Efficient least-squares migration with *sparsity* promotion

Felix J. Herrmann & Xiang Li

SLIM

Seismic Laboratory for Imaging and Modeling
the University of British Columbia

SLIM

Drivers

Recent push for *high-fi* images calls for

- high-quality *broad-band* data *volumes* (100k channels)
- *larger* offsets & *full* azimuth

Exposes vulnerabilities in our *ability* to control

- *acquisition* costs / time
- *processing* costs / time

Drivers

Complexity of *inversion* algorithms also *exposes* the "*curse of dimensionality*" in

- **sampling**: *exponential* growth of # *samples* for *high* dimensions
- **optimization**: *exponential* growth of # parameter *combinations* that need to be evaluated to minimize our *objective* functions

Wish list

Imaging costs determined by *transform-domain sparsity* of the *subsurface*

- *computational costs* that are *dictated by structure* 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

Prior art

Simultaneous sources & phase encoding [Beasley, '98, Berkhout, '08] [Morton, '98, Romero, '00]

- supershots [Krebs et.al., '09, Operto et. al., '09, Herrmann et.al., '08-10']

Stochastic optimization & machine learning [Bertsekas, '96, Nemirovsky, '08]

- stochastic gradient descent

Compressive sensing [Candès et.al, Donoho, '06]

- *sparse* recovery & *randomized* subsampling

In today's talk

Randomized-dimensionality reduction techniques aim to deal with these *impediments* by

- **removing** *linear* dependence on # sources
- **turning** NP-hard algorithms into *convex* computationally *tractable* optimization problems
- **avoiding** iterative algorithms that require multiple *passes* through *all* data

SLIM

[Li & FJH et. al. ’10-]

Compressive imaging

Challenge:

Least-squares migration requires *multiple passes & PDE solves*

Key idea:

- combine *Compressive Sensing* & ‘*Phase encoding*’
- turn “*overdetermined*” imaging problem into *underdetermined* problem with *randomized* supershots
- use *curvelet*-based *sparse* recovery to remove *crosstalk*

Computational *gain* when costs *least-squares* with *all data* >> costs of *sparse recovery* of the *reduced* problem

Wave-equation migration

Solution of a large *'overdetermined'* system

$$\min_{\mathbf{x}} \frac{1}{2K} \sum_{i=1}^{K} \|\mathbf{b}_i - \mathbf{A}_i \mathbf{x}\|_2^2,$$

$$\begin{aligned} \mathbf{b}_i &= \text{monochromatic shot record} \\ \mathbf{A}_i &= \text{demigration operator} \\ K &= n_f \times n_s (\text{batch size}) \end{aligned}$$

Wave-equation migration

Solution of a large *'overdetermined'* system

$$\min_{\mathbf{x}} \frac{1}{2K} \sum_{i=1}^{K} \|\mathbf{b}_i - \mathbf{A}_i\mathbf{x}\|_2^2,$$

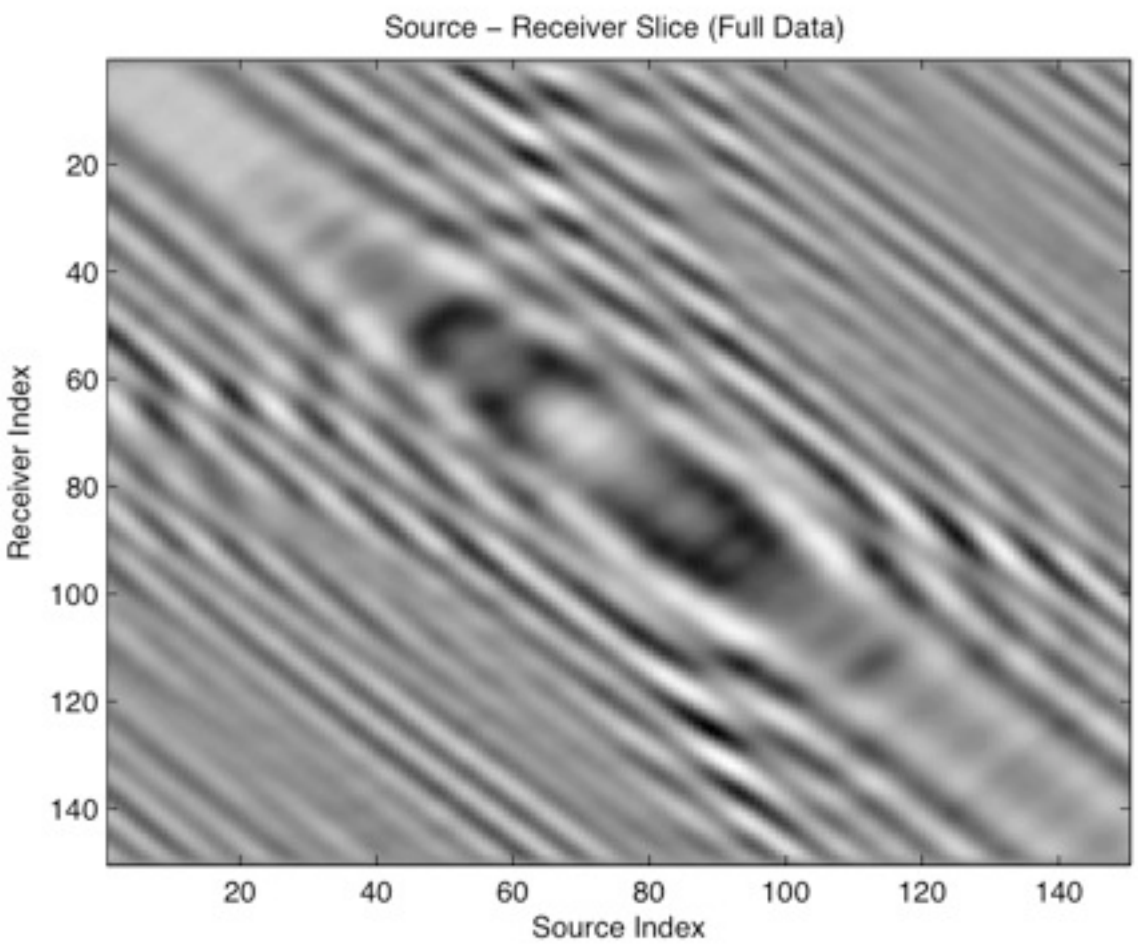
Iterations of the solver requires 4 PDE solves for each *source*

- use *linearity* of the *source* to turn *sequential* sources into *simultaneous* sources
- use *fewer* simultaneous sources

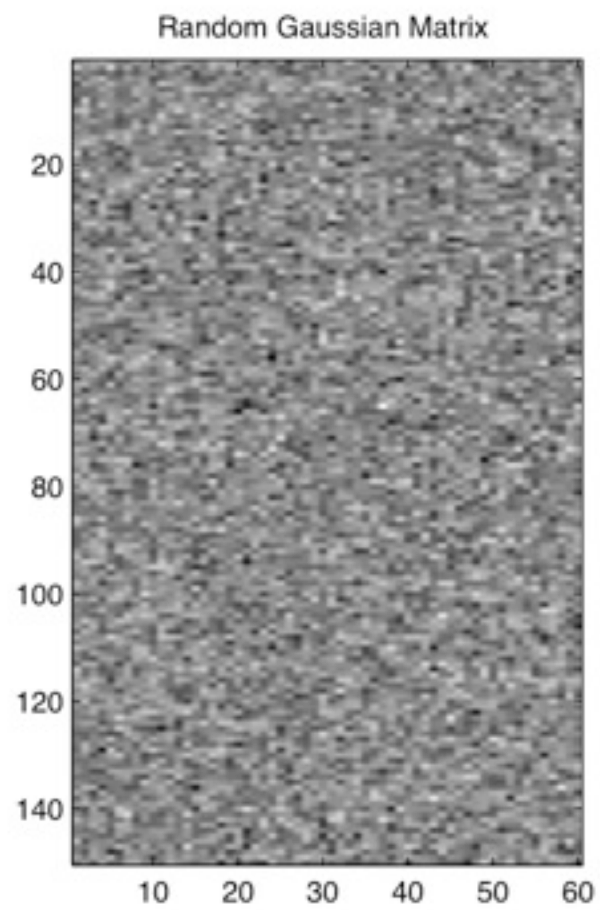
Study behavior as # of sources *increases*

Randomized source superposition

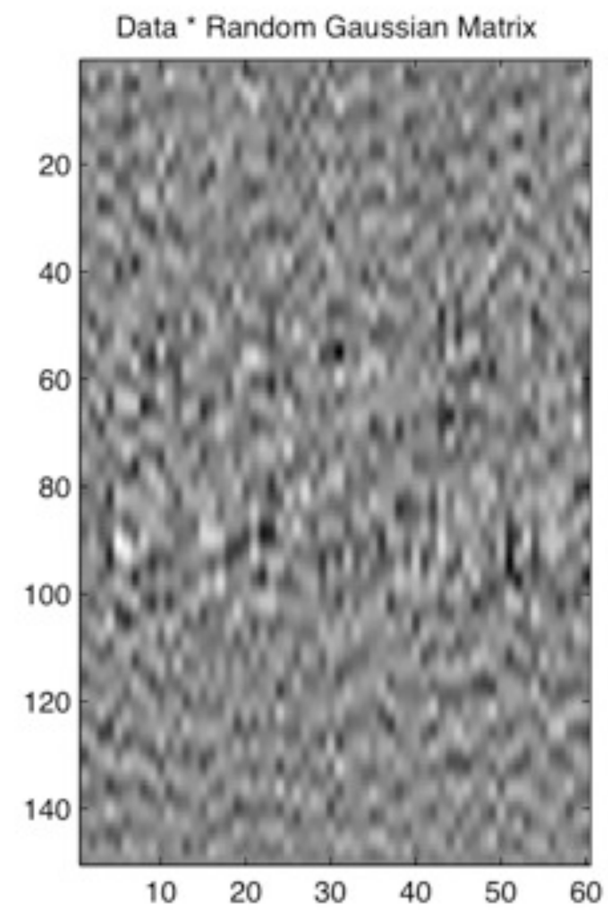
$$[\mathbf{b}_1, \cdots, \mathbf{b}_{n_s}]$$



