Combined Non-convex Second-order Total Variation with Overlapping Group Sparsity for Full Waveform Inversion

Hongsun fu (fuhongsun@dlmu.edu.cn)
Dalian Maritime University

Hongyu Qi
Harbin Institute of Technology

Ruixue Gu
Dalian Maritime University

Research Article

Keywords: Full waveform inversion, Non-convex second-order total variation, Overlapping group sparsity, Alternating direction multiplier method

Posted Date: October 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2142256/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Combined Non-convex Second-order Total Variation with Overlapping Group Sparsity for Full Waveform Inversion

Hongsun Fu\(^1\)*, Hongyu Qi\(^2\)† and Ruixue Gu\(^1\)†

\(^1\)*School of Science, Dalian Maritime University, Dalian, 116026, China.
\(^2\)Department of Mathematics, Harbin Institute of Technology, Harbin, 150001, China.

*Corresponding author(s). E-mail(s): fuhongsun@dlmu.edu.cn;
Contributing authors: 756106217@qq.com;
ruixue.gu@dlmu.edu.cn;
†These authors contributed equally to this work.

Abstract

Full Waveform Inversion (FWI) can provide an accurate velocity model by matching observed and simulated seismograms. Mathematically, FWI is a highly ill-posed inverse problem that the inversion results lack dependence on the observed data. For a stable and reasonable inversion, proper regularization methods have to be taken into account. We propose a novel composite regularization for frequency-domain FWI problem, which uses a non-convex second-order total variation (TV) term and an overlapping group sparse TV (OGS-TV) regularization term. Compared with the conventional TV regularization, our method has better accuracy and robustness, and could effectively make use of the structural sparsity in the velocity model. Furthermore, the alternating direction multiplier algorithm with the adaptive selection of the penalty parameters is developed to solve this composite constraint problem, which can improve the stability of the FWI process. To illustrate the superior performance of the proposed FWI method both visually and quantitatively, we present several numerical examples through the comparison between our FWI method and conventional FWI method with TV regularization.

Keywords: Full waveform inversion, Non-convex second-order total variation, Overlapping group sparsity, Alternating direction multiplier method
1 Introduction

Full waveform inversion (FWI) is a promising technique that uses full waveforms acquired from the ground or wells to image complex subsurface structures\cite{1}. The FWI problems are, however, often highly nonlinear and ill-posed, meaning that there can be multiple model parameters consistent with the observed data and/or model parameter estimation is highly sensitive to small changes in the data. Therefore, the regularization methods are necessary to stabilize the numerical computations and determine the high-resolution model structures \cite{2–4}.

An approximate stable solution can be obtained by using the meaningful prior or statistical information of the model \cite{5, 6}. The well-known total variation (TV) regularization has been broadly used in the FWI community because of its effectiveness in preserving sharp interfaces and spatial piecewise smoothness \cite{7–9}. Nonetheless, this regularization often suffers from staircase artifacts and the loss of structural details. So many variants of TV regularization techniques are developed. For example, Esser et al.\cite{10} proposed an extended FWI formulation to remove some of these artifacts while improving the inversion result significantly, taking advantage of a combination of the box constraints, TV-norm, and relaxed asymmetric TV constraints in succession. Yong et al.\cite{11} constructed an adaptive primal-dual hybrid gradient method, which integrates the TV-norm and box constraints into FWI, to mitigate the local minima and reconstruct the salt body effectively. Li et al.\cite{12} developed an effective sparsity-promoting TV regularization method for FWI that incorporates nonlocal similarity in the model, which can be seen as a generalization of TV regularization with adaptive dictionary learning. Recently, some researchers found that the second-order formula for the total generalized variation can generate superior results compared with the classical first-order TV\cite{13–15}. Gao et al.\cite{16} introduced a total generalized $p$-variation regularization scheme into acoustic and elastic-waveform inversion, and proposed an efficient iterative algorithm to implement the defined regularization scheme by using the split-Bregman strategy.

More recently, a new regularization scheme has been obtained by utilizing non-convex high-order TV and overlapping group sparsity (OGS) prior, which could effectively deal with image denoising/deblurring\cite{17}. The underlying idea of the grouping sparsity is the first-order derivative of the unknown model parameters is not only sparse, but also exhibits local dependence between adjacent derivatives. In general, large gradient values are near to the other large values rather than occur in isolation. One key point of this regularization scheme is thus to integrate the structure information and sparsity of the model gradient into one framework to substantially suppress the stair-step artifacts in the traditional TV regularization\cite{18}. In the context of FWI problems, OGS is utilized as a regularization term to constraint grouping or clustering behavior of a model parameter in its gradient domain\cite{19}. In this paper, an effective strategy is designed for solving the FWI problem that uses non-convex second-order TV regularization with OGS prior and the alternating direction method.
Combined regularization for Full Waveform Inversion

