Efficient Full Waveform Inversion Subject To A Total Variation Constraint

Yudai INADA, Shingo TAKEMOTO, and Shunsuke ONO

Institute of Science Tokyo

Abstract   Full waveform inversion (FWI) aims to reconstruct subsurface properties from observed seismic data. Since FWI is an ill-posed inverse problem, appropriate regularizations or constraints are useful approaches to achieve accurate reconstruction. The total variation (TV) -type regularization or constraint is widely known as a powerful prior that models the piecewise smoothness of subsurface properties. However, the optimization problem of the TV-type regularized or constrained FWI is difficult to solve due to the nonlinearity of the observation process and the non-smoothness of the TV-type regularization or constraint. Conventional approaches to solve the problem rely on an inner loop and/or approximations, resulting in high computational cost and/or inappropriate solutions. In this paper, we develop an efficient algorithm with neither an inner loop nor approximations to solve the problem based on a primal-dual splitting method. We also demonstrate the effectiveness of the proposed method through experiments using the SEG/EAGE Salt and Overthrust Models.

1   Introduction

Full waveform inversion (FWI) [1], [2] is a task that aims to reconstruct subsurface properties from observed seismic data. These subsurface properties are used for geological research and resource exploration, including deposits of gas, oil, mineral, and groundwater [2]–[4]. In addition to resource exploration, FWI has also been successfully applied to non-destructive testing [5], [6].

FWI is an ill-posed problem, and observed seismic data are generated by subsurface properties through a nonlinear and complex observation process [2]. To address this, FWI is formulated as an optimization problem [1], [7]–[12] such as minimizing the squared error between observed and modeled data. To mitigate the risk of local minima and improve accuracy, regularization techniques are useful approaches. Many studies have introduced regularizations [13]–[17], such as Tikhonov [18] and Total Variation (TV)-type [19], [20] regularization, based on the prior knowledge that subsurface properties exhibit piecewise smoothness.

However, regularizations often require careful tuning of balance parameters to trade off between data fitting and smoothness. In contrast to regularizations, constraints have the advantage that their parameters can be determined based only on prior knowledge of the subsurface properties [21]. In addition, formulating the problem with constraints makes the problem and solution more interpretable, which is beneficial for practical applications. Therefore, it has also been proposed to incorporate TV as a constraint into the problem [22]–[24].

However, the TV-constrained problem is difficult to solve not only because of the nonlinearity of the observation process, but also because of the non-smoothness of the TV constraint. As a result, conventional methods [22]–[24] to solve it adjust the parameter updates to satisfy the constraint at each step of an iterative optimization algorithm. This requires an additional optimization, resulting in an inner loop and increased computational cost. In addition, approximations are introduced to incorporate the constraint, such as treating nonlinear transformations as linear and imposing constraints outside the optimization method.

If the TV-constrained problem could be solved with neither an inner loop nor these approximations, more efficient and accurate reconstructions of subsurface properties would be possible. In this paper, we propose a novel algorithm to solve the problem based on the primal-dual splitting (PDS) method, which eliminates the need for an inner loop and approximations that are common in conventional approaches. We demonstrate that our algorithm efficiently handles constraints without significantly increasing the computational cost, while achieving accurate reconstruction.

2   Preliminaries

2.1   Mathematical Tools

Throughout this paper, we denote vector and matrix by bold lowercase letters (e.g., $\boldsymbol{x}$) and bold uppercase

letters (e.g., $\boldsymbol{X}$), respectively. The operator l_X norm of a vector and matrix is denoted by $\|\cdot\|_X$.

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

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

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 $\boldsymbol{x} \in \mathbb{R}^N$, the total variation (TV) [19] is defined as follows:

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

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.

