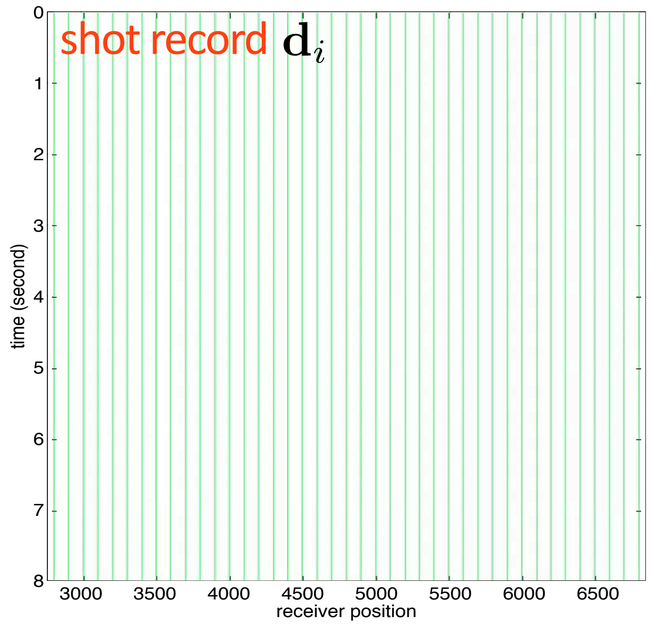
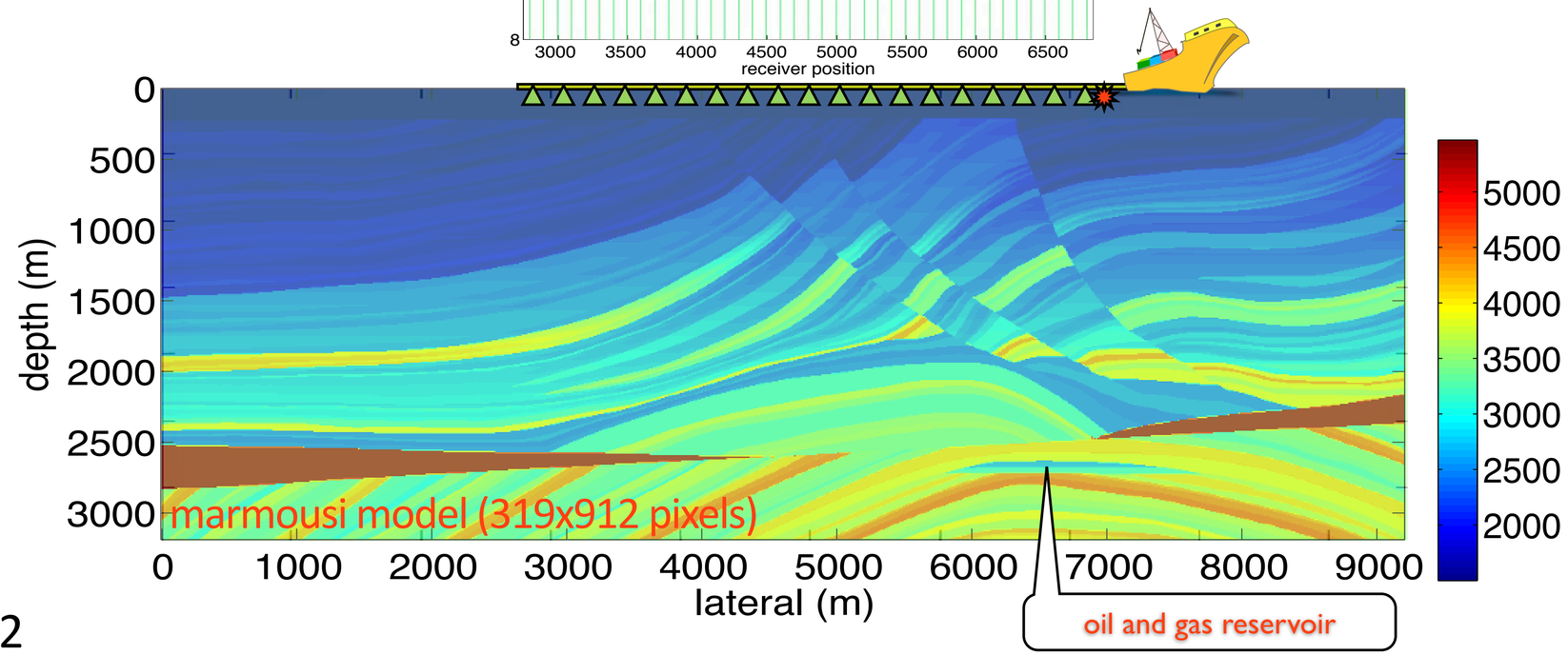
Sparsity-promoting Seismic Imaging and Full-waveform Inversion

Xiang Li

SLIM
University of British Columbia

Exploration seismology





forward modeling

$$\mathbf{d}_i = \mathcal{F}_i(\mathbf{m}) = \mathbf{P}_i\mathbf{A}^{-1}(\mathbf{m})\mathbf{q}_i$$

$\mathbf{d}_i$: vectorized shot record

$\mathbf{P}_i$: receiver restriction operator

$\mathbf{A}$: discretized wave equation stencil matrix

$\mathbf{m}$: model parameters

$\mathbf{q}_i$: source term

$\mathbf{A}$

time domain: 2.4e+09 X 2.4e+09

frequency domain: 2.9e+05 X 2.9e+05

Exploration seismology

forward modeling

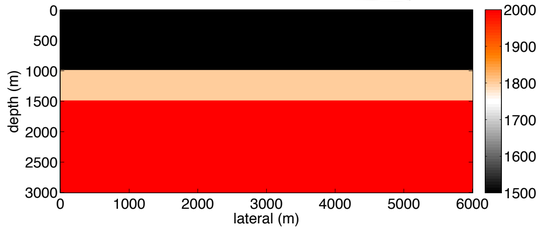
$$\mathbf{d}_i = \boldsymbol{\mathcal{F}}_i(\mathbf{m})$$

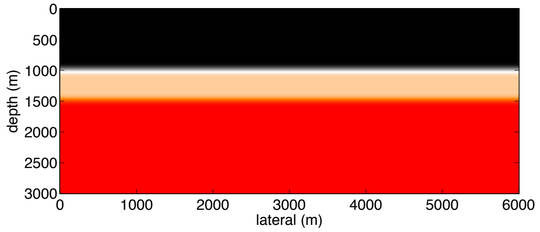
each evaluation of $\mathbf{J}_i$ requires two PDE solvers

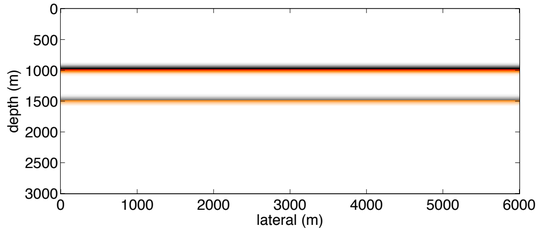
$$\mathbf{J}_i\delta\mathbf{m} = \mathbf{vec}\left[-\mathbf{P}_i\mathbf{A}^{-1}\left[\mathrm{diag}[\frac{\partial\mathbf{A}}{\partial\mathbf{m}}(\mathbf{A}^{-1}\mathbf{q_i})](\mathbf{I}_{nt}\otimes\delta\mathbf{m})\right]\right]$$

consider the Taylor approximation

$$\boldsymbol{\mathcal{F}}_i(\mathbf{m}) = \boldsymbol{\mathcal{F}}_i(\mathbf{m}_0) + \mathbf{J}_i(\mathbf{m}_0)\delta\mathbf{m} + \boldsymbol{\mathcal{O}}(\|\delta\mathbf{m}\|_2^2)$$







Purposes of thesis

Outline:

- linearized seismic imaging with given smooth background velocity model
- estimate velocity with modified Gauss-Newton full-waveform inversion (FWI)

Challenge:

- computational cost associated w/ large number of wave-equation (PDE) solves

Key idea:

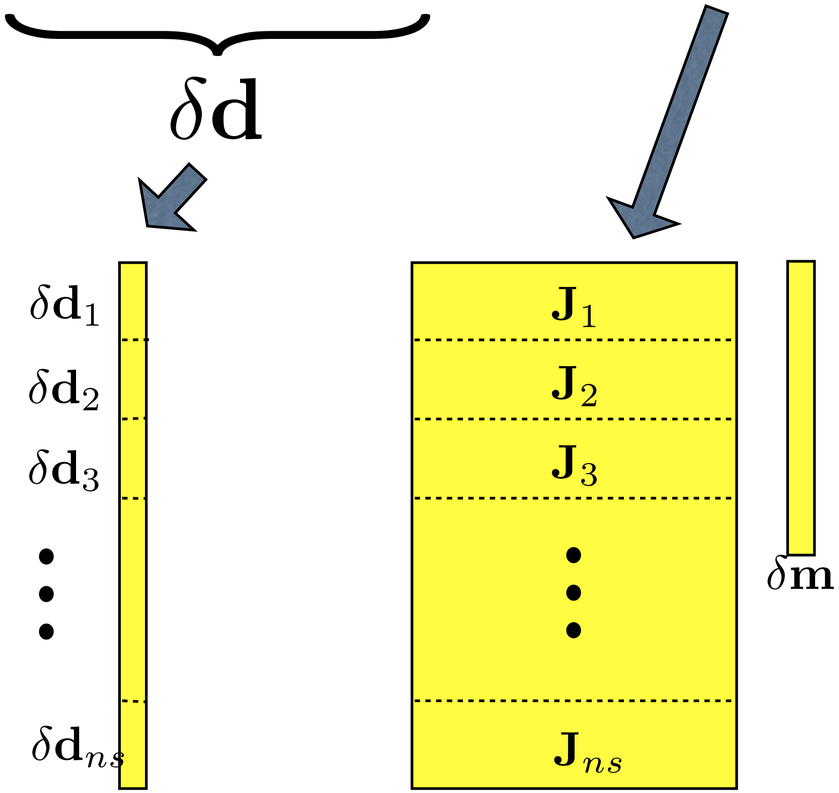
- combine *randomized dimensionality reduction* w/ sparse inversion to reduce
  - # of PDE solves
  - artifacts
- use curvelets to promote sparsity

[H. Kuehl & M. Sacchi '03]
[X. Li & FJH, '10]

SLIM

Seismic imaging
– linearized inversion

$$\delta\mathbf{m} = \arg\min_{\delta\mathbf{m}} \|\underbrace{\boldsymbol{\mathcal{F}}(\mathbf{m}) - \boldsymbol{\mathcal{F}}(\mathbf{m}_0)}_{\delta\mathbf{d}} - \mathbf{J}(\mathbf{m}_0)\delta\mathbf{m}\|_2^2$$



analytic solution

$$\delta\mathbf{m} = \underbrace{(\mathbf{J}^T\mathbf{J})}_{\text{Gauss-Newton Hessian of FWI objective}}{}^{-1}\underbrace{\mathbf{J}^T\delta\mathbf{d}}_{\text{RTM}}$$

[H. Kuehl & M. Sacchi '03]
[X. Li & FJH, '10]

Seismic imaging
– linearized inversion

$$\delta\mathbf{m} = \arg\min_{\delta\mathbf{m}} \|\delta\mathbf{d} - \mathbf{J}(\mathbf{m}_0)\delta\mathbf{m}\|_2^2$$

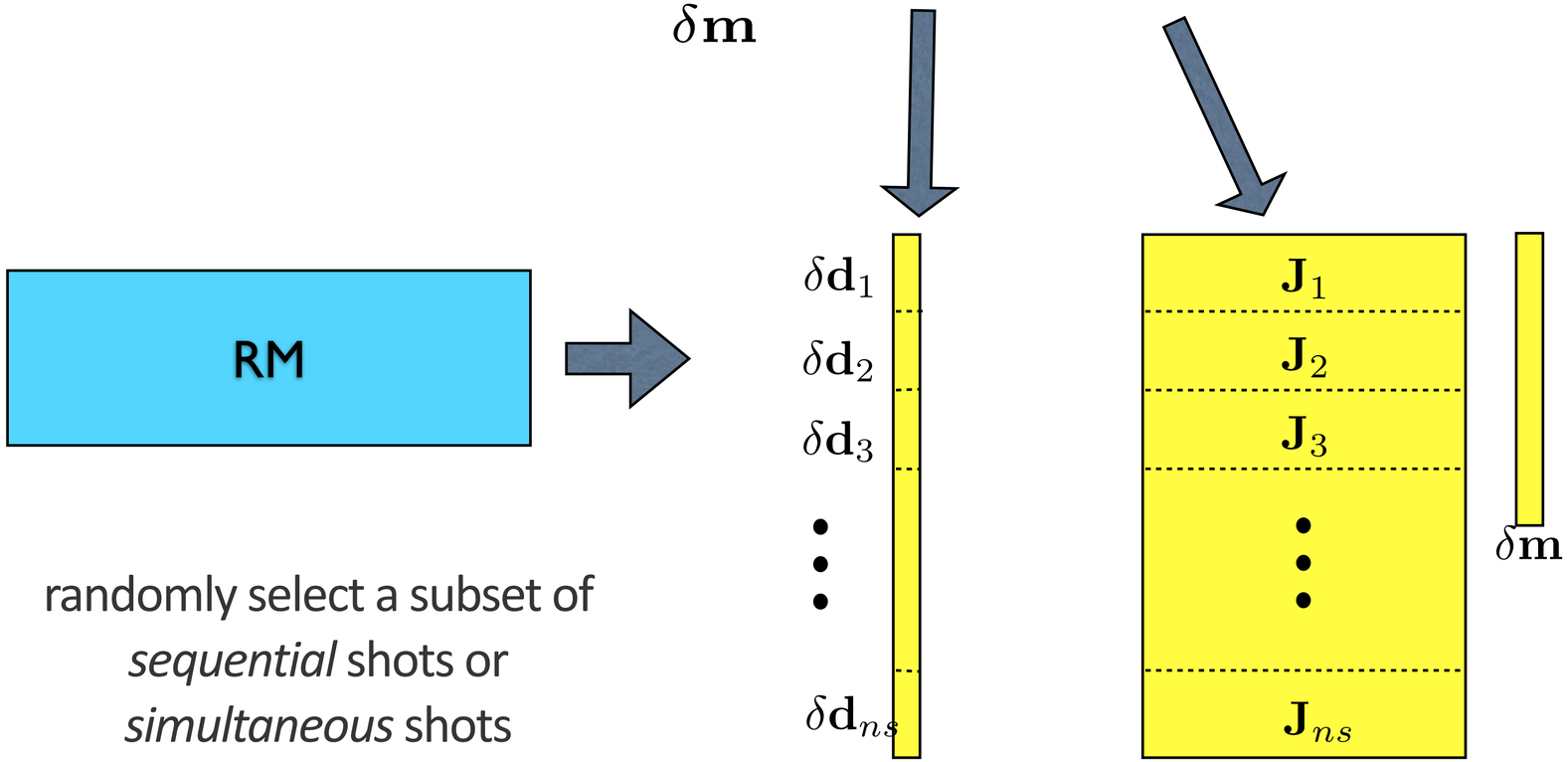
- requires iterative solver
- each iteration requires at least 2 evaluations of $\mathbf{J}$ and $\mathbf{J}^T$
- each evaluation of $\mathbf{J}$ or $\mathbf{J}^T$ requires 2 wave-equation (PDE) solves of all sources

expensive!!!

