Novel Optimization Schemes for Full Waveform Inversion: Optimal Transport and Inexact Gradient Projection

Da Li

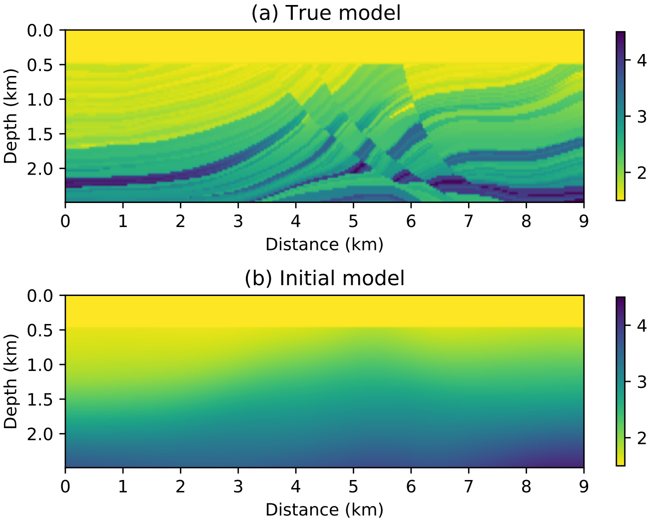
Department of Mathematics and Statistics
University of Calgary

February 18, 2021

Table of Contents

Table of Contents

Full waveform inversion problem



Formulation of full waveform inversion problem

Full waveform inversion (FWI) is a PDE constrained optimization problem:

$$\min_{(y,u)\in Y\times U_{\text{ad}}} J(y,u) = \frac{1}{2}\|Qy - y_d\|_Y^2 + \lambda_r \mathcal{R}(u), \quad \text{such that } e(y,u) = 0. \tag{1}$$

The constraint PDE $e(y,u) = 0$ is well-posed, the parameter-to-state map can be defined as $F(u) = y$.

The PDE constrained optimization problem (1) has a reduced form:

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

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

Improving the inverse result

The parameter-to-state map $F(u) = y$ is nonlinear, the objective function $f(u)$ is non-convex.

Only local minima can be expected.

We hope the local minimum is close to the global minimum, several methods can be considered:

- Enlarge the search space: wavefield reconstruction inversion (from U to $Y \times U$), etc.
- Change the distance in the objective function: optimal transport based distance (Wasserstein distance), etc.
- Add regularization term to the objective function: total variation regularization, adaptive regularization, etc.
- Solve the constraint optimization problem: box constraint set, etc.

Table of Contents

Optimal transport problem

Define discrete probability measure:

$$\alpha = \sum_{i=1}^{n} a_i \delta_{x_i}, \quad \beta = \sum_{i=1}^{m} b_i \delta_{y_i},$$

where $a_i > 0$, $b_i > 0$, $\sum_i a_i = \sum_i b_i = 1$.

Optimal transport distance

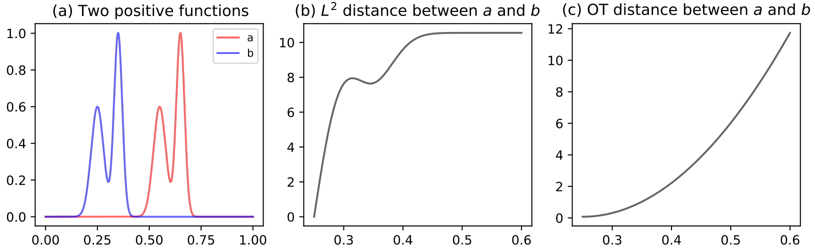
$$W_2^2(\alpha, \beta) = \min_{P \in \Pi_d(\alpha,\beta)} \langle P, C \rangle = \min_{P \in \Pi_d(\alpha,\beta)} \sum_{i,j} C_{i,j} P_{i,j},$$

where the cost matrix $C_{i,j} = |x_i - y_j|^2$, the set of discrete transport plan is

$$\Pi_d(\alpha, \beta) = \left\{ P \in \mathbb{R}_+^{n \times m} \;\middle|\; P\mathbf{1}_m = a \text{ and } P^{'}\mathbf{1}_n = b \right\}.$$

Initial idea and limitations

Convexity with respect to shift of 2-Wasserstein distance:



Limitations of OT distance:

- Mass equality condition ($\sum_i a_i = \sum_i b_i = 1$): unbalanced optimal transport distance, mixed L^1/Wasserstein distance;
- Positive vectors are needed ($a_i > 0$, $b_i > 0$): normalization methods.

