Accelerated large-scale inversion with *message* passing

Felix J. Herrmann

thanks to Xiang Li

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

Drivers

Recent technology push calls for collection

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

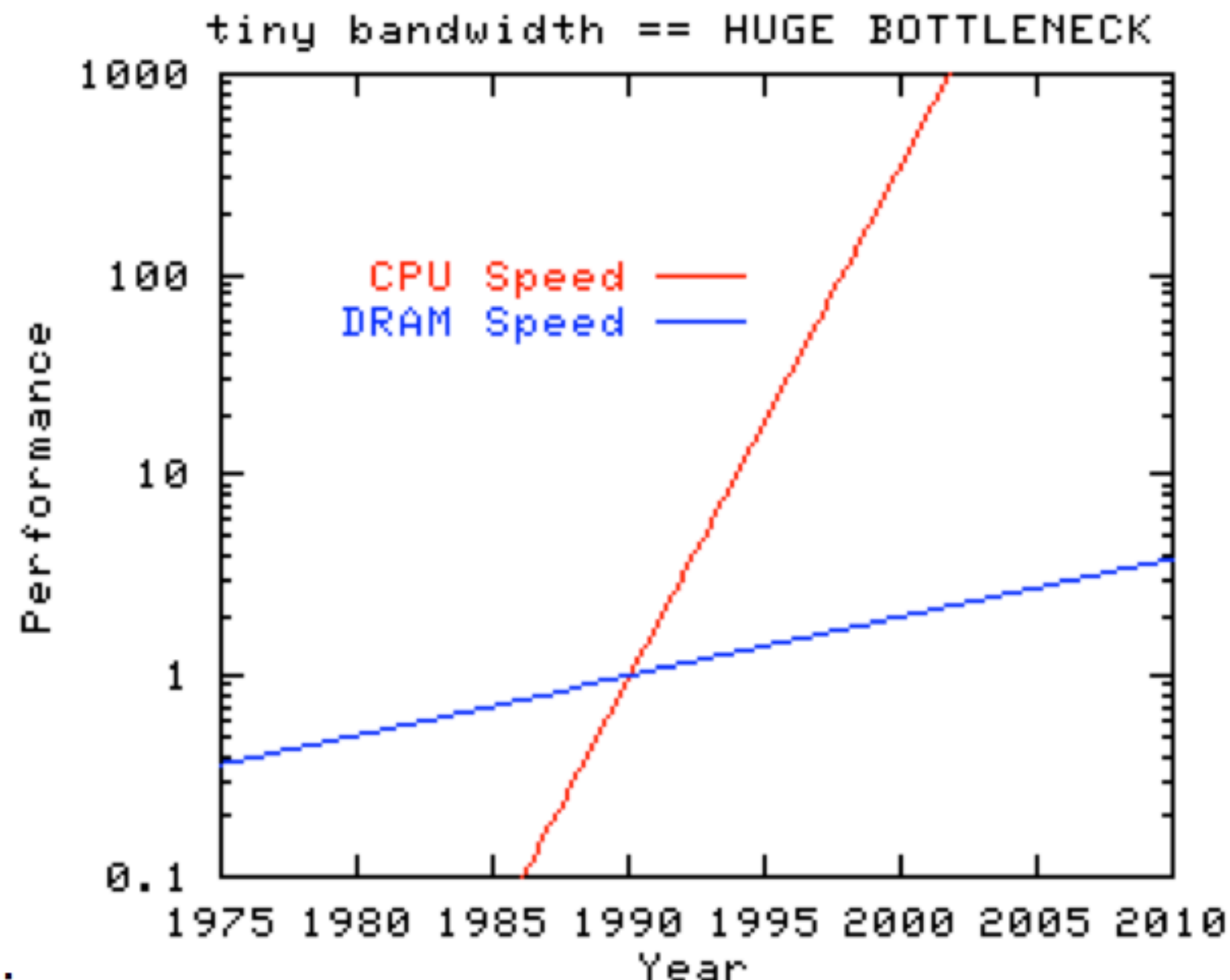
Exposes vulnerabilities in our *ability* to control

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

SLIM

Drivers cont'd

Problems exacerbated by IO bottleneck:



Goals

Replace a ‘*sluggish*’ inversion *paradigm* that

- relies on *touching* **all** *data* all the *time*

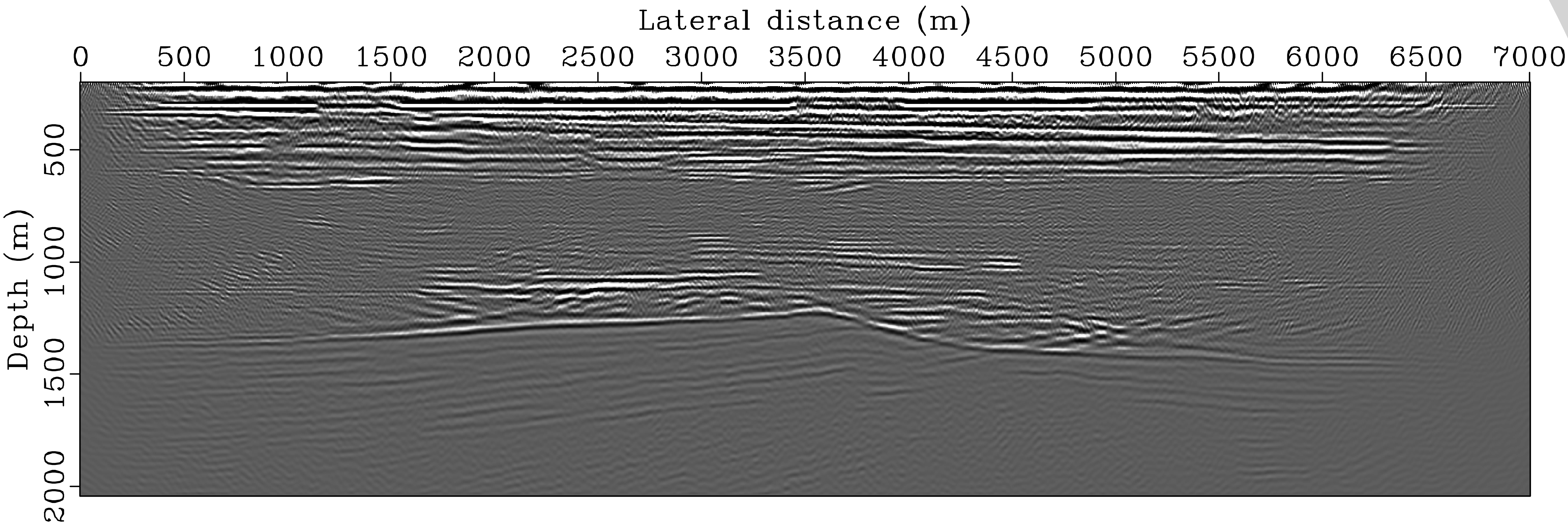
by an *agile* optimization *paradigm* that works on

- **small** *randomized* subsets of data *iteratively*

Confront “*data explosion*” by

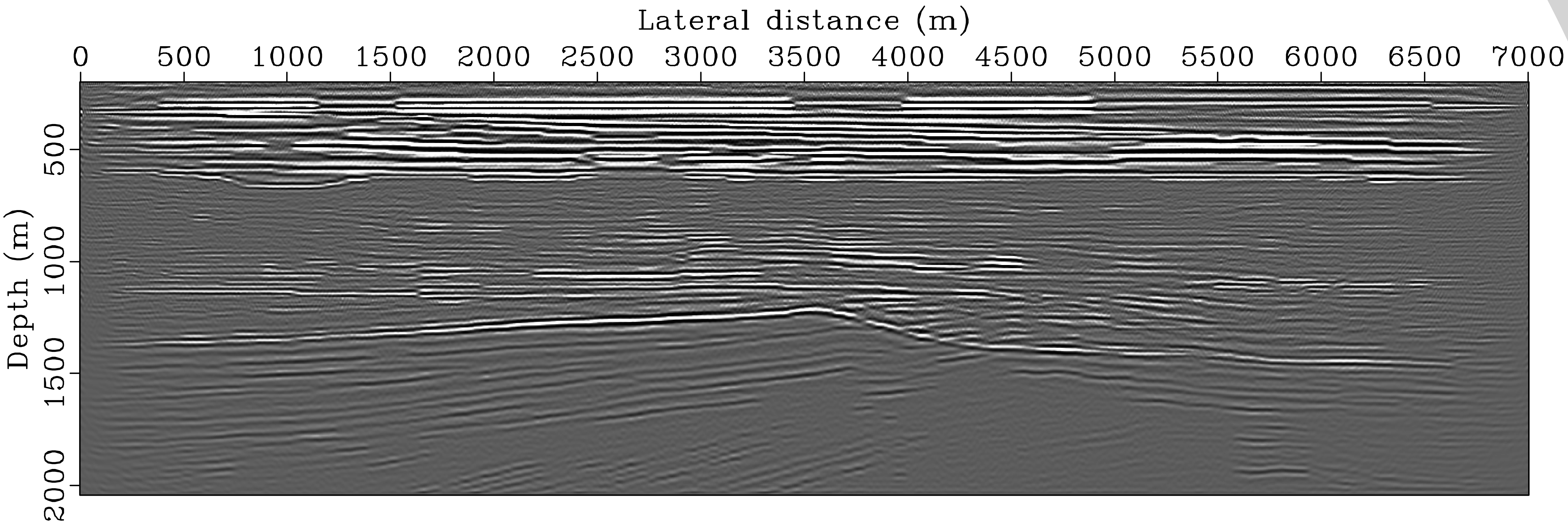
- *reducing* acquisition *costs*
- *removing* IO & PDEs-solve *bottlenecks*

Imaging results
[migration with "*all*" data]



Imaging results

[*linearized* inversion with *small subsets*]



Key technologies

Fast imaging with Stochastic optimization / Compressive Sensing:

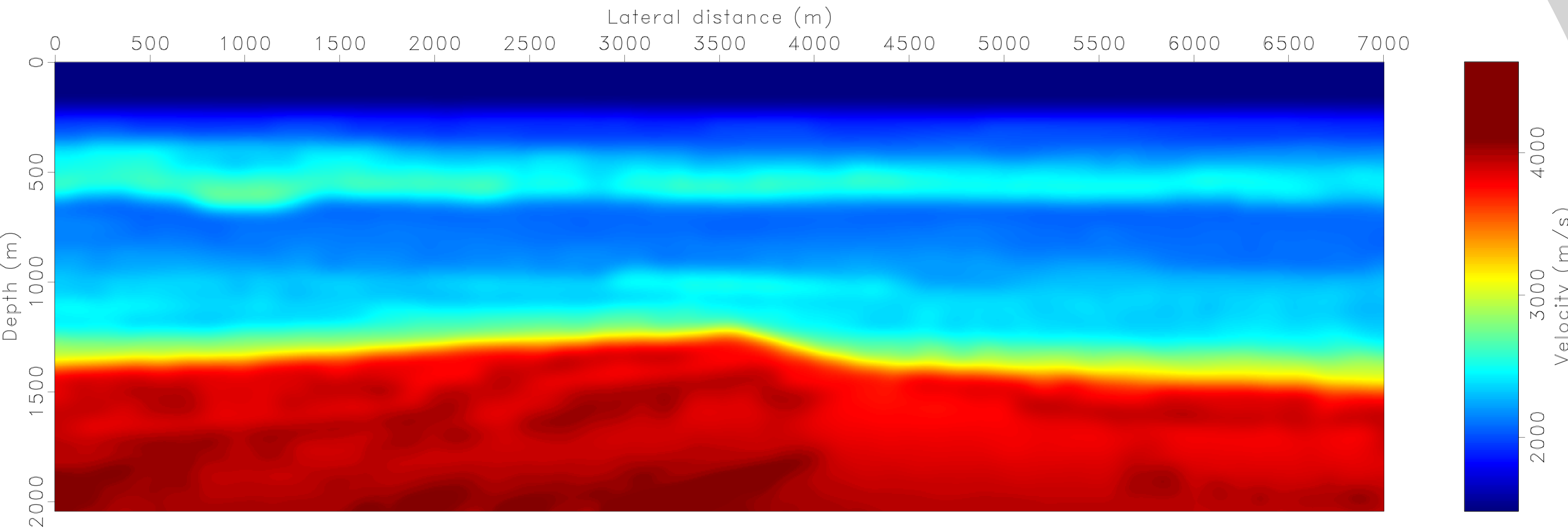
- *subsets* of *simultaneous* sources – *supershots* generated by *random* amplitude-weighted *superpositions*
- *random* subsets of *sequential* sources

Imaging via large-scale *curvelet*-domain *sparsity* promoting *convex* optimization with *cooling*

