Computational Complexity versus Statistical Performance for Sparse Recovery Problems

Vincent Roulet

June 5, 2020

Sparse recovery problems

Goal: Recover a signal $x_* \in \mathbb{R}^d$ from n noisy linear observations

$$b_i = a_i^T x_* + \eta_i$$

where η_i are bounded i.i.d. noise

Assumption:
x_* is s-sparse i.e.

$$\|x\|_0 = \mathbf{Card}(\{i \in \{1, \ldots, d\} : x_i \neq 0\}) \leq s \ll d$$

Applications:

- Coding/Decoding audio signals, images, ...
- Find explanatory variables for an experiment

Sparse recovery problems

Decoding procedures:

Given $b = (b_1, \dots, b_n)^T \in \mathbb{R}^n$ and $A = (a_1, \dots, a_n)^T \in \mathbb{R}^{n \times d}$

- In absence of noise ($\eta = 0$), solve

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = b \end{array} \qquad \text{(Exact recovery)}$$

- In presence of noise, solve

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & \|Ax - b\|_2 \le \delta \|A\|_2 \end{array} \qquad \text{(Robust recovery)}$$

  where δ is an estimation of the level of noise and $\|A\|_2$ is the spectral norm of the observation matrix A

Other procedures : Lasso, Dantzig selector,...

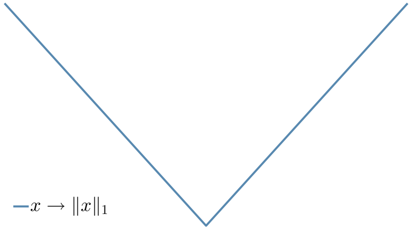
Sparse recovery problems

Resolution of

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = b \end{array} \qquad \text{(Exact recovery)}$$

Assumption: Observations are orthonormal, $AA^T =_n$ so projection on feasible set is available

Problem: $\|x\|_1$ is convex but non smooth.



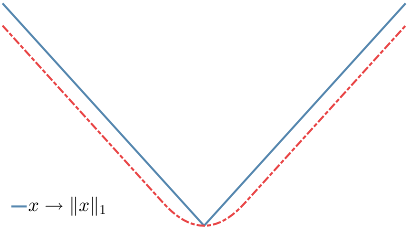
Recovery procedures

Resolution of

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = b \end{array} \qquad \text{(Exact recovery)}$$

Assumption: Observations are orthonormal, $AA^T =_n$ so projection on feasible set is available

Problem: $\|x\|_1$ is convex but non smooth. Yet it can be smoothed !



Recovery procedures

Smoothing technique:

$$\|x\|_1 = \sup_{\|y\|_\infty \leq 1} \langle x, y \rangle \quad \rightarrow \quad f_\varepsilon(x) = \sup_{\|y\|_\infty \leq 1} \langle x, y \rangle - \frac{\varepsilon}{2d}\|y\|_2^2$$

f_ε is

- smooth with constant d/ε
- approximates $\|.\|_1$ uniformly up to $\frac{\varepsilon}{2}$

Optimal algorithm Beck11:
Use accelerated gradient algorithm on f_ε to solve exact recovery at $\varepsilon/2$
Starting from x_0 overall procedure outputs for an accuracy ε after t iterations $\hat{x} = \mathcal{S}(x_0, \varepsilon, t)$ s.t.

$$\|\hat{x}\|_1 - \|x^*\|_1 \leq \frac{2d\|x_0 - x^*\|_2^2}{\varepsilon t} + \frac{\varepsilon}{2}$$

Given an estimate $R \geq \|x_0 - \hat{x}\|$, with appropriate ε,

$$\|\hat{x}\|_1 - \|x^*\|_1 = O(1/t)$$

Sharpness

Idea:
$\|x\|_1$ is sharp so exploit it on set $\{x : Ax = b\}$ by restarts to

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = b \end{array} \tag{Exact recovery}$$

In the following:
For x^* s-sparse such that $Ax^* = b$, assume

$$\|x\|_1 - \|x^*\|_1 > \gamma \|x - x^*\|_1 \tag{Sharp}$$

for any $x \neq x^*$ such that $Ax = b$, and some $0 \leq \gamma < 1$.
Remark: x^* unique minimizer of the decoding procedure

Plan:

- Analyze restart schemes to converge to x^*
- Link sharpness to recovery performance by computing γ

Scheduled restarts

Scheme: Run universal smoothing procedure $\mathcal{S}$, stop it, restart it from last iterate with new target accuracy

Schedule both times t_k of restart and target accuracies ε_k, from $x_0 \in \mathbb{R}^d$,

$$x_k = \mathcal{S}(x_{k-1}, \varepsilon_k, t_k)$$

Scheduled restarts

Optimal schedule *Roul17a

Given $A \in \mathbb{R}^{n \times p}$ and $x^* \in \mathbb{R}^p$, assume there exists $\gamma > 0$

$$\|x\|_1 - \|x^*\|_1 > \gamma \|x - x^*\|_1,$$

for any $x \neq x^*$ such that $Ax = Ax^* = b$
Given $\varepsilon_0 \geq \|x_0\|_1 - \|x^*\|_1$, run scheduled restarts with

$$\varepsilon_k = e^{-1}\varepsilon_{k-1}, \quad t_k = 2e\sqrt{p}/\gamma$$

Then after R restarts and $N = \sum_{i=1}^{R} t_k$ total iterations, it outputs $\hat{x}$ s.t.

