Full Waveform Inversion and Penalty Methods

Da Li
Supervisor: Dr. Michael Lamoureux
Dr. Wenyuan Liao

February 15, 2019

Outline

Full Waveform Inversion

A Reduced Penalty Method (by Tristan van Leeuwen and Felix J Herrmann)

Current and Future Works

Full Waveform Inversion Problem

Consider the discrete full waveform inversion (FWI) problem in frequency domain:

$$\begin{aligned} \min_{M\times U} J(m,u) &= \frac{1}{2}\|Pu-d\|_2^2+\mu r(m), \\ \text{s.t. } A(m)u &= (\omega^2 Im+\tilde{L})u = q,\ 0<a\le \min(m)\le \max(m)\le b. \end{aligned} \tag{1}$$

Here,

- $u \in \mathbb{C}^{N_u} = U$ is the wavefield in the earth. $m \in \mathbb{R}^{N_m} = M$ is the physical parameter which value is between a and b. $d \in \mathbb{R}^{N_d}$ is the received data. $q \in \mathbb{C}^{N_q}$ is the source term which is known in the problem.
- P is the projection operator, measures the wavefield at the location of receiver. And $A(m)$ is a numerical Helmholtz operator, here I is the identity matrix and $\tilde{L}$ is the discrete Laplace operator with proper boundary condition or perfectly matched layer.
- $r(\cdot)$ is a regularization term which we do not consider in this talk.

Full Waveform Inversion Problem

The wavefield u can be solved through the Helmholtz equation, we can have the corresponding reduced problem,

$$\begin{aligned} &\min_{M} \hat{J}(m) = J(m, u(m)) = \frac{1}{2}\|PA^{(-1)}(m)q - d\|_2^2 + \mu r(A^{(-1)}(m)q). \\ &\text{s.t. } 0 < a \le \min(m) \le \max(m) \le b. \end{aligned} \tag{2}$$

Since we only have the equality restriction in the full problem (1), we can have the Lagrangian,

$$L(m, u, v) = \frac{1}{2}\|Pu - d\|_2^2 + v^T(A(m)u - q). \tag{3}$$

Here v is the Lagrange multiplier, which is also the solution of adjoint equation in FWI problem.

Full Waveform Inversion Problem

First Order Necessary Optimality Condition

A solution $(\bar{m}, \bar{u}, \bar{v})$ is a stationary point of equation (3) as,

$$\nabla L(\bar{m}, \bar{u}, \bar{v}) = \begin{bmatrix} L_m \\ L_u \\ L_v \end{bmatrix} = \begin{bmatrix} \left(\frac{\partial A(\bar{m})\bar{u}}{\partial m}\right)^T \bar{v} \\ A(\bar{m})^T \bar{v} + P^T(P\bar{u} - d) \\ A(\bar{m})\bar{u} - q \end{bmatrix} = 0. \tag{4}$$

Remarks:

1. The Lagrange multiplier exists (proved by the KKT theory).
2. The discrete frequency domain FWI problem has an optimal solution $(\bar{m}, \bar{u}, \bar{v})$ which satisfies the first order necessary optimality condition.
3. Reduced problem $\hat{J}(m) = L(m, u(m), v)$, then we have the adjoint method.

A Reduced Penalty Problem

Consider a reduced penalty problem [van Leeuwen and Herrmann, 2015],

$$\min_{M\times U} \ J(m,u) = \frac{1}{2}\|Pu - d\|_2^2 + \frac{\lambda}{2}\|A(m)u - q\|_2^2. \tag{5}$$

Remarks:

- As $\lambda \to \infty$, the penalty problem (5) will coincide with the constraint problem (1).
- Comparing to the reduced problem (2), problem (5) enlarges the feasible space from subspace of M to subspace of $M \times U$.
- The problem (5) is similar with the Lagrange function of problem (1). In that case, we can explain the term $v = \lambda(A(m)u - q)$ as the Lagrange multiplier.