Acceleration with *approximate* message passing

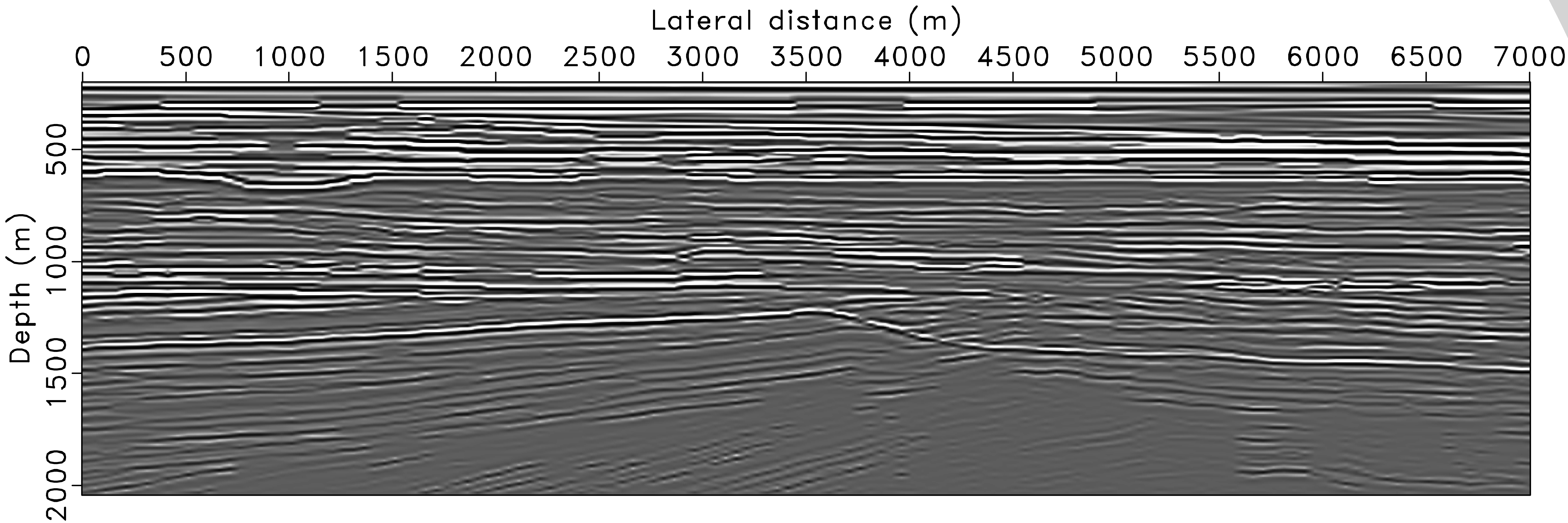
Imaging

[background model]



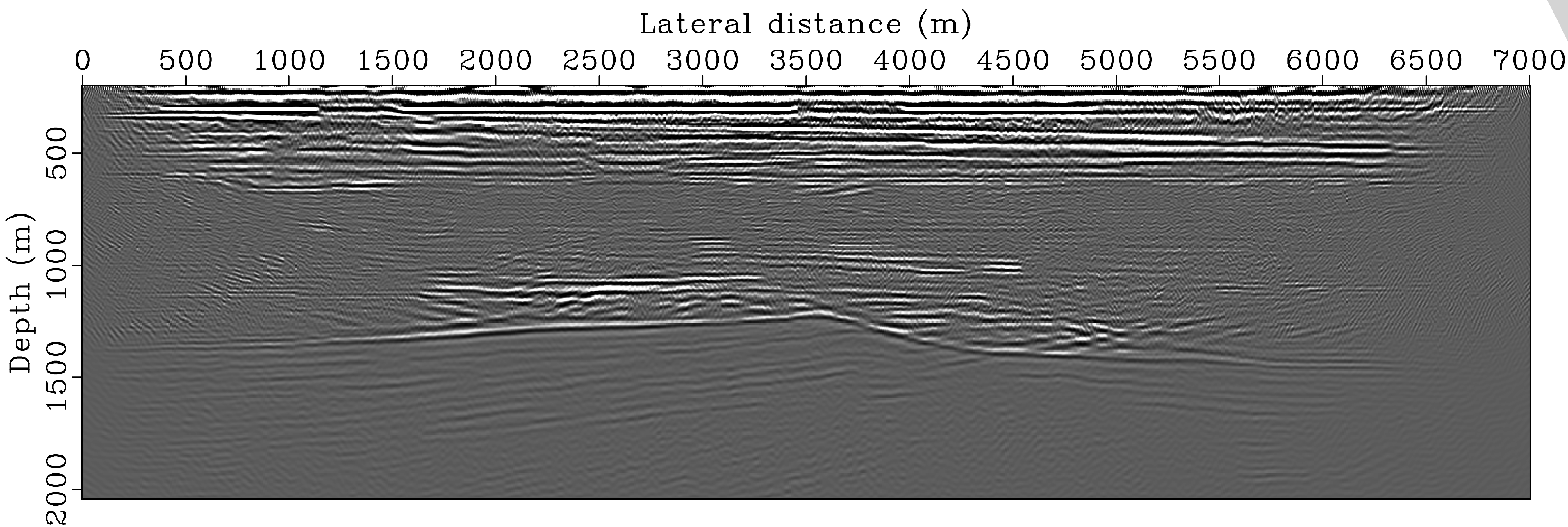
Imaging

[*true* perturbation]



Migration

[single migration with "*all*" data]



Too expensive to *invert* with "*all*" data...

Fast imaging

[*via* stochastic optimization]

*R*erandomized sampling

- *linear* speed up by *reducing* # PDE solves
- *increases* convergence but may *fail* to *converge*

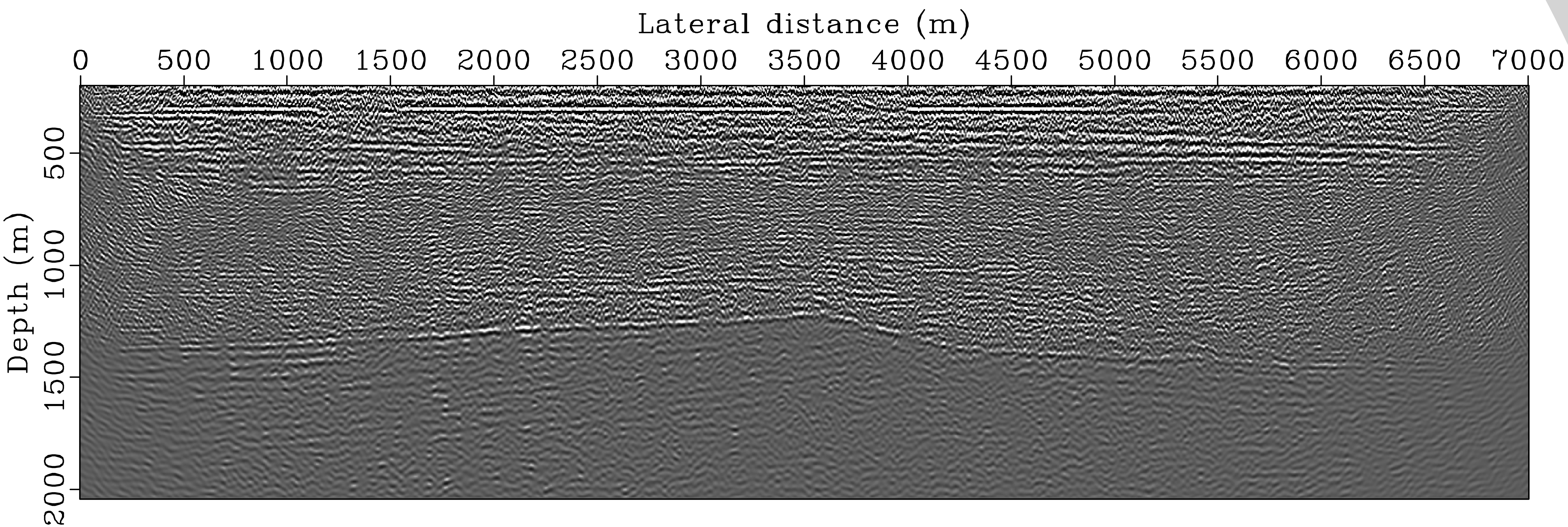
Exploits multi-experiment *redundancy* of seismic data volumes

- *regularly* draw *independent* subsets of shots
- cancels *crosstalk* by *rerandomization*

Heuristic of current phase-encoding migration/FWI methods

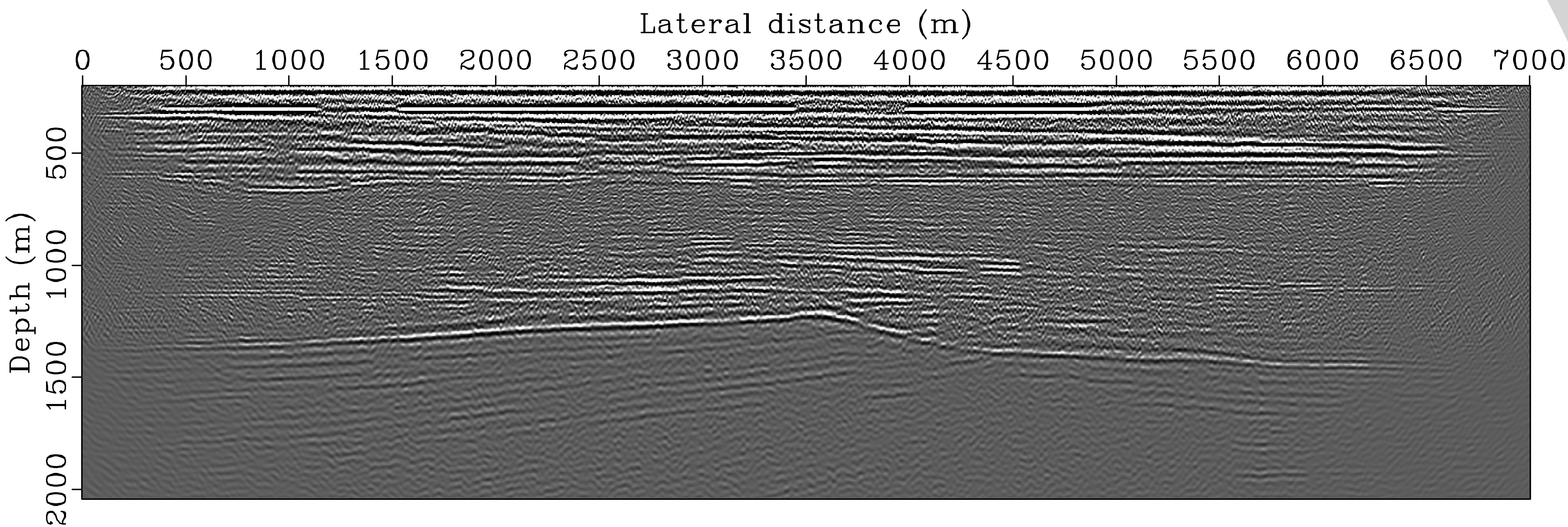
Linearized inversion

[ℓ_2 *without* rerandomization 3 *super* shots]



Linearized inversion

[ℓ_2 *with* rerandomization 3 *super* shots]



Fast imaging

[*via* compressive sensing]

Incoherent randomized sampling

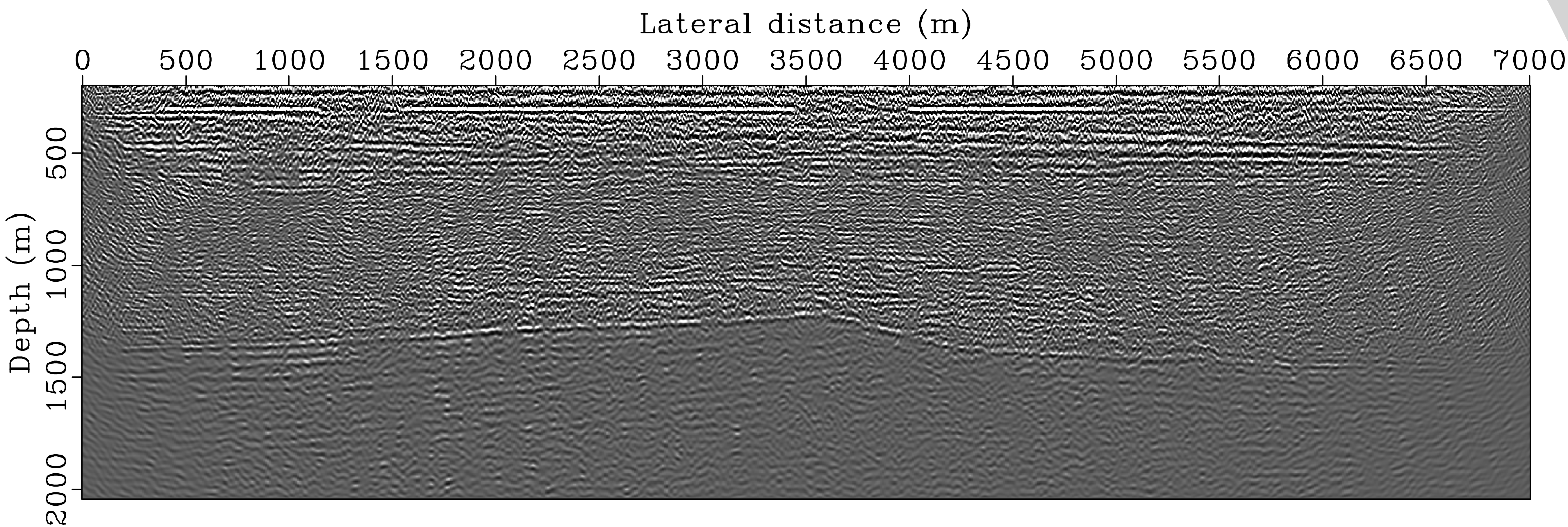
- *linear* speed up by *reducing* # PDE solves
- *coherent* source crosstalk turns into ***non-sparse*** *incoherent* noise

