

Incorporating multiple a priori information for full waveform inversion

Da Li, Michael P. Lamoureux and Wenyuan Liao

ABSTRACT

Full waveform inversion (FWI) is a powerful data-fitting procedure for the seismic inversion problem. However, it suffers the local minimum problem when the accurate initial model is not available because of the nonlinear nonconvex structure of the objective function. Our initial idea in this work is that a better inverse result can be expected when more a priori information of the physical model is provided. We propose a numerical scheme which can incorporate multiple a priori information to the optimization problem. First, a scaled gradient projection method on adaptive constraint sets is provided which is compatible with the inexact projection algorithm. Next, we incorporate a priori information as convex constraint sets. Then, the FWI problem is solved as a constraint optimization problem on the intersection of the constraint sets as the feasible set. Numerical examples with box constraints, total variation constraints, and l^1 constraints on the cross-well model and the reflective seismic wave model are provided.

INTRODUCTION

Full waveform inversion (FWI) is a high-resolution seismic inversion technique that has been widely studied and applied in both academia and industry. It is a data-fitting procedure by minimizing the distance between the recorded seismic data and the forward modeling seismic data which is generated through the simulation of the physical model. The physical model can be chosen from the acoustic wave equation to the elastic wave equation, depends on how accurate the physical model is expected.

FWI was introduced by (Lailly and Bednar, 1983) and (Tarantola, 1984) in the early 1980s. The development of scientific computing performance and the improvement of seismic data quality make the FWI one of the most important geophysical exploration methods. Also, new improvements have been made such as frequency domain inversion (Pratt, 1999; Pratt and Shipp, 1999), new regularization technique (Esser et al., 2018; Yong et al., 2018; Aghamiry et al., 2020), optimal transport distance (Engquist et al., 2016; Yang and Engquist, 2017; Yong et al., 2019; Métivier et al., 2018b,a), etc. We refer to (Virieux and Operto, 2009) and (Virieux et al., 2017) for a recent general review of the FWI problem.

From the mathematic point of view, FWI is a PDE-constrained optimization problem with a nonlinear nonconvex objective function. The nonconvex property determines that the local minima are unavoidable with the optimization algorithm. When the initial model is far away from the true model, especially on the large scale, the result is usually trapped in local minima which can be explained as the cycle-skipping issue. Although the local minima are unavoidable, there are several methods can be considered to improve the inverse result. First, change the distance in the objective function. The optimal transport distance is one choice since it is convex with respect to shift and dilation (Yang and Engquist, 2017; Yang et al., 2018). Second, we can expand the optimization space, for example as wavefield

reconstruction inversion (WRI) (van Leeuwen and Herrmann, 2015) and source extension method (Symes, 2008). Third, to describe the special properties of the model space, the regularization term can be added to the objective function. For example, the total variation regularization will lead to a piecewise constant structure of the inverse result. Last, the FWI problem can be implemented as a constraint optimization problem. The box constraint set is one of the most popular feasible set for the PDE-constrained optimization problem. The third and last method are equivalent based on the dual structure of the Lagrangian. Comparing to the regularization coefficients, the size of constraint sets can describe the a priori information directly and accurately.

Suppose for the grid point x_0 in the physical model, we know the average value of the ball $B(x_0, r)$ which centered at x_0 with the radius r . And suppose for each of the grid points in the physical domain, we know this kind of average value information a priori. As $r \rightarrow 0$, we actually have the values at all grid points. In other words, we already know the global solution to the inverse problem. This toy example suggests that, with more and more a priori information, a more accurate inverse result can be expected. The initial idea of this work is, the a priori information can be formulated as several convex constraint sets. And then the optimization problem can be formulated as a constraint optimization problem with the intersection of the above convex constraint sets as the feasible set. A scaled gradient projection method on the adaptive constraint sets are provided to fulfill our need. And the projection algorithm proposed in the work (Combettes, 2000, 2003) are integrated into this work.

Formulation of the FWI problem

The numerical scheme of the main algorithm is introduced in section 2 and 3. We review the FWI problem in a compact form first. Denote $U_{\text{ad}} \subset U$ be the feasible set of the physical parameters such as velocity or density, and Y as the space of the seismic wavefield. Given the recorded data $y_d \in Y$, and the record operator $Q : U \rightarrow Y$, the FWI problem can be formulated as a PDE-constrained optimization problem as:

$$\begin{aligned} \min_{(y,u) \in Y \times U_{\text{ad}}} J(y, u) &= \frac{1}{2} \|Qy - y_d\|_Y^2, \\ \text{such that } e(y, u) &= \mathcal{L}(u)y - s = 0. \end{aligned} \quad (1)$$

Here the constraint PDE is written in a compact form, which can be acoustic wave equation, elastic wave equation, etc. The differential operator $\mathcal{L} : U \times Y \rightarrow Z$ is linear with respect to y but nonlinear with respect to u . This nonlinearity is the reason for the local minima issue discussed above, and without change the equation, the nonlinearity will not change. Since the PDE $e(y, u) = 0$ is well posed, a parameter-to-state map is well defined as $F(u) = y$, then the reduced form of problem (1) is

$$\min_{u \in U_{\text{ad}}} f(u) = J(F(u), u). \quad (2)$$

The gradient of $f(u)$ can be achieved through the adjoint state method:

$$\nabla f(u) = \iint v(x, t) \partial_{tt} y(x, t) \, dx dt, \quad (3)$$

where $y(x, t)$ is the solution of the constraint equation, and $v(x, t)$ is the adjoint wavefield which is the solution of the adjoint equation

$$\mathcal{L}(u)v = Q'(Qy - y_d). \quad (4)$$

For more information of the adjoint state method in FWI problem, we refer to the paper (Plessix, 2006).

SCALED GRADIENT PROJECTION METHOD ON ADAPTIVE CONSTRAINT SETS

Before introducing the scaled gradient projection (SGP) method on adaptive constraint sets, we review the scaled gradient projection method for the general case. Consider the constraint optimization problem:

$$\min_{u \in \mathbb{R}^n} f(u), \quad \text{such that } u \in U_{\text{ad}}. \quad (5)$$

Here f is the nonlinear (possibly nonconvex) and smooth enough objective function. The feasible $U_{\text{ad}} \in \mathbb{R}^n$ is nonempty, convex, and closed. Since we are working with both n -dimensional Euclidean space and the scaled Euclidean space, we denote $\mathcal{H} = \mathbb{R}^n$ be the n dimensional Euclidean space with inner product $\langle x, y \rangle = x'y$ and norm $\|x\| = \sqrt{x'x}$. Given symmetric positive definite matrix B , let the space \mathcal{H}_B be the scaled Euclidean space with element in \mathbb{R}^n , the inner product $\langle x, y \rangle_B = \langle Bx, y \rangle = x'By$ and norm $\|x\|_B = \sqrt{x'Bx}$.

The scaled gradient projection method at k -th iteration is given by:

$$\bar{u}^k = \arg \min_{u \in U_{\text{ad}}} \langle \nabla f(u^k), u - u^k \rangle + \frac{1}{2\beta^k} \langle B_k(u - u^k), u - u^k \rangle, \quad (6)$$

$$u^{k+1} = u^k + \alpha^k(\bar{u}^k - u^k). \quad (7)$$

Here the matrix B_k can be chosen as an approximation of Hessian matrix $\nabla^2 f(x^k)$. When the Hessian matrix is symmetric positive definite, and $B_k = \nabla^2 f(x^k)$, the scaled gradient projection method is equivalent to the constrained Newton's method. We set the scale parameter $\beta^k = 1$ in this work. The linesearch parameter α^k can be achieved through the linesearch method like Armijo rule and Wolfe conditions.