Overcome the mass equality condition

Unbalanced optimal transport distance [Chizat et al., 2018]

When the mass equality condition is not satisfied ($\sum_i a_i \neq \sum_i b_i$):

$$W^2_{2,\varepsilon_u}(\alpha,\beta) = \min_{P \in \mathbb{R}_+^{n\times m}} \langle P, C\rangle + \varepsilon_u KL(P\mathbf{1}_m|a) + \varepsilon_u KL(P^{'}\mathbf{1}_n|b),$$

where the Kullback–Leibler (KL) divergence is defined by

$$KL(a|b) = \sum_{i=1}^{n} a_i \log\left(\frac{a_i}{b_i}\right).$$

Mixed L^1/Wasserstein distance

Normalized discrete measure $\hat{\alpha} = \frac{1}{\|a\|_1}\sum_{i=1}^{n} a_i \delta_{x_i}$ and $\hat{\beta} = \frac{1}{\|b\|_1}\sum_{i=1}^{m} b_i \delta_{y_i}$,

$$\bar{W}_2(\alpha,\beta) = W_2(\hat{\alpha},\hat{\beta}) + \left|\|a\|_1 - \|b\|_1\right|. \quad (3)$$

Overcome the mass equality condition

Proposition 1

The mixed L^1/Wasserstein distance defined in equation (3) is a distance over the set of positive discrete measures.

Since,

$$\bar{W}_2^2(\alpha, \beta) = \left(W_2(\hat{\alpha}, \hat{\beta}) + \left| \|a\|_1 - \|b\|_1 \right| \right)^2 \leq 2 \left(W_2^2(\hat{\alpha}, \hat{\beta}) + \left(\|a\|_1 - \|b\|_1 \right)^2 \right),$$

define,

$$J(\alpha; \beta) = W_2^2(\hat{\alpha}, \hat{\beta}) + \left(\|a\|_1 - \|b\|_1 \right)^2.$$

Proposition 2

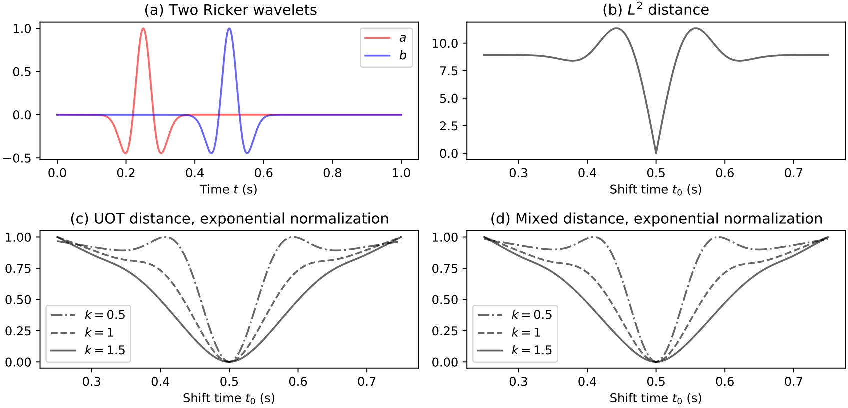
The objective function $J(\alpha; \beta)$ is convex with respect to the shift, dilation, and mass change of α.

