Compressive Sensing and Sparse Recovery in Exploration Seismology

Felix J. Herrmann

SLIM
Seismic Laboratory for Imaging and Modeling
the University of British Columbia

Compressive Sensing and Sparse Recovery in Exploration Seismology

Aleksandr Aravkin
Tristan van Leeuwen
Xiang Li

SLIM

Seismic Laboratory for Imaging and Modeling
the University of British Columbia

Drivers

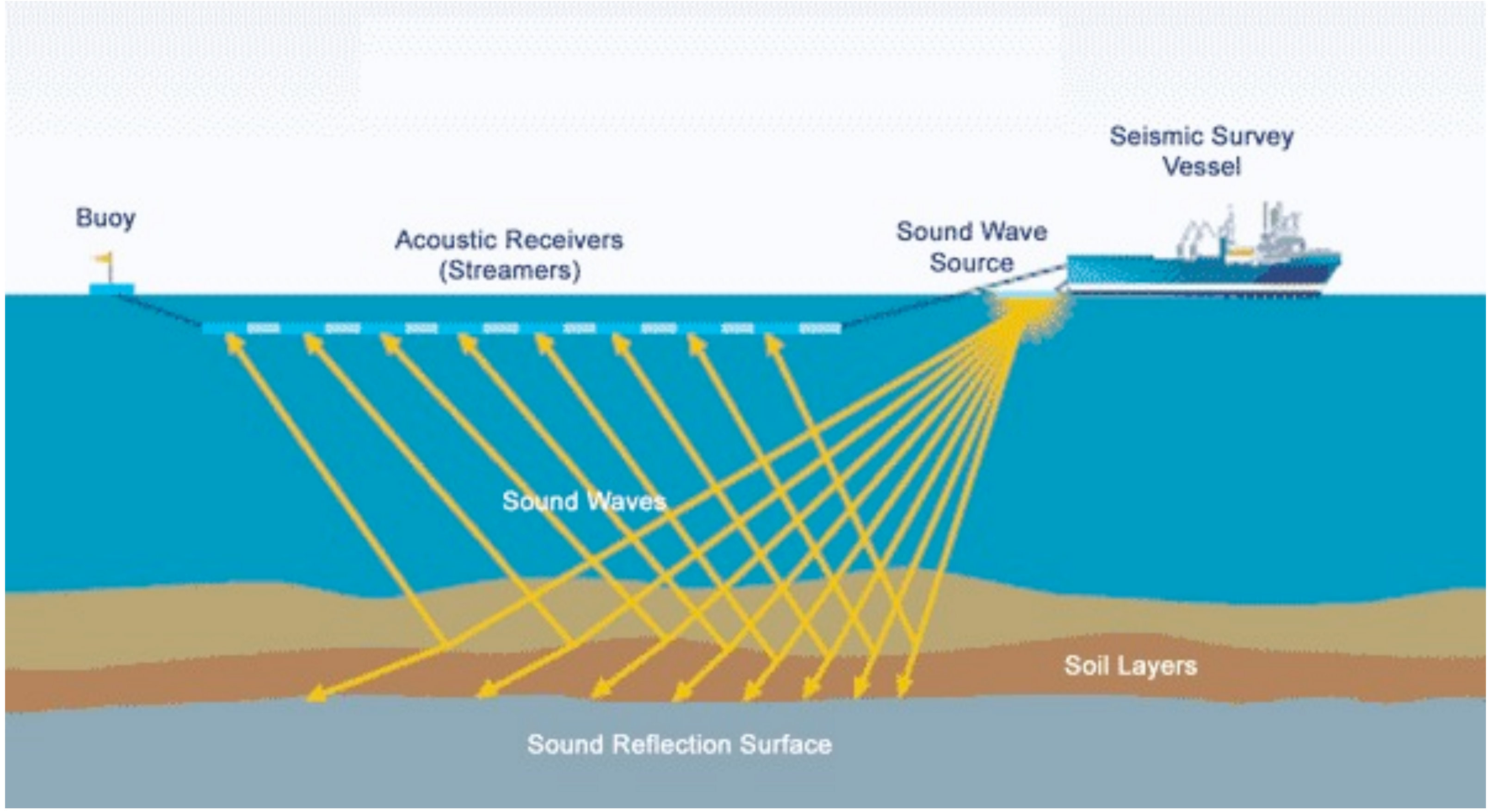
Our incessant

- demand for *hydrocarbons* while we are *no* longer finding oil...
- desire to understand the Earth's inner workings

Push for improved *seismic inversion* to

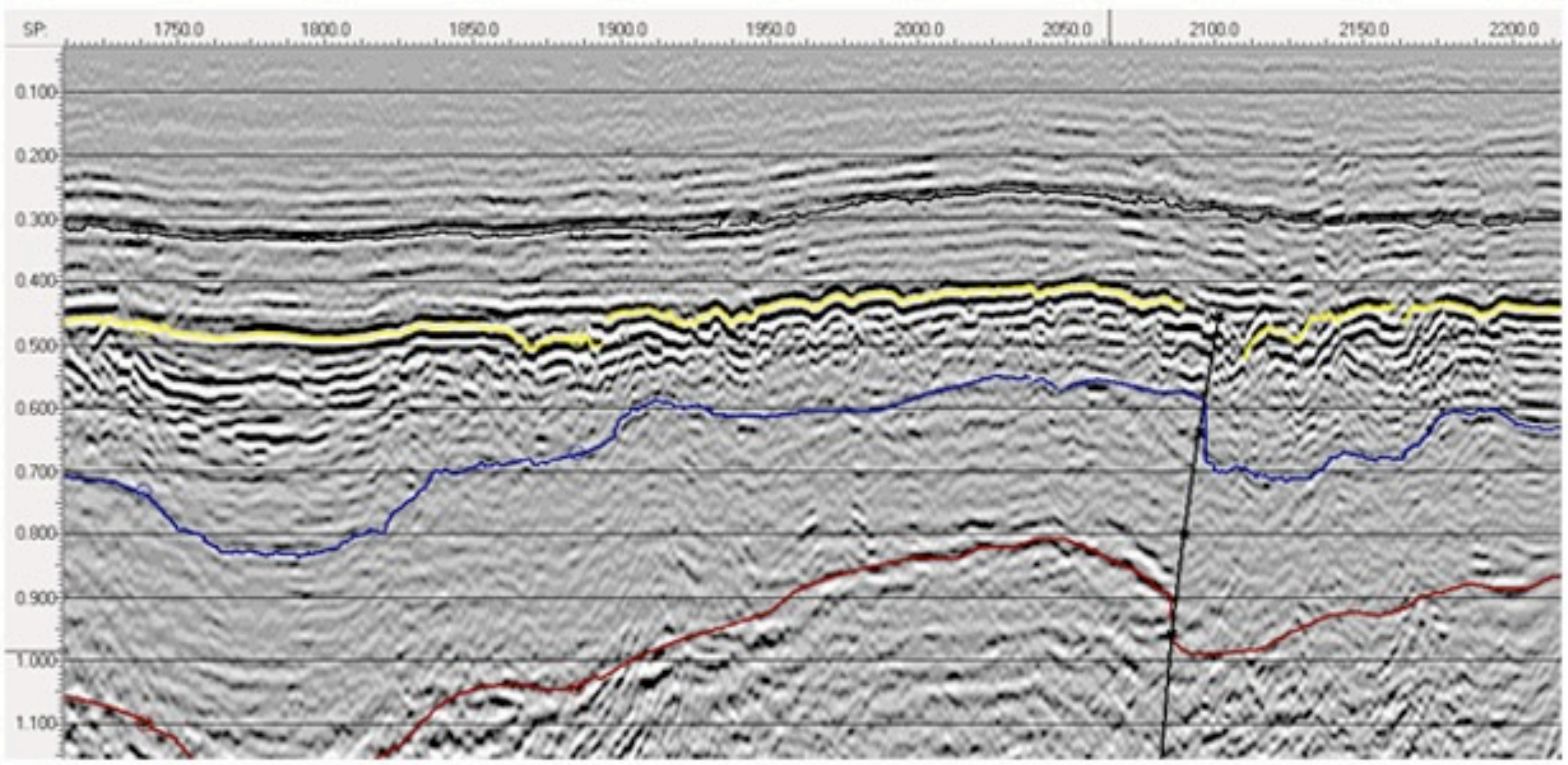
- create *more high-resolution* information
- from *more* and *more data*... (moving to 100k channel systems)

Seismic survey



http://fishsafe.eu/en/offshore-structures/seismic-surveys.aspx

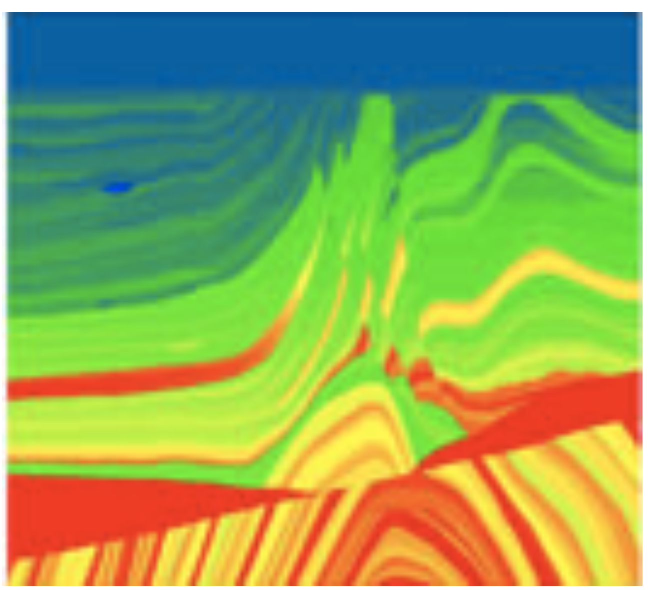
Seismic image



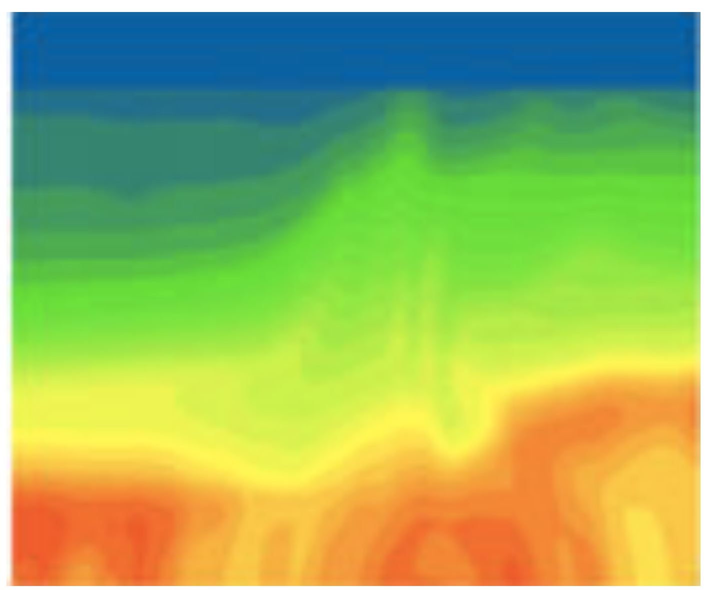
http://www.gentechintl.com/seismic.htm

Full-waveform inversion (FWI)

