COMS E6998-15: Algorithms for Massive Data (Fall'2023) Sep 6, 2023

Lecture 9: Compressed Sensing

Instructor: *Alex Andoni* Scribes: *Yasaman Mahdaviyeh*

1 Overview

Recall that in compressed sensing we have a vector $x \in \mathbb{R}^n$ that is approximately (at most) k sparse, in the sense that the error

$$\mathrm{Err}_1^k(x) \triangleq \min_{x' \text{ is k-sparse}} \left\|x - x'\right\|_1$$

is small. The goal is to design an $m \times n$ matrix of measurements A, where $m < n$ such that given A and $y = Ax$, we can recover a vector x' such that

$$\left\|x - x'\right\|_1 \leq c\, \mathrm{Err}_1^k(x), \tag{1}$$

for some small constant c. We considered the following L_1 algorithm.

L_1 algorithm:

$$\begin{aligned} \text{Minimize} \quad & \|x^*\|_1 \\ \text{such that} \quad & Ax^* = y \end{aligned}$$

Note that the algorithm above is a linear program and can be solved in polynomial time in n. We define the following property of matrix A that would be sufficient to get guarantees of the type eq. (1) for this algorithm.

Definition 1 (Restricted Isometry Property)**.** *We say that the matrix A is (k, ε)-RIP if for all k-sparse x*

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

2 Guarantee on L_1 algorithm

The main focus of the class today is to prove the following theorem. Let x^* be a solution to L_1 algorithm.

Theorem 2 (L_1 theorem)**.** *If A is $(4k, \varepsilon)$-RIP then*

$$\|x^* - x\|_1 \leq \mathcal{O}(1)\ \mathit{Err}_1^k(x).$$

Before we prove the theorem above we will show that there are matrices that satisfy RIP.

Theorem 3. *If A is an OSE for dimension k with probability of at least $1-\delta$ where $\delta \leq \frac{0.1}{\binom{n}{k}}$, then with probability of at least 0.9, A is (k,ε)-RIP.*

Proof. On a high level, since the set of vectors supported on a fixed set of coordinates of size at most k form a subspace of dimension k and there are at most $\binom{n}{k}$ such subspaces, we can apply OSE to each subspace and then do a union bound.

Fix any set $U \subset [n]$ of size k, by OSE,

$$\Pr\left[\forall x,\ \mathrm{supp}(x) \subseteq U, 1-\varepsilon \leq \frac{\|Ax\|_2}{\|x\|_2} \leq 1+\varepsilon\right] \geq 1-\delta.$$

By union bound over all $U \subset [n]$ of size k,

$$\Pr\left[\forall x,\ x \text{ is k-sparse}, 1-\varepsilon \leq \frac{\|Ax\|_2}{\|x\|_2} \leq 1+\varepsilon\right] \geq 1-\binom{n}{k}\cdot\delta \geq 0.9.$$

$\square$

Corollary 4. *Existence of OSE (see HW2 problem 1) with dimension $m = \mathcal{O}(k+\log(1/\delta))$ implies that there exists A that is (k,ε)-RIP with $m = \mathcal{O}(k+\log\binom{n}{k}) = \mathcal{O}(k\log(n/k))$.*

Normally we would upper bound $\log\binom{n}{k}$ by $\mathcal{O}(k\log n)$, but here we have written the tighter bound above because in a lot of the applications of compressed sensing, k can be cn for a small constant c, so the difference between $\log n/k$ and $\log n$ becomes significant.

It is known that if our goal is to achieve eq. (1) the bound above is tight. But if the original x is k-sparse and our goal is to recover it exactly, the bound above for m is not tight, as we saw in the last question of HW1.

Note that RIP itself is deterministic, while OSE is a random map. So we could take an OSE and check whether it is actually RIP, if it is not, then we can sample again, and we know that we won't have to resample too many times. One issue is that it is hard to check whether a matrix is RIP (can't be done in polynomial time). Another option is to hand design a deterministic matrix A and know for sure that it is RIP.

Now we prove the L_1 Theorem.

Proof of theorem 2 The proof will rely on the following intermediate property.

Definition 5. *We say that A satisfies (k,ε)-nullspace property if for all $T \subset [n], |T| \leq k$, and for all $\eta \in \mathbb{R}^n$, if $A\eta = 0$ then $\|\eta\|_1 \leq (1+\varepsilon)\|\eta_{-T}\|_1$ or equivalently $\|\eta_T\|_1 \leq \varepsilon\|\eta_{-T}\|_1$. Here we have used η_T to denote the restriction of η to coordinates that are in T, and η_{-T} to denote the restriction to the rest of the coordinates.*

Proof of the theorem would then follow from the two following Lemmas which we will prove next.

Lemma 6. *If A is $((2+r)k, \varepsilon$-RIP then A satisfies $(2k, \sqrt{\frac{2}{r}}\cdot\frac{1+\varepsilon}{1-\varepsilon})$-nullspace property.*

Lemma 7. *If A satisfies $(2k,\varepsilon)$-nullspace property with $\varepsilon < 1/2$, then $\|x^*-x\|_1 \leq 2\cdot\frac{1+\varepsilon}{1-\varepsilon}\cdot Err_1^k(x)$.*

We will first prove the lemma above.

