SINBAD Consortium 2010

Sparsity Promoting Formulations and Algorithms for FWI

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

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: billions 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

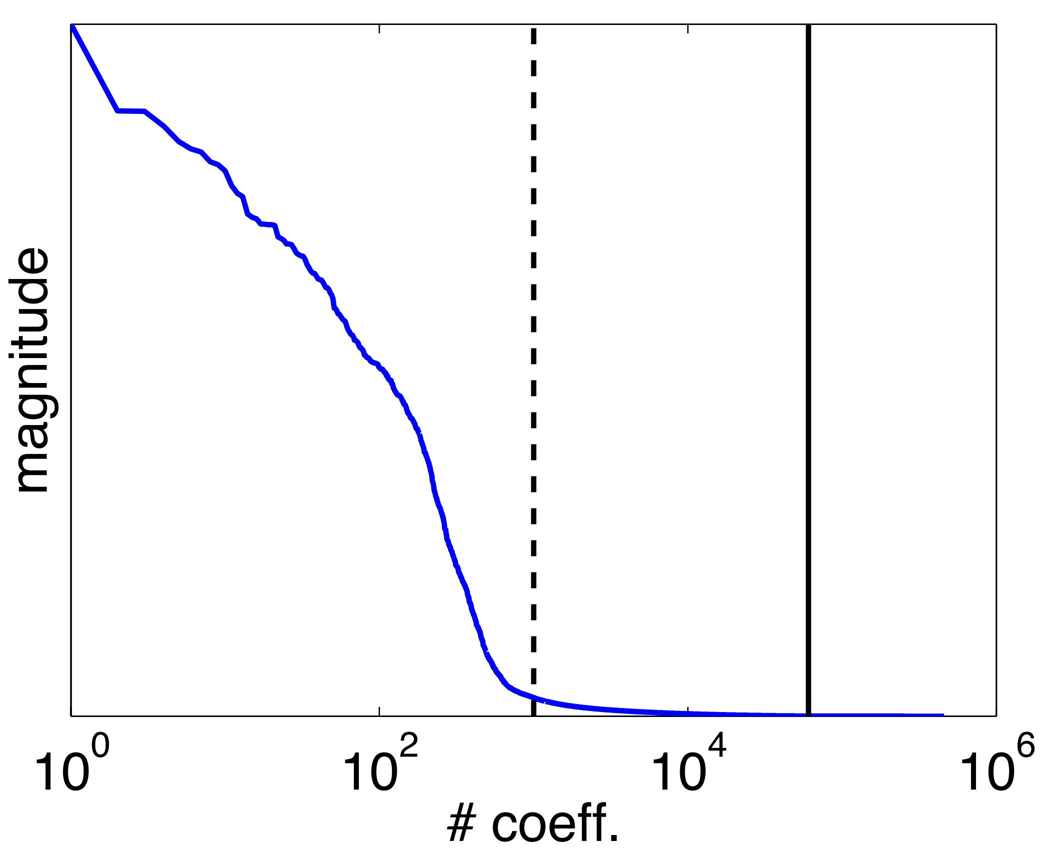
- Here we focus on sparse formulations to address some of these problems.

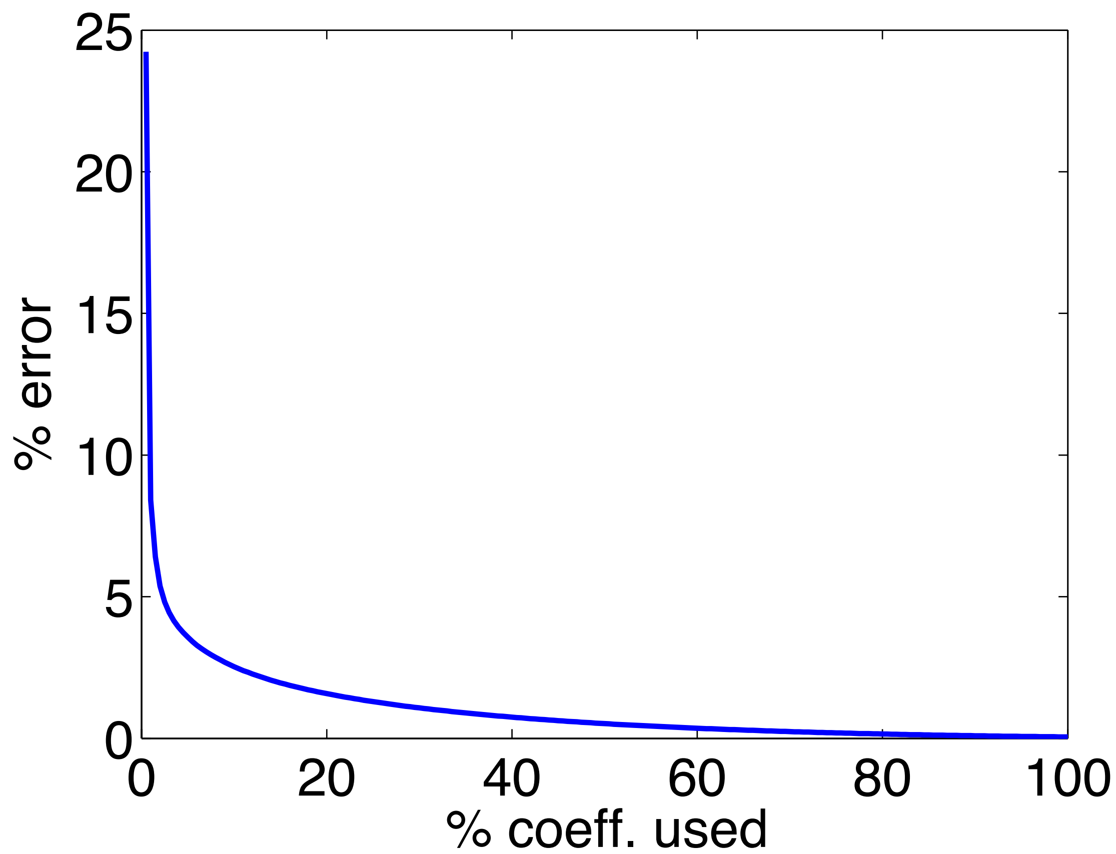
[Virieux ’09; Symes ’09; Symes ’08]

Compressibility in Curvelets

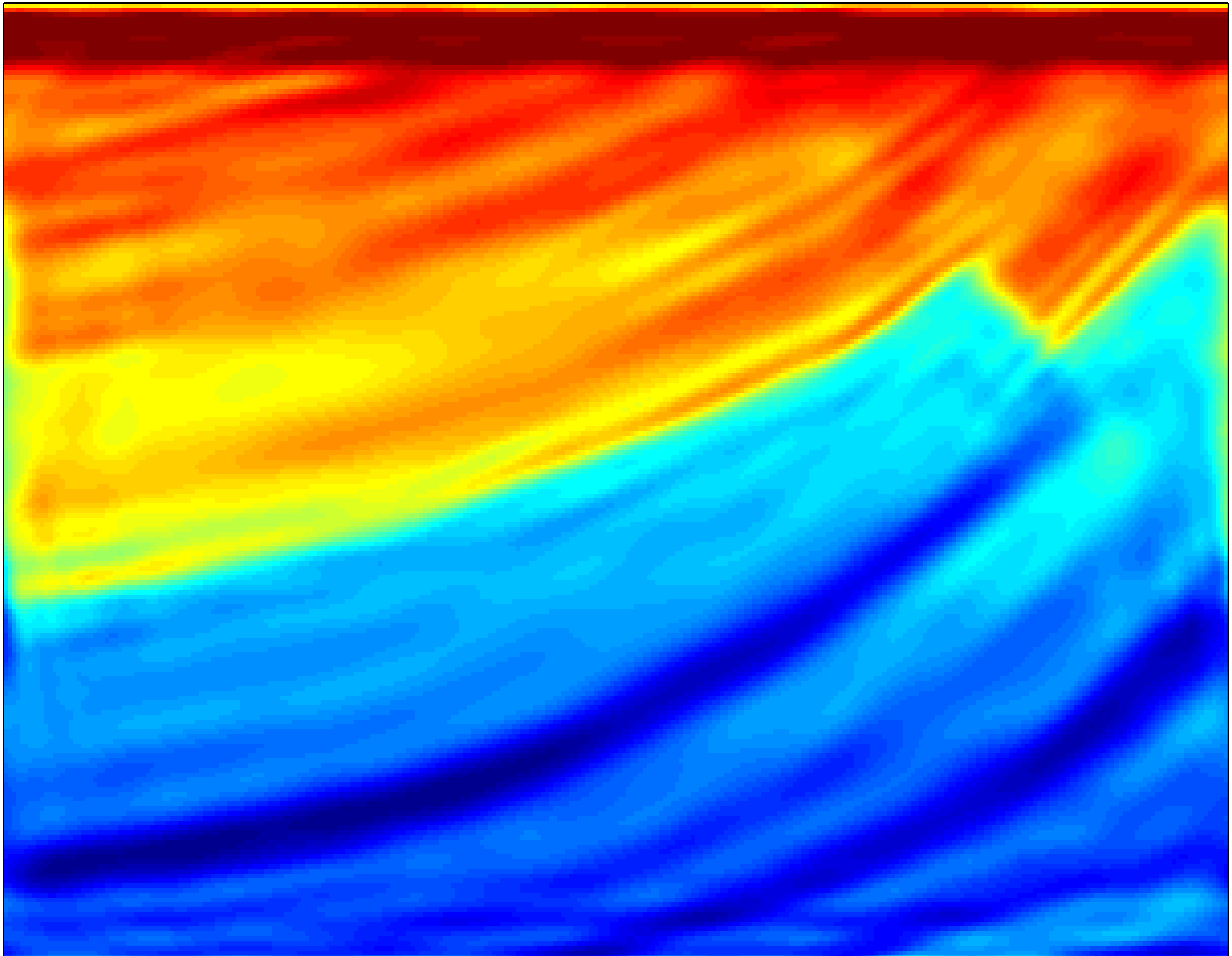
- **Velocity models are compressible in Curvelets.**
- **Geophysical images are layered, and may me modeled as objects with edges. Curvelets provide sparse representations for such images.**

[Candes '00]

