The Difference of L_1 and L_2 norms for Compressive Sensing and Image Processing

Yifei Lou

Department of Mathematical Sciences
University of Texas Dallas

May 25, 2016

Joint with Penghang Yin and Jack Xin

Partially supported by NSF DMS 1522786

Outline

1. Introduction to Compressive Sensing
2. A non-convex approach to promote sparsity: L_1-L_2
3. One application: point-source super-resolution
4. Conclusions and future works

To find a sparse vector x,

$$\hat{x} = \text{argmin}_x \|x\|_0 \quad \text{s.t.} \quad Ax = b.$$

To find a sparse vector x,

$$\hat{x} = \text{argmin}_x \|x\|_0 \quad \text{s.t.} \quad Ax = b.$$

This is NP-hard.

To find a sparse vector x,

$$\hat{x} = \text{argmin}_x \|x\|_0 \quad \text{s.t.} \quad Ax = b.$$

This is NP-hard.

One popular trend in CS is to replace L_0 by L_1, i.e.,

$$\hat{x} = \text{argmin}_x \|x\|_1 \quad \text{s.t.} \quad Ax = b.$$

To find a sparse vector x,

$$\hat{x} = \text{argmin}_x \|x\|_0 \quad \text{s.t.} \quad Ax = b.$$

This is NP-hard.

One popular trend in CS is to replace L_0 by L_1, i.e.,

$$\hat{x} = \text{argmin}_x \|x\|_1 \quad \text{s.t.} \quad Ax = b.$$

The big bang of CS started when restricted isometry property (RIP) was derived to guarantee the success of L_1 minimization.

Candes-Romberg-Tao (2006)

Coherence

Another sparse recovery guarantee is based on coherence.

$$\|x\|_0 \leqslant \frac{1}{2}(1 + \mu(A)^{-1}),$$

where coherence of a matrix $A = [a_1, \cdots, a_n]$ is defined as

$$\mu(A) = \max_{i \neq j} \frac{|a_i^T a_j|}{\|a_i\| \|a_j\|} .$$

Coherence

Another sparse recovery guarantee is based on coherence.

$$\|x\|_0 \leqslant \frac{1}{2}(1+\mu(A)^{-1}),$$

where coherence of a matrix $A = [a_1, \cdots, a_n]$ is defined as

$$\mu(A) = \max_{i \neq j} \frac{|a_i^T a_j|}{\|a_i\| \|a_j\|} \, .$$

Two extreme cases are

- $\mu \sim 0 \Rightarrow$ incoherent matrix
- $\mu \sim 1 \Rightarrow$ coherent matrix

Coherence

Another sparse recovery guarantee is based on coherence.

$$\|x\|_0 \leqslant \frac{1}{2}(1+\mu(A)^{-1}),$$

where coherence of a matrix $A = [a_1, \cdots, a_n]$ is defined as

$$\mu(A) = \max_{i \neq j} \frac{|a_i^T a_j|}{\|a_i\| \|a_j\|}.$$

Two extreme cases are

- $\mu \sim 0 \Rightarrow$ incoherent matrix
- $\mu \sim 1 \Rightarrow$ coherent matrix

What if the matrix is coherent?

Outline

1. Introduction to Compressive Sensing
2. A non-convex approach to promote sparsity: L_1-L_2
3. One application: point-source super-resolution
4. Conclusions and future works

- Nonconvex regularizations: L_0, L_p for $p \in (0, 1)$

- Nonconvex regularizations: L_0, L_p for $p \in (0, 1)$
- We consider $L_1 - L_2$, solved by difference of convex algorithm (DCA)

E. Esser, Y. Lou and J. Xin, SIAM on Imaging Sciences 2013
P. Yin, Y. Lou, Q. He and J. Xin, SIAM Sci. Comput., 2015
Y. Lou, P. Yin, Q. He and J. Xin, J. Sci. Comput., 2015

- Nonconvex regularizations: L_0, L_p for $p \in (0, 1)$
- We consider $L_1 - L_2$, solved by difference of convex algorithm (DCA)

Lipschitz continuous

Free of parameter

good for coherent compressive sensing

E. Esser, Y. Lou and J. Xin, SIAM on Imaging Sciences 2013
P. Yin, Y. Lou, Q. He and J. Xin, SIAM Sci. Comput., 2015
Y. Lou, P. Yin, Q. He and J. Xin, J. Sci. Comput., 2015

Algorithms

We solve the $L_1 - L_2$ minimization via DCA.

$$\min_{x \in \mathrm{R}^N} F(x) = \frac{1}{2}\|Ax - b\|_2^2 + \lambda(\|x\|_1 - \|x\|_2)$$

Decompose $F(x) = G(x) - H(x)$, where

$$\begin{cases} G(x) = \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1 \\ H(x) = \lambda\|x\|_2 \end{cases}$$

An iterative scheme is,

$$x^{n+1} = \arg\min_{x \in \mathrm{R}^N} \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1 - \langle x, \frac{\lambda x^n}{\|x^n\|_2}\rangle$$

We consider an over-sampled DCT matrix

$$A = [\mathbf{a}_1, \cdots, \mathbf{a}_N] \in \mathbb{R}^{M \times N}$$

with

$$\mathbf{a}_j = \frac{1}{\sqrt{N}} \cos(\frac{2\pi j \mathbf{w}}{F}) \ , j = 1, \cdots, N$$

where $\mathbf{w}$ is a random vector of length M.

