Aleksandr Y. Aravkin

Theory
James V. Burke
Michael P. Friedlander

R-BPDN
Anais Tamalet
Tristan van Leeuwen

Source estimation
Tristan van Leeuwen
Ning Tu

Rank Opt.
Rajiv Mittal
Hassan Mansour
Ben Recht
Felix Herrmann

SLIM Consortium Meeting, December 5, 2012

Outline

- General Regularized Inverse Problems
  - Optimization formulation
  - Inverse function theorem
  - Overview of key ingredients

Outline

- General Regularized Inverse Problems
  - Optimization formulation
  - Inverse function theorem
  - Overview of key ingredients

- Robust BPDN
  - Formulation & implementation
  - Simple spike train example
  - Seismic example

Outline

- General Regularized Inverse Problems
  - Optimization formulation
  - Inverse function theorem
  - Overview of key ingredients
- Robust BPDN
  - Formulation & implementation
  - Simple spike train example
  - Seismic example
- Source estimation
  - Variable projection (review)
  - Implementation details
  - Numerical experiments

Outline

- General Regularized Inverse Problems
  - Optimization formulation
  - Inverse function theorem
  - Overview of key ingredients
- Robust BPDN
  - Formulation & implementation
  - Simple spike train example
  - Seismic example
- Source estimation
  - Variable projection (review)
  - Implementation details
  - Numerical experiments
- (Robust) rank optimization
  - Implementation details for factorization approach (Rajiv)
  - Robust rank formulation & results

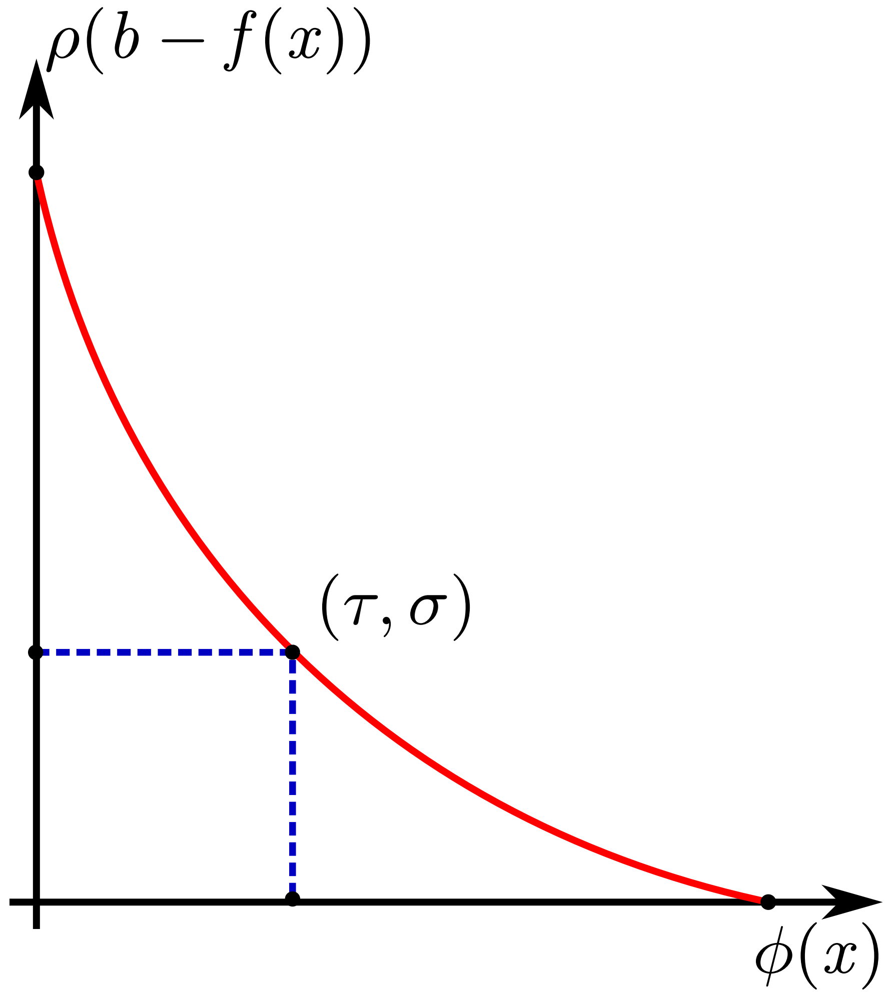
Regularized Inverse Problems

General Formulation

Consider the pair of problems

$$\mathcal{P}_\sigma : \left\{ \begin{array}{l} \min \phi(x) \\ \quad \text{s.t.} \quad \rho(b - f(x)) \le \sigma \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{r} v(\tau) := \min \rho(b - f(x)) \\ \text{s.t.} \quad \phi(x) \le \tau \end{array} \right\} : \mathcal{P}_\tau$$

- $\phi(x)$ can be
  - $\|x\|_1$, $\|x\|_2$, $\|X\|_*$, $\|x\| + \delta\left(x \mid \mathbb{R}^n_+\right)$
- $\rho(\cdot)$ can be
  - $\|\cdot\|_2^2$, Huber, Student's t penalty
- $f(x)$ can be
  - linear or nonlinear forward model



General Formulation

Main results:

$$\mathcal{P}_\sigma : \left\{ \begin{array}{ll} \min \phi(x) & \\ \quad \text{s.t.} \quad \rho(b - f(x)) \le \sigma \end{array} \right\} \quad \longleftrightarrow \quad \left\{ \begin{array}{r} v(\tau) := \min \rho(b - f(x)) \\ \text{s.t.} \quad \phi(x) \le \tau \end{array} \right\} : \mathcal{P}_\tau$$

- We can solve $\mathcal{P}_\sigma$ while working with $\mathcal{P}_\tau$ alone, **if we solve** $v(\tau) = \sigma$.
- We can characterize $v'(\tau)$ for any linear f, convex ϕ and convex ρ, allowing Newton's method to be used in most practically useful cases.
- We obtain explicit formulae for broad function classes, and these work even when the linearity of f or convexity of ρ are violated.