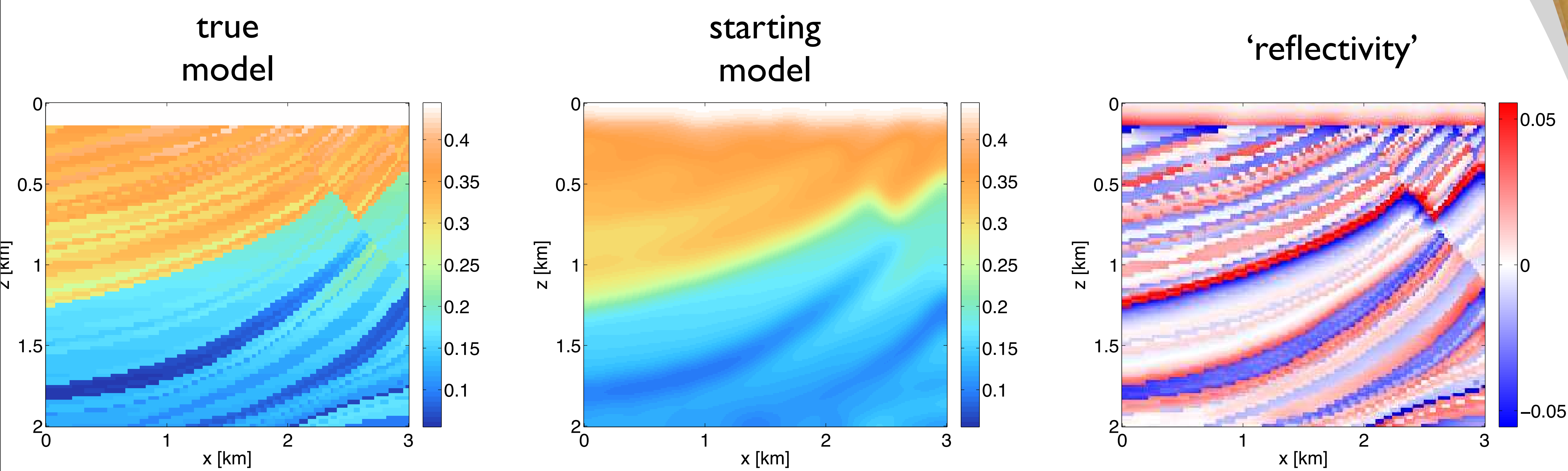
$$\mathbf{W}$$



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



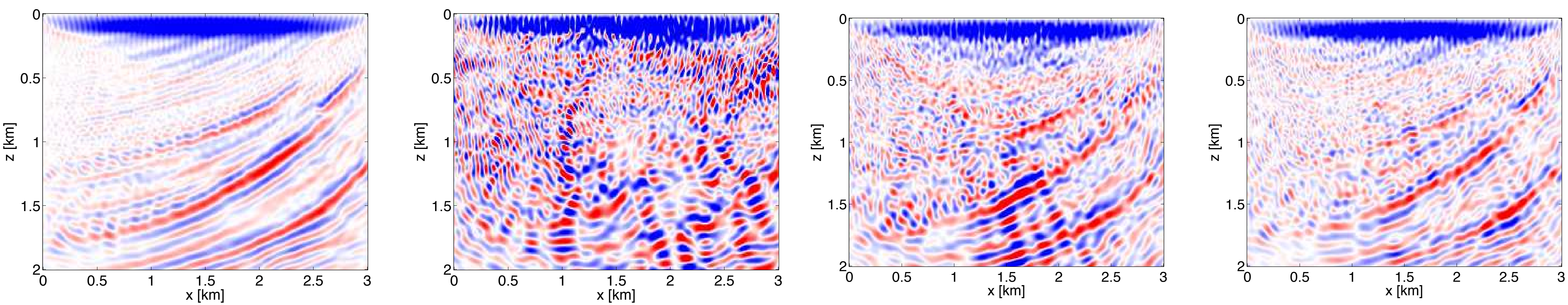
Stylized example



Migration

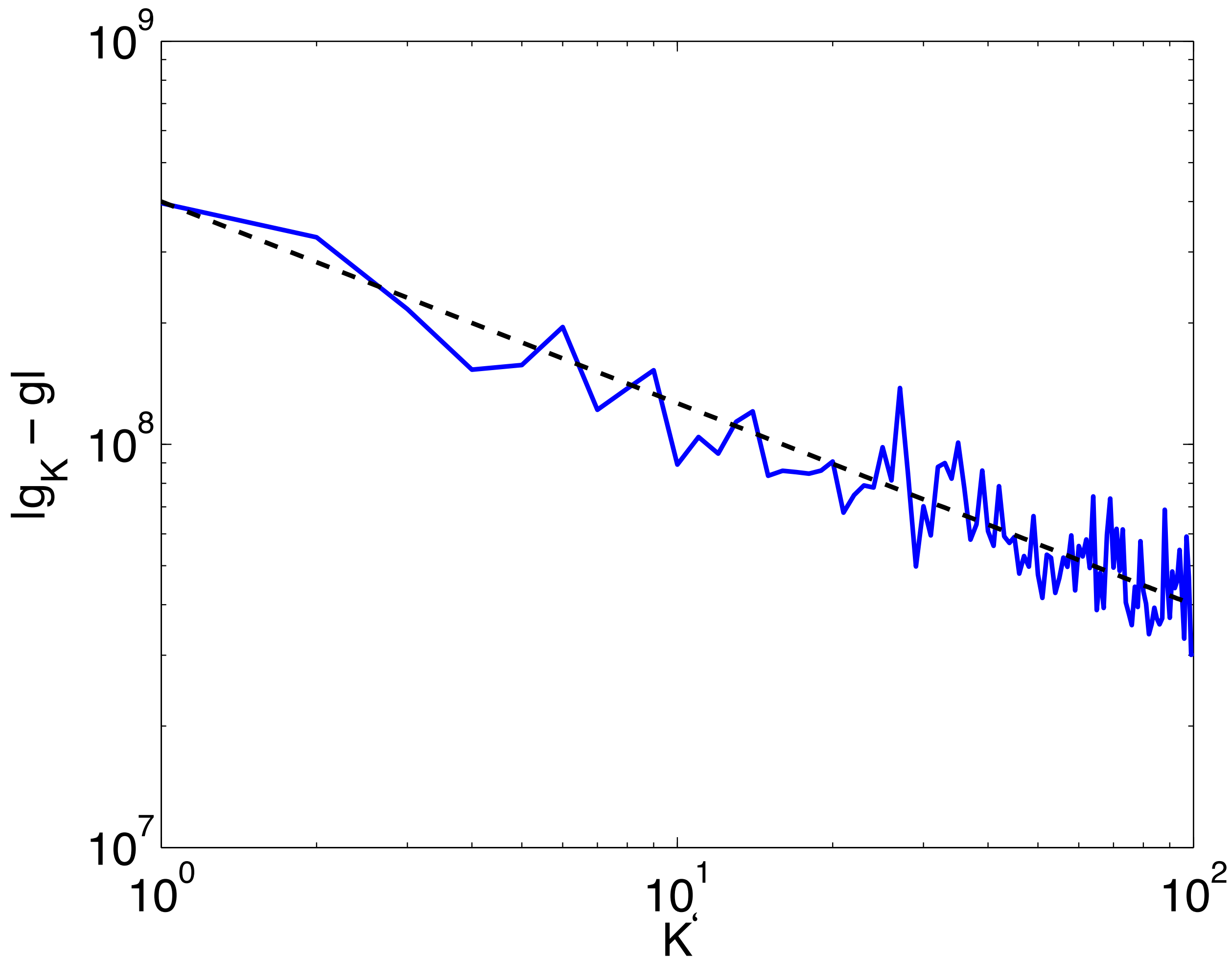
Search direction for *increasing* batch size K:

$$\mathbf{g}_{K'} \approx \frac{1}{K'} \sum_{j=1}^{K'} \mathbf{A}_j^* \underline{\mathbf{b}}_j \qquad K' = n'_f \times n'_s$$



full | K'=1 | K'=5 | K'=10

Decay



error between full and sampled migration

[Nemirovsky, '08]

SLIM

Stochastic *average* approximation (SAA)

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

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

Can be used with *arbitrary* optimization methods

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

Experience shows *benefits of* '*redraws* & *warm* restarts'

Algorithm

Algorithm 1: Stochastic-average approximation with warm restarts

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

SLIM

Subproblems

least-squares migration

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

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

Subproblems

sparsity-promoting migration

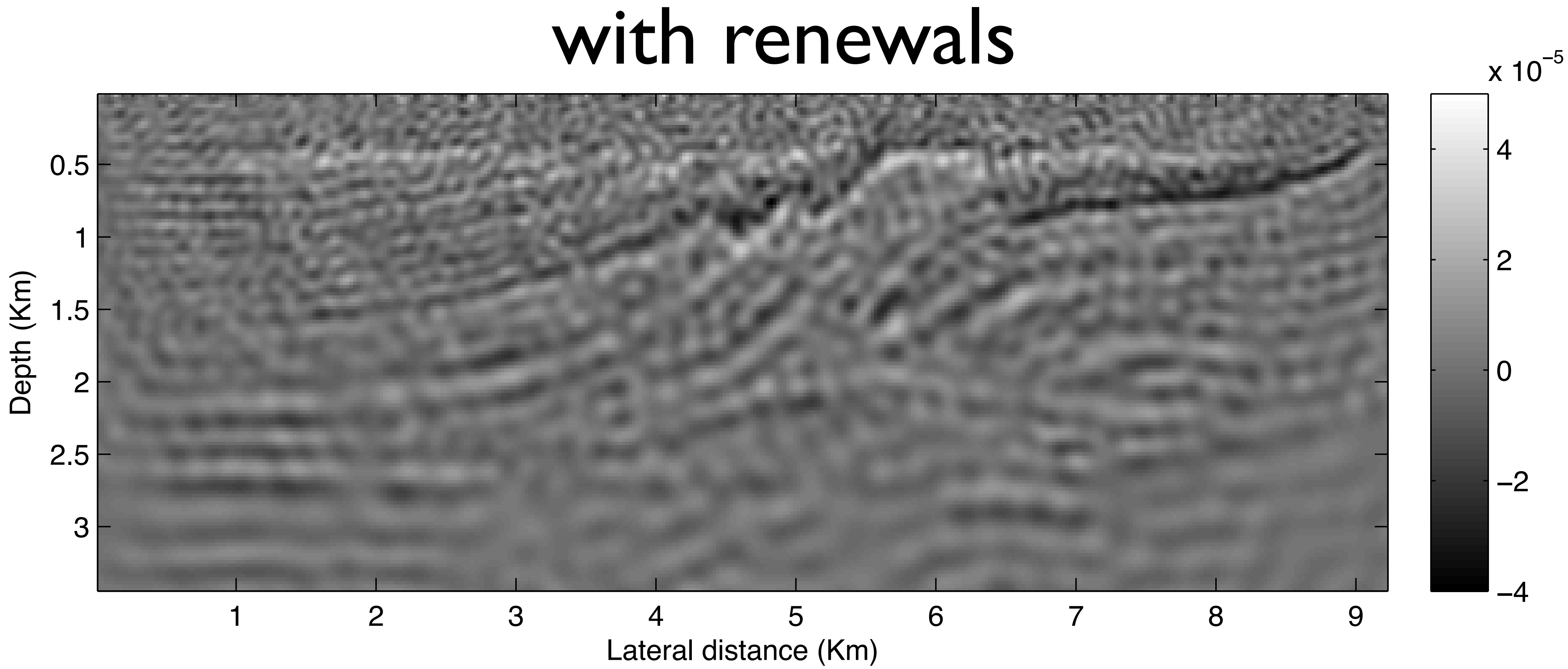
$$\mathbb{P}_{\ell_1}(\mathbf{W}^k;\mathbf{x}_0) \quad \min_{\mathbf{x}} \frac{1}{2K'}\sum_{j=1}^{K'} \|\underline{\mathbf{b}}_j^k - \mathbf{A}_j^k\mathbf{x}\|_2 \quad \text{subject to} \quad \|\mathbf{x}\|_{\ell_1} \leq \tau^k$$

