Source estimation and uncertainty quantification for wave-equation based seismic imaging and inversion

Zhilong Fang

PhD Defense - April 27, 2018
Seismic exploration

Main questions:

1. Where is the oil and gas reservoir?

2. What is the volume of the oil and gas reservoir?

Seismic exploration

Seismic exploration survey

Seismic data volume
Seismic exploration

Objective

Seismic exploration survey

Seismic data volume
Seismic exploration

Forward modeling:
\[ d = F(m) = PA(m)^{-1}q \]

A : discretized wave equation stencil matrix
P : receiver restriction operator
Seismic exploration

Inversion or imaging:

\[ \mathbf{m} = F^{-1}(\mathbf{d}) \]
Seismic exploration — unknown source

Inversion or imaging:

\[ m = F^{-1}(d, q) \]

if source is wrong?
Seismic exploration — uncertainty quantification

Uncertainty in data

Uncertainty in model

Confidence of the model
Purposes of thesis

Outline:

• **Chapter 2**: Source estimation for time-domain sparsity promoting least-squares reverse-time migration (LS-RTM);

• **Chapter 3-5**: Source estimation for wavefield reconstruction inversion (WRI);

• **Chapter 6**: Uncertainty quantification for inverse problems with weak wave-equation constraints.
Chapter 2

Source estimation for time-domain least-squares reverse-time migration
Time domain LS-RTM

Model separation

True model $m_t$

Smooth background model $m_s$

Model perturbation $\delta m$

Data separation

$F(m_t)$

$F(m_s)$

$\delta d$
Time domain LS-RTM

Taylor expansion:

\[ F(m_t) = F(m_s) + \nabla F(m_s) \delta m + O(\delta m^2) \]

Linearized approximation or Born modeling:

\[ F(m_t) \approx F(m_s) + \nabla F(m_s) \delta m \quad \delta d \approx \nabla F(m_s) \delta m \]

Forward operator:

\[ F(m) = PA(m)^{-1} q, \]

it is related to the acoustic wave equation:

\[ (m \frac{\partial^2}{\partial t^2} + \Delta) u = q. \]

Born modeling operator:

\[ \nabla F(m) = -PA(m)^{-1} (\nabla A(m) u), \]

where \( \nabla A(m) u \) is the Jacobian matrix.

[12]

[N. Aoki and G. T. Schuster, 2009] [X. Li and F. Herrmann, 2012], [N. Tu and F. Herrmann, 2015]
Time domain LS-RTM

Obtain the model perturbation $\delta m$ from the data perturbation $\delta d$:

$$\delta m \approx \nabla F(m_s)^{-1} \delta d.$$ 

In practice, we have to solve the following optimization problem:

$$\min_{\delta m} \frac{1}{2n_{\text{src}}} \sum_{i=1}^{n_{\text{src}}} \| \nabla F_i(m_s) \delta m - \delta d_i \|^2 \quad (\text{LS-RTM})$$

1. Very large overdetermined system
2. Computationally expensive
Time domain LS-RTM

Dimensional reduction + sparsity promoting optimization:

\[
\min_{\delta m} \frac{1}{2n_{\text{src}}} \sum_{i=1}^{n_{\text{src}}} \| \nabla F_i(m_s) \delta m - \delta d_i \|^2
\]

To

\[
\min_x |x|_1
\]

s.t.

\[
\frac{1}{2|\mathcal{I}|} \sum_{i \in \mathcal{I}} \| \nabla F_i(m_s) C^\top x - \delta d_i \|^2 \leq \sigma
\]

where \( \mathcal{I} \subset [1, 2, \cdots, n_{\text{src}}] \), and \(|\mathcal{I}| << n_{\text{src}}\).

\[\nabla F(m_s) \delta m = \delta d \]
Time domain LS-RTM

Linearized Bregman method:

\[
\begin{align*}
\text{from} & \quad \min_x \|x\|_1 \\
\text{to} & \quad \min_x \lambda_1 \|x\|_1 + \frac{1}{2} \|x\|_2^2 \\
\text{s.t.} & \quad \frac{1}{2|\mathcal{I}|} \sum_{i \in \mathcal{I}} \|\nabla F_i(m_s) C^T x - \delta d_i\|^2 \leq \sigma
\end{align*}
\]

Algorithm:

