Efficient and Accurate Full-Waveform Inversion with Total Variation Constraint

Yudai INADA
Institute of Science Tokyo
inada.y.de71@m.isct.ac.jp

Shingo TAKEMOTO
Institute of Science Tokyo
takemoto.s.e908@m.isct.ac.jp

Shunsuke ONO
Institute of Science Tokyo
ono.s.5af2@m.isct.ac.jp

Abstract—**This paper proposes a computationally efficient algorithm to address the Full-Waveform Inversion (FWI) problem with a Total Variation (TV) constraint, designed to accurately reconstruct subsurface properties from seismic data. FWI, as an ill-posed inverse problem, requires effective regularizations or constraints to ensure accurate and stable solutions. Among these, the TV constraint is widely known as a powerful prior for modeling the piecewise smooth structure of subsurface properties. However, solving the optimization problem is challenging because of the nonlinear observation process combined with the non-smoothness of the TV constraint. Conventional methods rely on inner loops and/or approximations, which lead to high computational cost and/or inappropriate solutions. To address these limitations, we develop a novel algorithm based on a primal-dual splitting method, achieving computational efficiency by eliminating inner loops and ensuring high accuracy by avoiding approximations. We also demonstrate the effectiveness of the proposed method through experiments using the SEG/EAGE Salt and Overthrust Models.**

Index Terms—**full-waveform inversion, total variation, primal-dual splitting method.**

I. INTRODUCTION

Full-Waveform Inversion (FWI) [1, 2] is a technique used to reconstruct subsurface properties from seismic data measured at multiple observation points. The subsurface properties obtained through FWI are essential in geological research and resource exploration, such as locating gas and oil reservoirs, characterizing mineral deposits, and assessing groundwater flow patterns [2–4]. FWI extends beyond geological applications to non-destructive testing in medical imaging, such as brain tissue analysis, and industrial inspections, including material defect detection [5, 6].

FWI faces significant challenges in directly recovering subsurface properties from seismic data because of the nonlinear wave propagation and the inherent complexity of the observation process [2]. To address this issue, FWI is typically formulated as an optimization problem that minimizes the squared error between observed and simulated seismic data [1, 7–12], which serves as the standard approach. Nevertheless, the inherent ill-posedness of FWI necessitates the incorporation of regularization techniques. Among these, Tikhonov regularization [13] and Total Variation (TV)-based methods [14, 15] are widely used to promote piecewise smoothness in the reconstructed subsurface properties, improving stability and accuracy [16–20]. While these regularization techniques are effective, they often involve the careful tuning of a balance parameter that determines the trade-off between the FWI objective function and the regularization term. To overcome this limitation, an alternative approach has been proposed: incorporating the TV prior as a constraint rather than as a regularization term [21–24]. This formulation decouples the parameter for the TV constraint from the FWI objective, allowing it to be determined independently based on prior knowledge of subsurface properties [25–30]. By doing so, this approach not only simplifies parameter selection but also enhances the interpretability of both the mathematical formulation and the reconstructed subsurface models.

Despite its advantages, solving the TV-constrained FWI problem presents considerable difficulties due to the interplay between the nonlinear observation process and the non-smoothness of the TV term. Conventional methods [21–24] attempt to address these issues by incorporating inner loops to enforce the constraint at each optimization step or by employing linear or quadratic approximations. However, these approaches come with notable drawbacks: inner loops substantially increase computational cost, and approximations compromise reconstruction accuracy. This raises a crucial question: *Can we develop an algorithm that effectively solves the TV-constrained FWI problem without relying on inner loops or approximations?*

In this paper, we introduce a novel algorithm for solving the TV-constrained FWI problem using a primal-dual splitting method [31]. The proposed algorithm effectively addresses the intertwined issues of the nonlinear observation process and the non-smoothness of the TV constraint, achieving accurate reconstructions without relying on approximations. Additionally, by eliminating the need for inner loops, our approach is markedly more computationally efficient than existing methods. We validate the performance of our algorithm through numerical experiments on the SEG/EAGE Salt and Overthrust Models, demonstrating its capability to efficiently enforce the constraints while delivering high-quality reconstructions.