- 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

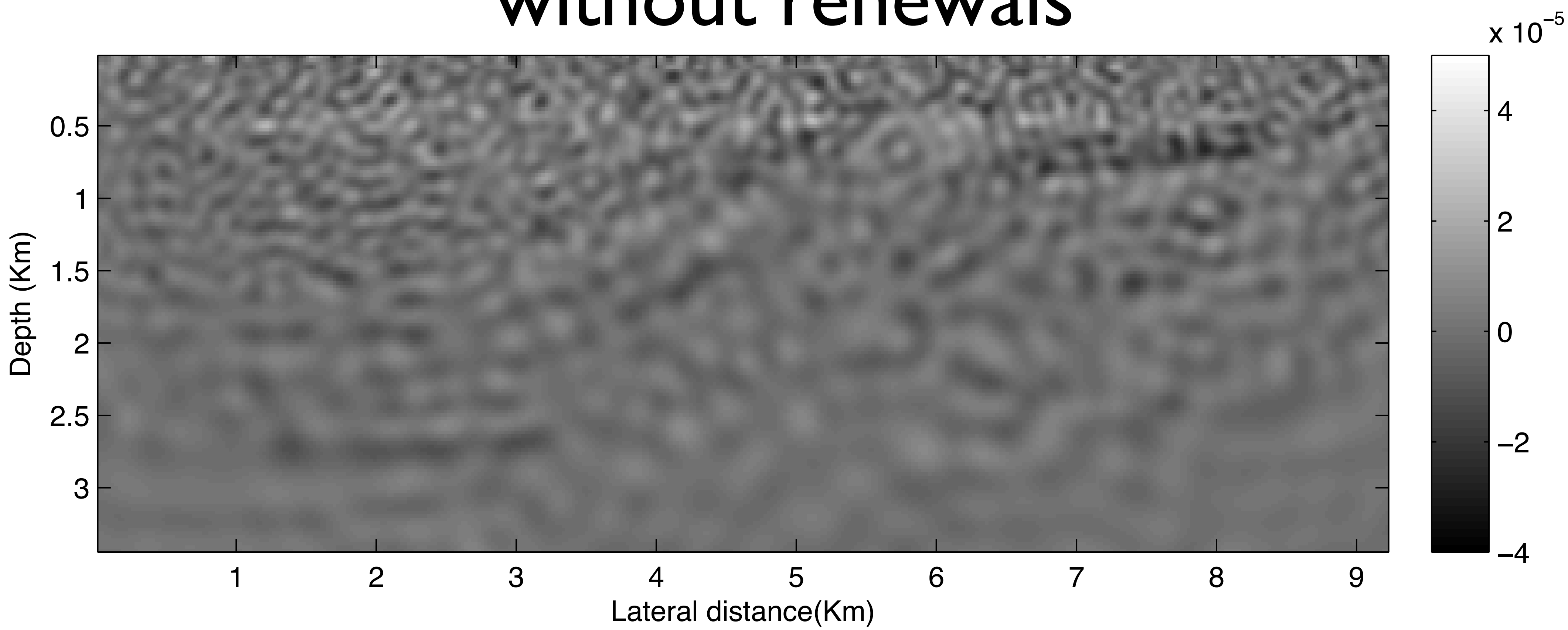
8 supershots w 3 frequencies



Least-squares migration

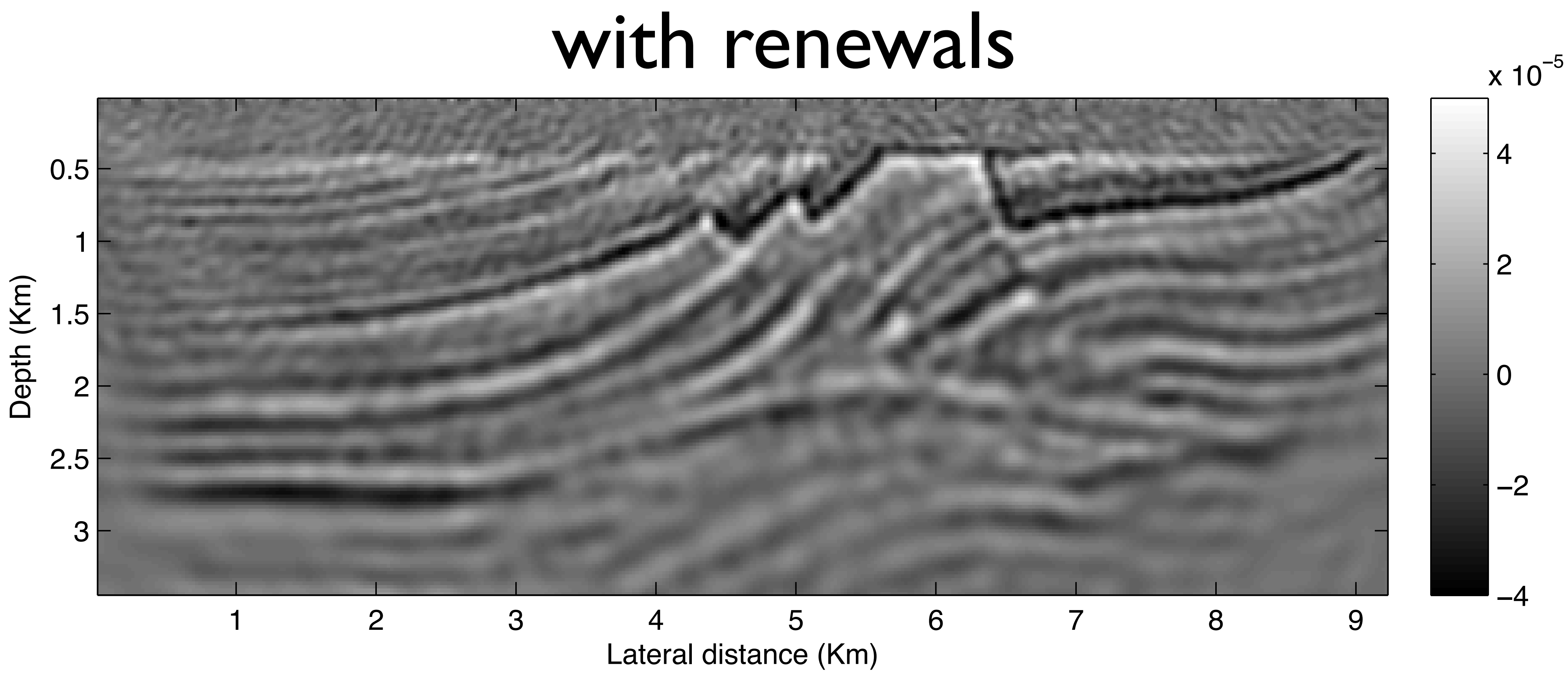
8 supershots w 3 frequencies

without renewals



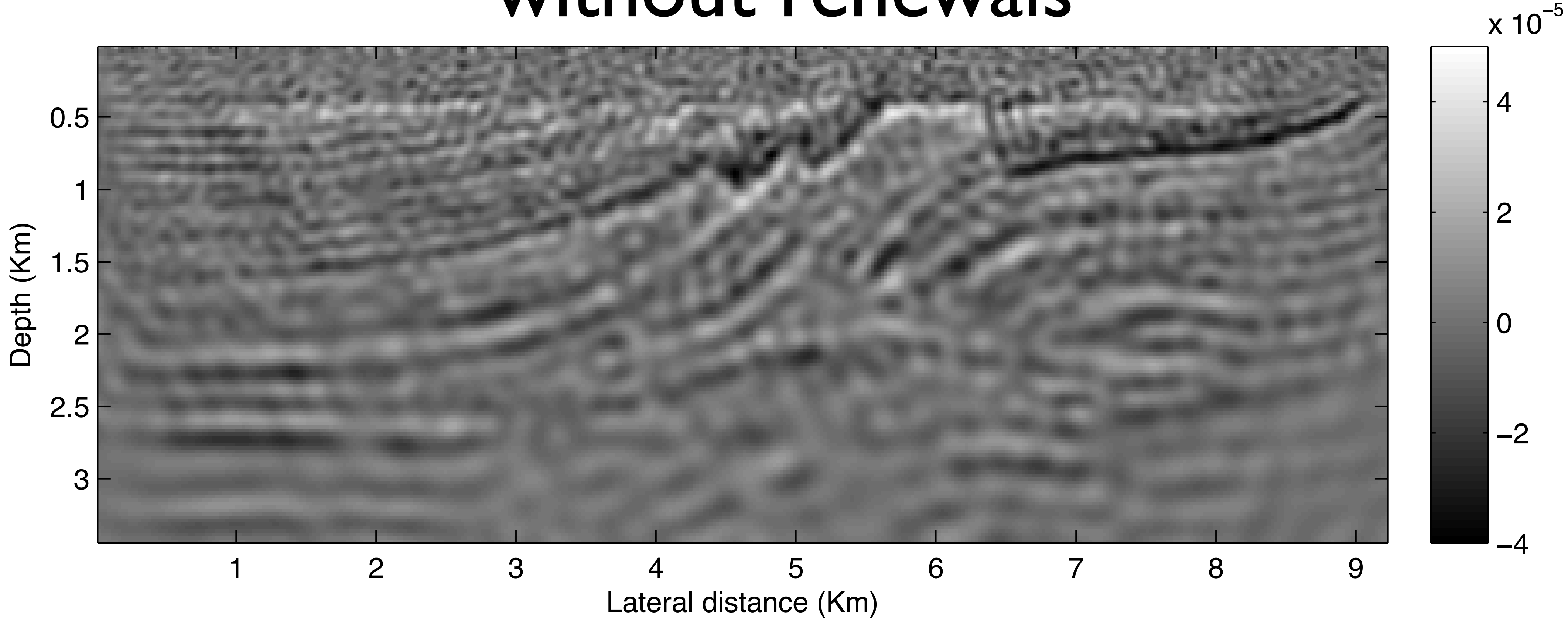
Sparse migration

8 supershots w 3 frequencies



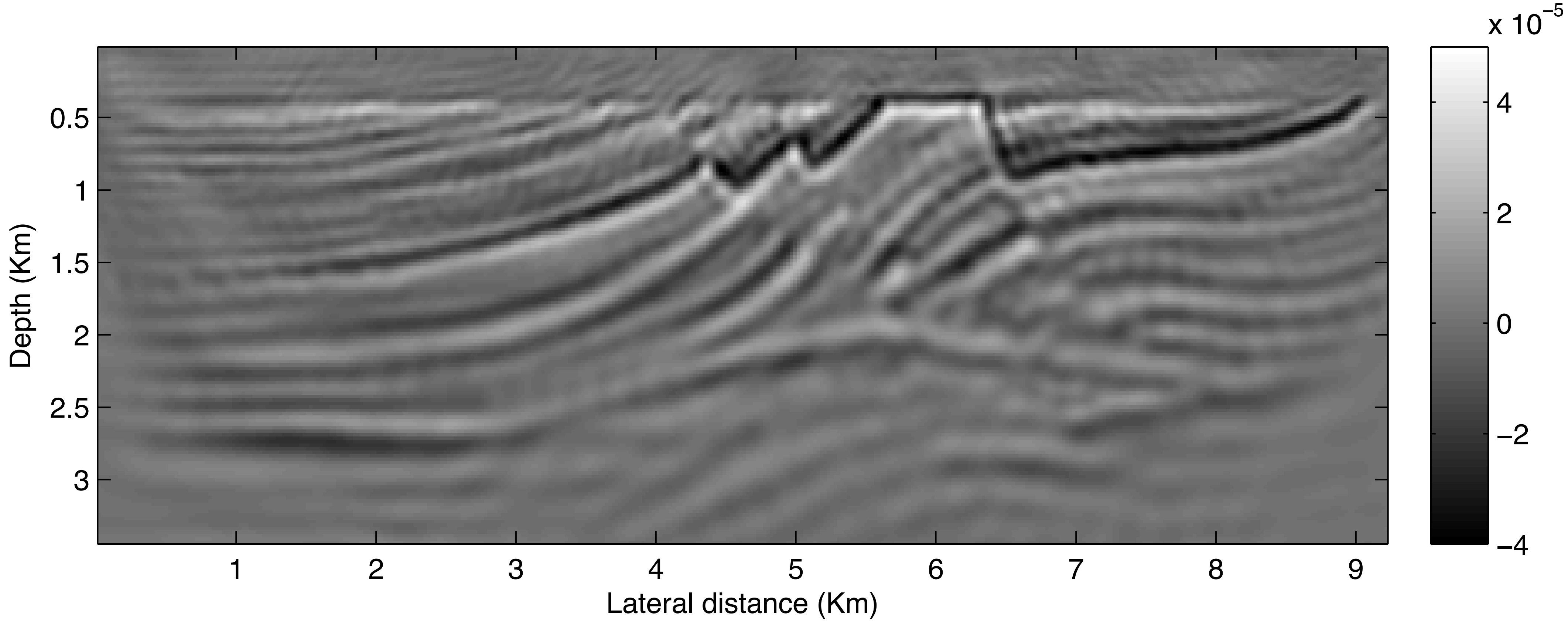
Sparse migration
8 supershots w 3 frequencies

without renewals



Least-squares migration

all 192 shots w 10 frequencies



SLIM

Why does this work?

Geophysics perspective:

Richer wavenumber *content* of the *randomized* simultaneous sources

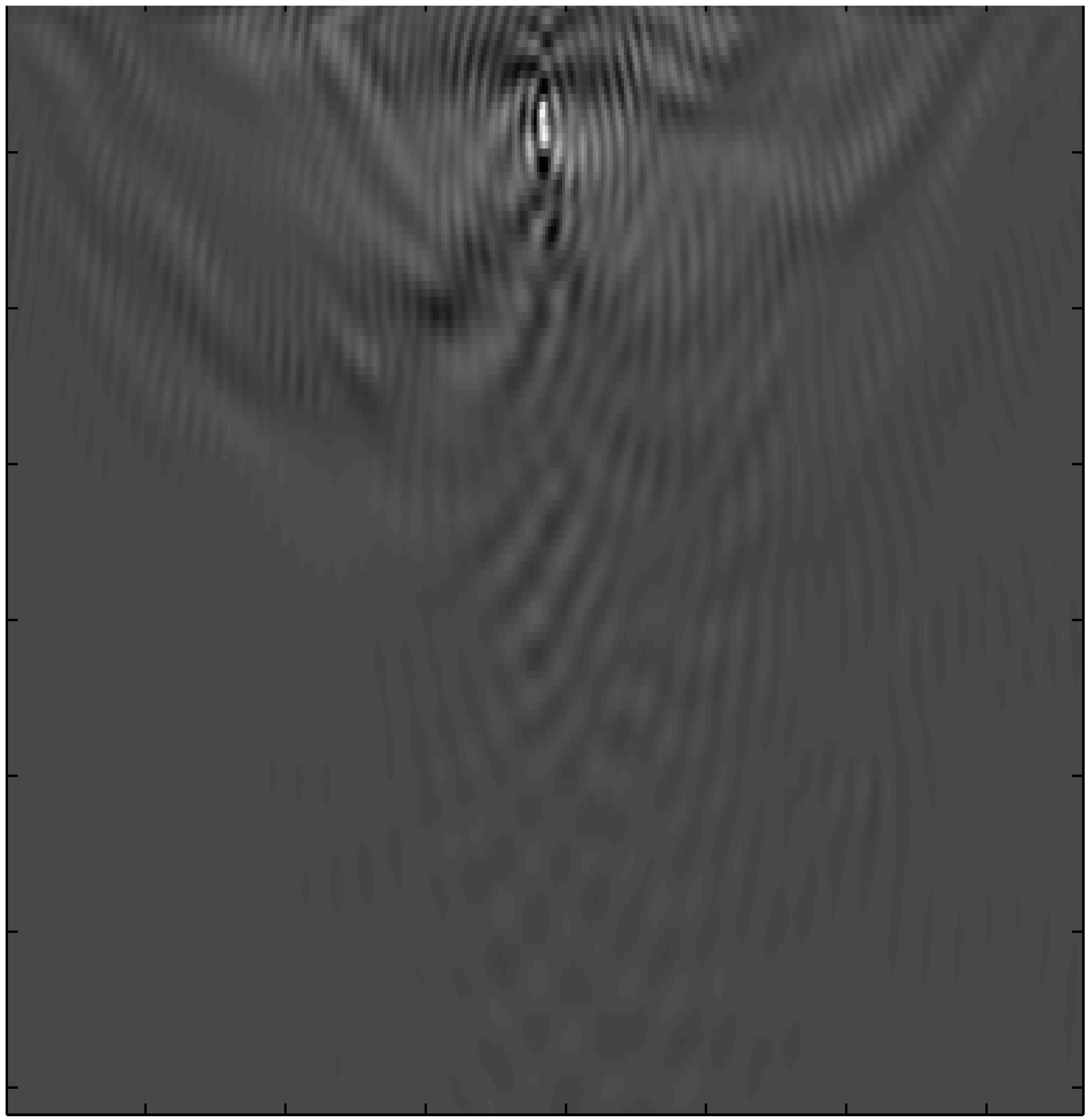
This is the premise of '*phase encoding*'

