

Regularized Full-Waveform Inversion With Shearlet Transform and Total Generalized Variation

Han Wang^{ID} and Siwei Yu^{ID}, *Member, IEEE*

Abstract—Full-waveform inversion (FWI) is a powerful method of reconstructing subsurface properties during seismic exploration. However, it is difficult for FWI to accurately describe a subsurface model with sharp surfaces and smooth variations because of the highly nonlinear and ill-posed problems associated with FWI. We first propose a novel FWI with shearlet transform and total generalized variation (TGV) regularization on a subsurface model to alleviate this challenge. Shearlet transform is particularly well adapted to preserve the abundant geometric information of models by representing anisotropic features such as curves and edges; however, it often induces the boundary effect, leading to a resolution reduction. To address shearlet transform drawbacks, we employ TGV to reduce the artifacts by involving various order derivatives to adjust different degrees of smoothness. The proposed regularization scheme is robust to the noise of the observed data during the inversion process. Using the simple synthetic, Society of Exploration Geophysicists (SEG)/European Association of Geoscientists and Engineers (EAGE) overthrust, Marmousi, and modified 2004 BP models, we demonstrate that the proposed method reconstructs subsurface geophysical models with sharp interfaces and smooth background variations more accurately than the conventional methods without any regularization and those with only the total variation (TV) regularization, TGV regularization, and shearlet transform regularization.

Index Terms—Full-waveform inversion (FWI), shearlet transform, total generalized variation (TGV).

I. INTRODUCTION

FULL-WAVEFORM inversion (FWI) has been used to obtain high-resolution subsurface physical properties by minimizing the misfit between the observed and modeled seismograms [1], [2]. FWI is a strongly nonlinear and ill-posed inverse problem and may cause multiple solutions and uncertainties. It is highly dependent on the initialization and may produce an inversion result near the starting velocity model. Cycle skipping occurs in FWI when the initial guess differs significantly from the true subsurface model and the error in predicting arrival times is larger than half of the period [2]. The objective function has been improved with a more convex nature to alleviate the most challenging problems. Over the past decade, many more convex distance measures have emerged to calculate the misfit between the observed and

Manuscript received 13 March 2024; revised 20 June 2024; accepted 4 August 2024. Date of publication 27 August 2024; date of current version 9 September 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFA1003502, in part by the NSFC under Grant 42074156, and in part by the Fundamental Research Funds for the Central Universities under Grant 2022FRFK060027. (*Corresponding author: Siwei Yu.*)

The authors are with the School of Mathematics, Institute of Artificial Intelligence, Harbin Institute of Technology, Harbin 150001, China (e-mail: wanghan_jlu@163.com; siweiyu@hit.edu.cn).

Digital Object Identifier 10.1109/TGRS.2024.3449244

modeled data, such as envelope and instantaneous phase [3], [4], [5], [6], optimal transport [7], [8], matching filters [9], and so on. More ambitious strategies have been proposed in recent years. Huang et al. [10] applied a dynamic matching method to maximize the cross correlation of the observed and modeled data instead of the least-squares scheme to provide a high-resolution image. Yang and Ma [11] proposed a novel method called FWIGAN to mitigate the cycle-skipping problem. The method combines the advantages of the Wasserstein distance and generative adversarial learning (GAN) induced by optimal transport theory and deep learning, respectively. Izzatullah and Alkalifah [12] developed a Fourier-based misfit function for FWI, which minimizes the difference between the power spectrum information of the observed and modeled data. Further strategies broaden the search space in the model or data domains to extend the linear regime of FWI for addressing the cycle-skipping problem [13], [14], [15], [16], [17], [18].

Regularized methods have been applied to mitigate the ill-posed nature of FWI. Tikhonov regularization [19] and total variation (TV) [20] have been commonly applied to geophysical inversion. Asnaashari et al. [21] employed the prior model information derived from wells as a penalty term in regularized FWI with Tikhonov constraints to help converge to the expected models. TV [22] was used as a model penalty term in the objective function of FWI to obtain high-resolution results with sharp interfaces. To improve TV-regularized methods for describing high-contrast medium anomalies, asymmetric TV regularization [23] was applied to FWI to penalize negative velocity excursions. Aghamiry et al. [24] proposed a hybrid Tikhonov and TV-regularization method to recover large-contrast physical properties in FWI. The application of Tikhonov and TV regularization accounted for both sharp interfaces and smoothly varying background parts of the example. Qu et al. [25] developed a novel FWI scheme with directional TV based on a local dip field to optimally reveal geological information. The local dip field was estimated from the initial velocity using plane-wave destruction. Du et al. [26] designed a high-order TV-regularized FWI method to facilitate the convergence of the inversion process and achieve more focused sharp interfaces based on the sparsity of the high-order term.

Total generalized variation (TGV) [27] was first proposed to alleviate the staircasing effects of first-order TV methods by inducing various order derivatives. The method inverts for both sharp interfaces and smooth variations in the expected example. Inspired by image reconstruction using the TGV constraint, Gao and Huang [28] developed a novel FWI scheme with TGV regularization for acoustic and elastic waves to obtain stable and unique solutions, which could powerfully

suppress noise in the inversion process. Huber et al. [29] employed coupled TGV regularization to multimodal electron tomography to provide high-resolution image quality for all data. Chen et al. [30] employed the structural similarity together with a TGV scheme to develop a TGV denoising effect by inducing overlapping group sparsity.

In contrast to the above-regularized methods, other advanced regularizations are also applied to seismic inversion. The nonlocal regularization is a more recent technique for reconstructing complex models with piecewise smooth features. Aghamiry et al. [31] applied the block matching and 3-D filtering (BM3D) regularized method to FWI to reconstruct high-resolution piecewise smooth models. Fu et al. [32] proposed a novel regularization combining dictionary learning and graph Laplacian matrix used in FWI to obtain an accurate result, which induces the nonlocal self-similarity information to improve the stability of the inversion process. The deep learning technology is a powerful regularizer to improve the inversion results. Zhu et al. [33] integrated a deep neural network with FWI to suppress noise in the gradients and mitigate local minima, in which neural networks employs spatial correlations to constrain the velocity model. Sun et al. [34] used a pre-trained deep convolutional neural network in FWI, which is regarded as a regularizer to decrease artifacts in the inversion process. These regularized methods can improve the resolution of inversion results well, but the computational cost increases significantly compared with the local regularization.

Shearlet transform and TGV regularization [35] were developed to mitigate the ill-posedness of linear inverse problems based on the advantages of sparse representation and various-order derivatives. The shearlet transform was applied to extract multiscale and multidirectional information to complete the model constraint. It does not contain the inverse shearlet transform and reduce the computational cost. TGV was used to improve the inversion result by suppressing boundary effects caused by shearlet transform. For reconstructing images from blurry, noisy, incomplete measurements, or other challenging tasks, the combined regularization can preserve a variety of image features such as edges and textures. Compared with other conventional regularized methods, the above regularization represents better reconstruction accuracy and feature-preserving ability. This regularized scheme has been applied effectively to image reconstruction, deconvolution, and denoising in recent years. In seismic data process fields, Yang et al. [36] employed shearlet transform and TGV to seismic reconstruction to reveal better image results than other L_1 -norm regularized methods, which could consider signal sparsity and subsurface structure complexity. To improve the signal-to-noise ratios of seismograms, the combination of shearlet transform and TGV was also used in seismic noise attenuation [37], [38]. Liu [39] proposed a novel Poisson image deconvolution approach based on the advantages of shearlet transform and TGV to obtain the expected result. These studies show that shearlet transform and TGV regularization can produce more accurate images than other conventional approaches and have great application potential.

Inspired by shearlet transform and TGV regularization used in image processing, we propose a novel FWI scheme with combined regularization to obtain high-resolution results. The

method reduces the ill-posedness of the nonlinear inverse problem. Shearlet transform has multiscale and multidirectional properties that preserve the boundaries and details of the model. However, some artifacts may be regarded as useful information to retain, resulting in that the background variation of the model is typically not smooth. TGV is a powerful tool for smoothing the model and suppressing the artifacts of the inversion process. Therefore, the proposed method can reveal the subsurface geophysical characteristics well, with sharp interfaces and background variations. We test the proposed method on the simple synthetic, Society of Exploration Geophysicists (SEG)/European Association of Geoscientists and Engineers (EAGE) overthrust, Marmousi, and modified 2004 BP models. Numerical experiments demonstrate that the regularized FWI method with shearlet transform and TGV is highly effective.

This article is organized as follows. First, we introduce the construction of the objective function in the proposed FWI scheme and the mathematical framework of shearlet transform and TGV. Second, we focus on numerical experiments of the simple synthetic, SEG/EAGE overthrust, Marmousi, and modified 2004 BP models. Next, we discuss parameter selection, computational cost, and high-dimensional applications of the method. Finally, we conclude this article.

II. METHODOLOGY

FWI minimizes the misfit between the observed and modeled data by efficient optimization to find a subsurface model. Based on shearlet transform and TGV regularization, we formulate the proposed objective function as follows:

$$J = \frac{1}{2} \|(\mathbf{d}_{\text{obs}} - \mathbf{d}(\mathbf{m}))\|_2^2 + \lambda \sum_{j=1}^N \|\mathcal{SH}_j(\mathbf{m})\|_1 + \mathcal{T}\mathcal{G}\mathcal{V}_\alpha^2(\mathbf{m}) \quad (1)$$

where $\mathbf{d}_{\text{obs}} \in \mathbb{R}^V$ with $V = n_s \times n_t \times n_r$ represents the observed data; n_s , n_t , and n_r are the number of shots, time sampling points, and geophones, respectively; $\mathbf{d}(\mathbf{m})$ denotes the modeled seismogram; $\mathbf{d}(\cdot)$ represents a constant-density acoustic finite-difference model for wave propagation; $\mathbf{m} \in \mathbb{R}^{n_x \times n_y}$ is the subsurface model and n_x and n_y denotes the number of grids in the horizontal and vertical directions, respectively; μ represents the weight factor of the data-fidelity term; and λ is the tradeoff parameter of the regularized terms. To directly introduce the mathematics of the proposed method, we convert the model from the matrix to the vector $\mathbf{m} \in \mathbb{R}^{n_x n_y}$. Next, the objective function is introduced in detail.

The second term of the objective function represents the shearlet transform regularization. Considering the numerical computations, a fast finite shearlet transform (FFST) [40] is adopted to realize the sparse transform using Meyer scaling and wavelet functions. Where $\mathcal{SH}_j(\mathbf{m})$ represents the j_{th} shearlet coefficients of the discrete shearlet transform of \mathbf{m} , and N denotes the total number of subbands related to the number of scales, n , by $N = 2^{n+2} - 3$. The discrete shearlet transform [35] can be calculated using the discrete Fourier and discrete inverse Fourier transform, which is defined as

$$\mathcal{SH}_j(\mathbf{m}) = \mathbf{F}^* \text{diag}(\text{vec}(\mathbf{H}_j)) \mathbf{F} \mathbf{m} = \mathbf{M}_{H_j} \mathbf{m} \quad (2)$$

$$\mathbf{M}_{H_j} = \mathbf{F}^* \text{diag}(\text{vec}(\mathbf{H}_j)) \mathbf{F} \quad (3)$$

where \mathbf{F} and \mathbf{F}^* are discrete Fourier and discrete inverse Fourier transform matrices, respectively; \mathbf{H}_j denotes the j th discrete shearlet used in $\mathcal{SH}_j(\mathbf{m})$; $\text{vec}(\cdot)$ represents the matrix vectorization; and $\text{diag}(\cdot)$ denotes the operation that creates a diagonal matrix based on the vector.