II. PRELIMINARIES

A. Mathematical Tools

Throughout this paper, we denote vectors and matrices by bold lowercase letters (e.g., x) and bold uppercase letters (e.g., X), respectively.

For $\boldsymbol{x} \in \mathbb{R}^{\mathrm{N}}$, the mixed $\ell_{1,2}$ norm is defined as follows:

$$\|\boldsymbol{x}\|_{1,2} := \sum_{\mathfrak{g} \in \mathfrak{G}} \|\boldsymbol{x}_{\mathfrak{g}}\|_2, \qquad (1)$$

where $\mathfrak{G}$ is a set of disjoint index sets, and $\boldsymbol{x}_{\mathfrak{g}}$ is the subvector of $\boldsymbol{x}$ indexed by $\mathfrak{g}$.

For $\boldsymbol{x} \in \mathbb{R}^{\mathrm{N}}$, the total variation (TV) [14] is defined as follows:

$$\mathrm{TV}(\boldsymbol{x}) := \|\mathbf{D}\boldsymbol{x}\|_{1,2} = \sum_{i=1}^{\mathrm{N}} \sqrt{d_{h,i}^2 + d_{v,i}^2}, \qquad (2)$$

where $d_{h,i}$ and $d_{v,i}$ are the horizontal and vertical differences of the i-th element of $\boldsymbol{x}$, respectively, when the vector $\boldsymbol{x}$ is considered as a matrix.

B. Proximal Tools

We denote by $\Gamma_0(\mathbb{R}^N)$ the set of proper lower-semicontinuous convex functions $\mathbb{R}^N \to (-\infty, \infty]$.

For $\gamma > 0$ and $f \in \Gamma_0(\mathbb{R}^N)$, the proximity operator is defined as follows:

$$\mathrm{prox}_{\gamma f}(\boldsymbol{x}) := \underset{\boldsymbol{y} \in \mathbb{R}^N}{\mathrm{argmin}} \; f(\boldsymbol{y}) + \frac{1}{2\gamma} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2. \qquad (3)$$

For $f \in \Gamma_0(\mathbb{R}^N)$, the convex conjugate function f^* is defined as follows:

$$f^*(\boldsymbol{x}) := \sup_{\boldsymbol{y} \in \mathbb{R}^N} \; \boldsymbol{y}^T \boldsymbol{x} - f(\boldsymbol{y}). \qquad (4)$$

The proximity operator for the convex conjugate function is expressed as follows [32, Theorem 3.1 (ii)]:

$$\mathrm{prox}_{\gamma f^*}(\boldsymbol{x}) = \boldsymbol{x} - \gamma \mathrm{prox}_{\frac{1}{\gamma} f}\left(\frac{1}{\gamma}\boldsymbol{x}\right). \qquad (5)$$

For a nonempty closed convex set $C \subset \mathbb{R}^N$, the indicator function $\iota_C : \mathbb{R}^N \to (-\infty, \infty]$ is defined as follows:

$$\iota_C(\boldsymbol{x}) := \begin{cases} 0 & \text{if } \boldsymbol{x} \in C, \\ \infty & \text{otherwise.} \end{cases} \qquad (6)$$

The proximity operator of ι_C is the projection onto C, given by

$$\mathrm{prox}_{\gamma \iota_C}(\boldsymbol{x}) = P_C(\boldsymbol{x}) := \underset{\boldsymbol{y} \in C}{\mathrm{argmin}} \; \|\boldsymbol{y} - \boldsymbol{x}\|_2. \qquad (7)$$

C. Primal-Dual Splitting Algorithm

The Primal-Dual Splitting algorithm (PDS) [31] is applied to the following problem:

$$\min_{\boldsymbol{x} \in \mathbb{R}^N} \{f(\boldsymbol{x}) + g(\boldsymbol{x}) + h(\boldsymbol{L}\boldsymbol{x})\}, \qquad (8)$$

where $\boldsymbol{L} \in \mathbb{R}^{\mathrm{M} \times \mathrm{N}}$, f is a differentiable convex function and g, h are convex functions whose proximity operator can be computed efficiently.