Gradient of the objective function:

$$\nabla J = \begin{bmatrix} J_m \\ J_u \end{bmatrix} = \begin{bmatrix} \lambda\left(\frac{\partial A(m)u}{\partial m}\right)^T (A(m)u - q) \\ P^T(Pu - d) + \lambda A(m)^T(A(m)u - q) \end{bmatrix}.$$

A Reduced Penalty Problem

We can solve the reduced penalty problem in an iterative way,

1. Let $J_u = 0$, solve u_λ with an augmented linear system,

$$\begin{bmatrix} \lambda A(m)^T A(m) \\ P^T P \end{bmatrix} u_\lambda = \begin{bmatrix} \lambda A(m)^T q \\ P^T d \end{bmatrix}.$$

2. Fix u_λ, optimize m with J_m,

$$J_m = \lambda \left(\frac{\partial A(m) u_\lambda}{\partial m} \right)^T (A(m)u_\lambda - q) = \left(\frac{\partial A(m) u_\lambda}{\partial m} \right)^T v,$$

   where v is the Lagrange multiplier.

This method is called waveform reconstruction inversion (WRI) in some literature.

Theory of Reduced Penalty Problem

Assume (m^k, u_r^k, v_r^k) is a minimizing sequence of reduced problem (2), and $(m^k, u_\lambda^k, v_\lambda^k)$ is a minimizing sequence of the penalty problem (5).

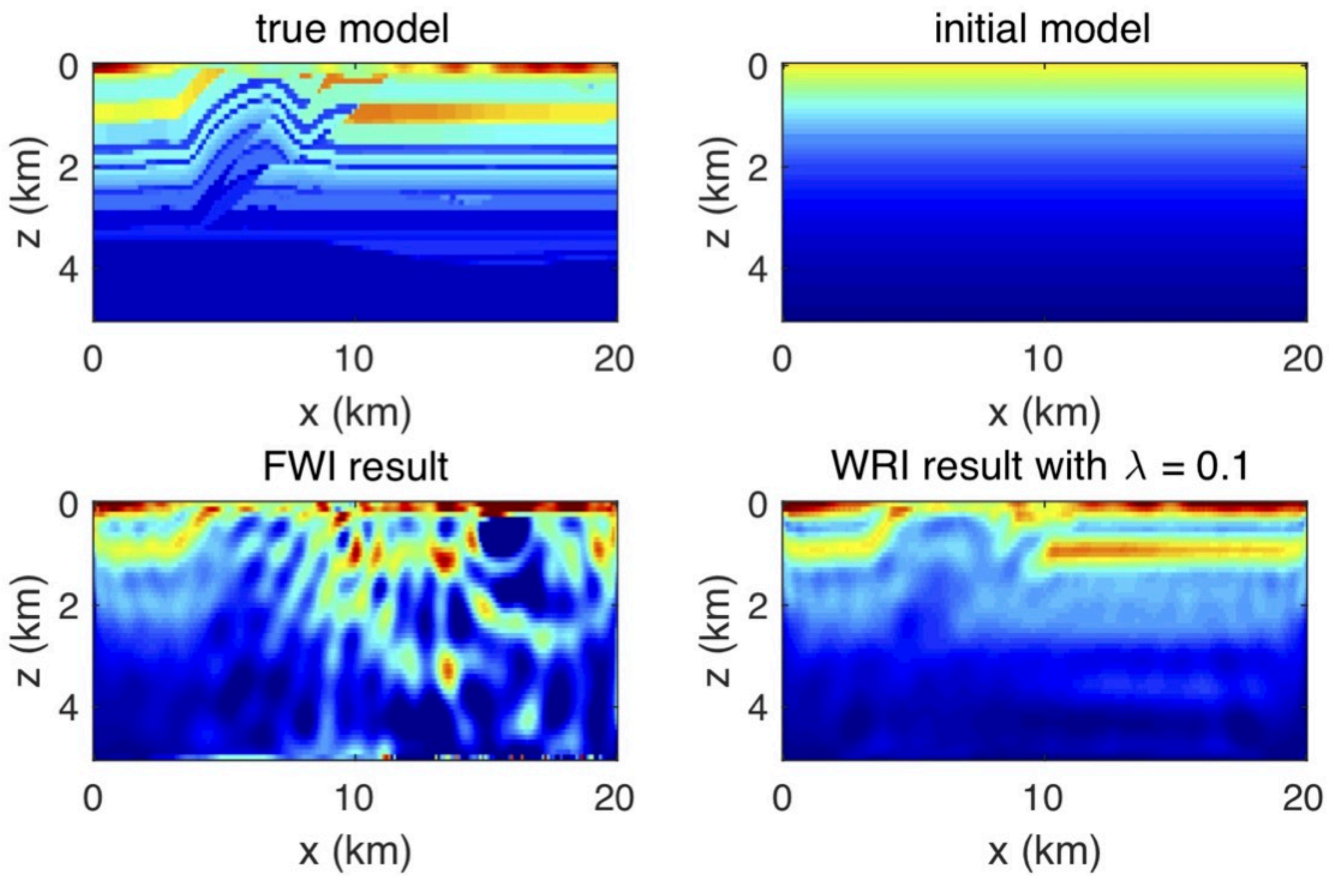
Reduced Problem

Recall in the reduced problem, at each iteration we assume $\|L_u(m^k, u_r^k, v_r^k)\|_2 = 0$ and $\|L_v(m^k, u_r^k, v_r^k)\|_2 = 0$. The gradient $\hat{J}_m(m^k) = L_m(m^k, u_r^k, v_r^k)$.