[X. Li & FJH, '10]
[N. Tu & FJH, '15]
[R. N. Neelamani, etc, '10]
[T. Nemeth, etc, '99]

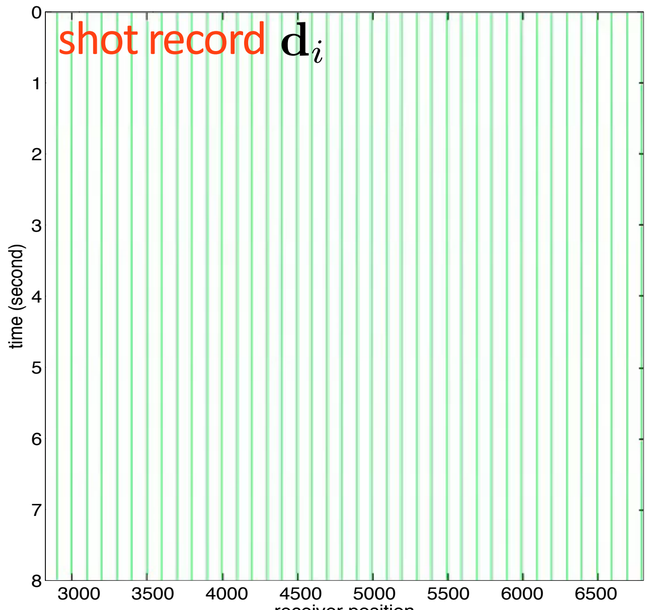
Dimensionality reduction

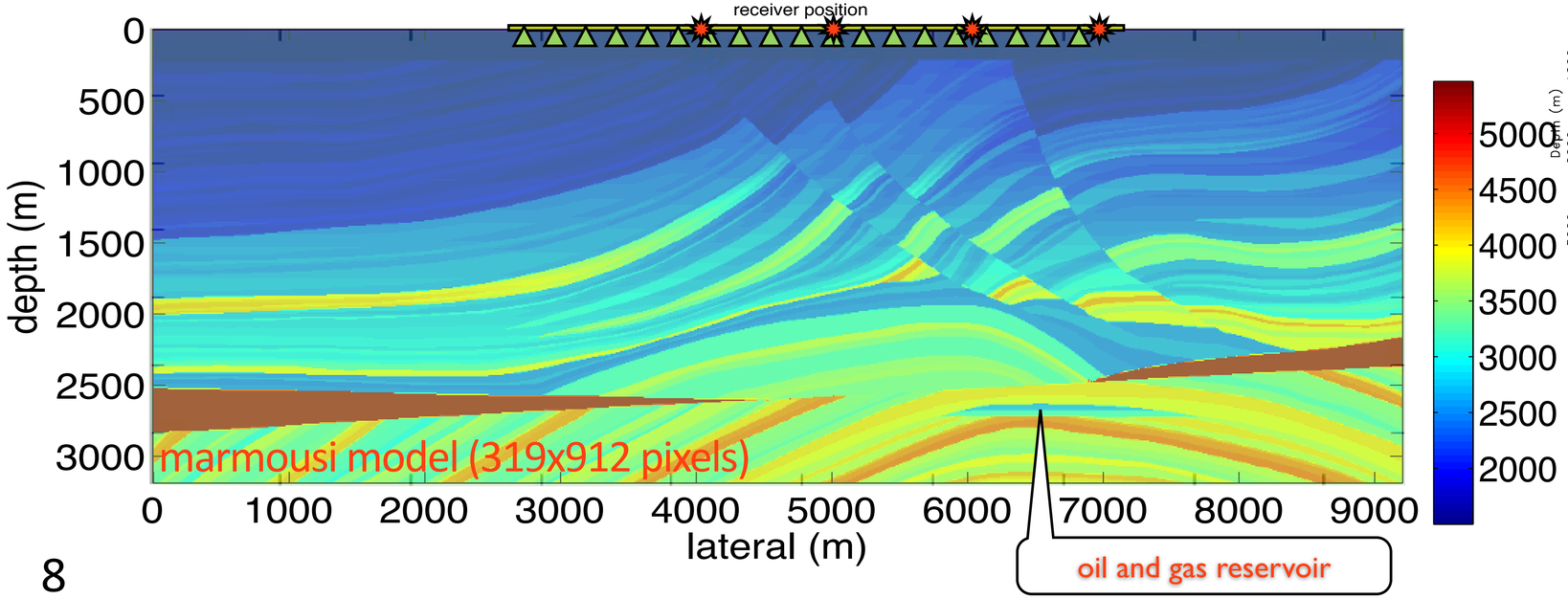
$$\delta\mathbf{m} = \arg\min_{\delta\mathbf{m}} \|\delta\mathbf{d} - \mathbf{J}(\mathbf{m}_0)\delta\mathbf{m}\|_2^2$$



randomly select a subset of *sequential* shots or *simultaneous* shots

Simultaneous shot





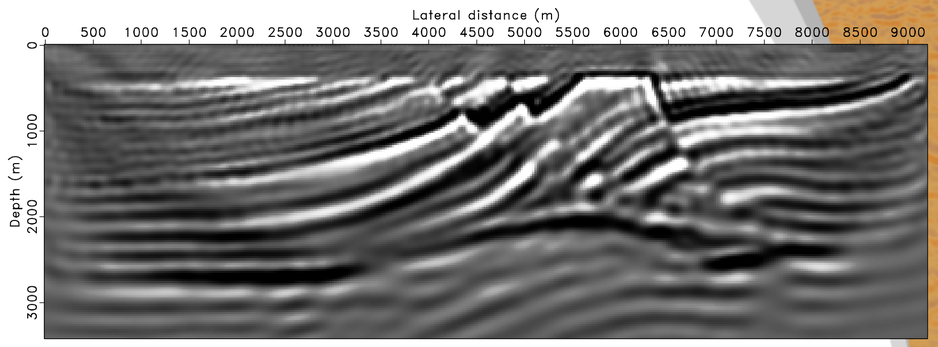


image of 192 *sequential* shots

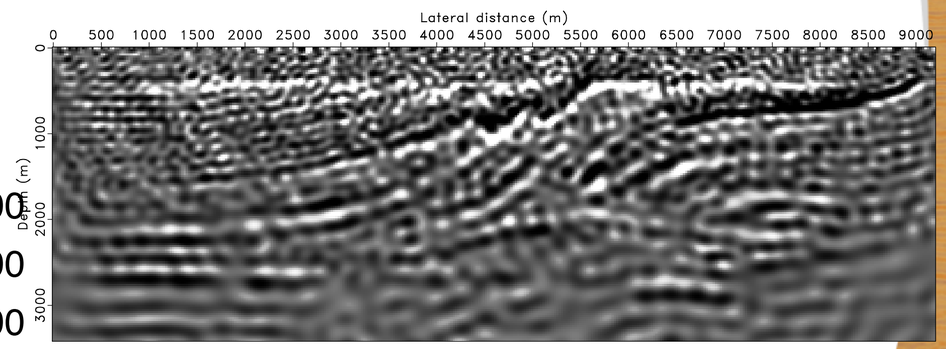


image of 8 *simultaneous* shots

[X. Li & FJH, '10]
[E. van den Berg and M. P. Friedlander. '06]

Sparsity-promoting linearized inversion

$$\min_{\delta\mathbf{x}} \|\delta\mathbf{x}\|_1 \quad \text{subject to} \quad \|\delta\mathbf{d} - \mathbf{J}(\mathbf{m}_0)\mathbf{C}^T\delta\mathbf{x}\|_2^2 \leq \sigma$$

$$\delta\mathbf{m} = \mathbf{C}^T\delta\mathbf{x}$$

$\mathbf{C}^T$: inverse curvelet transform

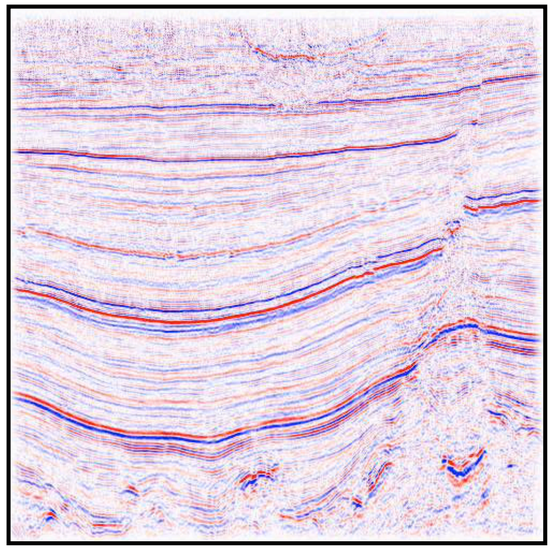
$\delta\mathbf{x}$: curvelet coefficients of imaging result

- solving an *intelligent* series of relaxed LASSO *subproblems* for *decreasing* sparsity levels (convex constraint in terms of the one-norm)
- slowly allowing *components* to *enter* into the *solution*

[Demanet et. al., '06]
[Hennenfent & FJH, '06]

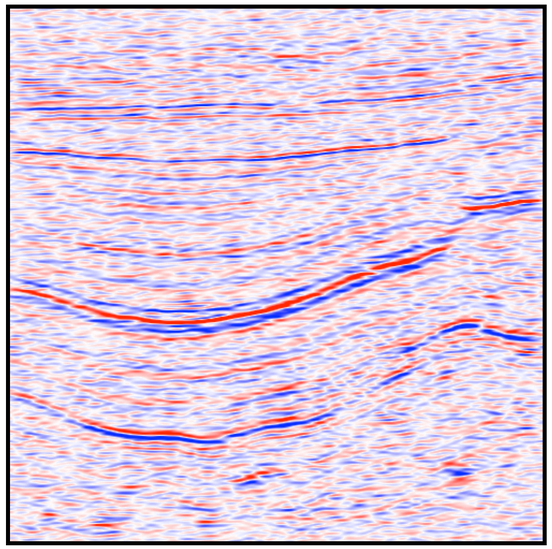
Motivation

exploit multiscale and multi-angle structure of real seismic images



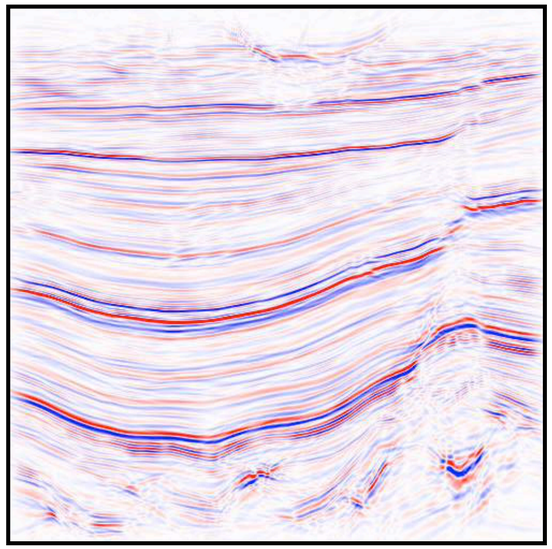
seismic image

SNR 2.1 dB



nonlinear approximation w/ Fourier

SNR 6.0 dB



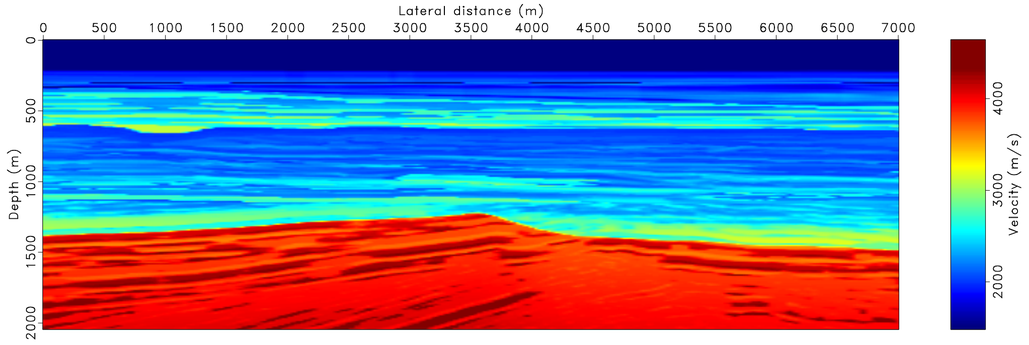
nonlinear approximation w/ curvelets

1 % of coefficients

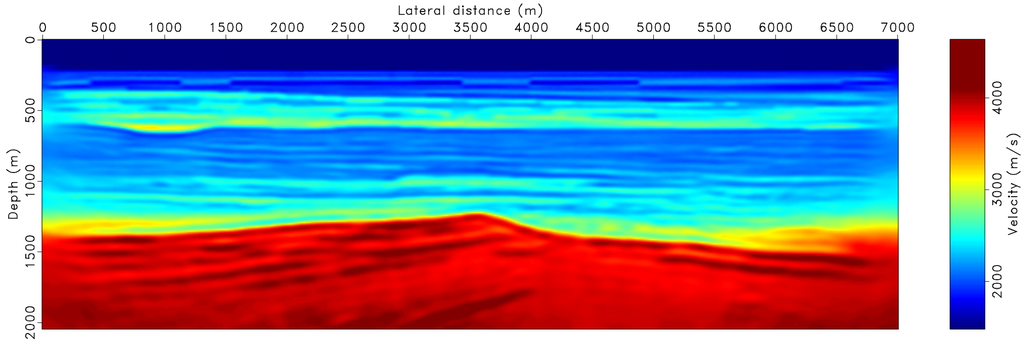
BG model example

BG Compass model

- 2 x 7 km, 5m grid interval
- 350 shot positions, 700 fixed receivers
- 20-50Hz, 10 randomly frequency bands
- 10 iterations
- observed data is generated with the same modeling kernel



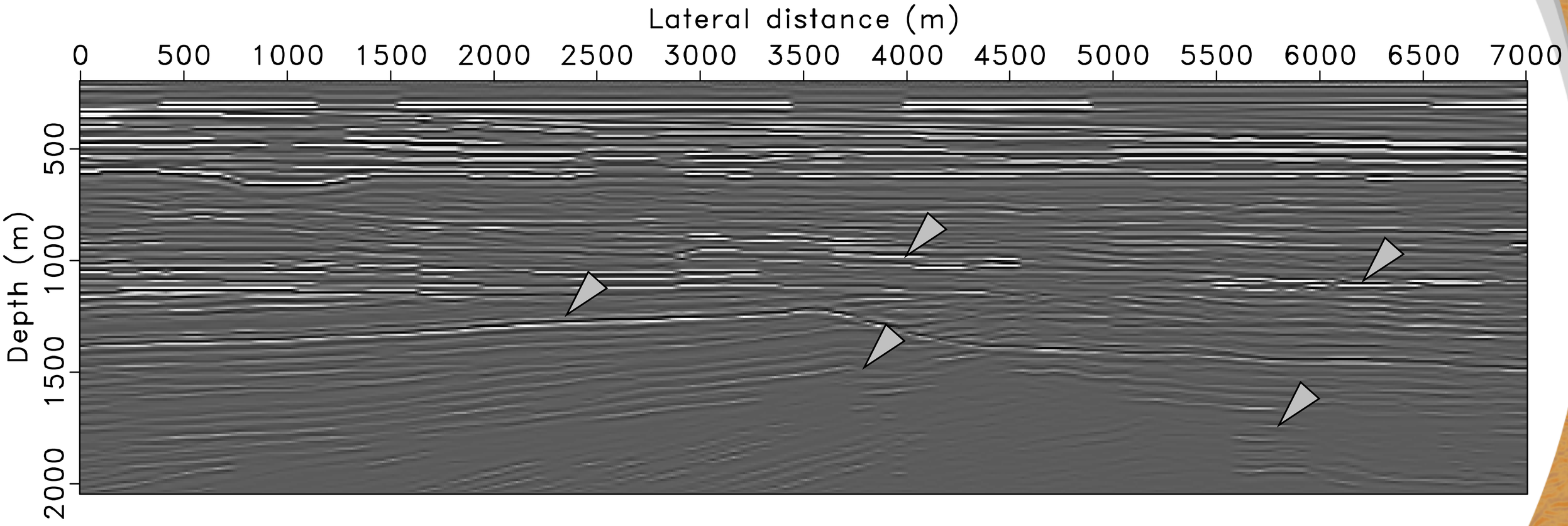
true model



background velocity model

Experiments

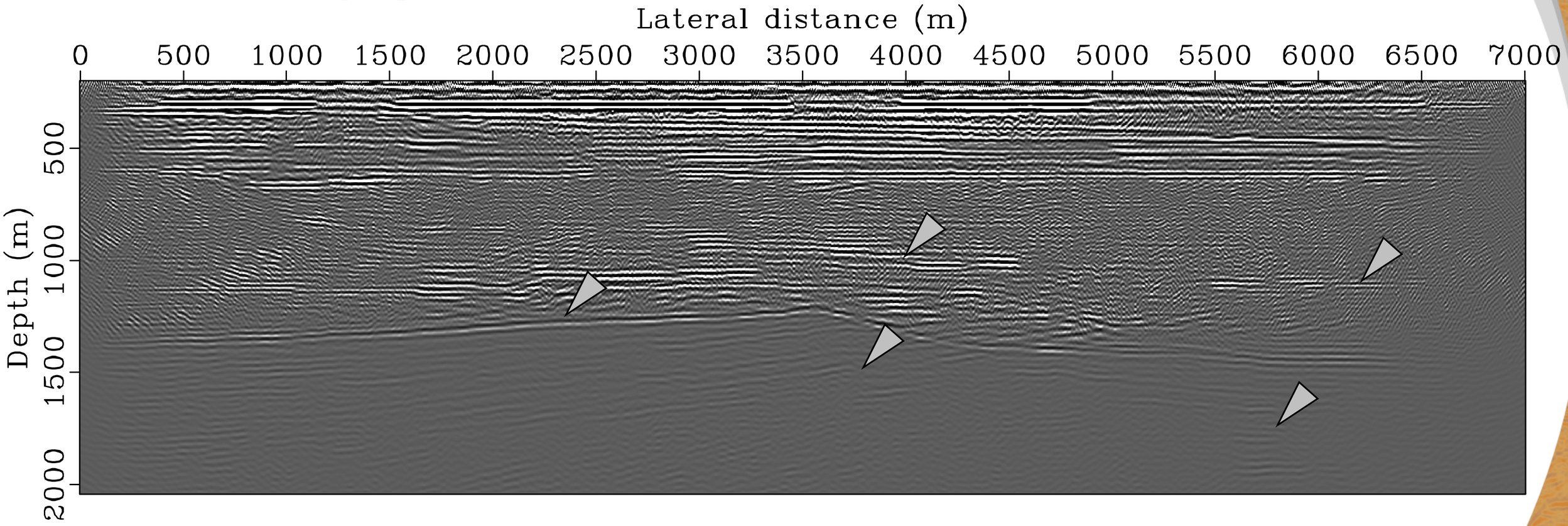
true perturbation



Reverse-time migration

$$\delta \mathbf{m} = \mathbf{J}^T \delta \mathbf{d}$$

with all 350 sequential shots and 10 randomly selected frequencies (2%)

