SINBAD Consortium 2010

FWI with Compressive Updates

Aleksandr Aravkin, Felix Herrmann, Tristan van Leeuwen, Xiang Li, James Burke

SLIM
University of British Columbia

SLIM

Full Waveform Inversion

- **The Full Waveform Inversion (FWI) problem is to find solutions to the Helmholtz PDE that match data from source experiments on the surface**
- **Problems are typically very large: trillions of variables and terabytes of data.**
- **Typically formulated as a Nonlinear Least Squares (NLLS) problem:**

$$\min_{\mathbf{m}} \; \left\{ f(\mathbf{m}) := \|\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m}; \mathbf{Q}]\|_F^2 \right\}$$

$$\begin{aligned} \mathbf{D} &:= \text{data} \\ \mathbf{m} &:= \text{model parameters (speed or slowness squared)} \\ \mathbf{Q} &:= \text{multiple source experiments} \\ \boldsymbol{\mathcal{F}} &:= \text{solution operator of Helmholtz eqn. with absorbing boundary} \end{aligned}$$

Difficulties with NLLS

- The size of FWI requires algorithms that reduce computation time, e.g. by working on reduced data volumes.

- In addition to size, there are problems with the NLLS formulation:
  1) Local minima (missing low frequency information, model misspecification, cycle skipping)

  2) Insufficient data (multiple models fit the same data)

  3) Inadequate data (data not in the range of modeling operator)

  4) Sensitivity - small changes in data yield large changes in the model estimate

- Both types of issues need to be addressed.

[Virieux ’09; Symes ’09; Symes ’08]

Stochastic Optimization

- **Stochastic optimization is a promising approach for FWI.**

- **Suppose** W **is a random matrix with** $E[WW^{\mathrm{T}}] = I$ **:**

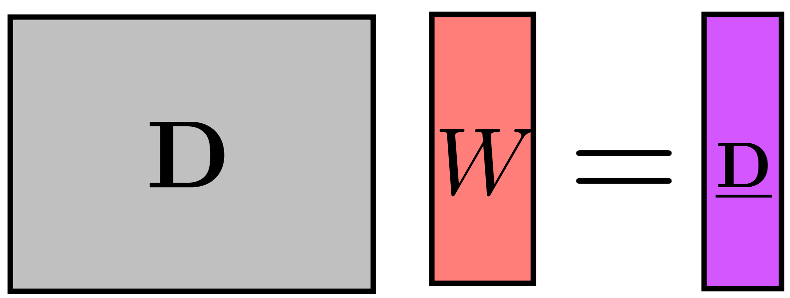
$$\|A\|_F^2 = \mathrm{trace}(A^{\mathrm{T}}A) = E\{\mathrm{trace}(A^{\mathrm{T}}AWW^{\mathrm{T}})\}$$
$$= E\{\mathrm{trace}(W^{\mathrm{T}}A^{\mathrm{T}}AW)\} = E\|AW\|_F^2$$

- **With above identity, FWI can be viewed as stochastic optimization problem.**

$$\begin{aligned} f(\mathbf{m}) &= E_W\|\underline{\mathbf{D}} - \boldsymbol{\mathcal{F}}[\mathbf{m};\underline{\mathbf{Q}}]\|_F^2 \\ \underline{\mathbf{D}} &:= \mathbf{D}W \\ \underline{\mathbf{Q}} &:= \mathbf{Q}W \end{aligned}$$

Stochastic Optimization

- Stochastic optimization provides a dimensionality reduction technique, since randomization (simultaneous shots) compress data and sources:



[Haber '10]

$$\underline{f}(\mathbf{m}) := \|\underline{\mathbf{D}} - \boldsymbol{\mathcal{F}}[\mathbf{m};\underline{\mathbf{Q}}]\|_F^2$$

- SAA approach: replace f by $\underline{f}$ with large W (many shots)
- SA approach: use descent directions of $\underline{f}$ with small W (few shots) to iteratively minimize f [Shapiro '03 , Shapiro '05]

Gauss-Newton Method

- **Objective:** $\underline{f}(\mathbf{m}) := \|\underline{\mathbf{D}} - \mathcal{F}[\mathbf{m}; \underline{\mathbf{Q}}]\|_F^2$
- **Iterative algorithm:** $\mathbf{m}^{\nu+1} = \mathbf{m}^{\nu} + \gamma_{\nu} \overline{\boldsymbol{\delta}\mathbf{m}}$
- **Direction** $\overline{\boldsymbol{\delta}\mathbf{m}}$ **solves** $\min_{\boldsymbol{\delta}\mathbf{m}} \|\underline{\mathbf{D}} - \mathcal{F}[\mathbf{m}^{\nu}; \underline{\mathbf{Q}}] - \nabla\mathcal{F}[\mathbf{m}^{\nu}; \underline{\mathbf{Q}}]\boldsymbol{\delta}\mathbf{m}\|_F^2$
- **The subproblem for** $\overline{\boldsymbol{\delta}\mathbf{m}}$ **is convex, and** $\overline{\boldsymbol{\delta}\mathbf{m}}$ **is a descent direction:**

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

Compressibility in Curvelets

- The Gauss-Newton subproblem can be seen as the Born scattering problem, where $\mathbf{m}^\nu$ is a background velocity and $\delta\mathbf{m}$ is a model perturbation.
- The gradient of FWI $\nabla \underline{f}(\mathbf{m}^\nu)$ can be interpreted as a perturbation wavefield scattered by missing heterogeneities in the staring model $\mathbf{m}^\nu$.
- Wavefields have been shown to have compressible representations in the Curvelet frame (coefficients decay exponentially fast). [Candes 2004]
- We exploit this idea by placing a Lasso (1-norm) constraint on the Gauss-Newton update representation in the Curvelet frame $\mathcal{C}$.

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**

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

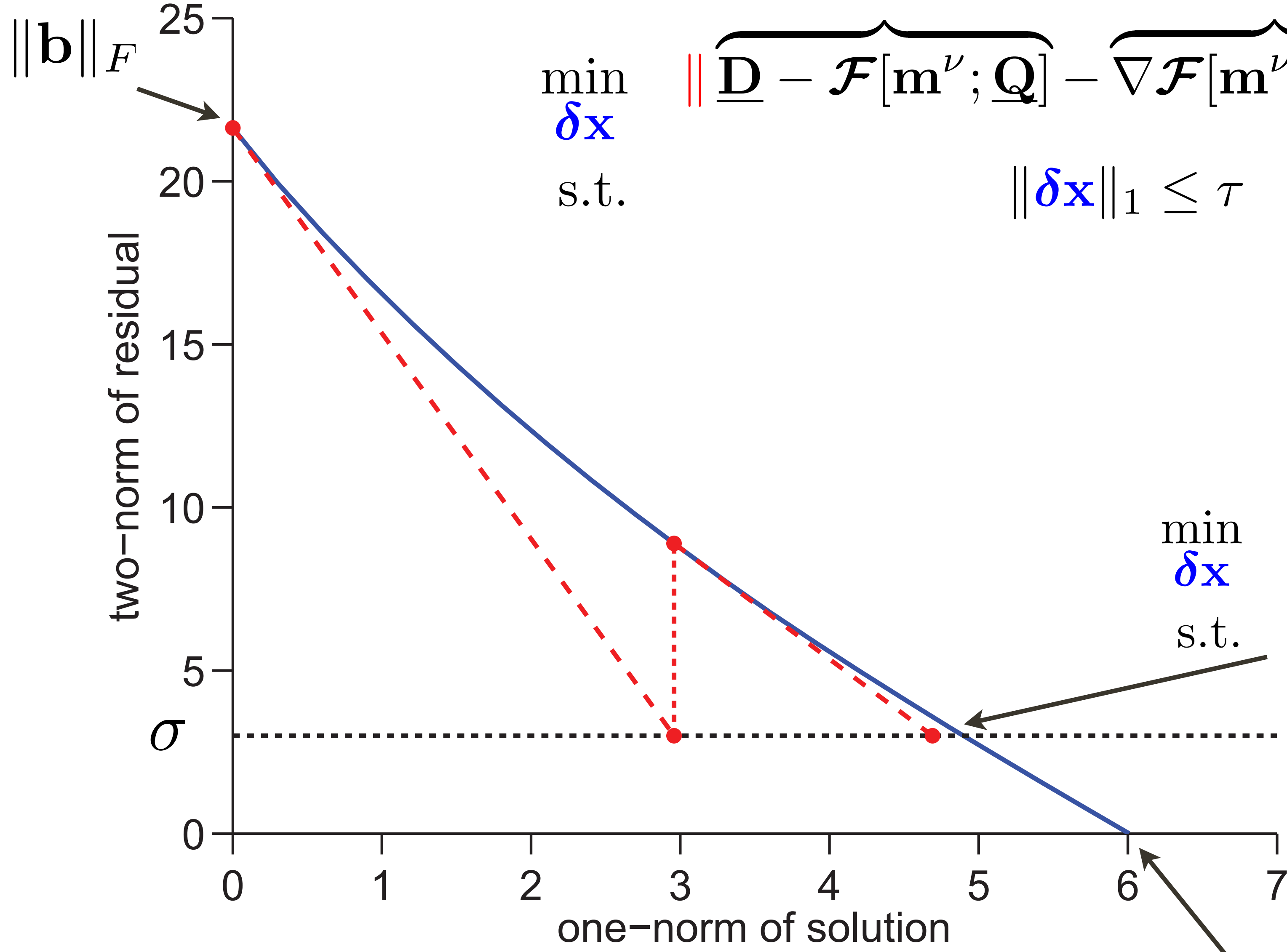
- **The subproblem for** $\overline{\boldsymbol{\delta}\mathbf{x}}$ **is convex, and** $\boldsymbol{\mathcal{C}}^*\overline{\boldsymbol{\delta}\mathbf{x}}$ **is a descent direction:**

