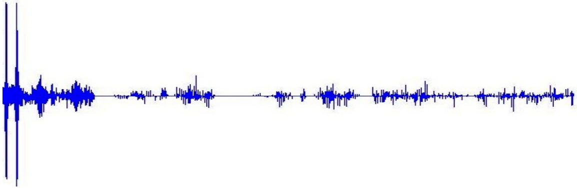
Putting the Curvature Back into Sparse Solvers

Julie Nutini
University of British Columbia



SINBAD 2013 Fall Consortium Meeting, Whistler BC
December 3rd, 2013

MOTIVATION

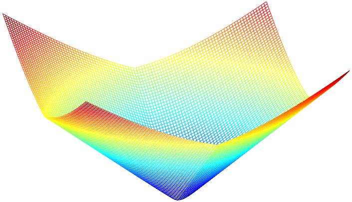
The *basis pursuit denoising problem*:

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2}\|Ax-b\|_2^2 + \lambda\|x\|_1$$

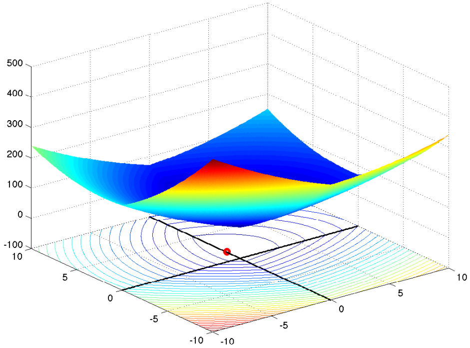
- e.g., SPGL1 (first-order method)
- ill-conditioned, may be slow to converge

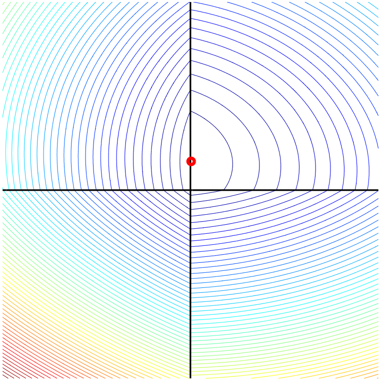
GOAL:

- Exploit curvature
- Create second-order sparse solver
- Improve convergence rates of existing solvers



Visualize





APPROACH

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1$$

A two-phase algorithm that combines 2 well-known methods:

- **Phase 1:** Proximal Gradient Method (**soft-thresholding**)
- **Phase 2:** Conjugate Gradient Method (e.g., **LSQR**)

NOTE: No additional storage overhead required like quasi-Newton

- e.g., [Schmidt, van den Berg, Friedlander, and Murphy, 2009], [Miao and Herrmann, 2013]

THE PROXIMAL GRADIENT METHOD

The *proximal operator* for $\|x\|_1$ is **soft-thresholding**:

$$\mathbf{prox}_{\lambda\|\cdot\|_1}(x) = \mathrm{sgn}(x).*\left[|x| - \lambda\right]_+$$

A *proximal gradient iteration* is applying soft-thresholding to a **steepest descent** step on the least-squares term:

$$\begin{aligned} \bar{x} &= x_k - \alpha\, A^\top(Ax_k - b), \quad (\alpha > 0) \\ x_{k+1} &= \mathrm{sgn}(\bar{x}).*\left[|\bar{x}| - \lambda\right]_+ \end{aligned}$$

THE CONJUGATE GRADIENT METHOD

Solves (iteratively) linear systems of the form

$$A^\top Ax = A^\top b, \qquad \text{i.e.,} \quad \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2}\|Ax-b\|_2^2.$$

We want to solve a linear system of the form

$$A^\top Ax = A^\top b + \lambda\bar{e}, \qquad \text{i.e.,} \quad \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2}\|Ax-b\|_2^2 + \lambda\bar{e}^\top x,$$

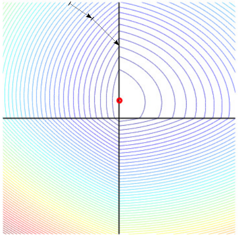
where $\bar{e} = \text{sgn}(x) = \begin{cases} +1, & \text{if } x > 0 \\ -1, & \text{if } x < 0 \end{cases}.$

PROBLEM: The function $\text{sgn}(x)$ is undefined when $x = 0$.

ALGORITHM

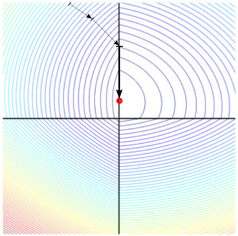
Phase 1: Use soft-thresholding to find an *active set*:

$$\mathcal{A}(x) = \{i : x_i = 0\}.$$



Phase 2: Apply conjugate gradient to the *support*:

$$\mathcal{S}(x) = \{i : x_i \neq 0\}.$$



- Check if updated active set is optimal.
  - **Yes**: Continue exploring in Phase 2.
  - **No**: Define new active set in Phase 1.

