WAVES 2011

Robust FWI Using Student's t-distribution

Aleksandr Aravkin, Tristan van Leeuwen, Felix Herrmann

SLIM
University of British Columbia

Full Waveform Inversion

- The Full Waveform Inversion (FWI) problem is to estimate subsurface velocity parameters for which solutions to the corresponding Helmholtz PDE best match data from source experiments.

$$\mathbf{H}_{\omega}[\mathbf{m}]\mathbf{u} = [\omega^2\mathbf{m} + \nabla^2]\mathbf{u}$$

- Problems are very large: billions of variables and terabytes of data.

- FWI is typically formulated as a Nonlinear Least Squares (NLLS) problem

Single source monochromatic:

$$\min_{\mathbf{m},\mathbf{u}} \quad \frac{1}{2}\|\mathbf{P}\mathbf{u} - \mathbf{d}_\omega\|_2^2 \quad \text{subject to} \quad \mathbf{H}_\omega[\mathbf{m}]\mathbf{u} = \mathbf{q}_\omega$$

Variable	**Type**	**Dimension**	**Description**
$\mathbf{m}$	$\mathbb{R}$	$n_x n_z$	Model (slowness squared)
$\mathbf{H}_\omega[\mathbf{m}]$	$\mathbb{C}$	$n_x n_z \times n_x n_z$	Discrete Helmholz with boundary
$\mathbf{P}$	$\mathbb{R}$	$n_r \times n_x n_z$	Sampling operator
$\mathbf{d}_\omega$	$\mathbb{C}$	n_r	Data vector
$\mathbf{q}_\omega$	$\mathbb{C}$	$n_x n_z$	Source
$\mathbf{u}$	$\mathbb{C}$	$n_x n_z$	Wavefield

$$\min_{\mathbf{m}} \quad \phi(\mathbf{m}) := \frac{1}{2}\|\mathbf{P}\mathbf{H}_\omega[\mathbf{m}]^{-\mathbf{1}}\mathbf{q}_\omega - \mathbf{d}_\omega\|_2^2$$

Multi-source, multi-frequency FWI

$$\min_{\mathbf{m},\mathbf{U}} \quad \frac{1}{2}\|\mathcal{P}(\mathbf{U}) - \mathbf{D}\|_F^2 \quad \text{subject to} \quad \mathbf{H}[\mathbf{m}]\mathbf{U} = \mathbf{Q}$$

Variable	Type	Dimension	Description
$\mathbf{H}[\mathbf{m}]$	$\mathbb{C}$	$n_f(n_x n_z \times n_x n_z)$	$\mathrm{diag}[\mathbf{H}_{\omega_1}[\mathbf{m}], \ldots, \mathbf{H}_{\omega_{n_f}}[\mathbf{m}]]$
$\mathbf{D}$	$\mathbb{C}$	$n_f(n_r \times n_s)$	$\mathrm{stack}[\mathbf{D}_{\omega_1}, \ldots, \mathbf{D}_{\omega_{n_f}}]$
$\mathcal{P}$	$\mathbb{R}$	$n_f(n_x n_z \times n_s) \rightarrow n_f(n_r \times n_s)$	Applies $\mathcal{P}_f$ to each frequency
$\mathbf{Q}$	$\mathbb{C}$	$n_f(n_x n_z \times n_s)$	$\mathrm{stack}[\mathbf{Q}_{\omega_1}, \ldots, \mathbf{Q}_{\omega_{n_f}}]$
$\mathbf{U}$	$\mathbb{C}$	$n_f(n_x n_z \times n_s)$	$\mathrm{stack}[\mathbf{U}_{\omega_1}, \ldots, \mathbf{U}_{\omega_{n_f}}]$

$$\min_{\mathbf{m}} \quad \phi(\mathbf{m}) := \frac{1}{2}\| \underbrace{\mathcal{P}(\mathbf{H}[\mathbf{m}]^{-1}\mathbf{Q})}_{\mathcal{F}[\mathbf{m},\mathbf{Q}]} - \mathbf{D}\|_F^2$$

Statistical Implications

- The NLLS formulation is equivalent to the following statistical model:

$$\begin{aligned} \mathbf{D} &= \boldsymbol{\mathcal{F}}[\mathbf{m};\mathbf{Q}] + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &\sim \mathbf{N}(0, I) \end{aligned}$$

- Equivalence follows from maximum likelihood estimate for model parameters:

$$\boldsymbol{\mathcal{L}}(\mathbf{m}) \quad \propto \quad \exp\left(-\frac{1}{2}\left\|\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m};\mathbf{Q}]\right\|_F^2\right)$$

- Minimizing the negative log likelihood is exactly the FWI problem.

- Q: So what?

Practical Consequences

- Large deviations from the mean are VERY unlikely in the Gaussian model:

	Gaussian
$p(\|x-\mu\| > 4\sigma)$	6.3×10^{-5}
$p(\|x-\mu\| > 8\sigma)$	1.3×10^{-15}

- Observations more than 4 standard deviations away from the mean occur less than .006 percent of the time.
- As we get further away, the likelihood shrinks astronomically.
- Low likelihood values correspond to HIGH penalties for outliers.

SLIM

‘Outliers’ in FWI??

- Mathematical model cannot distinguish ‘artifacts’ from ‘outliers’. Any unexplained events in the residual will have a strong effect on the final image.