The third objective function term is TGV regularization. TGV contains various-order derivatives to alleviate the TV staircasing effect, whereas the TV only uses the first-order derivative. We use the second-order TGV to constrain the velocity model, defined as

$$\mathcal{T}\mathcal{G}\mathcal{V}_\alpha^2(\mathbf{m}) = \alpha_2 \|\nabla \mathbf{m} - \mathbf{p}\|_1 + \alpha_3 \|\mathcal{E}(\mathbf{p})\|_1 \quad (4)$$

$$\nabla \mathbf{m} = \begin{bmatrix} \mathbf{D}_x \mathbf{m} \\ \mathbf{D}_y \mathbf{m} \end{bmatrix} \quad (5)$$

$$\mathcal{E}(\mathbf{w}) = \begin{bmatrix} \mathbf{D}_{xx} \mathbf{m} & \mathbf{D}_{yx} \mathbf{m} \\ \mathbf{D}_{yx} \mathbf{m} & \mathbf{D}_{yy} \mathbf{m} \end{bmatrix} \quad (6)$$

where \mathbf{p} is the estimated first-order difference, α_2 and α_3 are the positive weights to balance the first- and second-order terms, respectively. Here, \mathbf{D}_x and \mathbf{D}_y are the circulant matrices corresponding to the first-order forward finite-difference operators in the horizontal and vertical directions, respectively. \mathbf{D}_{xx} , \mathbf{D}_{yx} , and \mathbf{D}_{yy} are the circulant matrices corresponding to the second-order forward finite-difference operators in different directions. Next, we define $\mathbf{p} = (1/2)\nabla \mathbf{m}$, (4) can be simplified as

$$\begin{aligned} \mathcal{T}\mathcal{G}\mathcal{V}_\alpha^2(\mathbf{m}) &= \frac{\alpha_2}{2} \|\nabla \mathbf{m}\|_1 + \frac{\alpha_3}{2} \|\mathcal{E}(\nabla \mathbf{m})\|_1 \\ &= \alpha_1 \|\nabla \mathbf{m}\|_1 + \alpha_0 \|\mathcal{E}(\nabla \mathbf{m})\|_1. \end{aligned} \quad (7)$$

If α_0 equals 0, TGV is deduced to TV. By substituting (7) into (1), the proposed objective function can be reformulated as

$$\begin{aligned} J = \frac{1}{2} \|(\mathbf{d}_{\text{obs}} - \mathbf{d}(\mathbf{m}))\|_2^2 + \lambda \sum_{j=1}^N \|\mathbf{M}_{H_j} \mathbf{m}\|_1 \\ + \alpha_1 \|\nabla \mathbf{m}\|_1 + \alpha_0 \|\mathcal{E}(\nabla \mathbf{m})\|_1. \end{aligned} \quad (8)$$

To ensure the differentiable property of (1), The gradient of L1 regularization at 0 is 0. We define a sign function to approximate L1-norm derivatives, which is shown in the following equation:

$$\text{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0. \end{cases} \quad (9)$$

According to the chain rule, we can easily obtain the gradient of the proposed objective function concerning the model parameters, defined as

$$\begin{aligned} \mathcal{G}_m = \mathbf{W}_d \left[\left(\frac{\partial \mathbf{d}(\mathbf{m})}{\partial \mathbf{m}} \right)^T (\mathbf{d}_{\text{obs}} - \mathbf{d}(\mathbf{m})) \right] + \lambda \sum_{j=1}^N \mathbf{M}_{H_j}^T \text{sgn}(\mathbf{M}_{H_j} \mathbf{m}) \\ + \alpha_1 [\mathbf{D}_x^T \text{sgn}(\mathbf{D}_x \mathbf{m}) + \mathbf{D}_y^T \text{sgn}(\mathbf{D}_y \mathbf{m})] \\ + \alpha_0 [\mathbf{D}_{xx}^T \text{sgn}(\mathbf{D}_{xx} \mathbf{m}) + 2\mathbf{D}_{yx}^T \text{sgn}(\mathbf{D}_{yx} \mathbf{m}) \\ + \mathbf{D}_{yy}^T \text{sgn}(\mathbf{D}_{yy} \mathbf{m})] \end{aligned} \quad (10)$$

Algorithm 1: Regularized FWI Using Adam Optimization

Require: \mathbf{m}_0 shows the starting velocity model
Require: $\mathbf{p}_0 \leftarrow 0$ is the starting 1st moment
Require: $\mathbf{q}_0 \leftarrow 0$ denotes the starting 2nd moment
Require: $t_0 \leftarrow 0$ represents the starting timestep
Require: α is learning rate
Require: $\beta_1 = 0.9$, $\beta_2 = 0.999$ represent exponential decay rates for the moment estimates, β_1^t and β_2^t denote β_1 and β_2 to the power t
Require: J denotes the objective function with respect to the velocity model \mathbf{m}

- 1: **while** \mathbf{m}_t not converged **do**
- 2: $t \leftarrow t + 1$
- 3: $\mathbf{g}_t \leftarrow \frac{\partial J_t}{\partial \mathbf{m}_t}$ (calculated by Equation 10)
- 4: $\mathbf{p}_t \leftarrow \beta_1 \mathbf{p}_{t-1} + (1 - \beta_1) \mathbf{g}_t$
- 5: $\mathbf{q}_t \leftarrow \beta_2 \mathbf{q}_{t-1} + (1 - \beta_2) \mathbf{g}_t^2$ (where $\mathbf{g}_t^2 = \mathbf{g}_t \odot \mathbf{g}_t$)
- 6: $\hat{\mathbf{p}}_t \leftarrow \mathbf{p}_t / (1 - \beta_1^t)$
- 7: $\hat{\mathbf{q}}_t \leftarrow \mathbf{q}_t / (1 - \beta_2^t)$
- 8: $\mathbf{m}_t \leftarrow \mathbf{m}_{t-1} - \alpha \hat{\mathbf{p}}_t / (\sqrt{\hat{\mathbf{q}}_t} + \epsilon)$ (Where $\epsilon = 10^{-8}$)
- 9: **end while**

where the data-fidelity gradient $((\partial \mathbf{d}(\mathbf{m}) / \partial \mathbf{m}))^T$ is efficiently calculated using an adjoint scheme [41] without an explicit computation. \mathbf{W}_d denotes a coefficient matrix that maintains the order of magnitude with each gradient in the data-fidelity term consistently, which ensures the effectiveness of shearlet transform regularization. In Section III, we provide the specific definitions of \mathbf{W}_d .

The optimization algorithms used in FWI include the alternating-direction method of multipliers [42], conjugate gradient [43], Limited-Memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) [44], split Bregman method [45], etc. In recent years, the adaptive moment estimation (Adam) algorithm [46] has been employed to solve the inverse problem with GPU acceleration technology [47], [48]. Here, we choose the Adam method to address the FWI constrained by the shearlet transform and TGV and the Adam optimization framework is shown in Algorithm 1. The parameters β_1 and β_2 are initial decay rates used when estimating the first and second moments of the gradient, which provide some momentum to cross a local minima or saddle point to impact the convergence of Adam optimizer. Here, we employ the exponential attenuation to perform the change about the parameters β_1 and β_2 . Refer to many numerical examples and the work of Kingma and Ba [46], Richardson [48], Manu et al. [49], Deng et al. [50], and Feng et al. [51], we set β_1 and β_2 to 0.9 and 0.999, respectively.

III. NUMERICAL EXPERIMENTS

A. Simple Synthetic Model

To illustrate the effectivenesses of the shearlet transform and TGV in the combined regularization, respectively, we generate a simple synthetic model, which is 1 km long and 1 km deep with a spatial grid of 10 m. It is composed of 101 and 101 grid points in the horizontal and vertical directions, respectively. Meanwhile, the background velocity of the model is 1.8 km/s, and the square is regarded as an anomaly with a velocity of

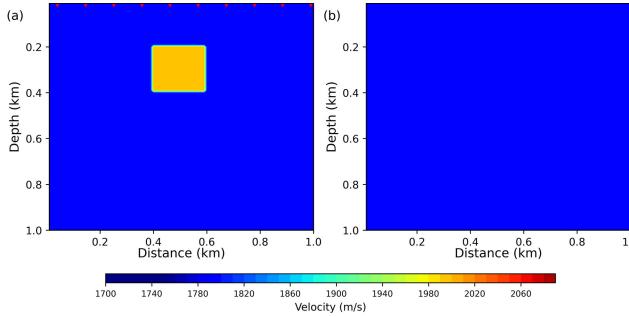


Fig. 1. Simple synthetic model. (a) True model with ten equally spaced shots (red triangles). (b) Starting model.

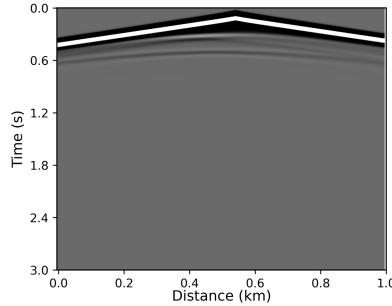


Fig. 2. Middle-shot gather of the simple synthetic model.

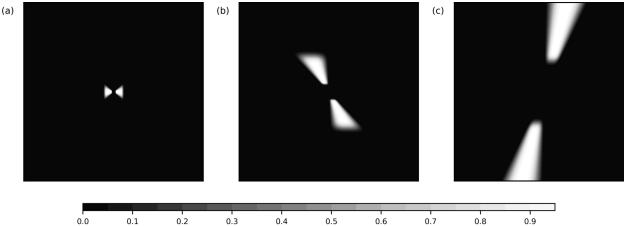


Fig. 3. Different scale shearlets in the frequency domain. (a) First-scale shearlet. (b) Second-scale shearlet. (c) Third-scale shearlet.

2.0 km/s. Fig. 1 shows the true model and starting model, we use a uniform model with a velocity of 1.8 km/s as the starting model. It is noted that the true model is not feasible for realistic geological structures. We employ a Ricker wavelet with a dominant frequency of 10 Hz as the source wavelet and use DeepWave [48] to complete the wave propagation for constant-density acoustic finite-difference modeling. Next, we generate ten equally spaced shots represented by the red triangle in Fig. 1(a) with 100 equally spaced receivers for each shot. The shot interval and the receiver spacing are 100 and 10 m, respectively. The middle-shot gather is represented in Fig. 2.

Next, we compare five inversion methods: traditional FWI without regularization (TRAFWI), FWI with TV regularization (TVFWI), FWI with TGV regularization (TGVFWI), FWI with shearlet transform regularization (STFWI), and FWI with shearlet transform and TGV regularization (SGVFWI). We set the tradeoff parameter of TVFWI to 0.0005. The parameters α_1 and α_0 of the TGVFWI are 0.0001 and 0.00001, respectively. We employ FFST to obtain three different scale shearlet coefficients by three scale shearlets (see Fig. 3), which contain multiscale model information to constrain the velocity model. However, because the true model is too simple, it is difficult

to find that the shearlet coefficients include more model information with the increase of the scale parameter shown in Fig. 4. In addition, we design a matrix \mathbf{W}_d to improve the effectiveness of the shearlet transform regularization, which is obtained by the Hadamard product of two matrices \mathbf{W}_m and \mathbf{W}_n . The matrix \mathbf{W}_m assigns large values from the first column to the 30th column and from the 71st column to the 100th column, and the values at the first column and the 100th column decrease from 2 to 1 along the row and gradually decay inward to 1. The row of the matrix \mathbf{W}_n linearly increases from 1 to 2000 along the column from 1 to 100. The more details about the matrix \mathbf{W}_d are described in the next numerical example. The tradeoff parameter of the shearlet transform regularization is 0.0008. The SGVFWI parameters λ , α_1 and α_0 are 0.001, 0.05, and 0.025, respectively. The matrix \mathbf{W}_d is the same as STFWI. We use the Adam optimizer with a learning rate of 30 to solve FWI.

Fig. 5 shows the inversion results with different methods after 400 iterations, which clearly denote the contribution of the different parts contained in the proposed method. TGVFWI and TVFWI significantly reduce the artifacts of the inversion results to get a smooth background. Moreover, TGV induces the second-order derivative to further improve the smoothing ability compared with TV. STFWI applies the shearlet transform to extract different scale model information to constrain the velocity model, which retains the main model details. However, because FWI is a strongly nonlinear inverse problem, the inversion result always contains some wrong structures to confuse the shearlet transform regularization to reserve them. Therefore, it is necessary to combine the shearlet transform with TGV to improve the accuracy of the inversion result. Compared with TRAFWI, TVFWI, TGVFWI, and STFWI, SGVFWI achieves the best performance in the test of the simple synthetic model. Figs. 6 and 7 represent the convergence diagram of data misfits and model misfits using the following L_2 -norm and L_1 -norm to calculate the misfits, respectively,

$$J_d = \frac{\|(\mathbf{d}_{\text{obs}} - \mathbf{d}(\mathbf{m}))\|_2^2}{\max(\mathbf{d}_{\text{obs}} - \mathbf{d}(\mathbf{m}_{\text{initial}}))} \quad (11)$$

$$J_m = \frac{\|(\mathbf{m}_{\text{true}} - \mathbf{m})\|_1^1}{\max(\mathbf{m}_{\text{true}} - \mathbf{m}_{\text{initial}})} \quad (12)$$

where $\mathbf{m}_{\text{initial}}$ represents the initial velocity model. It is seen that SGVFWI has the lowest model misfit and a similar data misfit compared with other methods, which further demonstrates the effectiveness of the proposed method. In the next numerical experiment, we employ the shearlet transform and TGV regularization in a more complex velocity model.

