov and TV regularization in IR-WRI (Gholami & Hosseini 2013; Aghamiry *et al.* 2018a, 2019a).

Other perspective developments involve extension to multiparameter reconstruction, 3-D geometries and application to real data to further assess the potential and limits of IR-WRI. For 3-D applications, the wavefield reconstruction in the frequency domain requires to solve a large-scale linear algebra problem (eq. 16). Operto *et al.* (2015), Amestoy *et al.* (2016) and Operto & Miniussi (2018) have shown the computational efficiency of 3-D frequency-domain FWI based on sparse direct solvers in the 3.5–10 Hz frequency band for dense stationary-recording ocean-bottom cable acquisitions. Mary (2017) showed that block low-rank multifrontal solver allows one to tackle numerical problems involving up to 100 million unknowns. The symmetry of the normal operator, eq. (16), should balance the computational overhead resulting from the higher number of non-zero coefficients relative to the impedance matrix \mathbf{A} . Note that \mathbf{A} was processed as an unsymmetric matrix in the above-mentioned references due to discretization issues, although more suitable

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

5 CONCLUSION

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

ACKNOWLEDGEMENTS

We would like to thank the Editor Hervé Chauris, Tariq Alkhalifah and one anonymous reviewer for their comments. This study was partially funded by the SEISCOPE consortium (<http://seiscope2.osug.fr>), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXON-MOBIL, JGI, PETROBRAS, SCHLUMBERGER, SHELL, SINOPEC and TOTAL. This study was granted access to the HPC resources of SIGAMM infrastructure (<http://crimson.oca.eu>), hosted by Observatoire de la Côte d'Azur and which is supported by the Provence-Alpes Côte d'Azur region, and the HPC resources of CINES/IDRIS/TGCC under the allocation 0596 made by GENCI.