PDS solves Prob. (8) by iteratively updating the following:

$$\left\lfloor \begin{array}{l} \boldsymbol{x}^{(k+1)} = \mathrm{prox}_{\gamma_1 g}\left(\boldsymbol{x}^{(k)} - \gamma_1(\nabla f(\boldsymbol{x}^{(k)}) + \boldsymbol{L}^T \boldsymbol{y}^{(k)})\right), \\ \boldsymbol{y}^{(k+1)} = \mathrm{prox}_{\gamma_2 h^*}\left(\boldsymbol{y}^{(k)} + \gamma_2 \boldsymbol{L}(2\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)})\right), \end{array} \right.$$
$$(9)$$

where $\gamma_1, \gamma_2 > 0$ are step sizes.

D. Full-Waveform Inversion

Typically, FWI is treated as an optimization problem as follows [1]:

$$\underset{\boldsymbol{m} \in \mathbb{R}^N}{\mathrm{argmin}} \; E(\boldsymbol{m}) = \frac{1}{2} \|\boldsymbol{u}_{\mathrm{obs}} - \boldsymbol{u}_{\mathrm{cal}}(\boldsymbol{m})\|_2^2, \qquad (10)$$

where $\boldsymbol{m} \in \mathbb{R}^N$ is the velocity model representing subsurface properties, $\boldsymbol{u}_{\mathrm{obs}} \in \mathbb{R}^M$ is the observed seismic data, $\boldsymbol{u}_{\mathrm{cal}} : \mathbb{R}^N \to \mathbb{R}^M$ is the observation process, and $\boldsymbol{u}_{\mathrm{cal}}(\boldsymbol{m})$ is the modeled seismic data with the velocity model. N is the number of grid points, and M is the total data size of the observed seismic data, defined as the total product of the number of waveform sources, time samples, and receivers.

The observation process $\boldsymbol{u}_{\mathrm{cal}}$ is nonlinear and complex, making it difficult to analytically derive the optimal solution. However, the gradient ∇E can be computed numerically by simulating the wave equation using the adjoint-state method [33].

III. PROPOSED METHOD

We introduce the TV and box constraint into the FWI problem to achieve more accurate reconstruction. As shown in Fig. 1, the velocity model is piecewise smooth, thus introducing the TV constraint to achieve a more accurate reconstruction. The box constraint ensures that the velocity model remains within valid ranges.

The optimization problem of the TV and box constrained FWI is formulated as follows:

$$\underset{\boldsymbol{m} \in \mathbb{R}^N}{\mathrm{argmin}} \; E(\boldsymbol{m}) \quad \text{s.t.} \quad \begin{cases} \|\mathbf{D}\boldsymbol{m}\|_{1,2} \leq \alpha, \\ \boldsymbol{m} \in [l, u]^N, \end{cases} \qquad (11)$$

where $\alpha \geq 0$ is the upper bound of the $\ell_{1,2}$ norm, and $u > l > 0$ are the upper and lower bounds of the velocity model values, respectively. By incorporating TV as a constraint, the parameter α can be determined independently of other terms or constraints, which has been highlighted as an advantage in prior works [25–30]. This separation makes it possible to directly control smoothness according to α, providing a clearer interpretation of the reconstructed subsurface properties.