B. SEG/EAGE Overthrust Model

To test the effectiveness of the regularized FWI with shearlet transform and TGV, we adopt a portion of a 2-D SEG/EAGE overthrust model [52] as the velocity mode [see Fig. 8(a)]. Fig. 8(b) shows the starting model obtained by smoothing the velocity model with a Gaussian smoothed function where the standard deviation is 80. The starting model would be smoother with the standard deviation increases.

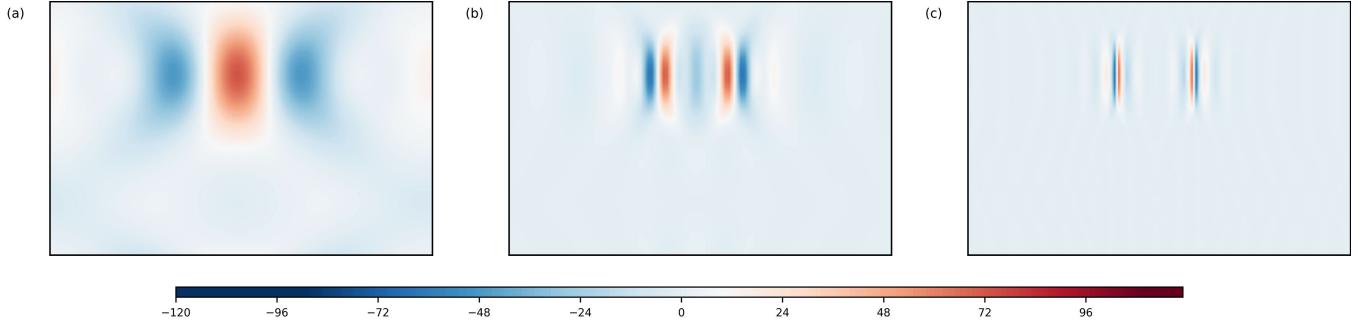


Fig. 4. Different scale shearlet coefficients of the simple synthetic model. (a) First-scale shearlet coefficient. (b) Second-scale shearlet coefficient. (c) Third-scale shearlet coefficient.

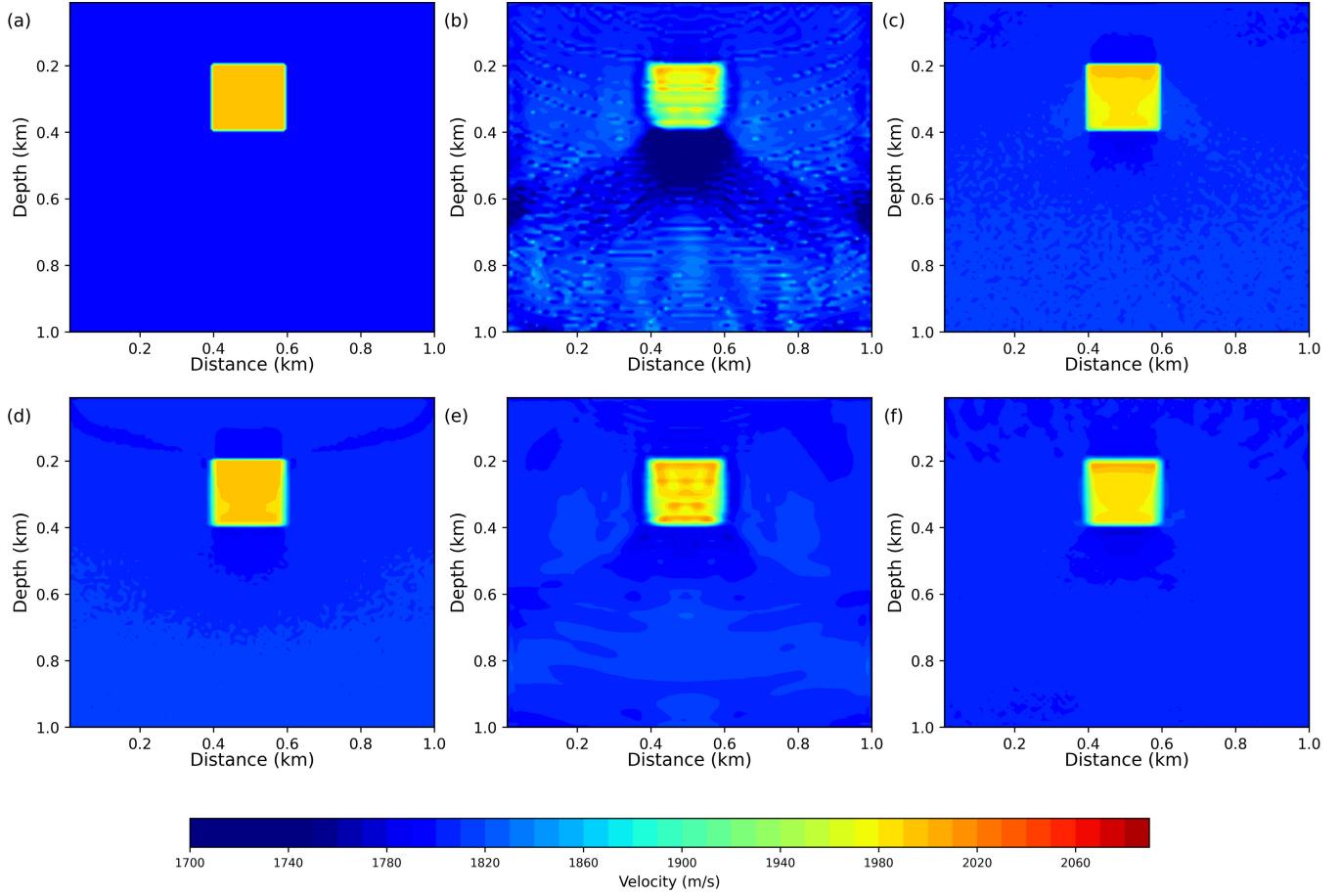


Fig. 5. Inversion results of the simple synthetic model with different methods. (a) True model. (b) TRAFWI. (c) TVFWI. (d) TGVFWI. (e) STFWI. (f) SGVFWI.

Considering the efficiency, calculation complexity, and performance of FWI, the velocity model is composed of 300 and 130 grid points in the horizontal and vertical directions, respectively, with a grid size of 25 m in both directions. A Ricker wavelet with a dominant frequency of 10 Hz is used as the source wavelet. We use DeepWave to perform the constant-density acoustic finite-difference modeling. Next, we generate 30 equally spaced shots—indicated by the red triangle in Fig. 8(a) with 300 equally spaced receivers for each shot. The shot interval and the receiver spacing are 250 and 25 m, respectively. Fig. 9 shows the middle-shot gather of the observed data without noise.

First, we test the effectiveness of STFWI. The FFST uses three scale shearlets (see Fig. 3) to extract multiscale

information from the velocity model. Fig. 10 shows different scale shearlet coefficients obtained by FFST to mitigate the ill-posedness of the problem. The shearlet coefficients can contain a larger range of the model with the increase of the scale parameter. Meanwhile, the third shearlet coefficients include the entire model, so the scale parameter of 3 is appropriate. We set the tradeoff parameter (λ) as 0.01, 0.001, and 0.0001, respectively, to find a perfect one for STFWI. Next, we use the Adam optimizer with a learning rate of 50 to solve the nonlinear inverse problem. From Fig. 11, STFWI with $\lambda = 0.0001$ produces the best result among these tradeoff parameters but has a small improvement compared with TRAFWI. If the tradeoff parameter is large, the shearlet transform regularization could cause the inversion to fall

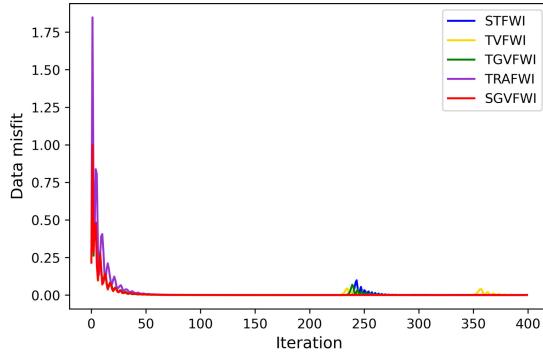


Fig. 6. Data misfits of the simple synthetic model.

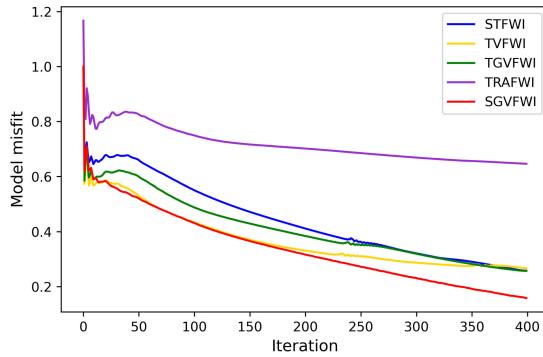


Fig. 7. Model misfits of the simple synthetic model.

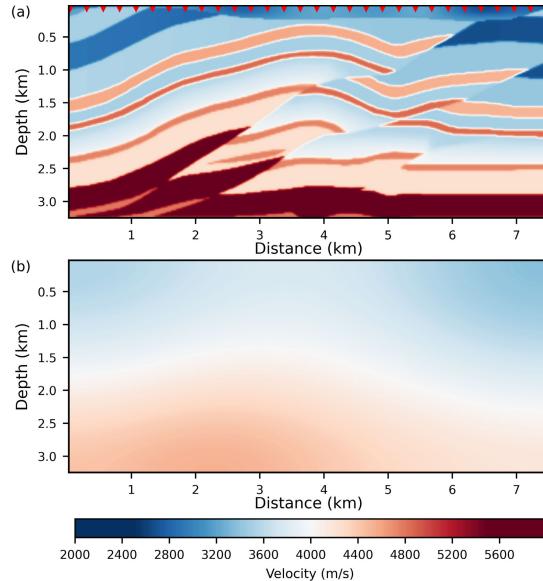


Fig. 8. Overthrust model. (a) True model with 30 equally spaced shots (red triangles). (b) Starting model generated by a Gaussian smoothed function.

into a local minima due to the magnitude of the gradient calculated by the data-fidelity term decreases with the increase of the model depth. However, the magnitude of the multiscale shearlets is in the same order. Therefore, the difference would finally result in a wrong velocity model, that is, the gradient of the data-fidelity term is significantly smaller than the one of the regularizations. To ensure the effectiveness of the shearlet transform regularization, we design a matrix called \mathbf{W}_d to improve the resolution of the inversion result obtained by the

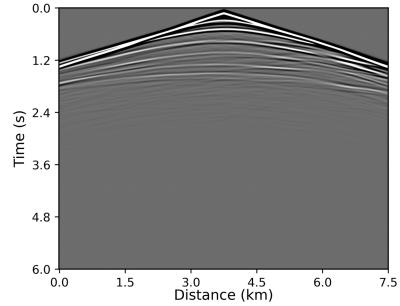


Fig. 9. Middle-shot gather of the overthrust model.

Hadamard product of two matrices shown in Fig. 12. The matrix \mathbf{W}_m assigns large values from the first column to the 50th column and from the 251st column to the 300th column, and the values at the first column and the 300th column decrease from 5 to 1 along the row and gradually decay inward to 1. The row of the matrix \mathbf{W}_n linearly increases from 1 to 1000 along the column from 1 to 90 and linearly increases from 1000 to 30 000 along the column from 91 to 130. The motivation for designing the matrix \mathbf{W}_d is to correct the gradient of the data-fidelity term, so the gradient in the deep and shallow regions of the model is in the same order, ensuring the effectiveness of shearlet transform regularization. Fig. 11(f) shows the inversion result of STFWI with \mathbf{W}_d , which produces the best resolution compared with other results. Figs. 13 and 14 show convergence diagrams of data misfits and model misfits after 500 iterations, respectively. Although the data misfits of these methods are similar to one another, the model misfits are significantly different, which demonstrates that \mathbf{W}_d is appropriate for STFWI. Therefore, we employ STFWI with \mathbf{W}_d in the next numerical experiments.

Next, we compare five inversion methods: TRAFWI, TVFWI, TGVFWI, STFWI, and SGVFWI. The tradeoff parameter of TVFWI is 0.005. We set parameters α_1 and α_0 of the TGVFWI to 0.003 and 0.0015, respectively. The SGVFWI method uses the same regularization as STFWI and TGVFWI, \mathbf{W}_d is shown in Fig. 12, and λ , α_1 , and α_0 are 1, 4, and 2, respectively. We use the Adam optimizer with a learning rate of 50 to solve FWI.