Normalization methods

Normalization methods are needed for signals:

$$h_{\mathsf{l}}(a,k)(t) = a(t) + k, \quad h_{\mathsf{e}}(a,k)(t) = e^{ka(t)}.$$

Normalization methods only partially solve the problem.
Shift signal a from left to right:



OT based distance for the FWI problem

Marmousi model inverse results:

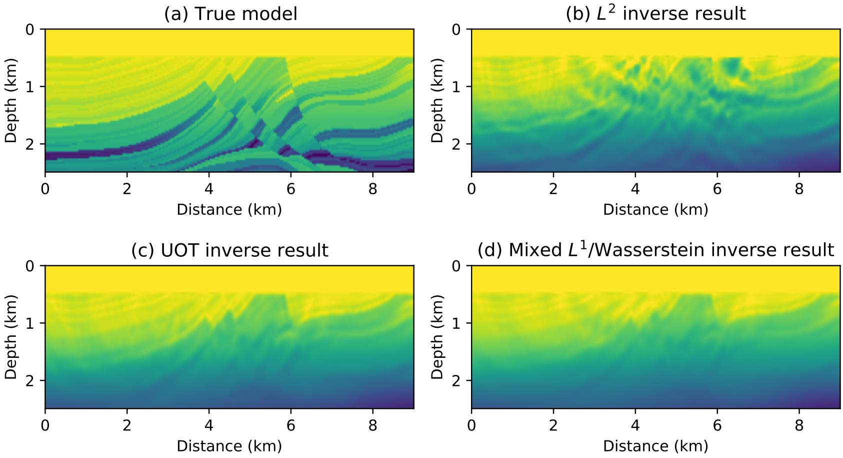


Table of Contents

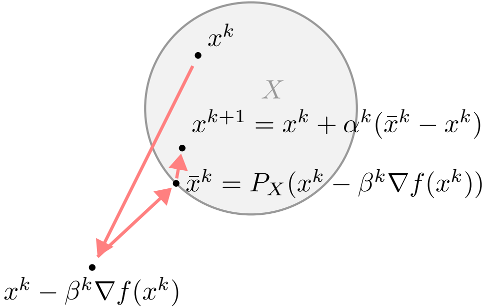
Initial idea

Objective function f is smooth and nonlinear, X is convex, consider the constrained optimization problem:

$$\min_x f(x), \quad \text{such that } x \in X.$$

Gradient projection method at the k-th iteration:

$$\begin{aligned} \bar{x}^k &= P_X(x^k - \beta^k \nabla f(x^k)), \\ x^{k+1} &= x^k + \alpha^k(\bar{x}^k - x^k). \end{aligned}$$



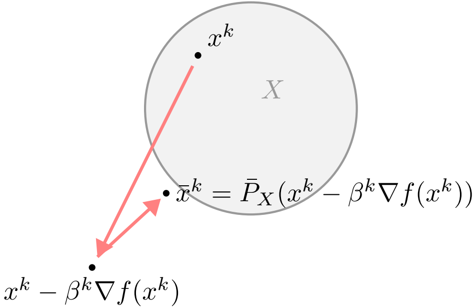
Initial idea

Objective function f is smooth and nonlinear, X is convex, consider the constrained optimization problem:

$$\min_x f(x), \quad \text{such that } x \in X.$$

Gradient projection method at the k-th iteration:

$$\begin{aligned} \bar{x}^k &= P_X(x^k - \beta^k \nabla f(x^k)), \\ x^{k+1} &= x^k + \alpha^k(\bar{x}^k - x^k). \end{aligned}$$



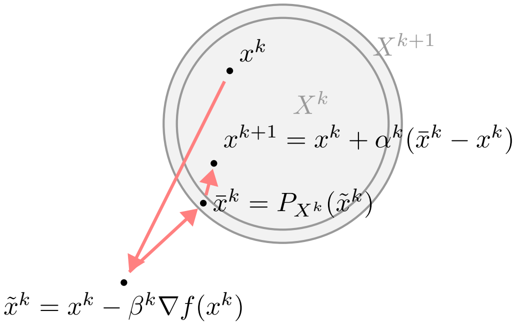
Set expanding strategy

Definition

Given a feasible set X is nonempty, closed, and convex, construct an expanding set sequence $\{X^k\}$ such that

$$X = \lim_{k \to \infty} X^k, \quad X^k \subset X^{k+1}, \quad X^k \neq X^{k+1}, \tag{4}$$

where X^k is nonempty, closed, and convex for each $k \in \mathbb{N}$.



