SINBAD Consortium 2010

Convex Composite Optimization

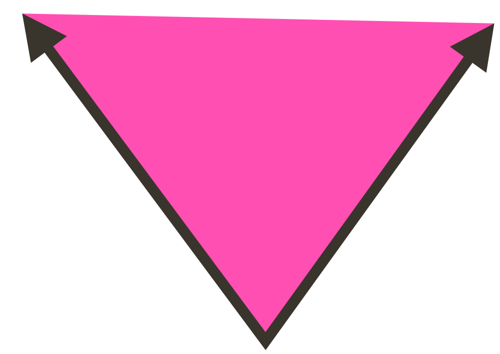
Aleksandr Aravkin

SLIM
University of British Columbia

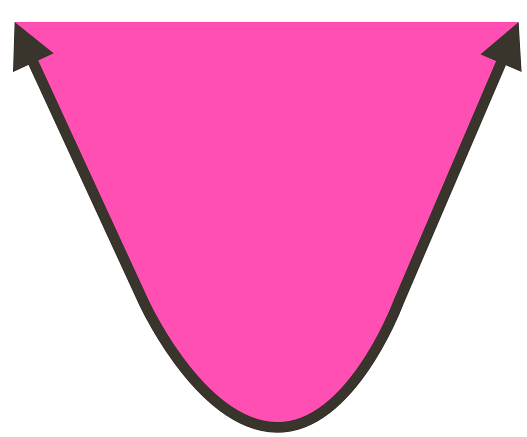
Convexity

- Key Concept in optimization, both in theory and algorithm design
- More important than differentiability
- Powerful algorithms and software for convex problems have been developed

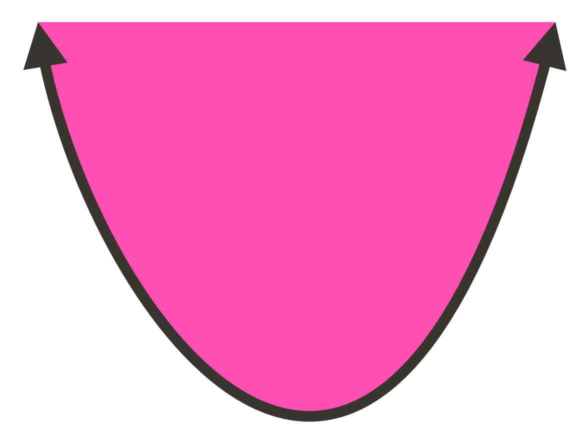
$f(x) = \|x\|_1$



$f(x) = \text{Huber}$



$f(x) = \frac{1}{2}\|x\|_2^2$



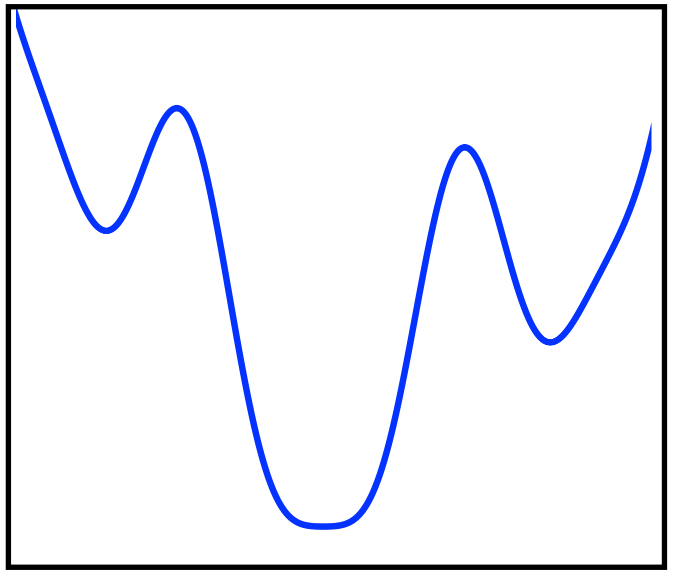
Beyond Convex

- Often we want to minimize functions that are NOT convex
- Example: $f(m) = \|D - RH^{-1}[m]Q\|_F^2$
- Even so, some convex structure remains:

$$f(m) = {\color{red}\|}{\color{blue}D - RH^{-1}[m]Q}{\color{red}\|_F^2}$$

RED: CONVEX
BLUE: DIFFERENTIABLE

SCALAR EXAMPLE



VELOCITY

Convex Composite

- Convex composite optimization:

$$f(m) = h(g(m))$$

h	CONVEX
g	DIFF.

- Includes all smooth functions: $h(x) = x$

- Includes all convex functions: $g(m) = m$

- Includes many non-smooth non-convex problems

- Why is this interesting?

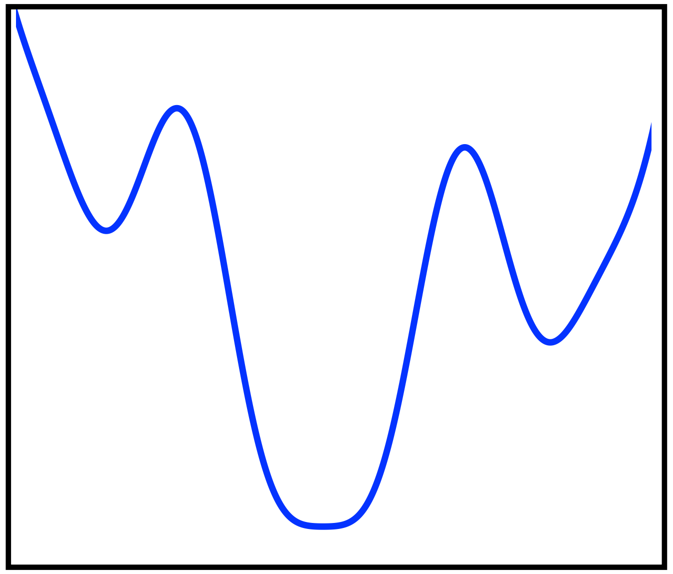
Extensions to FWI

- We would like to consider formulations

$$\min_m \rho_{\text{noise}}(D - RH^{-1}[m]Q) + \rho_{\text{reg}}(m)$$

- Two key applications: Robustness to outliers and sparsity regularization.

