Compound Regularization of Full-Waveform Inversion for Imaging Piecewise Media

Hossein Aghamiry, Ali Gholami, Stéphane Operto

Compound Regularization of Full-waveform Inversion for Imaging Piecewise Media

Hossein S. Aghamiry, Ali Gholami, Stéphane Operto,

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

I. Introduction

FULL waveform inversion (FWI) seeks to estimate constitutive parameters by nonlinear minimization of a distance between recorded and simulated wavefield measurements. This technology was originally developed in geophysical imaging [1], and has spread more recently into other fields of imaging sciences such as medical imaging [2] and oceanography [3]. This partial-differential equation (PDE)-constrained nonlinear inverse problem is classically solved with local reduced-space optimization methods [4]. In this linearized framework, a challenging source of non linearity is the so-called cycle skipping pathology which occurs when the initial model does not allow to match the data with a kinematic error smaller than half a period [5], [6]. Other sources of error are noise, approximate wave physics and ill-posedness resulting from parameter cross-talk, coarse acquisition sampling and uneven illumination of the targeted structure. Designing regularization techniques that mitigate these source of errors is therefore a key challenge for the success of FWI applications.

A proper regularization should be driven by the shape and statistical characteristics of the medium to be imaged. For example, in geophysical imaging, the subsurface can be represented by a piece-wise smooth medium, that is a model which contains smoothly varying and blocky components. The widespread Tikhonov regularizations [7] rely on the smoothness assumption and hence fail to recover sharp interfaces of such media. Conversely, first order total variation (TV) regularizations [8] are based on a blockiness assumption and hence are more suitable to image large contrasts. However, they generate undesirable staircase imprints in smooth regions [13]. Hereafter we refer TV to as first order TV regularization, unless we explicitly mention the order of the TV regularization. Regions characterized by smoothly-varying properties and those containing sharp contrasts have different statistical properties. The former are characterized by the normal prior, while the latter by a heavy tailed prior [9]. Consequently, simultaneous recovery of both properties is difficult when one type of regularization is used (Tikhonov, TV, etc). To overcome this issue, a combination of different regularizations can be used [10], [11], [12]. A naive approach consists of the simple additive coupling or convex combinations (CC) of regularizations. Alternatively, [13] proposed to explicitly decompose the model into several components of different statistical properties and use an appropriate regularization to reconstruct each component. Using this strategy, they combined Tikhonov and TV regularizations (referred to as TT regularization) to reconstruct piece-wise smooth media. The smooth components are captured by the Tikhonov regularization, while the blocky ones are determined by the TV counterpart. In many applications, it has been shown that a compound regularization based upon infimal convolution (IC) outperforms the one based upon additive coupling [14].

TT regularization based upon IC has been successfully applied to FWI for seismic subsurface imaging in the framework of iteratively-refined wavefield reconstruction inversion (IR-WRI) [15]. IR-WRI extends the search space of FWI and decreases cycle skipping through a relaxation of the wave-equation constraint [16], [17], [18]. Taking advantage of the bilinearity of the wave equation, IR-WRI breaks down FWI into two linear subproblems which are solved in an alternating mode: wavefield reconstruction driven by the observables and model-parameter estimation by minimization of the source

Hossein S. Aghamiry is with the Institute of Geophysics, University of Tehran, Tehran, Iran, and Université Côte d'Azur, CNRS, Observatoire de la Côte d'Azur, IRD , Géoazur, Valbonne, France, e-mail: h.aghamiry@ut.ac.ir, aghamiry@geoazur.unice.fr.

Ali Gholami is with the Institute of Geophysics, University of Tehran, Tehran, Iran, e-mail: agholami@ut.ac.ir.

Stéphane Operto is with Université Côte d'Azur, CNRS, Observatoire de la Côte d'Azur, IRD , Géoazur, Valbonne, France, e-mail: operto@geoazur.unice.fr.

residuals the relaxation generated. Furthermore, [19] extended the method to acoustic multiparameter inversion. The linearity of the parameter-estimation subproblem provides a suitable framework to implement sophisticated nonsmooth regularizations.

In this study, following [13] and [15], we develop a general framework to combine a couple of regularization terms in IR-WRI through IC. Then, we specifically develop this framework for Tikhonov and TV regularizations, which are suitable for seismic subsurface imaging applications. Compared to [15], we jointly update the blocky and the smooth components through a variable projection process rather than in an alternating mode. We first show that our new IC-based TT regularization outperforms the CC-based counterpart with several well-documented numerical benchmarks in the field of seismic imaging. We also compare the results obtained with these two TT regularizations with those obtained with total generalized variation (TGV) regularization, a combination of first and second order TV, and those obtained with Tikhonov and TV regularizations when used independently.

II. Notation

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

III. Method

In this section, we briefly review the frequency-domain FWI as a bi-convex feasibility problem and describe the extended forms of FWI. We show how the problem can be solved with the alternating direction method of multipliers (ADMM) [20] for a general regularization function.

A. Full-waveform inversion

Frequency-domain FWI with a general regularization term and bounding constraints can be formulated as [17], [21]

$$\begin{aligned} \min_{\mathbf{u},\mathbf{m}\in\mathcal{C}} \quad & \Phi(\mathbf{m}) \\ \text{subject to} \quad & \mathbf{A}(\mathbf{m})\mathbf{u} = \mathbf{b}, \\ & \mathbf{P}\mathbf{u} = \mathbf{d}, \end{aligned} \tag{1}$$

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

The first constraint in (1), $\mathbf{A}(\mathbf{m})\mathbf{u} = \mathbf{b}$, is a partial-differential equation (PDE) wherein $\mathbf{u} \in \mathbb{C}^{n\times 1}$ is the wavefield and $\mathbf{b} \in \mathbb{C}^{n\times 1}$ is the source term. In this study, $\mathbf{A}(\mathbf{m}) \in \mathbb{C}^{n\times n}$ is the discretized PDE Helmholtz operator [22], [23] given by

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

with ω the angular frequency and $\mathbf{\Delta}$ the discretized Laplace operator. The diagonal matrix $\mathbf{C}$ embeds boundary conditions and can be dependent or independent on $\mathbf{m}$ depending on the kinds of absorbing boundary conditions (radiation versus sponge) [17]. Also, $\mathbf{B}$ is used to spread the *mass* term $\omega^2\mathbf{C}(\mathbf{m})\text{diag}(\mathbf{m})$ over all the coefficients of the stencil to improve its accuracy following an anti-lumped mass strategy [24], [25], [26].

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

1) WRI approach to solving (1)*:* The extended approach, known as wavefield reconstruction inversion (WRI) [16], recasts the constrained optimization problem, equation (1), as an unconstrained problem where both constraints are implemented with quadratic penalty functions.

$$\min_{\mathbf{u},\mathbf{m}\in\mathcal{C}} \ \Phi(\mathbf{m}) + \frac{\lambda_0}{2}\|\mathbf{P}\mathbf{u} - \mathbf{d}\|_2^2 + \frac{\lambda_1}{2}\|\mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{b}\|_2^2, \tag{3}$$

where $\lambda_0, \lambda_1 > 0$ are the penalty parameters. For the unregularized case where $\Phi(\mathbf{m}) = 0$ and also without the bounding constraint, [16] solved this biconvex minimization problem with an alternating-direction algorithm, whereby the joint minimization over $\mathbf{u}$ and $\mathbf{m}$ is replaced by an alternating minimization over each variable separately. The main property of the penalty formulation given by equation (3) is that the PDE constraint in the original problem is replaced by a quadratic penalty term, which enlarges the search space and mitigates the inversion nonlinearity accordingly [16]. Its main drawback, however, is the difficulty related to the adaptive tuning of the penalty parameter, which is common to all penalty methods [27].

2) IR-WRI approach to solving (1)*:* To overcome the above limitation, the iteratively-refined WRI (IR-WRI) implements the original constrained problem (1) with the augmented Lagrangian (AL) method [27], [28].

$$\begin{aligned} \min_{\mathbf{u},\mathbf{m}\in\mathcal{C}} \max_{\mathbf{v}_0,\mathbf{v}_1} \ \Phi(\mathbf{m}) + \frac{\lambda_0}{2}\|\mathbf{P}\mathbf{u} - \mathbf{d}\|_2^2 + \frac{\lambda_1}{2}\|\mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{b}\|_2^2 \\ + \mathbf{v}_0^T[\mathbf{P}\mathbf{u} - \mathbf{d}] + \mathbf{v}_1^T[\mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{b}], \end{aligned} \tag{4}$$

where $\mathbf{v}_0 \in \mathbb{C}^{m\times 1}$ and $\mathbf{v}_1 \in \mathbb{C}^{n\times 1}$ are the dual variables (the Lagrangian multipliers). The min-max problem (4) can also be written in a more compact form (the scaled form AL) as

$$\begin{aligned} \min_{\mathbf{u},\mathbf{m}\in\mathcal{C}} \max_{\mathbf{v}_0,\mathbf{v}_1} \ \Phi(\mathbf{m}) + \frac{\lambda_0}{2}\|\mathbf{P}\mathbf{u} - \mathbf{d} + \frac{1}{\lambda_0}\mathbf{v}_0\|_2^2 - \frac{\lambda_0}{2}\|\frac{\mathbf{v}_0}{\lambda_0}\|_2^2 \\ + \frac{\lambda_1}{2}\|\mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{b} + \frac{1}{\lambda_1}\mathbf{v}_1\|_2^2 - \frac{\lambda_1}{2}\|\frac{\mathbf{v}_1}{\lambda_1}\|_2^2. \end{aligned} \tag{5}$$

Applying a gradient ascent to (5) with respect to the duals, after a simple change of variables $\mathbf{d}^k = -\mathbf{v}_0^k/\lambda_0$ and $\mathbf{b}^k =$