Exploits *structure* exhibited by migrated *images*

- leverages *curvelet*-domain ***sparsity*** promotion
- *maps* “noisy” crosstalk to *coherent* reflectors

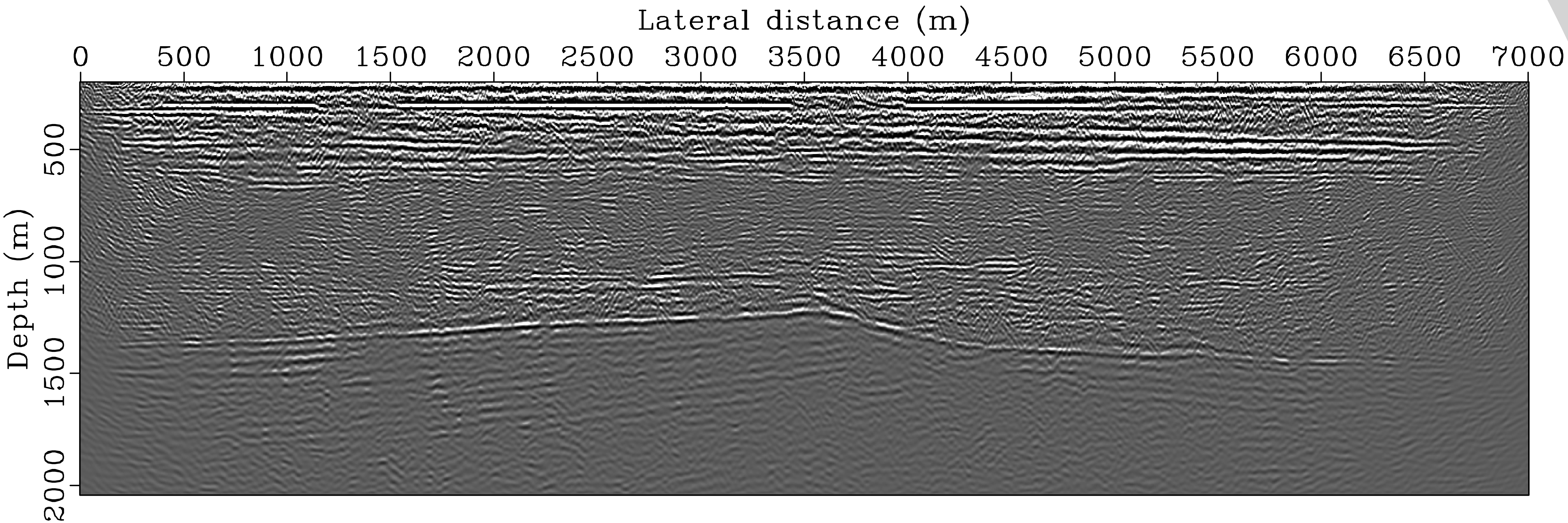
Linearized inversion

[ℓ_2 3 *super* shots]



Linearized inversion

[ℓ_1 3 *super* shots]



Observations
[reasonable PDE solve budget]

*R*erandomization and *curvelet*-domain *sparsity* promotion:

- *partly* eliminate "*noisy*" *crosstalk*
- *fail* to *remove* "small" incoherent crosstalk

Can we somehow combine these two methods?

- *continuation* method for large-scale *convex* optimization
- use *insights* from *approximate* message passing

[Daubechies et. al, ’04; Hennenfent et. al.,’08, Mallat, ’09, Donoho et. al, ’09]

[Montanari, ’12]

Convex optimization

Involves *iterations* of the type

soft threshold

$$\mathbf{x}^{t+1} = \eta_t \left(\mathbf{A}^* \mathbf{r}^t + \mathbf{x}^t \right)$$

$$\mathbf{r}^t = \mathbf{b} - \mathbf{A}\mathbf{x}^t$$

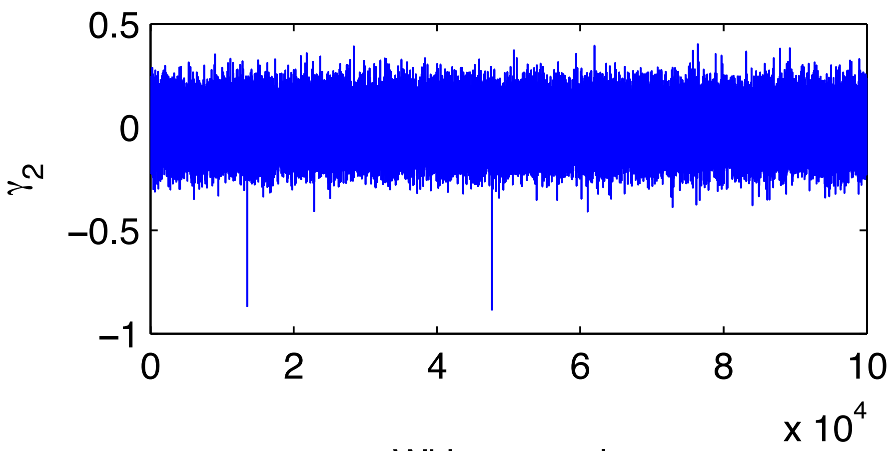
Corresponds to *vanilla* denoising if $\mathbf{A}$ is a Gaussian matrix.

But does the *same* hold for later (t>1) *iterations*...?