of multipliers (ADMM) algorithm. Some of the merits of this article are summarized as follows: (1) the second-order non-convex TV and TV with OGS (OGS-TV) prior are used as a composite regularization term to alleviate the ill-posed nature of FWI. This composite regularization scheme is beneficial to retain the effectiveness of TV and significantly suppress the unwanted artifacts on the reconstruction models to obtain high-resolution inversion results. (2) we apply the ADMM to our regularized objective function to derive an efficient FWI method. Then we introduce an adaptive parameter adjustment approach to guide the inversion towards a good direction during the ADMM iterations. (3) we focus on complex geological structures such as salt inversion and employ a multi-scale inversion strategy [20] from low frequency to high frequency in the experimental process. (4) Numerical experiments indicate that our FWI method has the advantage in retrieving the detailed velocity parameter. We call this method Non-convex Second-order Total Variation Overlapping Group Sparse FWI, or NSTVOGS-FWI method for short.

The remainder of the paper is as follows. Section 2 and Section 3 briefly introduce the mathematical expression of FWI and the details of the proposed NSTVOGS-FWI method. Experimental results are given in Section 4 to demonstrate the superiority of the proposed method. Finally, Section 5 presents concluding remarks.

2 Full-waveform inversion

FWI is a data fitting process, and the frequency-domain FWI has the following form

$$\arg \min_m \frac{1}{2} \| \mathbf{d}_{cal}(m) - \mathbf{d}_{obs} \|_2^2,$$  \hspace{1cm} (1)

where $\| \cdot \|_2$ is the $l^2$-norm, and $\mathbf{d}_{obs}$ denotes the observed wavefield data. The predicted data $\mathbf{d}_{cal}(m)$ is calculated by solving the Helmholtz equation with the boundary of the perfect matching layer (PML), and the model parameter is $m \in \mathbb{R}^{N_z \times N_x}$ in seconds$^2$/meters$^2$ (the squared slowness), where $N_z$ and $N_x$ express the number of grid points in $z$ and $x$ directions, respectively. Here, $\mathbf{P}$ projects the predicted wavefield $\mathbf{u}$ onto the receiver locations as a measurement operator. Then the predicted data $\mathbf{d}_{cal}(m)$ can be written as

$$\mathbf{d}_{cal}(m) = \mathbf{P}A_\omega(m)^{-1}\mathbf{s},$$  \hspace{1cm} (2)

where $\mathbf{s}$ denotes the source, and $A_\omega(m) = \omega^2 m + \nabla^2$ is the discretized Helmholtz operator for angular frequency $\omega$ and squared slowness $m$.

As mentioned above, FWI is an ill-posed problem that requires proper regularization techniques to receive satisfactory inversion results. Adding regularization term to the Eq.(1), we have

$$\arg \min_m \{ \frac{1}{2} \| \mathbf{d}_{cal}(m) - \mathbf{d}_{obs} \|_2^2 + \varphi(m) \},$$  \hspace{1cm} (3)
Combined regularization for Full Waveform Inversion

where $\varphi(\cdot)$ is the regularization function.

In many image reconstruction problems, TV regularization is a preferred choice because it effectively maintains the edges and smooth structure of the image. In TV regularization, the regularization function $\varphi(m) = \|\nabla m\|_1$, where $\nabla$ denotes the discrete first gradient operator with periodic boundary conditions. However, using the TV regularization typically induces staircase effects in the reconstructed images.

3 PROPOSED METHOD

3.1 Overlapping group sparsity and non-convex second-order TV regularization

For a vector $x = [x(0), x(1), \cdots, x(n-1)]^T \in \mathbb{R}^n$, the vector $x_{i,K} = [x(i), x(i+1), \cdots, x(i+K-1)] \in \mathbb{R}^K$ stands for the $K$-point group of the vector $x$, which is considered as a block of $K$ contiguous samples of $x$ starting with the current index $i$. A commonly group sparsity regularization, defined as

$$\varphi_1(x) = \sum_i \left[ \sum_{k=0}^{K-1} |x(i+k)|^2 \right]^{1/2}.$$  

(4)

For the 2-D case, the $K \times K$ group for the subsurface model $m \in \mathbb{R}^{N_x \times N_x}$ is

$$\tilde{m}(i,j)_K = \begin{bmatrix} m(i-c_1,j-c_1) & m(i-c_1,j-c_1+1) & \cdots & m(i-c_1,j-c_2) \\ m(i-c_1+1,j-c_1) & m(i-c_1+1,j-c_1+1) & \cdots & m(i-c_1+1,j-c_2) \\ \vdots & \vdots & \ddots & \vdots \\ m(i-c_2,j-c_1) & m(i-p_2,j-c_1+1) & \cdots & m(i-c_2,j-c_2) \end{bmatrix} \in \mathbb{R}^{K \times K},$$

(5)

where $c_1 = \lfloor (K-1)/2 \rfloor$, $c_2 = \lfloor K/2 \rfloor$, and $\lfloor \cdot \rfloor$ represents the floor function. Let $m(i,j)_K = \tilde{m}(i,j)_K(:)$ corresponds to the column-major vectorized form of $\tilde{m}(i,j)_K$.