1. Initialize \(x_0 = 0, z_0 = 0, q, \lambda_1, \) batchsize \(n'_{\text{src}} \ll n_{\text{src}}\)
2. for \(k = 0, 1, \ldots\)
3. Randomly choose shot subsets \(\mathcal{I} \in [1, \ldots, n_{\text{src}}], |\mathcal{I}| = n'_{\text{src}}\)
4. \(A_k = \{\nabla F_i(m_s) C^T\}_{i \in \mathcal{I}}\)
5. \(b_k = \{\delta d_i\}_{i \in \mathcal{I}}\)
6. \(z_{k+1} = z_k - t_k A_k^T P(\lambda_1) (A_k x_k - b_k)\)
7. \(x_{k+1} = S_{\lambda_1}(z_{k+1})\)
8. end

note: \(S_{\lambda_1}(z_{k+1}) = \text{sign}(z_{k+1}) \max\{0, |z_{k+1}| - \lambda_1\}\)
\(P(\lambda_1) (A_k x_k - b_k) = \max\{0, 1 - \frac{\sigma}{\|A_k x_k - b_k\|}\} \cdot (A_k x_k - b_k)\)
Time domain LS-RTM

Unknown source function:

\[
\begin{align*}
\min_{x,q} & \quad \lambda_1 \|x\|_1 + \frac{1}{2} \|x\|_2^2 \\
\text{s.t.} & \quad \frac{1}{2|\mathcal{I}|} \sum_{i \in \mathcal{I}} \|\nabla F_i(m_s, q)C^T x - \delta d_i\|_2^2 \leq \sigma
\end{align*}
\]

Subproblem:

\[
\begin{align*}
\min_{q} & \quad \frac{1}{2|\mathcal{I}|} \sum_{i \in \mathcal{I}} \|\nabla F_i(m_s, q)C^T x - \delta d_i\|_2^2 \\
\text{(Linear data fitting problem)}
\end{align*}
\]

Cons:
1. Expensive;
2. Unstable.
Time domain LS-RTM

Solutions: deconvolution + regularization
Let: \( q = w \ast q_0 \), \( q_0 \) is well defined.

**Property 1:**
\[
\nabla F_i(m_s, w \ast q_0)C^T x \approx w \ast \nabla F_i(m_s, q_0)C^T x;
\]

**Property 2:**
Source function should decay smoothly to zero with few oscillations within a short duration of time.

**Property 3:**
The energy of the source function can not explode.
Time domain LS-RTM

New subproblem:

$$\min_w \frac{1}{2|\mathcal{I}|} \sum_{i \in \mathcal{I}} \| w \ast \nabla F_i(m, q_0) - \delta d_i \|^2$$

$$+ \| \text{diag}(r)(w \ast q_0) \|^2 + \lambda_2 \| w \ast q_0 \|^2.$$

Penalty function:

$$r(t) = \log(1 + e^{\alpha(t-t_0)})$$

$t_0$: the approximate duration of the source signature

$\alpha$: controls the steepness of weight curve
Workflow for sparsity LS-RTM via LB with source estimation

1. Initialize $x_0 = 0$, $z_0 = 0$, $q_0$, $\lambda_1$, $\lambda_2$, batchsize $n'_\text{src} \ll n_{\text{src}}$, weights $r$
2. for $k = 0, 1, \cdots$
3. Randomly choose shot subsets $\mathcal{I} \in [1, \cdots, n_{\text{src}}], |\mathcal{I}| = n'_\text{src}$
4. $A_k = \{\nabla F_i(m_s, q_0)C^\top\}_{i \in \mathcal{I}}$
5. $b_k = \{\delta d_i\}_{i \in \mathcal{I}}$
6. $\tilde{d}_k = A_k x_k$
7. $w_k = \arg\min_w ||w * \tilde{d}_k - b_k||^2 + ||\text{diag}(r)(w * q_0)||^2 + \lambda_2||w * q_0||^2$
8. $z_{k+1} = z_k - t_k A_k^\top \left( w_k P_{\sigma}(w_k * \tilde{d}_k - b_k) \right)$
9. $x_{k+1} = S_\lambda(z_{k+1})$
10. end
Experiments

Data:
- 295 shots with shot interval 15m
- 295 receivers with receiver interval 15m
- 4s record, 15Hz peak frequency designed wavelet
- synthetic linearized data

Experiments:
- one pass through the data with batch sizes 2.5% data
- randomized subset of shots
- true source wavelet & initial guessed wavelet
Models and source functions

(a) Background model

(b) True model perturbation

(c) Source function

signal in time domain

frequency spectrum

Initial source
True source

Initial source
True source

Energy

0 1 2 3 4

0 10 20 30 40 50 60

0 0.1 0.2 0.3 0.4 0.5

0 1 2 3 4

0 10 20 30 40 50 60

0 0.1 0.2 0.3 0.4
Results

(a) Result w/ correct source

(b) Result w/ initial source

(c) Result w/ estimated source

(d) Vertical trace comparisons
Results

(a) Source comparisons

signal in time domain

frequency spectrum

(b) Residual decays

(c) Relative model error decays
Chapters 3, 4, 5
Source estimation for wavefield reconstruction inversion
Full-waveform inversion (FWI)