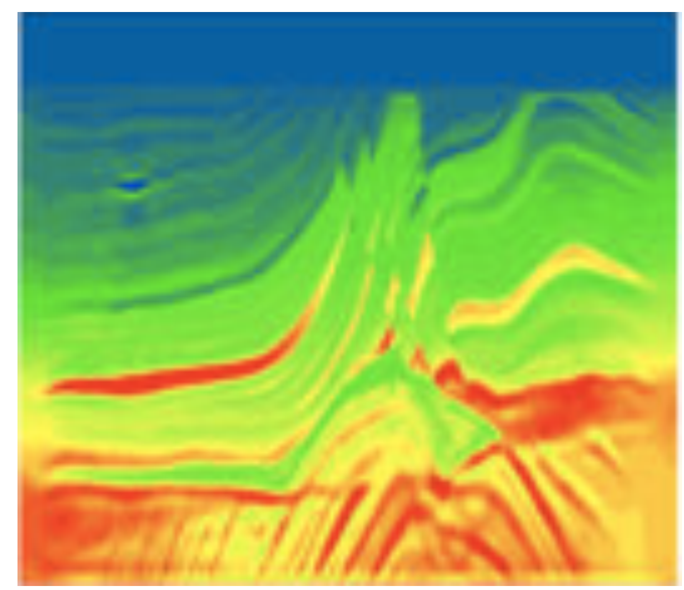
true model



starting model



inverted model



http://www.westerngeco.com/services/dp/omega/depth/tomoportfolio/fwi.aspx

Wish list

Inversion costs determined by *structure of data & complexity* of the *subsurface*

- sampling & computational *costs* that are *dictated by sparsity* and *not* by the *dimensionality* of the problem (e.g. *size* of the *discretization*)

Controllable error that depends on

- degree of *subsampling / dimensionality* reduction
- available computational resources

SLIM

[Tarantola, 84; Pratt, '98; Plessix, '06]

Problem statement

PDE-constrained optimization problem (*unconstrained form*):

$$\min_{\mathbf{m}} \frac{1}{2N} \sum_{j=1}^{n_f} \sum_{i=1}^{n_s} \|\mathbf{d}_{i,j} - \boldsymbol{\mathcal{F}}_{i,j}[\mathbf{m}, \mathbf{q}_{i,j}]\|_2^2 \quad \text{with} \quad \boldsymbol{\mathcal{F}}_{i,j}[\mathbf{m}; \mathbf{q}_{i,j}] := \mathbf{P}_i \mathbf{H}_j^{-1}[\mathbf{m}]\mathbf{q}_{i,j},$$

$$\begin{aligned}
\mathbf{d}_{i,j} &= \text{Monochromatic data from source } i \text{ and frequency } j \\
\mathbf{P}_i &= \text{Detection operator for source } i \\
\mathbf{H}_j^{-1} &= \text{Inverse of time-harmonic Helmholtz at frequency } j \\
\mathbf{q}_{i,j} &= \text{Seismic source } i \text{ at frequency } j \\
\mathbf{m} &= \text{Unknown model, e.g. } c^{-2}(x) \\
N &= n_s \times n_f \quad \text{('batch size')}
\end{aligned}$$

SLIM

[Tarantola, 84; Pratt, '98; Plessix, '06]

Simplification

Multiexperiment optimization problem:

$$\min_{\mathbf{m}\in\mathcal{M}} \frac{1}{2}\|\mathbf{D} - \mathcal{F}[\mathbf{m};\mathbf{Q}]\|_{2,2}^2 \quad \text{with} \quad \mathcal{F}[\mathbf{m};\mathbf{Q}] := \mathbf{P}\mathbf{H}^{-1}[\mathbf{m}]\mathbf{Q}$$

$$\begin{aligned} \mathbf{D} &= \text{Total multi-source and multi-frequency data volume} \\ \mathbf{P} &= \text{Single detection operator} \\ \mathbf{H}^{-1} &= \text{Inverse of time-harmonic Helmholtz} \\ \mathbf{Q} &= \text{Seismic sources} \\ \mathbf{m} &= \text{Unknown model, e.g. } c^{-2}(x) \end{aligned}$$

Properties

Multiexperiment optimization problem:

$$\min_{\mathbf{m}\in\mathcal{M}} \frac{1}{2}\|\mathbf{D}-\mathcal{F}[\mathbf{m};\mathbf{Q}]\|_{2,2}^2 \quad \text{and} \quad \mathbf{H}[\mathbf{m}]\cdot := [\omega^2 \mathrm{diag}\{\mathbf{m}\} + \nabla^2]\cdot$$

- hyperbolic PDE, non convex, 'over-' and 'underdetermined'
- wave-equation Hessian, $\nabla\mathcal{F}^H[\mathbf{m};\mathbf{Q}]\nabla\mathcal{F}[\mathbf{m};\mathbf{Q}]$, is *pseudo local*, i.e., 'preserves' singularities
- # PDE solves increases linearly with # of sources & frequencies
- *linear* in the *sources*

[Tarantola, 84; Pratt, '98; Plessix, '06; Symes '09]

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**

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

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

SLIM

Related work

Approximations of the Hessian

- Matrix probing: a randomized preconditioner for the wave-equation Hessian [FJH et. all, '03,'09; Demanet '08-'10]
- accurate linearization & high-frequency asymptotics
- redone for each GN iteration

Randomized-dimensionality reduction

- Randomized Kaczmarz [Strohmer & Vershynen, '09; Eldar & Needell '10]
- Faster Least Squares Approximation [Drineas, Mahoney, Muthukrishnan, and Sarlos, '07]
- Blendenpik: supercharging LAPACK's LS-solver [Avron et.al., '10]
- full overdetermined *explicit* matrices

Our approach

Combine techniques from

- compressive sensing (fast phase encoders)
- stochastic optimization (stochastic approximation)

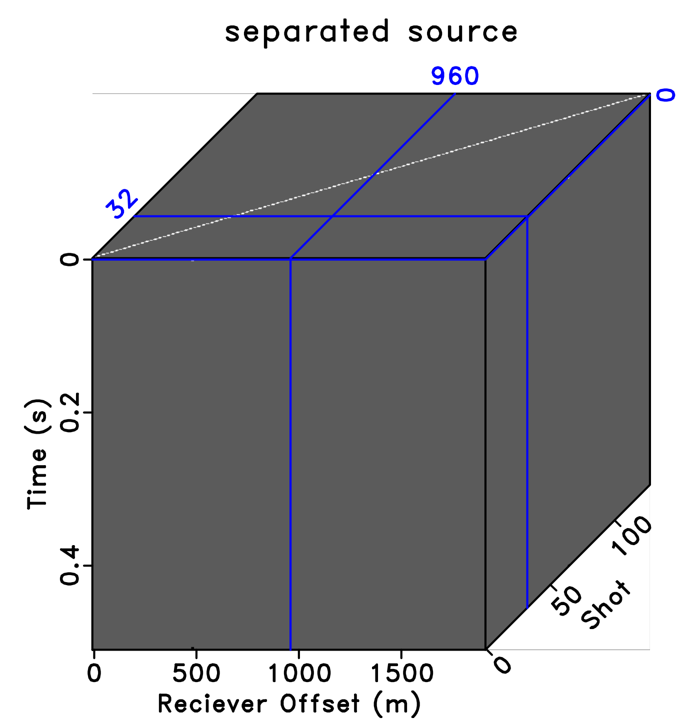
Exploit

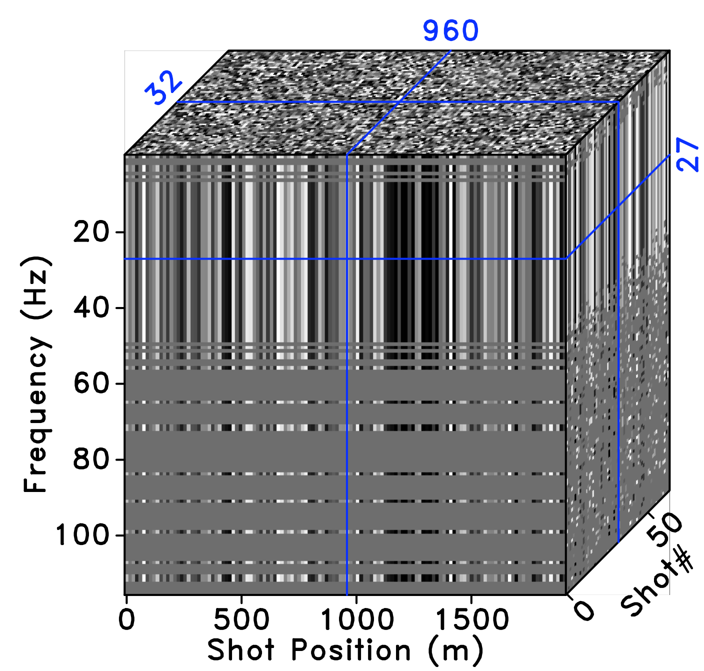
- block structure PDE-constrained optimization problem
- curvelet-domain sparsity
- convexity subproblems & properties Pareto curve

[FJH et. al. ’08-’10]

