Noise Reduction by Using Interferometric Measurements

Rongrong Wang, and Bas Peters

December 2nd, 2013

University of British Columbia

Problem

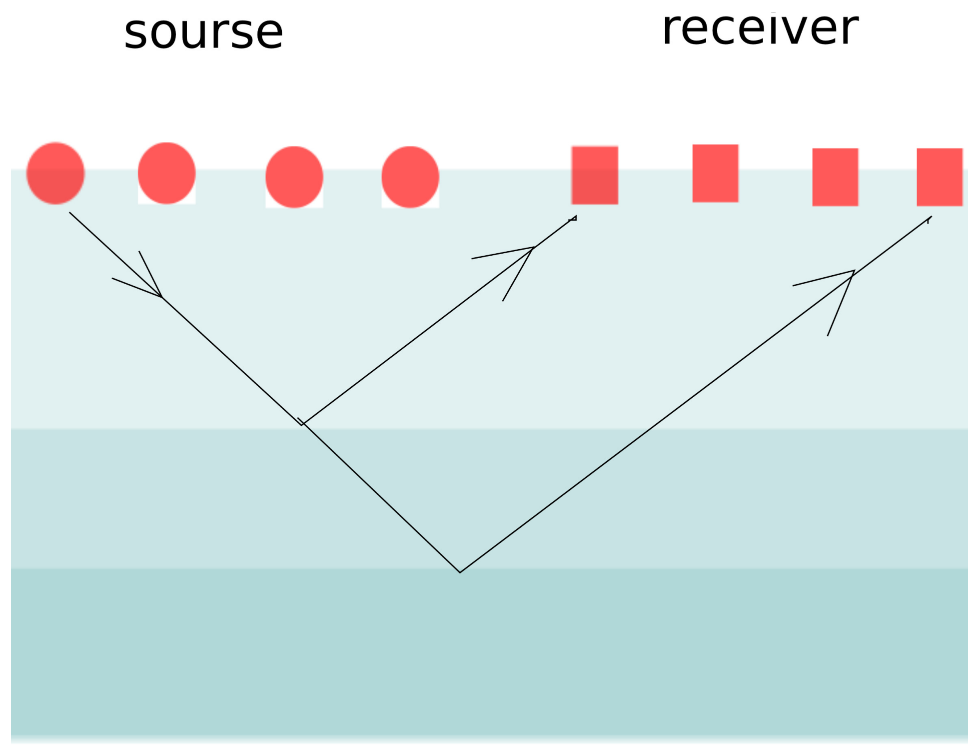


Figure : Source and receiver location in simulation

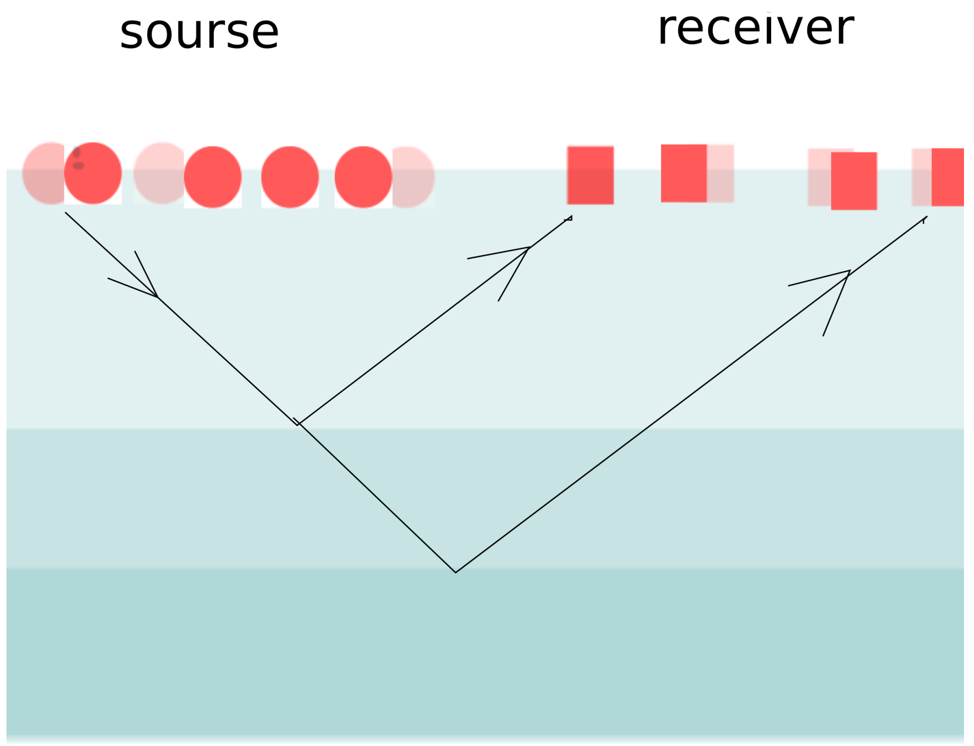


Figure : True location

The modeling error will cause: large phase shift in high frequency data.

Full waveform inversion in frequency domain

If the everything is accurate, the forward operator F maps the true model to the data

$$F_i(\mathbf{m}) = \mathbf{d_i}, \quad \text{for } \mathbf{i} = \mathbf{1}, ..., \mathbf{nf} * \mathbf{ns}, \tag{1}$$

where F_i satisfies

$$\begin{aligned} (\Delta + \mathbf{m}\omega^2)\mathbf{u_i} &= \mathbf{q_i}, \\ F_i(\mathbf{m}) &= P_i \mathbf{u_i}(\mathbf{m},\ \mathbf{q_i}), \end{aligned}$$

and where $\mathbf{q_i}$ is the source term, P_i is the projection operator on the receivers, $\mathbf{m}$ is the model parameter (s^2/km^2), and $\mathbf{u_i}$ is the wavefield.