Original problem:

\[
\min_{u,m} \frac{1}{2} \sum_{k,l} \|P_k u_{k,l} - d_{k,l}\|^2_2
\]

subject to \(A_{k,l}(m)u_{k,l} = q_{k,l}\),

where,

- \(u_{k,l}\) – Wavefield of the \(k\)th shot at \(l\)th frequency
- \(d_{k,l}\) – Observed data of the \(k\)th shot at \(l\)th frequency
- \(q_{k,l}\) – Source of the \(k\)th shot at \(l\)th frequency
- \(A_{k,l}\) – Helmholtz of the \(k\)th shot at \(l\)th frequency
- \(P_k\) – Receiver projection operator of the \(k\)th shot
- \(m\) – Squared-slowness

Reduced/adjoint-state method:

\[
\min_{u,m} \frac{1}{2} \sum_{k,l} \|P_k A_{k,l}(m)^{-1} q_{k,l} - d_{k,l}\|^2_2
\]

with the gradient given by

\[
g = \sum_{k,l} u^\top_{k,l} \frac{\partial A^\top_{k,l}}{\partial m} v_{k,l}
\]

\[
u_{k,l} = A_{k,l}(m)^{-1} q_{k,l}
\]

\[
v_{k,l} = A_{k,l}^{-\top}(m) P_k r_{k,l}
\]

\[
r_{k,l} = P_k A_{k,l}(m)^{-1} q_{k,l} - d_{k,l}
\]

Wavefield reconstruction inversion (WRI):

\[
\min_{u,m} \frac{1}{2} \sum_{k,l} (\|P_k u_{k,l} - d_{k,l}\|^2_2 + \lambda^2 \|A_{k,l}(m)u_{k,l} - q_{k,l}\|^2_2)
\]

Eliminating \(u\) w/ variable projection:

\[
\bar{u} = \arg \min_u \frac{1}{2} \sum_{k,l} (\|P_k u_{k,l} - d_{k,l}\|^2_2 + \lambda^2 \|A_{k,l}(m)u_{k,l} - q_{k,l}\|^2_2)
\]

Analytical solution:

\[
\left(\lambda^2 A_{k,l}^\top A_{k,l} + P_k^\top P_k\right) \bar{u}_{k,l} = \lambda^2 A_{k,l}^\top q_{k,l} + P_k^\top d_{k,l}
\]

Gradient:

\[
g = \sum_{k,l} \bar{u}^\top_{k,l} \frac{\partial A^\top_{k,l}}{\partial m} \bar{v}_{k,l}
\]

\[
\bar{v}_{k,l} = A_{k,l}(m) \bar{u}_{k,l} - q_{k,l}
\]
FWI vs WRI

1D velocity model: \( v(z) = v_0 + 0.75z \text{ km/s}, \) with \( v_0 = 2.5 \text{ km/s}. \)
WRI with source estimation

Triple parameters optimization problem:

\[
\min_{u, m, \alpha} \frac{1}{2} \sum_{k,l} \left( \| P_k u_{k,l} - d_{k,l} \|^2_2 + \lambda^2 \| A_{k,l}(m) u_{k,l} - \alpha_{k,l} e_{k,l} \|^2_2 \right)
\]

Two solutions:

- eliminate \( u \) first, and update \( (m, \alpha) \) simultaneously (WRI-SE-MS);
  - **Con:** the amplitudes of \( m \) and \( \alpha \) are different and the same for the gradients.

- simultaneously eliminate \( (u, \alpha) \) and then update \( m \) (WRI-SE-WS).
  - **Pro:** only need to update \( m \), so that it does not suffer from the different amplitudes.
WRI with source estimation

Eliminate $\mathbf{u}$ and $\alpha$ jointly w/ variable projection:

$$\begin{bmatrix} \mathbf{u}, \alpha \end{bmatrix} = \arg \min_{\mathbf{u}, \alpha} \frac{1}{2} \sum_{k,l} \left( \| \mathbf{P}_k \mathbf{u}_{k,l} - \mathbf{d}_{k,l} \|_2^2 + \lambda^2 \| \mathbf{A}_{k,l}(\mathbf{m}) \mathbf{u}_{k,l} - \alpha_{k,l} \mathbf{e}_{k,l} \|_2^2 \right)$$