- Examples:
  1) Modeling Inelastic/Anisotropic data with Acoustic PDE [Brossier et al. ’10]
  2) Ignoring Acquisition Models

- Key point: models are improving all the time, but are never perfect. It is worthwhile to have methods that still perform well when models are wrong.

- Q: How do we design such methods?

SLIM

Statistical Modeling

- We can tweak the assumptions on the noise in the model:

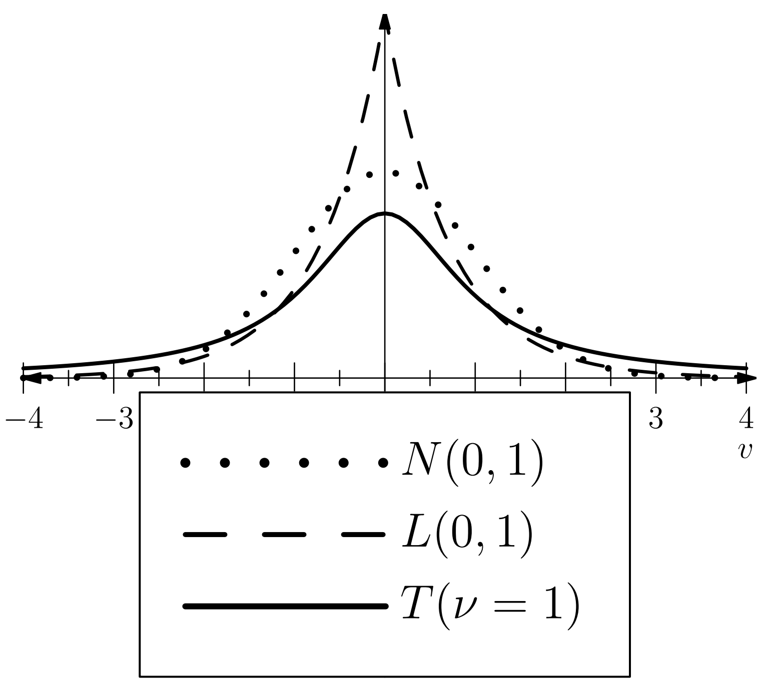
$$\begin{aligned} \mathbf{D} &= \mathcal{F}[\mathbf{m};\mathbf{Q}] + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &\sim \text{Heavy Tailed Distribution with density } \mathbf{p} \end{aligned}$$

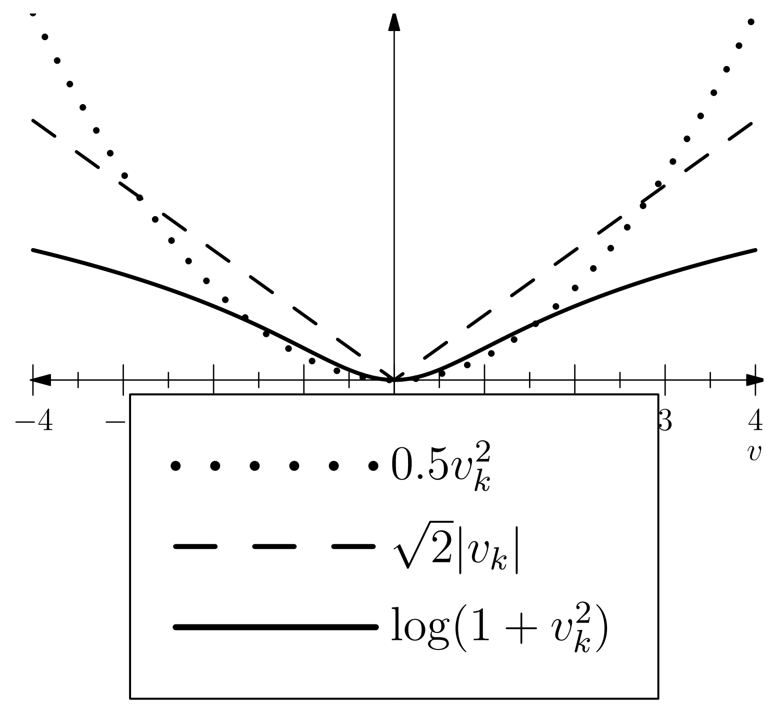
- The parametric form of the distribution then determines the optimization formulation:

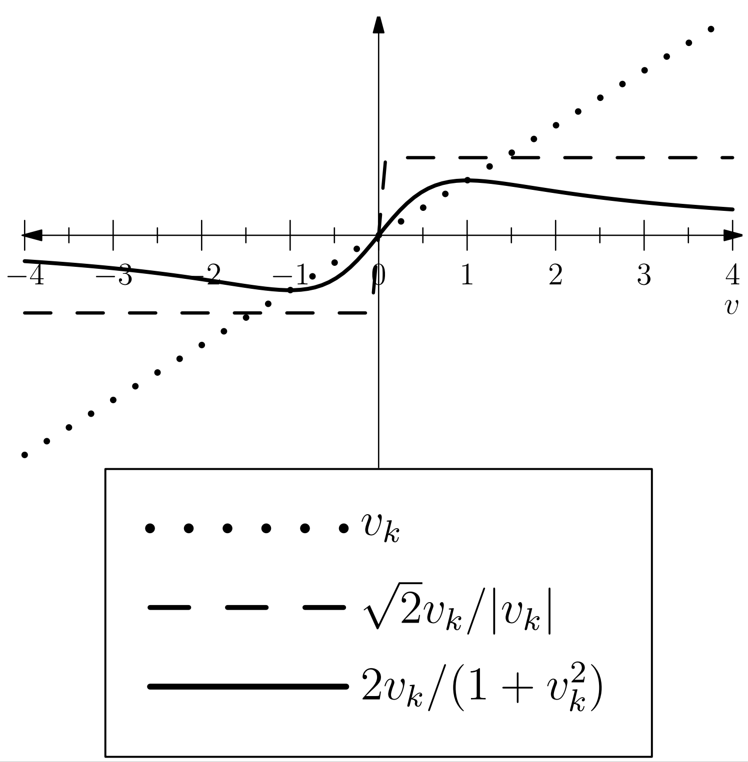
$$\min_{m} \quad -\log(\mathcal{L}(\mathbf{m})) := -\log\left[\mathbf{p}\Big(\mathbf{D} - \mathcal{F}[\mathbf{m};\mathbf{Q}]\Big)\right]$$

- Q: Which distribution do we choose, and how do we solve the problem?

Densities, Penalties, and Gradients







A Simple Comparison

- We present a comparison with two other distributions:

	Gaussian	$L(\lambda = 1)$	$T(df = 3)$
$p(\|x - \mu\| > 4\sigma)$	6.3×10^{-5}	1.8×10^{-2}	0.6×10^{-2}
$p(\|x - \mu\| > 8\sigma)$	1.3×10^{-15}	3.3×10^{-4}	8.1×10^{-4}

- The Laplace distribution corresponds to the L1 penalty on the misfit:

$$\|\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m}; \mathbf{Q}]\|_1$$

- In the class of CONVEX negative log likelihoods, it has the heaviest tail

Previous Work

	Gaussian	$L(\lambda = 1)$	$T(df = 3)$
$p(\|x-\mu\| > 4\sigma)$	6.3×10^{-5}	1.8×10^{-2}	0.6×10^{-2}
$p(\|x-\mu\| > 8\sigma)$	1.3×10^{-15}	3.3×10^{-4}	8.1×10^{-4}

- In the class of CONVEX negative log likelihoods, L1 has the heaviest tail: $e^{-\|x\|_1}$
- The likelihood corresponding to the Huber and to previous modifications have exactly the same tails.

Huber: Guitton & Symes, '03
Huber and L1: Brossier, Operto, Virieux '09, '10
Hybrid: Bube, '07.

- But the full problem is non-convex anyway, so let's consider Student's t!

New FWI Formulation

Density:

$$\mathbf{p}(\epsilon|\mu,\sigma,k) = \frac{\Gamma(\frac{k+1}{2})}{\sigma\Gamma(\frac{k}{2})\sqrt{\pi k}}\left(1+\frac{(\epsilon-\mu)^2}{k\sigma^2}\right)^{\frac{-(k+1)}{2}}$$

For FWI:

$$\mathbf{p}(\epsilon|\mu=0,\sigma=1,k) \propto \left(k+\epsilon^2\right)^{\frac{-(k+1)}{2}}$$

Robust Objective:

$$\min_{\mathbf{m}} \quad \phi_{St}(\mathbf{m}) := \frac{k+1}{2}\sum_{\omega}\sum_{s}\sum_{i}\log\left(k+(\mathbf{D}_{\omega,i}-\boldsymbol{\mathcal{F}}_{\omega}[\mathbf{m},\mathbf{Q}_i])^2\right)$$

Gradient Comparison:

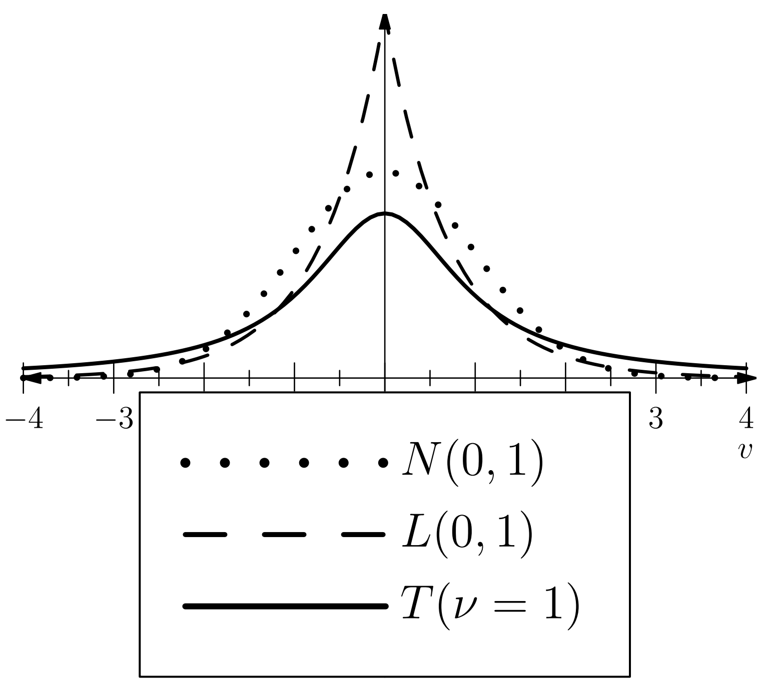
Least Squares:

