Iranian Journal of Geophysics, Vol 19 No 6, 2026, Page 71 – 83 (Research Article)
DOI: 10.30499/ijg.2024.457446.1602

Nonlinear travel-time cross-hole tomography with overlapping group sparse total variation regularization

Yaser Soufi[1], Mohammad Ali Riahi[2*], Reza Heidari[3] and Mahmoud Mehramuz[3]

[1] *Ph.D. Student, Islamic Azad University, Science and Research Branch, Tehran Iran*
[2] *Professor, Institute of Geophysics, University of Tehran, Iran*
[3] *Assistant Professor, Islamic Azad University, Science and Research Branch, Tehran Iran*

(Received: 14 May 2024, Accepted: 25 November 2024)

Abstract

To address the inherent ill-posedness of the geophysical inverse problems, it is necessary to include a suitable regularization function in the corresponding optimization framework. Typically, the choice of the regularization function depends on prior assumptions about the geometric characteristics of the unknown model parameters, e.g., smoothness or blockiness. First-order total variation regularization (TV) allows the reconstruction of well-defined edges and models exhibiting block-like characteristics. However, it is associated with the generation of undesirable staircase artifacts. This study applies a novel approach for removing staircase artifacts using a combined second-order non-convex total variation with overlapping group sparse regularizer. This regularizer aims to smooth out the staircase effect while still keeping the edges of the model. Moreover, the study applies the proposed method for the nonlinear seismic cross-hole tomography problems, where the goal is to reconstruct both smooth and blocky features of the model and avoid staircase artifacts of the TV regularization. The numerical examples indicate the efficiency of the proposed regularization method.

Keywords: Ill-posed problem, nonlinear travel-time tomography, total variation regularization, overlapping group sparse regularizer

[*]Corresponding author: mariahi@ut.ac.ir

1 Introduction

Seismic travel-time tomography is an imaging technique that uses seismic data to reconstruct physical subsurface parameters such as seismic velocity and attenuation coefficient. It is formulated as an inverse problem, to determine subsurface parameters based on the arrival time or amplitude information extracted from seismic waveform recordings (Wang and Rao, 2020). Regularization is critical in travel-time tomography, particularly when dealing with complex subsurface structures of a discontinuous nature such as faults or fractures. The regularization technique aids in balancing the trade-off between fitting the observed data and keeping the model in a desired shape (e.g., smooth or blocky), lowers the impact of noise in the inversion process and prevents overfitting. Moreover, it improves the quality of the estimated models (Lin et al., 2015; Jiang and Zhang, 2018).

One of the most effective regularization techniques in image processing is total variation (TV) regularization, as proposed by Rudin and Osher (1994). This particular approach is favoured due to its ability to effectively preserve the crucial sharp edges of the models throughout the reconstruction procedure. The TV regularization is appropriate for models that exhibit piecewise constant (blocky) features and is implemented using the L1-norm. In this scenario, the TV function encourages sparsity in a particular domain, such as the curvelet domain or the domain of the model gradient. Despite the success of TV regularization in inverse problems, its utilization is susceptible to a specific issue referred to as the staircase artifact, which manifests as visually displeasing blocky patterns in the resultant model (Xu et al., 2019). The use of TV regularization for seismic tomography problems suffers from this drawback, and proposing an efficient TV-based approach is the objective of this study.

Although higher-order techniques are found to help decrease staircase artifacts in the model, they often suffer from the drawback of causing blurring of model edges and structures (Xu et al., 2014). Hybrid methodologies that integrate higher-order TV by utilizing high-order derivatives and TV regularization (first-order derivative) have been proposed to address this problem. The total generalized variation (TGV) model, as presented by Bredies et al. (2010), and the combined model proposed by Papafitsoros and Schönlieb (2014) serve as two illustrative instances of such methodologies. The TGV technique effectively manages the trade-off between first-order and second-order total variation (TV) in the vicinity of model edges. In contrast, the hybrid approach proposed by Papafitsoros and Schönlieb (2014) tackles edge preservation and the reduction of staircase artifacts separately by employing first-order and second-order TV, respectively.

Several research studies have presented favourable outcomes in mitigating staircase artifacts through the utilization of hybrid Tikhonov and TV approaches (Gholami and Hosseini, 2013; Gholami and Gazzola, 2022). In addition, there is a growing interest in the image processing industry towards non-smooth, non-convex methods that can preserve sharp edges in the model (Nikolova et al., 2010). The utilization of methods incorporating structural sparsity has also been explored to address staircase artifacts when employing TV regularization. Previous studies (Selesnick and Chen, 2013) have demonstrated the advantages of utilizing group sparsity techniques in this particular context. To mitigate the presence of staircase artifacts, Liu et al. (2015) established the concept of overlapping group sparse total variation (OGSTV). Additionally, Shi et al. (2016) proposed a hybrid approach that combines the overlapping group sparse technique with a hyper Laplacian prior.

Drawing upon previous studies in the field of image processing, and following

the study conducted by Adam and Paramesran (2020), we apply a novel approach that integrates non-convex higher-order total variation with an overlapping group sparse regularizer specifically for seismic travel-time tomography. This combination effectively mitigates the problem of staircase artifacts while simultaneously preserving the integrity of model edges. The non-convex regularization technique known as higher-order TV is specifically developed to reduce blocky artifacts in restored images while preserving edge features. In contrast, the use of the non-convex higher-order TV term may inadvertently exacerbate speckle distortions due to its tendency to perceive them as edges. To tackle this matter, the utilization of OGSTV regularization is employed to effectively mitigate and control the consequences arising from the non-convex higher-order TV regularization.

2 Method

The primary objective of the majority of geophysical inversion techniques is to derive quantitative inferences regarding the physical characteristics of the Earth's subsurface (referred to as the model space) based on a limited number of recorded measurements (representing a specific subset of the data space). From a mathematical perspective, the problem can be approached as the task of determining a solution to a system of equations:

$$\mathbf{g}(\mathbf{m}) = \mathbf{d} \tag{1}$$

where $\mathbf{g}: \mathbb{R}^{\boldsymbol{m}} \rightarrow \mathbb{R}^{\boldsymbol{n}}$ governs the physics of the problem and is a function mapping from the set of real numbers m-dimensional space to another set. Depending on the problem at hand, $\mathbf{g}$ might be either linear or non-linear. The primary function of this entity is to establish a correspondence between the model space and the data space. The vector $\mathbf{m} \in \mathbb{R}^{\boldsymbol{m}}$ represents the desired model parameter (slowness) in the model space, whereas the vector $\mathbf{d} \in \mathbb{R}^{\boldsymbol{n}}$ represents the vector of observation (first arrival time) obtained when the mapping operator is applied to the true model of the subsurface.