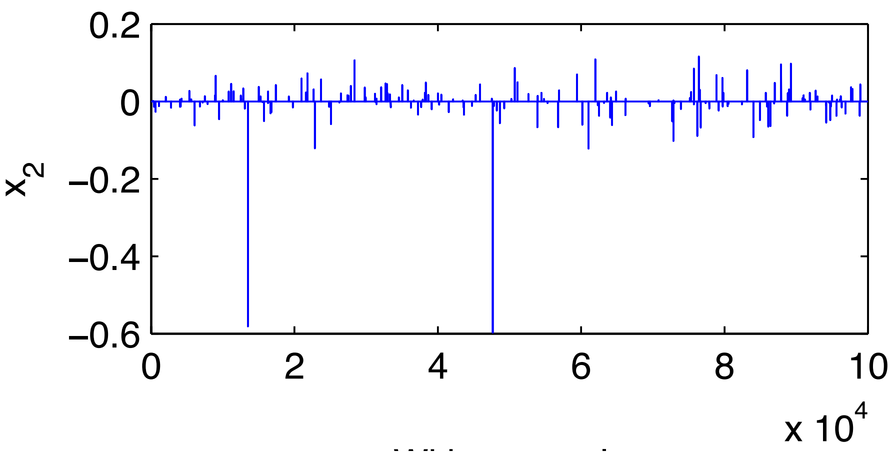
SLIM

Iteration t=1

$$\mathbf{A}^*\mathbf{r}^t + \mathbf{x}^t$$

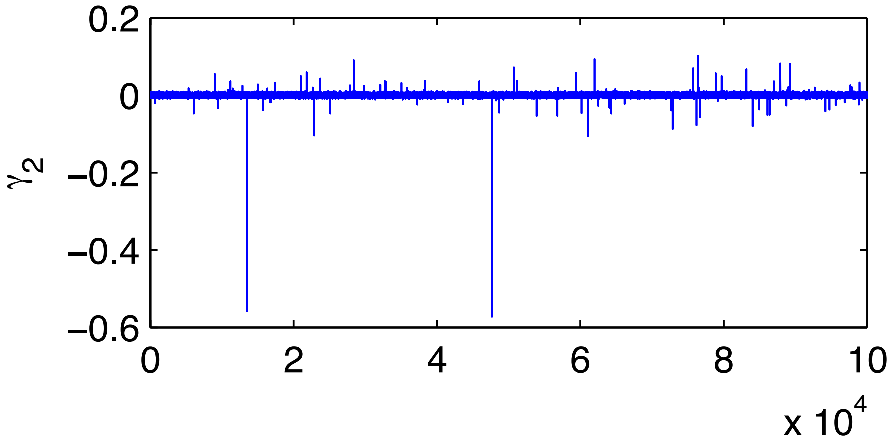


$$\eta_t(\mathbf{A}^*\mathbf{r}^t + \mathbf{x}^t)$$

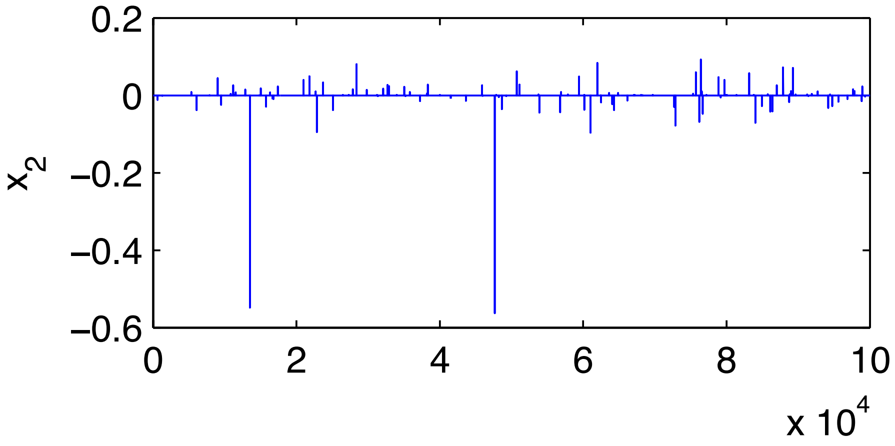


Iteration t=2

$$\mathbf{A}^*\mathbf{r}^t + \mathbf{x}^t$$

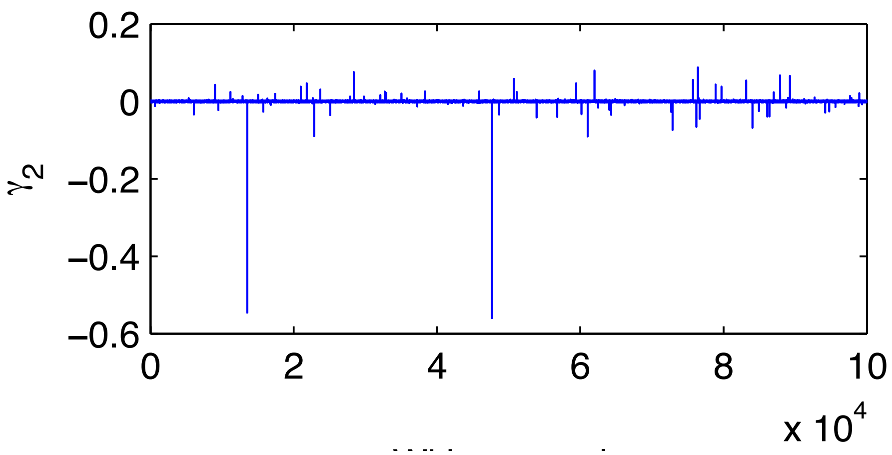


$$\eta_t(\mathbf{A}^*\mathbf{r}^t + \mathbf{x}^t)$$

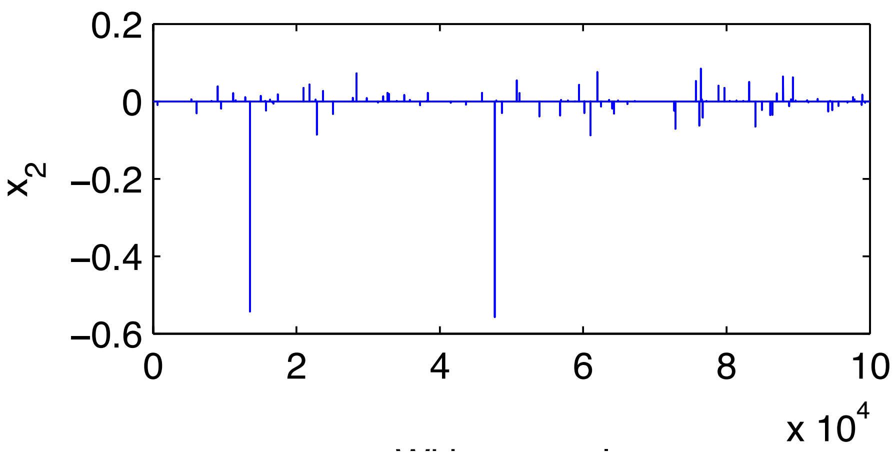


Iteration t=3

$$\mathbf{A}^*\mathbf{r}^t + \mathbf{x}^t$$

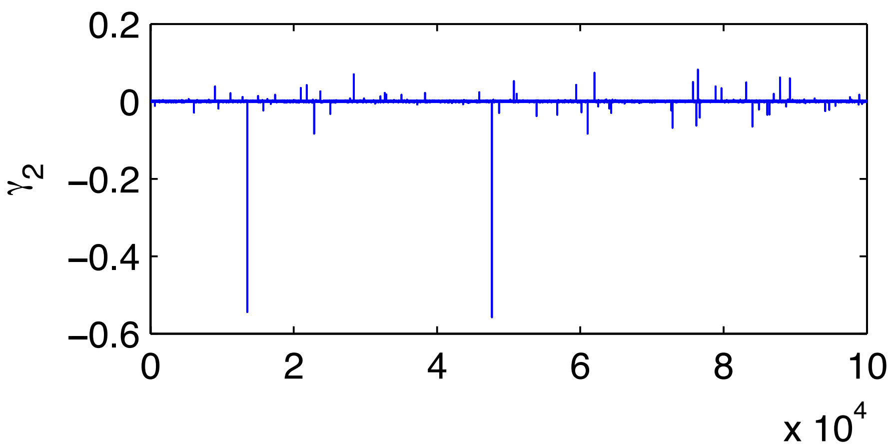


$$\eta_t(\mathbf{A}^*\mathbf{r}^t + \mathbf{x}^t)$$

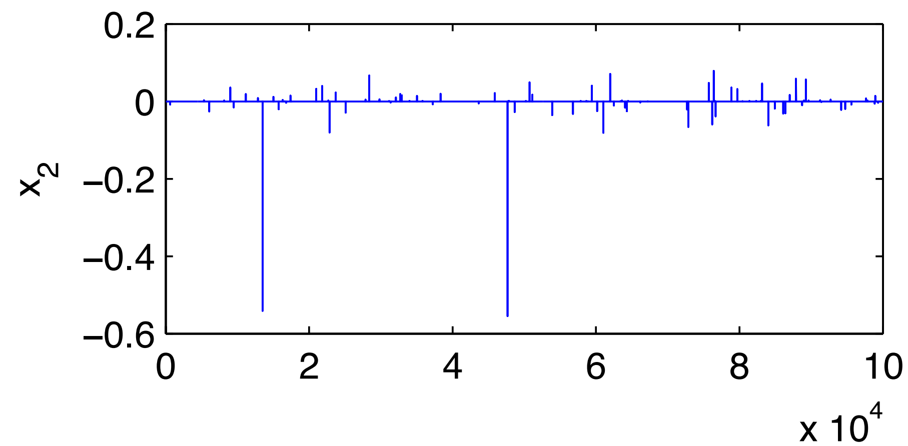


Iteration t=4

$$\mathbf{A}^*\mathbf{r}^t + \mathbf{x}^t$$



$$\eta_t(\mathbf{A}^*\mathbf{r}^t + \mathbf{x}^t)$$



Problem

After *first* iteration the *interferences* become 'spiky' because of *correlations* between model *iterate* $\mathbf{x}^t$ & the *matrix* $\mathbf{A}$

- *assumption* spiky vs Gaussian noise *no* longer *holds*
- renders soft *thresholding* less *effective*

Leads to *stalling* of sparsity-promoting *algorithms*...

[Donoho et. al, '09-'12; Montanari, '10-'12, Rangan, '11]

SLIM

Approximate message passing

Add a *term* to *iterative* soft *thresholding*, i.e.,

$$\mathbf{x}^{t+1} = \eta_t \left(\mathbf{A}^* \mathbf{r}^t + \mathbf{x}^t \right)$$

$$\mathbf{r}^t = \mathbf{b} - \mathbf{A}\mathbf{x}^t + \frac{\|\mathbf{x}^{t+1}\|_0}{n} \mathbf{r}^{t-1}$$

"message term"

Holds for

- *normalized* Gaussian matrices $\mathbf{A}_{ij} \in n^{-1/2} N(0,1)$
- large-scale limit and for specific thresholding *strategy*

[Montanari, '12]

SLIM

Approximate message passing

Statistically equivalent to

$$\mathbf{x}^{t+1} = \eta_t \left(\mathbf{A}_t^* \mathbf{r}^t + \mathbf{x}^t \right)$$
$$\mathbf{r}^t = \mathbf{b}_t - \mathbf{A}_t \mathbf{x}^t$$

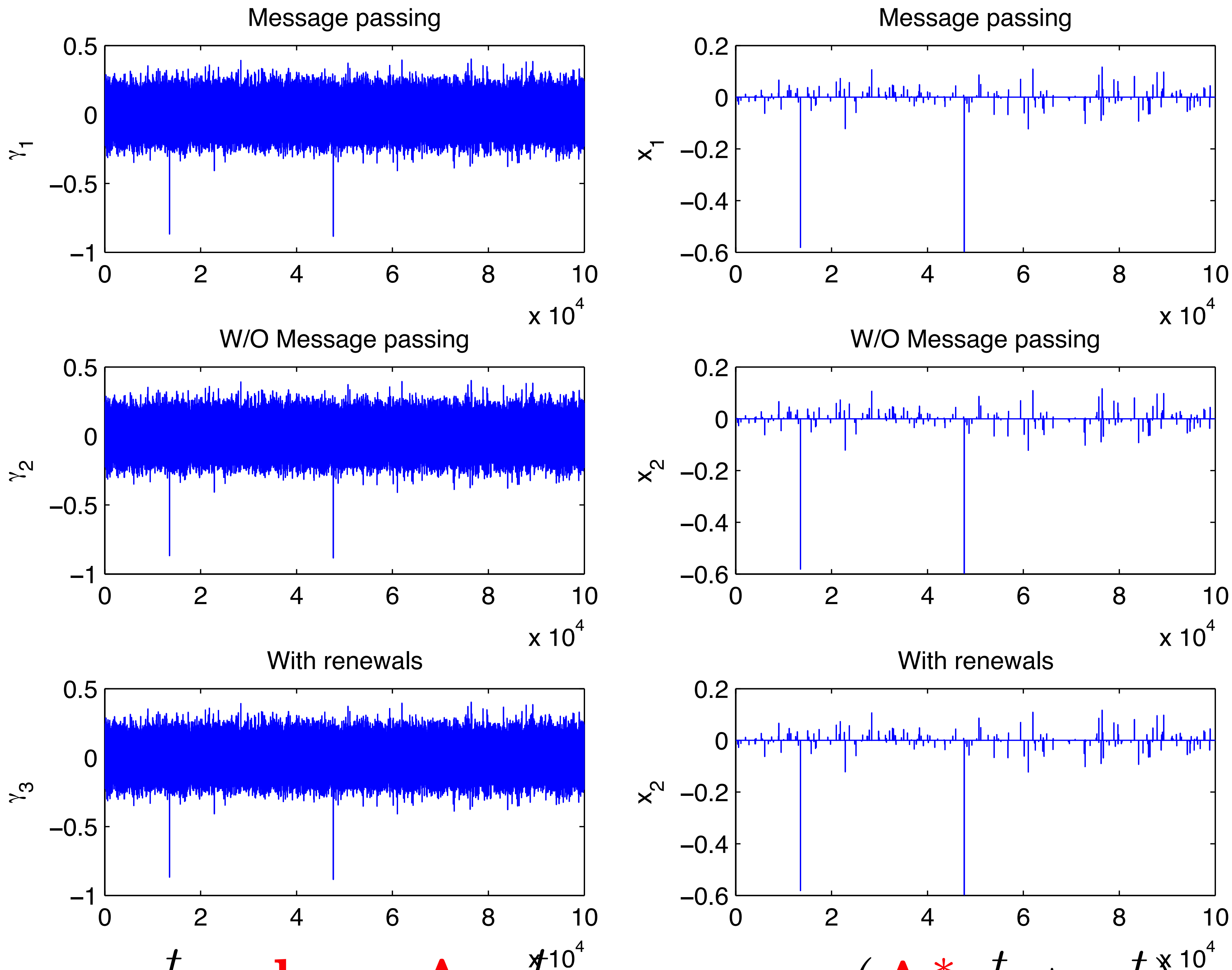
by drawing *new independent* pairs $\{\mathbf{b}_t, \mathbf{A}_t\}$ for *each* iteration

Changes the story *completely*

- breaks *correlation* buildup
- *faster* convergence

Iteration t=1

$$\mathbf{r}^t = \mathbf{b} - \mathbf{A}\mathbf{x}^t + \frac{\|\mathbf{x}^{t+1}\|_0}{\;}\mathbf{r}^{t-1} \qquad \eta_t(\mathbf{A}^*\mathbf{r}^t + \mathbf{x}^t)$$

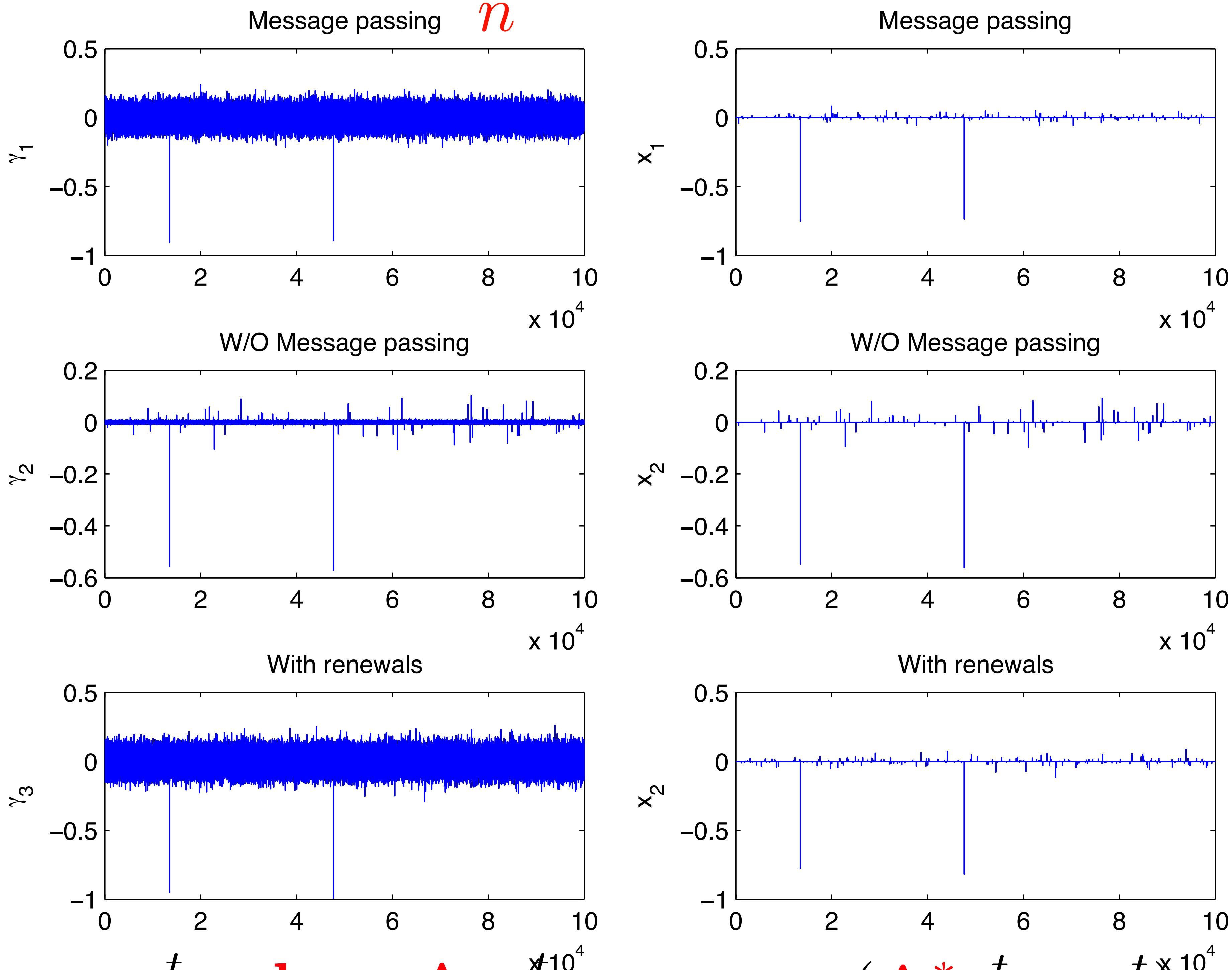


$$\mathbf{r}^t = \mathbf{b}_t - \mathbf{A}_t\mathbf{x}^t \qquad \eta_t(\mathbf{A}_t^*\mathbf{r}^t + \mathbf{x}^t)$$