Proof of lemma 7 Define the error vector $\eta \triangleq x^* - x$. Since x^* is a solution to $L1$ algorithm, we must have $Ax^* = y$ and consequently $A\eta = Ax^* - Ax = 0$. Set T to be the largest k coordinates of x so that $\text{support}(x') \subseteq T$ (where x' is a k-sparse vector that realizes $\text{Err}_1^k(x)$). By definition 5 we have

$$\|\eta_T\|_1 \leq \varepsilon \|\eta_{-T}\|_1 . \tag{2}$$

Since x^* a solution to L_1 algorithm we also have $\|x^*\|_1 \leq \|x\|_1$, or equivalently

$$\|x_T^*\|_1 + \left\|x_{-T}^*\right\|_1 \leq \|x_T\|_1 + \|x_{-T}\|_1 \tag{3}$$

By triangle inequality we have

$$\|x_T^*\|_1 = \|x_T^* - x_T + x_T\|_1 \geq \|x_T\|_1 - \|\eta_T\|_1 ,$$
and
$$\left\|x_{-T}^*\right\|_1 \geq \|\eta_{-T}\|_1 - \|x_{-T}\|_1 .$$

Plugging these lower bounds in eq. (3) we get

$$\|x_{-T}\|_1 \geq \|\eta_{-T}\|_1 - \|\eta_T\|_1 - \|x_{-T}\|_1 ,$$

which means that

$$\|\eta_{-T}\|_1 \leq 2\|x_{-T}\|_1 + \|\eta_T\|_1 \leq 2\text{Err}_1^k(x) + \varepsilon \cdot \|\eta_{-T}\|_1 ,$$

where we have applied eq. (2) and used the fact that $\|x_{-T}\|_1 = \text{Err}_1^k(x)$. Rearranging the terms in equation above we get

$$\|\eta_{-T}\|_1 \leq \frac{2}{1-\varepsilon} \text{Err}_1^k(x).$$

By the nullspace property we get

$$\|\eta\|_1 \leq 2 \cdot \frac{1+\varepsilon}{1-\varepsilon} \text{Err}_1^k(x).$$

□

Proof of lemma 6 Define $M \triangleq rk$, and let $T_0 = T$, Now let T_1 be the next largest M coordinates of η_{-T}, and so on up to T_s, which is the last at most M coordinates. Define $\eta_0 = \eta_{T_0} + \eta_T$ and $\eta_i = \eta_{T_{i+1}}$. Then we can write $\eta = \sum_{i\geq 0} \eta_i$. We have

$$\|\eta_T\|_2 \leq \|\eta_0\|_2 \leq \frac{\|A\eta_0\|_2}{1-\varepsilon}.$$

Since $A\eta = 0$, we can write $A\eta_0 = -A\eta_1 - A\eta_2 - \cdots - A\eta_{s-1}$. Combining this with equation above we get

$$\|\eta_T\|_2 \leq \frac{\|A\eta_1\|_2 + \|A\eta_2\|_2 + \cdots}{1-\varepsilon}.$$

Let $\eta^{(j)}$ denote the jth coordinate of η. For all $j \geq 1$, and all $j \in T_{j+1}$, $|\eta^{(i)}| \leq \frac{\|\eta_{j-1}\|_1}{M}$, which implies that

$$\|\eta_{j+1}\|_2 \leq \left(M \cdot (\frac{\|\eta_{j-1}\|_1}{M})^2\right)^{1/2} = \frac{\|\eta_{j-1}\|_1}{\sqrt{M}}$$

This implies that

$$\|\eta_T\|_2 \leq \frac{1+\varepsilon}{1-\varepsilon}\sum_{j\geq 1}\|\eta_j\|_2 \leq \frac{1+\varepsilon}{1-\varepsilon}\sum_{j\geq 1}\frac{\|\eta_{T_j}\|_1}{\sqrt{M}} = \frac{1+\varepsilon}{1-\varepsilon}\cdot\frac{1}{\sqrt{M}}\cdot\|\eta_{-T}\|_1.$$

We have

$$\|\eta_T\|_1 \leq \sqrt{2k}\,\|\eta_T\|_2 \leq \frac{1+\varepsilon}{1-\varepsilon}\cdot\frac{\sqrt{2k}}{\sqrt{rk}}\cdot\|\eta_{-T}\|_2 = \frac{1+\varepsilon}{1-\varepsilon}\cdot\sqrt{\frac{2}{r}}\cdot\|\eta_{-T}\|_2$$

□

□

The L_1 algorithm runs in polynomial time since it is a linear program. Can we get a solution with similar guarantees in $n^{\mathcal{O}(1)}$?

3 Iterative Hard Thresholding

In the next lecture we will go over Iterative Hard Thresholding algorithm and see that we can get similar guarantees with less run time. Here we give a brief overview of the algorithm. For a vector z, define projection $P_k(z) \triangleq \arg\min_{z' \text{ is k-sparse}} \|z - z'\|_1$, that is, projection into the top k coordinates of z.

Hard Thresholding Algorithm:

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

Theorem 8 (Blumensath, Davies' 09)**.** *If A is $(3k,\varepsilon)$-RIP, where $\varepsilon < 1/8$, suppose that $y = Ax + e$,* *where e is some error, then x^{T+1} satisfies*

$$\left\|x^{T+1} - x\right\|_2 \leq \mathcal{O}(1) \left[2^{-T} \left\|x\right\|_2 + \frac{Err_1^k(x)}{\sqrt{k}} + \left\|e\right\|_2\right]$$

Note the guarantee above, which is called L_2, L_1 guarantee is stronger than the guarantees of the type in eq. (1), and in fact implies a guarantee of type eq. (1). We will prove a slightly weaker theorem in the next class.