# Full-waveform inversion with randomized L1 recovery for the model updates

Xiang Li, Aleksandr Aravkin, Tristan van Leeuwen and Felix J. Herrmann

SLIM
University of British Columbia

# Motivation

*Curse of dimensionality* for d>2

- *Exponentially* increasing *data volumes*
- *Helmholtz* requires *iterative* solvers to address *bandwidth*
- Computational complexity grows *linearly* with # of frequencies & RHS's
- High-resolution image

# Wish list

An *imaging* technology that

- is based on a *time-harmonic* PDE solver, which is easily *parallelizable*, and *scalable* to 3D
- does *not* require many passes through *all* data
- removes the *linearly* increasing costs of *iterative* solvers for increasing numbers of frequencies & RHS's

# Key technologies

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]

- stochastic gradient descent

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

- *sparse* recovery & *randomized* subsampling

# 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

# Dimensionality reduction

Migration operator and its adjoint requires 2 PDE solves

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

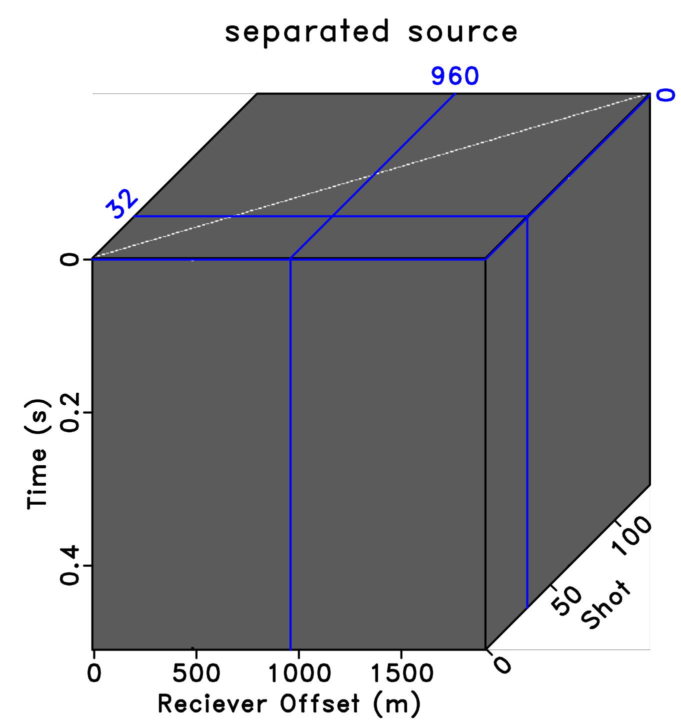
$$\delta\widetilde{\mathbf{m}} = \arg\min_{\delta\mathbf{m}} \frac{1}{2}\|\delta\underline{\mathbf{d}} - \nabla\mathcal{F}[\mathbf{m}_0;\underline{\mathbf{Q}}]\delta\mathbf{m}\|_2^2$$

$$\delta\underline{\mathbf{d}} = \text{Simultaneous-source data residue}$$

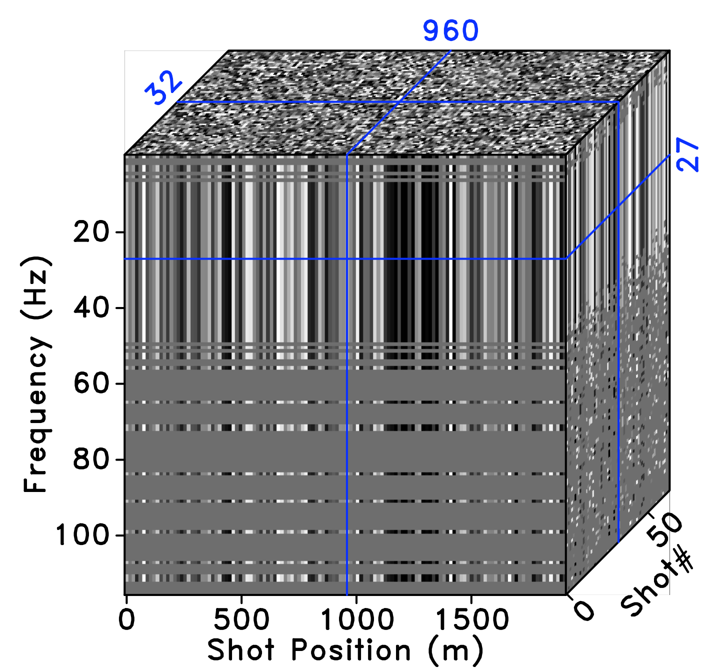
$$\underline{\mathbf{Q}} = \text{Simultaneous sources}$$

[Herrmann et. al. '08-'10]

# Supershot




$\mathbf{Q}$

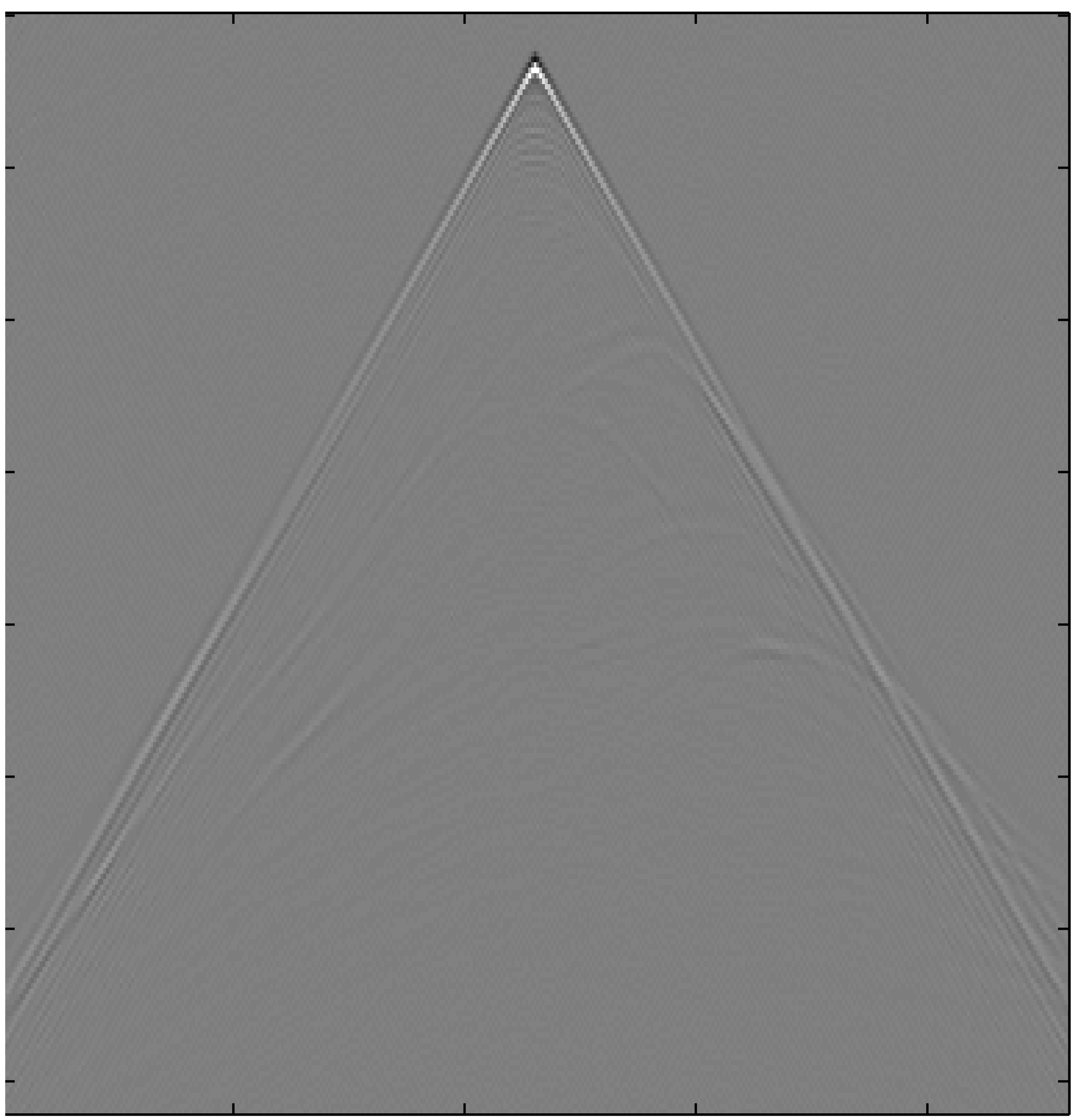



$\underline{\mathbf{Q}} = \mathbf{RMQ}$

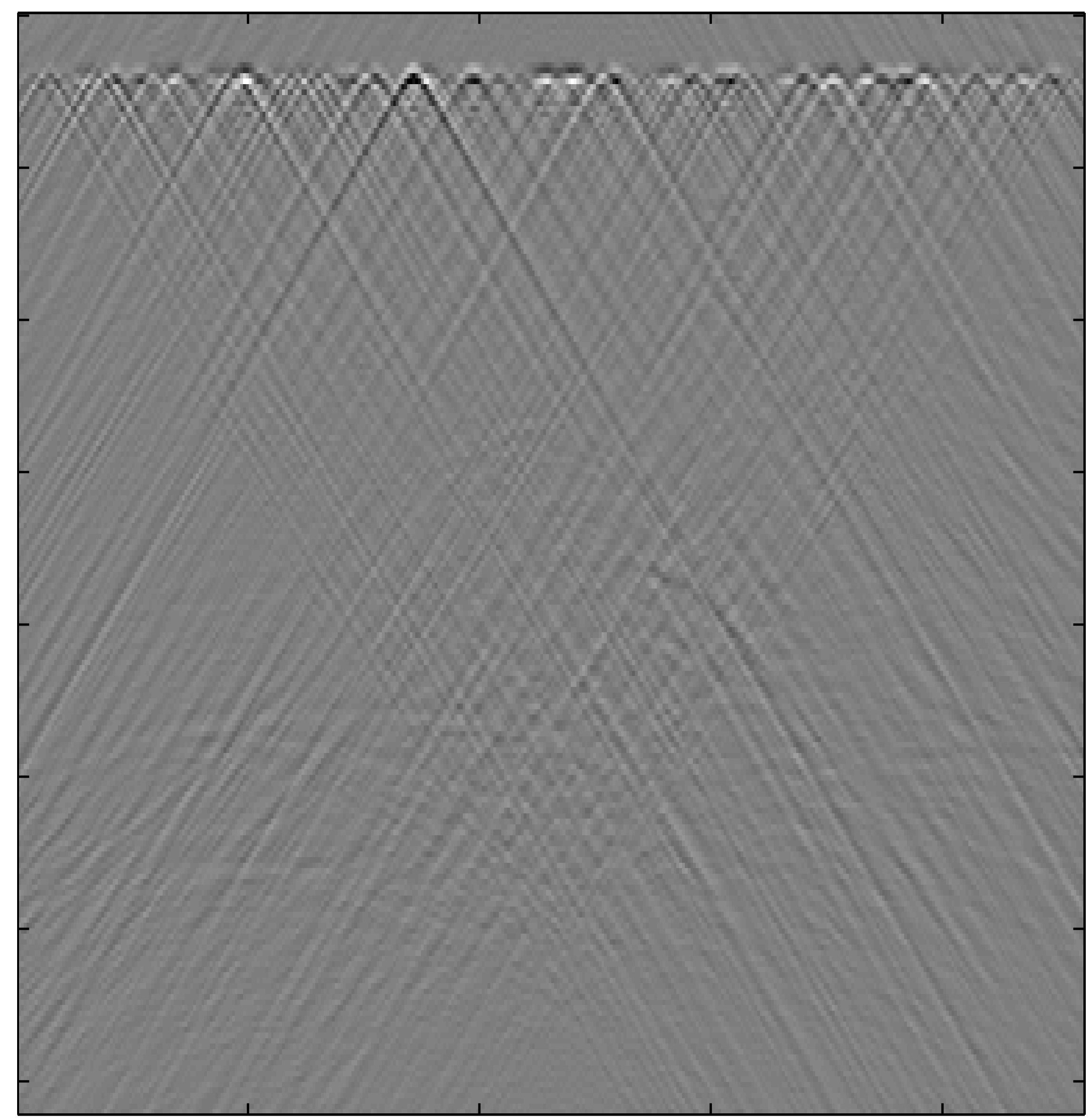
Collection of K *simultaneous-source* experiments with batch size $K \ll n_f \times n_s$

