University of Wrocław: Algorithms for Big Data (Fall'19) 16/12/2019

Lecture 10: Compressed Sensing

Lecturer: *Przemysław Uznański* Scribe: -

1 Overview

A version of sparse recovery: given linear measurements, recover the signal. In general impossible if $m < n$, but is possible if we make additional assumptions.

Definition 1. *Given $A \in \mathbb{R}^{m \times n}$, and $y = Ax \in \mathbb{R}^m$, recover $x \in \mathbb{R}^n$ given that x is k-sparse.*

Rephrasing problem:

$$\begin{aligned} \text{minimize} \quad & \|x'\|_0 \\ \text{subject to} \quad & Ax' = Ax \end{aligned}$$

(this is hard problem, NP-hard?)

The approach: solve the following program

$$\begin{aligned} \text{minimize} \quad & \|x'\|_1 \\ \text{subject to} \quad & Ax' = Ax \end{aligned}$$

The insight is that signals with small $\|x'\|_1$ should be sparse (the weight of x' is concentrated on a few "optimal" coordinates). Particularly, optimizing $\|x'\|_2$ would not work.

(Adding noise: subject to $\|Ax - Ax'\|_2 \le \varepsilon$.)

The program can be expressed as LP, with $n^{\Theta(1)}$ algorithm.

Theorem 2. *Let $m = \Theta(k \log(n/k))$. Pick $A \in \mathbb{R}^{m \times n}$ with entries $\mathcal{N}(0,1)$. W.h.p. the following holds for some C: the output x' of the LP satisfies*

$$\|x' - x\|_2 \le \frac{C}{\sqrt{k}} \min_{\|x''\|_0 \le k} \|x - x''\|_1. \tag{1}$$

In fact any JL-type distribution works.

1.1 Restricted Isometry Property

Definition 3. *We say that matrix A satisfies (k, ε)-RIP if for any k-sparse x there is*

$$\|Ax\|_2 = (1 \pm \varepsilon)\|x\|_2.$$

Theorem 4. *JL matrix with $m = \Theta(k \log(n/k))$ satisfies $(k, 1/3)$-RIP.*

Proof. Recall proof of subspace embeddings. Here as well it is enough to show the property for $\|x\|_2 = 1$. For any $K \subseteq [n]$, $|K| = k$, we take $B_K = \{x : \|x\|_2 = 1 \wedge \text{support}(x) \subseteq K\}$, that is vectors spanned on coordinates in K of length 1.

Pick $(1/7)$-net N_K for B_K of size $2^{\mathcal{O}(k)}$. For any $x \in B_K$ there is $x = \sum_{i\geq 0} \alpha_i x_i$ where $\alpha_i \leq 7^{-i}$ and $x_i \in B_K$, and $\alpha_0 = 1$.

For A properly set (JL with constant $1/7$), there is

$$\|Ax\|_2 = \|A\sum_{i\geq 0} \alpha_i x_i\|_2 \leq \sum_i \alpha_i \|Ax_i\|_2 \leq \sum_i \frac{8}{7} \cdot 7^{-i} \leq \frac{4}{3}$$

and

$$\|Ax\|_2 = \|A\sum_{i\geq 0} \alpha_i x_i\|_2 \geq \|Ax_0\|_2 - \sum_{i\geq 1} \alpha_i \|Ax_i\|_2 \geq \frac{6}{7} - \sum_{i\geq 1} \frac{8}{7} 7^{-i} = \frac{6}{7} - \frac{4}{3} \cdot \frac{1}{7} = \frac{2}{3}$$

How large the m needs to be? It needs to work for union of all 1/7-nets. We have $|\bigcup N_K| \leq \binom{n}{k} \cdot 2^{\mathcal{O}(k)} = 2^{\mathcal{O}(k\log(n/k))}$. □

1.2 Basis pursuit

Theorem 5. *$(4k, 1/3)$-RIP implies* (1).