For a proper lower-semicontinuous convex function $f \in \mathbb{R}^N \to \mathbb{R}$ and $\boldsymbol{x} \in \mathbb{R}^N$, the convex conjugate function is defined as follows:

$$f^*(\boldsymbol{x}) := \sup_{\boldsymbol{y} \in \mathbb{R}^N} \left\{ \boldsymbol{y}^T \boldsymbol{x} - f(\boldsymbol{y}) \right\}. \tag{3}$$

For a set $C \subset \mathbb{R}^N$ and $\boldsymbol{x} \in \mathbb{R}^N$, the indicator function is defined as follows:

$$\iota_C(\boldsymbol{x}) := \begin{cases} 0 & \text{if } \boldsymbol{x} \in C, \\ \infty & \text{otherwise.} \end{cases} \tag{4}$$

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

$$\mathrm{prox}_{\gamma f}(\boldsymbol{x}) := \underset{\boldsymbol{y} \in \mathbb{R}^N}{\mathrm{argmin}} \left\{ f(\boldsymbol{y}) + \frac{1}{2\gamma} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2 \right\}. \tag{5}$$

Define the proximity operator for the indicator function as P_C as follows.

$$\mathrm{prox}_{\gamma \iota_C(\cdot)}(\boldsymbol{x}) = P_C(\boldsymbol{x}) := \underset{\boldsymbol{y} \in C}{\mathrm{argmin}} \|\boldsymbol{y} - \boldsymbol{x}\|_2. \tag{6}$$

The proximity operator for the convex conjugate function is expressed as follows [25, Theorem 3.1 (ii)]:

$$\mathrm{prox}_{\gamma f^*}(\boldsymbol{x}) = \boldsymbol{x} - \gamma \mathrm{prox}_{f/\gamma}(\boldsymbol{x}/\gamma). \tag{7}$$

2.2 Primal-Dual Splitting Algorithm

The Primal-Dual Splitting (PDS) algorithm [26]–[29] is applied to the following problem:

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

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

The PDS algorithm solves by iteratively updating the following:

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

where $\gamma_1, \gamma_2 \in \mathbb{R}$ are step sizes.

2.3 Full Waveform Inversion

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

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

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, $\boldsymbol{u}_{\mathrm{cal}}$ 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 number of observed signals. In general, the velocity model is 2D or 3D grid data, but for simplicity we consider flattened 1D vector.

The observation process is nonlinear and complex, making it difficult to obtain $\boldsymbol{u}_{\mathrm{cal}(\cdot)}$ analytically. However, the gradient ∇E can be computed numerically using the adjoint-state method [30].

Therefore, the standard FWI minimizes the objective function and reconstructs the velocity model using the following procedures:

$$\boldsymbol{m}^{(k+1)} = \boldsymbol{m}^{(k)} - \gamma (\nabla E(\boldsymbol{m}^{(k)})), \tag{11}$$

where γ is the step size.

3 Proposed Method

As shown in Fig.1, the velocity model of the Salt is piecewise smooth. Therefore, we introduce the TV constraint to achieve more accurate reconstruction. Also, by introducing the box constraint, we can ensure that the velocity model does not take invalid values, and we show the flexibility of incorporating constraints using PDS.

We minimize the objective function of the TV and box constrained FWI, which is expressed as follows:

$$\underset{\boldsymbol{m} \in \mathbb{R}^N}{\mathrm{argmin}} \quad E(\boldsymbol{m}) \quad \text{s.t.} \quad \|\mathbf{D}\boldsymbol{m}\|_{1,2} \leq \alpha \ , \ \boldsymbol{m} \in [a, b]^N \tag{12}$$

where $\alpha \in \mathbb{R}$ is the upper bound of the $l_{1,2}$ norm, and $a, b \in \mathbb{R}$ are the lower and upper bounds of the velocity model value, respectively.
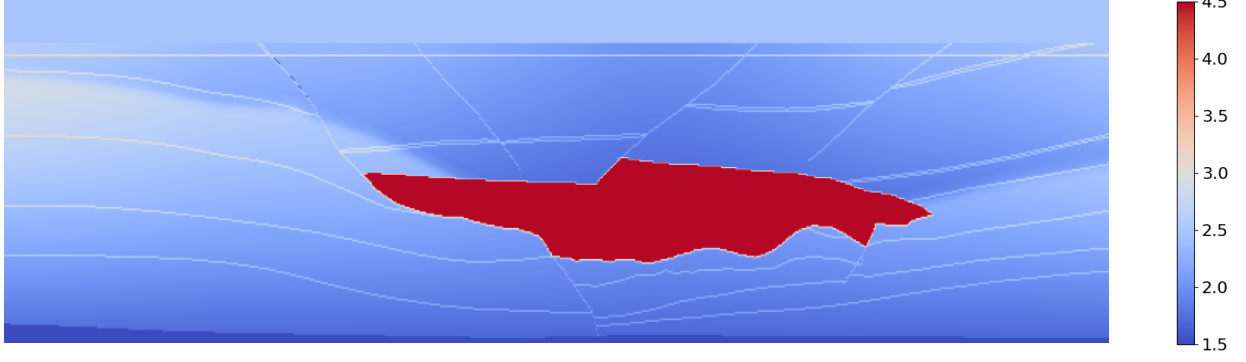
Fig. 1: the velocity model of the Salt [km/s]

