A numerical exploration of Compressed Sampling Recovery

Charles Dossal[1] Gabriel Peyré[2] Jalal Fadili[3]

[1]IMB CNRS-Université Bordeaux 1

[2]CEREMADE CNRS-Université Paris Dauphine

[3]GREYC CNRS-ENSICAEN-Université de Caen

SPARS'09 Saint-Malo

Setting

The problem

- Linear measurements $y = Ax \in \mathbb{R}^P$, $x \in \mathbb{R}^N$, $A \in \mathbb{R}^{P\times N}$.
- Problem : Recover x from y when $P < N$, ill-posed inversion.
- Prior : x is sparse $\Leftrightarrow$ $\|x\|_0 < P$ is small.

The framework

- x_0 is identifiable if and only if x_0 is the solution of

$$\min \|x\|_1 \text{ such that } Ax = y \tag{1}$$

  for $y = Ax_0$.
- We focus on
  - Conditions on $\|x\|_0$ ensuring identifiability.
  - Random matrices (Gaussian and USE).

The RIP approach

Definition of RIP

- For $A \in \mathbb{R}^{P \times N}$, $\delta_S^{\min}$ and $\delta_S^{\max}$ are the smallest real numbers such that for any x, $\|x\|_0 \leqslant S$,
$$(1 - \delta_S^{\min})\|x\|_2^2 \leqslant \|Ax\|_2^2 \leqslant (1 + \delta_S^{\max})\|x\|_2^2. \quad (2)$$

Theorem [Fourcart and Lai 08]

If
$$(4\sqrt{2} - 3)\delta_{2S}^{\min} + \delta_{2S}^{\max} < 4(\sqrt{2} - 1), \quad (3)$$

- All vectors x such that $\|x\|_0 \leqslant S$ are identifiable.
- Robustness to noise.
- Consistency if x is only compressible.

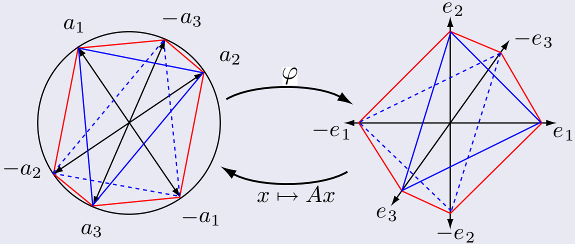
Random Matrices

- If A is a Gaussian matrix with iid entries, then with overwhelming probability, A satisfies (3) for $S = O\left(\frac{P}{\log(N/P)}\right)$.
- For $N/P = 2$, Blanchard et al. proved that (3) holds for $S \leqslant 0.003P$.
- This condition is global.

The geometrical viewpoint

Geometry of polytopes [Donoho, Donoho and Tanner]

- Identifiablility is a geometrical property.

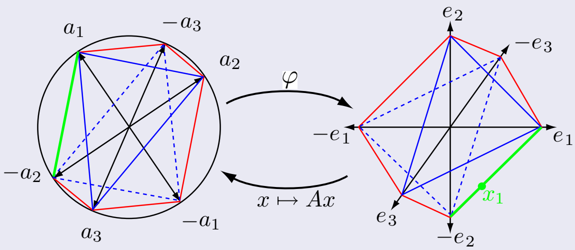


- x is identifiable if and only if $\frac{Ax}{\|x\|_1} \in \partial A(\mathcal{B}_{\ell_1})$.
- For $x \in \mathbb{R}^N$, $I = \{i,\ x_i \neq 0\}$, $f_x = \text{Conv.Hull}(\text{sign}\,(x_i)\, a_i)_{i \in I}$.
- x is identifiable $\Leftrightarrow$ f_x is an exterior facet of $A(\mathcal{B}_{\ell_1})$.

The geometrical viewpoint

Geometry of polytopes [Donoho, Donoho and Tanner]

- Identifiablility is a geometrical property.

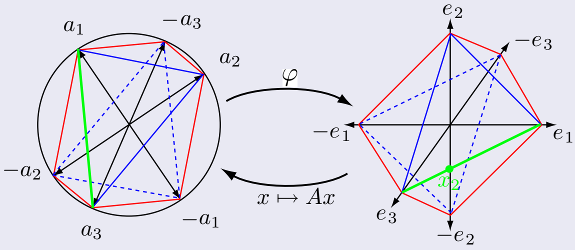


- x is identifiable if and only if $\frac{Ax}{\|x\|_1} \in \partial A(\mathcal{B}_{\ell_1})$.
- $x_1 = (1, -2, 0)/3$ is identifiable.

The geometrical viewpoint

Geometry of polytopes [Donoho, Donoho and Tanner]

- Identifiablility is a geometrical property.