# One shot record

Sequential-Shot



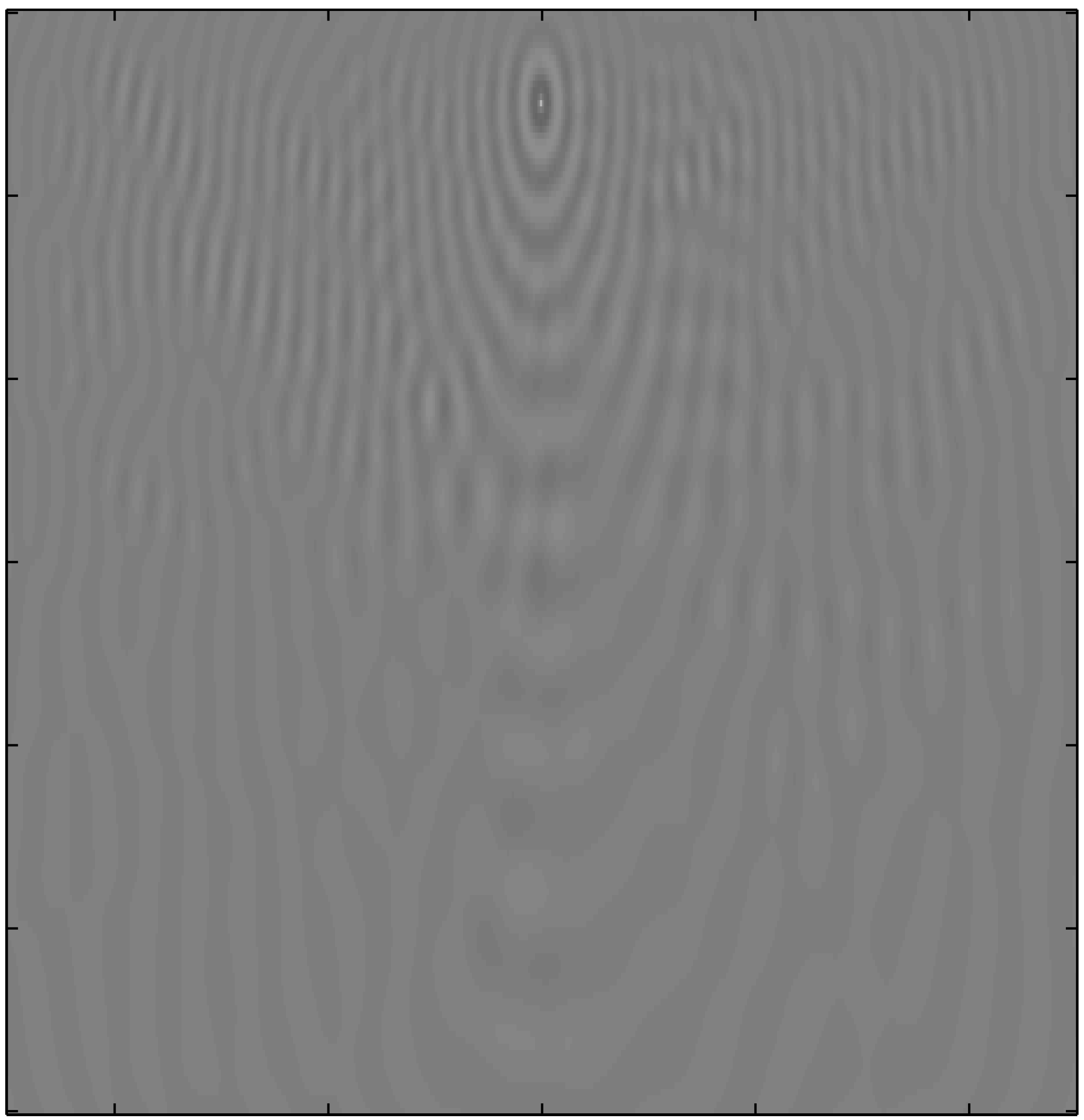
*Simultaneous*-Shot

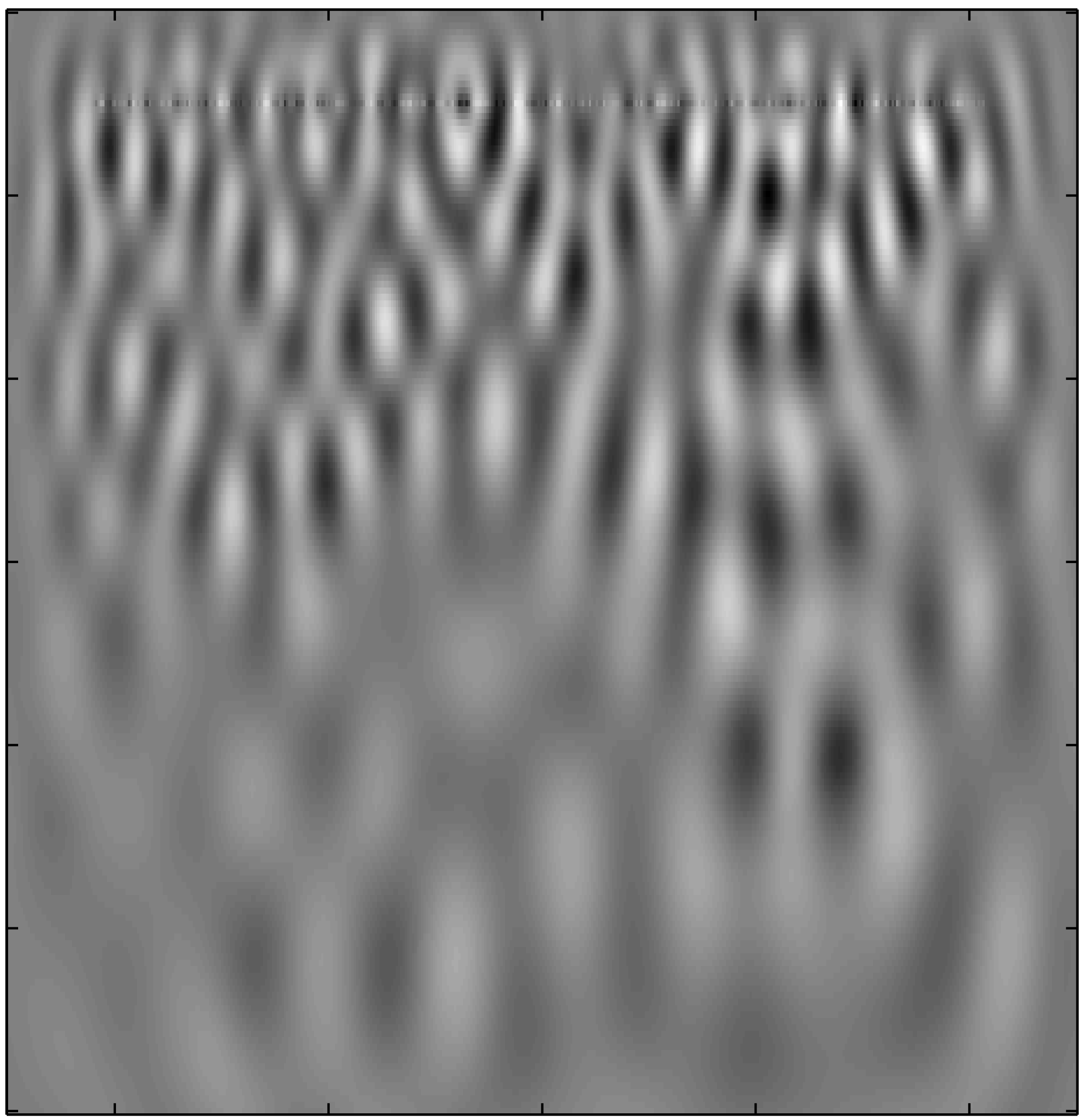


# Simultaneous shot

Sequential-source wavefield

*Simultaneous*-source wavefield

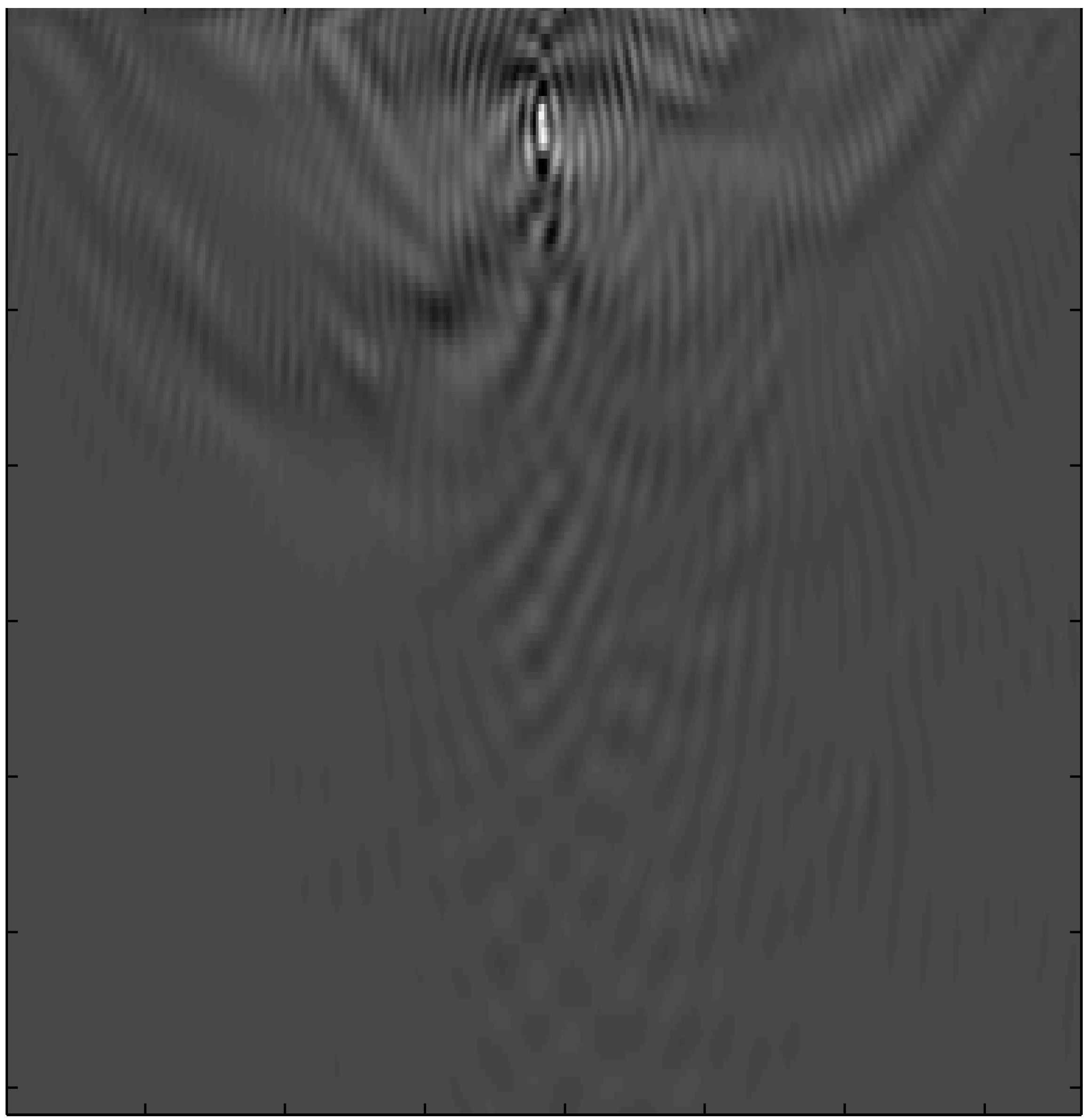




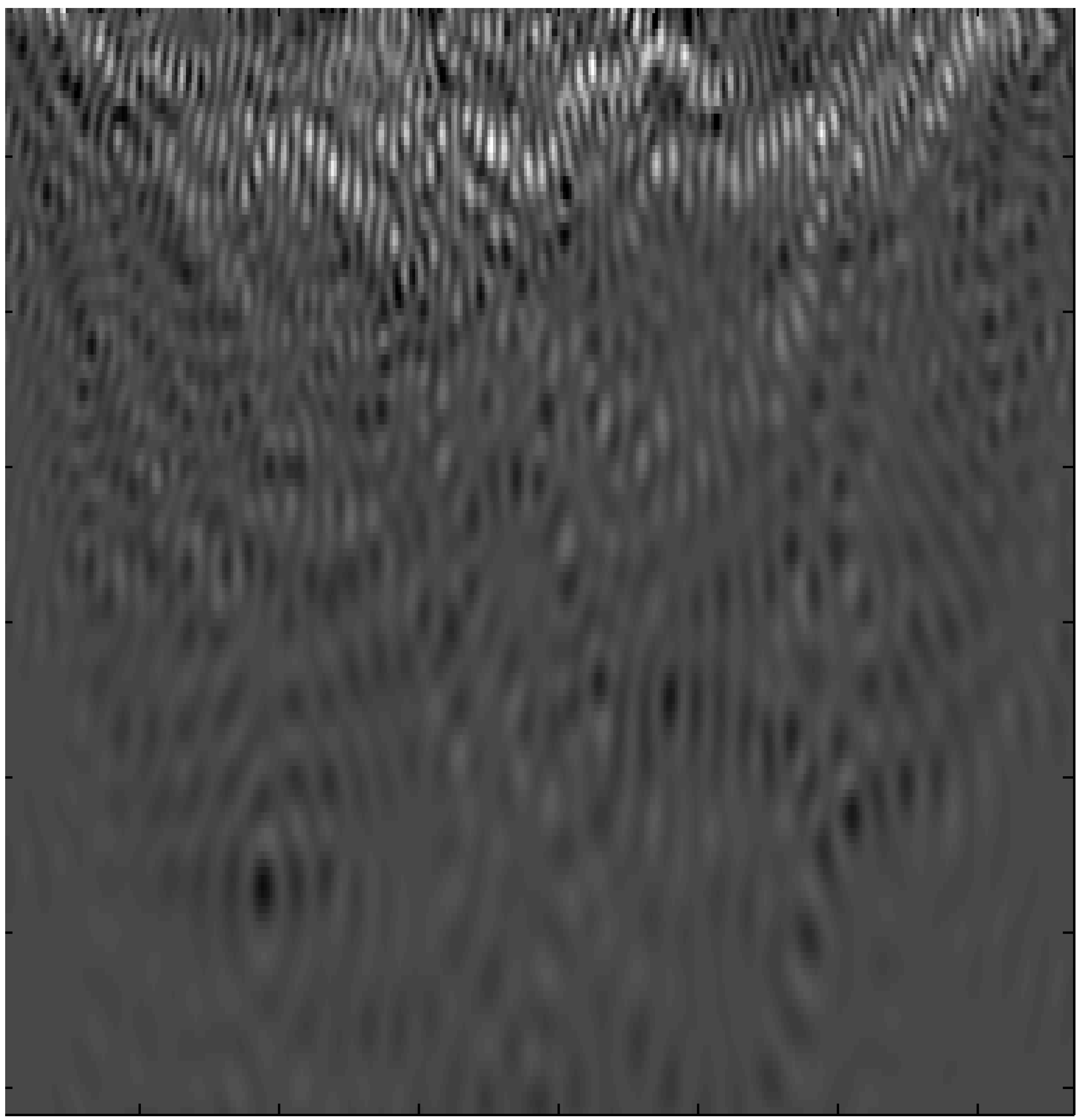
