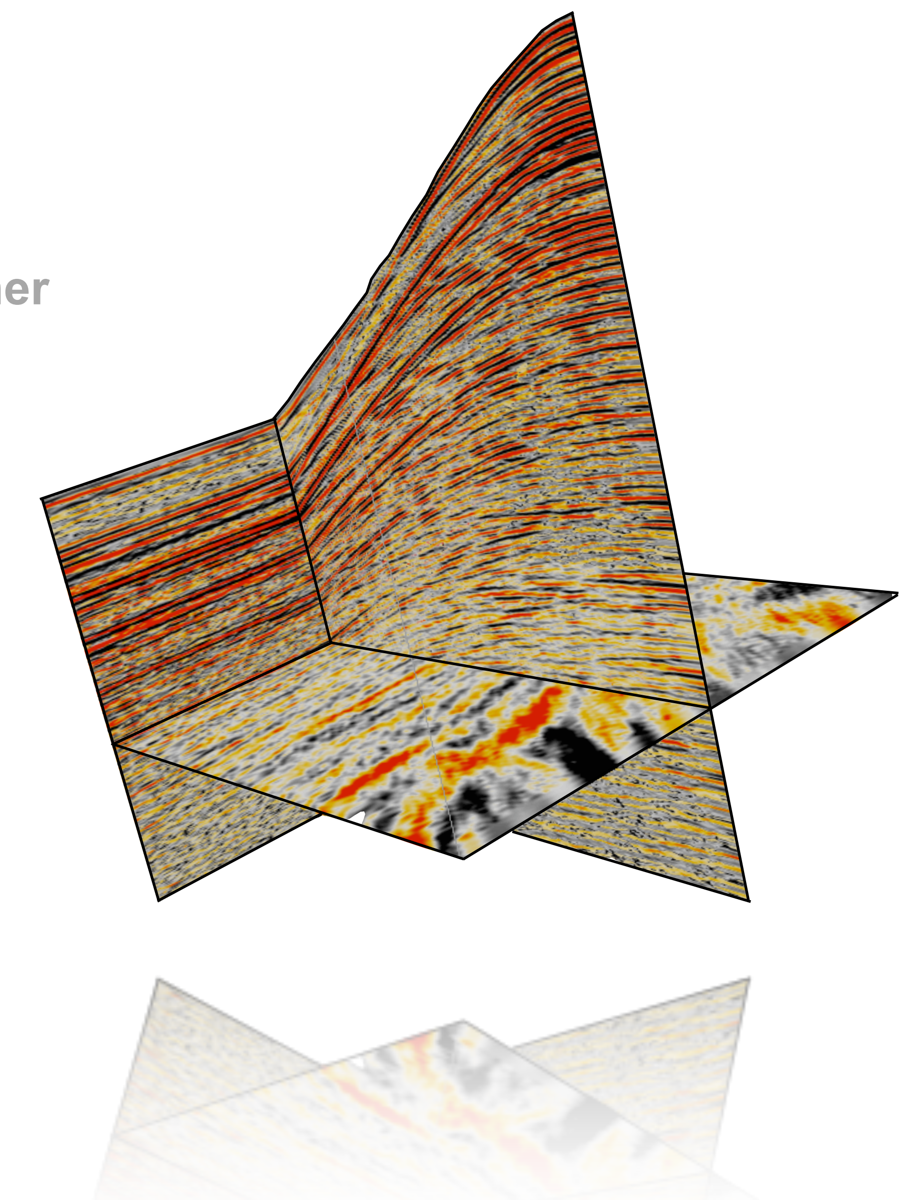
Full-waveform inversion

Felix J. Herrmann*
fherrmann@eos.ubc.ca

with help from

Tristan van Leeuwen, Sasha Aravkin, and other members of the SLIM team

***Seismic Laboratory for Imaging & Modeling**
Department of Earth & Ocean Sciences
The University of British Columbia



Golub Summer School
Vancouver, July 4-15, 2011

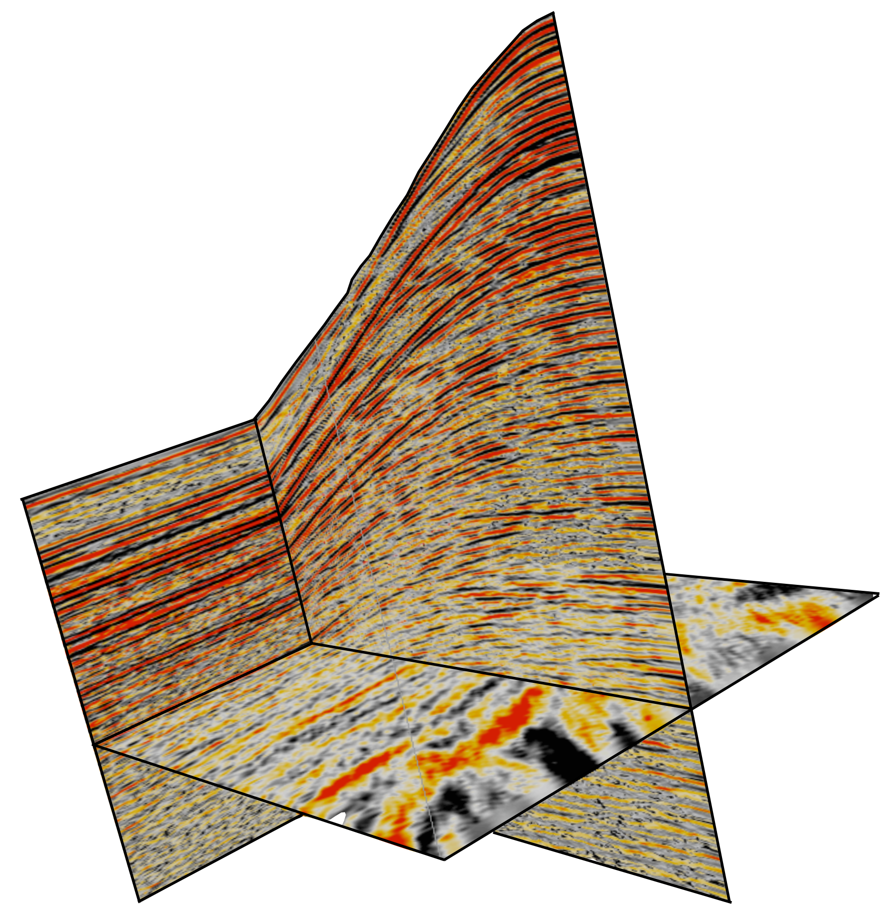
Outline

- Derive single-source monochromatic formulation of PDE constrained optimization
  - Lagrangian formulation
  - adjoint-state method
- Extend to multi-source and multi-frequency
  - multiple source
  - multiple source & multiple frequency
  - Gauss-Newton
- Discuss current-day cutting edge developments/applications of FWI
  - stochastic optimization
  - modified Gauss-Newton with sparsity promotion
- Related problems
  - source calibration
  - free surface

PDE-contrained optimization

Felix J. Herrmann*
fherrmann@eos.ubc.ca

***Seismic Laboratory for Imaging & Modeling**
Department of Earth & Ocean Sciences
The University of British Columbia



Golub Summer School
Vancouver, July 4-15, 2011

PDE-constrained optimization (monochromatic)

$$\min_{\mathbf{m},\mathbf{u}} \quad \frac{1}{2}\|\mathbf{P}\mathbf{u}-\mathbf{d}\|_2^2 \quad \text{subject to} \quad \mathbf{H}[\mathbf{m}]\mathbf{u}=\mathbf{q}$$

Variable	Type	Dimension	Description
n_x	$\mathbb{Z}_+$	1	Number of grid points in x
n_z	$\mathbb{Z}_+$	1	Number of grid points in z
n_r	$\mathbb{Z}_+$	1	Number of receivers
$\mathbf{m}$	$\mathbb{R}$	$n_x n_z$	Model (slowness squared)
$\mathbf{H}[\mathbf{m}]$	$\mathbb{C}$	$n_x n_z \times n_x n_z$	Discrete Helmholz with boundary
$\mathbf{P}$	$\mathbb{R}$	$n_r \times n_x n_z$	Sampling operator
$\mathbf{d}$	$\mathbb{C}$	n_r	Data vector
$\mathbf{q}$	$\mathbb{C}$	$n_x n_z$	Source
$\mathbf{u}$	$\mathbb{C}$	$n_x n_z$	Wavefield
$\mathbf{v}$	$\mathbb{C}$	$n_x n_z$	Adjoint Wavefield

Unconstrained formulation

$$\min_{\mathbf{m}} \quad \phi(\mathbf{m}) := \frac{1}{2}\| \overbrace{\mathbf{PH}[\mathbf{m}]^{-1}\mathbf{q}}^{\mathcal{F}[\mathbf{m},\mathbf{q}]} - \mathbf{d}\|_2^2$$

- interested in deriving the gradient for optimization

$$\mathbf{m}^{k+1} = \mathbf{m}^k - \gamma \nabla \phi(\mathbf{m}^k)$$

- matrix-free Jacobian ($\mathbf{J} = \nabla \mathcal{F}[\mathbf{m}, \mathbf{q}]$)

$$\nabla \phi(\mathbf{m}) = (\nabla \mathcal{F}[\mathbf{m}, \mathbf{q}])^* (\mathcal{F}[\mathbf{m}, \mathbf{q}] - \mathbf{d})$$

Gradient of the Lagrangian

- Adjoint formulation using the Lagrangian

$$\mathcal{L}(\mathbf{v},\mathbf{u},\mathbf{m}) := \frac{1}{2}\|\mathbf{P}\mathbf{u}-\mathbf{d}\|_2^2 + \mathbf{v}^*(\mathbf{H}[\mathbf{m}]\mathbf{u}-\mathbf{q})$$

- Gradient

$$\partial_{\mathbf{v}}\mathcal{L} = \mathbf{H}[\mathbf{m}]\mathbf{u}-\mathbf{q}$$

$$\partial_{\mathbf{u}}\mathcal{L} = \mathbf{P}^T(\mathbf{P}\mathbf{u}-\mathbf{d}) + \mathbf{H}[\mathbf{m}]^*\mathbf{v}$$

$$\partial_{\mathbf{m}_i}\mathcal{L} = \mathbf{v}^*\frac{\partial\mathbf{H}[\mathbf{m}]}{\partial\mathbf{m}_i}\mathbf{u}$$

- Put to zero top two equations uniquely defines **u, v**
- Solutions depend smoothly on **m**

Gradient calculation

$$f(\mathbf{m}) = \mathcal{L}(\mathbf{v}(\mathbf{m}), \mathbf{u}(\mathbf{m}), \mathbf{m})$$

where $\mathbf{v}(\mathbf{m})$ and $\mathbf{u}(\mathbf{m})$ are the solutions to $\partial_{\mathbf{v}}\mathcal{L} = \partial_{\mathbf{u}}\mathcal{L} = 0$

For a fixed value of $\mathbf{m}$, define

$$\begin{aligned} \bar{\mathbf{u}} &= \mathbf{H}[\mathbf{m}]^{-\mathbf{1}}\mathbf{q} \\ \bar{\mathbf{v}} &= -\mathbf{H}[\mathbf{m}]^{-*}\mathbf{P}^T(\mathbf{P}\bar{\mathbf{u}} - \mathbf{d}) \end{aligned}$$

$$\begin{aligned} \frac{d}{d\mathbf{m}} f(\mathbf{m}) &= \partial_{\mathbf{v}}\mathcal{L}(\bar{\mathbf{v}}, \bar{\mathbf{u}}, \mathbf{m})\frac{d\mathbf{v}}{d\mathbf{m}} + \partial_{\mathbf{u}}\mathcal{L}(\bar{\mathbf{v}}, \bar{\mathbf{u}}, \mathbf{m})\frac{d\mathbf{u}}{d\mathbf{m}} + \partial_{\mathbf{m}}\mathcal{L}(\bar{\mathbf{v}}, \bar{\mathbf{u}}, \mathbf{m}) \\ &= \partial_{\mathbf{m}}\mathcal{L}(\bar{\mathbf{v}}, \bar{\mathbf{u}}, \mathbf{m}) \end{aligned}$$

Gradient calculation cont'ed

$$
\begin{aligned}
f(\mathbf{m}) &= \tfrac{1}{2}\|\mathbf{P}\bar{\mathbf{u}} - \mathbf{d}\|_2^2 \\
&= \tfrac{1}{2}\|\mathbf{P}\mathbf{H}[\mathbf{m}]^{-\mathbf{1}}\mathbf{q} - \mathbf{d}\|_2^2 \\
&= \phi(\mathbf{m}) \ ,
\end{aligned}
$$

- corresponds to the unconstrained objective
- We obtain

$$\partial_{\mathbf{m}_i}\phi(\mathbf{m}) = \partial_{\mathbf{m}_i}\mathcal{L}(\bar{\mathbf{u}}, \bar{\mathbf{v}}, \mathbf{m}) = \bar{\mathbf{v}}^* \frac{\partial \mathbf{H}[\mathbf{m}]}{\partial \mathbf{m}_i}\bar{\mathbf{u}}$$

- or

$$\nabla\phi(\mathbf{m}) = \omega^2 \mathrm{diag}(\bar{\mathbf{u}}\bar{\mathbf{v}}^*)$$

Gradient calculation cont'ed

$$
\begin{aligned}
f(\mathbf{m}) &= \tfrac{1}{2}\|\mathbf{P}\bar{\mathbf{u}} - \mathbf{d}\|_2^2 \\
&= \tfrac{1}{2}\|\mathbf{P}\mathbf{H}[\mathbf{m}]^{-1}\mathbf{q} - \mathbf{d}\|_2^2 \\
&= \phi(\mathbf{m})\ ,
\end{aligned}
$$

- corresponds to the unconstrained objective
- We obtain

$$\partial_{\mathbf{m}_i}\phi(\mathbf{m}) = \partial_{\mathbf{m}_i}\mathcal{L}(\bar{\mathbf{u}}, \bar{\mathbf{v}}, \mathbf{m}) = \bar{\mathbf{v}}^* \frac{\partial \mathbf{H}[\mathbf{m}]}{\partial \mathbf{m}_i}\bar{\mathbf{u}}$$

- or

zero-'offset' imaging condition

$$\nabla\phi(\mathbf{m}) = \omega^2 \mathrm{diag}(\bar{\mathbf{u}}\bar{\mathbf{v}}^*)$$

Multisource FWI

- Constrained formulation

$$\min_{\mathbf{m},\mathbf{U}} \quad \frac{1}{2}\|\boldsymbol{\mathcal{P}}(\mathbf{U})-\mathbf{D}\|_F^2 \quad \text{subject to} \quad \mathbf{H}[\mathbf{m}]\mathbf{U}=\mathbf{Q}$$

- Unconstrained formulation

$$\min_{\mathbf{m}} \quad \phi(\mathbf{m}) := \frac{1}{2}\|\boldsymbol{\mathcal{P}}(\mathbf{H}[\mathbf{m}]^{-\mathbf{1}}\mathbf{Q})-\mathbf{D}\|_F^2$$

- Lagrangian

$$\mathcal{L}(\mathbf{V},\mathbf{U},\mathbf{m}) := \frac{1}{2}\|\boldsymbol{\mathcal{P}}(\mathbf{U})-\mathbf{D}\|_F^2 + \mathrm{tr}\left(\mathbf{V}^*(\mathbf{H}[\mathbf{m}]\mathbf{U}-\mathbf{Q})\right)$$

- Matrix-free Jacobian

$$\nabla\phi(\mathbf{m}) = (\nabla\boldsymbol{\mathcal{F}}[\mathbf{m},\mathbf{Q}])^*(\boldsymbol{\mathcal{F}}[\mathbf{m},\mathbf{Q}]-\mathbf{D})$$

Multisource FWI cont'nd

Variable	Type	Dimension	Description
n_x	$\mathbb{Z}_+$	1	Number of grid points in x
n_z	$\mathbb{Z}_+$	1	Number of grid points in z
n_r	$\mathbb{Z}_+$	1	Number of receivers
n_s	$\mathbb{Z}_+$	1	Number of sources
n_f	$\mathbb{Z}_+$	1	Number of frequencies
$\mathbf{m}$	$\mathbb{R}$	$n_x n_z$	Model (slowness squared)
$\mathbf{H}_{\boldsymbol{\omega}}[\mathbf{m}]$	$\mathbb{C}$	$n_x n_z \times n_x n_z$	Discrete Helmholz with boundary for ω
$\mathbf{H}[\mathbf{m}]$	$\mathbb{C}$	$n_f(n_x n_z \times n_x n_z)$	$\mathrm{diag}[\mathbf{H}_{\omega_1}[\mathbf{m}], \ldots, \mathbf{H}_{\omega_{n_f}}[\mathbf{m}]]$
$\mathbf{D}_\omega$	$\mathbb{C}$	$n_r \times n_s$	Data vector for ω
$\mathbf{D}$	$\mathbb{C}$	$n_f(n_r \times n_s)$	$\mathrm{stack}[\mathbf{D}_{\omega_1}, \ldots, \mathbf{D}_{\omega_{n_f}}]$
$\boldsymbol{\mathcal{P}}_f$	$\mathbb{R}$	$n_x n_z \times n_s \rightarrow n_r \times n_s$	Sampling operator
$\boldsymbol{\mathcal{P}}$	$\mathbb{R}$	$n_f(n_x n_z \times n_s) \rightarrow n_f(n_r \times n_s)$	Applies $\boldsymbol{\mathcal{P}}_f$ to each frequency
$\mathbf{Q}_\omega$	$\mathbb{C}$	$n_x n_z \times n_s$	Source for frequency ω
$\mathbf{Q}$	$\mathbb{C}$	$n_f(n_x n_z \times n_s)$	$\mathrm{stack}[\mathbf{Q}_{\omega_1}, \ldots, \mathbf{Q}_{\omega_{n_f}}]$
$\mathbf{U}_\omega$	$\mathbb{C}$	$n_x n_z \times n_s$	Wavefield for frequency ω
$\mathbf{U}$	$\mathbb{C}$	$n_\omega(n_x n_z \times n_s)$	$\mathrm{stack}[\mathbf{U}_{\omega_1}, \ldots, \mathbf{U}_{\omega_{n_f}}]$
$\mathbf{V}_\omega$	$\mathbb{C}$	$n_x n_z \times n_s$	Adjoint wavefield for frequency ω
$\mathbf{V}$	$\mathbb{C}$	$n_\omega(n_x n_z \times n_s)$	$\mathrm{stack}[\mathbf{V}_{\omega_1}, \ldots, \mathbf{V}_{\omega_{n_f}}]$

Multisource FWI cont'nd

- gradient of the Lagrangian

$$\partial_{\mathbf{V}}\mathcal{L} = \mathbf{H}[\mathbf{m}]\mathbf{U} - \mathbf{Q}$$

$$\partial_{\mathbf{U}}\mathcal{L} = \mathcal{P}^*(\mathcal{P}(\mathbf{U}) - \mathbf{D}) + \mathbf{H}[\mathbf{m}]^*\mathbf{V}$$

$$\partial_{\mathbf{m}_i}\mathcal{L} = \mathrm{tr}\left(\mathbf{V}^*\frac{\partial\mathbf{H}[\mathbf{m}]}{\partial\mathbf{m}_i}\mathbf{U}\right)$$

- or

$$\nabla\phi(\mathbf{m}) = \sum_{\omega}\omega^2\mathrm{diag}(\mathbf{U}\mathbf{V}^*)$$

- Corresponds to reverse-time migration

$$\nabla\phi(\mathbf{m}) = \underbrace{(\nabla\mathcal{F}[\mathbf{m},\mathbf{Q}])^*}_{\text{migration}}\ \overbrace{(\mathcal{F}[\mathbf{m},\mathbf{Q}] - \mathbf{D})}^{\text{residue/linearized data}}$$

Multisource FWI cont'nd

- gradient of the Lagrangian

$$\partial_{\mathbf{V}}\mathcal{L} = \mathbf{H}[\mathbf{m}]\mathbf{U} - \mathbf{Q}$$

$$\partial_{\mathbf{U}}\mathcal{L} = \mathcal{P}^*(\mathcal{P}(\mathbf{U}) - \mathbf{D}) + \mathbf{H}[\mathbf{m}]^*\mathbf{V}$$

$$\partial_{\mathbf{m}_i}\mathcal{L} = \mathrm{tr}\left(\mathbf{V}^*\frac{\partial \mathbf{H}[\mathbf{m}]}{\partial \mathbf{m}_i}\mathbf{U}\right)$$

- or

zero-'offset' imaging condition

$$\nabla\phi(\mathbf{m}) = \sum_{\omega}\omega^2 \mathrm{diag}(\mathbf{U}\mathbf{V}^*)$$