[Morton, '98, Romero, '00]

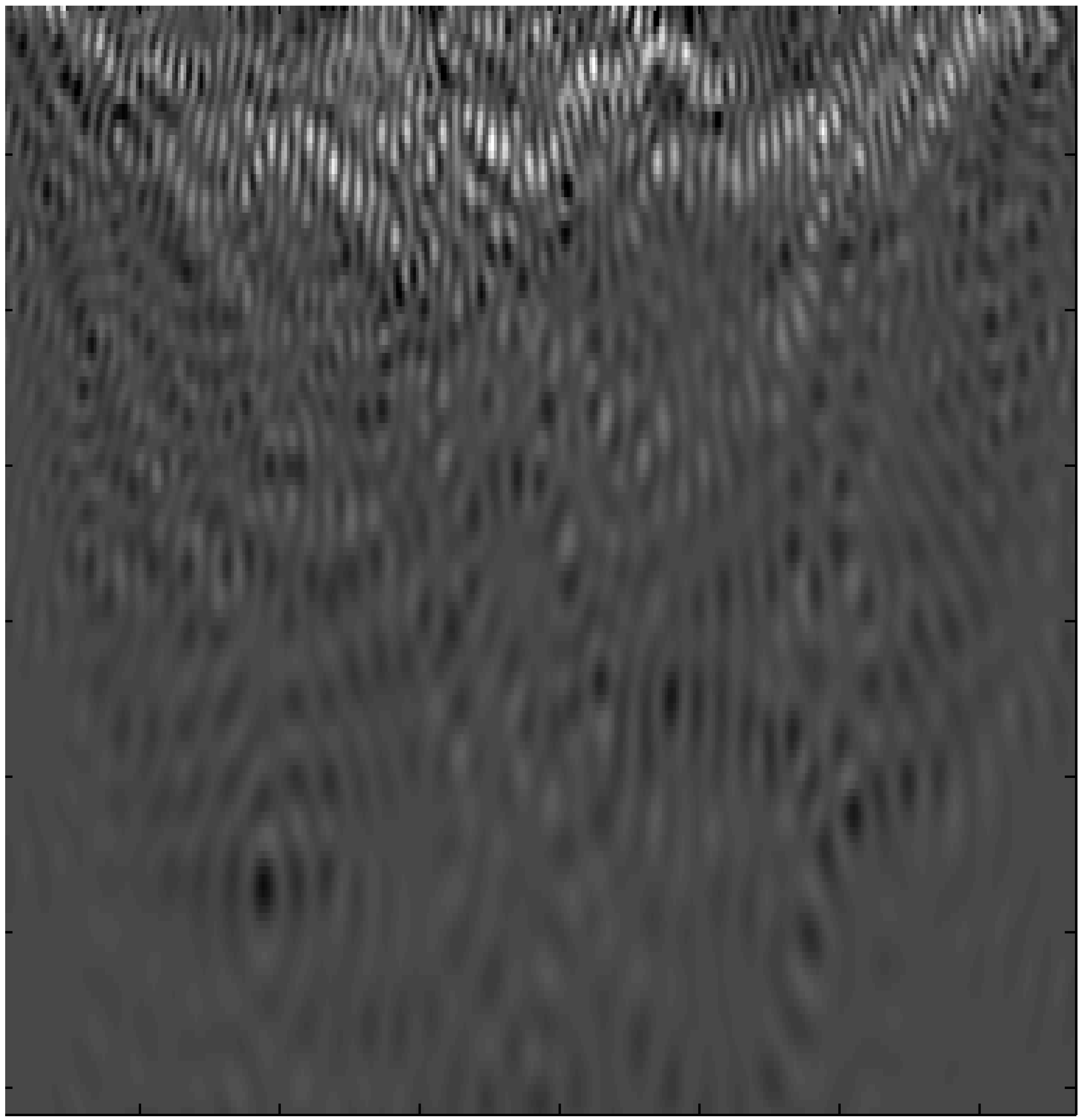
SLIM

Image from one shot

Sequential-source image



Simultaneous-source image



Why does this work?

Mathematician perspective:

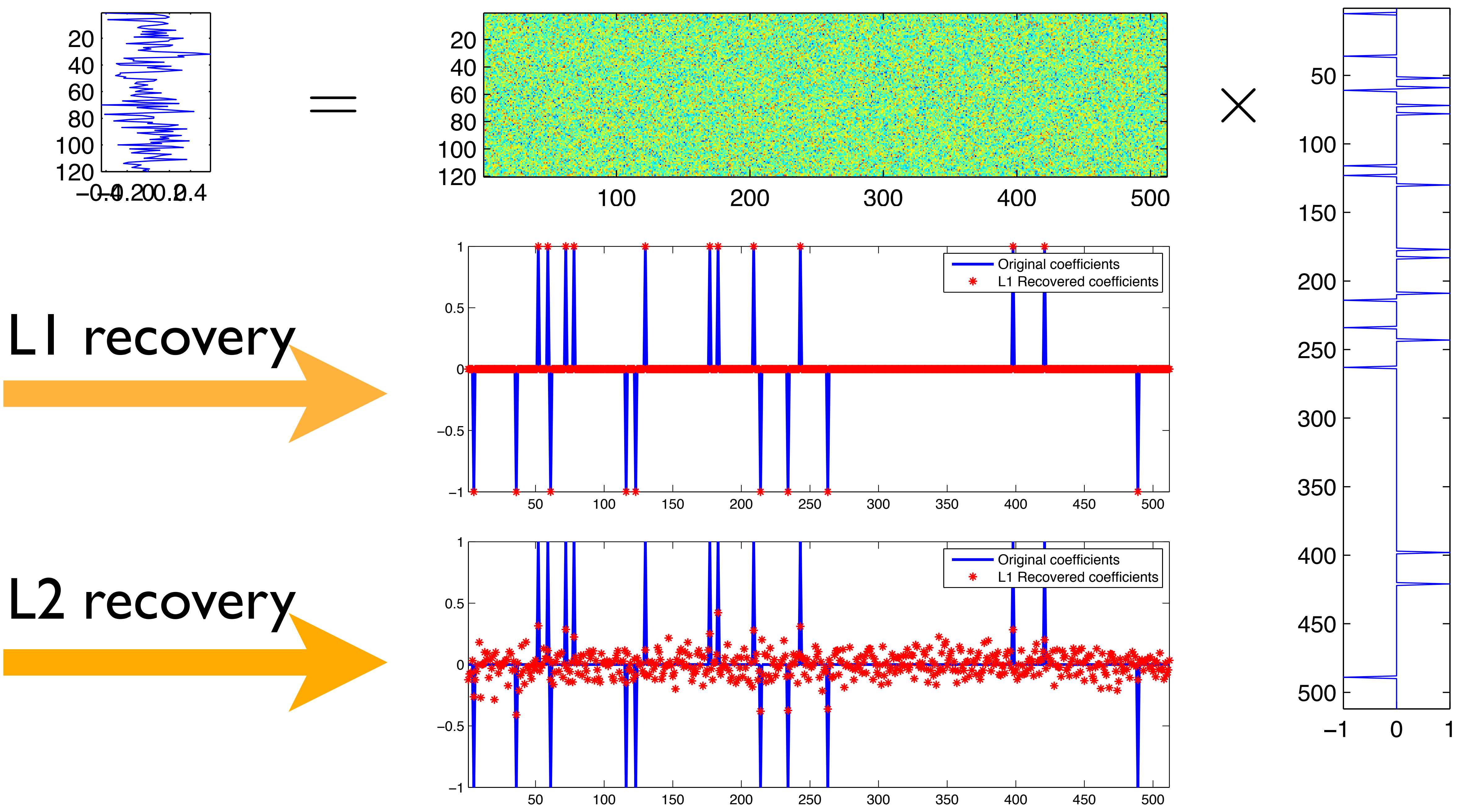
Random superposition is *incoherent* and gives rise to *Gaussian crosstalk*

This is the premise of Compressive Sensing & Stochastic Optimization

[Candès et.al, Donoho, ’06]

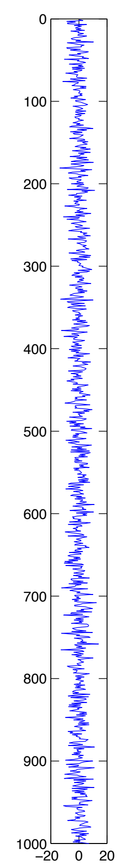
[Bertsekas, ’96, Nemirovsky, ’08]