Proof. Denote $h = x' - x$. Let us reorder x so that $x_1, \ldots, x_k$ are the largest magnitude coefficients, and then so that $h_{k+1}, \ldots$ are sorted (ignoring that $h_1, \ldots, h_k$ might be arbitrary). Denote $A = \{1, \ldots, k\}$ and $A_j = \{k + 5jk + 1, \ldots, 6k + 5jk\}$ for $j \geq 0$.

Denote $\eta = \min_{\|x''\|_0 \leq k} \|x - x''\|_1 = \|x_{A^c}\|_1$.

First, bound:

$$\|x\|_1 \geq \|x'\|_1 = \|x + h\|_1 = \|x_A + h_A\|_1 + \|x_{A^c} + h_{A^c}\|_1 \geq \|x_A\|_1 - \|h_A\|_1 - \|x_{A^c}\|_1 + \|h_{A^c}\|_1$$

so

$$\|h_{A^c}\|_1 \leq \|h_A\|_1 + 2\|x_{A^c}\| = \|h_A\|_1 + 2\eta \tag{2}$$

Then for any vector y with sorted coordinates, $|y_i| \leq \|y\|_1/i$. So

$$\|h_{(AA_0)^c}\|_2^2 \leq \|h_{A^c}\|_1^2 \sum_{j=5k}^{\infty} \frac{1}{j^2} \leq \|h_{A^c}\|_1^2 \frac{1}{\mathcal{O}(k)}$$

so, using (2)

$$\|h_{(AA_0)^c}\|_2 = \mathcal{O}(\frac{\|h_A\|_1 + \eta}{\sqrt{k}}) \tag{3}$$

Then

$$\begin{aligned} 0 = \|A(x' - x)\|_2 = \|Ah\|_2 &\geq \|Ah_{AA_0}\|_2 - \|\sum_{j\geq 1} Ah_{A_j}\|_2 \\ &\geq \|Ah_{AA_0}\|_2 - \sum_{j\geq 1} \|Ah_{A_j}\|_2 \\ &\geq \frac{2}{3}\|h_{AA_0}\|_2 - \frac{4}{3}\sum_{j\geq 1} \|h_{A_j}\|_2 \end{aligned}$$

Each term in $h_{A_{j+1}}$ is smaller than average term in h_{A_j}, so $\|h_{A_{j+1}}\|_2^2 \le (5k)\frac{\|h_{A_j}\|_1^2}{(5k)^2} = \frac{\|h_{A_j}\|_1^2}{5k}$. So

$$\sum_{j\ge 1}\|h_{A_j}\|_2 \le \sum_{j\ge 0}\frac{\|h_{A_j}\|_1}{\sqrt{5k}} = \frac{\|h_{A^c}\|_1}{\sqrt{5k}} = \frac{\|h_A\|_1 + 2\eta}{\sqrt{5k}} \le \sqrt{\frac{k}{5k}}\|h_A\|_2 + \frac{2}{\sqrt{5k}}\eta \tag{4}$$

Combining we have

$$\frac{2}{3}\|h_{AA_0}\|_2 \le \frac{4}{3\sqrt{5}}\|h_A\|_2 + \mathcal{O}(\frac{\eta}{\sqrt{k}})$$

so since $\|h_A\|_2 \le \|h_{AA_0}\|_2$ and $\frac{2}{3} < \frac{4}{3\sqrt{5}}$, there is

$$\|h_{AA_0}\|_2 = \mathcal{O}(\frac{\eta}{\sqrt{k}})$$

and this bound also applies in the form of

$$\frac{\|h_A\|_1}{\sqrt{k}} = \mathcal{O}(\frac{\eta}{\sqrt{k}}).$$

By triangle inequality

$$\|h\|_2 \le \|h_{AA_0}\|_2 + \|h_{(AA_0)^c}\|_2 = \mathcal{O}(\frac{\eta}{\sqrt{k}})$$

□

1.3 Iterative Hard Thresholding

Assume $y = Ax + e$, where e is some noise, and x is k-sparse.

The algorithm works as follow.

- $x^0 = 0$.
- $a^{i+1} = x^i + A^T(y - Ax^i)$
- $x^{i+1} = H_k[a^{i+1}]$ where H_k is projection onto top-k coefficients.