Let $\tilde{u}^k = u^k - B_k^{-1}\nabla f(u^k)$, equation (6) is equivalent to

$$\bar{u}^k = \arg \min_{u \in U_{\text{ad}}} \frac{1}{2} \|u - \tilde{u}^k\|_{B_k}^2 - \frac{1}{2} \langle B_k^{-1}\nabla f(u^k), \nabla f(u^k) \rangle. \quad (8)$$

In other words, the equation (6) is equivalent to compute \tilde{u}^k as the global minimum of the quadratic equation in (6) first, then project \tilde{u}^k to the feasible set U_{ad} in the metric of \mathcal{H}_{B_k} . This projection computation can be written as $\bar{u}^k = P_{B_k, U_{\text{ad}}}(\tilde{u}^k)$, where $P_{B_k, U_{\text{ad}}}$ is the projection operator onto set U_{ad} in the metric of \mathcal{H}_{B_k} . This process is demonstrated in Figure 1.

However, there is an issue when the projection of \tilde{u}^k can not be evaluated exactly. Some of the projection method like Dykstra's algorithm computing the projection by generating

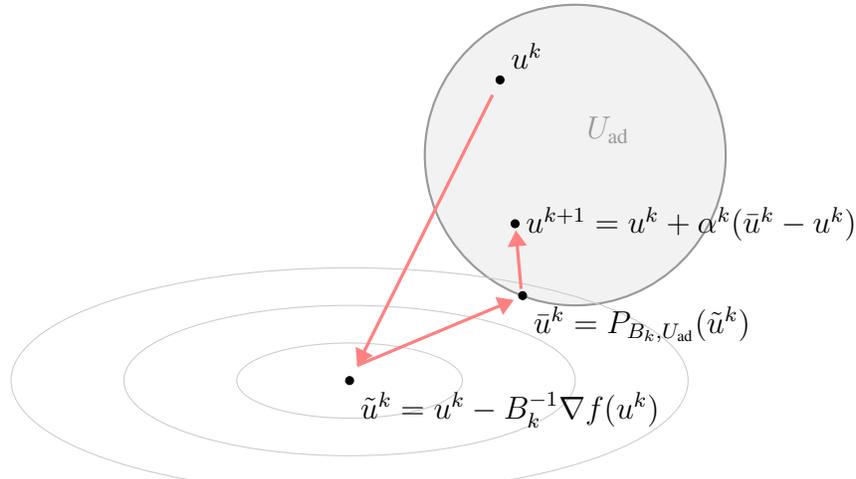


FIG. 1. SGP method at the k -th iteration.

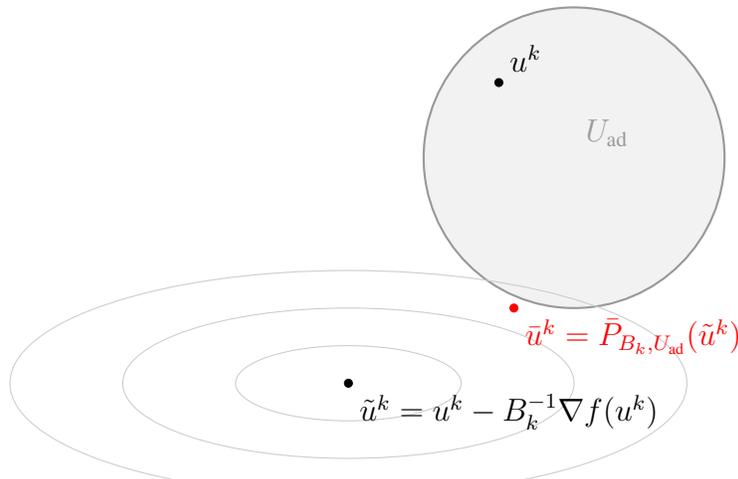


FIG. 2. When an inexact projection result is generated by the projection algorithm, \bar{u}^k may not in the feasible set U_{ad} , then u^{k+1} is not guaranteed in the feasible set U_{ad} .

a converging sequence which convergent to the exact projection. However, the convergent process has to be terminated after several iterations, and in this case, the result is different from the exact projection. Denote the inexact projection operator as $\bar{P}_{B_k, U_{\text{ad}}}$. Despite the inexact projection result that might be closed enough to the exact projection, the inexact projection might not fall in the feasible set U_{ad} . And this inexact projection might lead to that the update u^{k+1} will not stay in the feasible set. This problem is demonstrated in Figure 2.

To overcome the above problem, we design an increasing sequence of feasible sets, and the sets in the increasing sequence are assigned for each of the iterations of the algorithm. Given the feasible set U_{ad} , design the increasing sequence of the feasible set as

$$U_{\text{ad}} = \lim_{k \rightarrow \infty} U_{\text{ad}}^k, \text{ and } U_{\text{ad}}^k \subset \text{int}U_{\text{ad}}^{k+1}. \quad (9)$$

Then, the feasible set at k -th iteration is assigned as U_{ad}^k , and the feasible set at the next iteration is assigned as U_{ad}^{k+1} . The SGP method on the increasing sequence of feasible sets is: at k -th iteration, given symmetric positive definite matrix B_k ,

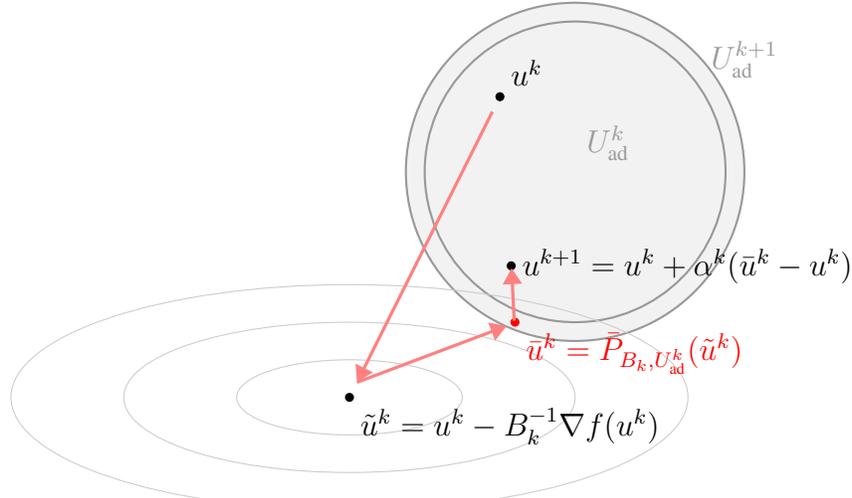


FIG. 3. To overcome the inexact projection issue, we expand the feasible set at each iteration. Find the inexact projection $\bar{u}^k \in U_{\text{ad}}^{k+1}$ first, then update $u^{k+1} \in U_{\text{ad}}^{k+1}$.

1. Compute $\tilde{u}^k = u^k - B_k^{-1} \nabla f(u^k)$.
2. Evaluate the inexact projection operator $\bar{u}^k = \bar{P}_{B_k, U_{\text{ad}}^k}(\tilde{u}^k)$, until the following equations are satisfied

$$\bar{u}^k \in U_{\text{ad}}^{k+1}, \quad (10)$$

$$\langle \tilde{u}^k - \bar{u}^k, u^k - \bar{u}^k \rangle_{B_k} \leq 0. \quad (11)$$

The equation (11) is a condition used in the convergence analysis and guarantees that the $\bar{u}^k - u^k$ is a decreasing direction.

3. Update $u^{k+1} = u^k + \alpha^k (\bar{u}^k - u^k)$, here α^k is determined by the linesearch algorithm. In this case, the update u^{k+1} is in the feasible set at $k + 1$ -th iteration U_{ad}^{k+1} .
4. Set $k = k + 1$, the feasible set at $k + 1$ -th iteration is U_{ad}^{k+1} .