Analytical solution:

$$\begin{bmatrix} \mathbf{u}_{k,l}(\mathbf{m}) \\ \alpha_{k,l}(\mathbf{m}) \end{bmatrix} = \left[ \begin{array}{ccc} \lambda^2 \mathbf{A}_{k,l}^{\top} \mathbf{A}_{k,l} & \mathbf{P}_k^{\top} \mathbf{P}_k & -\lambda^2 \mathbf{A}_{k,l}^{\top} \mathbf{e}_{k,l} \\ -\lambda^2 \mathbf{e}_{k,l}^{\top} \mathbf{A}_{k,l} & \lambda^2 \mathbf{e}_{k,l}^{\top} \mathbf{e}_{k,l} & 0 \end{array} \right]^{-1} \begin{bmatrix} \mathbf{P}_k \mathbf{d}_{k,l} \\ 0 \end{bmatrix}.$$
Fast solver

When $P_k = P$, $A_{k,l} = A_l$, 

$$
\begin{bmatrix}
\bar{u}_{k,l}(m) \\
\bar{\alpha}_{k,l}(m)
\end{bmatrix} = \begin{bmatrix}
\lambda^2 A_l^T A_l + P^T P & -\lambda^2 A_l^T e_{k,l} \\
-\lambda^2 e_{k,l}^T A_l & \lambda^2 e_{k,l}^T e_{k,l}
\end{bmatrix}^{-1} \begin{bmatrix}
P^T d_{k,l} \\
0
\end{bmatrix}.
$$

Solution —— block matrix inversion formula:

$$C_{k,l}(m) = \begin{bmatrix}
\lambda^2 A_l^T A_l + P^T P & -\lambda^2 A_l^T e_{k,l} \\
-\lambda^2 e_{k,l}^T A_l & \lambda^2 e_{k,l}^T e_{k,l}
\end{bmatrix} = \begin{bmatrix}
M_1 & M_2 \\
M_3 & M_4
\end{bmatrix}$$

Analytical solution:

$$\bar{x}_{k,l}(m) = C_{k,l}^{-1}(m) \begin{bmatrix}
P^T d_{i,j} \\
0
\end{bmatrix}$$

$$= \left( (I + M_1^{-1} M_2 (M_4 - M_3 M_1^{-1} M_2)^{-1} M_3) M_1^{-1} P^T d_{i,j} \right).$$

1. Differ from source to source; 2. Require different factorization and preconditioning matrix.
Example on BG Compass model
— WRI-SE-WS vs WRI-SE-MS

(a) WRI-SE-MS

(b) WRI-SE-WS

(c) Amplitude

(d) Phase
Example on BG Compass model
— WRI-SE-WS vs FWI-SE

Inversion information:
Frequency: 7~30 Hz
Optimization Solver: l-BFGS
Iterations per frequency band: 20
Penalty parameter $\lambda$: 1e0

(a) True model

(b) Bad initial model

(c) Real part

(d) Imaginary part
Example on BG Compass model
— WRI-SE-WS vs FWI-SE

(a) WRI-SE-WS

(b) FWI-SE

(c) Amplitude

(d) Phase
Example on BG Compass model
— WRI-SE-WS vs FWI-SE (Data comparison)

(a) Real part
WRI-SE-WS

(b) Imaginary part
WRI-SE-WS

(c) Real part
FWI-SE

(d) Imaginary part
FWI-SE
Chevron blind test data

Data-set information:
1. 1600 shots: $ds = 25$ m, Source depth = 15 m;
2. 321 recs/shot: $dr = 25$ m, Receiver depth = 15 m;
3. Maximum offset = 8000 m;
4. Record time = 8.0 s, sample rate 4 ms;
5. $V_p$ water = constant = 1510 m/s;
6. With free surface multiples present in the data;
WRI with minimum smoothness constraint

\[
\min_{u(m), m, \alpha(m)} \sum_{k,l} \left\| P_k u_{k,l} - d_{k,l} \right\|^2_2 + \lambda^2 \left\| A_{k,l}(m) u_{k,l} - \alpha_{k,l} e_{k,l} \right\|^2_2
\]

subject to \quad m \in C_1 \cap C_2

\[C_1 \equiv \{ m \mid b_l \leq m \leq b_u \}\]

\[C_2 \equiv \{ m \mid E^T F^T (I - S) F E m = 0 \}\]
WRI with minimum smoothness constraint

\[
\min_{u(m), m, \alpha(m)} \sum_{k,l} \| P_k u_{k,l} - d_{k,l} \|_2^2 + \lambda^2 \| A_{k,l}(m) u_{k,l} - \alpha_{k,l} e_{k,l} \|_2^2
\]

subject to \( m \in C_1 \cap C_2 \)

\[ C_1 \equiv \{ m | b_l \leq m \leq b_u \} \]

\[ C_2 \equiv \{ m | E^T F^T (I - S) F E m = 0 \} \]