SLIM

# Image from one shot

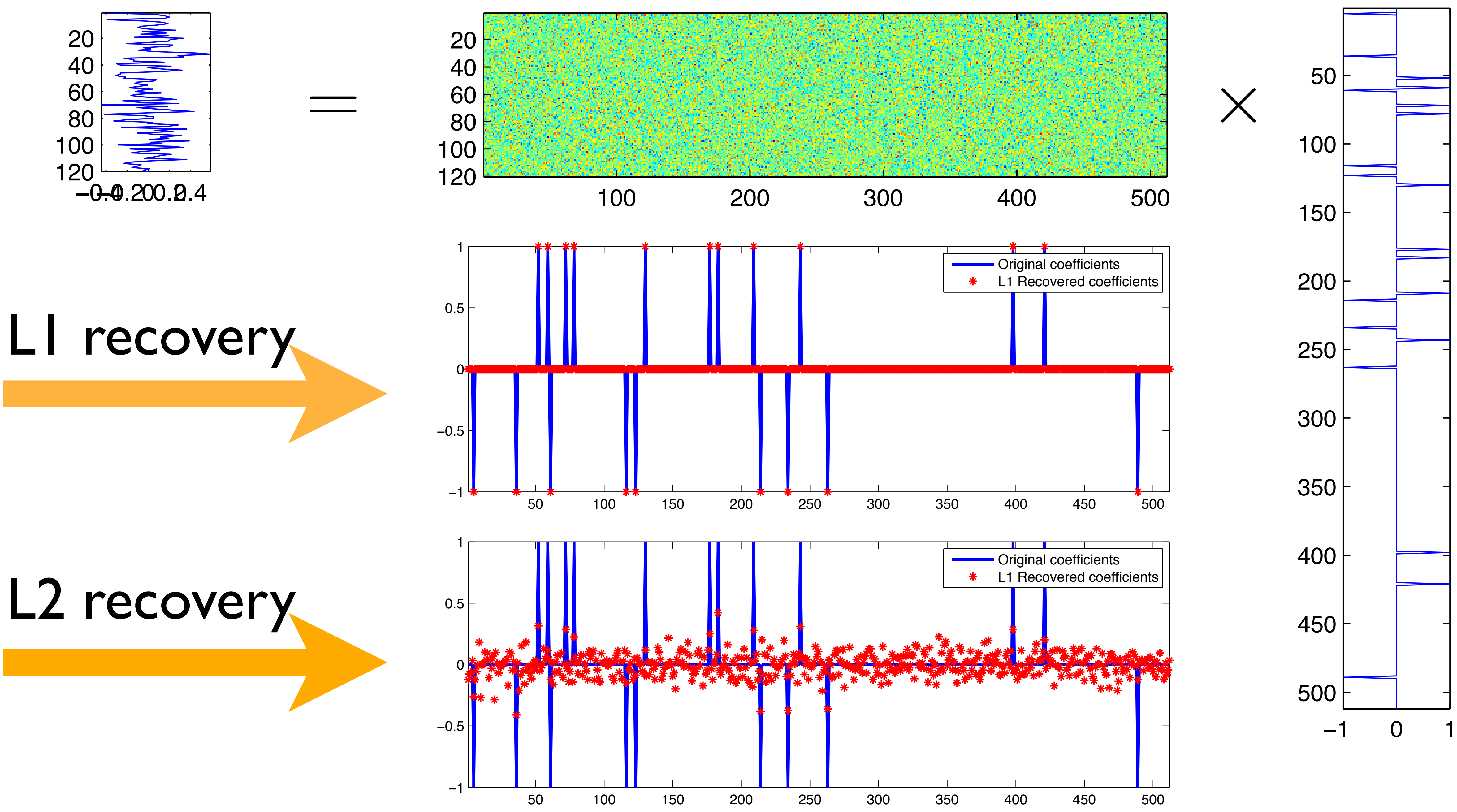
Sequential-source image



*Simultaneous*-source image

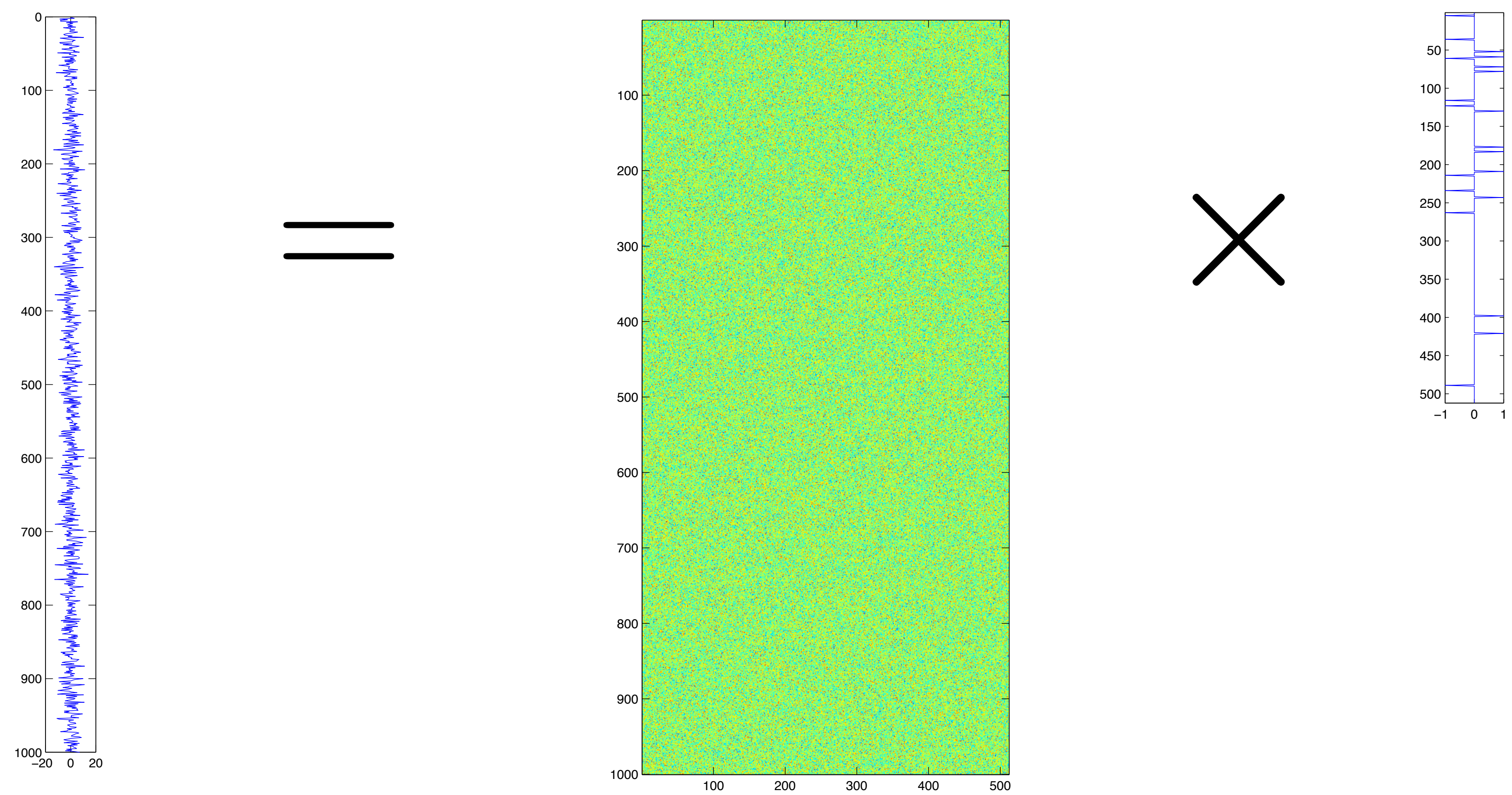


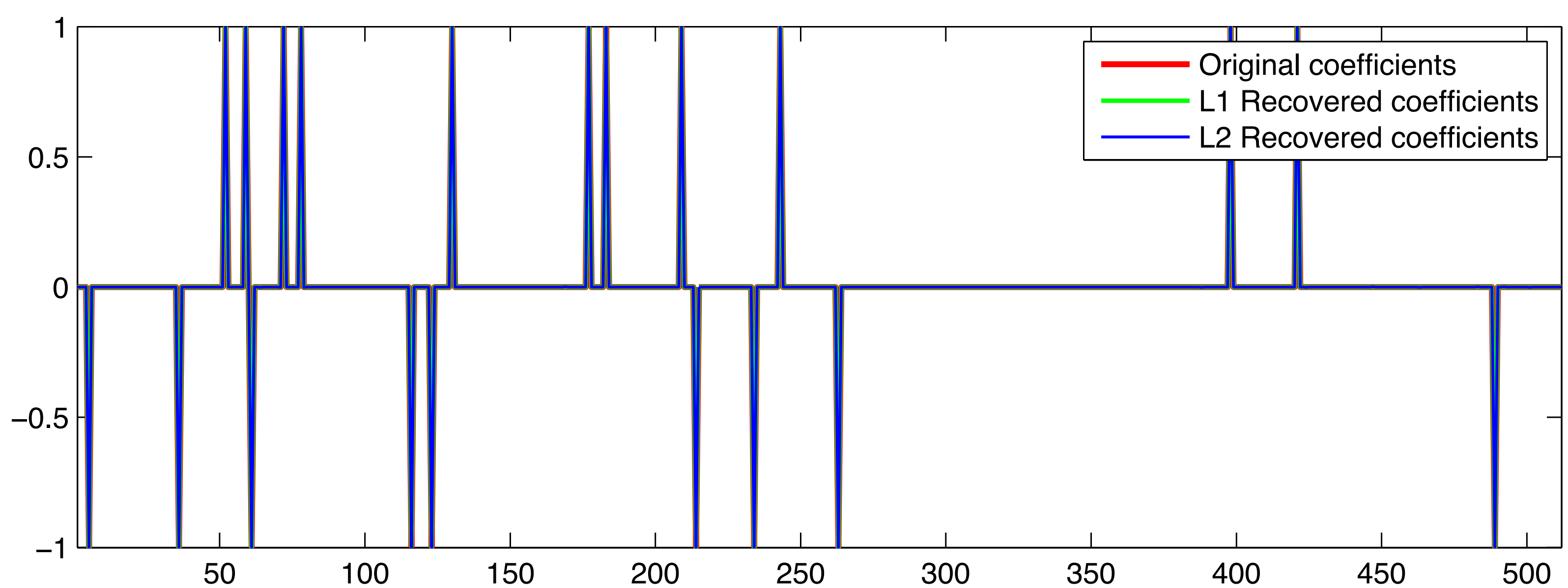
# Sparse recovery



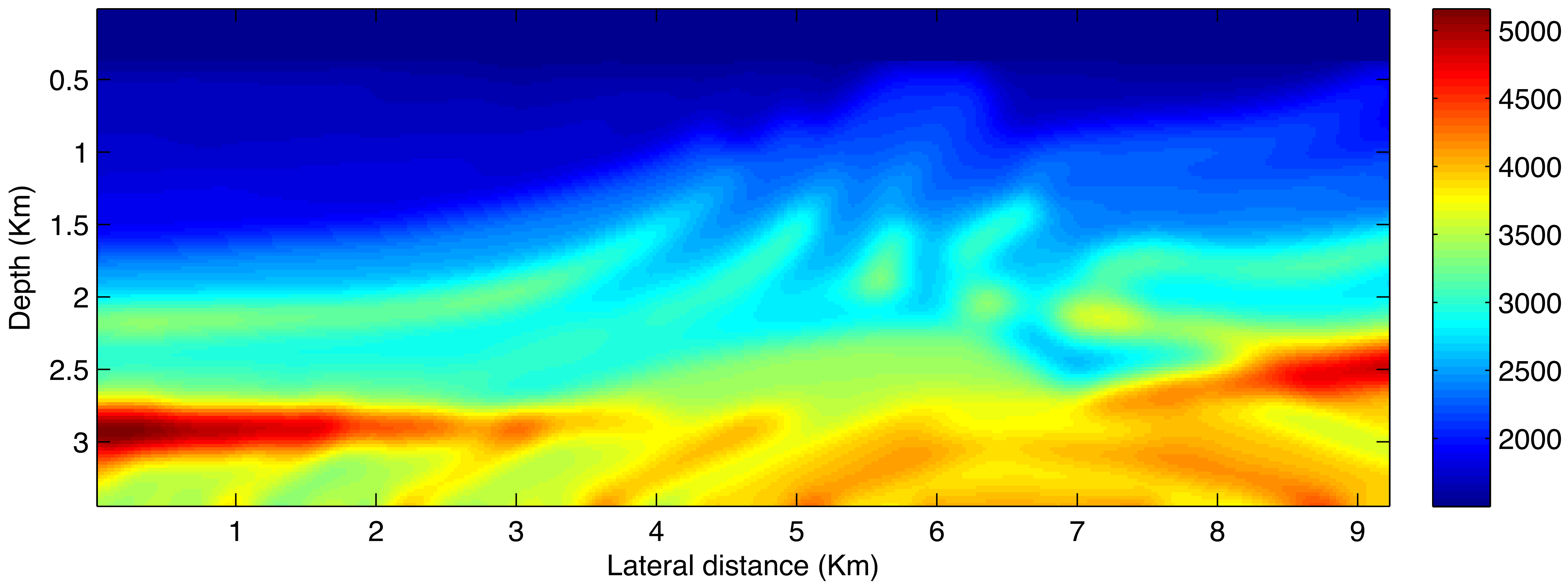

[van den Berg & Friedlander, ’08]