We expect x^t to converge onto $x \pm \mathcal{O}(e)$, in logarithmic number of iterations. Proof:
denote $r^i = x - x^i$. Our goal is to show that r^i decreases by a constant factor at each step.
Denote $S = \text{support}(x)$ and $S^i = \text{support}(x^i)$, and $B^i = S \cup S^i$.
We assume we have $(3k, \varepsilon)$-RIP for sufficiently small ε (like, small constant).
Properties: (A_S denotes matrix with only columns in S kept, and all others zeroed out).

Lemma 6. *For S such that $|S| \le 3k$:*

$$\|(I - A_S^T A_S)x_S\|_2 \le \varepsilon\|x_S\|_2$$

Proof. Follows from A_S being subspace embedding for $3k$-sparse vectors. □

Lemma 7. *for any disjoint S, S' with $|S| + |S'| \le 3k$*

$$\|A_S^T A_{S'} x_{S'}\|_2 \le \varepsilon\|x_{S'}\|_2$$

Proof.

$$\|A_S^T A_{S'}\|_2 = \sup_{\|y\|=1} \|A_S^T A_{S'} y\|_2 = \sup_{\|x\|=1,\|y\|=1} x^T A_S^T A_{S'} y = \sup_{\|x\|=1,\|y\|=1} \langle Ax_S, Ay_{S'}\rangle$$

$$= \sup_{\|x\|=1,\|y\|=1} (\langle x_S, y_{S'}\rangle \pm \varepsilon) \leq \varepsilon$$

where we used that $\langle x_S, y_{S'}\rangle = 0$ and that linear combinations of $x_S, y_{S'}$ are also sparse. □

Lemma 8.

$$\|A_S^T y\|_2^2 \leq (1+\varepsilon)\|y\|_2^2$$

Proof. Follows from $\|A_S^T\|_2 = \|A_S\|_2 \leq (1+\varepsilon)$, which follows from:

$$\|A_S\|_2 = \sum_{\|x\|_2=1} \|A_S x_S\|_2 \leq \sqrt{1+\varepsilon}\|x_S\| = \sqrt{1+\varepsilon}$$

□

Now we bound $\|r^{i+1}\|_2$:

$$\|r^{i+1}\|_2 = \|x - x^{i+1}\|_2 = \|x_{B^{i+1}} - x^{i+1}_{B^{i+1}}\|_2 \leq \|x_{B^{i+1}} - a^{i+1}_{B^{i+1}}\|_2 + \|x^{i+1}_{B^{i+1}} - a^{i+1}_{B^{i+1}}\|_2$$

since $x^{i+1}_{B^{i+1}}$ is the best possible k-sparse approximation to $a^{i+1}_{B^{i+1}}$, and $x_{B^{i+1}}$ is k-sparse, we have

$$\|r^{i+1}\|_2 \leq 2\|x_{B^{i+1}} - a^{i+1}_{B^{i+1}}\|_2$$

$$\leq 2\|x_{B^{i+1}} - x^i_{B^{i+1}} - A^T_{B^{i+1}} A r^i - A^T_{B^{i+1}} e\|_2$$
$$\leq 2\|r^i_{B^{i+1}} - A^T_{B^{i+1}} A r^i\|_2 + 2\|A^T_{B^{i+1}} e\|_2$$
$$\leq 2\|(I - A^T_{B^{i+1}} A_{B^{i+1}}) r^i_{B^{i+1}}\|_2 + 2\|A^T_{B^{i+1}} A_{B^i - B^{i+1}} r^i_{B^i - B^{i+1}}\|_2 + 2\sqrt{1+\varepsilon}\|e\|_2$$
$$\leq 2\varepsilon\|r^i_{B^{i+1}}\|_2 + 2\varepsilon\|r^i_{B^i - B^{i+1}}\|_2 + 2\sqrt{1+\varepsilon}\|e\|_2$$
$$\leq 4\varepsilon\|r^i\|_2 + 2\sqrt{1+\varepsilon}\|e\|_2$$

set $\varepsilon = 1/8$

$$\leq \frac{1}{2}\|r^i\|_2 + 3\|e\|_2$$