One can effectively incorporate regularization by leveraging the benefits of Occam's inversion in the context of nonlinear problems. The concept of Occam's inversion entails the application of local linearization to the situation at hand through the use of Taylor expansion, as described by Aster (2018):

$$\mathbf{g}(\mathbf{m^k} + \boldsymbol{\delta}\mathbf{m}) \approx \mathbf{g}(\mathbf{m^k}) + \mathrm{J}(\mathbf{m}^k)\boldsymbol{\delta}\mathbf{m}, \tag{2}$$

where Jacobian matrix, denoted as $\mathrm{J}(\mathbf{m}^k) = \frac{\partial \mathbf{g}(\mathbf{m})}{\partial \mathbf{m}}$ represents the derivative of the modeled data, $\mathbf{g}(\mathbf{m})$, concerning the model parameter $\mathbf{m}$. This matrix quantifies the sensitivity of the modeled data $\mathbf{g}(\mathbf{m})$ to changes in the model parameters. $\boldsymbol{\delta}\mathbf{m}$ represents the model perturbation. By inserting Eq. (2) into Eq. (1), we obtain:

$$\mathrm{J}(\mathbf{m}^k)\delta\mathbf{m} = \mathbf{d} - \mathbf{g}(\mathbf{m^k}). \tag{3}$$

Presently, we are faced with a linear equation about the variable $\delta\mathbf{m}$. It should be noted that $\delta\mathbf{m} = \mathbf{m}^{k+1} - \mathbf{m}^{\mathrm{k}}$, and we have:

$$\mathrm{J}(\mathbf{m}^k)\,(\mathbf{m^{k+1}} - \mathbf{m^k}) = \mathbf{d} - \mathbf{g}(\mathbf{m^k}), \tag{4-1}$$

$$\mathrm{J}(\mathbf{m}^k)\,(\mathbf{m^{k+1}}) = \underbrace{\mathbf{d} - \mathbf{g}(\mathbf{m}^k) + \mathrm{J}(\mathbf{m}^k)\,(\mathbf{m}^k)}_{\hat{\boldsymbol{d}}}. \tag{4-2}$$

We can see that $\mathbf{J}(\mathbf{m}^k)\,\mathbf{m^{k+1}} = \hat{\mathbf{d}}$ is linear concerning $\mathbf{m}^{k+1}$. In what follows, we use $\mathbf{L} \equiv \mathbf{J}(\mathbf{m}^k)$ to have a consistent notation with linear inverse problem literature.

2-1 Non-linear second-order total variation overlapping group sparse total variation regularization

To formulate the proposed approach for nonlinear seismic tomography using a combination of non-linear second-order total variation and group sparsity, we introduce the following optimization problem (Adam and Paramesran, 2020):

$$\min_{\mathbf{m}} \frac{1}{2}\left\|\mathbf{L}\mathbf{m} - \hat{\mathbf{d}}\right\|_2^2 + \phi(\nabla\mathbf{m}) + \gamma\|\nabla^2\mathbf{m}\|_p^p + \mathcal{I}_C(\mathbf{m}). \tag{5}$$

In this context, the expressions $\phi(\mathbf{m})$ and $\|.\|_p^p$ represent the regularization function for overlapping group sparsity, and the

non-convex ℓp norm, respectively. The variable $\gamma > 0$, serves as parameters for balancing the data fidelity term and the non-convex second-order term. Additionally, the function $\mathcal{I}_C(\mathbf{m})$ functions as a characteristic (indicator) function, enabling us to enforce strict constraints within the objective function. Specifically, it assigns a value of 0 to $\mathbf{m}$ if it belongs to set C, and assigns an infinitely large value if $\mathbf{m}$ is not in set C. Symbols ∇ and ∇^2 represent the first-order difference operator and the second-order difference operator, respectively. The alternating direction method of multipliers (ADMM) (Boyd et al., 2010) can be used to minimize problem (5). By introducing the auxiliary variables $\mathbf{p} = \nabla\mathbf{m}, \mathbf{q} = \nabla^2\mathbf{m}, \boldsymbol{z} = \mathbf{m}$, problem (5) can be recast as the following constraint optimization:

$$\min_{m} \frac{1}{2}\left\|\mathbf{L}\mathbf{m} - \hat{\mathbf{d}}\right\|_2^2 + \phi(\mathbf{p}) + \gamma\|\mathbf{q}\|_p^p + I_c(\mathbf{z}), \quad subject\ to.\ \ \mathbf{p} = \nabla\mathbf{m}, \mathbf{q} = \nabla^2\mathbf{m}, \mathbf{z} = \mathbf{m} \tag{6}$$

The constraint optimization defined in (6) can be replaced with the corresponding augmented Lagrangian function as:

$$\begin{aligned} L = \frac{1}{2}\left\|\mathbf{L}\mathbf{m} - \hat{\mathbf{d}}\right\|_2^2 + \phi(\mathbf{p}) + \gamma\|\mathbf{q}\|_p^p + I_c(\mathbf{z}) \\ -\boldsymbol{\lambda}_1^T(\mathbf{p} - \nabla\mathbf{m}) + \frac{\alpha_1}{2}\|\mathbf{p} - \nabla\mathbf{m}\|_2^2 \\ - \boldsymbol{\lambda}_2^T(\mathbf{q} - \nabla^2\mathbf{m}) \\ + \frac{\alpha_2}{2}\|\mathbf{q} - \nabla^2\mathbf{m}\|_2^2 \\ - \boldsymbol{\lambda}_3^T(\mathbf{z} - \mathbf{m}) \\ + \frac{\alpha_3}{2}\|\mathbf{z} - \mathbf{m}\|_2^2, \end{aligned} \tag{7}$$

where $\alpha_1, \alpha_2, \alpha_3 > 0$ are the penalty parameters associated with penalty terms, and $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3$ are the vector of Lagrange multipliers. The ADMM approach to solve Eq. (7) leads to the following iterations (Boyd et al., 2010):

$$\mathbf{m}^{k+1} = \min_{\mathbf{m}} L(\mathbf{m}, \mathbf{p}^k, \mathbf{q}^k, \boldsymbol{z}^k, \boldsymbol{\lambda}_1^k, \boldsymbol{\lambda}_2^k, \boldsymbol{\lambda}_3^k) \tag{8-1}$$

$$\mathbf{p}^{k+1} = \min_{\boldsymbol{p}} L(\mathbf{m}^{k+1}, \mathbf{p}, \mathbf{q}^k, \boldsymbol{z}^k, \boldsymbol{\lambda}_1^k, \boldsymbol{\lambda}_2^k, \boldsymbol{\lambda}_3^k) \tag{8-2}$$