With these definitions, the TV regularization term with overlapping groups is represented as

$$\varphi_1(m) = \sum_{i,j=1} \|\nabla m(i,j)_K\|_2.$$  

(6)

OGS-TV provides more structural information about the original model through the grouping sparsity in the model-derivative domain, and can to some
Combined regularization for Full Waveform Inversion

extent suppress the staircase artifacts in the inversion results.

It is a known fact that high-order total variational regularization can more effectively eliminate artifacts by alleviating the under smooth characteristics of the model [21–23]. Recently, Adam et al. [24] proposed a new regularizer to combine the benefits of non-convex second-order TV and OGS-TV. More precisely, the non-convex second-order TV regularization with the $l^p$-norm ($0 < p < 1$) is presented as

$$\varphi_2(m) = \|\nabla^2 m\|_p^p,$$

where $\nabla^2$ is a second-order discrete operator of $m$. Especially, scholars demonstrate the superior ability of non-convex high-order TV regularization with OGS to recover high-resolution images from noise data [25, 26].

### 3.2 The NSTVOGS-FWI method

Using the combination of the non-convex second-order TV and OGS-TV, the composite regularization problem for FWI has the following form

$$\arg \min_m \{ \frac{1}{2} \|d_{cal}(m) - d_{obs}\|_2^2 + \lambda_1 \varphi_1(m) + \lambda_2 \varphi_2(m) \},$$

where the regularization parameters $\lambda_1$ and $\lambda_2$ control the balance of importance between these terms. We employ the ADMM algorithm to solve Eq.(8).

Eq.(8) can be equivalently rewritten as the following constraint optimization problem

$$\arg \min_q \{ \frac{1}{2} \|d_{cal}(m) - d_{obs}\|_2^2 + \lambda_1 \varphi_1(q) + \lambda_2 \varphi_2(q) \},$$

s.t. $q = m$.

According to the ADMM framework, we can alternately solve for the following sub-problems

$$m^{l+1} = \arg \min_m \{ \frac{1}{2} \|d_{cal}(m) - d_{obs}\|_2^2 + \frac{\gamma}{2} \|m - q^l + p^l\|_2^2 \};$$

$$q^{l+1} = \arg \min_q \{ \frac{\gamma}{2} \|m^{l+1} - q + p^l\|_2^2 + \varphi(q) \};$$

$$p^{l+1} = p + m^{l+1} - q^{l+1},$$

where $\varphi(q) = \lambda_1 \varphi_1(q) + \lambda_2 \varphi_2(q)$, $\gamma$ is a penalty parameter, $q$ denotes an auxiliary variable, and $p$ represents a dual variable.

To be specific, the first sub-problem (10a) is conventional FWI with $l^2$-norm regularization term. As a nonlinear optimization problem, any gradient-based optimization algorithm (steepest descent, conjugate gradient, or quasi-Newton) to solve this sub-problem. Considering the balance between
Combined regularization for Full Waveform Inversion

computational memory and the results accuracy, we utilize the limited memory quasi-Newton method (L-BFGS) \cite{27} to solve Eq.\eqref{10a}. The non-convex sub-problem \eqref{10b} is equal to a denoising operation, which is effectively solved by the majorization minimization based ADMM scheme \cite{28}. The details of solving Eq.\eqref{10b} are given in Appendix A for the reader reference. Finally, the dual variables are updated by the Eq.\eqref{10c}.

In particular, we point out that the penalty parameters $\lambda_1$ and $\lambda_2$ are selected according to \cite{29}, i.e., $\lambda_1 = \lambda_2 = \beta$, to simplify the parameter selection. In this paper, $\beta$ is obtained by an adaptive adjustment strategy. To be more specific, set $\chi = \frac{\|\nabla \Phi(m; q^l, p^l)\|}{\|m - q^l + p^l\|}$; if the condition $\chi^{l+1} < \tau \chi^l$ is satisfied, we have

$$\beta^{l+1} = \beta^0 \frac{\|\nabla \Phi(m^l; q^l, p^l)\|}{\|m^l - q^l + p^l\|};$$

otherwise $\beta^{l+1} = \tau \beta^l$. Here $\Phi(m; q^l, p^l) := \frac{1}{2} \|d_{cal}(m) - d_{obs}\|_2^2 + \frac{\gamma}{2} \|m - q^l + p^l\|_2^2$ and the parameter $0 < \tau < 1$. Following in the flowchart shown in Fig.1, we obtain the Algorithm 1 for solving the composite regularization problem \eqref{8} of FWI in detail.