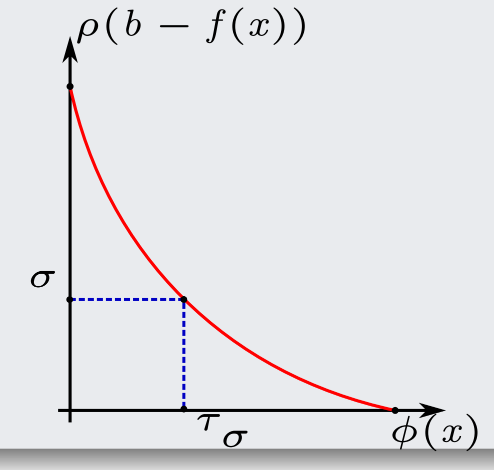
Inverse Function Theorem

$$\mathcal{P}_\sigma : \left\{ \begin{array}{rl} v_1(\sigma) = \min \phi(x) & \\ \text{s.t.} & \rho(b - f(x)) \le \sigma \end{array} \right\} \leftrightarrow \left\{ \begin{array}{rl} v_2(\tau) = \min \rho(b - f(x)) & \\ \text{s.t.} & \phi(x) \le \tau \end{array} \right\} : \mathcal{P}_\tau$$

Theorem (A., Burke, Friedlander)

Define $S_\sigma = \{\sigma : \emptyset \neq \arg\min \mathcal{P}_\sigma \subset \{x : \rho(b - f(x)) = \sigma\}\}$. Then for each $\sigma \in S_\sigma$, we have

- $v_2(\underbrace{v_1(\sigma)}_{\tau_\sigma}) = \sigma$
- $\arg\min \mathcal{P}_\sigma = \arg\min \mathcal{P}_{\tau_\sigma}$



- We can indeed solve one problem by working with the other
- What are the ingredients to make this a practical approach?

- Penalty ρ should be *differentiable*.
- We need a fast projection onto $\{x : \phi(x) \leq \tau\}$.
- We need to compute $v'(\tau)$, in order to solve $v(\tau) = \sigma$.

Theorem (A., Burke, Friedlander)

For convex ρ and linear forward model $b = Ax$, if $\bar{x}$ solves $\mathcal{P}_\tau$, then

$$v'(\tau) = -\arg\min_{\mu} \left\{ \tau\mu + \mu\phi^* \left(\frac{1}{\mu}(A^T \nabla\rho(b - A\bar{x}) \right) \right\}$$

- If $\phi(x) = \|x\|$, $v'(x) = -\|A^T \nabla\rho(b - A\bar{x})\|_*$.
- Even when ρ is nonconvex, formula is useful.

Robust BPDN

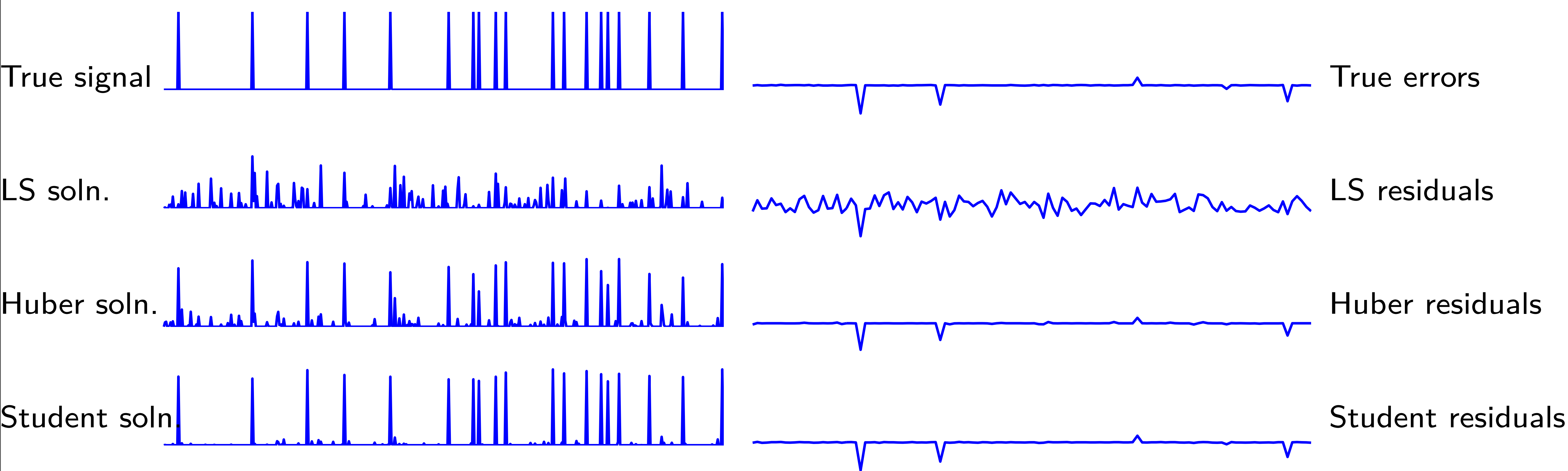
Formulation

Consider the sparse recovery problem:

$$\min_x \|x\|_1 \qquad \text{s.t.} \qquad \rho(b - Ax) \leq \sigma.$$

- The parameter σ encodes the allotted error level below which we do not want to fit the data.
- However, in the presence of outliers, different penalties ρ will behave differently, even when the *true error* level is known.
- Two penalties we can immediately try are the *huber* and *Student's t* penalty. Even though the latter isn't convex, it turns out we can find the root just fine when we use the formula for $v'(\tau)$.

Non-negative spike train recovery

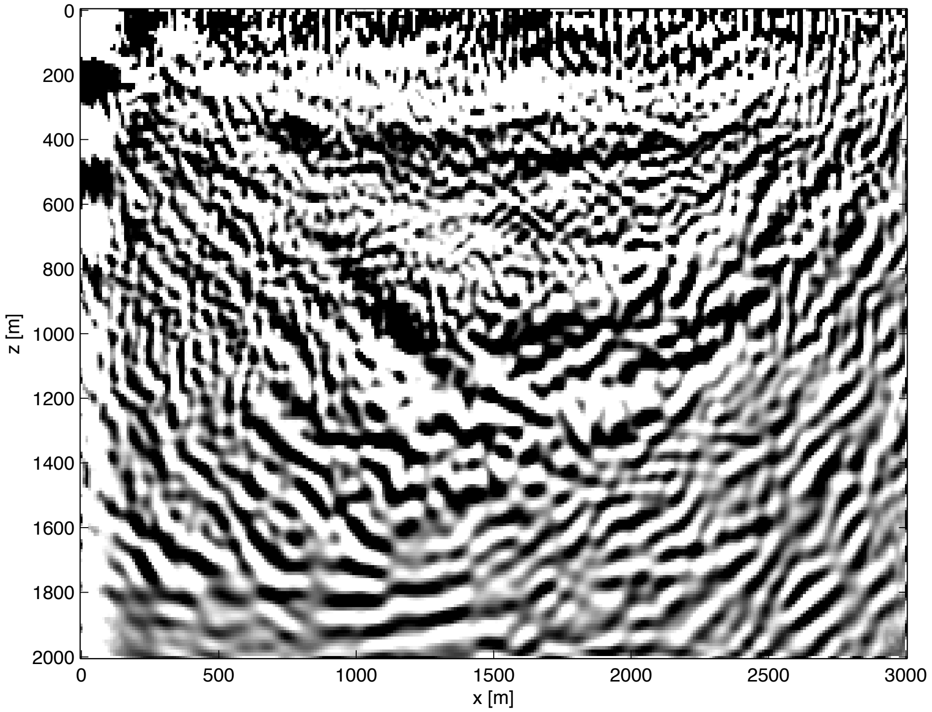


- In each case, $\sigma := \rho(b - A\bar{x}) = \rho(\bar{\epsilon})$.
- Huber does better than LS, and Student's t does better than Huber in this example.
- New SPGℓ_1 works just fine when you pass it the Student's t penalty.