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

$$\underset{\boldsymbol{m}\in\mathbb{R}^N}{\operatorname{argmin}}\ \ E(\boldsymbol{m}) + \iota_{\|\cdot\|_{1,2}\leq\alpha}(\mathbf{D}\boldsymbol{m}) + \iota_{[a,b]^N}(\boldsymbol{m}) \quad (13)$$

The proximity operator of $\iota_{\|\cdot\|_{1,2}\leq\alpha}$ and $\iota_{[a,b]^N}$ can be computed efficiently. Therefore, these functions of E, $\iota_{[a,b]^N}$ and $\iota_{\|\cdot\|_{1,2}\leq\alpha}$ correspond to f, g and h in (8), respectively, $\mathbf{D}$ is corresponds to $\boldsymbol{L}$, and the problem (13) can be solved using PDS. The iterative procedures are as follows:

$$\left| \begin{array}{l} \widetilde{\boldsymbol{m}}^{(k+1)} = \boldsymbol{m}^{(k)} - \gamma_1(\nabla E(\boldsymbol{m}^{(k)}) + \boldsymbol{D}^T\boldsymbol{y}^{(k)}) \\ \boldsymbol{m}^{(k+1)} = P_{[a,b]^N}(\widetilde{\boldsymbol{m}}^{(k+1)}) \\ \widetilde{\boldsymbol{y}}^{(k+1)} = \boldsymbol{y}^{(k)} + \gamma_2\boldsymbol{D}(2\boldsymbol{m}^{(k+1)} - \boldsymbol{m}^{(k)}) \\ \boldsymbol{y}^{(k+1)} = \widetilde{\boldsymbol{y}}^{(k+1)} - \gamma_2 P_{\{\boldsymbol{a}|\|\boldsymbol{a}\|_{1,2}\leq\alpha\}}(\frac{1}{\gamma_2}\widetilde{\boldsymbol{y}}^{(k+1)}) \end{array} \right.$$

The following are the proximity operators of indicator function of the box constraint and the $l_{1,2}$ norm upper bound constraint.

$$P_{[a,b]^N}(\boldsymbol{x}) = \min(\max(\boldsymbol{x},a),b). \quad (14)$$

$$(P_{\{\boldsymbol{a}|\|\boldsymbol{a}\|_{1,2}\leq\alpha\}}(\boldsymbol{x}))_{\mathfrak{g}_i} = \begin{cases} 0 & \text{if } \|\boldsymbol{x}_{\mathfrak{g}_i}\|_2 = 0, \\ \beta_i\frac{\boldsymbol{x}_{\mathfrak{g}_i}}{\|\boldsymbol{x}_{\mathfrak{g}_i}\|_2} & \text{otherwise,} \end{cases}, \quad (15)$$

where

$$\boldsymbol{\beta} = P_{\{\boldsymbol{a}|\|\boldsymbol{a}\|_1\leq\alpha\}}([\|\boldsymbol{x}_{\mathfrak{g}_1}\|_2,\ldots,\|\boldsymbol{x}_{\mathfrak{g}_N}\|_2]^T).$$