$$\mathbf{q}^{k+1} = \min_{\boldsymbol{q}} L(\mathbf{m}^{k+1}, \mathbf{p}^{k+1}, \mathbf{q}, \boldsymbol{z}^k, \boldsymbol{\lambda}_1^k, \boldsymbol{\lambda}_2^k, \boldsymbol{\lambda}_3^k) \tag{8-3}$$

$$\boldsymbol{z}^{k+1} = \min_{\boldsymbol{z}} L(\mathbf{m}^{k+1}, \mathbf{p}^{k+1}, \mathbf{q}^{k+1}, \boldsymbol{z}, \boldsymbol{\lambda}_1^k, \boldsymbol{\lambda}_2^k, \boldsymbol{\lambda}_3^k) \tag{8-4}$$

$$\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_1 + \alpha_1(\mathbf{p}^{k+1} - \nabla\mathbf{m}^{k+1}) \tag{8-5}$$

$$\boldsymbol{\lambda}_2 = \boldsymbol{\lambda}_2 + \alpha_2(\nabla^2\mathbf{m}^{k+1} - \mathbf{q}^{k+1}) \tag{8-6}$$

$$\boldsymbol{\lambda}_3 = \boldsymbol{\lambda}_3 + \alpha_3\left(\mathbf{m}^{k+1} - \boldsymbol{z}^{k+1}\right). \tag{8-7}$$

2-2 Optimization over m

The subproblem over $\mathbf{m}$ [Eq. (8-1)] is quadratic which leads to the following least squares optimization:

$$\begin{aligned} \mathbf{m}^{k+1} = \min_{\mathbf{m}} \frac{\lambda}{2}\|\mathbf{L}\mathbf{m} - \hat{\mathbf{d}}\|_2^2 - \boldsymbol{\lambda}_1^\top(\mathbf{p}^k - \nabla\mathbf{m}) \\ + \frac{\alpha_1}{2}\|\mathbf{p}^k - \nabla\mathbf{m}\|_2^2 - \boldsymbol{\lambda}_2^\top(\mathbf{q}^k - \nabla^2\mathbf{m}) \\ + \frac{\alpha_2}{2}\|\mathbf{q}^k - \nabla^2\mathbf{m}\|_2^2 - \boldsymbol{\lambda}_3^\top(\mathbf{z}^k - \mathbf{m}) + \frac{\alpha_3}{2}\|\mathbf{z}^k - \mathbf{m}\|_2^2. \end{aligned} \tag{9}$$

The above minimization problem has a closed-form solution, which is the solution of the following linear system of equations:

$$\begin{aligned} (\mathbf{L}^\top\mathbf{L} + \alpha_1\nabla^\top\nabla + \alpha_2(\nabla^2)^\top\nabla^2 + \alpha_3\mathbf{I})\mathbf{m}^{k+1} = \\ \left(\mathbf{L}^\top\hat{\mathbf{d}} - \nabla^\top\boldsymbol{\lambda}_1^k + \alpha_1\nabla^\top\mathbf{p}^k - (\nabla^2)^\top\boldsymbol{\lambda}_2^k \right. \\ \left. + \alpha_2(\nabla^2)^\top\mathbf{q}^k - \boldsymbol{\lambda}_3^k + \alpha_3\mathbf{z}^k\right). \end{aligned} \tag{10}$$

2-3 Optimization over p

The subproblem over $\mathbf{p}$ [Eq. (8-2)] is an overlapping group sparse total variation (OGSTV) denoising problem (Liu et al., 2015), and has the following form:

$$\begin{aligned} \mathbf{p}^{k+1} = \min_{\mathbf{p}} \frac{\alpha_1}{2}\|\mathbf{p} - \nabla\mathbf{m}^{k+1}\|_2^2 - \boldsymbol{\lambda}_1^\top(\mathbf{p} - \nabla\mathbf{m}^{k+1}) \\ + \phi(\mathbf{p}), \end{aligned} \tag{11}$$

It can be rewritten as follows:

$$\mathbf{p}^{k+1} = \min_{\mathbf{p}} \frac{\alpha_1}{2}\left\|\mathbf{p} - \left(\nabla\mathbf{m}^{k+1} + \frac{\boldsymbol{\lambda}_1^k}{\alpha_1}\right)\right\|_2^2 + \phi(\mathbf{p}) \tag{12}$$

The GSTV regularization function, $\phi(\mathbf{p})$, is defined as:

$$\phi(\mathbf{p}) = \sum_i \left[\sum_{j=0}^{K-1} |p(i+j)|^2\right]^{\frac{1}{2}}. \quad (13)$$

Here, K represents the group's size, which can be thought of as a contiguous window of size K beginning at index i (see e.g., Chen and Selesnick, 2014). Following Adam and Paramesran (2020), we employ the majorization-minimization (MM) strategy to address the OGSTV problem (12). The MM approach involves bounding $\phi(\mathbf{p})$ using a quadratic surrogate function during each iteration of the algorithm designed to solve Eq. (12). Consequently, in each step of the MM algorithm, we tackle a convex quadratic problem, which ultimately leads to a convergent solution for Eq. (12).

2-4 Optimization over q

The optimization over $\mathbf{q}$ [Eq. (8-3)] is a minimization task of the following form:

$$\mathbf{q}^{k+1} = \min_{\mathbf{q}} \frac{\alpha_2}{2} \| \mathbf{q} - \nabla^2 \mathbf{m}^{k+1} \|_2^2 - \boldsymbol{\lambda}_2^{\top} (\mathbf{q} - \nabla^2 \mathbf{m}^{k+1}) + \gamma \| \mathbf{q} \|_p^p, \quad (14)$$

which is a problem involving second-order total variation denoising, and is not convex in nature. It can be reformulated as:

$$\mathbf{q}^{k+1} = \min_{\mathbf{q}} \frac{\alpha_2}{2} \left\| \mathbf{q} - \left(\nabla^2 \mathbf{m}^{k+1} + \frac{\boldsymbol{\lambda}_2^k}{\alpha_2}\right) \right\|_2^2 + \gamma \| \mathbf{q} \|_p^p. \quad (15)$$

The lack of convexity of the problem arises from the ℓp norm, with a value between 0 and 1. To minimize problem (15), the iterative re-weighted ℓ1 (IRL1) algorithm can be utilized (Aster, 2018). During every step of the IRL1 algorithm, problem (15) is estimated as a weighted ℓ1 norm problem with the following structure:

$$\mathbf{q}^{k+1} = \arg\min_{\mathbf{q}} \frac{\alpha_2}{2} \left\| \mathbf{q} - \left(\nabla^2 \mathbf{m}^{k+1} + \frac{\boldsymbol{\lambda}_2^k}{\alpha_2}\right) \right\|_2^2 + \sum_i w_i |\mathrm{q}_i|, \quad (16)$$

where

$$w_i = \gamma \frac{p}{(|\mathrm{q}_i^k| + \epsilon)^{1-p}}. \quad (17)$$

Here, $\epsilon > 0$ is a small value. Now, the proximity operator can be used to solve Eq. (16):

$$\mathbf{q}^{k+1} = \mathrm{shrink}\left(\nabla^2 \mathbf{m}^{k+1} + \boldsymbol{\lambda}_2^k/\alpha_2, \frac{w_i \gamma}{\alpha_2}\right),$$

$$= max\left\{\left|\nabla^2 \mathbf{m}^{k+1} + \boldsymbol{\lambda}_2^k/\alpha_2\right| - \frac{w_i \gamma}{\alpha_2}, 0\right\} \cdot \mathrm{sgn}\,(\nabla^2 \mathbf{m}^{k+1} + \boldsymbol{\lambda}_2^k/\alpha_2), \quad (18)$$

where $\mathrm{sgn}(\cdot)$ is a sign function.