SLIM

Iteration t=2

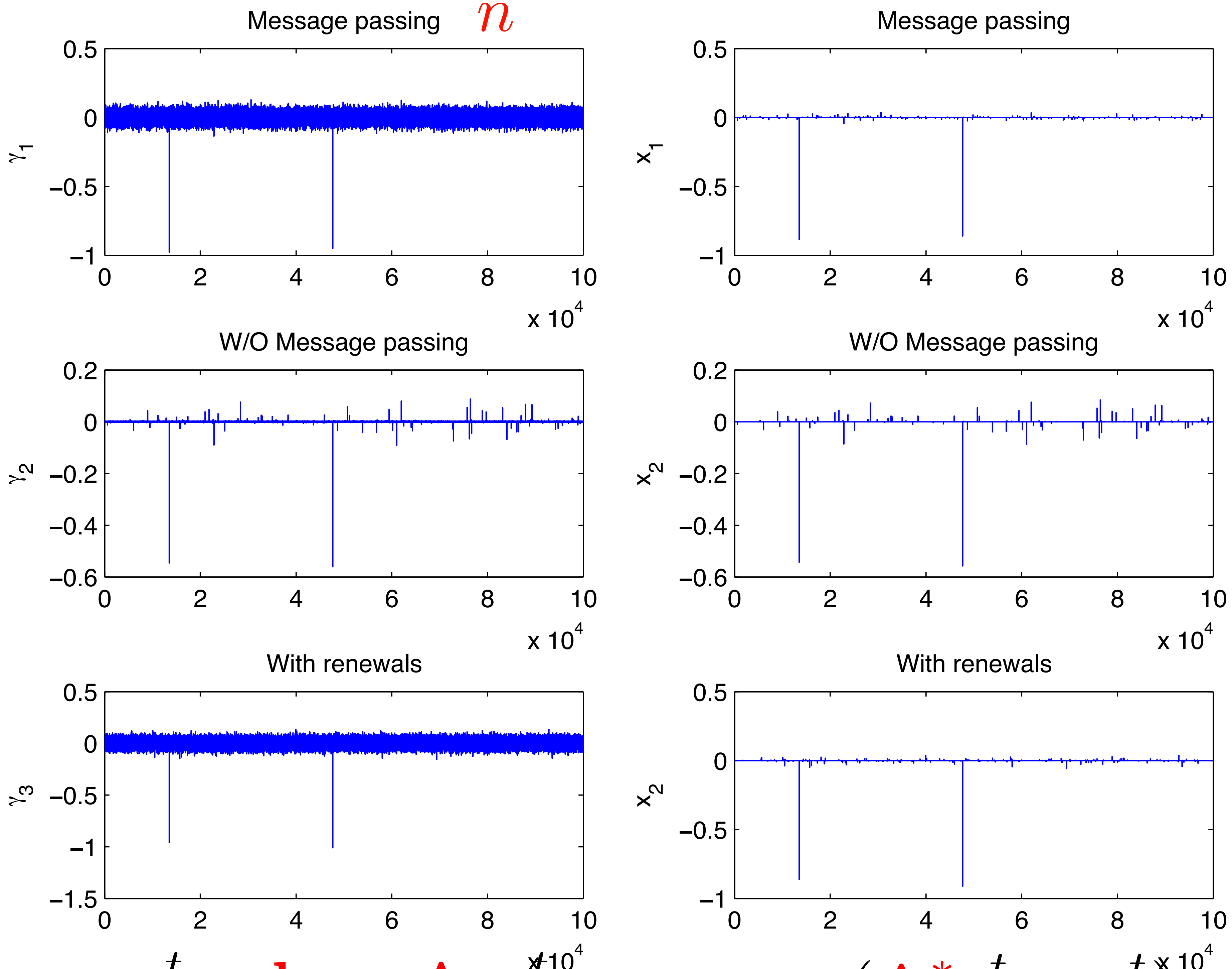
$$\mathbf{r}^t = \mathbf{b} - \mathbf{A}\mathbf{x}^t + \frac{\|\mathbf{x}^{t+1}\|_0}{n}\mathbf{r}^{t-1} \qquad \eta_t(\mathbf{A}^*\mathbf{r}^t + \mathbf{x}^t)$$



$$\mathbf{r}^t = \mathbf{b}_t - \mathbf{A}_t\mathbf{x}^t \qquad \eta_t(\mathbf{A}_t^*\mathbf{r}^t + \mathbf{x}^t)$$

Iteration t=3

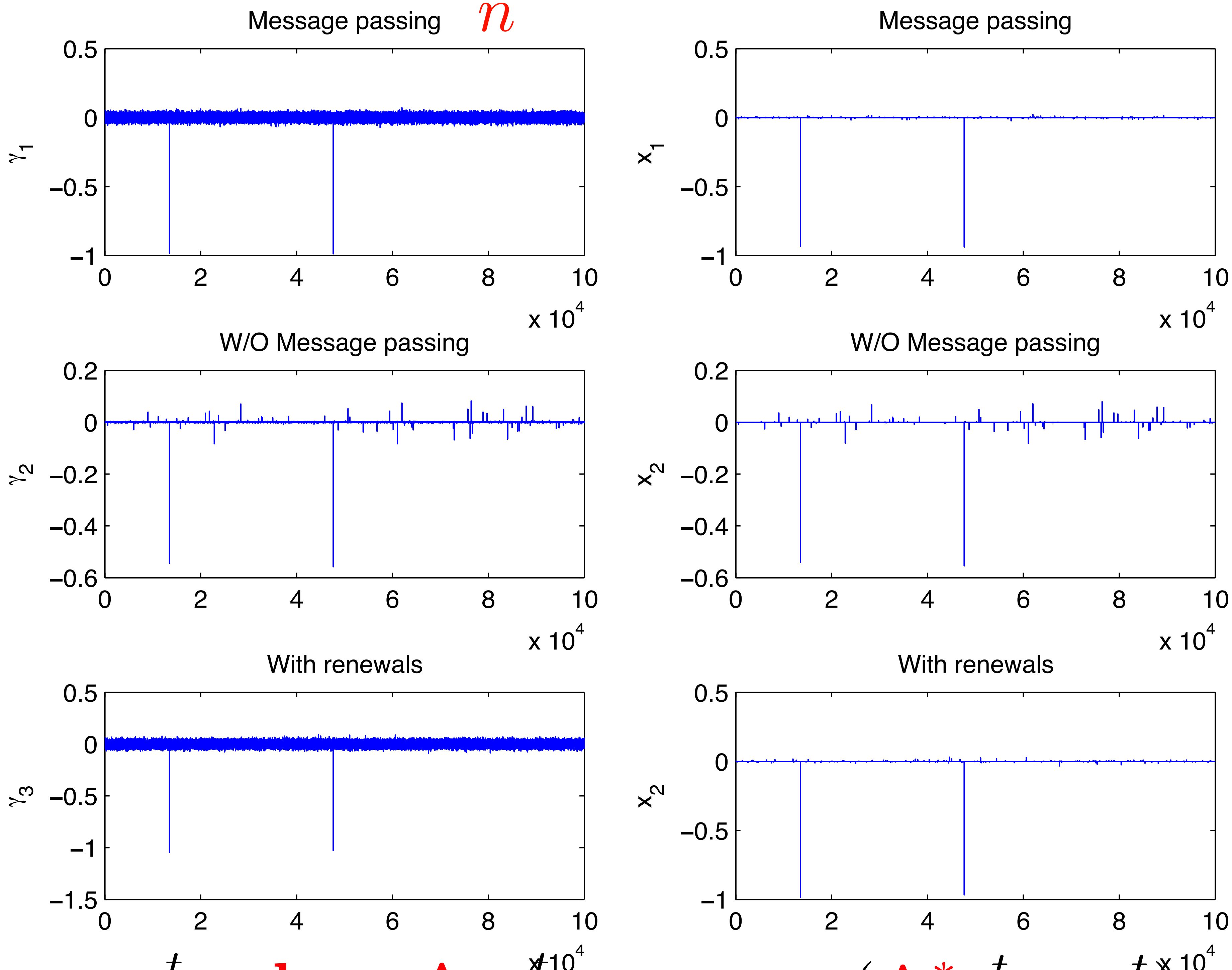
$$\mathbf{r}^t = \mathbf{b} - \mathbf{A}\mathbf{x}^t + \frac{\|\mathbf{x}^{t+1}\|_0}{n}\mathbf{r}^{t-1} \qquad \eta_t(\mathbf{A}^*\mathbf{r}^t + \mathbf{x}^t)$$



$$\mathbf{r}^t = \mathbf{b}_t - \mathbf{A}_t\mathbf{x}^t \qquad \eta_t(\mathbf{A}_t^*\mathbf{r}^t + \mathbf{x}^t)$$

Iteration t=4

$$\mathbf{r}^t = \mathbf{b} - \mathbf{A}\mathbf{x}^t + \frac{\|\mathbf{x}^{t+1}\|_0}{n}\mathbf{r}^{t-1} \qquad \eta_t(\mathbf{A}^*\mathbf{r}^t + \mathbf{x}^t)$$



$$\mathbf{r}^t = \mathbf{b}_t - \mathbf{A}_t\mathbf{x}^t \qquad \eta_t(\mathbf{A}_t^*\mathbf{r}^t + \mathbf{x}^t)$$

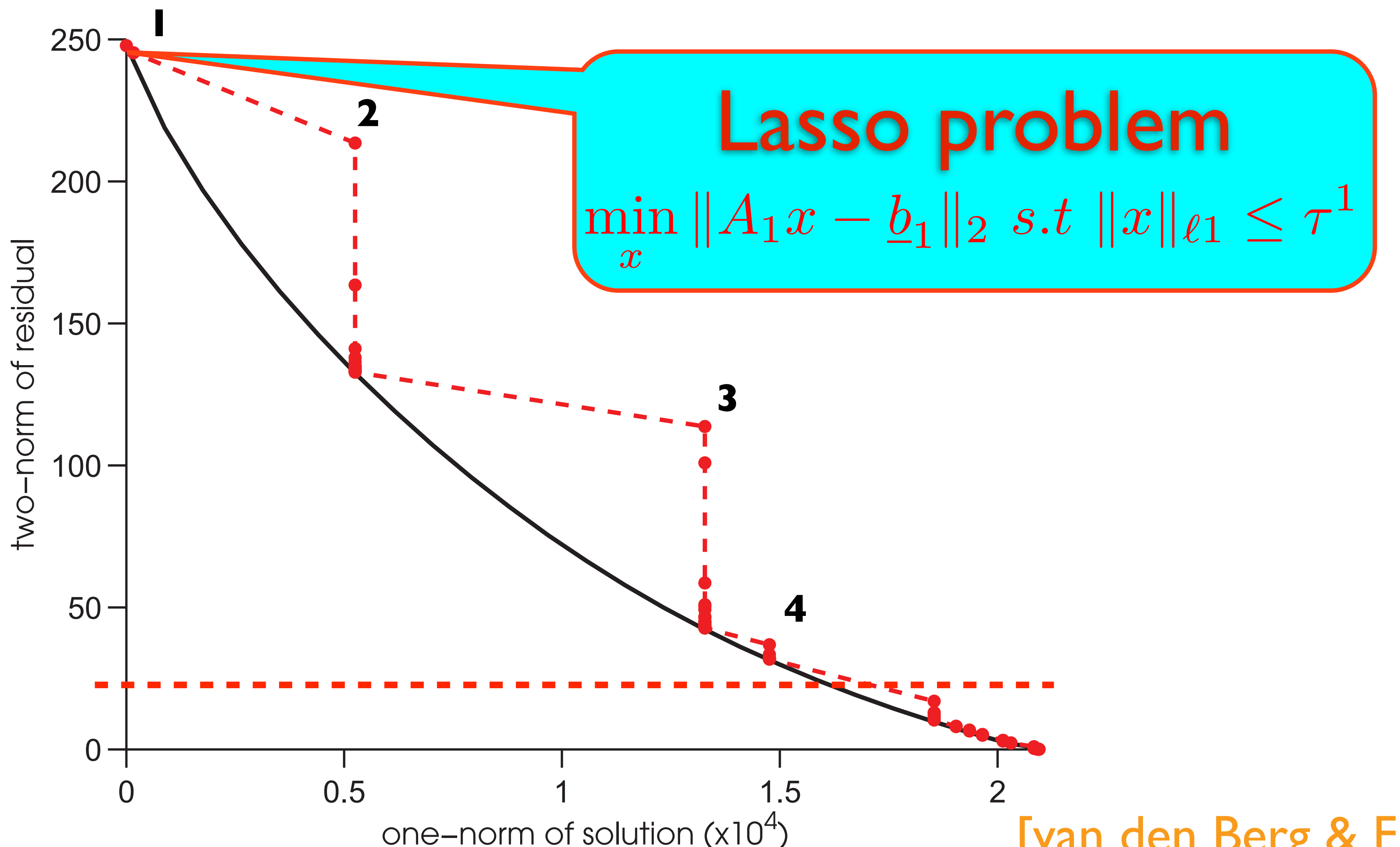
Supercooling

Break *correlations* between the model *iterate* and matrix **A** by *rerandomization*