Now $\mathbf{q_i}$ and P_i are inaccurate: $F_i \to \widetilde{F}_i$, and

$$\widetilde{F}_i(\mathbf{m}) \approx \mathbf{d_i} \circ e^{j\theta_\mathbf{i}}.$$

The least square solution

The least square solver solves

$$\hat{\mathbf{m}} = \arg\min \sum_{i=1}^{nf*ns} \|\widetilde{F}_i(\mathbf{m}) - \mathbf{d_i}\|_2^2.$$

Assume $nf = 1$, this solver becomes

$$\hat{\mathbf{m}} = \arg\min \|\widetilde{F}(\mathbf{m}) - \mathbf{d}\|_2^2,$$

where $\widetilde{F} = \begin{bmatrix} -- & \widetilde{F}_1 & -- \\ -- & \widetilde{F}_2 & -- \\ -- & \vdots & -- \\ -- & \widetilde{F}_{ns} & -- \end{bmatrix}$ and $\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_{ns} \end{bmatrix}$.

We are directly fitting $\mathbf{d}$ with $\widetilde{F}(m)$.

Interferometric lifting: another way of fitting the data

[Demanet et. al, 2013] Lift a vector to a matrix:

$$\mathbf{x} \equiv \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \rightarrow \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \left[\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n\right] \equiv \begin{bmatrix} |x_1|^2 & x_1\bar{x}_2 & \cdots & x_1\bar{x}_n \\ x_2\bar{x}_1 & |x_2|^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & x_{n-1}\bar{x}_n \\ x_n\bar{x}_1 & \cdots & x_n\bar{x}_{n-1} & |x_n|^2 \end{bmatrix} \equiv X$$

Benefit of the lifting: $\|x\|^2 = trace(X)$, changing the nonlinear function $\|\cdot\|_2^2$ of $\mathbf{x}$ to

the linear function $trace(\cdot)$ of X.

Interferometric lifting: another way of fitting the data

Use the lifting on the seismic model:

- $\mathbf{d} \to \mathbf{d}\mathbf{d}^{\mathbf{T}}$;
- $\widetilde{F}(\mathbf{m}) \to \widetilde{F}(\mathbf{m})\widetilde{F}(\mathbf{m})^T$;
- Fitting $\mathbf{d}$ with $\widetilde{F}(\mathbf{m}) \to$ fitting $\mathbf{d}\mathbf{d}^{\mathbf{T}}$ with $\widetilde{F}(\mathbf{m})\widetilde{F}(\mathbf{m})^T$;
- Solving $\min \|\widetilde{F}(\mathbf{m}) - \mathbf{d}\|_2^2 \to$ solving $\min \|\widetilde{F}(\mathbf{m})\widetilde{F}(\mathbf{m})^T - \mathbf{d}\mathbf{d}^{\mathbf{T}}\|_{\mathbf{F}}^{\mathbf{2}}$.

By lifting the data vector to a matrix, we

- used the same amount of data;
- increased the computational complexity;
- hope to increase the accuracy.

A quick observation: If F is linear, then $\arg\min \|\widetilde{F}(\mathbf{m}) - \mathbf{d}\|_2^2 = \arg\min \|\widetilde{F}(\mathbf{m})\widetilde{F}(\mathbf{m})^T - \mathbf{d}\mathbf{d}^{\mathbf{T}}\|_{\mathbf{F}}^{\mathbf{2}}$, the two problems are equivalent.

Interferometric lifting: another way of fitting the data

What if we fit only a portion of $\mathbf{dd^T}$?

Define P_Ω: $bb^T \to \mathbb{C}^{n,n}$ by

$$\begin{bmatrix} |\mathbf{d}_1|^2 & \mathbf{d}_1\bar{\mathbf{d}}_2 & \mathbf{d}_1\bar{\mathbf{d}}_3 & \mathbf{d}_1\bar{\mathbf{d}}_4 \\ \mathbf{d}_2\bar{\mathbf{d}}_1 & |\mathbf{d}_2|^2 & \mathbf{d}_2\bar{\mathbf{d}}_3 & \mathbf{d}_2\bar{\mathbf{d}}_4 \\ \mathbf{d}_3\bar{\mathbf{d}}_1 & \mathbf{d}_3\bar{\mathbf{d}}_2 & |\mathbf{d}_3|^2 & \mathbf{d}_3\bar{\mathbf{d}}_4 \\ \mathbf{d}_4\bar{\mathbf{d}}_1 & \mathbf{d}_4\bar{\mathbf{d}}_2 & \mathbf{d}_4\bar{\mathbf{d}}_3 & |\mathbf{d}_4|^2 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} |\mathbf{d}_1|^2 & 0 & 0 & \mathbf{d}_1\bar{\mathbf{d}}_4 \\ \mathbf{d}_2\bar{\mathbf{d}}_1 & |\mathbf{d}_2|^2 & 0 & 0 \\ \mathbf{d}_3\bar{\mathbf{d}}_1 & 0 & |\mathbf{d}_3|^2 & 0 \\ 0 & \mathbf{d}_4\bar{\mathbf{d}}_2 & 0 & |\mathbf{d}_4|^2 \end{bmatrix}$$