Furthermore, an adaptive enlarge procedure can be designed for the feasible sets of each iteration. Suppose the increasing sequence of constraint sets is designed as $\{U_{\text{ad}}^h\}$, and the feasible set at k -th iteration is $U_{\text{ad}}^{h_0}$. Then in step 3 and 4 of the above algorithm, the feasible set of $k + 1$ -th iteration can be determined as

$$U_{\text{ad}}^{k+1} = \begin{cases} U_{\text{ad}}^{h_0+1}, & \text{if } \bar{u}^k \notin U_{\text{ad}}^{h_0}, \\ U_{\text{ad}}^{h_0}, & \text{if } \bar{u}^k \in U_{\text{ad}}^{h_0}. \end{cases} \quad (12)$$

The SGP method on adaptive constraint sets is given by Algorithm 1.

L-BFGS Hessian approximation of the scaling matrix

A scaling matrix B_k which approximates the Hessian matrix accurately at each iteration is important to increase the convergence speed. On the other hand, efficient evaluations of matrix-vector productions of both B_k and B_k^{-1} are needed in the projection algorithm

Algorithm 1: Scaled gradient projection method on adaptive constraint sets

Given the feasible set U_{ad} , construct the increasing sequence of constraint sets $\{U_{\text{ad}}^h\}$.

while *Not converge* **do**

Step 1: Compute $\tilde{u}^k = u^k - B_k^{-1} \nabla f(u^k)$.

Step 2: Evaluate the inexact projection operator $\bar{u}^k = \bar{P}_{B_k, U_{\text{ad}}^k}(\tilde{u}^k)$, until the equation (10) and (11) are satisfied.

Step 3: Update $u^{k+1} = u^k + \alpha^k (\bar{u}^k - u^k)$ with the linesearch algorithm.

Step 4: Update the feasible set U_{ad}^{k+1} with equation (12), set $k = k + 1$.

end

introduced in the following section. In this case, the L-BFGS matrix is an ideal choice for this work.

Denote $H_k = B_k^{-1}$, the L-BFGS matrix in a compact form is introduced in the work (Byrd et al., 1994). Denote

$$s_k = u^{k+1} - u^k, \quad y_k = \nabla f(u^{k+1}) - \nabla f(u^k). \quad (13)$$

Then,

$$B_k = \sigma_k I - [\sigma_k S_k \quad Y_k] \begin{bmatrix} \sigma_k S_k^T S_k & U_k \\ U_k^T & -D_k \end{bmatrix}^{-1} \begin{bmatrix} \sigma_k S_k^T \\ Y_k^T \end{bmatrix}. \quad (14)$$

$$H_k = \gamma_k I + [S_k \quad \gamma_k Y_k] \begin{bmatrix} R_k^{-T} (D_k + \gamma_k Y_k^T Y_k) R_k^{-1} & -R_k^{-T} \\ -R_k^{-1} & 0 \end{bmatrix} \begin{bmatrix} S_k^T \\ \gamma_k Y_k^T \end{bmatrix}. \quad (15)$$

Here the coefficients are given by

$$S_k = [s_{k-m}, \dots, s_{k-1}], \quad Y_k = [y_{k-m}, \dots, y_{k-1}], \quad (16)$$

$$(R_k)_{i,j} = \begin{cases} (s_{k-m-1+i})^T (y_{k-m-1+j}), & \text{if } i \leq j, \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

$$D_k = \text{diag}[s_{k-m}^T y_{k-m}, \dots, s_{k-1}^T y_{k-1}], \quad (18)$$

$$(U_k)_{i,j} = \begin{cases} (s_{k-m-1+i})^T (y_{k-m-1+j}), & \text{if } i > j, \\ 0, & \text{otherwise.} \end{cases} \quad (19)$$

A PRIORI INFORMATION AS CONVEX CONSTRAINT SETS

As we discussed in section 1 and 2, the initial idea of this work is to transform the a priori information into convex constraint sets. Then solve the FWI problem as the constraint optimization on the intersection of these convex constraint sets. A projection algorithm is needed to use the SGP method we developed in the previous section. In this section, we discuss the convex constraint sets first, and then introduce a projection algorithm developed by (Combettes, 2000, 2003) which can project a point onto the intersection of convex sets. The main algorithm for the FWI problem is provided at the end of this section.

Before the discussion of convex constraint sets, we define a threshold function that is useful for constructing the increase sequence. Denote $\varepsilon > 0$ be a threshold and $\eta \in (0, 1)$ is a parameter to control the sequence of sets expanding speed. We need the sequence to be expanding as iteration goes on, but not expand to infinity large. To control the increasing sequence of sets not expand to infinity large, define the threshold function as

$$\theta(h) = \begin{cases} 0, & \text{if } h = 0, \\ \sum_{i=1}^h \eta^i \varepsilon, & \text{if } h \geq 1, \\ \frac{\eta}{1-\eta} \varepsilon, & \text{if } h \rightarrow \infty, \end{cases} \quad (20)$$

Here $h \in \mathbb{N}$ is the index controlling the sequence increasing.

Convex sets with closed form projection function

There are two kinds of convex constraint sets that are discussed in this work. We start with the convex sets with the closed form projection function, for example as box constraint, hyperplane constraint, etc. One of the most commonly used constraint sets in the constraint optimization problem is the box constraint which can provide the physical bounds of the model. Given $a, b \in \mathbb{R}$, with $a \leq b$, the box constraint set is given by:

$$U_{\text{box}} = \{u \in \mathbb{R}^n \mid a \leq u_i \leq b, i = 1, \dots, n\}. \quad (21)$$

The closed form projection function of the box constraint set is given by:

$$P_{\text{box}}(u)_i = \begin{cases} a, & \text{if } u_i < a, \\ u_i, & \text{if } a \leq u_i \leq b, \\ b, & \text{if } b < u_i, \end{cases} \quad \text{or } P_{\text{box}}(u)_i = \max(a, \min(u_i, b)). \quad (22)$$

Then the increasing sequence of box constraint set is given by:

$$U_{\text{box}}^h = \{u \in \mathbb{R}^n \mid a - \theta(h) \leq u_i \leq b + \theta(h), i = 1, \dots, n\}, \quad h \in \mathbb{N}, \quad (23)$$

where the threshold function $\theta(h)$ is defined by equation (20).

To represent the average value of a certain area, the affine hyperplane constraint can be used. Given $p \in \mathbb{R}^n$ and $\kappa \in \mathbb{R}$, the affine hyperplane is defined by

$$U_{\text{plane}} = \{u \in \mathbb{R}^n \mid \langle u, p \rangle = \kappa\}. \quad (24)$$

The closed form projection function of the hyperplane constraint set is:

$$P_{\text{plane}}(u) = u + \frac{\kappa - \langle u, p \rangle}{\|p\|^2} p, \quad (25)$$

and the increasing sequence of hyperplane sets can be defined by:

$$U_{\text{plane}}^h = \{u \in \mathbb{R}^n \mid \|u - P_{\text{plane}}(u)\| \leq \theta(j)\}, \quad h \in \mathbb{N}. \quad (26)$$

Convex sets with subgradient projection

The subgradient projection method can be implemented for some convex sets without a closed form projection function, for example as total variation constraints, l^1 constraints, etc. Instead of the exact projection onto the convex set, the subgradient projection project the point onto an outer approximation of the convex set. To have the subgradient projection, we need to represent the convex set as lower level set of a continuous convex function.