CS experiment

adapted from FJH et. al. ,09

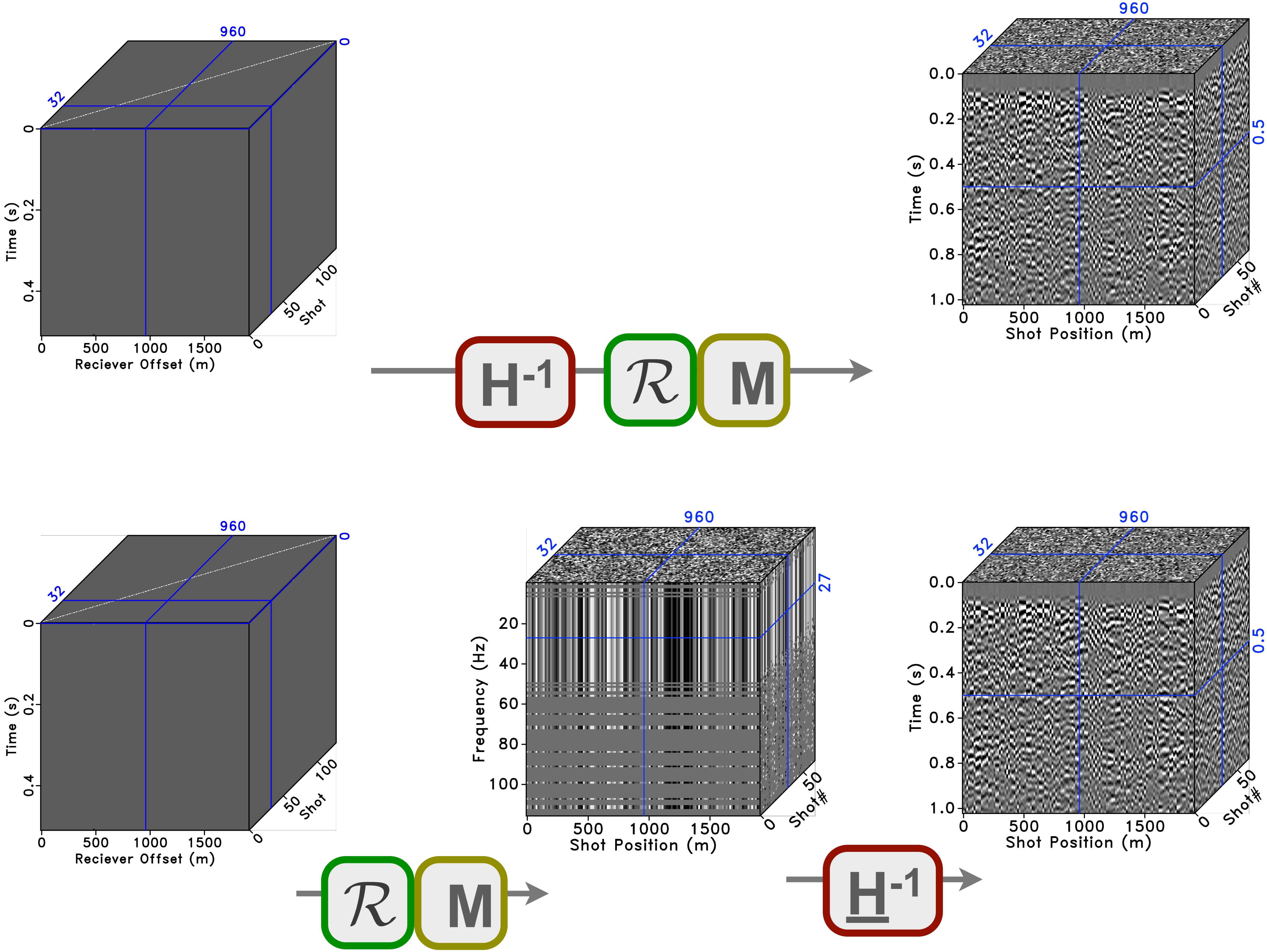




$$\mathbf{Q}$$

$$\underline{\mathbf{Q}} = \mathbf{RMQ}$$

- collection of K *simultaneous-source* experiments (*supershots*)
- $K = n'_f \times n'_s \ll n_f \times n_s$



960
32
0
Time (s)
0.2
0.4
100
50
Shot
0
500
1000
1500
Reciever Offset (m)
H-1
M
0.0
0.6
0.8
1.0
0.5
Shot#
Shot Position (m)
Frequency (Hz)
20
40
60
80
100
27

Math [Romberg, '07, FJH, '08-'10]

Fast ($n \log n$) compressive-sampling operator

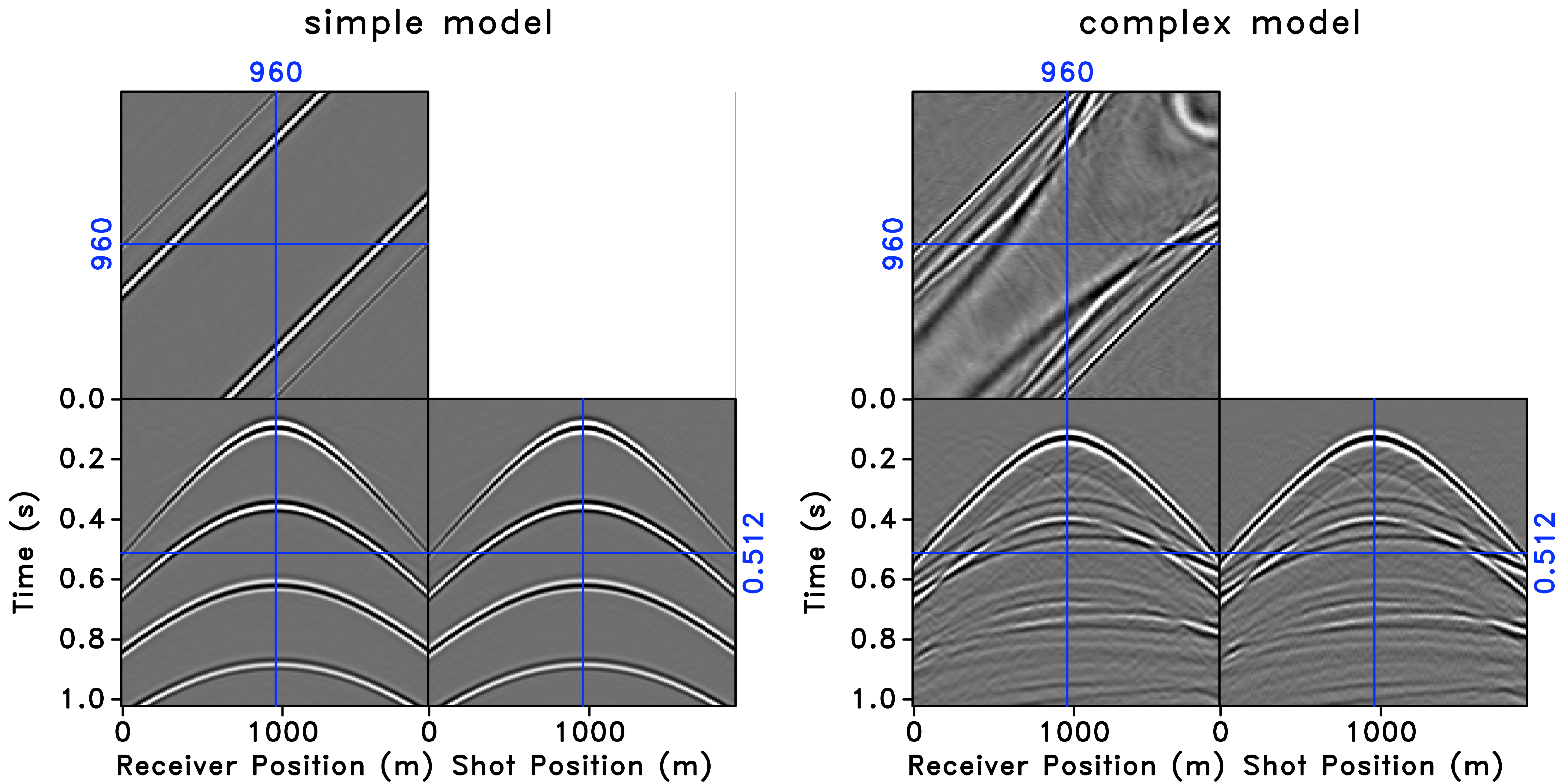
$$\mathbf{RM} = \mathrm{vec}^{-1}\left[(\mathbf{RM})_{1\cdots n_s'}\right]\mathrm{vec}$$

with
$$(\mathbf{RM})_k = (\mathbf{R^\Sigma}_k \mathbf{M^\Sigma} \otimes \mathbf{I} \otimes \mathbf{R^\Omega}_k)$$

and
$$\mathbf{M^\Sigma} = \overbrace{\mathrm{sign}(\eta) \odot \mathbf{F_\Sigma}^H e^{j\theta}\mathbf{F_\Sigma}}^{\text{'Gaussian matrix'}}$$

where $\theta \in \mathrm{Uniform}(-\pi,\, \pi]$, and $\eta \in \mathrm{Normal}(0, 1)$

Recovered Green's functions



28.1dB

18.2dB

300 SPGL1 iteration

Bottom line

Computational cost for the ℓ_1-solver is less $(\mathcal{O}(n^3 \log n)$ vs. $\mathcal{O}(n^4))$ than the cost of solving Helmholtz...

Problem:

- data space too large in 3D acquisition (1000^5 - $100k^5$)
- have to resimulate for each gradient update...

SLIM

Reduced FWI formulation

Multiexperiment simultaneous-source optimization problem:

$$\min_{\mathbf{m}\in\mathcal{M}} \frac{1}{2}\|\underline{\mathbf{D}} - \mathcal{F}[\mathbf{m};\underline{\mathbf{Q}}]\|_{2,2}^2 \quad \text{with} \quad \mathcal{F}[\mathbf{m};\underline{\mathbf{Q}}] := \mathbf{P}\underline{\mathbf{H}}^{-1}\underline{\mathbf{Q}}$$

- requires *smaller* number of PDE solves
- explores *linearity* in the *sources* & *block-diagonal* structure of the *Helmholtz system*
- uses *randomized* frequency *selection* and *phase encoding*

Interpretations

Consider *randomized*-dimensionality reduction as instances of

- *stochastic optimization* [Haber, Chung, and FJH, '10; van Leeuwen, Aravkin, FJH, '10]
  - *random*-trace estimates [Hutchinson, '90, Avron & Toledo, '10]
  - *stochastic* gradient descent [Bertsekas,' '96; Nemirovski, '09]
- *"compressive sensing"* [FJH et. al, '08-'10]

Stochastic optimization

Replace *deterministic*-optimization problem

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

with *sum* cycling over *different sources & corresponding monochromatic shot records* (columns of D & Q)

[Natterer, '01]

Stochastic *average* approximation [Haber, Chung, and FJH, '10]

by *a stochastic-optimization* problem

$$\min_{\mathbf{m}\in\mathcal{M}} \mathbf{E_w}\{f(\mathbf{m},\mathbf{w})\} = \frac{1}{2}\|\mathbf{Dw}-\mathcal{F}[\mathbf{m};\mathbf{Qw}]\|_2^2\}$$

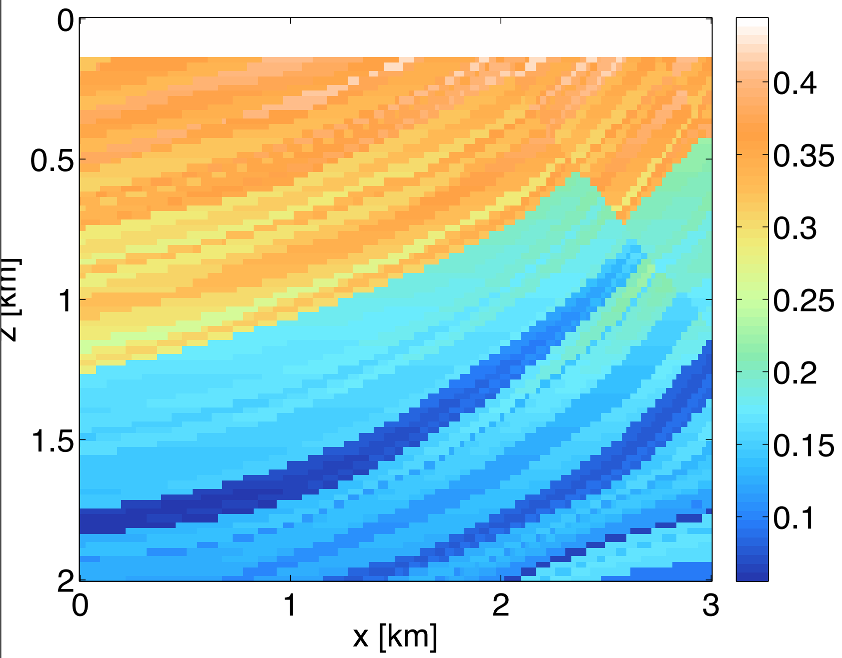
$$\approx \frac{1}{K}\sum_{j=1}^{K}\frac{1}{2}\|\underline{\mathbf{d}}_j-\mathcal{F}[\mathbf{m};\underline{\mathbf{q}}_j]\|_2^2$$

with $\mathbf{w}\in N(0,1)$ and $\mathbf{E_w}\{\mathbf{w}\mathbf{w}^H\}=\mathbf{I}$