1. 2D mirror extension of the model (to avoid periodic boundaries)
2. 2D DFT
3. Remove coefficients outside ellipse (highest spatial frequencies)
4. 2D inverse DFT

[B. Smithyman et al 2015]
WRI with minimum smoothness constraint

![Diagram showing E, F, E^T, F^T, and S]
Chevron blind test data

Inversion strategy:

1. Frequency domain WRI with Source estimation and smoothness constraint;
3. Batch sizes of random frequency subsets: 3, 6, 10, 10, 15, 15;
4. Batch size of random source subsets: 300;
5. Optimization solver: Projected quasi Newton with 20 iterations per frequency band;
6. 2 passes of WRI from frequency 3-19Hz;
7. Grid size: 20m for 3-11Hz and 12m for 3-19Hz;
8. No pre-processing !!!

[M. Schmidt et al, 2012]
Inversion results

(a) Result w/ smoothness constraint

(b) Result w/o smoothness constraint

(c) Well log comparison

(d) Source function comparison
Chapter 6

Uncertainty quantification for inverse problems with weak wave-equation constraints
Motivation

Uncertainty in data

Uncertainty in model

Confidence of the model
Bayesian inference

Prior probability density function (PDF):

\[ m \to \rho_{\text{prior}}(m) \]

Likelihood PDF: given data \( d \)

\[ m \to \rho_{\text{like}}(d|m) \]

Posterior PDF (Bayes’ rule):

\[ \rho_{\text{post}}(m|d) \propto \rho_{\text{like}}(d|m) \rho_{\text{prior}}(m) \]

[A. Tarantola and B. Valette, 1982]
[J. Kaipio and E. Somersalo, 2004]
Bayesian w/ FWI

\[ \mathbf{d}_{\text{obs}} = F(\mathbf{m}) + \epsilon \sim \mathcal{N}(0, \Sigma_{\text{noise}}) \]
Bayesian w/ FWI

Posterior PDF of FWI:

$$\rho_{\text{post}}(m|d) \propto \exp\left(-\frac{1}{2} \|PA(m)^{-1}q - d\|^2_{\Sigma^{-1}_{\text{noise}}} - \frac{1}{2} \|m - m_{\text{prior}}\|^2_{\Sigma^{-1}_{\text{prior}}}ight)$$
Bayesian w/ FWI

Posterior PDF of FWI:

\[ \rho_{\text{post}}(m|d) \propto \exp \left( -\frac{1}{2} \|PA(m)^{-1}q - d\|^2_{\Sigma_{\text{noise}}^{-1}} - \frac{1}{2} \|m - m_{\text{prior}}\|^2_{\Sigma_{\text{prior}}^{-1}} \right) \]

Strong nonlinearity
Many local minima
Bayesian w/ FWI

Posterior PDF of FWI:

\[
\rho_{\text{post}}(m|d) \propto \exp\left(-\frac{1}{2}||PA(m)^{-1}q - d||^2_{\Sigma^{-1}_{\text{noise}}} - \frac{1}{2}||m - m_{\text{prior}}||^2_{\Sigma^{-1}_{\text{prior}}}ight)
\]

Strong nonlinearity \(-\log \rho_{\text{post}}(m|d)\)

Many local minima
Extend the search space

Introduce auxiliary variable:

\[ \rho_{\text{post}}(u, m|d) \propto \rho(d|u, m)\rho(u, m) \]

\[ \rho(u, m) = \rho(u|m)\rho_{\text{prior}}(m) \]

Strict PDE constraints:

\[ \rho(d|u, m) \propto \exp\left(-\frac{1}{2}\|Pu - d\|^2_{\Sigma_{\text{noise}}^{-1}}\right) \]

\[ \rho(u|m) = \delta(A(m)u - q) \]
Extend the search space

Introduce auxiliary variable:

\[ \rho_{\text{post}}(u, m|d) \propto \rho(d|u, m)\rho(u, m) \]
\[ \rho(u, m) = \rho(u|m)\rho_{\text{prior}}(m) \]

Strict PDE constraints:

\[ \rho(d|u, m) \propto \exp\left(-\frac{1}{2}\|Pu - d\|_\Sigma^{-1}\right) \]
\[ \rho(u|m) = \delta(A(m)u - q) \]
Extend the search space

Relax PDE constraints:

\[ \rho(u|m) = \delta(A(m)u - q) \]

\[ \rho(u|m) \propto \det(\lambda^2 A(m)^\top A(m))^{\frac{1}{2}} \exp\left(-\frac{\lambda^2}{2} \|A(m)u - q\|^2\right) \]
Posterior distribution with weak PDE constraints