- draw new *independent* $\{\mathbf{b}_t, \mathbf{A}_t\}$ after each subproblem is solved
- brings in "*extra*" information *without* growing the *system*
- ***minimal*** *extra* computational & memory cost

Supercooled

spectral-projected gradients



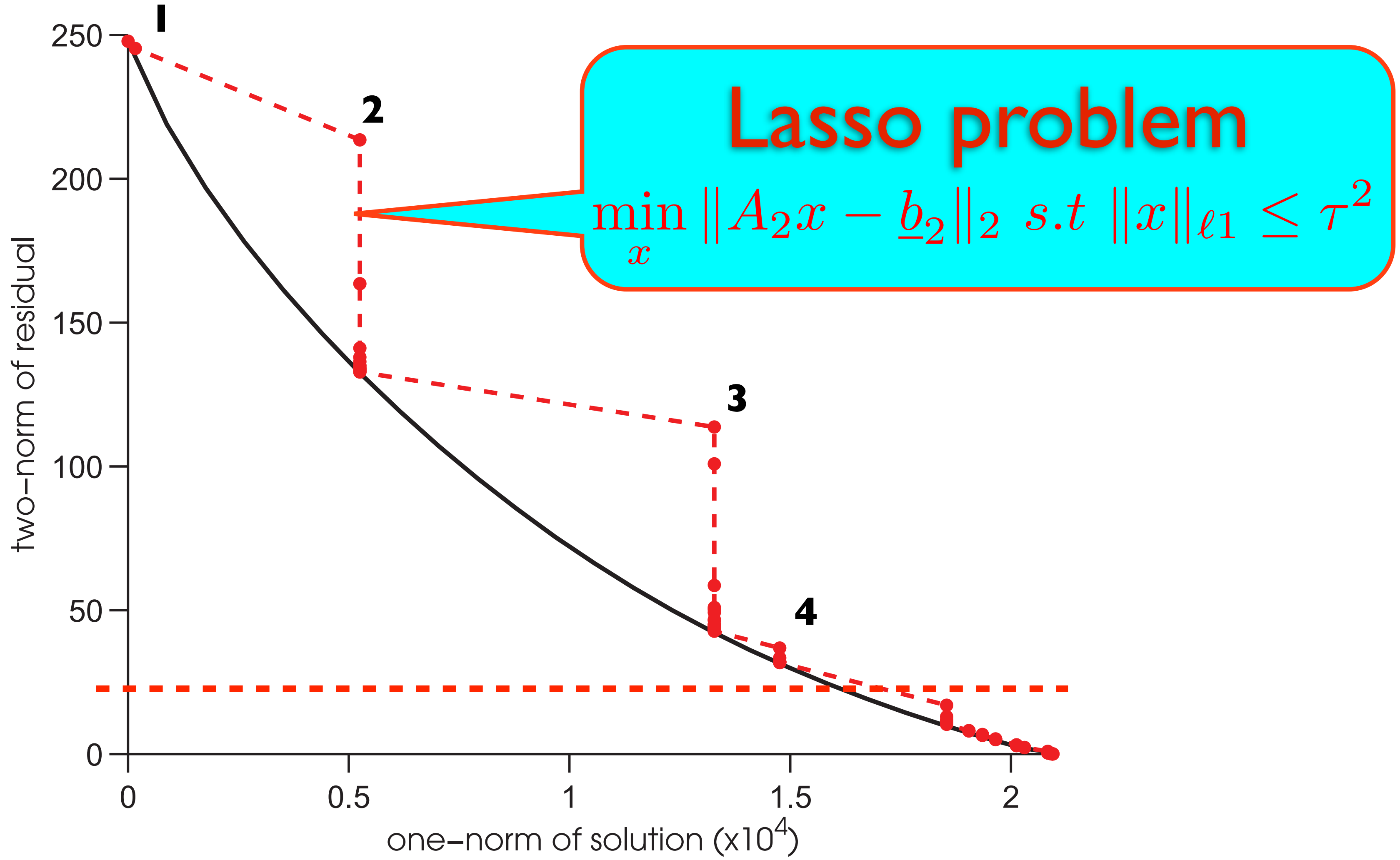
[van den Berg & Friedlander, ’08]

[Hennefent et. al., ’08]

[Lin & FJH, ’09-]

Supercooled

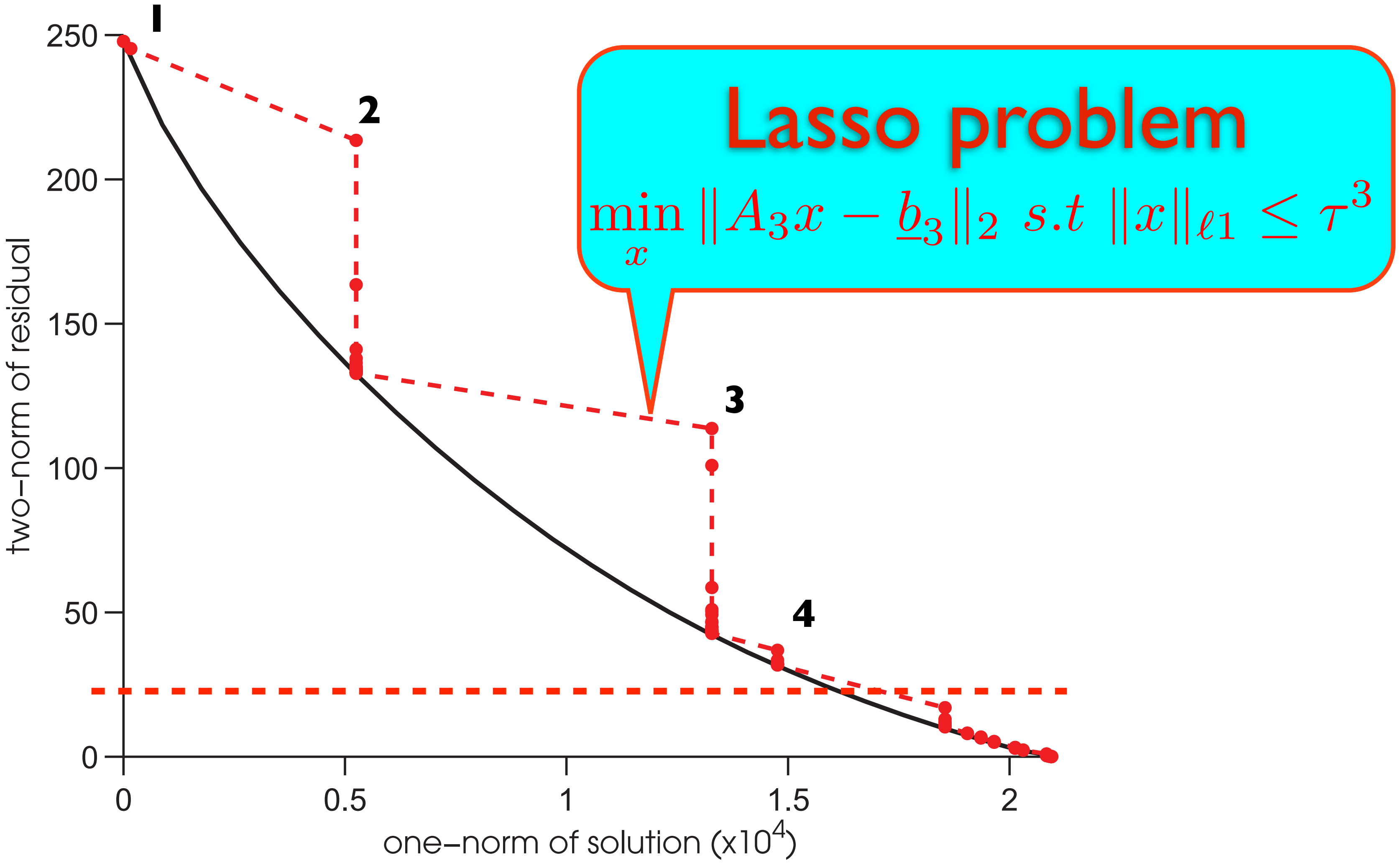
spectral-projected gradients



SLIM

Supercooled

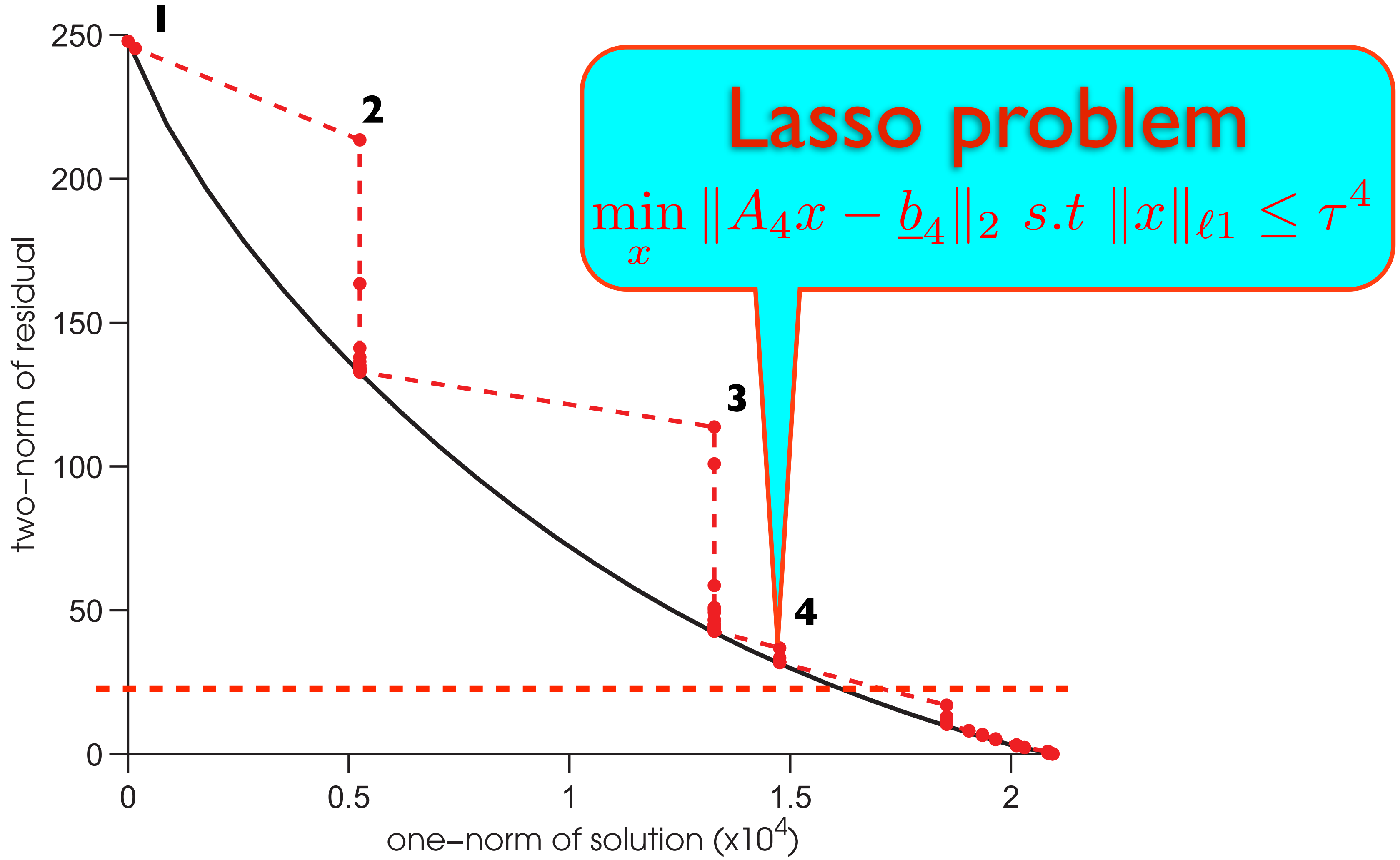
spectral-projected gradients



SLIM

Supercooled

spectral-projected gradients



Supercooled

spectral-projected gradients

Algorithm 1: Modified SPGℓ_1 with message passing.