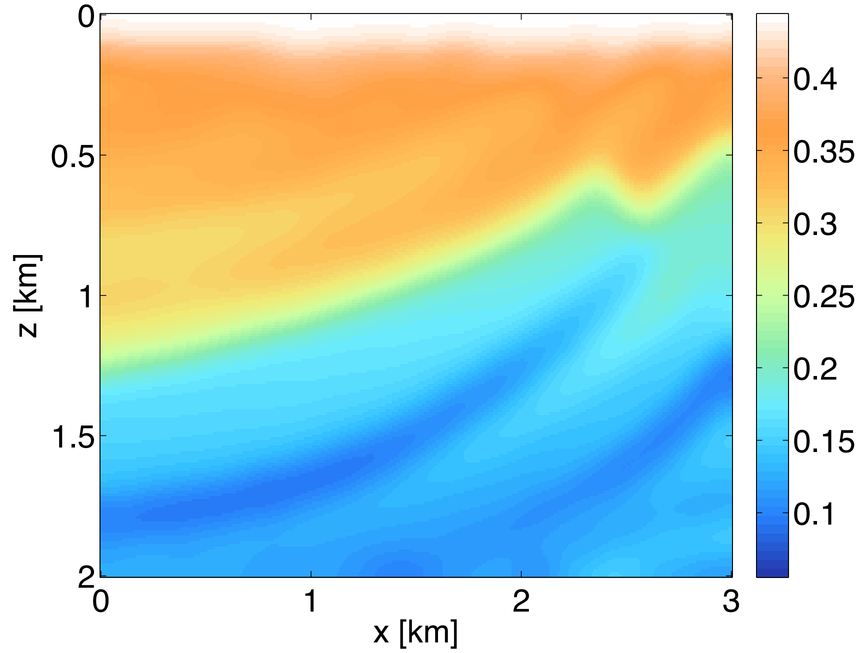
and $\underline{\mathbf{d}}_j=\mathbf{Dw}_j,\ \underline{\mathbf{q}}_j=\mathbf{Qw}_j$

Stylized example

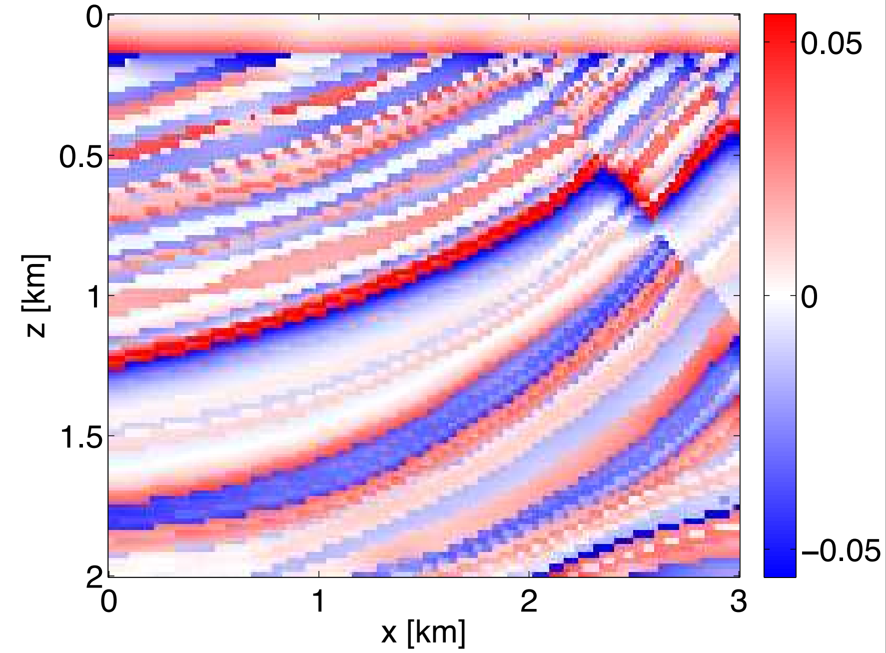
true model



starting model



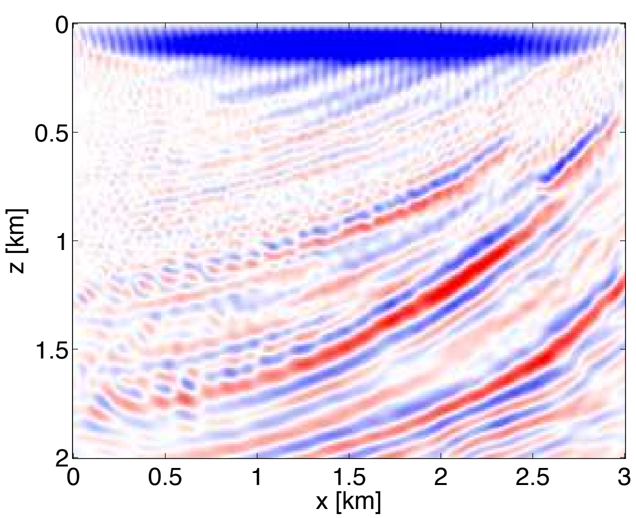
‘reflectivity’



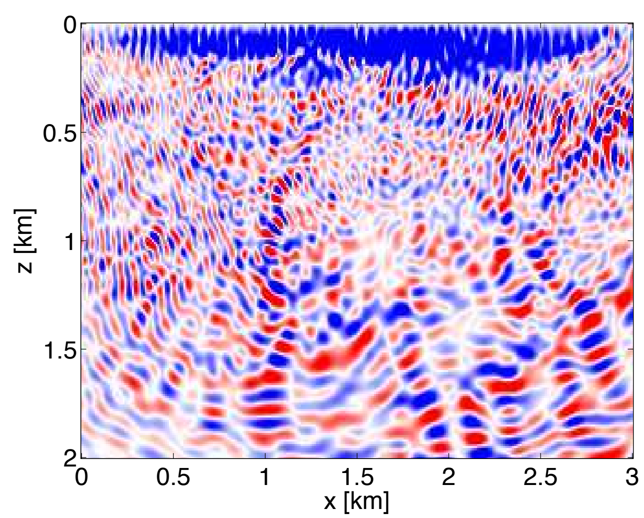
Gradients

Search direction for *increasing* batch size K:

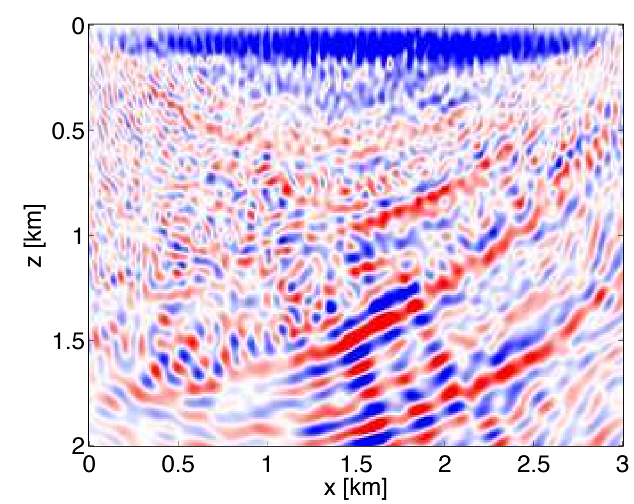
$$\mathbf{g}_K \approx \frac{1}{K}\sum_{j=1}^{K} \nabla\mathcal{F}^*[\mathbf{m};\underline{\mathbf{q}}_j]\delta\underline{\mathbf{d}}_j$$