The constraints can be incorporated into the objective function as indicator functions:

$$\underset{\boldsymbol{m} \in \mathbb{R}^N}{\mathrm{argmin}} \; E(\boldsymbol{m}) + \iota_{B_{\mathrm{box}}}(\boldsymbol{m}) + \iota_{B_{\ell_{1,2}}}(\mathbf{D}\boldsymbol{m}), \qquad (12)$$

where

$$\begin{aligned} B_{\mathrm{box}} &:= [l, u]^N, \\ B_{\ell_{1,2}} &:= \{\|\cdot\|_{1,2} \leq \alpha\}. \end{aligned} \qquad (13)$$

The proximity operator of $\iota_{B_{\mathrm{box}}}$ and $\iota_{B_{\ell_{1,2}}}$ can be computed efficiently. Therefore, these functions of E, $\iota_{B_{\mathrm{box}}}$ and $\iota_{B_{\ell_{1,2}}}$ correspond to f, g and h in (8), respectively, $\mathbf{D}$ is corresponds to $\boldsymbol{L}$, so the problem (12) can be solved using PDS. We show the detailed algorithm in Algorithm 1. By strictly following the

Algorithm 1 PDS based solver for (12)

> **Input:** $\boldsymbol{m}^{(0)}, \boldsymbol{y}^{(0)}, \gamma_0 > 0, \gamma_1 > 0$
> 1: **while** A stopping criterion is not satisfied **do**
> 2: $\widetilde{\boldsymbol{m}} \leftarrow \boldsymbol{m}^{(k)} - \gamma_1 \left(\nabla E(\boldsymbol{m}^{(k)}) + \boldsymbol{D}^T \boldsymbol{y}^{(k)} \right)$
> 3: $\boldsymbol{m}^{(k+1)} \leftarrow P_{B_{\text{box}}}(\widetilde{\boldsymbol{m}})$
> 4: $\widetilde{\boldsymbol{y}} \leftarrow \boldsymbol{y}^{(k)} + \gamma_2 \boldsymbol{D} \left(2\boldsymbol{m}^{(k+1)} - \boldsymbol{m}^{(k)} \right)$
> 5: $\boldsymbol{y}^{(k+1)} \leftarrow \widetilde{\boldsymbol{y}} - \gamma_2 P_{B_{\ell_{1,2}}} \left(\frac{1}{\gamma_2} \widetilde{\boldsymbol{y}} \right)$
> 6: **end while**
> **Output:** $\boldsymbol{m}^{(k)}$

PDS framework, our algorithm accurately solves the optimization problem without relying on approximations or relaxations.

The proximity operators of $\iota_{B_{\text{box}}}$, $\iota_{B_{\ell_{1,2}}}$, that is, the projection onto B_{box} and $B_{\ell_{1,2}}$ are calculated by

$$P_{B_{\text{box}}}(\boldsymbol{x}) = \min(\max(\boldsymbol{x}, l), u), \tag{14}$$

$$\left(P_{B_{\ell_{1,2}}}(\boldsymbol{x})\right)_{\mathfrak{g}_i} = \begin{cases} 0 & \text{if } \|\boldsymbol{x}_{\mathfrak{g}_i}\|_2 = 0, \\ \boldsymbol{\beta}_i \frac{\boldsymbol{x}_{\mathfrak{g}_i}}{\|\boldsymbol{x}_{\mathfrak{g}_i}\|_2} & \text{otherwise,} \end{cases} \tag{15}$$

where

$$\boldsymbol{\beta} = P_{\{\|\cdot\|_1 \leq \alpha\}} \left([\|\boldsymbol{x}_{\mathfrak{g}_1}\|_2, \dots, \|\boldsymbol{x}_{\mathfrak{g}_N}\|_2]^T \right), \tag{16}$$

and $\mathfrak{g}_i$ is an index set corresponding to the horizontal and vertical differences of the i-th element of $\boldsymbol{m}$. The proximity operator for the ℓ_1 norm upper bound constraint is provided in [34].