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

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

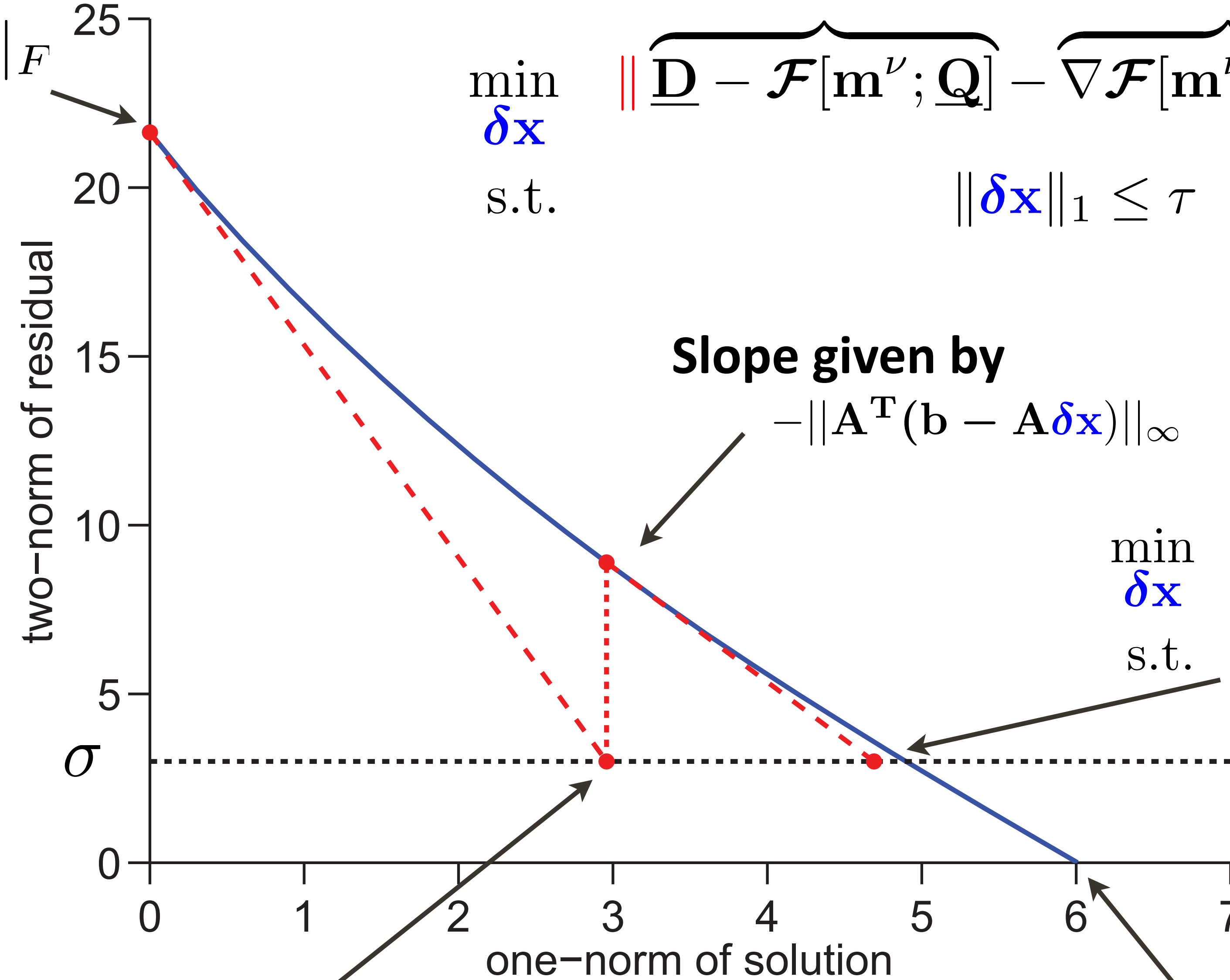
[van den Berg ’08]

Picking Lasso Parameter

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

Lasso

$\|\mathbf{b}\|_F$



Slope given by

$-\|\mathbf{A}^{\mathrm{T}}(\mathbf{b} - \mathbf{A}\delta\mathbf{x})\|_\infty$

$$\min_{\delta\mathbf{x}} \quad \|\delta\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{b} - \mathbf{A}\delta\mathbf{x}\|_F \leq \sigma$$

BPDN

$$\tau = \frac{\|\mathbf{b}\|_F - \sigma}{\|\mathbf{A}^{\mathrm{T}}\mathbf{b}\|_\infty}$$

$$\min_{\delta\mathbf{x}} \quad \|\delta\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{b} = \mathbf{A}\delta\mathbf{x}$$

BP

[van den Berg '08]

Modified GN with renewals

Algorithm 1: Modified Gauss-Newton with renewals

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

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

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

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

  for $i = 1 : N$ **do**

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

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

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

    $k \longleftarrow k + 1$

  end

end

SLIM

Example

Marmousi model:

- 128x384 with a mesh size of 24 meters
- 384 co-located shots and receivers with offset = 3 X depth
- 2.4s recording time for Marmousi

Explicit Time-harmonic Helmholtz solver

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

PDEs/Linearization

PDE solves for new method:

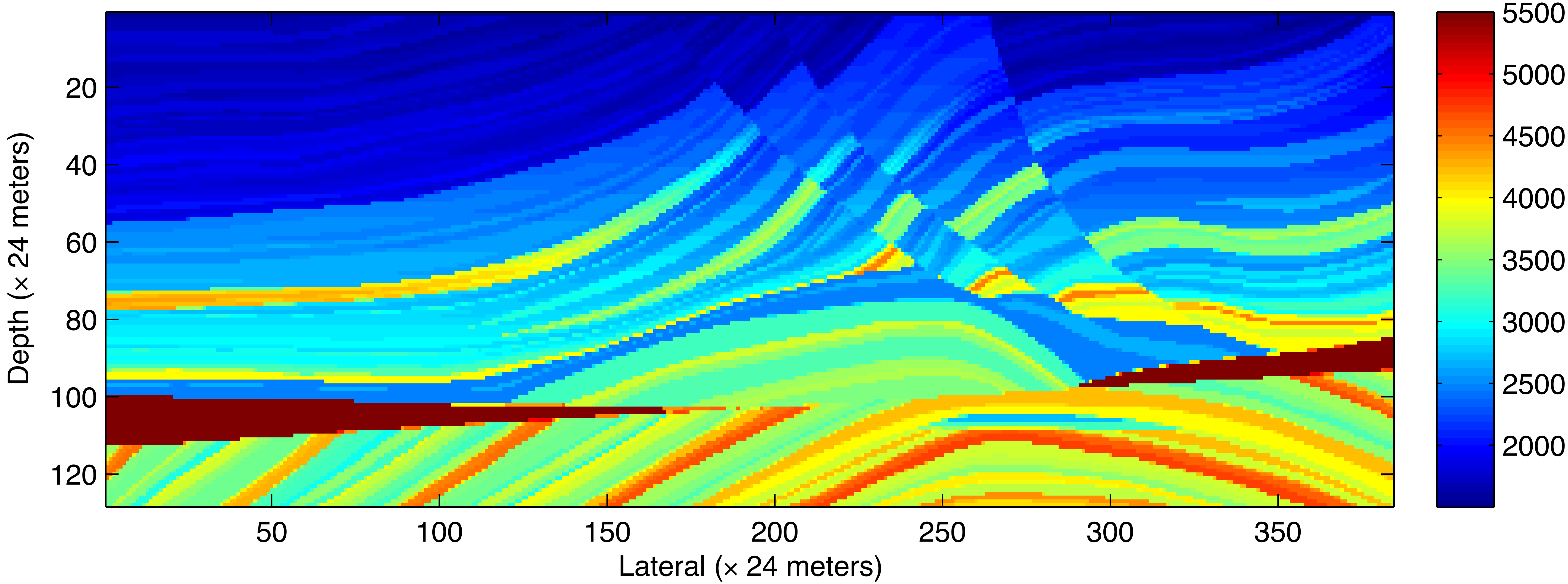
- 10 frequency bands, 10 frequencies in each
- 15 simultaneous shots
- 20 (average) iterations of SPGL1 solver
- 10 x 15 x 20 = 3000 PDE solves.

PDE solves for full Gauss-Newton subproblem:

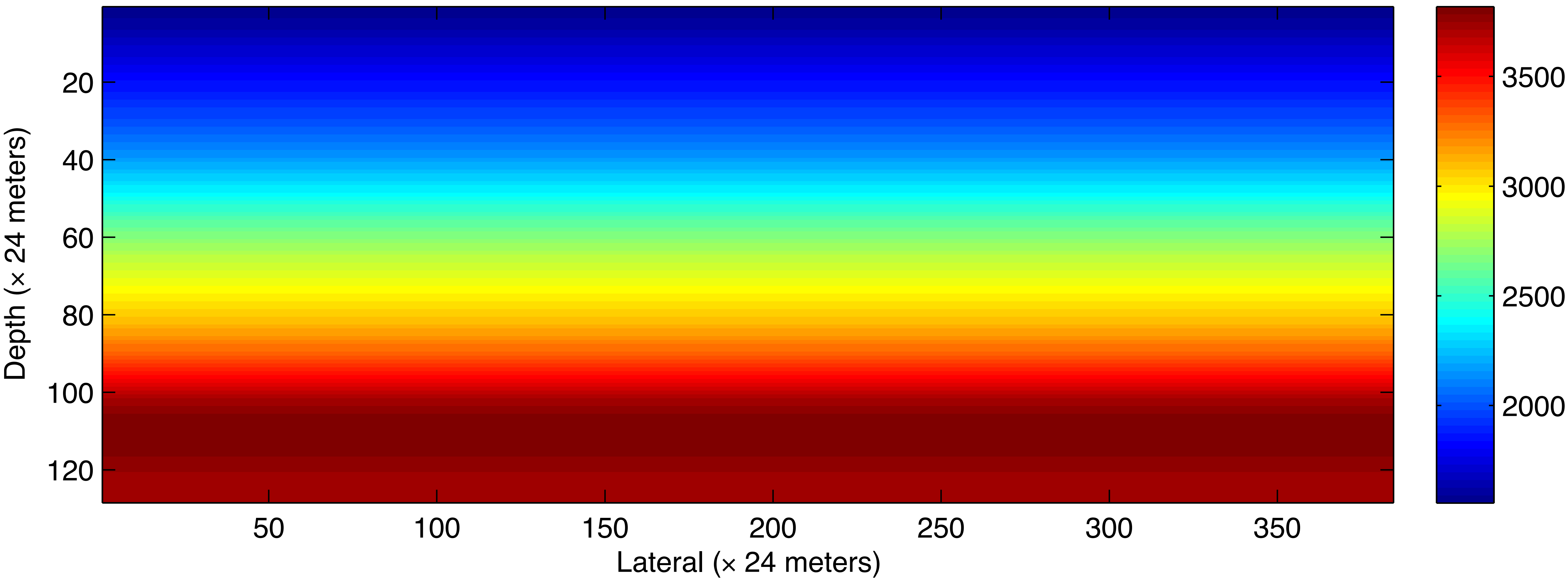
- 100 (freq) x 384 (shots) = 38400 PDE solves.