$$\|\hat{x}\|_1 - \|x^*\|_1 \leq \exp\left(-\frac{\gamma}{2\sqrt{p}} eN\right) \varepsilon_0.$$

So the optimal complexity is in $O(\gamma^{-1} \log \varepsilon)$ iterations to achieve ε accuracy

Scheduled restarts

In practice γ is unknown

Practical scheme (log-scale grid search):

- Given a fixed budget of iterations N, run for $i \in [1, ..., \lfloor \log_h N \rfloor]$

$$\text{Scheduled restart with } t_k = C_i \qquad \text{where } C_i = h^i$$

- Stop each scheme after N iterations

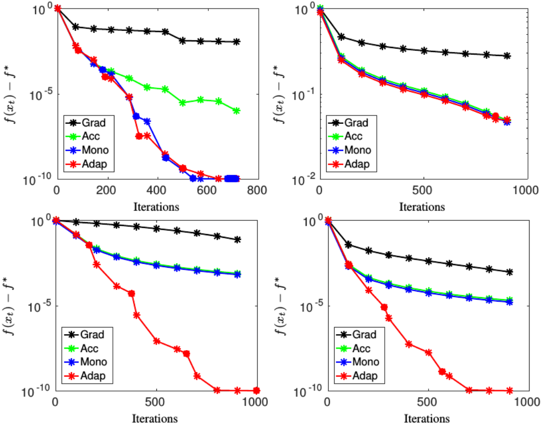
Analysis:

- Needs N big enough for restarts to be relevant
- Cost of the grid search $\log_h(N)$
- One scheme achieves

$$\|\hat{x}\|_1 - \|x^*\|_1 \leq \exp\left(-\frac{\gamma}{2h\sqrt{p}} eN\right) \varepsilon_0.$$

Numerical Illustration

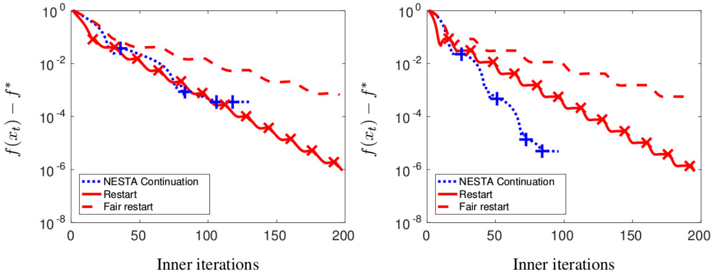
Comparison to plain smoothing technique implementation (called NESTA Beck11)



- Best restarted NESTA (solid red line) for a budget of 500 iterations
- Overall cost of the practical restart schemes (dashed red line)
- Plain NESTA implementation with low accuracy $\varepsilon = 10^{-1}$ (dotted black line) and higher accuracy $\varepsilon = 10^{-3}$ (dash-dotted black line)

Numerical Illustration

Comparison to heuristic of restarts based on stagnancy of objective values (called continuation steps Beck11)



- Best restarted NESTA (solid red line) for a budget of 500 iterations
- Overall cost of the practical restart schemes (dashed red line)
- NESTA with 5 continuation steps (dotted blue line)

Crosses represent the restart occurrences. Left: $n = 200$. Right : $n = 300$.

Recovery performance

Statistical point of view:
Given a s-sparse vector x^*, what are the assumptions on the observation matrix A such that x^* can be recovered by solving

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = b \end{array} \qquad \text{(Exact recovery)}$$

or such that x^* can be approximated by solving

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & \|Ax - b\|_2 \leq \delta \|A\|_2 \end{array} \qquad \text{(Robust recovery)}$$

where $b = Ax^*$

References: Most famous is the Restricted Isometric Property Cand06a, lots exist, here focus on Null Space Property and Minimal Conically Restricted Singular Values

Null Space Property

Null Space Property Cohe09

A matrix A satisfies the Null Space Property (NSP) at order s with constant $\alpha \geq 1$ if for any subset $S \subset \{1, \ldots d\}$, with **Card**$(S) \leq s$, and any $z \in \mathrm{Null}(A) \setminus \{0\}$,

$$\alpha \|z_S\|_1 < \|z_{S^c}\|_1,$$

where $\mathrm{Null}(A)$ is the null space of A, and $z_S \in \mathbb{R}^d$ denotes the vector obtained by zeroing all coefficients of z that are not in S.

Recovery guarantees:
Necessary and sufficient condition for exact recovery of any s-sparse vector

Proof for sufficiency:
Comes from sharpness !