$-\mathbf{v}_1^k/\lambda_1$, gives the following iteration:

$$\begin{aligned}\min_{\mathbf{u},\mathbf{m}\in\mathcal{C}}\ \ &\Phi(\mathbf{m})+\frac{\lambda_0}{2}\|\mathbf{P}\mathbf{u}-\mathbf{d}-\mathbf{d}^k\|_2^2\\ &+\frac{\lambda_1}{2}\|\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{b}-\mathbf{b}^k\|_2^2, \qquad (6)\\ \mathbf{d}^{k+1}&=\mathbf{d}^k+\mathbf{d}-\mathbf{P}\mathbf{u},\\ \mathbf{b}^{k+1}&=\mathbf{b}^k+\mathbf{b}-\mathbf{A}(\mathbf{m})\mathbf{u},\end{aligned}$$

beginning with $\mathbf{d}^0=0$ and $\mathbf{b}^0=0$. Capitalizing on the bilinearity of the wave equation in $\mathbf{m}$ and $\mathbf{u}$, ADMM [20] is a powerful method to solve this kind of multivariate optimization problem. ADMM updates $\mathbf{m}$ and $\mathbf{u}$ separately through a Gauss-Seidel like iteration, i.e., fixing $\mathbf{m}$ and solving for $\mathbf{u}$ and vice versa. Accordingly, beginning with an initial guess $\mathbf{m}^0$, we end up with the following iteration to solve (6) [17], [21]:

$$\mathbf{u}^{k+1}=\arg\min_{\mathbf{u}}\ \left\|\begin{bmatrix}\sqrt{\frac{\lambda_0}{\lambda_1}}\mathbf{P}\\ \mathbf{A}(\mathbf{m}^k)\end{bmatrix}\mathbf{u}-\begin{bmatrix}\sqrt{\frac{\lambda_0}{\lambda_1}}(\mathbf{d}+\mathbf{d}^k)\\ \mathbf{b}+\mathbf{b}^k\end{bmatrix}\right\|_2^2 \qquad (7a)$$

$$\mathbf{m}^{k+1}=\arg\min_{\mathbf{m}\in\mathcal{C}}\ \Phi(\mathbf{m})+\frac{\lambda_1}{2}\|\mathbf{A}(\mathbf{m})\mathbf{u}^{k+1}-\mathbf{b}-\mathbf{b}^k\|_2^2, \qquad (7b)$$

$$\mathbf{d}^{k+1}=\mathbf{d}^k+\mathbf{d}-\mathbf{P}\mathbf{u}^{k+1}, \qquad (7c)$$

$$\mathbf{b}^{k+1}=\mathbf{b}^k+\mathbf{b}-\mathbf{A}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}. \qquad (7d)$$

The subproblem (7a) associated with the wavefield reconstruction is quadratic and admits a closed-form solution. It relaxes the requirement to satisfy exactly the wave equation ($\mathbf{A}(\mathbf{m}^k)\mathbf{u}=\mathbf{b}$) for the benefit of improved data fitting ($\mathbf{P}\mathbf{u}=\mathbf{d}$). This is achieved by reconstructing the wavefields that best jointly fit the observations and satisfy the wave equation in a least-squares sense. While wavefields generated by the reduced approach (Appendix A) satisfy exactly the wave equation, $\mathbf{u}_r=\mathbf{A}(\mathbf{m}^k)^{-1}\mathbf{b}$, this makes classical FWI highly non-convex.

Equation (6) shows that the duals are updated with the running sum of the data and source residuals in iterations and are used to update the right-hand sides in the penalty functions of the scaled AL. These error correction terms in the AL method are the key ingredients that allow for a constant penalty parameter to be used in iterations, while guaranteeing convergence to accurate minimizer [27]. In the next section we focus on the solution of the model subproblem (7b) when compound regularizations are used as the regularization term.

IV. The model Subproblem

This section presents compound regularization functionals and the details of our approach to solve the model subproblem (7b) with these functionals.

A. Compound regularizers

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

1) Convex combination of simple regularizers: A CC of r simple regularizer functionals $\Phi_1,...,\Phi_r$ is a compound regularizer functional of the form

$$\Phi_\alpha(\mathbf{x})=\alpha_1\Phi_1(\mathbf{x})+...+\alpha_r\Phi_r(\mathbf{x}), \qquad (8)$$

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

$$\alpha_1+\alpha_2,...,+\alpha_r=1. \qquad (9)$$

Definitely, if all of the functions $\Phi_1,...,\Phi_r$ are convex then Φ is so. In CC models, the regularized solution is forced to satisfy the individual priors simultaneously. As an example, a compound regularizer functional constructed by a CC of ℓ_1- and squared ℓ_2-norms ($\ell_1+\ell_2$), known as an elastic net [29], [30], is

$$\Phi_\alpha(\mathbf{x})=(1-\alpha)\|\mathbf{x}\|_2^2+\alpha\|\mathbf{x}\|_1, \qquad (10)$$

with $0\leq\alpha\leq 1$. The convexity of ℓ_1- and ℓ_2-norms implies that $\Phi_\alpha(\mathbf{x})$ in (10) is convex. One may also construct a compound regularizer functional by a CC of two ℓ_1-norms which are applied in different domains, such as those spanned by two different wavelet transforms, or those spanned by a wavelet transform and the gradient operator [10].

2) Infimal convolution of simple regularizers: In IC models, the solution is decomposed into simple components and then each component is regularized by an appropriate prior. Accordingly, the IC of r simple regularizer functionals $\Phi_1,...,\Phi_r$ is a compound functional of the form

$$\Phi_\alpha(\mathbf{x})=\min_{\mathbf{x}=\mathbf{x}_1+...+\mathbf{x}_r}\{\alpha_1\Phi_1(\mathbf{x}_1)+...+\alpha_r\Phi_r(\mathbf{x}_r)\}. \qquad (11)$$

In the case of two functionals, Φ_α in (11) takes the form

$$\Phi_\alpha(\mathbf{x})=\min_{\mathbf{z}}\{(1-\alpha)\Phi_1(\mathbf{x}-\mathbf{z})+\alpha\Phi_2(\mathbf{z})\}, \qquad (12)$$

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

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

$$\Phi_\alpha(\mathbf{x})=\min_{\mathbf{z}}\{(1-\alpha)\|\mathbf{x}-\mathbf{z}\|_2^2+\alpha\|\mathbf{z}\|_1\}, \qquad (13)$$

which is a denoising problem whose solution is unique and is given by the well-known soft-threshold function [31]:

$$\mathbf{z}=\max\left(1-\frac{\alpha}{2(1-\alpha)|\mathbf{x}|},0\right)\circ\mathbf{x}, \qquad (14)$$

Putting $\mathbf{z}$ from (14) into (13) gives that

$$\Phi_\alpha(\mathbf{x})=\begin{cases}(1-\alpha)|\mathbf{x}|^2 & \text{if } |\mathbf{x}|\leq\frac{\alpha}{2(1-\alpha)}\\ \alpha|\mathbf{x}|-\frac{\alpha^2}{4(1-\alpha)} & \text{if } |\mathbf{x}|>\frac{\alpha}{2(1-\alpha)}\end{cases}, \qquad (15)$$

which is nothing other than the Huber function [32]. As seen, this function has a hybrid behavior: it has a quadratic behavior for small values of $|\mathbf{x}|$ and linear behavior for large values. The parameter $\frac{\alpha}{2(1-\alpha)}$ determines where the transition from quadratic to linear behavior takes place.

Geometrical illustrations of the ℓ_1-norm, ℓ_2-norm, $(\ell_1+\ell_2)$-norm, and $(\ell_1\square\ell_2)$-norm for $\alpha = 0.7$ are shown in Fig. 1. This figure shows that the ℓ_1- and ℓ_2-norms have a uniform behavior for all values, while the CC norm (the $(\ell_1 + \ell_2)$-norm) has a hybrid behavior: it approaches the ℓ_1-norm near zero, where it behaves as a linear function, but approaches the ℓ_2-norm for large values, where it behaves as a quadratic function. Unlike $\ell_1 + \ell_2$, the IC function $\ell_1\square\ell_2$ approaches the ℓ_2-norm near zero but is linear and approaches the ℓ_1-norm for large values.

In this paper, we consider (12) in the following settings, though other configurations are possible:

$$\Phi_\alpha^{\text{TT}}(\mathbf{x}) = \min_{\mathbf{x}=\mathbf{x}_1+\mathbf{x}_2} (1-\alpha)\|\nabla^2\mathbf{x}_2\|_2^2 + \alpha\|\nabla\mathbf{x}_1\|_1, \quad (16)$$

and

$$\Phi_\alpha^{\text{TGV}}(\mathbf{x}) = \min_{\mathbf{x}=\mathbf{x}_1+\mathbf{x}_2} (1-\alpha)\|\nabla^2\mathbf{x}_2\|_1 + \alpha\|\nabla\mathbf{x}_1\|_1, \quad (17)$$

where, in both (16) and (17), the norms are applied on the absolute valued components of $\nabla^2\mathbf{x}_2$ ((39) and (41)) and $\nabla\mathbf{x}_1$ (37). The compound regularizer Φ_α^{TT} is a combination of the second order Tikhonov and TV (TT) regularizations [13] and Φ_α^{TGV} is a combination of the first and second order TV regularizations, called total generalized variation (TGV) [33], [34]. The former is suitable for recovering piecewise-smooth models, while the latter is better suited for piecewise linear models. The next section gives a solution procedure to solve (7b) with these regularizers.

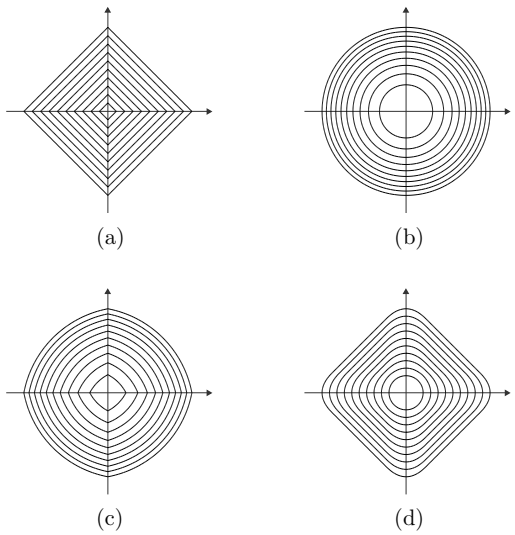

Fig. 1: Geometrical illustration of different regularizers. (a) the ℓ_1-norm, (b) the ℓ_2-norm, (c) the $(\ell_1 + \ell_2)$-norm, and (d) the $(\ell_1\square\ell_2)$-norm.