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, Society of Exploration Geophysicists.
- 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)*, EAGE.
- Aghamiry, H., Gholami, A. & Operto, S., 2019a. Compound regularization of full-waveform inversion for imaging piecewise media, preprint ([arXiv:1903.04405](https://arxiv.org/abs/1903.04405)).
- Aghamiry, H., Gholami, A. & Operto, S., 2019b. Improving full-waveform inversion by wavefield reconstruction with alternating direction method of multipliers, *Geophysics*, **84**(1), R139–R162.
- Alkhalifah, T., Sun, B.B. & Wu, Z., 2018. Full model wavenumber inversion: identifying sources of information for the elusive middle model wavenumbers, *Geophysics*, **83**(6), R597–R610.
- Amestoy, P., Brossier, R., Buttari, A., L'Excellent, J.-Y., Mary, T., Métivier, L., Miniussi, A. & Operto, S., 2016. Fast 3D frequency-domain full waveform inversion with a parallel Block Low-Rank multifrontal direct solver: application to OBC data from the North Sea, *Geophysics*, **81**(6), R363–R383.
- Anagaw, A. & Sacchi, M., 2011. Regularized 2D acoustic full waveform inversion, *73rd EAGE Conference and Exhibition Incorporation SPE EUROPEC*, EAGE, doi:10.3997/2214-4609.20149386.
- Aravkin, A.Y., Drusvyatskiy, D. & van Leeuwen, T., 2017. Efficient quadratic penalization through the partial minimization technique, *IEEE Trans. Autom. Control*, **63**, 2131–2138.
- Askan, A., Akcelik, V., Bielak, J. & Ghattas, O., 2007. Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures, *Bull. seism. Soc. Am.*, **97**(6), 1990–2008.
- Benning, M. & Burger, M., 2018. Modern regularization methods for inverse problems, *Acta Numer.*, **27**, 1–111.
- Bérenger, J.-P., 1994. A perfectly matched layer for absorption of electromagnetic waves, *J. Comput. Phys.*, **114**, 185–200.
- Bertsekas, D.P., 2016. *Nonlinear Programming*, Thirdedn, Athena Scientific.
- Billette, F.J. & Brandsberg-Dahl, S., 2004. The 2004 BP velocity benchmark, in *Extended Abstracts, 67th Annual EAGE Conference and Exhibition, Madrid, Spain*, B035, EAGE.
- Boyd, S.P. & Vandenberghe, L., 2004. *Convex Optimization*, Cambridge Univ. Press.
- Boyd, S., Parikh, N., Chu, E., Peleato, B. & Eckstein, J., 2010. Distributed optimization and statistical learning via the alternating direction of multipliers, *Found. Trends Mach. Learn.*, **3**(1), 1–122.
- Boyle, J.P. & Dykstra, R.L., 1986. A method for finding projections onto the intersection of convex sets in Hilbert spaces, in *Advances in Order Restricted Statistical Inference: Proceedings of the Symposium on Order Restricted Statistical Inference held in Iowa City, Iowa, September 11–13, 1985*, pp. 28–47, eds Dykstra, R., Robertson, T. & Wright, F.T., Springer.
- Brandsberg-Dahl, S., Chemingui, N., Valenciano, A., Ramos-Martinez, J. & Qiu, L., 2017. FWI for model updates in large-contrast media, *Leading Edge, Special section: Full Waveform Inversion Part II*(1), pp. 81–87.
- Chen, Z., Cheng, D., Feng, W. & Wu, T., 2013. An optimal 9-point finite difference scheme for the Helmholtz equation with PML, *Int. J. Numer. Anal. Model.*, **10**(2), 389–410.
- 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**, Springer Optimization and Its Applications, 185–212, eds Bauschke, H.H., Burachik, R.S., Combettes, P.L., Elser, V., Luke, D.R. & Wolkowicz, H., Springer.
- Dolean, V., Jolivet, P. & Nataf, F., 2015. *An Introduction to Domain Decomposition Methods—Algorithms, Theory, and Parallel Implementation*, SIAM.
- Engquist, B. & Majda, A., 1977. Absorbing boundary conditions of the numerical simulation of waves, *Math. Comput.*, **31**, 629–651.
- Epanomeritakis, I., Akcelik, V., Ghattas, O. & Bielak, J., 2008. A Newton-CG method for large-scale three-dimensional elastic full waveform seismic inversion, *Inverse Probl.*, **24**, pp. 1–26.
- Esser, E., 2009. Applications of lagrangian-based alternating direction methods and connections to split Bregman, Tech. Rep., UCLA.
- Esser, E., Guasch, L., van Leeuwen, T., Aravkin, A.Y. & Herrmann, F.J., 2018. Total variation regularization strategies in full-waveform inversion, *SIAM J. Imag. Sci.*, **11**(1), 376–406.
- Frick, K., Lorenz, D.A. & Resmerita, E., 2011. Morozov's principle for the augmented lagrangian method applied to linear inverse problems, *Multiscale Model. Simul.*, **9**(4), 1528–1548.
- Gholami, A., 2015. Nonlinear multichannel impedance inversion by total-variation regularization, *Geophysics*, **80**(5), R217–R224.
- Gholami, A., 2016. A fast automatic multichannel blind seismic inversion for high-resolution impedance recovery, *Geophysics*, **81**(1), 1–8.
- Gholami, A. & Hosseini, S.M., 2013. A balanced combination of Tikhonov and total variation regularizations for reconstruction of piecewise-smooth signals, *Signal Process.*, **93**, 1945–1960.
- Gholami, A. & Sacchi, M.D., 2013. Fast 3D blind seismic deconvolution via constrained total variation and GCV, *SIAM J. Imag. Sci.*, **6**(4), 2350–2369.
- Gholami, A. & Siahkoobi, H., 2010. Regularization of linear and non-linear geophysical ill-posed problems with joint sparsity constraints, *Geophys. J. Int.*, **180**(2), 871–882.
- Gholami, A., Aghamiry, H. & Abbasi, M., 2018. Constrained nonlinear AVO inversion using Zoeppritz equations, *Geophysics*, **83**(3), R245–R255.
- 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 J. Imag. Sci.*, **2**(2), 323–343.
- Goldstein, T., Li, M. & Yuan, X., 2015. Adaptive primal-dual splitting methods for statistical learning and image processing, in *Advances in Neural Information Processing Systems*, pp. 2089–2097, The MIT Press.
- Guittou, A., 2012. Blocky regularization schemes for full waveform inversion, *Geophys. Prospect.*, **60**(5), 870–884.
- Haber, E., Ascher, U.M. & Oldenburg, D., 2000. On optimization techniques for solving nonlinear inverse problems, *Inverse Probl.*, **16**(5), 1263.
- He, B., Liu, H., Wang, Z. & Yuan, X., 2014. A strictly contractive Peaceman–Rachford splitting method for convex programming, *SIAM J. Optim.*, **24**(3), 1011–1040.
- Hestenes, M.R., 1969. Multiplier and gradient methods, *J. Optim. Theor. Appl.*, **4**(5), 303–320.
- Hustedt, B., Operto, S. & Virieux, J., 2004. Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling, *Geophys. J. Int.*, **157**, 1269–1296.
- Jiao, Y., Jin, Q., Lu, X. & Wang, W., 2016. Alternating direction method of multipliers for linear inverse problems, *SIAM J. Numer. Anal.*, **54**(4), 2114–2137.
- Jo, C.H., Shin, C. & Suh, J.H., 1996. An optimal 9-point, finite-difference, frequency-space 2D scalar extrapolator, *Geophysics*, **61**, 529–537.
- Loris, I. & Verhoeven, C., 2012. Iterative algorithms for total variation-like reconstructions in seismic tomography, *Int. J. Geomath.*, **3**(2), 179–208.
- Maharramov, M. & Biondi, B., 2015. Robust simultaneous time-lapse full-waveform inversion with total-variation regularization of model difference, in *77th EAGE Conference and Exhibition, Expanded Abstracts, We P3 09*, EAGE.
- Maharramov, M. & Levin, S.A., 2015. Total-variation minimization with bound constraints, preprint ([arXiv:1505.05694](https://arxiv.org/abs/1505.05694)).
- Marfurt, K., 1984. Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations, *Geophysics*, **49**, 533–549.
- Mary, T., 2017. Block Low-Rank multifrontal solvers: complexity, performance and scalability, *PhD thesis*, Université de Toulouse.
- Métivier, L., Brossier, R., Méritot, Q., Oudet, E. & Virieux, J., 2016. Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion, *Geophys. J. Int.*, **205**, 345–377.
- Nocedal, J. & Wright, S.J., 2006. *Numerical Optimization*, 2ndedn. Springer.