Sharpness and Null Space Property

Null Space Property $\Leftrightarrow$ Sharpness *Roul17a

Given an observation matrix $A \in \mathbb{R}^{n \times p}$ satisfying NSP at order s with constant $\alpha \geq 1$, if the original signal x^* is s-sparse, then for any $x \in \mathbb{R}^p$ satisfying $Ax = Ax^*$, $x \neq x^*$, we have

$$\|x\|_1 - \|x^*\|_1 > \frac{\alpha - 1}{\alpha + 1}\|x - x^*\|_1 \qquad \text{(Sharp)}$$

Conversely, if A is such that (Sharp) is satisfied on every s-sparse vectors, then it satisfies NSP at order s.

Consequences:

- NSP controls optimal convergence rate of restart
- Can use statistical analysis of NSP

Recovery threshold

Recovery threshold

Denote $s_A = 1/\,\mathbf{diam}(B_1^d \cap \mathrm{Null}(A))^2$ where B_1^d is the ℓ_1 unit ball in $\mathbb{R}^d$, then A satisfies (NSP) at any $s < s_A$ with

$$\alpha = 2\sqrt{s_A/s} - 1 > 1$$

Consequence: Optimal rate of convergence of restart scheme in terms of recovery threshold

$$\|\hat{x}\|_1 - \|x^*\|_1 \leq \exp\left(-\left(1 - \sqrt{s/s_A}\right)\frac{e}{2\sqrt{p}}N\right)\varepsilon_0,$$

Oversampling ratio

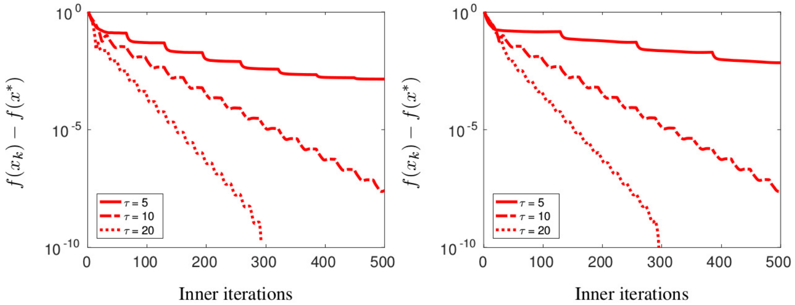
Random observations:
For random observation matrix A, $\textbf{diam}(B_1^d \cap \mathrm{Null}(A)) \leq c\sqrt{\frac{\log d}{n}}$ with high probability, where $c > 0$ is an absolute constant. Therefore

- $s_A \geq n/(c^2 \log d)$
- $O(s \log d)$ observations are sufficient to recover any s-sparse signals

Consequence: Optimal rate of convergence of restart scheme to decode s-sparse signals in terms of oversampling ratio

$$\|\hat{x}\|_1 - \|x^*\|_1 \leq \exp\left(-\left(1 - c\sqrt{\frac{s \log d}{n}}\right)\frac{e}{2\sqrt{p}}N\right)\varepsilon_0$$

Numerical Illustration



Best restart scheme found by grid search for increasing values of the oversampling ratio $\tau = n/s$.
Left : sparsity $s = 20$ fixed. Right : number of samples $n = 200$ fixed.

Robust Recovery

Robust Recovery Performance Chan12

Given a coding matrix $A \in \mathbb{R}^{n \times d}$ and an original s-sparse signal x^*, suppose we observe $b = Ax^* + w$ where $\|w\|_2 \leq \delta \|A\|_2$ and denote $\hat{x}$ an optimal solution of

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & \|Ax - b\|_2 \leq \delta \|A\|_2 \end{array} \qquad \text{(Robust recovery)}$$

If the Minimal Conically Restricted Singular Value (MCRSV)

$$\mu_s(A) = \min_{\substack{S \subset \{1,\ldots,d\} \\ \mathbf{Card}(S) \leq s}} \min_{\substack{\|z_{S^c}\|_1 \leq \|z_S\|_1 \\ \|z\|_2 = 1}} \|Az\|_2$$

is positive, the following error bound holds:

$$\|\hat{x} - x^*\|_2 \leq 2 \frac{\delta \|A\|_2}{\mu_s(A)}.$$

Restricted Singular values and Null Space Property

MCRSV and NSP *Roul17a

Given a matrix $A \in \mathbb{R}^{n \times p}$ and a sparsity level $1 \leq s \leq d$, if the MCRSV $\mu_s(A)$ is positive, then A satisfies NSP at order s for any constant

$$\alpha \leq \left(1 - \frac{\mu_s(A)}{\|A\|_2}\right)^{-1}$$

Consequence:
MCRSV $\mu_s(A)$ controls

- sharpness for exact recovery
- optimal rate of convergence of restart schemes for exact recovery of s-sparse signals

Additional remarks

Link to previous computational analysis
MCRSV matches Renegar condition number that measures complexity of optimality certificates of exact recovery of s-sparse signals

Beyond ℓ_1 norm
Results generalize to other sparse structures, i.e. group norms and low rank matrices. One retrieves

- Sharpness of exact recovery problem
- Generalized Null Space Property
- Generalized Minimal Conically Restricted Singular Values