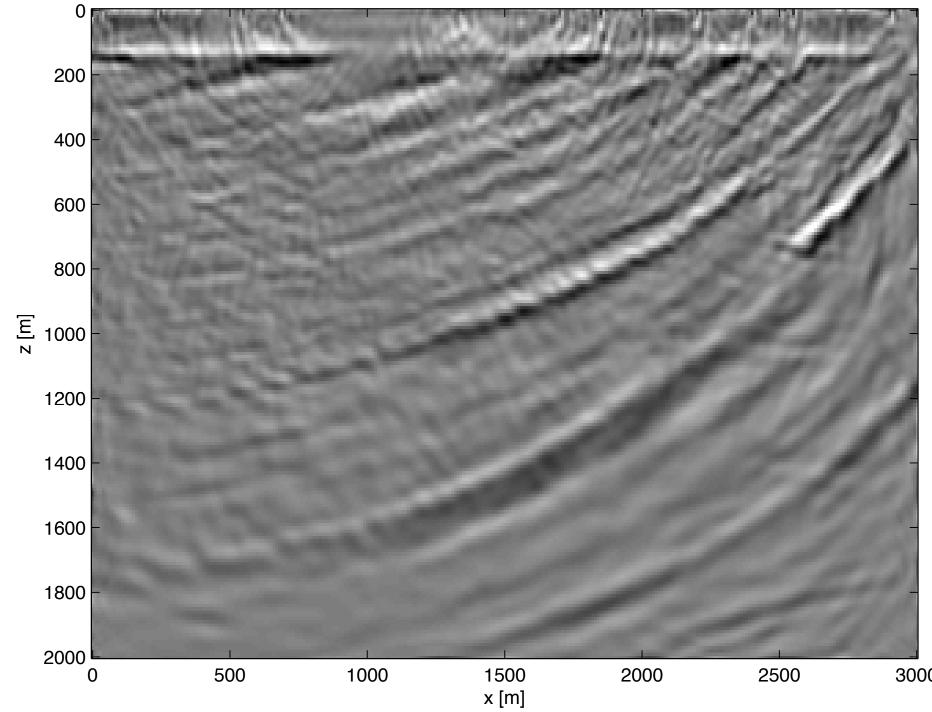
A.Y. Aravkin, J.V. Burke, M.P. Friedlander, *Variational Properties of Value Functions*, submitted to Siam J. Opt. 11/2012.

Sparse Robust Seismic Imaging (10% outlier error)

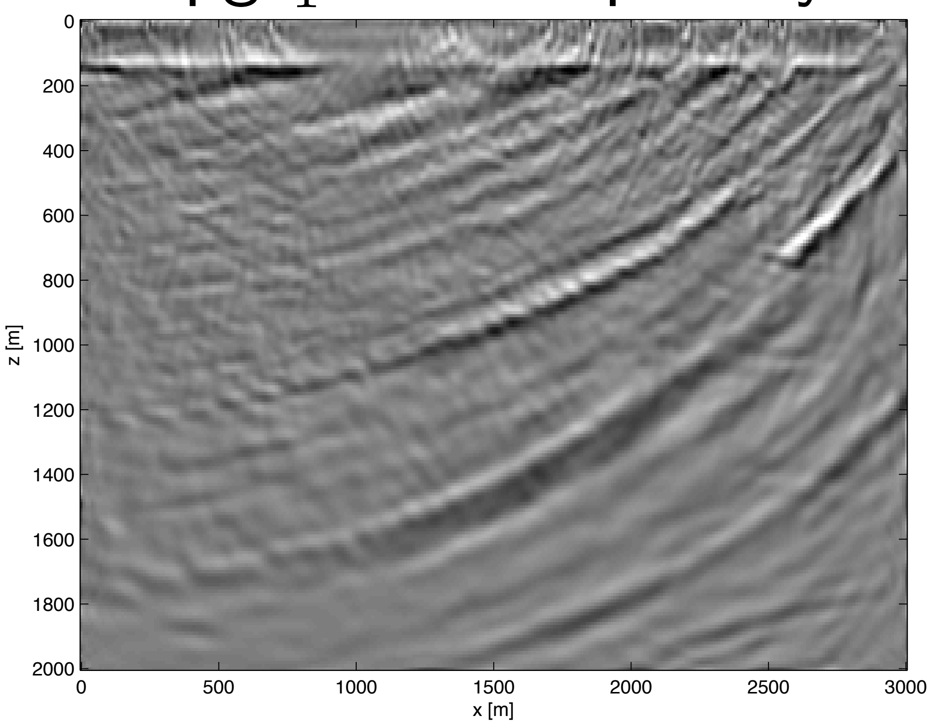
spgℓ_1 with LS penalty



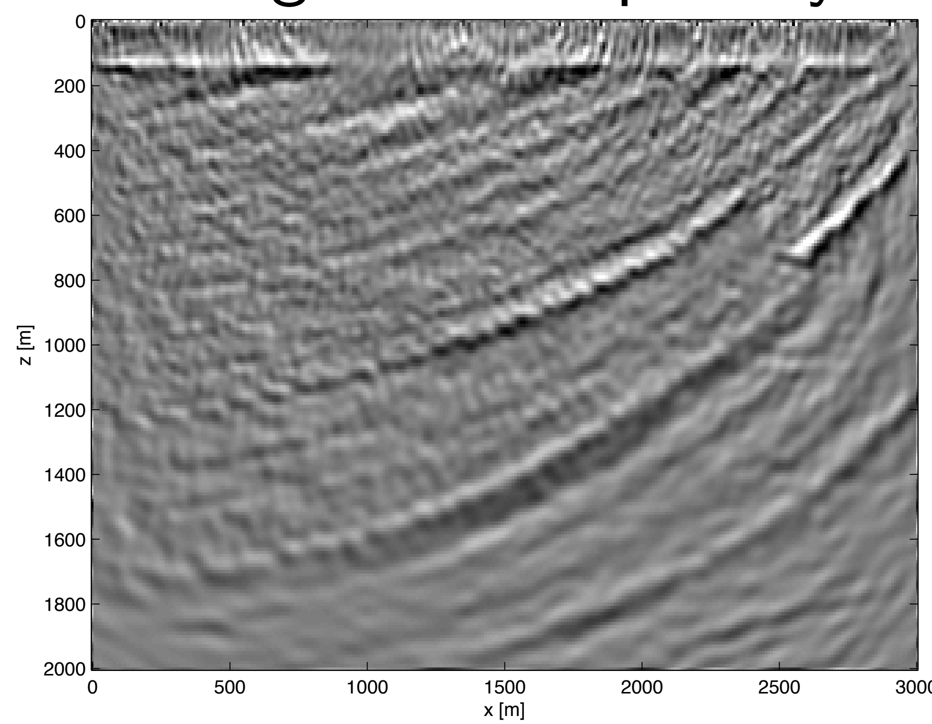
spgℓ_1 with St penalty



spgℓ_1 with St penalty



l-bfgs with St penalty



SPARSE IMAGING WITH SOURCE ESTIMATION

Formulation

Consider the seismic imaging problem with unknown source weights:

$$\min_{x,\alpha} \|x\|_1 \qquad \text{s.t.} \qquad \sum_i \|d_i - \alpha_i F_i C^T x\|_2^2 \leq \sigma^2$$

where

- x is the vector of Curvelet coefficients to recover
- α_i are unknown source weights
- C is the Curvelet transform
- d is frequency-domain data

We solve this problem by using a *variable projection* based forward model in extended SPGℓ_1.