# Recovery

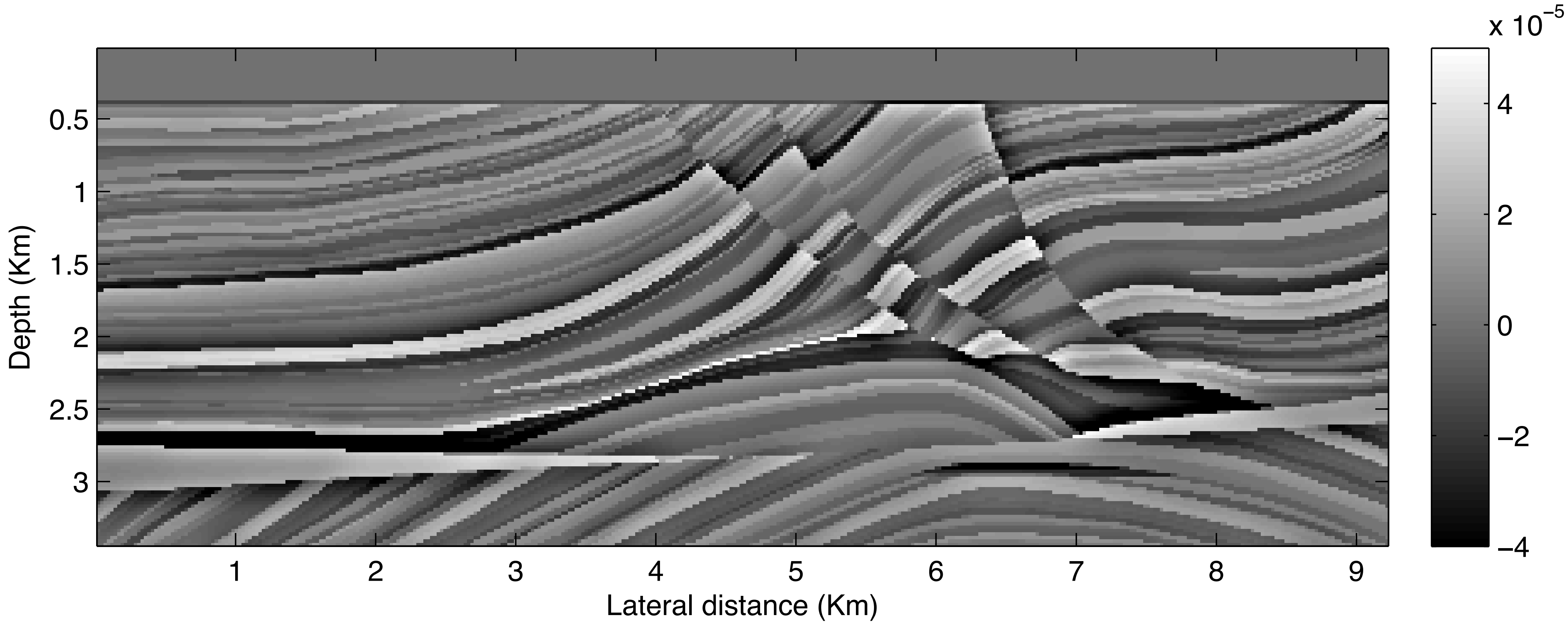



Original coefficients
L1 Recovered coefficients
L2 Recovered coefficients

# Initial model

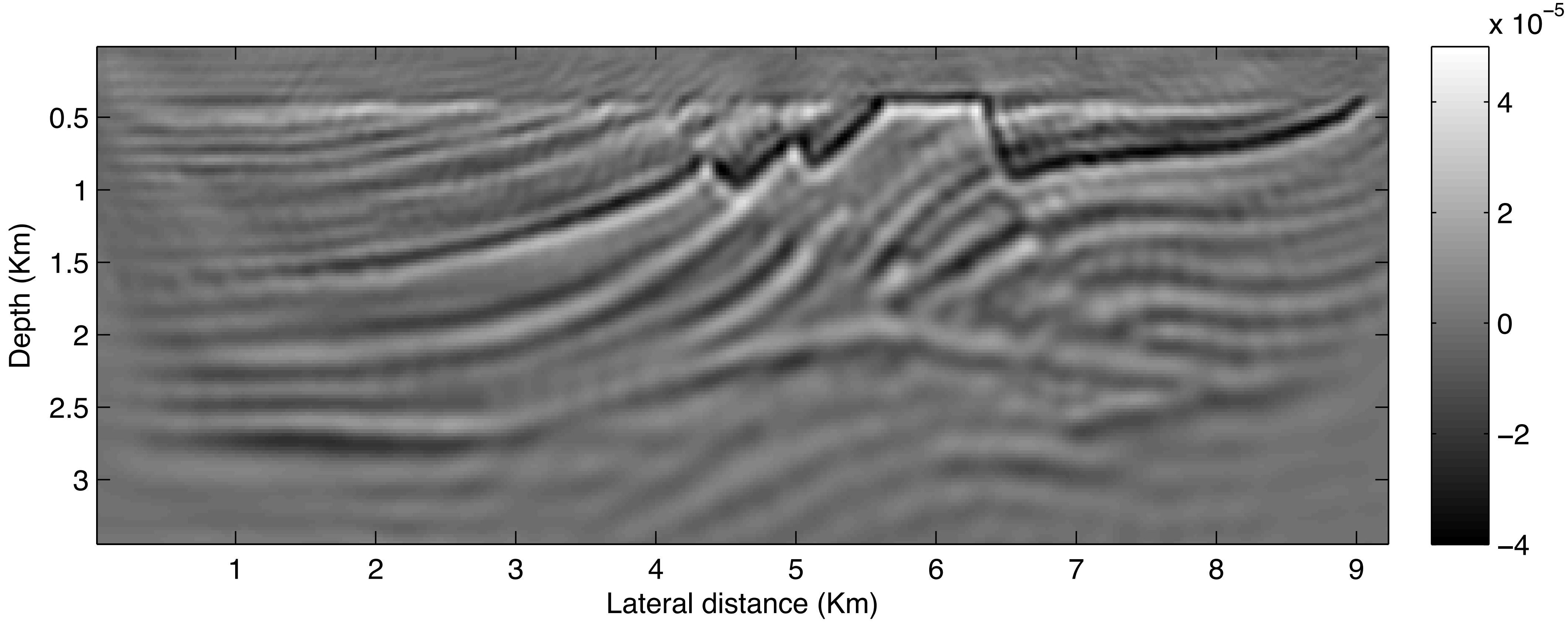


# True perturbation



# L2 for all shots

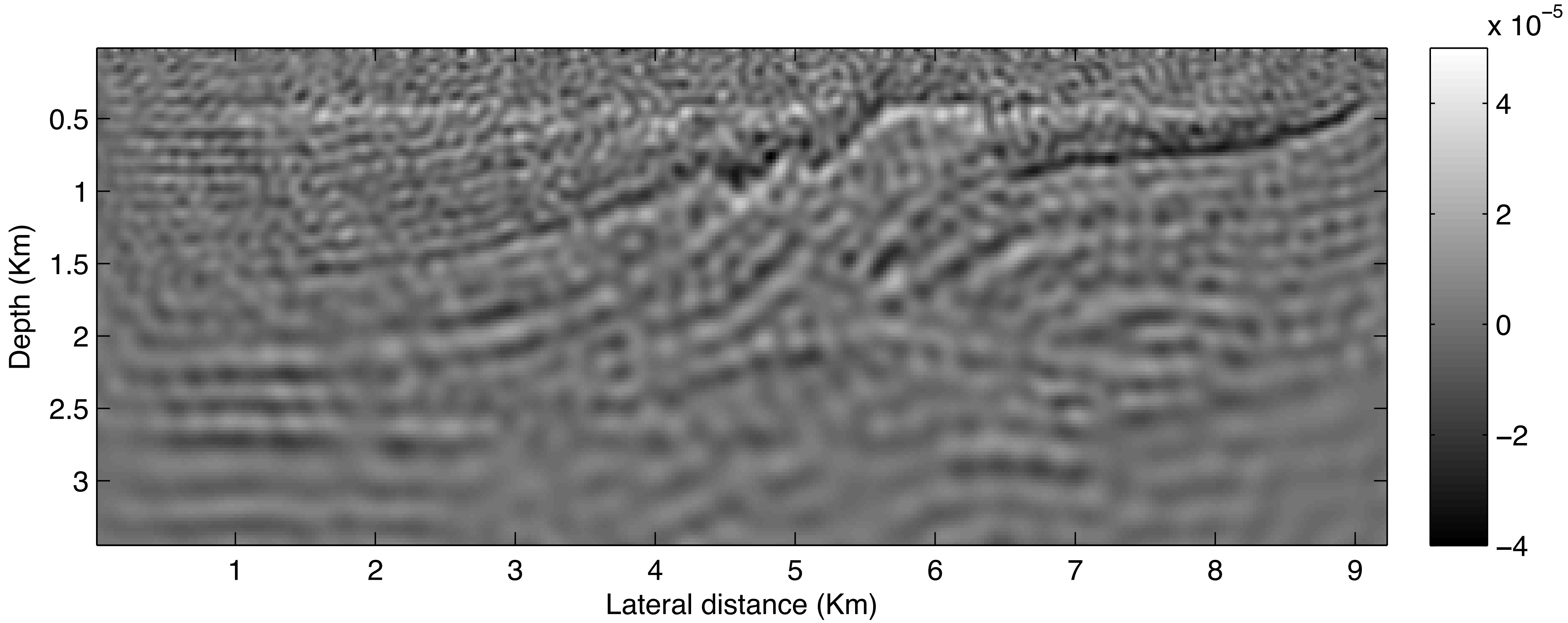
192 shots and 10 frequencies




Size of sampling matrix: 737280 x 49152

# L2 for supershots

8 supershots and 3 frequencies are used




Size of sampling matrix: 9216X49152

[Wang & Sacchi, '07]

# Sparse recovery

Least-squares migration with *sparsity* promotion

$$\delta\widetilde{\mathbf{m}} = \mathbf{S}^* \arg\min_{\delta\mathbf{x}} \frac{1}{2}\|\delta\mathbf{x}\|_{\ell_1} \quad \text{subject to} \quad \|\delta\underline{\mathbf{d}} - \nabla\boldsymbol{\mathcal{F}}[\mathbf{m}_0;\underline{\mathbf{Q}}]\mathbf{S}^*\delta\mathbf{x}\|_2 \leq \sigma$$

$$\delta\mathbf{x} \quad = \quad \text{Sparse curvelet-coefficient vector}$$

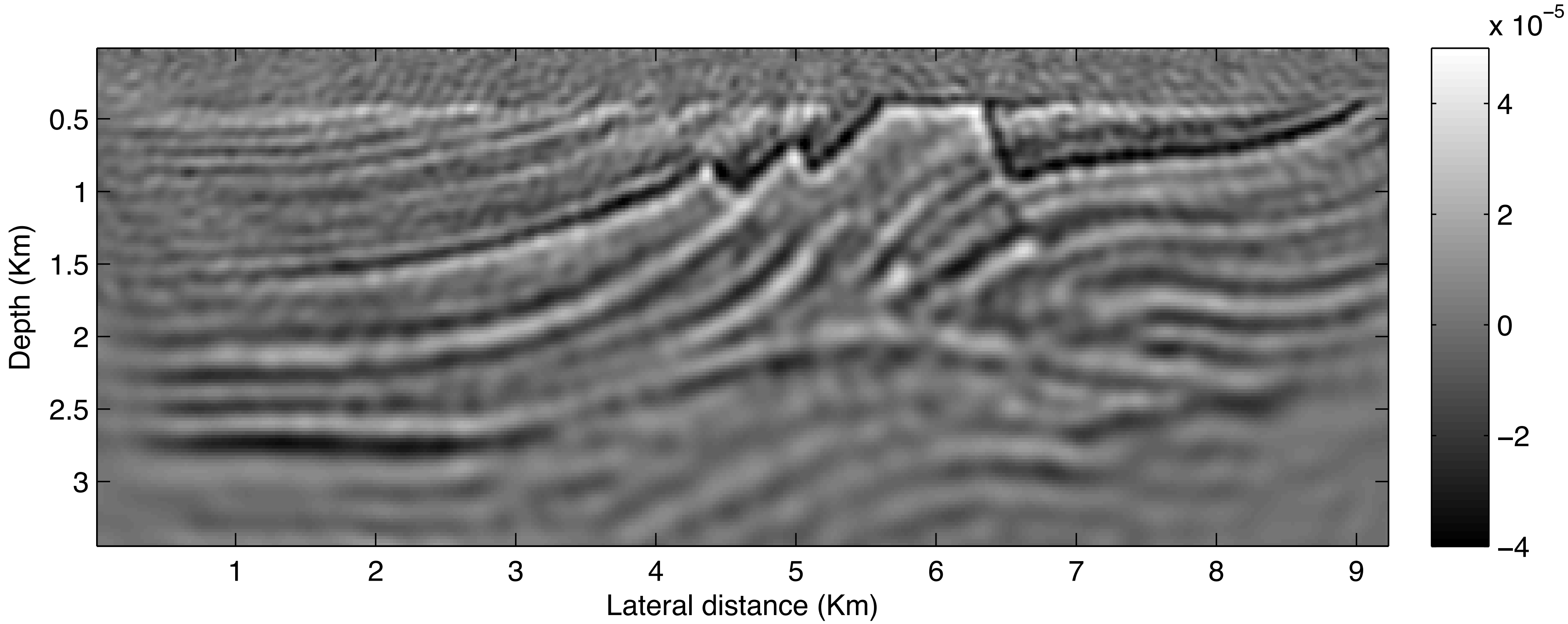
$$\mathbf{S}^* \quad = \quad \text{Curvelet synthesis}$$