Scalar Example



Robust FWI

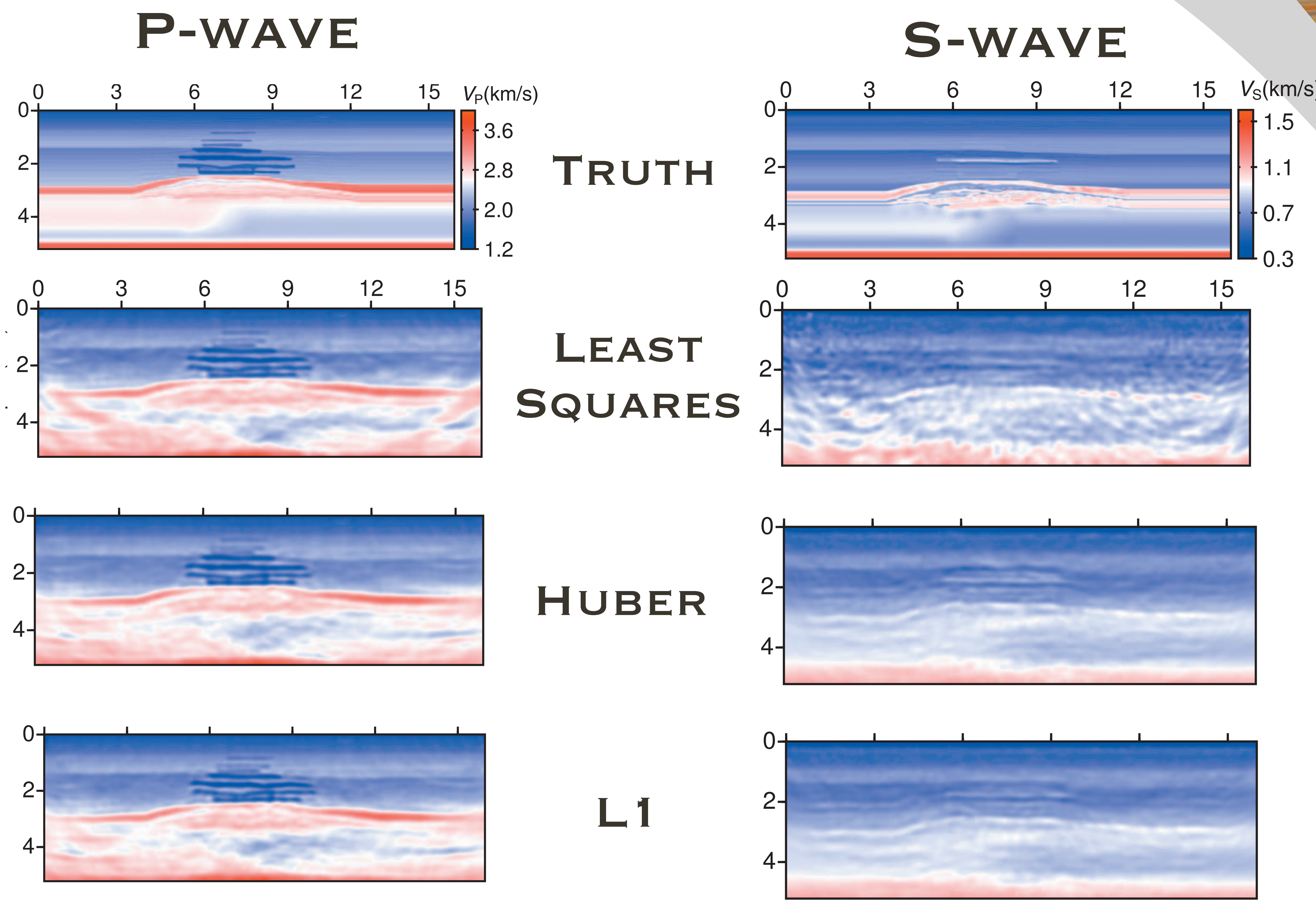
- The standard FWI formulation $(\rho_{\text{noise}} = \| \cdot \|_F^2)$ is equivalent to a normal error model for measurement errors:

$$D = RH^{-1}[m]Q + \epsilon$$

$$\epsilon \sim \mathcal{N}(0, I)$$

- Data error in the real world may not have this distribution, may have large outliers, or may have large systematic features we cannot model.

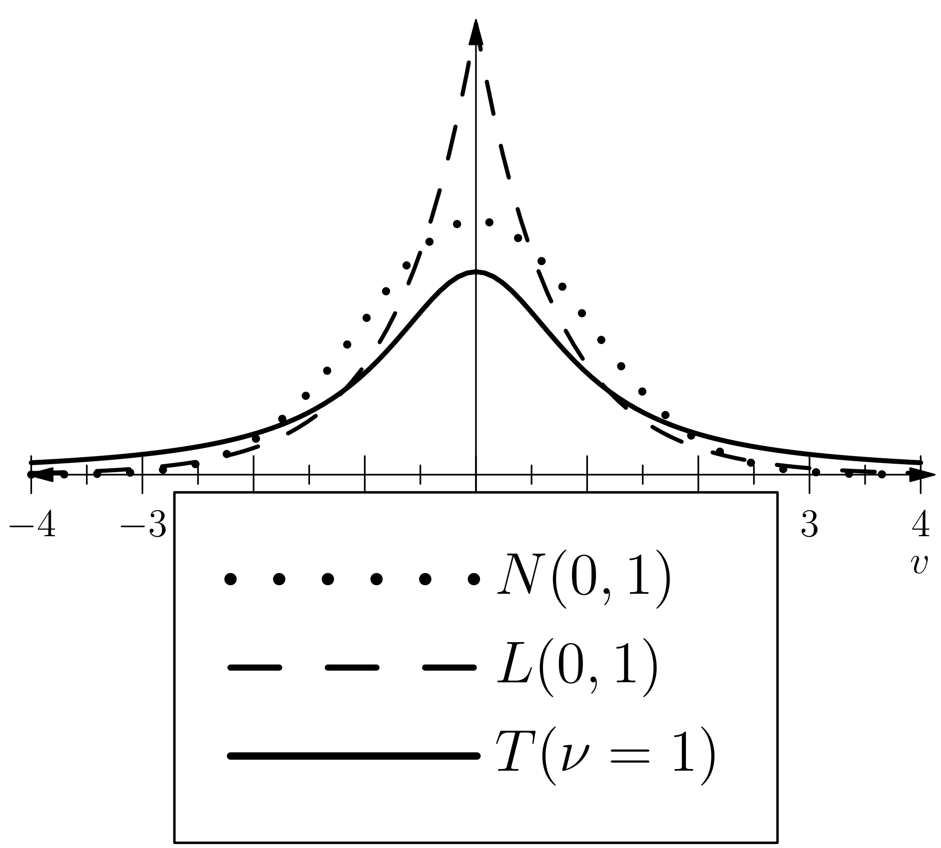
[Brossier et al. 2010, Guitton & Symes 2003]

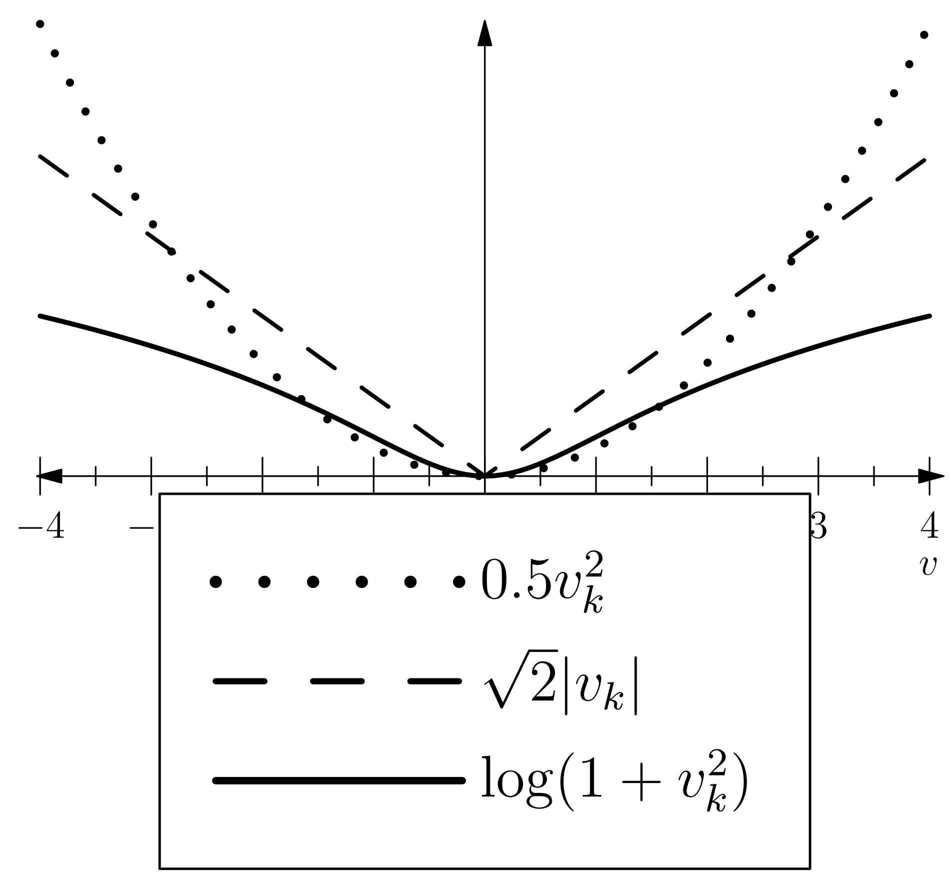


- When there are outliers in the data, Huber and L1 recover better.

