Noisy Signal Recovery via Iterative Reweighted L1-Minimization

Deanna Needell

UC Davis / Stanford University

Asilomar SSC, November 2009

Setup

1. Suppose x is an unknown signal in $\mathbb{R}^d$.
2. Design measurement matrix $\Phi : \mathbb{R}^d \to \mathbb{R}^m$.
3. Collect noisy measurements $u = \Phi x + e$.

$$\begin{bmatrix} u \end{bmatrix} = \begin{bmatrix} \quad & \Phi & \quad \end{bmatrix} \begin{bmatrix} x \\ \\ \end{bmatrix} + \begin{bmatrix} e \end{bmatrix}$$

4. Problem: Reconstruct signal x from measurements u
5. Wait, isn't this impossible?
   - Assume x is s-sparse: $\|x\|_0 \stackrel{\text{def}}{=} |\text{supp}(x)| \leq s \ll d$.

Setup

1. Suppose x is an unknown signal in $\mathbb{R}^d$.
2. Design measurement matrix $\Phi : \mathbb{R}^d \to \mathbb{R}^m$.
3. Collect noisy measurements $u = \Phi x + e$.

$$\begin{bmatrix} u \end{bmatrix} = \begin{bmatrix} \quad & \Phi & \quad \end{bmatrix} \begin{bmatrix} x \\ \\ \\ \end{bmatrix} + \begin{bmatrix} e \end{bmatrix}$$

4. Problem: Reconstruct signal x from measurements u
5. Wait, isn't this impossible?
   - Assume x is s-sparse: $\|x\|_0 \stackrel{\text{def}}{=} |\operatorname{supp}(x)| \le s \ll d$.

Setup

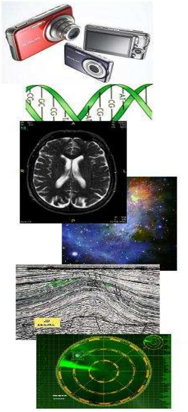
1. Suppose x is an unknown signal in $\mathbb{R}^d$.
2. Design measurement matrix $\Phi : \mathbb{R}^d \to \mathbb{R}^m$.
3. Collect noisy measurements $u = \Phi x + e$.

$$\begin{bmatrix} u \end{bmatrix} = \begin{bmatrix} & \Phi & \end{bmatrix} \begin{bmatrix} x \end{bmatrix} + \begin{bmatrix} e \end{bmatrix}$$

4. Problem: Reconstruct signal x from measurements u
5. Wait, isn't this impossible?
   - Assume x is s-sparse: $\|x\|_0 \stackrel{\text{def}}{=} |\text{supp}(x)| \le s \ll d$.

Applications

- Compressive Imaging
- Computational Biology
- Medical Imaging
- Astronomy
- Geophysical Data Analysis
- Compressive Radar
- Many more (see www.dsp.ece.rice.edu)



How can we reconstruct?

Obvious way:

Suppose the matrix Φ is one-to-one on the set of sparse vectors and $e = 0$. Set

$$\hat{x} = \operatorname{argmin} ||z||_0 \quad \text{such that} \quad \Phi z = u.$$

Then $\hat{x} = x$!

Bad news:

This would require a search through $\binom{d}{s}$ subspaces! Not numerically feasible.

How can we reconstruct?

Obvious way:

Suppose the matrix Φ is one-to-one on the set of sparse vectors and $e = 0$. Set

$$\hat{x} = \text{argmin}\, ||z||_0 \quad \text{such that} \quad \Phi z = u.$$

Then $\hat{x} = x$!

Bad news:

This would require a search through $\binom{d}{s}$ subspaces! Not numerically feasible.

How else can we reconstruct?

Geometric Idea

Minimizing the ℓ_0-ball is too hard, so let's try a different one.

Our favorites...

- Least Squares
- L1-Minimization (using Linear Programming)

Which one?

How else can we reconstruct?

Geometric Idea

Minimizing the ℓ_0-ball is too hard, so let's try a different one.

Our favorites...

- Least Squares
- L1-Minimization (using Linear Programming)

Which one?

Which one?