Remark: Computational Cost of Our Algorithm

Our algorithm efficiently solves the optimization problem with the Total Variation (TV) constraint without relying on inner loops. The additional computational cost to enforce the TV and box constraints at each iteration of our algorithm is equal to or less than the cost at each inner-loop iteration of existing methods. Thus, the iteration-dependent factor in the computational cost is eliminated, and the computational complexity to enforce the TV constraint is reduced to $O(N \log N)$, where N is the number of grid points in the velocity model. Consequently, the computational complexity of our algorithm at each iteration is dominated by the gradient computation ∇E, which involves simulating the wave equation along the time axis. This step has a complexity of $O(STN)$, where S is the number of waveform sources, T is the number of time samples.

IV. EXPERIMENTS

A. Experimental Setup

To demonstrate the effectiveness of the TV and box constrained FWI, we conducted FWI experiments where we compared with the standard FWI method[1][1], using the

---

The standard FWI method uses the following procedures:

[1]
$$\boldsymbol{m}^{(k+1)} = \boldsymbol{m}^{(k)} - \gamma \left(\nabla E(\boldsymbol{m}^{(k)}) \right) \tag{17}$$

where $\gamma > 0$ is the step size.
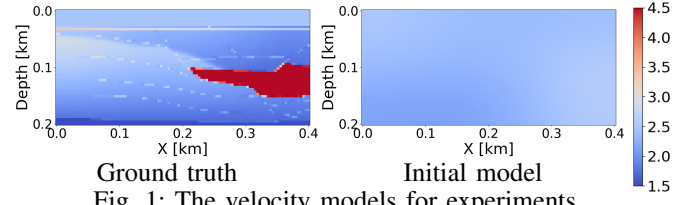
Ground truth   Initial model
Fig. 1: The velocity models for experiments

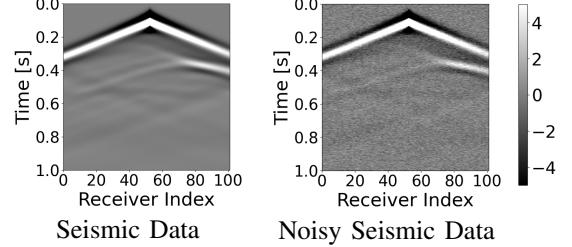Seismic Data   Noisy Seismic Data
Fig. 2: The sinthesized seismic data
corresponding to a single source waveform

SEG/EAGE Salt and Overthrust Models.

The velocity model consists of 50×100 grid points. Fig. 1 shows the ground truth velocity model and the initial velocity model, which is created by smoothing the ground truth with a Gaussian filter with a standard deviation of 80. The source waveform is a Ricker wavelet with a peak wavelet frequency of 10 Hz. The number of waveform sources and receivers is 20 and 101, respectively, and they are placed on the surface at equal intervals. The gradient ∇E is computed numerically using the Devito framework [35]. The number of iterations is set to 5000. Experiments are conducted with and without noise in the observed data, as shown in Fig. 2. The noise is Gaussian noise with mean 0 and variance 1. In our algorithm, the step sizes γ_1 and γ_2 are set to 1.0×10^{-4} and 1.0×10^2, respectively. The lower and upper bounds of the velocity model l, u are set to 1.5 [km/s] and 4.5 [km/s], respectively. The experiments are conducted using α values ranging from 100 to 700 in steps of 50, representing the upper bound of the $\ell_{1,2}$ norm. In the standard FWI method, the step size γ is set to 1.0×10^{-4}.

B. Results and Discussion

First, we present the experimental results without noise in the observed seismic data. Fig. 3 shows the reconstructed velocity models using the standard FWI method and the proposed methods with $\alpha = 150$, 350, and 550. The best parameter is $\alpha = 350$, where $\alpha = 150$ represents a stronger TV constraint and $\alpha = 550$ represents a weaker one. The standard FWI method generates wave-like artifacts throughout. In contrast, our proposed method with the best parameter, which is $\alpha = 350$, achieves the accurate velocity model reconstruction without these artifacts. As shown in Fig. 6 and Fig. 7, the proposed method archive accurate reconstruction performance, as evidenced by quantitative evaluations using the Root Mean Squared Error (RMSE) and the Structural Similarity Index Measure (SSIM). The case of $\alpha = 150$, 550 will be discussed later.