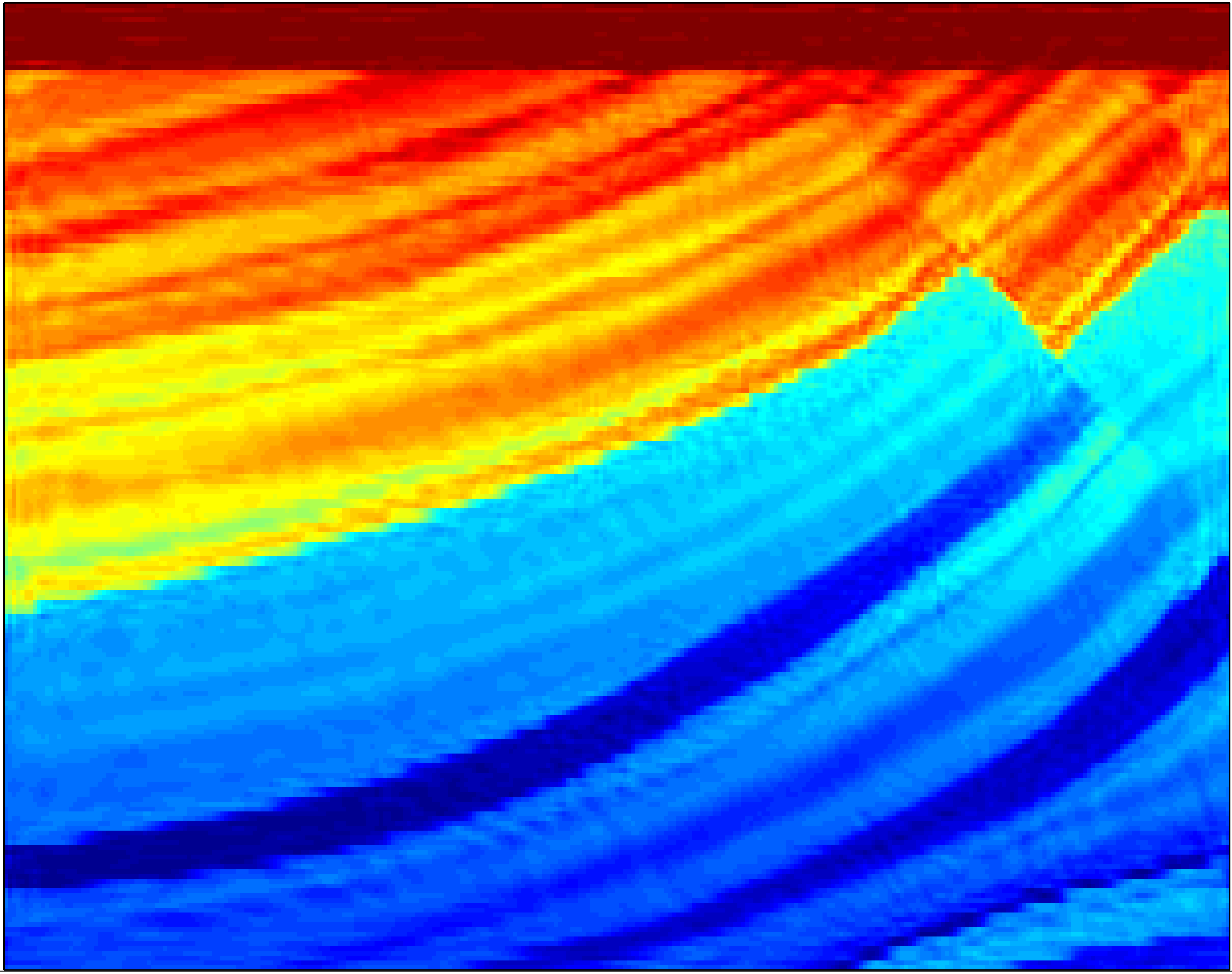




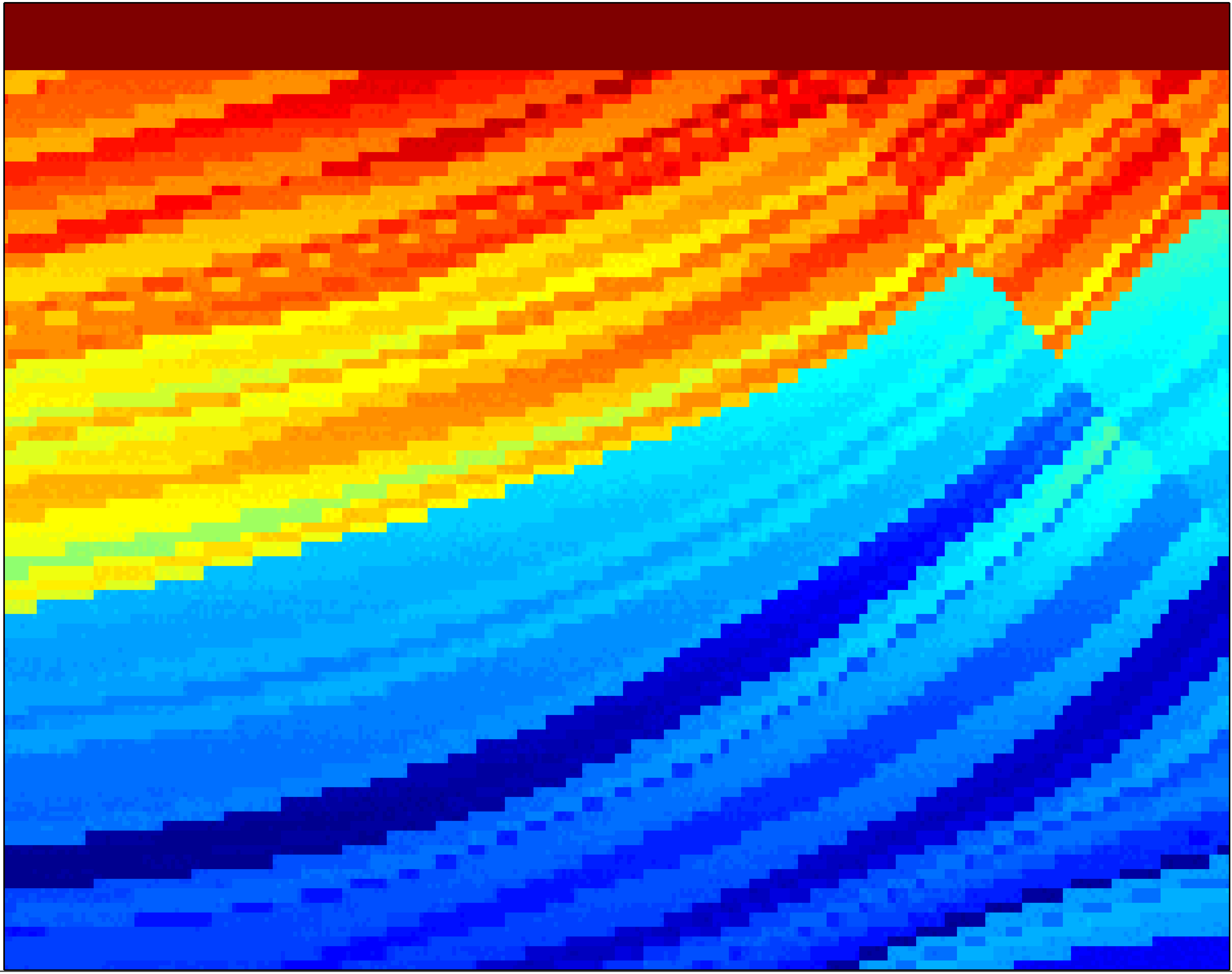
1% of coeff.



5% of coeff.



50% of coeff.



FWI: Sparsity Regularization

Sparsity-promoting formulations:

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

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

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

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

Algorithms I

For now we focus on the nonlinear LASSO formulation:

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

A Limited Memory Projected Quasi-Newton method has recently been proposed for optimization problems of the form

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in \mathbf{C}$$

[Schmidt et al. '09]

Matlab code is available from
http://www.cs.ubc.ca/~schmidtm/Software/PQN.html

Algorithms II

The BFGS method solves QP subproblems, and at each iteration updates the Hessian approximation B_k using rank 2 updates

L-BFGS keeps only the most recent vectors, allowing a compact representation of B_k that stores a few (10 or 20) of the most recent vectors.

The Schimdt et al. algorithm uses Spectral Projected Gradient (SPG) to solve the constrained subproblem

$$\begin{array}{ll} \min & f_k + (x - x_k)^T g_k + \frac{1}{2}(x - x_k)^T B_k (x - x_k) \\ \text{s.t.} & x \in C \end{array}$$

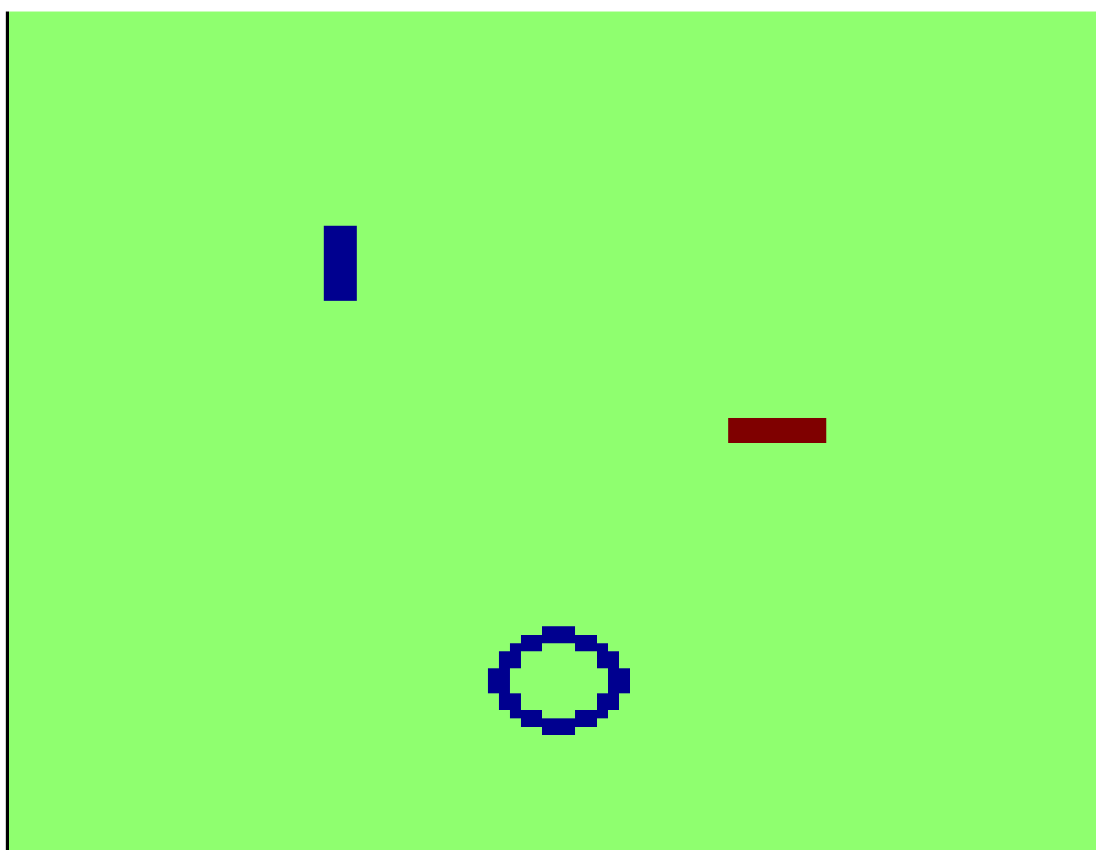
SLIM

Proof of Concept

- 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
- Sample of 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