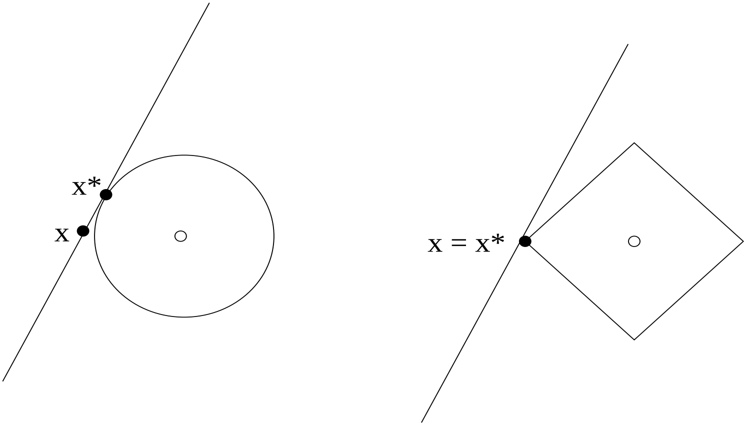


Figure: Minimizing the ℓ_2 versus the ℓ_1 balls.

What do we assume about Φ?

Restricted Isometry Property (RIP)

- The s^{th} restricted isometry constant of Φ is the smallest δ_s such that

$$(1-\delta_s)\|x\|_2 \leq \|\Phi x\|_2 \leq (1+\delta_s)\|x\|_2 \quad \text{whenever } \|x\|_0 \leq s.$$

- For Gaussian or Bernoulli measurement matrices, with high probability

$$\delta_s \leq c < 1 \quad \text{when } m \gtrsim s \log d.$$

- Random Fourier and others with fast multiply have similar property.

What do we assume about Φ?

Restricted Isometry Property (RIP)

- The s^{th} restricted isometry constant of Φ is the smallest δ_s such that

$$(1-\delta_s)\|x\|_2 \le \|\Phi x\|_2 \le (1+\delta_s)\|x\|_2 \quad \text{whenever } \|x\|_0 \le s.$$

- For Gaussian or Bernoulli measurement matrices, with high probability

$$\delta_s \le c < 1 \quad \text{when } m \gtrsim s \log d.$$

- Random Fourier and others with fast multiply have similar property.

Proven Results

L1-Minimization [Candès-Tao]

Assume that the measurement matrix Φ satisfies the RIP with $\delta_{2s} < \sqrt{2} - 1$. Then every s-sparse vector x can be exactly recovered from its measurements $u = \Phi x$ as a unique solution to the linear optimization problem:

$$\hat{x} = \operatorname{argmin} ||z||_1 \quad \text{such that} \quad \Phi z = u.$$

Numerical Results

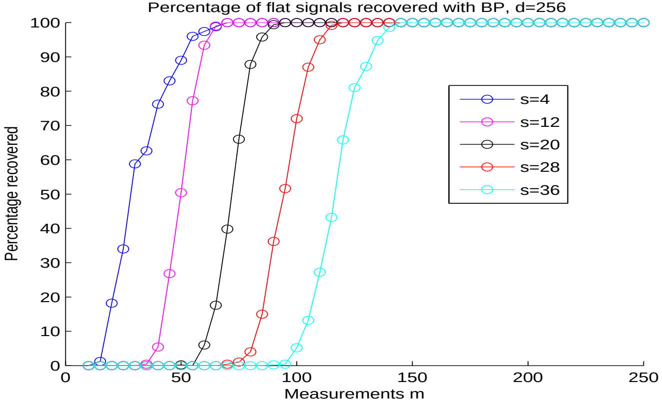


Figure: The percentage of sparse flat signals exactly recovered by Basis Pursuit as a function of the number of measurements m in dimension $d = 256$ for various levels of sparsity s.

What about noise?

Noisy Formulation

For a non-sparse vector x with noisy measurements $u = \Phi x + e$,

$$\hat{x} = \operatorname{argmin} ||z||_1 \quad \text{such that} \quad ||\Phi z - u||_2 \leq \varepsilon. \tag{1}$$

L1-Minimization [Candès-Romberg-Tao]

Let Φ be a measurement matrix satisfying the RIP with $\delta_{2s} < \sqrt{2} - 1$. Then for any *arbitrary* signal and corrupted measurements $u = \Phi x + e$ with $||e||_2 \leq \varepsilon$, the solution $\hat{x}$ to (1) satisfies

$$||\hat{x} - x||_2 \leq C_s \cdot \varepsilon + C_s' \cdot \frac{||x - x_s||_1}{\sqrt{s}}.$$

Note: As $\delta_{2s} \to \sqrt{2} - 1$, $C_s, C_s' \to \infty$!!