- 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, *Geophys. J. Int.*, **213**, 2037–2059.
- Operto, S., Miniussi, A., Brossier, R., Combe, L., Métivier, L., Monteiller, V., Ribodetti, A. & Virieux, J., 2015. Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: application to Valhall in the visco-acoustic vertical transverse isotropic approximation, *Geophys. J. Int.*, **202**(2), 1362–1391.
- Peaceman, D.W. & Rachford, H.H., Jr, 1955. The numerical solution of parabolic and elliptic differential equations, *J. Soc. Indus. Appl. Math.*, **3**(1), 28–41.
- Peters, B. & Herrmann, F.J., 2017. Constraints versus penalties for edge-preserving full-waveform inversion, *Leading Edge*, **36**(1), 94–100.
- Plessix, R.E., 2007. A Helmholtz iterative solver for 3D seismic-imaging problems, *Geophysics*, **72**(5), SM185–SM194.
- Pratt, R.G. & Smithyman, B., 2018. Full waveform inversion for density: reciprocity, divergence, parameter resolution, and other issues, in *Expanded Abstracts, 80th Annual EAGE Meeting (Copenhagen)*, EAGE.
- Pratt, R.G., Shin, C. & Hicks, G.J., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophys. J. Int.*, **133**, 341–362.
- Rudin, L., Osher, S. & Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms, *Phys. D*, **60**, 259–268.
- Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**(8), 1259–1266.
- van Leeuwen, T. & Herrmann, F., 2016. A penalty method for PDE-constrained optimization in inverse problems, *Inverse Probl.*, **32**(1), 1–26.
- van Leeuwen, T. & Herrmann, F.J., 2013. Mitigating local minima in full-waveform inversion by expanding the search space, *Geophys. J. Int.*, **195**(1), 661–667.

- Virieux, J. & Operto, S., 2009. An overview of full waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26.
- Yong, P., Liao, W., Huang, J. & Li, Z., 2018. Total variation regularization for seismic waveform inversion using an adaptive primal dual hybrid gradient method, *Inverse Probl.*, **34**(4), 045006, doi.

APPENDIX A: Scaled form of augmented Lagrangian

Let us start with the following constrained problem

$$\min_{\mathbf{x}} \quad \|P(\mathbf{x})\|_2^2 \quad \text{subject to} \quad Q(\mathbf{x}) = \mathbf{0}. \quad (\text{A1a})$$

The AL function for the problem is (Nocedal & Wright 2006)

$$\mathcal{L}_A(\mathbf{x}, \mathbf{v}) = \|P(\mathbf{x})\|_2^2 + \mathbf{v}^T Q(\mathbf{x}) + \frac{\lambda}{2} \|Q(\mathbf{x})\|_2^2. \quad (\text{A2})$$

The problem (A2) can be written in a more compact form by introducing the scaled dual variable $\bar{\mathbf{q}} = -\frac{\mathbf{v}}{\lambda}$ and adding and subtracting the term $\frac{\lambda}{2} \|\bar{\mathbf{q}}\|_2^2$ to the AL (A2). In this case, we arrive at the following scaled form of the method of multipliers:

$$\begin{aligned} \mathcal{L}_A(\mathbf{x}, \bar{\mathbf{q}}) &= \|P(\mathbf{x})\|_2^2 - \lambda \bar{\mathbf{q}}^T Q(\mathbf{x}) + \frac{\lambda}{2} \|Q(\mathbf{x})\|_2^2 + \frac{\lambda}{2} \|\bar{\mathbf{q}}\|_2^2 - \frac{\lambda}{2} \|\bar{\mathbf{q}}\|_2^2 \\ &= \|P(\mathbf{x})\|_2^2 + \frac{\lambda}{2} \|Q(\mathbf{x}) - \bar{\mathbf{q}}\|_2^2 - \frac{\lambda}{2} \|\bar{\mathbf{q}}\|_2^2. \end{aligned} \quad (\text{A3})$$