leads to *significant* speedup as long as

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

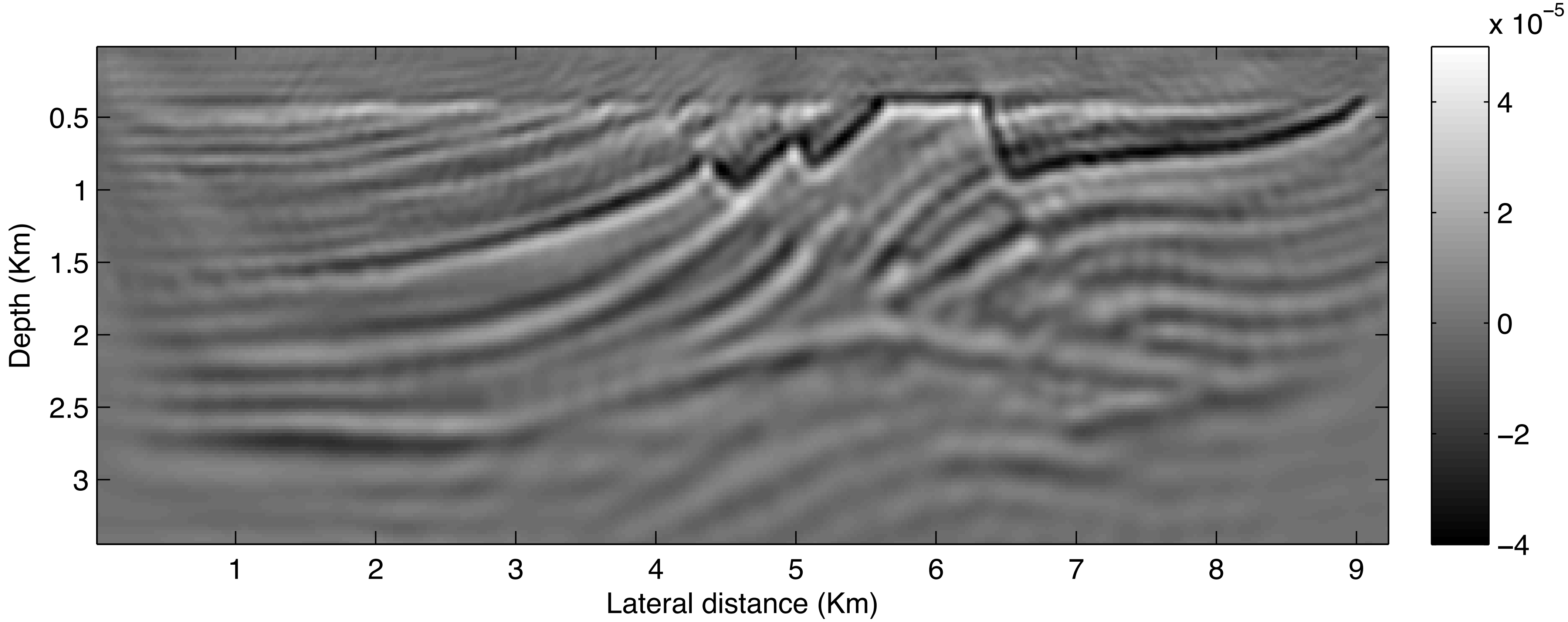
# L1 with small batch

8 supershots and 3 frequencies are used



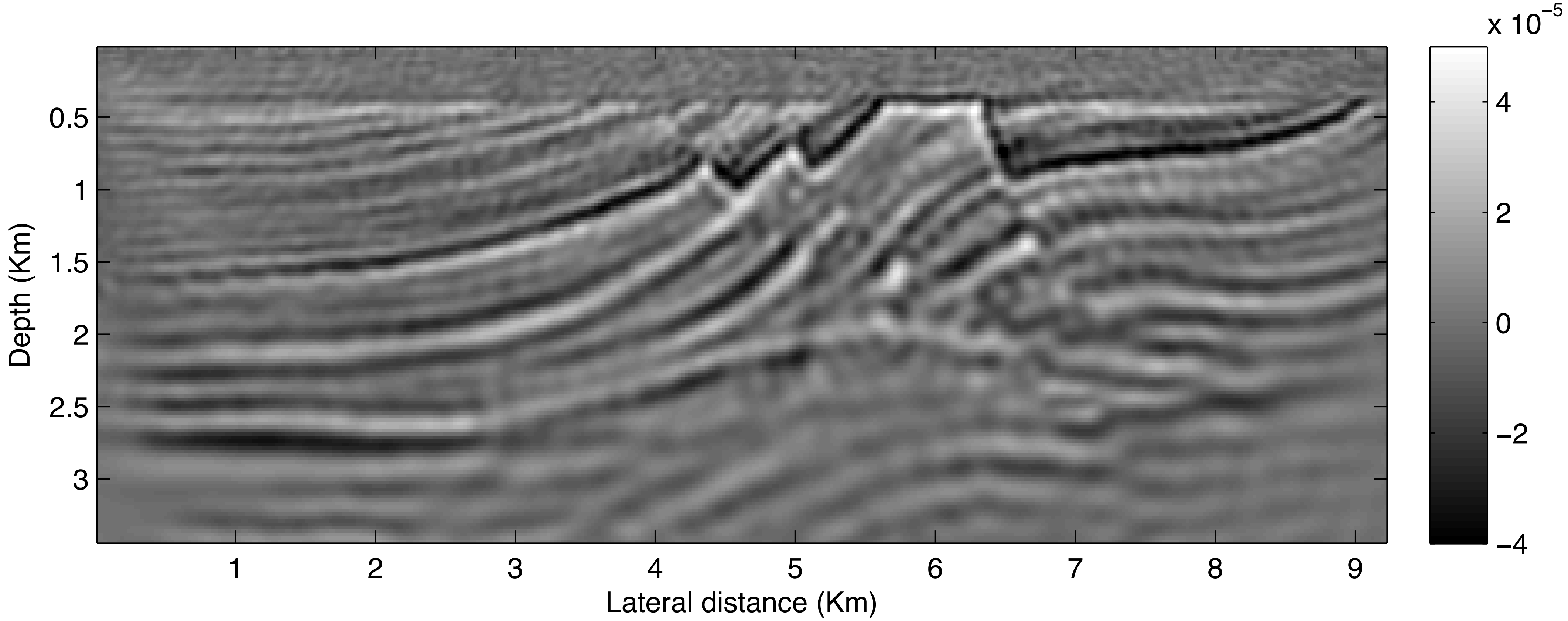
# L2 for all shots

192 shots and 10 frequencies



# Different batch size

15 supershots and 7 frequencies



# Observations

- Reconstruct images from smaller *under-determined* system with *sparse* recovery
- Sparse optimization is possible because of small batch size

# Compressive updates

**Algorithm 1**: Gauss Newton with sparse updates

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

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

**while** not converged **do**

$\quad \delta\mathbf{m}^k \longleftarrow \mathbf{S}^* \arg\min_{\mathbf{x}} \frac{1}{2}\|\delta\underline{\mathbf{d}}^k - \nabla\boldsymbol{\mathcal{F}}[\mathbf{m}^k;\underline{\mathbf{Q}}^k]\mathbf{S}^*\mathbf{x}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_1 \leq \tau^k$

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

$\quad k \longleftarrow k + 1$;

**end**

[van den Berg & Friedlander, '08]

# Renewals

Use *different* batch of sources and frequencies for each *subproblem*

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

- motivated by *stochastic approximation* [Nemirovski, '09]
- *justifies* approach by Krebs *et.al.*, 2009

# 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
- remove *bias* by selecting *new supershots* for each GN *iteration*

# Example

Marmousi model:

- 128x384 with a mesh size of 24 meters
- 192 shots and 384 receivers with offset = 3 X depth
- 3.6s recording time for Marmousi

*Explicit* Time-harmonic Helmholtz solver

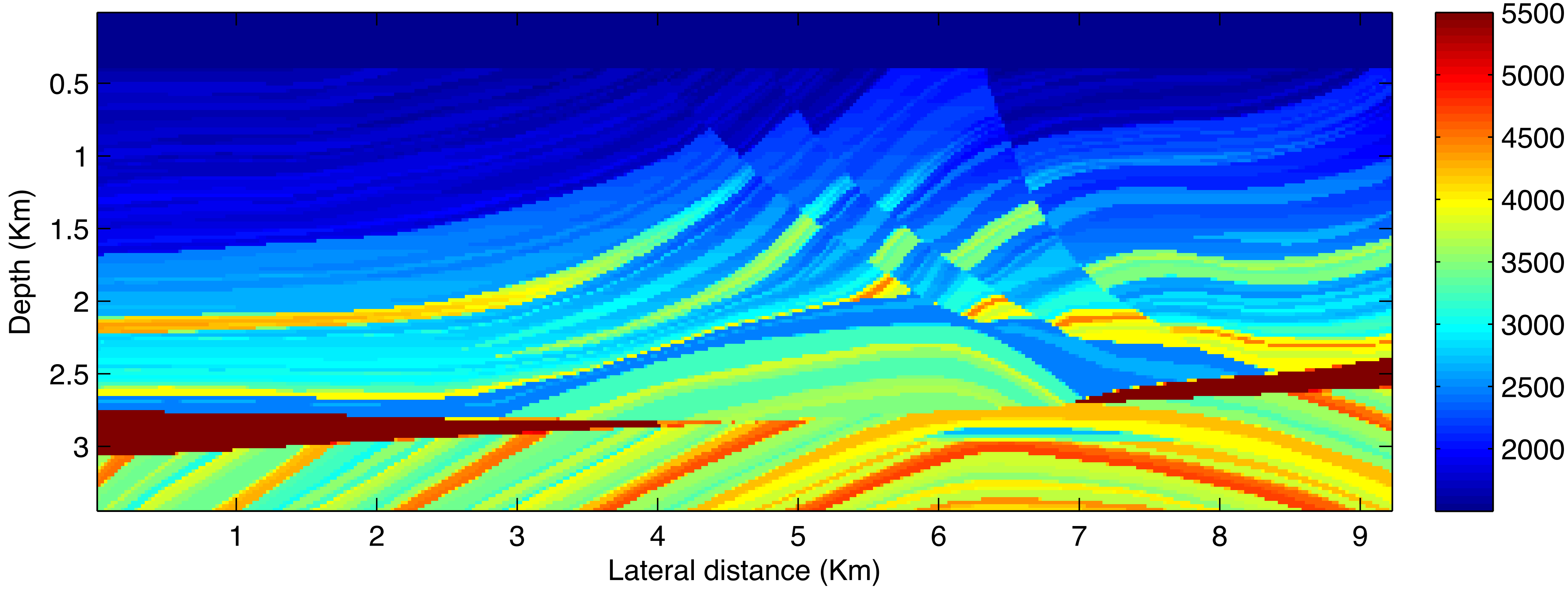
- 9-point finite difference
- absorbing boundary condition
- 12 Hz Ricker source wavelet

# Example

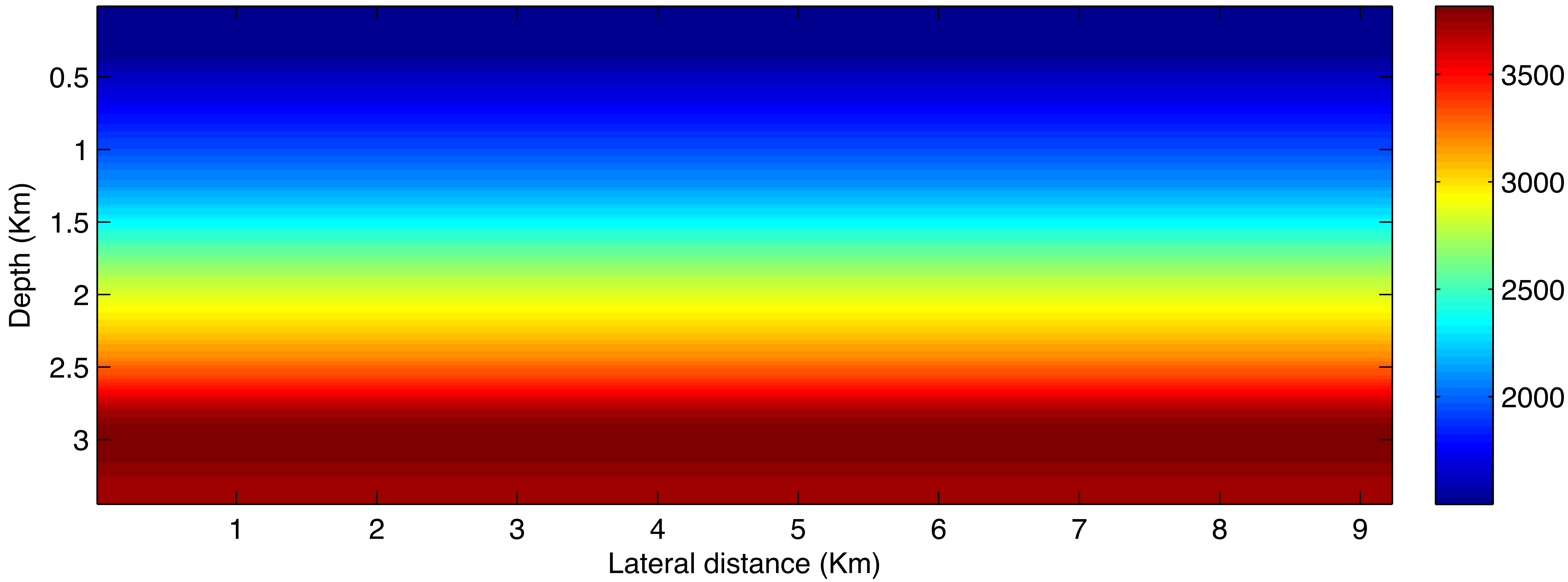
FWI specs:

- Frequency continuation over 10 bands
- 10 *frequencies* for each frequency band
- Start from 2.2 Hz

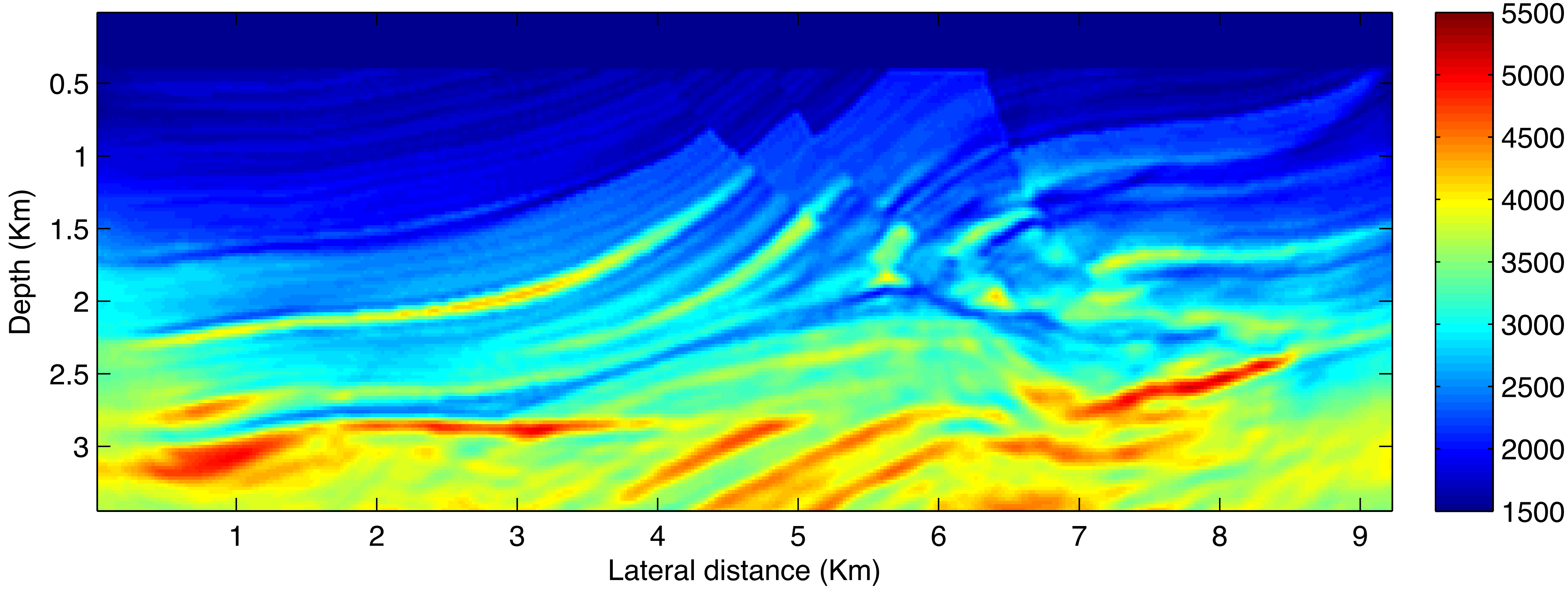
# True model


0.5
1
1.5
2
2.5
3
Depth (Km)
1
2
3
4
5
6
7
8
9
Lateral distance (Km)
5500
5000
4500
4000
3500
3000
2500
2000