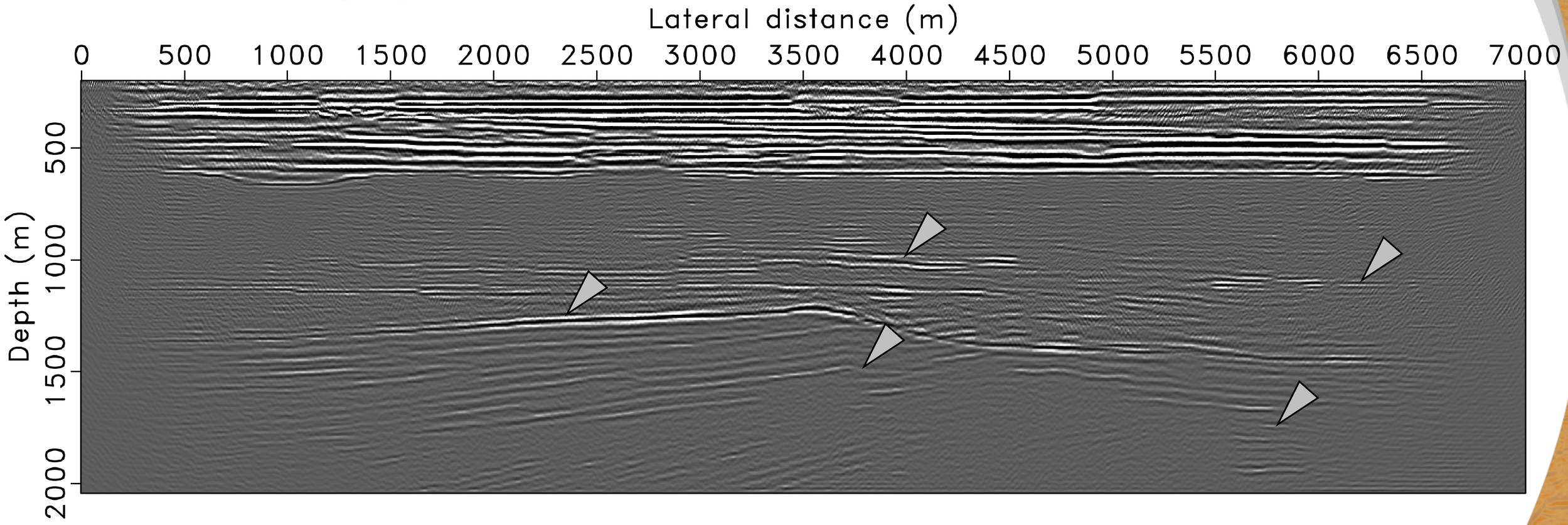


10,500 PDE solves

Linearized inversion

with all 350 sequential shots and 10 randomly selected frequencies (2%)

$$\delta\mathbf{m} = (\mathbf{J}^T\mathbf{J})^{-1}\mathbf{J}^T\delta\mathbf{d}$$

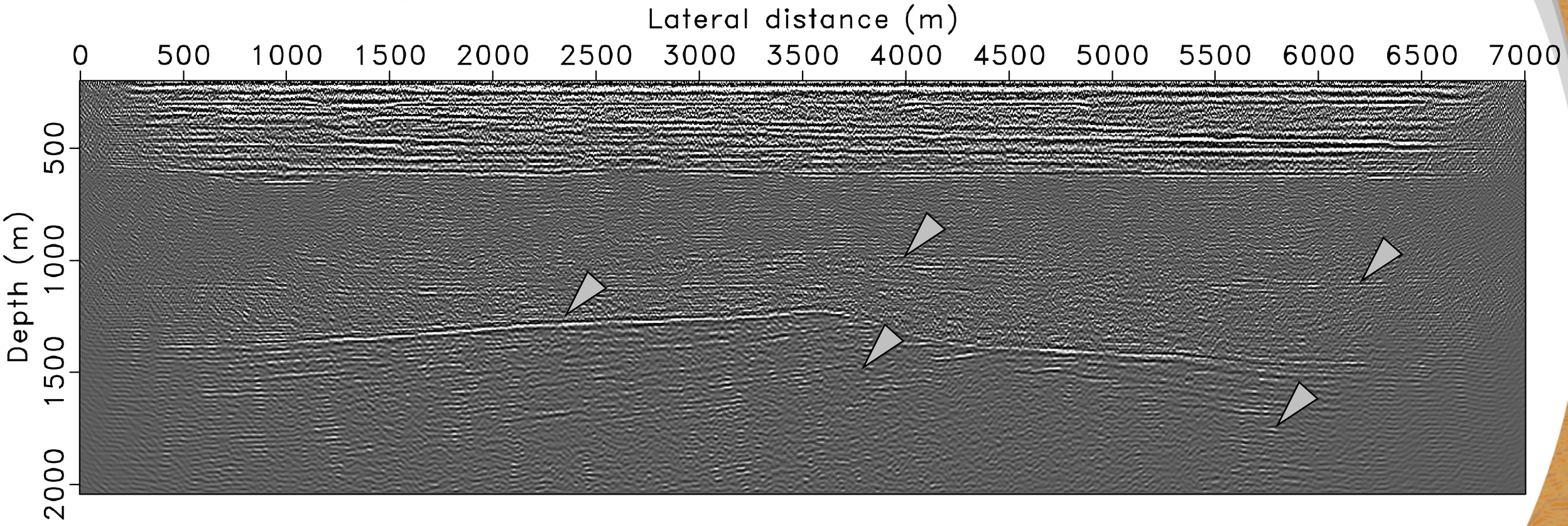


203,500 PDE solves

[X. Li & FJH, '11]

Linearized inversion

with 17 simultaneous shots and 10 randomly selected frequencies (2%)



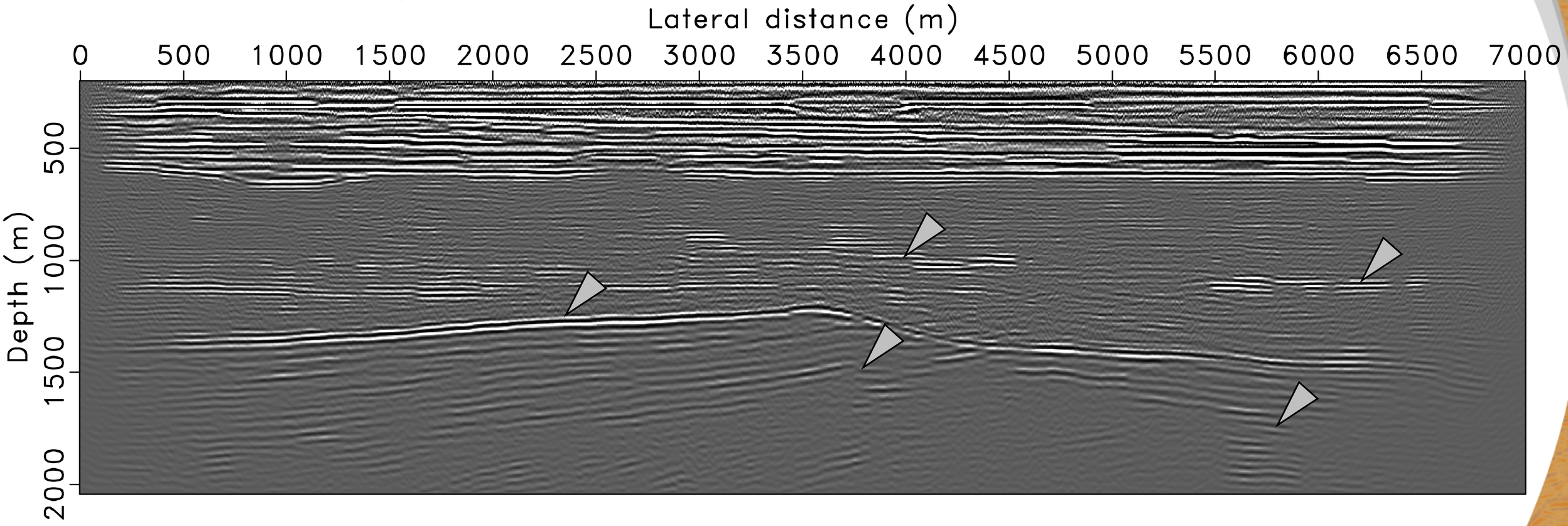
speedup: x29 times

6,970 PDE solves

[X. Li & FJH, '11]

Sparsity promoting linearized inversion

with 17 simultaneous shots and 10 randomly selected frequencies (2%)



speedup: x29 times

6,970 PDE solves

Observations

- computational cost can be *reduced* significantly by using *randomized dimensionality reduction*
- *curvelet-domain* sparsity promotion can suppress the subsampling related artifacts & noise

We still need an accurate velocity model !!!

[R.G. Pratt, '98]
[R.-E. Plessix, etc, '06]
[X. Li & FJH, '12]

Full-waveform inversion

FWI objective

$$\Phi(\mathbf{m}) = \sum_{i=1}^{n_s} \frac{1}{2} \|\mathbf{d}_i - \boldsymbol{\mathcal{F}}_i(\mathbf{m})\|_2^2$$

Gauss-Newton update

$$\delta\mathbf{m} := \arg\min_{\delta\mathbf{m}} \|\delta\mathbf{d} - \mathbf{J}\delta\mathbf{m}\|_2^2$$

the modified Gauss-Newton update

$$\delta\mathbf{m} := \mathbf{C}^T \arg\min_{\delta\mathbf{x}} \|\underline{\delta\mathbf{d}} - \underline{\mathbf{J}}\mathbf{C}^T\delta\mathbf{x}\|_2^2 \quad \text{s.t.} \quad \|\delta\mathbf{x}\|_1 \leq \tau$$

[X. Li & FJH, '12]

The modified Gauss-Newton algorithm

Algorithm 1: Modified Gauss-Newton with curvelet-domain sparsity promotion and randomization.

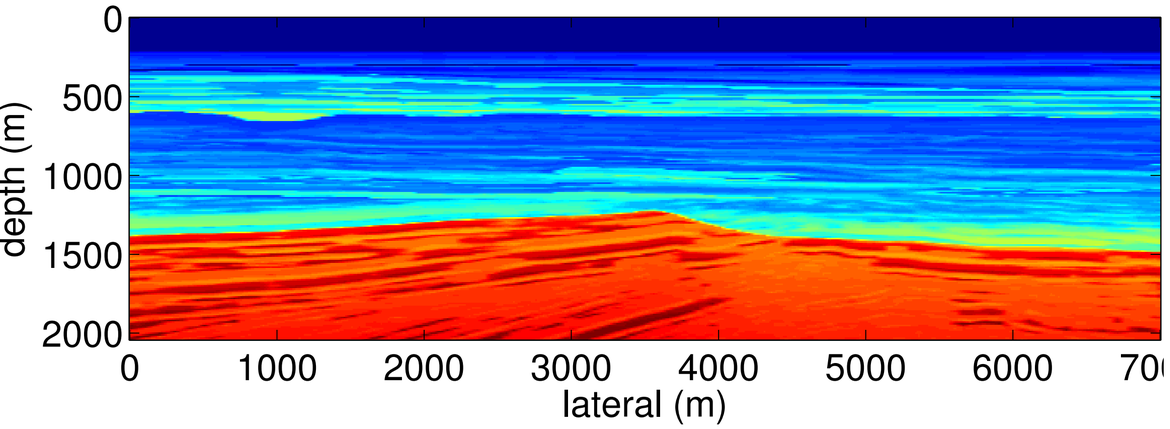
Output: Solution $\widetilde{\mathbf{m}}$ of the randomized modified Gauss-Newton problem for starting model $\mathbf{m}_0$, tolerance ξ, and step length α.

1. $\widetilde{\mathbf{m}} \longleftarrow \mathbf{m_0}$, and ξ `// initial guess and expected residual`
2. **while** $\|\delta\mathbf{d}_k\|_2 \geq \xi$ **do**
4. $\tau_k = \|\delta\mathbf{d}_k\|_2 / \|\mathbf{C}\mathbf{J}(\mathbf{m}_k)^T \delta\mathbf{d}_k\|_\infty$
5. $\delta\mathbf{m}_k = \mathbf{C}^T \arg\min_{\delta\mathbf{x}} \|\delta\mathbf{d}_k - \mathbf{J}(\mathbf{m}_k)\mathbf{C}^T\delta\mathbf{x}\|_2^2$ subject to $\|\delta\mathbf{x}\|_1 \leq \tau_k$.
6. $\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha\mathbf{C}^T\delta\mathbf{x}_k$ `// update with linesearch`
7. **end**

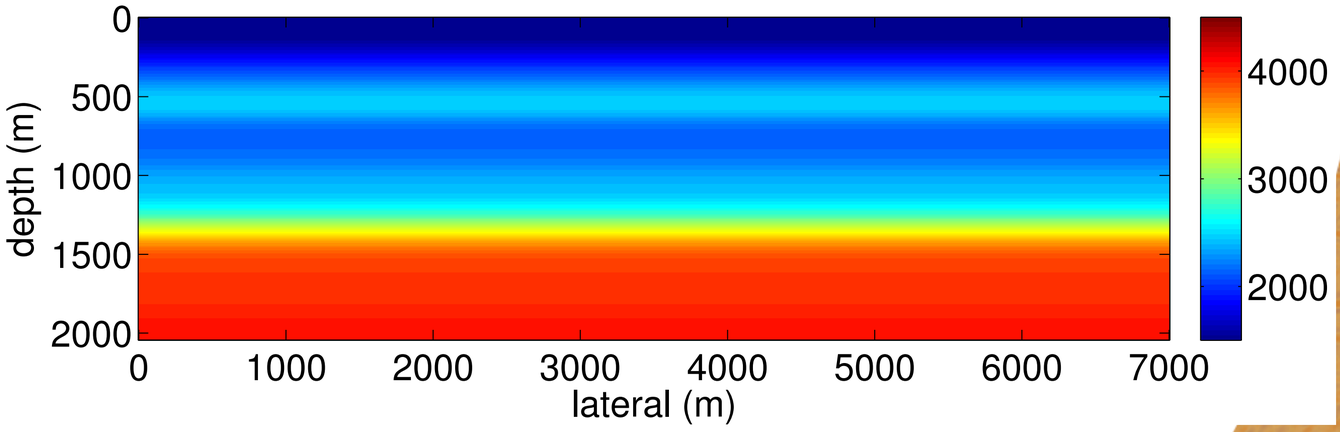
BG model example

BG Compass model

- 2 x 7 km
- 350 shot positions, 700 fixed receivers
- 3-15Hz, 10 frequency bands
- 5 GN updates for each band
- observed data is from time domain finite difference