Joint posterior distribution:
\[ \rho_{\text{post}}(u, m|d) \propto \rho(d|u, m) \rho(u, m) \quad \text{with} \]
\[ \rho(u, m) = \rho(u|m) \rho_{\text{prior}}(m), \]
\[ \rho(d|u, m) \propto \exp \left( -\frac{1}{2} \| Pu - d \|_{\Sigma^{-1}_{\text{noise}}}^2 \right), \]
\[ \rho(u|m) \propto \det \left( \lambda^2 A(m)^T A(m) \right)^{\frac{1}{2}} \exp \left( -\frac{\lambda^2}{2} \| A(m)u - q \|^2 \right). \]

Marginal distribution:
\[ \rho_{\text{post}}(m|d) = \int \rho_{\text{post}}(u, m|d) du \]
\[ \propto \det(H(m))^{-\frac{1}{2}} \det \left( \lambda^2 A(m)^T A(m) \right)^{\frac{1}{2}} \]
\[ \times \exp \left( -\frac{1}{2} \left( \lambda^2 \| A(m)\tilde{u}(m) - q \|^2 + \| Pu(m) - d \|_{\Gamma^{-1}_{\text{noise}}}^2 + \| m - \tilde{m} \|_{\Gamma^{-1}_{\text{prior}}}^2 \right) \right) \]
where
\[ H(m) = -\nabla_u^2 \log \rho_{\text{post}}(u, m|d)|_{u=\tilde{u}(m)} \]
\[ = \lambda^2 A(m)^T A(m) + P^T \Gamma^{-1}_{\text{noise}} P, \]
\[ \tilde{u}(m) = H(m)^{-1} \left( \lambda^2 A(m)^T q + P^T \Gamma^{-1}_{\text{noise}} d \right). \]
Selection of $\lambda$

Negative logarithm function of posterior distribution:

$$
\phi(m) = -\log \rho_{\text{post}}(m|d) = \phi_1(m) + \phi_2(m), \quad \text{with}
$$

$$
\phi_1(m) = \frac{1}{2} \log \det \left( I + \frac{1}{\lambda^2} \Gamma^{-\frac{1}{2}}_{\text{noise}} P A(m)^{-1} A(m)^{-T} P^T \Gamma^{-\frac{1}{2}}_{\text{noise}} \right)
$$

$$
\phi_2(m) = \frac{1}{2} \left( \lambda^2 \| A(m) \bar{u}(m) - q \|_2^2 + \| P \bar{u}(m) - d \|_{\Gamma^{-1}_{\text{noise}}}^2 + \| m - \hat{m} \|_{\Gamma^{-1}_{\text{prior}}}^2 \right).
$$

when $\lambda \to 0$, $\bar{u}(m)$ tends to fit the data and

$$
\phi_2(m) \to \frac{1}{2} \| m - \hat{m} \|_{\Gamma^{-1}_{\text{prior}}}^2,
$$

$$
\phi_1(m) \to \infty.
$$

when $\lambda \to \infty$,

$$
\phi_1(m) \to 0
$$

$$
\phi_2(m) \to \frac{1}{2} \| P A(m)^{-1} q - d \|_{\Gamma^{-1}_{\text{noise}}}^2 + \frac{1}{2} \| m - \hat{m} \|_{\Gamma^{-1}_{\text{prior}}}^2
$$

Strict constraint
Selection of $\lambda$

Select $\lambda$ according to the largest eigenvalue $\mu_1$ of the matrix $A(m)^{-\top}P^\top\Gamma_{\text{noise}}^{-1}PA(m)^{-1}$:

(1) $\lambda^2 \gg \mu_1$, $\lambda$ is large; 
(2) $\lambda^2 \ll \mu_1$, $\lambda$ is small.

Selection of $\lambda$:

$$\lambda^2 = 0.01\mu_1.$$ 

Approximate the posterior distribution by:

$$\rho_{\text{post}}(m|d) \approx \bar{\rho}_{\text{post}}(m|d) \propto \exp \left( -\frac{\lambda^2}{2} \|A(m)\bar{u}(m) - q\|^2 - \frac{1}{2} \|P\bar{u}(m) - d\|_{\Gamma_{\text{noise}}^{-1}}^2 - \frac{1}{2} \|m - \tilde{m}\|_{\Gamma_{\text{prior}}^{-1}}^2 \right).$$

\[\text{Likelihood } \rho_{\text{like}} \quad \text{Prior } \rho_{\text{prior}}\]
Gaussian Approximation