# Initial model

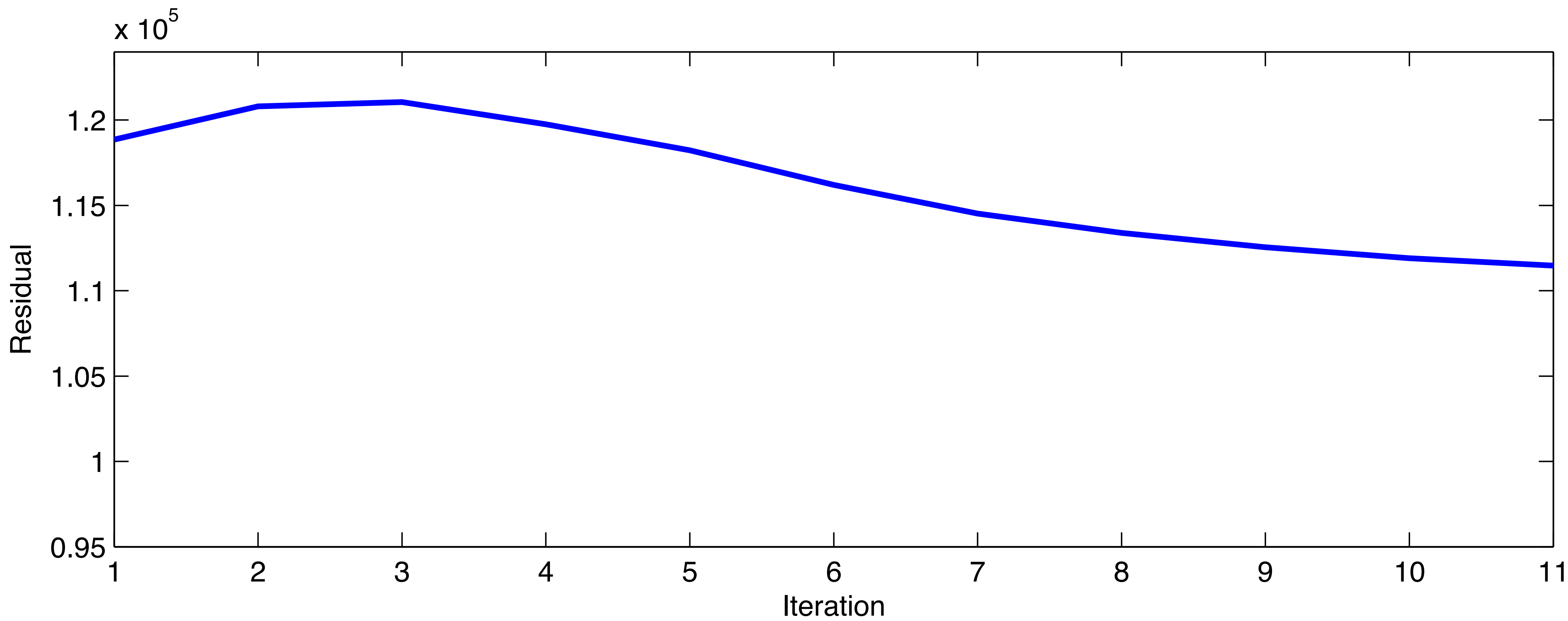


# L2 with *all* shots

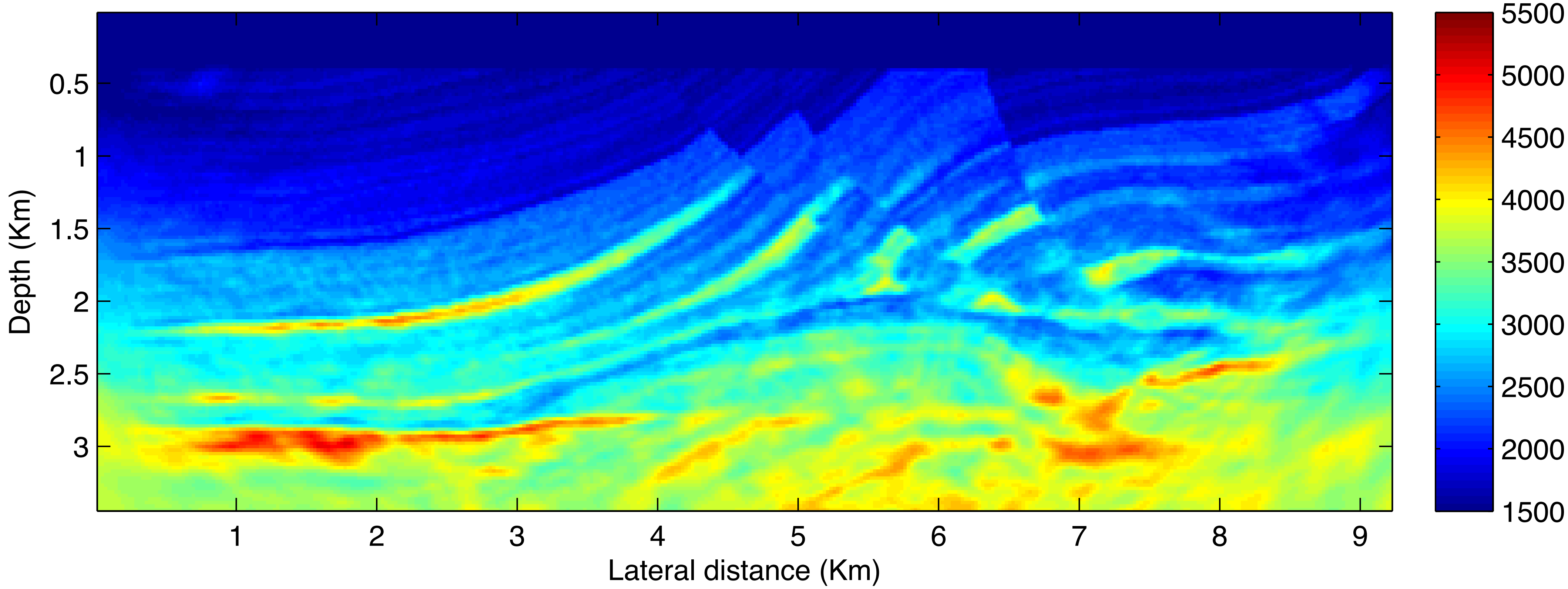



10f X192s X 45 X 2 = 172800 PDEs

# *Model* residual



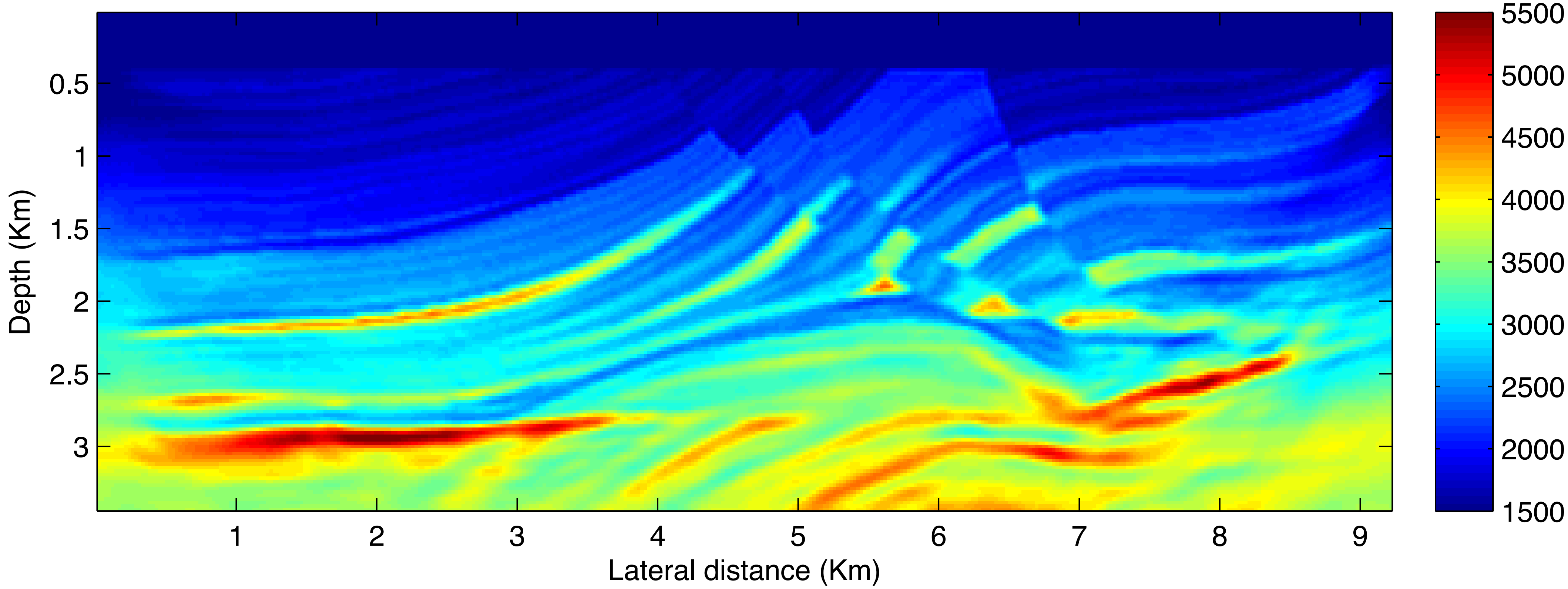
# L1 without renewals




Speed up: X10

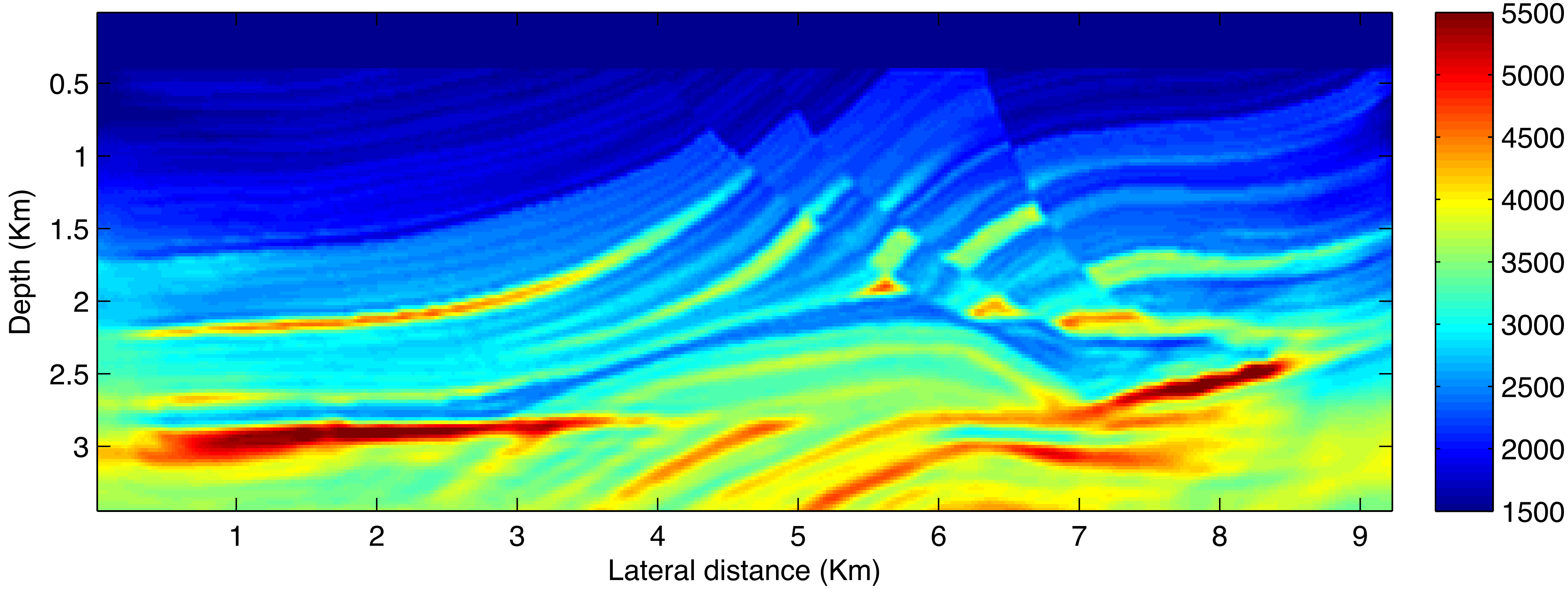
SLIM

# L1 with renewals



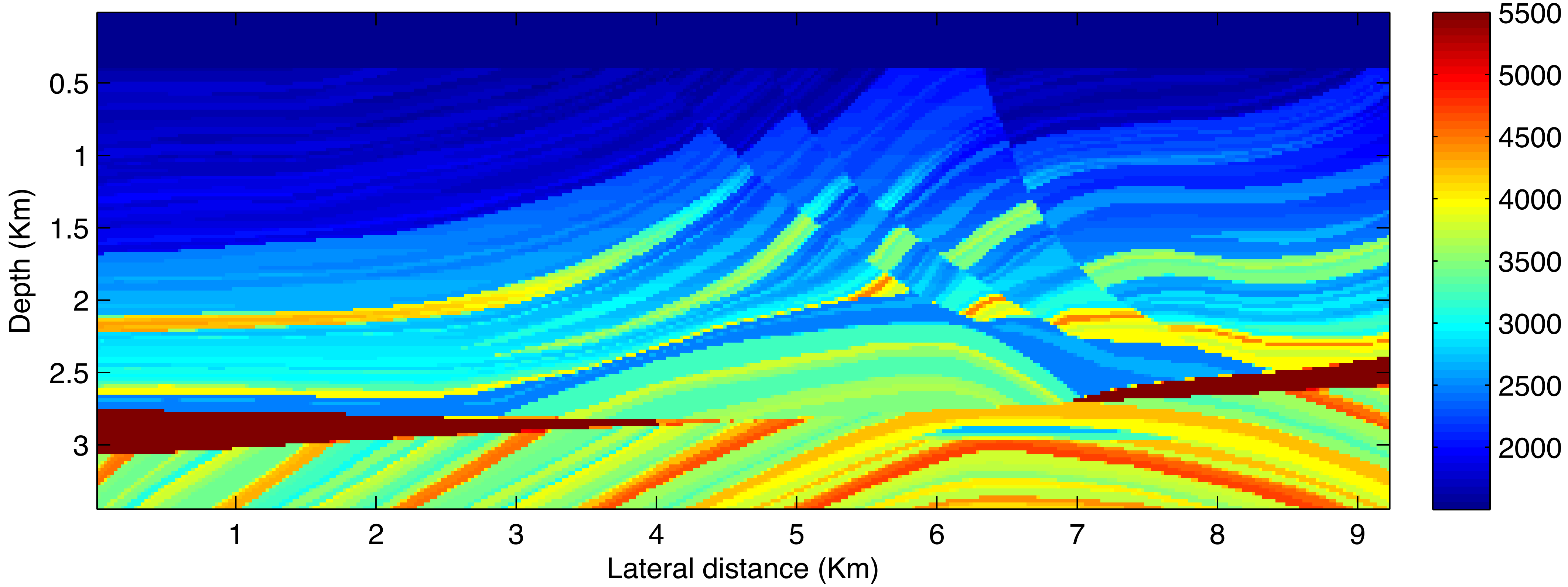

Speed up: X10

# L1 with renewals



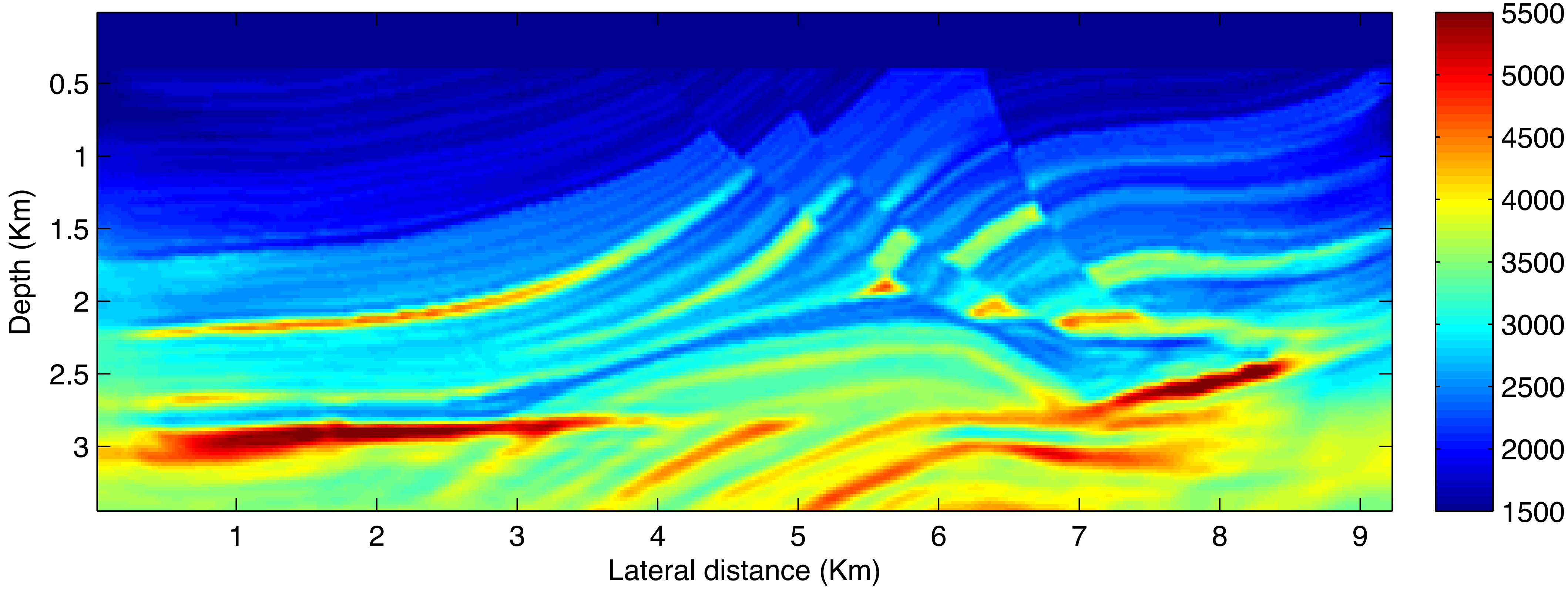

Speed up: X 3

# True result



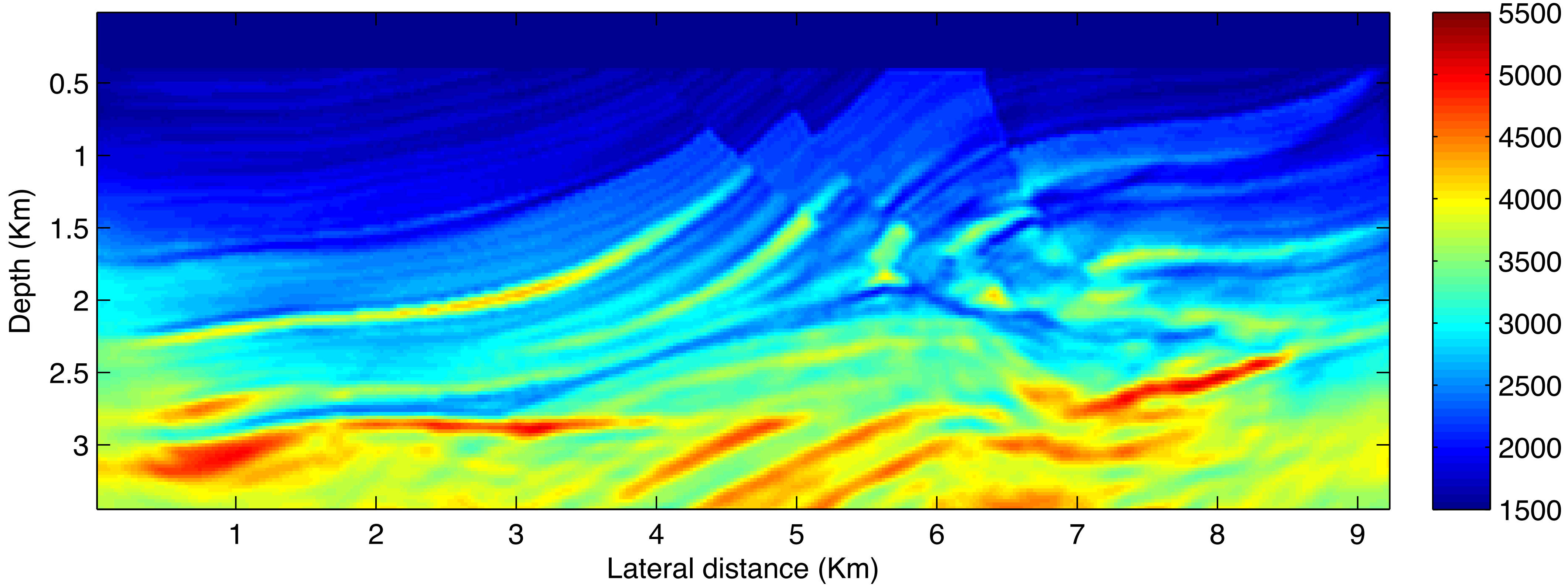
Depth (Km)
0.5
1
1.5
2
2.5
3
1
2
3
4
5
6
7
8
9
Lateral distance (Km)
5500
5000
4500
4000
3500
3000
2500
2000