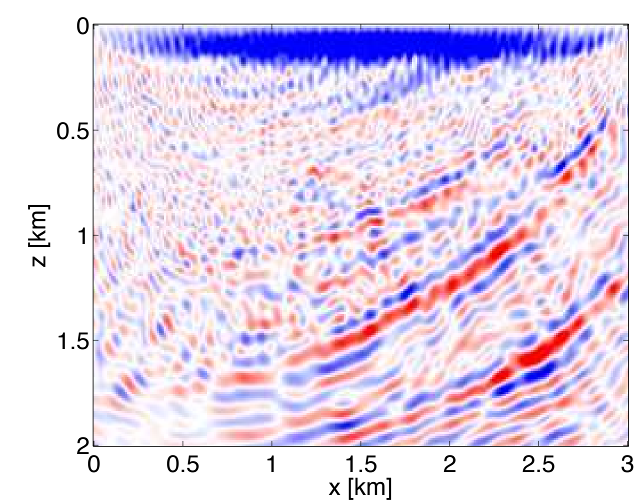
full



K=1

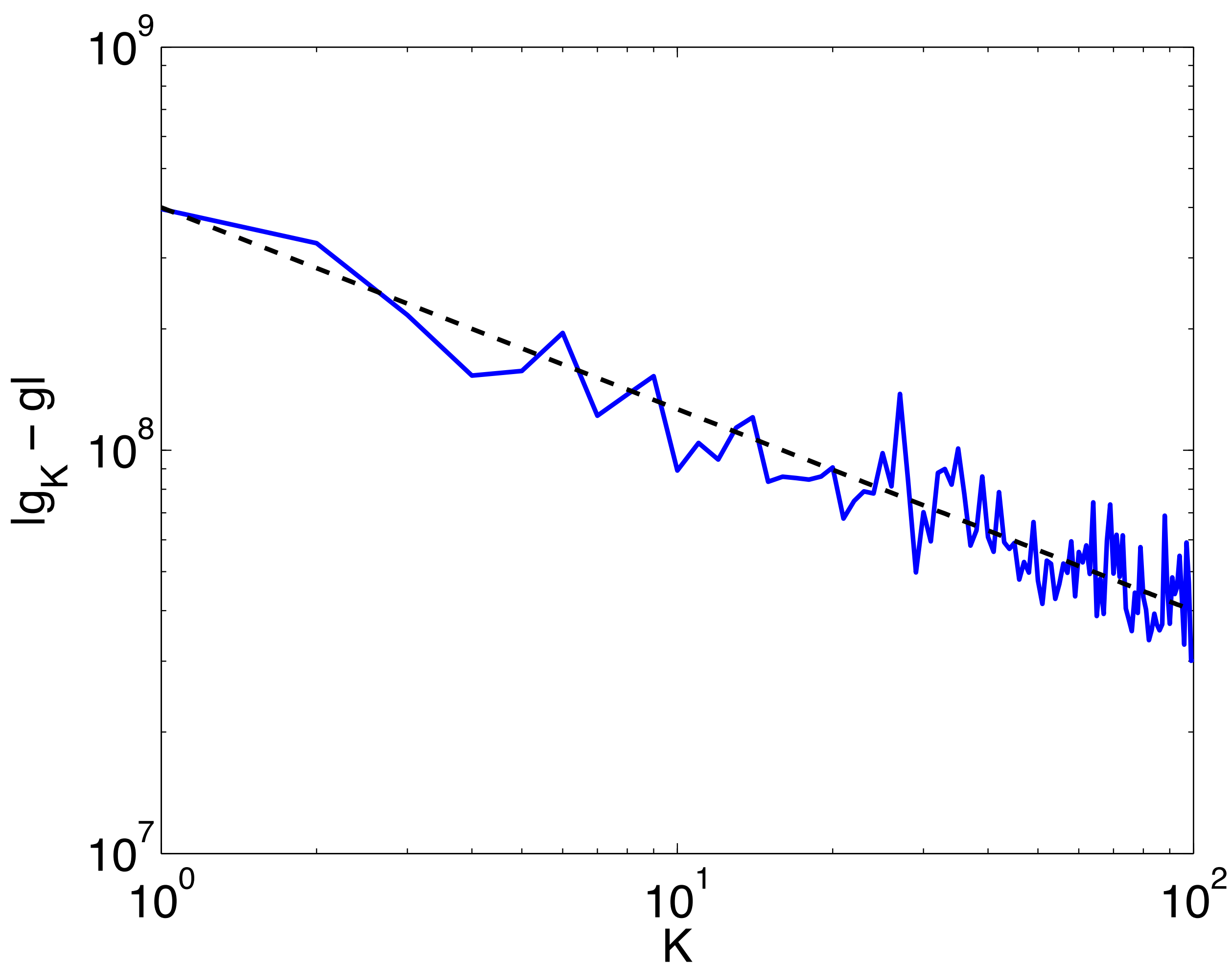


K=5



K=10

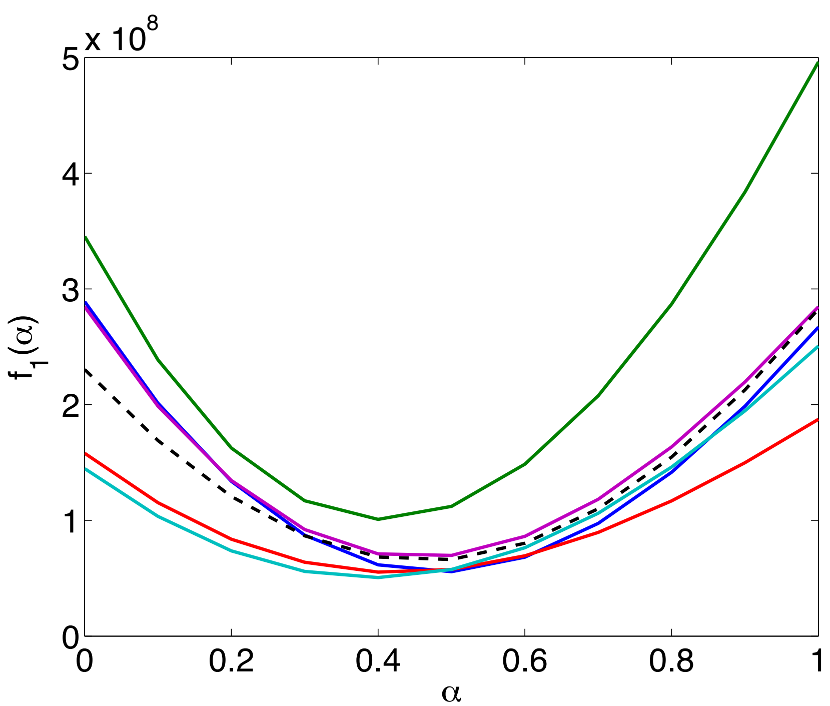
Decay



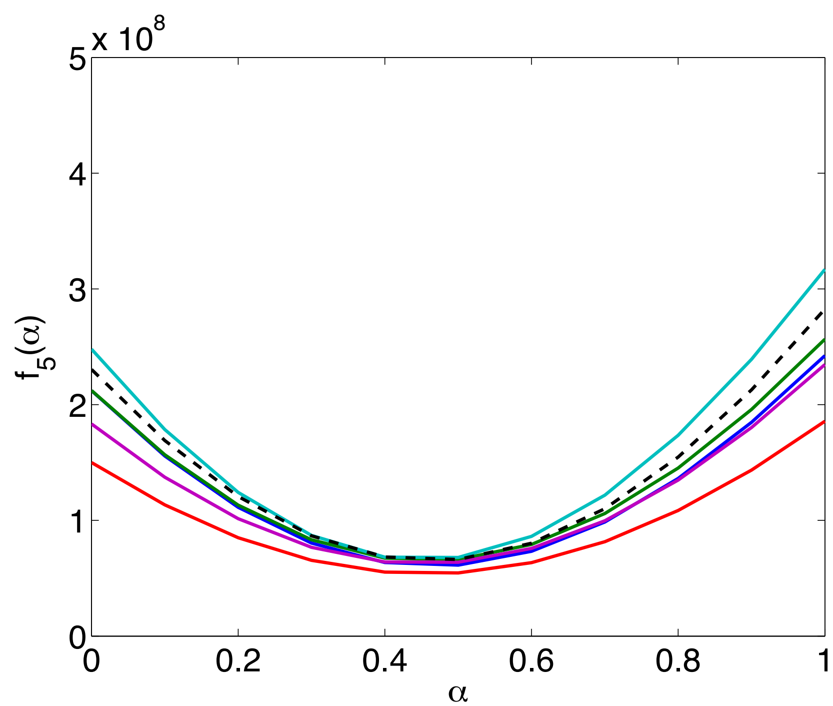
error between full and sampled gradient

Misfit functional

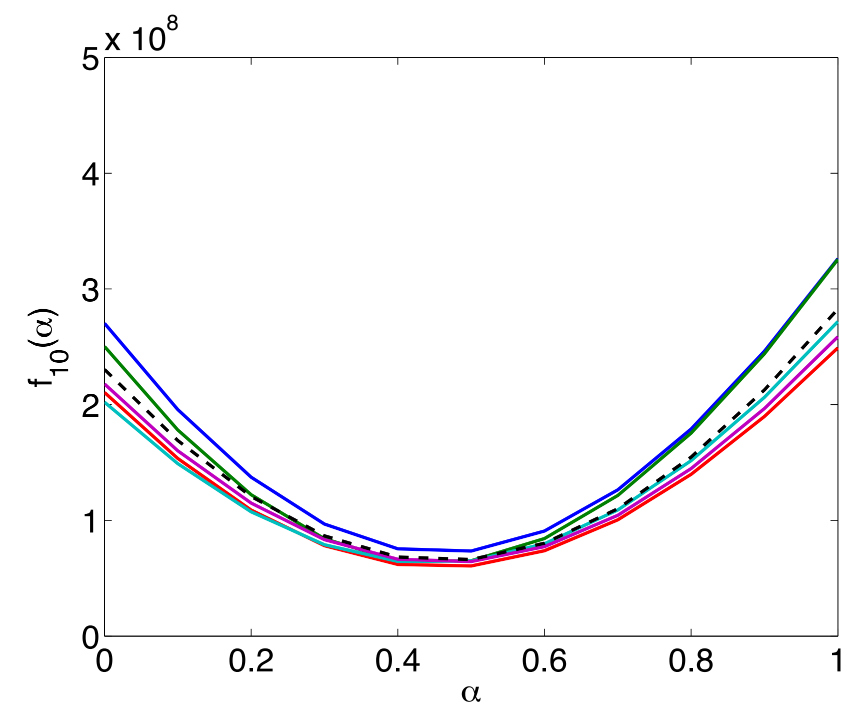
$$f_K(\mathbf{g}_K) = \frac{1}{K}\sum_{j=1}^{K}\frac{1}{2}\|\underline{\mathbf{d}}_j - \mathcal{F}[\mathbf{m}+\alpha\mathbf{g}_K;\underline{\mathbf{q}}_j]\|_2^2$$



K=1



K=5



K=10

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

Stochastic *average* approximation

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

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 [Bertsekas,' '96; Nemirovski, '09]

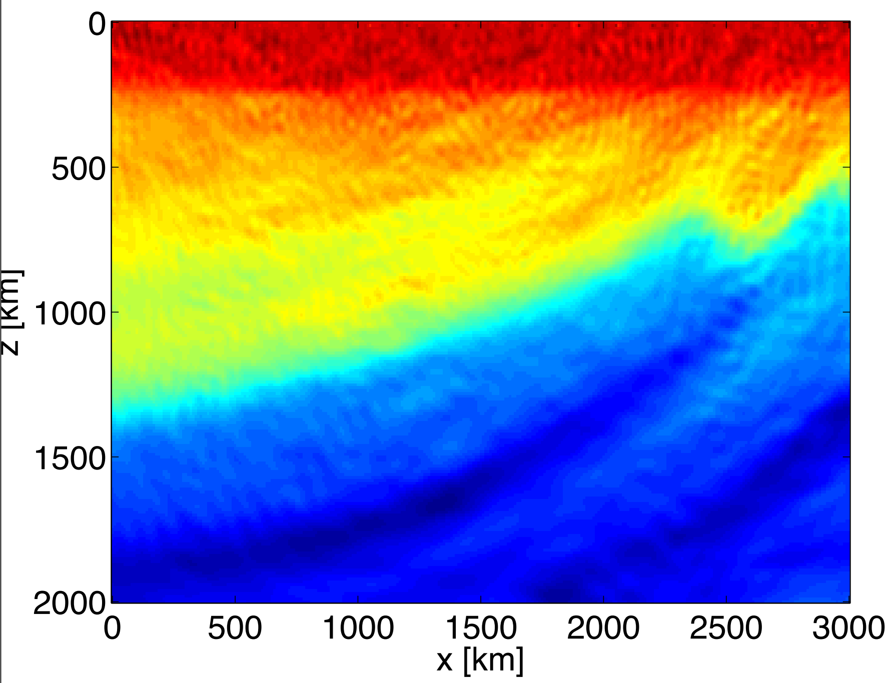
Use *different* simultaneous shots for each *subproblem*, i.e.,

$$\underline{\mathbf{Q}} \quad \mapsto \quad \underline{\mathbf{Q}}^k$$

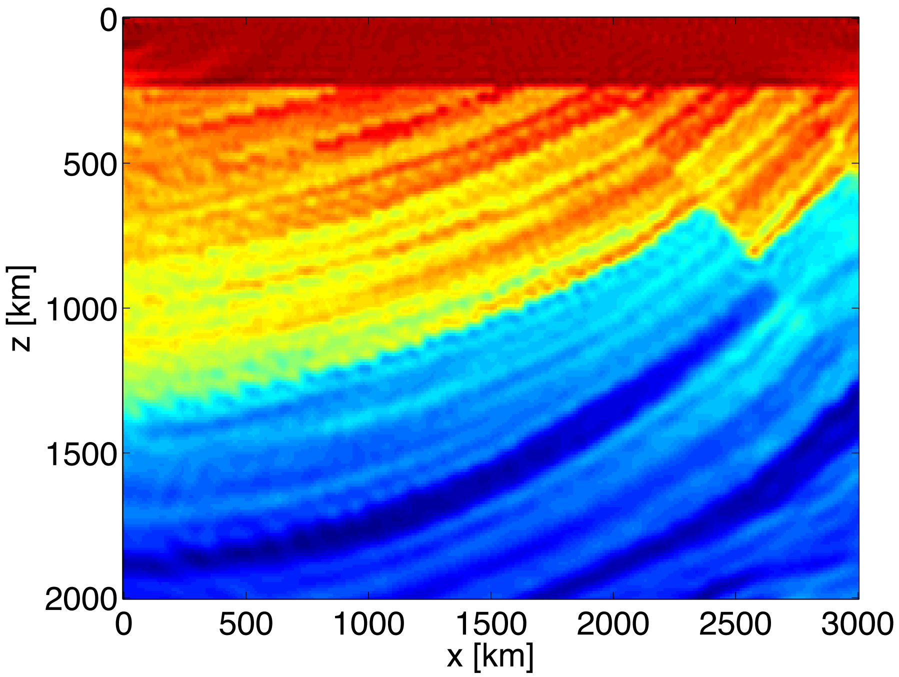
Requires *fewer* PDE solves for each *subproblem*...