Formulation

$$\left\{\begin{array}{ll} \min_{x,\alpha} \|x\|_1 & \\ \quad \text{s.t.} & \sum_i \|d_i - \alpha_i F_i C^T x\|_2^2 \le \sigma^2 \end{array}\right\} \longleftrightarrow \left\{\begin{array}{ll} \min_{x,\alpha} \sum_i \|d_i - \alpha_i F_i C^T x\|_2^2 & \\ \quad \text{s.t.} \quad \|x\|_1 \le \tau & \end{array}\right\}$$

- The problem on the right can be solved by *variable projection*, where we compute

$$\bar{\alpha}_i(x) = \arg\min_{\alpha_i} \|d_i - \alpha_i F_i C^T x\|_2^2.$$

- The gradient is then easily computed using values of $\bar{\alpha}$, and the SPG method can be used to solve the objective on the right.

$$\left\{\begin{array}{ll} \min_{x} \|x\|_1 & \\ \quad \text{s.t.} & \sum_i \|d_i - \alpha_i(x) F_i C^T x\|_2^2 \le \sigma^2 \end{array}\right\} \longleftrightarrow \left\{\begin{array}{ll} \min_{x} \sum_i \|d_i - \alpha_i(x) F_i C^T x\|_2^2 & \\ \quad \text{s.t.} \quad \|x\|_1 \le \tau & \end{array}\right\}$$

Results

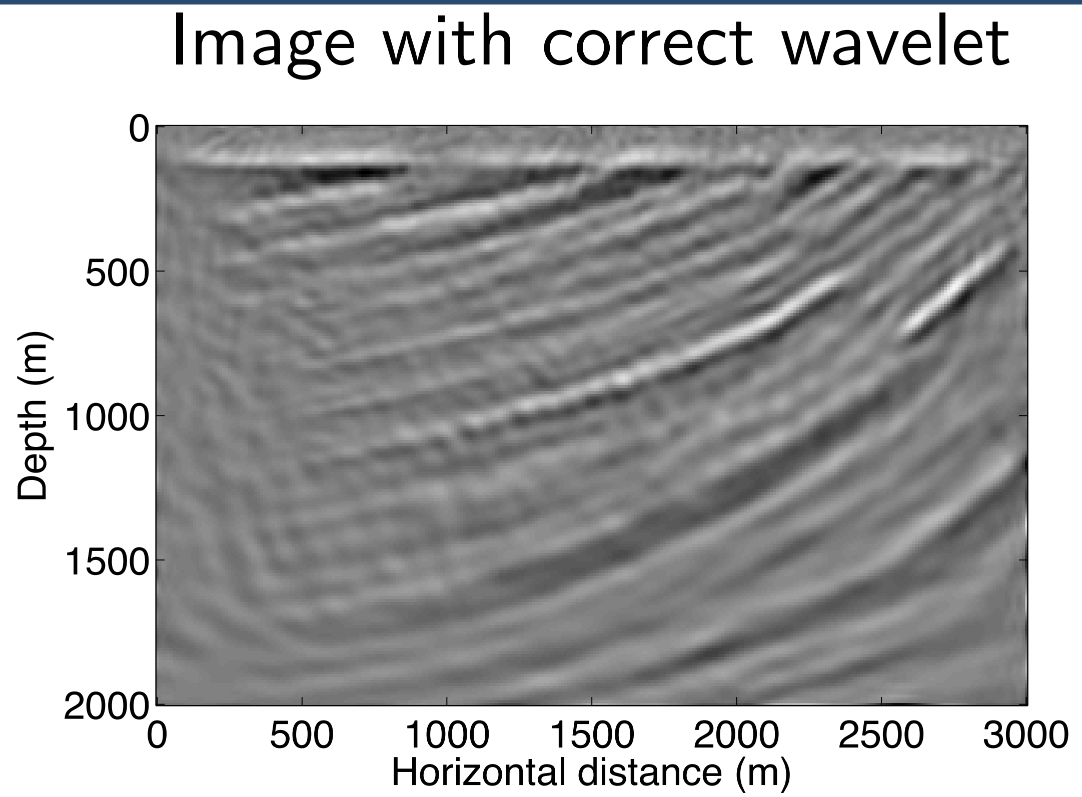


Image with correct wavelet
Depth (m)
0
500
1000
1500
2000
0
500
1000
1500
2000
2500
3000
Horizontal distance (m)