The proximity operator for the l_1 norm upper bound constraint is expressed as follows [31](faster algorithms are also proposed [32]):

$$P_{\{\boldsymbol{a}|\|\boldsymbol{a}\|_1\leq\alpha\}}(\boldsymbol{x}) = \text{SoftThrethold}(\boldsymbol{x},\beta), \quad (16)$$

where

$$\boldsymbol{x}_{\text{abs}} = \text{abs}(\boldsymbol{x}),$$

$$\boldsymbol{y} = \text{sort}_{\text{desc}}(\boldsymbol{x}_{\text{abs}}),$$

$$\beta' = \max\{\frac{1}{i}((\sum_{j=1}^{i}\boldsymbol{y}_j) - \alpha) \mid i = 1,\ldots,N\},$$

$$\beta = \max\{\beta',0\}.$$

4 Experiments

4.1 Experimental Setup

To demonstrate the effectiveness of the TV and box constrained FWI, we conducted experiments where we compared with the standard FWI with gradient method (11), using the SEG/EAGE Salt and Overthrust Models. The velocity model consists of 101×51 grid points. The ground truth velocity model is generated by zooming and cropping Fig.1, and the initial velocity model is generated by smoothing the ground truth velocity model with a Gaussian function with a standard deviation of 80. The number of receivers and source shots are 101 and 20, respectively, and are placed on the surface at equal intervals. The source waveform is a Ricker wavelet with a peak wavelet frequency of 10 Hz. The gradient of E is computed numerically using the Devito framework[33]. In the standard FWI, the step size γ is set to 1.0×10^{-4}. In the TV and box constrained FWI, the step size γ_1 and γ_2 are set to 1.0×10^{-4} and 1.0×10^{2}, respectively, the upper bound of the $l_{1,2}$ norm α is set to 340, and the lower and upper bounds of the velocity model a, b are set to 1.5[km/s] and 4.5[km/s], respectively. The number of iterations is set to 5000.

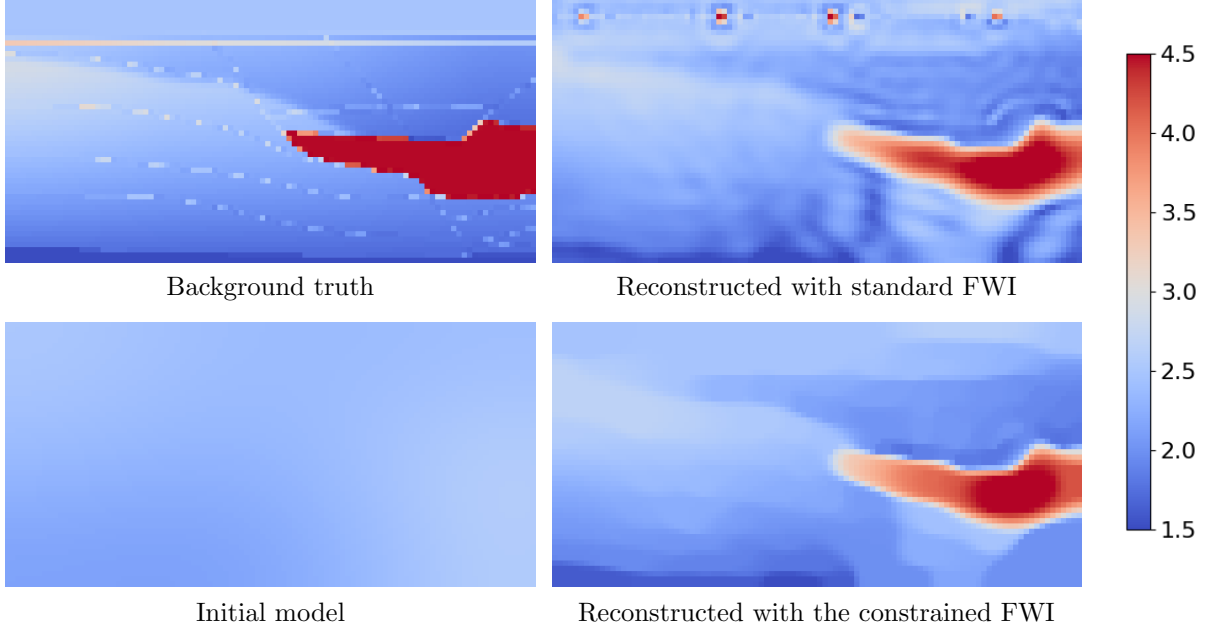| Background truth | Reconstructed with standard FWI |
| Initial model | Reconstructed with the constrained FWI |

Fig. 2: Velocity models and their corresponding reconstructions.