2-5 Optimization over z

The optimization over z [Eq. (8-4)] can be rewritten as:

$$\boldsymbol{z}^{k+1} = \min_{\mathbf{z}} \frac{\alpha_3}{2} \left\| \mathbf{z} - \left(\mathbf{m}^{k+1} + \frac{\boldsymbol{\lambda}_3^k}{\alpha_3}\right) \right\|_2^2 + \mathcal{I}_C(\mathbf{z}), \quad (19)$$

That is a projection problem as:

$$\boldsymbol{z}^{k+1} = \mathrm{proj}_C \left(\mathcal{I}_C{}^{k+1} + \frac{\boldsymbol{\lambda}_3^k}{\alpha_3}\right), = min\left(\mathbf{m}^{*\,max}, max\left(\mathcal{I}_C{}^{k+1} + \frac{\boldsymbol{\lambda}_3^k}{\alpha_3}, \mathbf{m}^{*\,min}\right)\right), \quad (20)$$

where $\mathbf{m}^{*\,min}$ and $\mathbf{m}^{*\,max}$ are the predefined minimum and maximum bound of the desired model according to the prior information.

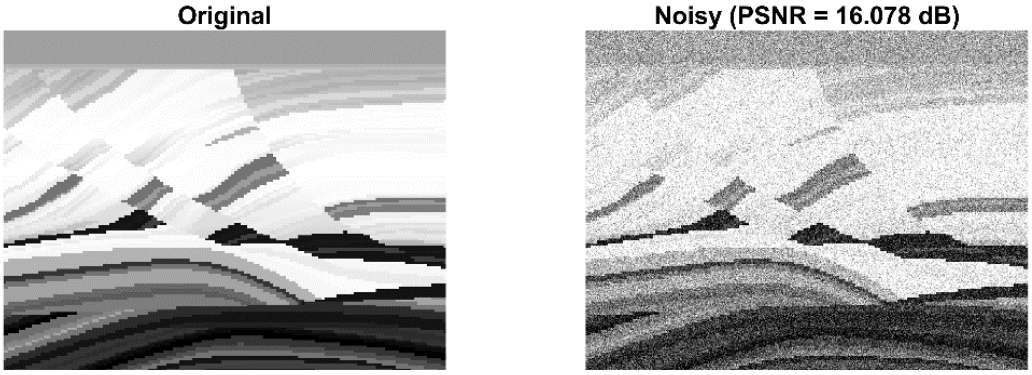
The subproblem over the Lagrange multipliers are solved by the gradient ascent approach according to Eqs. (8-5) to (8-7).

3 Numerical examples

3-1 Example 1: Choosing the optimal group size based on an image denoising problem

The determination of the group size K is a crucial component in the OGS-TV approach. The group size of one is equivalent to the usual TV method. Opting for larger values has the potential to increase CPU time. To comprehend the influence of group size, an experiment is undertaken

on an image-denoising problem. Specifically, a portion of the Marmousi model is sliced for analysis (Fig. 1a). In this seismic-like portion, Gaussian-distributed random noise with a peak signal-to-noise ratio (PSNR) of approximately 16 dB is added to the original image (Fig. 1b).



(a) (b)

Figure 1. Denoising seismic section. (a) Original seismic section. (b) Gaussian noise contaminated section with PSNR of 16 dB.

The suggested technique is not sensitive to the choice of penalty parameters $\alpha_1, \alpha_2, \alpha_3$, owing to the inclusion of the Lagrange multipliers. The values for all examples were picked based on a process of trial and error. The identical methodology applies to the variable γ. The method is executed with varying group sizes ranging from 2 to 7. The denoising results after 400 iterations are depicted in Fig. 2. The improvement in the quality of the denoised portions, as measured by the labeled PSNR value, is evident with an increase in the group size. This is confirmed by the calculated relative model error during iterations that is shown in Fig. 3. It can be seen that a higher group size leads to a higher convergence rate. Nevertheless, there is also an observable increase in CPU time. Consequently, a group size of 5 was chosen as the ideal value for the nonlinear travel-time tomography problem, as discussed in the subsequent section.

3-2 Example 2: Nonlinear travel-time tomography

In this section, a comparative study will be conducted between the proposed method and the traditional TV regularization solved by the iterative reweighted least squares (IRLS) method. The objective of this study is to evaluate the performance and efficacy of both methodologies using a simulated numerical illustration.

The primary emphasis of this study is on nonlinear seismic tomography, yet it is worth noting that the technique employed in this research has the potential to be utilized in several other seismic inversion applications. We deploy a configuration consisting of 14 seismic sources placed within one borehole, along with 14 receivers located in a separate borehole. The ground truth model for this synthetic example is shown in Fig. 4a, which includes a sharp high-velocity anomaly embedded in a smooth background model. It can be seen that the model contains both smooth and

