

Support identification of sparse vectors from random and noisy measurements.

Charles Dossal¹ Marie-Line Chabanol¹ Gabriel Peyré²
Jalal Fadili³

¹IMB CNRS-Université Bordeaux 1

²CEREMADE CNRS-Université Paris Dauphine

³GREYC CNRS-ENSICAEN-Université de Caen

Congrès SMAI Guidel 2011

Outline

- Setting
- ℓ_1 recovery: Overview
 - ℓ_1 minimization and geometry of polytopes.
 - Restricted Isometry Property.
 - Exact support recovery using LASSO.
- Contributions.
- Sketch of proof of the main result.

Setting

Noisy Gaussian measurements of sparse vectors

- Linear random measurements $y = Ax + w \in \mathbb{R}^n$,
 $x \in \mathbb{R}^N$, $A = (a_{ij})_{i \leq n, j \leq N} \in \mathbb{R}^{n \times N}$, $a_{ij} \sim \mathcal{N}(0, 1/n)$ and iid,
 $w \in \mathbb{R}^n$ and $\|w\|_2 \leq \varepsilon$.
- x is sparse $\Leftrightarrow \|x\|_0 < N$ is small.
- x is weakly sparse (compressible).

Questions

- Estimate x from y when $n < N$, ill-posed inverse problem.
- Estimate the support I of x from y .
- Stability to noise and robustness to compressibility.

Sparse Recovery Algorithms

A large choice of methods

- Greedy methods : Matching Pursuit, OMP, Cosamp, MCA ...
- Non convex optimization :

$$\min_{x \in \mathbb{R}^N} \|x\|_p \quad s.t. \quad y - Ax \in C, \quad p \in (0, 1), \quad C = \{0\} \text{ or } C = \mathcal{B}_{\ell_2}(\sigma)$$

- Convex optimization :

$$\min_{x \in \mathbb{R}^N} \|x\|_1 \quad s.t. \quad y - Ax \in C,$$

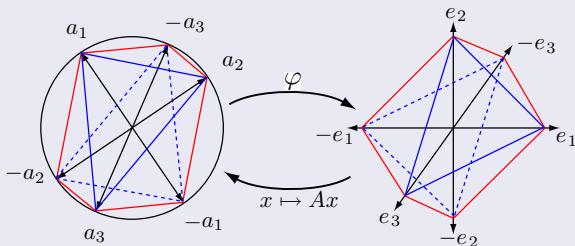
- $C = \{0\}$: exact ℓ_1 minimization (Basis Pursuit).
- $C = \{z \quad s.t. \quad \|A^t z\|_\infty \leq \tau\}$: Dantzig Selector.
- $C = \mathcal{B}_{\ell_2}(0, r)$: LASSO/BPDN equivalent to

$$\min_{x \in \mathbb{R}^N} \frac{1}{2} \|y - Ax\|_2^2 + \gamma \|x\|_1 \quad (\text{LASSO})$$

Noiseless observations : Geometry of centrosymmetric polytopes

Donoho [04], Donoho and Tanner [05-07]

- Identifiability is a geometrical property.



- x is identifiable if and only if $\frac{Ax}{\|x\|_1} \in \partial A(\mathcal{B}_{l_1})$.
- 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 $A(\mathcal{B}_{l_1})$.

The geometrical viewpoint

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

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

$$\|x\|_0 \leq \rho_N(n/N)n \text{ are } \ell_1 - \text{identifiable.} \quad (1)$$

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

$$\|x\|_0 \leq \rho_F(n/N)n \text{ are } \ell_1 - \text{identifiable.} \quad (2)$$

- No stability to noise.
- Sharp phase transition :

$$\rho_N(1/2) \sim 0.089, \quad \rho_F(1/2) \sim 0.38$$

$$\rho_N(1/4) \sim 0.065, \quad \rho_F(1/4) \sim 0.25.$$

Restricted Isometry Property

Definition of RIP

- For $A \in \mathbb{R}^{n \times N}$, δ_S^{\min} and δ_S^{\max} are the smallest real numbers in $(0, 1)$ such that for any x , $\|x\|_0 \leq S$,

$$(1 - \delta_S^{\min})\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_S^{\max})\|x\|_2^2.$$

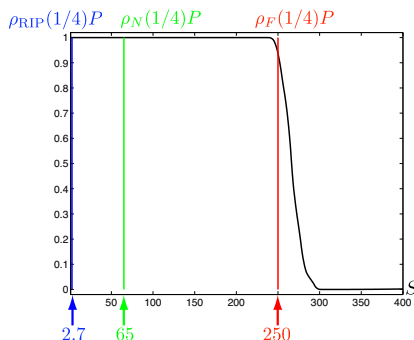
Theorem [Fourcart and Lai 08]

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

- All vectors x such that $\|x\|_0 \leq S$ are identifiable.
- Stability to noise, consistency if x is only compressible.
- There exist C_0 and C_1 depending on δ_{2S}^{\min} and δ_{2S}^{\max} such that the solution x^* of (LASSO) satisfies

$$\|x^* - x_0\|_2 \leq C_0 S^{-1/2} \|x - x_S\|_1 + C_1 \varepsilon.$$

Bounds for gaussian matrices, $N = 4000$, $n = 1000$



- If A is a Gaussian matrix with iid entries, then w.o.p. A satisfies (RIPFL) for $S = O\left(\frac{n}{\log(N/n)}\right)$.
- For $n/N = \frac{1}{4}$, (RIPFL) applies up to $S = 0.0027n$ [Blanchard et al. 09].
- but (RIPFL) doesn't apply if $S \geq 0.005n$. [D. et al 09].