- Corresponds to reverse-time migration

$$\nabla\phi(\mathbf{m}) = \underbrace{(\nabla\mathcal{F}[\mathbf{m},\mathbf{Q}])^*}_{\text{migration}} \overbrace{(\mathcal{F}[\mathbf{m},\mathbf{Q}] - \mathbf{D})}^{\text{residue/linearized data}}$$

Multisource FWI cont'nd

- gradient of the Lagrangian

$$\partial_{\mathbf{V}}\mathcal{L} = \mathbf{H}[\mathbf{m}]\mathbf{U} - \mathbf{Q}$$

$$\partial_{\mathbf{U}}\mathcal{L} = \boldsymbol{\mathcal{P}}^*(\boldsymbol{\mathcal{P}}(\mathbf{U}) - \mathbf{D}) + \mathbf{H}[\mathbf{m}]^*\mathbf{V}$$

$$\partial_{\mathbf{m}_i}\mathcal{L} = \mathrm{tr}\left(\mathbf{V}^*\frac{\partial\mathbf{H}[\mathbf{m}]}{\partial\mathbf{m}_i}\mathbf{U}\right)$$

- or

zero-'offset' imaging condition

$$\nabla\phi(\mathbf{m}) = \sum_{\omega}\omega^2\mathrm{diag}(\mathbf{U}\mathbf{V}^*)$$

zero-'time' imaging condition

- Corresponds to reverse-time migration

$$\nabla\phi(\mathbf{m}) = \underbrace{(\nabla\boldsymbol{\mathcal{F}}[\mathbf{m},\mathbf{Q}])^*}_{\text{migration}}\overbrace{(\boldsymbol{\mathcal{F}}[\mathbf{m},\mathbf{Q}] - \mathbf{D})}^{\text{residue/linearized data}}$$

Gauss-Newton

Algorithm 1: Gauss Newton

Result: Output estimate for the model $\mathbf{m}$

$\mathbf{m} \longleftarrow \mathbf{m}_0$; $k \longleftarrow 0$; `// initial model`

while not converged **do**

$\delta\mathbf{m}^k \longleftarrow \arg\min_{\delta\mathbf{m}} \frac{1}{2}\|\mathbf{D} - \mathcal{F}[\mathbf{m}^k;\mathbf{Q}] - \nabla\mathcal{F}[\mathbf{m}^k;\mathbf{Q}]\delta\mathbf{m}\|_F^2$

$\mathbf{m}^{k+1} \longleftarrow \mathbf{m}^k + \gamma^k \delta\mathbf{m}^k$; `// update with linesearch`

$k \longleftarrow k+1$;

end

Evaluation of $\nabla\mathcal{F}^H[\mathbf{m};\mathbf{Q}]$ and $\nabla\mathcal{F}[\mathbf{m};\mathbf{Q}]$ each require ***two*** PDE solves for *each* source & *angular* frequency

Involves inversion of a ***tall*** linear system of equations

Gauss-Newton

Evaluation of $\nabla\mathcal{F}^H[\mathbf{m};\mathbf{Q}]$ and $\nabla\mathcal{F}[\mathbf{m};\mathbf{Q}]$ each require ***two*** PDE solves for *each* source & *angular* frequency

Involves inversion of a ***tall*** linear system of equations

Gauss-Newton

Algorithm 1: Gauss Newton

Result: Output estimate for the model $\mathbf{m}$

$\mathbf{m} \longleftarrow \mathbf{m}_0;\ k \longleftarrow 0$; `// initial model`

while not converged **do**

$$\delta\mathbf{m}^k \longleftarrow \arg\min_{\delta\mathbf{m}} \frac{1}{2} \| \overbrace{\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m}^k;\mathbf{Q}]}^{\mathbf{b}} - \underbrace{\nabla\boldsymbol{\mathcal{F}}[\mathbf{m}^k;\mathbf{Q}]\delta\mathbf{m}}_{\mathbf{Ax}} \|_F^2$$

$\mathbf{m}^{k+1} \longleftarrow \mathbf{m}^k + \gamma^k \delta\mathbf{m}^k$; `// update with linesearch`

$k \longleftarrow k+1;$

end

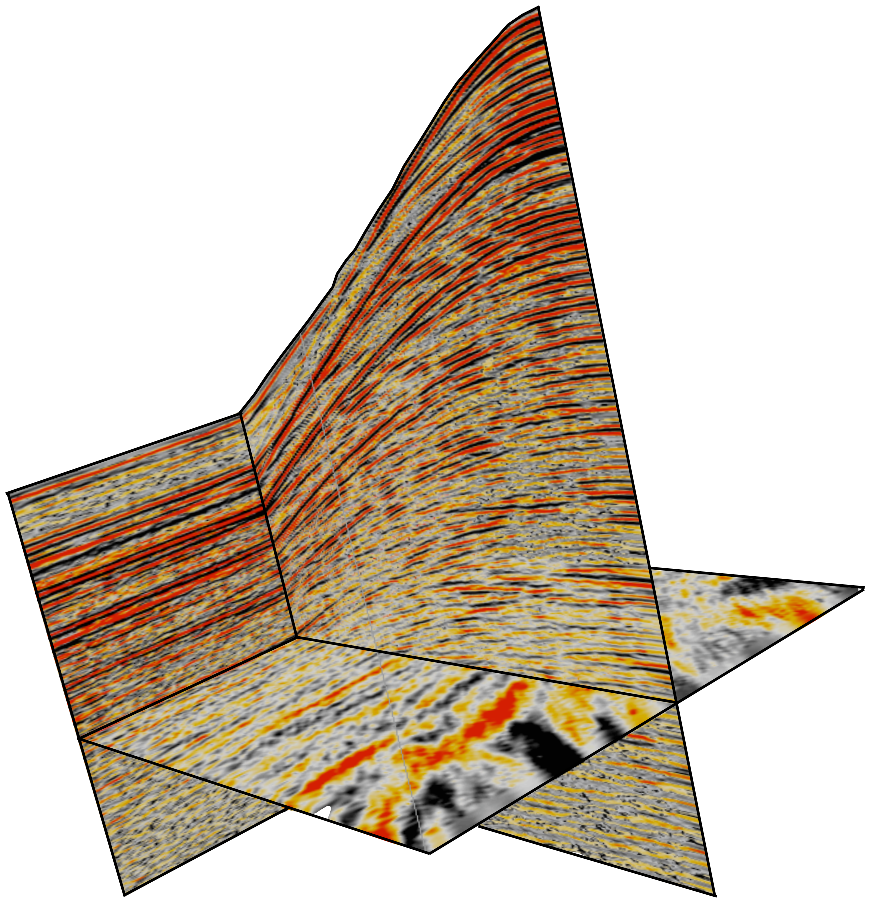
Evaluation of $\nabla\boldsymbol{\mathcal{F}}^H[\mathbf{m};\mathbf{Q}]$ and $\nabla\boldsymbol{\mathcal{F}}[\mathbf{m};\mathbf{Q}]$ each require ***two*** PDE solves for *each* source & *angular* frequency

Involves inversion of a ***tall*** linear system of equations

Developments/applications FWI

Felix J. Herrmann*
fherrmann@eos.ubc.ca

***Seismic Laboratory for Imaging & Modeling**
Department of Earth & Ocean Sciences
The University of British Columbia



Golub Summer School
Vancouver, July 4-15, 2011

Stochastic optimization

[Haber, Chung, and FJH, '10]

[Bertsekas, '96, Nemirovsky, '08]

Replace deterministic-optimization problem with *sum* cycling over *different sources & corresponding monochromatic shot records* (columns of D & Q):

$$\min_{\mathbf{m}} \phi(\mathbf{m}) = \frac{1}{N} \sum_{i=1}^{n_s} \frac{1}{2} \|\mathbf{d}_i - \mathcal{F}[\mathbf{m}; \mathbf{q}_i]\|_2^2$$

Stochastic *average* approximation

[Haber, Chung, and FJH, '10]

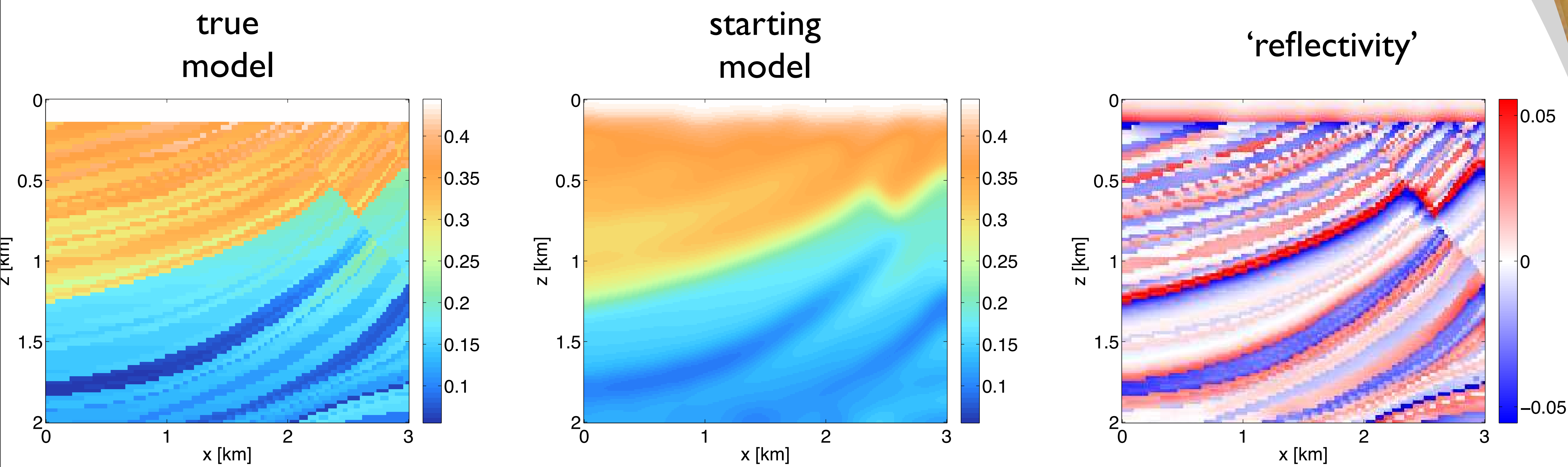
by *a stochastic-optimization* problem (SAA)

$$\begin{aligned}\min_{\mathbf{m}} \mathbf{E}_{\mathbf{w}}\{\phi(\mathbf{m},\mathbf{w}) &= \frac{1}{2}\|\mathbf{D}\mathbf{w} - \mathcal{F}[\mathbf{m};\mathbf{Q}\mathbf{w}]\|_2^2\} \\ &= \min_{\mathbf{m}} \phi(\mathbf{m}) \\ &\approx \min_{\mathbf{m}} \frac{1}{K}\sum_{j=1}^{K}\frac{1}{2}\|\underline{\mathbf{d}}_j - \mathcal{F}[\mathbf{m};\underline{\mathbf{q}}_j]\|_2^2\end{aligned}$$

with $\mathbf{E}_{\mathbf{w}}\{\mathbf{w}\mathbf{w}^H\} = \mathbf{I}$

and $\underline{\mathbf{d}}_j = \mathbf{D}\mathbf{w}_j,\ \underline{\mathbf{q}}_j = \mathbf{Q}\mathbf{w}_j$

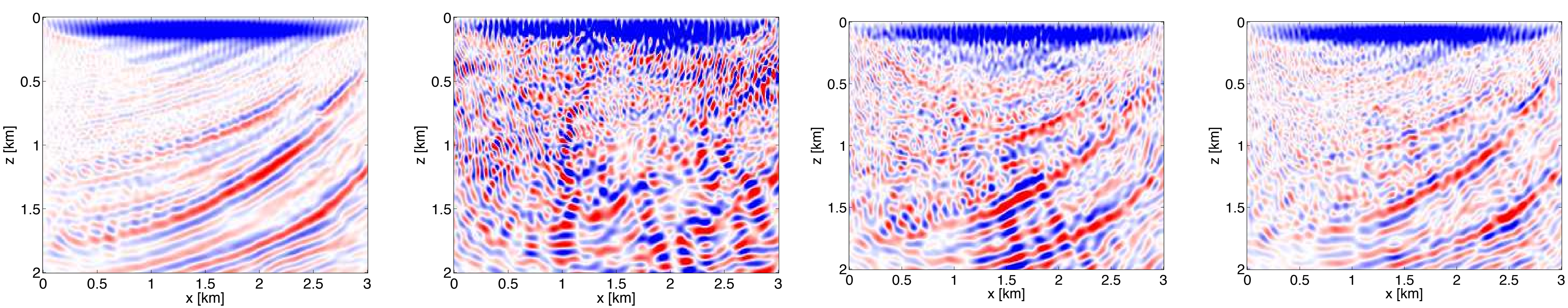
Stylized example



Gradients

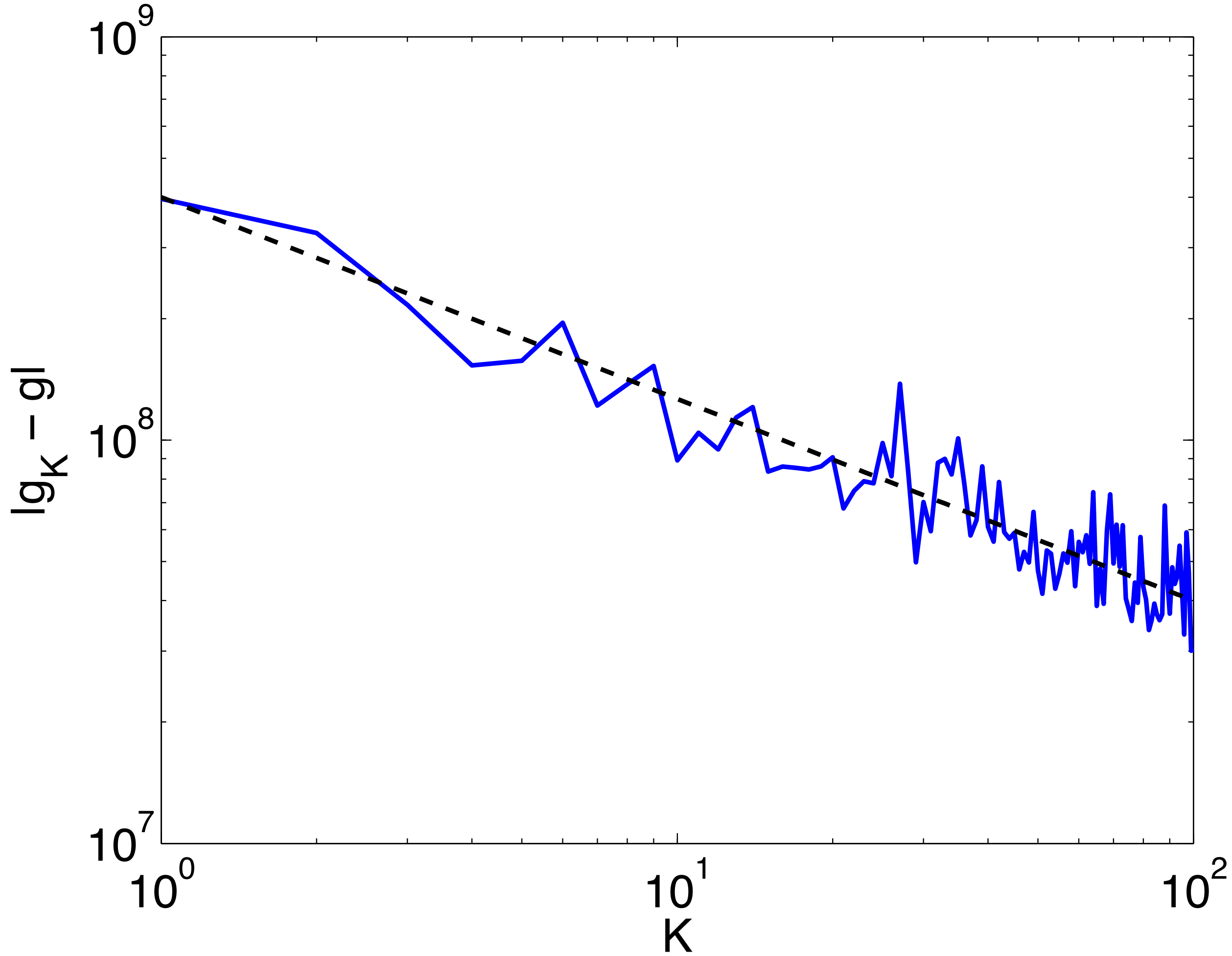
Search direction for *increasing* batch size K:

$$\mathbf{g}_K \approx \frac{1}{K}\sum_{j=1}^{K} \nabla\mathcal{F}^*[\mathbf{m};\underline{\mathbf{q}}_j]\delta\underline{\mathbf{d}}_j$$



full K=1 K=5 K=10

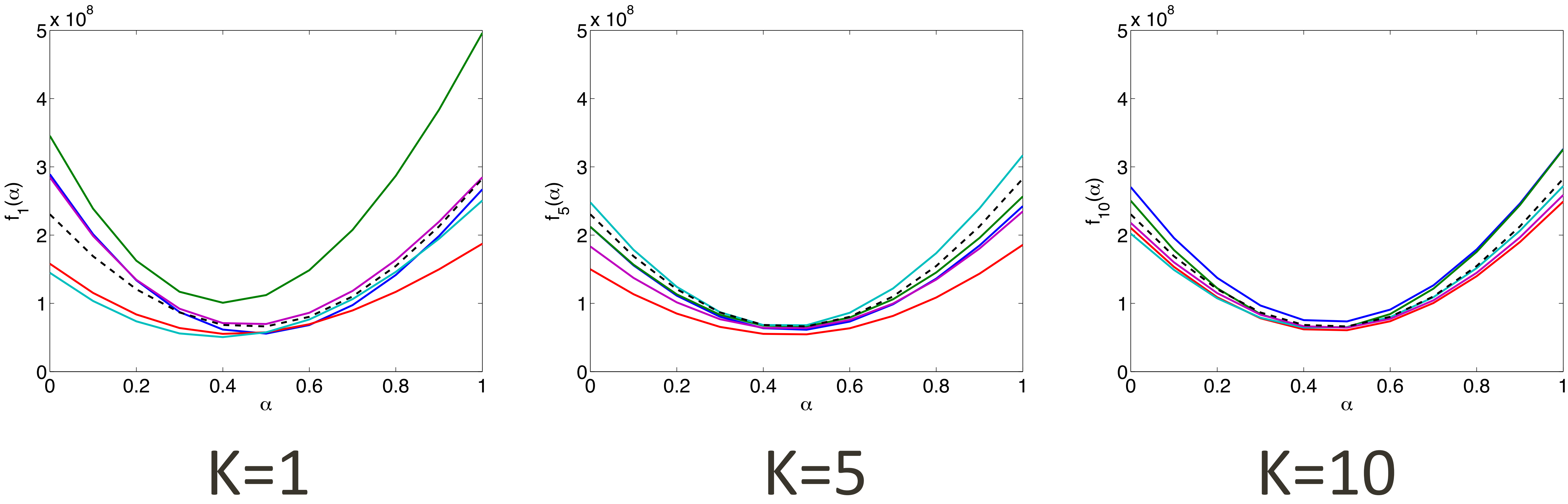
Decay



error between full and sampled gradient

Misfit functional

$$\phi_K(\mathbf{g}_K) = \frac{1}{K}\sum_{j=1}^{K}\frac{1}{2}\|\underline{\mathbf{d}}_j - \mathcal{F}[\mathbf{m} + \alpha\mathbf{g}_K; \underline{\mathbf{q}}_j]\|_2^2$$



[Haber, Chung, and FJH, '10; van Leeuwen, Aravkin, FJH, '10]

SLIM

Stochastic *average* approximation

In the *limit* $K \to \infty$, *stochastic* & *deterministic* formulations are *identical*

Applicable to arbitrary optimization problems of the form

$$\min_{\mathbf{m}} \phi(\mathbf{m}) = \sum_{i=1}^{K} \phi_i(\mathbf{m})$$

We *gain* as long as $K \ll N$ ...