What about noise?

Noisy Formulation

For a non-sparse vector x with noisy measurements $u = \Phi x + e$,

$$\hat{x} = \text{argmin}\, ||z||_1 \quad \text{such that} \quad \|\Phi z - u\|_2 \leq \varepsilon. \tag{1}$$

L1-Minimization [Candès-Romberg-Tao]

Let Φ be a measurement matrix satisfying the RIP with $\delta_{2s} < \sqrt{2} - 1$. Then for any *arbitrary* signal and corrupted measurements $u = \Phi x + e$ with $\|e\|_2 \leq \varepsilon$, the solution $\hat{x}$ to (1) satisfies

$$\|\hat{x} - x\|_2 \leq C_s \cdot \varepsilon + C_s' \cdot \frac{\|x - x_s\|_1}{\sqrt{s}}.$$

Note: As $\delta_{2s} \to \sqrt{2} - 1$, $C_s, C_s' \to \infty$!!

What about noise?

Noisy Formulation

For a non-sparse vector x with noisy measurements $u = \Phi x + e$,

$$\hat{x} = \text{argmin}\, ||z||_1 \quad \text{such that} \quad \|\Phi z - u\|_2 \leq \varepsilon. \tag{1}$$

L1-Minimization [Candès-Romberg-Tao]

Let Φ be a measurement matrix satisfying the RIP with $\delta_{2s} < \sqrt{2} - 1$. Then for any *arbitrary* signal and corrupted measurements $u = \Phi x + e$ with $\|e\|_2 \leq \varepsilon$, the solution $\hat{x}$ to (1) satisfies

$$\|\hat{x} - x\|_2 \leq C_s \cdot \varepsilon + C_s' \cdot \frac{\|x - x_s\|_1}{\sqrt{s}}.$$

Note: As $\delta_{2s} \to \sqrt{2} - 1$, $C_s, C_s' \to \infty$!!

What about noise?

Noisy Formulation

For a non-sparse vector x with noisy measurements $u = \Phi x + e$,

$$\hat{x} = \operatorname{argmin} ||z||_1 \quad \text{such that} \quad \|\Phi z - u\|_2 \leq \varepsilon. \tag{1}$$

L1-Minimization [Candès-Romberg-Tao]

Let Φ be a measurement matrix satisfying the RIP with $\delta_{2s} < \sqrt{2} - 1$. Then for any *arbitrary* signal and corrupted measurements $u = \Phi x + e$ with $\|e\|_2 \leq \varepsilon$, the solution $\hat{x}$ to (1) satisfies

$$\|\hat{x} - x\|_2 \leq C_s \cdot \varepsilon + C'_s \cdot \frac{\|x - x_s\|_1}{\sqrt{s}}.$$

Note: As $\delta_{2s} \to \sqrt{2} - 1$, $C_s, C'_s \to \infty$!!

Numerical Results

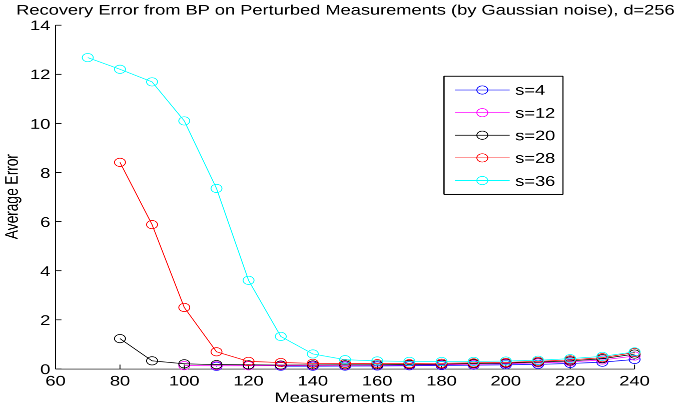


Figure: The recovery error of L1-Minimization under perturbed measurements ($\|e\|_2 = 0.5$) as a function of the number of measurements m in dimension $d = 256$ for various levels of sparsity s.

What if we are close?

- Suppose we recover $\hat{x} \approx x$
- Most likely, this means $\hat{x}_i \approx x_i$
- In particular, $\hat{x}_i$ is small/large when x_i is small/large

Weighted L1

$$\hat{x}^{(2)} = \operatorname*{argmin}_z \sum_{i=1}^{d} \left| \frac{z_i}{\hat{x}_i} \right| \quad \text{such that} \quad \|\Phi z - u\|_2 \leq \varepsilon$$

What if we are close?