Speed-up: 38400/3000 = 12.8.

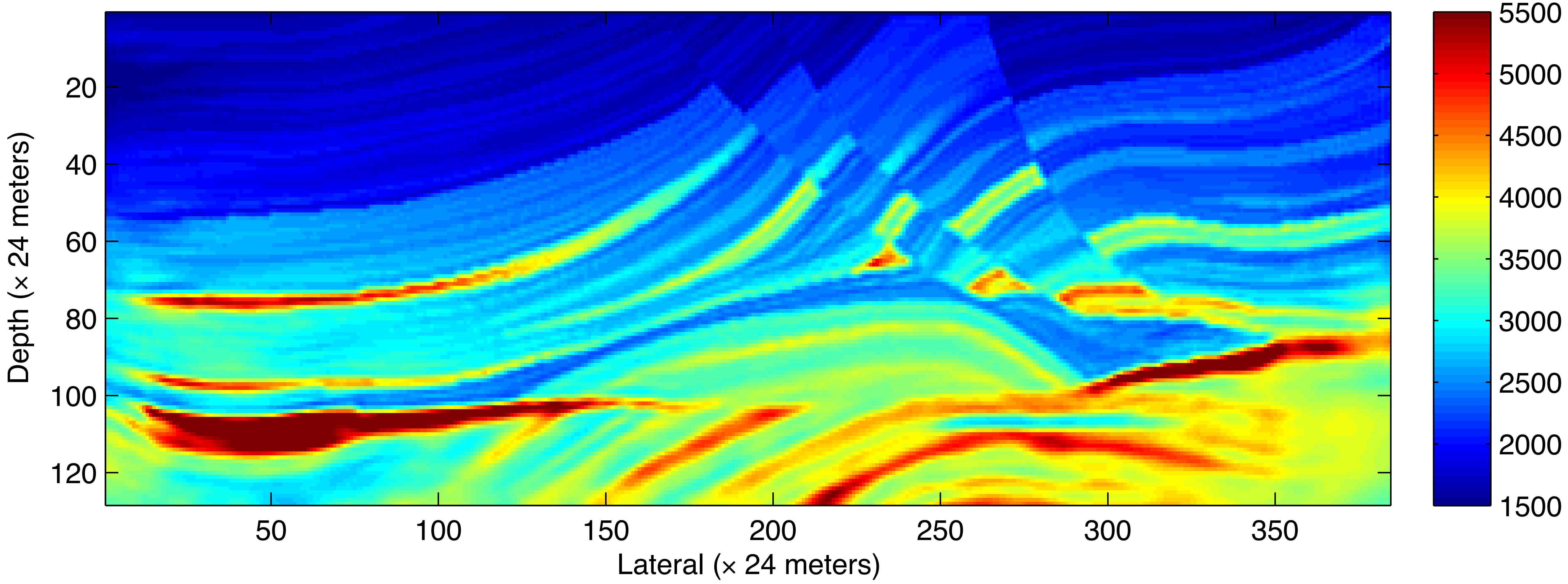
True model


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

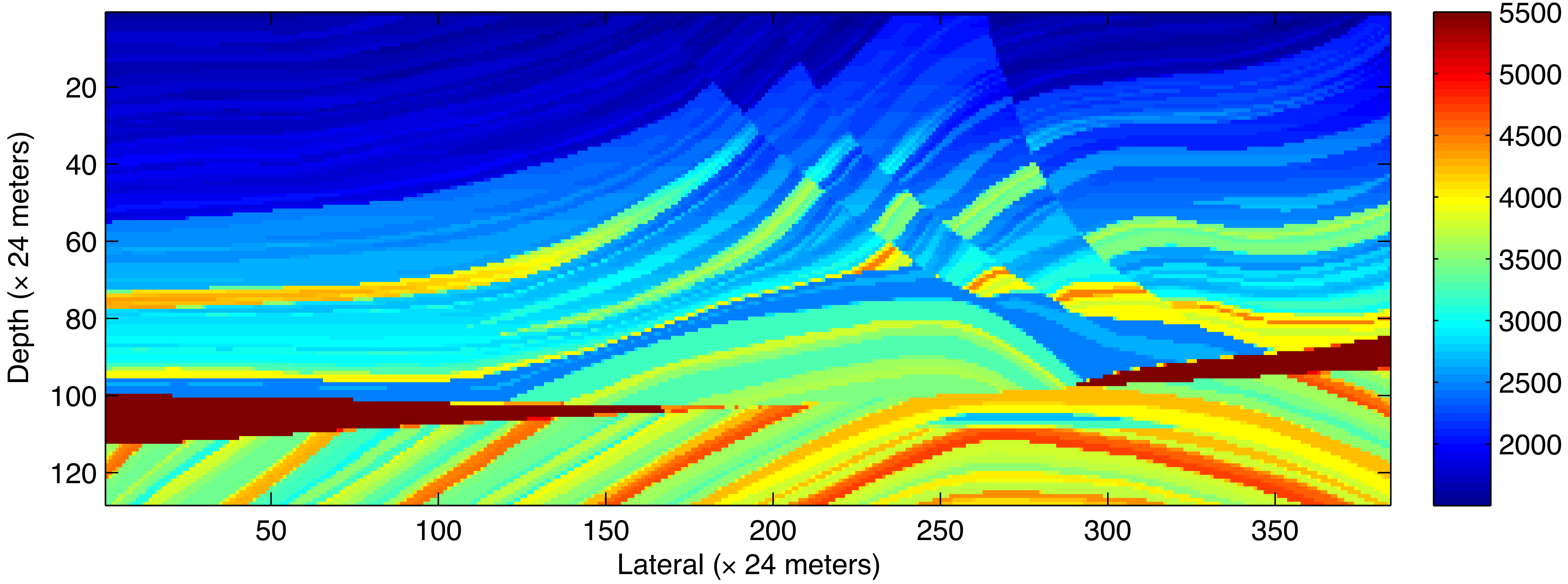
Initial model



Inverted model



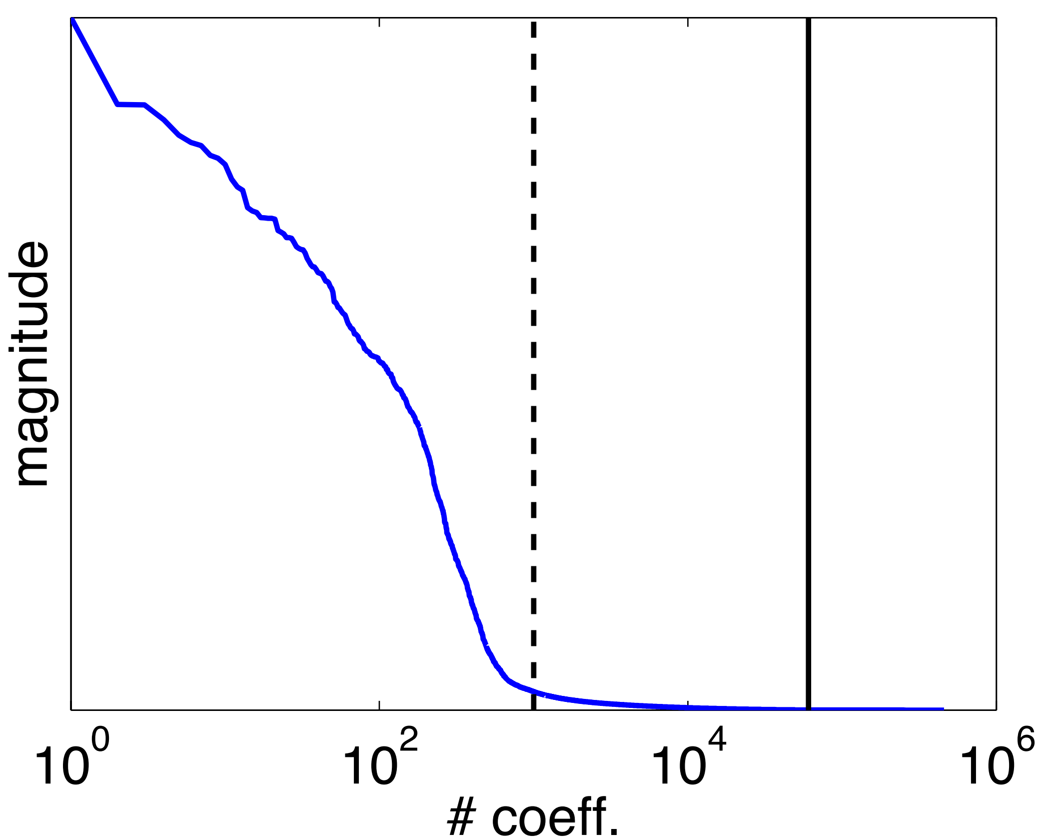
True model

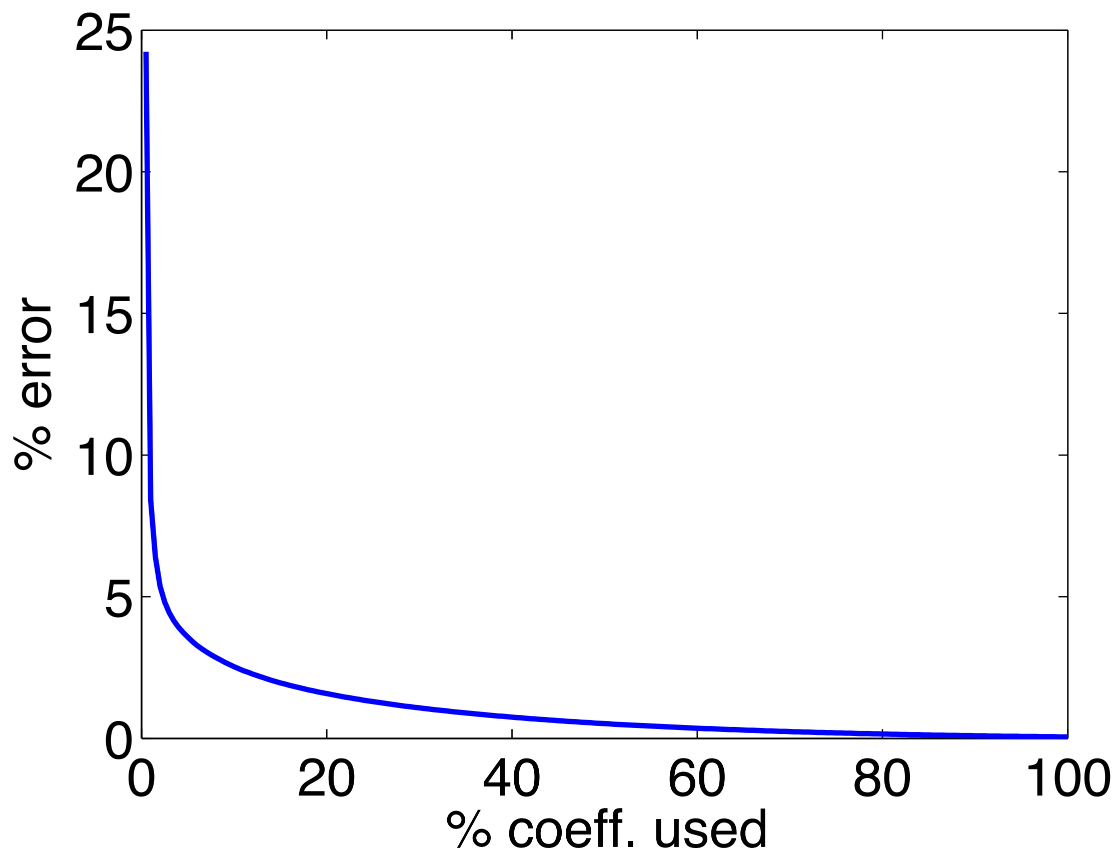


Depth (× 24 meters)
20
40
60
80
100
120
50
100
150
200
250
300
350