The larger F is, the more coherent the matrix. Take a 100×1000 matrix for an example:

F	coherence
1	0.3981
10	0.9981
20	0.9999

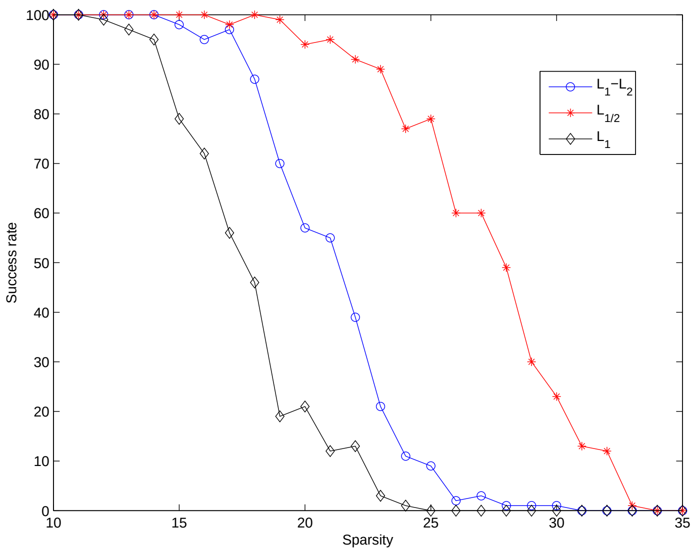


Figure: Success rates of incoherent matrices, $F = 1$.

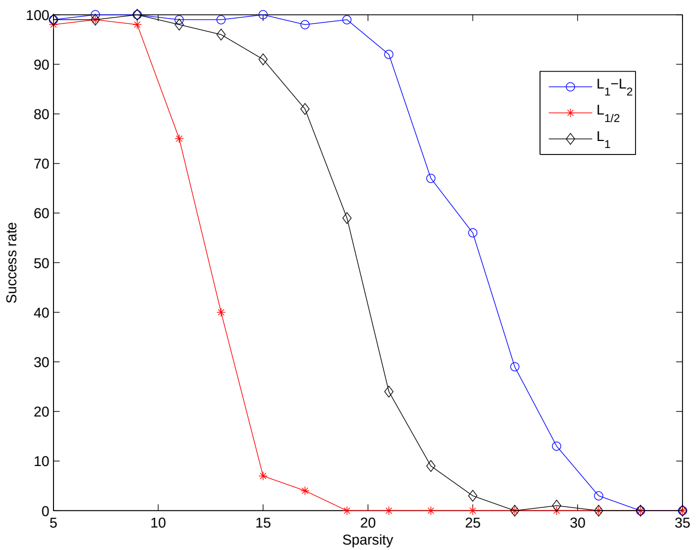


Figure: Success rates of coherent matrices, $F = 20$.

RIP for L_1-L_2

L_1-L_2 has RIP but more stringent than the one for L_1.

Convergence

The limit point of DCA minimizing sequence is a stationary point.

Rank property

The sparsity of any local minimizer of L_1-L_2 is less than or equal to the rank of A.

Outline

Problem setup

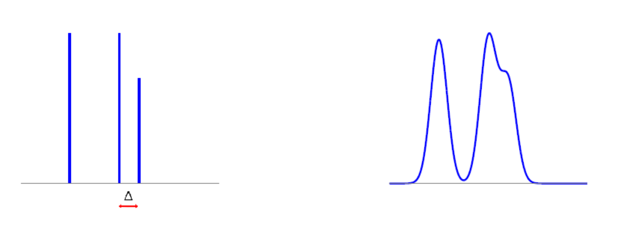
The super-resolution problem discussed here is different to image zooming or magnification, but aiming to recover a real-valued signal from its low-frequency measurements.

A mathematical model is expressed as

$$b_k = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x_t e^{-i2\pi kt/N}, \qquad |k| \leq f_c,$$

where $x \in \mathbb{R}^N$ is a vector of interest, and $b \in \mathbb{C}^n$ is the given low frequency information with $n = 2f_c + 1$ $(n < N)$.