Gradient projection method with inexact projection

Algorithm 1: Gradient projection method with inexact projection

Initialization: Given feasible set X, construct set sequence $\{X^k\}$ satisfies (4). Given the initial point $x^0 \in X^0$.

while *not convergent* **do**

Step 1: Compute $\tilde{x}^k = x^k - \beta\nabla f(x^k)$;

Step 2: Project $\tilde{x}^k$ to X^k, until equation $\bar{P}_{X^k}(\tilde{x}^k) \in X^{k+1}$, $\left\langle \tilde{x}^k - \bar{P}_{X^k}(\tilde{x}^k), x^k - \bar{P}_{X^k}(\tilde{x}^k) \right\rangle \leq 0$ are satisfied.

Step 3: Evaluate the line search stepsize with Armijo rule or Wolfe conditions, update with $x^{k+1} = x^k + \alpha^k(\bar{x}^k - x^k)$;

Step 4: Enlarge the feasible set $X^k = X^{k+1}$, let $k = k + 1$.

end

Theorem 1

Under proper assumptions, let $\{x^k\}$ be the sequence generated by Algorithm 1. Then every limit point of $\{x^k\}$ is stationary.

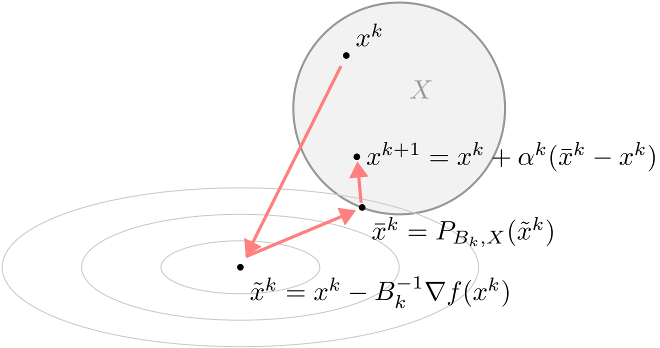
Scaled gradient projection method

The scaled gradient projection method is given by:

$$\bar{x}^k = \arg\min_{x \in X} \left\langle \nabla f(x^k), x - x^k \right\rangle + \frac{1}{2} \left\langle B_k(x - x^k), x - x^k \right\rangle,$$
$$x^{k+1} = x^k + \alpha^k(\bar{x}^k - x^k).$$

The first equation equivalents to

$$\tilde{x}^k = x^k - B_k^{-1} \nabla f(x^k), \quad \bar{x}^k = P_{B_k, X}(\tilde{x}^k).$$



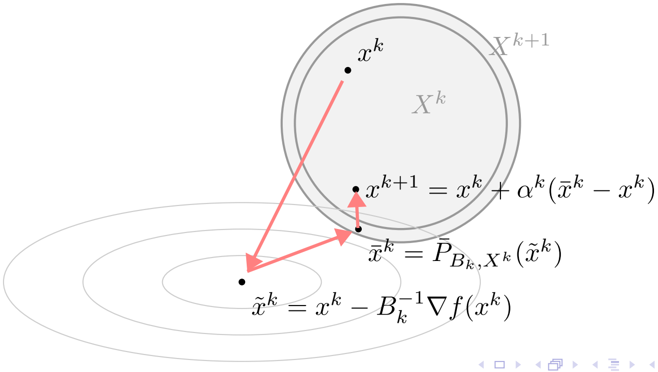
Scaled gradient projection method

The scaled gradient projection method is given by:

$$\bar{x}^k = \arg\min_{x \in X} \left\langle \nabla f(x^k), x - x^k \right\rangle + \frac{1}{2} \left\langle B_k(x - x^k), x - x^k \right\rangle,$$
$$x^{k+1} = x^k + \alpha^k(\bar{x}^k - x^k).$$