Definition 1 (lower level set) *Given a continuous convex function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, the lower level set function of f with a height $\eta \in \mathbb{R}$ is given by*

$$\text{lev}_{\leq \eta} f = \{x \in \mathbb{R}^n \mid f(x) \leq \eta\}. \quad (27)$$

For the case when C is some lower level set of a continuous convex function f and height η , i.e.

$$C = \text{lev}_{\eta} f = \{x \in \mathbb{R}^n \mid f(x) \leq \eta\}, \quad (28)$$

the subgradient projection is an efficient way to approximate the projection. The following proposition provides an outer approximation of convex set C .

Proposition 1 *Given continuous convex function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, vector $x \in \mathbb{R}^n$ and $x \notin C$, and the subgradient $x^* \in \partial f(x)$, the lower level set C is defined by equation (28). The half-space set*

$$H_x = \{z \in \mathbb{R}^n \mid f(x) + \langle x^*, z - x \rangle \leq \eta\}. \quad (29)$$

is an approximation of set C . Also, we have $x \notin H_x$.

Then, the subgradient projection is given by the following definition.

Definition 2 *Given continuous convex function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, vector $x \in \mathbb{R}^n$, and the subgradient $x^* \in \partial f(x)$, the lower level set C is defined by equation (28). The subgradient projection function which project x towards C is given by*

$$\tilde{P}_C(x) = \begin{cases} x + \frac{\eta - f(x)}{\|x^*\|^2} x^*, & \text{if } f(x) > \eta, \\ x, & \text{if } f(x) \leq \eta. \end{cases} \quad (30)$$

Total variation regularization is a popular technique in imaging problems to provide a piecewise constant result. The total variation constraint set is constructed by the above method. Consider a two dimensional digital image $u \in \mathbb{R}^{N_x \times N_y}$ with N_x rows and N_y columns, and $n = N_x \times N_y$. It is equivalent to consider u is a vector in \mathbb{R}^n . Define the discrete gradient operator $D : \mathbb{R}^{N_x \times N_y} \rightarrow \mathbb{R}^{N_x \times N_y \times 2}$ with

$$(Du)_{i,j,1} = \begin{cases} u_{i+1,j} - u_{i,j}, & \text{if } 0 \leq i < N_x, \\ 0, & \text{if } i = N_x, \end{cases} \quad (31)$$

$$(Du)_{i,j,2} = \begin{cases} u_{i,j+1} - u_{i,j}, & \text{if } 0 \leq j < N_y, \\ 0, & \text{if } j = N_y. \end{cases} \quad (32)$$

Then the discrete total variation norm is given by the TV function $f_{\text{tv}} : \mathbb{R}^n \rightarrow \mathbb{R}$

$$f_{\text{tv}}(u) = \|u\|_{\text{tv}} = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} |(Du)_{i,j}|. \quad (33)$$

Given the radius τ_{tv} , the increasing sequence of the total variation constraint set can be constructed as

$$U_{\text{tv}}^h = \{u \in \mathbb{R}^n \mid f_{\text{tv}}(u) \leq \theta(h) + \tau_{\text{tv}}\}. \quad (34)$$

The subgradient projection function for the sequence of total variation constraint sets U_{tv}^h is

$$\tilde{P}_{U_{\text{tv}}^h}(u) = \begin{cases} u + \frac{\theta(h) + \tau_{\text{tv}} - f_{\text{tv}}(u)}{\|u^*\|^2} u^*, & \text{if } f_{\text{tv}}(u) > \theta(h) + \tau_{\text{tv}}, \\ u, & \text{if } f_{\text{tv}}(u) \leq \theta(h) + \tau_{\text{tv}}. \end{cases} \quad (35)$$

Next, we discuss how to transform the sparsity a priori information to the convex constraint set with l^1 ball. Given a matrix $\Phi \in \mathbb{R}^{n \times m}$, with $\Phi = [\phi_1, \dots, \phi_m]$, here each ϕ_i , $i = 1, \dots, m$ is a n -dimensional row vector represents some basis of \mathbb{R}^n . Here Φ represents a linear transformation that maps the signal u to the coefficient spaces \mathbb{R}^m . Typical chooses of Φ can be Fourier transform, wavelet transform, or curvelet transform, etc. We define the l^1 function with linear operator Φ as

$$f_{l^1}(u) = \|\Phi u\|_1 = \sum_{j=1}^m |\langle \phi_j, u \rangle| = \sum_{j=1}^m \left| \sum_{i=1}^n \Phi_{i,j} u_i \right|. \quad (36)$$

Given initial radius τ_{l^1} , by the same method, the increasing sequence of l^1 constraint sets can be defined as

$$U_{l^1}^h = \{u \in \mathbb{R}^n \mid f_{l^1}(u) \leq \theta(h) + \tau_{l^1}\}, \quad (37)$$

The subgradient projection function for the l^1 constraint set $U_{l^1}^h$ is

$$\tilde{P}_{U_{l^1}^h}(u) = \begin{cases} u + \frac{\theta(h) + \tau_{l^1} - f_{l^1}(u)}{\|u^*\|^2} u^*, & \text{if } f_{l^1}(u) > \theta(h) + \tau_{l^1}, \\ u, & \text{if } f_{l^1}(u) \leq \theta(h) + \tau_{l^1}. \end{cases} \quad (38)$$

Projection onto intersection of convex sets

Given a symmetric positive definite matrix B , convex sets U_1, \dots, U_{N_c} , and the initial value $u^0 \notin \cap_{i=1}^{N_c} U_i$. Let $U = \cap_{i=1}^{N_c} U_i$, the projection problem which project the point u^0 onto U_i in the metric of \mathcal{H}_B is given by

$$P_{B,U}(u^0) = \arg \min_{u \in U} \|u - u^0\|_B^2. \quad (39)$$

To solve the above projection problem, we introduce the projection algorithm developed in (Combettes, 2000, 2003) to our main algorithm. Algorithm 2 is the projection algorithm used in this work.

Notice that, the matrix-vector productions of both B and B^{-1} are needed in the Algorithm 2. For computation efficiency, the L-BFGS approximation of the Hessian matrix is the ideal choice of this numerical scheme.

Algorithm 2: The projection algorithm

Initialization: Given a symmetric positive definite matrix B , a family of nonempty convex closed set $U_i, i = 1, \dots, N_c$, initial point u^0 , weight parameter ω_i , with $\sum_{i=1}^m \omega_i = 1$.

At k -th iteration:

while Not converge **do**

Step 1: compute the projection of u^k onto each of U_i with:

$$p_i = \begin{cases} P_{U_i}(u^k), & \text{if } U_i \text{ is simple,} \\ \tilde{P}_{U_i}(u^k), & \text{if } U_i \text{ is not simple and have subgradient projection.} \end{cases} \quad (40)$$

Step 2: set $z^k = u^k + \lambda_k B^{-1} (\sum_{i \in I} \omega_i p_i - u^k)$, where λ_k is given by equation:

$$\lambda_k = \begin{cases} \frac{\sum_{i \in I} \omega_i \|p_i - u^k\|^2}{\|u^k - \sum_{i \in I} \omega_i p_i\|_{B^{-1}}}, & \text{if } u^k \notin \cap_{i=1}^m U_i, \\ 1/\|B^{-1}\|, & \text{otherwise.} \end{cases} \quad (41)$$

Step 3: Set $\pi_k = \langle u^0 - u^k, u^k - z^k \rangle_B$, $\mu_k = \|u^0 - u^k\|_B^2$, $\nu_k = \|u^k - z^k\|_B^2$, and $\rho_k = \mu_k \nu_k - \pi_k^2$, update $u^{k+1} = Q_B(u^0, u^k, z^k)$ with equation:

$$Q_B(u^0, u^k, z^k) = \begin{cases} z^k, & \text{if } \rho_k = 0, \pi_k \geq 0, \\ u^0 + \left(1 + \frac{\pi_k}{\nu_k}\right) (z^k - u^k), & \text{if } \rho_k > 0, \pi_k \nu_k \geq \rho_k, \\ u^k + \frac{\nu_k}{\rho_k} (\pi_k (u^0 - u^k) + \mu_k (z^k - u^k)), & \text{if } \rho_k > 0, \pi_k \nu_k < \rho_k. \end{cases} \quad (42)$$

