CS 395T: Sublinear Algorithms	Fall 2014

Lecture 13 — Oct. 9, 2014

Prof. Eric Price *Scribe: Udit Agarwal*

In today's lecture, we will cover the following topics:

1. l_2/l_1 upper bounds [CRT06]
2. l_2/l_2 lower bounds [CDD09] :- To perform l_2/l_2 recovery deterministically, at least $\Omega(n)$ samples are required.
3. l_1/l_1 lower bounds [DIPW10] :- To perform l_1/l_1 recovery deterministically or with a randomized algorithm, at least $\Omega(k \log \frac{n}{k})$ samples are required.

Consider the problem of stable sparse recovery: given a matrix $A \in \mathbb{R}^{m \times n}$ and a k-sparse vector x and given $y = Ax + e$, with e as the error term, we wish to recover $\hat{x}$ such that

$$||\hat{x} - x||_2 \leq C||e||_2$$

.

Ques: What about if x is not k-sparse??

Then the problem becomes: given Ax, $\forall x \in \mathbb{R}^n$, recover $\hat{x}$ such that

$$||\hat{x} - x||_p \leq C \min_{k\text{-sparse } x'} ||x - x'||_q$$

for some norm parameters p and q and an approximation factor C.

Thus, the error term depends only on the top k terms of x. Some of the l_p/l_q recovery guarantees are as follows:

- l_2/l_2 :
$$||\hat{x} - x||_2 \leq C||x - x_k||_2$$
- l_2/l_1 :
$$||\hat{x} - x||_2 \leq \frac{C}{\sqrt{k}}||x - x_k||_1$$
- l_1/l_1 :
$$||\hat{x} - x||_1 \leq C||x - x_k||_1$$

where x_k contains the top k terms of x.

Now, we'll talk about the bounds on the number of samples required to perform each of these l_p/l_q guarantees deterministically:

- l_2/l_2 : To perform l_2/l_2 recovery deterministically, at least $\Omega(n)$ samples are required.

- l_2/l_1 : To perform l_2/l_1 recovery deterministically, at least $O(k \log \frac{n}{k})$ samples are required.
- l_1/l_1 : To perform l_1/l_1 recovery deterministically, at least $\Omega(k \log \frac{n}{k})$ samples are required and can be done in $O(n \log n)$ time.

1 l_2/l_1 Recovery Upper Bound

We are given a matrix $A \in \mathbb{R}^{m \times n}$ that satisfies RIP and $Y = Ax_{2k} + e$, where e is the error term. Then, we have

$$||\hat{x} - x_k|| \leq C||e||_2 \tag{1}$$

Ques: What about if x is non-sparse??

We have, $Ax = Ax_{2k} + A(x - x_{2k})$.

From (1), running with $k' = 2k$, we get that

$$||\hat{x} - x_{2k}||_2 \leq C||A(x - x_{2k})||_2 \tag{2}$$

Now, we'll use a shelling argument, similar to one we described in the last class.

Split x into blocks $x^{(1)}, x^{(2)}, ...$ of size k, so that $x^{(1)}$ has the largest k coordinates, and each next $x^{(i)}$ has the next largest k coordinates. Then, we have

$$x - x_{2k} = x^{(3)} + x^{(4)} + ...$$

Then,

$$\begin{aligned}
||A(x - x_{2k})||_2 &= ||A. \sum_{i=3}^{\infty} x^{(i)}||_2 \\
&\leq \sum_{i=3}^{\infty} ||Ax^{(i)}||_2 \\
&\leq \sum_{i=3}^{\infty} (1+\epsilon)||x^{(i)}||_2 && \text{(As } A \text{ satisfies the RIP)} \\
&\leq (1+\epsilon) \sum_{i=3}^{\infty} \sqrt{k}||x^{(i)}||_\infty \\
&\leq (1+\epsilon) \sum_{i=3}^{\infty} \sqrt{k} \frac{||x^{(i-1)}||_1}{k} \\
&= \frac{(1+\epsilon)}{\sqrt{k}} \sum_{i=2}^{\infty} ||x^{(i)}||_1 \\
&= \frac{(1+\epsilon)}{\sqrt{k}} ||x - x_k||_1
\end{aligned}$$