![Flowchart of the NSTVOGS-FWI method.](image-url)

Fig. 1 Flowchart of the NSTVOGS-FWI method.
Algorithm 1 The NSTVOGS-FWI method

Input: the observed data $d_{\text{obs}}$, $m_0$, $\beta^0$, $\gamma$, $L$, $\iota$.
Repeat
1. For $l = 0, 1, 2, \cdots L - 1$ do
2. if $l==1$
3. $\beta = \beta^0$;
4. else
5. $\beta = \beta^{l+1}$;
6. end
7. $m^{l+1} = \arg\min_m \{ \frac{1}{2} \| d_{\text{cal}}(m) - d_{\text{obs}} \|^2 + \frac{\gamma}{2} \| m - q^l + p^l \|^2 \}$;
8. $q^{l+1} = \arg\min_q \{ \frac{2}{2} \| m^{l+1} - q + p^l \|^2 + \phi(q) \}$;
9. if $\chi^{l+1} <= \iota \chi^l$
10. $\beta^{l+1} = \beta^0 \| \nabla \Phi(m^l; q^l, p^l) \| / \| m^l - q^l + p^l \|$;
11. else
12. $\beta^{l+1} = \iota \beta^l$;
13. end
14. $p^{l+1} = p + m^{l+1} - q^{l+1}$.
End for
Output: the inversion result $m^L$.

4 Numerical results

In the experiments, we employ the well-known Marmousi model, the modified Marmousi model with the irregular salt body, and the 2004 BP benchmark model to demonstrate the superiority of the NSTVOGS-FWI method. The standard TV regularized FWI method (called TV-FWI method for short) is considered a comparison method and defined as

$$
\arg\min_m \{ \frac{1}{2} \| d_{\text{cal}}(m) - d_{\text{obs}} \|^2 + \lambda_{\text{TV}} \| \nabla m \|_1 \},
$$

where $\lambda_{\text{TV}}$ is the regularization parameter. For a fair comparison, the objective function (11) is also calculated by the L-BFGS algorithm. In FWI methods, the multiscale strategy in frequency domain is adopted to decrease the nonlinearity of waveform inversion, which can effectively improve the problem of FWI methods falling into local minima. For each frequency, the stop criterion for the L-BFGS algorithm is set to satisfy the $l^\infty$ norm of the gradient less than $1 \times 10^{-5}$ or the maximum number 15 of iterations is reached.

Through many experiments and attempts, we set the following parameters of the three experiments to be consistent. We pick the overlapping group size $K = 10$ and set the outer loop as $L = 3$. The second-order non-convex parameter $p$ is 0.9 in the first two experiments and 0.5 in the third experiment. In addition, other parameters set by the NSTVOGS-FWI and TV-FWI methods
are given in different examples. We utilize three quality measures such as peak signal-to-noise ratio (PSNR), root mean square error (RMSE), and running time for the quantitative evaluation. All experiments are running in Matlab 2018a on a PC equipped with an Intel Core i7-7700 Eight-core (8 Core) 3.60 GHz Processor and 16 GB of RAM to maintain fairness.

4.1 Marmousi model

In the first numerical example, we consider a 2D Marmousi velocity model (see Fig. 2(a)) with size $176 \times 302$, where the grid spacing is 0.02km in each direction. The true velocity model is first smoothed by the Gaussian kernel and then averaged along the horizontal direction to obtain the 1-D initial velocity model (see Fig. 2(b)). We model 59 sources and 172 receivers in the horizontal direction from 0.1km to 5.92km located at surface $z=0$km. For our NSTVOGS-FWI method, we use the settings $\beta^0 = 1 \times 10^{-3}$, $\gamma = 0.38$ and $\iota = 0.83$. For the TV-FWI method, Fig. 3(a) shows the peak signal-to-noise ratio (PSNR) of the inversion results obtained by different $\lambda_{TV}$. Therefore, we choose $\lambda_{TV} = 1 \times 10^{-2}$ to obtain the better inversion result.

Both FWI methods are carried out by 3 overlapping frequency bands, $2 \sim 8$Hz, $4 \sim 14$Hz, and $6 \sim 18$Hz, and add SNR=15dB Gaussian noise to the observed data. Fig. 4 displays the inversion results for the NSTVOGS-FWI and TV-FWI methods. Both of them can obtain a good reconstructed velocity model, but the proposed NSTVOGS-FWI method is better at handling the details in the middle and lower parts with a large degree (such as the black dotted box and the red ellipse) of the model. Particularly, we represent in Fig. 5 a vertical trace at $x=2.3$km in the Marmousi model. As expected, the NSTVOGS-FWI result matched the actual model better. For each FWI method, we compute the PSNR as well as the RMSE, which better reflects the visual quality of the inversion results. Table 1 summarizes the quantitative results on the different noise levels. Our method has the best performance except for the slight increase in running time, which suggests the effectiveness and feasibility of the proposed FWI method.
Fig. 2  (a) The true velocity and (b) The initial velocity for the Marmousi model.