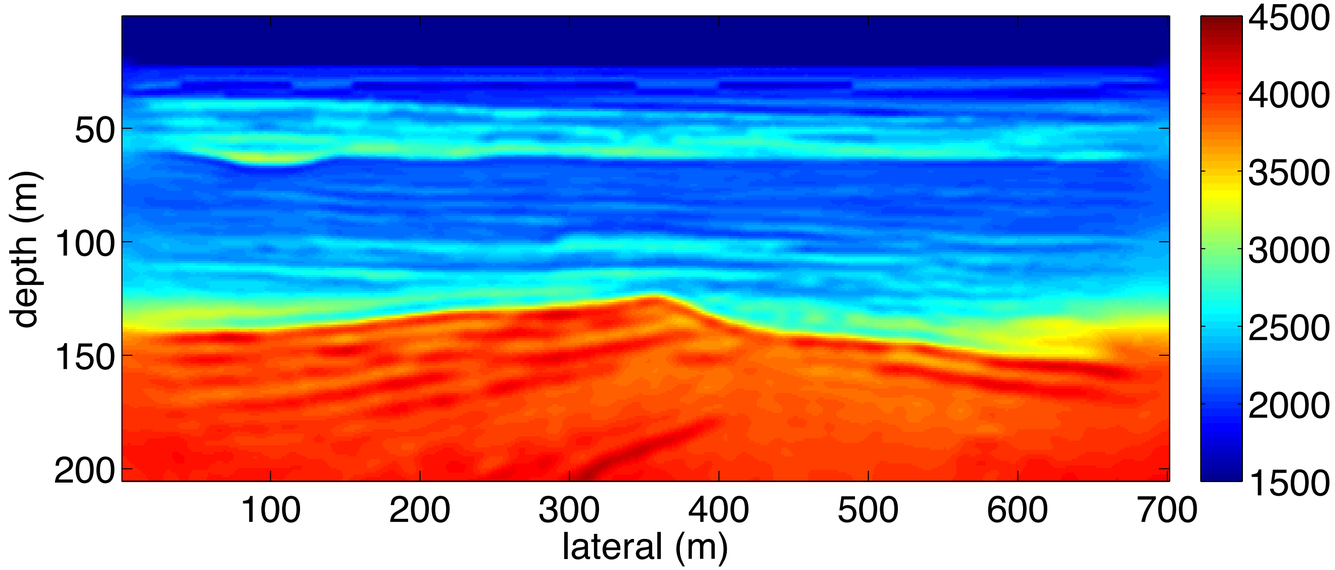


true model

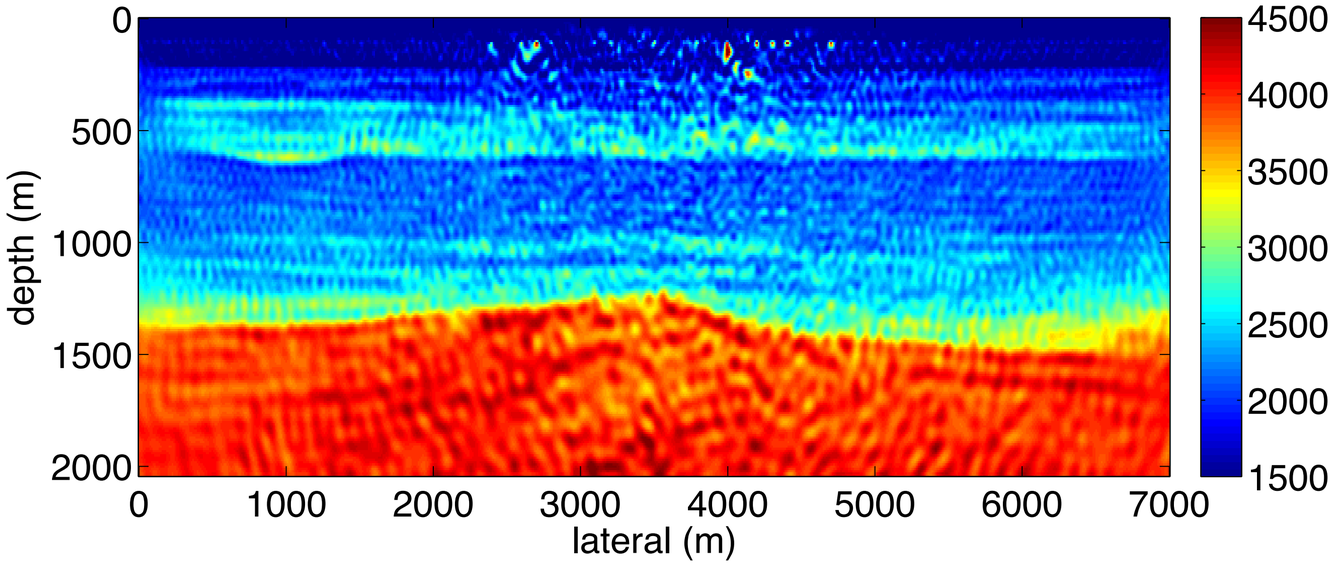


initial model

Inversion results



the modified Gauss-Newton method with L1 constraint



the modified Gauss-Newton method with L2 constraint

Observation & question

Modified Gauss-Newton:

- only promotes sparsity on individual updates
- does NOT change the FWI objective function

Why

- **would the *sum* of all *sparse* updates still be sparse?**
- **is promoting sparsity on the Gauss-Newton updates a good idea?**

Least-squares optimization problem

Unconstrained objective function:

$$\min_{\mathbf{m}} \Phi(\mathbf{m}) := \left\{ \frac{1}{2} \|\mathbf{d} - \boldsymbol{\mathcal{F}}(\mathbf{m})\|_2^2 \right\}$$

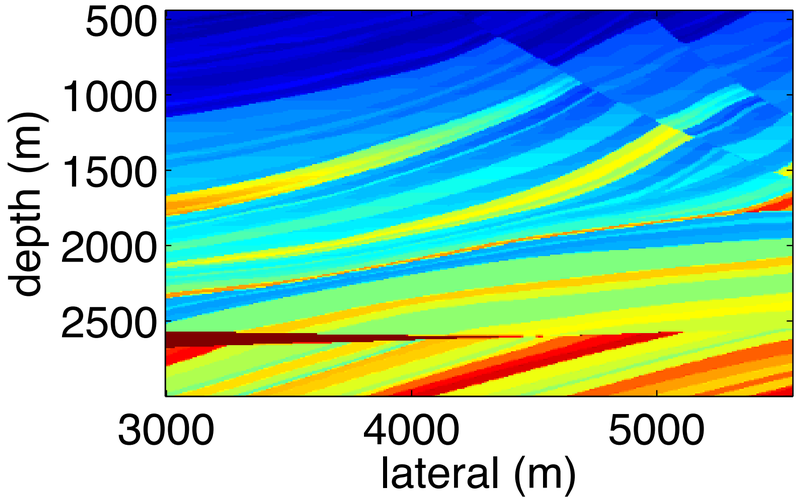
Gauss-Newton update:

$$\delta\mathbf{m} = \arg\min_{\delta\mathbf{m}} \|\delta\mathbf{d} - \mathbf{J}(\mathbf{m}_k)\delta\mathbf{m}\|_2 \qquad 1$$

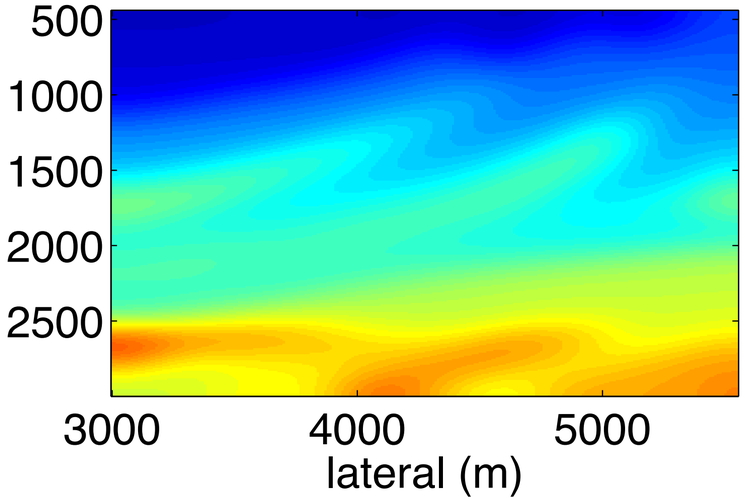
Modified Gauss-Newton update:

$$\delta\mathbf{m} = \mathbf{C}^T \arg\min_{\delta\mathbf{x}} \|\delta\mathbf{d} - \mathbf{J}(\mathbf{m}_k)\mathbf{C}^T\delta\mathbf{x}\|_2^2 \quad \text{subject to} \quad \|\delta\mathbf{x}\|_1 \leq \tau \qquad 2$$

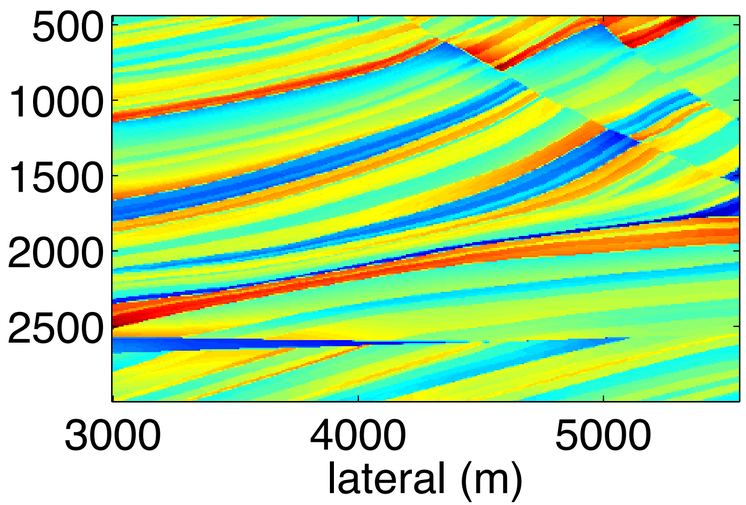
Sparse velocity perturbation



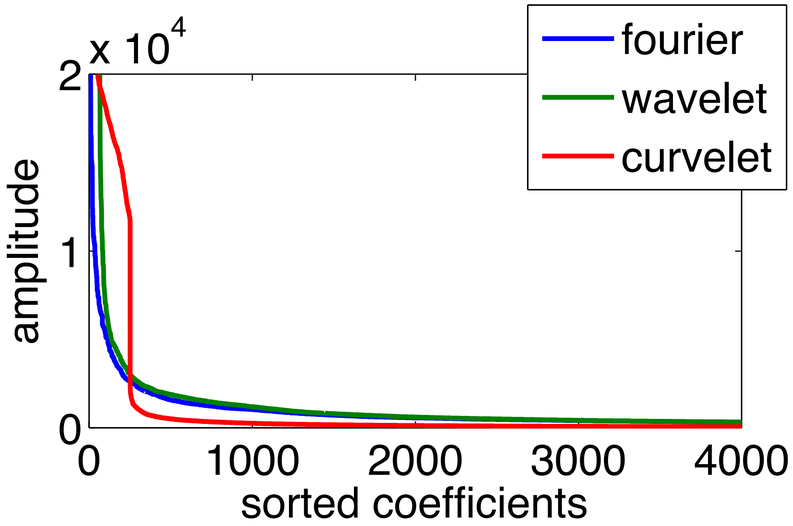
true velocity model

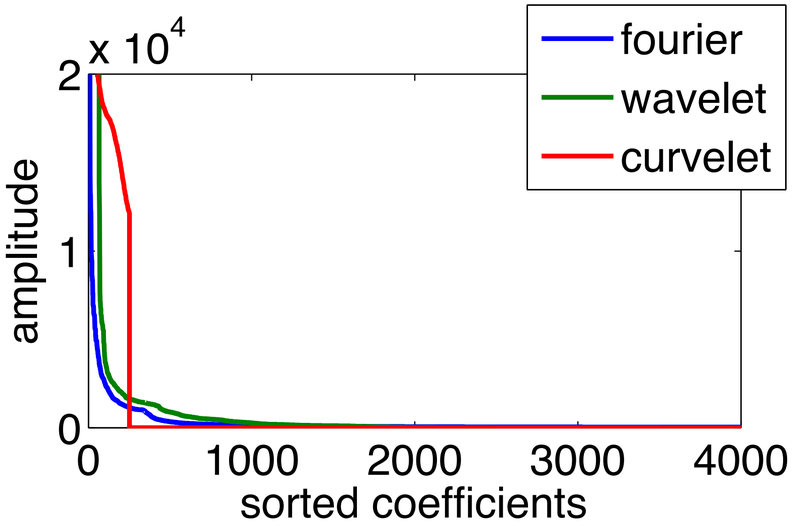


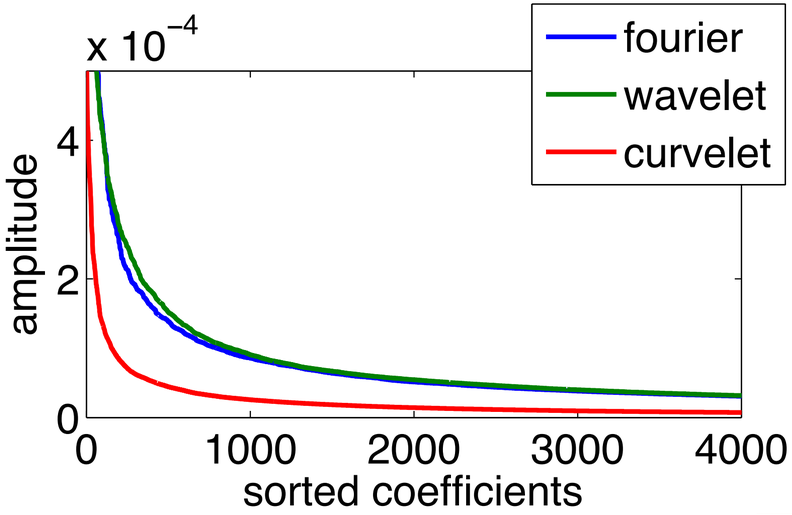
initial velocity model



velocity perturbation







Least-squares optimization problem
– w/ sparse constraint

Objective function with sparse constraint:

$$\min_{\mathbf{x}} \Phi(\mathbf{x}) := \left\{ \frac{1}{2} \|\mathbf{d} - \boldsymbol{\mathcal{F}}(\mathbf{C}^T \mathbf{x})\|_2^2 \right\} \quad \text{subject to} \quad \|\mathbf{x} - \mathbf{x}_0\|_1 \leq \tau$$

Gauss-Newton update:

$$\delta \mathbf{m} = \mathbf{C}^T \arg\min_{\delta \mathbf{x}} \|\delta \mathbf{d} - \mathbf{J}(\mathbf{m}_k) \mathbf{C}^T \delta \mathbf{x}\|_2^2 \quad \text{subject to} \quad \|\delta \mathbf{x} + \mathbf{x}_k - \mathbf{x}_0\|_1 \leq \tau$$