The comparisons among these methods after 500 iterations are shown in Fig. 15. TRAFWI well reveals the velocity model represented in Fig. 15(b) but still contains some artifacts in the inversion result. Regularized FWI can smooth the velocity model and generates more accurate inversion results than TRAFWI. Fig. 15(c) shows the TVFWI inversion result, which constrains the velocity model using the model gradients in the horizontal and vertical directions. TVFWI improves the deep structures of the inversion results but causes staircasing effects. TGVFWI overcomes the staircasing effect by introducing second-order derivatives based on first-order derivatives. Fig. 15(d) shows that the TGVFWI inversion result is smoother than that of TVFWI. The STFWI adopts the advantages of multiscale and multidirectional properties from the shearlet transform to obtain a high-resolution inversion result [see Fig. 15(e)]. The STFWI result has accurate interfaces and smooth variations; however, there are some banded anomalies around the boundary, called boundary effects. SGVFWI

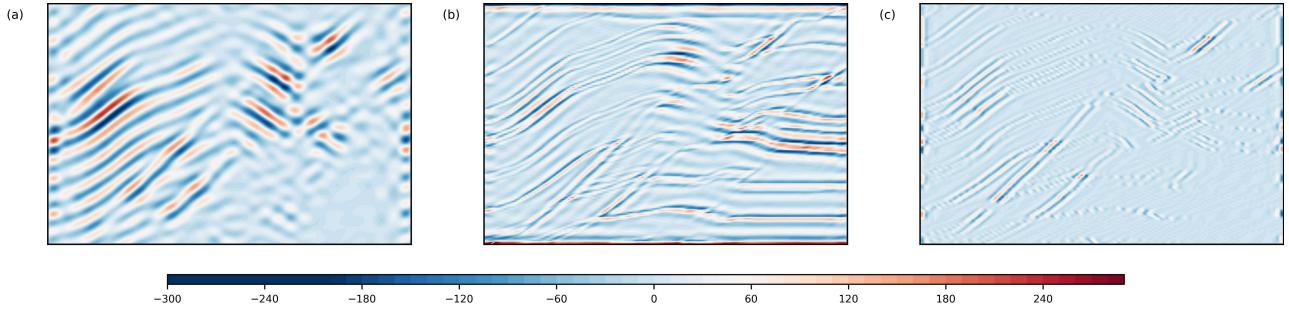


Fig. 10. Different scale shearlet coefficients of the overthrust model. (a) First-scale shearlet coefficient. (b) Second-scale shearlet coefficient. (c) Third-scale shearlet coefficient.

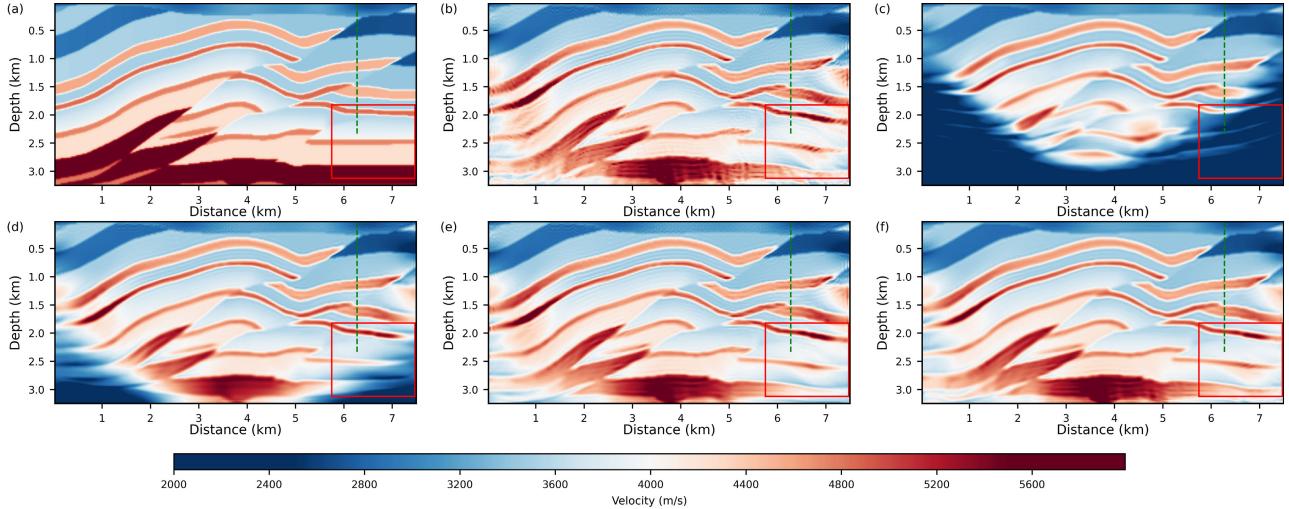


Fig. 11. Inversion results of the overthrust model with different parameters. (a) True model. (b) TRAFWI. STFWI with (c) $\lambda = 0.01$, (d) $\lambda = 0.001$, and (e) $\lambda = 0.0001$. (f) Matrix \mathbf{W}_d .

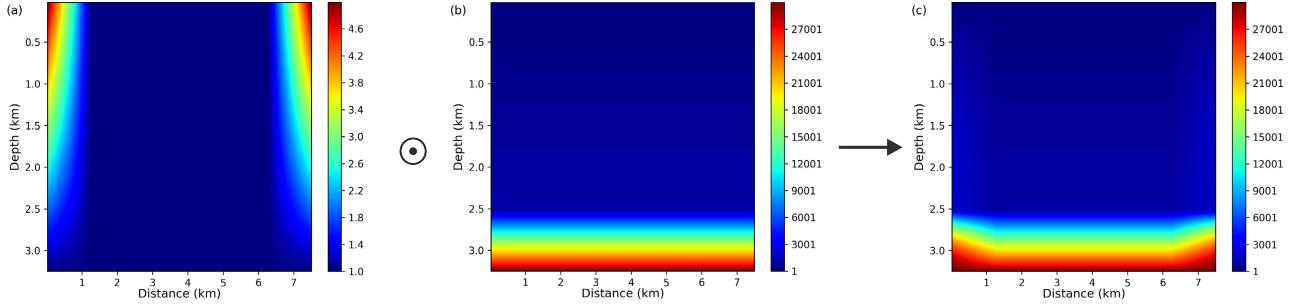


Fig. 12. Generation of the matrix \mathbf{W}_d . (a) Matrix \mathbf{W}_m . (b) Matrix \mathbf{W}_n . (c) Matrix \mathbf{W}_d .

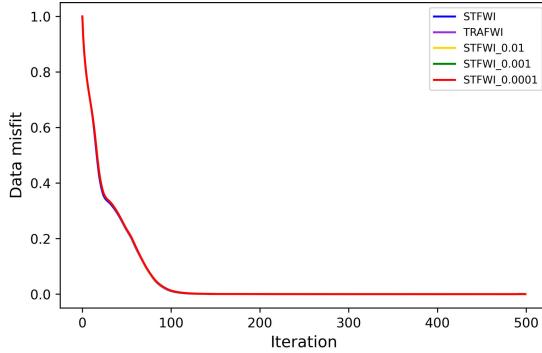


Fig. 13. Data misfits of the overthrust model.

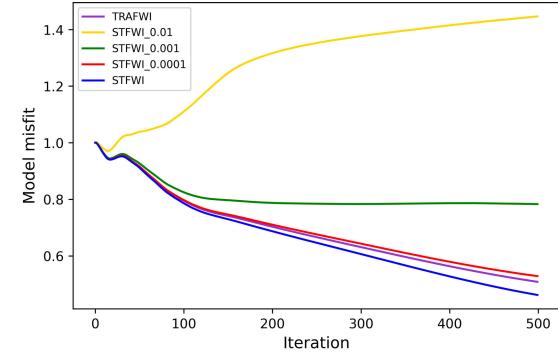


Fig. 14. Model misfits of the overthrust model.

reconstructs the best velocity model [see Fig. 15(f)], which overcomes the TVFWI, TGVFWI, and STFWI drawbacks. This method employs the nature of TGV to address the

boundary effect of the shearlet transform, resulting in more accurate inversion results.

Figs. 16 and 17 show the comparisons of different methods at red rectangle and green line in Fig. 15. We select a

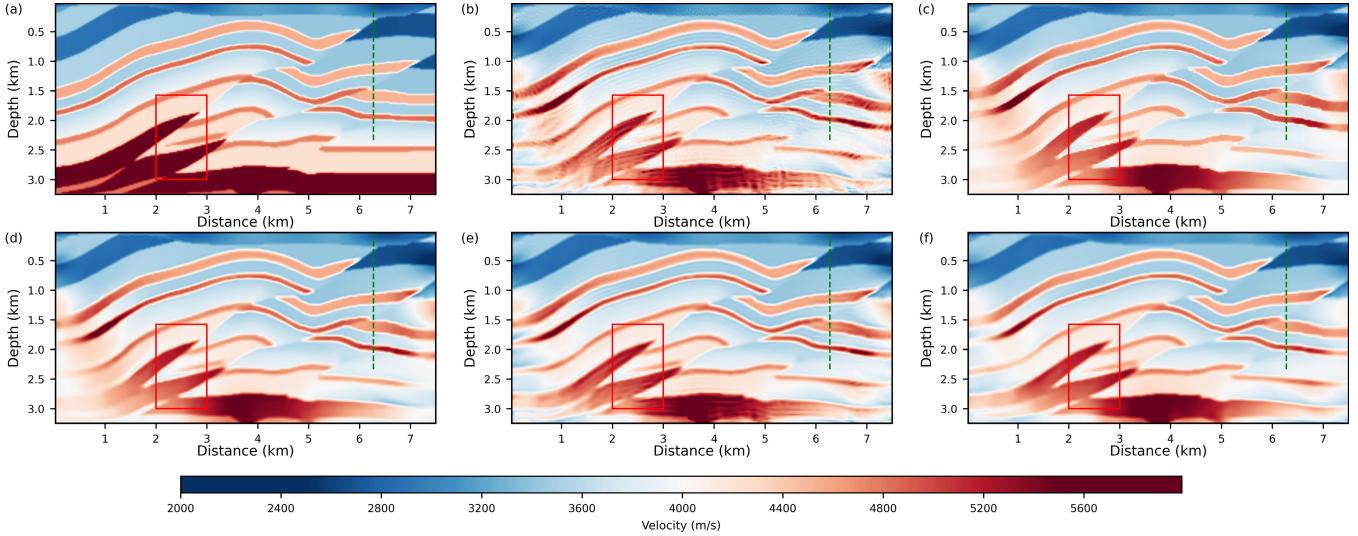


Fig. 15. Inversion results of the overthrust model with different methods. (a) True model. (b) TRAFWI. (c) TVFWI. (d) TGVFWI. (e) STFWI. (f) SGVFWI.

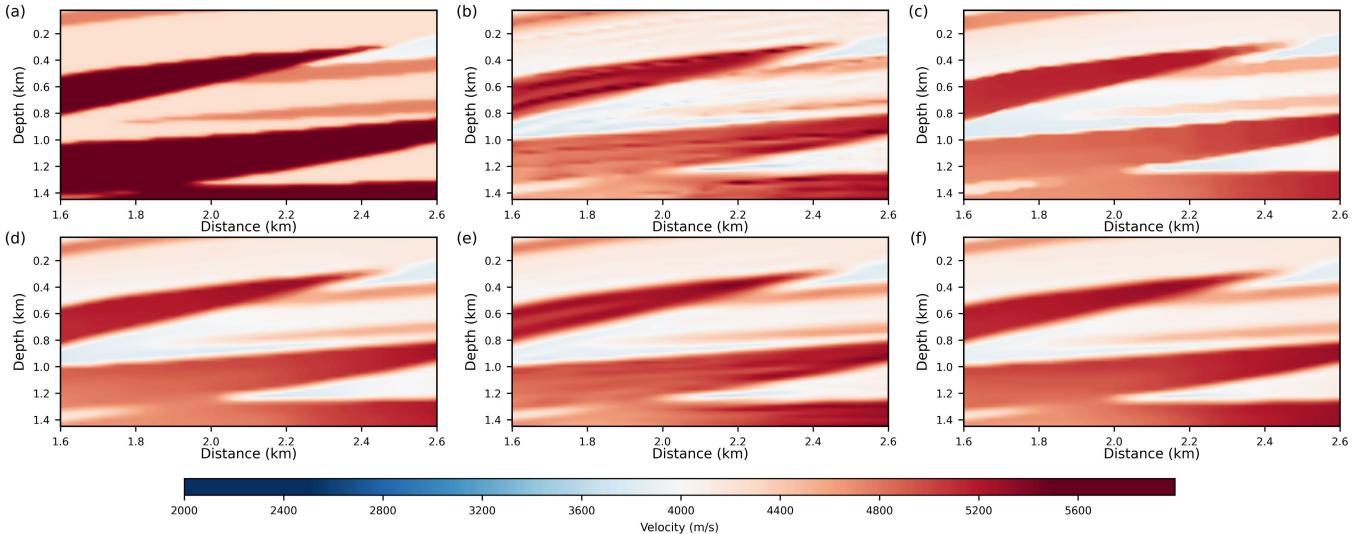


Fig. 16. Portions of inversion results for the overthrust model with different methods. (a) True model. (b) TRAFWI. (c) TVFWI. (d) TGVFWI. (e) STFWI. (f) SGVFWI.

region with full data coverage for 2-D comparison, and the results between SGVFWI and TGVFWI are similar. For 1-D examples, SGVFWI can produce the most similar inversion result to the true one. Fig. 18 represents the data misfits of the above five methods. They all converge to a similar value. SGVFWI gives the minimal model misfit between these methods shown in Fig. 19, demonstrating that the proposed method can mitigate the ill-posed nature of FWI. However, these regularized FWI methods produce similar inversion results, and SGVFWI significantly improves the resolution of the velocity model at the boundaries. This numerical experiment cannot convincingly illustrate the advantages of SGVFWI, so we use the following complex model to test the method.