Now, plugging this inequality in equation (2), we have

$$||\hat{x} - x_k||_2 \leq C\frac{(1+\epsilon)}{\sqrt{k}}||x - x_k||_1 \tag{3}$$

Now, we have

$$||\hat{x} - x||_2 \leq ||\hat{x} - x_{2k}||_2 + ||x - x_{2k}||_2 \tag{4}$$

Also, by plugging $A = I$ in the previous argument, we have

$$||x - x_{2k}||_2 \leq \frac{1}{\sqrt{k}}||x - x_k||_1 \tag{5}$$

Now using equations (4) & (5) in equation (3), we have

$$||\hat{x} - x_k||_2 \leq (C\frac{(1+\epsilon)}{\sqrt{k}} + \frac{1}{\sqrt{k}})||x - x_k||_1$$

2 l_1/l_1 Recovery Algorithm

We have seen in Problem 2 of Problem Set 2 that $(k, C/\sqrt{k})$ l_2/l_1 recovery implies $(k, O(C))$ l_1/l_1 recovery. Hence, l_1/l_1 recovery guarantee is taken care of by the results in the previous section.

3 l_2/l_2 Recovery: Deterministic Lower Bound

We will show that deterministic l_2/l_2 recovery requires $\Omega(n)$ samples even for $k = 1$. So let's think about the $k = 1$ case.

Now, suppose we are given $y = Ax$ for some $A \in \mathbb{R}^{m \times n}$ and $x \in \mathbb{R}^n$, and can recover $\hat{x}$ such that

$$||\hat{x} - x||_2 \leq C \min_{i \in [n]} ||x - x_i||_2$$

where x_i contains the top i terms of x.

If $y = 0$, then $\hat{x}$ must also be zero vector.

Thus, $\forall x \in \mathcal{N} := nullspace(A)$, we need 0 to be an OK output.

Then $\forall j \in [n]$ and $x \in \mathcal{N}$,

$$\sum_i x_i^2 \leq C^2 \sum_{i \neq j} x_i^2$$
$$x_j^2 \leq (C^2 - 1).(||x||_2^2 - x_j^2)$$

$$x_j^2 \leq \underbrace{(1 - \frac{1}{C^2})}_{\alpha < 1}.||x||_2^2 \tag{6}$$

Our goal is to show that (6) implies that the dimension of $\mathcal{N}$ must be small.

Let $v_1, \ldots, v_{n-m}$ be the orthogonal basis for $\mathcal{N}$. Thus, $(n-m)$ is the dimension of the null space $\mathcal{N}$.

Let $e_i \in \mathbb{R}^n$ such that it's i-th entry is 1 and the rest of the entries are 0.

Then, $Proj_{\mathcal{N}}(e_i)$ (the orthogonal projection of e_i onto $\mathcal{N}$) $= \sum_{j=1}^{n-m} v_j v_j^T e_i$

Since $Proj_{\mathcal{N}}(e_i) \in \mathcal{N}$, using (6) we have for the ith coordinate that

$$(Proj_{\mathcal{N}}(e_i))_i = \sum_{j=1}^{n-m} e_i^T v_j v_j^T e_i \leq \sqrt{\alpha.||Proj_{\mathcal{N}}(e_i)||_2^2} \tag{7}$$

$$\sum_{j=1}^{n-m} |\langle v_j, e_i \rangle|^2 \leq \sqrt{\alpha} \tag{8}$$

Now, sum equation (8) over $i \in \{1, \ldots, n\}$ and find

$$\begin{aligned} n - m &= \sum_{j=1}^{n-m} \|v_j\|_2^2 \\ &= \sum_{j=1}^{n-m} \sum_{i=1}^{n} |\langle v_j, e_i \rangle|^2 \\ &\leq n\sqrt{\alpha} \\ &< (1 - \frac{1}{2C^2})n \end{aligned}$$

using (6). Hence, $m \geq \frac{n}{2C^2}$.

This was proved by Albert Cohen, Wolfgang Dahmen, and Ronald DeVore [CDD09] .

4 Deterministic l_1/l_1 lower bound [DIPW10]

Idea: We need to find a large set of well-separated sparse points and we should be able to cover them even in presence of lot of noise.

We'll use a Volume Argument to find such a set of points.