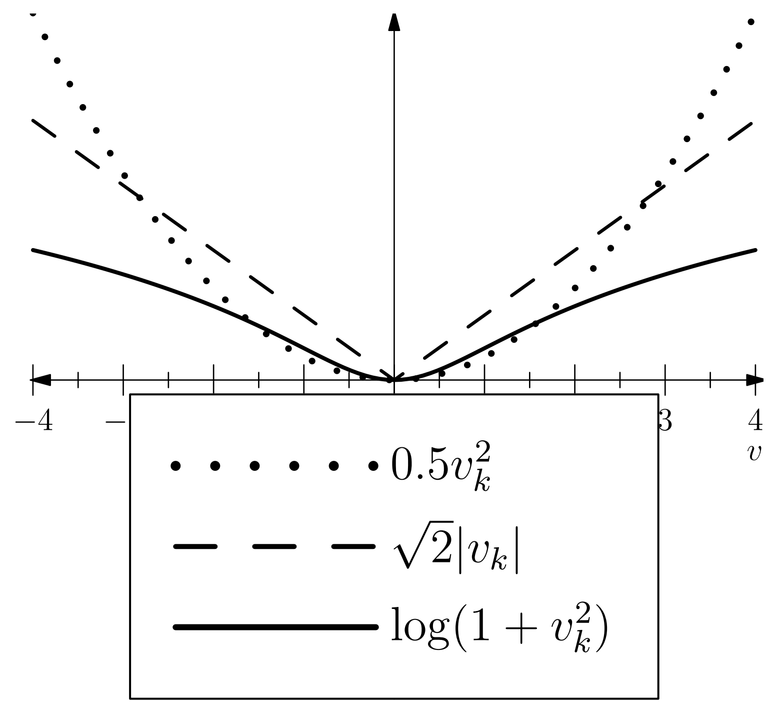
$$\nabla \phi(\mathbf{m}) = \frac{1}{2} \sum_{\omega} \sum_{s} \sum_{i} \nabla \boldsymbol{\mathcal{F}}_{\omega}[\mathbf{m}, \mathbf{Q}_i]^T \left(\mathbf{D}_{\omega,i} - \boldsymbol{\mathcal{F}}_{\omega}[\mathbf{m}, \mathbf{Q}_i]\right)$$

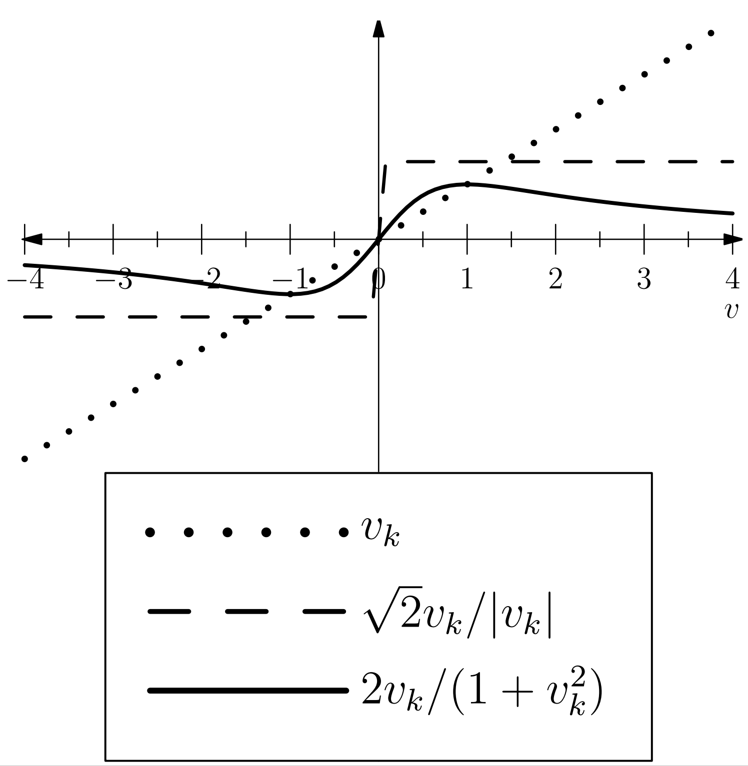
Student's t:

$$\nabla \phi_{St}(\mathbf{m}) = \frac{k+1}{2} \sum_{\omega} \sum_{s} \sum_{i} \frac{\nabla \boldsymbol{\mathcal{F}}_{\omega}[\mathbf{m}, \mathbf{Q}_i]^T \left(\mathbf{D}_{\omega,i} - \boldsymbol{\mathcal{F}}_{\omega}[\mathbf{m}, \mathbf{Q}_i]\right)}{k + \left(\mathbf{D}_{\omega,i} - \boldsymbol{\mathcal{F}}_{\omega}[\mathbf{m}, \mathbf{Q}_i]\right)^2}$$