Approximate the posterior distribution by a Gaussian distribution:

\[ \bar{\rho}_{\text{post}}(m|d) \approx \rho_{\text{Gauss}}(m) = \mathcal{N}(m_*, \tilde{H}^{-1}_{\text{post}}) = \mathcal{N}(m_*, (\tilde{H}_{\text{like}} + \Gamma^{-1}_{\text{prior}})^{-1}), \]

with the maximum a posterior estimate \( m_* \), and the Gauss-Newton Hessian \( \tilde{H}_{\text{like}} \) of the negative logarithm likelihood function \(-\log \rho_{\text{like}}(m)\) given by:

\[
\tilde{H}_{\text{like}} = \begin{bmatrix} G^\top & A^{-\top} P^\top \end{bmatrix} \begin{bmatrix} (\Gamma_{\text{noise}} + \frac{1}{\lambda^2} P A^{-1} A^{-\top} P^\top)^{-1} P A^{-1} G. \end{bmatrix}
\]

where \( G = \frac{\partial A \bar{u}}{\partial m}. \)

\[
\tilde{H}_{\text{like}} = \begin{bmatrix} S & W \end{bmatrix} \begin{bmatrix} S^\frac{1}{2} W \end{bmatrix} = (S^\frac{1}{2} W)^\top S^\frac{1}{2} W = R_{\text{like}}^\top R_{\text{like}}.
\]

**Computational cost:**
\( n_{\text{freq}} \times (n_{\text{src}} + n_{\text{rcv}}) \)

**storage cost:**
\( n_{\text{freq}} \times n_{\text{grid}} \times (n_{\text{src}} + n_{\text{rcv}}) \)

in parallel !!
Sampling method

Sample Gaussian distribution:

\[ \tilde{\rho}_{\text{post}}(\mathbf{m}|\mathbf{d}) \approx \rho_{\text{Gauss}}(\mathbf{m}) = \mathcal{N}(\mathbf{m}^*_s, \tilde{\mathbf{H}}^{-1}_{\text{post}}) = \mathcal{N}(\mathbf{m}^*_s, (\tilde{\mathbf{H}}_{\text{like}} + \Gamma^{-1}_{\text{prior}})^{-1}). \]

Cholesky factorization method:

\[ \tilde{\mathbf{H}}_{\text{post}} = \mathbf{R}_{\text{post}}^\top \mathbf{R}_{\text{post}}, \]

\[ \mathbf{m}_s = \mathbf{m}^*_s + \mathbf{R}^{-1}_{\text{post}} \mathbf{r}, \mathbf{r} \sim \mathcal{N}(0, \mathcal{I}). \]

Cons: 1. Explicit formulation of \( \tilde{\mathbf{H}}_{\text{post}} \)

requires a storage cost of

\[ \mathcal{O}(n_{\text{grid}}^2). \]

2. Computational cost is \( \mathcal{O}(n_{\text{grid}}^3) \).

Randomize then optimize (RTO) method:

Let \( \tilde{\mathbf{H}}_{\text{like}} = \mathbf{R}_{\text{like}}^\top \mathbf{R}_{\text{like}} \), and \( \Gamma_{\text{prior}} = \mathbf{R}_{\text{prior}}^\top \mathbf{R}_{\text{prior}} \),

we have:

\[ \mathbf{m}_s = \arg \min_{\mathbf{m}} \| \mathbf{R}_{\text{like}} \mathbf{m} - \mathbf{R}_{\text{like}} \mathbf{m}^*_s - \mathbf{r}_{\text{like}} \|^2 \]

\[ + \| \mathbf{R}_{\text{prior}}^{-\top} \mathbf{m} - \mathbf{R}_{\text{prior}}^{-\top} \mathbf{m}^*_s - \mathbf{r}_{\text{prior}} \|^2; \]

\[ \mathbf{r}_{\text{like}} \sim \mathcal{N}(0, \mathcal{I}_{n_{\text{data}} \times n_{\text{data}}}); \]

\[ \mathbf{r}_{\text{prior}} \sim \mathcal{N}(0, \mathcal{I}_{n_{\text{grid}} \times n_{\text{grid}}}). \]

Pros: 1. Do not need the explicit Hessian matrix;

2. Can use iterative methods to solve it.
Workflow

Likelihood distribution → MAP estimate → GN Hessian and Gaussian distribution

Prior distribution

Gaussian approximation + RTO (GA-RTO)

Compute statistics from samples → Sample the Gaussian distribution by RTO

Workflow

Likelihood distribution → MAP estimate → GN Hessian and Gaussian distribution

Prior distribution

Gaussian approximation + RTO (GA-RTO)

Compute statistics from samples → Sample the Gaussian distribution by RTO
Numerical examples
— investigate $\lambda$

1D velocity model: $v(z) = v_0 + 0.75z \text{ km/s}$, with $v_0 = 2.5 \text{ km/s}$.