Fig. 3  Different parameter $\lambda_{TV}$ of TV-FWI corresponding to the peak signal-to-noise ratio (PSNR) values of inversion results. (a) The parameter selection of the Marmousi model; (b) The parameter selection of the 2004 BP benchmark model.

Table 1  Quantitative results of the Marmousi model on different noise levels.

<table>
<thead>
<tr>
<th>Initialization</th>
<th>Noise (SNR=15)</th>
<th>Noise (SNR=5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR</td>
<td>RMSE</td>
</tr>
<tr>
<td>NSTVOGS-FWI</td>
<td>13.66</td>
<td>0.37</td>
</tr>
<tr>
<td>TV-FWI</td>
<td>17.08</td>
<td>0.28</td>
</tr>
<tr>
<td>NSTVOGS-FWI</td>
<td>17.66</td>
<td>0.24</td>
</tr>
<tr>
<td>TV-FWI</td>
<td>16.08</td>
<td>0.28</td>
</tr>
</tbody>
</table>
Combined regularization for Full Waveform Inversion

![Fig. 4](image)

**Fig. 4** Inversion results of the Marmousi model with SNR=15dB noise. (a) The inversion result with NSTVOGS-FWI; (b) The inversion result with TV-FWI.

![Fig. 5](image)

**Fig. 5** The velocity profile at x=2.3km from the Marmousi model.

### 4.2 The Marmousi model with salt body

We use the Marmousi model as the velocity background, then add an irregular salt body on top of it. The salt body in this model has a constant velocity of 4.5km/s, which the reader can refer to [30]. The true velocity model is shown in Fig.6(a). The initial velocity as shown in Fig.6(b) from 1.5km/s to 3.84km/s. In this experiment, we redefine the frequency band, as 2 ∼ 5Hz, 2.5 ∼ 7Hz, 3 ∼ 10Hz, respectively. For TV-FWI, the regularization parameter is $2 \times 10^{-3}$.

Fig.7(a), (c) and (e) show the NSTVOGS-FWI results at 2Hz of the first frequency band, 4Hz of the second frequency band, and 6Hz of the third frequency band, respectively. With the increasing frequency band, the texture of the geological model is gradually clear, and the feasible solution is closer to the true velocity model, as the red arrow in Fig.7. Similarly, the TV-FWI results at the above three frequencies are displayed in Fig.7(b), (d) and (f). Fig.7(g) and (h) show the final inversion results of the two FWI methods, indicating that the NSTVOGS-FWI and TV-FWI methods can construct the salt body and Marmousi background model. Our proposed method can maintain the fine structure inside the model as the red rectangles in Fig.7, while TV-FWI only protect the boundary information of the model. Observing that the salt body part of the inversion results, the salt body obtained by our proposed method is complete without missing parts compared with the TV-FWI method as the ellipse in Fig.7.
Fig. 6 (a) The true velocity and (b) The initial velocity for the Marmousi model with salt body.
Fig. 7 Inversion results of the Marmousi model with salt body (SNR=15dB noise). The first column shows the inversion results with NSTVOGS-FWI and the second column shows the inversion results with TV-FWI. (a), (c) and (e) are the inversion results with NSTVOGS-FWI at 2Hz (first frequency band), 4Hz (second frequency band) and 6Hz (third frequency band), respectively; (b), (d) and (f) are the inversion results with TV-FWI at 2Hz (first frequency band), 4Hz (second frequency band) and 6Hz (third frequency band), respectively. (g) and (h) are the final inversion results with two FWI methods.
Combined regularization for Full Waveform Inversion

4.3 2004 BP Model

The 2004 BP benchmark model (as shown in Fig. 8(a)) is a challenge for FWI attributing to a large-scale contrast salt body. The initial velocity is shown in Fig. 8(b), which is horizontal linearly increasing with depth. Similarly, on the surface, we fix 82 sources and 298 receivers. In this example, we solve all these FWI problems by sequentially inverting for 3 frequency bands of $2 \sim 5\text{Hz}$, $2.5 \sim 7\text{Hz}$, and $3 \sim 11\text{Hz}$. For the NSTVOGS-FWI method, we set $\beta^0 = 1 \times 10^{-3}$, $\gamma = 0.45$ and $\iota = 0.75$. According to Fig. 3(b), we choose $\lambda_{TV} = 1 \times 10^{-3}$.

The NSTVOGS-FWI and TV-FWI results are exhibited in Fig. 9(d1) and (e1). Both of them can obtain acceptable inversion results for the large contrast salt body model. Fig. 9 shows the NSTVOGS-FWI results at $2\text{Hz}$ of the first frequency band, $2.5\text{Hz}$ of the second frequency band, and $3\text{Hz}$ of the third frequency band. As expected, the reconstruction results show a gradually good trend during inversion. The “tooth” boundary of the velocity model is gradually revealed, and the related details are clear with the advance of inversion. Compared with the TV-FWI method, the NSTVOGS-FWI method can reduce artifacts as red rectangle in Fig. 9(d1) and (e1). Fig. 9(a2), (b2), (c2) and (d2) are transverse velocity profiles at $z=1.02\text{km}$ about the true velocity and inverted velocity corresponding to the first column. By observing Fig. 9(d2) and (e2), our proposed method (red line) is closer to the true velocity (blue line) than the TV-FWI method (green line), which verifies that the NSTVOGS-FWI method can obtain better inversion results. Table 2 shows the NSTVOGS-FWI method achieves the best performance for PSNR and RMSE, which quantitatively indicates the superiority of the proposed method.