- Suppose we recover $\hat{x} \approx x$
- Most likely, this means $\hat{x}_i \approx x_i$
- In particular, $\hat{x}_i$ is small/large when x_i is small/large

Weighted L1

$$\hat{x}^{(2)} = \operatorname*{argmin}_z \sum_{i=1}^{d} \left| \frac{z_i}{\hat{x}_i} \right| \quad \text{such that} \quad \|\Phi z - u\|_2 \leq \varepsilon$$

What if we are close?

- Suppose we recover $\hat{x} \approx x$
- Most likely, this means $\hat{x}_i \approx x_i$
- In particular, $\hat{x}_i$ is small/large when x_i is small/large

Weighted L1

$$\hat{x}^{(2)} = \underset{z}{\operatorname{argmin}} \sum_{i=1}^{d} \left| \frac{z_i}{\hat{x}_i + a} \right| \quad \text{such that} \quad \|\Phi z - u\|_2 \leq \varepsilon$$

Weighted Geometry

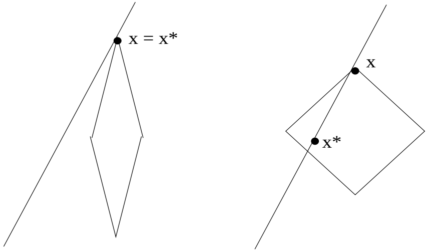


Figure: The geometry of the weighted ℓ_1-ball.

- Noise-free case: In cases where $\hat{x} \neq x$, we should have that $\hat{x}^{(2)}$ is closer to x, or even *equal*.
- Noisy case: This implies $\hat{x}^{(2)}$ should be closer to x than $\hat{x}$ was.
- Can we repeat this again?

Reweighted ℓ_1-minimization (RWL1)

INPUT: Measurement vector $u \in \mathbb{R}^m$, stability parameter a
OUTPUT: Reconstructed vector $\hat{x}$

Initialize Set the weights $w_i = 1$ for $i = 1 \ldots d$.

Approximate Solve the reweighted ℓ_1-minimization problem:

$$\hat{x} = \underset{\hat{x} \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{i=1}^{d} w_i \hat{x}_i \text{ subject to } \|\Phi \hat{x} - u\|_2 \leq \varepsilon.$$

Update Reset the weights:

$$w_i = \frac{1}{|\hat{x}_i| + a}.$$

Numerical Results

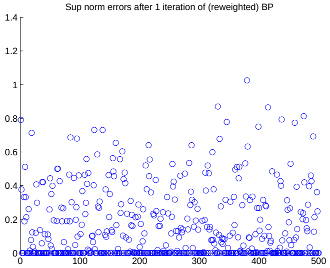


Figure: ℓ_∞-norm error for reweighted L1 in the noise-free case

Numerical Results

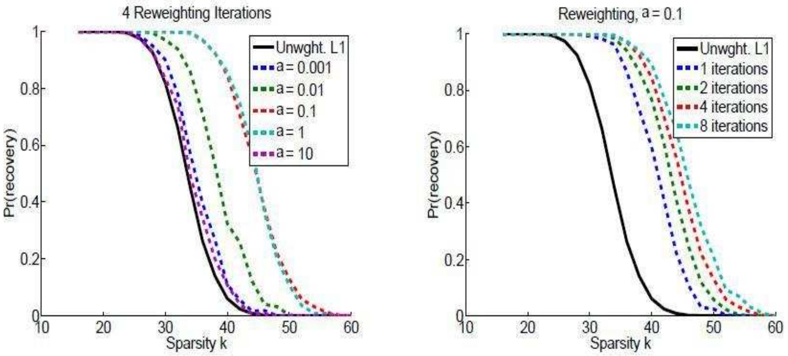


Figure: Probability of reconstruction [Candès-Wakin-Boyd].