Underdetermined recovery $K' \ll K$

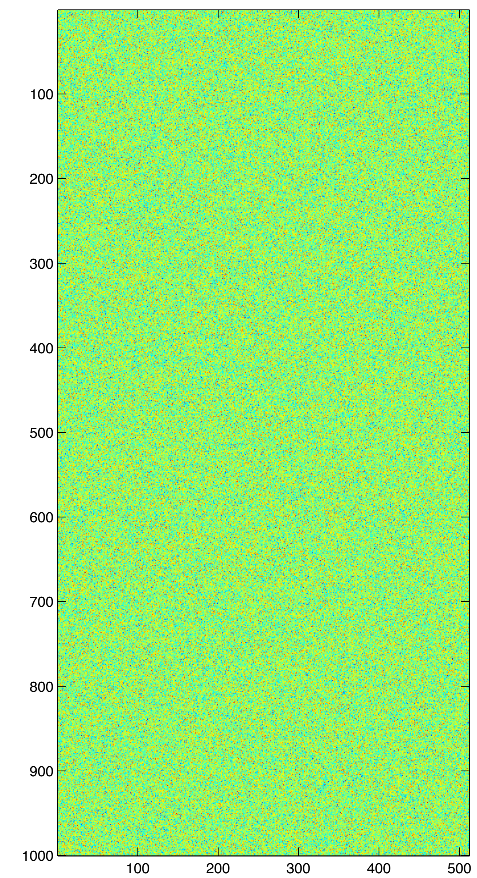


Overdetermined recovery

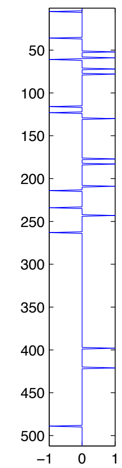
$$K' \approx K$$



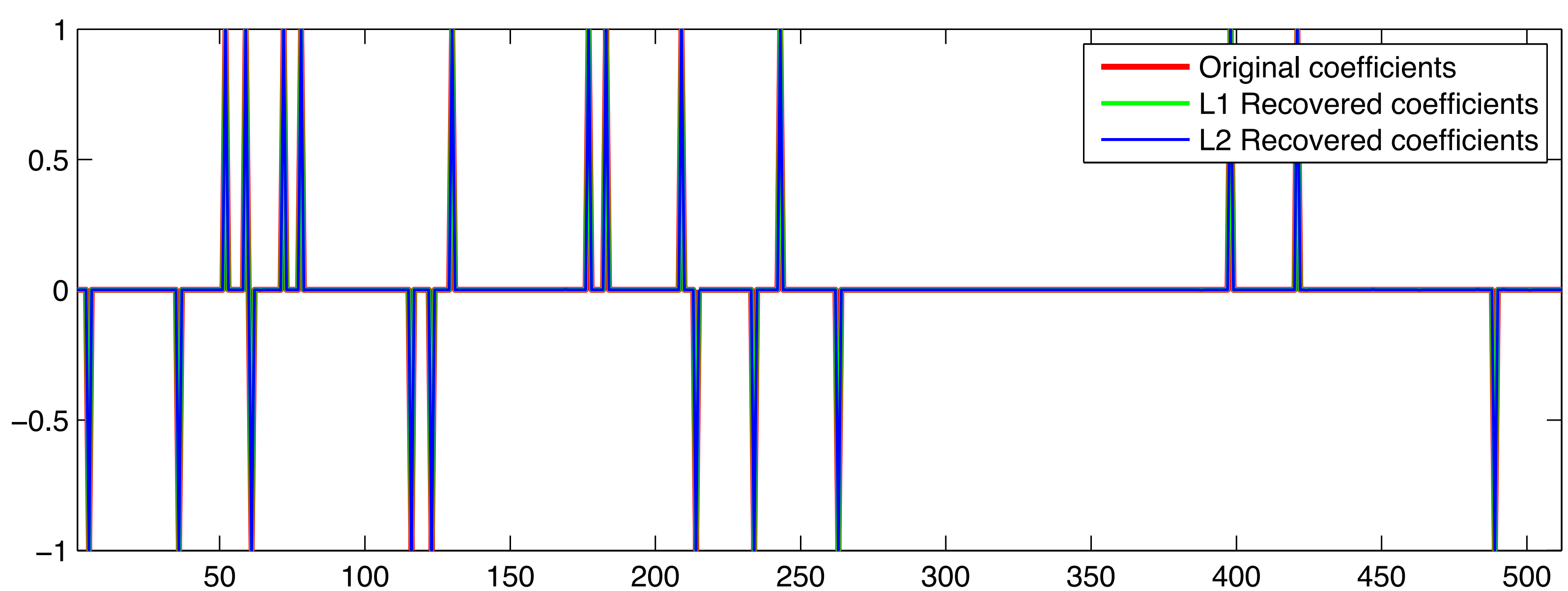
=



×



L1/L2
recovery



SLIM

Why does this work?

Inversion perspective:

Sparsity promotion acts as a *regularization*

This is the premise of Tikhonov regularization

Why does this work?

Computational perspective:

Random superposition makes the sparsity-promoting program computationally *tractable*

This is the premise of *randomized* dimensionality reduction

SLIM

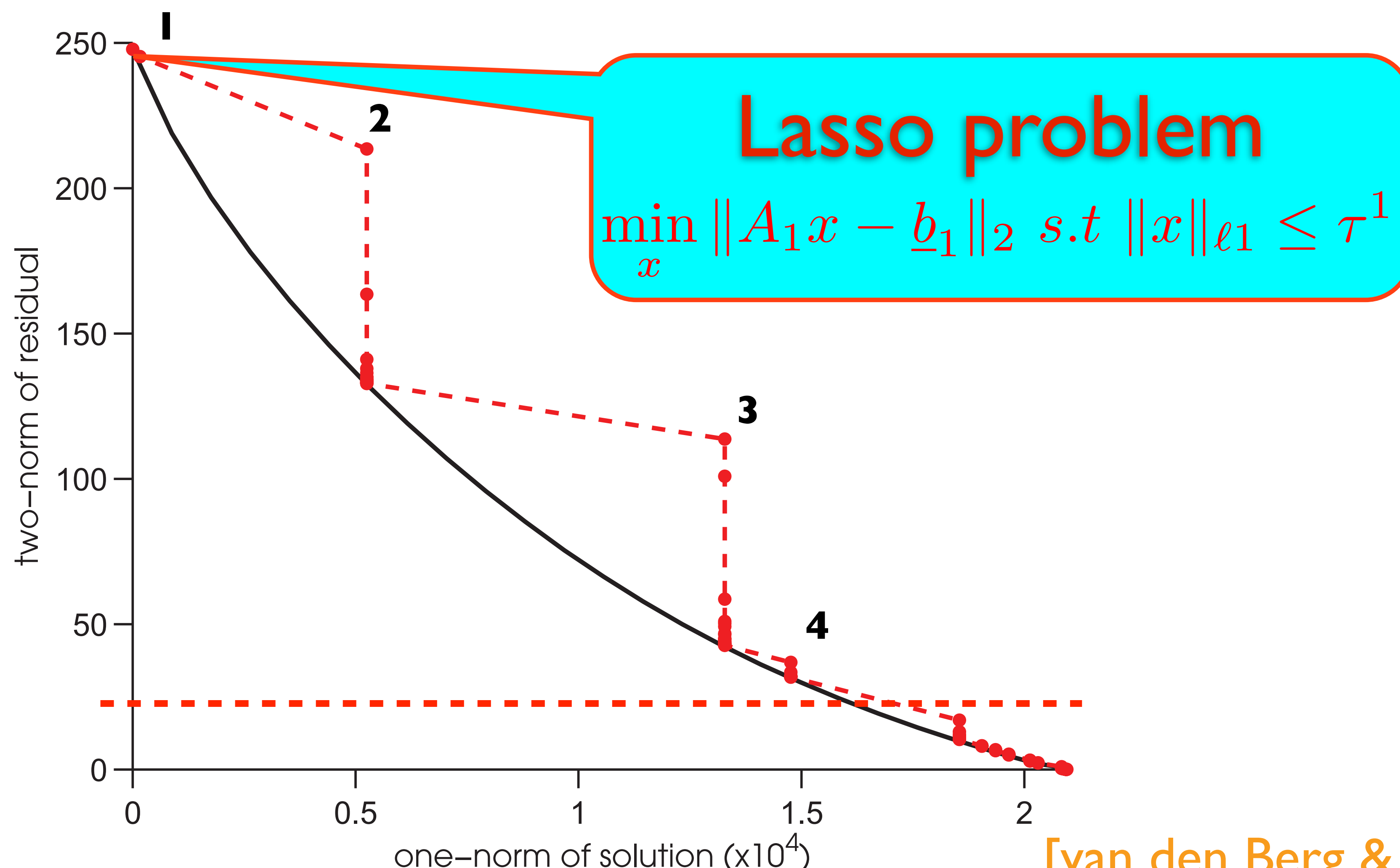
Continuation methods

Large-scale *sparsity*-promoting solvers *limit* the number of *matrix-vector* multiplies by

- slowly allowing *components* to *enter* into the *solution*
- solving an *intelligent* series of LASSO *subproblems* for *decreasing* sparsity levels
- exploring properties of the Pareto trade-off curve

Pareto curve

subproblems



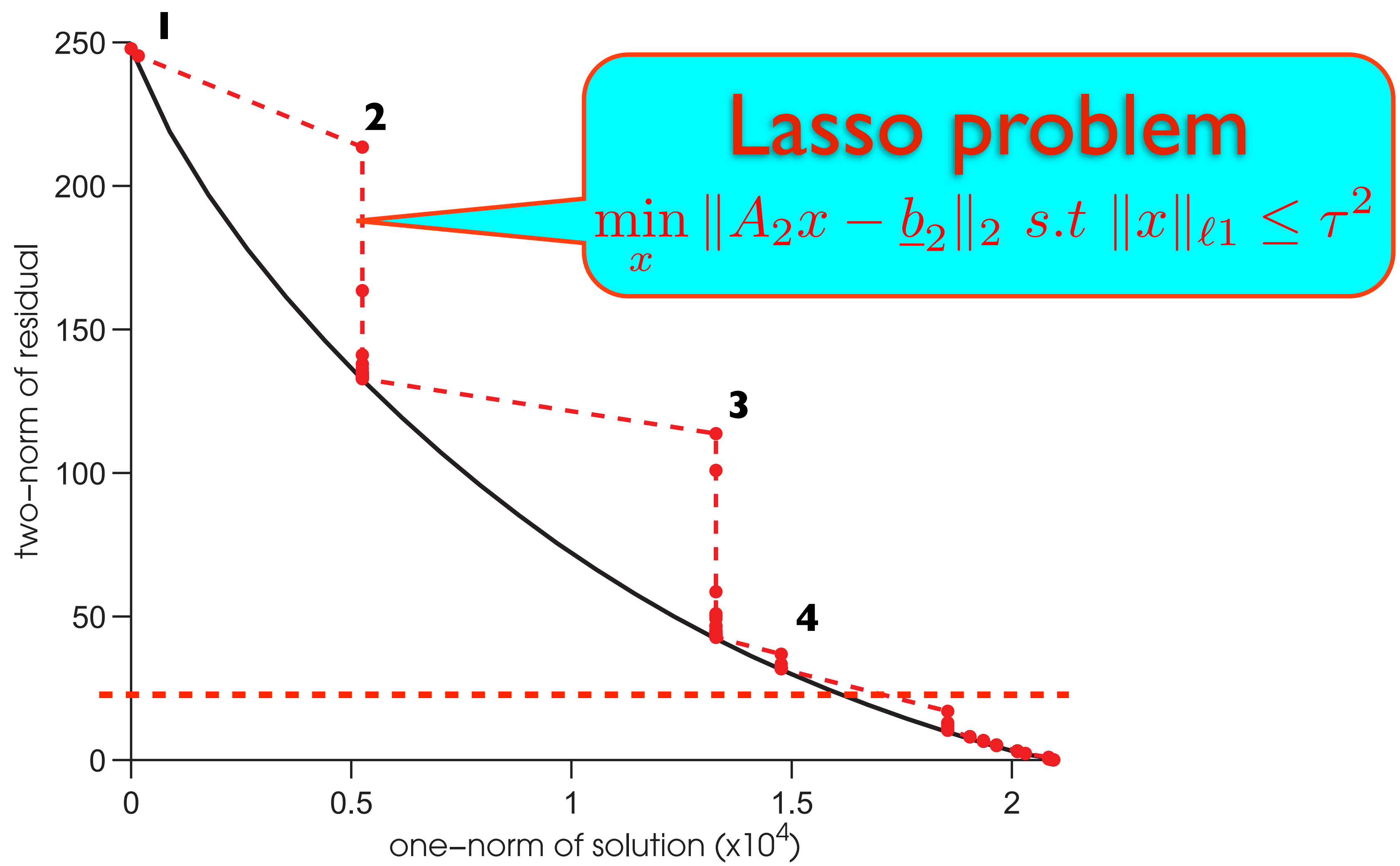
[van den Berg & Friedlander, '08]

[Hennefent et. al., '08]

[Lin & FJH, '09-]