SEISMIC DATA INTERPOLATION

Find sparse representation of y in curvelet operator C:

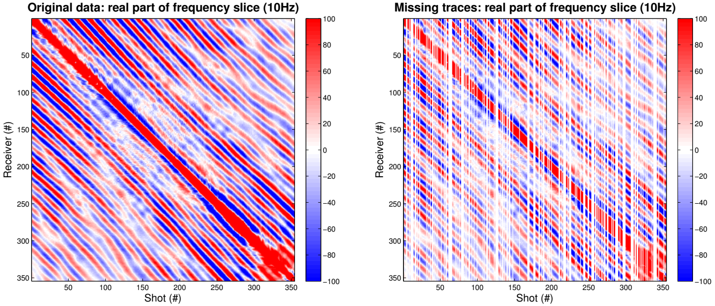
$$\underset{x \in \mathbb{R}^m}{\text{minimize}} \quad \frac{1}{2}\left\|RMC^\top x - y\right\|_2^2 + \lambda\left\|x\right\|_1,$$

where

- RM is a restriction matrix operator (restricts to 60% original data set), and
- y is the vectorized restricted data.

SEISMIC DATA INTERPOLATION

- Frequency slice (10 Hz) for sequential source acquisition from Gulf of Suez
  - e.g., [Kumar, Aravkin, and Herrmann, 2012]



SEISMIC DATA INTERPOLATION

Compare to SPGL1, which solves

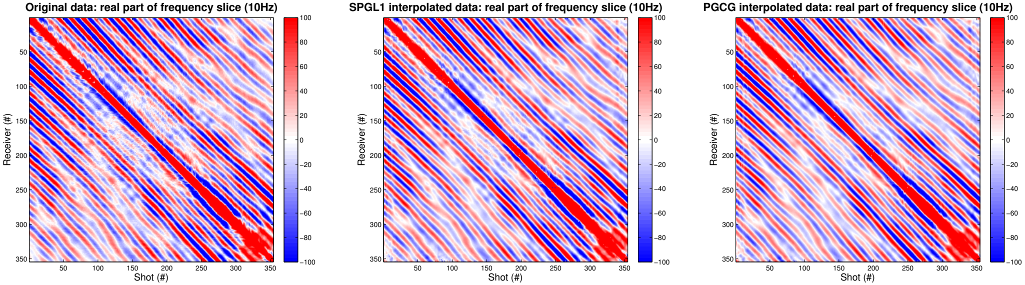
$$\underset{x\in\mathbb{R}^m}{\text{minimize}} \quad \|x\|_1 \quad \text{s.t. } \|RMC^\top x - y\|_2 \leq \sigma. \tag{1}$$

To ensure a valid comparison, we choose regularization parameter

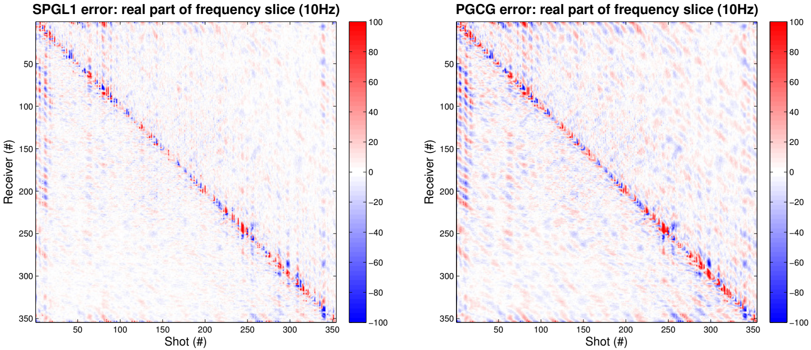
$$\lambda = \|C(RM)^\top r_\sigma\|_\infty,$$

where r_σ is the residual corresponding to the solution of (1) found by SPGL1.

SEISMIC DATA INTERPOLATION



SEISMIC DATA INTERPOLATION



Matrix-vector products		
	A	A^T
SPGL1	139	102
GPCG	147	148

FUTURE WORK

- Accelerated methods for Phase 1
  - e.g., FISTA [Beck and Teboulle, 2009]
- Continuation on λ, where $\lambda_k \to \lambda$
- Extending to complex valued data: $\|z\|_1$, where $z \in \mathbb{C}^n$

Thank you!

REFERENCES

- A. Beck and M. Teboulle. *A fast iteration shrinkage-thresholding algorithm for linear inverse problems*. SIAM J. Img. Sci., 2009.

- E. van den Berg and M. P. Friedlander. *SPGL1: A solver for large-scale sparse reconstruction*. http://www.cs.ubc.ca/labs/scl/spgl1, 2007.

- R. Kumar, A. Y. Aravkin and F. J. Herrmann. *Fast methods for rank minimization with applications in seismic-data interpolation*. Technical Report, 2012.

- L. Miao and F. J. Herrmann. *Acceleration on Sparse Promoting Seismic Applications*. CSEG conference proceedings, 2013.

- M. Schmidt, E. van den Berg, M. P. Friedlander and K. Murphy. *Optimizing costly functions with simple constraints: a limited-memory projected quasi-Newton algorithm*. Proceedings of the 12^{th} International Conference on Artificial Intelligence and Statistics (AISTATS), 2009