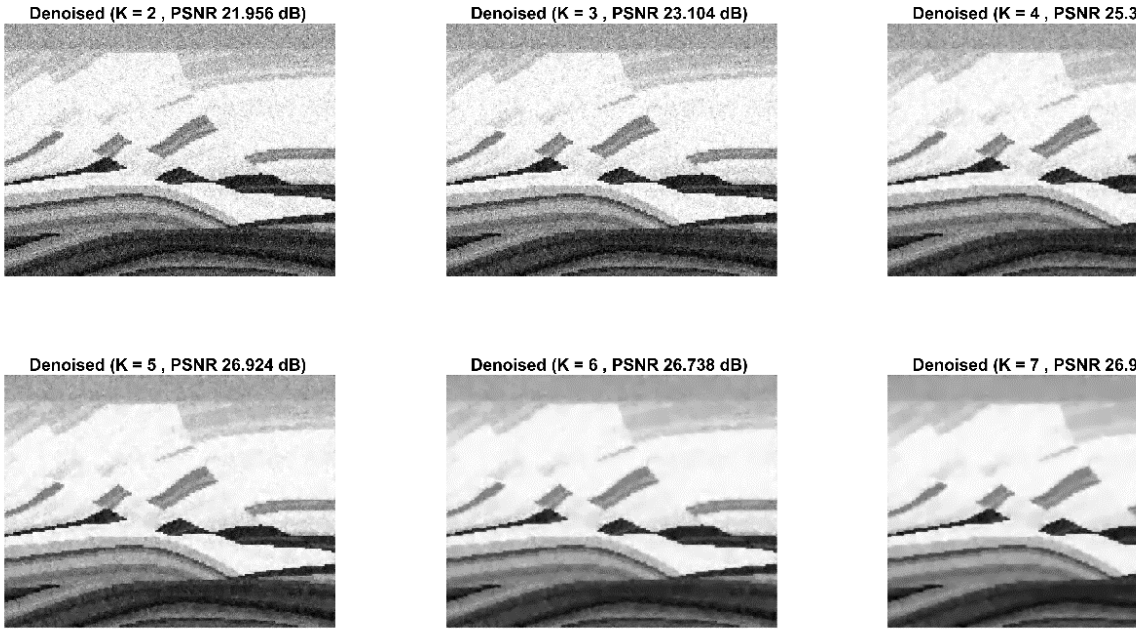


Figure 2. Denoising seismic section. Denoised sections using different group sizes are labelled by K.

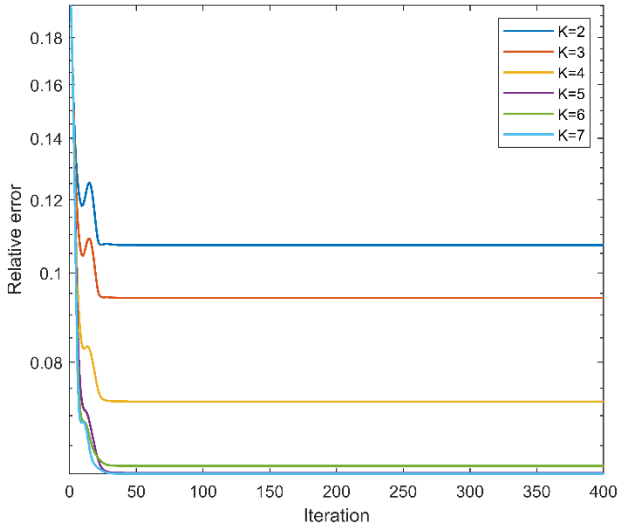


Figure 3. The relative model error versus iteration for different values of group size.

3-2 Example 2: Nonlinear travel-time tomography

In this section, a comparative study will be conducted between the proposed method and the traditional TV regularization solved by the iterative reweighted least squares (IRLS) method. The objective of this study is to evaluate the performance and efficacy of both methodologies using a simulated numerical illustration.

The primary emphasis of this study is on nonlinear seismic tomography, yet it is worth noting that the technique employed in this research has the potential to be utilized in several other seismic inversion applications. We deploy a configuration consisting of 14 seismic sources placed within one borehole, along with 14 receivers located in a separate borehole. The ground truth model for this synthetic example is

shown in Fig. 4a, which includes a sharp high-velocity anomaly embedded in a smooth background model. It can be seen that the model contains both smooth and blocky features as demonstrated by a vertical velocity profile extracted at X=100 m (Fig. 4b). The Fast Marching Method (FMM) is employed for the computation of seismic data, which is a crucial component of our research. This method, as described by Sethian (1996, 1999a,b) and Sethian and Popovici (1999), is utilized for forward modeling purposes. Fig. 4 depicts the accurate representation of the model alongside a vertical profile log that is extracted from the central portion of the model. The present model has a distinct geometric irregularity that is juxtaposed against a uniform and smooth background, rendering it well-suited for the examination and evaluation of various regularization techniques. The original model functions as a point of reference or ground truth. Subsequently, the collecting of data between boreholes is conducted through the utilization of a forward modeling engine, yielding first arrival time data.

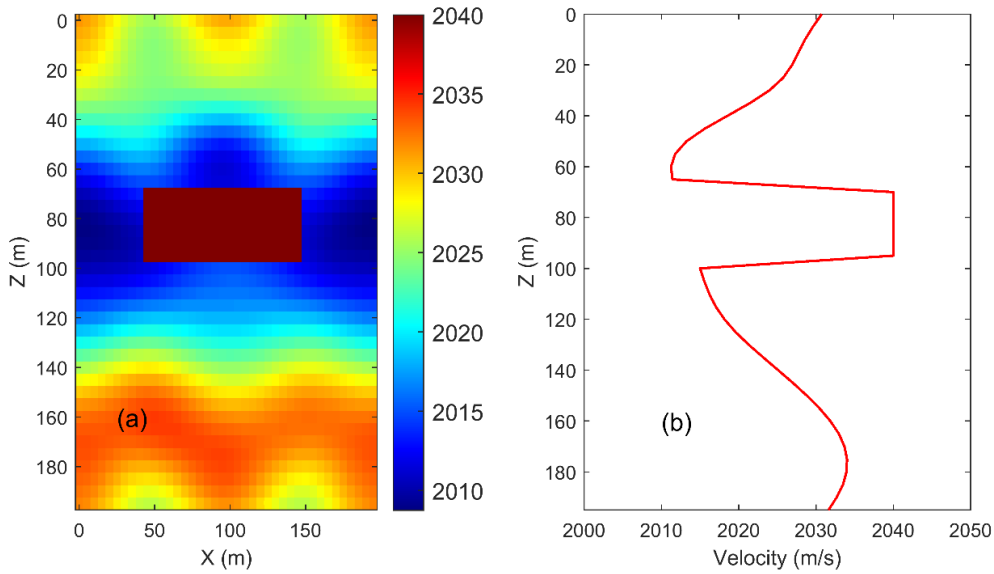


Figure 4. (a) True velocity model. (b) Vertical velocity profile extracted at X = 100 m.

Fig. 5a depicts the ray tracing overlaid onto the model, providing a visual representation of the ray paths (from source to receiver) within the geological structure. Meanwhile, Fig. 5b showcases the results of the first arrival time (as observed data), offering insights into the propagation of seismic waves within the subsurface. To simulate real-world conditions, we add random noise to data with an SNR of 10 dB. This allows us to assess the robustness of our methodology in the presence of noise. Fig. 6 illustrates the comparison between noise-free data and data with noise.