[Brossier et al. 2010]

Heavy Tailed Modeling



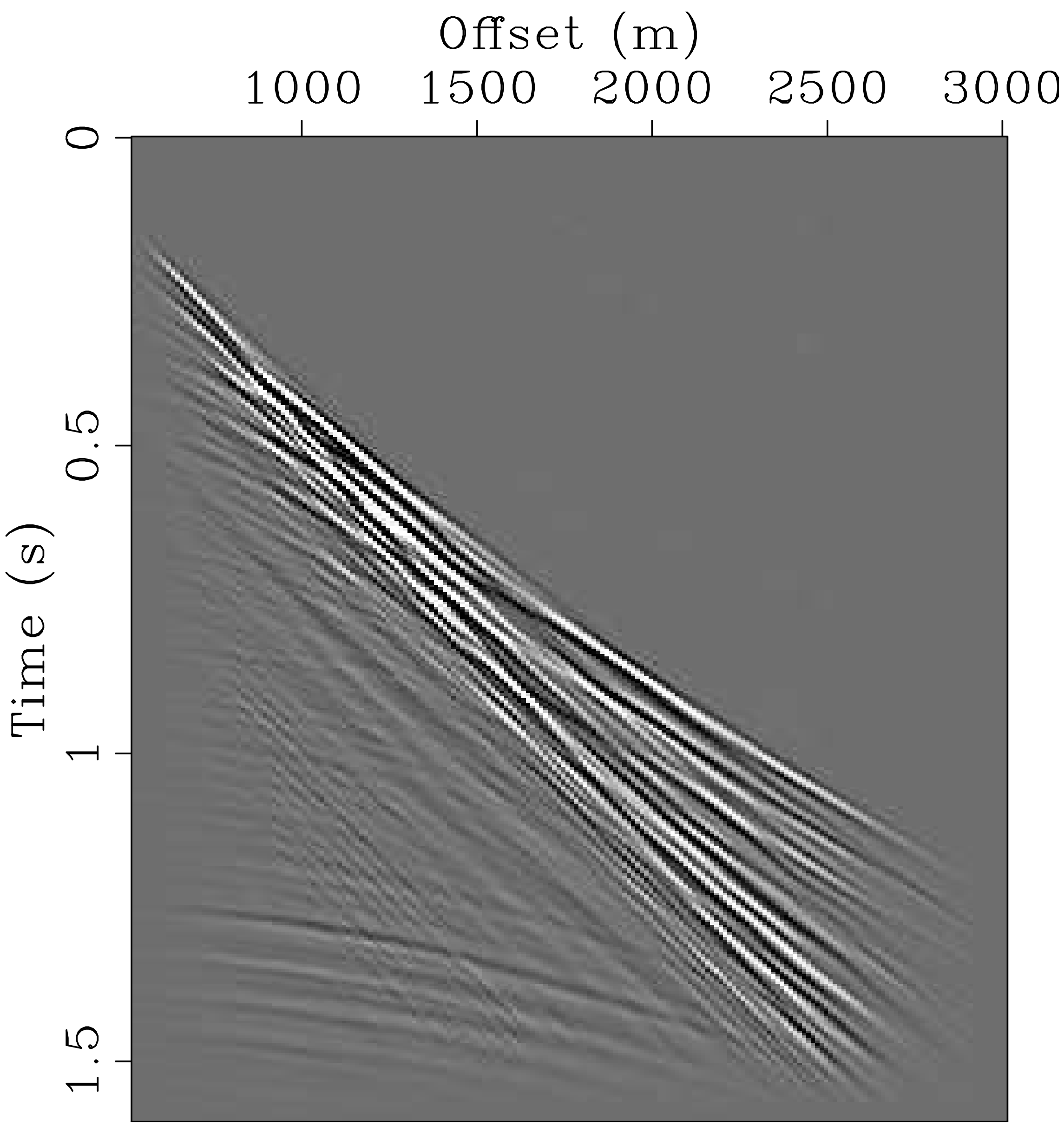


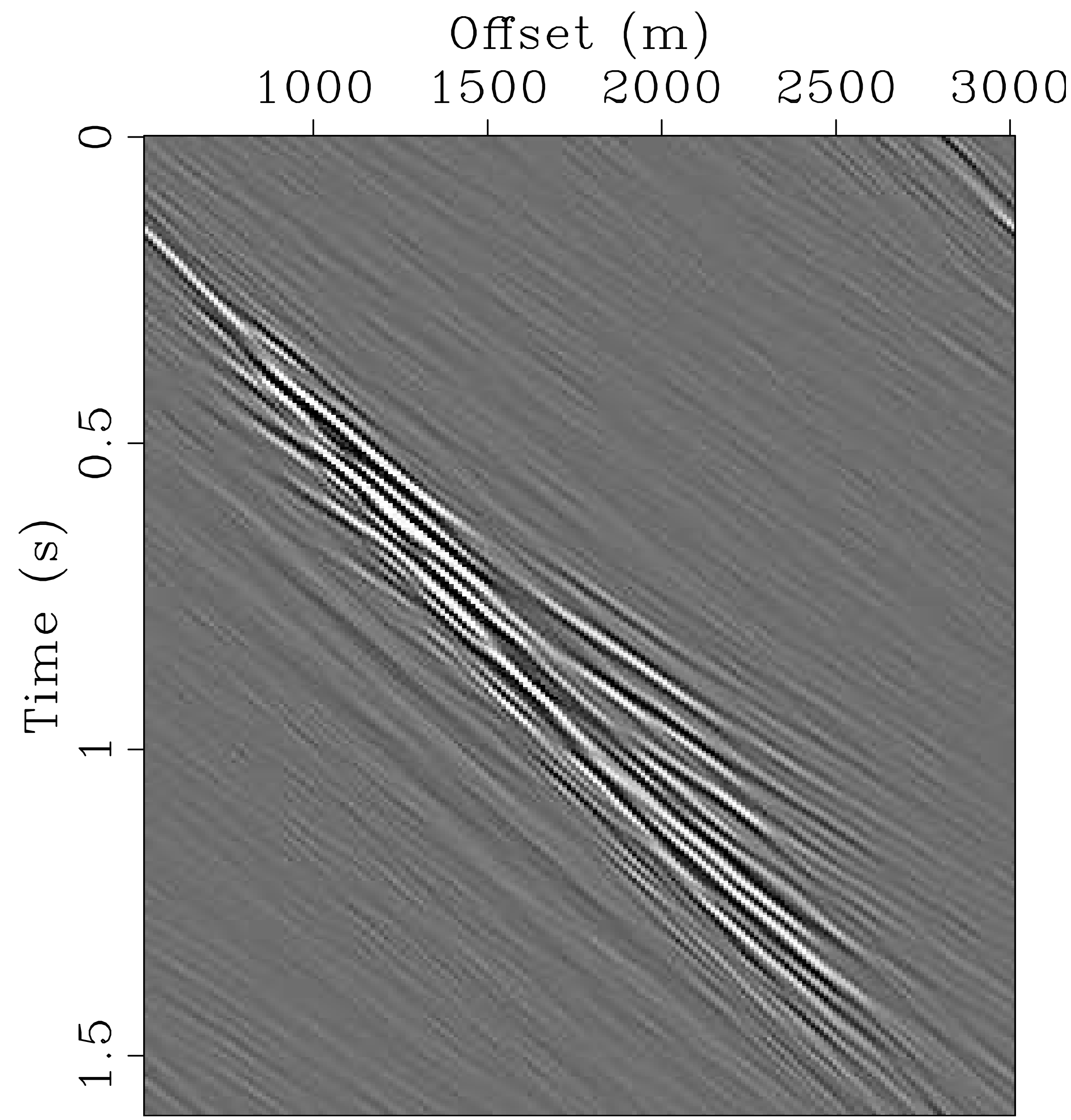
[Aravkin 2010]