3

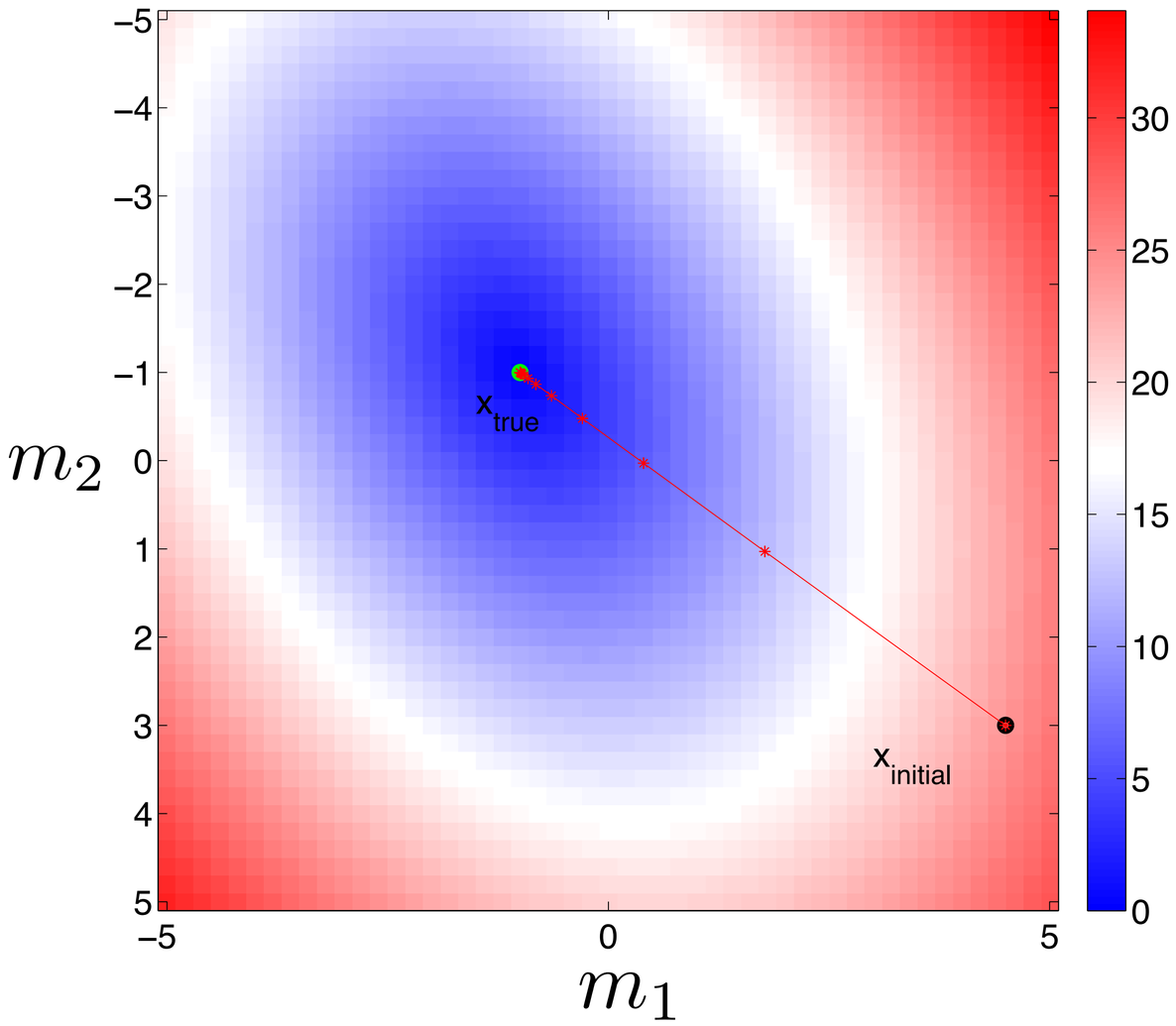
Convex problem
w/ unique solution

$$\Phi(\mathbf{m}) := \left\{ \frac{1}{2} \|\mathbf{d} - \mathbf{A}\mathbf{m}\|_2^2 \right\}$$

$$\mathbf{A} = \begin{bmatrix} 2 & 4 \\ 6 & -3 \end{bmatrix}$$

$$\mathbf{d} = \begin{bmatrix} -6 \\ -3 \end{bmatrix}$$

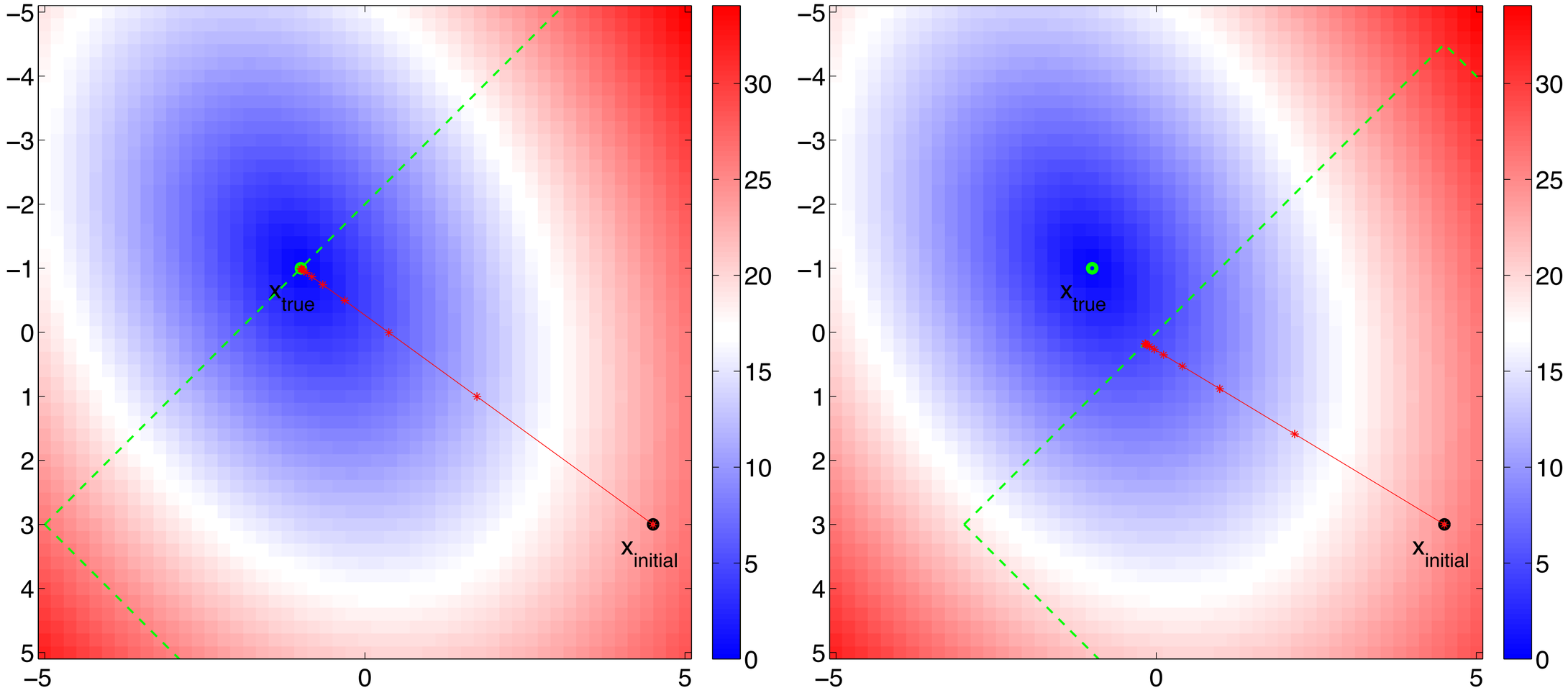
$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$



GN with unconstrained objective

GN – w/ sparse constrained objective function

$$\min_{\mathbf{x}} \Phi(\mathbf{x}) := \left\{ \frac{1}{2} \|\mathbf{d} - \boldsymbol{\mathcal{F}}(\mathbf{S}^{\mathrm{H}}\mathbf{x})\|_2^2 \right\} \quad \text{subject to} \quad \|\mathbf{x} - \mathbf{x}_0\|_1 \leq \tau$$



with right constraint

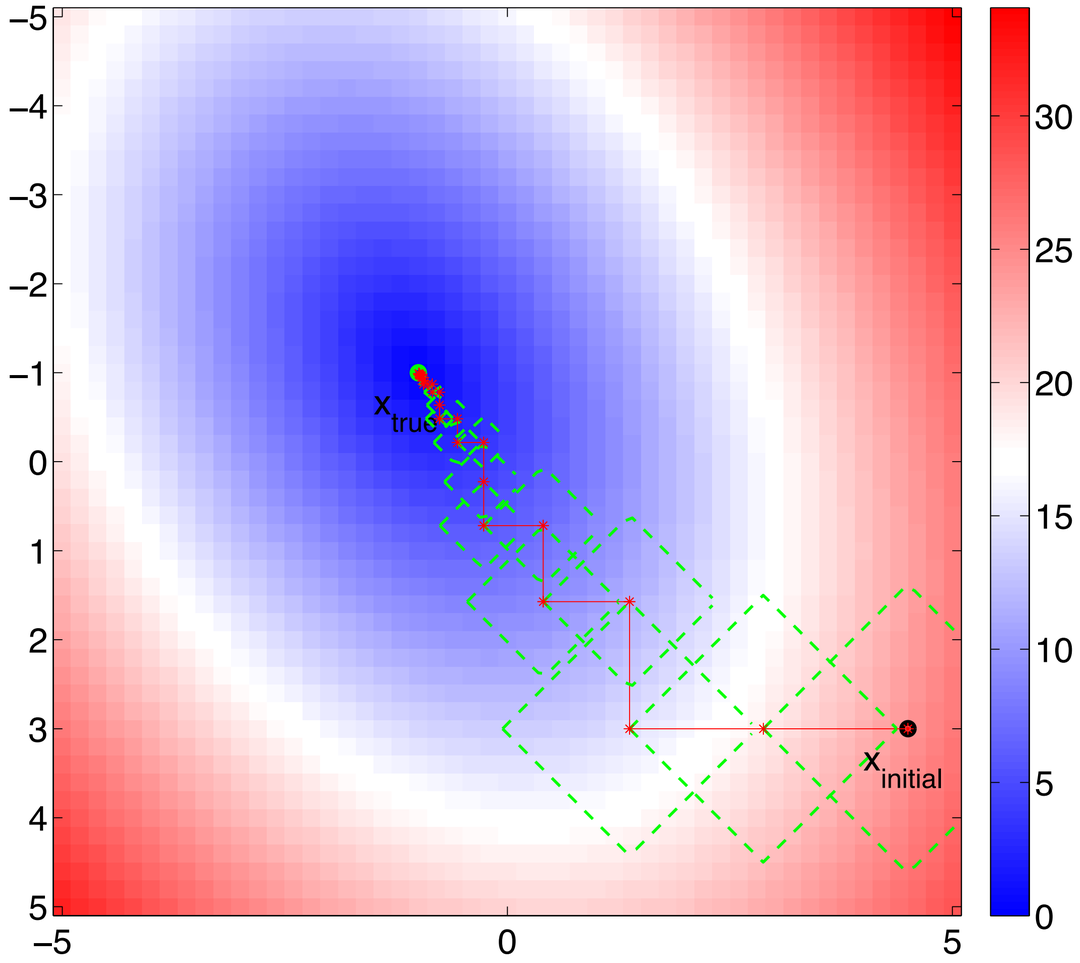
with wrong constraint

[J. Burke, '92]

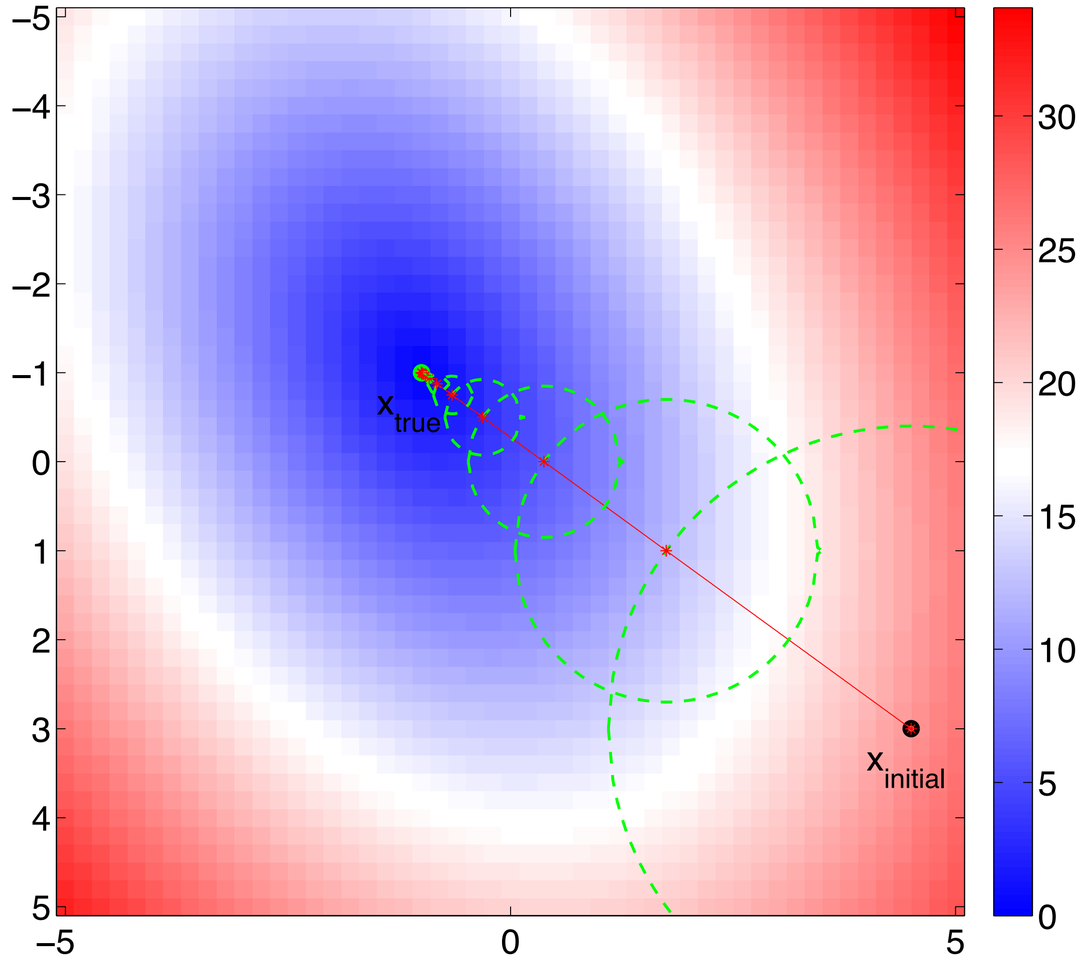
Modified Gauss-Newton

$$\min_{\mathbf{m}} \Phi(\mathbf{m}) := \left\{ \frac{1}{2} \|\mathbf{d} - \mathcal{F}(\mathbf{m})\|_2^2 \right\}$$

$$\delta\mathbf{m} = \mathbf{S}^{\mathrm{H}} \arg\min_{\delta\mathbf{x}} \|\delta\mathbf{d} - \mathbf{J}(\mathbf{m}_k)\mathbf{S}^{\mathrm{H}}\delta\mathbf{x}\|_2^2 \quad \text{subject to} \quad \|\delta\mathbf{x}\|_1 \leq \tau$$