Set $k = k + 1$.

end

Main algorithm

The main algorithm in this work is a combination of the SGP method on adaptive constraint sets, the L-BFGS Hessian approximation, and the projection algorithm. Given convex constraint sets U_1, \dots, U_{N_c} , let $I = \{1, \dots, N_c\}$ be the index set, denote the feasible set U_{ad} as the intersection of these sets:

$$U_{\text{ad}} = \cap_{i \in I} U_i. \quad (43)$$

We rewrite the constraint algorithm as

$$\min_{u \in \mathbb{R}^n} f(u), \quad \text{such that } u \in U_{\text{ad}}. \quad (44)$$

First, the increasing sequence of constraint sets is need to be fixed. For each of the constraint set U_i , design an increasing sequence $\{U_i^h\}$ satisfies

$$U_i = \lim_{h \rightarrow \infty} U_i^h, \quad \text{and } U_i^h \subset \text{int}U_i^{h+1}. \quad (45)$$

Then the feasible set of each iteration can be constructed with the increase sequences $\{U_i^h\}$. For example as

$$U_{\text{ad}}^k = \bigcap_{i \in I} U_i^{h_i}. \quad (46)$$

The main algorithm is given by Algorithm 3.

Algorithm 3: Scaled gradient projection on sequence of multiple adaptive constraint sets

Given: the objective function f and initial value u^0 ; a family of nonempty, closed, convex constraint sets U_i , for $i \in I$,

Construct: for each U_i construct an increasing set sequence $\{U_i^h\}_{h \in \mathbb{N}}$; set

$$U_{\text{ad}}^0 = \bigcap_{i \in I} U_i^0.$$

while *Not converge* **do**

Step 1: Compute the gradient $\nabla f(u^k)$.

Step 2: Update s_k and y_k with equation (13), S_k, Y_k, R_k, U_k with equation (16).

Step 3: Compute $\tilde{u}^k = u^k - H_k \nabla f(u^k)$ with equation (15).

Step 4: Compute $\bar{u}^k = P_{B_k, U_{\text{ad}}^k}(\tilde{u}^k)$, i.e., project \tilde{u}^k to U_{ad}^k in \mathcal{H}_{B_k} with the Algorithm 2, until the stopping criteria equation (10) (11) are satisfied. The multiplication between B_k, H_k and vectors are evaluated with equation (14) and (15).

Step 5: Update $u^{k+1} = u^k + \alpha_k(\bar{u}^k - u^k)$, here α_k is the linesearch parameter achieved with the Wolfe conditions.

Step 6: For each $i = 1, \dots, N_c$, update the constraint sets:

$$U_i^{k+1} = \begin{cases} U_i^h, & \text{if } u^{k+1} \in \text{int}U_i^h, \\ U_i^{h+1}, & \text{if } u^{k+1} \notin \text{int}U_i^h. \end{cases} \quad (47)$$

Step 7: Construct $U_{\text{ad}}^{k+1} = \bigcap_{i \in I} U_i^{k+1}$, set $k = k + 1$.

end

NUMERICAL EXAMPLES

Three numerical examples of the FWI problem are provided in this section, with example 1 and 2 are the cross-well model, and example 3 are the reflective wave model. The box constraint, hyperplane constraint, total variation constraint, and l^1 constraint are discussed in this section.

Before the numerical examples, we fix the physical model in the FWI problem (1). The wave equation is used as the constraint PDE, and the corresponding physical model is the velocity field. To simulate the wave propagation in free space, the perfectly matched layer technique is applied. And the finite difference method is used for the numerical computation of the PDE.

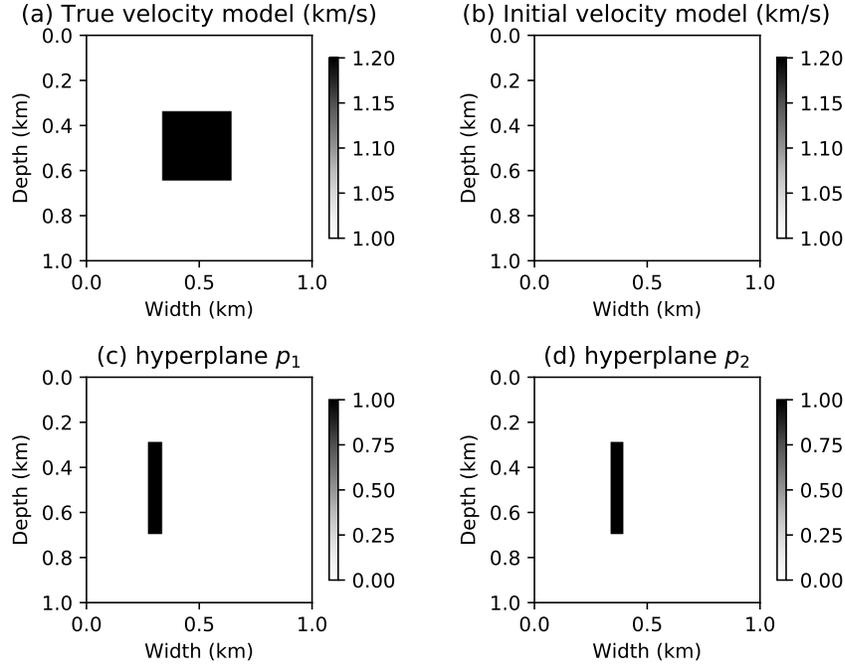


FIG. 4. (a): true velocity model. (b): initial velocity model used in the seismic inverse problem. (c): the hyperplane p_1 . (d): the hyperplane p_2 .

Example 1

A cross-well model is studied in this example as shown in Figure 4 (a), denoted as u_{true} . The initial velocity model is shown in Figure 4 (b). There 6 equally spaced sources in the left boundary of the domain, and there are 51 equally spaced receivers in the right boundary of the domain. The model is 1 km by 1 km and is discretized with model size 101×101 . A second-order finite difference scheme is used to discrete the wave equation with spatial step size 0.01 km and temporal step size 0.0005 s. And the perfectly matched layer technique is used to simulate the wave propagation in a boundary-free domain. The source is a Ricker wavelet with a peak frequency of 5 Hz.