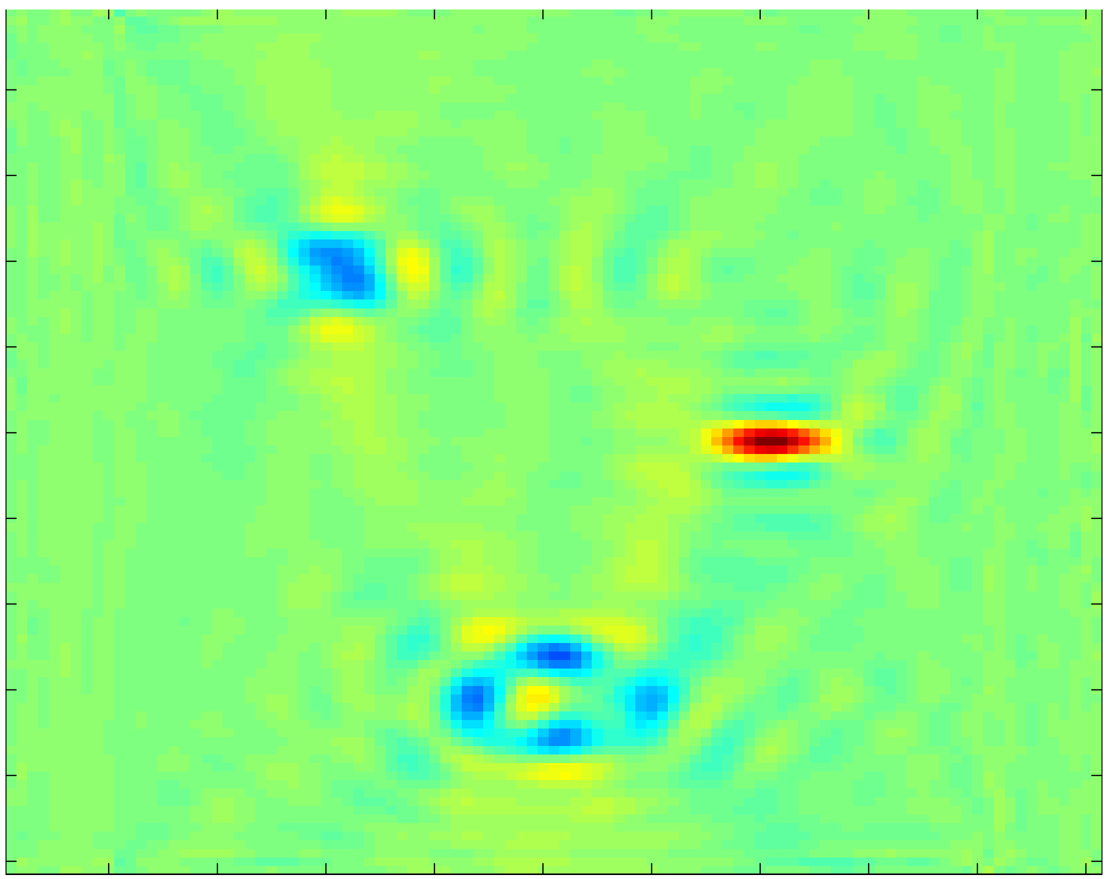


True l1-norm: 5.7

Initial Model

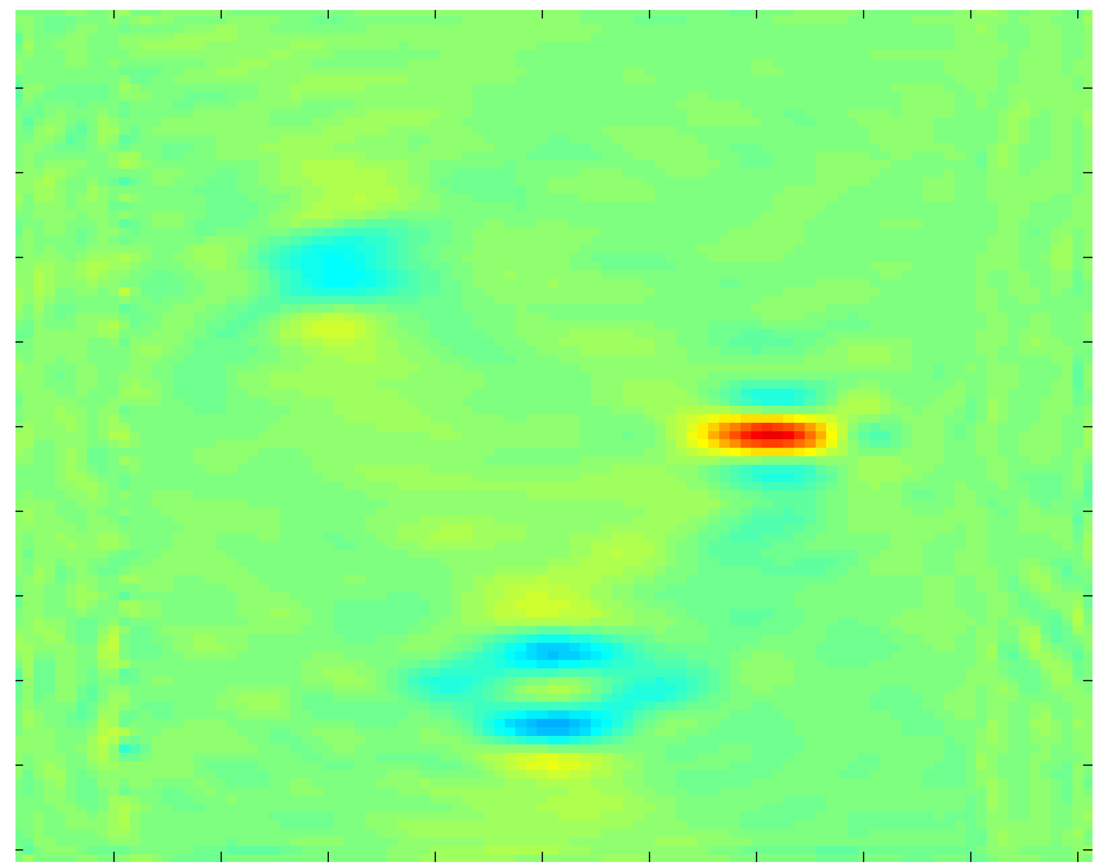
Least Squares Results:

Full model, LBFGS (500)