4.1 Gilbert-Varshamov Bound

They showed that $\forall q, k \in \mathbb{Z}^+$, $\epsilon \in \mathbb{R}^+$ with $0 < \epsilon < 1 - \frac{1}{q}$, $\exists$ a set $S \subseteq [q]^k$ such that S has minimum Hamming Distance ϵk and

$$\log |S| \geq (1 - H_q(\epsilon))k \log q$$

where H_q is the q-ary entropy function

$$H_q(\epsilon) = -\epsilon \log_q(\frac{\epsilon}{q-1}) - (1-\epsilon)\log_q(1-\epsilon)$$

If we set $q = \frac{n}{k}$ and $\epsilon = \frac{1}{2}$, then $S \subseteq [\frac{n}{k}]^k$ and has minimum hamming distance equal to $\frac{k}{2}$ and $\log |S| \gtrsim k \log \frac{n}{k}$.

We can transform the set $[q]^k$ to $\{0,1\}^k$ by taking each character and writing it into a unit. For example,

$$5 \to (0,0,0,0,1,0,\ldots,0)$$

$$6 \to (0,0,0,0,0,1,0,\ldots,0)$$

This gives us a set $S \subseteq \{0,1\}^n$ consisting of only k-sparse vectors with minimum l_1-distance k and $\log |S| \gtrsim k \log \frac{n}{k}$.

Now, suppose $x \in S$, $||w||_1 \leq \frac{k}{10}$, and we recover $\hat{x}$ from $y = A(x+w)$.

We know that

$$\begin{aligned} ||\hat{x} - (x+w)||_1 &\leq 2 \min_{k\text{-sparse } x'} ||(x+w) - x'||_1 \\ &\leq 2 \cdot \frac{k}{10} \quad \text{(can be achieved by plugging } x' = x) \\ &= \frac{k}{5} \end{aligned}$$

Now, we have

$$\begin{aligned} ||\hat{x} - x||_1 &\leq ||w||_1 + ||\hat{x} - (x+w)||_1 \\ &\leq \frac{3}{10}k \\ &< \frac{k}{2} \end{aligned}$$

We have bunch of points $x \in S$ and $S \subset B_1.k$, where B_1 is the l_1 ball in $\mathbb{R}^n$.

Now, $\forall x_i \in S$, consider a ball $(x_i + \frac{k}{10}B_1)$. For any given real matrix $A \in \mathbb{R}^{m \times n}$, we can project the ball $(x_i + \frac{k}{10}B_1)$ to $A(x_i + \frac{k}{10}B_1)$ and these balls are disjoint for different $x_i \in S$. And as $\cup_{x_i \in S}(x_i + \frac{k}{10}B_1) \subset \frac{11}{10}kB_1$, all these projected balls lies inside $A(\frac{11}{10}kB_1)$.

Now, the volume of each of the projected small balls is equal to $Vol(A(\frac{k}{10}B_1))$ and that of the bigger ball inside which each of the disjoint smaller balls lie is equal to $Vol(A(\frac{11}{10}kB_1))$. And, we have

$$\frac{Vol(A(\frac{11}{10}kB_1))}{Vol(A(\frac{k}{10}B_1))} = 11^m \tag{9}$$

Note: AB_1 is some convex shape in $\mathbb{R}^m$.

As the smaller balls are disjoint and they lie inside the bigger ball, we have

$$\begin{aligned}
|S| Vol(A(\frac{k}{10}B_1)) &\leq Vol(A(\frac{11}{10}kB_1)) \\
|S| &\leq 11^m \quad \text{(from equation (9))} \\
m &\geq \log_{11}|S| \\
m &\gtrsim k \log \frac{n}{k}
\end{aligned}$$

5 Randomized Lower Bound [DIPW10]

We'll show that any matrix $A \in \mathbb{R}^{m \times n}$ which is used for randomized l_1/l_1 recovery must have at least $m = \Omega(k \log \frac{n}{k})$ rows. We'll first assume that each of the entries A_{ij} is an integer with $O(\log n)$ bits.

Thus, the vector Ax requires $O(m \log n)$ bits. Thus, in total $\Omega(k \log \frac{n}{k} \log n)$ bits must be stored for Ax where each x_i is poly-precision ($\log n$ bits per entry).

Let S be a set of k-sparse binary vectors and has minimum hamming distance k and $\log |S| \gtrsim k \log \frac{n}{k}$.