Result: Estimate for the model $\mathbf{x}^{t+1}$

$\mathbf{x}^0, \widetilde{\mathbf{x}} \longleftarrow \mathbf{0}$ and $t, \tau^0 \longleftarrow 0$; `// Initialize`

while $t \leq T$ **do**

- $\mathbf{A} \longleftarrow \mathbf{A} \sim P(\mathbf{A})$; `// Draw new sensing matrix`
- $\mathbf{b} \longleftarrow \mathbf{A}\mathbf{x}$; `// Collect new data`
- $\mathbf{x}^{t+1} \longleftarrow \mathrm{spgl1}(\mathbf{A}, \mathbf{b}, \tau^t, \sigma = 0, \mathbf{x}^t)$; `// Reach Pareto`
- $\tau^t \longleftarrow \|\mathbf{x}^{t+1}\|_1$; `// New initial` τ `value`
- $t \longleftarrow t + \Delta T$; ; `// Add # of iterations of spgl1`

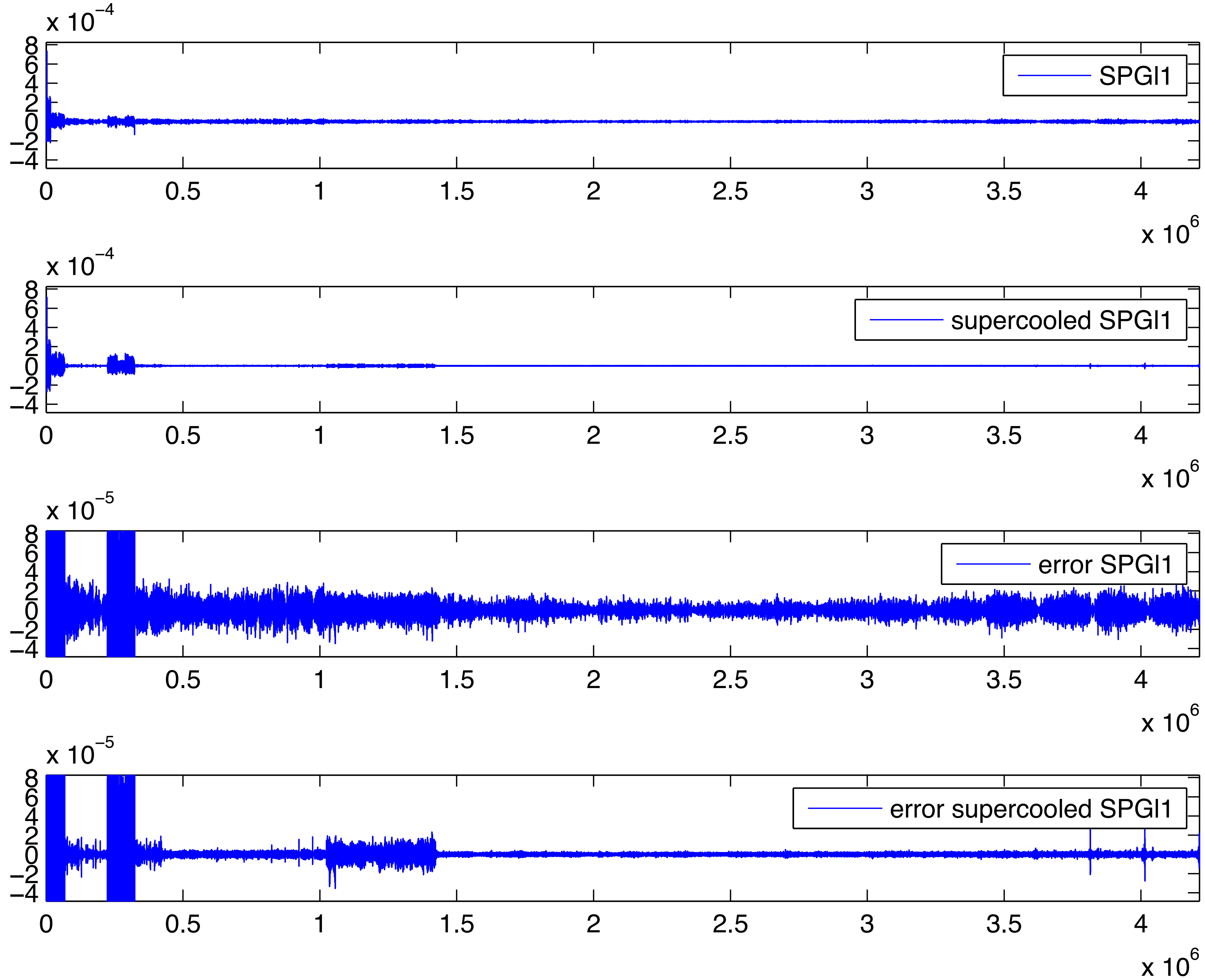
end

Linearized inversion

[estimated coefficients]

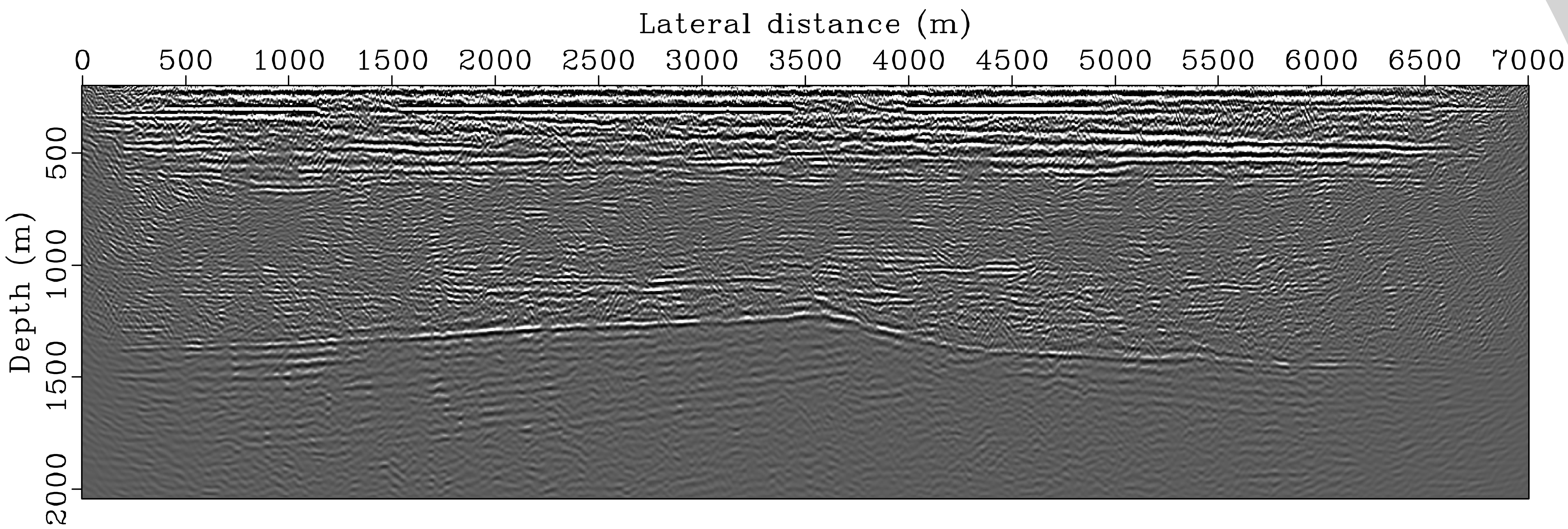
10 X

10 X



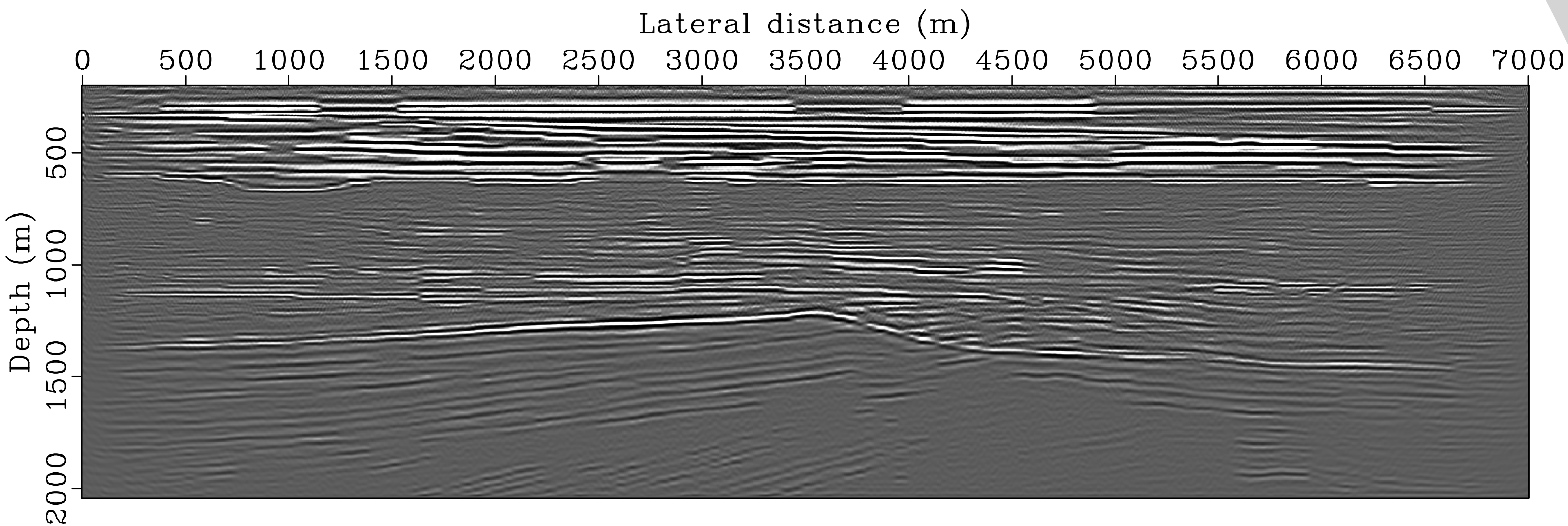
Linearized inversion

[ℓ_1 *without* rerandomization 3 *super* shots]



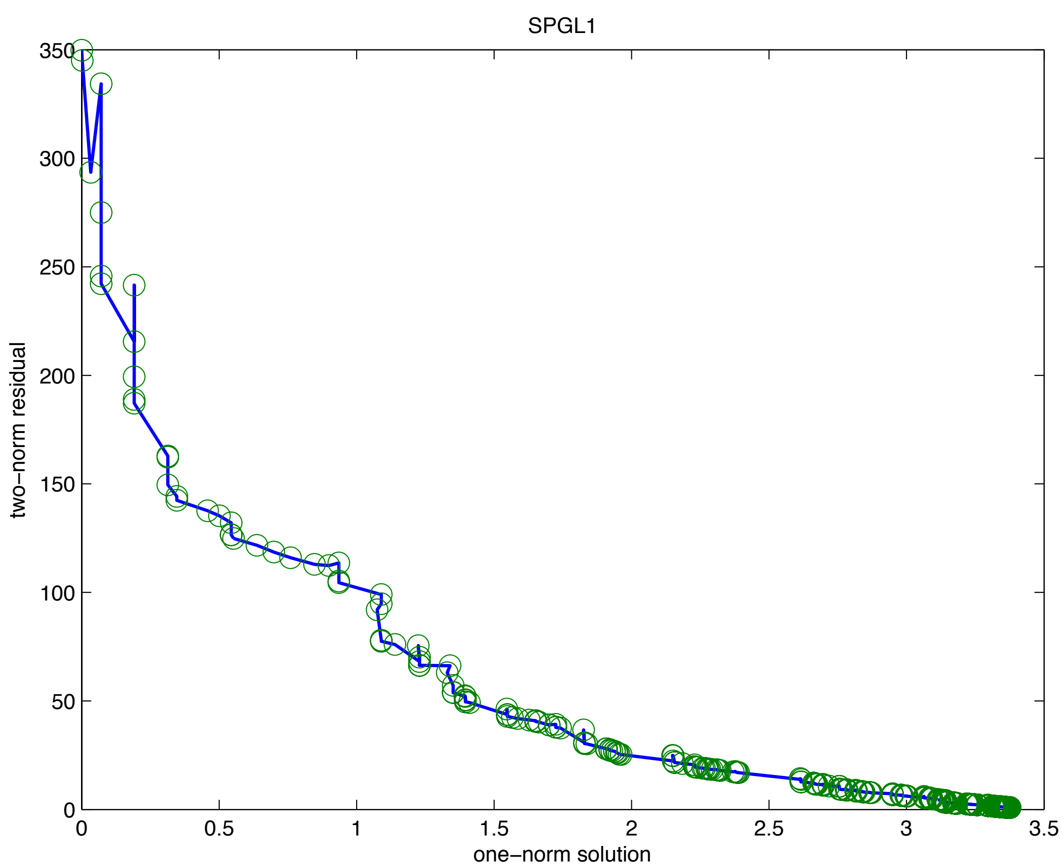
Linearized inversion

[ℓ_1 *with* rerandomization 3 *super* shots]