To verify further the robustness of the NSTVOGS-FWI method, the FWI results (SNR=5dB) are shown in Fig. 10. For the Marmousi model, our NSTVOGS-FWI method can capture the details in the lower part of the model,
and the inversion results are relatively complete (see Fig.10(a)). Clearly, the BP 2004 model gained by the NSTVOGS-FWI method (see Fig.10(c)) contains more structural details and fewer inversion artifacts than the TV-FWI method (Fig.10(b)). Quantitative indexes (PSNR, RMSE and Time) of inversion results under SNR=5dB are shown in Table 1 and Table 2 respectively. Obviously, the NSTVOGS-FWI method maintains the optimal quantitative results except for time, which verifies the proposed method suppresses the influence of noise.

Table 2 Quantitative results of the BP 2004 model on different noise levels.

<table>
<thead>
<tr>
<th>Initialization</th>
<th>Noise (SNR=15)</th>
<th>Noise (SNR=5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NSTVOGS-FWI</td>
</tr>
<tr>
<td>PSNR</td>
<td>13.22</td>
<td>19.73</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.53</td>
<td>0.14</td>
</tr>
<tr>
<td>Times/s</td>
<td>–</td>
<td>397</td>
</tr>
</tbody>
</table>
Fig. 9 Inversion results of the 2004 BP model with SNR=15dB noise. The first column shows the NSTVOGS-FWI results for at 2Hz, 2.5Hz and 3Hz, respectively; (a2), (b2), (c2), (d2) and (e2) are a fit of true velocity (blue line) and the inversion results(red or green line) at z=1.02km. (d1) The NSTVOGS-FWI result; (e1) The TV-FWI result.
5 Conclusion

In this paper, we propose a new composite regularization FWI strategy based on an adaptive ADMM framework. The motivation for the proposed composite regularization is the fact that the advantages of different regularization methods are fully exploited to alleviate artifacts and depict delicate features of geological models. Comparative results show that the proposed NSTVOGS-FWI method is robust to noise and competitive with current FWI at reasonable computational efficiency. In the future, an exciting work would be to develop a fast computation method to reduce the time of the NSTVOGS-FWI method.

Appendix A

The minimization problem Eq. (10b) is difficult to be solved by simple numerical methods due to the non-convex high-order TV regularization terms. The ADMM has been applied to solve this problem as an efficient algorithm [17]. Then, the problem Eq. (10b) is reformulated as

$$\min_{q} \frac{\gamma}{2} \| m^{t+1} - q + p^t \|_2^2 + \lambda_1 \varphi_1(v) + \lambda_2 \varphi_2(w) + I_H(z),$$

s.t. $v = \nabla q, w = \nabla^2 q, z = q$.  \hspace{1cm} (A1)
Combined regularization for Full Waveform Inversion

Here, the $\mathcal{I}$ is the indicator function and $\mathcal{H}$ is the feasible set of the variable $z$.

Then, the augmented Lagrangian (AL) function is defined

$$
\mathcal{L}_A(q, v, w, z; \mu^1, \mu^2, \mu^3) = \frac{\gamma}{2} \| m^{t+1} - q + p \|_2^2 + \lambda_1 \varphi_1(v) + \lambda_2 \varphi_2(w) + \lambda_3 \varphi_3(z) - (\mu^1)^T (v - \nabla q) + \frac{\rho}{2} \| v - \nabla q \|_2^2 - (\mu^2)^T (w - \nabla^2 q) + \frac{\rho}{2} \| w - \nabla^2 q \|_2^2 - (\mu^3)^T (z - q) + \frac{\rho}{2} \| z - q \|_2^2,
$$

where $\rho$ represents the penalty parameter, $\mu^1$, $\mu^2$ and $\mu^3$ are Lagrange multipliers. $v$, $w$ and $z$ are auxiliary variables corresponding to $\nabla q$, $\nabla^2 q$, and $q$ respectively.