with ℓ_1 constraint



with ℓ_2 constraint

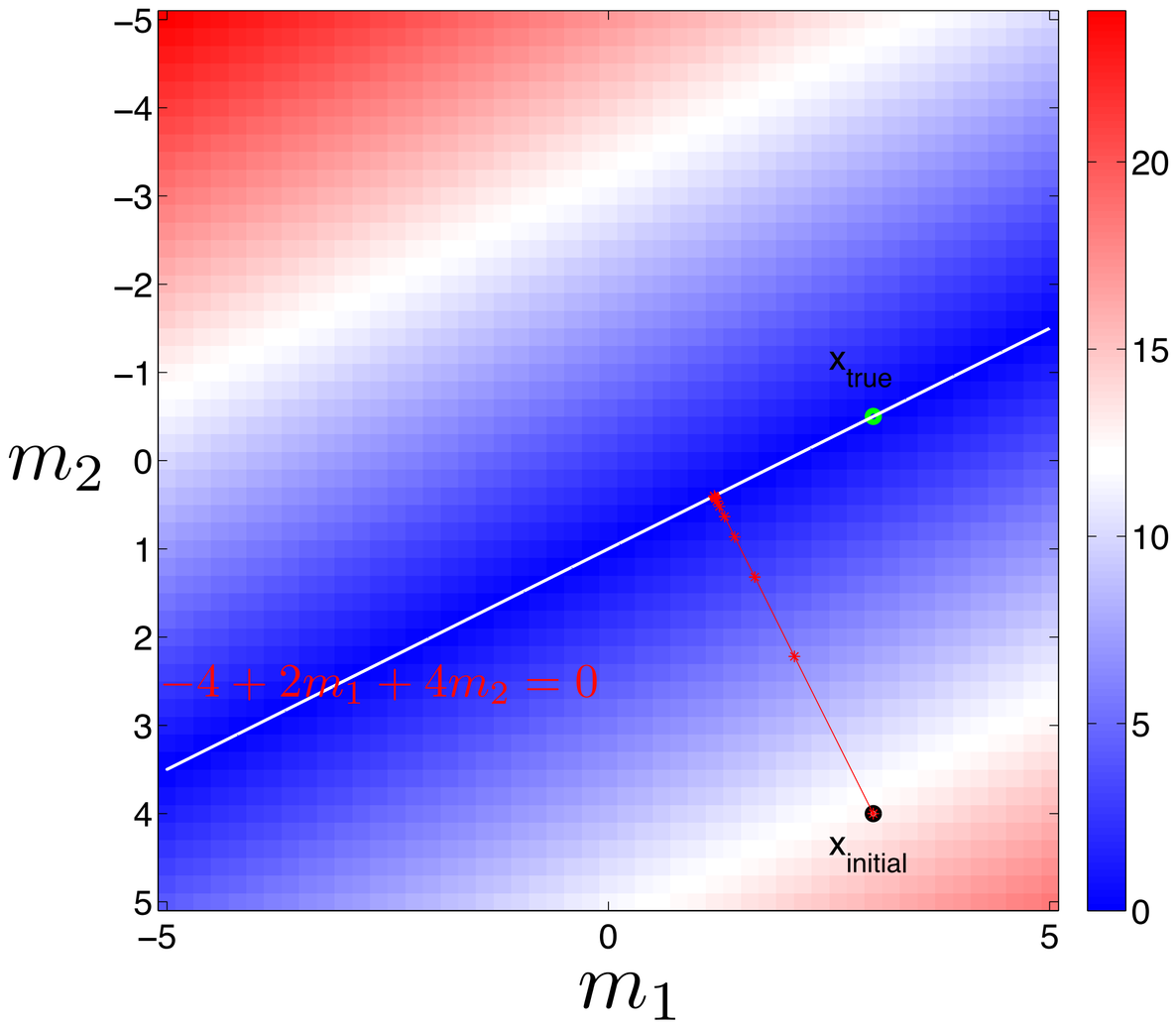
Linear example
w/ multiple solutions

$$\Phi(\mathbf{m}) := \left\{ \frac{1}{2} \|\mathbf{d} - \mathbf{A}\mathbf{m}\|_2^2 \right\}$$

$$\mathbf{A} = \begin{bmatrix} 2 & 4 \end{bmatrix}$$

$$\mathbf{d} = -4$$

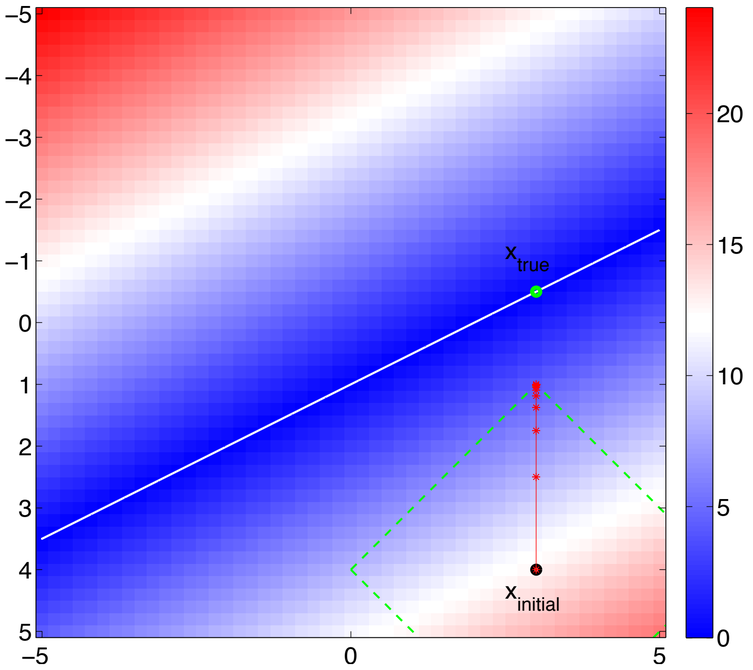
$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$



GN with unconstrained objective

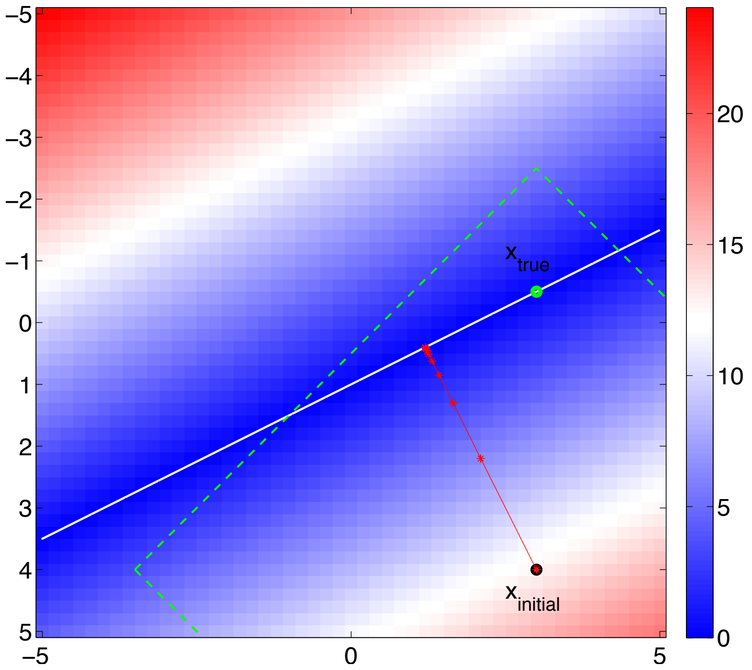
GN – w/ sparse constrained objective function

$$\min_{\mathbf{x}} \Phi(\mathbf{x}) := \left\{ \frac{1}{2} \|\mathbf{d} - \boldsymbol{\mathcal{F}}(\mathbf{S}^{\mathrm{H}}\mathbf{x})\|_2^2 \right\} \quad \text{subject to} \quad \|\mathbf{x} - \mathbf{x}_0\|_1 \leq \tau$$



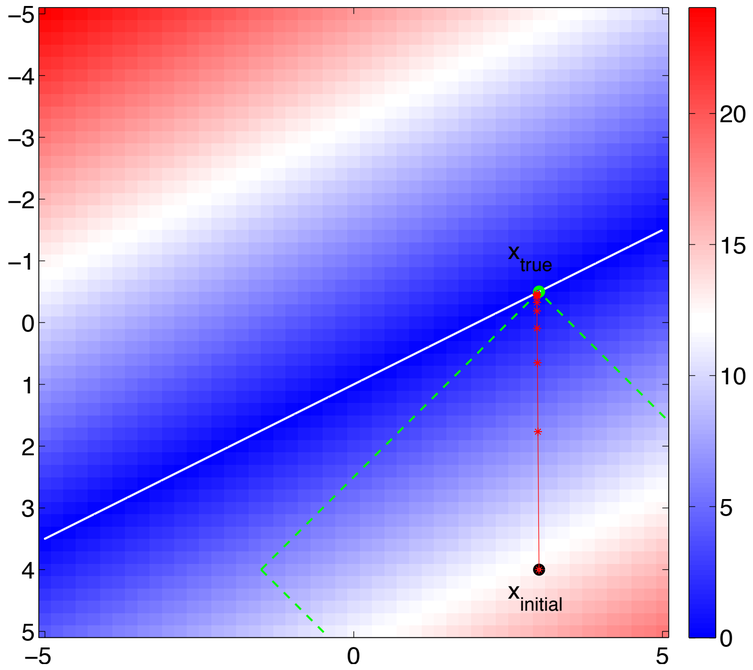
with wrong constraint

$\tau < \tau_{true}$



with wrong constraint

$\tau > \tau_{true}$

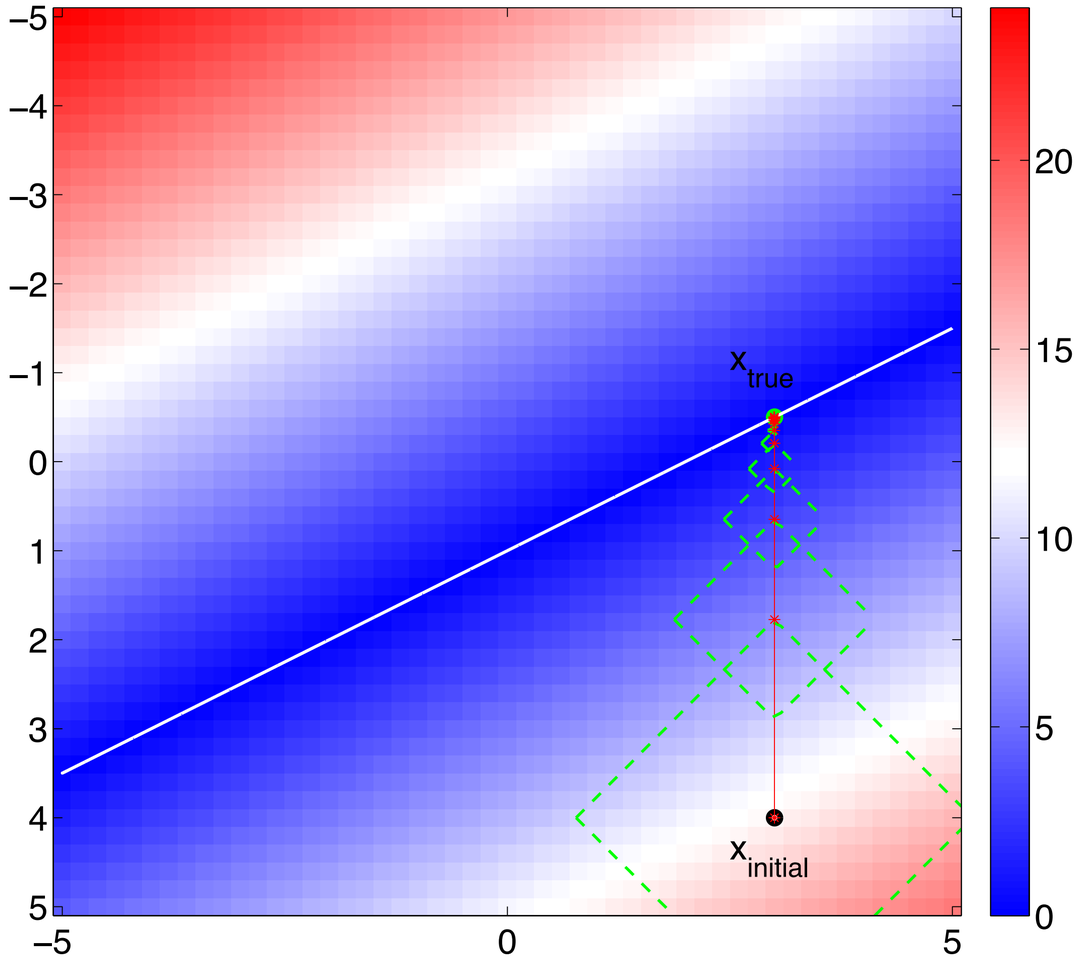


with right constraint

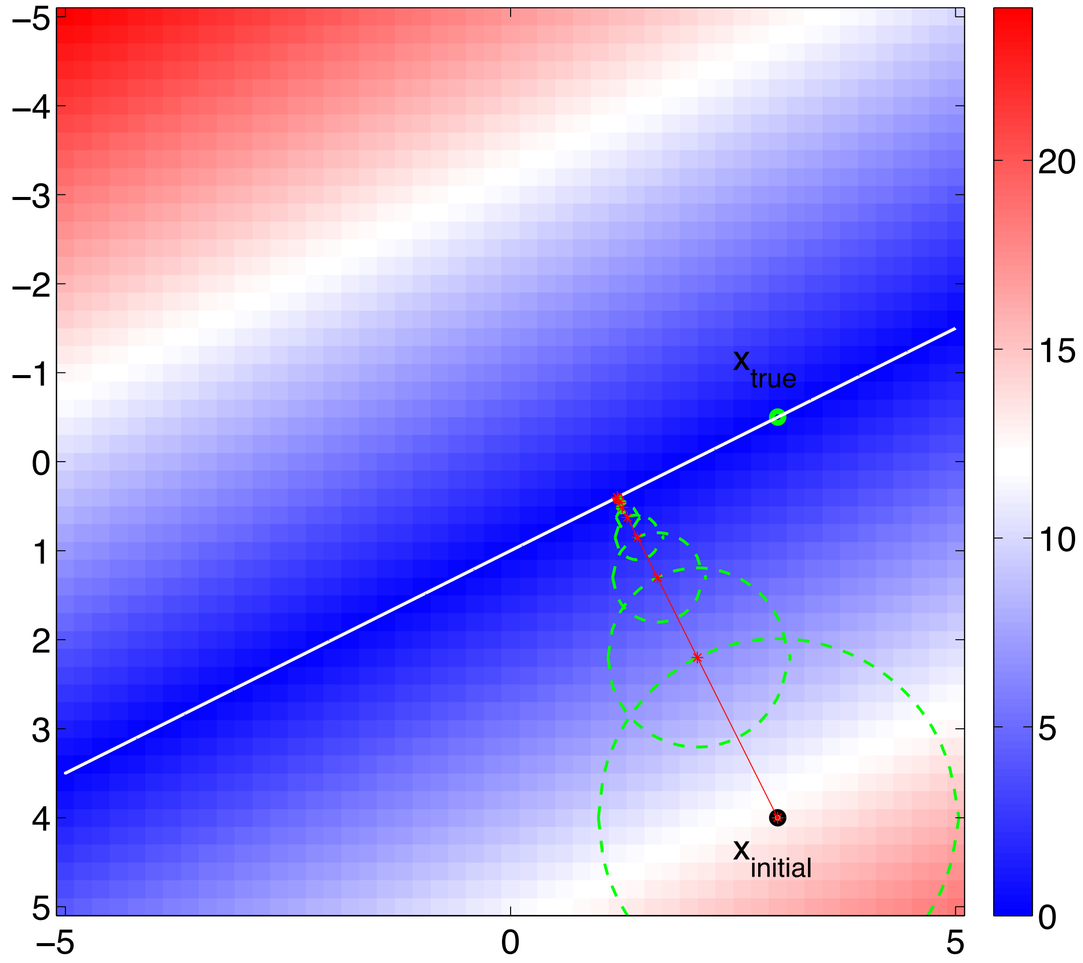
$\tau = \tau_{true}$

Modified Gauss-Newton

$$\min_{\mathbf{m}} \Phi(\mathbf{m}) := \left\{ \frac{1}{2} \|\mathbf{d} - \mathcal{F}(\mathbf{m})\|_2^2 \right\} \qquad \delta\mathbf{m} = \mathbf{S}^{\mathrm{H}} \arg\min_{\delta\mathbf{x}} \|\delta\mathbf{d} - \mathbf{J}(\mathbf{m}_k)\mathbf{S}^{\mathrm{H}}\delta\mathbf{x}\|_2^2 \quad \text{subject to} \quad \|\delta\mathbf{x}\|_1 \leq \tau$$



with ℓ_1 constraint



with ℓ_2 constraint

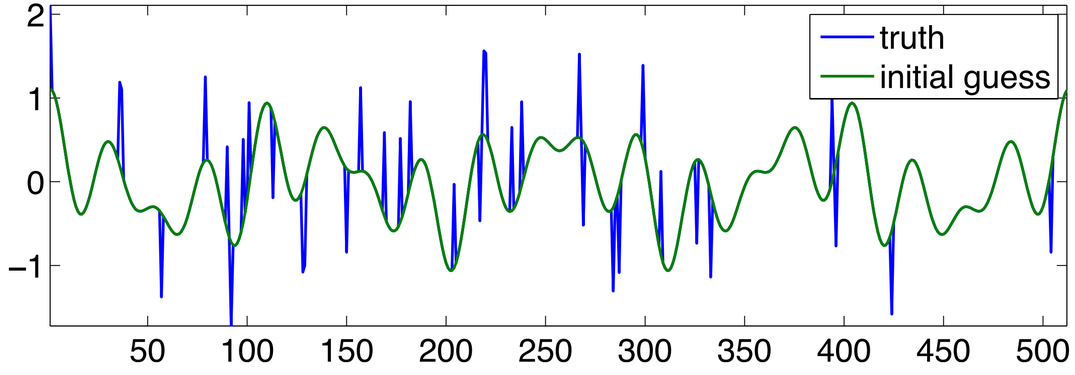
Phase retrieval problem

Objective function:

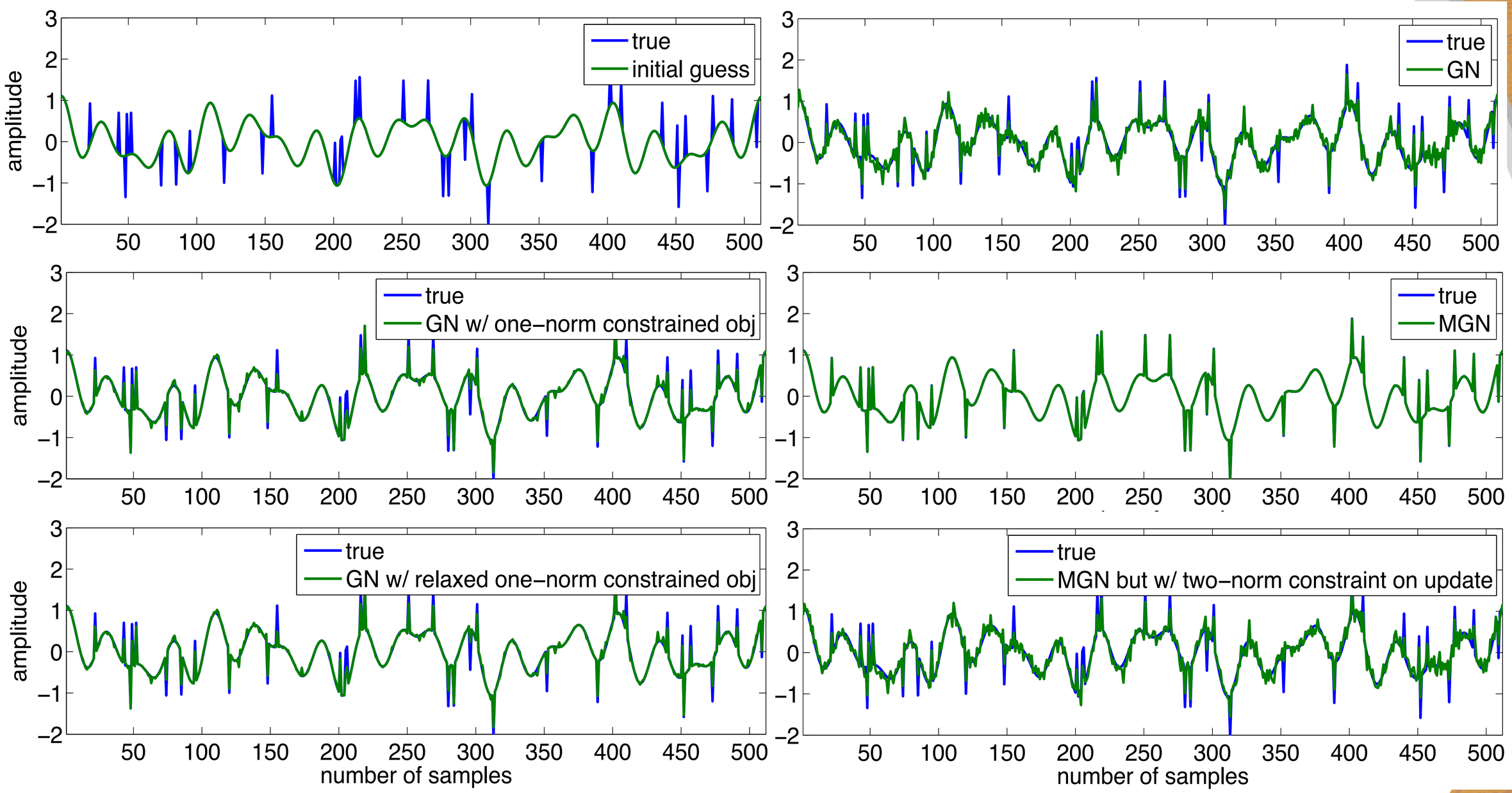
$$\Phi(\mathbf{x}) := \left\{ \frac{1}{2} \|\mathbf{d} - \mathrm{diag}(\mathbf{A}\mathbf{x})(\mathbf{A}\mathbf{x})\|_2^2 \right\}$$

$$\mathbf{d} = \mathrm{diag}(\mathbf{A}\mathbf{x}_{true})(\mathbf{A}\mathbf{x}_{true})$$

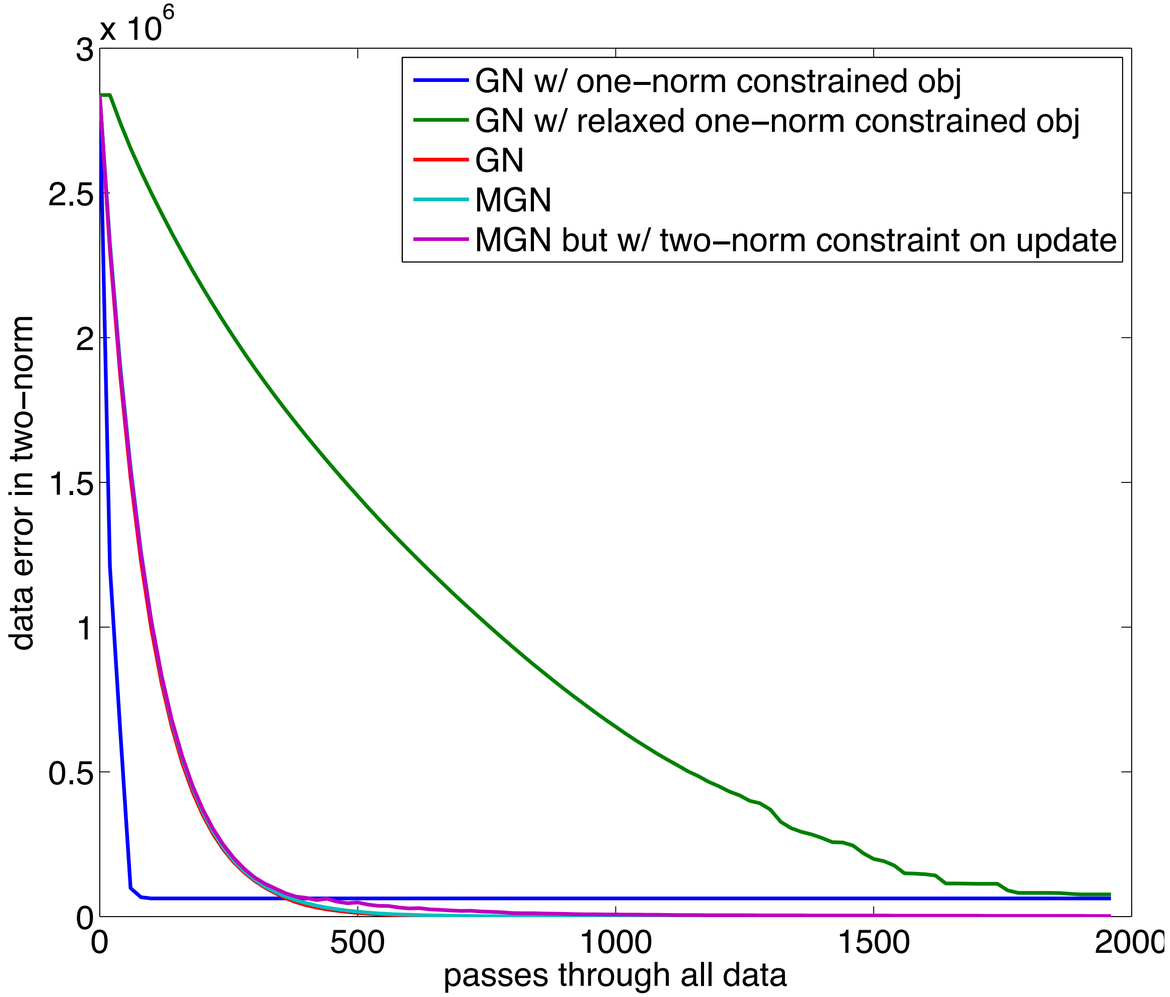
$$\begin{aligned} \mathbf{A} &: 400 \times 512 \text{ matrix} \\ \mathbf{d} &: 400 \times 1 \text{ vector} \\ \mathbf{x} &: 512 \times 1 \text{ unknown vector} \end{aligned}$$



Results

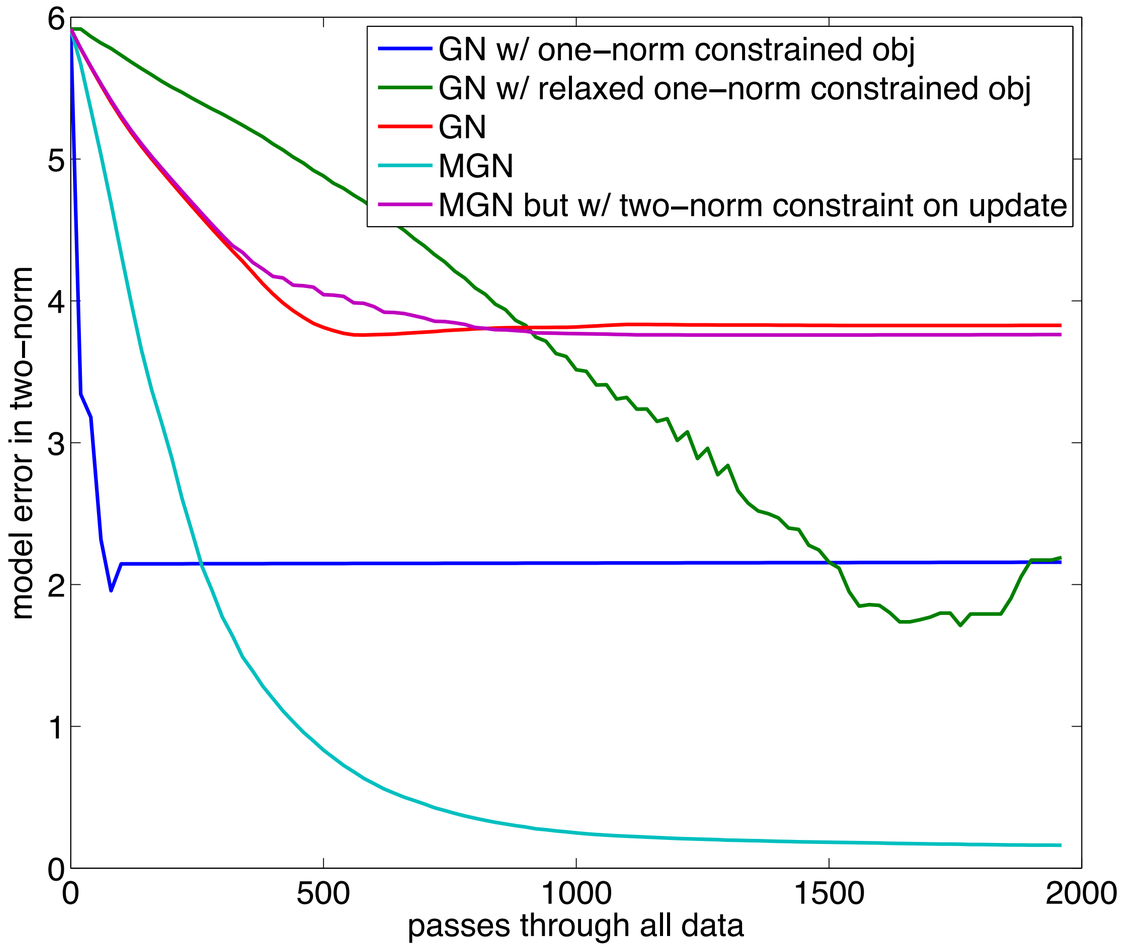

true
initial guess
true
GN
true
GN w/ one−norm constrained obj
true
MGN
true
GN w/ relaxed one−norm constrained obj
true
MGN but w/ two−norm constraint on update
amplitude
number of samples

Convergence



objective

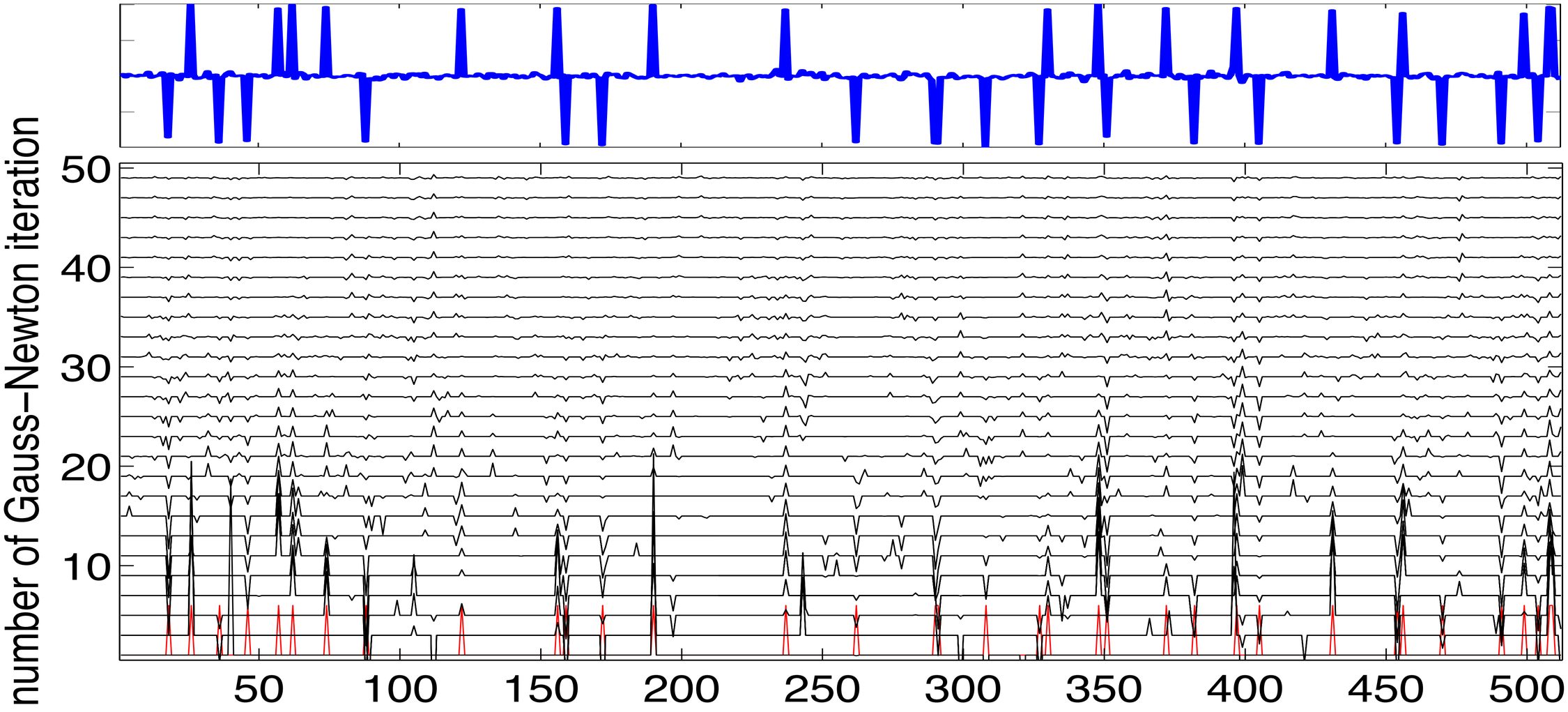
$$\frac{1}{2}\|\mathbf{d} - \mathrm{diag}(\mathbf{A}\mathbf{x}_k)(\mathbf{A}\mathbf{x}_k)\|_2^2$$



relative model error

$$\|\mathbf{x}_k - \mathbf{x}_{true}\|_2$$

Modified Gauss-Newton updates



red is the position of sparse support

Observations

Modified Gauss-Newton:

- can find the solution as other methods with unconstrained objective for convex problems with unique solution
- can find a solution with sparse perturbation of the initial guess for problems with multiple solutions, if updates share the same support.