Regularized FWI

- The standard FWI formulation $(\rho_{\text{error}} = 0)$ does not exploit the fact that the solutions to FWI are sparse in Fourier, Wavelets, and Curvelets
- Adding a sparsity regularization term $(\rho_{\text{error}} = \|C * x\|_1)$ is a good regularization strategy for FWI
- Resulting problems are again Convex Composite!

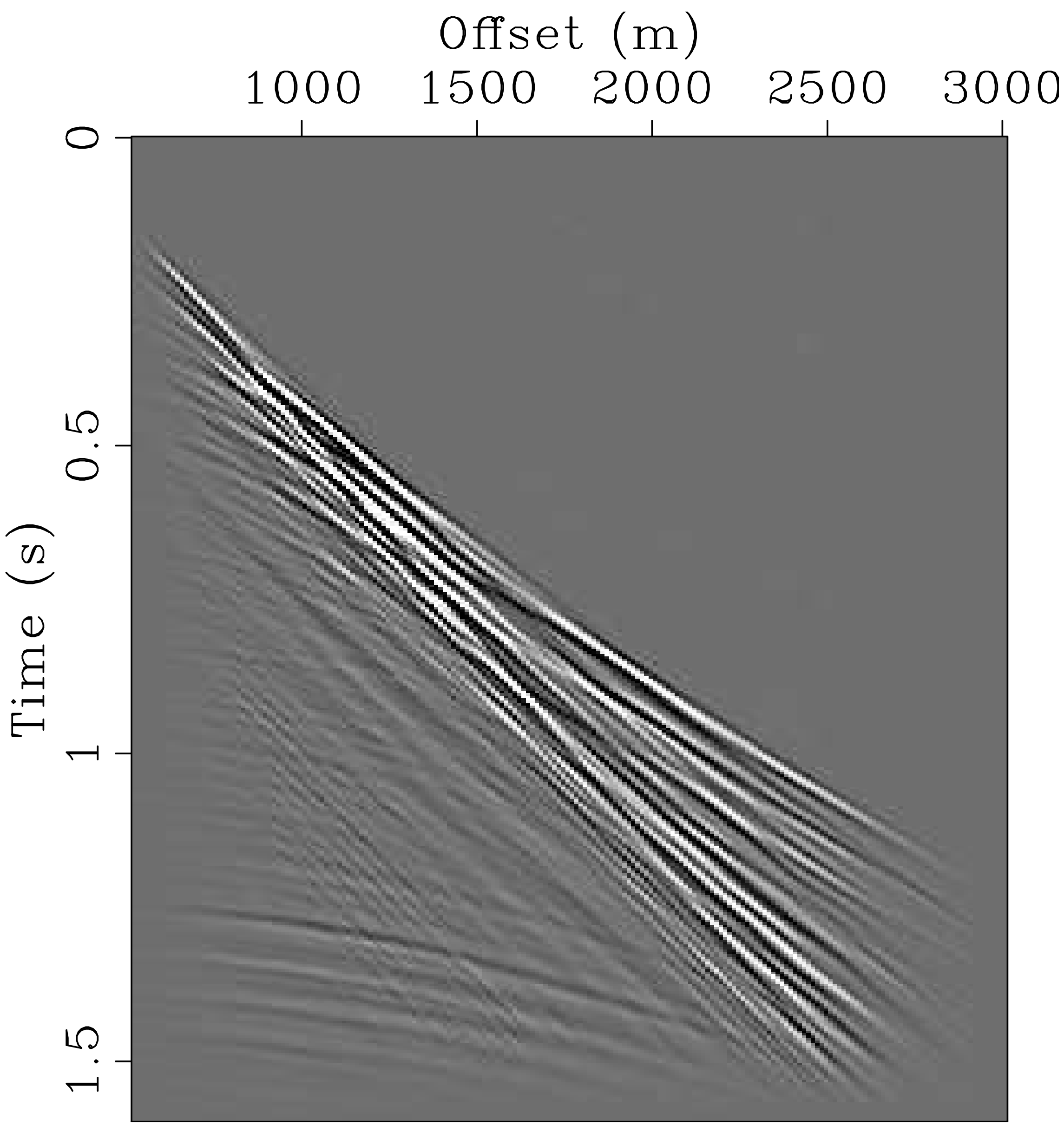
Fourier reconstruction

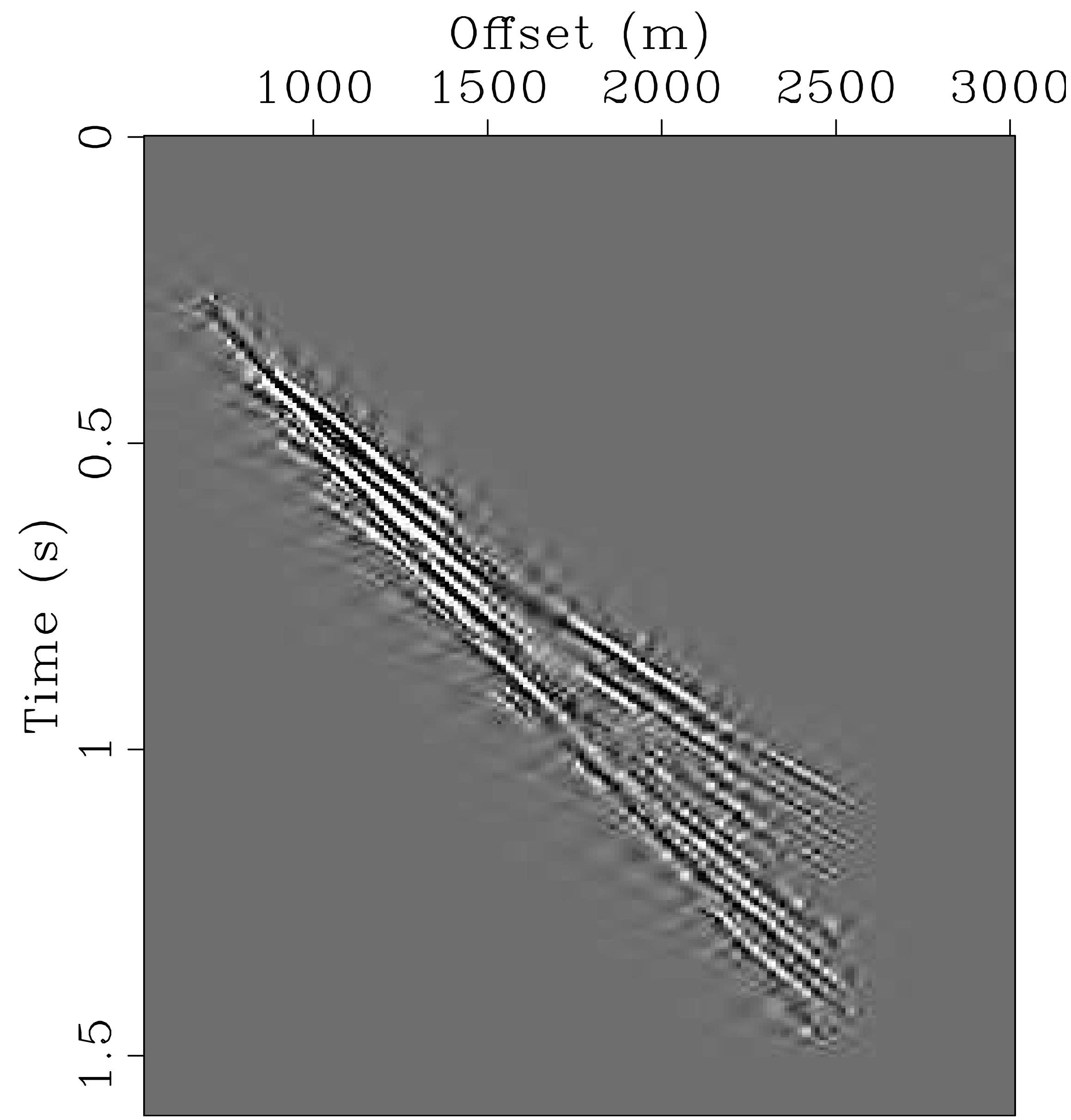




1 % of coefficients

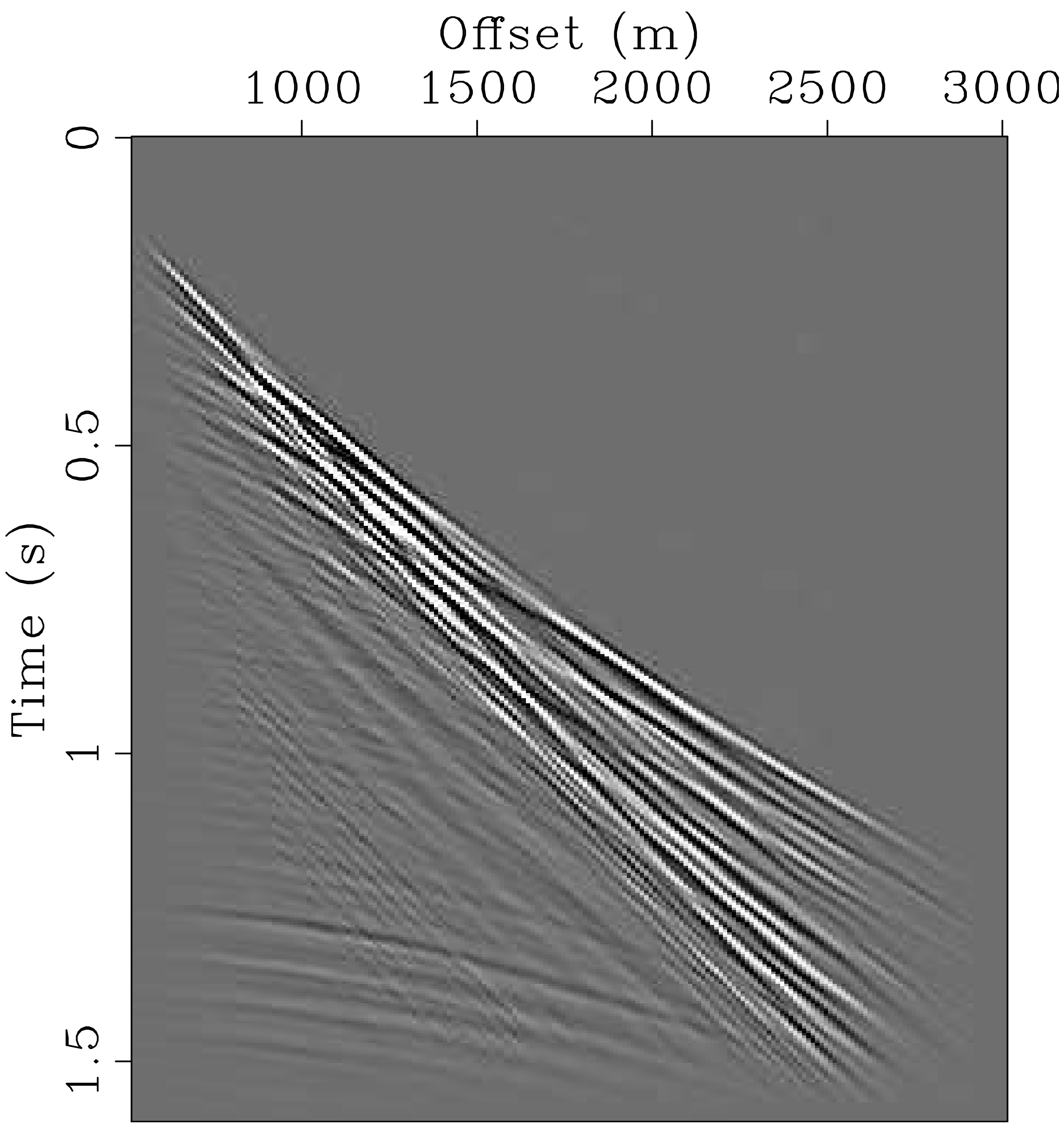
Wavelet reconstruction

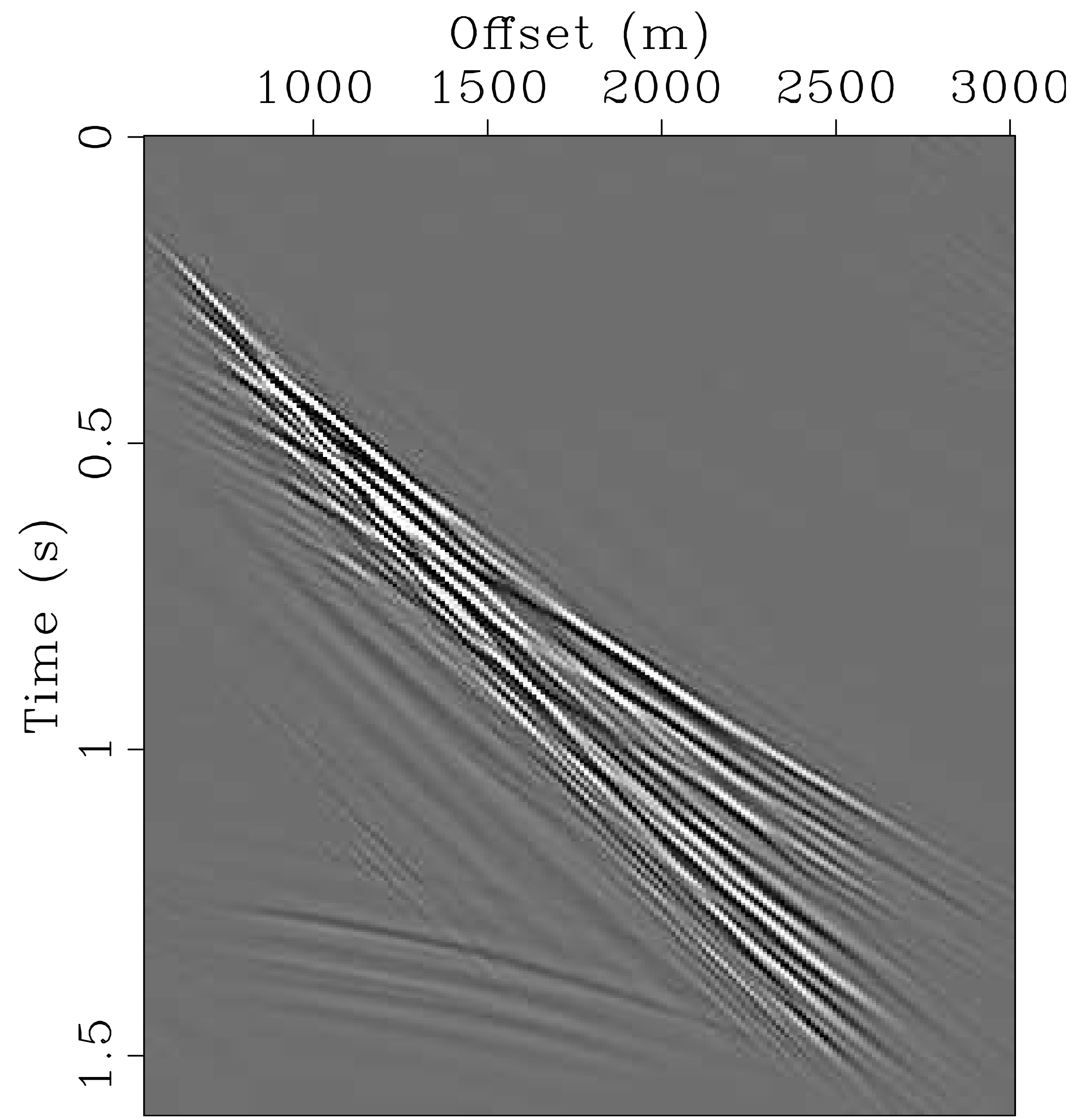




1 % of coefficients

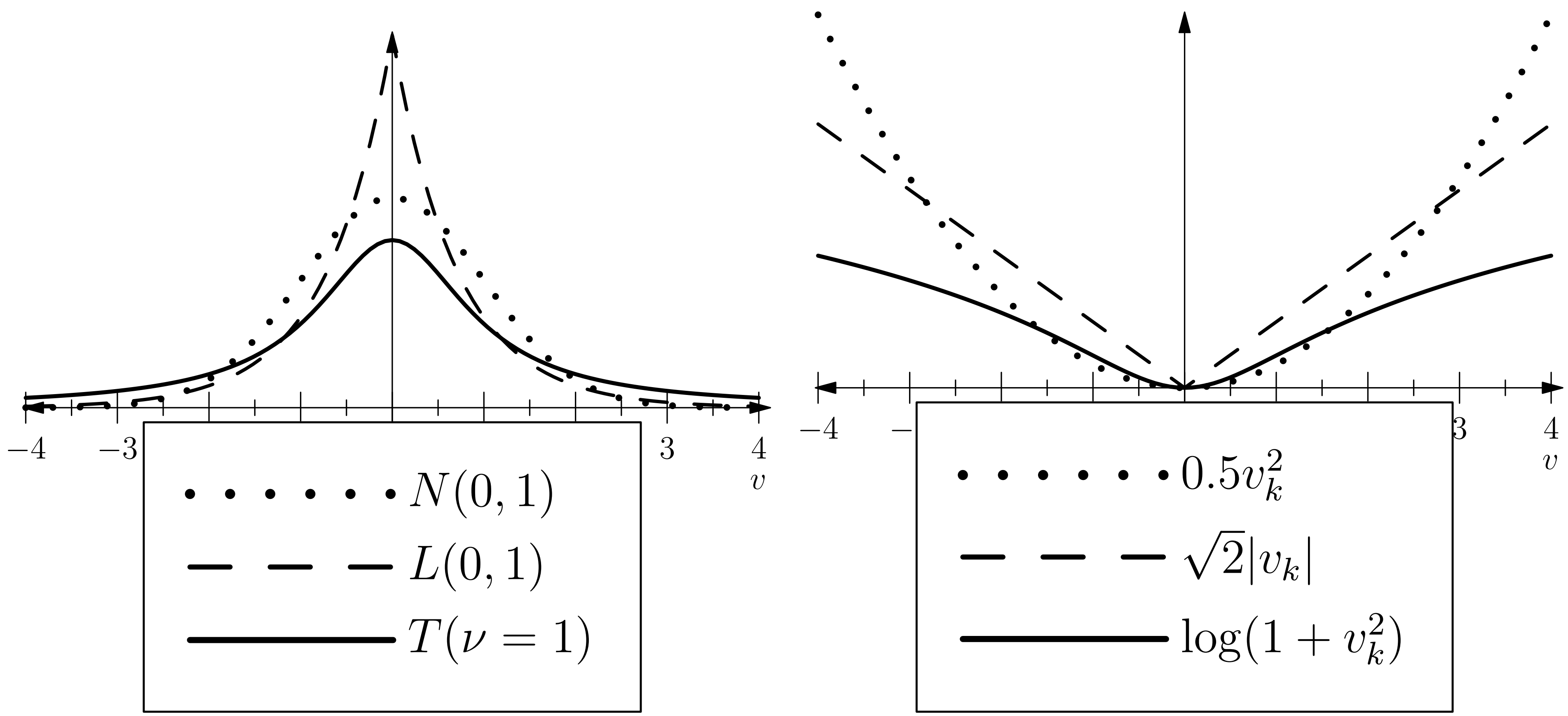