Next, $\mathcal{L}_A$ is split into subproblems to solve

$$
q_{t+1} = \arg \min_q \mathcal{L}_A(q, v_t, w_t, z_t; \mu^1_t, \mu^2_t, \mu^3_t); \quad (A3a)
$$

$$
v_{t+1} = \arg \min_v \mathcal{L}_A(q_{t+1}, v, w_t, z_t; \mu^1_t, \mu^2_t, \mu^3_t); \quad (A3b)
$$

$$
w_{t+1} = \arg \min_w \mathcal{L}_A(q_{t+1}, v_{t+1}, w, z_t; \mu^1_t, \mu^2_t, \mu^3_t); \quad (A3c)
$$

$$
z_{t+1} = \arg \min_z \mathcal{L}_A(q_{t+1}, v_{t+1}, w_{t+1}, z; \mu^1_t, \mu^2_t, \mu^3_t); \quad (A3d)
$$

$$
\mu^1_{t+1} = \mu^1_t + (v_{t+1} - \nabla q_{t+1}); \quad (A3e)
$$

$$
\mu^2_{t+1} = \mu^2_t + \rho (\nabla^2 q_{t+1} - w_{t+1}); \quad (A3f)
$$

$$
\mu^3_{t+1} = \mu^3_t + (q_{t+1} - z_{t+1}). \quad (A3g)
$$

(1) $q$–subproblems

Subproblem (A3a) is a least squares problem

$$
q_{t+1} = \arg \min_q \frac{\gamma}{2} \| m^{t+1} - q + p \|_2^2 - (\mu^1_t)^T (v_t - \nabla q) + \frac{\rho}{2} \| v_t - \nabla q \|_2^2 - (\mu^2_t)^T (w_t - \nabla^2 q) + \frac{\rho}{2} \| w_t - \nabla^2 q \|_2^2 - (\mu^3_t)^T (z_t - q) + \frac{\rho}{2} \| z_t - q \|_2^2. \quad (A4)
$$

This problem can be solved as a normal equation to give the following closed solution

$$
q_{t+1} = (\gamma I + \rho \nabla^2 \nabla + \rho (\nabla^2)^T \nabla^2 + \rho I)^{-1} (\gamma (m^{t+1} + p) - \nabla^T \mu^1_t + \rho \nabla^T v_t - (\nabla^2)^T \mu^2_t + \rho (\nabla^2)^T w_t - \mu^3_t + \rho z_t). \quad (A5)
$$

(2) $v$–subproblems

The $v$ subproblem is the calculation of the TV with overlapping group
sparse. Selesnick and Chen et al. [18] proposed the majorization-minimization (MM) method to solve the first-order TV problem with overlapping group sparse and successfully apply it in signal processing. We utilize the MM algorithm to solve the subproblem (A3b) by setting $K > 1$

$$
\mathbf{v}_{t+1} = \arg \min_{\mathbf{v}} \frac{\rho}{2} \left\| \mathbf{v} - \nabla \mathbf{q}_{t+1} \right\|_2^2 - (\mathbf{\mu}_t^1)^T (\mathbf{v} - \nabla \mathbf{q}_{t+1}) + \lambda_1 \varphi_1(\mathbf{v})
$$

$$
= \arg \min_{\mathbf{v}} \frac{\rho}{2} \left\| \mathbf{v} - (\nabla \mathbf{q}_{t+1} + \frac{\mathbf{\mu}_t^1}{\rho}) \right\|_2^2 + \lambda_1 \varphi_1(\mathbf{v}).
$$

We do not show the steps of the MM algorithm, readers can refer to [16] for more details.

(3) $w$–subproblems

The $w$–subproblem (A3c) is a non-convex second-order TV denoising problem

$$
\mathbf{w}_{t+1} = \arg \min_{\mathbf{w}} \frac{\rho}{2} \left\| \mathbf{w} - \nabla^2 \mathbf{q}_{t+1} \right\|_2^2 - (\mathbf{\mu}_t^2)^T (\mathbf{w} - \nabla^2 \mathbf{q}_{t+1}) + \lambda_2 \left\| \mathbf{w} \right\|_p^p
$$

$$
= \arg \min_{\mathbf{w}} \frac{\rho}{2} \left\| \mathbf{w} - (\nabla^2 \mathbf{q}_{t+1} + \frac{\mathbf{\mu}_t^2}{\rho}) \right\|_2^2 + \lambda_2 \left\| \mathbf{w} \right\|_p^p.
$$

(A6)

We use the iterative re-weighting (IRL1) algorithm to solve this non-convex problem, and $\mathbf{d}_{t+1} = \nabla^2 \mathbf{q}_{t+1} + \frac{\mathbf{\mu}_t^2}{\rho}$ into the Eq.(A6)

$$
\mathbf{w}_{t+1} = \arg \min_{\mathbf{w}} \frac{\rho}{2} \left\| \mathbf{w} - \mathbf{d}_{t+1} \right\|_2^2 + \lambda_2 \left\| \mathbf{w} \right\|_p^p.
$$

(A7)

According to the IRL1 algorithm, the convex problem of approximate Eq.(A7) is found as follows

$$
\mathbf{w}_{t+1} = \arg \min_{\mathbf{w}} \frac{\rho}{2} \left\| \mathbf{w} - \mathbf{d}_{t+1} \right\|_2^2 + \lambda_2 p ((|w_i|) + \varepsilon)^{p-1} |w_i|,
$$

(A8)

where $\varepsilon$ is to avoid having a zero in the denominator.