The first equation equivalents to

$$\tilde{x}^k = x^k - B_k^{-1} \nabla f(x^k), \quad \bar{x}^k = P_{B_k, X}(\tilde{x}^k).$$



Scaled gradient projection method with inexact projection

Algorithm 2: Scaled gradient projection method with inexact projection

Initialization: Given feasible set X, construct set sequence $\{X^k\}$ satisfies (4). Given the initial point $x^0 \in X^0$.

while *not convergent* **do**

Step 1: Find a scaling matrix B_k, compute $\tilde{x}^k = x^k - B_k^{-1}\nabla f(x^k)$;

Step 2: Project $\tilde{x}^k$ towards X^k, until equation $\bar{P}_{B_k,X^k}(\tilde{x}^k) \in X^{k+1}$ and $\left\langle \tilde{x}^k - \bar{P}_{B_k,X^k}(\tilde{x}^k), x^k - \bar{P}_{B_k,X^k}(\tilde{x}^k) \right\rangle_{B_k} \leq 0$ are satisfied.

Step 3: Evaluate the line search stepsize with Armijo rule or Wolfe conditions, update with $x^{k+1} = x^k + \alpha^k(\bar{x}^k - x^k)$;

Step 4: Enlarge $X^k = X^{k+1}$, let $k = k+1$.

end

Theorem 2

Under certain assumptions, let $\{x^k\}$ be the sequence generated by Algorithm 2. Then every limit point of $\{x^k\}$ is stationary.

Table of Contents

Initial idea

Consider the constrained optimization problem with multiple constraints:

$$\min_u f(u), \quad \text{such that } u \in U_{\text{ad}} = \cap_i U_i,$$

where U_i is nonempty, closed, and convex, U_{ad} is nonempty.

A priori information can be described by constraint sets.

Two kinds of constraint sets:

- Constraint set with closed-form projection function: box constraint, hyperplane constraint, etc.
- Constraint set with subgradient projection: total variation constraint, l_1 constraint.

Proposed algorithm is a combination of scaled gradient projection method with inexact projection, L-BFGS Hessian approximation, an iterative projection method [Combettes, 2003].

Main algorithm

Algorithm 3: Scaled gradient projection method with multiple constraints

Given: the objective function f and initial value u^0; a family of nonempty, closed, convex constraint sets U_i.

For each U_i construct the set sequence $\{U_i^j\}_{j\in\mathbb{N}}$; set $U_{\text{ad}}^k = \cap_{i\in I} U_i^k$ for all k.

while *Not converge* **do**

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

Step 2: Update the L-BFGS coefficients s_k, y_k, S_k, Y_k, R_k, D_k, U_k.

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

Step 4: Compute $\bar{u}^k = \bar{P}_{B_k, U_{\text{ad}}^k}(\tilde{u}^k)$, i.e., project $\tilde{u}^k$ to U_{ad}^k in $\mathcal{H}_{B_k}$ with the inexact projection algorithm [Combettes, 2003], until the following equations are satisfied:

$$\bar{u}^k \in U_{\text{ad}}^{k+1},$$

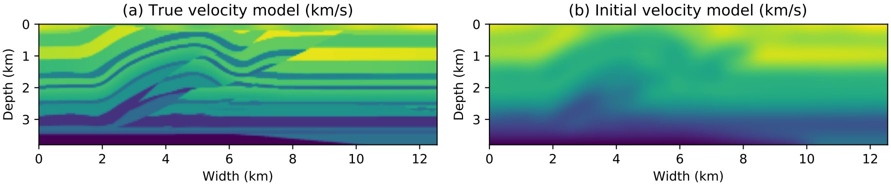
$$\left\langle \tilde{u}^k - \bar{u}^k, u^k - \bar{u}^k \right\rangle_{B_k} \leq 0.$$

Step 5: Update $u^{k+1} = u^k + \alpha^k(\bar{u}^k - u^k)$, here α^k is the linesearch parameter.