L1-norm: 19.2

5 Shots, LBFGS (200)



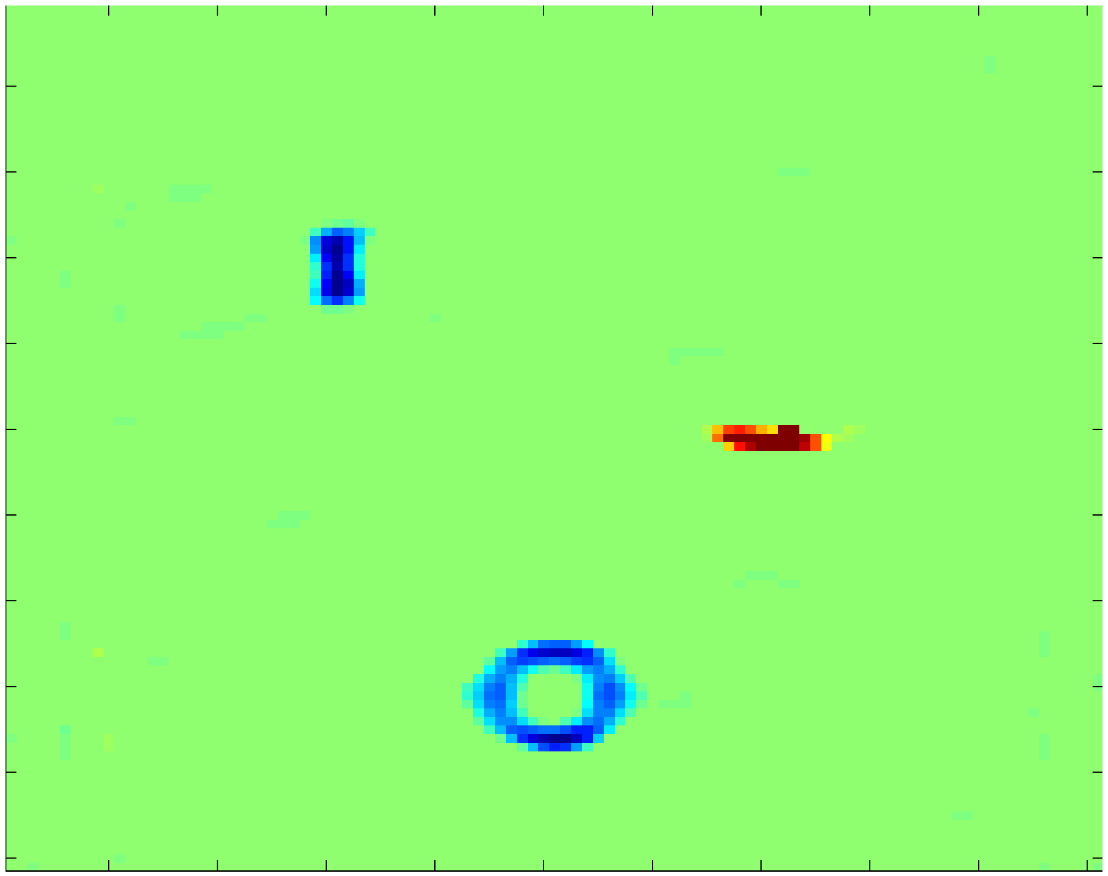
L1-norm: 22.7

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

Marmoussi Example

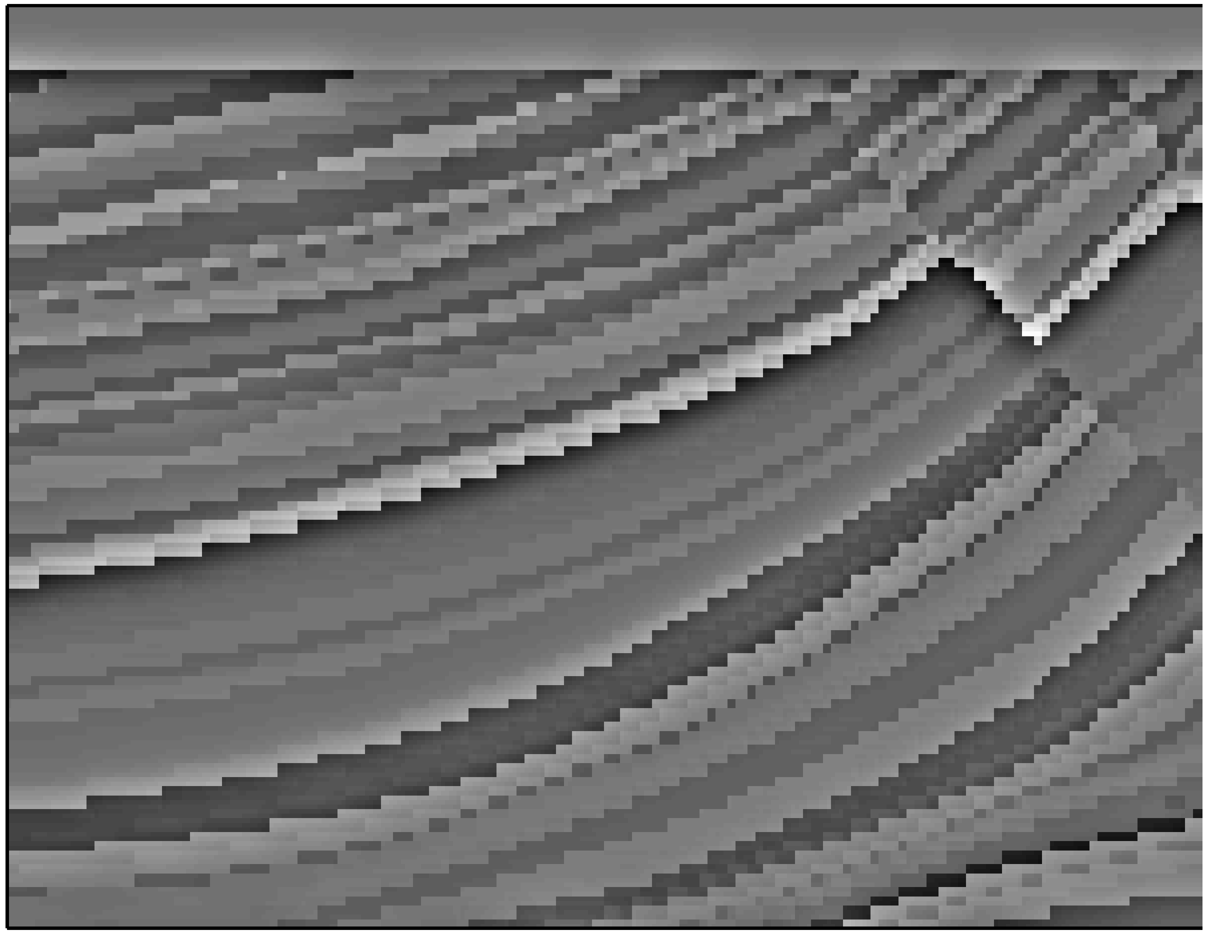
- We consider a subset of the Marmoussi model
- 151 shots, 301 receivers
- 9 pt. discretization of Helmholtz operator with absorbing boundary; 10 m. spacing on grid
- Sample of Frequencies [5.0, 6.0, 11.5, 14.0, 15.5, 17.5, 23.5] Hz
- We consider subsampling with 5 sim. shots

Curvelet Example

Curvelet Lasso formulation

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

True Reflectivity



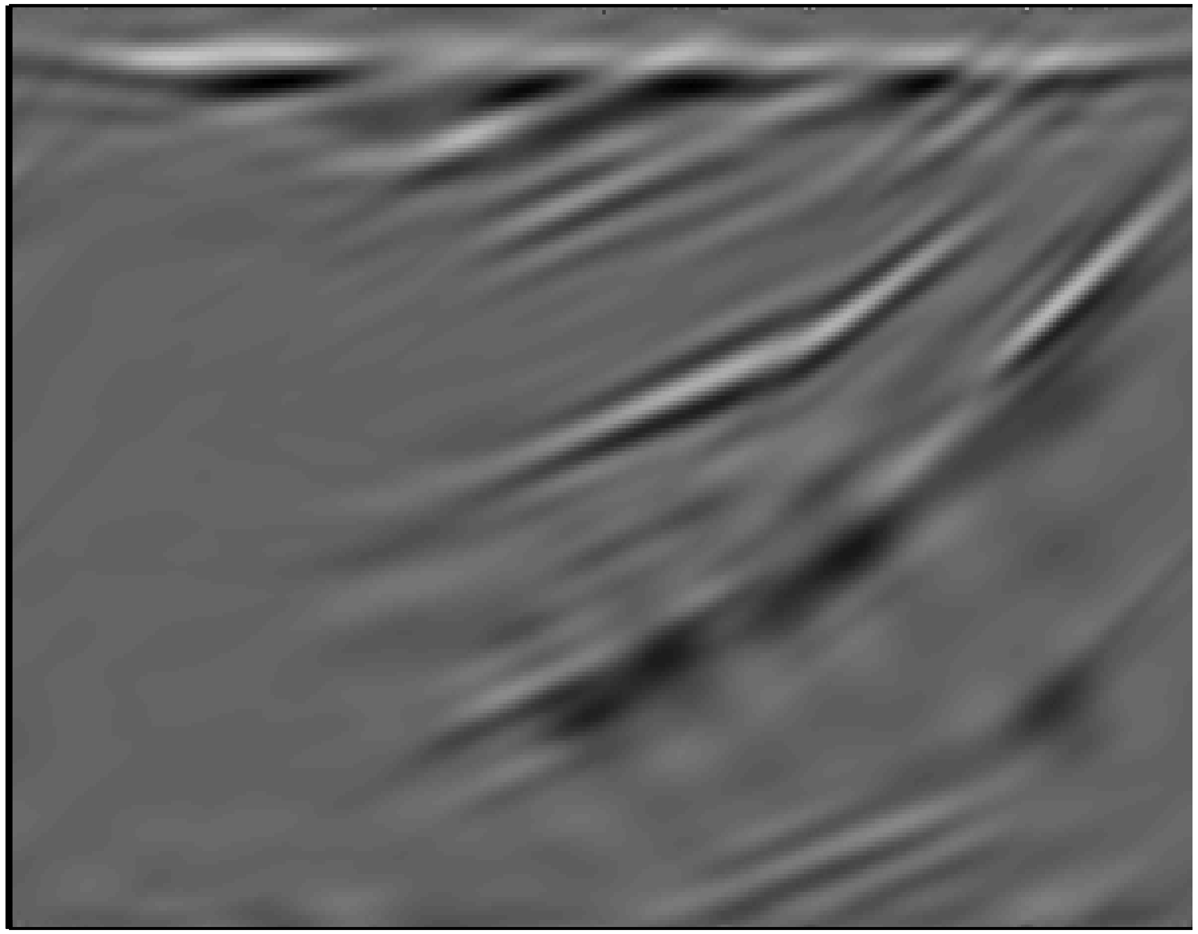
Curvelet Results

CURVELET LASSO FORMULATION

$$\min_{\mathbf{x}} \quad \|\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m_0} + \overbrace{C^*\mathbf{x}}^{\mathbf{m}}; \mathbf{Q}]\|_F^2$$