Numerical Results with noise

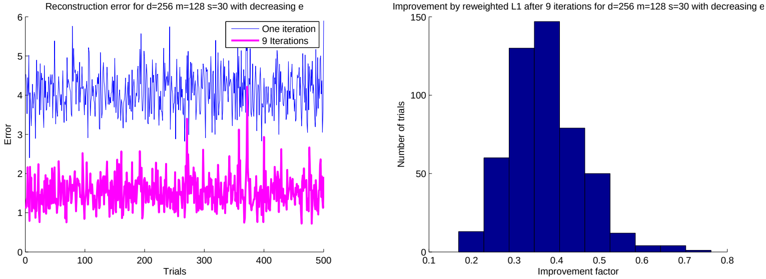


Figure: Improvements in the ℓ_2 reconstruction error using reweighted ℓ_1-minimization versus standard ℓ_1-minimization for sparse Gaussian signals.

Numerical Results with noise

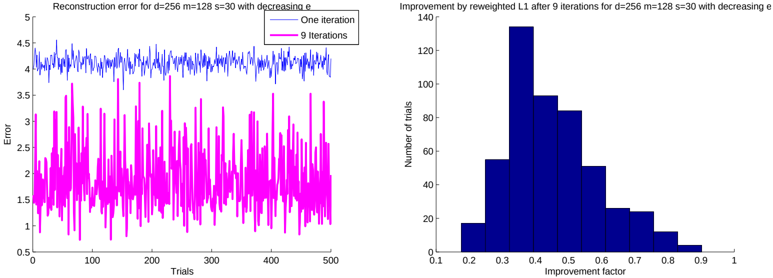


Figure: Improvements in the ℓ_2 reconstruction error using reweighted ℓ_1-minimization versus standard ℓ_1-minimization for sparse Bernoulli signals.

Observations

- The noiseless case suggests that an ℓ_∞-norm bound may be required for RWL1 to succeed.
- In the noisy case it is clear that we cannot recover signal coordinates that are below some threshold.
- If each iteration of RWL1 improves the error, perhaps we should take $a \to 0$. (Recall $w_i = \frac{1}{|\hat{x}_i|+a}$).

Observations

- The noiseless case suggests that an ℓ_∞-norm bound may be required for RWL1 to succeed.
- In the noisy case it is clear that we cannot recover signal coordinates that are below some threshold.
- If each iteration of RWL1 improves the error, perhaps we should take $a \to 0$. (Recall $w_i = \frac{1}{|\hat{x}_i| + a}$).

Observations

- The noiseless case suggests that an ℓ_∞-norm bound may be required for RWL1 to succeed.
- In the noisy case it is clear that we cannot recover signal coordinates that are below some threshold.
- If each iteration of RWL1 improves the error, perhaps we should take $a \to 0$. (Recall $w_i = \frac{1}{|\hat{x}_i| + a}$).

Main Results

RWL1 - Sparse case [N]

Assume Φ satisfies the RIP with $\delta_{2s} \leq \delta$ where $\delta < \sqrt{2} - 1$. Let x be an s-sparse vector with noisy measurements $u = \Phi x + e$ where $\|e\|_2 \leq \varepsilon$. Assume the smallest nonzero coordinate μ of x satisfies $\mu \geq \frac{4\alpha\varepsilon}{1-\rho}$. Then the limiting approximation from reweighted ℓ_1-minimization satisfies

$$\|x - \hat{x}\|_2 \leq C''\varepsilon,$$

where $C'' = \frac{2\alpha}{1+\rho}$, $\rho = \frac{\sqrt{2}\delta}{1-\delta}$ and $\alpha = \frac{2\sqrt{1+\delta}}{1-\delta}$.

Remarks

- Without noise, this result coincides with previous results on L1.
- The key improvement: As $\delta \to \sqrt{2} - 1$, C'' remains bounded.
- The error bound is the *limiting* bound, but a recursive relation in the proof gives exact improvements per iteration. We show in practice it is attained quite quickly.
- For signals whose smallest non-zero coefficient μ does not satisfy the condition of the theorem, we may apply the theorem to those coefficients that do satisfy this requirement, and treat the others as noise...

Remarks

- Without noise, this result coincides with previous results on L1.
- The key improvement: As $\delta \to \sqrt{2} - 1$, C'' remains bounded.
- The error bound is the *limiting* bound, but a recursive relation in the proof gives exact improvements per iteration. We show in practice it is attained quite quickly.
- For signals whose smallest non-zero coefficient μ does not satisfy the condition of the theorem, we may apply the theorem to those coefficients that do satisfy this requirement, and treat the others as noise...