C. Marmousi Model

To further demonstrate the effectiveness of the proposed method, we test it on the velocity model by selecting a portion

of the Marmousi model [53]. The velocity model is composed of 330 and 140 grid points in the horizontal and vertical directions, respectively, with a grid size of 20 m in both directions. Subsequently, we use a Gaussian smoothed function to smooth the velocity model to obtain the starting model where the standard deviation is 80. Fig. 20(a) and (b) shows the velocity and starting models, respectively. A Ricker wavelet with a dominant frequency of 8 Hz is used as the source wavelet. We employ DeepWave to perform constant-density acoustic finite-difference modeling. We then generate 32 equally spaced shots [red triangles in Fig. 20(a)] with 330 equally spaced receivers for each shot. The shot spacing and the receiver interval are 200 and 20 m, respectively. We obtain the observed data by adding random noise with a signal-to-noise ratio of 7 dB to the synthetic data. Fig. 21(a) and (b) shows the middle-shot gather of the synthetic and observed data. Fig. 22 shows different scale shearlet coefficients to constrain the velocity model.

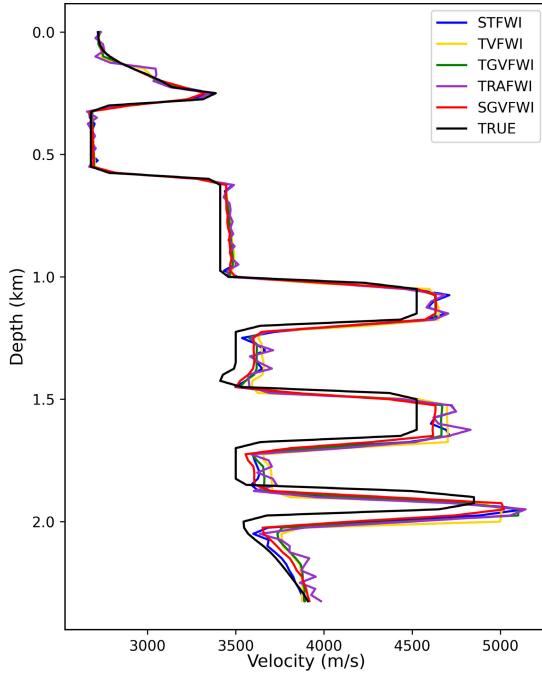


Fig. 17. Comparison of velocities for the overthrust model at Distance = 6.30 km.

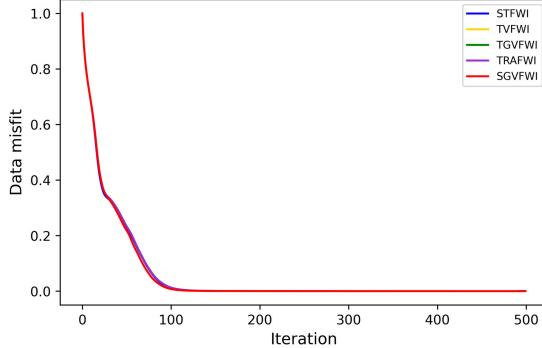


Fig. 18. Data misfits of the overthrust model.

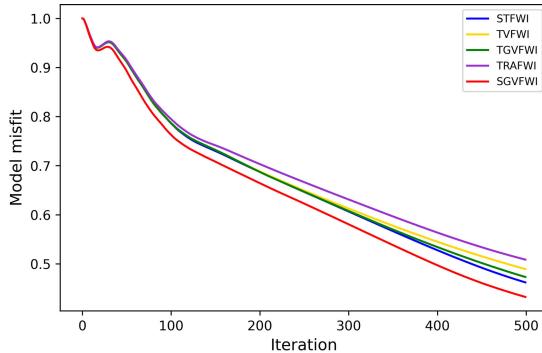


Fig. 19. Model misfits of the overthrust model.

Next, we adopt the same five methods as described in Section III-B to compare the inversion results. For TGVFWI, parameters α_1 and α_0 are 0.05 and 0.025, respectively. The tradeoff parameter of TVFWI is 0.1. STFWI sets the tradeoff parameter to 1 and designs the matrix \mathbf{W}_d by the Hadamard product of the matrices \mathbf{W}_m and \mathbf{W}_n . \mathbf{W}_m contains large values from the first column to the 50th column and from

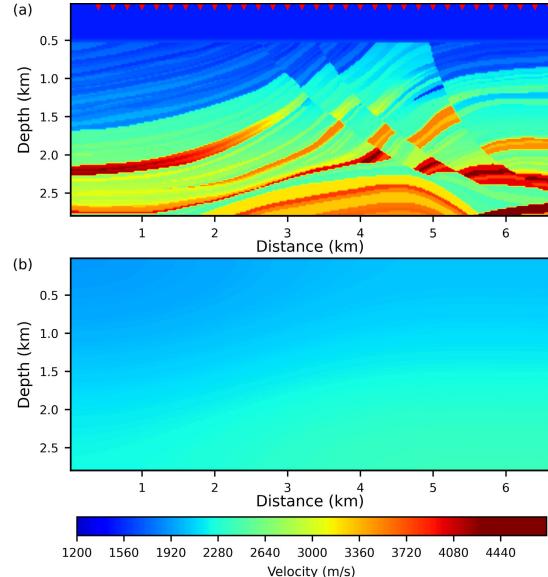


Fig. 20. Marmousi model. (a) True model with 32 equally spaced shots (red triangles). (b) Starting model generated by a Gaussian smoothed function.

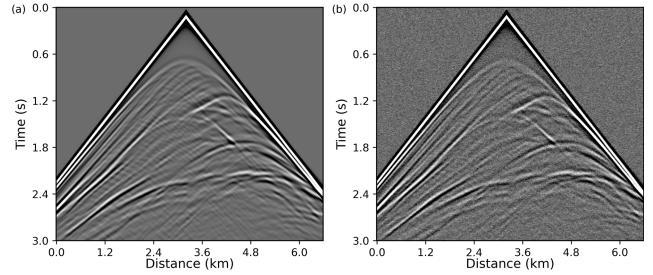


Fig. 21. Middle-shot gathers of the Marmousi model. (a) Synthetic data. (b) Observed data.

the 281st column to the 330th column, and the values at the first column and the 330th column decrease from 5 to 1 along the row and gradually decay inward to 1. The row of \mathbf{W}_n linearly increases from 1 to 100 along the column from 1 to 100 and linearly increases from 100 to 10 000 along the column from 101 to 140. For SGVFWI, \mathbf{W}_d is the same as the one used in STFWI, λ , α_1 and α_0 are 1, 5, and 2.5, respectively. The Adam algorithm is used to solve FWI with a learning rate of 100.

After 500 iterations, we obtain the inversion results, as shown in Fig. 23. TRAFWI is seriously trapped in local minima, which cannot reconstruct the velocity model [see Fig. 23(b)]. Compared with TRAFWI, regularized FWI can produce more accurate velocity models. Fig. 23(c) shows the inversion result of TVFWI, which uses the model gradients in the horizontal and vertical directions to constrain the velocity model. TVFWI suppresses discontinuities in the model but causes staircasing effects, while TGVFWI uses second-order derivatives based on first-order derivatives to produce a smoother result than that of TVFWI [see Fig. 23(d)]; however, the above two regularized methods cannot effectively address the ill-posedness of the nonlinear inverse problem, which means that the inversion results are not accurate. The STFWI uses the advantages of multiscale and multidirectional properties from the shearlet transform to improve the inversion

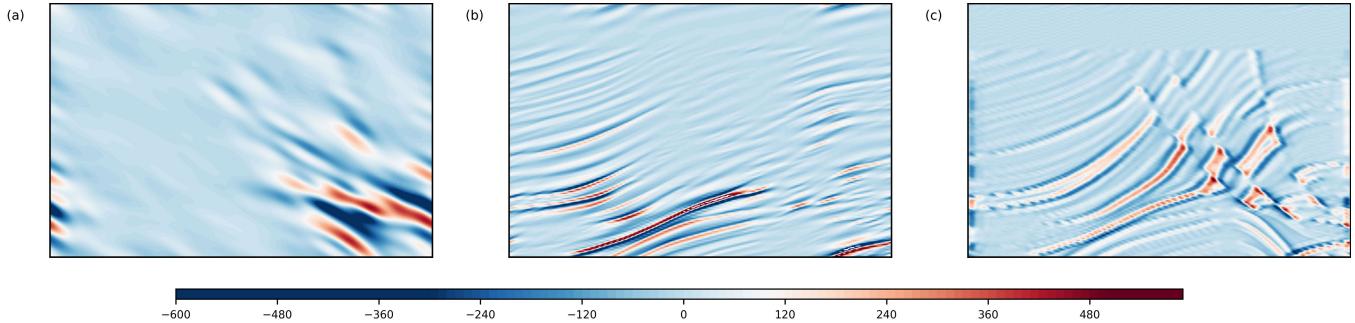


Fig. 22. Different scale shearlet coefficients of the Marmousi model. (a) First-scale shearlet coefficient. (b) Second-scale shearlet coefficient. (c) Third-scale shearlet coefficient.

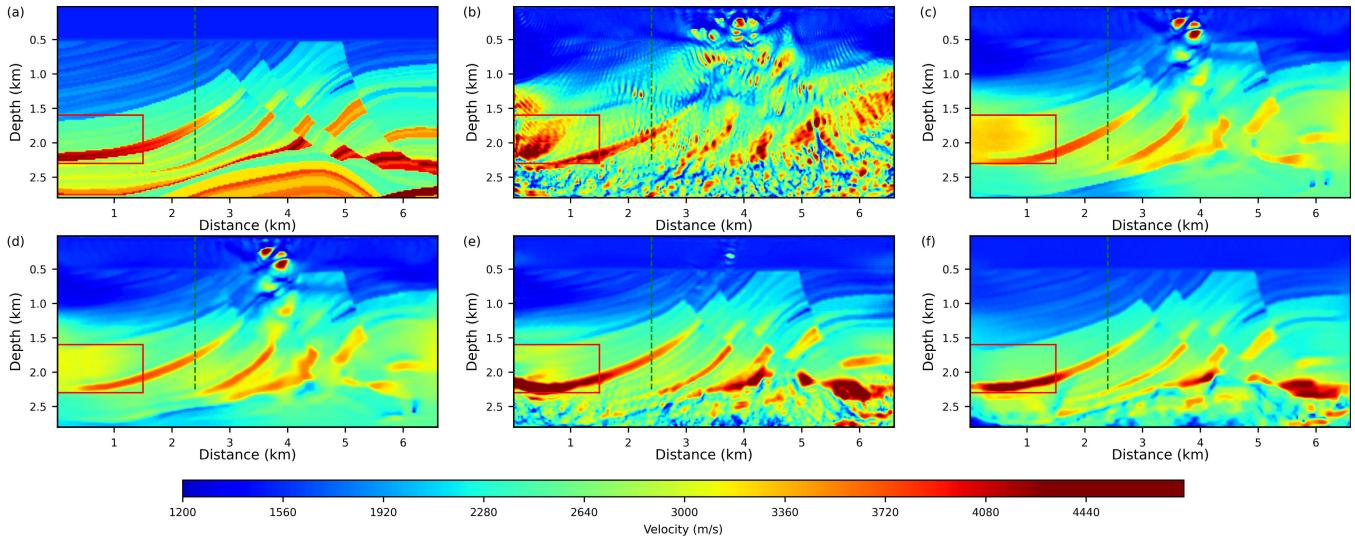


Fig. 23. Inversion results of the Marmousi model with different methods. (a) True model. (b) TRAFWI. (c) TVFWI. (d) TGVFWI. (e) STFWI. (f) SGVFWI.