Second, we present the experimental results with the noisy observed seismic data. Fig. 4 shows the reconstructed velocity

(a) Standard FWI   (b) Proposed, $\alpha = 150$ (c) Proposed, $\alpha = 350$ (d) Proposed, $\alpha = 550$
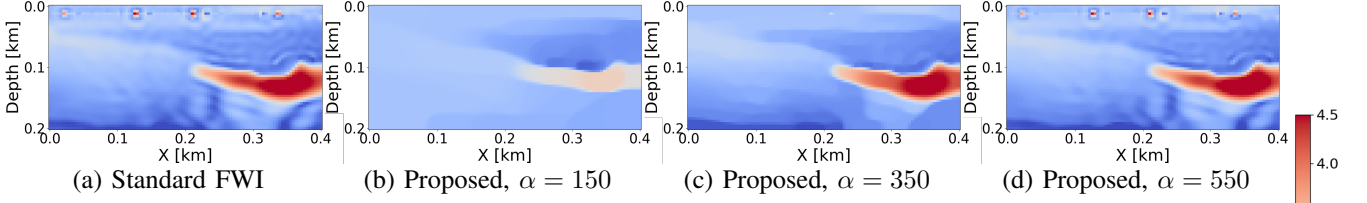
Fig. 3: Velocity models [km/s] and their corresponding reconstructions. (c) is the best reconstruction result. (b) is over-smoothed with a stronger TV constraint, and (d) similar results to (a) with a weaker one.
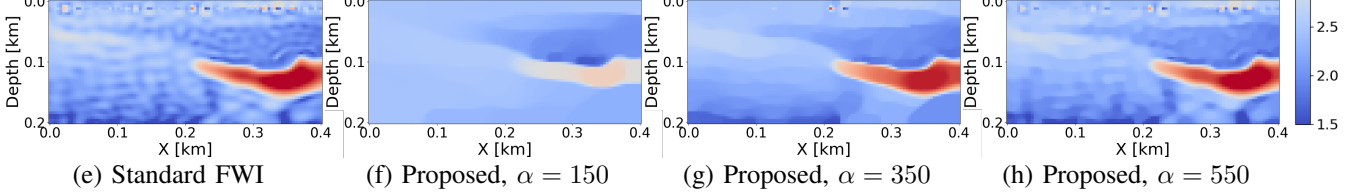

(e) Standard FWI   (f) Proposed, $\alpha = 150$ (g) Proposed, $\alpha = 350$ (h) Proposed, $\alpha = 550$

Fig. 4: Velocity models [km/s] and their corresponding reconstructions (with the noisy data). Similar to Fig. 3, (g) is the best result, and (f)/(h) are the results with stronger/weaker TV constraint.
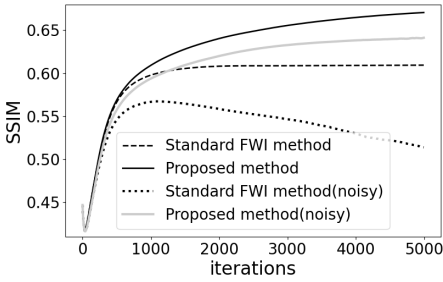

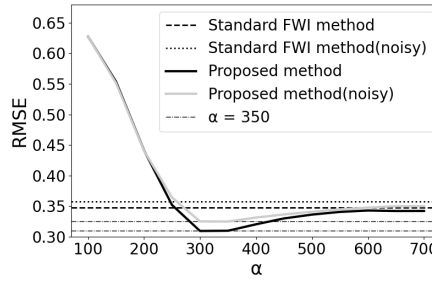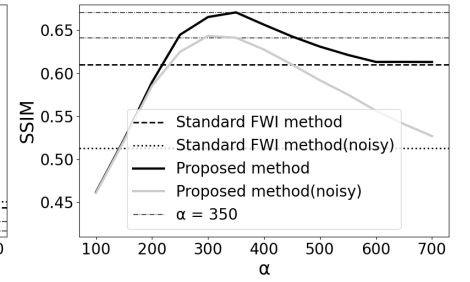Fig. 5: SSIM against iters.   Fig. 6: RMSE against α.   Fig. 7: SSIM against α.