We only fit the chosen entries

$$\min \left\| \begin{bmatrix} \lambda|\widetilde{F}(\mathbf{m})_1|^2 & 0 & 0 & \widetilde{F}(\mathbf{m})_1\overline{\widetilde{F}(\mathbf{m})_4} \\ \widetilde{F}(\mathbf{m})_2\overline{\widetilde{F}(\mathbf{m})_1} & \lambda|\widetilde{F}(\mathbf{m})_2|^2 & 0 & 0 \\ \widetilde{F}(\mathbf{m})_3\overline{\widetilde{F}(\mathbf{m})_1} & 0 & \lambda|\widetilde{F}(\mathbf{m})_3|^2 & 0 \\ 0 & \widetilde{F}(\mathbf{m})_4\overline{\widetilde{F}(\mathbf{m})_2} & 0 & \lambda|\widetilde{F}(\mathbf{m})_4|^2 \end{bmatrix} - \begin{bmatrix} \lambda|\mathbf{d}_1|^2 & 0 & 0 & \mathbf{d}_1\bar{\mathbf{d}}_4 \\ \mathbf{d}_2\bar{\mathbf{d}}_1 & \lambda|\mathbf{d}_2|^2 & 0 & 0 \\ \mathbf{d}_3\bar{\mathbf{d}}_1 & 0 & \lambda|\mathbf{d}_3|^2 & 0 \\ 0 & \mathbf{d}_4\bar{\mathbf{d}}_2 & 0 & \lambda|\mathbf{d}_4|^2 \end{bmatrix} \right\|_F^2$$

or in compact form

$$\min \|(P_\Omega + \lambda(P_D))(\widetilde{F}(\mathbf{m})\widetilde{F}(\mathbf{m})^T - \mathbf{dd^T})\|_{\mathbf{F}}^{\mathbf{2}}, \tag{2}$$

where D represent the diagonal indices and $\lambda > 1$.

Interferometric lifting: another way of fitting the data

How to choose the set of fitting indices Ω?

- The diagonal entries are always included;
- Put more weight on the diagonal entries when phase error is dominant;
- Each of the off diagonal entries is chosen independently with probability p. This is called the Erdös-Rényi model.

Why the Erdös-Rényi model? A short review of interferometry

Borcea et al. (2005) proposed the Coherent INTerfermetric imaging functional (CINT)

$$I_{\mathrm{CINT}} = \mathrm{diag}\{\widetilde{F}^*(P_\Omega(\mathbf{dd}^*)\widetilde{F}\}$$

Jugnon et al. (2013) extend the model to solve inversion problem

$$\text{find} \;\; \mathbf{m} \; s.t. \; \|P_\Omega[\widetilde{F}\mathbf{mm}^*\widetilde{F} - \mathbf{dd}^*]\|_1 < \sigma$$

and prove that

$$\frac{\|\hat{\mathbf{m}} - e^{i\alpha}\mathbf{m}\|}{\|\mathbf{m}\|} \leq 15\kappa(\widetilde{F})^2\sqrt{\frac{\sigma}{\lambda_2}} + \kappa(\widetilde{F})\frac{\|(F - \widetilde{F})\mathbf{m}\|}{\|\mathbf{b}\|}.$$

The Erdös-Rényi model will guarantee λ_2 to be large.

Fitting indices chosen in the Erdös-Rényi model

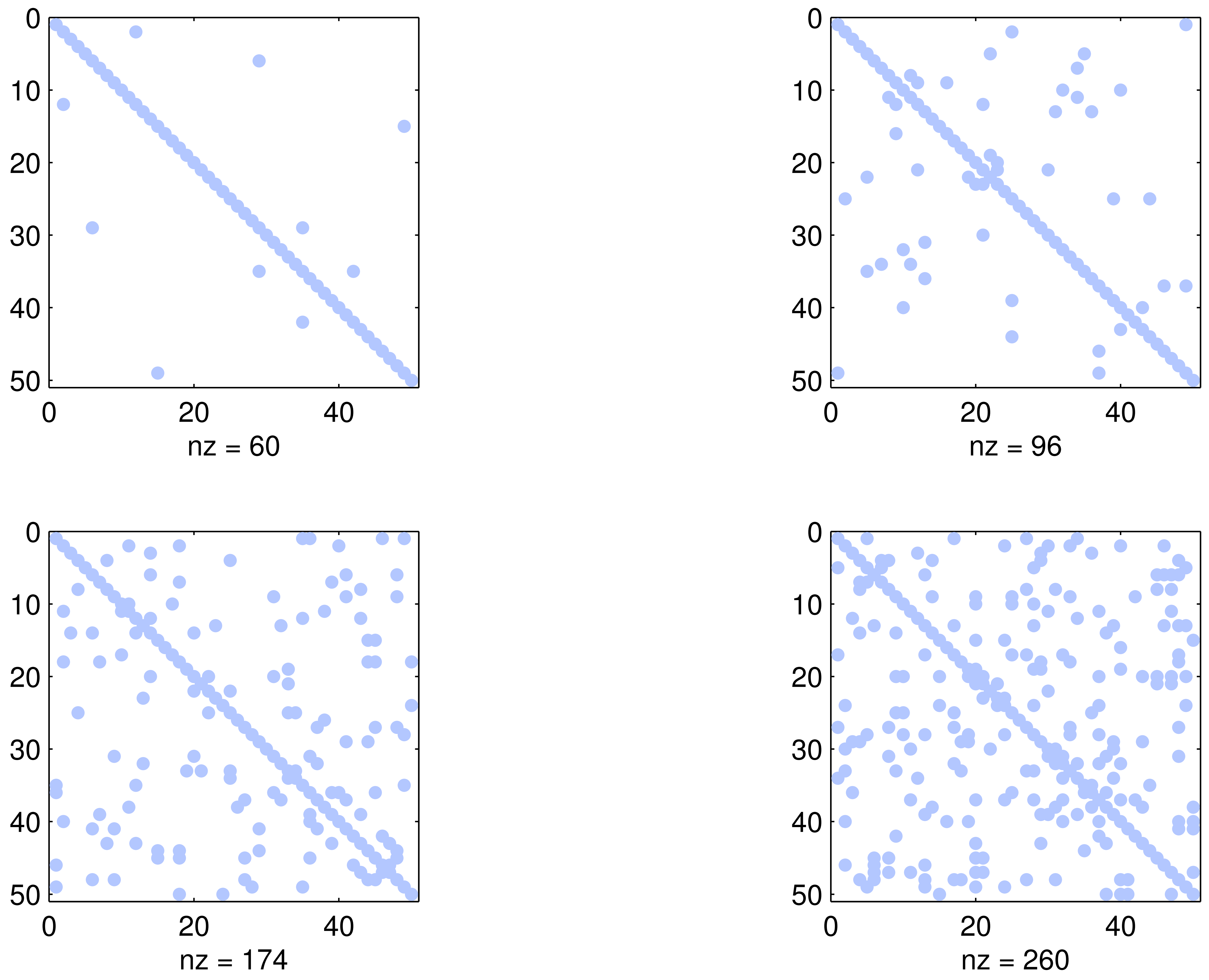


Figure : The active indices for p=0.005,0.01,0.05,0.1

How to choose p?