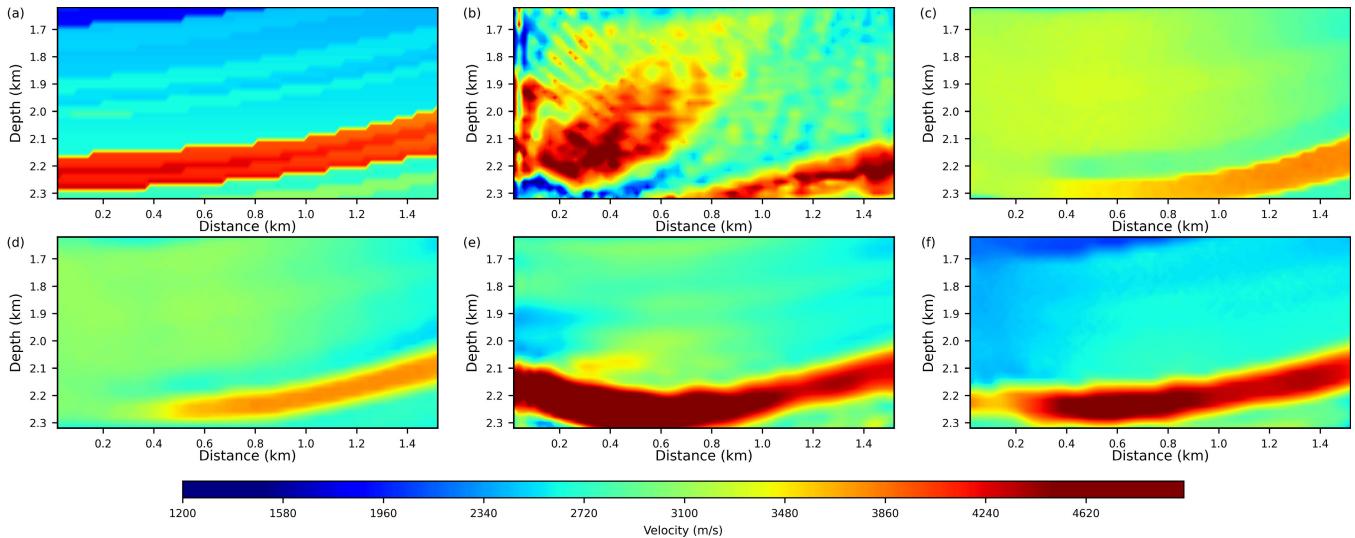


Fig. 24. Portions of inversion results for the Marmousi model with different methods. (a) True model. (b) TRAFWI. (c) TVFWI. (d) TGVFWI. (e) STFWI. (f) SGVFWI.

result [see Fig. 23(e)]. The velocity model of STFWI is higher resolution than the others, but there are still some banded anomalies in the deep area and boundary. Therefore, introducing TGV to smooth the discontinuities is important to obtain an accurate inversion result. Compared with TVFWI, TGVFWI, and STFWI, SGVFWI reconstructs the best velocity

model [see Fig. 23(f)]. This method combines the advantages of TGV and shearlet transform to ensure the accuracy of the inversion result.

To further illustrate the contribution of different regularized methods, we show 2-D and 1-D comparisons at red rectangle and green line in Fig. 23. From Figs. 24 and 25, we find

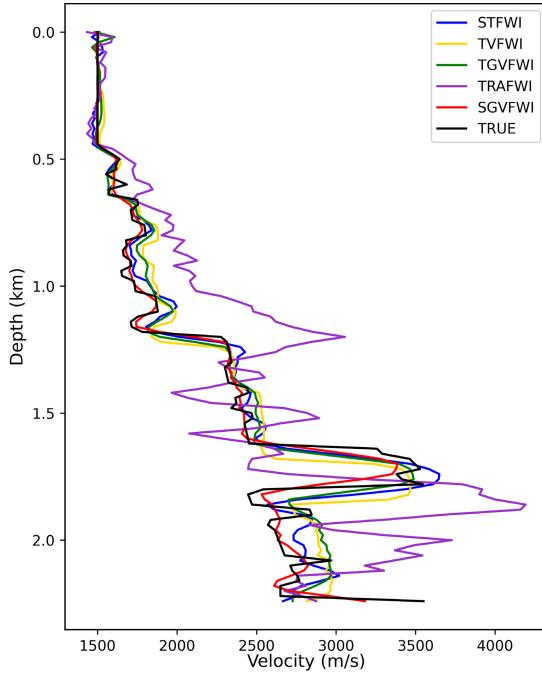


Fig. 25. Comparison of velocities for the Marmousi model at Distance = 2.26 km.

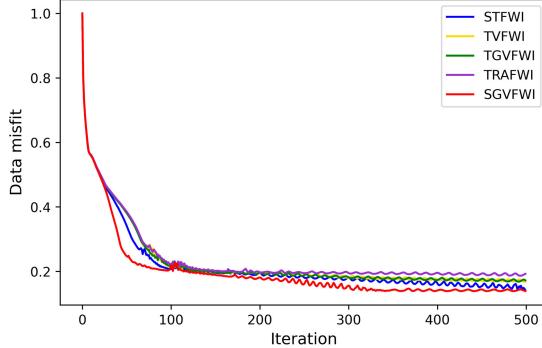


Fig. 26. Data misfits of the Marmousi model.

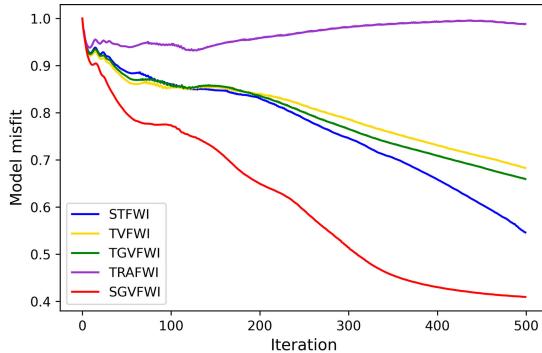


Fig. 27. Model misfits of the Marmousi model.

that SGVFWI produces the most accurate inversion result, which further demonstrates the advantages of the shearlet transform and TGV regularization. Figs. 26 and 27 represent the convergence diagrams of data and model misfits, respectively. SGVFWI gives the minimal data misfit and model misfit compared with other regularized methods, which shows that the proposed method can mitigate the ill-posedness

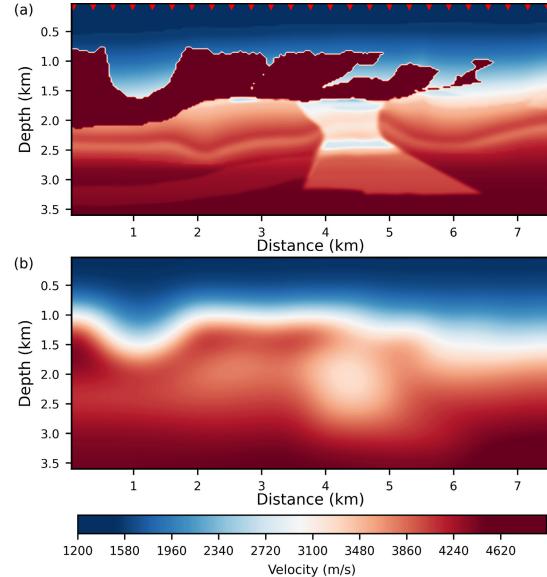


Fig. 28. Modified 2004 BP model. (a) True model with 25 equally spaced shots (red triangles). (b) Starting model generated by a Gaussian smoothed function.

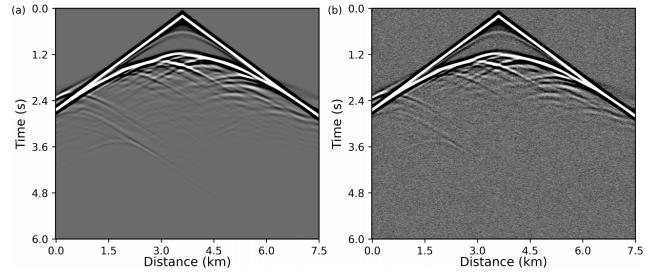


Fig. 29. Middle-shot gathers of the modified 2004 BP model. (a) Synthetic data. (b) Observed data.

of FWI. Therefore, if the inversion condition is difficult as inaccurate initial models and strong noise pollution, SGVFWI could achieve higher-resolution inversion results than other regularized methods for the velocity model of layered medium such as the Marmousi model. In addition, these regularized methods can all reasonably reconstruct the velocity model when TRAFWI is not seriously trapped in the local minima, which is unnecessary to select the combined regularization to improve the inversion result.

D. Modified 2004 BP Model

We select a modified 2004 BP model [47] to further test the effectiveness of the proposed method. The velocity model is composed of 120 and 250 grid points in the vertical and horizontal directions, respectively, with a grid size of 30 m in both directions. Due to the difficulty of the time-domain FWI based on the modified 2004 BP model, a Gaussian smoothed function is applied here to smooth the velocity model to generate the starting model where the standard deviation is 15. The velocity and starting models are shown in Fig. 28. We select the Ricker wavelet with a dominant frequency of 6 Hz as the source wavelet and use DeepWave to perform the constant-density acoustic finite-difference modeling. Next, 25 equally spaced shots are generated [red triangles in Fig. 28(a)] with 250 equally spaced receivers for each

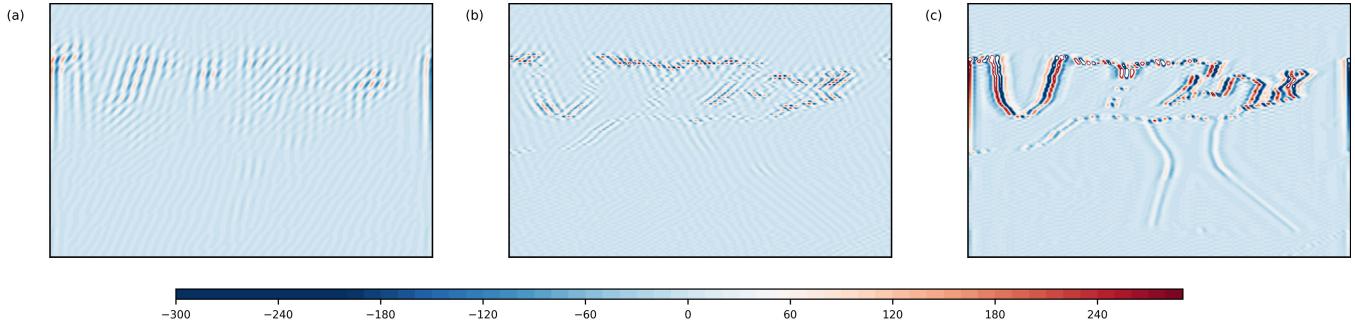


Fig. 30. Different scale shearlet coefficients of the modified 2004 BP model. (a) First-scale shearlet coefficient. (b) Second-scale shearlet coefficient. (c) Third-scale shearlet coefficient.