Curvelet reconstruction





1 % of coefficients

Behavior at 0



[Aravkin 2010]

Convex Composite

- Convex composite optimization:

$$f(m) = h(g(m))$$

h	Convex
g	Diff.

- Includes many non-smooth non-convex problems, e.g. Robust/Sparsity Regularized FWI

- How do we exploit this structure to design algorithms?

Gauss-Newton Method

- Objective: $$\min_{m} \|d - g(m)\|_2^2$$
- Iterative algorithm: $$m^{\nu+1} = m^{\nu} + \gamma_{\nu} s^{\nu}$$
- Direction s^{ν} solves $$\min_{\delta m} \|d - g(m^{\nu}) - \nabla g(m^{\nu})\delta m\|_2^2$$
- The subproblem for s^{ν} is convex!

CONVEX IN δm

SLIM

Extension of Gauss-Newton

- Objective: $$\min_{m} h(d - g(m))$$
- Iterative algorithm: $$m^{\nu+1} = m^{\nu} + \gamma_{\nu} s^{\nu}$$
- Direction s^{ν} solves $$\min_{\delta m} h(d - g(m^{\nu}) - \nabla g(m^{\nu})\delta m)$$
- The subproblem for s^{ν} is convex! **Convex in δm**

[Burke & Ferris 1993]

Examples

Convex Composite Objective	Convex Subproblem	
$\|d - g(m)\|_2^2 + \lambda\|m\|_1$	$\|d - g(m^\nu) - \nabla g(m^\nu)\delta m\|_2^2 + \lambda\|m + \delta m\|_1$	Sparsity promotion
$\rho_H(d - g(m))$	$\rho_H(d - g(m^\nu) - \nabla g(m^\nu)\delta m)$	Robust Optimization
$\rho_H(d - g(m)) + \lambda\|m\|_1$	$\rho_H(d - g(m^\nu) - \nabla g(m^\nu)\delta m) + \lambda\|m + \delta m\|_1$	Sparse & Robust