Chevron 2012-2013 benchmark (blind test)

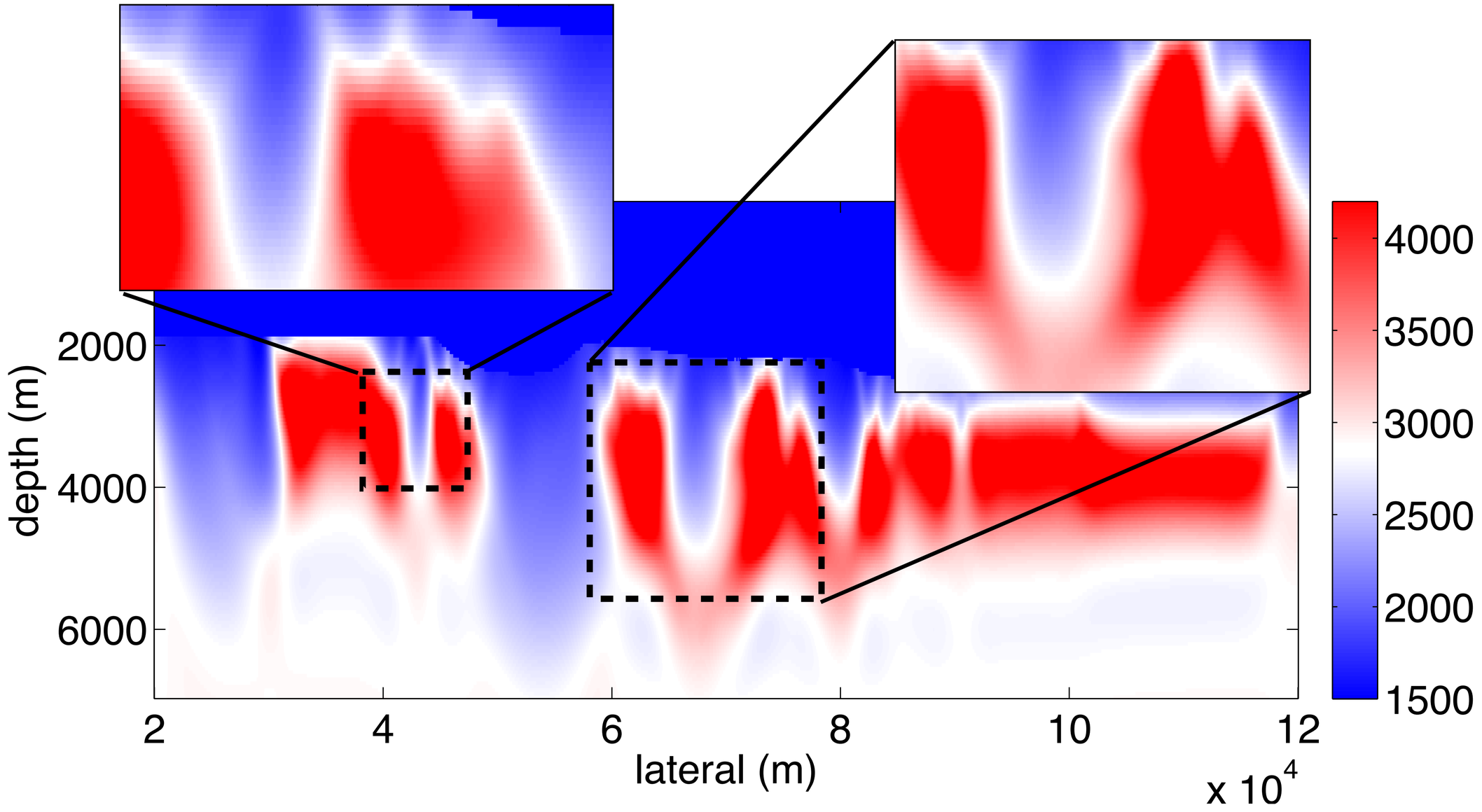
- 3201 shots with interval 25 m
- 801 receivers with interval 25 m, yielding 20km offset
- record time 14s, sample rate 4ms
- free surface
- isotropic elastic

Inversion setting

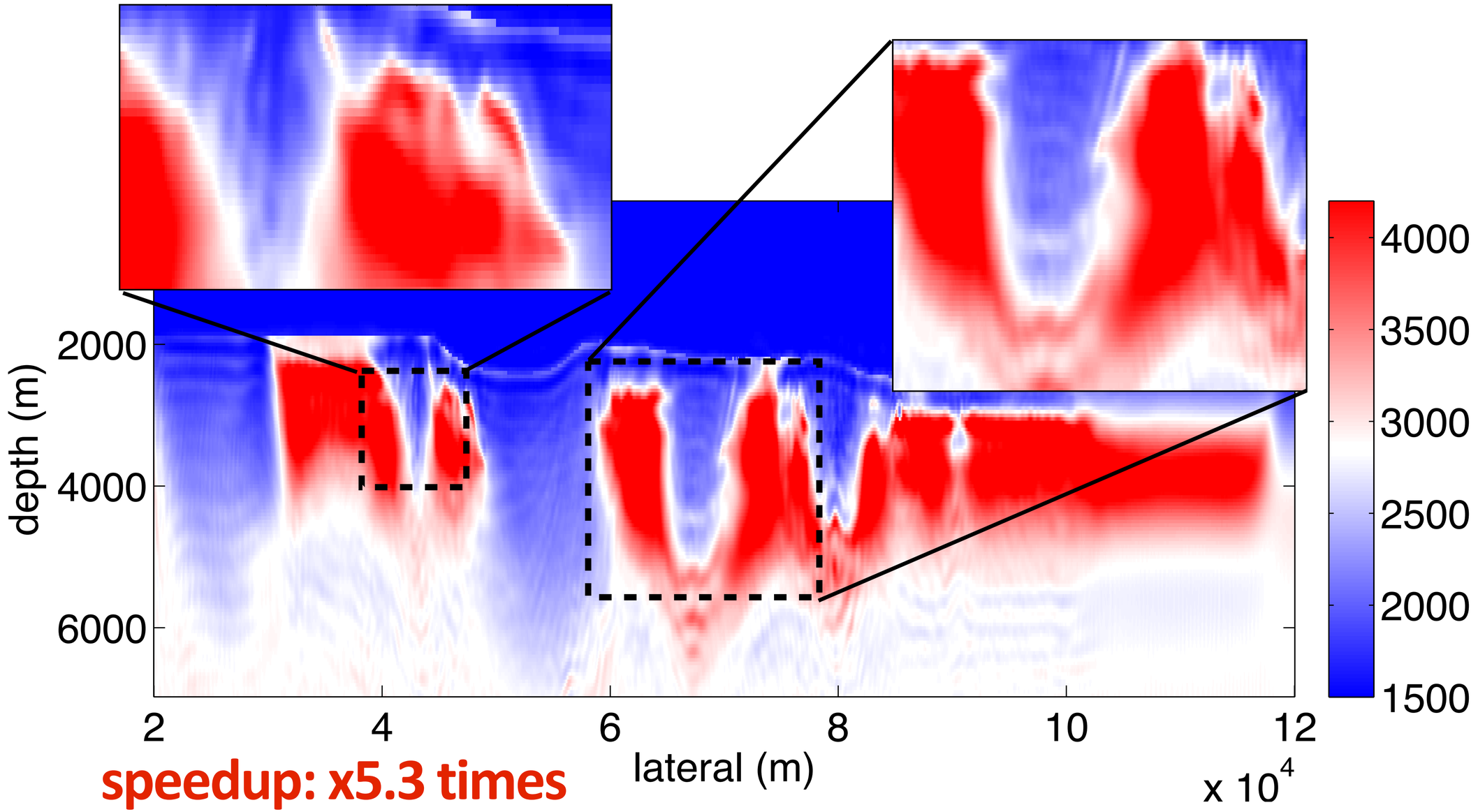
- 7 frequency bands (2-5Hz)
- 6 GN iterations per frequency band, with 600 randomly selected shots

Initial model
ray base tomography

Andrew J. Calvert

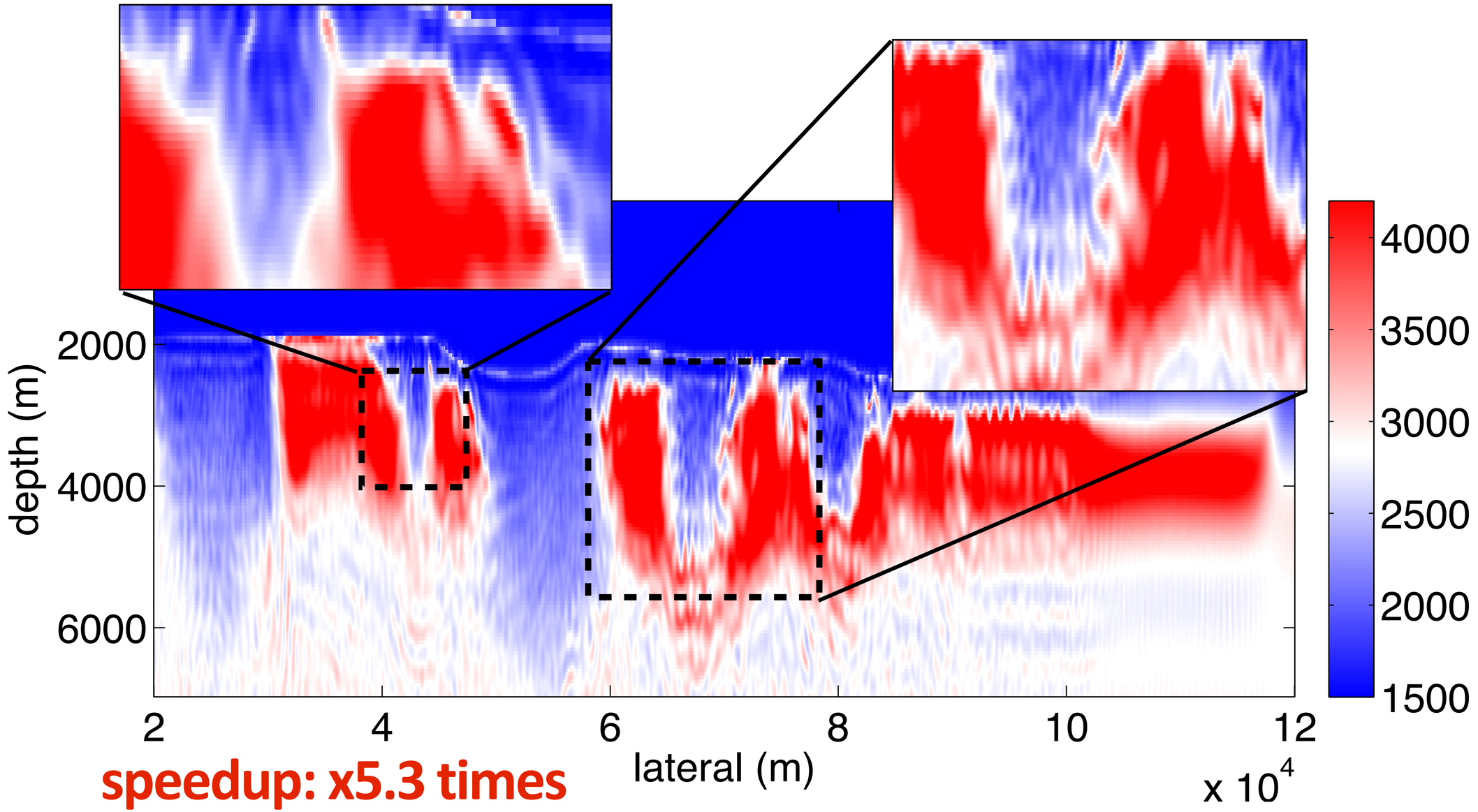


Modified Gauss-Newton inversion result



speedup: x5.3 times

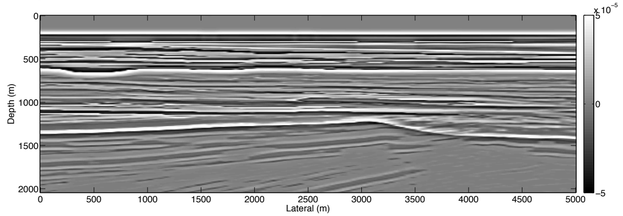
Gauss-Newton inversion result



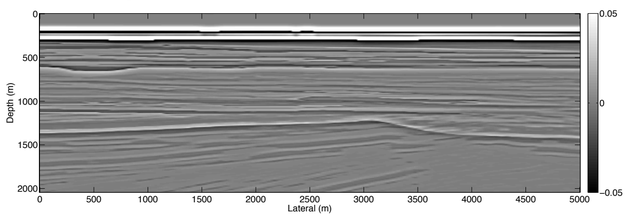
speedup: x5.3 times

Future work

- multi-parameter inversion with joint-sparsity

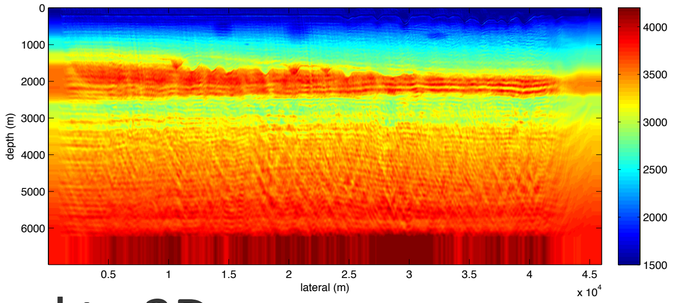


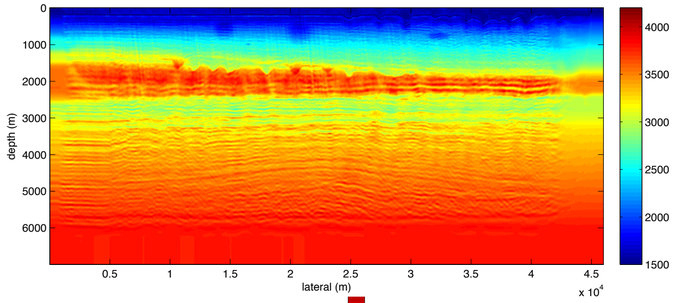
velocity



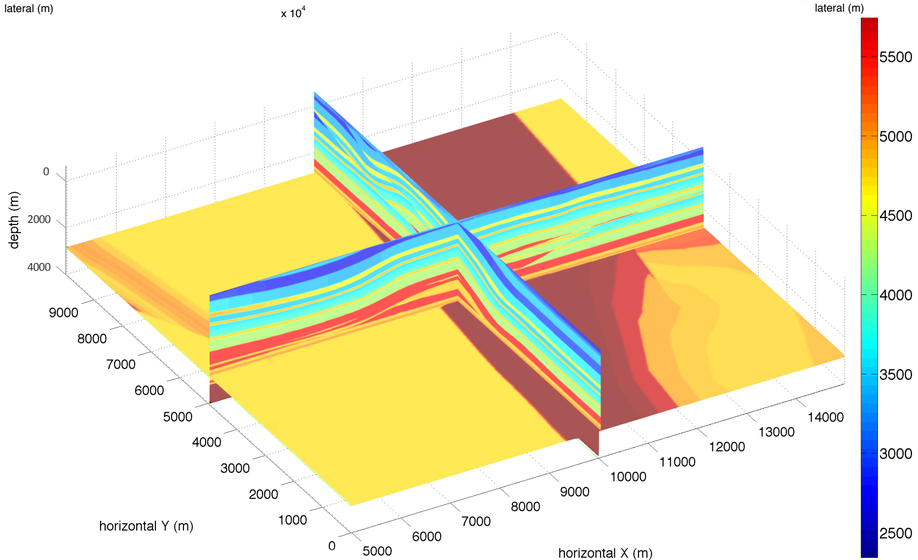
density

- time domain approach





- extend to 3D



Conclusions

- Geological structures are sparse in the curvelet domain
- Computational cost can be reduced significantly by using randomized dimensionality reduction
- Sparsity constraints on the objective do not necessarily generate solutions with sparse cumulative updates w.r.t. starting model
- Modified Gauss-Newton method yields solutions with sparse cumulative, if most updates share the same sparsity support

Acknowledgements

Thank you for your attention !

https://www.slim.eos.ubc.ca/

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, BGP, CGG, Chevron, ConocoPhillips, DownUnder Geosolutions, Hess, Petrobras, PGS, Sub Salt Solutions, WesternGeco, and Woodside.