But the error in *Monte-Carlo* methods decays only slowly ($\mathcal{O}(K^{-1/2})$)

Stochastic approximation (SA)

Algorithm 1: Stochastic gradient descent

Result: Output estimate for the model $\mathbf{m}$

$\mathbf{m} \longleftarrow \mathbf{m}_0$; $k \longleftarrow 0$; `// initial model`

while not converged **do**

$\{\underline{\mathbf{d}}^k, \underline{\mathbf{q}}^k\} \longleftarrow \{\mathbf{D}\mathbf{w}^k, \mathbf{Q}\mathbf{w}^k\}$ with $\mathbf{w}^k \in N(0,1)$; `// draw sim. exp.`

$\mathbf{g}^k \longleftarrow \nabla\mathcal{F}^*[\mathbf{m}^{k-1}, \underline{\mathbf{q}}^k](\underline{\mathbf{d}}^k - \mathcal{F}[\mathbf{m}^{k-1}, \underline{\mathbf{q}}^k])$; `// gradient`

$\underline{\mathbf{m}}^{k+1} \longleftarrow \mathbf{m}^k - \gamma^k \mathbf{g}^k$; `// update`

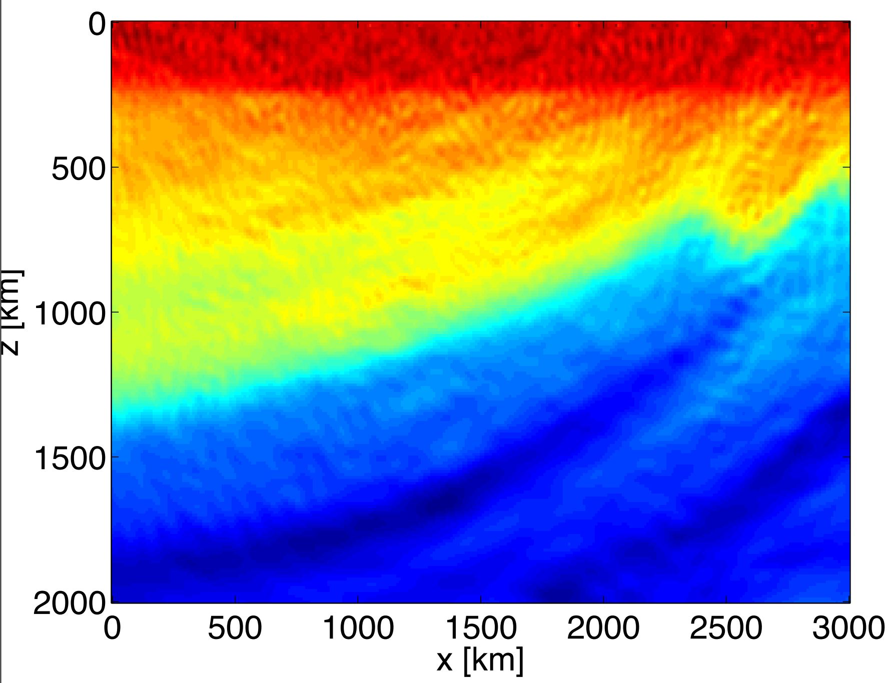
$\mathbf{m}^{k+1} = \frac{1}{k+1}\left(\sum_{i=1}^{k} \mathbf{m}^i + \underline{\mathbf{m}}^{k+1}\right)$; `// average`

$k \longleftarrow k+1$;

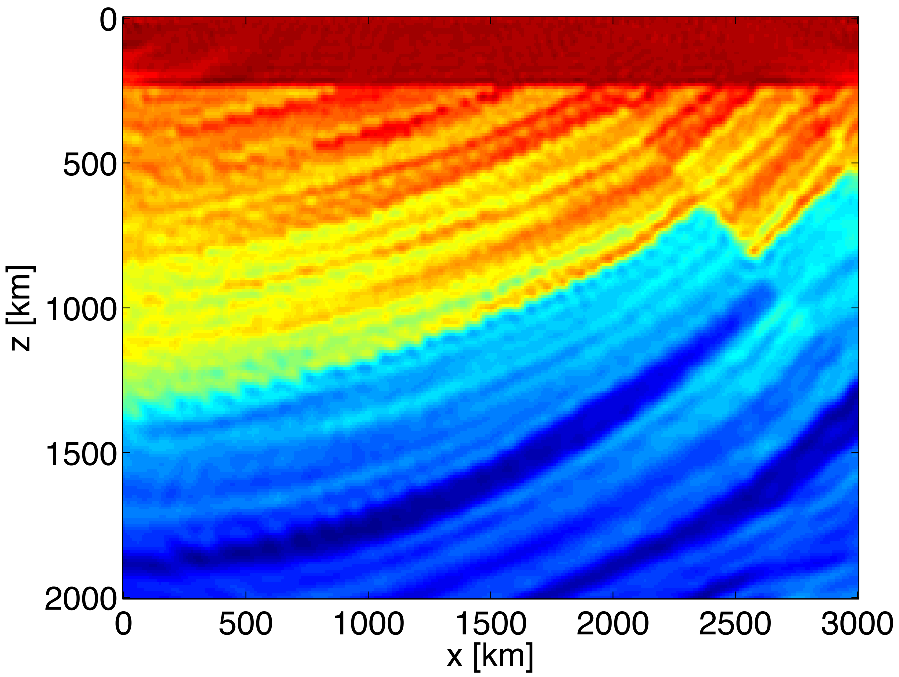
end

[Bertsekas, '96; Haber, Chung, and FJH, '10]

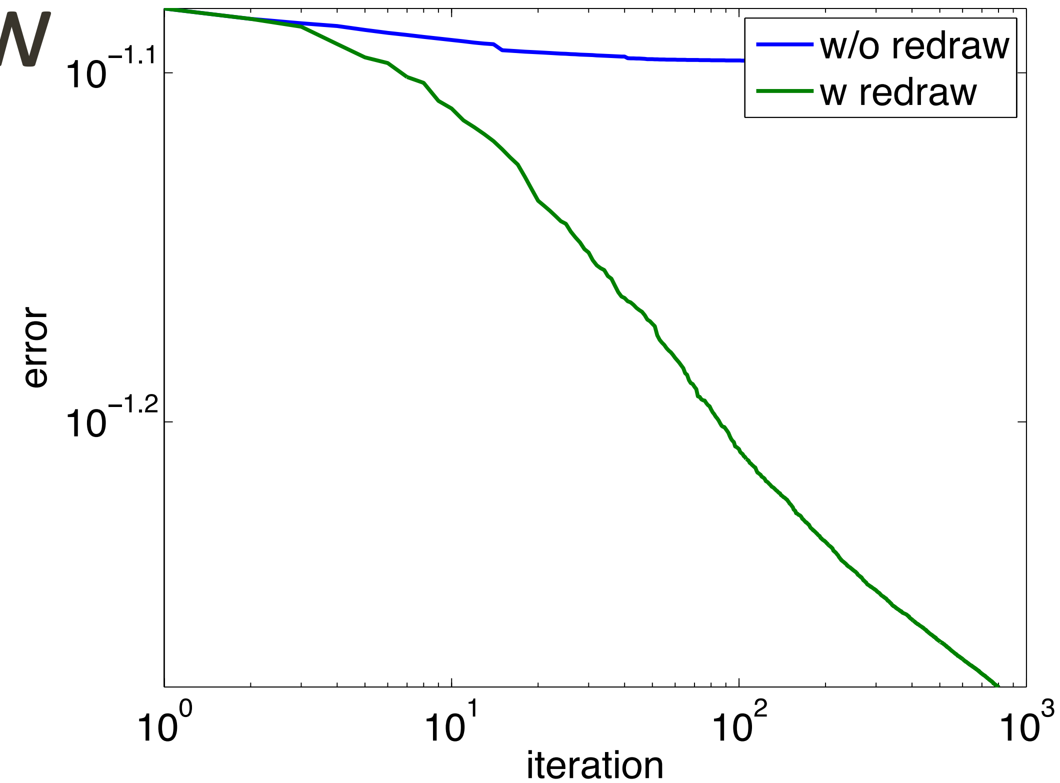
K=1 w and w/o redraw [noise-free case]



w/o redraw



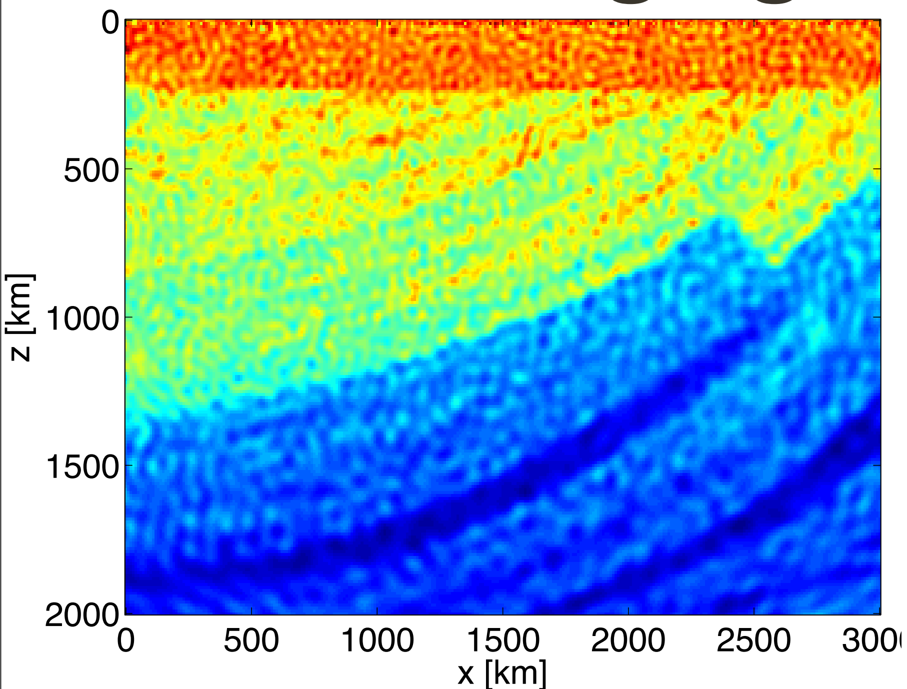
w redraw



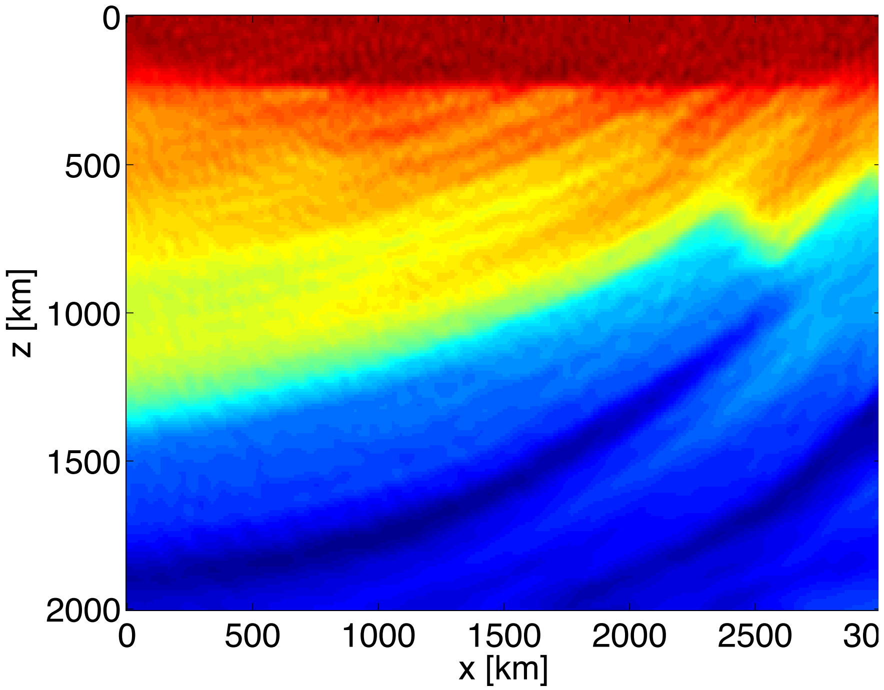
model error K=1

K=1 [noisy case]

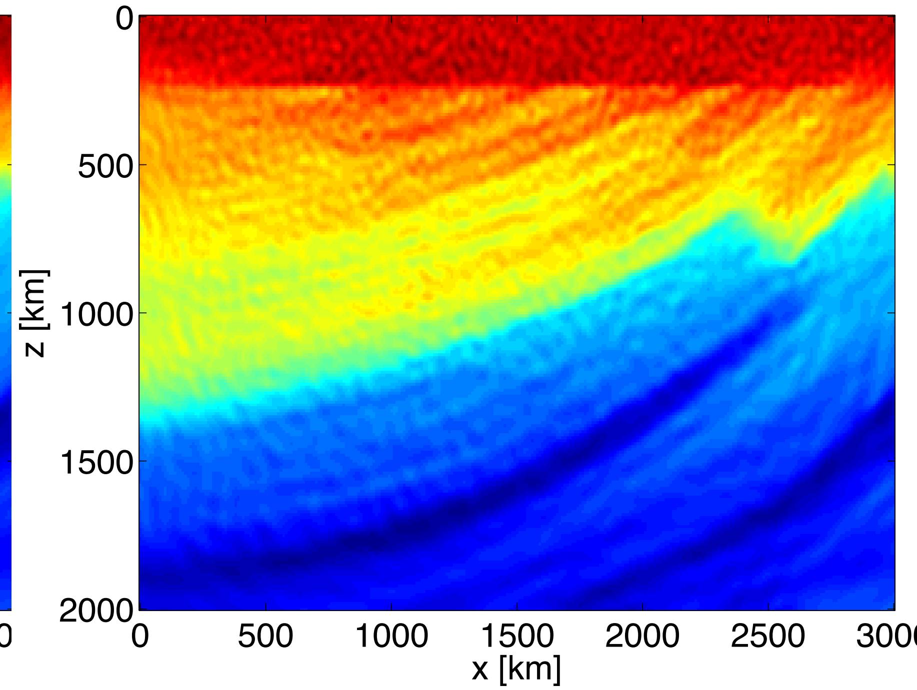
w/o averaging

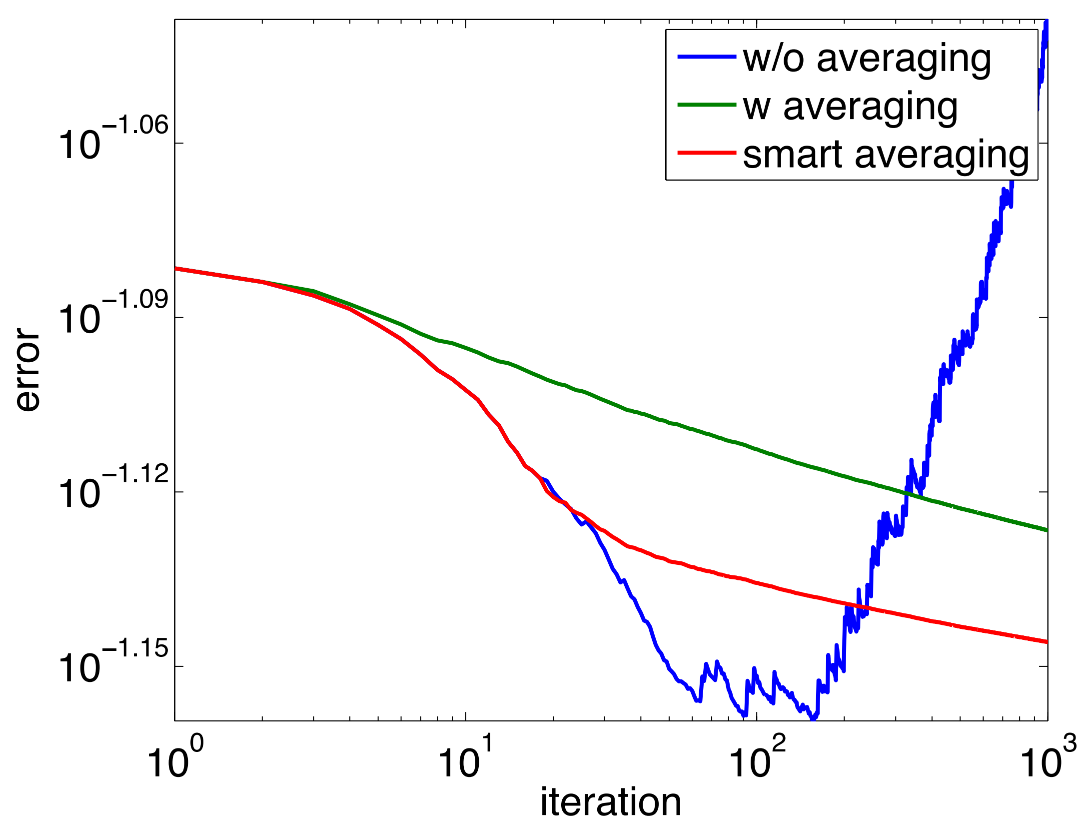


w averaging



smart averaging





Observations

Stochastic-average approximation (**SAA**):

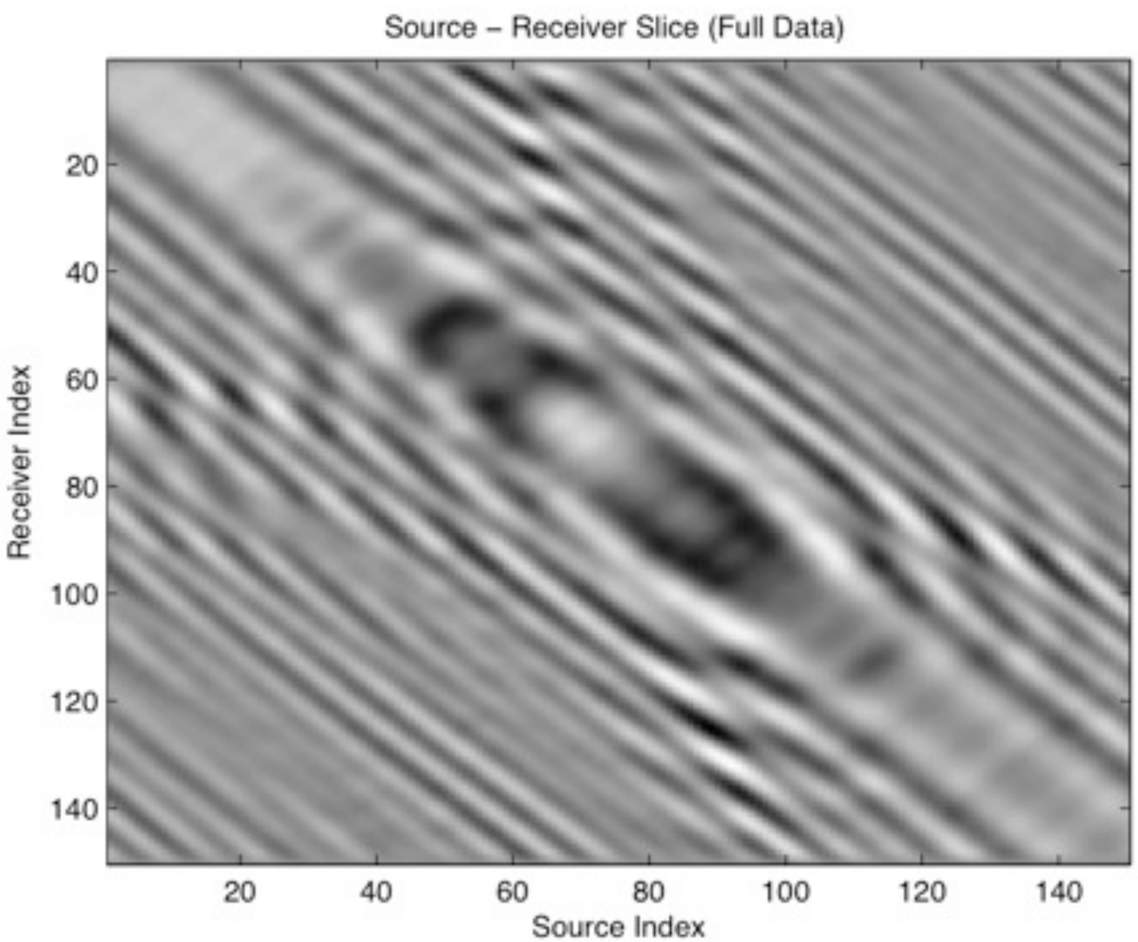
- Error decays slowly with batch size K
- Works for separable optimization problems

Stochastic approximation (**SA**):