Lateral (× 24 meters)
5500
5000
4500
4000
3500
3000
2500
2000

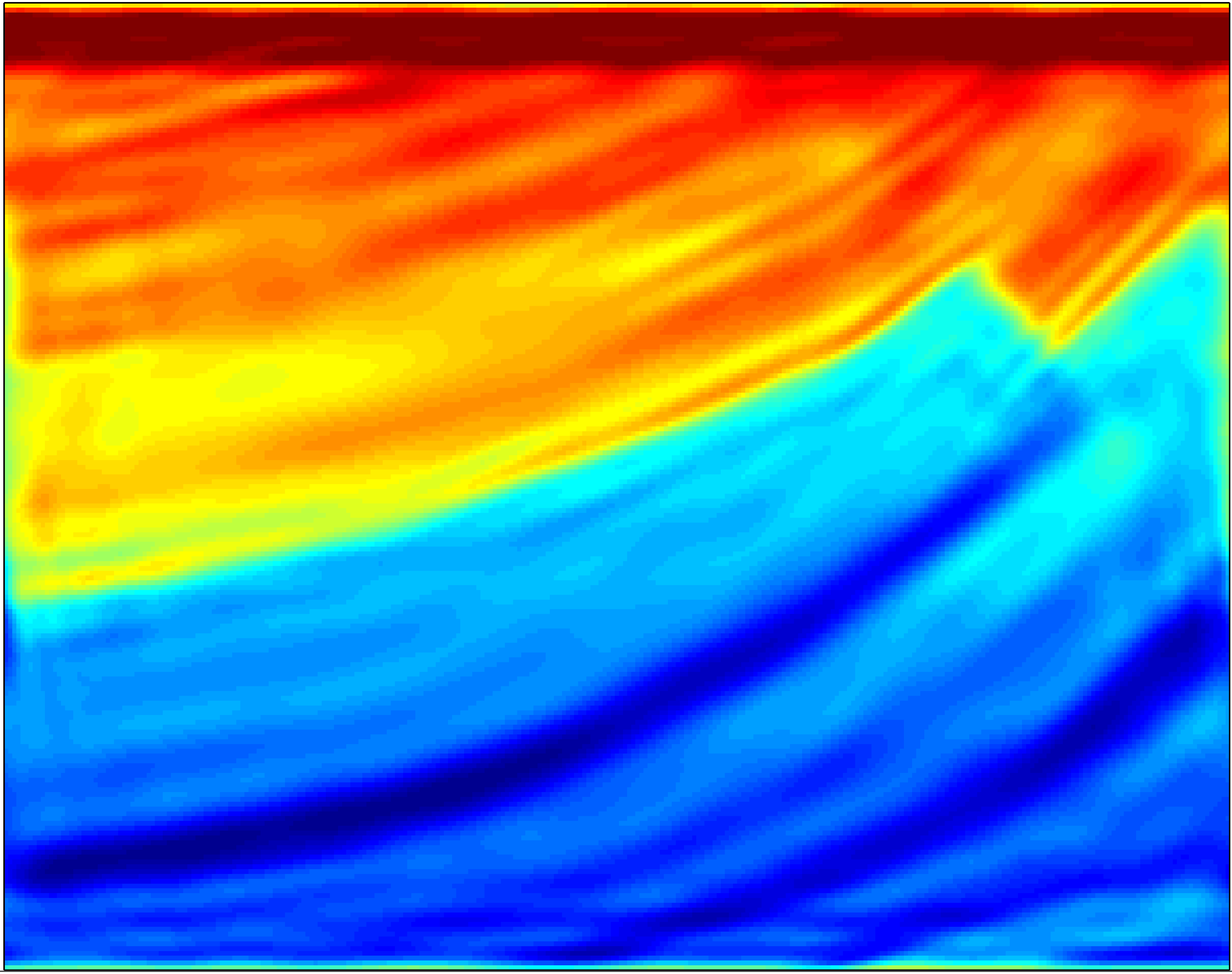
Compressibility in Curvelets

- Velocity models are also compressible in Curvelets. $\mathbf{m} = \mathcal{C}^*\mathbf{x}$
- Geophysical images are layered, and may me modeled as objects with edges. Curvelets provide sparse representations for such images. [Candes ’00]