B. Solving the subproblem (7b)

In this section we present how to solve the subproblem (7b) with TT regularization. The solution procedure for the TGV regularizer follows easily. From the definition of $\mathbf{A}$ in (2), we get that

$$\mathbf{A}(\mathbf{m})\mathbf{u} \quad = \quad \boldsymbol{\Delta}\mathbf{u} + \mathbf{L}\mathbf{m}, \quad (18)$$

where

$$\mathbf{L} = \frac{\partial\mathbf{A}(\mathbf{m})}{\partial\mathbf{m}}\mathbf{u} = \omega^2\mathbf{C}\text{diag}(\mathbf{B}\mathbf{u}), \quad (19)$$

and we assume that $\mathbf{C}$ does not depend on $\mathbf{m}$ (this is the case for perfectly-matched absorbing boundary conditions [17]). From the explicit decomposition $\mathbf{m} = \mathbf{m}_1 + \mathbf{m}_2$ and (18), the solution of the optimization problem (7b) can be expressed as

$$\underset{\substack{\mathbf{m}=\mathbf{m}_1+\mathbf{m}_2\\ \mathbf{m}\in\mathcal{C}}}{\arg\min} \quad \Phi_\alpha^{\text{TT}}(\mathbf{m}_1,\mathbf{m}_2) + \frac{\lambda_1}{2}\|\mathbf{L}[\mathbf{m}_1+\mathbf{m}_2] - \mathbf{y}\|_2^2, \quad (20)$$

where $\mathbf{y} = \mathbf{b} + \mathbf{b}^k - \boldsymbol{\Delta}\mathbf{u}^{k+1}$. Defining auxiliary variables $\mathbf{p} = \nabla\mathbf{m}_1 \in \mathbb{R}^{2n\times 1}$ and $\mathbf{q} = \mathbf{m}_1 + \mathbf{m}_2 \in \mathbb{R}^{n\times 1}$ and recasting (20) as a constrained problem and then applying ADMM leads to the following iteration [21, Section 2.2.2]

$$\begin{bmatrix}\mathbf{m}_1^{k+1}\\ \mathbf{m}_2^{k+1}\end{bmatrix} = \underset{\mathbf{m}_1,\mathbf{m}_2}{\arg\min}\ C(\mathbf{m}_1,\mathbf{m}_2,\mathbf{q}^k,\tilde{\mathbf{q}}^k,\mathbf{p}^k,\tilde{\mathbf{p}}^k), \quad (21a)$$

$$\mathbf{p}^{k+1} = \underset{\mathbf{p}}{\arg\min}\ \alpha\|\mathbf{p}\|_1 + \frac{\gamma_1}{2}\|\nabla\mathbf{m}_1^{k+1} - \mathbf{p} - \tilde{\mathbf{p}}^k\|_2^2, \quad (21b)$$

$$\mathbf{q}^{k+1} = \underset{\mathbf{q}\in\mathcal{C}}{\arg\min}\ \frac{\gamma_0}{2}\|\mathbf{m}_1^{k+1} + \mathbf{m}_2^{k+1} - \mathbf{q} - \tilde{\mathbf{q}}^k\|_2^2, \quad (21c)$$

where $\gamma_0, \gamma_1 > 0$ and

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

The auxiliary primal variables $\mathbf{p}$ and $\mathbf{q}$ are introduced to decouple the ℓ_1 and the ℓ_2 minimization problems and solve the former ones with proximal algorithms following operator splitting methods. The dual variables $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{q}}$ are updated through a gradient ascent step according to the method of multipliers [27]

$$\tilde{\mathbf{p}}^{k+1} \quad = \quad \tilde{\mathbf{p}}^k + \mathbf{p}^{k+1} - \nabla\mathbf{m}_1^{k+1}, \quad (22a)$$

$$\tilde{\mathbf{q}}^{k+1} \quad = \quad \tilde{\mathbf{q}}^k + \mathbf{q}^{k+1} - [\mathbf{m}_1^{k+1} + \mathbf{m}_2^{k+1}], \quad (22b)$$

We now discuss how to solve the subsubproblems given in (21).

1) The subsubproblem (21a)*:* A solution of subsubproblem (21a) occurs at the point where the derivatives of the objective function C with respect to $\mathbf{m}_1$ and $\mathbf{m}_2$ vanish simultaneously. Accordingly, we end up with the following linear system of equations:

$$\begin{bmatrix}\mathbf{G}_{11} & \mathbf{G}_{12}\\ \mathbf{G}_{21} & \mathbf{G}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{m}_1\\ \mathbf{m}_2\end{bmatrix} = \begin{bmatrix}\mathbf{h}_1\\ \mathbf{h}_2\end{bmatrix}, \quad (23)$$

with

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

and

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

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

[15] broke down the $2n \times 2n$ problem (23) into two smaller $n \times n$ systems and updates $\mathbf{m}_1$ and $\mathbf{m}_2$ in alternating mode at the expense of convergence speed [15, their eqs. 10 and 11]. Instead, we solve here the original system exactly using a variable projection scheme, thus leading to faster convergence and more accurate results. From the first equation of (23), we find that

$$\mathbf{m}_2 = \mathbf{G}_{12}^{-1}[\mathbf{h}_1 - \mathbf{G}_{11}\mathbf{m}_1] \tag{24}$$

and plugging this into the second equation of (23) we get the following:

$$\mathbf{m}_1 = [\mathbf{G}_{11} - \mathbf{G}_{22}\mathbf{G}_{12}^{-1}\mathbf{G}_{11}]^{-1}[\mathbf{h}_2 - \mathbf{G}_{22}\mathbf{G}_{12}^{-1}\mathbf{h}_1]. \tag{25}$$

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

2) The subsubproblem (21b)*:* The sub-problem for $\mathbf{p}$, equation (21b), is a denoising problem and is straightforward to solve. Note that $\mathbf{p}$ has two components associated with the gradient in each direction:

$$\mathbf{p} = \begin{bmatrix} \mathbf{p}_x \\ \mathbf{p}_z \end{bmatrix}. \tag{26}$$

Equation (21b) is solved with a generalized proximity operator [35] leading to

$$\mathbf{p}^{k+1} = \text{prox}_{\alpha/\gamma_1}(\mathbf{z}) = \begin{bmatrix} \xi \circ \mathbf{z}_x \\ \xi \circ \mathbf{z}_z \end{bmatrix}, \tag{27}$$

where

$$\mathbf{z} = \nabla \mathbf{m}_1^{k+1} - \tilde{\mathbf{p}}^k = \begin{bmatrix} \mathbf{z}_x \\ \mathbf{z}_z \end{bmatrix}, \tag{28}$$

and

$$\xi = \max(1 - \frac{\alpha}{\gamma_1 \sqrt{\mathbf{z}_x^2 + \mathbf{z}_z^2}}, 0). \tag{29}$$

3) The subsubproblem (21c)*:* The optimization problem (21c) also has an entrywise solution given by

$$\mathbf{q}^{k+1} = \text{proj}_{\mathcal{C}}(\mathbf{m}_1^{k+1} + \mathbf{m}_2^{k+1} - \tilde{\mathbf{q}}^k), \tag{30}$$

where the projection operator projects its argument onto the desired box $[\mathbf{m}_l, \mathbf{m}_u]$ according to $\text{proj}_{\mathcal{C}}(\bullet) = \min(\max(\bullet, \mathbf{m}_l), \mathbf{m}_u)$. Based on the above, the proposed ADMM based TT regularized IR-WRI algorithm is summarized in Algorithm 1. It should be noted that the total algorithm consists of two levels of iterations: an outer iteration (lines 4-13) given in (7) and an inner iteration (lines 5-11) given in (21) corresponding to the model subproblem (7b). Numerical results, however, show that only one inner iteration suffices for convergence of the algorithm, hence significantly reducing the total computational cost [36], [17]. The main computational cost of the algorithm is on lines 4 and 5 where we need to solve the augmented PDE and the $n \times n$ sparse system given in equation 25. The other steps don't introduce significant computational overheads. The computational overhead introduced by compound regularizer compared to single regularizer results from the larger numerical bandwidth of the matrix $[\mathbf{G}_{11} - \mathbf{G}_{22}\mathbf{G}_{12}^{-1}\mathbf{G}_{11}]$ in equation 25 compared to that associated with a single regularizer, either $\mathbf{G}_{11}$ or $\mathbf{G}_{22}$.