Three constraints are considered: box constraint, total variation constraint, l^1 ball constraint. The sequence of box constraint sets is given by

$$U_1^h = \{u \in \mathbb{R}^n \mid 1 - \theta_1(h) \leq u_i \leq 1.2 + \theta_1(h), i = 1, \dots, n\}, \quad (48)$$

$$\text{where } \theta_1(h) = \begin{cases} 0, & \text{if } h = 0, \\ \sum_{i=1}^h 0.001 \times 0.9^i, & \text{otherwise.} \end{cases} \quad (49)$$

The sequence of total variation constraint sets is given by

$$U_2^h = \{u \in \mathbb{R}^n \mid f_{\text{tv}}(u) \leq 24 + \theta_2(h)\}, \quad (50)$$

$$\text{where } \theta_2(h) = \begin{cases} 0, & \text{if } h = 0, \\ \sum_{i=1}^h 0.24 \times 0.9^i, & \text{otherwise.} \end{cases} \quad (51)$$

Here the TV function f_{tv} is given in equation (33). The sequence of first hyperplane con-

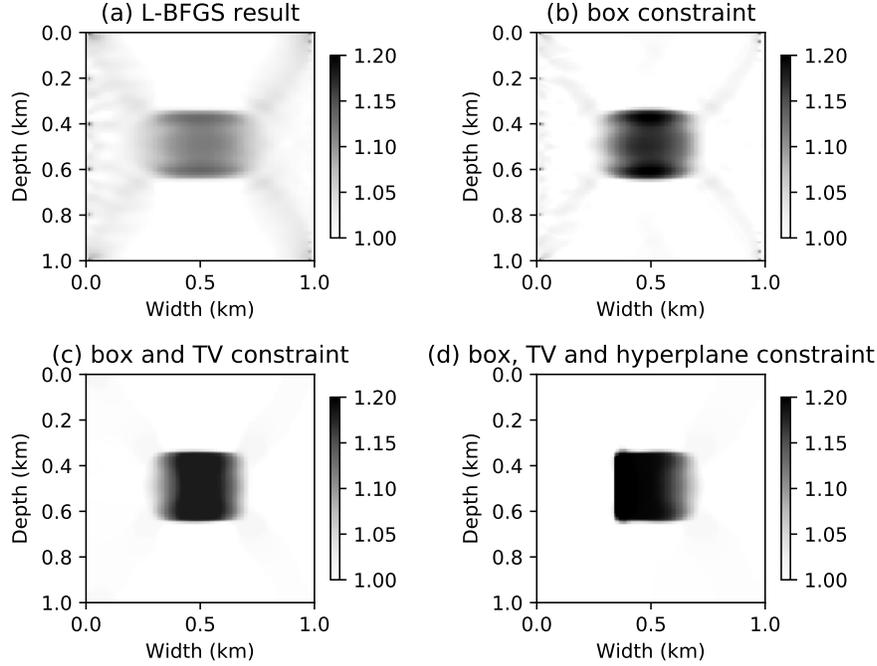


FIG. 5. (a): unconstraint result. (b): inverse result with box constraint. (c): inverse result with box and total variation constraint. (d): inverse result with box, total variation and hyperplane constraint.

straint sets is given by

$$U_3^h = \{u \in \mathbb{R}^n \mid \|u - P_1(u)\| \leq \theta_3(h) + 0.01\}, \quad (52)$$

$$\text{where } P_1(u) = u + \frac{\langle u_{\text{true}}, p_1 \rangle - \langle u, p_1 \rangle}{\|p_1\|^2} p_1, \quad (53)$$

$$\theta_3(h) = \begin{cases} 0, & \text{if } h = 0, \\ \sum_{i=1}^h 0.01 \times 0.9^i, & \text{otherwise.} \end{cases} \quad (54)$$

And sequence of second hyperplane constraint sets is given by

$$U_4^h = \{u \in \mathbb{R}^n \mid \|u - P_2(u)\| \leq \theta_4(h) + 0.01\}, \quad (55)$$

$$\text{where } P_2(u) = u + \frac{\langle u_{\text{true}}, p_2 \rangle - \langle u, p_2 \rangle}{\|p_2\|^2} p_2, \quad (56)$$

$$\theta_4(h) = \begin{cases} 0, & \text{if } h = 0, \\ \sum_{i=1}^h 0.01 \times 0.9^i, & \text{otherwise.} \end{cases} \quad (57)$$

The vector $p_1, p_2 \in \mathbb{R}^n$ are shown in Figure 4 (c) and (d). The hyperplane constraints used here are to provide the average value information of the two areas near the left boundary of the velocity perturbation. In this case, an inverse result with a sharp left boundary is expected.

The inverse results are shown in figure 5. For the unconstraint result in figure (a), the L-BFGS method is performed with 20 iterations. The scaled gradient projection method with adaptive constraints is performed with 20 iterations for the results in figure (b), (c),

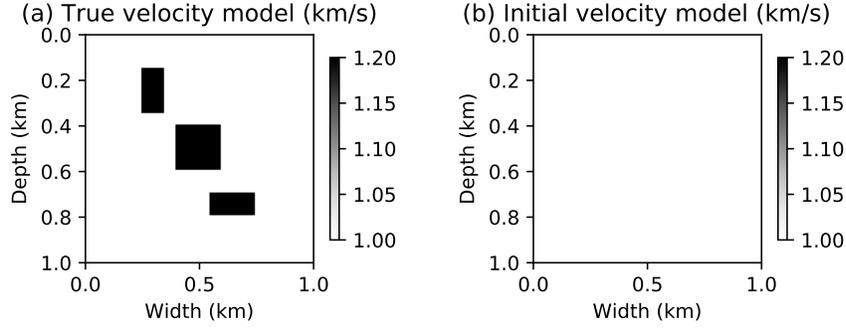


FIG. 6. (a): true velocity model. (b): initial velocity model.

and (d). Figure (b) provides the inverse result with only box constraint sequence $\{U_1^h\}$ with the index set $I = \{1\}$ and weight parameter $\omega_1 = 1$. Figure (c) provides the inverse result with box and total variation constraint sequences, with the index set $I = \{1, 2\}$ and the weight parameter $\omega_1 = \omega_2 = 1/2$. The inverse result with all four constraint sequences of sets is provided in figure (d), with the index set $I = \{1, 2, 3, 4\}$ and the weight parameter $\omega_1 = \omega_2 = \omega_3 = \omega_4 = 1/4$. All the constraints play a role in improving the inverse result comparing with the unconstraint case. Although the result of hyperplane constraints is artificial, a sharp left boundary of the velocity perturbation is inverted. This provides a way to increase local inverse results with accurate a priori information. The above examples show that the proposed method can handle multiple constraint sets at the same time. With more information provided for the optimization algorithm, a more accurate image can achieve.

Example 2

In this example, we incorporate the sparsity constraint with the proposed method, a cross-well model similar to example 1 is provided. The true velocity model and initial velocity model are shown in figure 6 (a) and (b). We use the initial velocity model as a reference model in the l^1 fidelity constraint sets, denoted as u_{ref} . The acquisition is the same as example 1.

Next, we denote the sequences of constraint sets. The sequence of box constraint sets is given by

$$U_1^h = \{u \in \mathbb{R}^n \mid 1 - \theta_1(h) \leq u_i \leq 1.2 + \theta_1(h), i = 1, \dots, n\}, \quad (58)$$

$$\text{where } \theta_1(h) = \begin{cases} 0, & \text{if } h = 0, \\ \sum_{i=1}^h 0.001 \times 0.9^i, & \text{otherwise.} \end{cases} \quad (59)$$

The sequence of total variation constraint sets is given by

$$U_2^h = \{u \in \mathbb{R}^n \mid f_{\text{tv}}(u) \leq 39.5 + \theta_2(h)\}, \quad (60)$$

$$\text{where } \theta_2(h) = \begin{cases} 0, & \text{if } h = 0, \\ \sum_{i=1}^h 0.395 \times 0.9^i, & \text{otherwise.} \end{cases} \quad (61)$$