Densities, Penalties, and Gradients



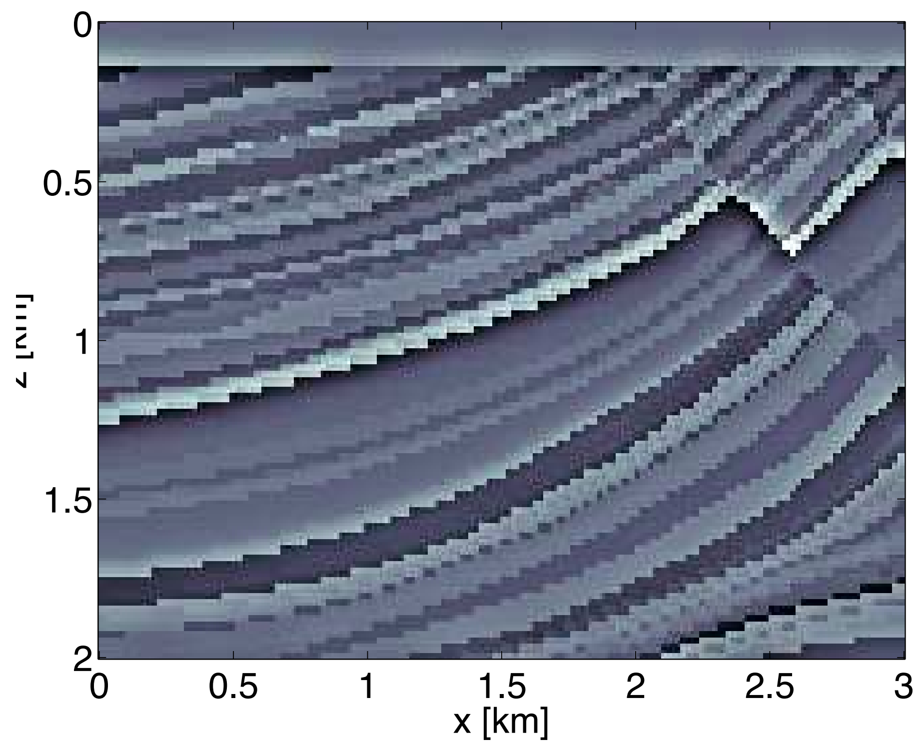




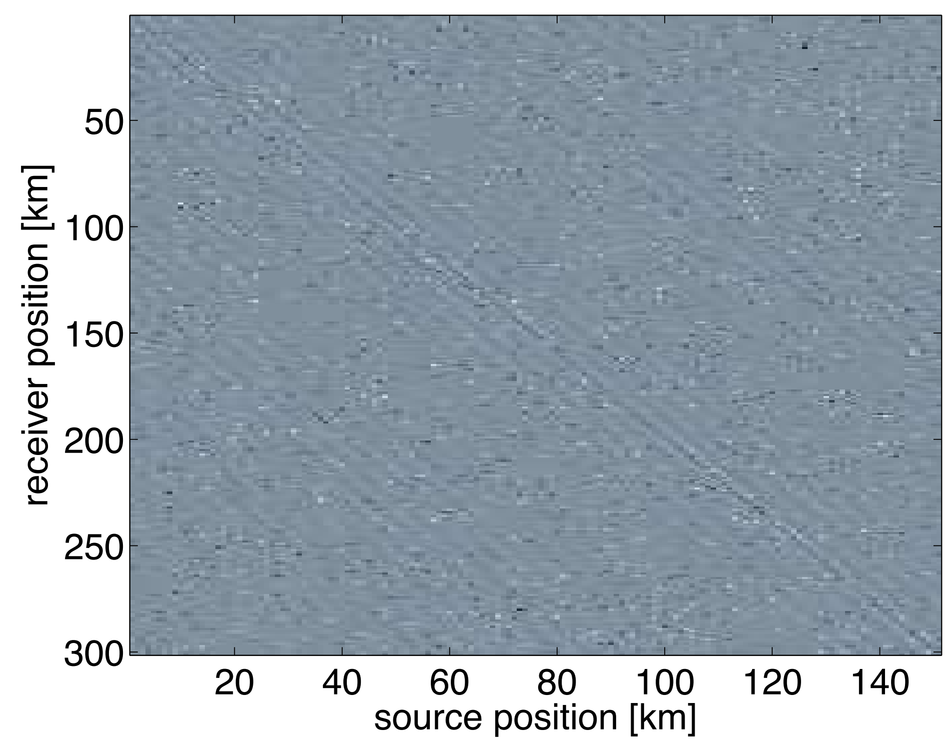
Marmoussi Example

- 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

Experiment I

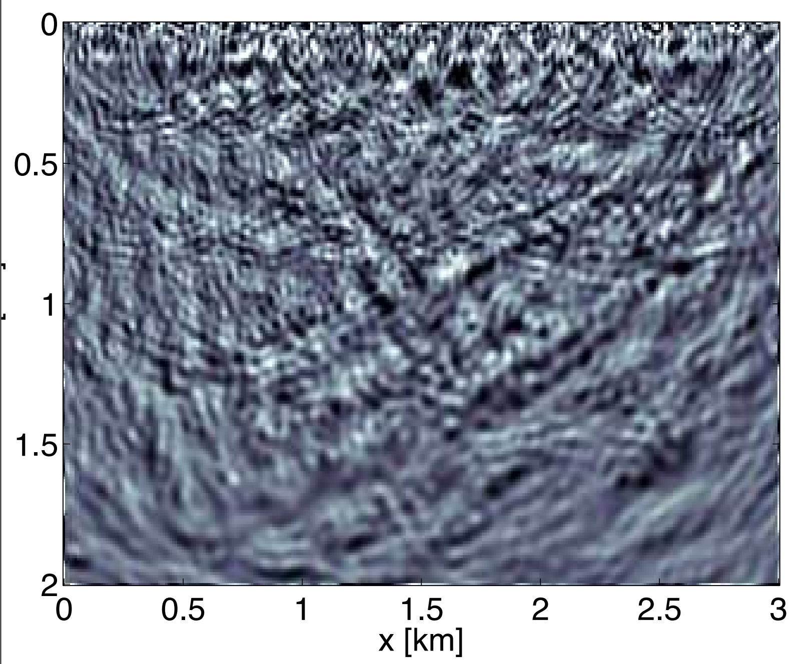


True Reflectivity

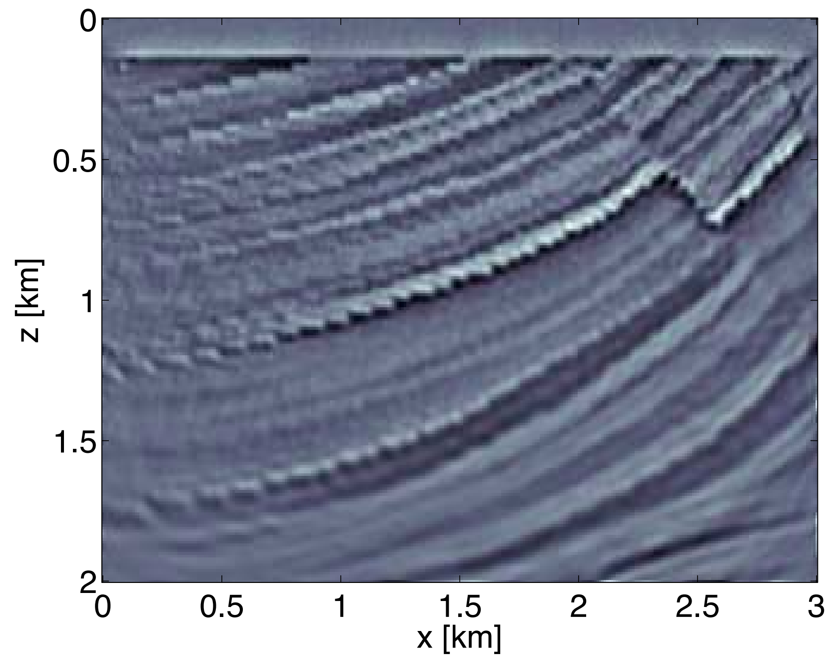


15 hz data slice with spiky noise

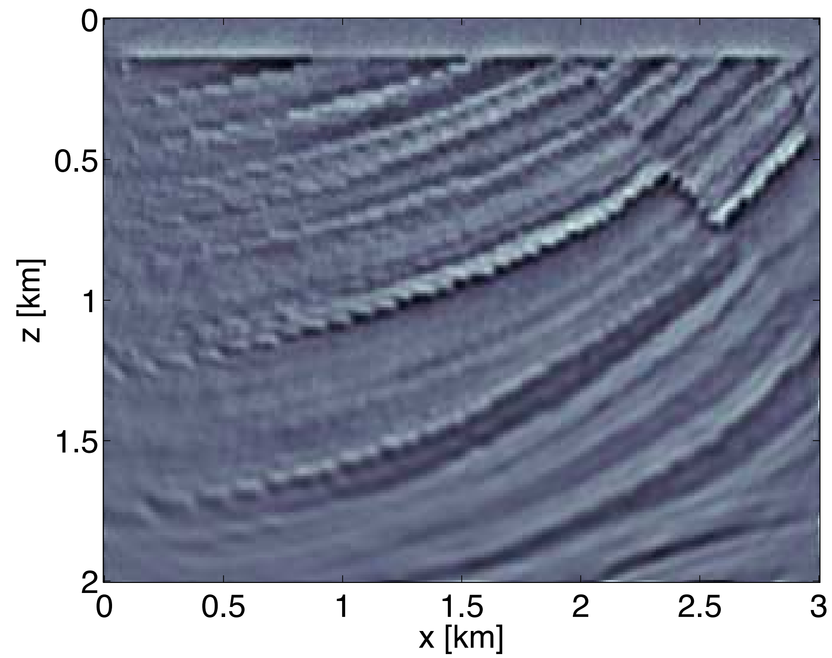
Results I



NLLS

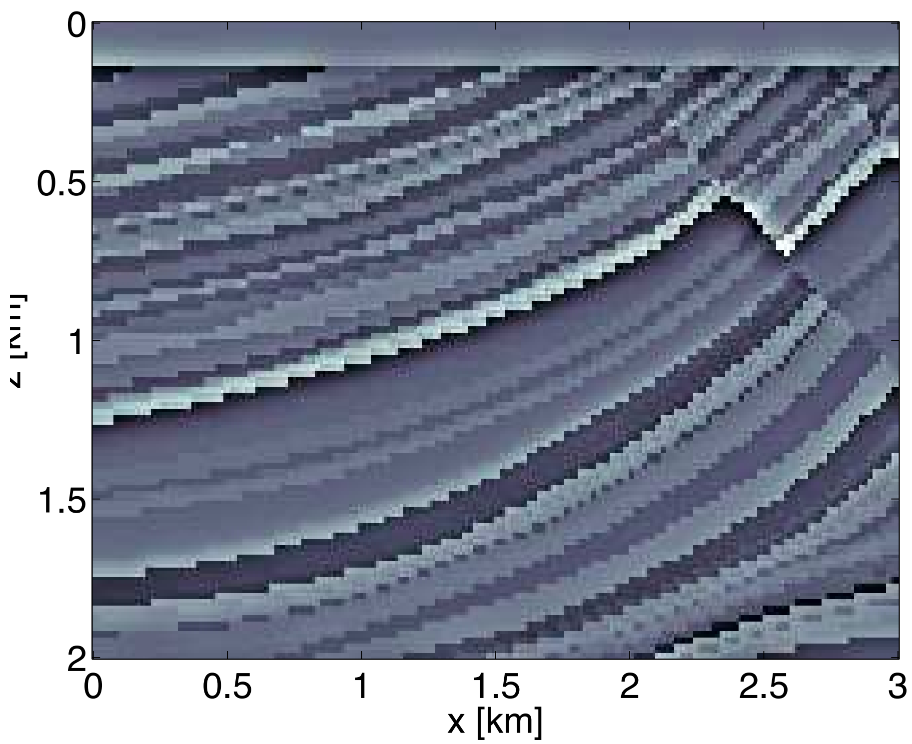


HUBER