Algorithm 1 ADMM based TT regularized IR-WRI algorithm

1: Initialize: set the dual variables $\mathbf{b}^0$, $\mathbf{d}^0$, $\tilde{\mathbf{p}}^0$ and $\tilde{\mathbf{q}}^0$ equal to $\mathbf{0}$
2: Input: $\mathbf{m}^0$ (initial model parameters)
3: **while** convergence criteria not satisfied **do**
4: $\quad \mathbf{u}^{k+1} \leftarrow \left[\frac{\lambda_0}{\lambda_1}\mathbf{P}^T\mathbf{P} + \mathbf{A}(\mathbf{m}^k)^T\mathbf{A}(\mathbf{m}^k)\right]^{-1}\left[\frac{\lambda_0}{\lambda_1}\mathbf{P}^T[\mathbf{d} + \mathbf{d}^k] + \mathbf{A}(\mathbf{m}^k)^T[\mathbf{b} + \mathbf{b}^k]\right]$
5: $\quad \mathbf{m}_1^{k+1} \leftarrow$ update according to eq. (25)
6: $\quad \mathbf{m}_2^{k+1} \leftarrow$ update according to eq. (24)
7: $\quad \mathbf{m}^{k+1} \leftarrow \mathbf{m}_1^{k+1} + \mathbf{m}_2^{k+1}$
8: $\quad \mathbf{p}^{k+1} \leftarrow \text{prox}_{\alpha/\gamma_1}(\nabla \mathbf{m}_1^{k+1} - \tilde{\mathbf{p}}^k)$
9: $\quad \mathbf{q}^{k+1} \leftarrow \text{proj}_{\mathcal{C}}(\mathbf{m}^{k+1} - \tilde{\mathbf{q}}^k)$
10: $\quad \tilde{\mathbf{p}}^{k+1} \leftarrow \tilde{\mathbf{p}}^k + \mathbf{p}^{k+1} - \nabla \mathbf{m}_1^{k+1}$
11: $\quad \tilde{\mathbf{q}}^{k+1} \leftarrow \tilde{\mathbf{q}}^k + \mathbf{q}^{k+1} - \mathbf{m}^{k+1}$
12: $\quad \mathbf{d}^{k+1} \leftarrow \mathbf{d}^k + \mathbf{d} - \mathbf{P}\mathbf{u}^{k+1}$
13: $\quad \mathbf{b}^{k+1} \leftarrow \mathbf{b}^k + \mathbf{b} - \mathbf{A}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}$
14: **end while**

V. Numerical examples

We assess the performance of our algorithm against 1D and 2D mono-parameter synthetic examples. In Table I we give different regularization functions which are applied for stabilizing the FWI solution. We start with zero-offset Vertical-Seismic-Profiling (VSP) examples (1D IR-WRI) where the targeted wave speed profiles are selected from well-documented 2D benchmark subsurface velocity models in exploration seismic. To tackle more realistic applications, we proceed with a target of the 2D challenging 2004 BP salt model [37] with noiseless and noisy data when a crude initial model and realistic frequencies are used as starting points.

A. Performance comparison using 1D test on benchmark models

First, we assess the performance of our regularized IR-WRI against 1D mono-parameter synthetic examples when the true models are 100 vertical profiles selected from the 2004 BP salt [37], Marmousi II [38], SEG/EAGE overthrust [39], SEG/EAGE salt [39] and synthetic Valhall [40] benchmark velocity models (we extracted 20 profiles from each benchmark model). For all of the experiments, a single source is used at the surface and the receivers are evenly deployed along the entire profile. A single frequency, whose value is set so that the reduced-space inversion is prone to cycle skipping, is considered for inversion. The model dimension, the inverted frequency and the receiver spacing are outlined for each model in Table II. We perform forward modeling with a 3-point $\mathcal{O}(\Delta x^2)$ staggered-grid finite-difference stencil and PML absorbing boundary conditions at the two ends of the model. The starting model for IR-WRI is a homogeneous velocity model in which the velocity is the mean value of each profile. We set the penalty parameters according to the guideline given in Appendix C. Moreover, for a fair comparison of the compound regularizers (JTT, TT and TGV), we select for each of them the optimum value of α among a range of preset values that minimizes the error in the models estimated by the IR-WRI. Also, we set the parameter bounds $\mathbf{m}_l$ and $\mathbf{m}_u$ equal to 50% and 150% of the minimum and maximum velocities of the true model, respectively. The monochromatic inversion

TABLE I: Different regularization functions.

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

TABLE II: Experimental setup of 1D model tests

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

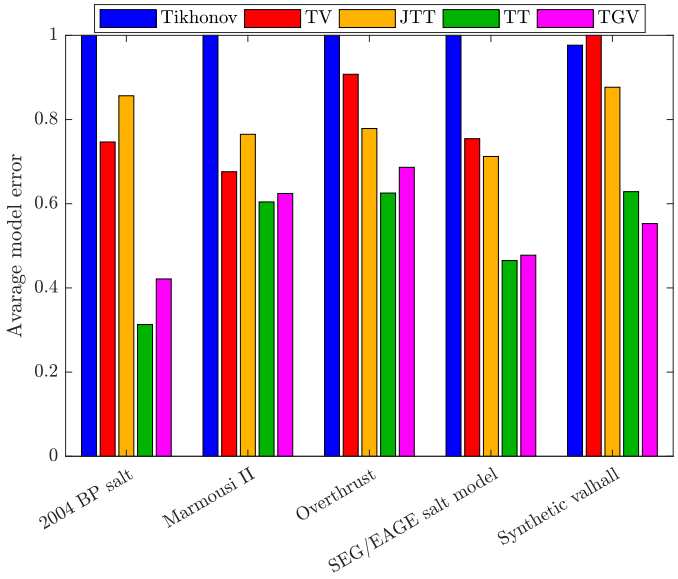

Fig. 2: Zero offset VSP test. Average model error in estimated 1D profiles for different velocity models and different regularization functions.

is performed with noiseless data when a maximum number of iterations, equal to 100, is used as a stopping criterion. The average error of the estimated velocity profiles for the five benchmark models and the different regularizations are plotted in Fig. 2. In this paper, the model error is defined as the energy of the difference between the true model and the estimated one compared to the energy of the true model. The errors in each model for different regularizations are normalized to 1 for sake of clarity (the error of DMP regularizer is not shown because of its worse performance). Fig. 2 clearly shows that the compound regularizations based upon infimal convolution (TT and TGV) always behave better than the CC regularization and the single regularization functionals (TV and Tikhonov). To emphasize the effects of the different regularization functions, we plot some close-ups of the reconstructed profiles in Fig. 3. These results show that TT provides the most accurate reconstruction for the 2004 BP salt (Fig. 3a) and Overthrust (Fig. 3c) models. This is consistent with the fact that the velocity trends of these two models match well the piecewise smooth prior. In contrast, TGV behaves slightly better than TT for the Valhall model, whose velocity trend is the closest one to the piecewise linear prior (Fig. 3d). For Marmousi II (Fig. 3b), TT and TGV give similar results.

B. 2004 BP salt model

We now consider a more realistic application with a target of the challenging 2004 BP salt model [37]. The 2004 BP salt model is representative of the geology of the deep offshore Gulf of Mexico and mainly consists of a simple background with a complex rugose multi-valued salt body, sub-salt slow velocity anomalies related to over-pressure zones and a fast velocity anomaly to the right of the salt body. The selected subsurface model is 16250 m wide and 5825 m deep, and is discretized with a 25 m grid interval (Fig. 4a). We used 108 sources spaced 150 m apart on the top side of the model. We perform forward modeling with a staggered-grid 9-point finite-difference method [23] with PML boundary conditions along the four edges of the model. The source signature is a 10 Hz Ricker wavelet. A line of receivers with a 25 m spacing are deployed all along the surface leading to a long-offset stationary-receiver acquisition. We used small batches of two frequencies with one frequency overlap between two consecutive batches, moving from the low frequencies to the higher ones according to a classical frequency continuation strategy. We use $\alpha = 0.7$ and set the rest of hyperparameters according to the guidelines reviewed in appendix C. We set the parameter bounds $\mathbf{m}_l$ and $\mathbf{m}_u$ equal to the minimum and maximum velocity of the true model, respectively. The starting and final frequencies are 3 Hz and 13 Hz and the sampling interval in one batch is 0.5 Hz. The initial velocity model is a crude laterally-homogeneous velocity-gradient model with velocities ranging between 1.5 to 4.5 km/s (Fig. 4b). We start with inverting the first batch of frequencies ($\{3, 3.5\}$ Hz) with noiseless data using a maximum number of iterations equal to 45 as a stopping criterion. To highlight the specific role of bound constraints, we activate them after 20 iterations. To emphasize the effect of regularization, the result of bound constrained IR-WRI with a simple DMP regularization is shown in Fig. 5a, while the bound-constrained IR-WRI results with Tikhonov and TV regularizations are shown in Figs. 5b and 5c, respectively. Although the TV reconstruction is better than the Tikhonov one, it provides a velocity model which is far from the optimal one. A direct comparison between the true model, the starting model and the estimated models is shown in Fig. 6a along three vertical logs at 2.5, 9.0 and 15.0 km distance (as depicted with dashed white lines in Fig. 4a). We continue with compound regularization results which are shown in Fig. 5d-f and Fig. 6b. Clearly, the TT regularizer better captures the long wavelengths of the salt body and the smooth subsalt

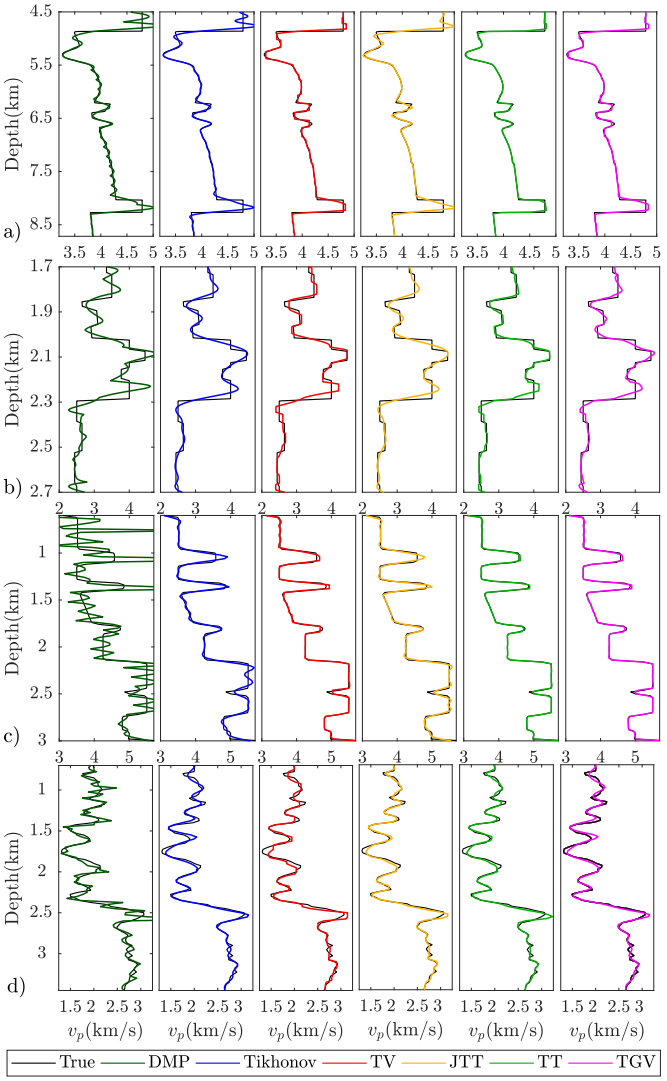

Fig. 3: Zero offset VSP test. Some part of (a) 2004 BP salt, (b) Marmousi II, (c) Overthrust, and (d) synthetic Valhall models estimated with different regularizations.