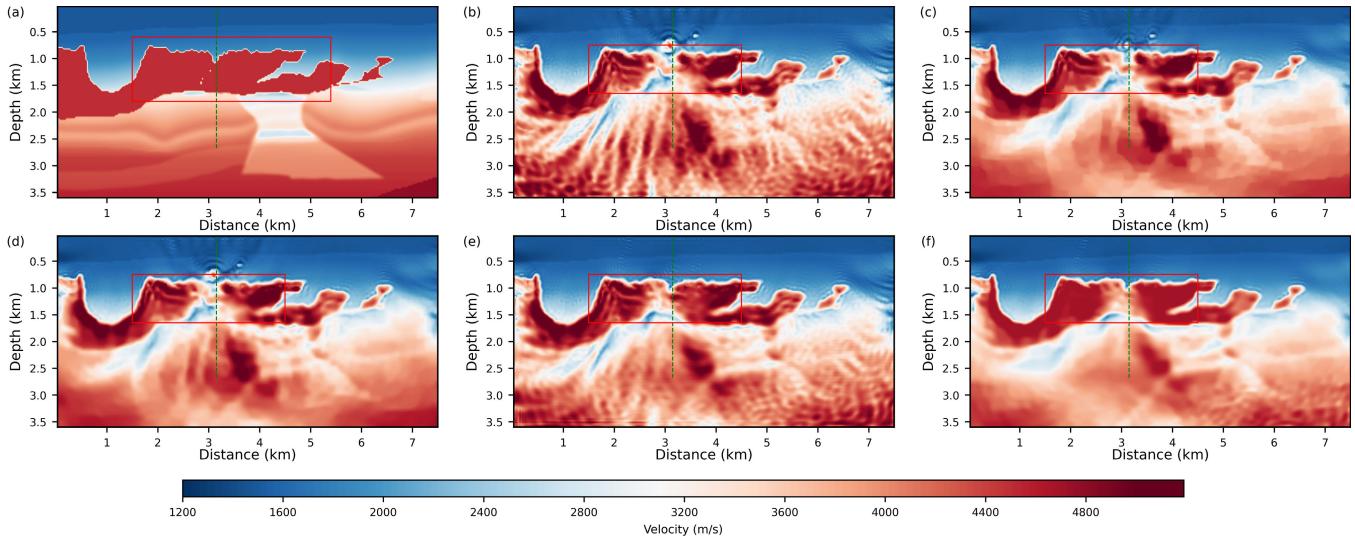


Fig. 31. Inversion results of the modified 2004 BP model with different methods. (a) True model. (b) TRAFWI. (c) TVFWI. (d) TGVFWI. (e) STFWI. (f) SGVFWI.

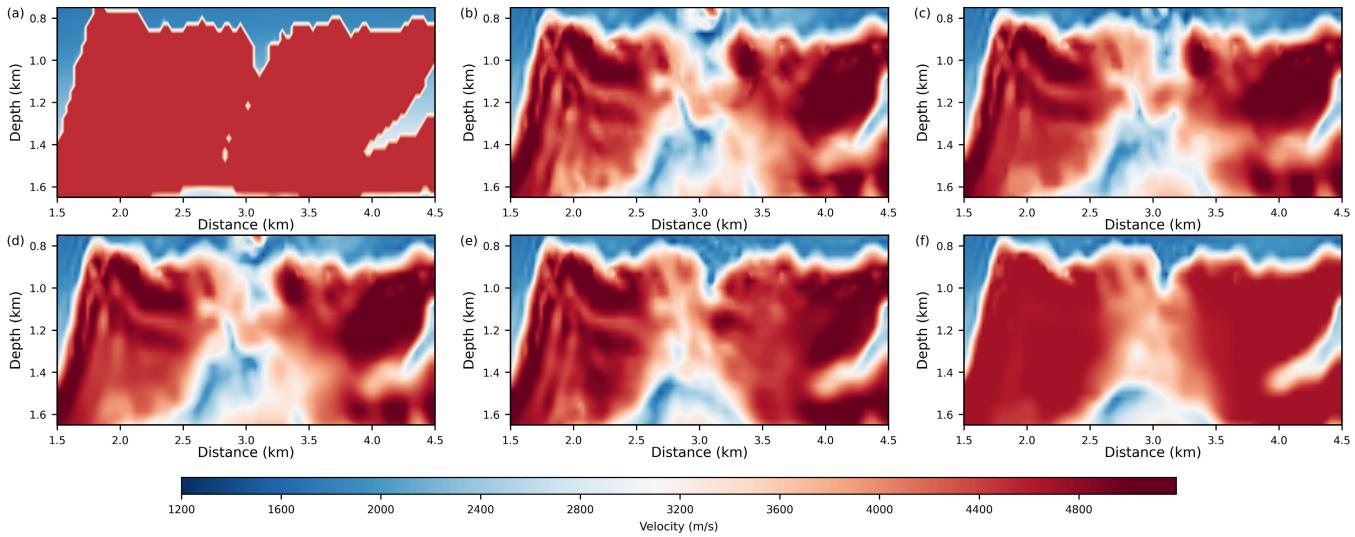


Fig. 32. Portions of inversion results for the modified 2004 BP model with different methods. (a) True model. (b) TRAFWI. (c) TVFWI. (d) TGVFWI. (e) STFWI. (f) SGVFWI.

shot. We set the shot spacing and the receiver interval to 300 and 30 m and add random noise with a signal-to-noise ratio of 10 dB to the synthetic data to produce the observed data. Fig. 29(a) and (b) shows the middle-shot gather of the synthetic and observed data, respectively. To illustrate the mechanism of the shearlet transform regularization, different

scale shearlet coefficients are shown in Fig. 30 to constrain the velocity model. The Adam optimizer is used to solve the nonlinear inverse problem with a learning rate of 30.

We choose the above five methods to compare the inversion results. For TVFWI, the tradeoff parameter is defined as 0.01. TGVFWI sets α_1 and α_0 to 0.005 and 0.0025, respectively. For

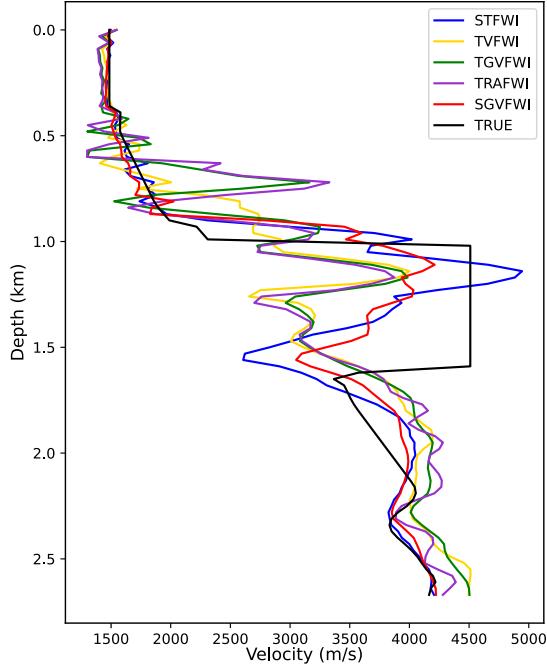


Fig. 33. Comparison of velocities for the modified 2004 BP model at Distance = 3.15 km.

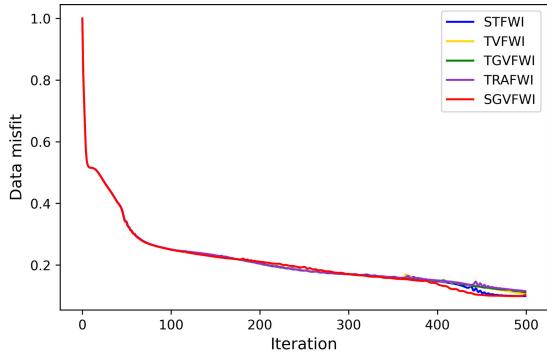


Fig. 34. Data misfits of the modified 2004 BP model.

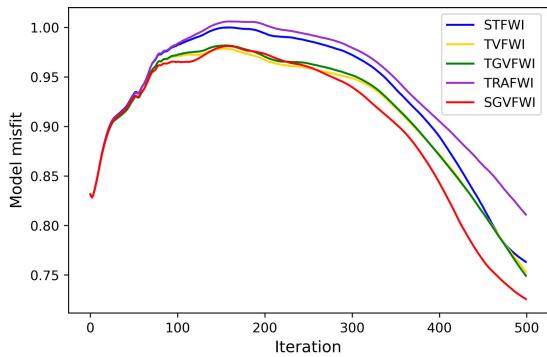


Fig. 35. Model misfits of the modified 2004 BP model.

STFWI, we define the tradeoff parameter as 1 and generate the matrix \mathbf{W}_d by the Hadamard product of the matrices \mathbf{W}_m and \mathbf{W}_n . The value of the \mathbf{W}_m matrix at the first column and the 250th column decreases from 20 to 1 along the row and gradually decays inward to 1. The row of \mathbf{W}_n linearly increases from 1 to 300 along the column from 1 to 90 and linearly increases from 300 to 3000 along the column from 91 to 120.

SGVFWI employs the same \mathbf{W}_d as the one used in STFWI and λ , α_1 , and α_0 are 1, 3.5, and 1.75, respectively.

Fig. 31 shows the inversion results with different methods after 500 iterations. Although we have generated a relatively accurate starting guess, TRAFWI is still trapped in the local minima and contains strong artifacts, demonstrating that the modified 2004 BP model is difficult for the time-domain FWI without any technology to alleviate the cycle skipping. Regularized FWI methods reduce the artifacts to obtain more accurate velocity models compared with TRAFWI, but still cannot well reconstruct the deep sediment of the model. Fig. 31(c) and (d) represents similar inversion results of TVFWI and TGVFWI, but TGV can reduce the staircasing effect to produce a smoother result. However, the above velocity models are not accurate to reveal the high-velocity body and the deep sediment. STFWI applies multiscale and multi-directional properties of shearlet transform to extract abundant information to improve the resolution of the high-velocity body [see Fig. 31(e)]. However, the inversion result contains many artifacts that smear the deep structure. Therefore, we combine TGV with shearlet transform to decrease the artifacts of the velocity model shown in Fig. 31(f). SGVFWI achieves the best high-velocity body among these regularized methods, but there is still a large difference in the deep structure compared with the true model.

Figs. 32 and 33 show 2-D and 1-D comparisons at red rectangle and green line in Fig. 31 to illustrate the results of different methods. SGVFWI reveals the best result of the high-velocity body compared with other regularized methods. The convergence plots of data and model misfits are shown in Figs. 34 and 35. We find that the data misfits of these methods are all converged to a similar value, but the model misfits increase first and then decrease. Fig. 35 illustrates that the regularized methods are first trapped in a local minima and then mitigate the ill-posedness to close to the true model, which shows that the benchmark model is a major challenge for FWI and the regularized methods are hard to obtain an accurate inversion result because of the serious cycle skipping.

IV. DISCUSSION

Here, we introduce a novel regularized method for FWI that uses shearlet transform and TGV to constrain the velocity model. Below, we discuss the details of parameter selection, computational cost, and high-dimensional FWI for our new method.

Considering the selection of parameters, we manually adjust more parameters of the proposed method than those of conventional regularized inversion methods. The gradient calculation of the shearlet transform is more complicated than that of TV and TGV, which involves multiscale information to constrain the velocity model. Therefore, we have to design a matrix \mathbf{W}_d to correct the gradient of the data-fidelity term to ensure the similar magnitude of the gradient, which is the most important parameter in (10). Numerous numerical experiments have proven that \mathbf{W}_d is suitable for STFWI, and we can generate \mathbf{W}_d by the Hadamard product of two matrices, both of the matrices can be easily produced linearly increasing or decreasing. Regarding the TGV parameters, α_1 and α_0 control

TABLE I
COMPUTATIONAL COST OF DIFFERENT METHODS FOR EACH ITERATION

Models	TRAFWI	TVFWI	TGVFWI	STFWI	SGVFWI
Overthrust model	0.907s	0.910s	0.916s	0.924s	0.928s
Marmousi model	0.358s	0.360s	0.362s	0.398s	0.399s
Benchmark model	0.392s	0.394s	0.396s	0.410s	0.418s

the weights of the first- and second-order model derivatives, respectively. We can increase the value of α_0 if the model needs to be smoother. The proposed method aims to employ TGV to improve the boundary effect of the shearlet transform and obtain high-resolution inversion results. Therefore, the parameter λ is usually smaller than those of α_1 and α_0 . Because the parameters provided by the proposed methods are different in each numerical experiment, we do not define the specific value of each parameter but share the tuning experience that we have summarized.

Regarding the computational cost of FWI, DeepWave is used to implement the proposed method to perform wave propagation and FWI with the GPU acceleration technology. Meanwhile, DeepWave replaces traditional gradient calculation with automatic differentiation technology to easily produce gradients based on Pytorch [54]. To improve computational efficiency, we have completed the tests of all inversion methods on the RTX4080 GPU, the computational cost of different methods for each iteration is shown in Table I. The computational cost of the proposed method and comparison methods is in the same order. Although the proposed method has the highest computational cost, it is acceptable for the increased cost with no more than 1.12 times that of TRAFWI.

Considering the 3-D FWI, the proposed method can theoretically obtain accurate inversion results. Both the 3-D shearlet transform and the 3-D TGV have complete descriptions that have been used in some fields to achieve ideal results. Although the proposed method can be used for 3-D cases, some challenges remain in solving 3-D FWI problems. First, the computational cost of 3-D FWI is expensive, and the proposed method has many manually adjusted parameters. Second, the construction of \mathbf{W}_d in (10) is more complex than that in the 2-D cases, which cannot be obtained using a linearly increasing method. We must reasonably generate \mathbf{W}_d from the gradient of the data-fidelity term to maintain a consistent order of magnitude for the gradient.