# L1 with small batch

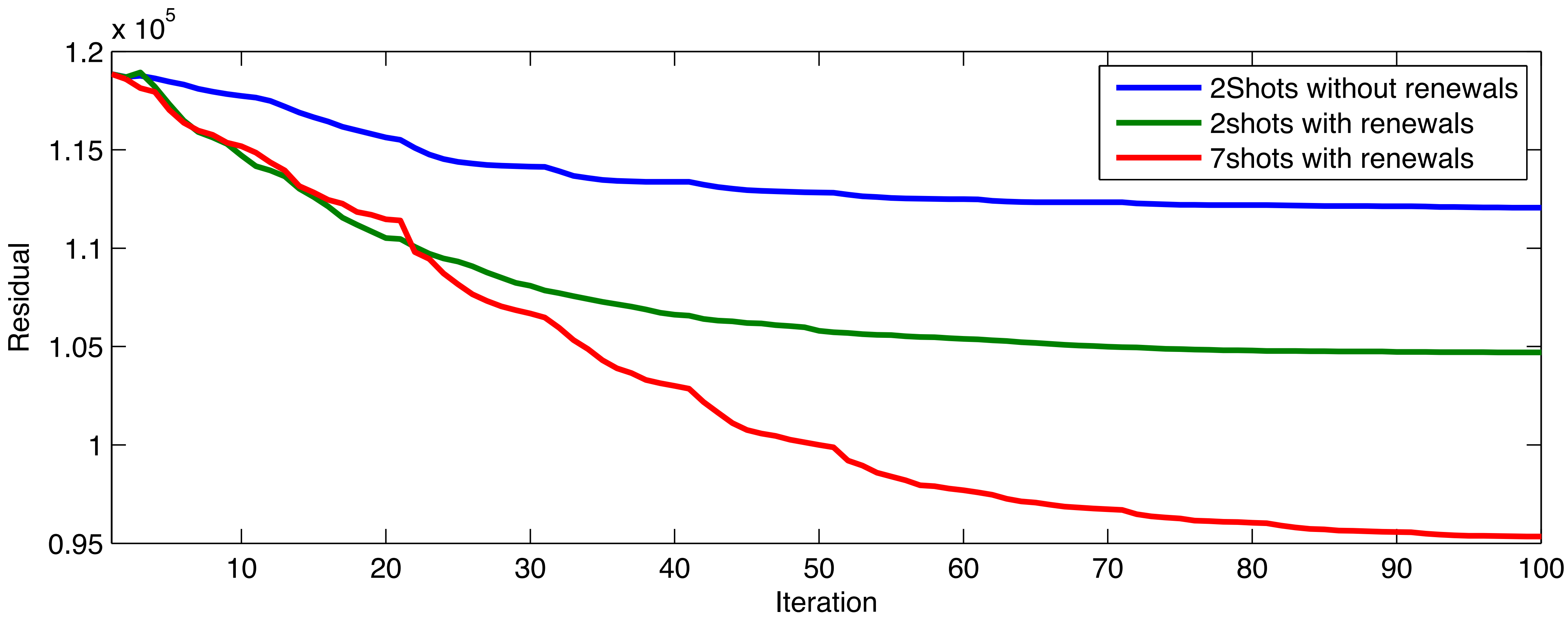



Speed up: X 3

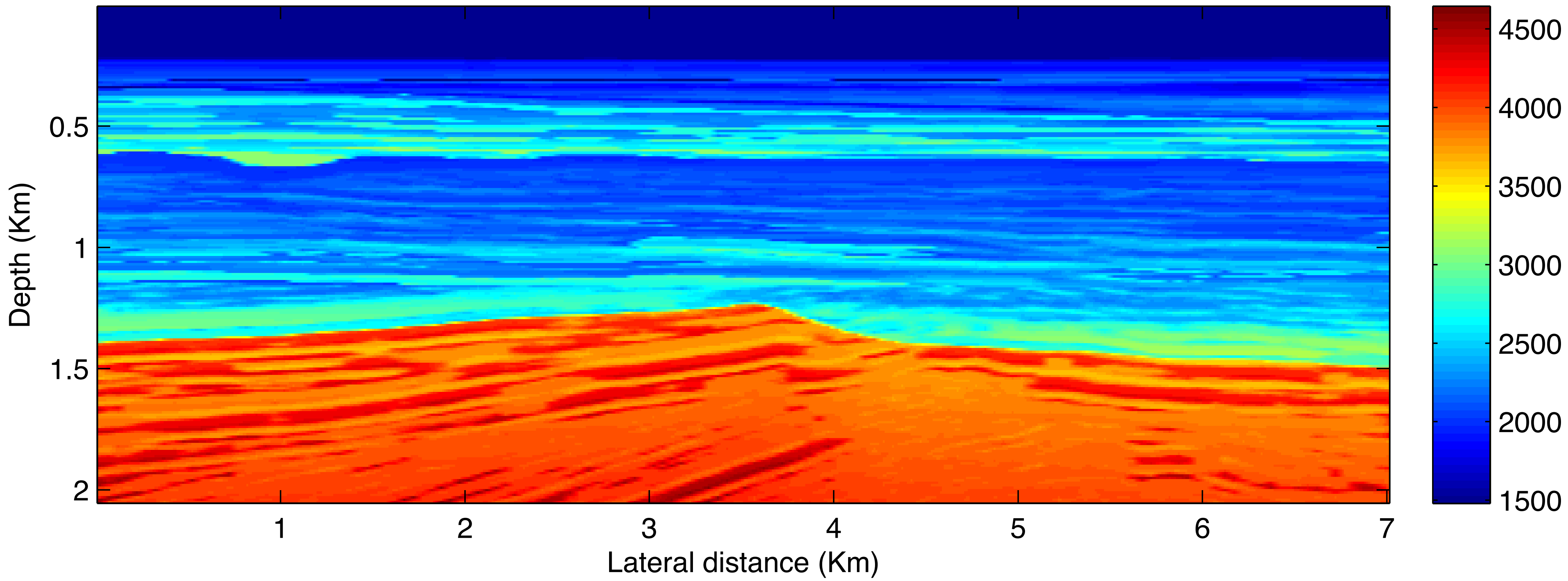
# L2 with *all* shots



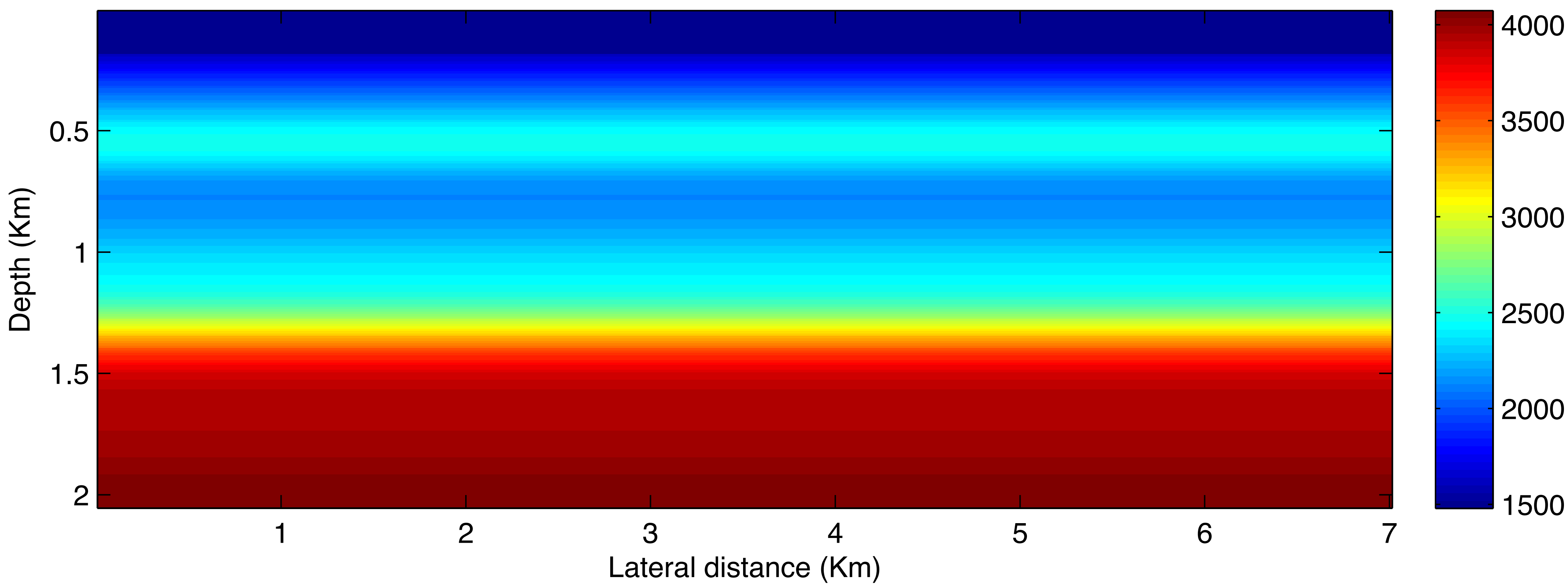
# Model residual



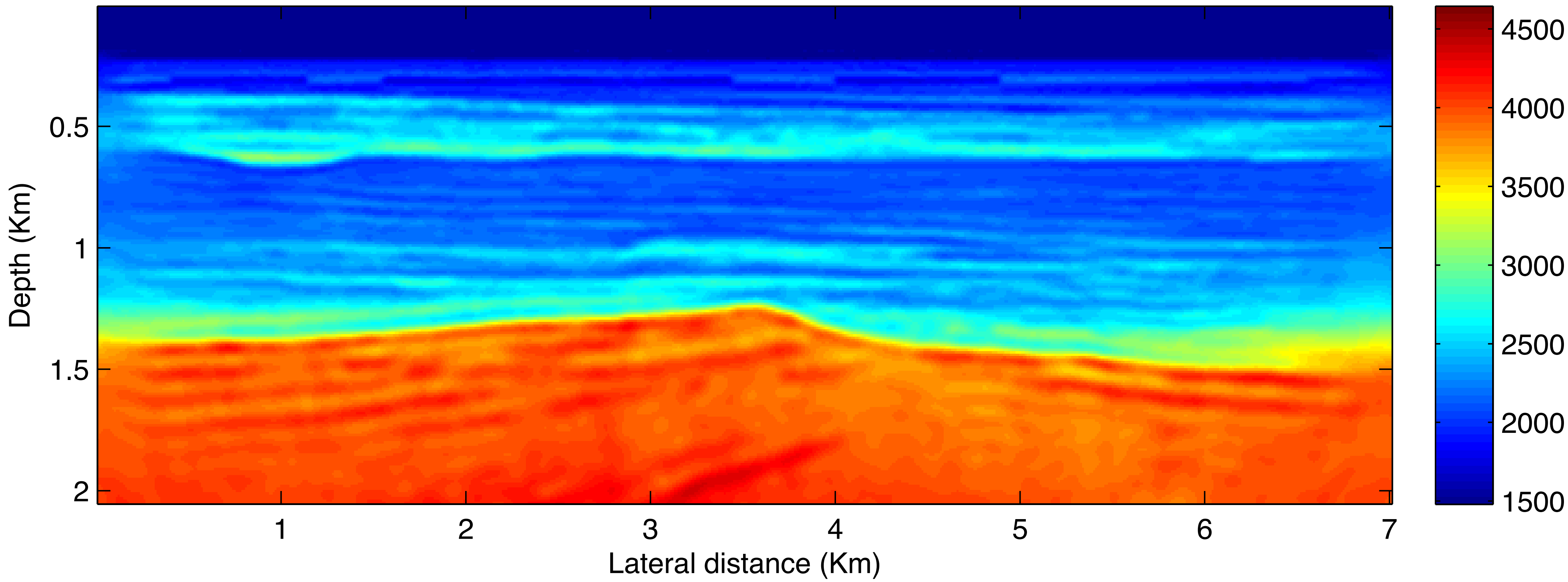
# True model


0.5
1
1.5
2
Depth (Km)
1
2
3
4
5
6
7
Lateral distance (Km)
4500
4000
3500
3000
2500
2000
1500

# Initial model

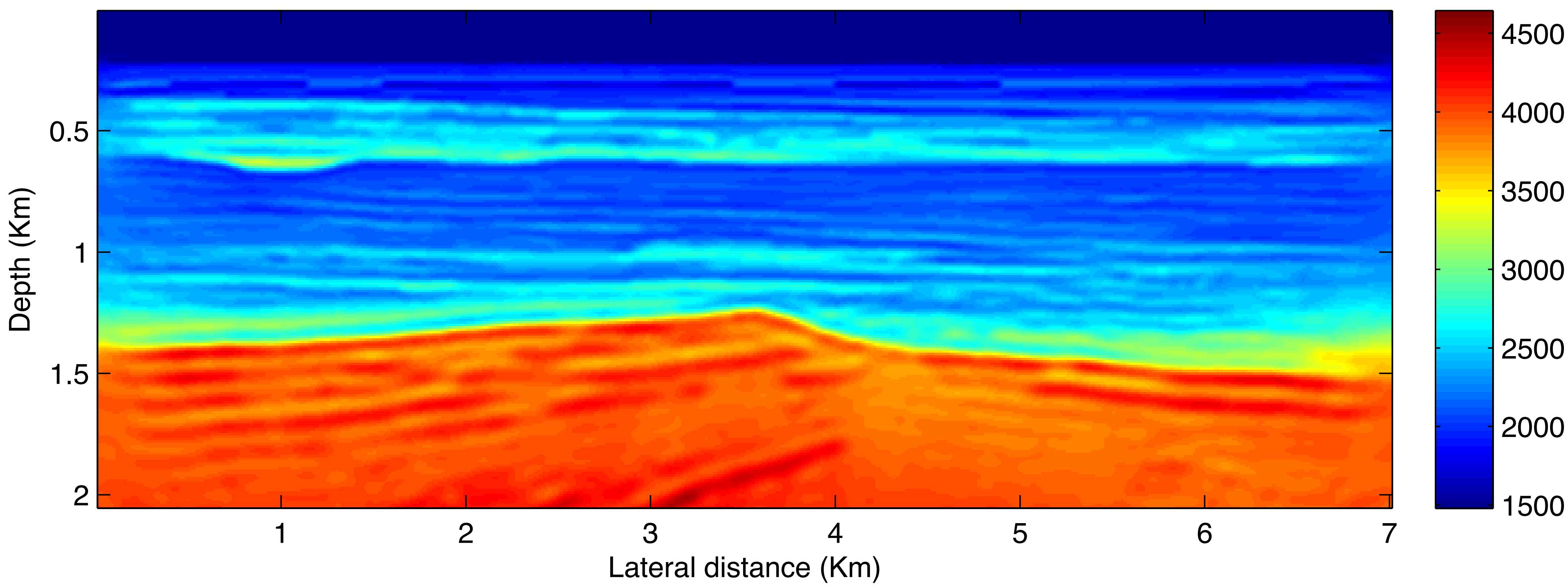


# Inverted result with 1 super shot



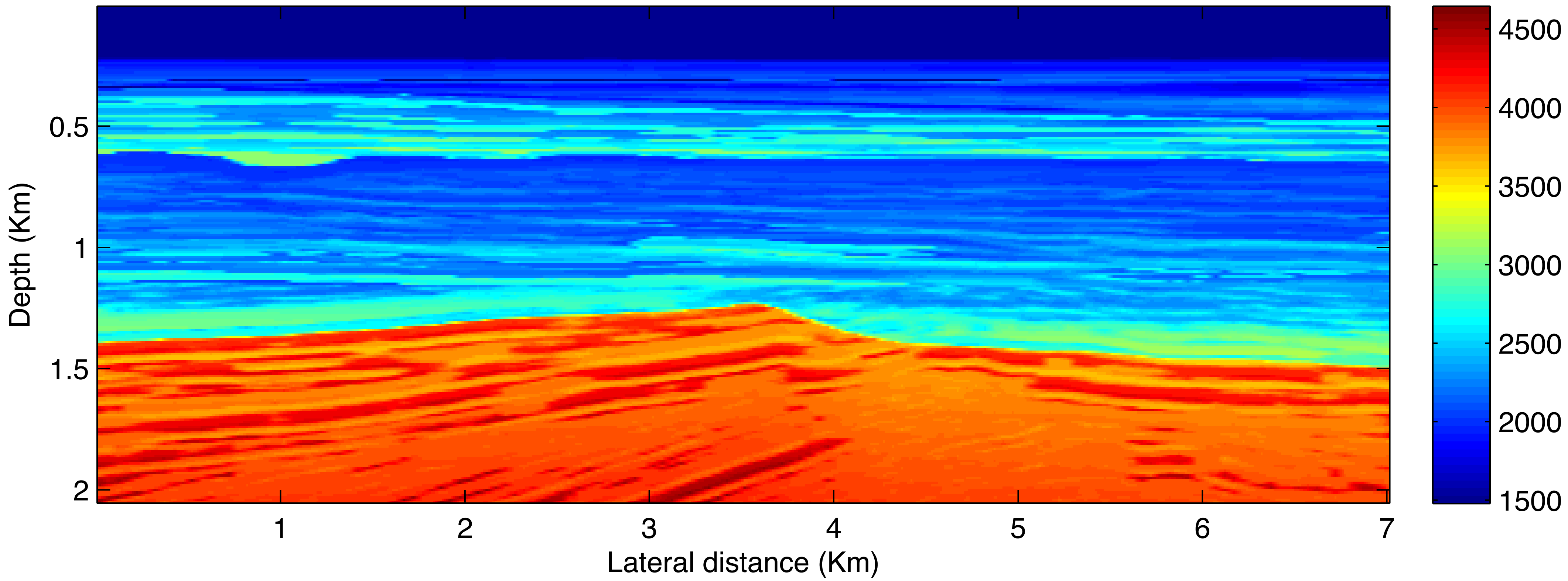

Start from 1.7 Hz

# Inverted result with 7 super shot




Start from 1.7 Hz

# True model


0.5
1
1.5
2
Depth (Km)
1
2
3
4
5
6
7
Lateral distance (Km)
4500
4000
3500
3000
2500
2000
1500

SLIM

# Observations

- Reduce the cost of FWI by using super shots
- Recover high resolution Gauss-Newton updates by imposing Curvelet domain sparsity

# Conclusions

- *sparse* recovery offsets *random interferences* due to *source encoding*
- *high-quality & high-resolution inversions* have been *achieved with significant accelerations*
- *no* need for additional *migration step*
- *improvements* come from *sparsity promotion & curvelets*
- indications that the *curse of dimensionality* can be removed...

SLIM

# Acknowledgements

- Authors of CurveLab.