- corresponds to the *stochastic approximation*
- related to Kaczmarz ('37) method applied by Natterer, '01
- *supersedes ad hoc* approach by Krebs *et.al.*, '09

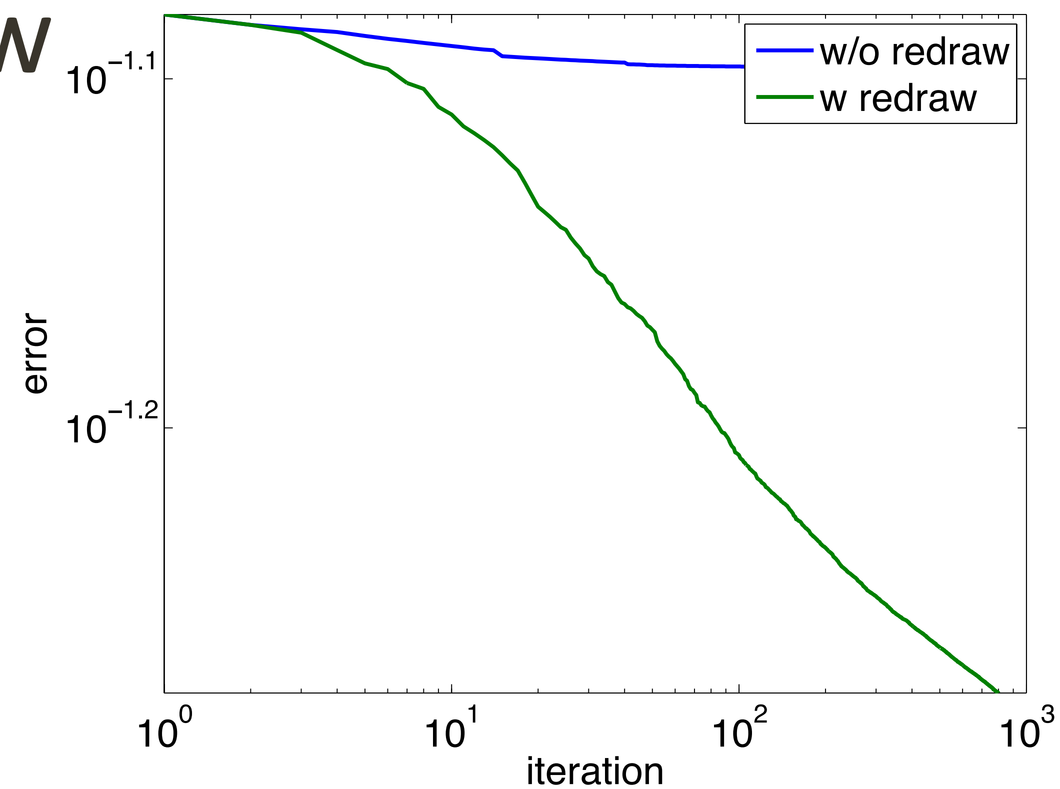
K=1 w and w/o redraw [noise-free case]



w/o redraw



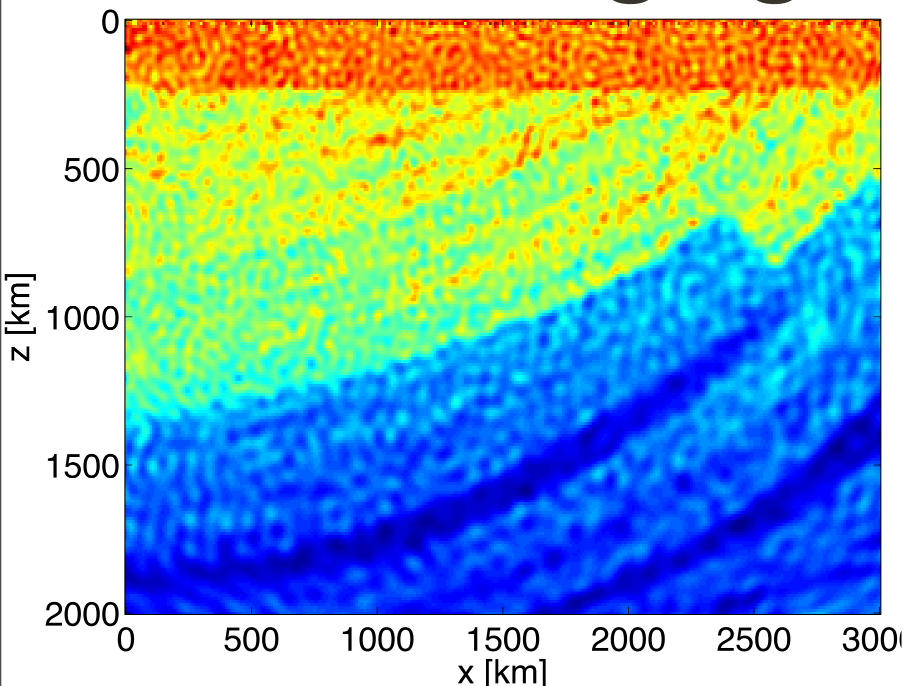
w redraw



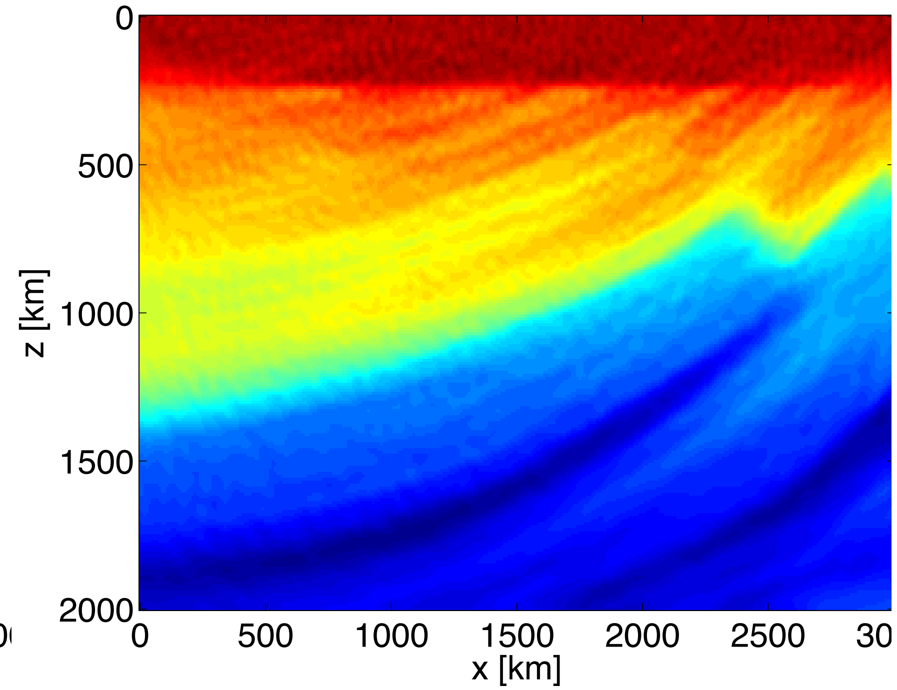
model error K=1

K=1 [noisy case]

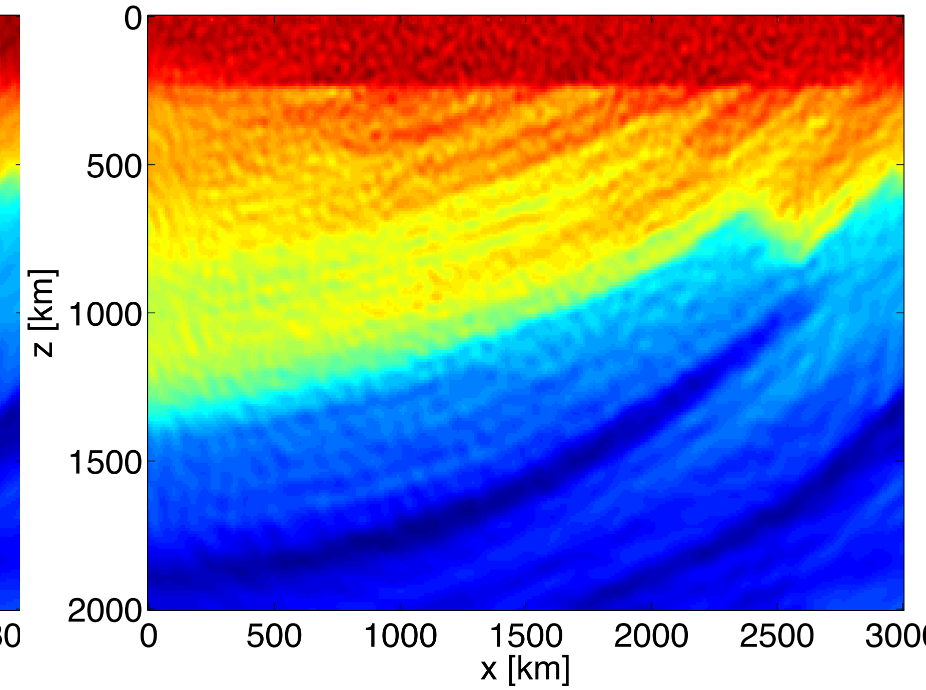
w/o averaging

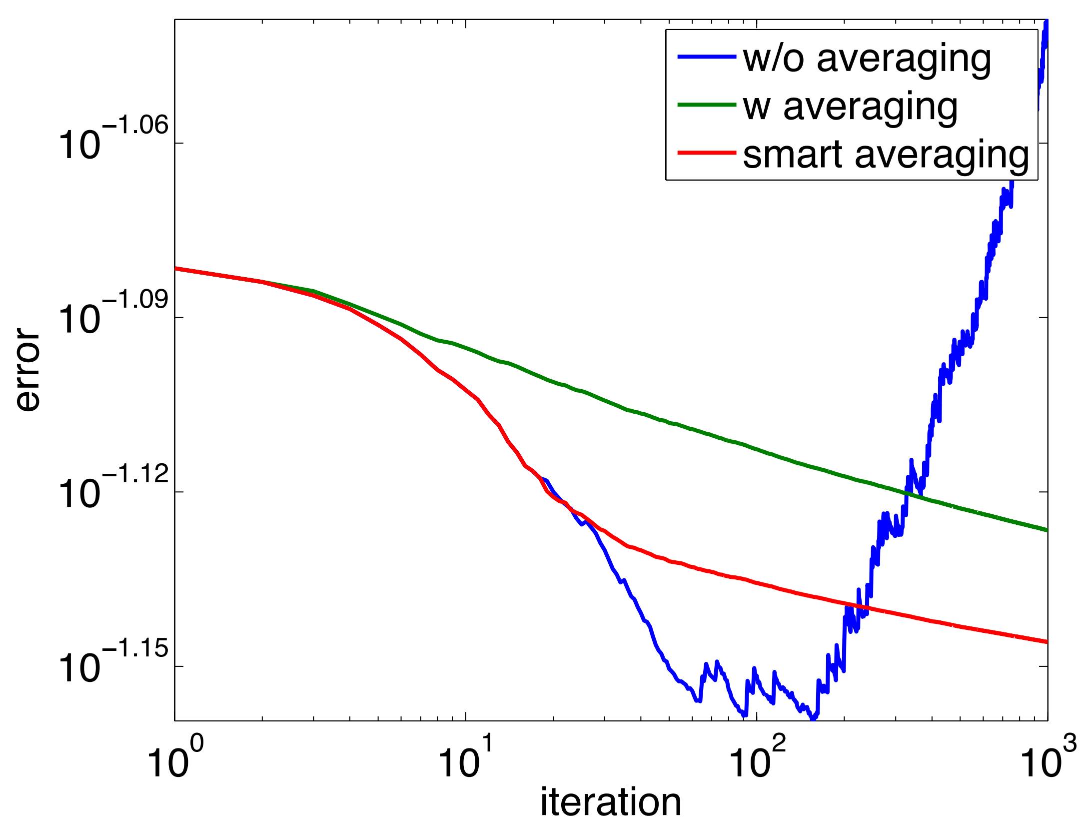


w averaging



smart averaging





Observations

SAA:

- Error decays slowly with batch size K
- becomes worse when noisy

SA

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

Dimensionality reduction gives '*noisy*' results ... Sounds *familiar?*

Combined approach

Leverage findings from *sparse recovery* & *compressive sensing*

- consider *phase-encoded* Gauss-Newton updates as separate "*compressive-sensing* /$\ell_1$regularized experiments"
- remove *interferences* by *curvelet-domain sparsity* promotion
- exploit properties of Pareto curves in combination with stochastic optimization
- turn 'overdetermined' problems with large matrix-setup costs into 'undetermined' problems via *randomization*

Rationale [Smith, '97; Candes & Demanet, '03]

Wavefields are *compressible* in curvelet *frames*

- *correlations* between *source & residual wavefields* are *compressible*
- *velocity* distributions of *sedimentary basins* are *also compressible*

Linearized subproblems are *convex*

Assume proximity Pareto *curves* amongst *successive GN iterations*

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

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

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

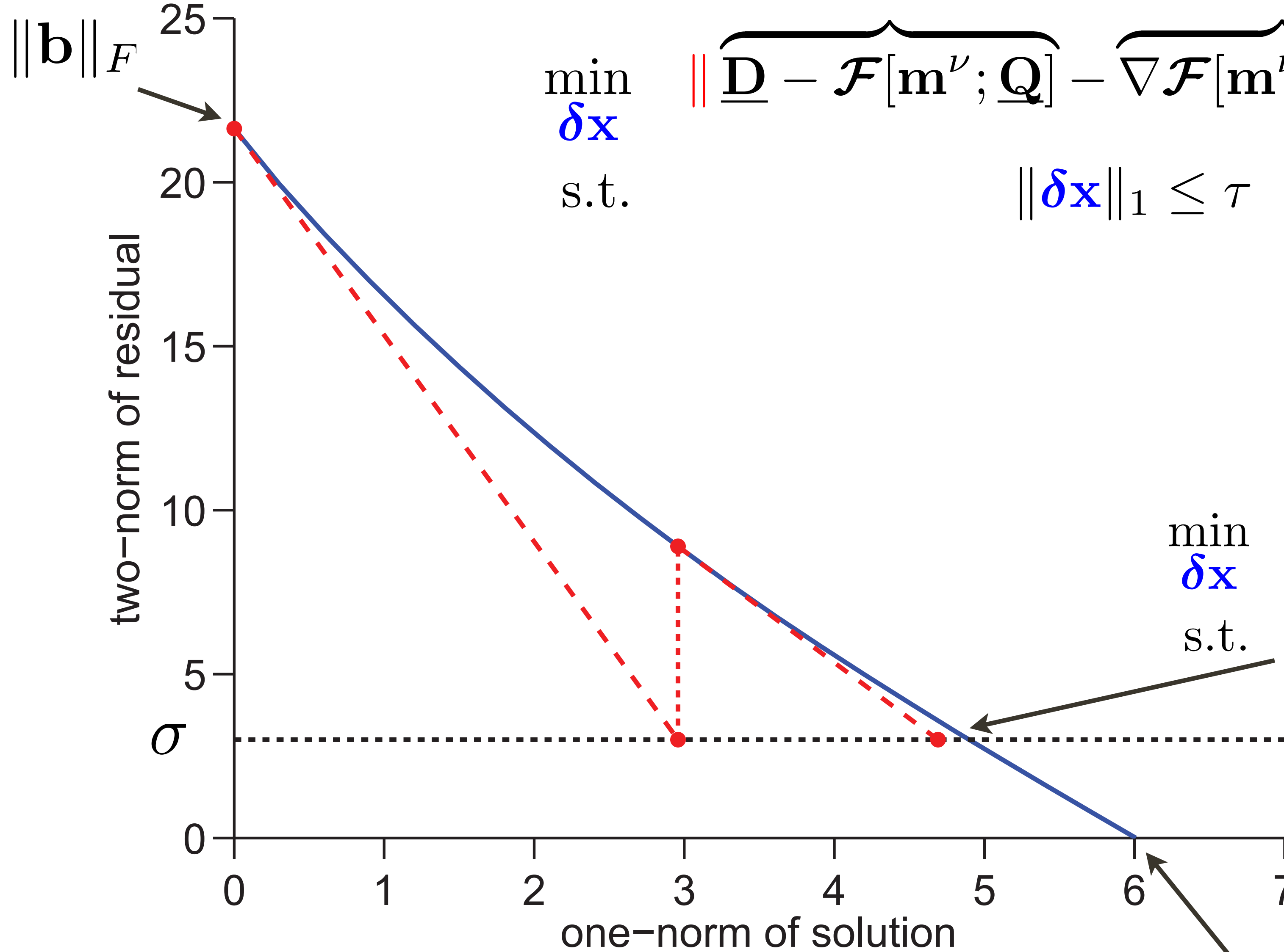
$$\underline{f}'(\mathbf{m}^\nu; \boldsymbol{\mathcal{C}}^* \overline{\boldsymbol{\delta}\mathbf{x}}) \leq \underline{f}(\mathbf{m}^\nu) - \| \underbrace{\underline{\mathbf{D}} - \boldsymbol{\mathcal{F}}[\mathbf{m}^\nu; \mathbf{Q}]}_{\underline{f}(\mathbf{m}^\nu)} - \nabla \boldsymbol{\mathcal{F}}[\mathbf{m};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

[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]

Modified GN with renewals

Algorithm 1: Modified Gauss-Newton with renewals

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

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

for $j = 1 : M$ **do**

Obtain frequency band j, corresponding data slice $\mathbf{D}$ and operator $\boldsymbol{\mathcal{F}}$

while *not converged* **do**

Randomly subsample to obtain $\underline{\mathbf{D}}^k, \underline{\mathbf{Q}}^k$.

Solve with warm start $\overline{\delta\mathbf{x}}$

$$\overline{\delta\mathbf{x}} \longleftarrow \begin{cases} \underset{\delta\mathbf{x}}{\arg\min} \quad \|\underline{\mathbf{D}}^k - \boldsymbol{\mathcal{F}}[\mathbf{m}^k; \underline{\mathbf{Q}}^k] - \nabla\boldsymbol{\mathcal{F}}[\mathbf{m}^k; \underline{\mathbf{Q}}^k]\boldsymbol{\mathcal{C}}^*\delta\mathbf{x}\|_F \\ \qquad\qquad \text{subject to} \quad \|\delta\mathbf{x}\|_1 \leq \tau^k \end{cases}$$