Theorem 1 ([van Leeuwen and Herrmann, 2015])

At each iteration of the reduced penalty method (5), the iterates satisfies $\|L_u(m^k, u_\lambda^k, v_\lambda^k)\|_2 = 0$ and $\|L_v(m^k, u_\lambda^k, v_\lambda^k)\|_2 = O(\lambda^{(-1)})$. Moreover if the reduced penalty method (5) terminates successfully at $\bar{m}$ for which $\|L_m(\bar{m}, \bar{u}_\lambda, \bar{v}_\lambda)\|_2 \leq \epsilon$, then we have $\|\nabla L(\bar{m}, \bar{u}_\lambda, \bar{v}_\lambda)\|_2 \leq \epsilon + O(\lambda^{(-1)})$.

Numerical Result



Grid size: 401 by 101
99 sources and 100 receivers on the first line.

Current Work: Partial Penalty Method

Suppose the state space can be divided two parts $U = U_1 \oplus U_2$, we minimize the cost function in a smaller space $M \times U_2$.
One spacial case is U_1 is the data space. In this case, we fix $\hat{u} = (d, u_2)$. And from the projection operator $P : U \to U_1$, we have $d = P\hat{u}$.
Replace u with $\hat{u}$ in the penalty problem:

$$\min_{M \times U} J(m, \hat{u}) = \frac{1}{2}\|P\hat{u} - d\|_2^2 + \frac{\lambda}{2}\|A(m)\hat{u} - q\|_2^2.$$

Then we have a new problem,

$$\begin{aligned} \min_{M \times U_2} J(m, u_2) &= \frac{\lambda}{2}\|A(m)\hat{u} - q\|_2^2 = \frac{\lambda}{2}\|A(m)(d, u_2) - q\|_2^2 \\ &= \frac{\lambda}{2}\|A(m)R^T R\hat{u} - q\|_2^2 = \frac{\lambda}{2}\|(A_1 \ A_2)\begin{pmatrix} d \\ u_2 \end{pmatrix} - q\|_2^2. \end{aligned} \tag{6}$$

Here R is a pivoting matrix, $(A_1 \ A_2) = A(m)R^T$ and $\begin{pmatrix} d \\ u_2 \end{pmatrix} = R\hat{u}$.

Current Work: Partial Penalty Method

Gradient of the cost function in partial penalty problem (6),

$$\nabla J = \begin{bmatrix} J_m \\ J_{u_2} \end{bmatrix} = \begin{bmatrix} \lambda \left(\frac{\partial A(m)\hat{u}}{\partial m} \right)^T (A(m)\hat{u} - q) \\ A_1 d + A_2 d - q \end{bmatrix}.$$