models using the standard FWI method and the proposed methods with α = 150, 350 and 550. In the FWI problem with the noisy observed seismic data, the standard FWI method generates stronger wave-like artifacts throughout. In contrast, our proposed method with α = 350 as the appropriate parameter still achieves accurate velocity model reconstruction without these artifacts and achieves almost the same level of performance as in the noiseless case. This demonstrates the robustness of the reconstruction accuracy to noise in the observed seismic data.

Fig. 5 shows the SSIM against the number of iterations for our proposed method and the standard FWI method. The standard FWI method with the noisy observed seismic data degrades the SSIM after a certain number of iterations. On the other hand, the proposed method maintains a consistently high value without degrading the SSIM even after a large number of iterations. This shows that the proposed method effectively reduces the risk of overfitting to noisy data, providing stable and accurate performance even as the number of iterations increases.

For a more detailed analysis of the TV constraint parameter α, we plot the RMSE and SSIM of our proposed method against the parameter α and the standard FWI method in Fig. 6 and Fig. 7. When around $\alpha = 350$, the proposed method achieves the best RMSE and SSIM. When α is smaller, the TV constraint becomes stronger, RMSE and SSIM worsens, and over-smoothing occurs, as in Fig. 3 and Fig. 4 for α =

150. On the other hand, when α is larger, the TV constraint becomes weaker and the result is similar to the standard FWI method, as in Fig. 3 and Fig. 4 for α = 550. However, thanks to the box constraint, the proposed method still outperforms the standard FWI method. This demonstrates that the parameter α has a clear and predictable effect on the reconstructed velocity model, which can be easily adjusted to achieve accurate results.

V. CONCLUSION

In this paper, we developed an efficient algorithm to solve the TV and box constrained FWI problem based on PDS. Our algorithm does not require approximations when incorporating the constraints, leading to more accurate reconstructions. Furthermore, the algorithm significantly enhances computational efficiency without inner loops. Experimental results demonstrate that our method successfully eliminates wave-like artifacts and noise present in the standard FWI method, resulting in a more accurate velocity model and a superior RMSE and SSIM value regardless of the presence of noise in the observed seismic data.

REFERENCES

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

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

[3] A. Klotzsche, J. van der Kruk, G. A. Meles, J. Doetsch, H. Maurer, and N. Linde, "Full-waveform inversion of cross-hole ground-penetrating radar data to characterize a gravel aquifer close to the thur river, switzerland," *Near surface geophysics*, vol. 8, no. 6, pp. 635–649, 2010.

[4] A. Klotzsche, H. Vereecken, and J. van der Kruk, "Gpr full-waveform inversion of a variably saturated soil-aquifer system," *J. Appl. Geophysics*, vol. 170, p. 103823, 2019.

[5] L. Guasch, O. Calderón Agudo, M.-X. Tang, P. Nachev, and M. Warner, "Full-waveform inversion imaging of the human brain," *NPJ Digit. Med.*, vol. 3, no. 1, p. 28, 2020.

[6] J. Rao, J. Yang, M. Ratassepp, and Z. Fan, "Multi-parameter reconstruction of velocity and density using ultrasonic tomography based on full waveform inversion," *Ultrasonics*, vol. 101, p. 106004, 2020.

[7] C. Shin and D.-J. Min, "Waveform inversion using a logarithmic wavefield," *Geophysics*, vol. 71, no. 3, pp. R31–R42, 2006.

[8] E. Bozdağ, J. Trampert, and J. Tromp, "Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements," *Geophys. J. Int.*, vol. 185, no. 2, pp. 845–870, 2011.