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

$\tau = 30$

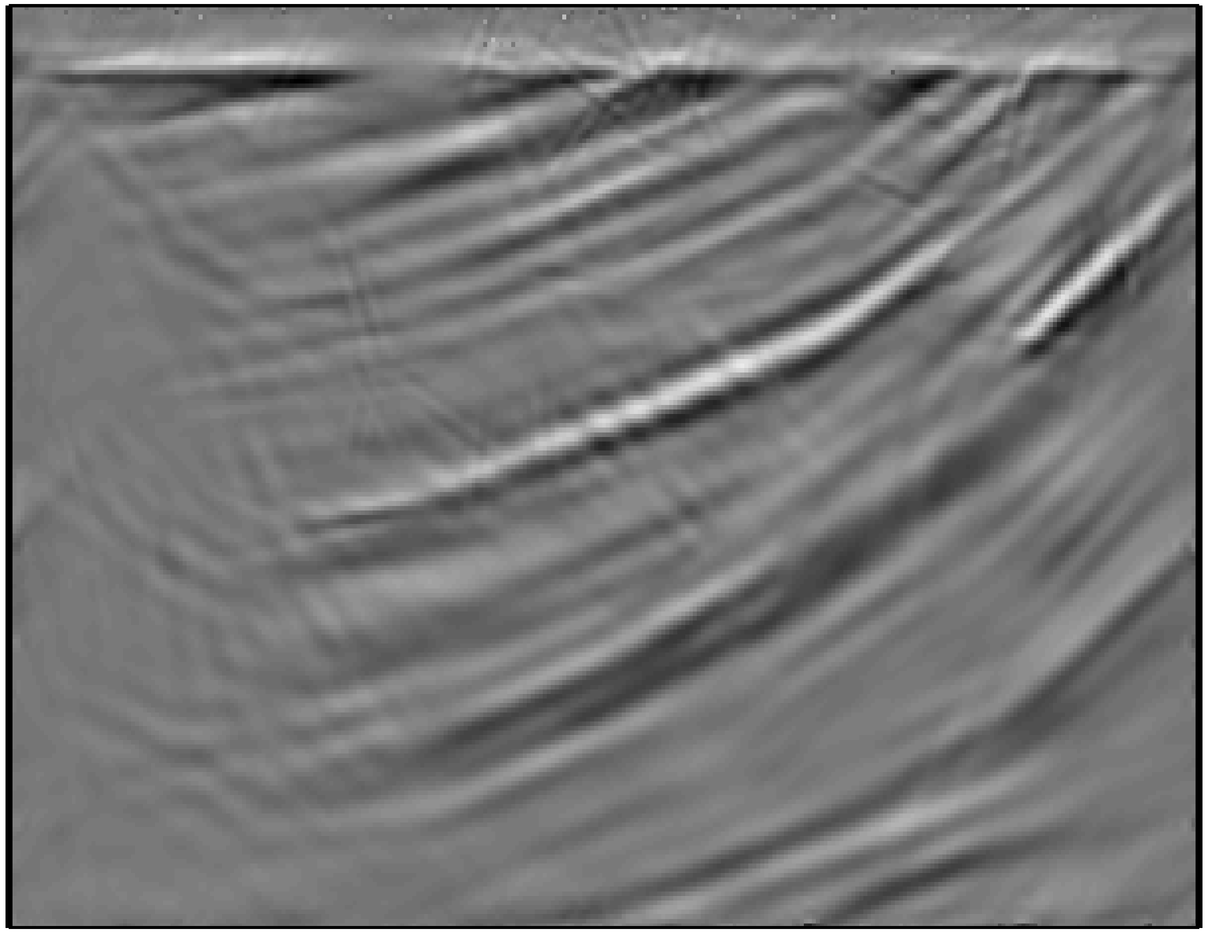


Curvelet Results

CURVELET LASSO FORMULATION

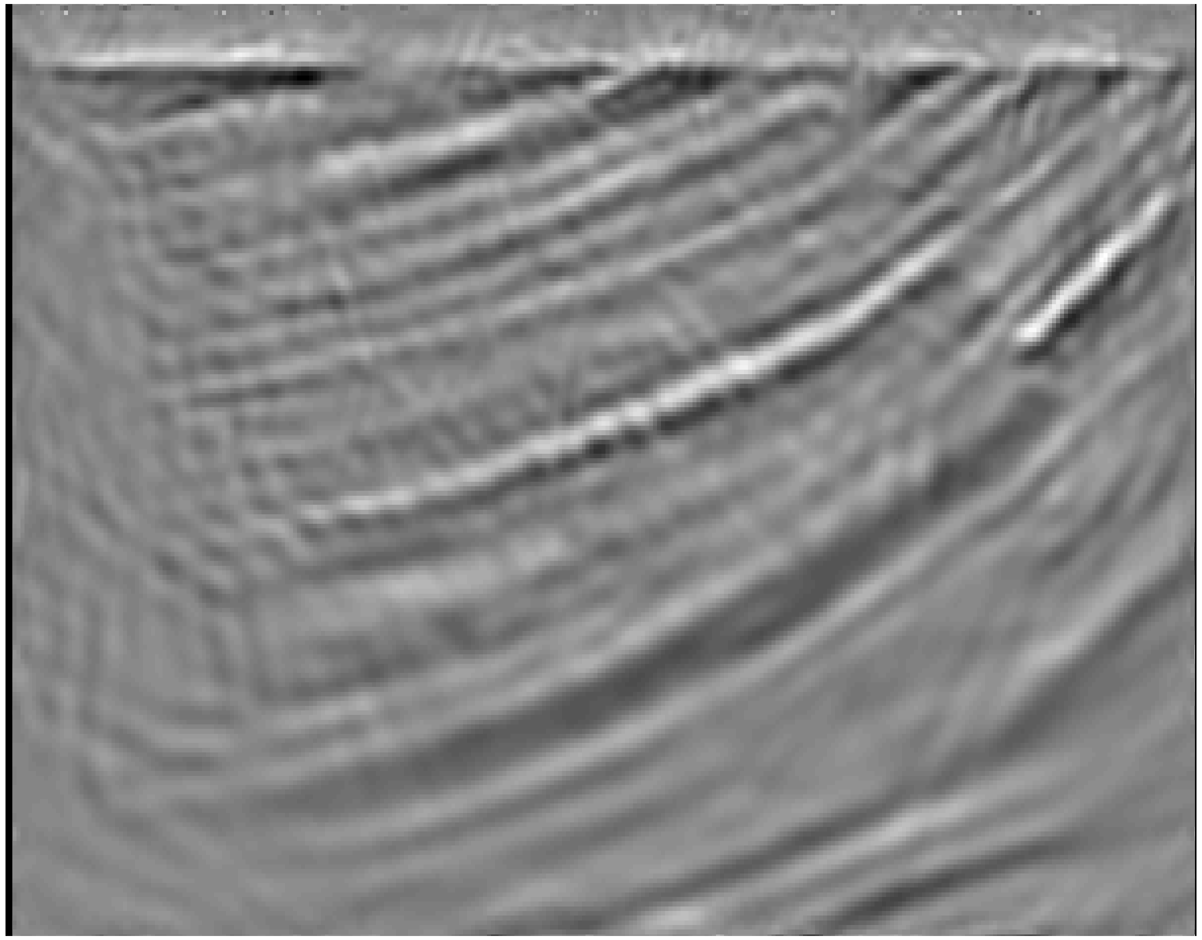
$$\min_{\mathbf{x}} \quad \|\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m_0} + \overbrace{C^*\mathbf{x}}^{\mathbf{m}}; \mathbf{Q}]\|_F^2$$
$$\text{s.t.} \quad \|\mathbf{x}\|_1 \leq \tau$$

$\tau = 100$



Curvelet Results

$\tau = 170$



Curvelet Lasso formulation

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

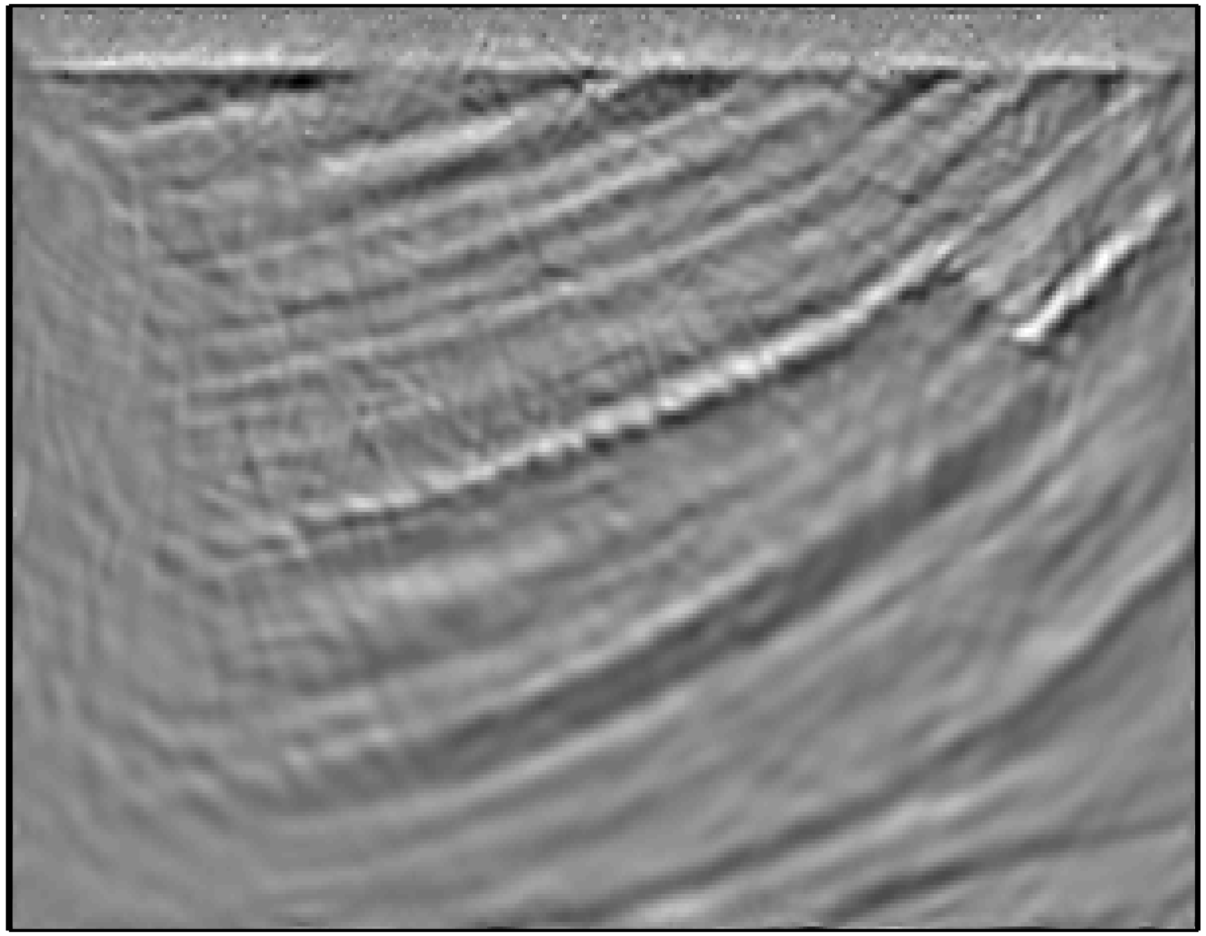
Curvelet Results

Curvelet Lasso formulation

$$\min_{\mathbf{x}} \quad \|\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m_0} + \overbrace{C^*\mathbf{x}}^{\mathbf{m}}; \mathbf{Q}]\|_F^2$$