Theorem (R. Wang, and B. Peters, 2013)

Suppose F is linear, Ω is symmetric and the error only comes in the phase of the data. Let $\hat{\mathbf{m}}$ be the solution to (2), then we have

$$(\hat{\mathbf{m}}-\mathbf{m})^T(F^T\mathcal{L}F)(\hat{\mathbf{m}}-\mathbf{m}) \leq 2\left|Trace(\mathcal{L}\circ(\mathbf{d}\mathbf{d}^T - \widetilde{F}(\mathbf{m})\widetilde{F}(\mathbf{m})^T))\right|,$$

where $\mathcal{L}$ is the Graph laplacian depending on both Ω and $\mathbf{d}$.

Theorem

If we further assume that $\hat{m}$ is close to $\mathbf{m}$, then it holds

$$\|\hat{\mathbf{m}}-\mathbf{m}\|_2^2 \lesssim 2\frac{\left|Trace(\mathcal{L}\circ(\mathbf{d}\mathbf{d}^T - \widetilde{F}(\mathbf{m})\widetilde{F}(\mathbf{m}))^T)\right|}{\lambda_{\min}(R^T\mathcal{L}R)}, \tag{3}$$

where $R = F\cdot Null\,(diag[Re(F\mathbf{m})]Re(F) + diag[Im(F\mathbf{m})]Im(F))$. If $R=0$, then the magnitude $|\mathbf{d}|$ alone is enough to recover the model $\mathbf{m}$.

Choose p such that the bound in (3) is minimized.

Implementation issues

In practice, both F and $\mathbf{m}$ are unknown and F is nonlinear. We propose to approximate the ratio $\frac{|Trace(\mathcal{L}\circ(\mathbf{d}\mathbf{d}^T - \widetilde{F}(\mathbf{m})\widetilde{F}(\mathbf{m})^T))|}{\lambda_{\min}(R^T\mathcal{L}R)}$, by

- replacing $\widetilde{F}(\mathbf{m})$ with $\widetilde{F}(\mathbf{m_0})$, where $\mathbf{m_0}$ is the initial guess;
- replacing $R = F \cdot Null\,(diag[Re(F\mathbf{m})]Re(F) + diag[Im(F\mathbf{m})]Im(F))$ with

$$\widetilde{R} = F\left\{Null\left[diag[Re(F(\mathbf{m}_0))]Re(\frac{\mathrm{d}F}{\mathrm{d}\mathbf{m}}(\mathbf{m}_0)) + diag[Im(F(\mathbf{m}_0)]Im(\frac{\mathrm{d}F}{\mathrm{d}\mathbf{m}}(\mathbf{m}_0))\right]\right\}.$$

Let p_0 be the minimizer of $\frac{|Trace(\mathcal{L}\circ(\mathbf{d}\mathbf{d}^T - \widetilde{F}(\mathbf{m})\widetilde{F}(\mathbf{m})^T))|}{\lambda_{\min}(\widetilde{R}^T\mathcal{L}\widetilde{R})}$, we denote this ratio by $\frac{\sigma}{\lambda}$. p_0 is data dependent so it needs to be calculated every time.

Summarizing the algorithm

Input: data $\mathbf{d}$, inaccurate forward operator $\tilde{F}$, initial guess $\mathbf{m_0}$.