Point source with minimum separation



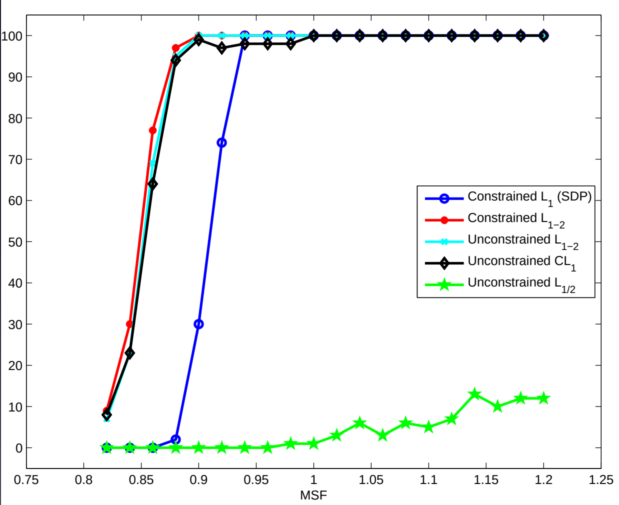
Theorem by Candés and Fernandez-Granda 2012

Let $T = \{t_j\}$ be the support of x. If the minimum distance obeys

$$\triangle(T) \geq 2 \cdot N/f_c,$$

then x is the unique solution to L_1 minimization. If x is real-valued, then the minimum gap can be lowered to $1.26 \cdot N/f_c$.

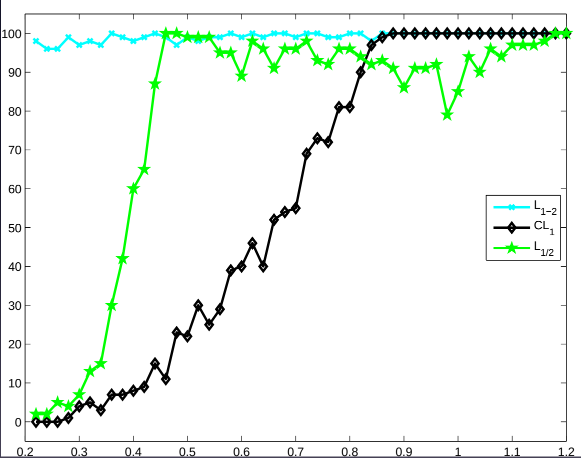
Success rates



Y. Lou, P. Yin and J. Xin, J. Sci. Comput., 2016 to appear

Rank property

The sparsity of any local minimizer of L_1-L_2 is smaller than or equal to the rank of A.



Define an objective function for an unconstrained minimization

$$f(x) = \lambda(\|x\|_1 - \|x\|_2) + \frac{1}{2}\|Ax - b\|_2^2$$

Theorem

Suppose $\lambda < \min\{\frac{\|A^T b\|_2}{\sqrt{N}+\|A\|^2}, \frac{\|A^T b\|_2}{\sqrt{N}+1}\}$. Let x^* be any limit point of the DCA minimizing sequence. Then we have either $\|x^*\|_0 \leq n$ (rank property) or there exists $d \in \mathbb{R}^N$ such that $F(x^* + d) < F(x^*)$ and supp$(d) \subseteq$ supp(x^*) (so we can choose d s.t. $x^* + d$ is sparser than x^*.)

This theorem motivates us to study nonconvex optimization, regarding how to jump among stationary points.

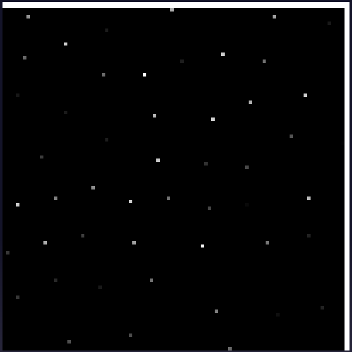


Figure: A 2D point-source example with minimum separation.

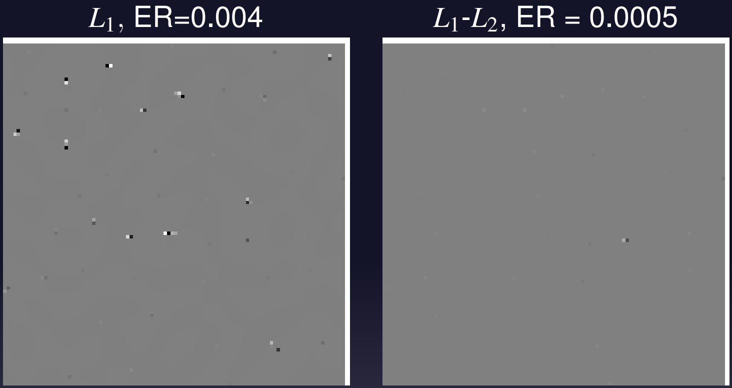


Figure: Comparison of error map, i.e. the difference between reconstruction and ground-truth.

Conclusions

1. L_1-L_2 is always better than L_1, and is better than L_p for highly coherent matrices.

2. For super-resolution, non-convex methods outperform standard L_1 when a necessary condition for perfect reconstruction is not met.

3. We can expect L_1-L_2 to lead to sparse solutions.

Thank you!