- x is identifiable if and only if $\frac{Ax}{\|x\|_1} \in \partial A(\mathcal{B}_{\ell_1})$.
- $x_1 = (1, -2, 0)/3$ is identifiable.
- $x_2 = (2, 0, 5)/7$ is not.
- This condition is local.

The geometrical viewpoint

Counting k-faces of centro-symmetric polytopes [Donoho 04]

- If A is gaussian or USE, there is a function $\rho_N(.)$ such that w.o.p. on A, all sparse x with

$$\|x\|_0 \leqslant \rho_N(P/N)P \text{ are identifiable.} \quad (4)$$

- If A is gaussian or USE, x with randomly chosen support and sign, there is $\rho_F(.)$ such that w.o.p. on A, most sparse x with

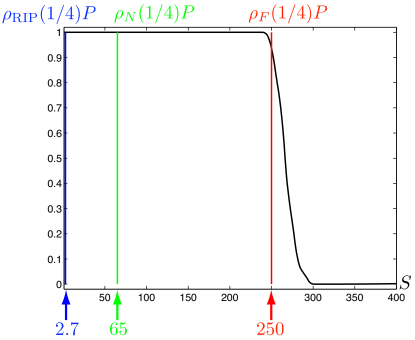
$$\|x\|_0 \leqslant \rho_F(P/N)P \text{ are identifiable.} \quad (5)$$

- No robustness to noise, no consistency if x is compressible.

Constants ρ_N and ρ_F

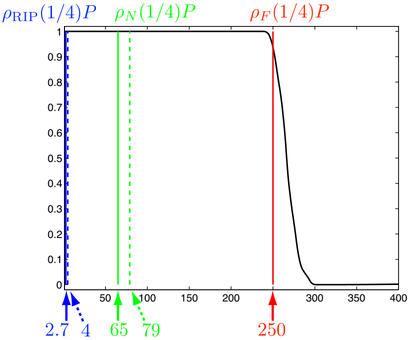
$$\rho_N(1/2) \sim 0.089,\ \rho_F(1/2) \sim 0.38$$
$$\rho_N(1/4) \sim 0.065,\ \rho_F(1/4) \sim 0.25.$$

Contributions



An efficient algorithm for two tasks

Contributions



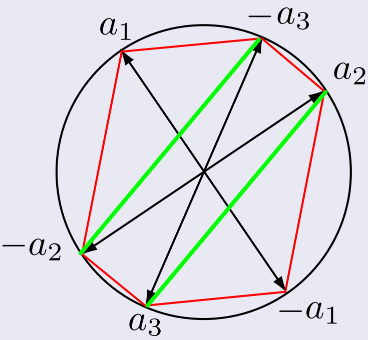
An efficient algorithm for two tasks

- The algorithm builds a non-identifiable vector x such that $\|x\|_0 = 79$.
- The algorithm reveals that $0.58 \leqslant \delta_{10}^{\max}$ and $0.42 \leqslant \delta_{10}^{\min}$ which implies that $(4\sqrt{2}-3)\delta_{2S}^{\min} + \delta_{2S}^{\max} > 4(\sqrt{2}-1)$.
- RIP is not verified for $S = 5$.
- $\mathrm{Proba}_{|I|=10}(\max(eig(A_I^t A_I)) \geqslant 1.58) \leqslant 4 \times 10^{-6}$.

Looking for interior facets

Identifiability and facets

- Non-identifiable vectors correspond to interior facets.



- We look for facets whose images are as close as possible to the origin.
- Exploit the rotation invariance of the matrix distribution.

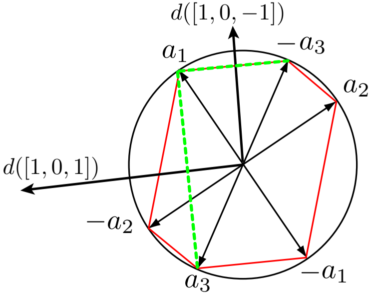
The distance between facets and the origin

Proposition (The vector $d(x)$)

The distance of f_x to the origin is equal to $\frac{1}{\|d(x)\|_2}$ where

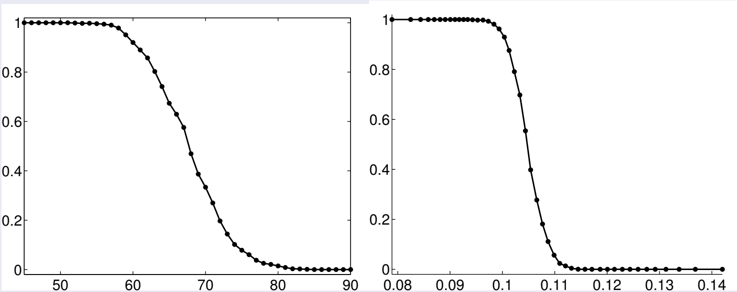
$$d(x) = A_I(A_I^t A_I)^{-1} sign(x_I)$$

with $A_I = A(i)$ and $x_I = x(i)$, for $i \in I$



$\|d\|_2$, A new indicator for ℓ_1 recovery.