V. CONCLUSION

We have developed a novel regularized FWI method with shearlet transform and TGV that uses Adam optimizer to obtain accurate inversion results. We have applied the proposed method to the inversion of the observed data based on the simple synthetic, 2-D SEG/EAGE overthrust, Marmousi, and modified 2004 BP models, and compared it with TRAFWI, TVFWI, TGVFWI, and STFWI to demonstrate the effectiveness of the method. Numerical experiments show that the proposed method combines the advantages of shearlet transform and TGV; that is, shearlet transform extracts multiscale and multidirectional information to constrain the velocity model, which reduces artifacts with retaining abundant structure details; TGV alleviates the staircasing effect of

TV regularization and can smooth the boundary effect caused by shearlet transform regularization. Therefore, the proposed method can mitigate the ill-posedness of FWI and produce a high-resolution velocity model with a smooth background variation and accurate interfaces. However, the proposed method still hard to reveal a high-resolution inversion result under serious cycle skipping.

REFERENCES

- [1] A. Tarantola, "Inversion of seismic reflection data in the acoustic approximation," *Geophysics*, vol. 49, no. 8, pp. 1259–1266, Aug. 1984.
- [2] J. Virieux and S. Operto, "An overview of full-waveform inversion in exploration geophysics," *Geophysics*, vol. 74, no. 6, pp. 1–26, Nov. 2009.
- [3] C. Shin and D.-J. Min, "Waveform inversion using a logarithmic wavefield," *Geophysics*, vol. 71, no. 3, pp. 31–42, May 2006.
- [4] E. Bozdag, J. Trampert, and J. Tromp, "Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements: Misfit functions for full waveform inversion," *Geophys. J. Int.*, vol. 185, no. 2, pp. 845–870, May 2011.
- [5] J. Luo and R. Wu, "Seismic envelope inversion: Reduction of local minima and noise resistance," *Geophys. Prospecting*, vol. 63, no. 3, pp. 597–614, May 2015.
- [6] T. Van Leeuwen and W. A. Mulder, "A correlation-based misfit criterion for wave-equation traveltime tomography," *Geophys. J. Int.*, vol. 182, no. 3, pp. 1383–1394, Sep. 2010.
- [7] B. Engquist and B. D. Froese, "Application of the wasserstein metric to seismic signals," *Commun. Math. Sci.*, vol. 12, no. 5, pp. 979–988, 2014.
- [8] L. Metivier, R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, "Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion," *Geophys. J. Int.*, vol. 205, no. 1, pp. 345–377, Feb. 2016.
- [9] M. Warner and L. Guasch, "Adaptive waveform inversion: Theory," *Geophysics*, vol. 81, no. 6, pp. R429–R445, Nov. 2016.
- [10] Y. Huang et al., "Toward high-fidelity imaging: Dynamic matching FWI and its applications," *Lead. Edge*, vol. 42, no. 2, pp. 124–132, Feb. 2023.
- [11] F. Yang and J. Ma, "FWIGAN: Full-waveform inversion via a physics-informed generative adversarial network," *J. Geophys. Research: Solid Earth*, vol. 128, no. 4, Apr. 2023, Art. no. e2022JB025493.
- [12] M. Izzatullah and T. Alkhalifah, " ω -FWI: Robust full-waveform inversion with Fourier-based metric," 2022, *arXiv:2205.09234*.
- [13] W. W. Symes, "Migration velocity analysis and waveform inversion," *Geophys. Prospecting*, vol. 56, no. 6, pp. 765–790, Nov. 2008.
- [14] A. Abubakar, W. Hu, T. M. Habashy, and P. M. van den Berg, "Application of the finite-difference contrast-source inversion algorithm to seismic full-waveform data," *Geophysics*, vol. 74, no. 6, pp. 47–58, Nov. 2009.
- [15] T. van Leeuwen and F. J. Herrmann, "Mitigating local minima in full-waveform inversion by expanding the search space," *Geophys. J. Int.*, vol. 195, no. 1, pp. 661–667, Oct. 2013.
- [16] B. Biondi and A. Almomin, "Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion," *Geophysics*, vol. 79, no. 3, pp. 129–140, May 2014.
- [17] H. S. Aghamiry, A. Gholami, and S. Operto, "Improving full-waveform inversion by wavefield reconstruction with the alternating direction method of multipliers," *Geophysics*, vol. 84, no. 1, pp. R125–R148, Jan. 2019.
- [18] G. Huang, R. Nammour, and W. W. Symes, "Source-independent extended waveform inversion based on space-time source extension: Frequency-domain implementation," *Geophysics*, vol. 83, no. 5, pp. R449–R461, Sep. 2018.
- [19] A. N. Tikhonov, A. Goncharsky, V. Stepanov, and A. G. Yagola, *Numerical Methods for the Solution of Ill-Posed Problems*. Dordrecht, The Netherlands: Springer, 1995.
- [20] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, Nov. 1992.
- [21] 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, Mar. 2013.
- [22] A. Y. Anagaw and M. D. Sacchi, "Full waveform inversion with total variation regularization," in *Proc. CSPG CSEG CWLS Joint Annu. Conv.*, 2011, pp. 1–4.

- [23] E. Esser, L. Guasch, F. J. Herrmann, and M. Warner, "Constrained waveform inversion for automatic salt flooding," *Lead. Edge*, vol. 35, no. 3, pp. 235–239, Mar. 2016.
- [24] H. S. Aghamiry, A. Gholami, and S. Operto, "Hybrid Tikhonov + total-variation regularization for imaging large-contrast media by full-waveform inversion," in *Proc. SEG Tech. Program Expanded Abstr.*, Aug. 2018, pp. 1253–1257.
- [25] S. Qu, E. Verschuur, and Y. Chen, "Full-waveform inversion and joint migration inversion with an automatic directional total variation constraint," *Geophysics*, vol. 84, no. 2, pp. 175–183, Mar. 2019.
- [26] Z. Du, D. Liu, G. Wu, J. Cai, X. Yu, and G. Hu, "A high-order total-variation regularisation method for full-waveform inversion," *J. Geophys. Eng.*, vol. 18, no. 2, pp. 241–252, Apr. 2021.
- [27] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM J. Imaging Sci.*, vol. 3, no. 3, pp. 492–526, Jan. 2010.
- [28] K. Gao and L. Huang, "Acoustic- and elastic-waveform inversion with total generalized p-variation regularization," *Geophys. J. Int.*, vol. 218, no. 2, pp. 933–957, Aug. 2019.
- [29] R. Huber, G. Haberfehlner, M. Holler, G. Kothleitner, and K. Bredies, "Total generalized variation regularization for multi-modal electron tomography," *Nanoscale*, vol. 11, no. 12, pp. 5617–5632, 2019.
- [30] Y. Chen, Z. Peng, M. Li, F. Yu, and F. Lin, "Seismic signal denoising using total generalized variation with overlapping group sparsity in the accelerated ADMM framework," *J. Geophys. Eng.*, vol. 16, no. 1, pp. 30–51, Feb. 2019.
- [31] H. S. Aghamiry, A. Gholami, and S. Operto, "Full waveform inversion by proximal Newton method using adaptive regularization," *Geophys. J. Int.*, vol. 224, no. 1, pp. 169–180, Jul. 2020.
- [32] H. Fu, H. Qi, S. Dong, and R. Hua, "Sparse graph-regularized dictionary learning for full waveform inversion," *Comput. Geosci.*, vol. 180, Nov. 2023, Art. no. 105449.
- [33] W. Zhu, K. Xu, E. Darve, B. Biondi, and G. C. Beroza, "Integrating deep neural networks with full-waveform inversion: Reparameterization, regularization, and uncertainty quantification," *Geophysics*, vol. 87, no. 1, pp. 93–109, Dec. 2021.
- [34] P. Sun, F. Yang, H. Liang, and J. Ma, "Full-waveform inversion using a learned regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5920715.
- [35] W. Guo, J. Qin, and W. Yin, "A new detail-preserving regularization scheme," *SIAM J. Imag. Sci.*, vol. 7, no. 2, pp. 1309–1334, Jan. 2014.
- [36] G. Yang, X. Luan, F. Meng, and J. Huang, "Seismic data reconstruction based on Shearlet transform and total generalized variation regularization," *Chin. J. Geophys.*, vol. 63, no. 9, pp. 3465–3477, 2020.
- [37] D. Kong and Z. Peng, "Seismic random noise attenuation using Shearlet and total generalized variation," *J. Geophys. Eng.*, vol. 12, no. 6, pp. 1024–1035, Dec. 2015.
- [38] X. Wang, J. Zhang, and H. Cheng, "A joint framework for seismic signal denoising using total generalized variation and Shearlet transform," *IEEE Access*, vol. 9, pp. 6661–6673, 2021.
- [39] X. Liu, "Total generalized variation and Shearlet transform based poissonian image deconvolution," *Multimedia Tools Appl.*, vol. 78, no. 13, pp. 18855–18868, Jul. 2019.
- [40] S. Häuser and G. Steidl, "Fast finite Shearlet transform," 2012, *arXiv:1202.1773*.
- [41] R.-E. Plessix, "A review of the adjoint-state method for computing the gradient of a functional with geophysical applications," *Geophys. J. Int.*, vol. 167, no. 2, pp. 495–503, Nov. 2006.
- [42] K. Aghazade, H. S. Aghamiry, A. Gholami, and S. Operto, "Randomized source sketching for full waveform inversion," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5909412.
- [43] J. Fang, H. Zhou, Y. E. Li, and Y. Shi, "Time-domain elastic full waveform inversion with frequency normalization," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5903917.
- [44] D. Feng, X. Wang, and X. Wang, "New dynamic stochastic source encoding combined with a minmax-concave total variation regularization strategy for full waveform inversion," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 7753–7771, Nov. 2020.
- [45] Y. Lin, H. Liu, L. Xing, and H. Lin, "Time-domain wavefield reconstruction inversion solutions in the weighted full waveform inversion form," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5923514.
- [46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [47] F. Yang and J. Ma, "Wasserstein distance-based full-waveform inversion with a regularizer powered by learned gradient," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5904813.
- [48] A. Richardson, "Seismic full-waveform inversion using deep learning tools and techniques," 2018, *arXiv:1801.07232*.
- [49] D. Manu, P. M. Tshakwanda, Y. Lin, W. Jiang, and L. Yang, "Seismic waveform inversion capability on resource-constrained edge devices," *J. Imag.*, vol. 8, no. 12, p. 312, Nov. 2022.
- [50] C. Deng et al., "OpenFWI: Large-scale multi-structural benchmark datasets for seismic full waveform inversion," 2021, *arXiv:2111.02926*.
- [51] Y. Feng, Y. Chen, P. Jin, S. Feng, Z. Liu, and Y. Lin, "Auto-linear phenomenon in subsurface imaging," 2023, *arXiv:2305.13314*.
- [52] H. S. Aghamiry, A. Gholami, L. Combe, and S. Operto, "Accurate 3D frequency-domain seismic wave modeling with the wavelength-adaptive 27-point finite-difference stencil: A tool for full-waveform inversion," *Geophysics*, vol. 87, no. 3, pp. 305–324, May 2022.
- [53] G. S. Martin, R. Wiley, and K. J. Marfurt, "marmousi2: An elastic upgrade for marmousi," *Lead. Edge*, vol. 25, no. 2, pp. 156–166, Feb. 2006.
- [54] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," 2019, *arXiv:1912.01703*.



Han Wang received the B.S. and M.S. degrees from the College of Geo-exploration Science and Technology, Jilin University, Changchun, China, in 2019 and 2022, respectively. He is currently pursuing the Ph.D. degree in mathematics with Harbin Institute of Technology, Harbin, China.

His research interests include optimization algorithm, full-waveform inversion, magnetotelluric inversion, and deep learning.



Siwei Yu (Member, IEEE) received the B.S. and M.S. degrees in aerospace engineering from Tsinghua University, Beijing, China, in 2010 and 2013, respectively, and the Ph.D. degree in mathematics from Harbin Institute of Technology, Harbin, China, in 2017.

He was a Visiting Scholar with the Mathematics Department, University of California at Los Angeles, Los Angeles, CA, USA, from 2014 to 2016. He is currently a Professor with the School of Mathematics, Harbin Institute of Technology. His research interests include optimization algorithm, compressive sensing, dictionary learning, deep learning, and seismic data processing.