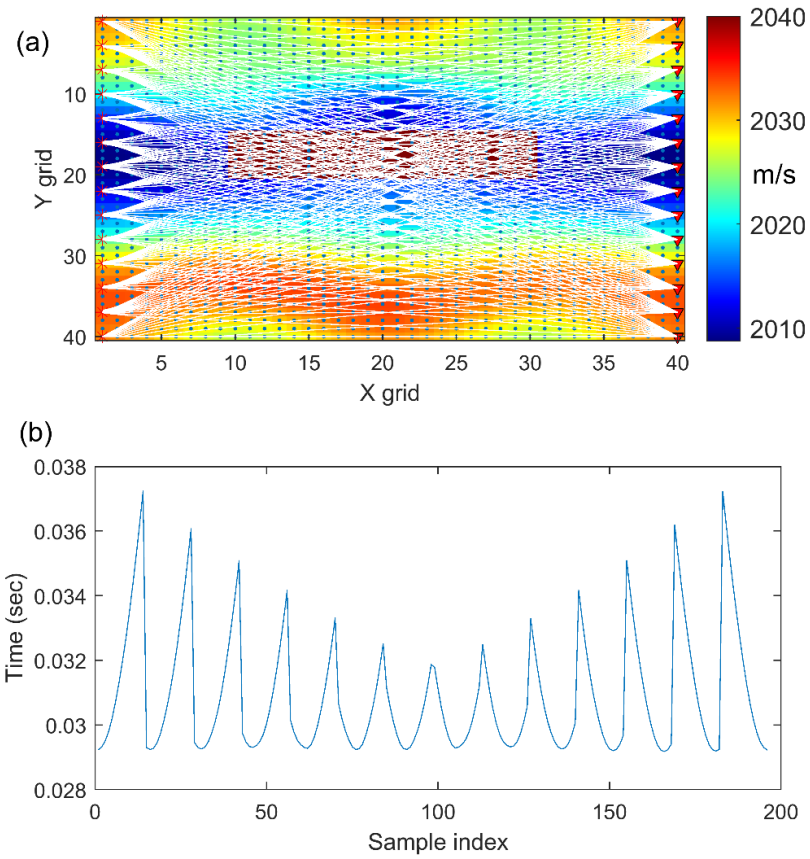


Figure 5. (a) Ray paths from source to receiver overlaid onto the model. (b) First arrival time recorded by receivers.

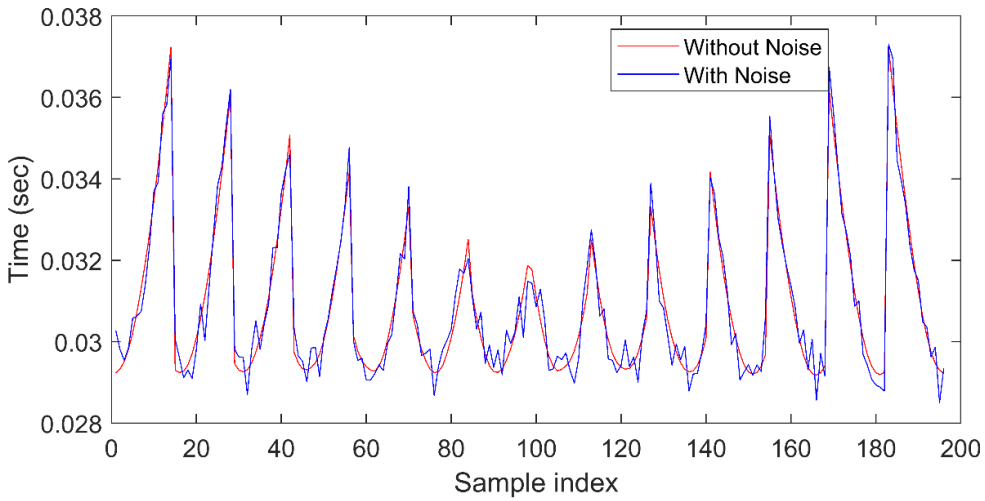


Figure 6. Comparison between noise-free data (red) and noise-contaminated data (blue).

Given the observed data, the next step is the inversion process. In doing so, two kinds of regularization schemes are incorporated into the optimization procedure. The first one is TV regularization, which is solved by the iteratively reweighted least squares (IRLS) method (Aster, 2018). The second one is the proposed

method based on the OGS method, which is combined with a hybrid non-convex second-order total variation. For each method, the inversion is performed through 20 iterations. Fig. 7 displays the inversion results obtained using both the proposed method and a conventional TV regularization method. It can be seen that the proposed method (Fig. 7b) outperforms conventional TV regularization (Fig. 7a), and is successful in capturing both smooth background and blocky anomalies. For a better comparative view, the difference between the true model (Fig. 4a) and the estimated model is shown in Fig. 7c (for conventional TV) and Fig. 7d (for the proposed OGS method). For the sake of a comprehensive assessment of the effectiveness and performance of our proposed approach in contrast to the conventional TV, some vertical profiles at different locations of the true model, initial model, and estimated models (shown in Figs. 8a,b) are extracted. The extracted velocity profiles are shown in Fig. 8. Obviously, the proposed method is successful in reconstructing both smooth features (left and right panels) and blocky features embedded in a smooth background (middle panel).

During the iterations, the data misfit error as well as relative model error are computed as:

Data misfit: $\|\mathbf{g}(\mathbf{m}^k) - \mathbf{d}\|_2$

Relative model error: $\dfrac{\|\mathbf{m}^k - \mathbf{m}^{true}\|_2}{\|\mathbf{m}^{true}\|_2}$

The computed error curves are shown in Fig. 9. It can be seen that after some iterations, the conventional TV regularization (red curves) is unable to reduce both data and model errors. In contrast, the proposed method (blue curve) is successful in reducing the mentioned errors.