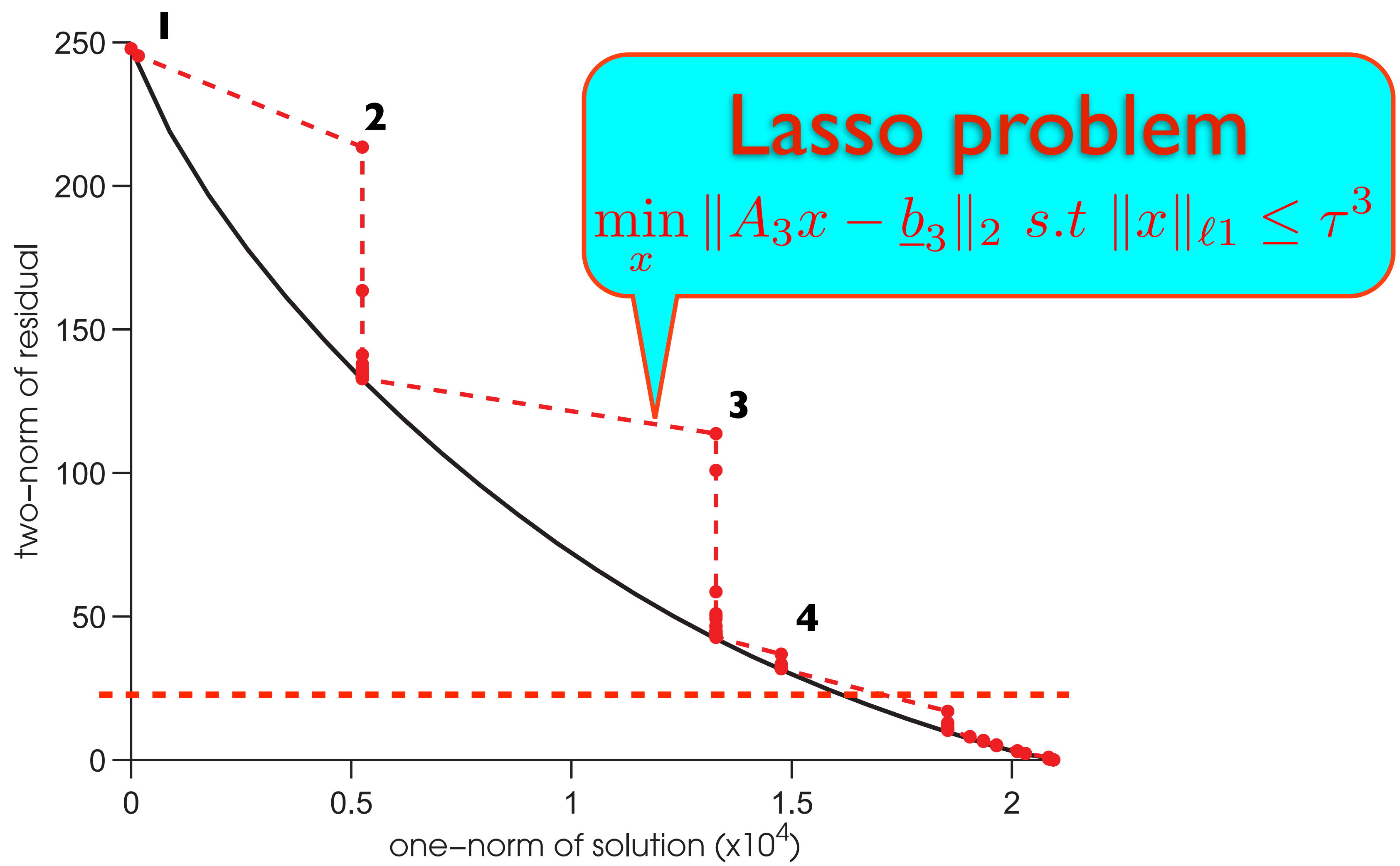
Pareto curve

subproblems



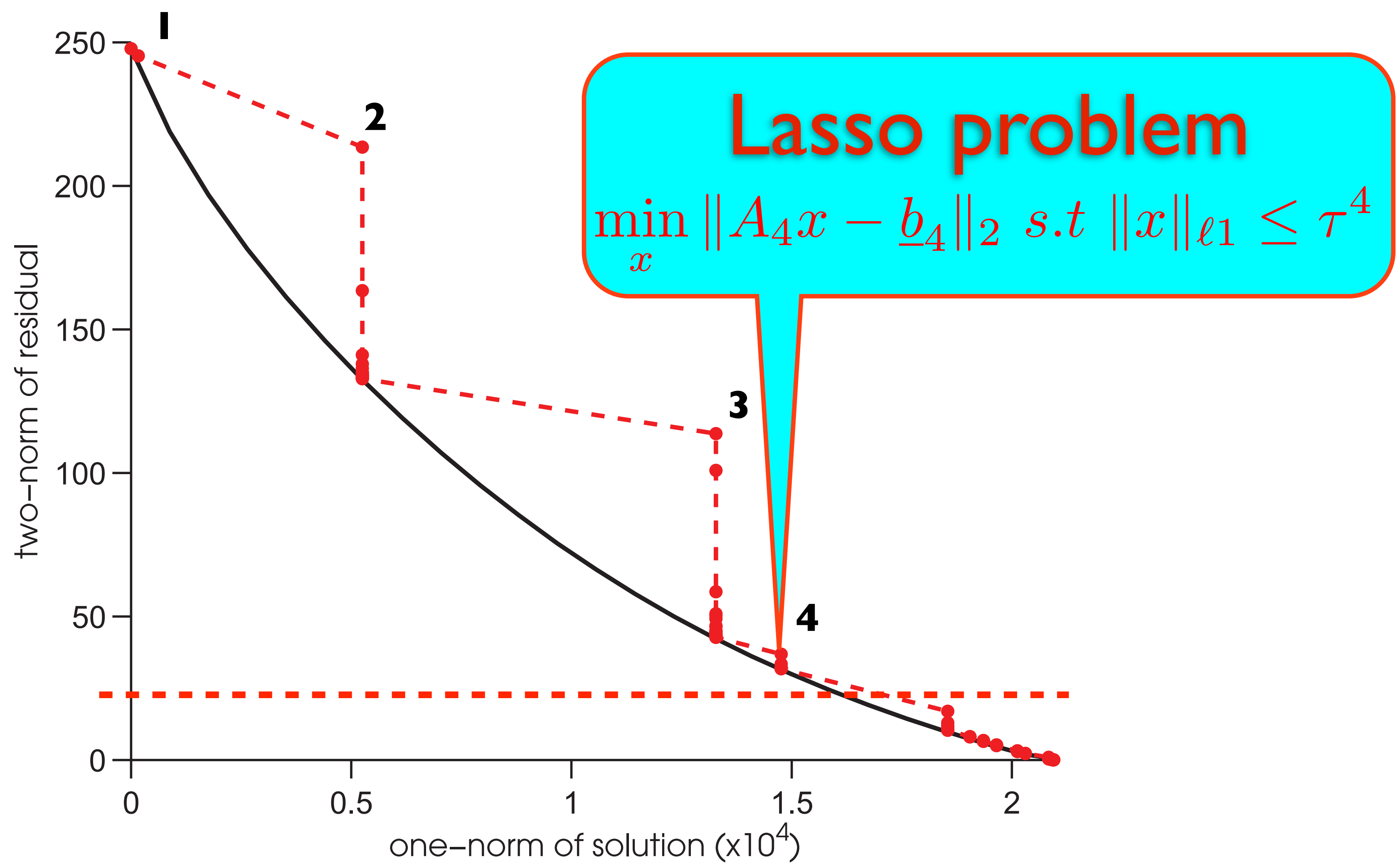
Pareto curve

subproblems



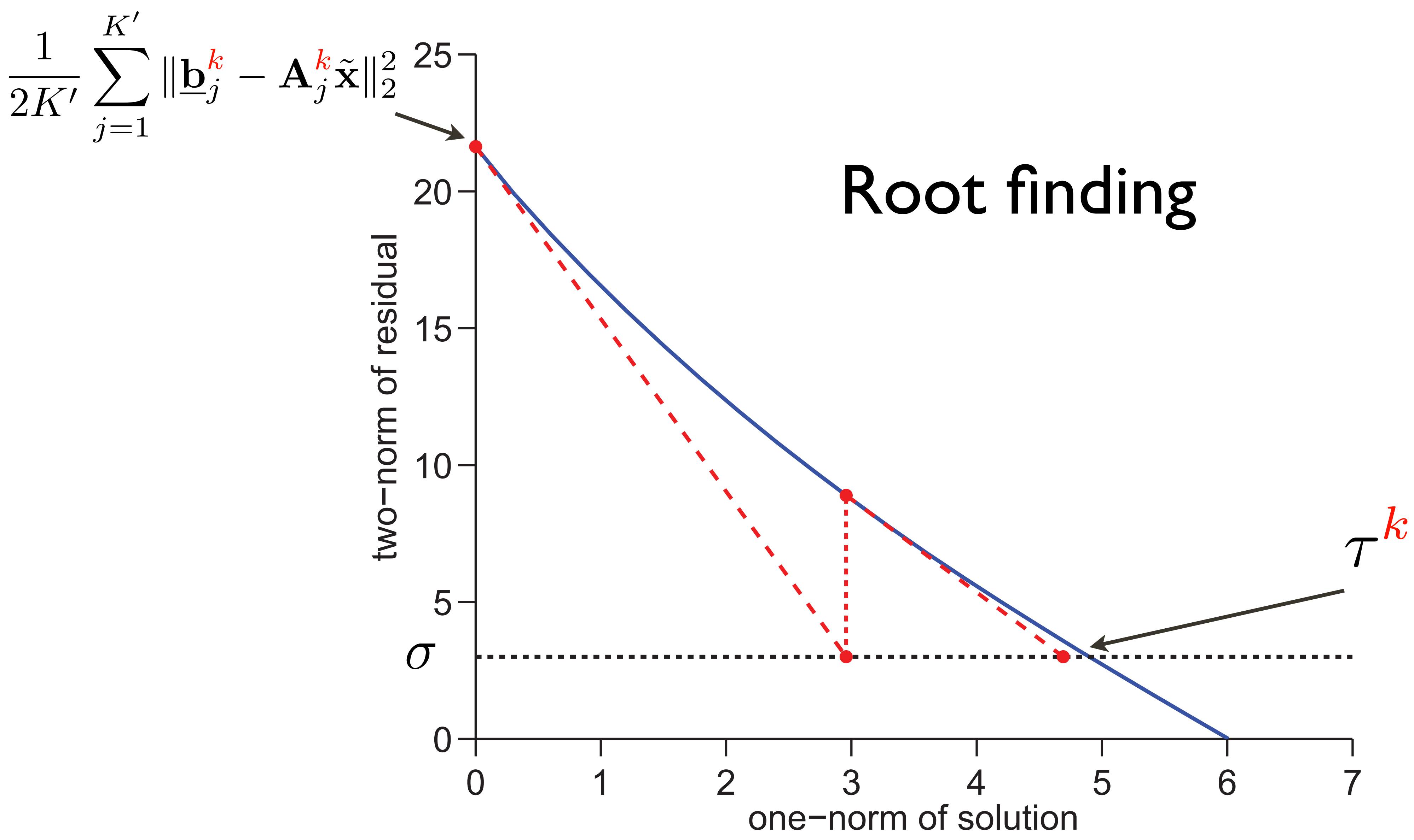
Pareto curve

subproblems



Picking Lasso Parameter

with warm starts



Continuation methods & *renewals*

Underlying assumption is that Pareto curves are similar

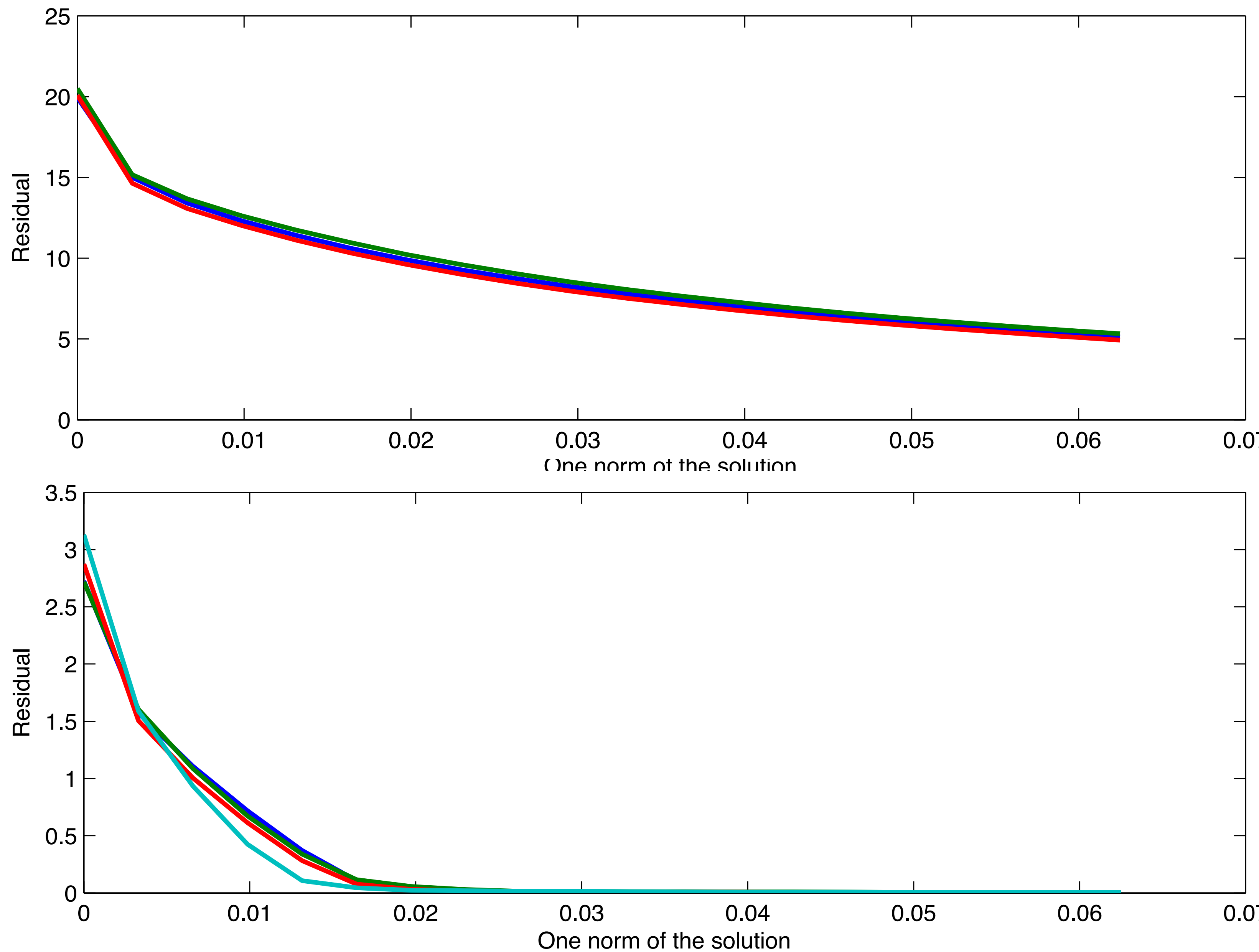
- for large enough batch sizes

In that case the *warm starts* are effective

Renewals remove biases

SLIM

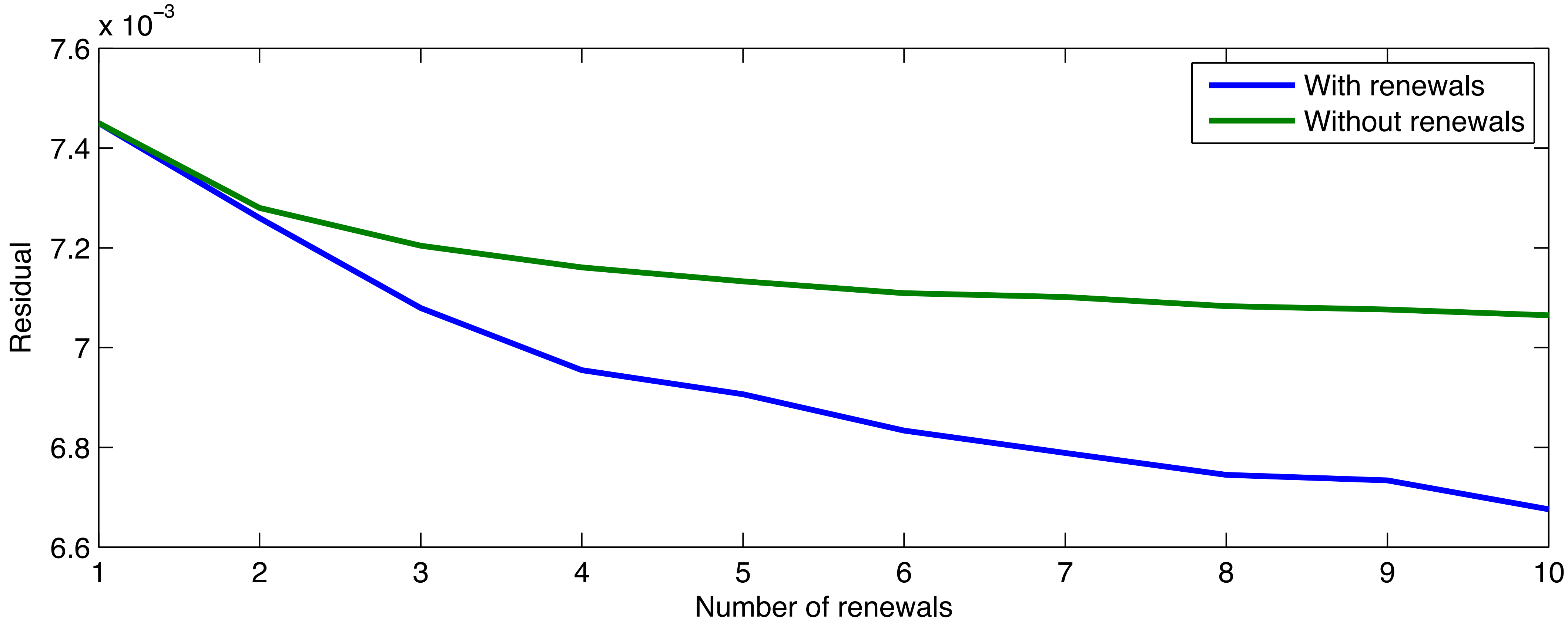
Pareto curves



'large' batch

small batch

Model-space residue



Example

Survey & modeling specifications:

- Model size 143 x 384; mesh distance 24m
- 12 Hz Ricker wavelet; time sample interval 6ms; 192 source locations for each sim-shot
- Input *residual* wavefield computed by subtracting full modeled data and response smooth background model
- Frequencies are selected *randomly* amongst 30 frequencies in the band of the wavelet