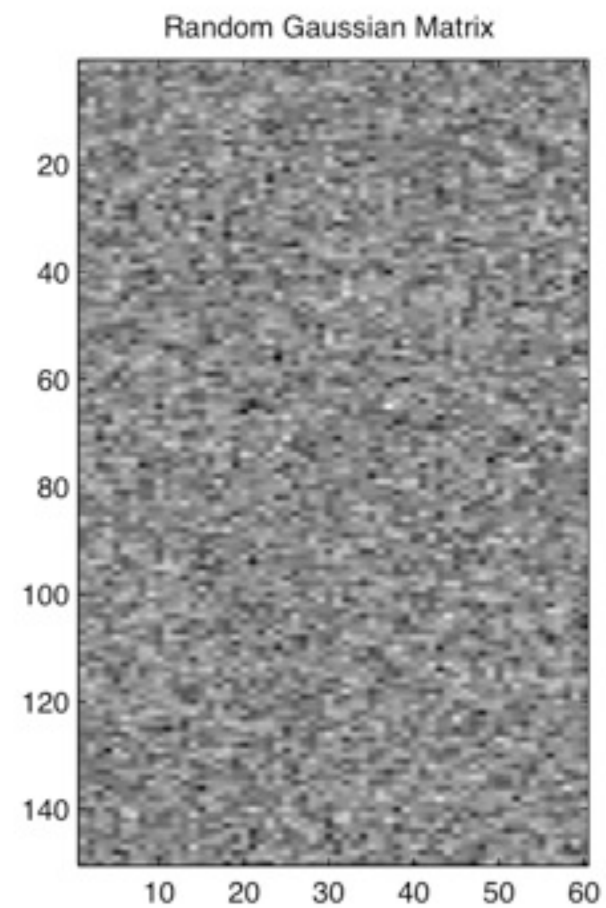
- Renewals improve convergence *significantly*
- Requires *averaging* to remove noise *instability*, which is *detrimental* to the *convergence*

Randomized source superposition

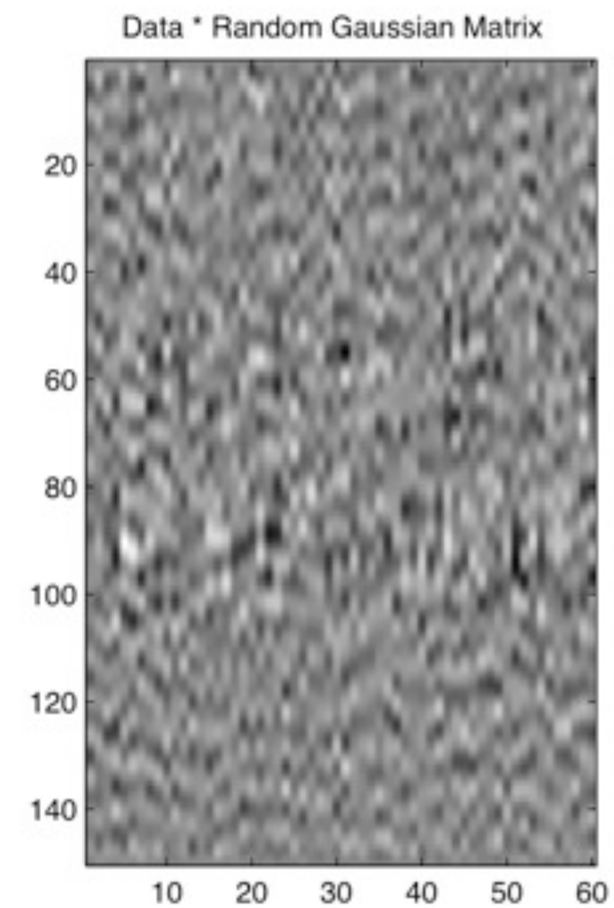
$$[\mathbf{b}_1, \cdots, \mathbf{b}_{n_s}]$$



$$\mathbf{W}$$



$$[\underline{\mathbf{b}}_1, \cdots, \underline{\mathbf{b}}_{n'_s}]$$



Heuristic algorithm [Haber, Chung, and FJH, '10]

Algorithm 1: Stochastic-average approximation with warm starts

```
x_0 ⟵ 0; k ⟵ 0 ;                                  // initialize
while ||x_0 − x̃||_2 ≥ ε do
    k ⟵ k + 1;                                    // increase counter
    x̃ ⟵ x_0;                                     // update warm start
    W ⟵ Draw(W);                                  // draw new subsampler
    x_0 ⟵ Solve(P(W); x̃);                        // solve the subproblem
end
```

$\mathbf{x}_0 \longleftarrow \mathbf{0}; \mathbf{k} \longleftarrow \mathbf{0}$; // initialize

while $\|\mathbf{x}_0 - \widetilde{\mathbf{x}}\|_2 \geq \epsilon$ **do**

- $k \longleftarrow k + 1$; // increase counter
- $\widetilde{\mathbf{x}} \longleftarrow \mathbf{x}_0$; // update warm start
- $\mathbf{W} \longleftarrow \mathrm{Draw}(\mathbf{W})$; // draw new subsampler
- $\mathbf{x}_0 \longleftarrow \mathrm{Solve}(\mathbb{P}(\mathbf{W}); \widetilde{\mathbf{x}})$; // solve the subproblem

end

Subproblems

least-squares migration

$$\mathbb{P}_{\ell_2}(\mathbf{W}^k; \mathbf{x}_0): \quad \min_{\mathbf{x}} \frac{1}{2K} \sum_{j=1}^{K} \|\underline{\mathbf{b}}_j^k - \underline{\mathbf{A}}_j^k \mathbf{x}\|_2^2$$

- solve with *limited* # of iterations of LSQR
- initialize solver with *warm* start
- solves *damped* least-squares problem

Subproblems

sparsity-promoting migration

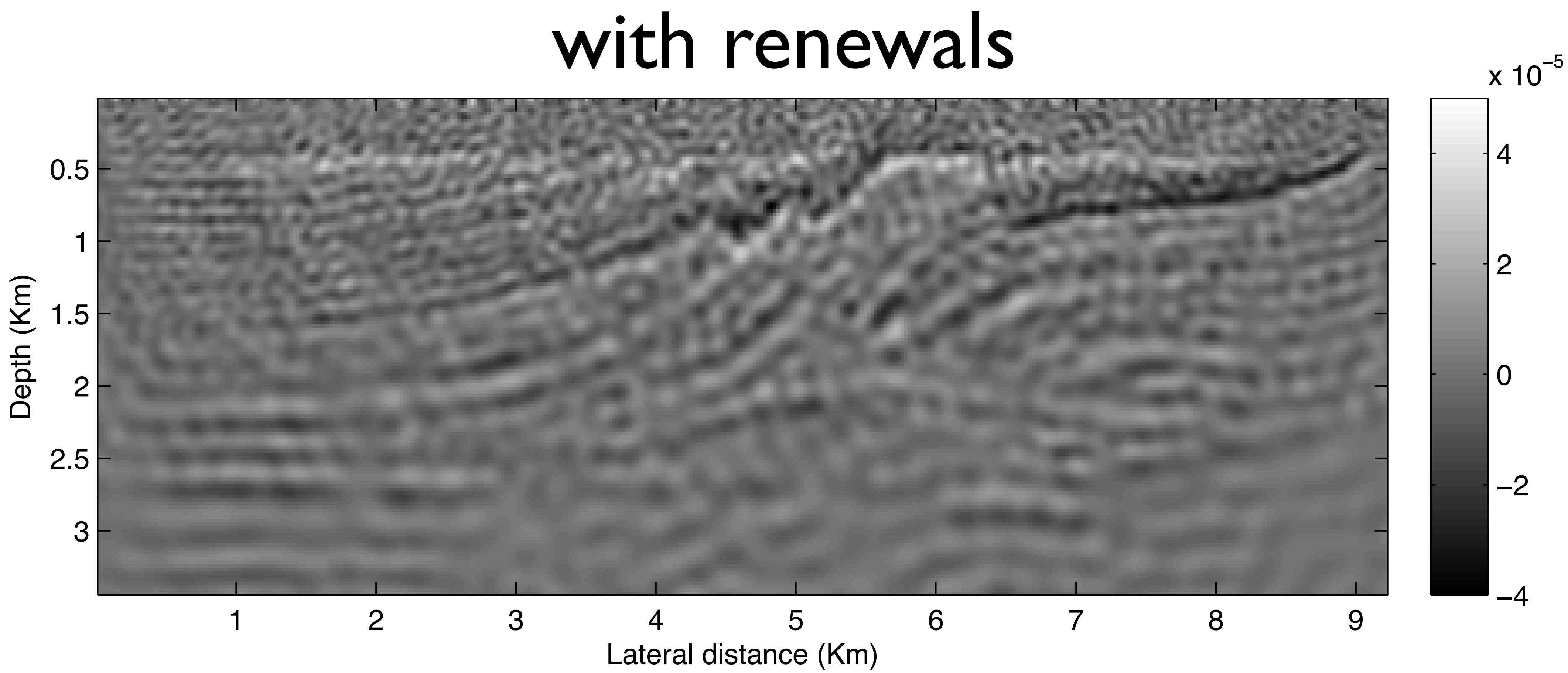
$$\mathbb{P}_{\ell_1}(\mathbf{W}^k;\mathbf{x}_0) \quad \min_{\mathbf{x}} \frac{1}{2K}\sum_{j=1}^{K}\|\underline{\mathbf{b}}_j^k - \underline{\mathbf{A}}_j^k\mathbf{x}\|_2 \quad \text{subject to} \quad \|\mathbf{x}\|_{\ell_1} \leq \tau^k$$

- solve LASSO problem for a given *sparsity* level using the *spectral-gradient* method ($\mathrm{SPG}\ell_1$)
- initialize solver with *warm* start
- solves *sparsity-promoting* subproblem

[van den Berg & Friedlander, '08]

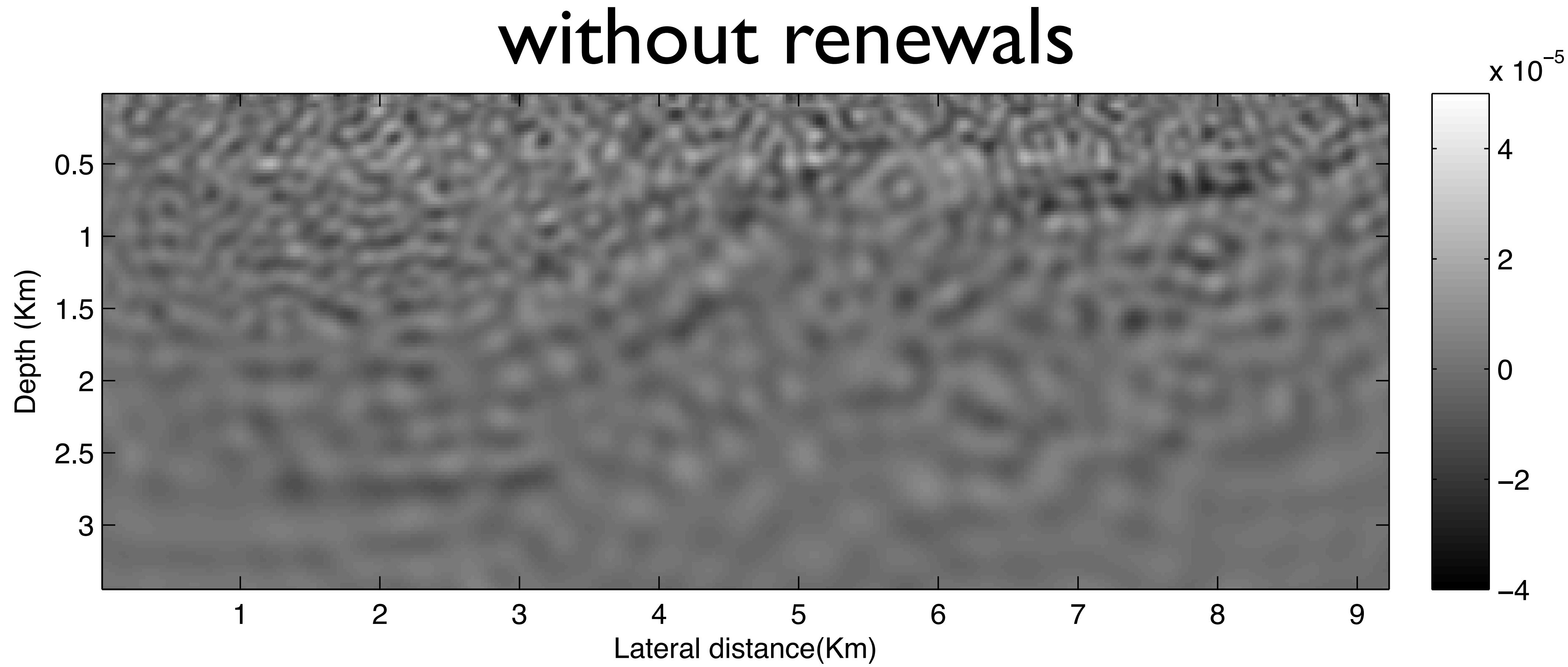
Least-squares migration

8 supershots w 3 frequencies



Least-squares migration

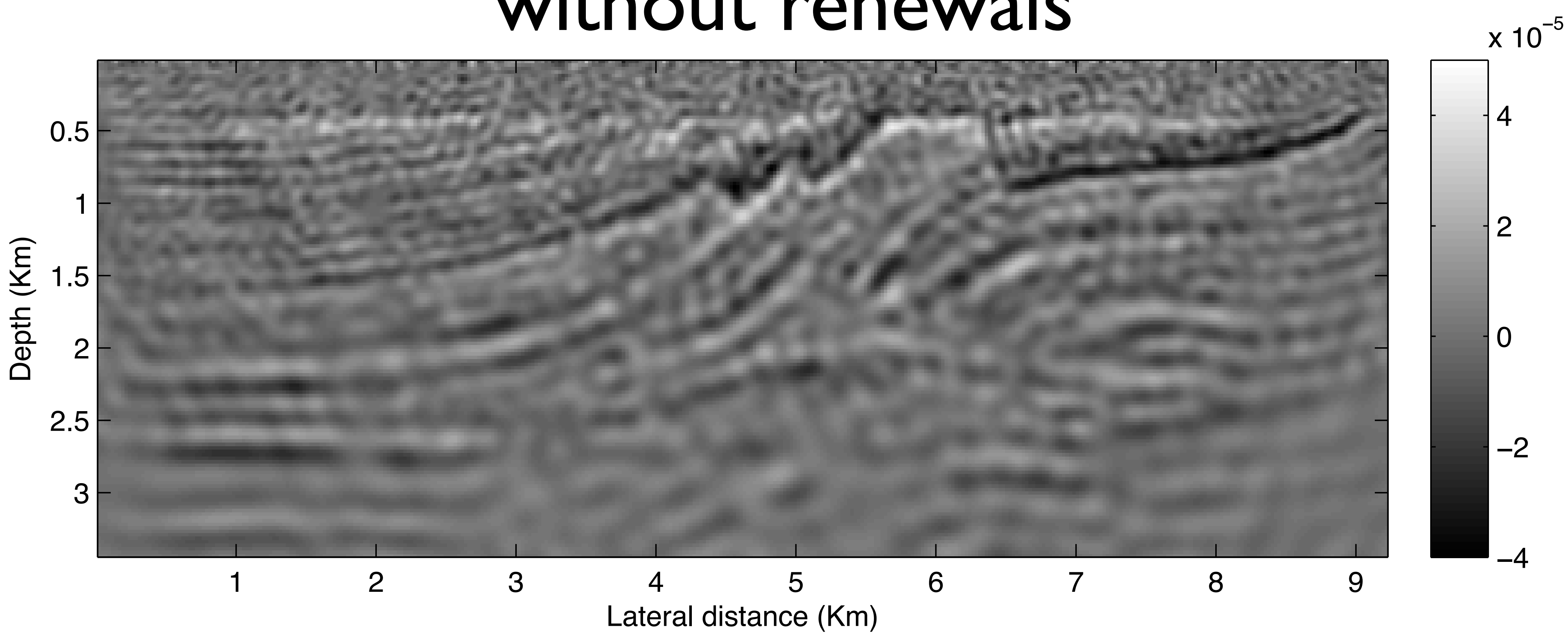
8 supershots w 3 frequencies



Sparse migration

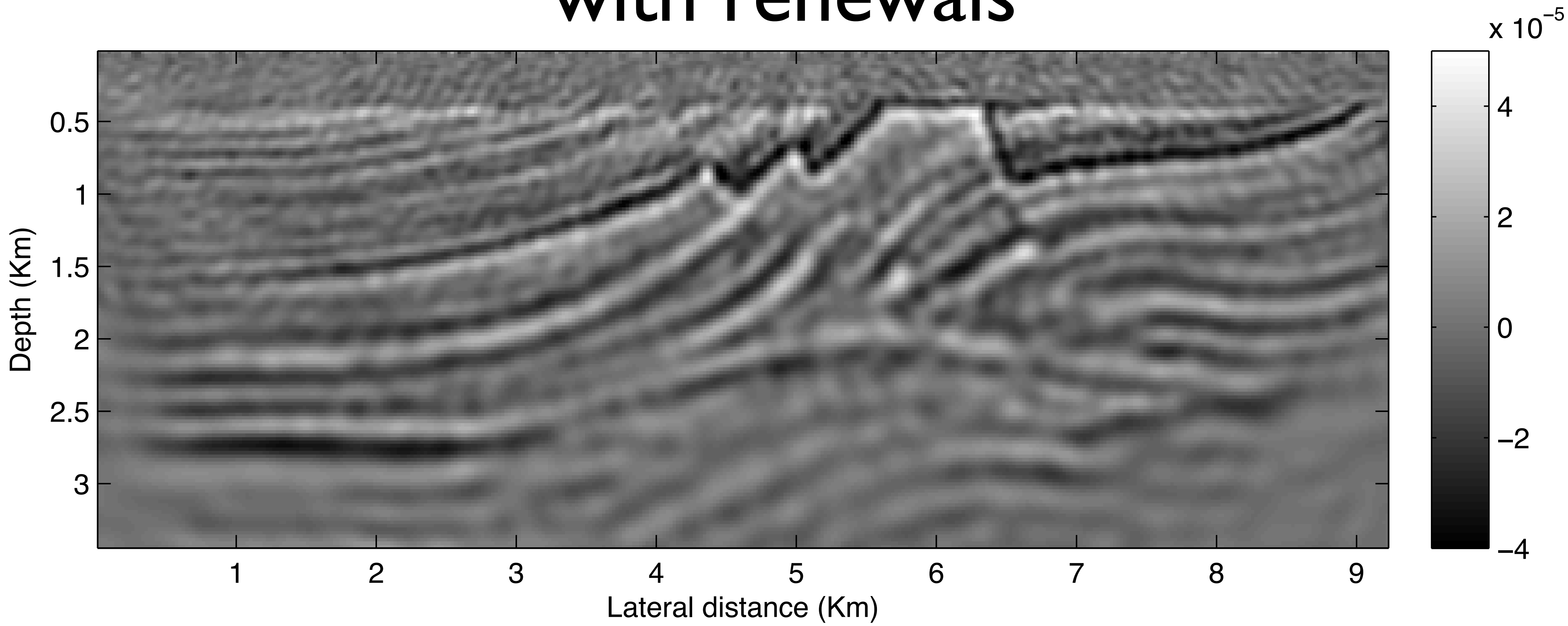
8 supershots w 3 frequencies

without renewals

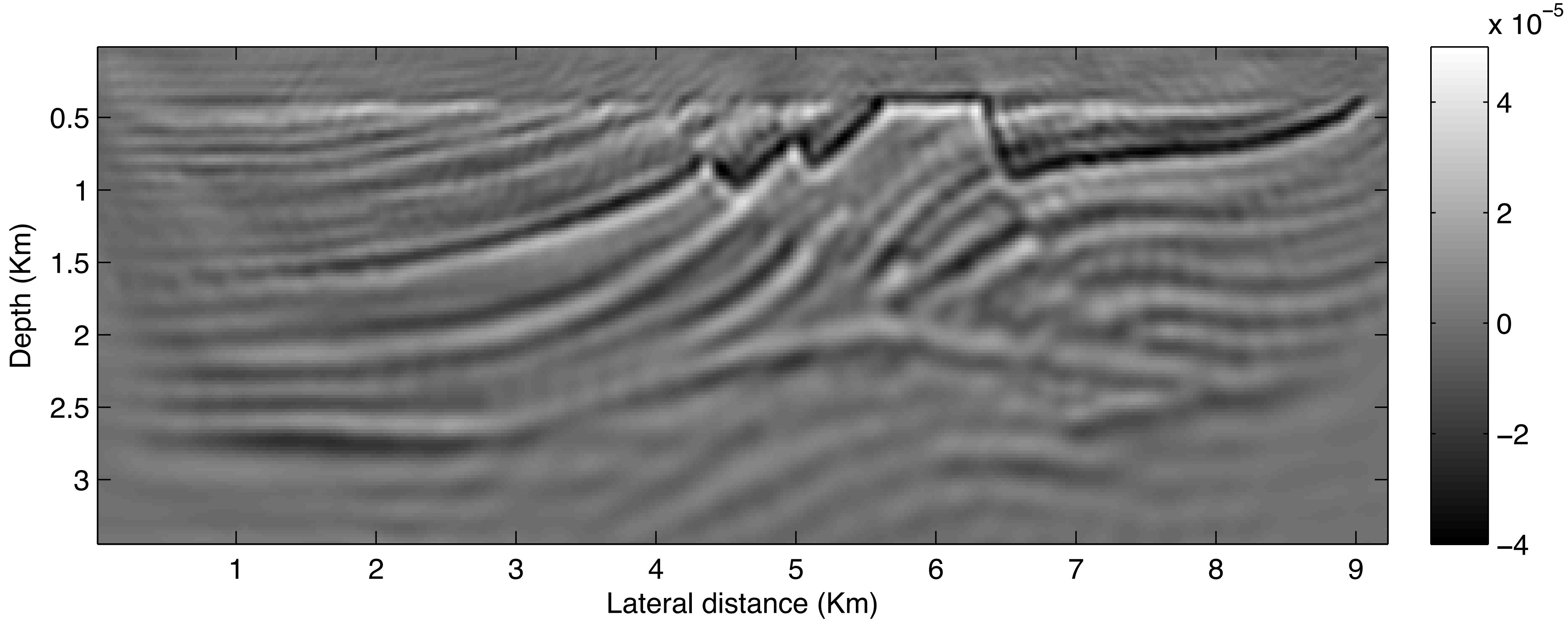


Sparse migration
8 supershots w 3 frequencies

with renewals



Least-squares migration
all 192 shots w 10 frequencies



Combined approach

Leverage findings from *stochastic* & *compressive sensing*

- consider *dimensionality* reduced Gauss-Newton updates as separate "*compressive-sensing* /$\ell_1$regularized experiments"
- turn *large* 'overdetermined' problems with large matrix-setup costs into *small* 'undetermined' problems via *randomization*