SINBAD

NSERC CRSNG

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). 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.

# Thank you

slim.eos.ubc.ca

# Further reading

**Compressive sensing**

- *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information by Candes, 06*.
- *Compressed Sensing* by D. Donoho, '06
- Curvelets and Wave Atoms for Mirror-Extended Images by L. Demanet, L. Ying, 07.

**Simultaneous 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.
- *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

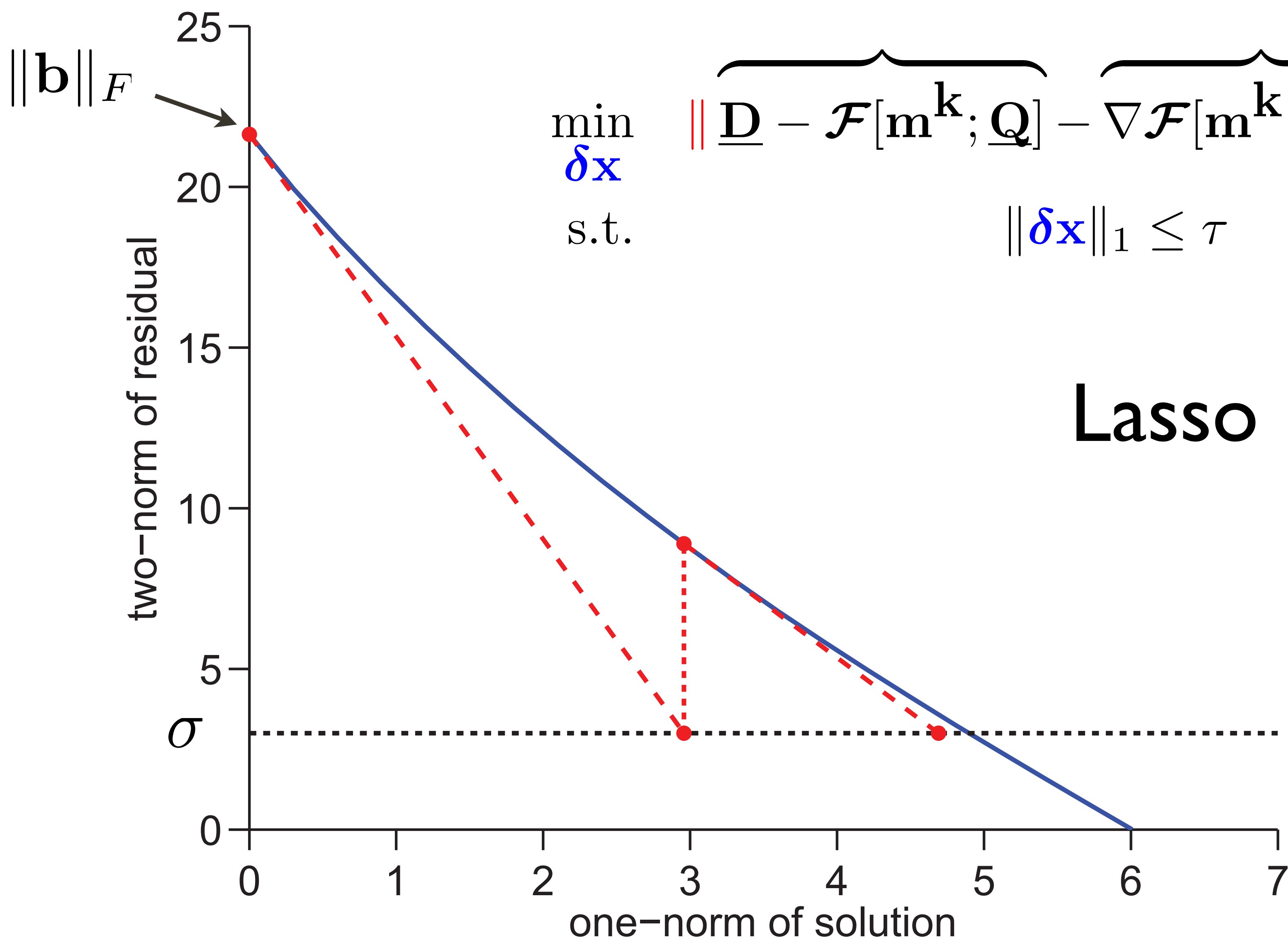
**Stochastic optimization and machine learning:**

- *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 and Recursive Algorithms and Applications* by Kushner and Lin
- *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

# Picking Lasso Parameter

$$\min_{\boldsymbol{\delta}\mathbf{x}} \; \| \overbrace{\underline{\mathbf{D}} - \mathcal{F}[\mathbf{m}^{\mathbf{k}}; \underline{\mathbf{Q}}]}^{\mathbf{b}} - \overbrace{\nabla\mathcal{F}[\mathbf{m}^{\mathbf{k}}; \underline{\mathbf{Q}}]\mathbf{S}^*}^{\mathbf{A}} \, \boldsymbol{\delta}\mathbf{x} \|_F^2$$

$$\text{s.t.} \quad \|\boldsymbol{\delta}\mathbf{x}\|_1 \leq \tau$$




Lasso