Exact Support and sign pattern recovery with LASSO

Theorem [Candes Plan 07]

Let $A \in \mathcal{M}_{n,N}(\mathbb{R})$ which columns are normed and such that $\mu(A) \leq \frac{c_1}{\ln N}$. Let $w \in \mathbb{R}^n$ such that $w(i) \sim \mathcal{N}(0, \frac{\varepsilon^2}{n})$. Let $x_0 \in \mathbb{R}^N$ and $T = \min_{i \in I} |x_0(i)|$.

For sufficiently small constant c_0

- if x_0 is randomly chosen among vectors such that $|I| \leq \frac{c_0 N}{\|A\|_2^2 \ln N}$, (support : uniform and sign : Rademacher).
- if $T \geq 8\varepsilon \sqrt{\frac{2 \ln N}{n}}$ and $\gamma = 2\varepsilon \sqrt{\frac{2 \ln N}{n}}$

the solution x^* of

$$\min_{x \in \mathbb{R}^N} \frac{1}{2} \|y - Ax\|_2^2 + \gamma \|x\|_1 \quad (\text{LASSO})$$

satisfies $\text{Supp}(x^*) = \text{Supp}(x_0)$ and $\text{sign}(x^*) = \text{sign}(x_0)$ w.o.p.

Contributions

Results for Gaussian matrices

- Refinement of Theorem [Candes Plan 07] for Gaussian matrices
 - without any prior on the distribution of x_0 and w .
 - with explicit and optimal constants.
 - robustness to compressibility.
- Without any hypothesis on $\frac{T}{\varepsilon}$
 - $Supp(x^*)$ is controlled.
 - l_2 consistency results.
- Explicit bounds may be better than the ones derived from RIP.
- Justify debiasing.

Support and sign pattern identification

Theorem 1

Let $(a, b) \in (0, 1)^2$, $N > n$, $y = Ax_0 + w$ where A is a Gaussian matrix and $\|w\|_2 \leq \varepsilon$.

- If $\|x_0\|_0 = S \leq \frac{ab}{2} \frac{n}{\ln N}$,

- if $\gamma = \frac{\varepsilon}{\sqrt{1-a}} \sqrt{\frac{2 \ln N}{n}}$ and if $T \geq \frac{6\varepsilon}{\sqrt{1-a}} \sqrt{\frac{2 \ln N}{n}}$

then w.o.p. $\text{Supp}(x^*) = \text{Supp}(x_0)$ and $\text{sign}(x^*) = \text{sign}(x_0)$ and

$$\|x^* - x_0\|_2 \leq \varepsilon \left(\sqrt{\frac{a}{1-a}} + 1 \right)$$

Support inclusion

Theorem 2

Let $(a, b) \in (0, 1)^2$, $N > n$, $y = Ax_0 + w$ where A is a Gaussian matrix and $\|w\|_2 \leq \varepsilon$.

- If $\|x_0\|_0 = S \leq \frac{ab}{2} \frac{n}{\ln N}$,

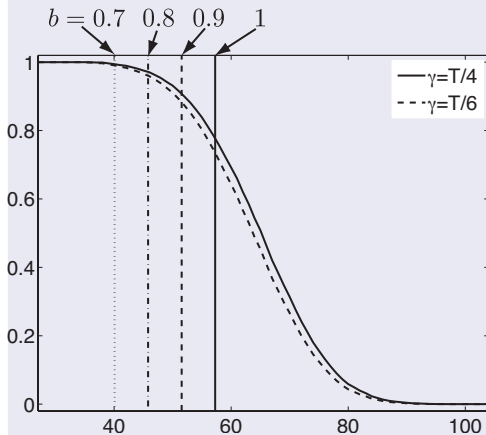
- if $\gamma = \frac{\varepsilon}{\sqrt{1-a}} \sqrt{\frac{2 \ln N}{n}}$,

then w.o.p $\text{Supp}(x^*) \subset \text{Supp}(x_0)$ and

$$\|x^* - x_0\|_2 \leq \varepsilon \left(\sqrt{\frac{a}{1-a}} + 1 \right)$$

Numerical experiments

Example with $a = 0.95$, $n = 1000$ and $N = 4000$



- Rates of exact support recovery versus sparsity level.