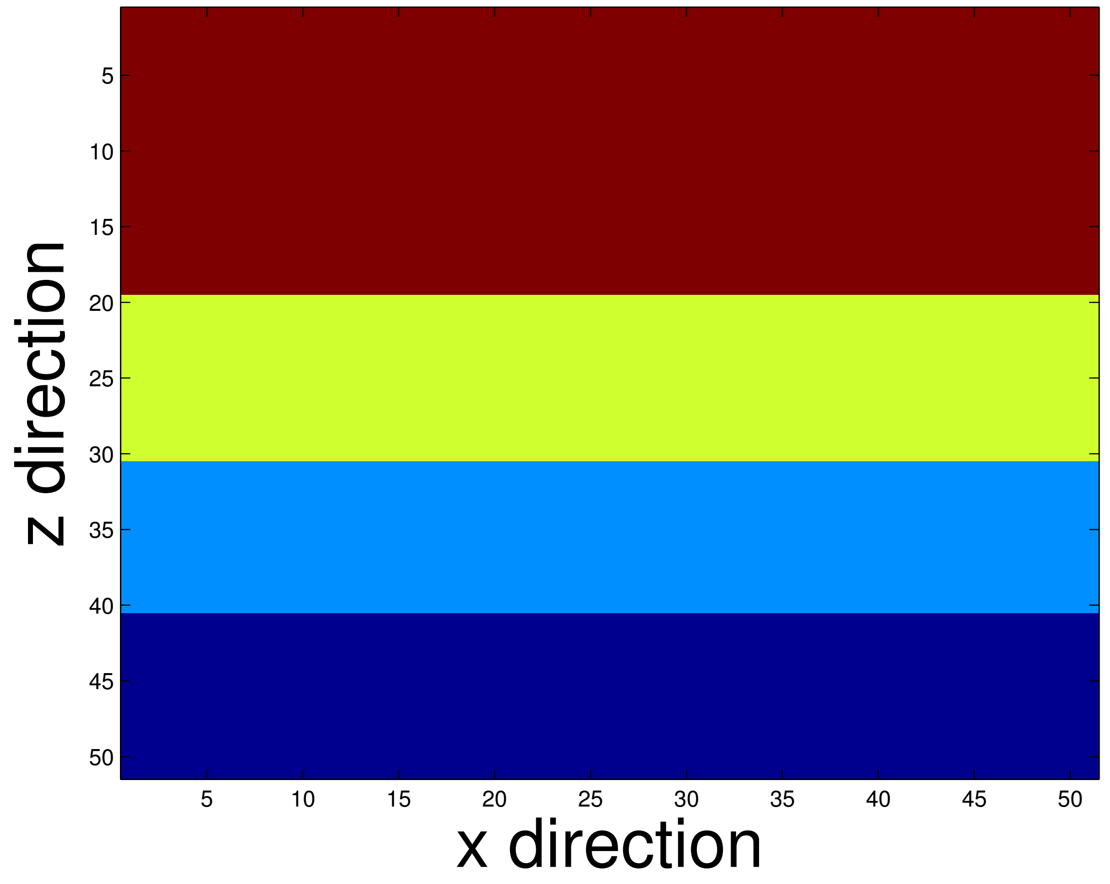
- lift the date to a matrix $\mathbf{dd^T}$.
- Find the p that minimizes the ratio $\frac{\left|Trace(P_\Omega(\mathbf{dd}^T - \widetilde{F}(\mathbf{m})\widetilde{F}(\mathbf{m})^T))\right|}{\lambda_{\min}(\widetilde{R}^T\mathcal{L}\widetilde{R})}$.
- Use the minimizer p_0 to generate the random fitting index set Ω.
- Solve $\min L \equiv \|(P_\Omega + \lambda(P_D))(\widetilde{F}(\mathbf{m})\widetilde{F}(\mathbf{m})^T - \mathbf{dd^T})\|_\mathbf{F}^\mathbf{2}$, by the L-BFGS algorithm with hand calculated gradient

$$\frac{\mathrm{d}L}{\mathrm{d}\mathbf{m}} = 2\left(\frac{\mathrm{d}F}{\mathrm{d}\mathbf{m}}\right)^T((P_\Omega + \lambda(P_D))(\widetilde{F}(\mathbf{m})\widetilde{F}(\mathbf{m})^T - \mathbf{dd}^T)\widetilde{F}(\mathbf{m}).$$

An example

2-D frequency domain full waveform inversion:

- grid size 51×51, model size $510m \times 510m$;
- equally distributed sources and receivers near surface;
- Narrowbanded data: $f = [15, 20, 25]$;
- Horizontal perturbation of source and receiver with variance $2m$;
- 30Hz peak frequency of the Ricker wavelet;
- Vertical variable velocity



- Use L-BFGS for both least square and interferometric methods.