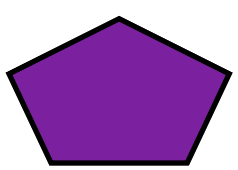
Constrained Optimization

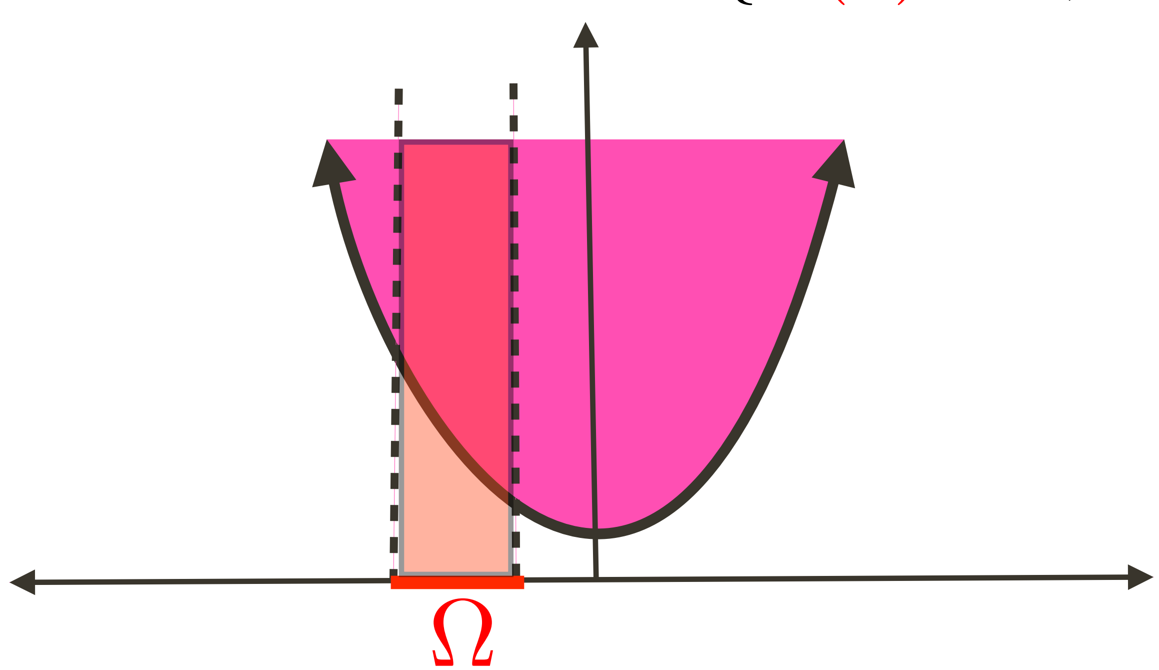
- Convex programs:

$$\begin{aligned} \min \quad & h(x) \\ \text{s.t.} \quad & x \in \Omega \end{aligned}$$

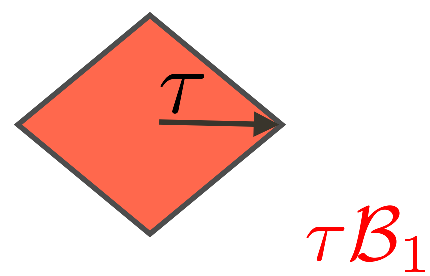
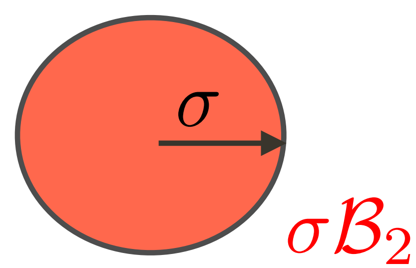
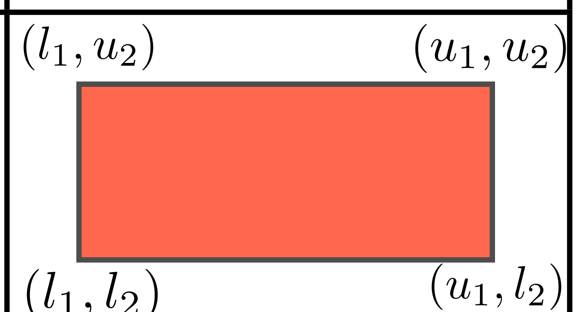
- More explicit representation: $\Omega = \{c_i(x) \leq 0,\ Ax = b\}$

- Polyhedral constraints: $\Omega = \{C(x) \leq c,\ Ax = b\}$





Examples

Formulation	Constraint Region	
$\min \; \|Ax-b\|_2^2$ s.t. $\|x\|_1 \leq \tau$ $(x \in \tau\mathcal{B}_{L_1})$		Lasso (Statistics)
$\min \; \|x\|_1$ s.t. $\|Ax-b\|_2 \leq \sigma$ $(Ax-b \in \sigma\mathcal{B}_2)$		Basis Pursuit Denoise
$\min \; \|Ax-b\|_2^2$ s.t. $l \leq x \leq u$		Regression w/Box Constr.

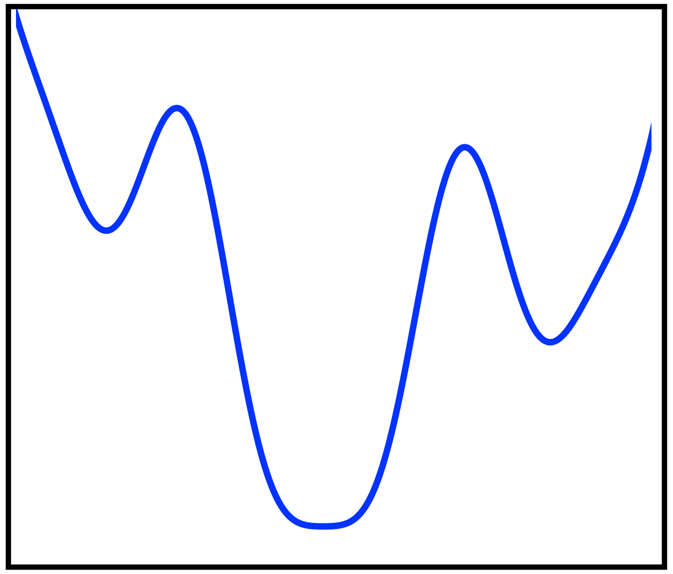
Back to FWI

- Suppose we wanted a LASSO type constraint for FWI in the Curvelet frame:

$$\begin{array}{ll} \min & \|D - RH^{-1}[C^* x]Q\|_F^2 \\ \text{s.t.} & \|x\|_1 \leq \tau \\ & (x \in \tau\mathcal{B}_1) \end{array}$$

- We are still in the convex composite setting!

Scalar Example



More Convex Composite

- Objective:

$$\begin{array}{ll} \min & h(d - g(m)) \\ \text{s.t.} & f(m) \in \Omega \end{array}$$

- Iterative algorithm:

$$m^{\nu+1} = m^{\nu} + \gamma_{\nu} s^{\nu}$$

- Direction s^{ν} solves

$$\begin{array}{ll} \min_{\delta m} & h(d - g(m^{\nu}) - \nabla g(m^{\nu})\delta m) \\ \text{s.t.} & f(m^{\nu}) + \nabla f(m^{\nu})\delta m \in \Omega \end{array}$$

- The subproblem for s^{ν} is convex!