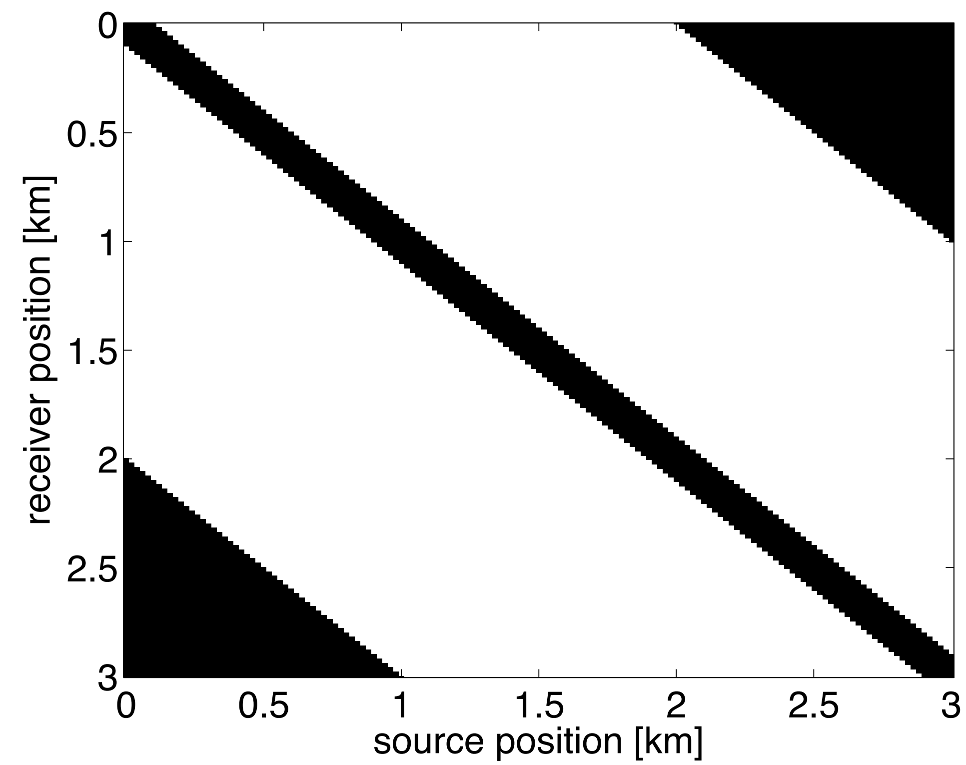


STUDENT

Experiment II

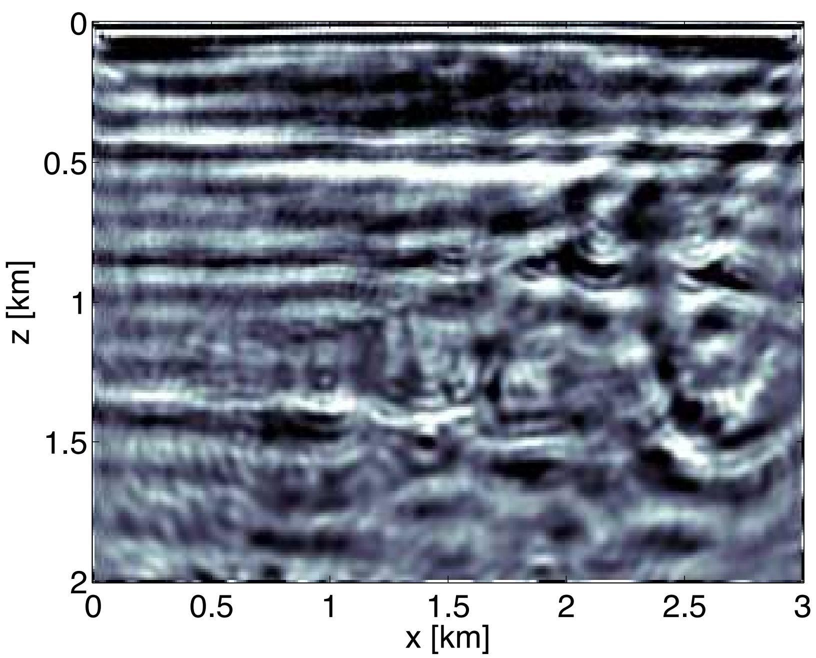


True Reflectivity

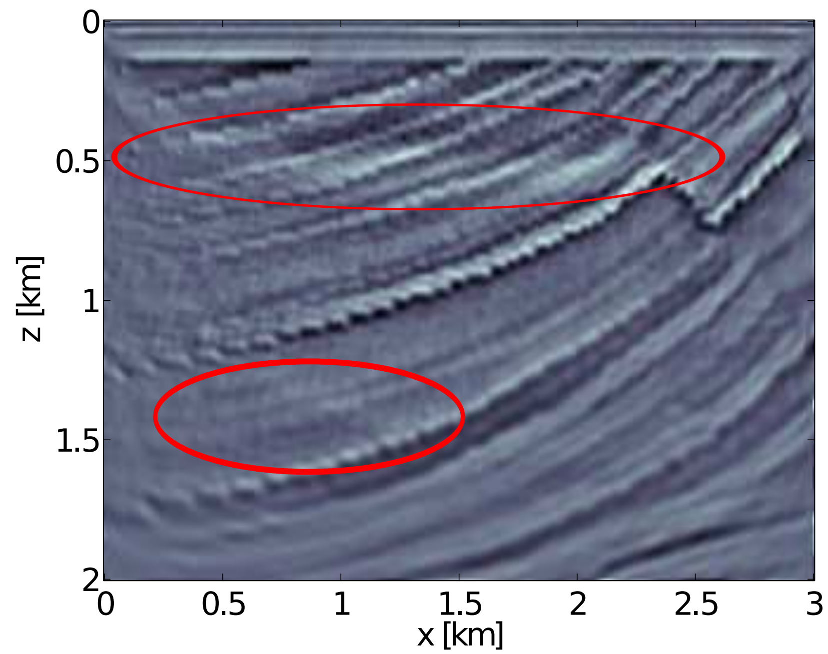


Marine Acquisition Mask

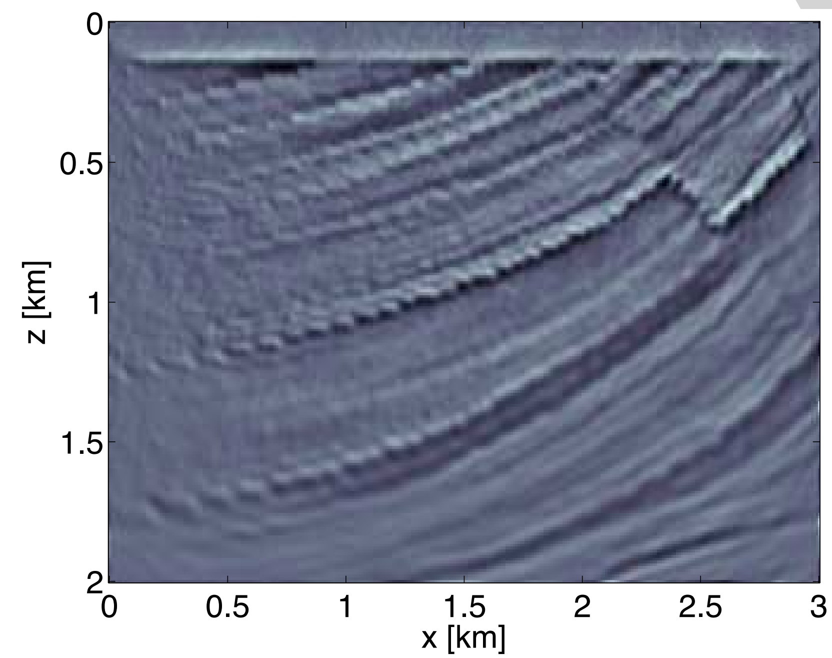
Results II



NLLS



Huber



Student

Conclusions

- Robust formulations allow good recovery even with poor modeling
- 'Mistakes' are typically thought of as 'outliers' in the data, but can also be events unexplained (or ignored) by the modeling
- Since the least-squares FWI problem is non-convex anyway, we can consider distributions with non-convex negative log likelihoods to see how well they work
- Future direction: combining robust and sparse recovery.

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.

SLIM

References

Bube, K. P. and Nemeth, T. Fast line searches for the robust solution of linear systems in the hybrid and Huber norms. *Geophysics; March-April 2007; v. 72; no. 2; p. A13-A17.*

Brossier, R., S. Operto and J. Virieux Robust frequency-domain full-waveform inversion using the L1 norm. *Geophysical Research Letters, 36 L20310 (2009).*

Brossier R., S. Operto and J. Virieux. Which data residual norm for robust elastic frequency-domain Full Waveform Inversion? *Geophysics 75 (3)*, *R37-R4 (2010).*

Guitton, A. and Symes, W. Robust inversion of seismic data using the Huber norm. *Geophysics; July-August 2003; v. 68; no. 4; p. 1310-1319*