The best p

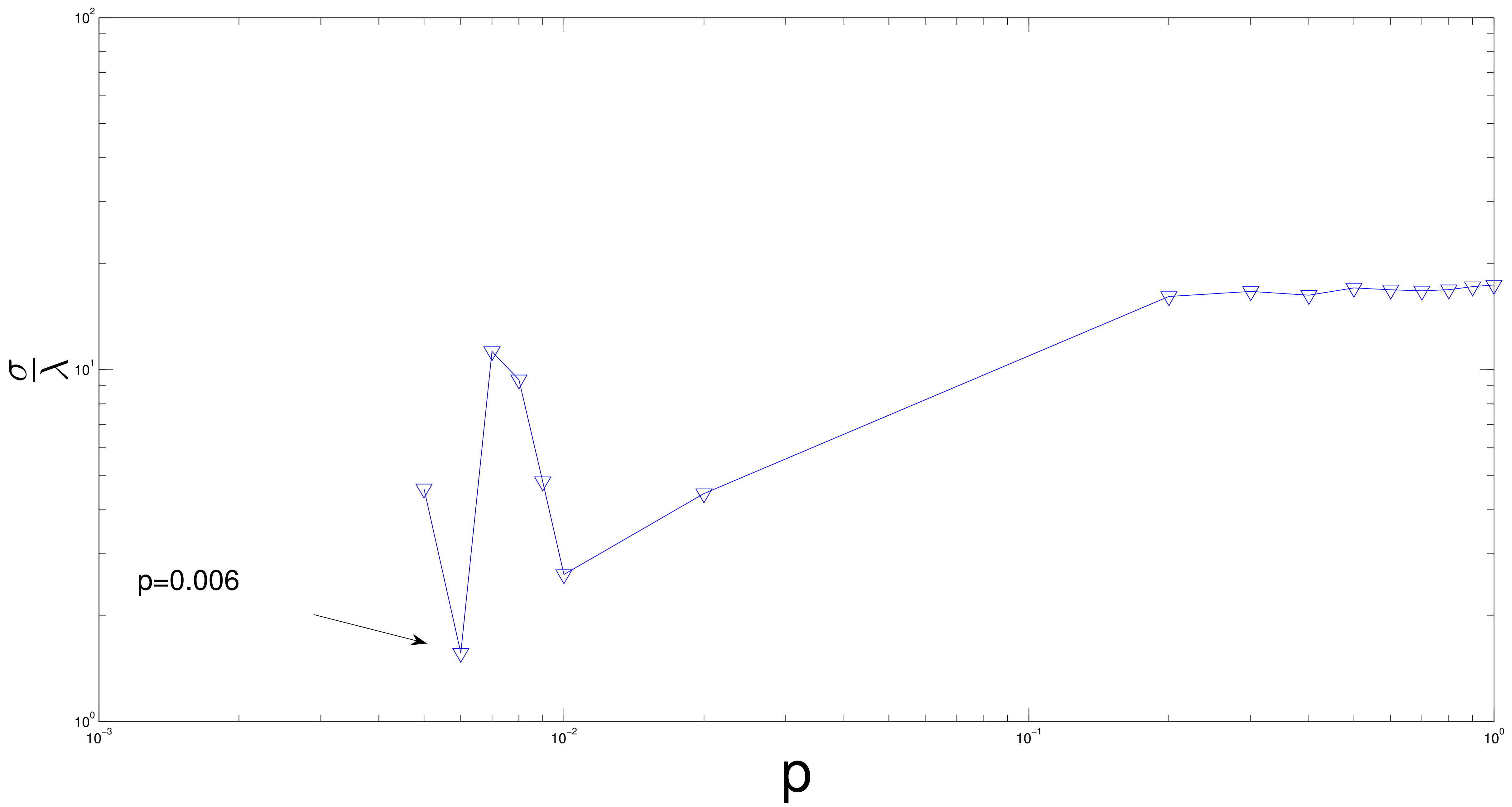


Figure : $\frac{\sigma}{\lambda}$ as a function of p.

Reconstruction Result

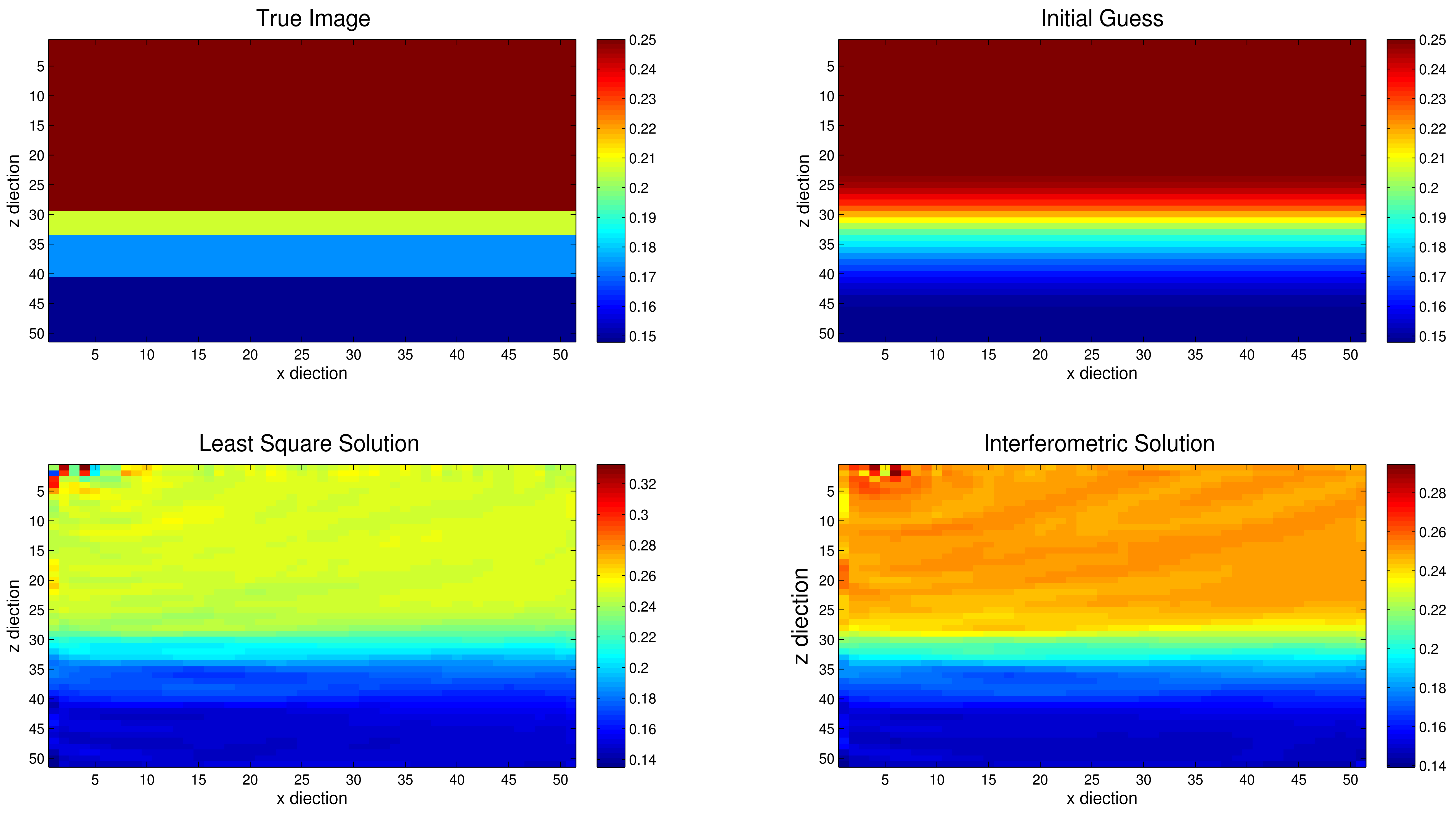


Figure : results after 10 iterations. 6 sources on the left and 800 receivers on the right, frequencies in use: [20,25,30], $\lambda = 100$.