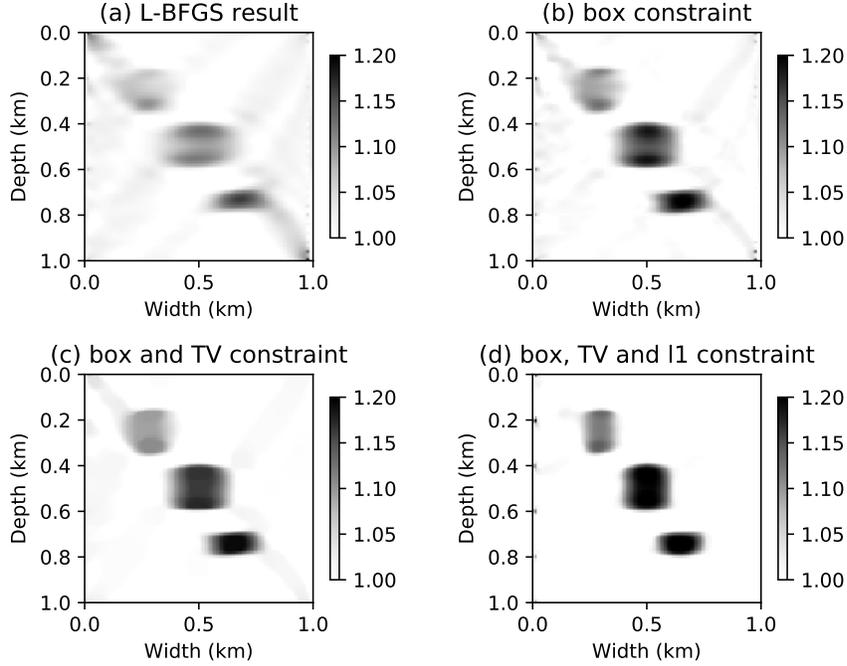


FIG. 7. (a): unconstraint result. (b): inverse result with box constraint. (c): inverse result with box and total variation constraint. (d): inverse result with box, total variation and l^1 constraint.

The sequence of l^1 constraint sets is given by

$$U_3^h = \{u \in \mathbb{R}^n \mid \|u - u_{\text{ref}}\|_1 \leq 128 + \theta_3(h)\}, \quad (62)$$

$$\text{where } \theta_3(h) = \begin{cases} 0, & \text{if } h = 0, \\ \sum_{i=1}^h 1.28 \times 0.9^i, & \text{otherwise.} \end{cases} \quad (63)$$

Numerical results are shown in Figure 7. For the unconstraint case, the L-BFGS method is performed 20 iterations and the result is shown in figure (a). The scaled gradient projection method with adaptive constraints is performed with 20 iterations for the results in figure (b), (c), and (d). Figure (b) provides the inverse result with only the sequence of box constraint U_1^h , with the index set $I = \{1\}$ and weight parameter $\omega_1 = 1$. Figure (c) provides the inverse result with the sequence of box and total variation constraint U_2^h , with the index set $I = \{1, 2\}$ and weight parameter $\omega_1 = \omega_2 = 1/2$. The inverse result with all three sequences of constraints is shown in figure (d), with the index set $I = \{1, 2, 3\}$ and weight parameter $\omega_1 = \omega_2 = \omega_3 = 1/3$. Comparing with different inverse results, both total variation constraint and l^1 constraint play an important role in providing a better inverse result.

Example 3

A more realistic velocity model is provided in this example as shown in Figure 8 (a). And the initial velocity model is shown in Figure 8 (b). With 3.8 km depth and 12.55 km width, the model is discretized into 76×251 points. There are 10 equally spaced sources and 126 equally spaced receivers on the top of the model. The second-order finite

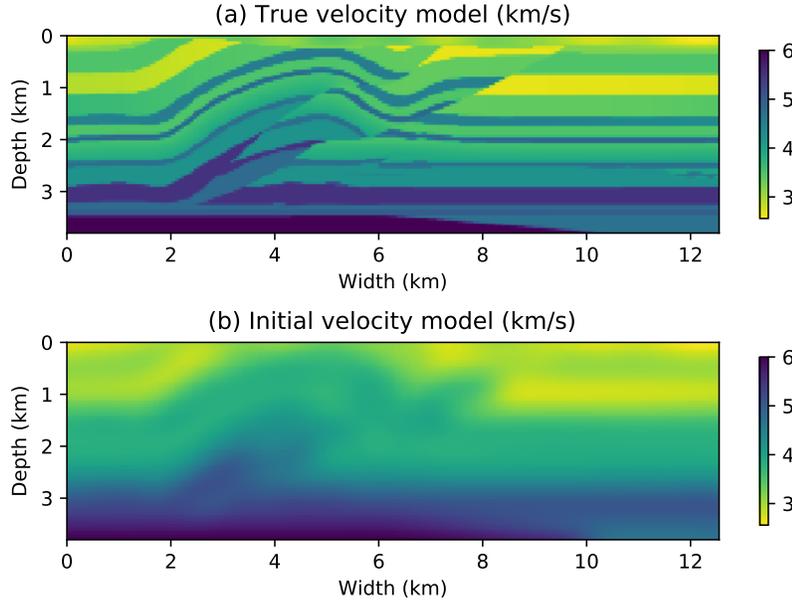


FIG. 8. (a): true velocity model. (b): initial velocity model.

difference method is used for the constraint equation, with spatial step size 0.05 km, and temporal step size 0.004 s. The perfectly matched layer technique is used to simulate the seismic wave propagating in the free domain. The 5 Hz Ricker wavelet is used for each of the sources.

In this example, we compare the inverse results with the different thresholds of the constraints. First, we fix the box constraint as

$$U_1^h = \{u \in \mathbb{R}^n \mid 2.5588 - \theta_1(h) \leq u_i \leq 6 + \theta_1(h), i = 1, \dots, n\}, \quad (64)$$

$$\text{where } \theta_1(h) = \begin{cases} 0, & \text{if } h = 0, \\ \sum_{i=1}^h 0.02 \times 0.9^i, & \text{otherwise.} \end{cases} \quad (65)$$

Then we define three total variation constraint sequences with different radius

$$U_2^h = \{u \in \mathbb{R}^n \mid f_{\text{tv}}(u) \leq 800 + \theta_2(h)\}, \quad (66)$$

$$\text{where } \theta_2(h) = \begin{cases} 0, & \text{if } h = 0, \\ \sum_{i=1}^h 40 \times 0.9^i, & \text{otherwise.} \end{cases} \quad (67)$$

$$U_3^h = \{u \in \mathbb{R}^n \mid f_{\text{tv}}(u) \leq 1000 + \theta_3(h)\}, \quad (68)$$

$$\text{where } \theta_3(h) = \begin{cases} 0, & \text{if } h = 0, \\ \sum_{i=1}^h 50 \times 0.9^i, & \text{otherwise.} \end{cases} \quad (69)$$

$$U_4^h = \{u \in \mathbb{R}^n \mid f_{\text{tv}}(u) \leq 1200 + \theta_4(h)\}, \quad (70)$$

$$\text{where } \theta_4(h) = \begin{cases} 0, & \text{if } h = 0, \\ \sum_{i=1}^h 60 \times 0.9^i, & \text{otherwise.} \end{cases} \quad (71)$$

For each of the following examples, 50 iterations are performed, and the inverse results are shown in Figure 9. The unconstrained result is shown in figure (a). In the case of figure (b), (c), and (d), both box constraint and total variation constraint are performed. For figure (b), the constraint sequence set index is $I = \{1, 2\}$, with weight parameter $\omega_1 = \omega_2 = 1/2$. For figure (c), the constraint sequence set index is $I = \{1, 3\}$, with weight parameter $\omega_1 = \omega_3 = 1/2$. For figure (d), the constraint sequence set index is $I = \{1, 4\}$, with weight parameter $\omega_1 = \omega_4 = 1/2$. Comparing with the above inverse results, as the total variation constraint radius is larger, the result is closer to the unconstrained case. The results show that the proposed method can control the inverse result by changing the radius of the sequence of constraint sets.

CONCLUSIONS

In this work, we use the viewpoint that a priori information can be incorporated in the FWI problem as convex constraint sets. And then the FWI problem is solved as a constrained optimization problem with the intersection of the convex constraint sets as the feasible set. A numerical framework is provided to solve the above problem which is a combination of scaled gradient projection method, L-BFGS Hessian approximation, and the projection algorithm. Numerical examples show that the proposed algorithm is flexible for integrating multiple constraints at the same time, and is easy to control the constraint effects by changing the size of constraint sets. Also, enhanced inverse results can be expected with appropriate a priori information been incorporated.