4.2 Results and Discussion

Fig.2 shows the ground truth, the initial model, and the reconstructed velocity models using the standard FWI and the TV and box constrained FWI. It can be observed that the TV and box constrained FWI successfully eliminates wave-like artifacts and noise that appear at the source positions, resulting in a more accurate velocity model reconstruction.

In Fig.3, we plot the Structural Similarity Index Measure (SSIM) against the number of iterations for both methods. The proposed method consistently achieves higher SSIM values than the standard FWI at every iteration, indicating improved reconstruction quality.

Furthermore, because the most computationally intensive part is the gradient computation of E, introducing the constraints does not significantly increase the overall computational cost. This demonstrates that the proposed method enhances reconstruction accuracy without incurring additional computational costs.

However, it should be noted that parameters such as α, a, and b were determined by referencing the ground truth data. While this experiment shows that good results can be achieved by appropriately setting these parameters, they need to be determined independently of this framework in practical applications.
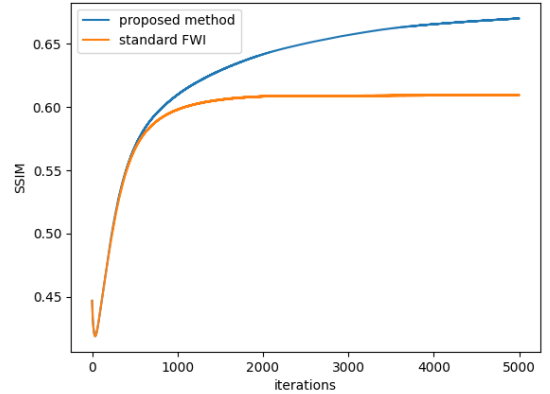

Fig. 3: SSIM against the number of iterations.

5 Conclusion

In this paper, we developed an efficient algorithm to solve the TV and box constrained FWI problem with neither an inner loop nor approximations based on PDS. We demonstrated that the constrained problem can be fully handled within the PDS algorithm. We also demonstrated that the piecewise smoothness by the TV constraint is well represented even when the PDS algorithm is used, and that efficient and accurate reconstruction is possible. Furthermore, the PDS framework allows for the incorporation of more complex constraints and regularizations, making it a valuable tool for future research.

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," *Journal of Applied 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 digital medicine*, 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," *Geophysical Journal International*, 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," *Geophysical 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. Métivier, R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, "Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion," *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, 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. 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.

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

[15] 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.

[16] Z. Du, D. Liu, G. Wu, J. Cai, X. Yu, and G. Hu, "A high-order total-variation regularisation method for full-waveform inversion," *Journal of Geophysics and Engineering*, vol. 18, no. 2, pp. 241–252, 2021.

[17] K. Gao and L. Huang, "Acoustic-and elastic-waveform inversion with total generalized p-variation regularization," *Geophysical Journal International*, vol. 218, no. 2, pp. 933–957, 2019.

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

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

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

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

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

[23] E. Esser, L. Guasch, F. J. Herrmann, and M. Warner, "Constrained waveform inversion for automatic salt flooding," *The Leading Edge*, vol. 35, no. 3, pp. 235–239, 2016.

[24] 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.

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

[26] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of mathematical imaging and vision*, vol. 40, pp. 120–145, 2011.

[27] P. L. Combettes and J.-C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, lipschitzian, and parallel-sum type monotone operators," *Set-Valued and variational analysis*, vol. 20, no. 2, pp. 307–330, 2012.

[28] L. Condat, "A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *Journal of optimization theory and applications*, vol. 158, no. 2, pp. 460–479, 2013.

[29] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Advances in Computational Mathematics*, vol. 38, no. 3, pp. 667–681, 2013.

[30] R.-E. Plessix, "A review of the adjoint-state method for computing the gradient of a functional with geophysical applications," *Geophysical Journal International*, vol. 167, no. 2, pp. 495–503, 2006.

[31] 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.

[32] L. Condat, "Fast projection onto the simplex and the l 1 ball," *Mathematical Programming*, vol. 158, no. 1, pp. 575–585, 2016.

[33] 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 Development*, vol. 12, no. 3, pp. 1165–1187, 2019.