Modified Gauss-Newton

- Objective:

$$\underline{f}(\mathbf{m}) := \|\underline{\mathbf{D}} - \boldsymbol{\mathcal{F}}[\mathbf{m};\underline{\mathbf{Q}}]\|_F^2$$

- Iterative algorithm:

$$\mathbf{m}^{\nu+1} = \mathbf{m}^{\nu} + \gamma_{\nu}\boldsymbol{\mathcal{C}}^*\overline{\boldsymbol{\delta}\mathbf{x}}$$

- Direction $\overline{\boldsymbol{\delta}\mathbf{x}}$ solves

$$\begin{array}{ll}\min\limits_{\boldsymbol{\delta}\mathbf{x}} & \|\underline{\mathbf{D}} - \boldsymbol{\mathcal{F}}[\mathbf{m}^{\nu};\underline{\mathbf{Q}}] - \nabla\boldsymbol{\mathcal{F}}[\mathbf{m}^{\nu};\underline{\mathbf{Q}}]\boldsymbol{\mathcal{C}}^*\boldsymbol{\delta}\mathbf{x}\|_F^2 \\ \text{s.t.} & \|\boldsymbol{\delta}\mathbf{x}\|_1 \leq \tau\end{array}$$

- The subproblem for $\overline{\boldsymbol{\delta}\mathbf{x}}$ is convex, and $\boldsymbol{\mathcal{C}}^*\overline{\boldsymbol{\delta}\mathbf{x}}$ is a ***descent*** direction:

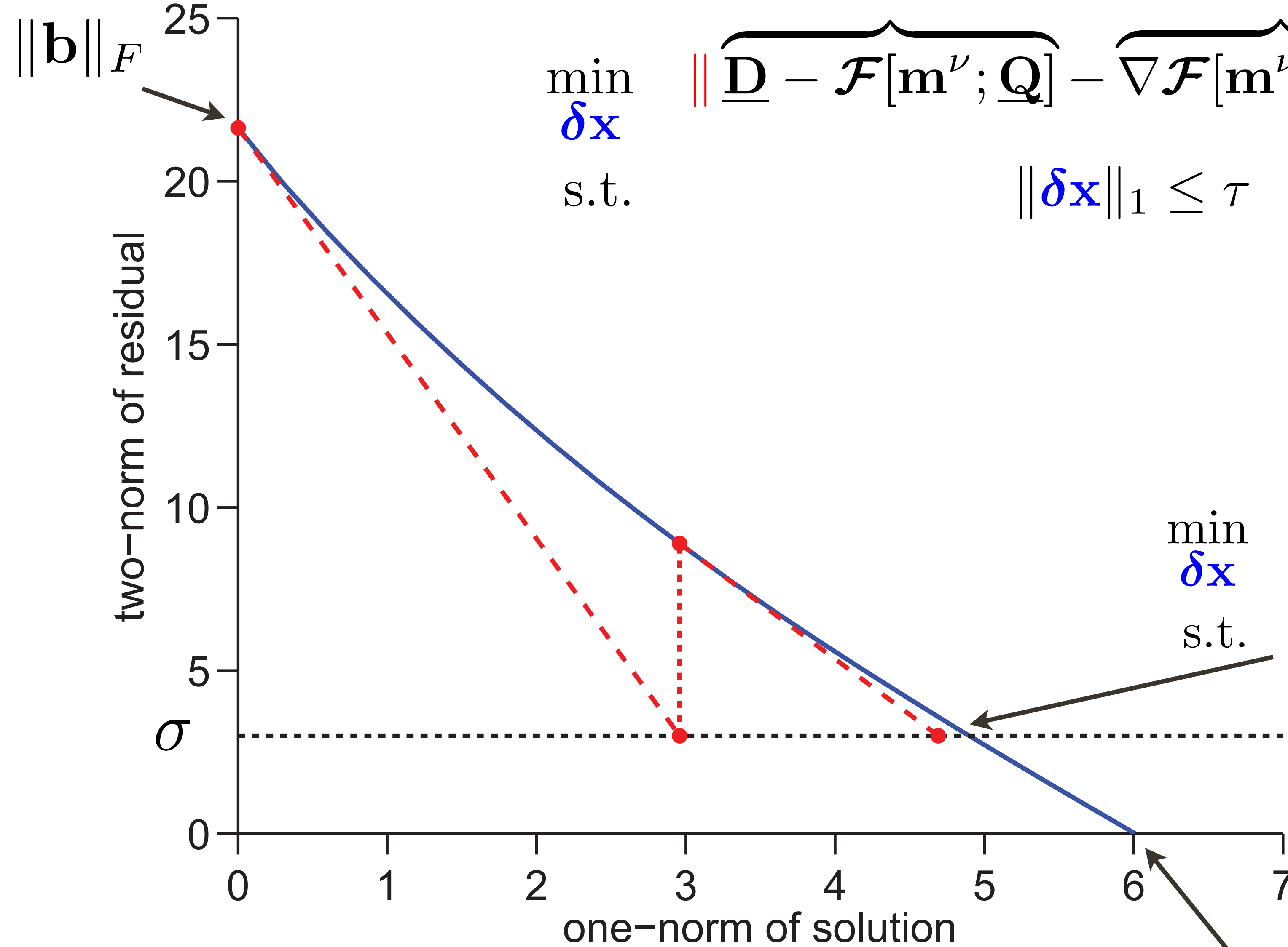
$$\underline{f}'(\mathbf{m}^{\nu};\boldsymbol{\mathcal{C}}^*\overline{\boldsymbol{\delta}\mathbf{x}}) \leq \underline{f}(\mathbf{m}^{\nu}) - \underbrace{\|\underline{\mathbf{D}} - \boldsymbol{\mathcal{F}}[\mathbf{m}^{\nu};\underline{\mathbf{Q}}]}_{\underline{f}(\mathbf{m}^{\nu})} - \nabla\boldsymbol{\mathcal{F}}[\mathbf{m};\underline{\mathbf{Q}}]\boldsymbol{\mathcal{C}}^*\overline{\boldsymbol{\delta}\mathbf{x}}\|_F^2 < 0$$

[Burke ’89, Burke ’92]

Picking Lasso Parameter

$$\min_{\boldsymbol{\delta}\mathbf{x}} \quad \| \overbrace{\underline{\mathbf{D}} - \mathcal{F}[\mathbf{m}^{\nu};\underline{\mathbf{Q}}]}^{\mathbf{b}} - \overbrace{\nabla\mathcal{F}[\mathbf{m}^{\nu};\underline{\mathbf{Q}}]\mathcal{C}^{*}}^{\mathbf{A}}\, \boldsymbol{\delta}\mathbf{x}\|_F^2 \quad \text{s.t.} \quad \|\boldsymbol{\delta}\mathbf{x}\|_1 \leq \tau$$

Lasso



$$\min_{\boldsymbol{\delta}\mathbf{x}} \quad \|\boldsymbol{\delta}\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{b} - \mathbf{A}\boldsymbol{\delta}\mathbf{x}\|_F \leq \sigma$$

BPDN

$$\min_{\boldsymbol{\delta}\mathbf{x}} \quad \|\boldsymbol{\delta}\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{b} = \mathbf{A}\boldsymbol{\delta}\mathbf{x}$$

BP

[van den Berg ’08]

Compressive inversion

Algorithm 1: Dimensionality-reduced Gauss Newton with sparsity

Result: Output estimate for the model $\mathbf{m}$

$\mathbf{m} \longleftarrow \mathbf{m}_0$; $k \longleftarrow 0$; `// initial model`

while not converged **do**

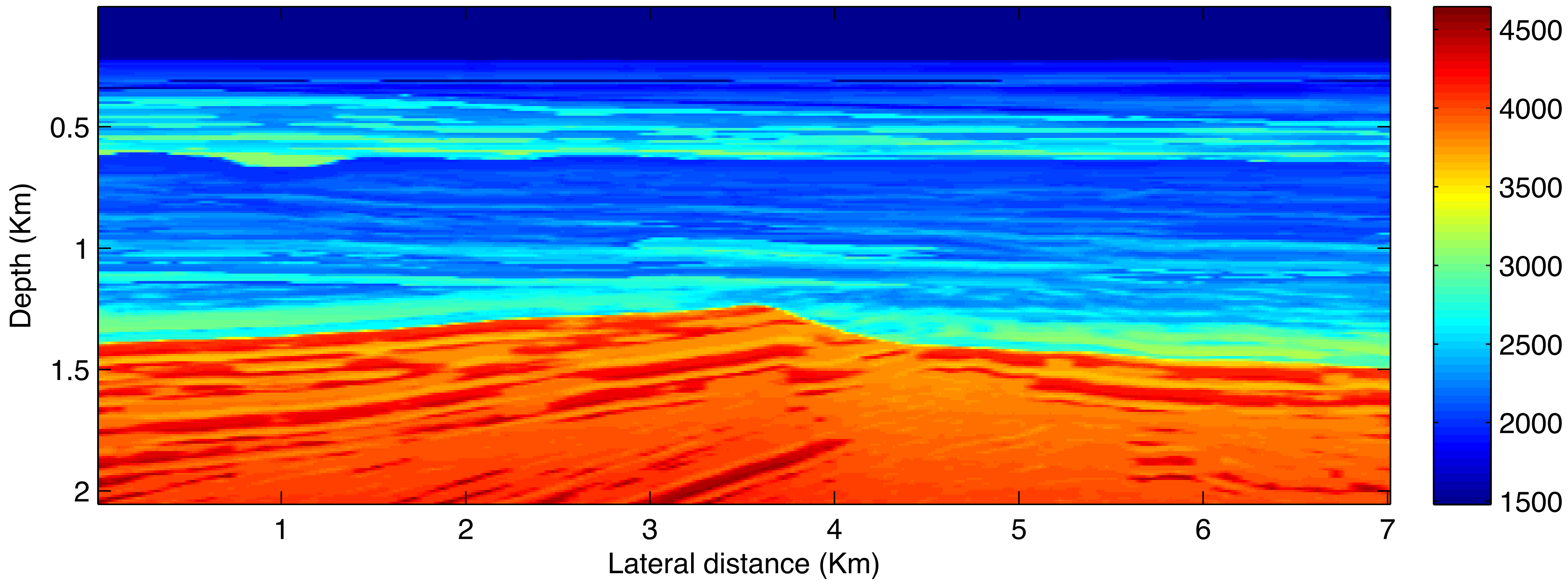
$$\delta\mathbf{x} \longleftarrow \begin{cases} \mathbf{S}^* \arg\min_{\delta\mathbf{x}} \frac{1}{2}\|\underline{\mathbf{D}}^k - \boldsymbol{\mathcal{F}}[\mathbf{m}^k; \underline{\mathbf{Q}}^k] - \nabla\boldsymbol{\mathcal{F}}[\mathbf{m}^k; \underline{\mathbf{Q}}^k]\delta\mathbf{x}\|_F^2 \\ \text{subject to} \quad \|\delta\mathbf{x}\|_1 \leq \tau^k \end{cases}$$

$\mathbf{m}^{k+1} \longleftarrow \mathbf{m}^k + \gamma^k \mathbf{S}^* \delta\mathbf{x}$; `// update with linesearch`

$k \longleftarrow k + 1$;