[9] J. Luo and R.-S. Wu, "Seismic envelope inversion: reduction of local minima and noise resistance," *Geophys. Prospecting*, vol. 63, no. 3, pp. 597–614, 2015.

[10] B. Engquist and B. D. Froese, "Application of the wasserstein metric to seismic signals," *arXiv preprint arXiv:1311.4581*, 2013.

[11] 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, 2016.

[12] M. Warner and L. Guasch, "Adaptive waveform inversion: Theory," *Geophysics*, vol. 81, no. 6, pp. R429–R445, 2016.

[13] A. Tikhonov, A. Goncharsky, V. Stepanov, and A. Yagola, *Numerical methods for the approximate solution of ill-posed problems on compact sets*. Springer, 1995.

[14] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[15] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp. 492–526, 2010.

[16] A. Asnaashari, R. Brossier, S. Garambois, F. Audebert, P. Thore, and J. Virieux, "Regularized seismic full waveform inversion with prior model information," *Geophysics*, vol. 78, no. 2, pp. R25–R36, 2013.

[17] A. Y. Anagaw and M. D. Sacchi, "Full waveform inversion with total variation regularization," in *Recovery-CSPG CSEG CWLS Convention*, 2011, pp. 1–4.

[18] 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. R175–R183, 2019.

[19] 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. Geophysics and Eng.*, vol. 18, no. 2, pp. 241–252, 2021.

[20] 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, 2019.

[21] E. Esser, L. Guasch, T. van Leeuwen, A. Y. Aravkin, and F. J. Herrmann, "Total variation regularization strategies in full-waveform inversion," *SIAM J. Imag. Sci.*, vol. 11, no. 1, pp. 376–406, 2018.

[22] 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, 2016.

[23] P. Yong, W. Liao, J. Huang, and Z. Li, "Total variation regularization for seismic waveform inversion using an adaptive primal dual hybrid gradient method," *Inverse Problems*, vol. 34, no. 4, p. 045006, 2018.

[24] B. Peters, B. R. Smithyman, and F. J. Herrmann, "Projection methods and applications for seismic nonlinear inverse problems with multiple constraints," *Geophysics*, vol. 84, no. 2, pp. R251–R269, 2019.

[25] M. V. Afonso, J. M. Bioucas-Dias, and M. Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, 2011.

[26] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, "Epigraphical projection and proximal tools for solving constrained convex optimization problems," *Signal Image Video Process.*, vol. 9, no. 8, pp. 1737–1749, 2015.

[27] S. Ono and I. Yamada, "Signal recovery with certain involved convex data-fidelity constraints," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 6149–6163, 2015.

[28] S. Ono, "Primal-dual plug-and-play image restoration," *IEEE Signal Process. Lett.*, vol. 24, no. 8, pp. 1108–1112, 2017.

[29] S. Ono, "Efficient constrained signal reconstruction by randomized epigraphical projection," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., (ICASSP)*, 2019, pp. 4993–4997.

[30] B. Peters and F. J. Herrmann, "Constraints versus penalties for edge-preserving full-waveform inversion," *Lead. Edge*, vol. 36, no. 1, pp. 94–100, 2017.

[31] L. Condat, "A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.

[32] P. L. Combettes and N. N. Reyes, "Moreau's decomposition in banach spaces," *Math. Program.*, vol. 139, no. 1, pp. 103–114, 2013.

[33] 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, 2006.

[34] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, "Efficient projections onto the l 1-ball for learning in high dimensions," in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 272–279.

[35] M. Louboutin, M. Lange, F. Luporini, N. Kukreja, P. A. Witte, F. J. Herrmann, P. Velesko, and G. J. Gorman, "Devito (v3. 1.0): an embedded domain-specific language for finite differences and geophysical exploration," *Geoscientific Model Develop.*, vol. 12, no. 3, pp. 1165–1187, 2019.