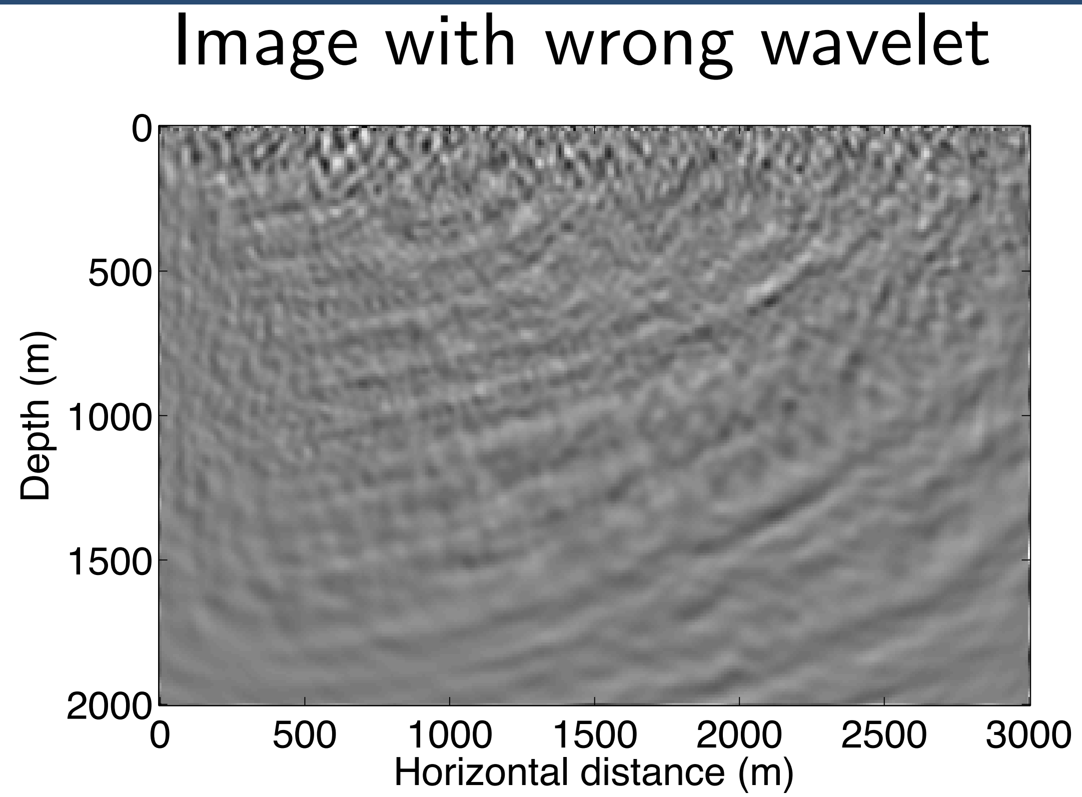


Image with wrong wavelet
Depth (m)
0
500
1000
1500
2000
0
500
1000
1500
2000
2500
3000
Horizontal distance (m)