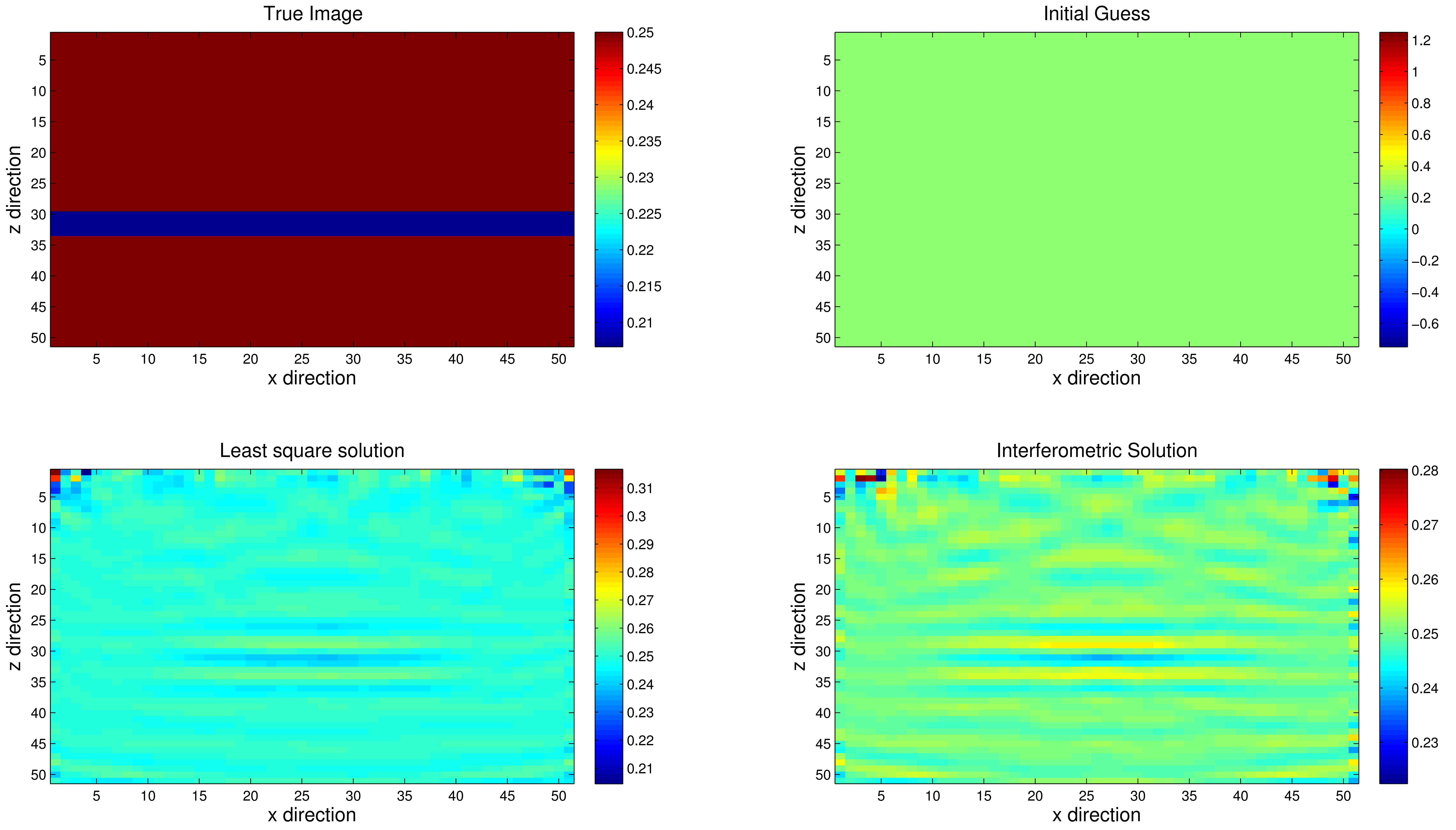


Figure : results after 10 iterations. 3 sources on the left, 3 sources on the right, and 400 receivers in the middle, frequencies in use: [20,25,30] ,$\lambda = 100$.