CONVEX IN δm

[Burke 1989]

More Examples

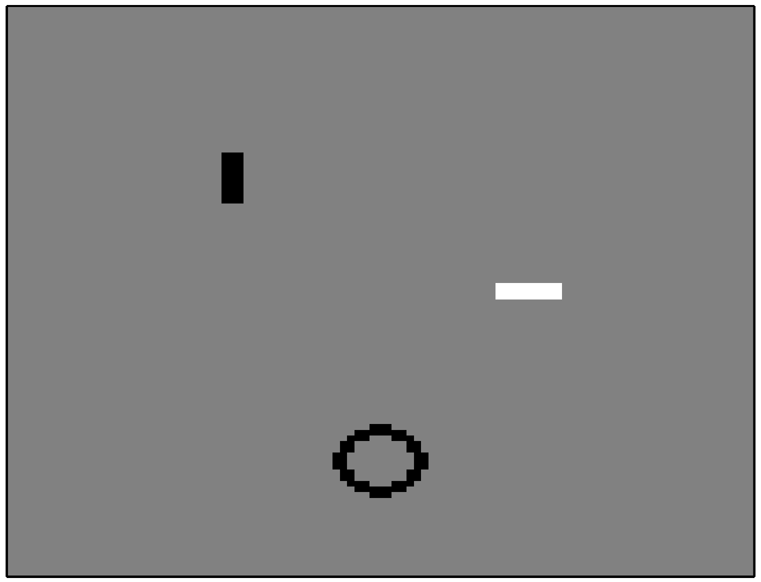
Convex Composite Objective	Convex Subproblem	
$\min_{m} \; \|d - g(m)\|_2^2$ s.t. $\|m\|_1 \leq \tau$	$\min_{\delta m} \; \|d - g(m^\nu) - \nabla g(m^\nu)\delta m\|_2^2$ s.t. $\|m^\nu + \delta m\|_1 \leq \tau$	Nonlinear Lasso
$\min_{m} \; \|m\|_1$ s.t. $\|d - g(m)\|_2 \leq \sigma$	$\min_{\delta m} \; \|m^\nu + \delta m\|_1$ s.t. $\|d - g(m^\nu) - \nabla g(m^\nu)\delta m\|_2 \leq \sigma$	Nonlinear BPDN
$\min_{m} \; \rho_H(d - g(m))$ s.t. $l \leq m \leq u$	$\min_{\delta m} \; \rho_H(d - g(m^\nu) - \nabla g(m^\nu)\delta m)$ s.t. $l \leq m^\nu + \delta m \leq u$	Robust & Box Constr.

Exploiting Sparsity in FWI: Nonlinear BPDN

SINBAD Consortium 2010

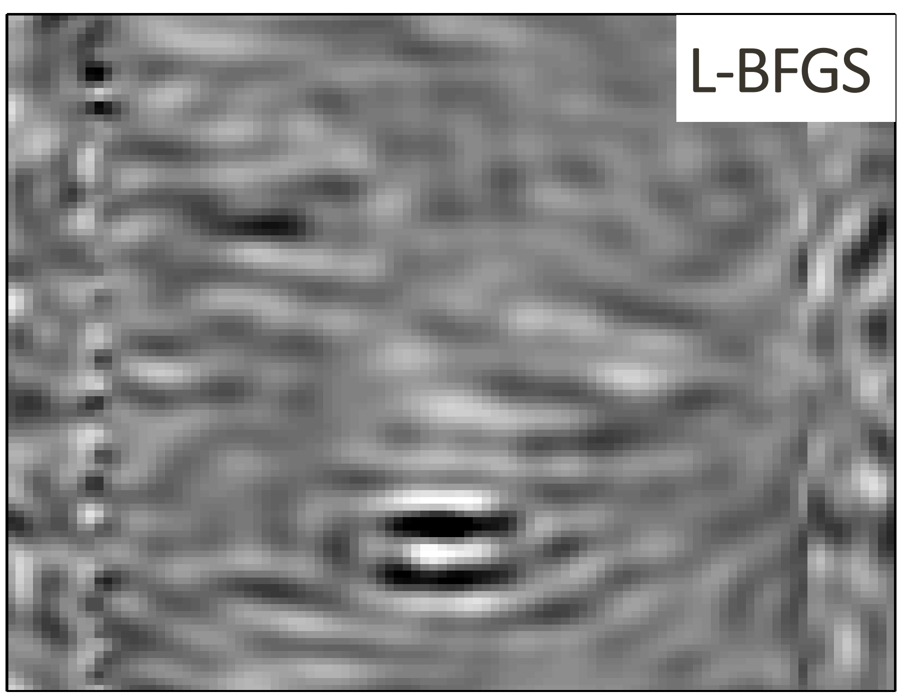
Presenter: Aleksandr Aravkin & Tristan van Leeuwen

True Model

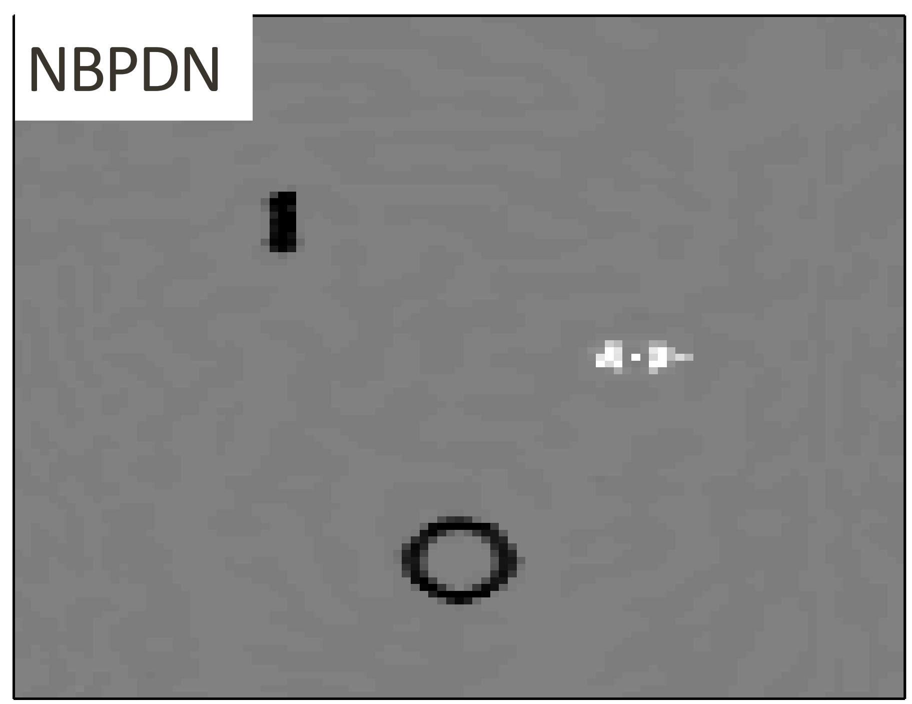


- Cross-well setting, 101 sources and receivers
- True model is sparse w.r.t to a constant reference model in pixel space
- Use of simultaneous shots to reduce computational load
- Compare to L-BFGS recovery without sparsity constraints

Simulateneous Shots: 5



Simulateneous Shots: 5



The Road Ahead

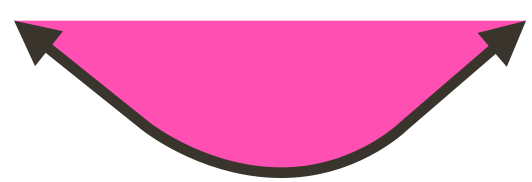
- Convex Composite optimization provides a solid theoretical framework for algorithm development

- Applications to FWI :

    1) Sparsity promotion (NBPDN)

$$\min_{m} \ \|m\|_1 \quad \text{s.t.} \quad \|d - g(m)\|_2 \leq \sigma$$

    2) Outlier-robust misfit (e.g. Huber)



    3) Using prior information (e.g. constraints for velocity modeling).

Acknowledgements

- James V. Burke (UW Mathematics)

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

Aravkin, Robust methods for Kalman filtering/smoothing and Bundle Adjustment, Ph.D. Thesis, University of Washington

Brossier, R., Operto, S., and Virieux, J., Which data residual norm for robust elastic frequency-domain full waveform inversion? *Geophysics, 75,3:37-46, 2010*

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*

Guitton, A., and Symes, W., Robust inversion of seismic data using the Huber norm, *Geophysics, 68,4:1310-1319, 2003*

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

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