COMS E6998-9: Algorithms for Massive Data (Fall'23) Oct 9, 2023

Lecture 10

Instructor: *Alex Andoni* Scribes: *Hantao Yu*

Recall our Iterative Hard Thresholding algorithm: given input $y = A \cdot x \in \mathbb{R}^m, x \in \mathbb{R}^n$,

- Set $x^1 = (0, \ldots, 0)$,
- For $t = 1, \ldots, T$, set $x^{t+1} = P_k(x^t + A^\top(y - Ax^t))$,
- Return x^{T+1},

where $P_k(x) := \arg\min_{k-\text{sparse} x'} \|x' - x\|$.

Theorem 1. *Let $x \in \mathbb{R}^n$ be k-sparse and A is $(3k, \varepsilon = \frac{1}{8})$-RIP, then*

$$\|x^{T+1} - x\|_2 \leq 2^{-T}\|x\|_2,$$

where x^{T+1} is the output of the Iterative Hard Thresholding algorithm.

Proof. First observe that $\|Ax\|_2^2 \in (1 \pm \varepsilon)\|x\|_2^2$ because A is $(3k, \varepsilon = \frac{1}{8})$-RIP. Since $\|Ax\|_2^2 = x^\top A^\top Ax$ and $\|x\|_2^2 = x^\top Ix^\top$, we have $\|Ax\|_2^2 - \|x\|_2^2 \in \pm\varepsilon\|x\|_2^2$, which implies $x^\top(A^\top A - I)x \in \pm\varepsilon\|x\|_2^2$.

We define $r^t := x - x^t, a^{t+1} := x^t + A^\top(y - Ax^t)$. We will show that $\|r^{t+1}\|_2 \leq \frac{1}{2}\|r^t\|_2$, which suffices to prove our theorem. The intuition of the proof is that we can write

$$a^{t+1} = x^t + A^\top(y - Ax^t) = x^t + A^\top A(x - x^t)$$

and this is an approximation of x because $A^\top A \approx I$.

Let $B^t := \text{supp}(x) \cup \text{supp}(x^t) \supseteq \text{supp}(r^t)$ $(|B^t| \leq 2k)$. Denote $B := B^{t+1}, B^- = B^t$. We now have

$$\begin{aligned}\|r^{t+1}\|_2 &= \|x - x^{t+1}\|_2 \\ &= \|x_B - x_B^{t+1}\|_2 \\ &\leq \|x_B - a_B^{t+1}\|_2 + \|a_B^{t+1} - x_B^{t+1}\|_2.\end{aligned}$$

Since x^{t+1} is the best k-sparse approximation to a^{t+1}, $\|x_B^{t+1} - a_B^{t+1}\|_2 \leq \|x_B - a_B^{t+1}\|_2$, which implies that $\|r^{t+1}\|_2 \leq 2\|x_B - a_B^{t+1}\|_2$. Let $A_B = A$ with columns not in B zeroed out, we have

$$a_B^{t+1} = x_B^t + (A_B)^\top A(x - x^t),$$

which implies (let $-B := B^t - B$)

$$\begin{aligned}\|x_B - a_B^{t+1}\|_2 &= \|x_B - x_B^t - A_B^\top A r^t\|_2 \\ &= \|r_B^t - A_B^\top A r^t\|_2 \\ &= \|r_B^t - A_B^\top A r_B^t - A_B^\top A r_{-B}^t\|_2 \\ &\le \|(I - A_B^\top A) r_B^t\|_2 + \|A_B^\top A r_{-B}^t\|_2 \\ &\le \varepsilon \|r_B^t\|_2 + \|A_B^\top A r_{-B}^t\|_2,\end{aligned}$$

where the last inequality follows from the following observation (using the RIPness of A):

$$\|(I - A_B^\top A) r_B^t\|_2 = \|(I_B - A_B^\top A_B) r_B^t\|_2 \le \max_{u,v:\|u\|=\|v\|=1} u^\top (I_B - A_B^\top A_B) v \cdot \|r_B^t\|_2 \le \epsilon \|r_B^t\|_2.$$

Next, we claim that $\|A_B^\top A_{-B}\|_2 \le 2\varepsilon$. This is because

$$\|A_B^\top A_{-B}\|_2 \le \max_{\substack{\|p\|=\|q\|=1 \\ \text{supp}(p)\subseteq B \\ \text{supp}(q)\subseteq -B}} p^\top A_B^\top A_{-B} q.$$

By RIP of A, we have $\|A(p-q)\|_2^2 \ge 2(1-\varepsilon)$. As a result,

$$\begin{aligned}2(1-\varepsilon) &\le (p-q)^\top A^\top A (p-q) \\ &= \|Ap\|_2^2 + \|Aq\|_2^2 - 2p^\top A^\top A q \\ &\le (1+\varepsilon) + (1+\varepsilon) - 2p^\top A^\top A q,\end{aligned}$$

which implies $p^\top A^\top A q \le 2\varepsilon$.

Finally, we can upper bound

$$\begin{aligned}\|r^{t+1}\|_2 &\le 2\|x_B - a_B^{t+1}\|_2 \\ &\le 2(\varepsilon \|r_B^t\|_2 + 2\varepsilon \|r^t\|_2) \\ &\le 6\varepsilon \|r^t\|_2 \qquad \text{assuming } \varepsilon \le \frac{1}{12} \\ &\le \frac{1}{2}\|r^t\|_2.\end{aligned}$$

An induction argument will finish the proof of $\|r^{t+1}\|_2 \le 2^{-t}\|r^1\|_2 = 2^{-t}\|x\|_2$.

□