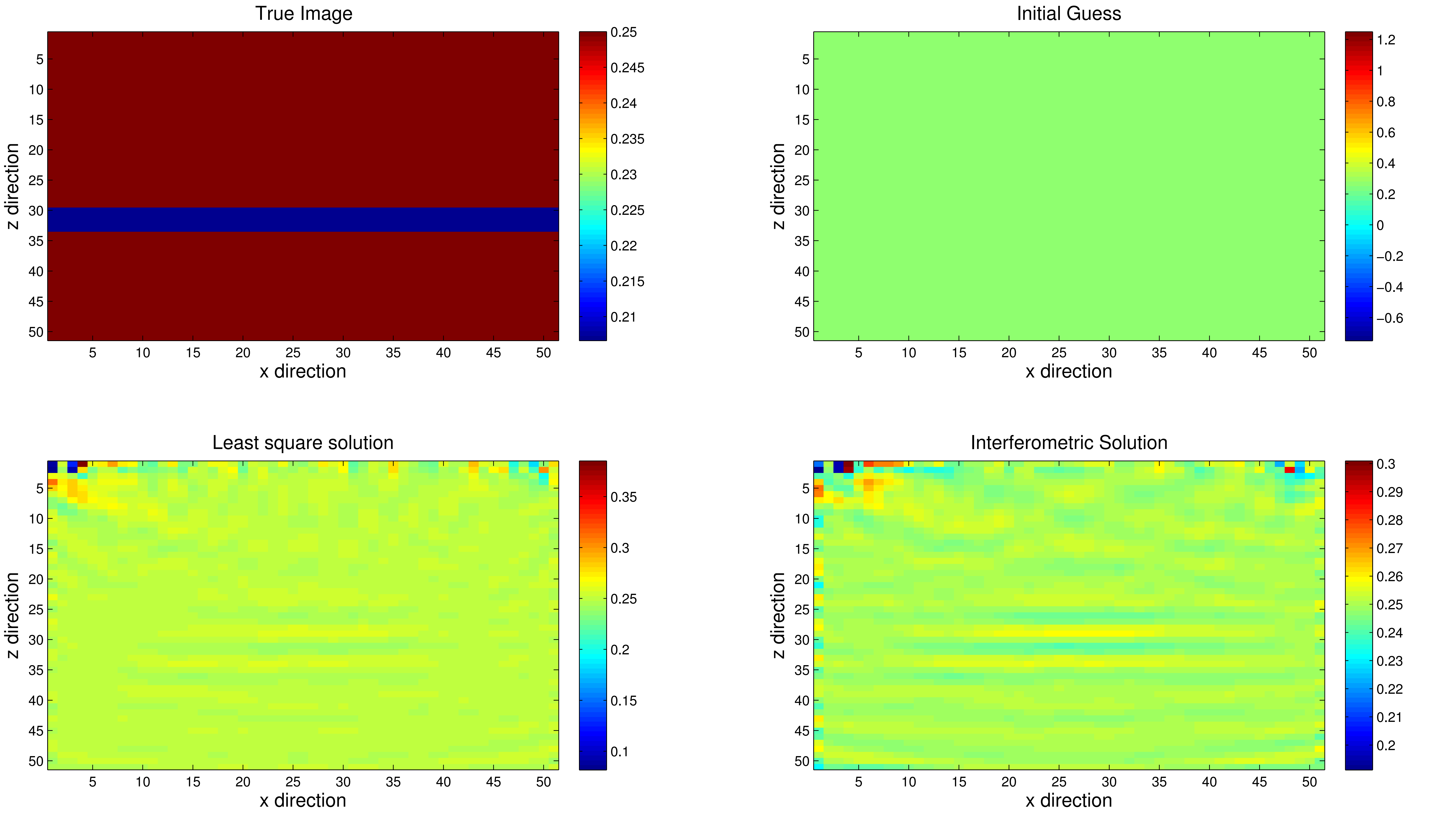


Figure : results after 20 iterations. 3 sources on the left, 3 sources on the right, and 400 receivers in the middle, frequencies in use: [20,25,30] ,$\lambda = 100$.

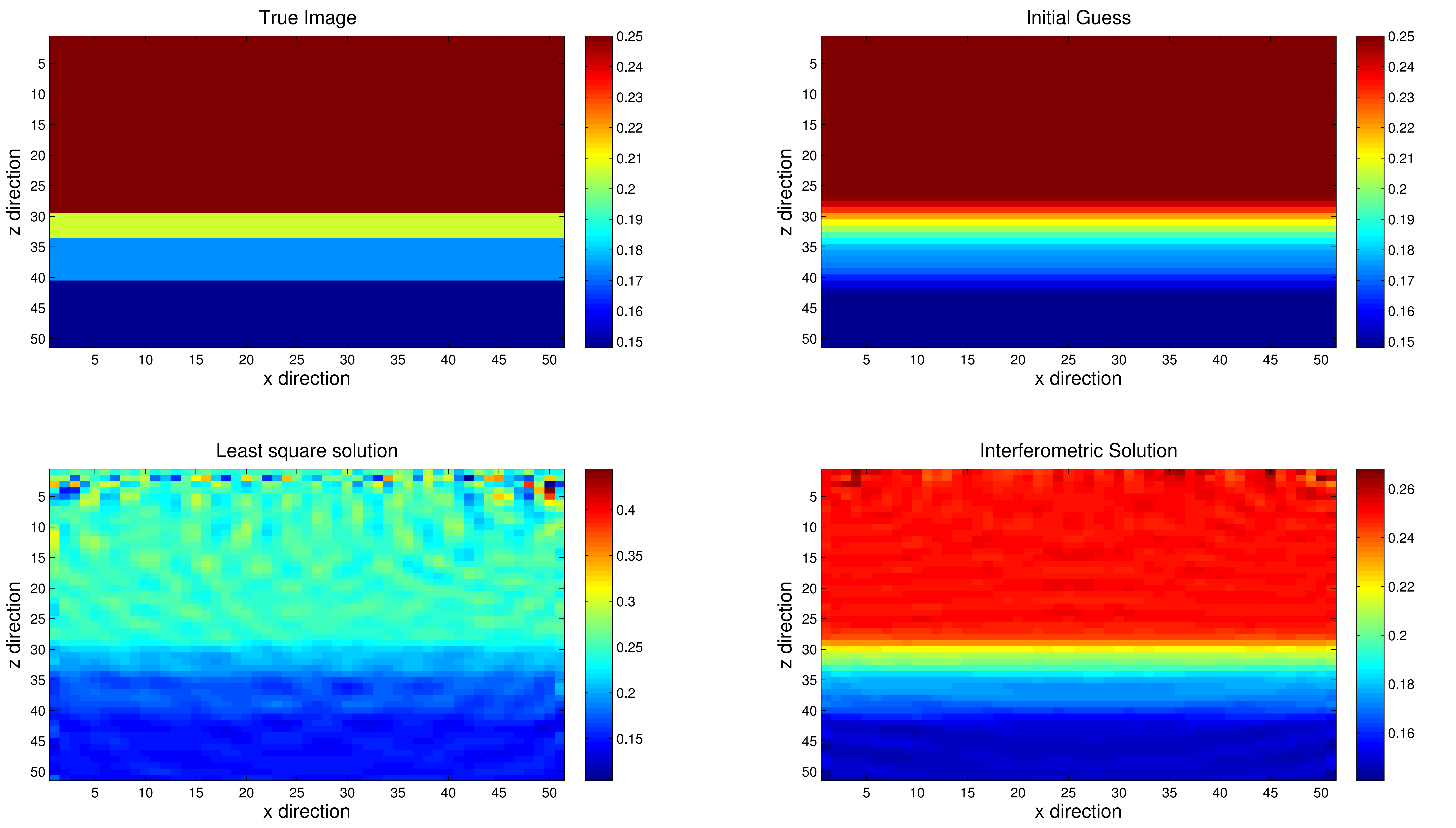


Figure : results when the algorithm stops. 3 sources on the left, 3 sources on the right, and 400 receivers in the middle, frequencies in use: [20,25,30], $\lambda = 10$. $e_{LS} \approx 3.5 e_{IF}$.

Future work

- Explore ways to reduce the computational cost;
- Extend the method to linear model;
- Extend the incoherent interferometric to coherent ones.

Acknowledgement

Thank you for your attention!