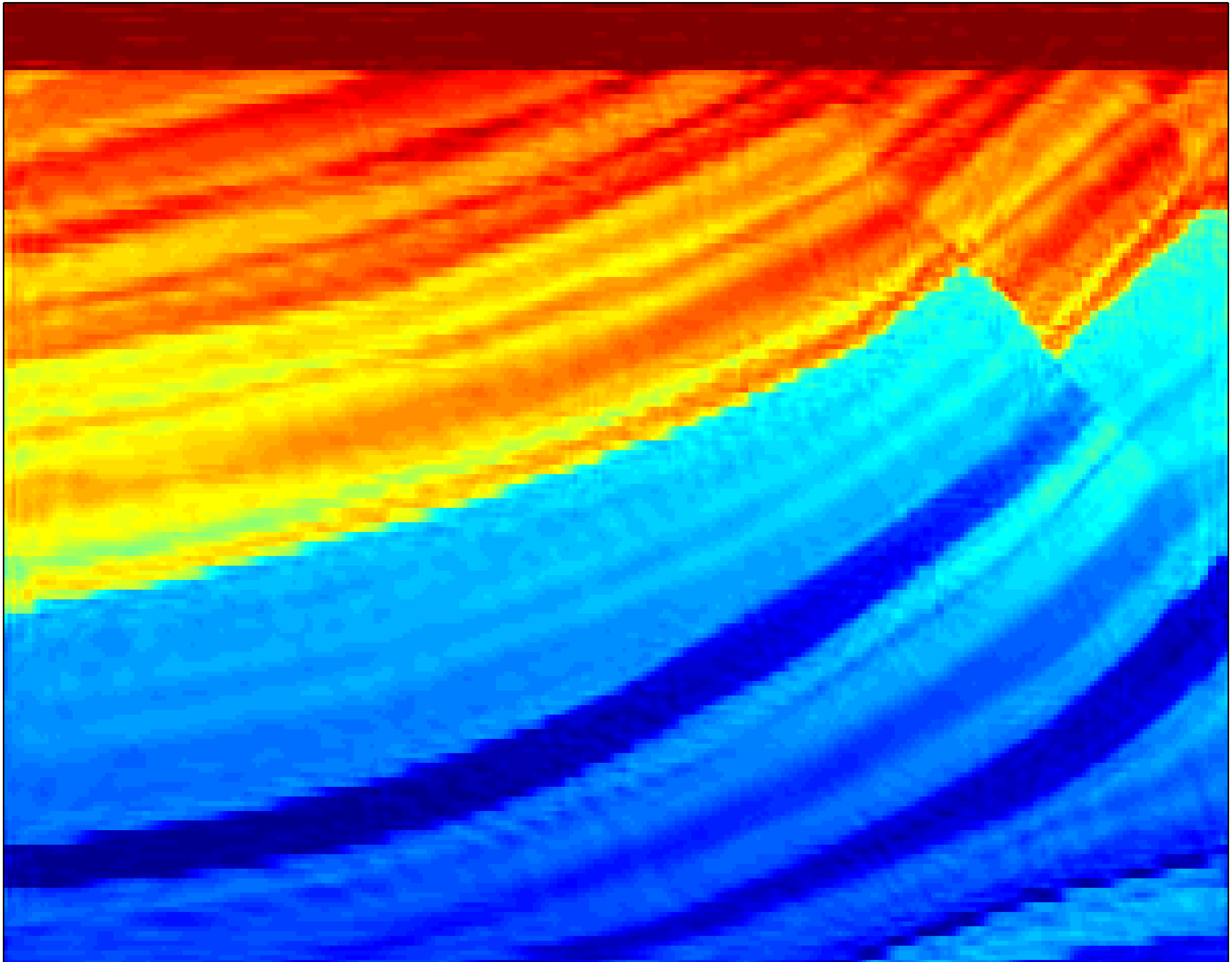




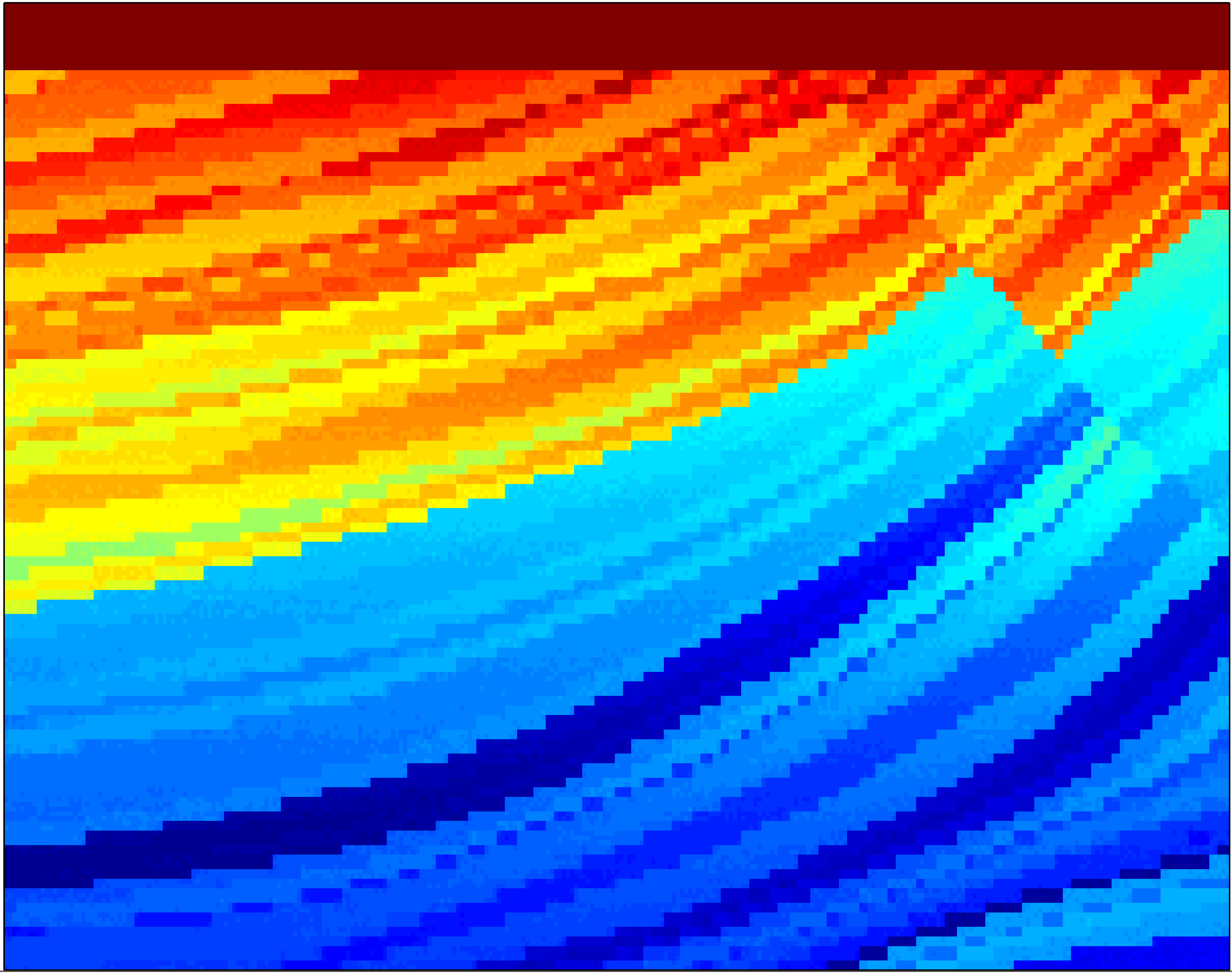
1% of coeff.



5% of coeff.



50% of coeff.



FWI: Sparsity Regularization

Sparsity-promoting formulations:

1: QP $$\min_{\mathbf{x}} \|\mathbf{D} - \boldsymbol{\mathcal{F}}[\boldsymbol{\mathcal{C}}^*\mathbf{x};\mathbf{Q}]\|_F^2 + \lambda\|\mathbf{x}\|_1$$

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

3: BPDN $$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{D} - \boldsymbol{\mathcal{F}}[\boldsymbol{\mathcal{C}}^*\mathbf{x};\mathbf{Q}]\|_F^2 \leq \sigma$$

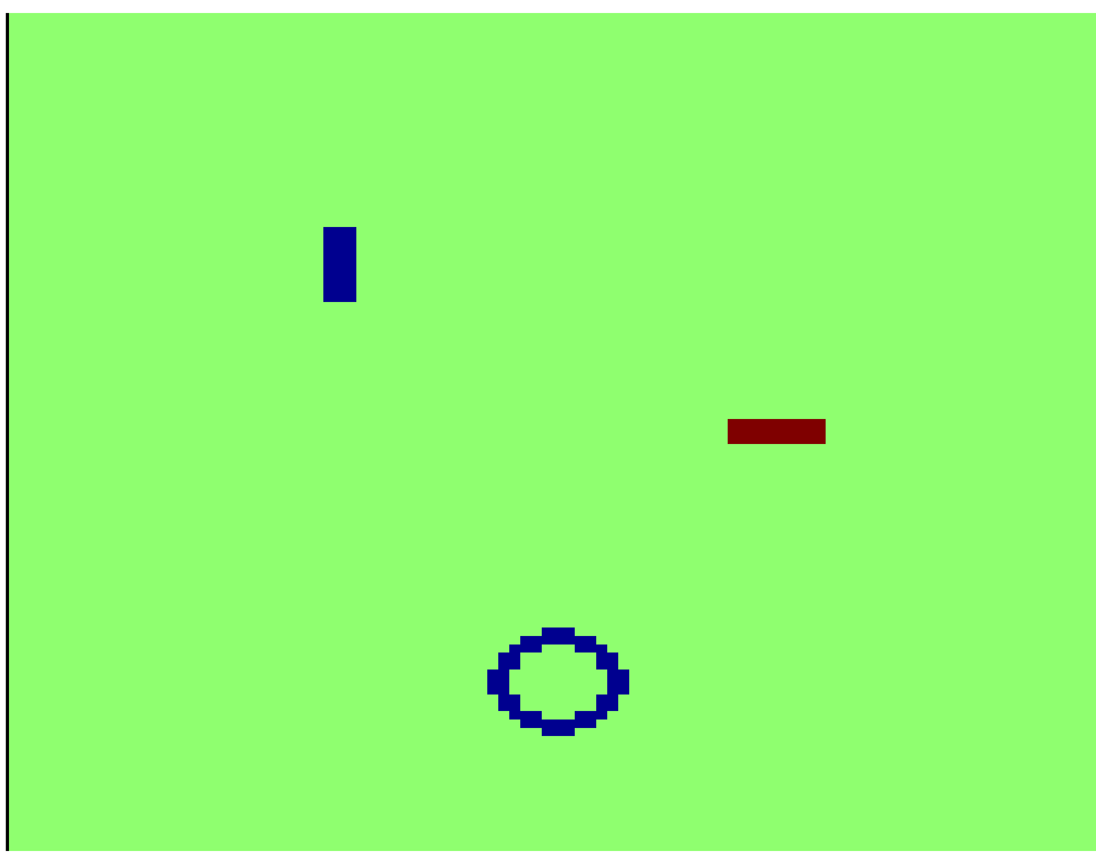
BPDN formulation looks promising from a scientific standpoint, but Lasso formulation is easier to optimize.

Case Study

- We consider a model that is sparse in physical domain: sparse perturbation of constant background velocity (2km/s)
- Cross-well setting, 101 sources and receivers in vertical wells 800 m. apart
- 9 pt. discretization of Helmholtz operator with absorbing boundary; 10 m. spacing on grid;
- Random frequencies [5.0, 6.0, 11.5, 14.0, 15.5, 17.5, 23.5] Hz.
- We consider full inversion, and subsampling with 5 sim. shots.