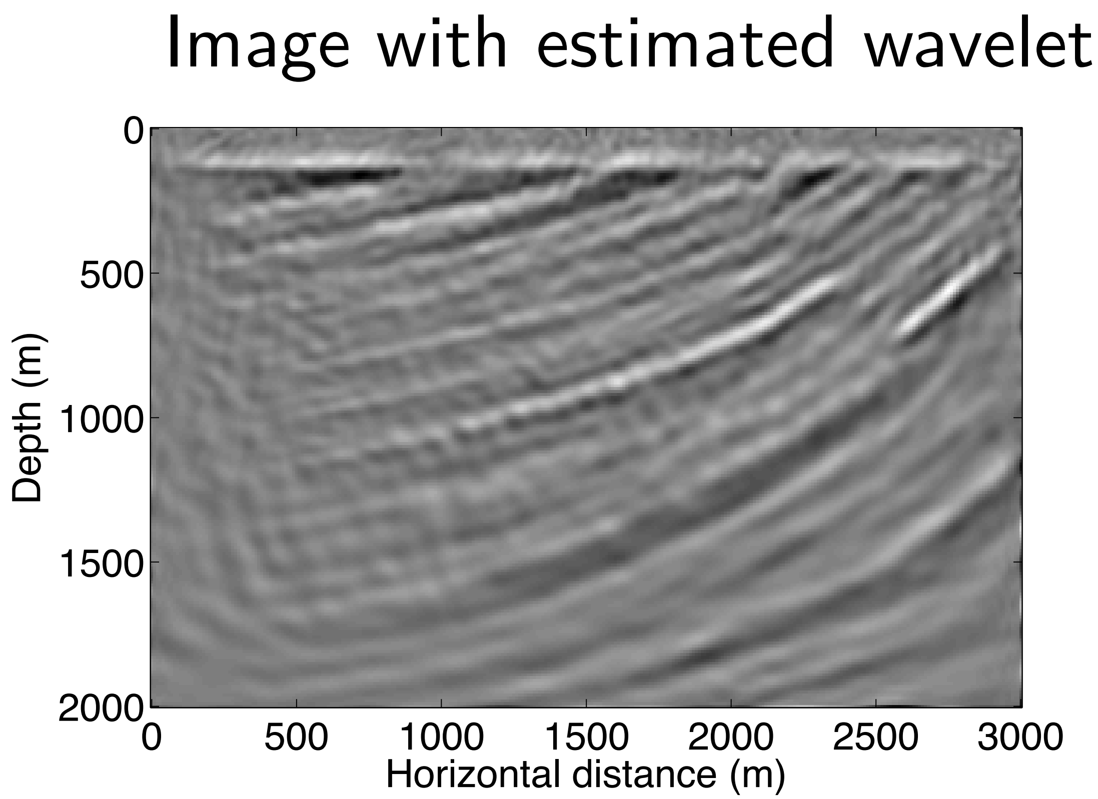
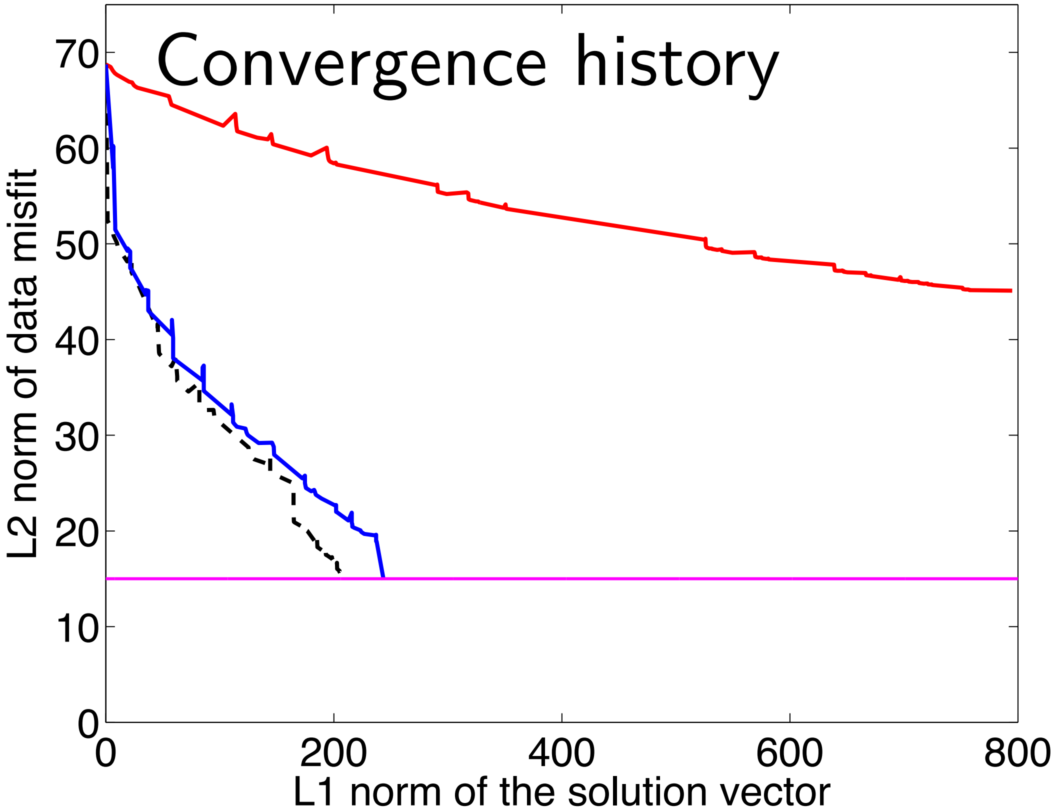
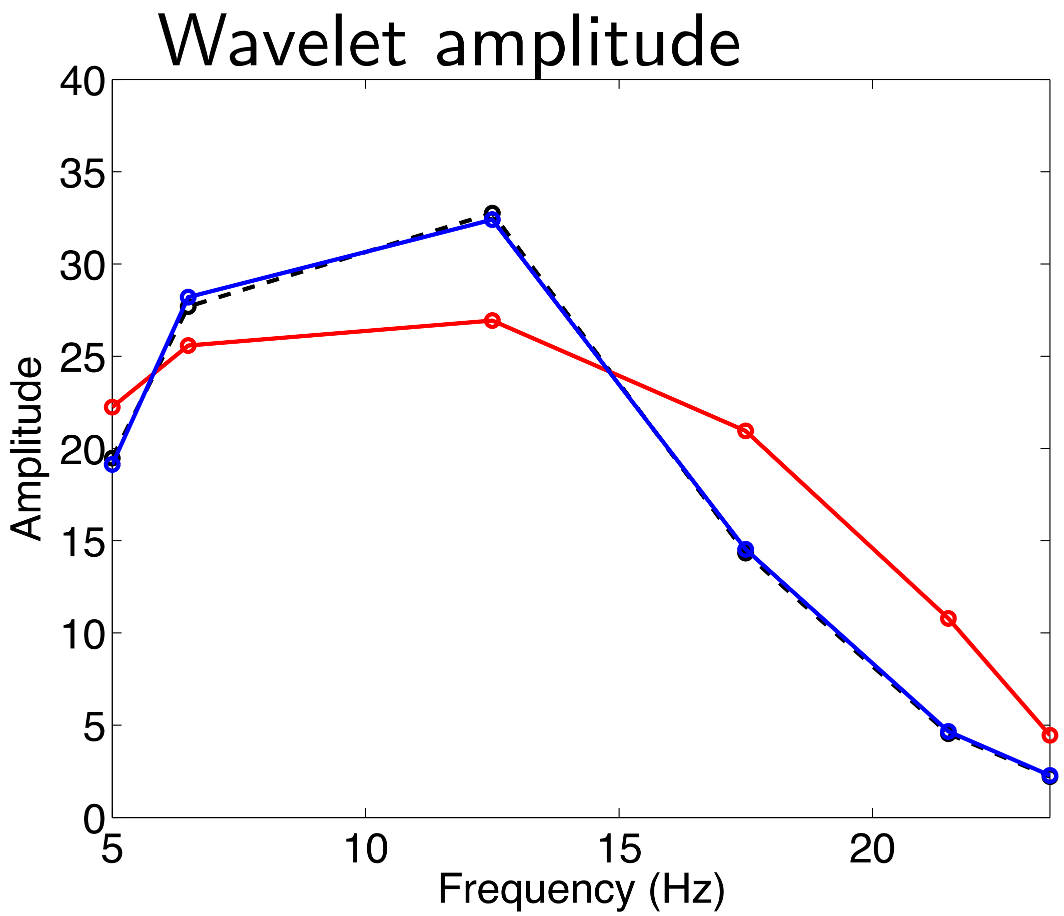
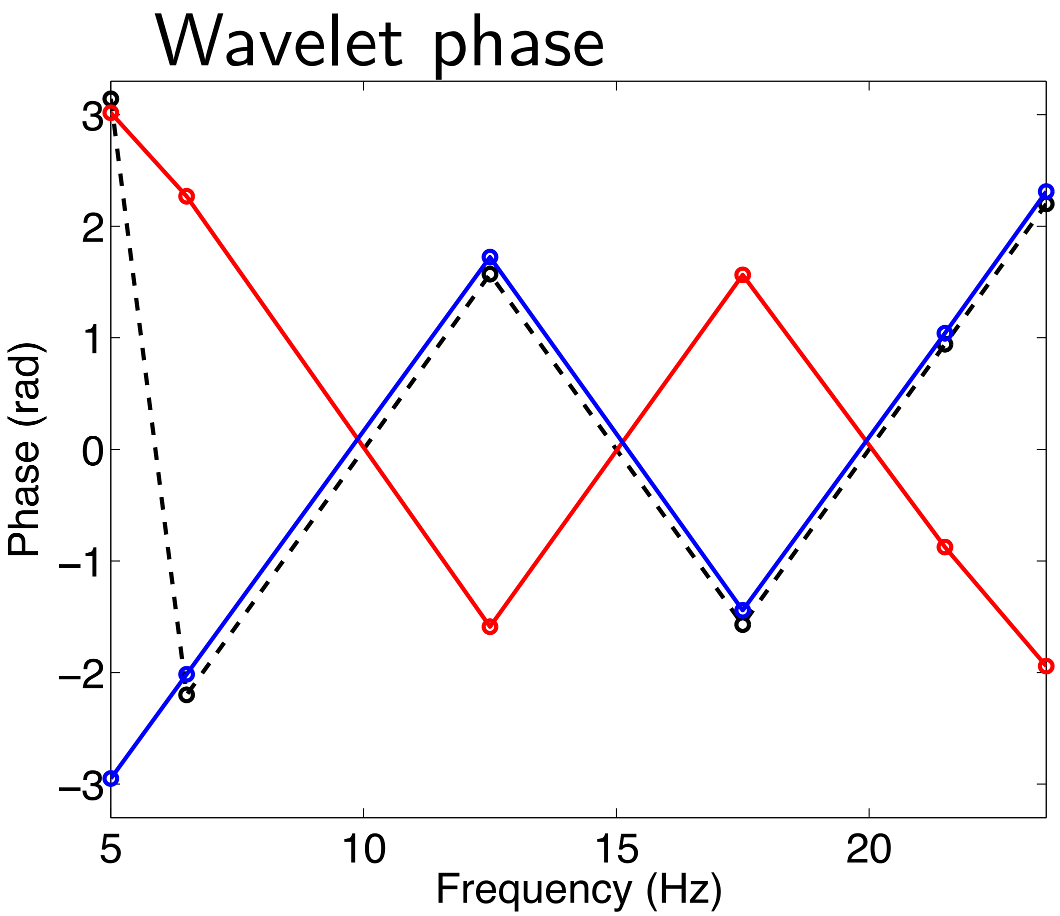


Image with estimated wavelet
Depth (m)
0
500
1000
1500
2000
0
500
1000
1500
2000
2500
3000
Horizontal distance (m)


Convergence history
L2 norm of data misfit
0
10
20
30
40
50
60
70
0
200
400
600
800
L1 norm of the solution vector

Results: Wavelet





- True wavelet amplitude and phase shown with dotted lines
- Wrong guess shown in red
- Estimation results from variable projection shown in blue.

A.Y. Aravkin, T. van Leeuwen, N. Tu, *Sparse Seismic Imaging Using Variable Projection*, submitted to ICASSP 11/2012.