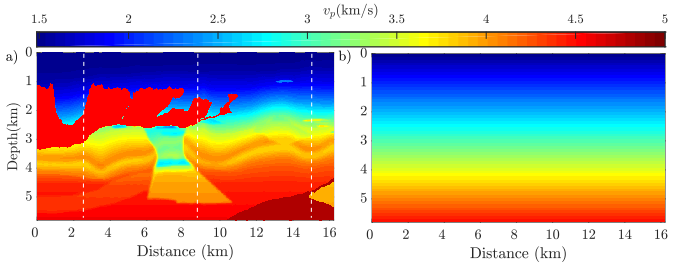

Fig. 4: 2004 BP salt case study. (a) True velocity model. The vertical dashed lines indicate the location of vertical logs of Figs. 6, 9 and 15. (b) The velocity-gradient initial model.

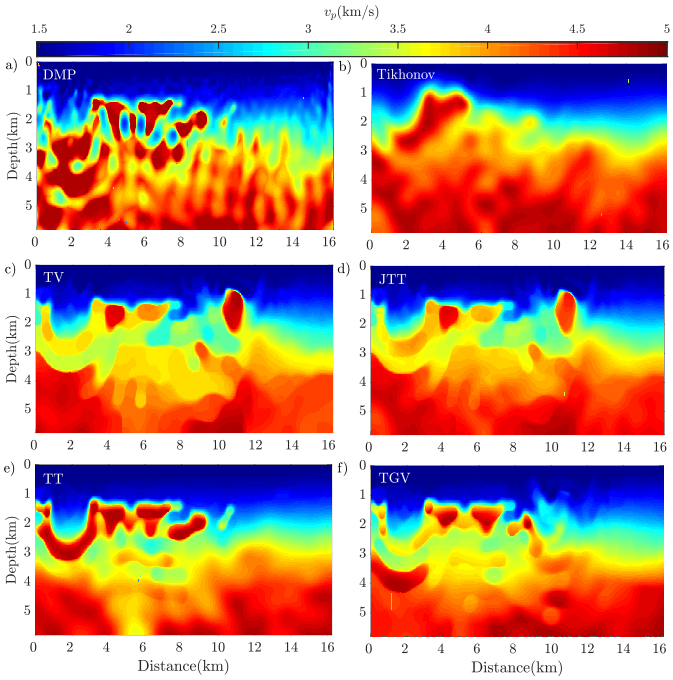

Fig. 5: Noiseless 2004 BP salt case study. Velocity models obtained after the $\{3, 3.5\}$ Hz inversion when the velocity-gradient model (Fig. 4b) is used as initial model. (a-f) Bound constrained IR-WRI with (a) DMP, (b) Tikhonov, (c) TV, (d) JTT, (e) TT and (f) TGV regularization.

background model. The joint evolution in iterations of the observation-equation ($\|\mathbf{Pu}-\mathbf{d}\|_2$) and wave-equation errors ($\|\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{b}\|_2$), Fig. 7a-b, and the relative model errors in iterations, Fig. 7c-d, further confirm the relative performance of each regularizer during the inversion of the first frequency batch. Note the complex zigzag path followed by the inversion to jointly minimize the data residuals and the wave equation error in Fig. 7a-b. As already highlighted by [17], this results from the dynamic balancing in iterations of the observation-equation and wave-equation constraints performed by the dual updates with the data and source residuals.

We continue the inversion at higher frequencies using the final models of the $\{3, 3.5\}$ Hz inversion, Fig. 5a-f, as initial models when the stopping criteria is either $k_{max} = 15$ or

$$\|\mathbf{A}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}-\mathbf{b}\|_2 \leq \varepsilon_b, \qquad \|\mathbf{P}\mathbf{u}^{k+1}-\mathbf{d}\|_2 \leq \varepsilon_d, \quad (31)$$

where k_{max} denotes the maximum iteration count, ε_b=1e-3, and ε_d=1e-5. We perform three paths through the frequency batches to improve the IR-WRI results, using the final model of one path as the initial model of the next one (these cycles can be viewed as outer iterations of IR-WRI). The starting and finishing frequencies of the paths are [3.5, 6], [4, 8.5], [6, 13] Hz respectively, where the first element of each pair shows the starting frequency and the second one is the finishing frequency. The bound-constrained IR-WRI models obtained from noiseless data are shown in Fig. 8. As for the inversion of the first batch, direct comparison between the true model, the starting model and the estimated models are shown in Fig. 9 along three vertical logs at 2.5 km, 9.0 km and 15 km distance (vertical dashed lines in Fig. 4a). The TT and TGV regularizers lead to high-quality velocity models, that capture both the fine-scale structure of the rugose large-contrast salt body and the high-velocity shallow anomaly on the right, as well as the smoother sub-salt background model including the low-velocity over-pressure structure. It is also

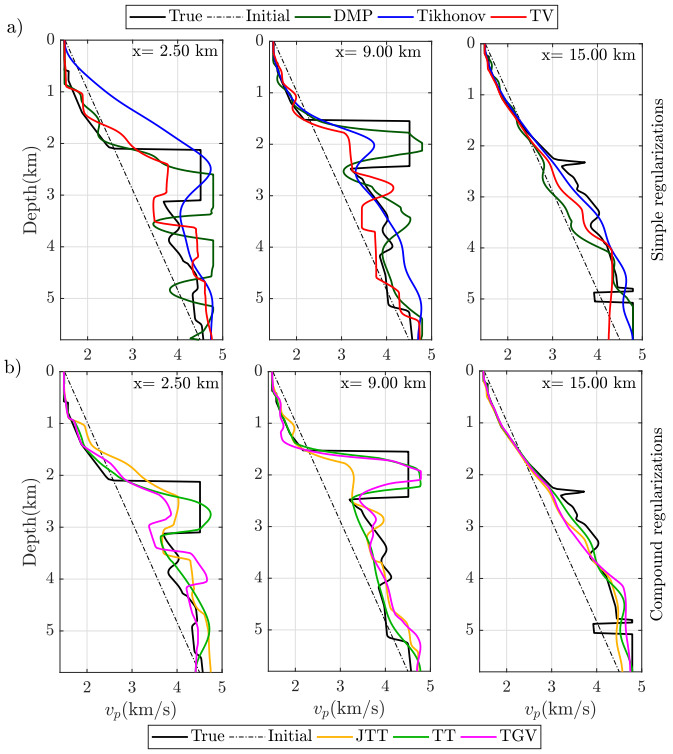

Fig. 6: Noiseless 2004 BP salt case study. Direct comparison along the logs shown in Fig. 4a between the true velocity model (black), the initial model (dashed line) and the estimated models obtained after the $\{3, 3.5\}$ Hz inversion (Fig. 5) with (a) simple regularizations (DMP in olive-green, Tikhonov in blue and TV in red) and (b) compound regularizations (JTT in orange, TT in green and TGV in pink).

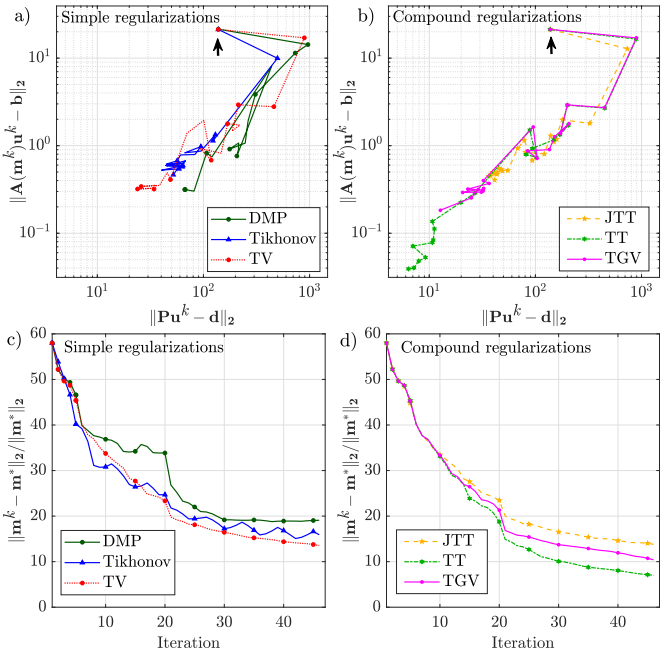

Fig. 7: Noiseless 2004 BP salt case study. $\{3, 3.5\}$ Hz inversion with the velocity-gradient initial model. (a-b) convergence history of the algorithm in the $(\|\mathbf{P}\mathbf{u}^k - \mathbf{d}\|_2 - \|\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b}\|_2)$ plane for (a) simple regularizations and (b) compound regularizations. The black arrow points the starting point. (c-d) evaluation of $\|\mathbf{m}^k - \mathbf{m}^*\|_2/\|\mathbf{m}^*\|_2$ during the iteration where $\mathbf{m}^*$ is the true model. The panels (a) and (b) as well as (c) and (d) are plotted with the same horizontal and vertical scale.

TABLE III: Total number of IR-WRI iterations for each regularization.

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

worth noting the significant differences between the JTT and TT IR-WRI models in particular in the deep part of the model. Moreover, the number of iterations performed by IR-WRI for each regularization shows that TT has the best convergence speed (Table III).