- $$\varepsilon = \frac{T}{6} \sqrt{\frac{(1-a)n}{2 \ln N}}.$$

Sketch of proof of Theorem 1

Notations

- For a vector x , let's denote I its support,
- A_I the associated active matrix and \bar{x} the restriction of x to I .
- We have $Ax = A_I\bar{x}$.
- Let's denote $P_{A_I^\perp}$ the orthogonal projection on V^\perp with $V = \text{Span}\{(a_i)_{i \in I}\}$.

Remarks

- If A is gaussian, $\forall (y, \gamma) \in \mathbb{R}^n \times \mathbb{R}^{+*}$, the solution x^* of

$$\min \frac{1}{2} \|y - Ax\|_2^2 + \gamma \|x\|_1 \quad (\text{LASSO})$$

- is unique with probability 1 and
- $(A_I^t A_I)$ associated to x^* is invertible with probability 1.

Sketch of proof of Theorem 1

A necessary condition

- If $\text{Supp}(x^*) = I = \text{Supp}(x_0)$ and $\text{sign}(\overline{x^*}) = \text{sign}(\overline{x_0})$ then the solution of (LASSO) is defined by

$$\overline{x^*} = \overline{x_0} - \gamma(A_I^t A_I)^{-1} \text{sign}(\overline{x_0}) + (A_I^t A_I)^{-1} A_I^t w. \quad (3)$$

A sufficient condition

- Let's denote $T = \min_{i \in I} |x_0(i)|$, if

$$\|\gamma(A_I^t A_I)^{-1} \text{sign}(\overline{x_0}) - (A_I^t A_I)^{-1} A_I^t w\|_\infty < T \quad (\text{SC1})$$

and

$$|\langle a_j, \gamma A_I (A_I^t A_I)^{-1} \text{sign}(\overline{x_0}) - P_{A_I^\perp}(w) \rangle| \leq \gamma, \forall j \notin I \quad (\text{SC2})$$

the vector x^* defined by (3) is the solution of (LASSO).

Sketch of proof of Theorem 1

$$\text{SC1 : } \left\| \gamma (A_I^t A_I)^{-1} \text{sign}(\bar{x}_0) - (A_I^t A_I)^{-1} A_I^t w \right\|_\infty < T$$

- If $|I| \leq \frac{abn}{2 \ln N}$
 - $\left\| (A_I^t A_I)^{-1} \text{sign}(x_0) \right\|_\infty \leq 1 + 3\sqrt{ab} \leq 4$ with w.o.p.
 - Properties of Wishart matrices (signs of coefficients and spectral norm).
 - $\left\| (A_I^t A_I)^{-1} A_I^t w \right\|_\infty \leq 2\varepsilon \sqrt{\frac{\ln N}{n}}$ with w.o.p.
 - Rotation Invariance of $(A_I^t A_I)^{-1} A_I^t$, χ^2 concentration lemmas, and spectral norm of Wishart matrices.
- If $\gamma \leq \frac{T}{6}$ and $\varepsilon \leq \frac{T}{6} \sqrt{\frac{n}{2 \ln N}}$ then condition (SC1) applies :

$$\left\| \gamma (A_I^t A_I)^{-1} \text{sign}(\bar{x}_0) - (A_I^t A_I)^{-1} A_I^t w \right\|_\infty < T$$

Sketch of proof of Theorem 1

$$\text{SC2 : } |\langle a_j, \gamma A_I (A_I^t A_I)^{-1} \text{sign}(\bar{x}_0) - P_{A_I^+}(w) \rangle| \leq \gamma, \forall j \notin I$$

- If u and a_j are independent, then $\langle a_j, u \rangle \sim \mathcal{N}(0, \frac{\|u\|_2}{n})$.
- If $j \notin I$, $u = \gamma A_I (A_I^t A_I)^{-1} \text{sign}(\bar{x}_0) - P_{A_I^+}(w)$ and a_j are independent.
- It follows that w.o.p.

$$\max_{j \notin I} |\langle a_j, u \rangle| \leq \sqrt{\frac{2 \ln N}{n}} \|u\|_2 \quad (4)$$

- $\|u\|_2^2 \leq \gamma^2 \|A_I (A_I^t A_I)^{-1} \text{sign}(\bar{x}_0)\|_2^2 + \varepsilon^2$ using Pythagore !!!
- $\|A_I (A_I^t A_I)^{-1} \text{sign}(\bar{x}_0)\|_2^2$ is bounded using a classical Wishart concentration lemma.
- It follows that if $|I| \leq \frac{abn}{2 \ln N}$ and $\gamma \geq \varepsilon \sqrt{\frac{(1-a)n}{2 \ln N}}$, condition SC2 applies.

Take Away Messages

Conclusions

- Optimal bounds for exact support recovery with Gaussian measurements.
- Partial support recovery if $\frac{T}{\varepsilon}$ is too small.
- New bounds for ℓ_2 recovery different from RIP.
- Robustness to noise and compressibility without RIP.

Going Further

- Extension to subgaussian matrices (USE and Bernoulli)
- [Paper available online.](#)