ACKNOWLEDGEMENTS

We thank the sponsors of CREWES for continued support. The first author thanks the China Scholarship Council (CSC) for supporting the research. The work of the first and second author was funded by NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 461179-13 and the Discovery grant RGPIN-2015-06038. The work of the first and third author was funded by NSERC through the grant CRDPJ 532227-18 and the Discovery grant RGPIN-2019-04830.

REFERENCES

- Aghamiry, H. S., Gholami, A., and Operto, S., 2020, Full waveform inversion with adaptive regularization: arXiv preprint arXiv:2001.09846.
- Byrd, R. H., Nocedal, J., and Schnabel, R. B., 1994, Representations of quasi-newton matrices and their use in limited memory methods: *Mathematical Programming*, **63**, No. 1-3, 129–156.
- Combettes, P. L., 2000, Strong convergence of block-iterative outer approximation methods for convex optimization: *SIAM Journal on Control and Optimization*, **38**, No. 2, 538–565.
- Combettes, P. L., 2003, A block-iterative surrogate constraint splitting method for quadratic signal recovery: *IEEE Transactions on Signal Processing*, **51**, No. 7, 1771–1782.
- Engquist, B., Froese, B. D., and Yang, Y., 2016, Optimal transport for seismic full waveform inversion: arXiv preprint arXiv:1602.01540.
- Esser, E., Guasch, L., van Leeuwen, T., Aravkin, A. Y., and Herrmann, F. J., 2018, Total variation regularization strategies in full-waveform inversion: *SIAM Journal on Imaging Sciences*, **11**, No. 1, 376–406.

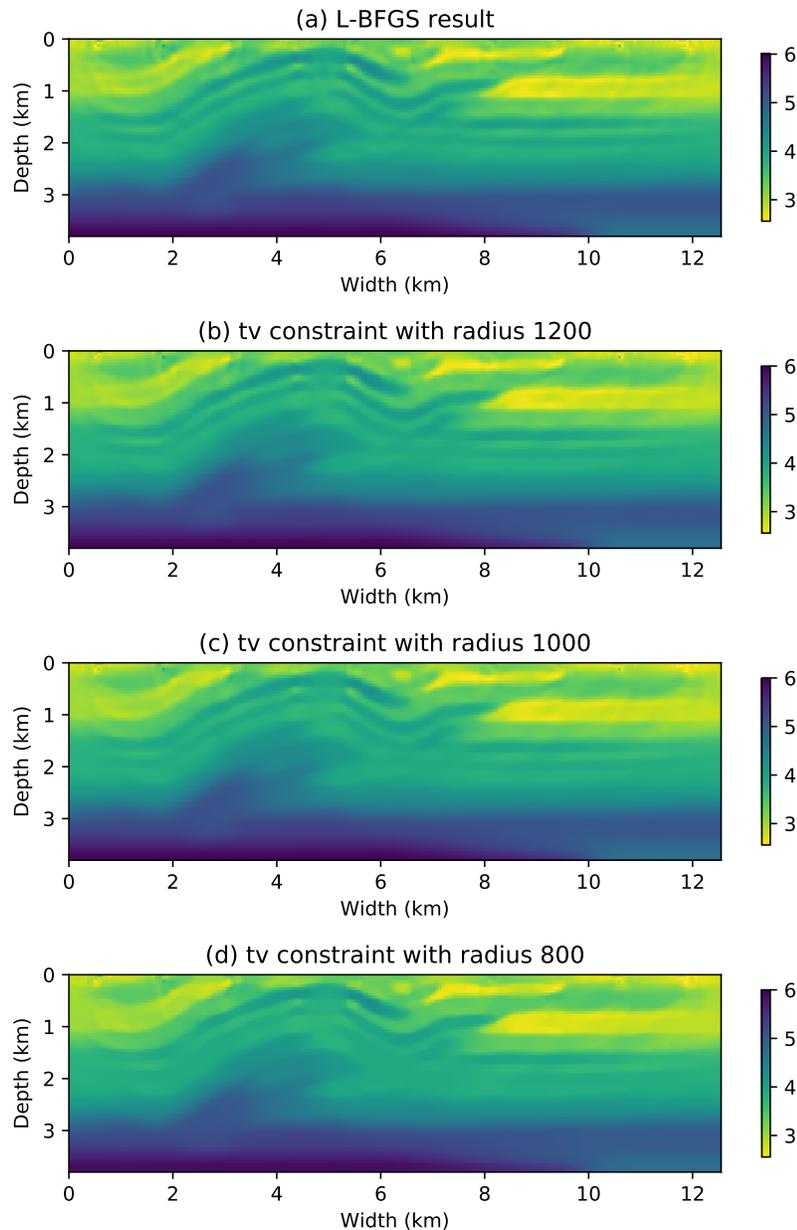


FIG. 9. (a): unconstraint result. (b): the inverse result with box constraint and total variation constraint with radius 1200. (c): the inverse result with box constraint and total variation constraint with radius 1000. (d): the inverse result with box constraint and total variation constraint with radius 800.

Lailly, P., and Bednar, J., 1983, The seismic inverse problem as a sequence of before stack migrations, *in* Conference on inverse scattering: theory and application, SIAM Philadelphia, PA, 206–220.

Métivier, L., Allain, A., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J., 2018a, A graph-space approach to optimal transport for full waveform inversion, *in* SEG Technical Program Expanded Abstracts 2018, Society of Exploration Geophysicists, 1158–1162.

Métivier, L., Allain, A., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J., 2018b, Optimal transport for mitigating cycle skipping in full-waveform inversion: A graph-space transform approach: *Geophysics*, **83**, No. 5, R515–R540.

- Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: *Geophysical Journal International*, **167**, No. 2, 495–503.
- Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain; part 1, theory and verification in a physical scale model: *Geophysics*, **64**, No. 3, 888–901.
- Pratt, R. G., and Shipp, R. M., 1999, Seismic waveform inversion in the frequency domain, part 2: Fault delineation in sediments using crosshole data: *Geophysics*, **64**, No. 3, 902–914.
- Symes, W. W., 2008, Migration velocity analysis and waveform inversion: *Geophysical prospecting*, **56**, No. 6, 765–790.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, No. 8, 1259–1266.
- van Leeuwen, T., and Herrmann, F. J., 2015, A penalty method for pde-constrained optimization in inverse problems: *Inverse Problems*, **32**, No. 1, 015,007.
- Virieux, J., Asnaashari, A., Brossier, R., Métivier, L., Ribodetti, A., and Zhou, W., 2017, An introduction to full waveform inversion, *in* *Encyclopedia of exploration geophysics*, Society of Exploration Geophysicists, R1–1.
- Virieux, J., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, No. 6, WCC1–WCC26.
- Yang, Y., and Engquist, B., 2017, Analysis of optimal transport and related misfit functions in full-waveform inversion: *Geophysics*, **83**, No. 1, A7–A12.
- Yang, Y., Engquist, B., Sun, J., and Hamfeldt, B. F., 2018, Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion: *Geophysics*, **83**, No. 1, R43–R62.
- Yong, P., Liao, W., Huang, J., and Li, Z., 2018, Total variation regularization for seismic waveform inversion using an adaptive primal dual hybrid gradient method: *Inverse Problems*, **34**, No. 4, 045,006.
- Yong, P., Liao, W., Huang, J., Li, Z., and Lin, Y., 2019, Misfit function for full waveform inversion based on the Wasserstein metric with dynamic formulation: *Journal of Computational Physics*, **399**, 108,911.