- Example of proportion of identifiable vectors depending on sparsity and $\|d\|_2$, ($P = 250$, $N = 1000$).



Maximisation of $\|d\|_2$

A greedy approach

- Interior Facet $\Rightarrow$ Maximize $\|d\|_2$.
- Difficult problem $\Rightarrow$ Greedy extension :

$$x_1 = x + \sigma\Delta_j, \quad j \notin I \quad \sigma \in \{-1, 1\}$$

- Choose $(\sigma, j) = \arg\max_{k \notin I} \|d(x + \sigma\Delta_k)\|_2$.

$$\|d(x + \sigma\Delta_j)\|_2^2 = \|d(x)\|_2^2 + \|\tilde{a}_j\|^2 |\sigma - \langle d(x), a_j\rangle|^2$$

where $\tilde{a}_j \in Span(a_k)_{k \in I \cup j}$ and $\langle \tilde{a}_j, a_k\rangle = \delta_{j,k}$.

Approximate optimization

$$(\sigma, j) = \arg\max_{k \notin I} |\sigma - \langle d(x), a_k\rangle|$$

$$j = \arg\max_{k \notin I} |\langle d(x), a_k\rangle| \text{ and } \sigma = -\mathrm{sign}\left(\langle d(x), a_j\rangle\right).$$

Algorithm

Support Extension Algorithm

Initialization $\mathcal{P}_1 = \{\Delta_1, \cdots, \Delta_N\}$.

For $s = 2$ **to** S

$\mathcal{P}_s = \emptyset$

For $x \in \mathcal{P}_s$

Compute $j = \arg\max_{k \notin I} |\langle d(x), a_k \rangle|$ **and** $\sigma = -\mathrm{sign}\left(\langle d(x), a_j \rangle\right)$.

$\mathcal{P}_s = \mathcal{P}_s \cup \{x + \sigma\Delta_j\}$.

end

end

Numerical Results

Comparison with the asymptotic bound.

- Donoho proved $\rho(1/4) \sim 0.065$.

P	125	250	500	1000
$s^\star(1/4, P)$	10	20	42	79
$\lceil \rho(1/4) P \rceil$	9	17	33	65

- $N = 1000$ and $P = 250$, we are able to find a 20-sparse vector that is non-identifiable.
- Testing randomly 1000 vectors for each sparsity, we find only identifiable vectors besides 54 non zero components.
- It's not clear if it exists sparser non identifiable vectors.

Lower bound of RIP constants

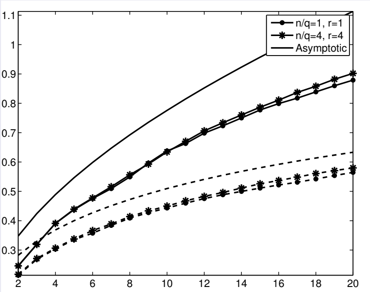
$d(x)$ and RIP constant.

$$d(x) = A_I^{+t}\mathrm{sign}\left(x_i\right) \Rightarrow 1 - \delta_S^{\min} \leqslant \frac{S}{\|d(x)\|_2^2} \leqslant 1 + \delta_S^{\max}$$

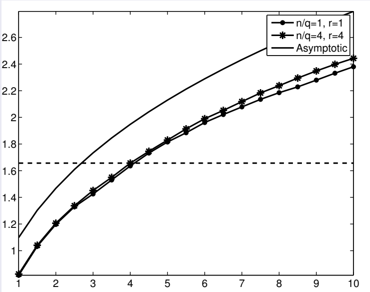
- Maximize $\|d(x)\|_2 \Rightarrow$ efficient lower bound of $\delta_S^{\min}$.
- Minimize $\|d(x)\|_2 \Rightarrow$ efficient lower bound of $\delta_S^{\max}$.
- $\|d(x)\|_2$ may be minimized using a similar algorithm.

Numerical Results

Comparison with Blanchard et al.



$\delta_{2S}^{\max}$ and $\delta_{2S}^{\min}$



$(4\sqrt{2}-3)\delta_{2S}^{\min}+\delta_{2S}^{\max}$

P	250	500	1000	2000
$s_0^\star(1/4, P)$	2	3	5	8
$\lceil \rho_{RIP}(1/4)P \rceil$	1	2	3	6

Take away messages

- Looking for sparse and non identifiable vectors,
  Looking for really ill-conditionned sub-matrices is ...
- Looking for a needle in a haystack.
- Geometry-inspired Greedy Algorithm

Main findings

- *Non-identifiable vectors*: Close to the theoretical bound of Donoho.
- *RIP constant estimation*: A numerical lower bound close the upper bound of Blanchard et al.