10% Gaussian noise.
Numerical examples

--- investigate $\lambda$

Note: $\psi_1 = -\log \rho_{\text{like}}$, $\psi_2 = -\log \bar{\rho}_{\text{like}}$, $\psi_3 = -\log \rho_{\text{Gauss,like}}$
**Numerical examples — Layer model**

**Depth of sources and receivers:** 50 m  
**Number of sources and receivers:** 61  
**Penalty parameter:** $\lambda^2 = 0.01\mu_1$

**Frequency:** 5, 6, and 7 Hz  
**15% Gaussian noise:** $\Gamma_{\text{noise}} = 175^2 I$

---

(a) True model  
(b) Prior model
Numerical examples
— Layer model

Covariance matrix of the prior distribution — Gaussian smoothness:

$$\Gamma_{\text{prior}}(k, l) = a \exp\left(\frac{-||s_k - s_l||^2}{2b^2}\right) + c\delta_{k,l},$$

where the vectors $s_k = (z_k, x_k)$ and $s_l = (z_l, x_l)$ denote the $k^{\text{th}}$ and $l^{\text{th}}$ elements in the vector $m$. Parameters $a$, $b$, and $c$ control the correlation strength, variance, and spatial correlation distance.

In this example, our selections are as follows:

$$a = 0.1 \text{ km}^2/\text{s}^2,$$
$$b = 0.65,$$
$$c = 0.01 \text{ km}^2/\text{s}^2.$$
Benchmark method
— Randomized Maximum Likelihood - RML

Generate independent samples from $\tilde{\rho}_{\text{post}}(m)$ by solving:

$$
\min_{m} \frac{1}{2} \left( \| \mathbf{P}\bar{u}(m) - d - r_d \|_{\Gamma_{\text{noise}}^{-1}}^2 + \lambda^2 \| \mathbf{A}(m)\bar{u}(m) - q - r_s \|_2^2 \right) \\
+ \frac{1}{2} \| m - m_p - r_p \|_{\Sigma_{\text{prior}}^{-1}}^2 ,
$$

where

$$
\begin{align*}
\mathbf{r}_d & \sim \mathcal{N}(0, \sigma^2 \mathcal{I}_{n_{\text{recV}} \times n_{\text{recV}}}) , \\
\mathbf{r}_s & \sim \mathcal{N}(0, \lambda^{-2} \mathcal{I}_{n_{\text{grid}} \times n_{\text{grid}}}) , \\
\mathbf{r}_p & \sim \mathcal{N}(0, \Sigma_{\text{prior}}).
\end{align*}
$$

Require solving an FWI type problem to draw one sample.
Posterior mean models and standard deviations of 10000 samples

(a) Mean model by GA-RTO

(b) Mean model by RML

(c) Standard deviation by GA-RTO

(d) Standard deviation by RML

36360 PDE solves

360 million PDE solves

1.5%

6%
95% Confidence intervals of 10000 samples

(a) $x = 500$ m

(b) $x = 1500$ m

(c) $x = 2500$ m
Marginal distributions of 10000 samples

(a) $x = 1500 \text{ m}, \ z = 200 \text{ m}$  
(b) $x = 1500 \text{ m}, \ z = 700 \text{ m}$  
(c) $x = 1500 \text{ m}, \ z = 1200 \text{ m}$
Numerical examples — BG compass model

Depth of sources and receivers: 50
Number of sources and receivers: 91 / 451
Central frequency: 15 Hz
Frequency: 2-31 Hz
15% Gaussian noise: $\Gamma_{\text{noise}} = 46^2 I$
Penalty parameter: $\lambda^2 = 0.01 \mu_1$

(a) True model

(b) Prior model

(c) STD of prior distribution
MAP and standard deviation of 2000 samples

We first compute the MAP estimate, and then generate 2000 samples.
Marginal distributions of 2000 samples

(a) $x = 2240 \text{ m}, z = 40 \text{ m}$

(b) $x = 2240 \text{ m}, z = 640 \text{ m}$

(c) $x = 2240 \text{ m}, z = 1240 \text{ m}$

(d) $x = 2240 \text{ m}, z = 1840 \text{ m}$
Cross section comparison
– 95% confidence interval vs 10 realizations by RML

(a) $x = 1000 \text{ m}$

(b) $x = 2500 \text{ m}$

(c) $x = 4000 \text{ m}$
Conclusion

In this thesis, I have developed

- a source estimation algorithm for time-domain sparsity promoting LS-RTM approaches

- a source estimation algorithm for wavefield reconstruction inversion

- an algorithm to quantify uncertainties for inverse problems with weak wave-equation constraints
Publications


Thank you!

- PhD Advisory Committee members
- PhD Exam Committee members
- SLIM Group Members
- SINBAD sponsors & NSERC
- Family and Friends