CS 395T: Sublinear Algorithms Fall 2016

Lecture 18 — 27th October, 2016

Prof. Eric Price *Scribe: Garrett Goble, Yanyao Shen*

1 Overview

Previously, we have shown that non-adaptive sparse recover requires $O(k \log \frac{n}{k})$ rows. Today, we discuss the problem in the adaptive setting, and show lower bounds for both deterministic and randomized setting.

Adaptive sparse recovery is done by picking a_i in a sequential manner. We first pick a_1, get $\langle a_1, x\rangle$, then select a_2, get $\langle a_2, x\rangle$, continue until we select a_m and get $\langle a_m, x\rangle$, we use these m measurements to give an estimate for x. A natural question to ask is, does this sequential procedure give us more power than ordinary method?

In deterministic setting, we can not do better than $\Omega(k \log \frac{n}{k})$, and for non-adaptive setting, we also need $O(k \log \frac{n}{k})$ rows. In randomized setting, a naive lower bound is $\Omega(k)$, since you need it to even knowing the whole support set. On the other hand, it is possible with $m = O(k \log \log \frac{n}{k})$, and in today's lecture, we will show that algorithm and as well as a lower bound $\Omega(k + \log \log n)$ for m, which is tight when the sparsity $k = O(1)$. In the homework, we will show that adaptive compressed sensing using Fourier matrix has a lower bound $\Omega(k \frac{\log(n/k)}{\log \log n})$. All the above results are shown in the setting:

$$\|\hat{x} - x\|_p \leq (1+\epsilon) \min_{k-\text{sparse } x'} \left\|x - x'\right\|_p$$

for $p = 2$, i.e., l_2/l_2 recovery.

2 Deterministic Setting

Let us take $k = 1$, and suppose $x = e_i + w$ for $w \sim \mathcal{N}(0, \frac{1}{100n} I_n)$ (easy to show that $\|w\|_2 \approx \frac{1}{10}$). In this case, l_2/l_2 recovery is equivalent to identifying $i \in [n]$. Based on this observation, we can use the Shannon-Hartley Theorem to find the lower bound.

Theorem 1 (Shannon-Hartley Theorem)**.**

$$I(x + w; x) \leq \frac{1}{2} \log(1 + \frac{\mathbb{E}[|x|^2]}{\mathbb{E}[|w|^2]})$$

Take any $a \in \mathbb{R}^n$, then $\langle a, x\rangle = a_i + \langle a, w\rangle = a_i + \mathcal{N}(0,1) \cdot \frac{\|a\|_2}{10\sqrt{n}}$. Intuitively, we can think of the measurement as a sample point in a n-mixture of gaussian distribution, since n is large, it is difficult to identify i. More formally, we will show that $\log n$ for m is required, hence getting the

lower bound for deterministic setting.

$$\begin{aligned} I(\langle a, x\rangle; i) =& I(\langle a, x\rangle; a_i) \\ \leq& \frac{1}{2}\log(1 + \frac{\mathbb{E}[a_i^2]}{\|a\|_2^2/(100n)}) \\ =& \frac{1}{2}\log 101 = O(1). \end{aligned}$$

In the above equation the expectation on a_i^2 is due to the randomness of x, not a. Therefore, $\mathbb{E}[a_i^2] = \frac{\|a\|_2^2}{n}$. This implies that with a single measurement, only constant bits of information is possible. Since identifying i needs at least $\log n$ bits, $m = O(\log n)$.

Rethink moment estimation lower bound. We used similar method in moment estimation lower bound. There, we were in the setting of knowing $b_i = c_i + d_i$, where d_i are independent of c_i and each other, and we bound the mutual information:

$$I(b; c) \leq \sum I(b_i; c_i),$$

where the left hand side is analogous to $I(Ax; i)$ in our context, and the right hand side is bounded by $mO(1)$. For more reference and details on applications of the Shannon-Hartley Theorem to data streams and sparse recovery, see [PW12].

3 Random Setting - 1-sparse

3.1 1-sparse lower bound - $x = e_i + w$

We wonder how big could $\mathbb{E}[a_i^2]$ be in any given round (assume $\|a\|_2^2 = 1$). Since in a non-deterministic setting, a_i can be dependent with measurements from previous rounds and calculating this quantity is not clear for now, we first consider simple cases, e.g., when we are at the first and the last round.

- First round: The calculation of $\mathbb{E}[a_i^2]$ is consistent with the deterministic setting, i.e., $\mathbb{E}[a_i^2] = \frac{\|a\|_2^2}{n}$, and the information gain is constant bits.
- Last round: Since we know enough information about i, $\mathbb{E}[a_i^2] \leq \|a\|_2^2 = 1$, and the information gain is bounded by $\frac{1}{2}\log(1 + 100n) = O(\log n)$.