As a final quality control of the different IR-WRI models, it is instructive to check the wave-equation and data residuals left by the different regularization methods for the starting 3-Hz frequency (Fig. 10 and 11). The real part of wave-equation error (Fig. 10) and data residuals (Fig. 11) are plotted at the first and final iterations of the inversion. Both of the final data and source residuals suggest that the TT regularizer slightly outperforms the TGV counterpart at low frequencies. To further illustrate the ability of compound regularizations to manage the blocky and smooth components of the subsurface, we show separately the two model components of IC based regularizations ($\mathbf{m}_1$ and $\mathbf{m}_2$) estimated with TT (Fig. 12a-b) and TGV regularizations (Fig. 12c-d) as well as histograms of $\nabla\mathbf{m}_1$ and $\nabla^2\mathbf{m}_2$ (Fig. 13). The trend of the reconstructed $\mathbf{m}_1$ and $\mathbf{m}_2$ shows that the IC based compound regularizations mostly succeeded in de-coupling the reconstruction of the blocky structure from that of the smooth background for TT regularization (Fig. 12a-b) and de-coupling the reconstruction of the blocky structure from that of the piecewise linear background for TGV regularization (Fig. 12c-d). This statement is further supported by the long-tail shape and the Gaussian shape of the histograms of the $\nabla\mathbf{m}_1$ and $\nabla^2\mathbf{m}_2$ components of TT regularization (Fig. 13a-b) and the long-tail shapes of the $\nabla\mathbf{m}_1$ and $\nabla^2\mathbf{m}_2$ components of TGV regularization (Fig. 13c-d).

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

$$\text{SNR} = 20\log\left(\frac{A_{signal}}{A_{noise}}\right), \tag{32}$$

in which A denotes root mean square (RMS) amplitude. We use the same setup and the same initial velocity model (Fig. 4b) as those used for the noiseless case. The stopping criterion is defined by (31), where ε_d is now set to the noise level. The final models of bound-constrained IR-WRI obtained from noisy data are shown in Fig. 14. The number of iterations performed by IR-WRI with the different regularizations are outlined in Table III. As for the inversion of the first batch,

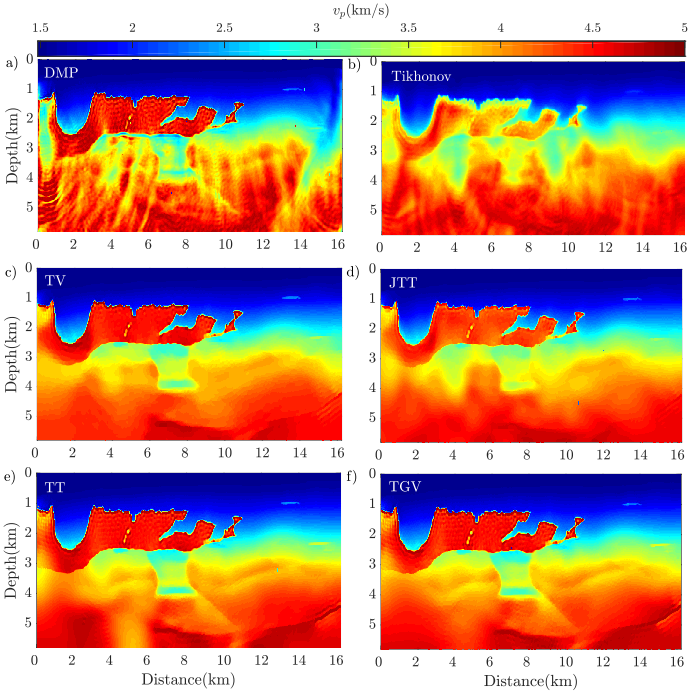

Fig. 8: Noiseless 2004 BP salt case study. Final inversion results with the velocity models of Fig. 5a-f as initial models. The display of the panels is the same as that of Fig. 5.

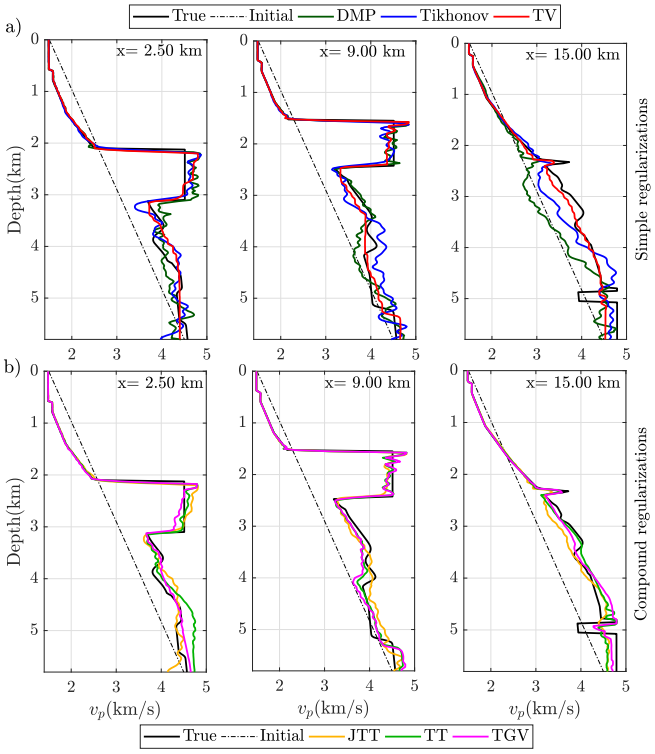

Fig. 9: Noiseless 2004 BP salt case study. Direct comparison of final inversion results with Fig. 5a-f as initial models. The panels are same as Fig. 6 for final results of Fig. 8.

direct comparison between the true model, the starting model and the estimated models are shown in Fig. 15 along three vertical logs at 2.5 km, 9.0 km and 15 km distance. The

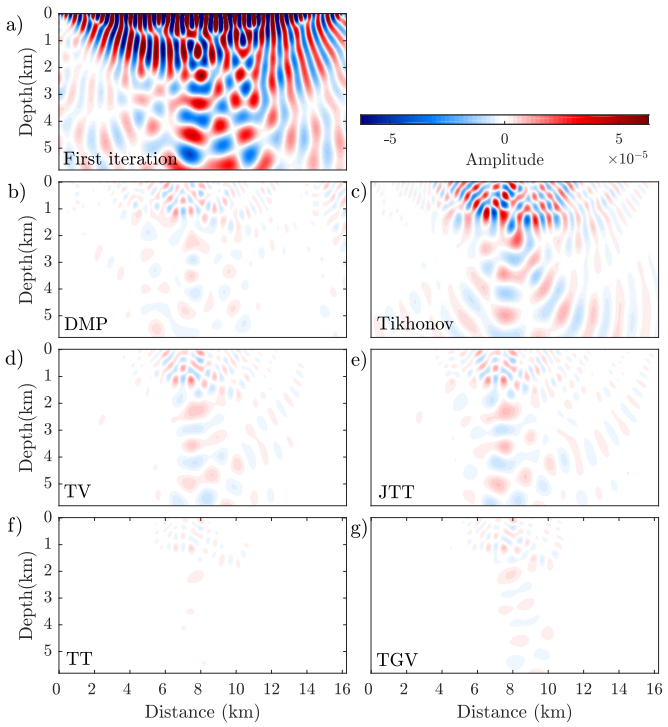

Fig. 10: Noiseless 2004 BP salt case study. For a source at x=8.12 km and the 3-Hz frequency: (a) Real part of wave-equation residual at first iteration of wavefield reconstruction, namely $(\mathbf{A}(\mathbf{m}^0)\mathbf{u}^1 - \mathbf{b})$. (b-g) Real part of wave-equation residual at the final iteration achieved respectively with DMP, Tikhonov, TV, JTT, TT, and TGV regularization.

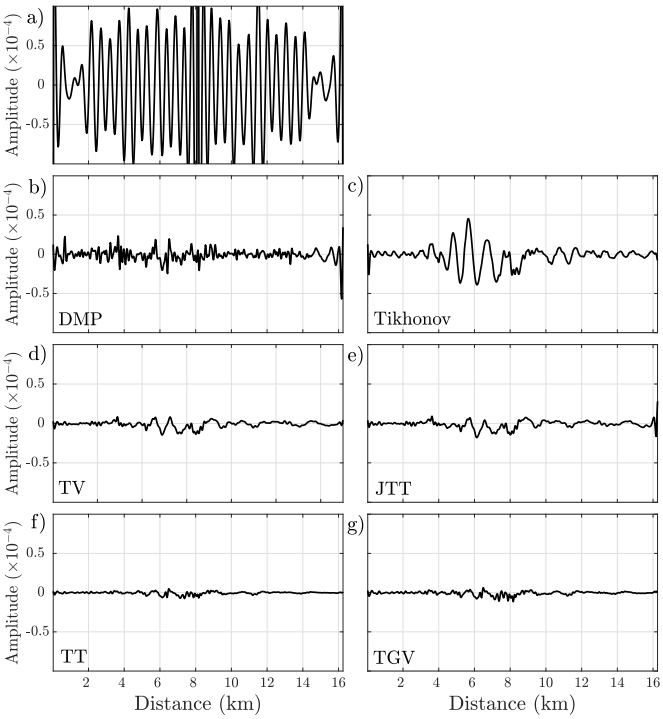

Fig. 11: Noiseless 2004 BP salt case study. For a source at x=8.12 km and the 3-Hz frequency: (a) Real part of data residual $(\mathbf{Pu} - \mathbf{d})$ at first iteration. (b-g) Real part of data residual at the final iteration achieved respectively with DMP, Tikhonov, TV, JTT, TT, and TGV regularization.