Geometric Setup

True Model

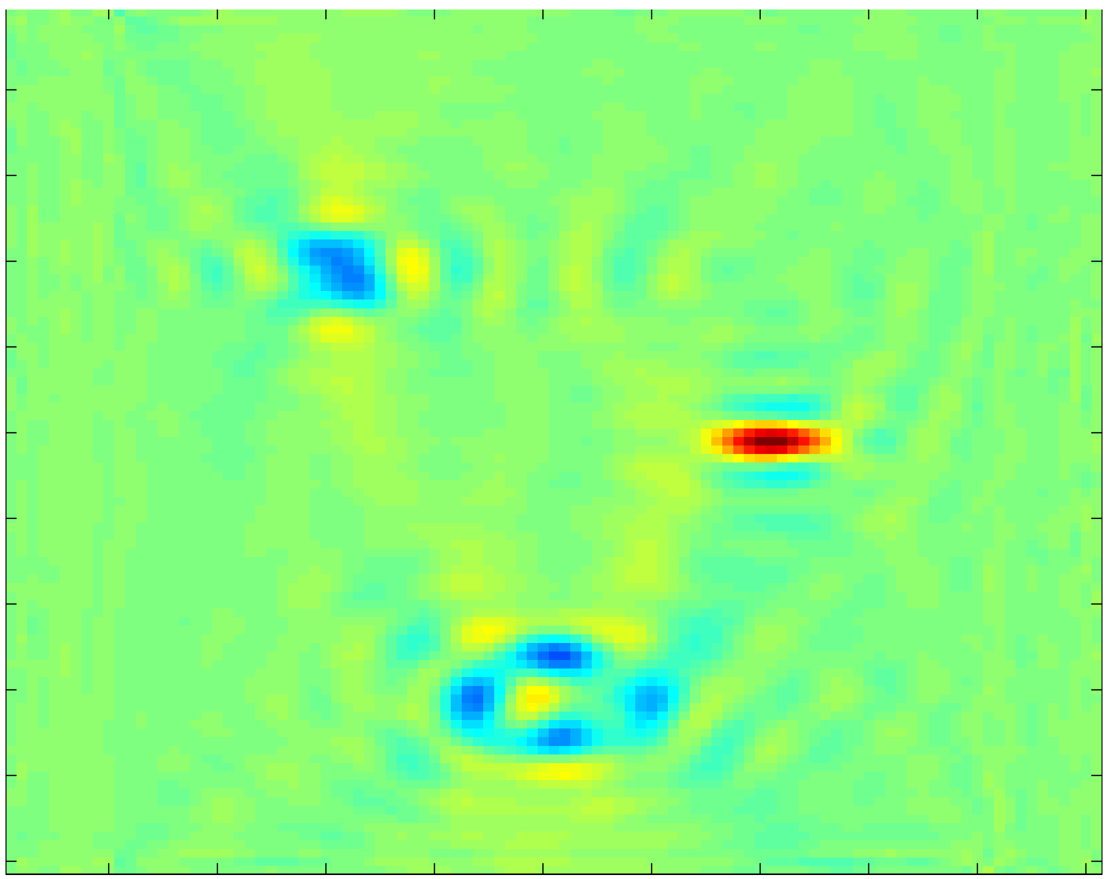


True L1-norm: 5.7
L2-error: 0

Initial Model

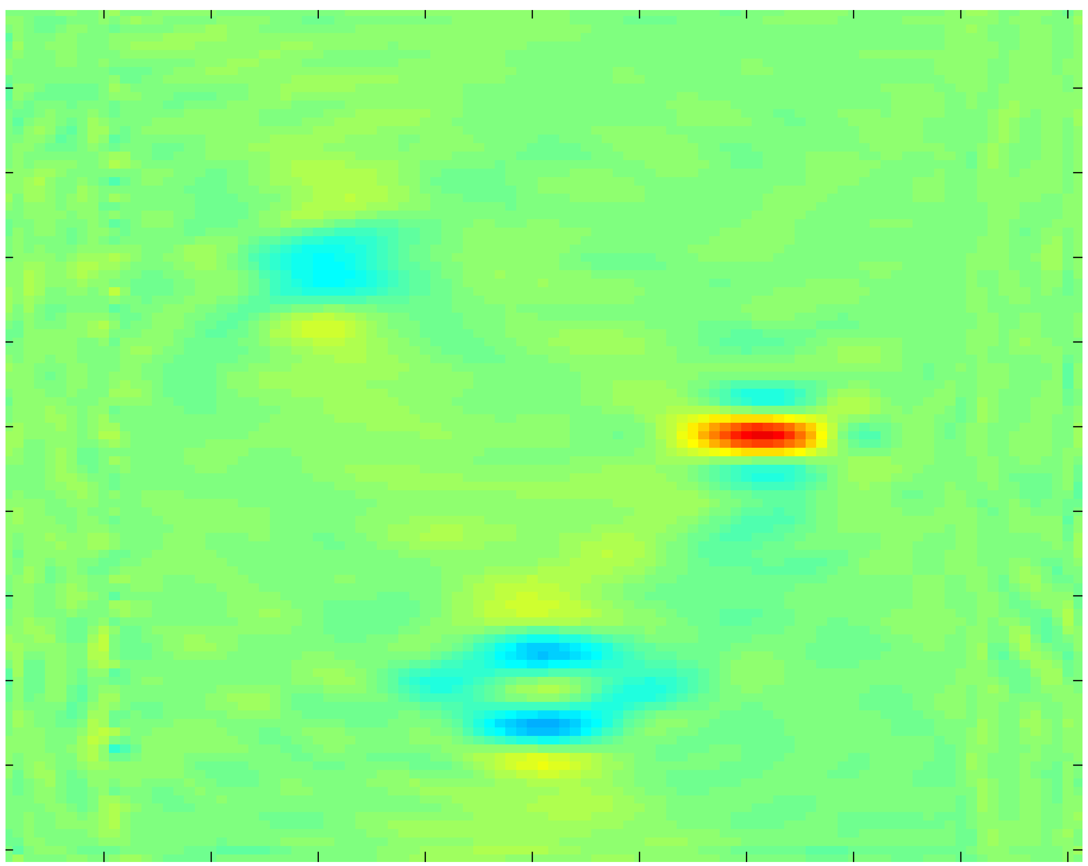
Least Squares Results:

Full model, LBFGS (500)



L1-norm: 19.2
L2 Relative Residual:1e-5

5 Shots, LBFGS (200)



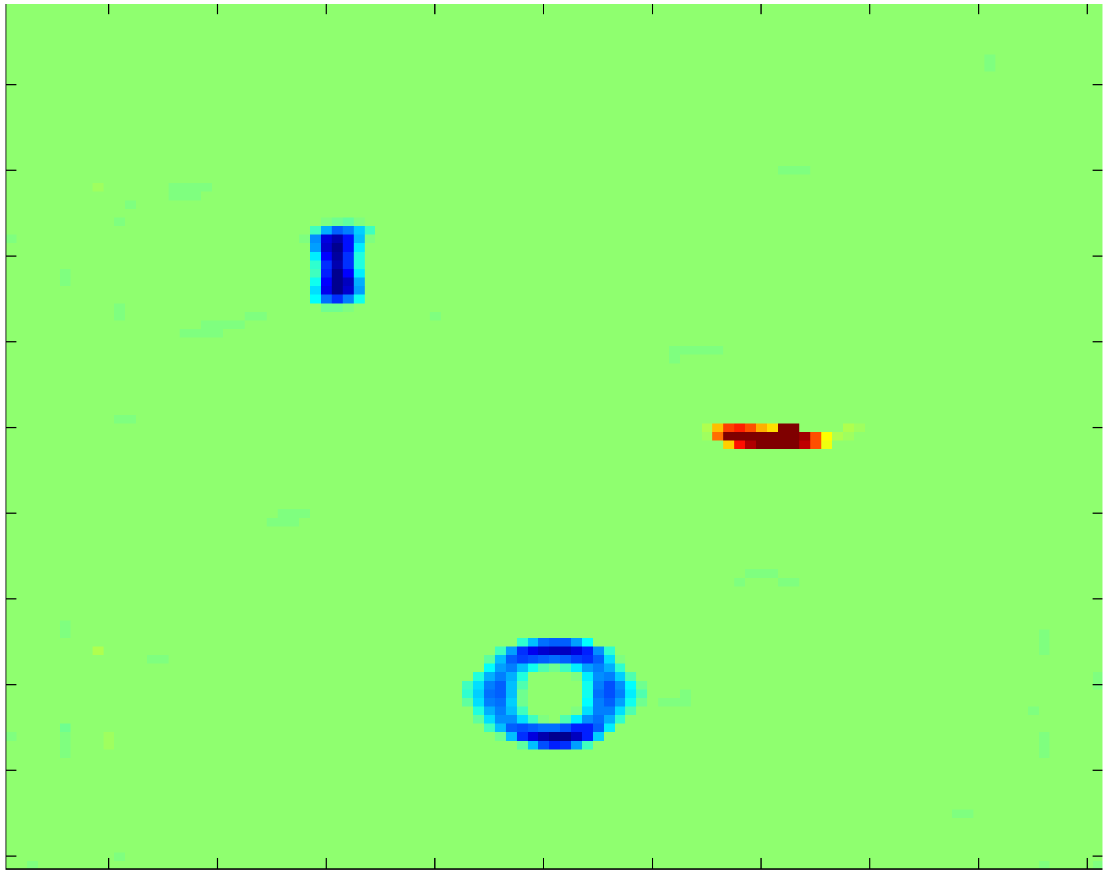
L1-norm: 22.7
L2 Relative Residual:1e-7

Lasso Results

LASSO FORMULATION

$$\min_{\mathbf{m}} \quad \|\mathbf{D} - \mathcal{F}[\mathbf{m_0} + \mathbf{m}; \mathbf{Q}]\|_F^2$$
$$\text{s.t.} \quad \|\mathbf{m}\|_1 \leq \tau$$

5 SHOTS, SPG (400)



L1-NORM: 5.7
L2 RELATIVE RESIDUAL:1E-4

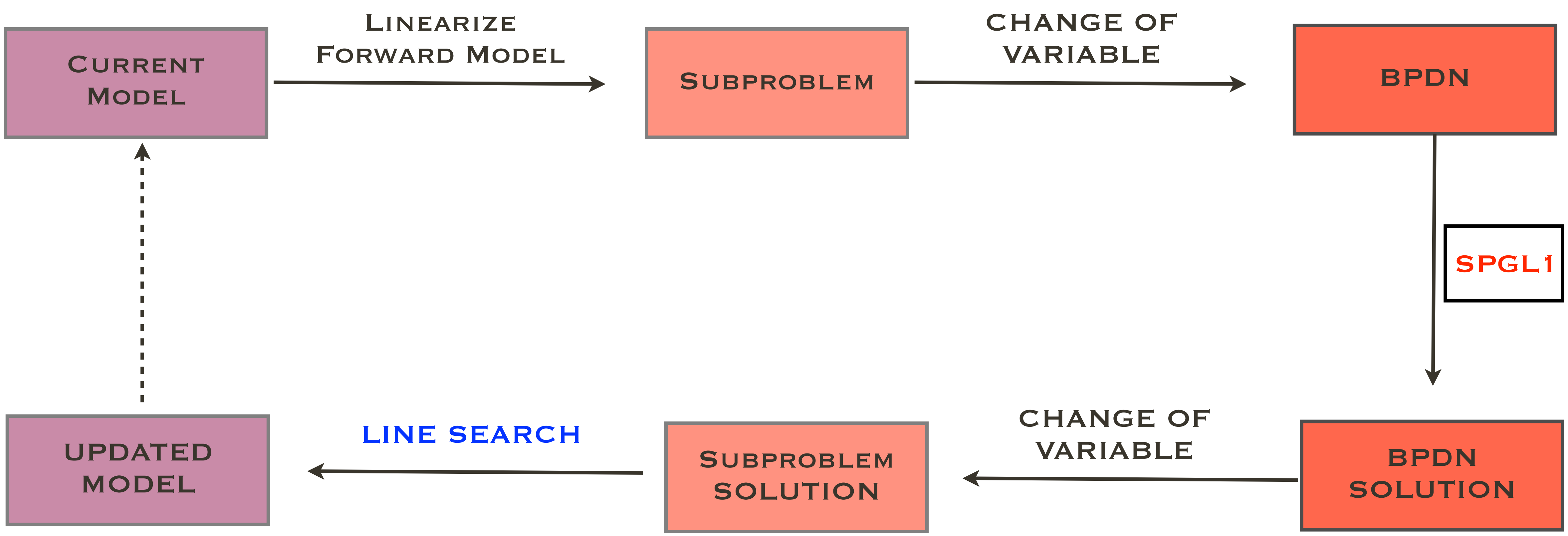
BPDN Algorithm

- **Optimization problem:** $$\min_{\mathbf{m}} \|\mathbf{m}\|_1 \quad \text{s.t.} \quad \|\mathbf{D} - \mathcal{F}[\mathbf{m_0} + \mathbf{m}; \mathbf{Q}]\|_F^2 \leq \sigma$$
- **Implement iterated algorithm:** $$\mathbf{m}^{\nu+1} = \mathbf{m}^{\nu} + \gamma_\nu \boldsymbol{\delta}\mathbf{m}$$
- **Direction $\boldsymbol{\delta}\mathrm{m}$ solves subproblem below using SPGL1 algorithm:**

$$\min_{\boldsymbol{\delta}\mathbf{m}} \|\mathbf{m}^{\nu} + \boldsymbol{\delta}\mathbf{m}\|_1$$
$$\text{s.t.} \quad \|\mathbf{D} - \mathcal{F}[\mathbf{m_0} + \mathbf{m}^{\nu}; \mathbf{Q}] - \nabla\mathcal{F}[\mathbf{m_0} + \mathbf{m}^{\nu}; Q]\boldsymbol{\delta}\mathbf{m}\|_F^2$$
$$\leq 0.95 \left(\|\mathbf{D} - \mathcal{F}[\mathbf{m_0} + \mathbf{m}^{\nu}; \mathbf{Q}]\|_F^2 - \sigma\right)_+$$