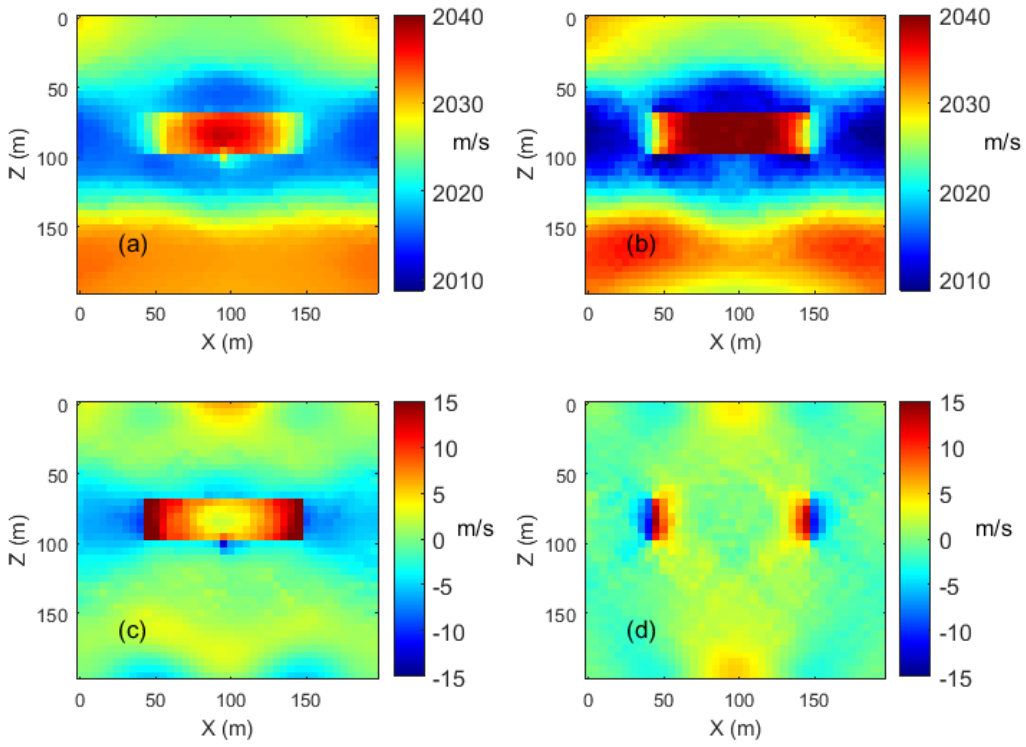


Figure 7. Nonlinear travel-time tomography inversion. The inversion results obtained using (a) conventional TV regularization, and (b) the proposed regularization method. (c) The difference between the true model and the estimated model using conventional TV regularization and (d) the proposed regularization.

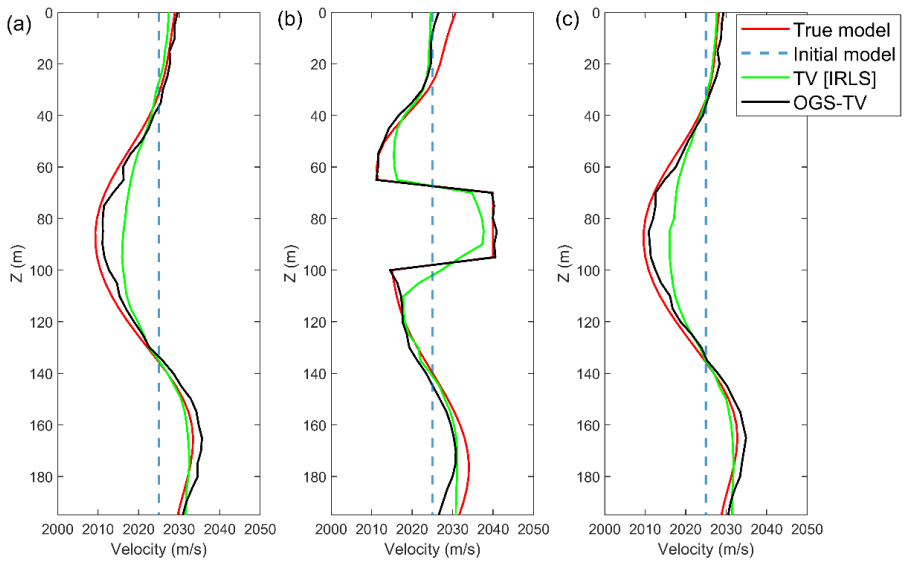


Figure 8. Comparison of extracted vertical profiles at (from left to right) X = 25, 100, 175 m.

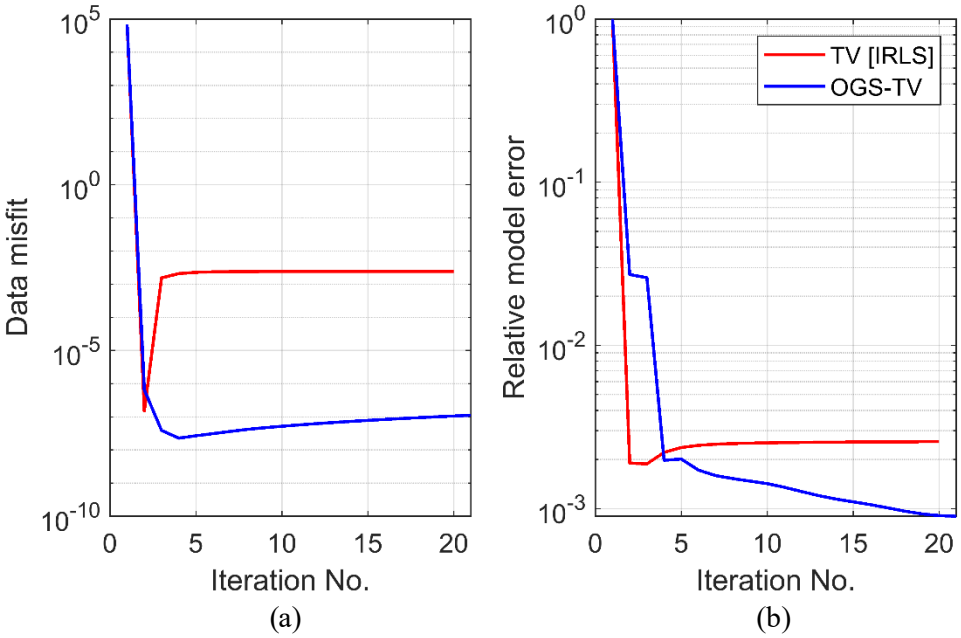


Figure 9. (a) Data misfit versus iteration. (b) Relative model error versus iteration.

The difference between noisy data (observed) and data computed from estimated models (Figs. 7a,b) is illustrated in Fig. 10. In comparison to the noise-free data (Fig. 5b), the two methods can fit the data well. However, from Fig. 9, still quantitatively, the proposed method outperforms TV regularization in fitting the data.