To solve the problem (6), we can use an iterative way,

1. Let $J_{u_2} = 0$, we can solve u_2 with

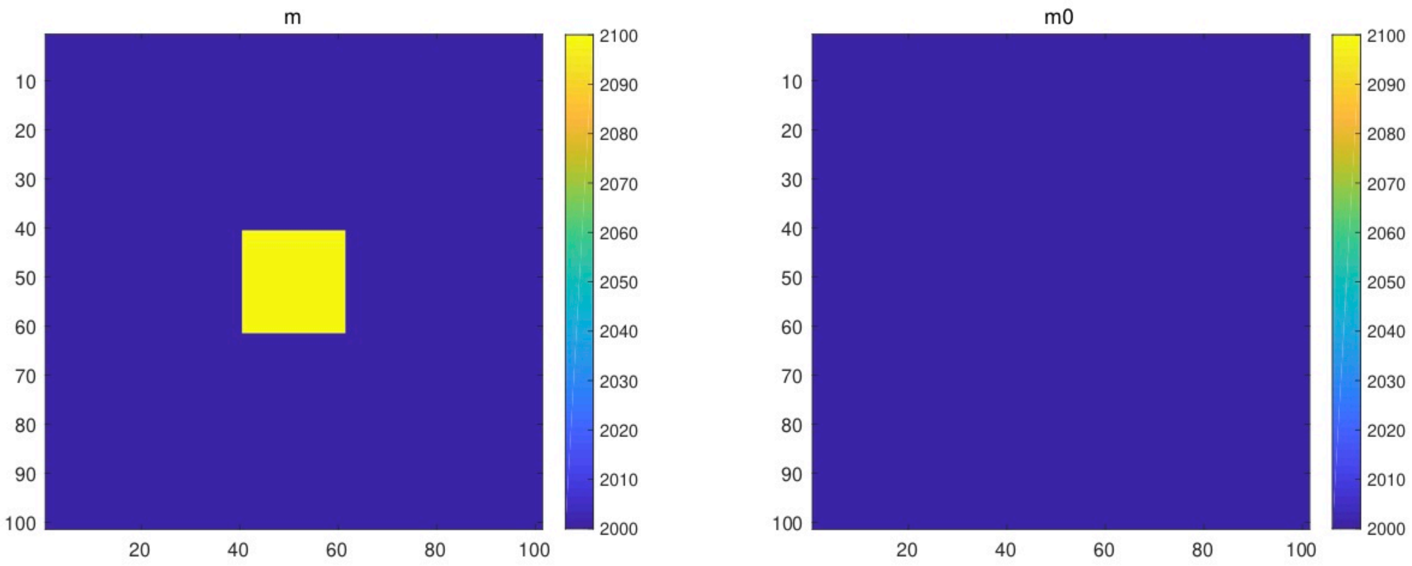
$$A_2 u_2 = q - A_1 d,$$

   then $\hat{u}_\lambda = R^T \begin{pmatrix} d \\ u_2 \end{pmatrix}$.

2. Fix $\hat{u}_\lambda$, optimize m with J_m,

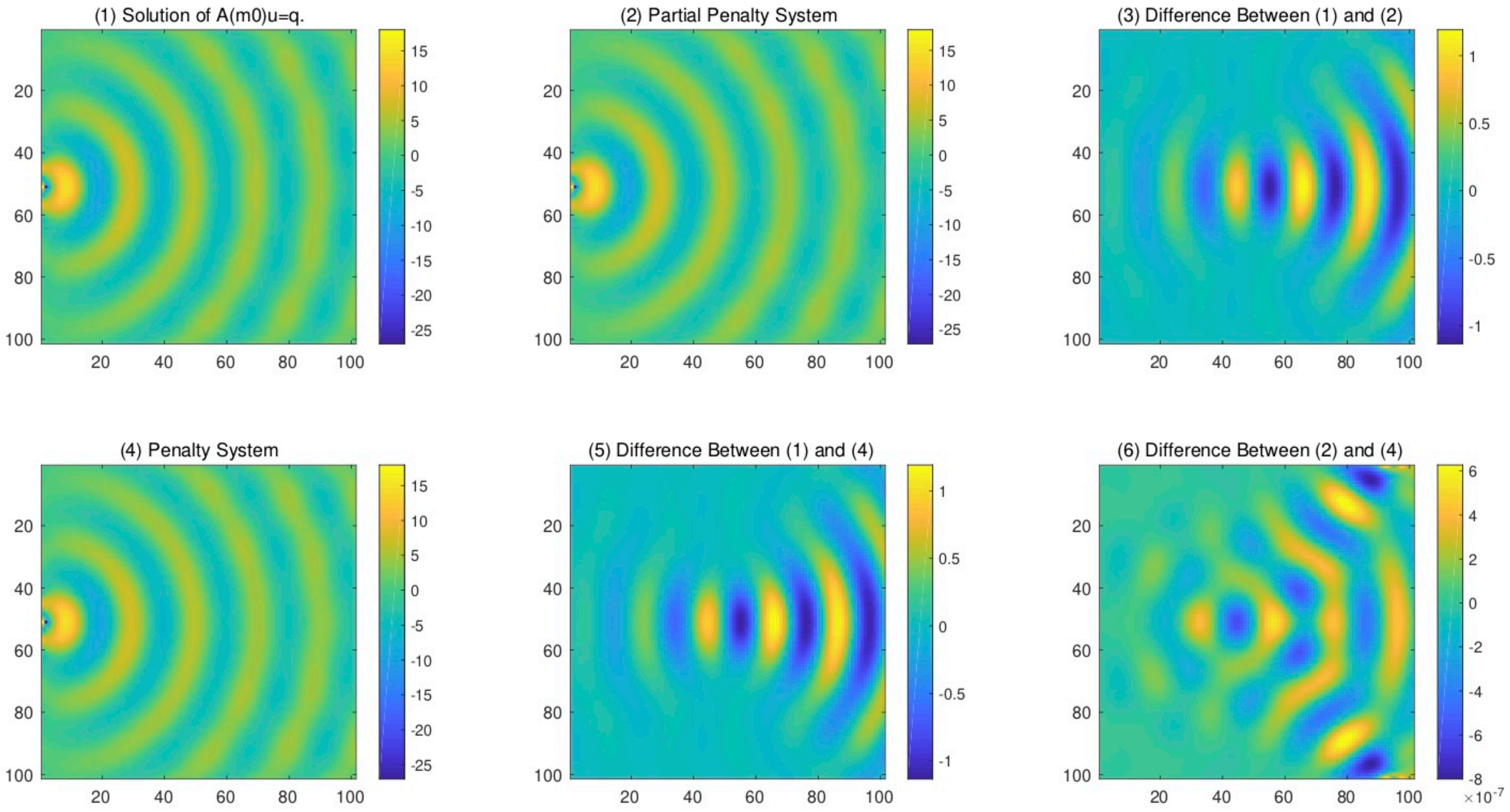
$$J_m = \lambda \left(\frac{\partial A(m)\hat{u}_\lambda}{\partial m} \right)^T (A(m)\hat{u}_\lambda - q).$$