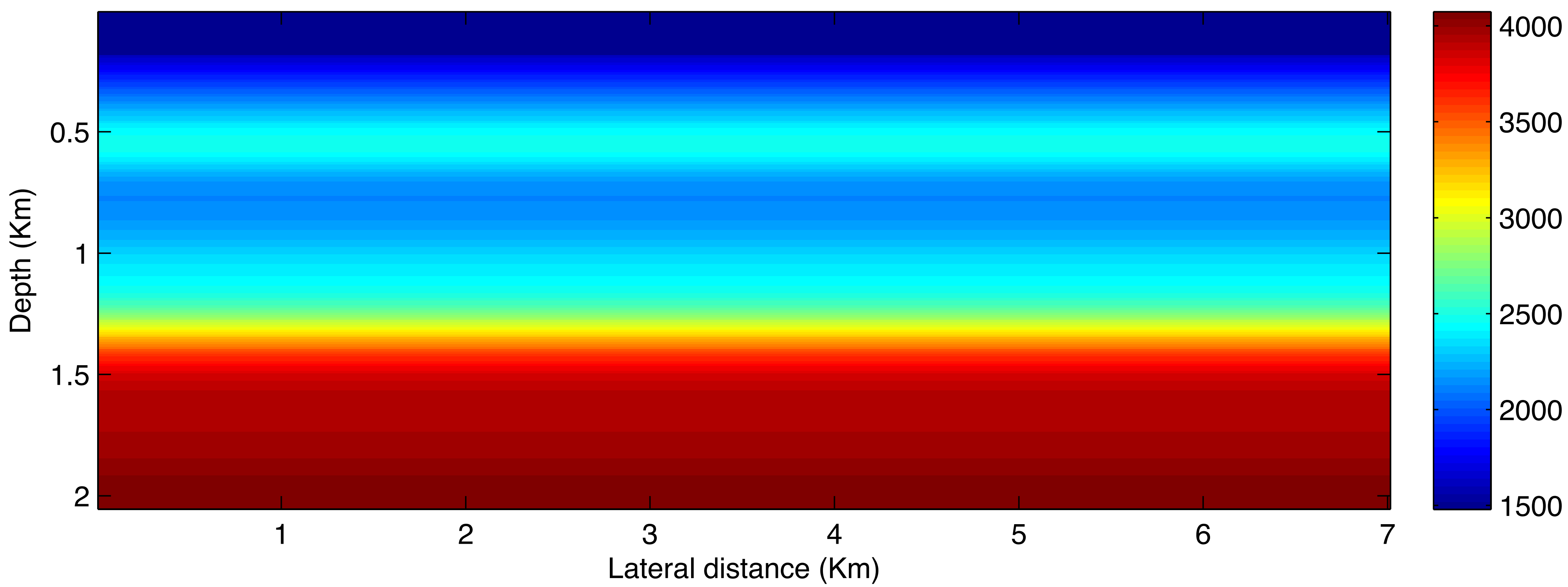
end

Example II
BG model

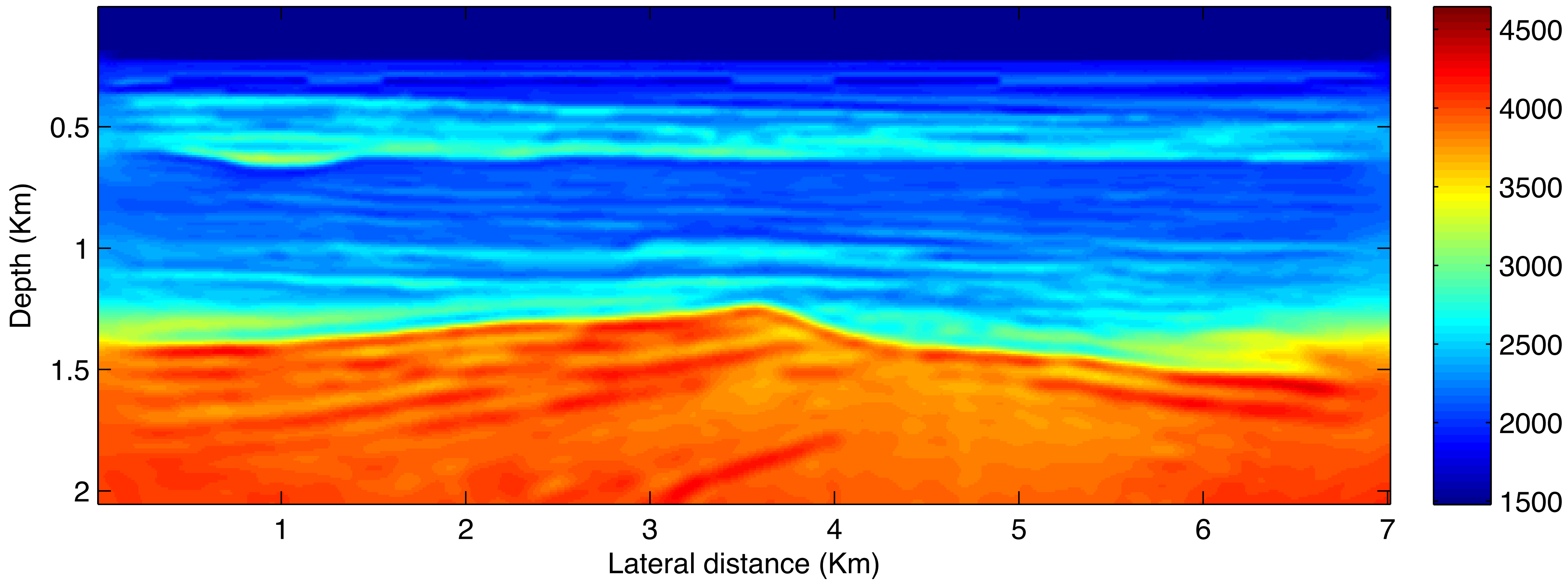


BG model

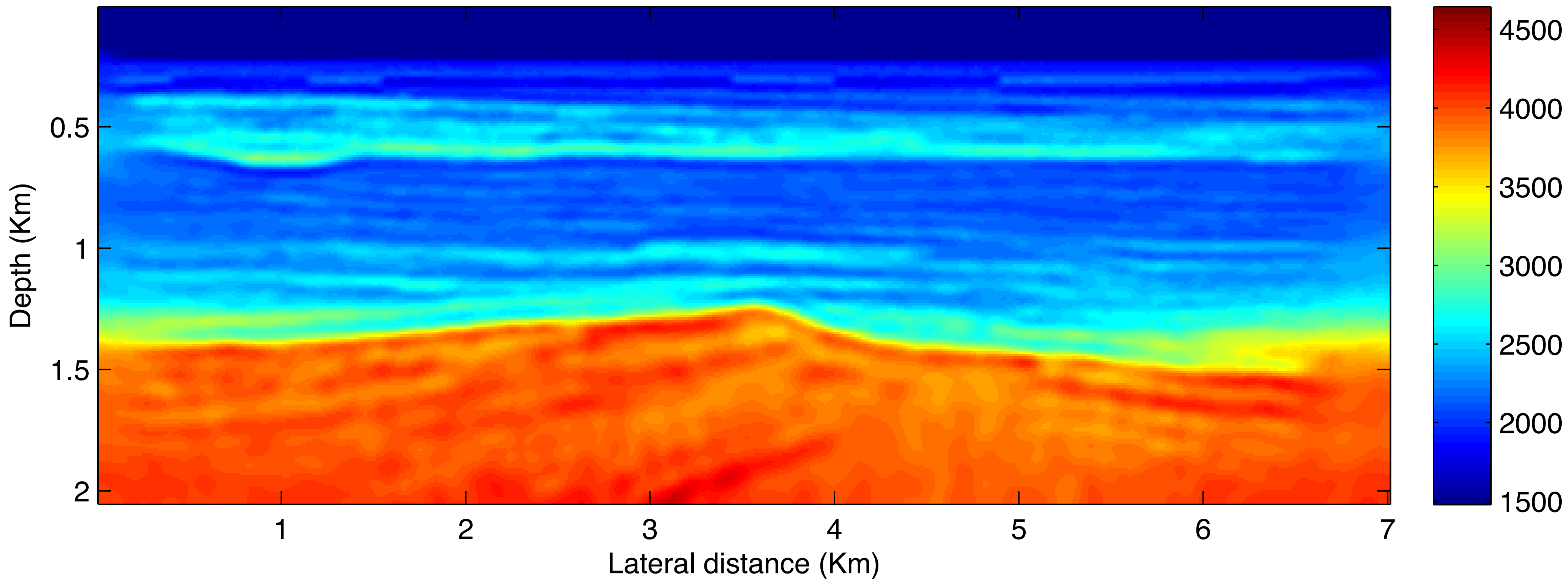
initial model



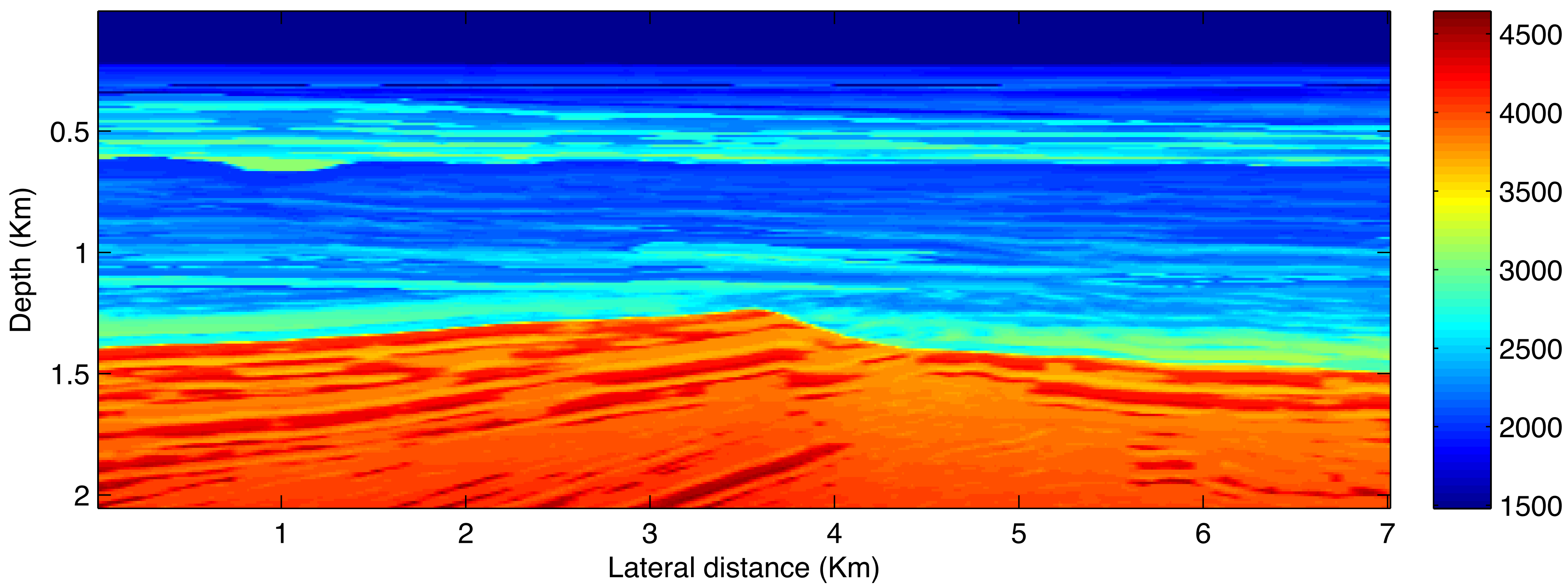
BG model
inverted model



BG model
inverted model w/o renewals



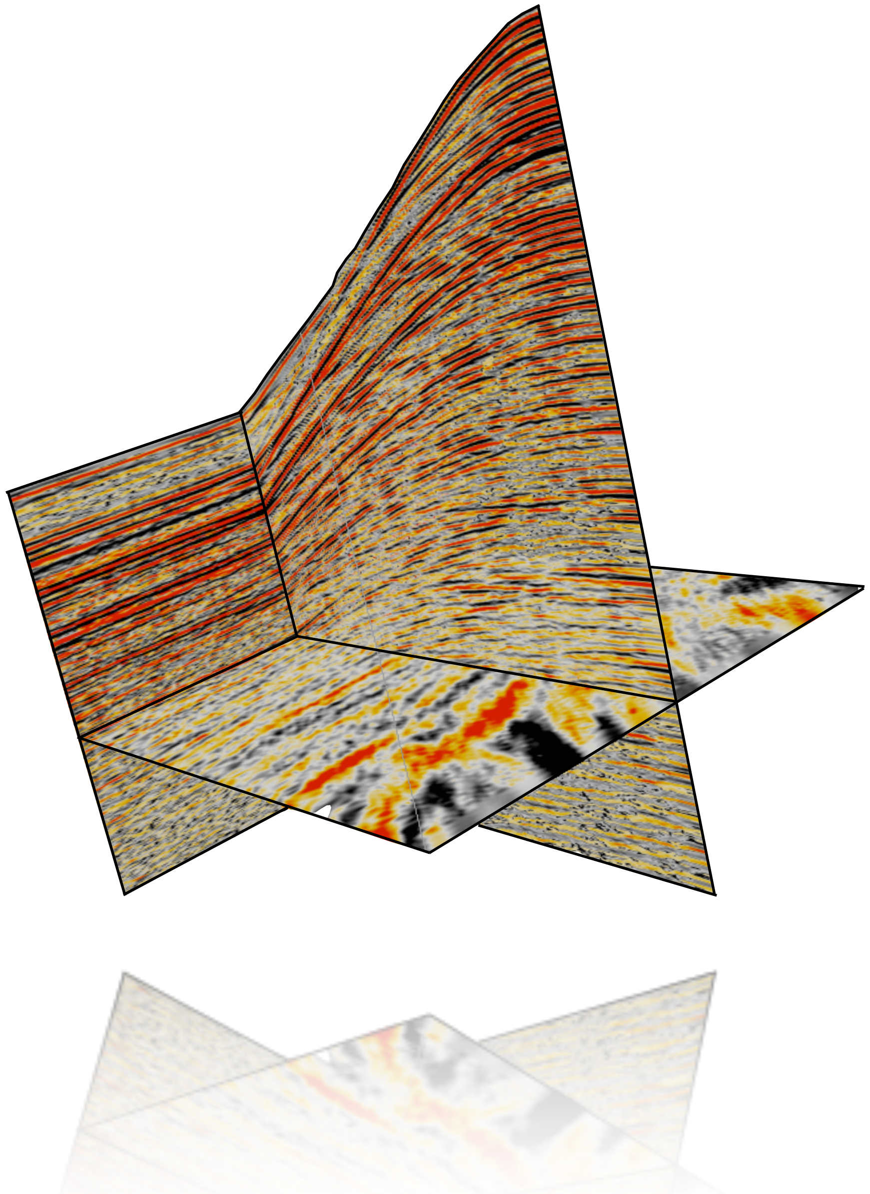
Example II
BG model



Related problems

Felix J. Herrmann*
fherrmann@eos.ubc.ca

***Seismic Laboratory for Imaging & Modeling**
Department of Earth & Ocean Sciences
The University of British Columbia



Golub Summer School
Vancouver, July 4-15, 2011

Free-surface mitigation

Robust estimation of primaries by sparse inversion

- alternating optimization
- curvelet-domain sparsity promotion
- informed blind deconvolution

Sparsity-promoting *imaging* with *multiples*

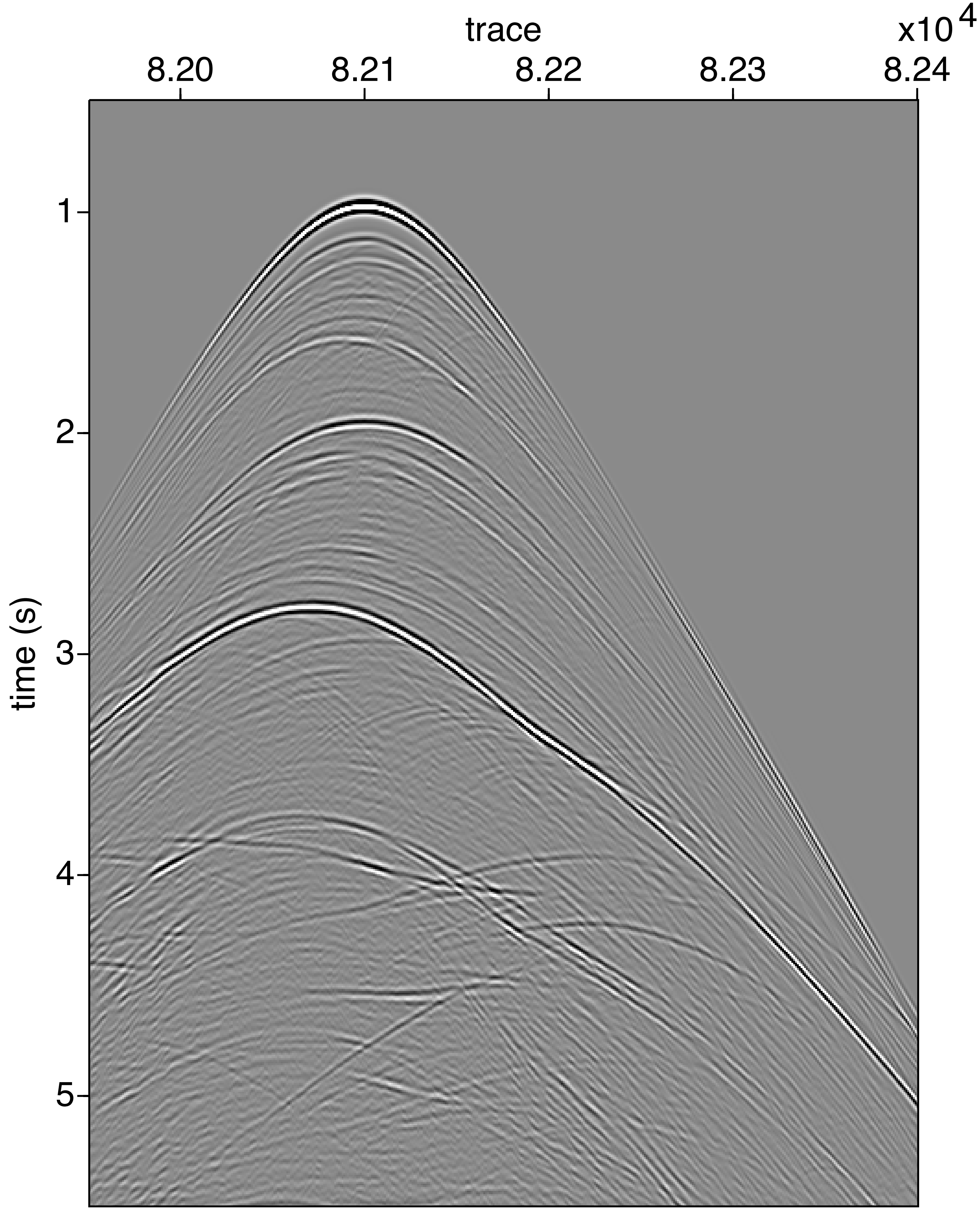
Robust EPSI

Estimation of *primaries* by *sparse* inversion

$$\overbrace{\mathbf{P}}^{\text{upgoing wavefield}} \approx \underbrace{\mathbf{G}}_{\text{surface-free impulse response}} \overbrace{[\mathbf{Q} - \mathbf{P}]}^{\text{downgoing wavefield}}$$

Involves the solution of a bi-convex optimization problem yielding *alternating* estimates

- for the source function **Q**
- for the surface-free Green's function **G**



SLIM

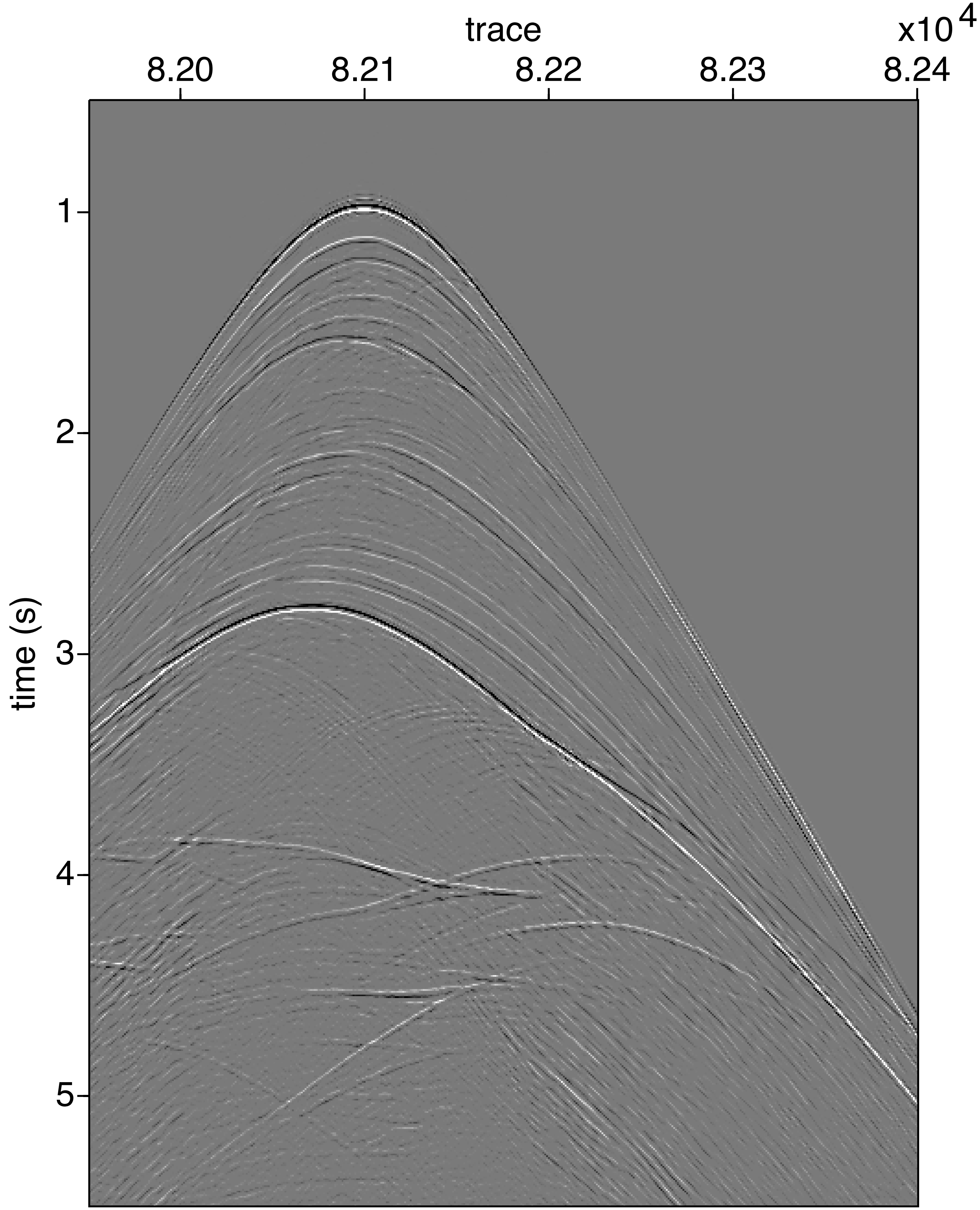
Pluto15

data

Elastic FD Modeling

muted

no deghosting



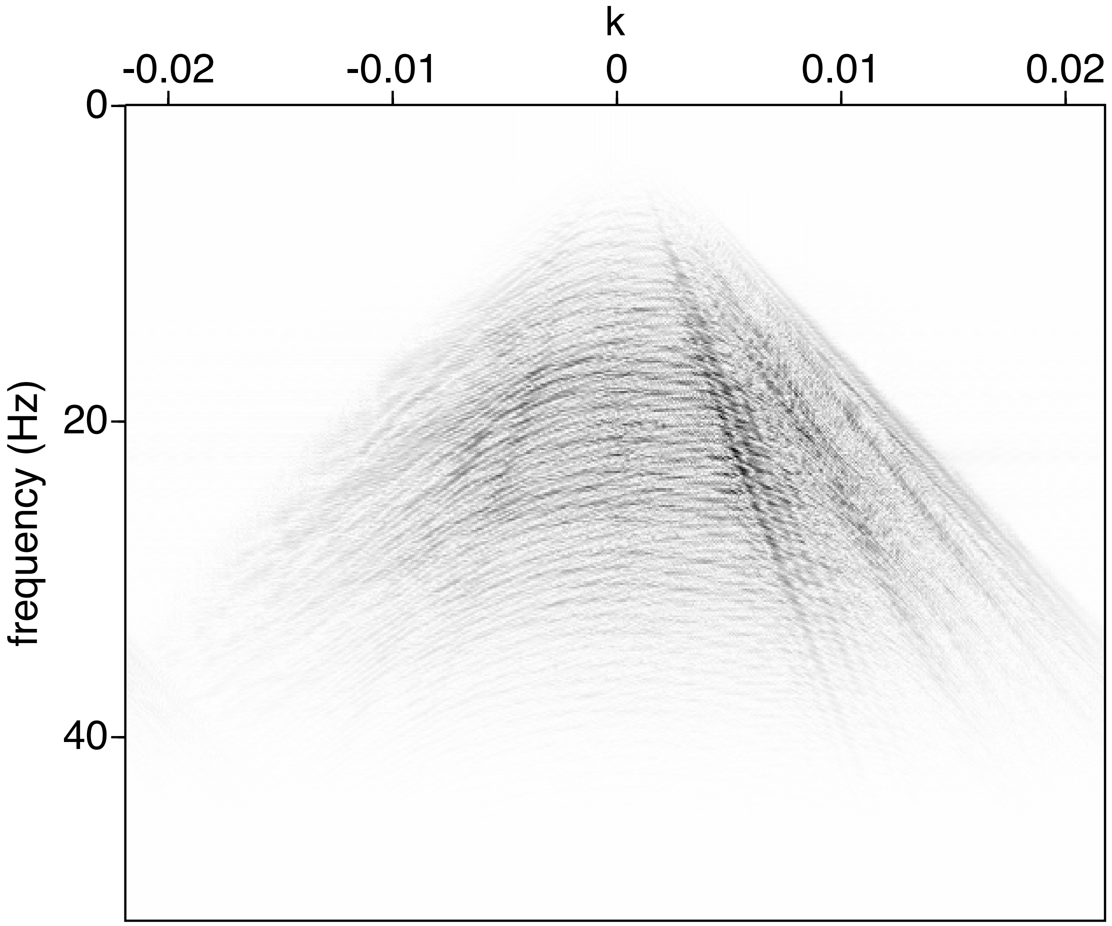
Pluto15

REPSI

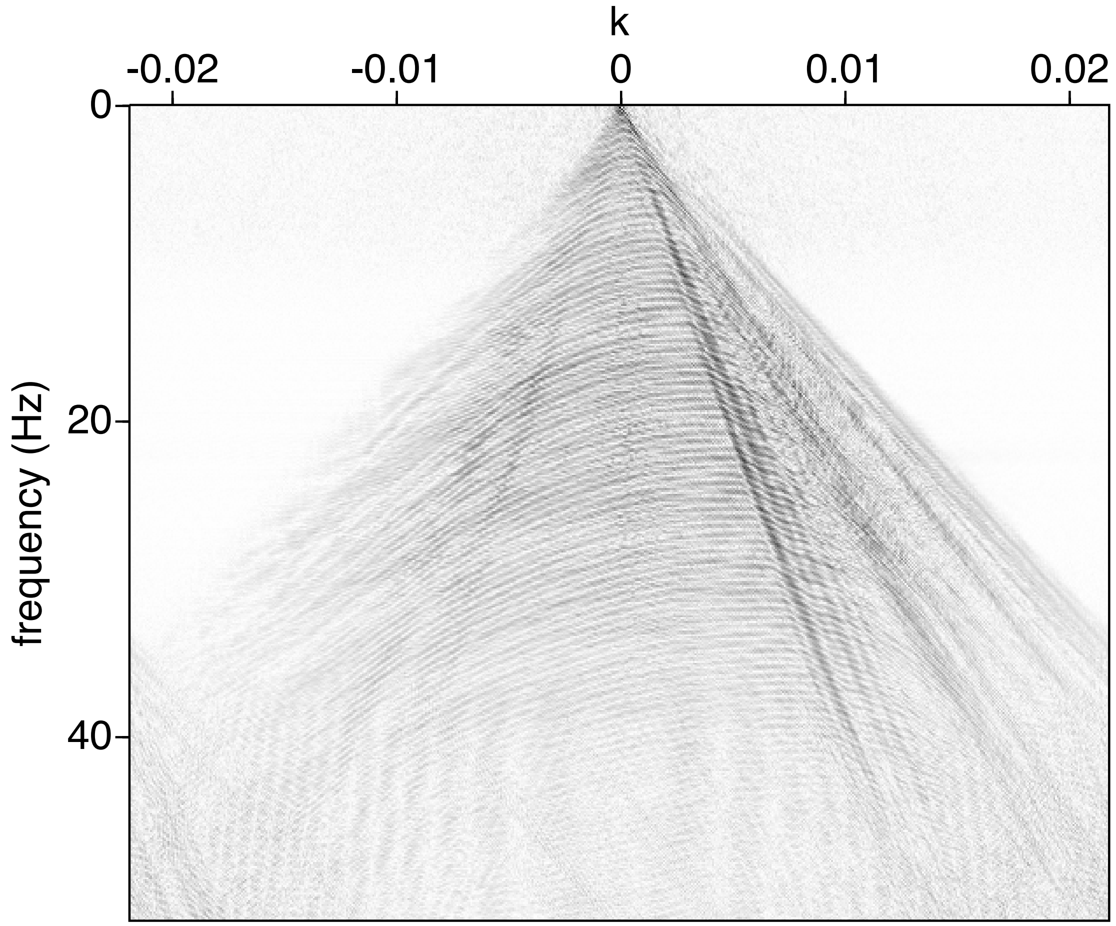
Primary IR (G)