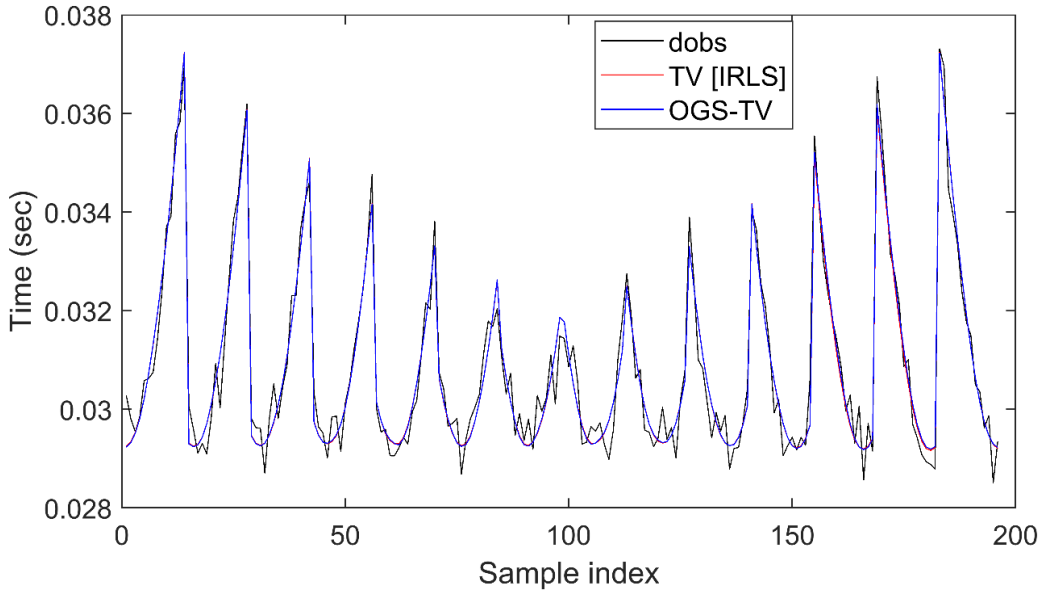


Figure 10. Comparison between noisy data (observed) and data computed from estimated models (Figs. 7a,b).

.

4 Conclusion

To address the problem of staircase artifacts in conventional total variation (TV) regularization applied to nonlinear travel-time tomography, we have employed a methodology that combines the benefits of non-convex higher-order total variation and overlapping group sparsity. This approach is necessary for regularization in seismic tomography inversion problems. Alternating direction methodology was used to efficiently tackle the proposed regularization approach. The current study utilized a synthetic model to aid in the rebuilding of the model during the comparative analysis. The findings indicate that the proposed strategy demonstrated improved performance in estimating both smooth and blocky features when compared to the conventional total variation (TV) regularization method.

5 Data and sharing resources

The Marmousi model used in this study was taken from the database of the University of Huston and can be downloaded at this address: http://www.agl.uh.edu/downloads/downloads.htm.

References

Adam, T., and Paramesran, R., 2020, Hybrid non-convex second-order total variation with applications to non-blind image deblurring: Signal, Image and Video Processing, **14**(1), 115-123.

Aster, R. C., Borchers, B., and Thurber, C. H., 2018, Parameter Estimation and Inverse Problems: Elsevier.

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J., 2010, Distributed optimization and statistical learning via the alternating direction method of multipliers: Foundations and Trends® in Machine Learning, **3**(1), 1-122.

Bredies, K., Kunisch, K., and Pock, T., 2010, Total generalized variation: SIAM Journal on Imaging Sciences, **3**(3), 492-526.

Chen, P. Y., and Selesnick, I. W., 2014, Translation-invariant shrinkage/thresholding of group sparse signals: Signal Processing, **94**, 476-489.

Du, H., and Liu, Y., 2018, Minmax-concave total variation denoising: Signal, Image and Video Processing, **12**, 1027-1034.

Gholami, A., and Gazzola, S., 2022, Automatic balancing parameter selection

for Tikhonov-TV regularization: BIT Numerical Mathematics, **62**(4), 1873-1898.

Gholami, A., and Hosseini, S. M., 2013, A balanced combination of Tikhonov and total variation regularizations for reconstruction of piecewise-smooth signals: Signal Processing, **93**(7), 1945-1960.

Guo, W., Qin, J., and Yin, W., 2014, A new detail-preserving regularization scheme: SIAM Journal on Imaging Sciences, **7**(2), 1309-1334.

Jiang, W., and Zhang, J., 2018, 3D first-arrival traveltime tomography with modified total variation regularization: Journal of Geophysics and Engineering, **15**(1), 207-223.

Lin, Y., Syracuse, E. M., Maceira, M., Zhang, H., and Larmat, C., 2015, Double-difference traveltime tomography with edge-preserving regularization and a priori interfaces: Geophysical Journal International, **201**(2), 574-594.

Liu, J., Huang, T. Z., Selesnick, I. W., Lv, X. G., and Chen, P. Y., 2015, Image restoration using total variation with overlapping group sparsity: Information Sciences, **295**, 232-246.

Lysaker, M., Lundervold, A., and Tai, X. C., 2003, Noise removal using a fourth-order partial differential equation with applications to medical magnetic resonance images in space and time: IEEE Transactions on Image Processing, **12**(12), 1579-1590.

Nikolova, M., Ng, M. K., and Tam, C. P., 2010, Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction: IEEE Transactions on Image Processing, **19**(12), 3073-3088.

Papafitsoros, K., and Schönlieb, C. B., 2014, A combined first and second-order variational approach for image reconstruction: Journal of Mathematical Imaging and Vision, **48**, 308-338.

Rudin, L. I., and Osher, S., 1994, Total variation-based image restoration with free local constraints: Proceedings of 1st International Conference on Image Processing, IEEE, **1**, 31-35.

Selesnick, I. W., and Chen, P. Y., 2013, Total variation denoising with overlapping group sparsity: IEEE International Conference on Acoustics, Speech and Signal Processing, 5696-5700.

Shi, M., Han, T., and Liu, S., 2016, Total variation image restoration using hyper-Laplacian prior with overlapping group sparsity: Signal Processing, **126**, 65-76.

Sethian, J., 1996, A fast marching level set method for monotonically advancing front: Proceeding of the National Academy of Sciences, **93**(4), 1591-1595.

Sethian, J., 1999a, Fast Marching Methods: SIAM REVIEW, **41**(2), 199-235.

Sethian, J., 1999b, Level Set Methods and Fast Marching Methods, Evolving Interfaces in Computational Geometery, Fluid Mechanics, Computer Vision, and Materials Science: Cambridge University Press.

Sethian, J. A., and Popovici, M., 1999, 3-D traveltime computation using the fast marching method: Geophysics, **64**, 516-523, https://doi.org/10.1190/1.1444558.

Wang, Y., and Rao, Y., 2020, Seismic, waveform modeling and tomography, in Encyclopedia of Solid Earth Geophysics (2nd edition): Springer, 1-15.

Xu, J., Feng, X., Hao, Y., and Han, Y., 2014, Image decomposition using adaptive second-order total generalized variation: Signal, Image and Video Processing, **8**, 39-47.

Xu, J., Hao, Y., Li, M., and Zhang, X., 2019, A novel variational model for image decomposition: Signal, Image and Video Processing, **13**, 967-974.

You, Y. L., and Kaveh, M., 2000, Fourth-order partial differential equations for noise removal: IEEE Transactions on Image Processing, **9**(10), 1723-1730.