Numerical Example

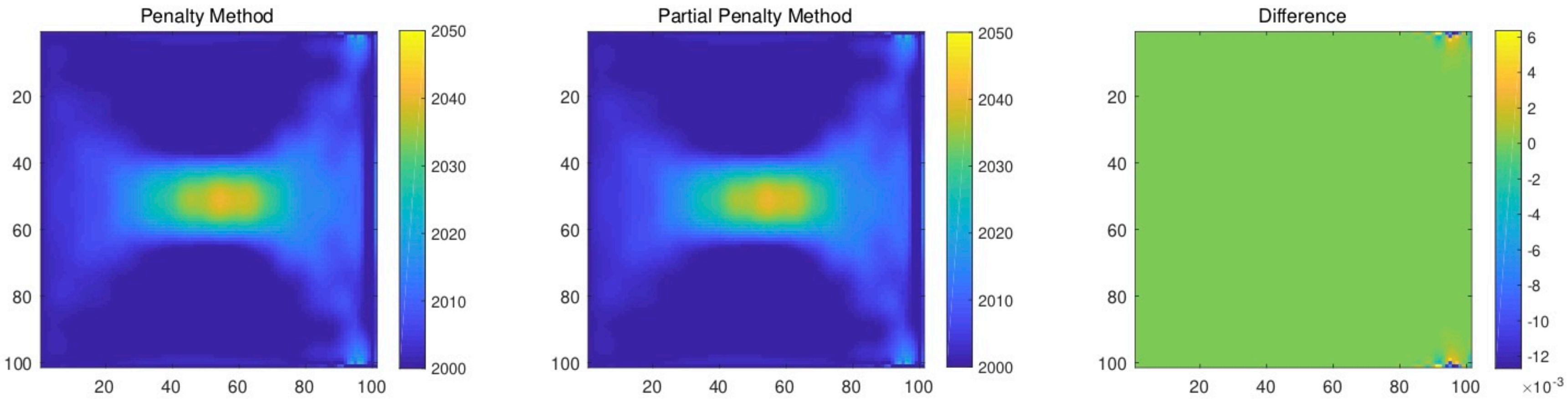


Grid size: 101 by 101
101 sources on the second column and 101 receivers on the 100th column.

Numerical Example



Numerical Example



Iterate 10 times.
CPU time for penalty method: 8.505432 s.
CPU time for partial penalty method: 7.5864 s.

Future Work

Future work:

- Add the penalty parameter λ. Consider $\tilde{u} = (\lambda u_1, u_2)$.
- Enlarge the space U_1, to increase the numerical efficiency and keep the stability at the same time.
- Analytical results between partial penalty method and reduced method.
- Consider the regularization term.

Thank You!

Reference

van Leeuwen, T. and Herrmann, F. J. (2015).
A penalty method for pde-constrained optimization in inverse problems.
Inverse Problems, 32(1):015007.