Then $\theta_i = \lambda_2 p ((|w_i|) + \varepsilon)^{-p}$ is used to simplify Eq.(A8)

$$
\mathbf{w}_{t+1} = \arg \min_{\mathbf{w}} \frac{\rho}{2} \left\| \mathbf{w} - \mathbf{d}_{t+1} \right\|_2^2 + \sum_i \theta_i |w_i|.
$$

(A9)

Therefore, we can use the threshold function to obtain the optimal solution of the Eq.(A8)

$$
\mathbf{w}_{k+1} = shrink(\mathbf{d}_{t+1}, \frac{\theta_i \lambda_2}{\rho})
$$

$$
= \max\{\frac{|\mathbf{d}_{t+1} - \theta_i \lambda_2}{\rho}, 0\} \cdot sign(\mathbf{d}_{t+1}).
$$

(A10)
(4) \( z \)-subproblems

The \( z \) subproblem is guaranteed to the solution in \( \mathcal{H} \), which is a feasible solution set about \( q \). We transform the problem (A3d) into

\[
    z_{t+1} = \arg \min_z \frac{\rho}{2} \left\| z - (q_{t+1} + \frac{\mu^3_t}{\rho}) \right\|_2^2 + I_\mathcal{H}(z). \tag{A11}
\]

Since problem (A11) is the proximal operator of indicator function, we can find \( z_{t+1} \) by a simple project on to the set \( \mathcal{H} \).

For subproblems (A3e)~(A3g), we use the solutions obtained above to update iteratively

\[
    \begin{align*}
    \mu^1_{t+1} &= \mu^1_t + (v_{t+1} - \nabla q_{t+1}); \\
    \mu^2_{t+1} &= \mu^2_t + \rho(\nabla_2 q_{t+1} - w_{t+1}); \\
    \mu^3_{t+1} &= \mu^3_t + (q_{t+1} - z_{t+1}).
    \end{align*}
\]

We show the ADMM framework for solving the Eq.(10b) in Algorithm 2. The value of \( \rho \) is 0.009 and the initial value of \( \mu^1, \mu^2, \mu^3 \) are the zero vector. The maximum number of iterations is 20, i.e. \( O = 20 \). \( \mathcal{H} \) is the set from the minimum to the maximum true velocity of the different media model in our paper.

**Algorithm 2** The ADMM method for Eq.(10b)

Repeat

Input: \( m^{l+1}, q^l, p^l \) and parameters \( \lambda_1, \lambda_2, \rho \) and \( O \)

1. \( q_0 = q^l \).
2. For \( t = 0, 1, 2, \ldots, O - 1 \) do
3. \( q_{t+1} = (\gamma I + \rho \nabla^T \nabla + \rho(\nabla^2)^T \nabla^2 + \rho I)^{-1}(\gamma(m^{l+1} + p^l) \\
- \nabla^T \mu^1_t + \rho \nabla T v_t - (\nabla^2)^T \mu^2_t + \rho(\nabla^2)^2 w_t - \mu^3_t + \rho z_t) \).
4. \( v_{t+1} = \arg \min_v \frac{\rho}{2} \left\| v - (\nabla q_{t+1} + \frac{\mu^1_t}{\rho}) \right\|_2^2 + \lambda_1 \varphi_1(v) \).
5. \( w_{t+1} = \max\{|d_{t+1}| - \frac{\mu^3_t}{\rho}, 0\} \cdot \text{sign}(d_{t+1}) \).
6. \( z_{t+1} = \arg \min_z \frac{\rho}{2} \left\| z - (q_{t+1} + \frac{\mu^3_t}{\rho}) \right\|_2^2 + I_\mathcal{H}(z) \).
7. \( \mu^1_{t+1} = \mu^1_t + (v_{t+1} - \nabla q_{t+1}) \).
8. \( \mu^2_{t+1} = \mu^2_t + \rho(\nabla^2 q_{t+1} - w_{t+1}) \).
9. \( \mu^3_{t+1} = \mu^3_t + (q_{t+1} - z_{t+1}) \).
11. End for

Output: the final result \( q_O \).

References

Combined regularization for Full Waveform Inversion


Combined regularization for Full Waveform Inversion


[26] Adam T, Paramesran R, Mingming Y, et al.. Combined higher order non-convex total variation with overlapping group sparsity for impulse noise


Statements and Declarations

Funding
This work is supported by the National Natural Science Foundation of China (No. 42274166) and the Fundamental Research Funds for the Central Universities (No. 3132022201).

Competing Interests

Author Contributions
All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Hongyu Qi and Ruixue Gu. The first draft of the manuscript was written by Hongsun Fu and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data Availability
The data that support the findings of this study are available from the corresponding author upon reasonable request.