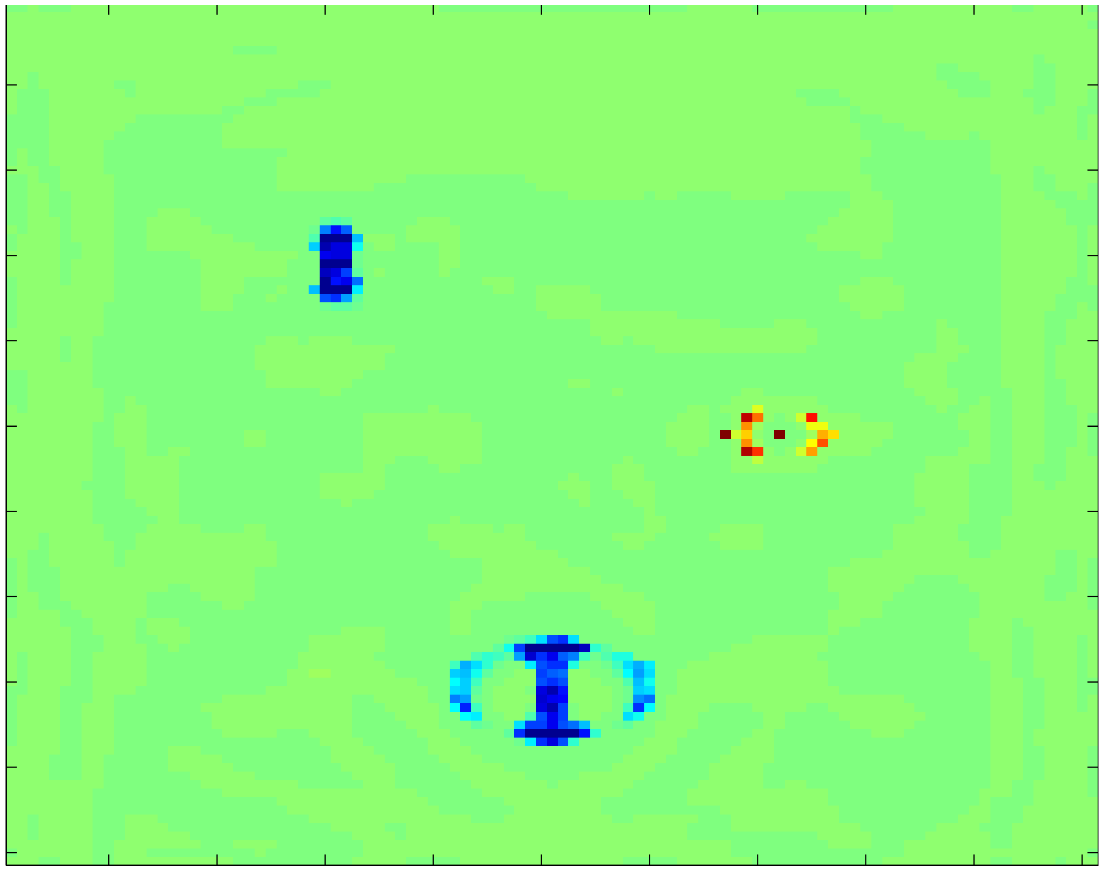
[Burke ’89, Burke ’92]

BPDN Algorithm



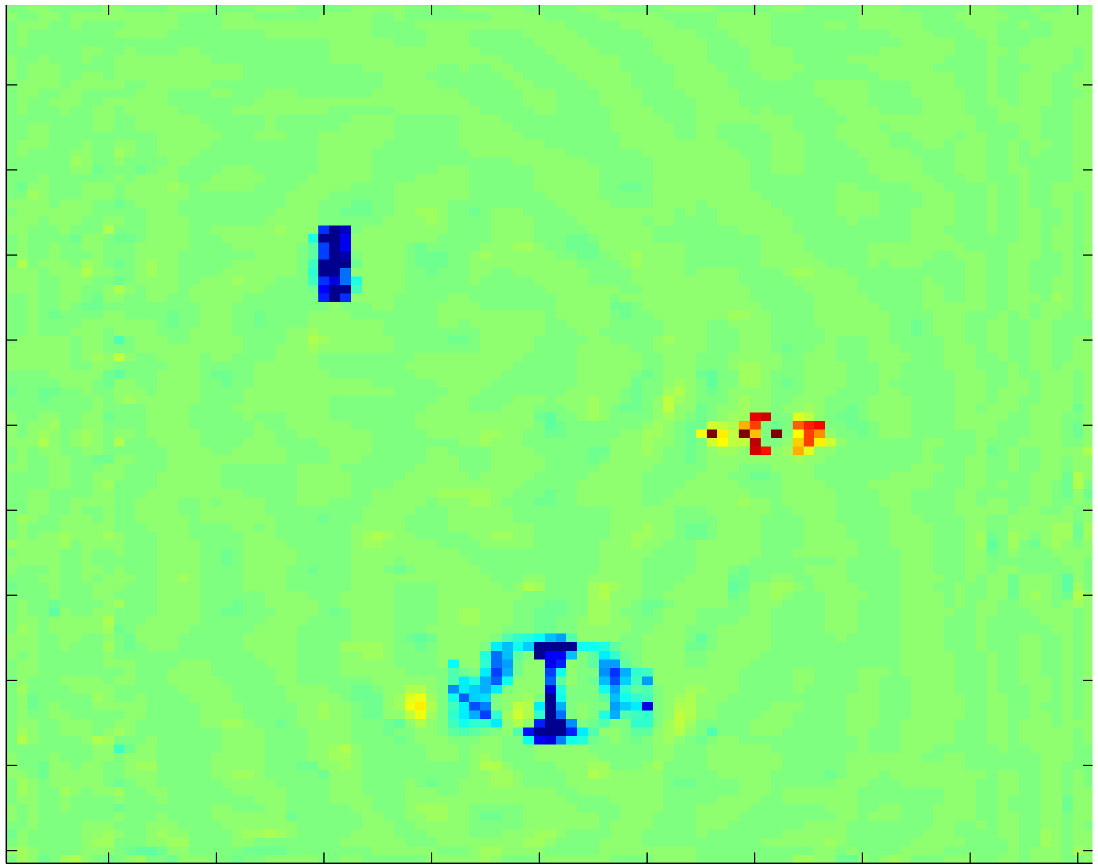
BPDN Results

Full model (200)



L1-norm: 5.85
L2 Relative Residual:1e-2

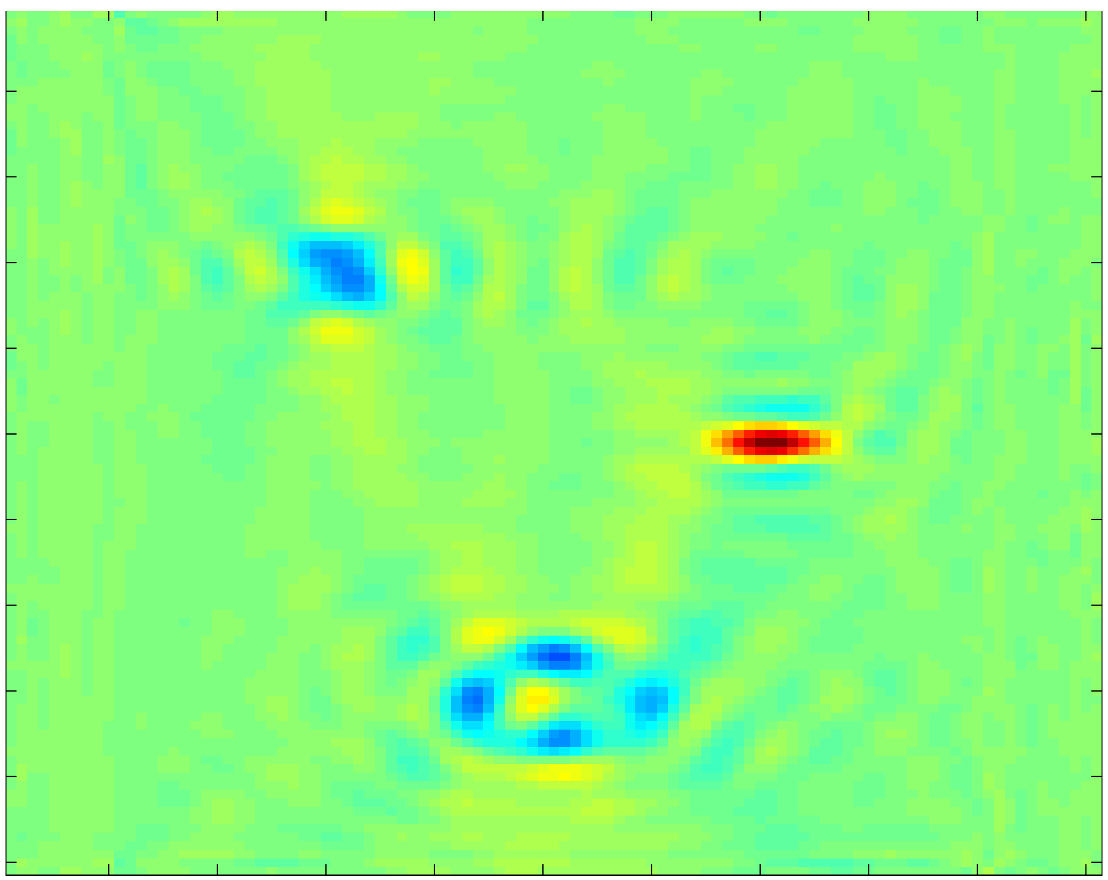
5 Shots (200)