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

$\tau = 250$

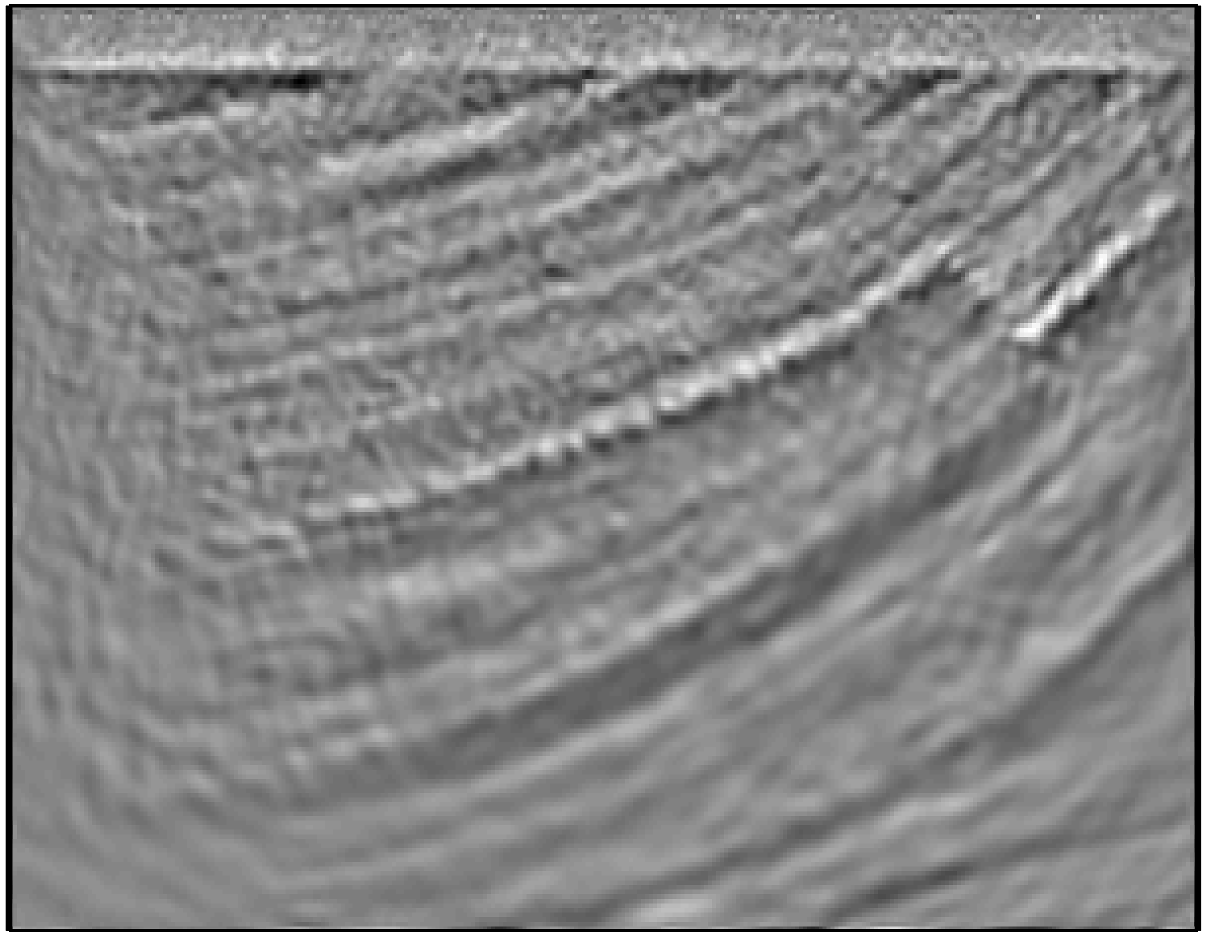


Curvelet Results

Curvelet Lasso formulation

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

$\tau = 400$



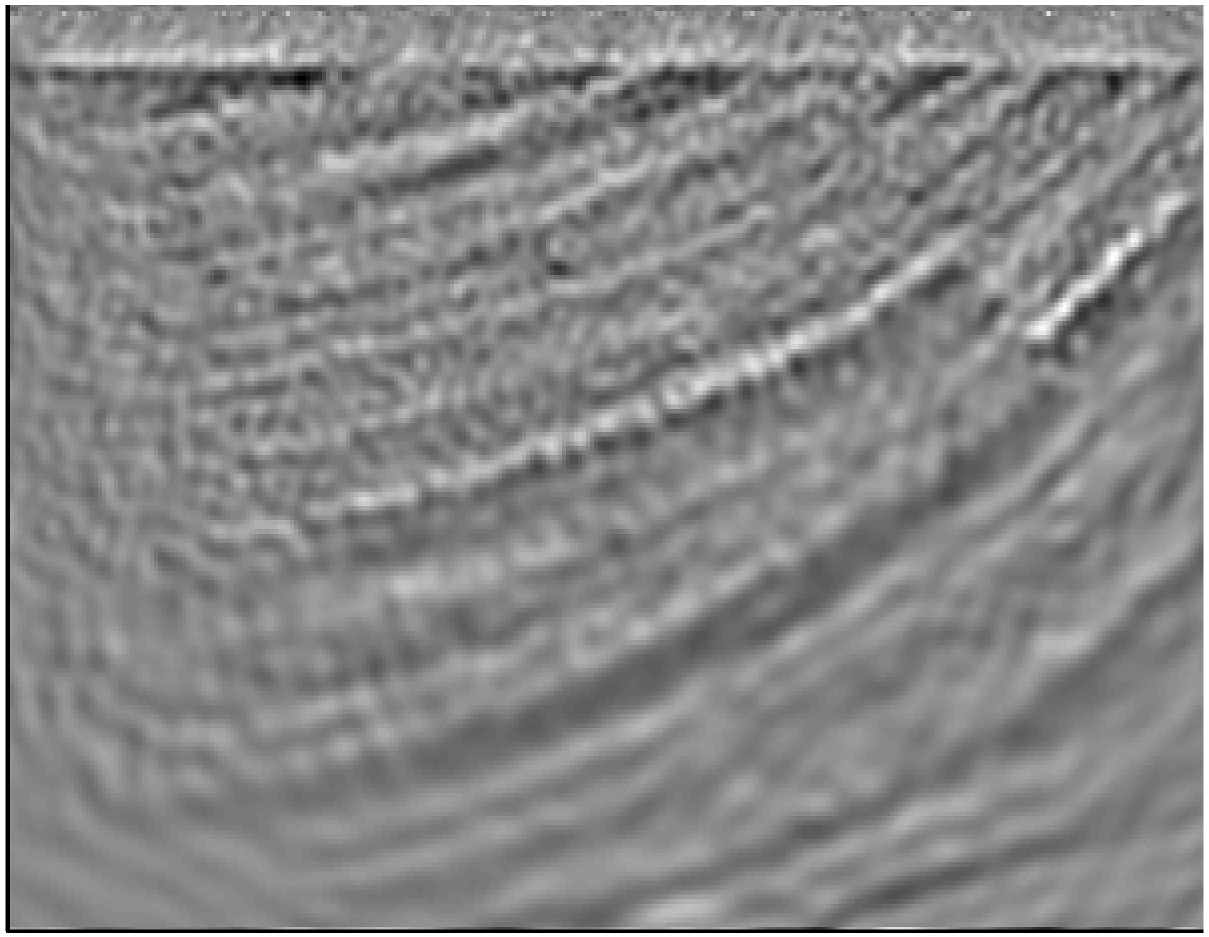
Curvelet Results

Curvelet Lasso formulation

$$\min_{\mathbf{x}} \quad \|\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m_0} + \overbrace{C^*\mathbf{x}}^{\mathbf{m}}; \mathbf{Q}]\|_F^2$$

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

$\tau = 760$

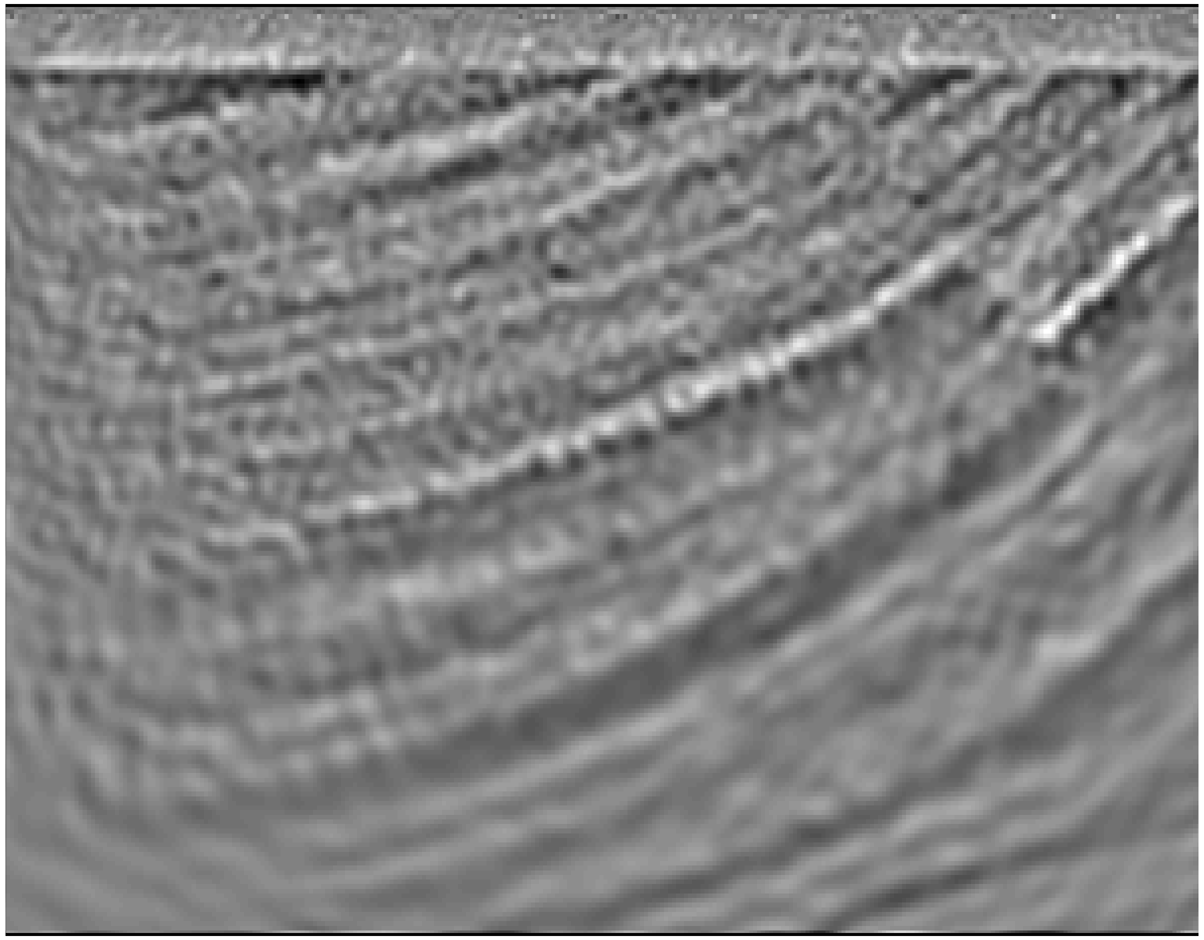


Curvelet Results

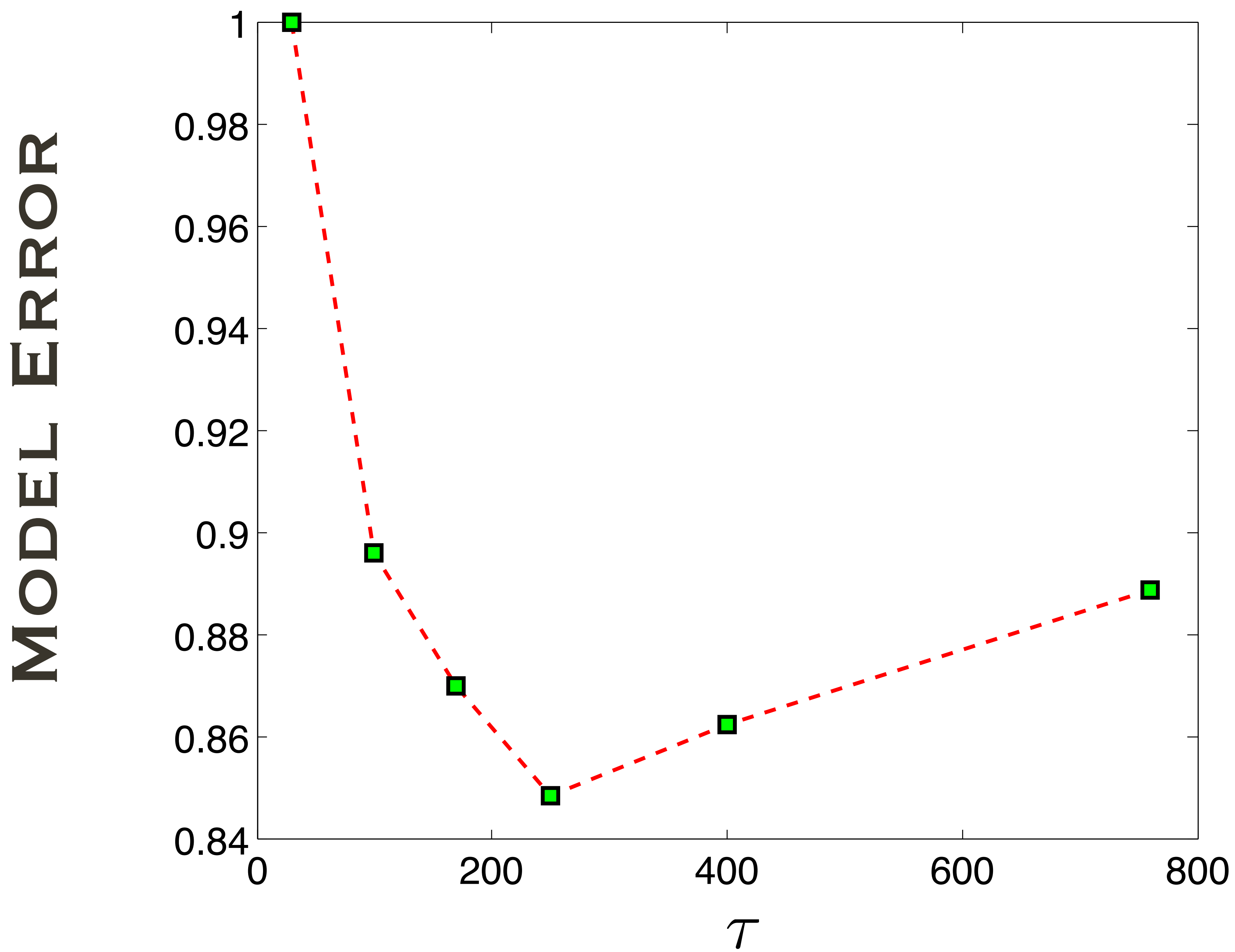
Standard FWI

$$\min_{\mathbf{m}} \quad \|\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m_0} + \mathbf{m}; \mathbf{Q}]\|_F^2$$