Step 6: Set $k = k+1$.

end

Numerical example



Settings of the sequences of constraint sets:

- $U_1^h = \{u \in \mathbb{R}^n \mid 2.5588 - \theta_1(h) \leq u_i \leq 6 + \theta_1(h),\ i = 1, \cdots, n\}$, where $\theta_1(h) = \sum_{i=1}^{h} 0.02 \times 0.9^i$.
- $U_2^h = \{u \in \mathbb{R}^n \mid f_{\mathsf{TV}}(u) \leq 800 + \theta_2(h)\}$, where $\theta_2(h) = \sum_{i=1}^{h} 40 \times 0.9^i$.
- $U_3^h = \{u \in \mathbb{R}^n \mid f_{\mathsf{TV}}(u) \leq 1000 + \theta_3(h)\}$, where $\theta_3(h) = \sum_{i=1}^{h} 50 \times 0.9^i$.
- $U_4^h = \{u \in \mathbb{R}^n \mid f_{\mathsf{TV}}(u) \leq 1200 + \theta_4(h)\}$, where $\theta_4(h) = \sum_{i=1}^{h} 60 \times 0.9^i$.

Numerical example

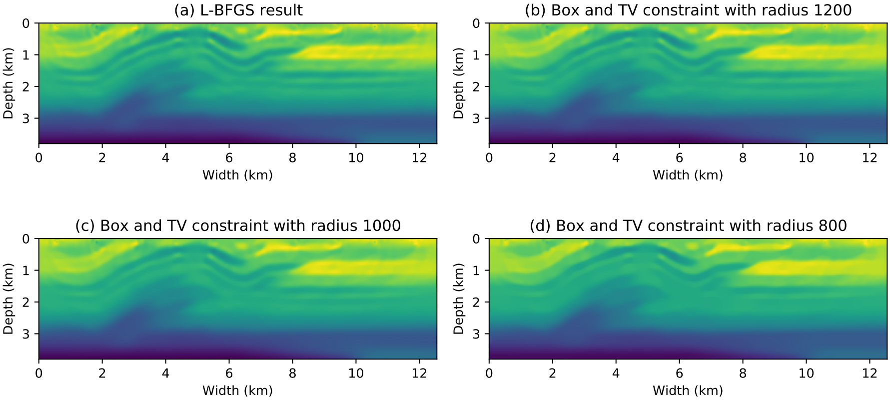


Table of Contents

Contributions

This thesis studies the novel optimization methods for the FWI problem:

- Introduce the UOT distance to the FWI problem.
- Construct a mixed L^1/Wasserstein distance and introduced to the FWI problem.
- A set expanding strategy is developed for the gradient projection methods when the inexact projection algorithm is used.
- The convergence results for the gradient projection methods with inexact projection are proved under proper assumptions.
- A new optimization scheme: scaled gradient projection method with multiple constraints. This optimization scheme is applied to the FWI problem.
- All computing programs were written by the author in Julia and are available on Github.

Future studies

- The application of UOT and mixed L^1/Wasserstein distance to the variational problem based on the positive quantities.
- A new optimization scheme can be designed with the combination of the spectral projected gradient method and the set expanding strategy.
- Incorporate the well-log data to the FWI problem with the proposed optimization scheme in Chapter 6.
- Find new constraints for the multi-parameter FWI problem to decrease the cross-talk issue.
- The set expanding strategy developed in Chapter 5 provides a method to analyze the following constrained optimization problem:

$$\min_x f(x), \quad \text{such that } x \in X^k.$$

This might be a new way to analysis the optimization problem with adaptive constraint (regularization).

Acknowledgment

Thank You!

References

Chizat, L., Peyré, G., Schmitzer, B., and Vialard, F.-X. (2018).
Scaling algorithms for unbalanced optimal transport problems.
Mathematics of Computation, 87(314):2563–2609.

Combettes, P. L. (2003).
A block-iterative surrogate constraint splitting method for quadratic signal recovery.
IEEE Transactions on Signal Processing, 51(7):1771–1782.