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

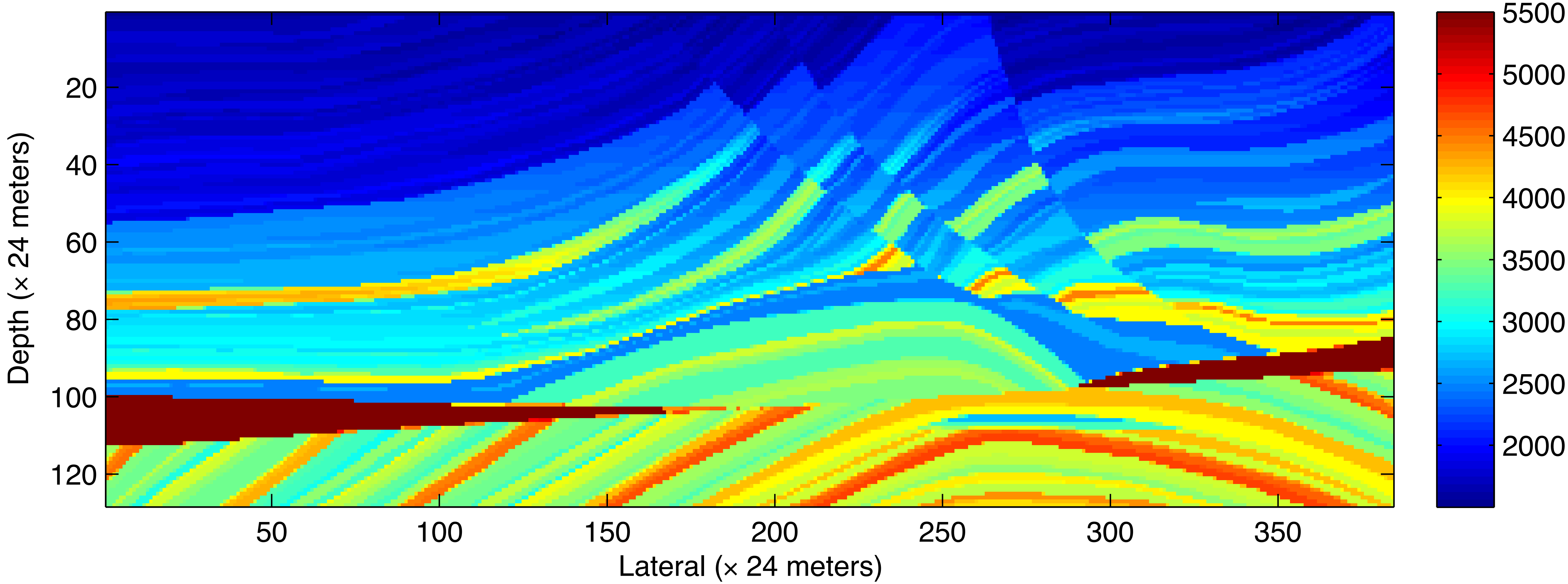
$k \longleftarrow k + 1$

end

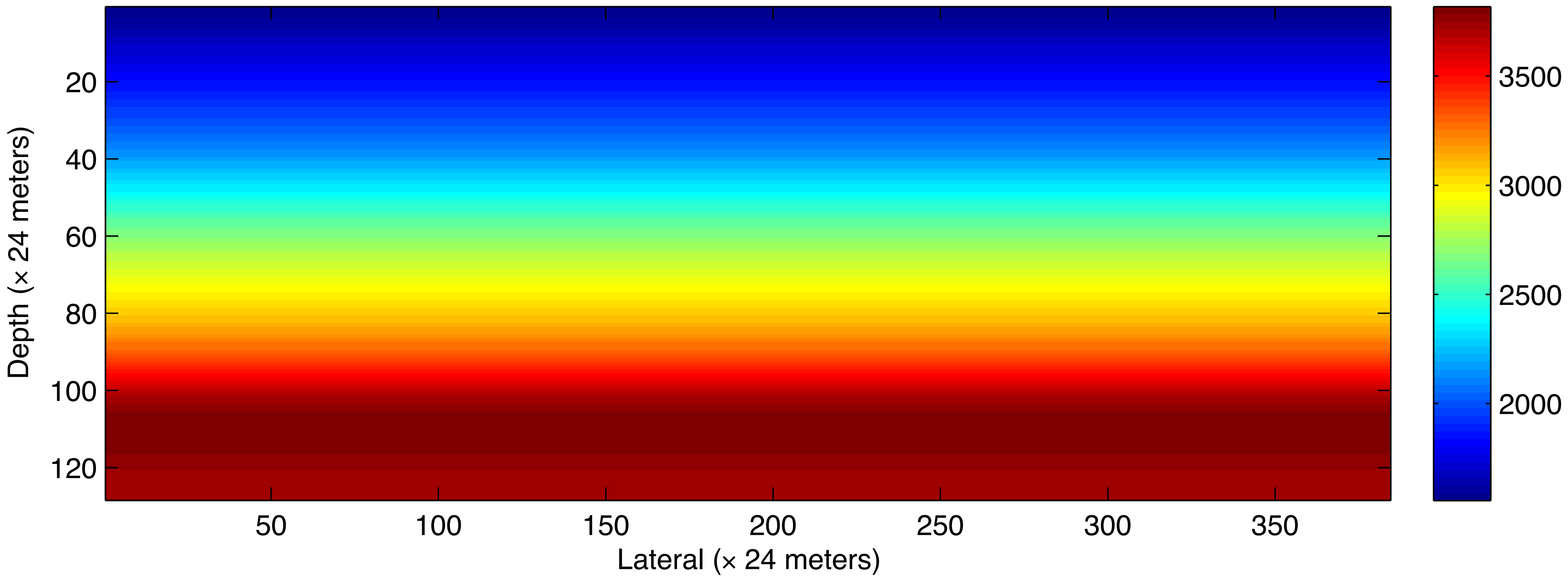
end

SLIM

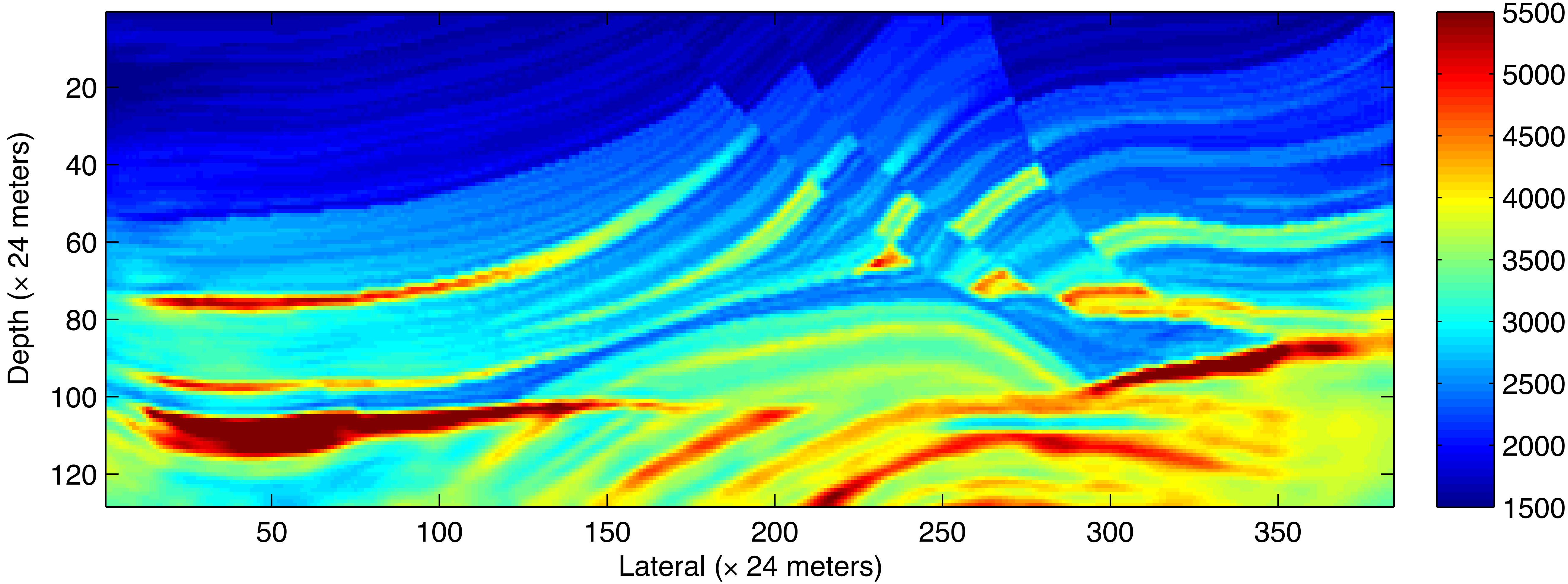
True model



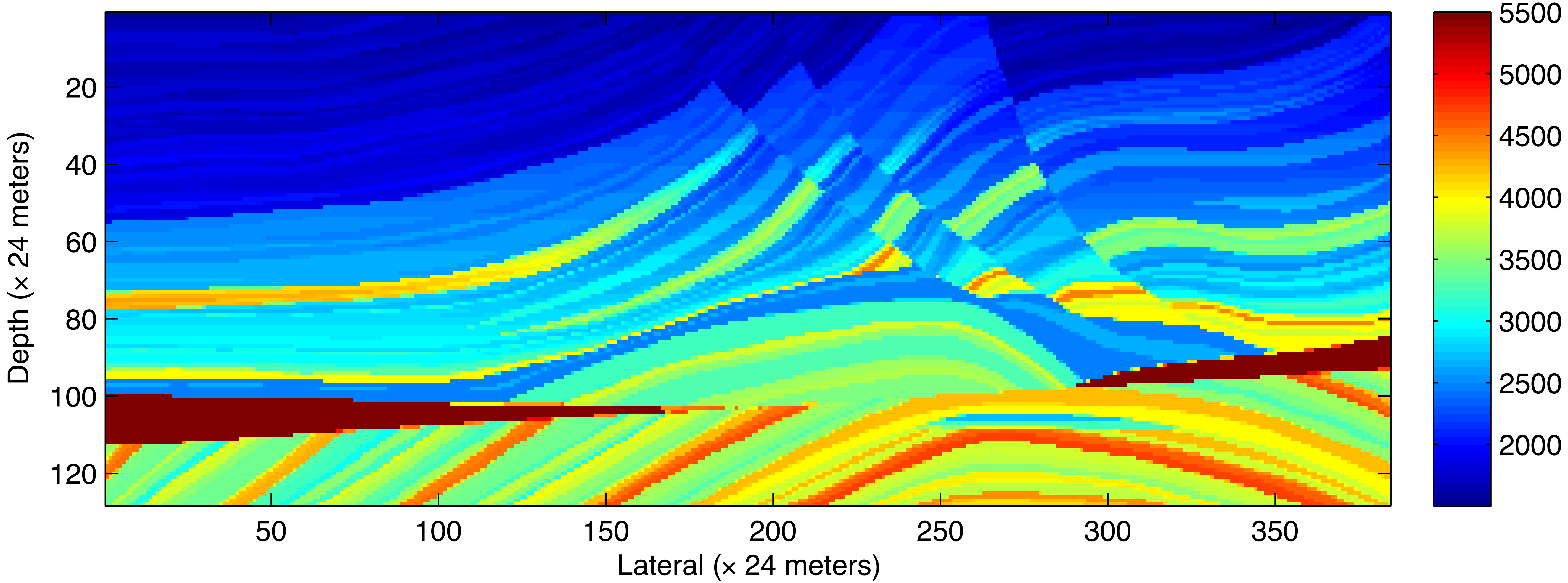
Initial model



Inverted model



True model


Depth (× 24 meters)
20
40
60
80
100
120
50
100
150
200
250
300
350
Lateral (× 24 meters)
5500
5000
4500
4000
3500
3000
2500
2000

Performance

Remember per *subproblem*

$$n_{PDE}^{\ell_1} \times K \ll n_{PDE}^{\ell_2} \times n_f \times n_s$$

$$\begin{aligned} n_{PDE}^{\ell_1} &\approx 200 \\ K &= 150 \end{aligned}$$

versus

$$\begin{aligned} n_{PDE}^{\ell_2} &\approx 10 \\ K &= 38400 \end{aligned}$$

SPEEDUP of 13 X

Conclusions

Leveraged

- curvelet-domain sparsity on the model
- invariance under solution operators <=> preservation of sparsity

Indications that *compressive sensing* supersedes the *stochastic approximation by sparse recovery of dimensionality reduced* subproblems

Extension to 3D (5D data) will lead to larger *improvements*...

Open problems
[some of them]

Preconditioner for *indirect* Helmholtz solvers in 3D

Extension to incomplete data, i,e, $\mathbf{P} \mapsto \mathbf{P}_i$ (Hadamard product)

Analysis of performance of the proposed algorithm

- extension to nonlinear problems
- behavior Pareto curves etc.

Non-convexity of FWI

- *'ad-hoc' multiscale* continuation methods

'Holy grail' [FWI with focusing]

Convexification by *extensions*

$$\tilde{\mathbf{X}} = \arg\min_{\mathbf{X}\in\boldsymbol{\mathcal{X}}} \|\mathbf{X}\|_{\boldsymbol{\mathcal{A}}} \quad \text{subject to} \quad \|\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{X};\mathbf{Q}]\|_{2,2} \le \sigma$$

$$\tilde{\mathbf{m}} = \mathrm{diag}\{\mathbf{S}^H\mathbf{X}\} \quad \text{with} \quad \mathbf{X} \quad \text{the extension}$$

$\boldsymbol{\mathcal{F}}[\mathbf{X};\mathbf{Q}] := \mathbf{P}\bar{\mathbf{H}}^{-1}[\mathbf{S}^H\mathbf{X}]\mathbf{Q}$, $\mathbf{S}^H\mathbf{X}$ *positive-definite matrix*

1. $\|\mathbf{X}\|_{\boldsymbol{\mathcal{A}}} = \|\overbrace{A_h}^{\text{annihilator}}\mathbf{X}\|_{1,2}$ [Symes, '09]

2. $\|\mathbf{X}\|_{\boldsymbol{\mathcal{A}}} = \|\mathbf{X}\|_{*}$ **?**

Acknowledgments

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08).

We also would like to thank the authors of CurveLab.

This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BP, Chevron, ConocoPhillips, Petrobras, Total SA, and WesternGeco.

NSERC
CRSNG

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.

SLIM

Thank you

slim.eos.ubc.ca