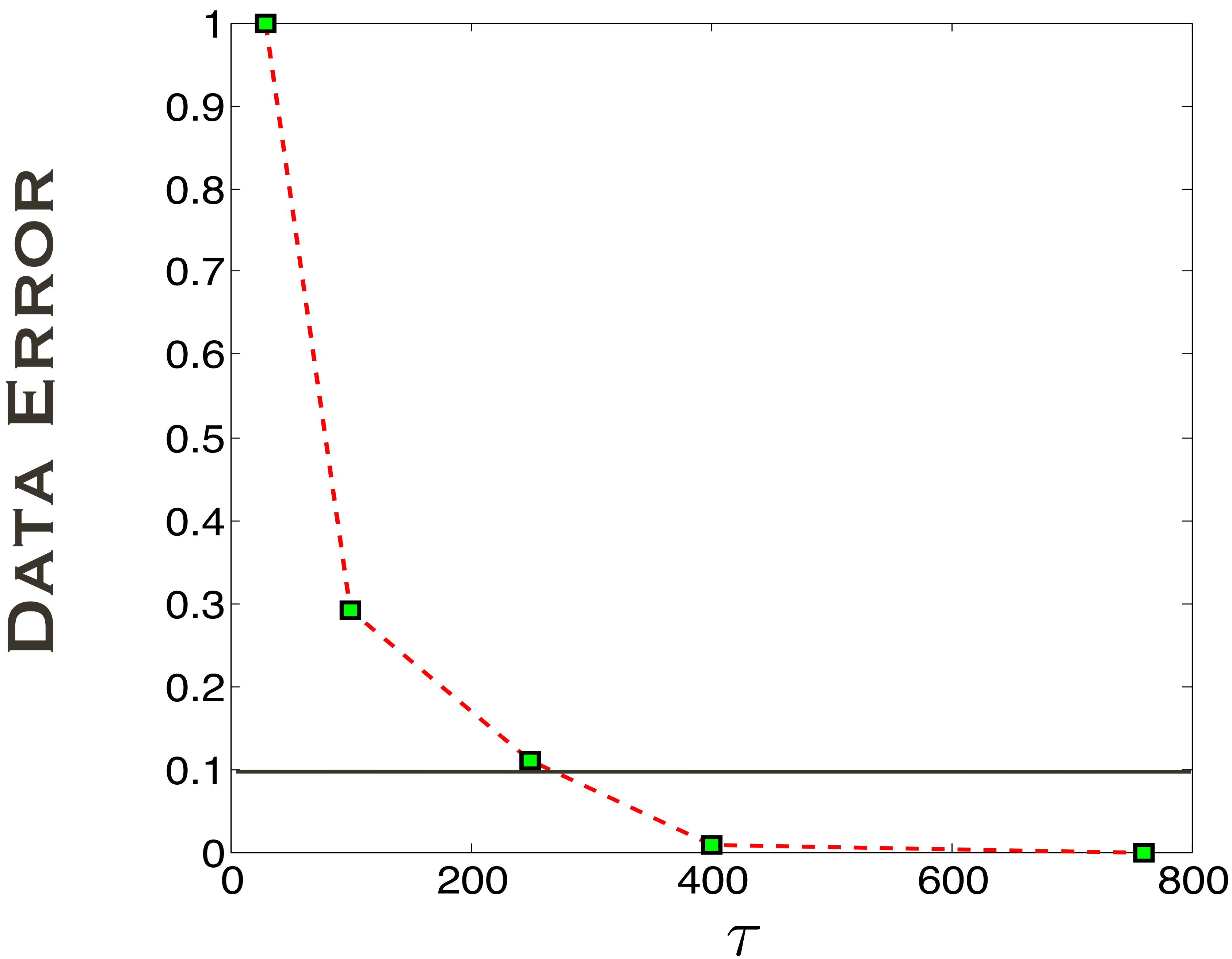
LBFGS



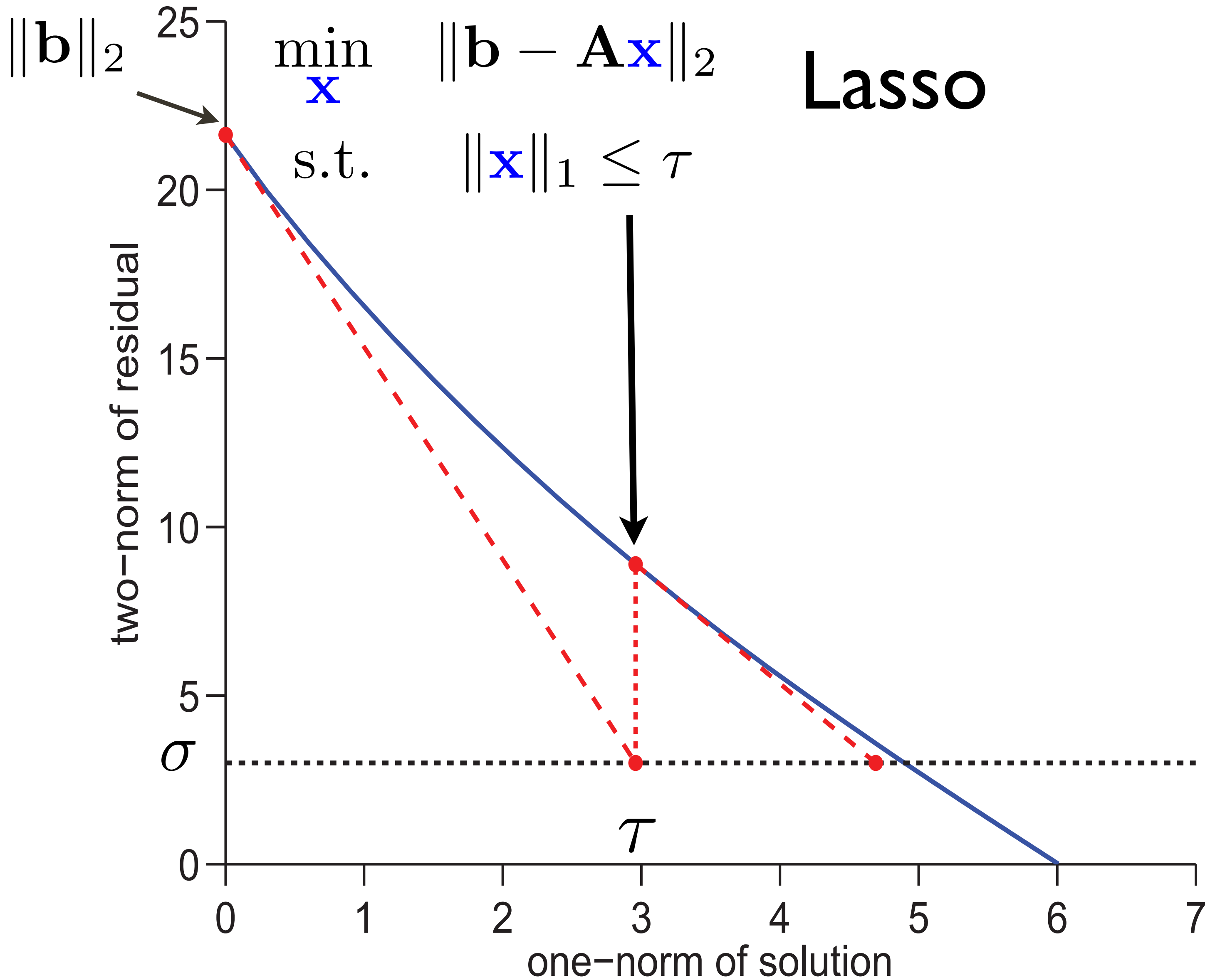
Model Error vs. Tau



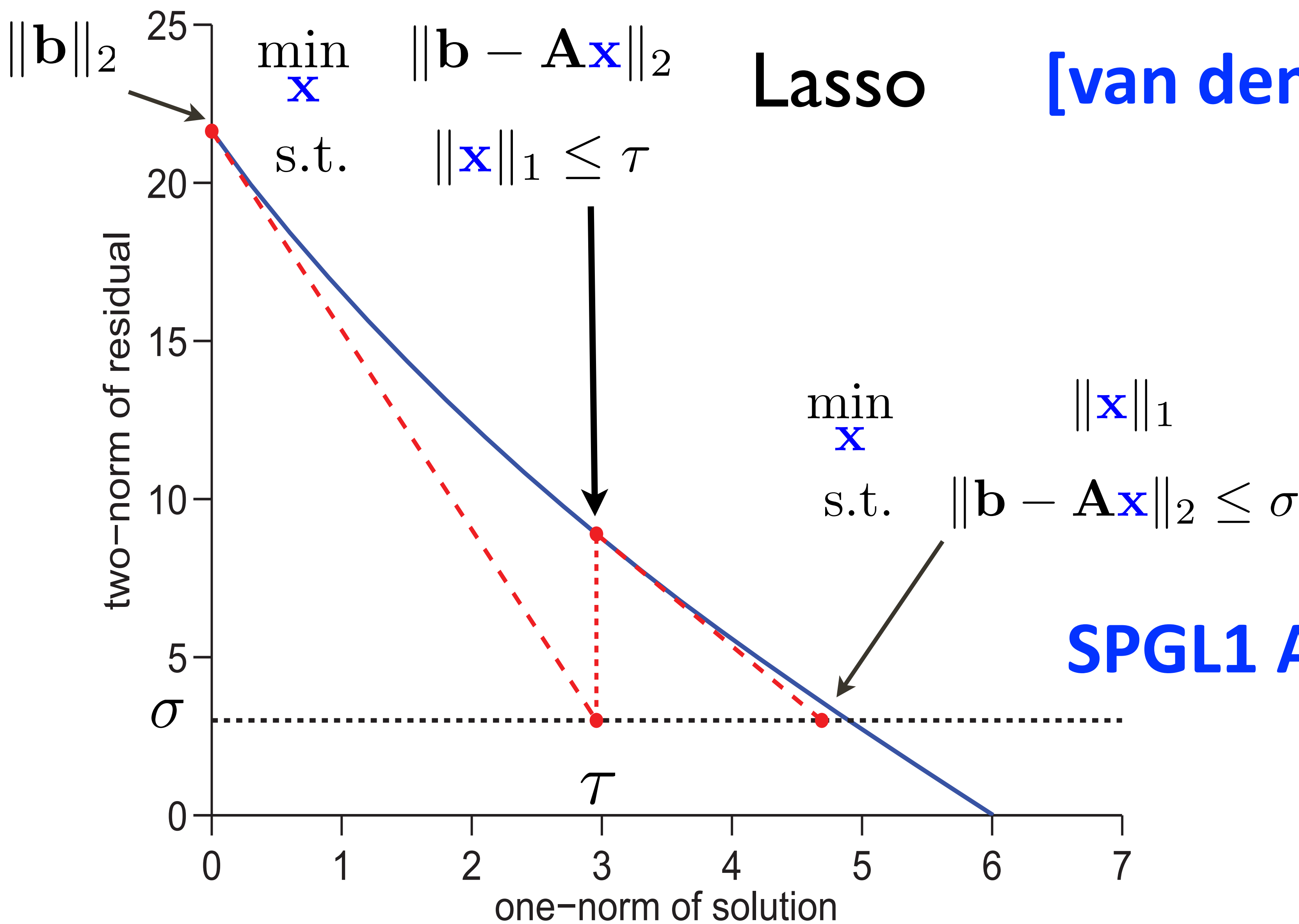
Data Error vs. Tau



Pareto Trade-Off Curve



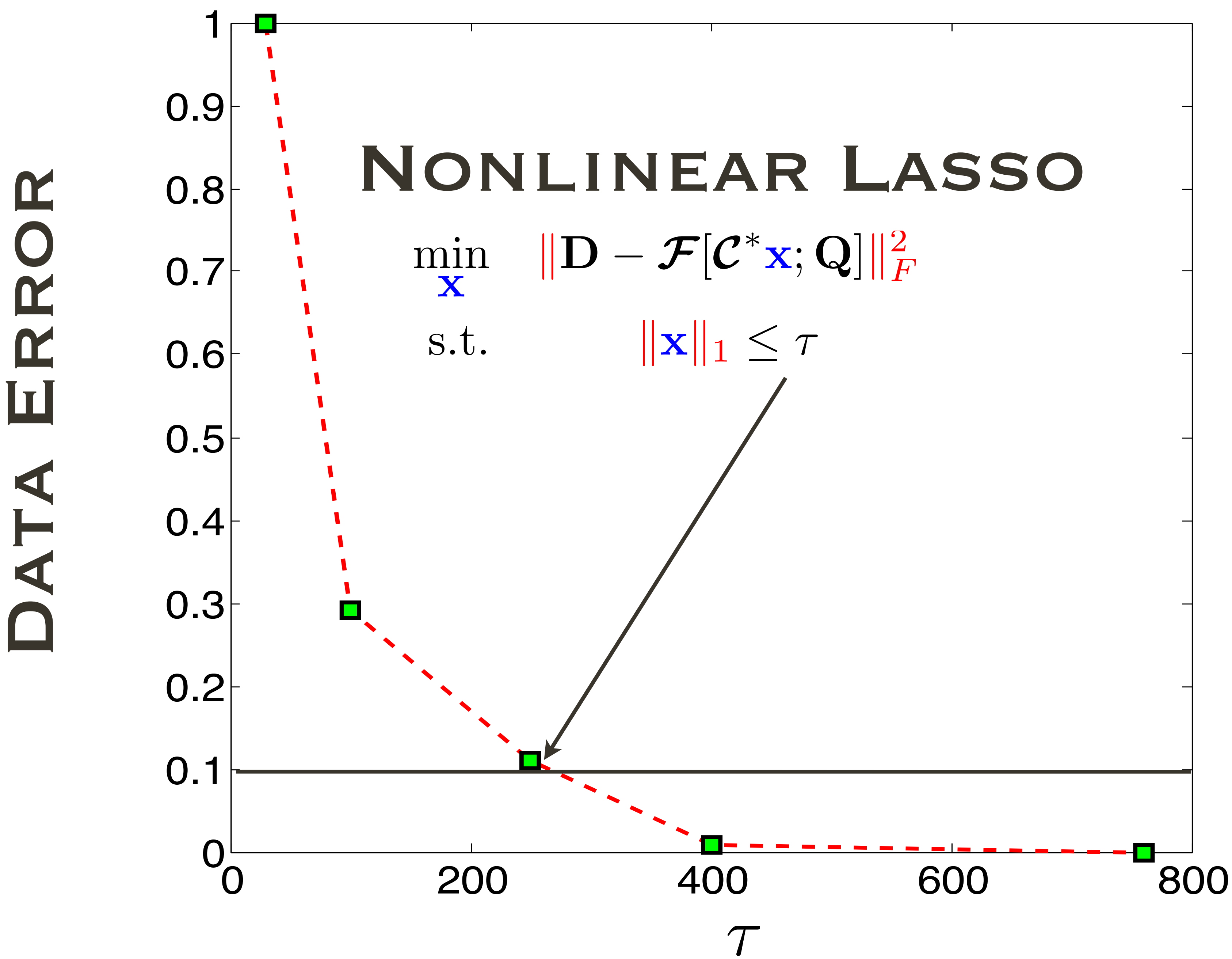
Basis Pursuit Denoise



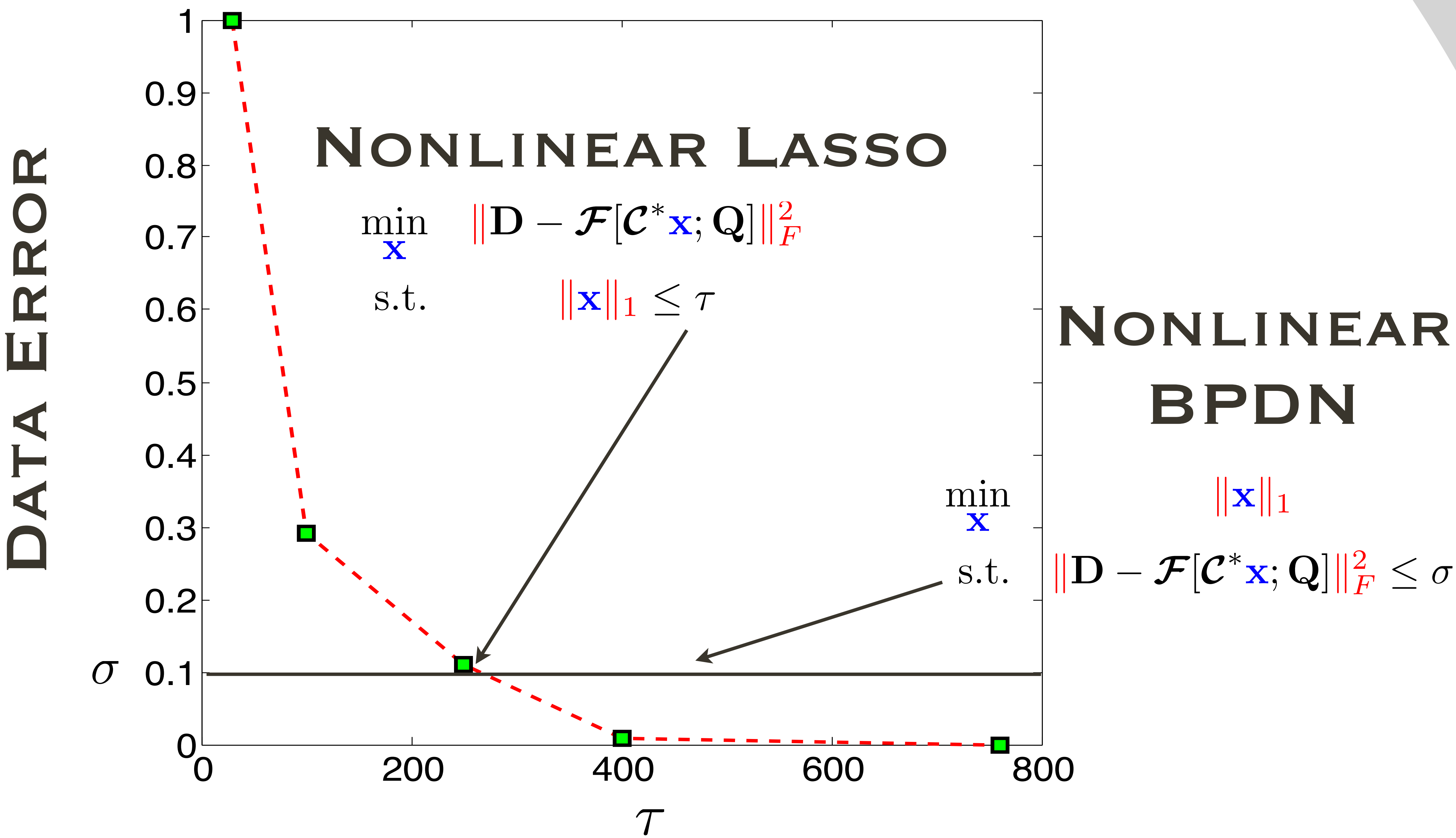
[van den Berg '08]

SPGL1 Algorithm

Nonlinear Lasso



Nonlinear BPDN



Algorithms III (Current)

- **Optimization problem:**

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

- **Implement iterated algorithm:**

$$\mathbf{m}^{\nu+1} = \mathbf{m}^{\nu} + \gamma_\nu \boldsymbol{\delta}\mathbf{m}$$

- **Direction $\boldsymbol{\delta}\mathbf{m}$ solves subproblem below using SPGL1 algorithm:**

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

[Burke ’89, Burke ’92]

Conclusions

- **Exploiting sparsity is a promising direction for modeling/regularization of FWI**
- **Preliminary results are promising: we can improve recovery from insufficient data with sparsity promotion.**
- **Understanding trade-off between NONLINEAR least-squares and model sparsity is our current focus in this work.**

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

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*

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.

M. Schmidt, E. van den Berg, M. P. Friedlander, and K. Murphy, Optimizing costly functions with simple constrains: a limited memory projected quasi-Newton algorithm. *Proc. of the 12th Inter. Conf. on Artificial Intelligence and Statistics (AISTATS) 2009, J. Machine Learning Research, W&CP 5, April 2009.*

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

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

van den 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*