ROBUST LOW RANK ESTIMATION

Matrix Completion

Consider the matrix completion problem

$$\min_X \|X\|_* \qquad \text{s.t.} \qquad \rho\left(\mathcal{A}(X) - b\right) \leq \sigma.$$

- We can assume that ρ is differentiable.
- Dual norm to $\|X\|_*$ is the spectral norm (maximal eigenvalue), relatively easy to compute.
- The main problem is *projection* onto $\{X : \|X\|_* \leq \tau\}$, since this requires SVD.

Matrix Factorization Idea (Recht et al.)

Let $X = LR^T$. Then

- We have the useful inequality

$$\|X\|_* = \|LR^T\|_* \leq \frac{1}{2}\|L\|_F^2 + \frac{1}{2}\|R\|_F^2$$

- Projection on the factors is easy, and

$$\frac{1}{2}\|L\|_F^2 + \frac{1}{2}\|R\|_F^2 \leq \tau \implies \|LR^T\|_* \leq \tau.$$

We can formulate LASSO-type matrix completion formulations

$$\min_{L,R} \rho(b - \mathcal{A}(LR^T))$$

$$\text{s.t.} \quad \frac{1}{2}\|L\|_F^2 + \frac{1}{2}\|R\|_F^2 \leq \tau$$

as well as penalized formulations

$$\min_{L,R} \rho(b - \mathcal{A}(LR^T)) + \lambda\left(\frac{1}{2}\|L\|_F^2 + \frac{1}{2}\|R\|_F^2\right)$$

Matrix Factorization in SPGℓ_1

We can also incorporate the idea into the extended SPGℓ_1 framework:

$$\left\{ \begin{array}{l} \min\limits_X \|X\|_* \\ \text{s.t.} \quad \rho(b - \mathcal{A}(X)) \le \sigma \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{l} \min\limits_{L,R} \rho(b - \mathcal{A}(LR^T)) \\ \text{s.t.} \quad \frac{1}{2}\|L\|_F^2 + \frac{1}{2}\|R\|_F^2 \le \tau \end{array} \right\}$$

We solve problem on the right with projected gradient — SVDs not required. Note that the forward model is nonlinear in L, R.

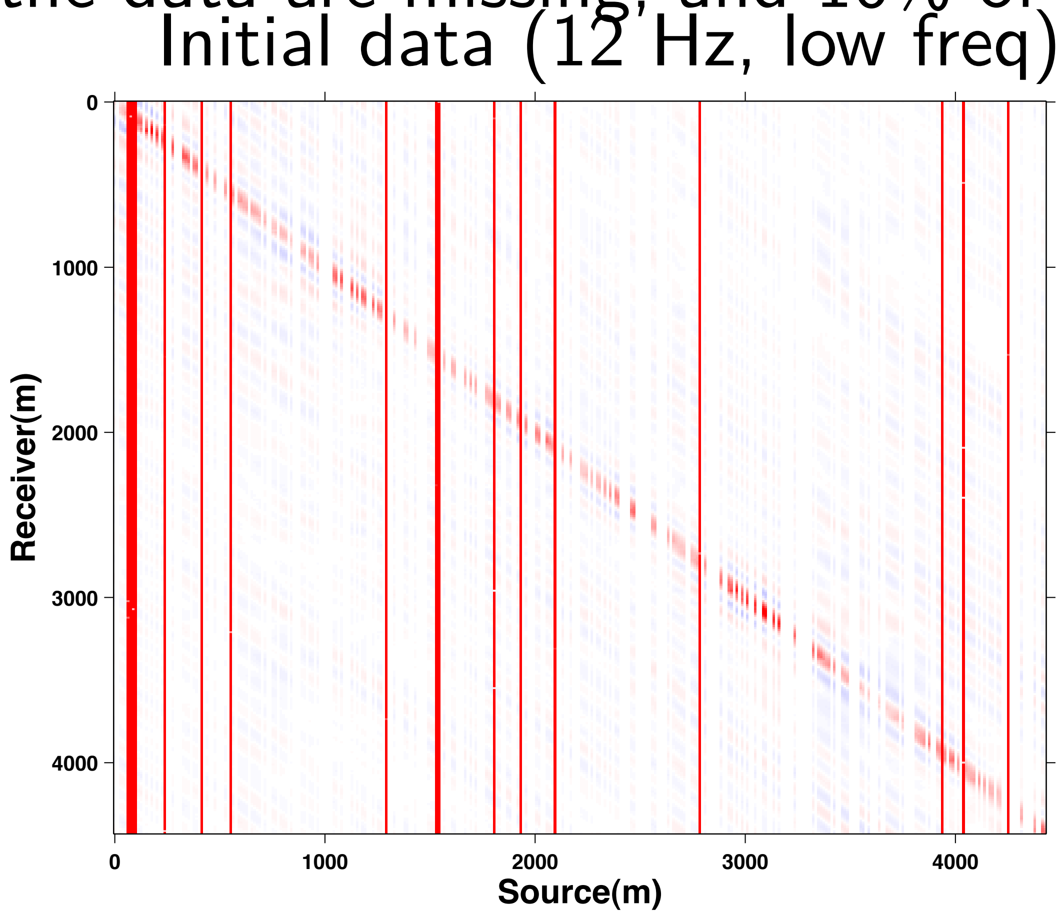
For Newton root finding, we form $X = LR^T$ and then forget the factors. The derivative of the value function is given by

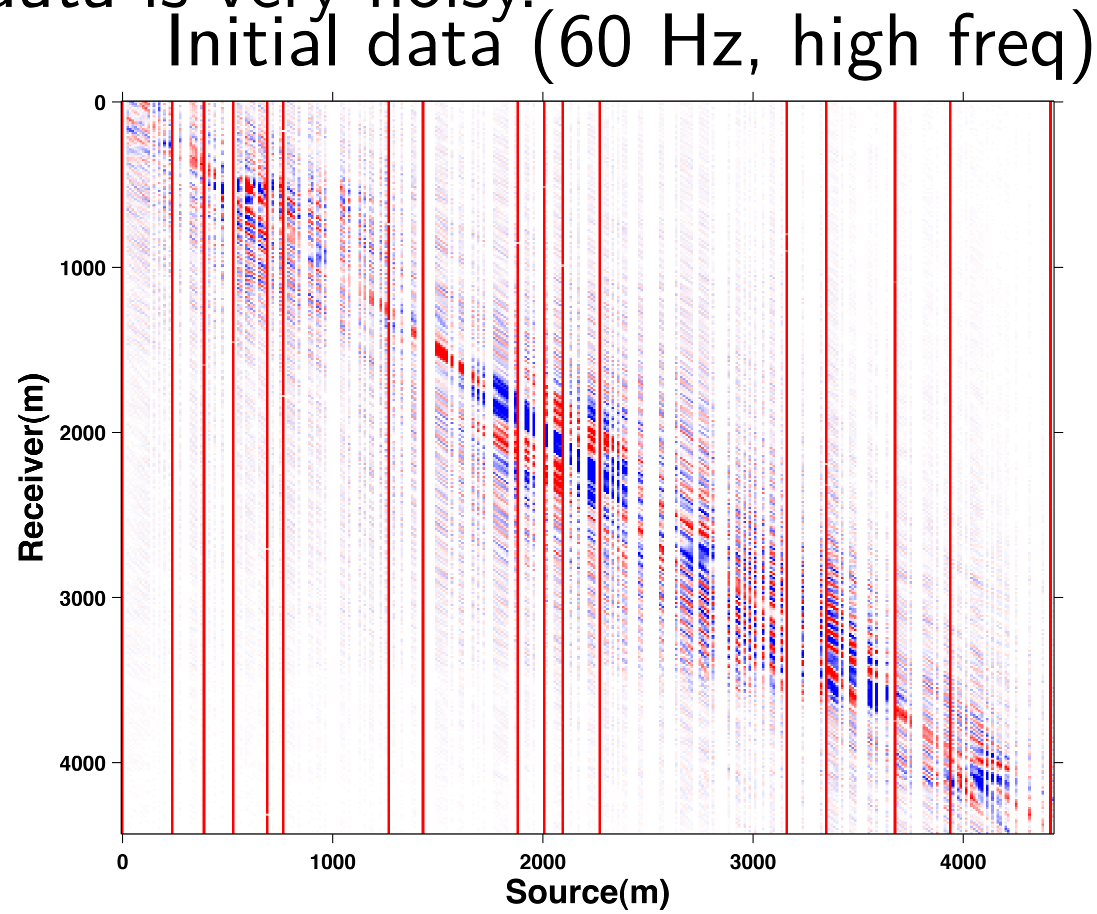
$$v'(\tau) = -\|\mathcal{A}^* \nabla \rho(b - \mathcal{A}(\bar{L}\bar{R}^T)\|_2,$$