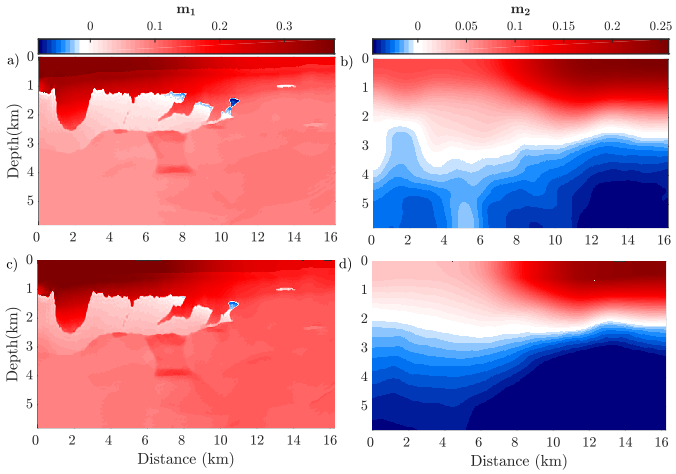

Fig. 12: Noiseless 2004 BP salt case study. Final blocky and smooth components, $\mathbf{m}_1$ and $\mathbf{m}_2$, reconstructed by IR-WRI with (a-b) TT and (c-d) TGV regularizations. The corresponding velocity models are shown in Figs. 8e and 8f. (a,c) $\mathbf{m}_1$. (b,d) $\mathbf{m}_2$. Note that $\mathbf{m}_1$ and $\mathbf{m}_2$ are parametrized with squared slownesses.

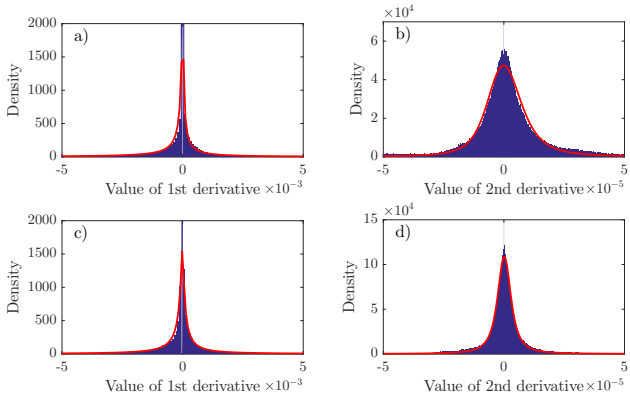

Fig. 13: Noiseless 2004 BP salt case study. Histograms of the model components shown in Fig. 12. The red lines show the probability density functions fitted to the histograms.

results further confirm that TT regularization provides the most reliable results and illustrate the resilience of this compound regularization to noise.

VI. Conclusions

In this study, we first show how to efficiently implement different kinds of regularization and bound constraints in the wavefield reconstruction inversion method with the alternating direction method of multipliers (ADMM). Then, we show the capability of IR-WRI when equipped with compound Tikhonov and TV regularizations to accurately reconstruct large-contrast subsurface media when starting from a very crude initial model. This compound regularization is suitable for seismic imaging of the subsurface as it can often be represented by piecewise smooth media. We show that the infimal convolution (IC) of the Tikhonov and TV regularizers captures much more accurately the blocky and smooth components of the subsurface than the convex combination of the two regularizers. It also outperforms the Tikhonov and TV regularizers when used alone. We also show how the infimal-convolution regularizer can be efficiently implemented by jointly updating the smooth and blocky subsurface components through variable projection. Alternatively, TGV regularized IR-WRI can be a suitable tool to reconstruct piecewise linear media and provides similar results than TT IR-WRI. We conclude that such hybrid regularizations in the extended search-space IR-WRI potentially provide a suitable framework to reconstruct, without cycle skipping, large-contrast subsurface media from ultra-long offset seismic data. It should also find applications in other fields of imaging sciences such as medical imaging.

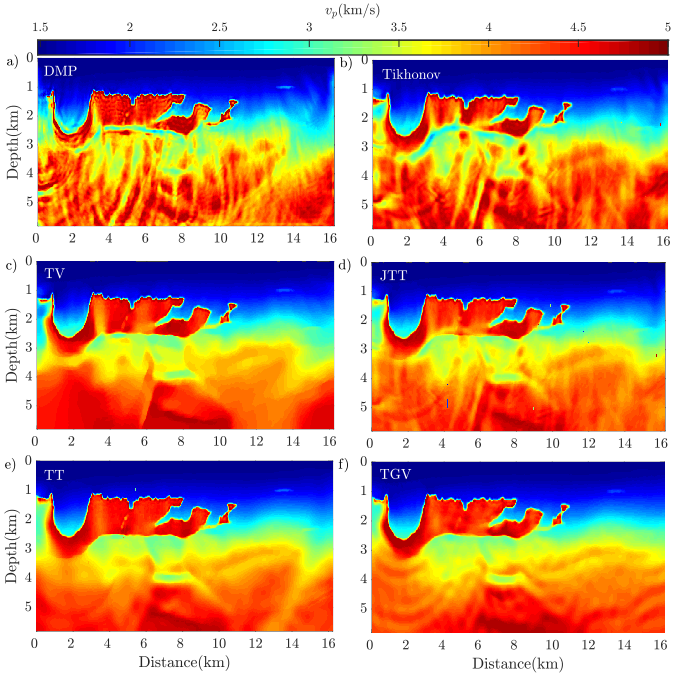

Fig. 14: Noisy 2004 BP salt case study. Final inversion results with Fig. 4b as initial model. (a-f) bound constrained IR-WRI with (a) DMP, (b) Tikhonov, (c) TV, (d) JTT, (e) TT and (f) TGV regularization.

VII. Acknowledgments

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

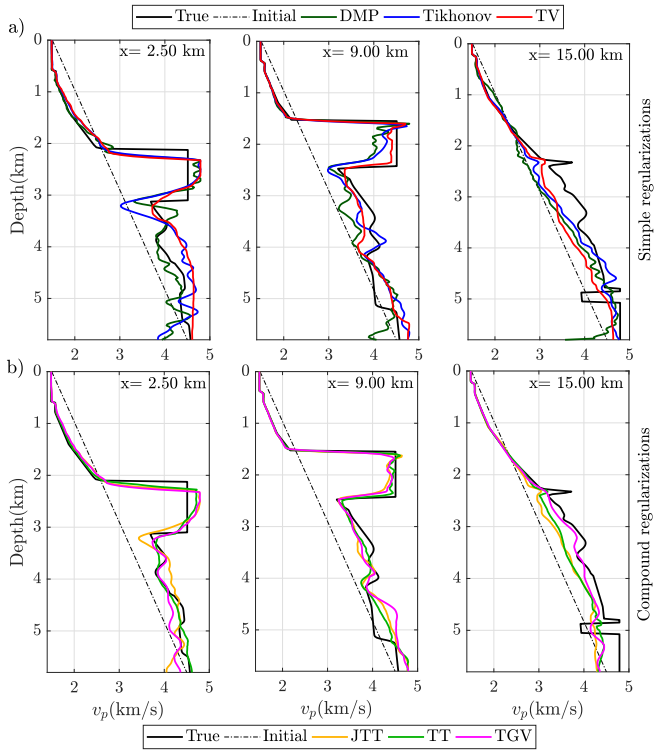

Fig. 15: Noisy 2004 BP salt case study. Direct comparison (along the logs shown in Fig. 4a) between the true velocity model (black), the initial model (dashed line) and the estimated models with (a) simple regularizations (DMP in olive-green, Tikhonov in blue and TV in red) and (b) compound regularizations (JTT in orange, TT in green and TGV in pink).

APPENDIX

A. Reduced approach to solving (1)

The reduced approach, which is more commonly used for sake of computational efficiency, strictly enforces the PDE constraint at each iteration by projection of the full space onto the parameter search space, leading to the following unconstrained optimization problem [5], [22], [41]

$$\min_{\mathbf{m}\in\mathcal{C}} \ \Phi(\mathbf{m}) + \frac{\lambda_0}{2}||\mathbf{P}\mathbf{A}^{-1}(\mathbf{m})\mathbf{b} - \mathbf{d}||_2^2, \tag{33}$$

where $\lambda_0 > 0$ is the penalty parameter. A number of methods have been proposed to solve the optimization problems of the form (33), either for unregularized form $\Phi(\mathbf{m}) = 0$ [22], or the regularized form [42], [43]. Although this reduced approach is more computationally tractable than the full-space approach, the highly-oscillating nature of the inverse PDE operator $\mathbf{A}^{-1}$ makes the inverse problem highly nonlinear, and hence prone to convergence to a spurious local minima when the initial $\mathbf{m}$ is not accurate enough [5], [44]. The extended approach described in this paper (section III) is an alternative way which is more immune to local minima.

B. Simple regularizers

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

$$\|\mathbf{x}\|_2^2 = \sum_{i=1}^{n} |\mathbf{x}_i|^2, \tag{34}$$

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

In contrast, the ℓ_1-norm, defined as

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |\mathbf{x}_i|, \tag{35}$$

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

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

$$\begin{cases} (\nabla_x f)_{i,j} = f_{i,j} - f_{i,j-1}, \\ (\nabla_z f)_{i,j} = f_{i,j} - f_{i-1,j}, \end{cases} \tag{36}$$

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

$$(|\nabla f|)_{i,j} = \sqrt{(\nabla_x f)_{i,j}^2 + (\nabla_z f)_{i,j}^2}. \tag{37}$$

The squared ℓ_2-norm of $|\nabla f|$ gives the first order Tikhonov regularization [7], which returns a flat regularized model (with a small gradient), while its ℓ_1-norm gives the total variation regularization [8], which returns a piecewise constant model (with a sparse gradient).

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

$$\begin{cases} (\nabla_{xx} f)_{i,j} = f_{i,j-1} - 2f_{i,j} + f_{i,j+1}, \\ (\nabla_{zz} f)_{i,j} = f_{i-1,j} - 2f_{i,j} + f_{i+1,j}, \end{cases} \tag{38}$$

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

$$(|\nabla^2 f|)_{i,j} = \sqrt{(\nabla_{xx} f)_{i,j}^2 + (\nabla_{zz} f)_{i,j}^2}. \tag{39}$$

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

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

$$(\nabla_{xz} f)_{i,j} = f_{i,j} - f_{i,j-1} - f_{i-1,j} + f_{i-1,j-1}. \tag{40}$$

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

$$(|\nabla^2 f|)_{i,j} = \sqrt{(\nabla_{xx} f)_{i,j}^2 + 2(\nabla_{xy} f)_{i,j}^2 + (\nabla_{yy} f)_{i,j}^2}, \quad (41)$$

which equals the Frobenius norm of the Hessian matrix [45], [46], [47].