no transform used

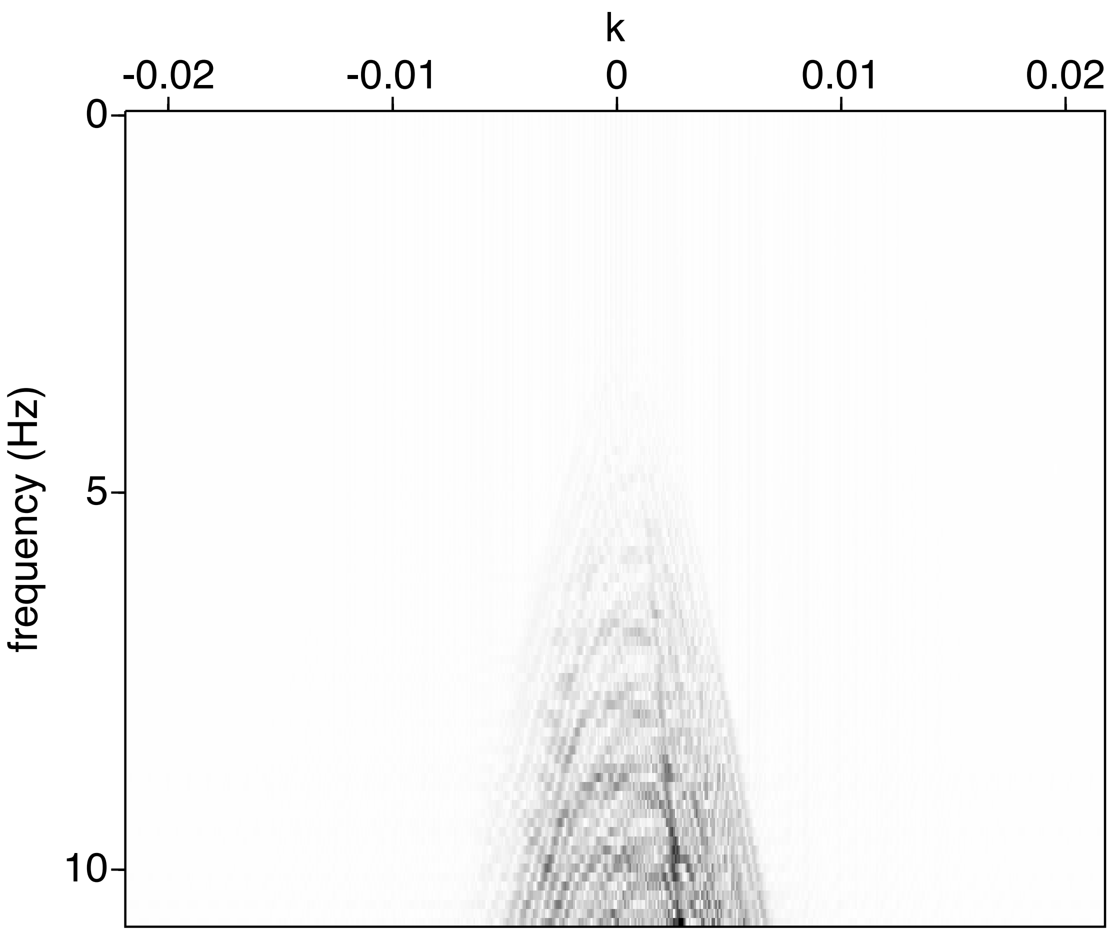
80 iters



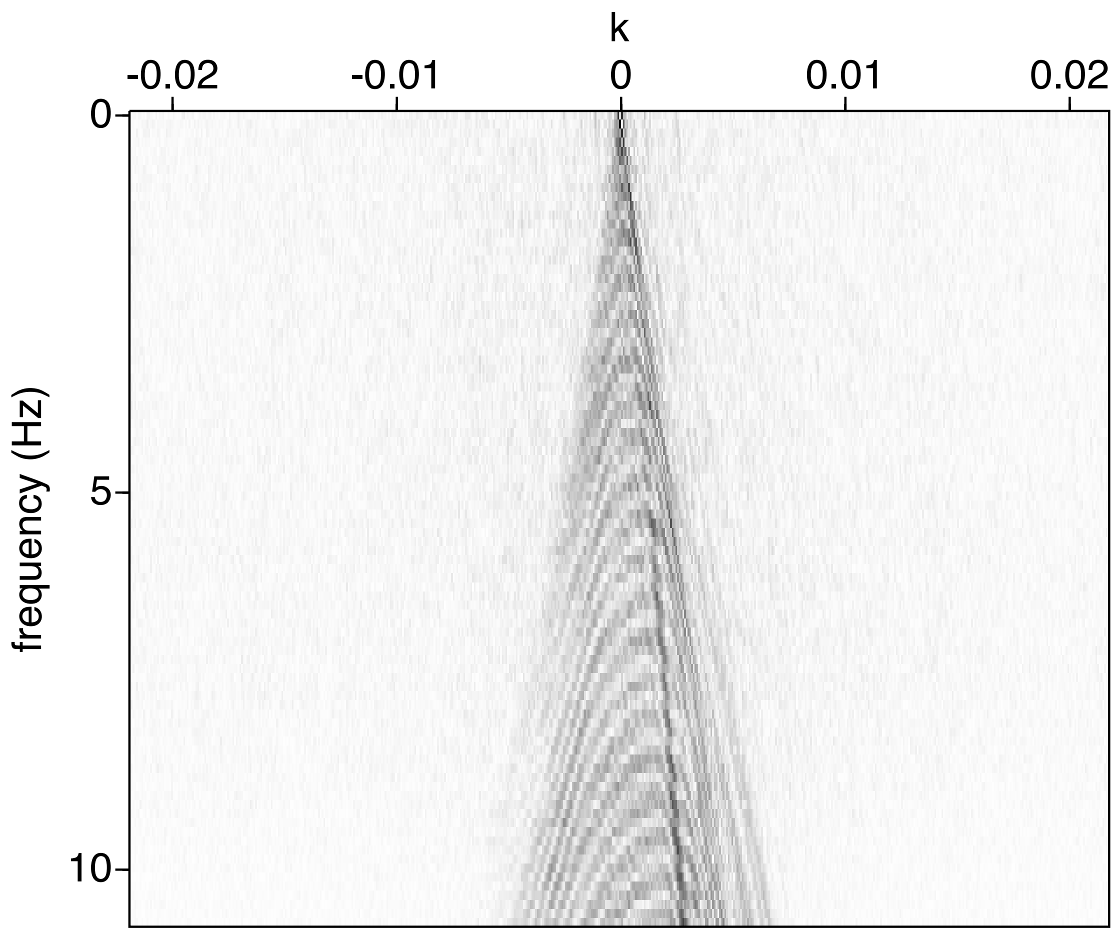
F-K Spectrum of data



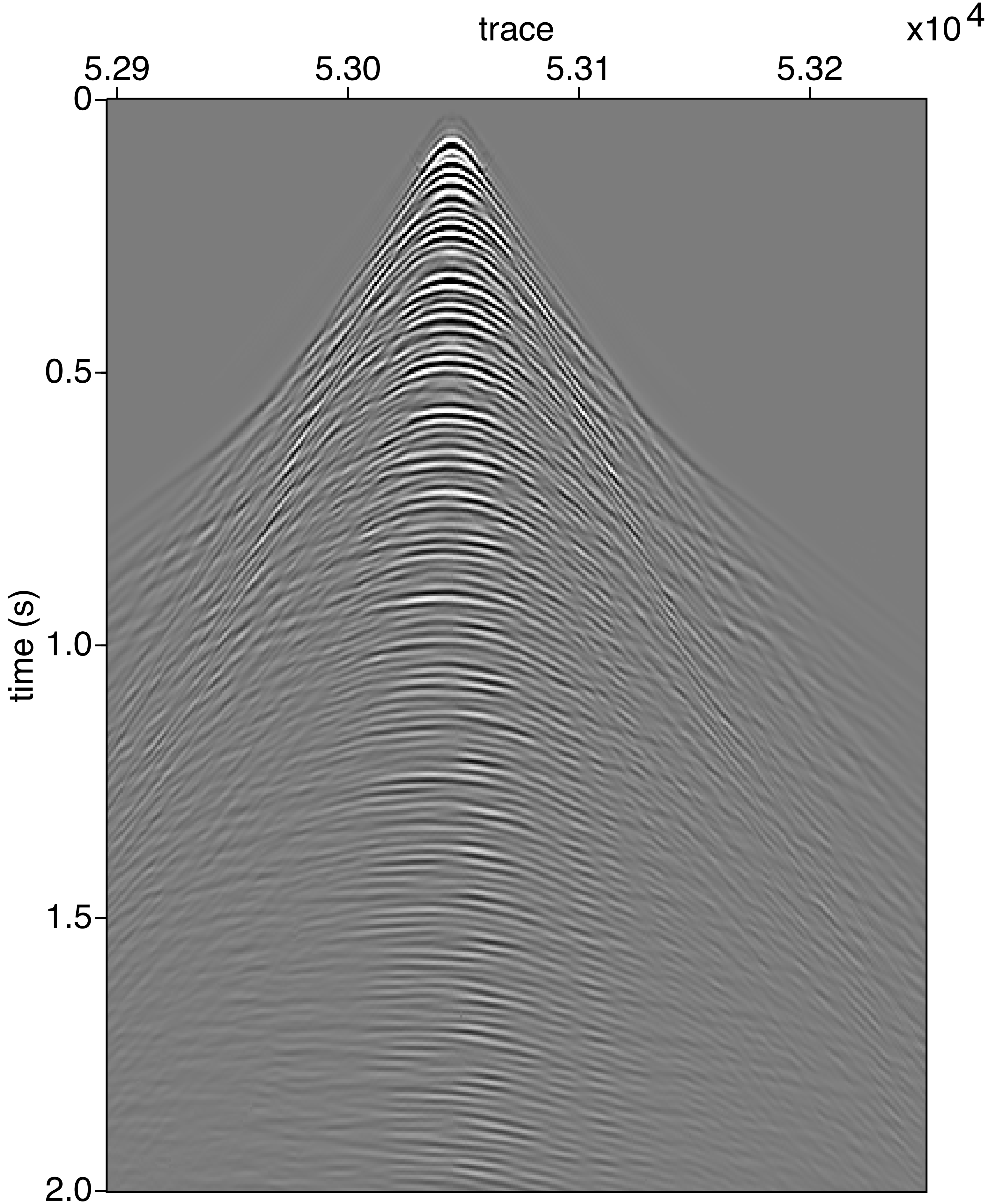
F-K Spectrum of REPSI Primary IR



F-K Spectrum of data



F-K Spectrum of REPSI Primary IR



SLIM

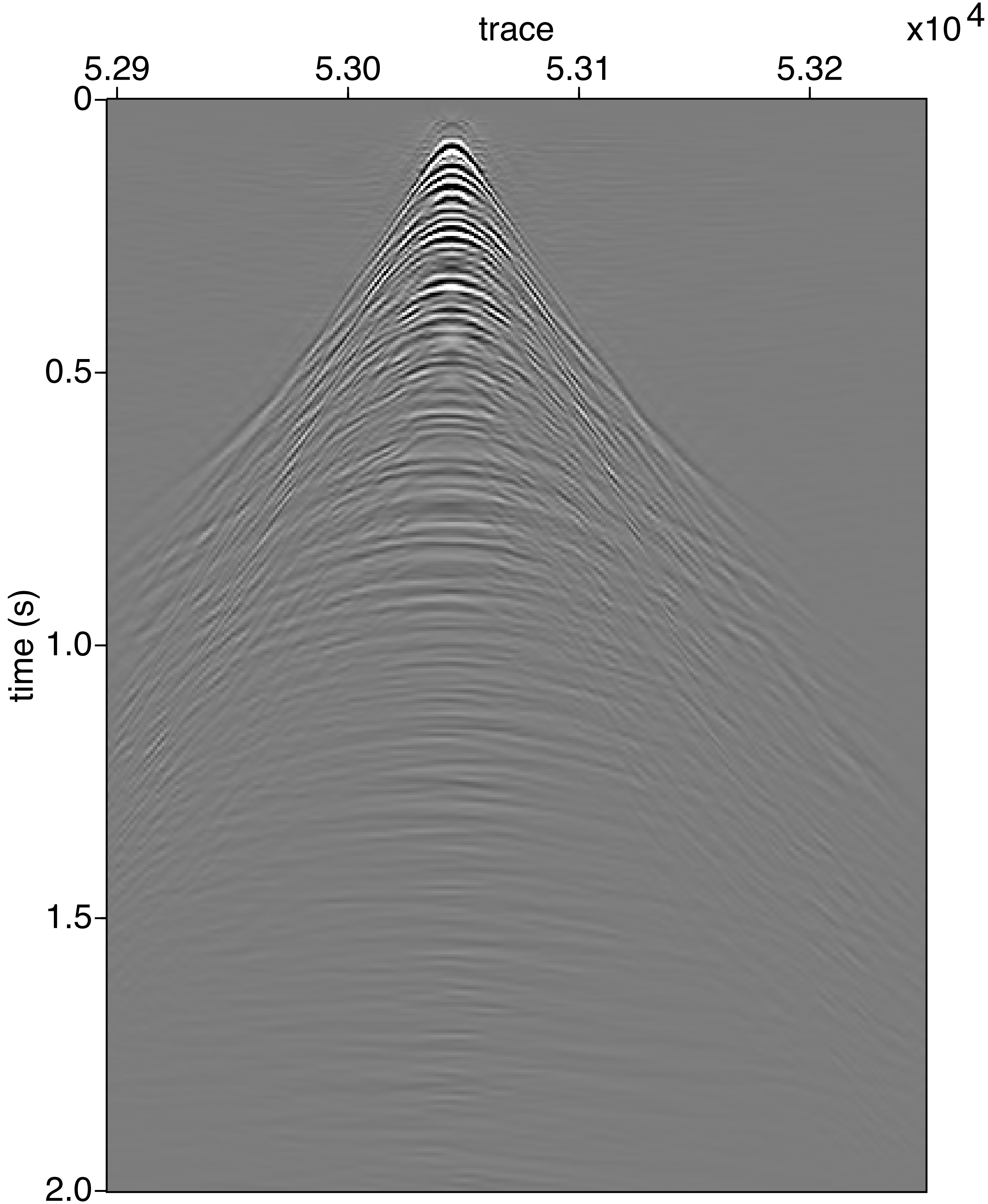
Gulf of Suez
data

shot gather

interpolated, muted

reciprocity

no deghosting



SLIM

Gulf of Suez

REPSI + Transform

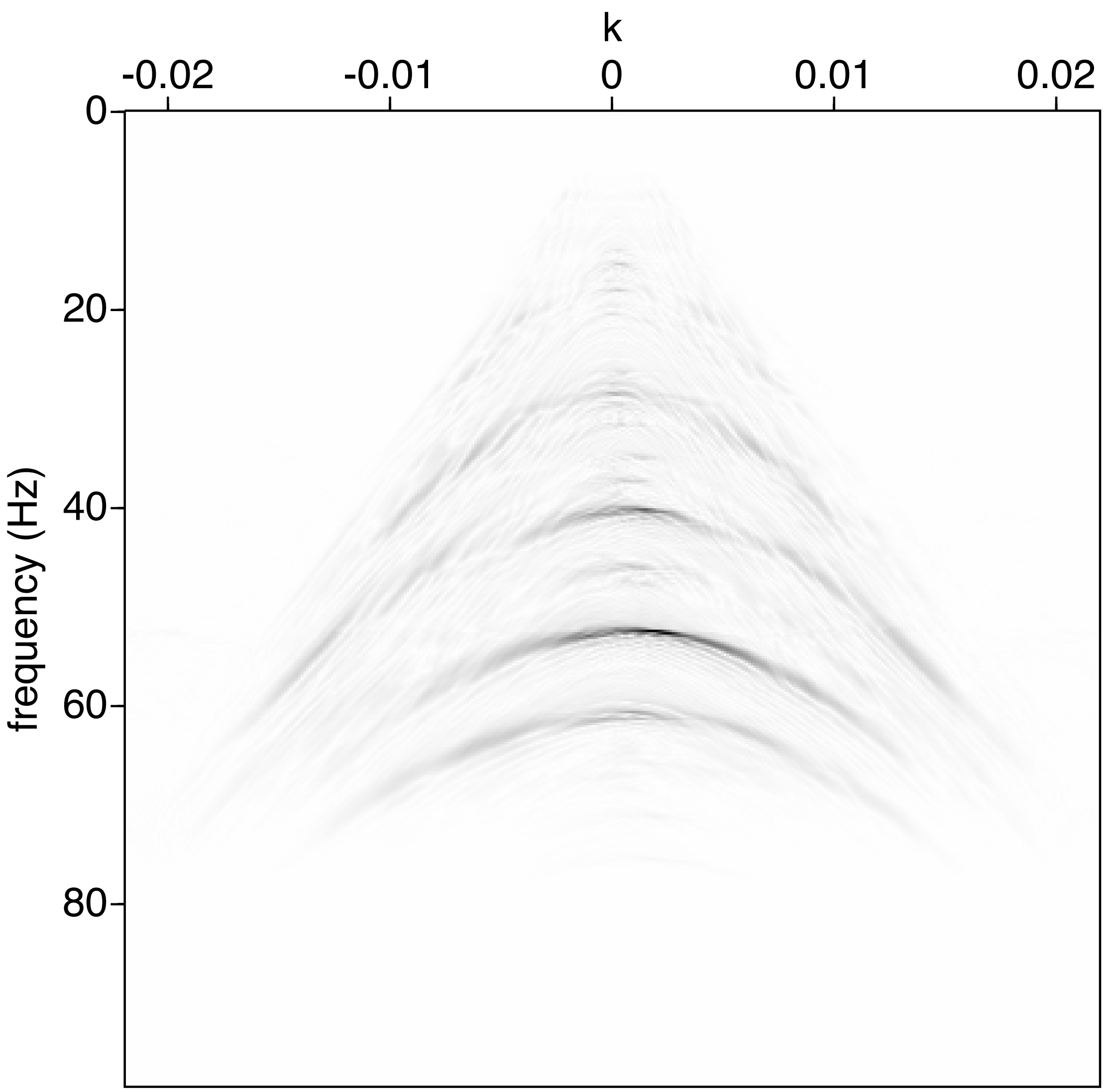
Primary IR (G)