L1-norm: 9.3
L2 Relative Residual:1e-3

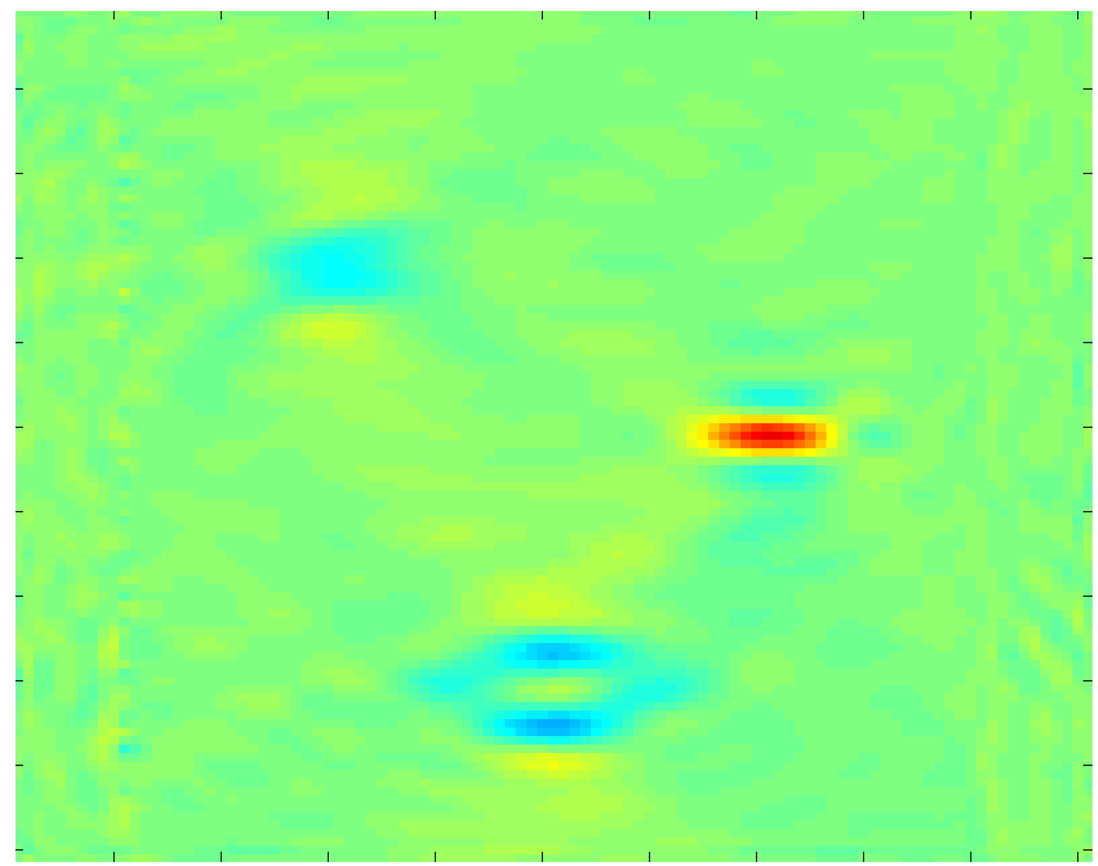
Least Squares Results:

Full model, LBFGS (500)



L1-norm: 19.2
L2 Relative Residual:1e-5

5 Shots, LBFGS (200)



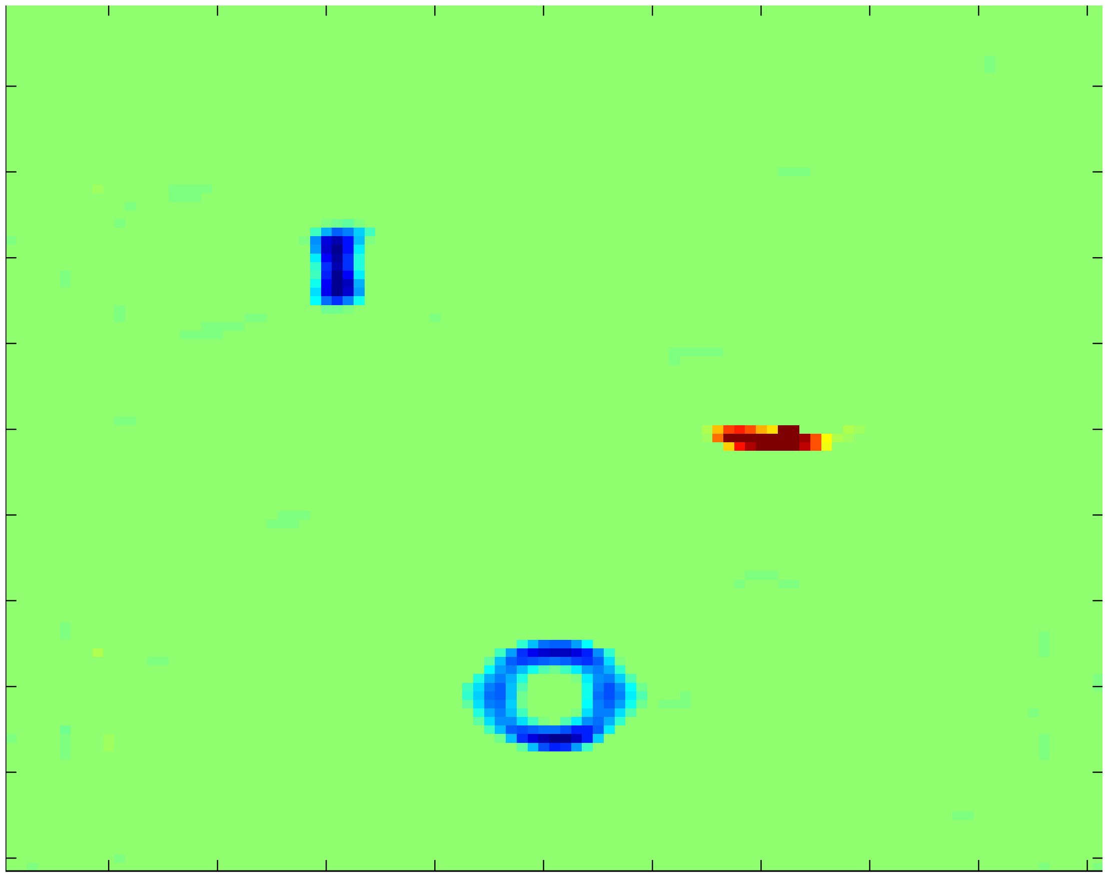
L1-norm: 22.7
L2 Relative Residual:1e-7

Lasso Results

Lasso formulation

$$\min_{\mathbf{m}} \quad \|\mathbf{D} - \mathcal{F}[\mathbf{m_0} + \mathbf{m}; \mathbf{Q}]\|_F^2$$
$$\text{s.t.} \quad \|\mathbf{m}\|_1 \leq \tau$$

5 Shots, SPG (400)



L1-norm: 5.7
L2 Relative Residual:1e-4

Conclusions

- Exploiting sparsity is useful for fast computation as well as for novel modeling/regularization of FWI

- Understanding trade-off between least-squares and sparsity promoting priors is important in modeling and algorithm design.

- Preliminary results are very promising: we can recover a sparse solution from insufficient data, and we can significantly improve speed of recovery.

The Road Ahead

- Test regularization approaches on seismic models using Curvelets
- Test all algorithms on problems with noisy data
- Implement renewal strategy for simultaneous shots in the regularization context
- Study the trade-off between sparsity and least-squares misfit in the nonlinear context

SLIM

Acknowledgements

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.

References I

Burke, J.V., 1989, A sequential quadratic programming method for potentially infeasible mathematical programs, *Journal of Mathematical Analysis and Applications, 139,2:319-351*

Burke, J.V., 1992, A robust trust region method for constrained nonlinear programming problems, *Siam J. Optimization, 2,2:325-347, 1992*

Burke, J.V., and Ferris, M.C, A Gauss-Newton method for convex composite optimization, *Technical Report, University of Wisconsin-Madison, 1993*

Candes, E. J., and Demanet, L., The curvelet representation of wave propagators is optimally sparse. *Technical Report, California Institute of Technology, 2004.*

Candes, E.J., and Donoho, D. L., *Curvelets - A Surprisingly Effective Nonadaptive Representation for Objects with Edges*, Saint-Malo Proceedings, Vanderbilt University Press.

R. Fletcher, Second order correction for nondifferentiable optimization, in *G. A. Watson, editor, Numerical Analysis, 85-114, Springer-Verlag, Berlin, 1982*

Haber, E., M. Chung, and Felix Herrmann, An effective method for parameter estimation with pde constraints with multiple right hand sides, *Technical Report TR-2010-4, UBC-EOS, 2010*

References II

Haber, E., M. Chung, and Felix Herrmann, An effective method for parameter estimation with pde constraints with multiple right hand sides, *Technical Report TR-2010-4, UBC-EOS, 2010*

Shapiro, A., Monte carlo sampling methods, *in Stochastic Programming, Volume 10 of Handbooks in Operation Research and Management Science: North Holland, 2003.*

Shapiro, A., and A. Nemirovsky, *On complexity of stochastic programming problems, in Continuous Optimization: Currenet Trends and Applications: Springer, New York, 2005.*

W.W. Symes, Migration velocity anallysis and waveform inversion, *Geophysical Prospecting, 56, 765-790, 2008*

W.W. Symes, The seismic reflection inverse problem, *Inverse Problems, 25, 2009*

van en Berg, E., and Friedlander, M.P., Probing the Pareto frontier for basis pursuit solutions, *Siam J. Sci Comput. Vol. 31, No.2, pp. 890-912, 2008*

J. Virieux and S. Operto, An overview of full-waveform inversion in exploration geophysics, *Geophysics, 74, 2009*

R. S. Womerseley, Local properties of algorithms for minimizing composite functions, *Mathematical Programming, 32:69-89, 1985*