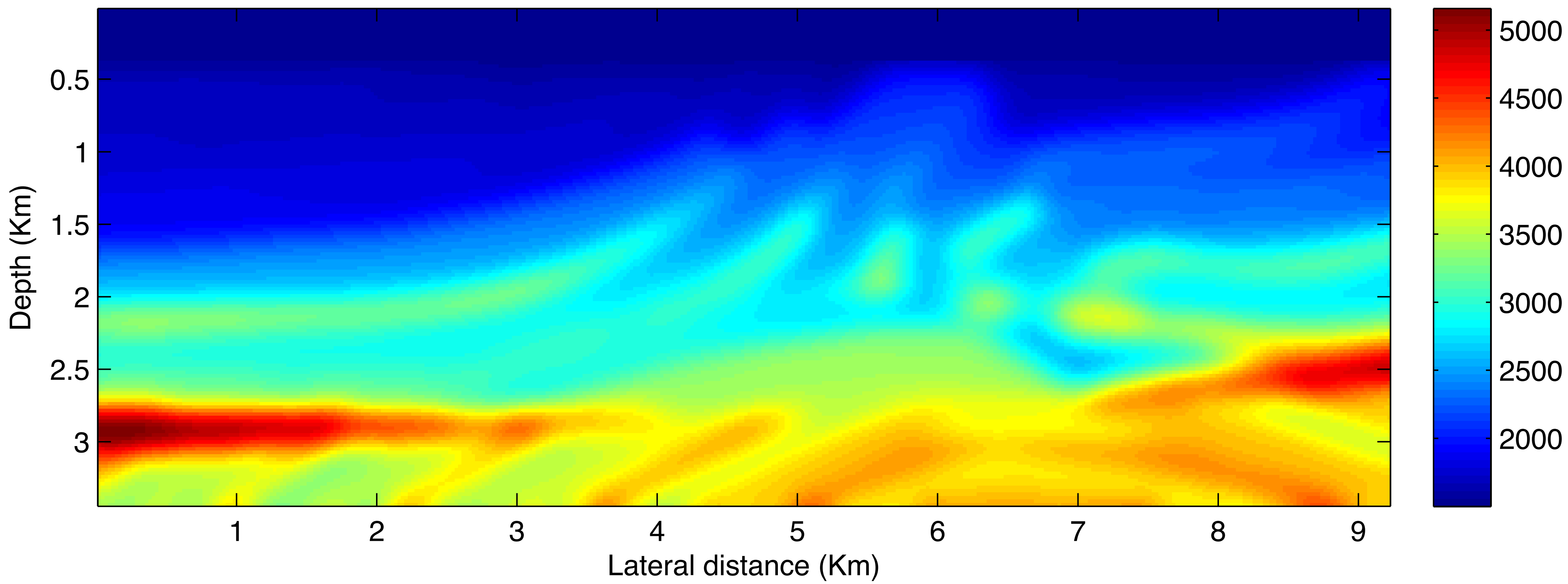
Helmholtz solver

Explicit time-harmonic Helmholtz solver

- 9-point finite difference
- absorbing boundary conditions all around

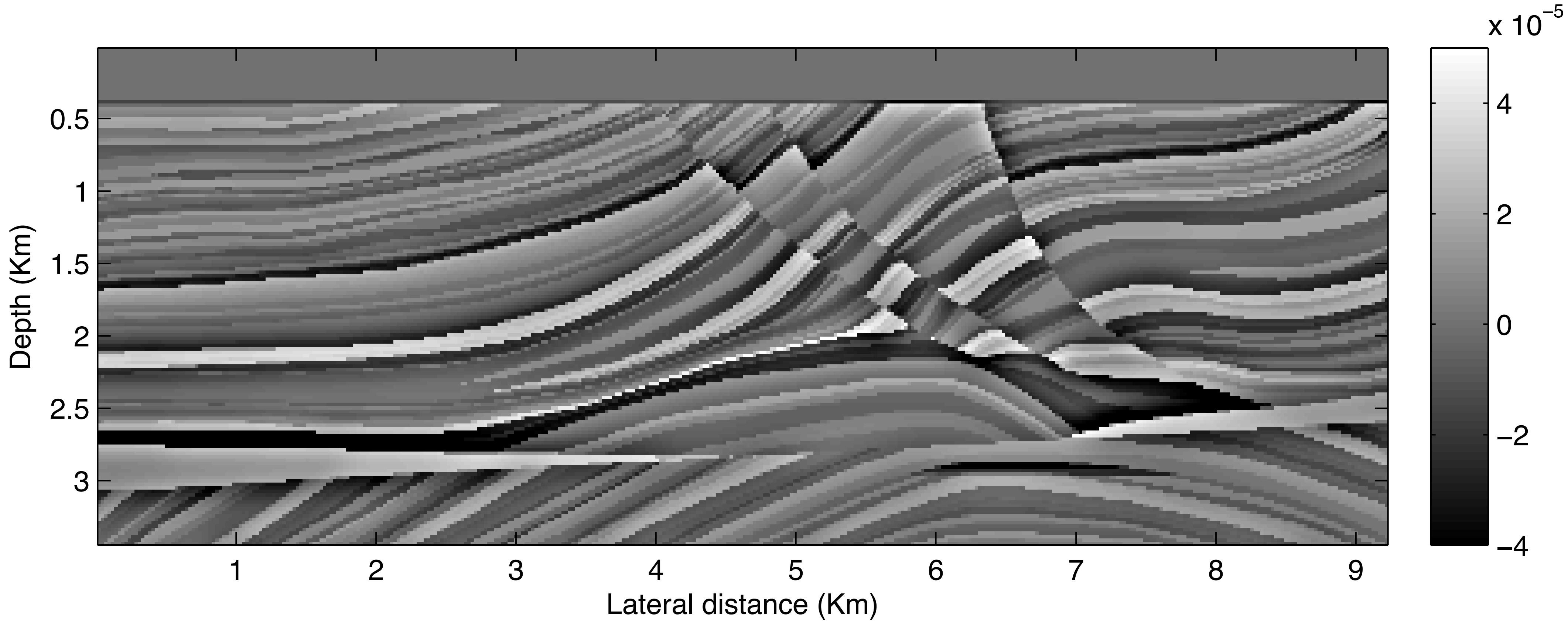
Marmousi model

background model



Marmousi model

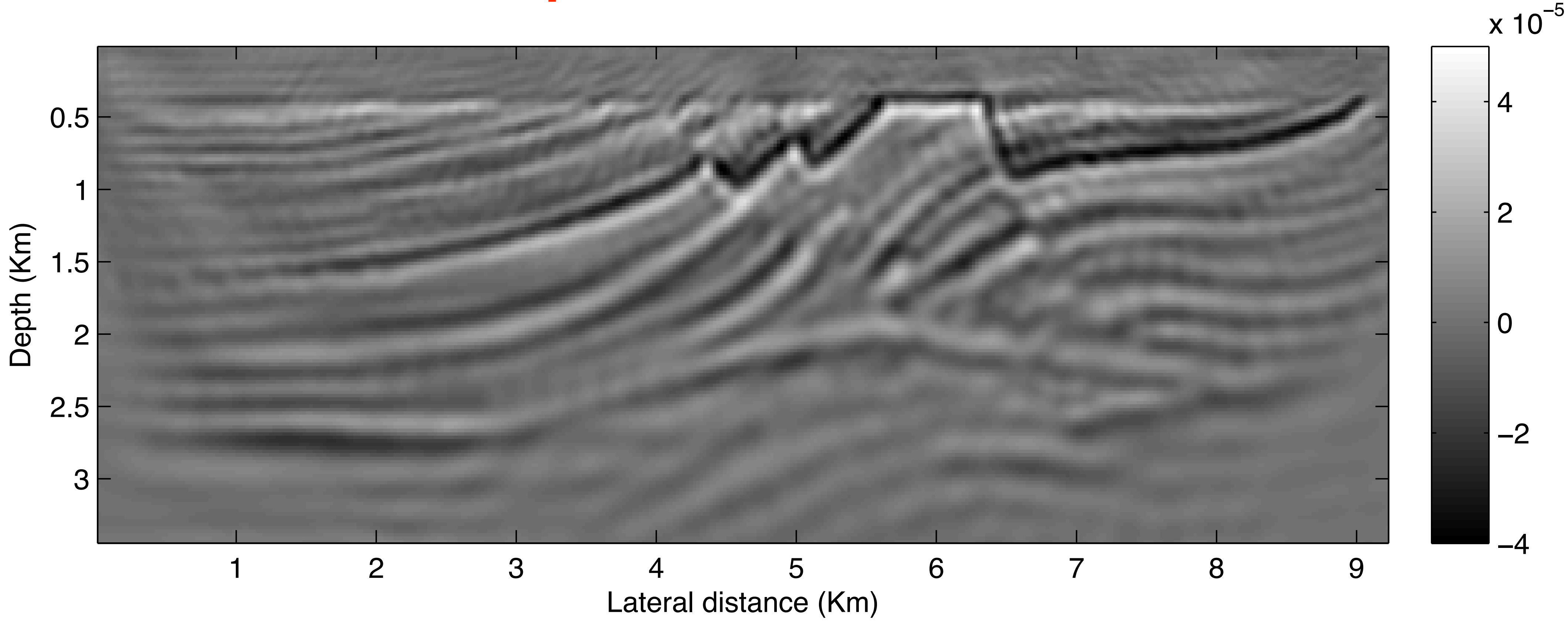
velocity perturbation



Least-squares migration

all 192 shots w 10 frequencies

Aspect ratio: 13

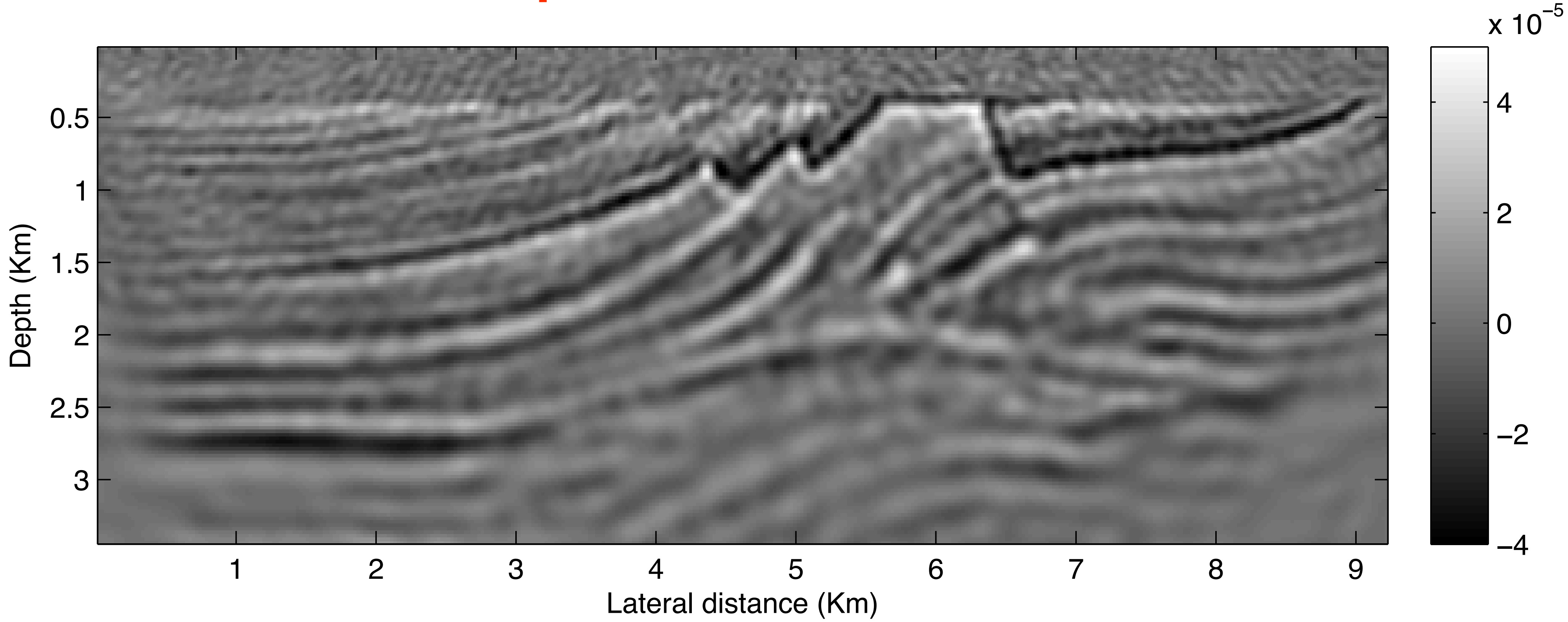


10 LSQR iterations

Sparse migration

8 supershots w 3 frequencies

Aspect ratio: 0.2

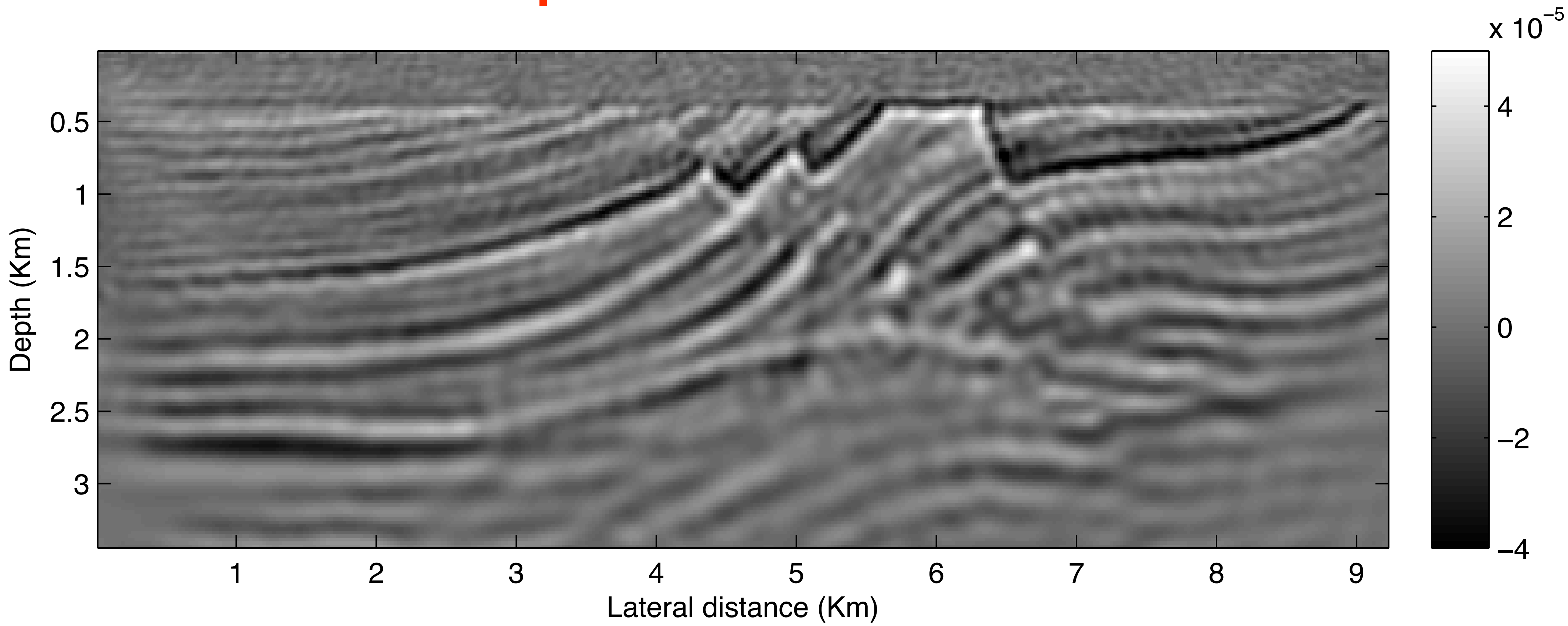


Speed up: x 5

Sparse migration

15 supershots w 8 frequencies

Aspect ratio: 0.8

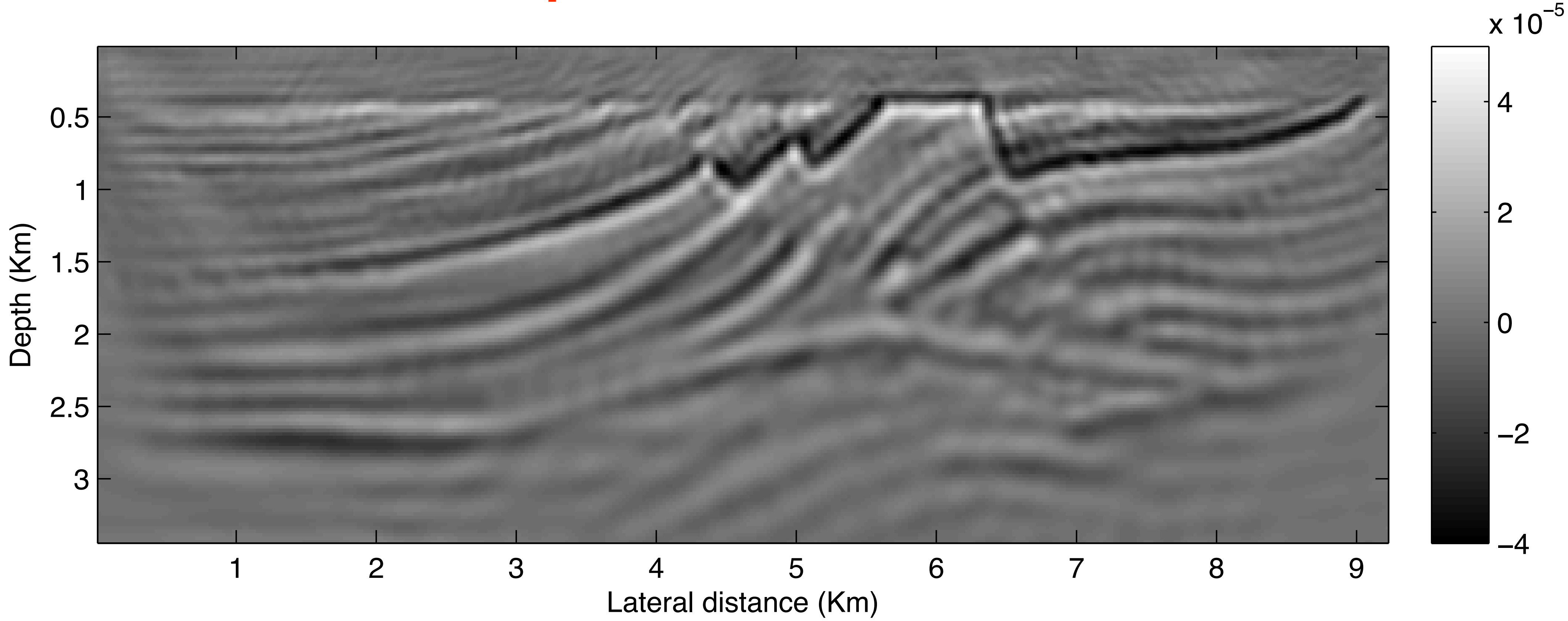


Speed up: x 1

Least-squares migration

all 192 shots w 10 frequencies

Aspect ratio: 13



10 LSQR iterations

SLIM

Conclusions

Sparsity-promotion serves two purposes

- removes *incoherent crosstalk* & *restores* amplitudes
- makes one-norm *regularization* computationally *feasible*
- *quality* of the *migration* results is *adequate* for GN *updates* in FWI

Renewals & *warm starts* are *essential* because they remove *bias*

Expect a bigger *uplift* in 3D

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

Simultaneous & continuous acquisition:

- *A new look at simultaneous sources* by Beasley et. al., '98.
- *Changing the mindset in seismic data acquisition* by Berkhout '08.

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.
- *How to choose a subset of frequencies in frequency-domain finite-difference migration* by Mulder & Plessix, '04.
- *Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies* by Sirque & Pratt, '04.
- *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.
- Randomized dimensionality reduction for full-waveform inversion by FJH & X. Li, '10
- Fast full-wavefield seismic inversion using encoded sources by Krebs et. al., '09
- *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.

Further reading

Compressive sensing & sparse solvers

- Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information by Candes, 06.
- Compressed Sensing by D. Donoho, ’06
- Probing the Pareto frontier for basis pursuit solutions by E. van den Berg and M. Friedlander, ’08

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

Thank you

slim.eos.ubc.ca