Remarks

- Without noise, this result coincides with previous results on L1.
- The key improvement: As $\delta \to \sqrt{2} - 1$, C'' remains bounded.
- The error bound is the *limiting* bound, but a recursive relation in the proof gives exact improvements per iteration. We show in practice it is attained quite quickly.
- For signals whose smallest non-zero coefficient μ does not satisfy the condition of the theorem, we may apply the theorem to those coefficients that do satisfy this requirement, and treat the others as noise...

Remarks

- Without noise, this result coincides with previous results on L1.
- The key improvement: As $\delta \to \sqrt{2} - 1$, C'' remains bounded.
- The error bound is the *limiting* bound, but a recursive relation in the proof gives exact improvements per iteration. We show in practice it is attained quite quickly.
- For signals whose smallest non-zero coefficient μ does not satisfy the condition of the theorem, we may apply the theorem to those coefficients that do satisfy this requirement, and treat the others as noise...

Extension

RWL1 - non-sparse extension [N]

Assume Φ satisfies the RIP with $\delta_{2s} \leq \sqrt{2} - 1$. Let x be an arbitrary vector with noisy measurements $u = \Phi x + e$ where $\|e\|_2 \leq \varepsilon$. Assume the smallest nonzero coordinate μ of x_s satisfies $\mu \geq \frac{4\alpha\varepsilon_0}{1-\rho}$, where $\varepsilon_0 = 1.2(\|x - x_s\|_2 + \frac{1}{\sqrt{s}}\|x - x_s\|_1) + \varepsilon$. Then the limiting approximation from reweighted ℓ_1-minimization satisfies

$$\|x - \hat{x}\|_2 \leq \frac{4.1\alpha}{1+\rho}\Big(\frac{\|x - x_{s/2}\|_1}{\sqrt{s}} + \varepsilon\Big),$$

and

$$\|x - \hat{x}\|_2 \leq \frac{2.4\alpha}{1+\rho}\Big(\|x - x_s\|_2 + \frac{\|x - x_s\|_1}{\sqrt{s}} + \varepsilon\Big),$$

where ρ and α are as before.

Theoretical Results

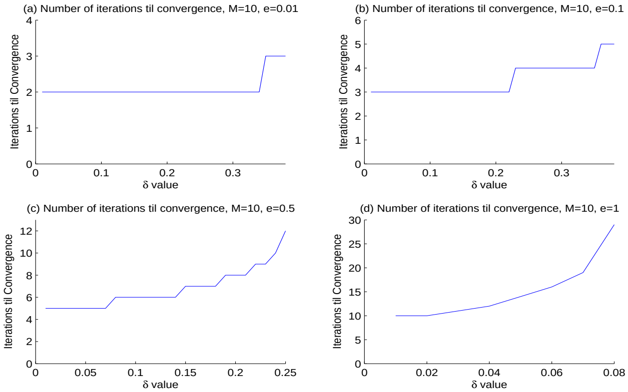


Figure: Number of iterations required for theoretical error bounds to reach limiting theoretical error when (a) $\mu = 10$, $\varepsilon = 0.01$, (b) $\mu = 10$, $\varepsilon = 0.1$, (c) $\mu = 10$, $\varepsilon = 0.5$, (d) $\mu = 10$, $\varepsilon = 1.0$.

Recent work

- Wipf-Nagarajan elaborate on convergence and show connections to reweighted ℓ_2-minimization.
- Wipf-Nagarajan also show that a non-separable variant has desirable properties.
- Xu-Khajehnejad-Avestimehr-Hassibi provide a theoretical foundation for the analysis of RWL1 and show that for a nontrivial class of signals, a variant of RWL1 indeed can improve upon L1 in the noiseless case.

For more information

E-mail:

- dneedell@stanford.edu

Web: http://www-stat.stanford.edu/~dneedell

References:

- Candes, Wakin, Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization", J. Fourier Anal. Appl., **14** 877-905.
- Needell, "Noisy signal recovery via iterative reweighted L1-minimization," 2009.
- Wipf and Nagarajan, "Solving Sparse Linear Inverse Problems: Analysis of Reweighted ℓ_1 and ℓ_2 Methods," *J. of Selected Topics in Signal Processing, Special Issue on Compressive Sensing*, 2010.
- Xu, Khajehnejad, Avestimehr, and Hassibi, "Breaking through the Thresholds: an Analysis for Iterative Reweighted ℓ_1 Minimization via the Grassmann Angle Framework," *Proc. Allerton Conference on Communication, Control, and computing*, Sept. 2009.