and requires finding the largest singular value of a matrix.

Robust Matrix Completion

We consider a joint **recovery** and **denoising** experiment, where 50% of the data are missing, and 10% of data is very noisy.



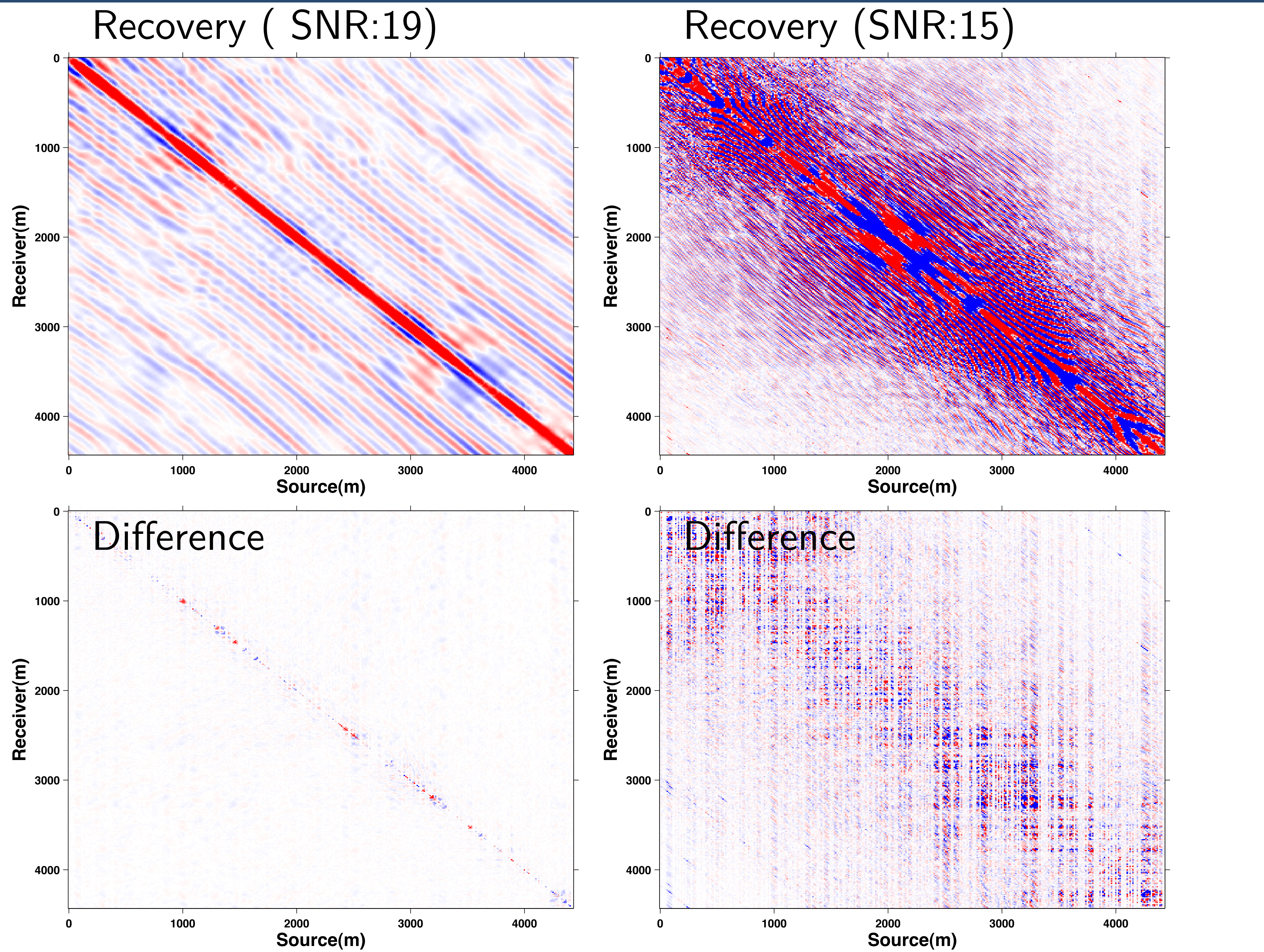


We solve the following problem:

$$\min_{L,R} \|LR^T\|_* \quad \text{s.t.} \quad \rho(b - \mathcal{A}(LR^T)) \leq \sigma$$

where ρ is the Student's t penalty. We use rank 5 for low frequency, and rank 30 for high frequency.

Robust Matrix Completion: Results



Summary and conclusions

In this talk we have presented the following applications:

- Robust BPDN (sparse imaging with outliers)
  - Regularization: $\|x\|_1$
  - Penalty: robust ρ
  - Forward model: linear
- Sparse imaging with source estimation
  - Regularization: $\|x\|_1$
  - Penalty: LS (but can be robust!!)
  - Forward model: nonlinear; uses *variable projection*
- Robust matrix completion
  - Regularization: $\|X\|_*$
  - Penalty: student's t (pick your own!)
  - Forward model: nonlinear because of matrix factors

Future work:

- Impact in EPSI formulation (Tim Lin & Ning Tu)
- Sparse dictionary learning (with Hassan and Tristan)
- Sparse FWI (?)

Thank you!

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, BP, Chevron, ConocoPhillips, Petrobras, PGS, Total SA, and WesternGeco.