shot gather

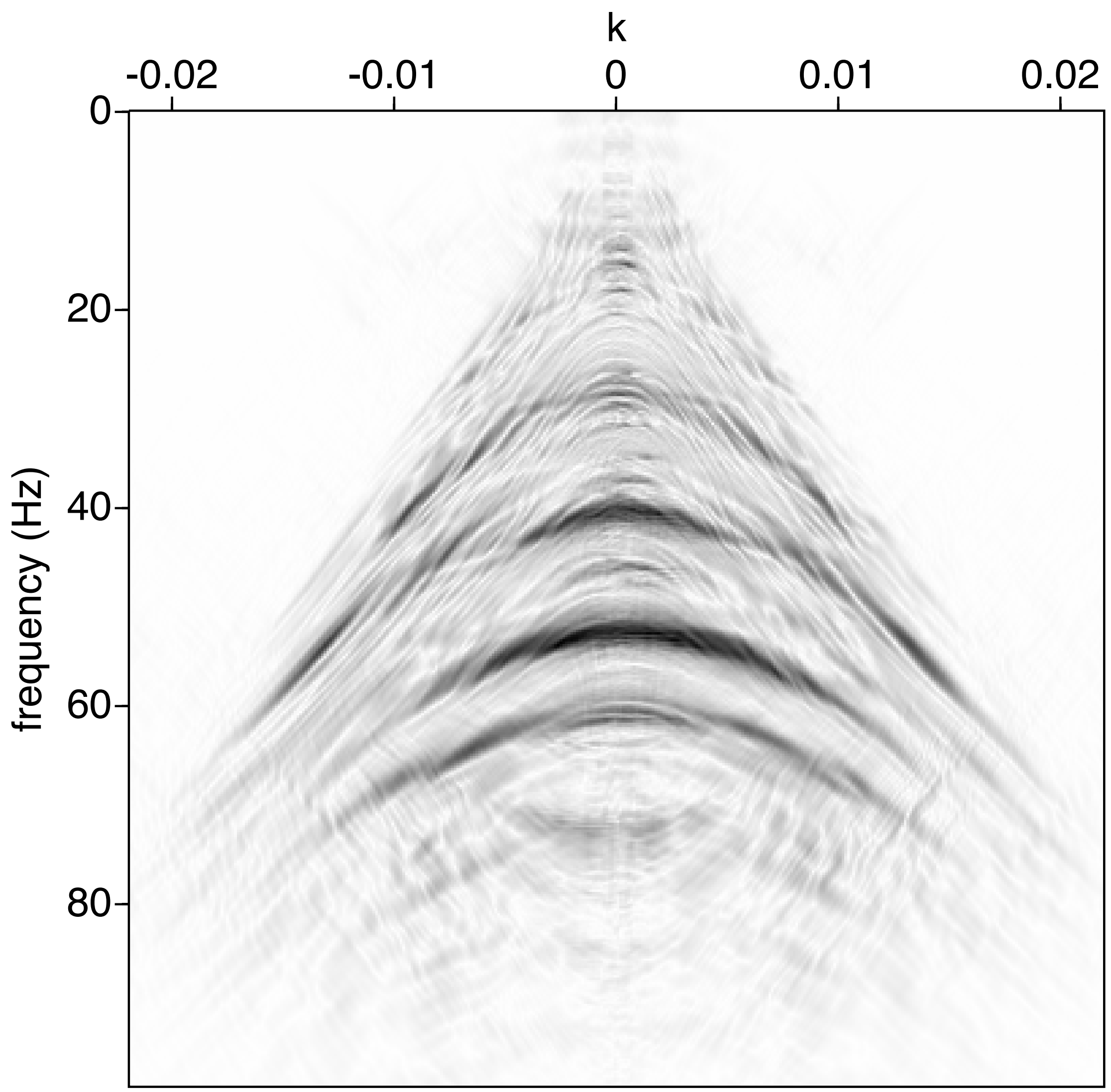
2D Curvelet (Src-Rcv)

Spline a=3.0 DWT (Time)

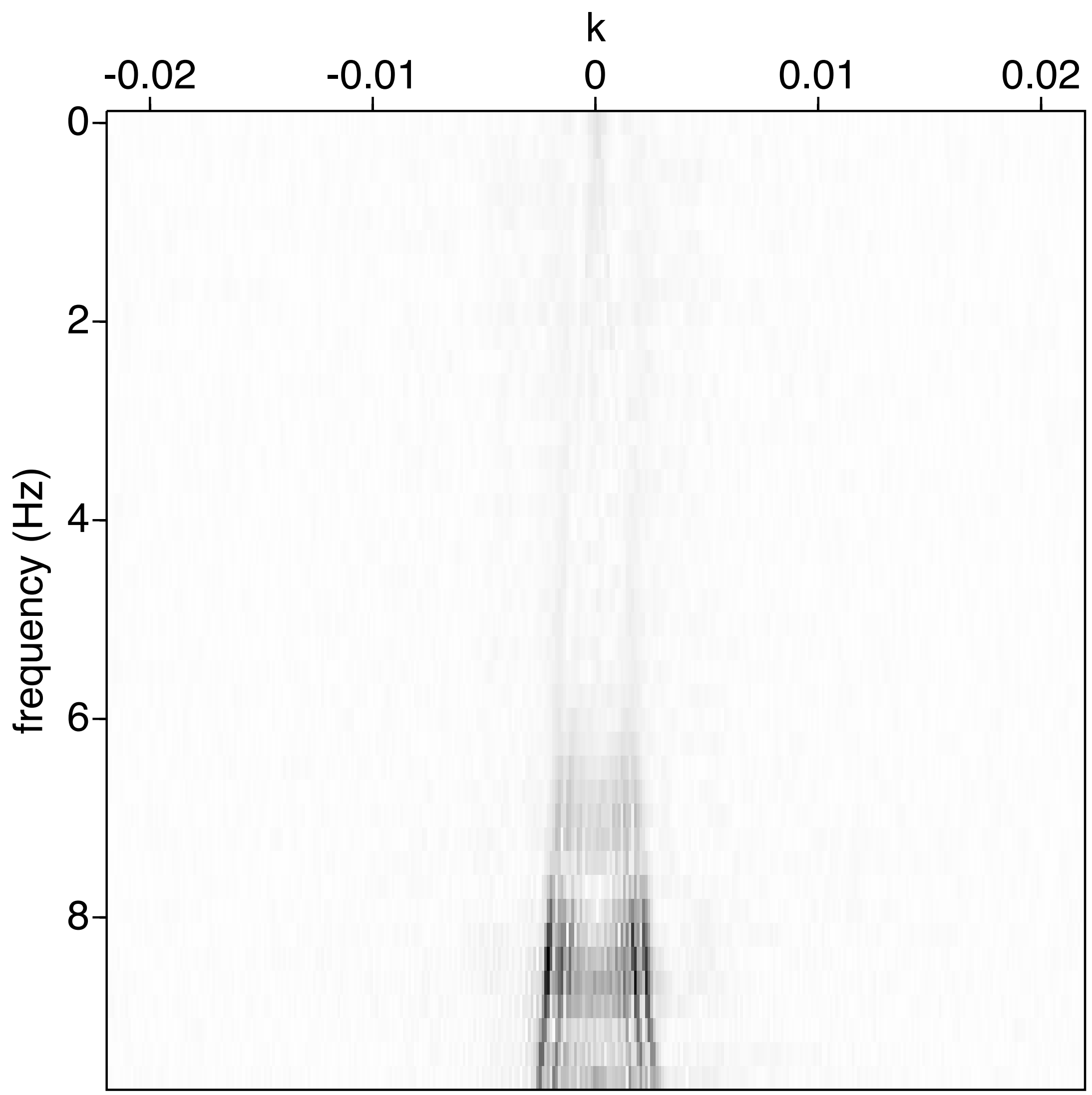
90 SPG grad. iterations



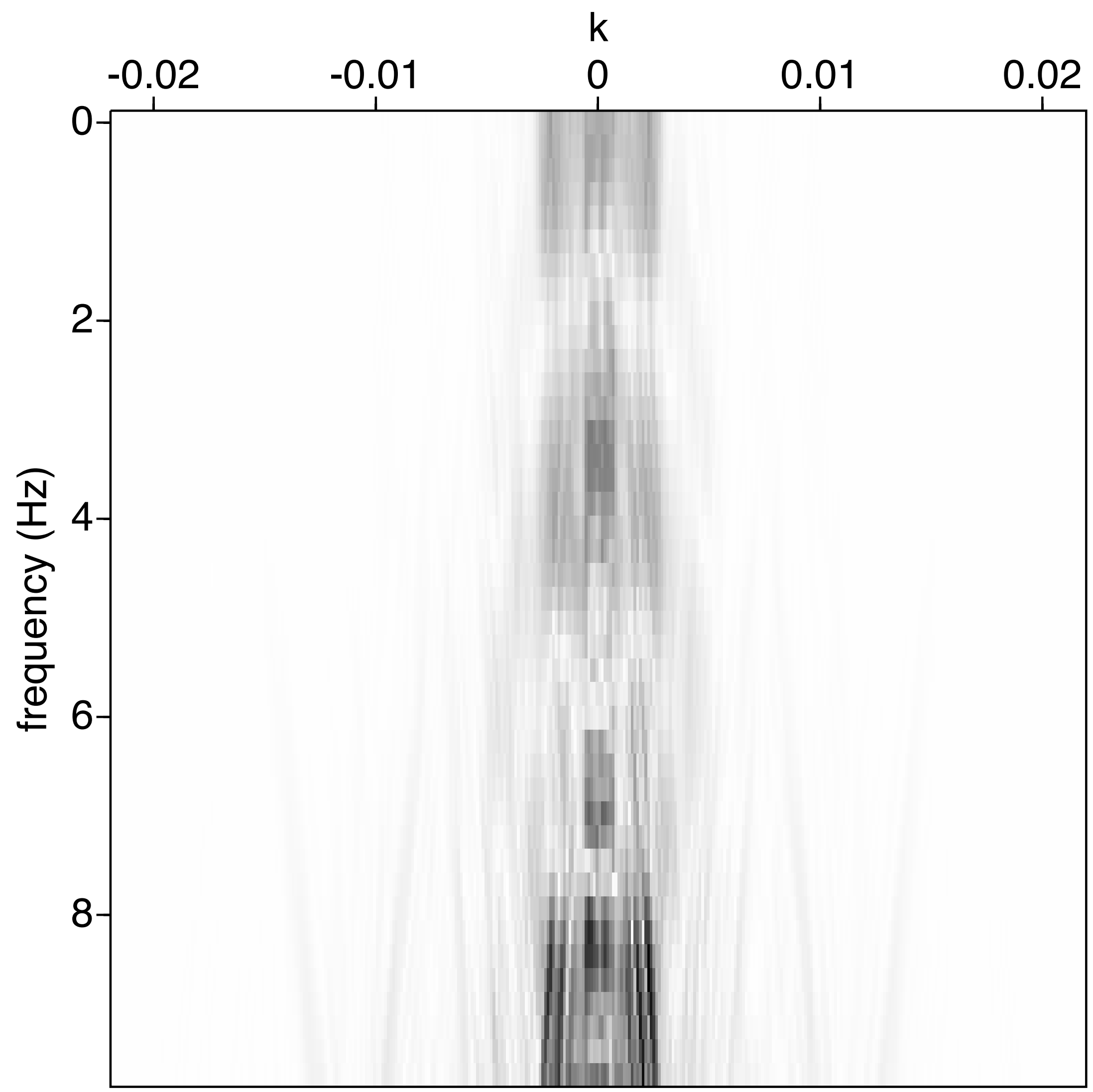
F-K Spectrum of data



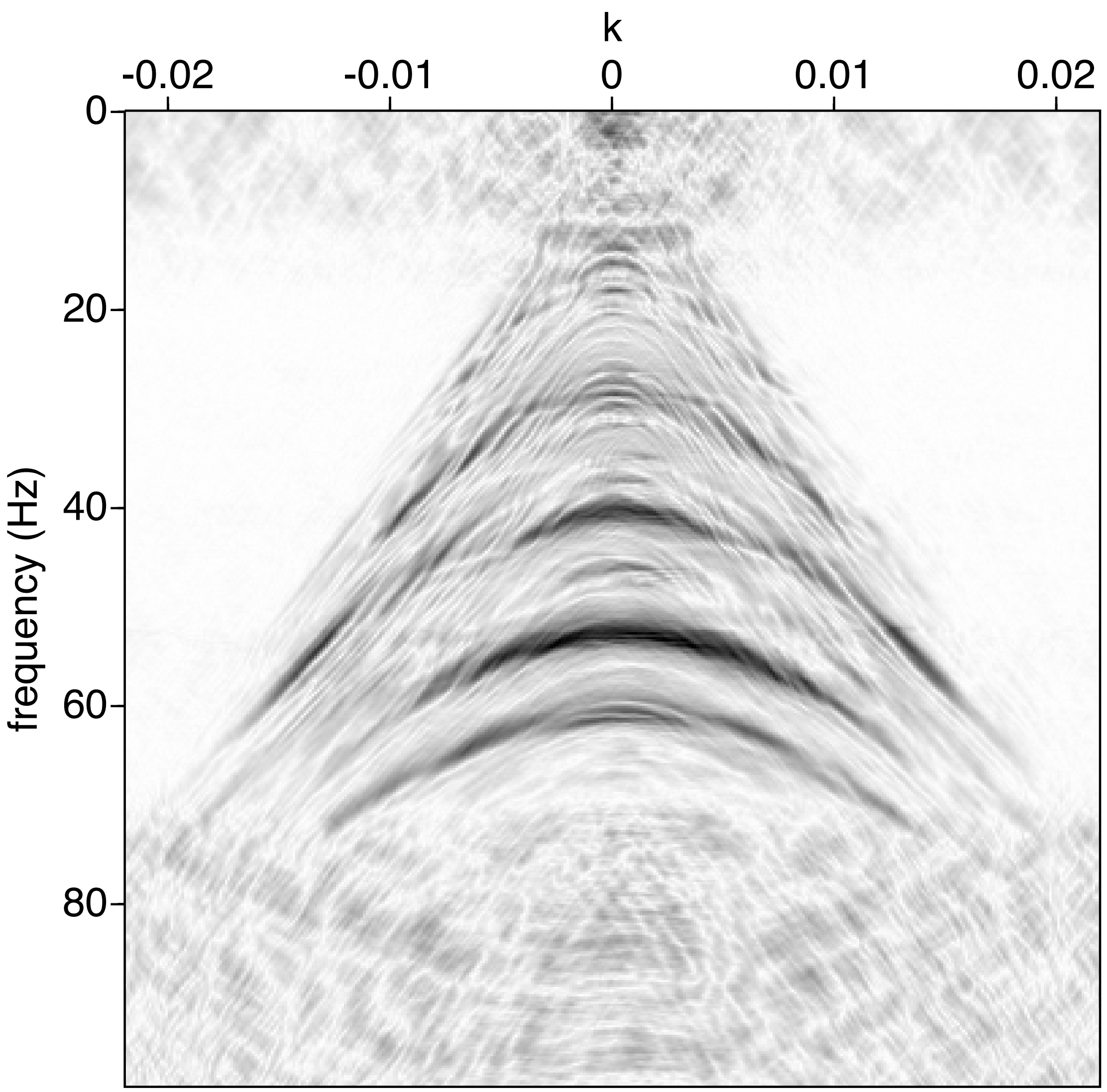
F-K Spectrum of REPSI+Transform Primary IR



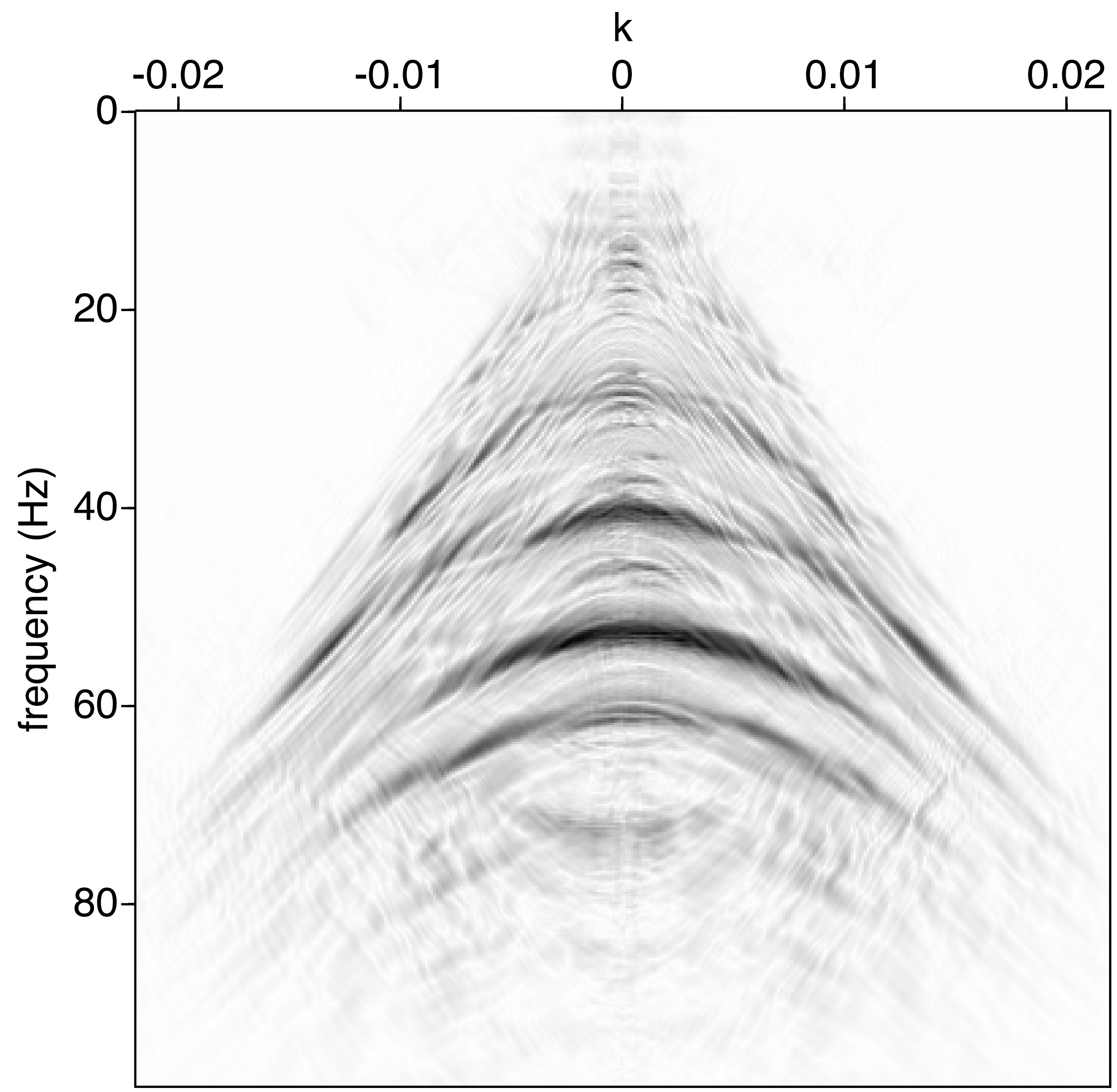
F-K Spectrum of data



F-K Spectrum of REPSI+Transform Primary IR



F-K Spectrum of REPSI



F-K Spectrum of REPSI+Transform Primary IR

Further reading

Compressive sensing

- *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information by Candes, 06*.
- *Compressed Sensing* by D. Donoho, '06

Simultaneous simulations, imaging, and full-wave inversion:

- *Faster shot-record depth migrations using phase encoding* by Morton & Ober, '98.
- *Phase encoding of shot records in prestack migration* by Romero et. al., '00.
- *High-resolution wave-equation amplitude-variation-with-ray-parameter (AVP) imaging with sparseness constraints* by Wang & Sacchi, '07
- Efficient Seismic Forward Modeling using Simultaneous Random Sources and Sparsity by N. Neelamani et. al., '08.
- *Compressive simultaneous full-waveform simulation* by FJH et. al., '09.
- *Fast full-wavefield seismic inversion using encoded sources* by Krebs et. al., '09
- *Randomized dimensionality reduction for full-waveform inversion* by FJH & X. Li, '10

Stochastic optimization and machine learning:

- *A Stochastic Approximation Method* by Robbins and Monro, 1951
- *Neuro-Dynamic Programming* by Bertsekas, '96
- *Robust stochastic approximation approach to stochastic programming* by Nemirovski et. al., '09
- *Stochastic Approximation approach to Stochastic Programming* by Nemirovski
- *An effective method for parameter estimation with PDE constraints with multiple right hand sides. by* Eldad Haber, Matthias Chung, and Felix J. Herrmann. '10
- Seismic waveform inversion by stochastic optimization. Tristan van Leeuwen, Aleksandr Aravkin and FJH, 2010.

Full-waveform inversion with extensions

- *Migration velocity analysis and waveform inversion* by Symes Geophysical Prospecting, 56: 765–790, 2008.
- *The seismic reflection inverse problem* by Symes, Inverse Problems 25, 2009.

Thank you

slim.eos.ubc.ca