Our guess for **round between** is that: Suppose at some certain round, we know B bits about i. Then, we should be able to have $\mathbb{E}[a_i^2] \approx \frac{2^B}{n}$ and accordingly, the information we get is bounded by $\frac{1}{2}\log(1 + 100 \cdot 2^B) = O(B)$. Typically, the bits of information we get after each round grows like this: $B \to (1+c)B \to (1+c)^2B \to \cdots$. Since the number of bits grows exponentially, then we would expect $O(\log\log n)$ rounds to identify i. A formal proof of this is shown in [PW13], section 3.

3.2 algorithm when $x = e_i + w$

Consider finding such an algorithm to match the lower bound, our goal is the following: given that we know i to $\frac{n}{2^B}$ size region, we need to find a $\frac{n}{2^{2B}}$ size region in $O(1)$ measurements.

The general idea of this algorithm is the following:

- Round 1: pick $a_i = \begin{cases} -1 & \text{if } i < \frac{n}{2} \\ +1 & \text{if } i \geq \frac{n}{2} \end{cases}$. For example, $\langle a, x \rangle > 0 \implies i \in [\frac{n}{2}, n]$ probably.
- After narrowing the possible region of index i, we can decrease the gap when selecting a_j, since the total noise for each measurement decreases. For example, when narrowing to a $\frac{n}{4}$ area, we can decrease the gap to 1 from 2.
- If we know $i \in R$ is in some region of size $|R| = \frac{n}{2^B}$, set $a_j \in [-1, 1]$ for $j \in R$, and 0 elsewhere. ($\mathbb{E}[a_j^2] \approx 1$ and $\mathbb{E}[\langle a, w \rangle^2] = \frac{\|a\|_2^2}{100n} \approx \frac{1}{2^B}$.)

3.3 1-sparse in general

In general, we may not get e_i as the 1-sparse signal, and may not get Gaussian distribution as noise. More specifically, the 1-sparse signal is:

$$x = \alpha a_i + w,$$

for some w satisfying $\|w\|_2 \leq \frac{\alpha}{R}$. To tackle this, we want to perform linear measurements, and find a small region of size $\frac{n}{R^{0.1}}$ where i must land.

In a noiseless setting, this is simplified to the problem of recovering the index of the unique element knowing that $\|x\|_0 = 1$, where our solution is:

- Pick $v = (1, \cdots, 1)$ and $v' = (1, 2, \cdots, n)$.
- Calculate $y_1 = v \cdot x$, and $y_2 = v' \cdot x$.
- Output y_2 / y_1.

Using this framework, consider if this still works in the noisy version, where $\|w\|_1 \leq \frac{\alpha}{R}$. Now,

$$y_1 = \alpha(1 \pm \frac{1}{R}),$$

and

$$y_2 = i\alpha \pm \frac{n\alpha}{R}.$$

Therefore,

$$\frac{y_2}{y_1} = i(\frac{1}{1 \pm \frac{1}{R}}) \pm \frac{n}{R(1 - \frac{1}{R})} = i \pm O(\frac{n}{R}).$$

Hence, the method still works with noise. Instead of directly calculating y_1 and y_2, we add randomness to each entry in y_1 and y_2 to cancel out the noise.

We can identify that B grows exponentially in the algorithm, and hence $O(\log \log n)$ rounds is needed.

4 k-sparse Recovery - An Algorithm with $O(k \log\log(n/k))$

In the k-sparse setting, we use the algorithm in 1-sparse recovery problem as a black box.

- Repeatedly do the following:
  - subsample $\frac{1}{k}$ fraction of coordinates
  - do 1-sparse recovery

Each round: $O(\log\log n)$ measurements finds 1 heavy hitter with $\geq \frac{3}{4}$ probability.

After k rounds, we will have found $\geq \frac{k}{2}$ heavy hitters with $\frac{3}{4}$ probability. With $O(k \log\log \frac{n}{k})$ measurements, we can find $\frac{1}{2}$ of the heavy hitters. Then, repeat on rest of the coordinates.

Repeat with $k \to \frac{k}{2} \to \frac{k}{4} \to \frac{k}{8}$. This process will have

$$k \log\log \frac{n}{k} + \frac{k}{2} \log\log \frac{n}{\frac{k}{2}} + \frac{k}{4} \log\log \frac{n}{\frac{k}{4}} + \cdots = O(k \log\log \frac{n}{k})$$

measurements combined [IPW11].

References

[PW12] Eric Price, David P. Woordruff. Applications of the Shannon-Hartley Theorem to Data Streams and Sparse Recovery. *ISIT*, 2012.

[PW13] Eric Price, David P. Woordruff. Lower Bounds for Adaptive Sparse Recovery *SODA*, 2013.

[IPW11] Piotr Indyk, Eric Price, David P. Woordruff. On the Power of Adaptivity in Sparse Recovery *FOCS*, 2011.