C. Parameter tuning

Here, we provide some guidelines to tune the different hyperparameters in the regularized IR-WRI method. The reader is also referred to [21] for more details. We start with γ_1 (step 8 of Algorithm 1), which controls the soft thresholding performed by the TV regularization, and set it equal to $0.02 \times \max|\nabla \mathbf{m}_1^k - \tilde{\mathbf{p}}^k|$. This tuning can be refined according to prior knowledge of the geological structure, coming from well logs for example. Also, we use the same weight for the bound constraints and the TV regularization: $\gamma_0 = \gamma_1$. Once we set γ_1, we define λ_1 such that γ_1/λ_1 is a percentage of mean absolute value of the diagonal coefficients of $\mathbf{L}^T\mathbf{L}$. Parameter λ_1 may be increased during iterations to reduce the weight of TV regularization and bound constraints near the convergence point. Finally, we set λ_0 such that $\lambda = \lambda_1/\lambda_0$ is a small fraction of the highest eigenvalue ξ of the normal operator $\mathbf{A}(\mathbf{m})^{-T}\mathbf{P}^T\mathbf{P}\mathbf{A}(\mathbf{m})^{-1}$ during the wavefield reconstruction subproblem according to the criterion proposed by [48]. In all the numerical tests, we use $\lambda =$ 1e-5ξ and $\lambda =$ 1e-3ξ for noiseless and noisy data, respectively.

REFERENCES

[1] A. Tarantola, "Inversion of seismic reflection data in the acoustic approximation," *Geophysics*, vol. 49, no. 8, pp. 1259–1266, 1984.

[2] G. Y. Sandhu, C. Li, O. Roy, S. Schmidt, and N. Duric, "Frequency domain ultrasound waveform tomography: breast imaging using a ring transducer," *Physics in Medicine and Biology*, vol. 60, pp. 5381–5398, 2015.

[3] W. T. Wood, W. S. Holbrook, M. K. Sen, and P. L. Stoffa, "Full waveform inversion of reflection seismic data for ocean temperature profiles," *Geophysical Research Letters*, vol. 35, p. L04608, 2008.

[4] E. Haber, U. M. Ascher, and D. Oldenburg, "On optimization techniques for solving nonlinear inverse problems," *Inverse problems*, vol. 16, no. 5, p. 1263, 2000.

[5] J. Virieux and S. Operto, "An overview of full waveform inversion in exploration geophysics," *Geophysics*, vol. 74, no. 6, pp. WCC1–WCC26, 2009.

[6] J. Virieux, A. Asnaashari, R. Brossier, L. Métivier, A. Ribodetti, and W. Zhou, "An introduction to Full Waveform Inversion," in *Encyclopedia of Exploration Geophysics*, V. Grechka and K. Wapenaar, Eds. Society of Exploration Geophysics, 2017, pp. R1–1–R1–40. [Online]. Available: http://library.seg.org/doi/abs/10.1190/1.9781560803027.entry6

[7] A. Tikhonov, A. Goncharsky, V. Stepanov, and A. Yagola, *Numerical methods for the solution of ill-posed problems*. Springer Science & Business Media, 2013.

[8] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[9] N. G. Polson and V. Sokolov, "Bayesian regularization: From Tikhonov to horseshoe," *Wiley Interdisciplinary Reviews: Computational Statistics*, p. e1463, 2019.

[10] A. Gholami and H. Siahkoohi, "Regularization of linear and non-linear geophysical ill-posed problems with joint sparsity constraints," *Geophysical Journal International*, vol. 180, no. 2, pp. 871–882, 2010.

[11] M. Benning and M. Burger, "Modern regularization methods for inverse problems," *Acta Numerica*, vol. 27, pp. 1–111, 2018.

[12] J. M. Bioucas-Dias and M. A. Figueiredo, "An iterative algorithm for linear inverse problems with compound regularizers," in *2008 15th IEEE International Conference on Image Processing*. IEEE, 2008, pp. 685–688.

[13] A. Gholami and S. M. Hosseini, "A balanced combination of Tikhonov and total variation regularizations for reconstruction of piecewise-smooth signals," *Signal Processing*, vol. 93, pp. 1945–1960, 2013.

[14] R. Bergmann, J. H. Fitschen, J. Persch, and G. Steidl, "Priors with coupled first and second order differences for manifold-valued image processing," *Journal of mathematical imaging and vision*, vol. 60, pp. 1459–1481, 2018.

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

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

[17] H. Aghamiry, A. Gholami, and S. Operto, "Improving full-waveform inversion by wavefield reconstruction with alternating direction method of multipliers," *Geophysics*, vol. 84(1), pp. R139–R162, 2019.

[18] ——, "Improving full-waveform inversion based on wavefield reconstruction via Bregman iterations," in *Expanded Abstracts, 80th Annual EAGE Meeting (Copenhagen)*, 2018.

[19] ——, "ADMM-based multi-parameter wavefield reconstruction inversion in VTI acoustic media with TV regularization," " *Geophysical Journal International*, vol. 219, no. 2, pp. 1316–1333, 2019.

[20] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction of multipliers," *Foundations and trends in machine learning*, vol. 3, no. 1, pp. 1–122, 2010.

[21] H. Aghamiry, A. Gholami, and S. 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*, vol. 218, no. 2, pp. 855–872, 2019.

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

[23] Z. Chen, D. Cheng, W. Feng, and T. Wu, "An optimal 9-point finite difference scheme for the Helmholtz equation with PML," *International Journal of Numerical Analysis & Modeling*, vol. 10, no. 2, 2013.

[24] K. Marfurt, "Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations," *Geophysics*, vol. 49, pp. 533–549, 1984.

[25] C. H. Jo, C. Shin, and J. H. Suh, "An optimal 9-point, finite-difference, frequency-space 2D scalar extrapolator," *Geophysics*, vol. 61, pp. 529–537, 1996.

[26] B. Hustedt, S. Operto, and J. Virieux, "Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling," *Geophysical Journal International*, vol. 157, pp. 1269–1296, 2004.

[27] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. Springer, 2006.

[28] M. R. Hestenes, "Multiplier and gradient methods," *Journal of optimization theory and applications*, vol. 4, no. 5, pp. 303–320, 1969.

[29] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the royal statistical society: series B (statistical methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[30] A. Gholami, "Sparse time–frequency decomposition and some applications," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 6, pp. 3598–3604, 2013.

[31] D. L. Donoho, "De-noising by soft-thresholding," *IEEE transactions on information theory*, vol. 41, no. 3, pp. 613–627, 1995.

[32] P. J. Huber, "Robust regression: Asymptotics, conjectures, and Monte Carlo," *The Annals of Statistics*, vol. 1, no. 5, pp. 799–821, 1973.

[33] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 492–526, 2010.

[34] S. Setzer, G. Steidl, and T. Teuber, "Infimal convolution regularizations with discrete 1-type functionals," *Communications in Mathematical Sciences*, vol. 9, no. 3, pp. 797–827, 2011.

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

[36] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.

[37] F. J. Billette and S. Brandsberg-Dahl, "The 2004 BP velocity benchmark," in *Extended Abstracts, 67th Annual EAGE Conference & Exhibition, Madrid, Spain*, 2004, p. B035.
[38] G. S. Martin, R. Wiley, and K. J. Marfurt, "Marmousi2: An elastic upgrade for Marmousi," *The Leading Edge*, vol. 25, no. 2, pp. 156–166, 2006. [Online]. Available: http://link.aip.org/link/?LEE/25/156/1
[39] F. Aminzadeh, J. Brac, and T. Kunz, *3-D Salt and Overthrust models*. SEG/EAGE 3-D Modeling Series No.1, 1997.
[40] V. Prieux, R. Brossier, Y. Gholami, S. Operto, J. Virieux, O. Barkved, and J. Kommedal, "On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study," *Geophysical Journal International*, vol. 187, pp. 1495–1515, 2011.
[41] R. E. Plessix, "A review of the adjoint-state method for computing the gradient of a functional with geophysical applications," *Geophysical Journal International*, vol. 167, no. 2, pp. 495–503, 2006.
[42] A. Asnaashari, R. Brossier, S. Garambois, F. Audebert, P. Thore, and J. Virieux, "Regularized seismic full waveform inversion with prior model information," *Geophysics*, vol. 78, no. 2, pp. R25–R36, 2013.
[43] E. Esser, L. Guasch, T. van Leeuwen, A. Y. Aravkin, and F. J. Herrmann, "Total variation regularization strategies in Full-Waveform Inversion," *SIAM Journal Imaging Sciences*, vol. 11, no. 1, pp. 376–406, 2018.
[44] W. W. Symes, "Migration velocity analysis and waveform inversion," *Geophysical Prospecting*, vol. 56, pp. 765–790, 2008.
[45] S. Lefkimmiatis, A. Bourquard, and M. Unser, "Hessian-based norm regularization for image restoration with biomedical applications," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 983–995, 2012.
[46] A. Gholami and M. D. Sacchi, "Fast 3D blind seismic deconvolution via constrained total variation and gcv," *SIAM Journal on Imaging Sciences*, vol. 6, no. 4, pp. 2350–2369, 2013.
[47] A. Gholami and E. Z. Naeini, "3D Dix inversion using bound-constrained TV regularization," *Geophysics*, vol. 84, no. 3, pp. 1–43, 2019.
[48] T. van Leeuwen and F. Herrmann, "A penalty method for PDE-constrained optimization in inverse problems," *Inverse Problems*, vol. 32(1), pp. 1–26, 2016.