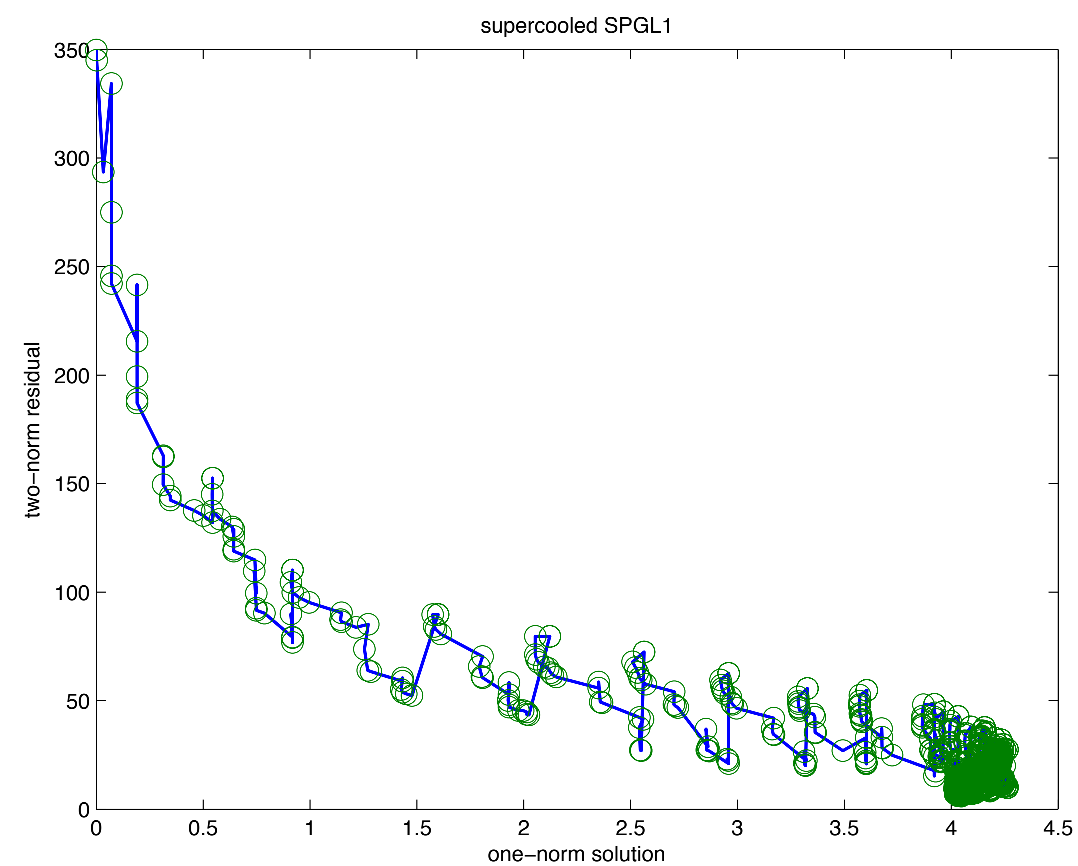


Linearized inversion

[solution paths ℓ_1]



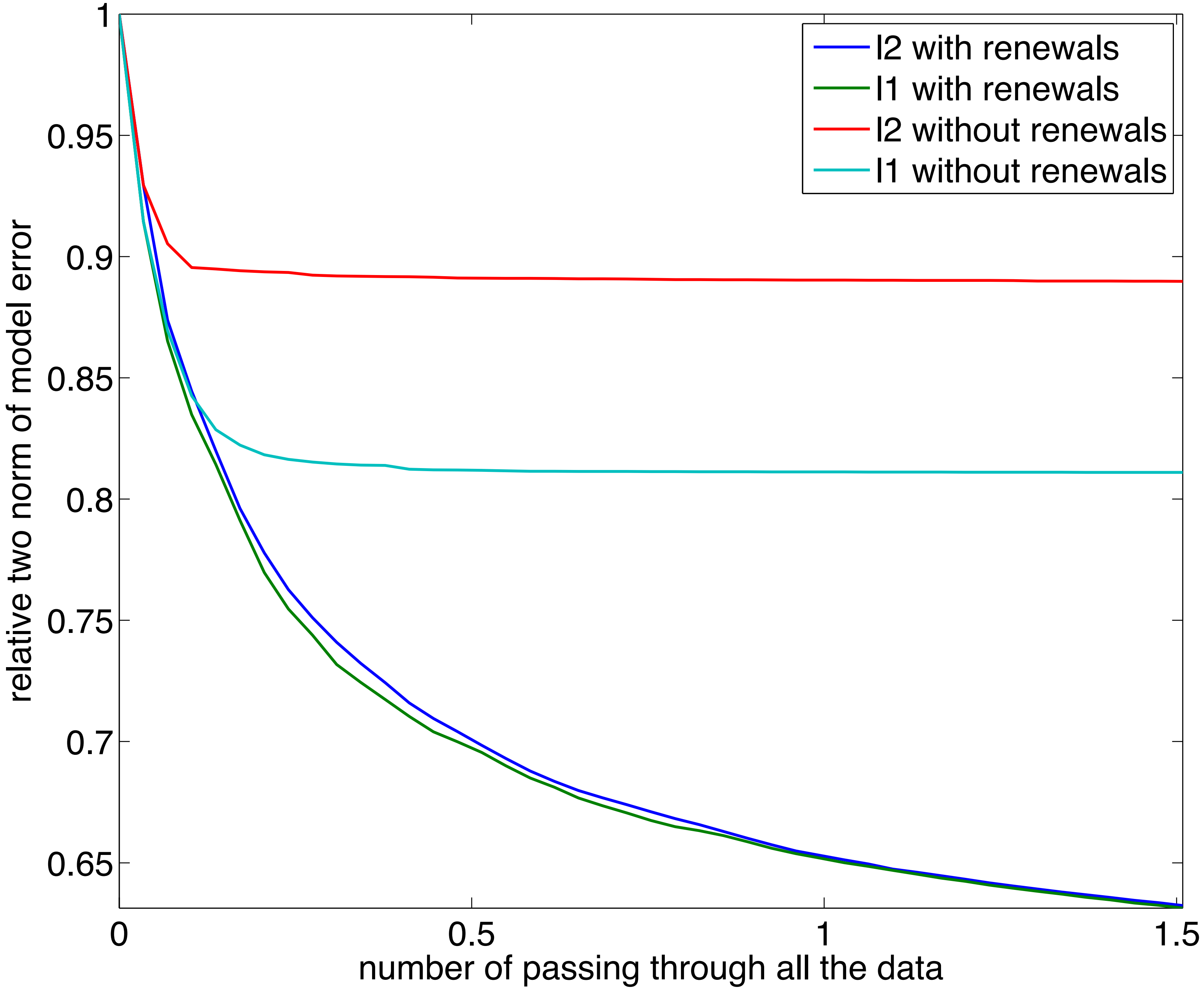
without rerandomization



with rerandomization

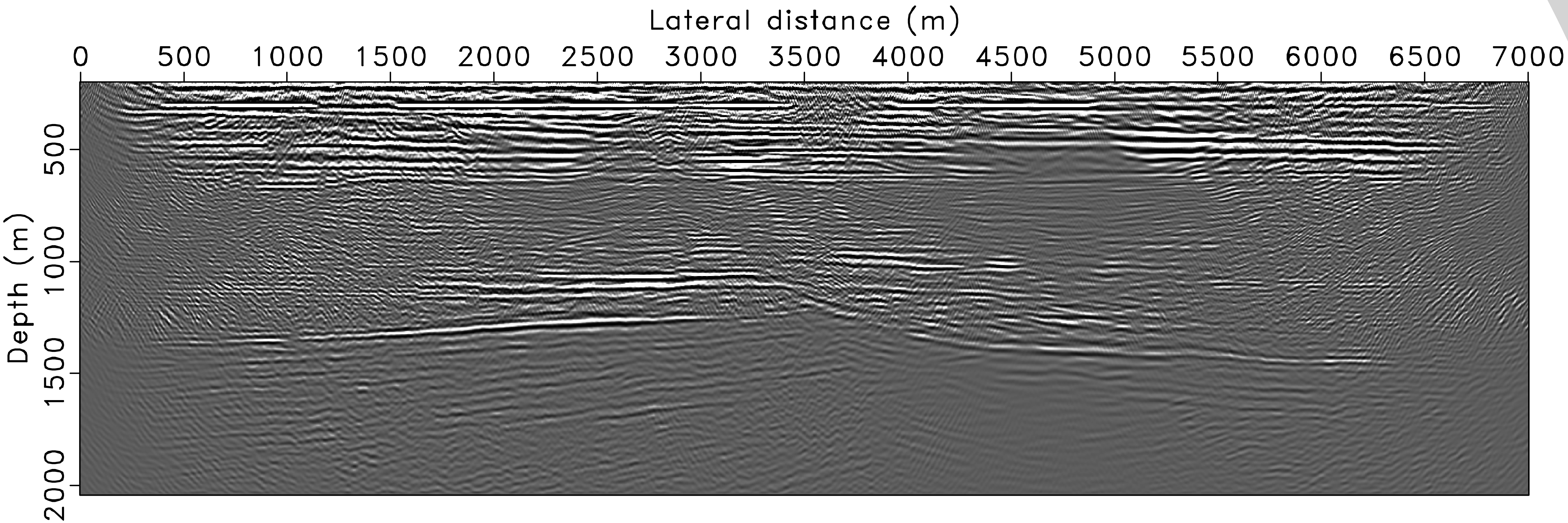
Linearized inversion

[model errors]



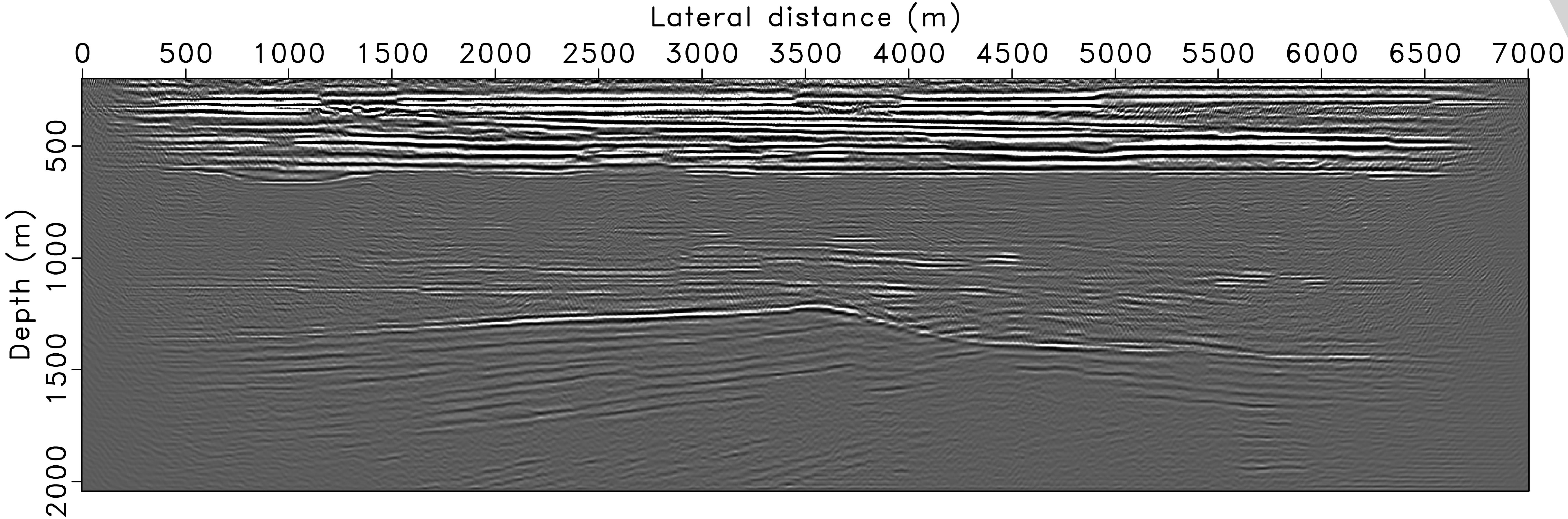
Marine linearized inversion

[ℓ_1 *without* rerandomization 17 shots]



Marine linearized inversion

[ℓ_1 *with* rerandomization 17 shots]



Conclusions

Message passing improves *image* quality

- computationally *feasible* one-norm *regularization*

Message *passing* via *rerandomization*

- *small* system *size* with *small* IO and memory *imprints*

Possibility to exploit *new* computer *architectures* that employ *model* space *parallelism* to *speed up* wavefield *simulations*...

Acknowledgments

We would like to thank Charles Jones from BG for providing us with the BG Compass model. 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, BGP, Chevron, ConocoPhillips, Petrobras, PGS, Total SA, and WesternGeco.

NSERC CRSNG

Thank you

slim.eos.ubc.ca