Now, consider $x_1, x_2, \ldots, x_R \in S$.

Let

$$z = x_1 + \underbrace{\frac{1}{11}x_2 + \frac{1}{11^2}x_3 + \cdots + \frac{1}{11^{R-1}}x_R}_{=w'(\text{let})}$$

We have,

$$||w'|| \leq k(\frac{1}{11} + \frac{1}{11^2} + \ldots) = \frac{k}{10}$$

Rounding the recovery z of $y = Az$ to S gives x_1.

Note: We can relate this problem to a Communication Complexity problem. Consider the following communication game. There are two parties, Alice and Bob. Alice is given the R vectors $x_1, x_2, x_3, \ldots, x_R$ from set S. Now, Alice sends the vector Ax as a message to Bob, who must recover the vectors $x_1, x_2, x_3, \ldots, x_R$ from Az, which implies that Az has indeed $\Omega(R \log S) = \Omega(Rk \log \frac{n}{k})$ bits.

Let

$$\begin{aligned}
y^{(2)} &= (y - Ax_1).11 \\
&= A.(x_2 + \frac{1}{11}x_3 + \frac{1}{11^2}x_4 + \ldots)
\end{aligned}$$

Now, rounding $y^{(2)}$ to S gives us x_2.

We can adopt the same strategy to recover all other x_i's for all $1 \leq i \leq R$.

If this algorithm works with probability $\geq 1 - \frac{1}{2R}$, then probably all stages succeed and we can recover all the x_i, which is $\Omega(Rk \log \frac{n}{k})$ bits.

If A has $\log n$ bits per coordinate, then Ax has $(R + \log n)$ bits per coordinate.

If $R \geq \log n$, then this means we have communicated $\Omega(Rk \log \frac{n}{k})$ bits of information using only $O(m \log n)$ bits of transmission. Hence

$$m \log n \gtrsim k \log \frac{n}{k} \log n$$
$$m \gtrsim k \log \frac{n}{k}$$

5.1 Removing the assumptions

The above proof had two flaws: it assumed that the entries of A were integers with $O(\log n)$ bits per entry, and it required the algorithm to succeed with probability $1 - \frac{1}{2 \log n}$ probability. Neither of these is necessary to the proof.

To decrease the probability requirement, consider the following communication game. There are two parties, Alice and Bob. Alice is given a string $z \in \{0, 1\}^n$. Bob is given an index $i \in [n]$, together with $z_1, z_2, \ldots, z_{i-1}$. Now Alice sends some message to Bob, who must output z_i with probability at least $\frac{3}{4}$. We refer to this problem as Augmented Indexing. It is known that solving Augmented Indexing requires lots of communication:

Theorem 5.1 ([BJKS02])**.** *Any protocol that solved Augmented Indexing requires $\Omega(n)$ bits of communication.*

In our current setting, Alice has a bit string of length $R \log S$, which she converts into vectors $x_1, x_2, \ldots, x_R \in S$. Bob converts his inputs into an index $i \in [R]$ and vectors $x_1, x_2, \ldots, x_{i-1}$, and wants to learn x_i. Now Alice sends the vector Az to Bob, who must recover the vector x_i.

Lemma 5.2. *[DIPW10] Consider any $m \times n$ matrix A with orthonormal rows. Let A' be the result of rounding A to $c \log n$ bits per entry. Then for any $x \in \mathbb{R}^n$ with $A'x = A(x + \epsilon)$ and $||\epsilon||_1 < n^{2-c}$*

References

[BJKS02] Z. Bar-Yossef, T.S. Jayram, R. Kumar, and D. Sivakumar. Information theory methods in communication complexity. In *Proceedings 17th Annual IEEE Conference on Computational Complexity*, pages 133–142, 2002.

[CRT06] Candes, Emmanuel J., Justin K. Romberg, and Terence Tao. "Stable signal recovery from incomplete and inaccurate measurements." Communications on pure and applied mathematics 59.8 (2006): 1207-1223.

[CDD09] Cohen, Albert, Wolfgang Dahmen, and Ronald DeVore. "Compressed sensing and best -term approximation." Journal of the American Mathematical Society 22.1 (2009): 211-231.

[DIPW10] Do Ba, K., Indyk, P., Price, E., & Woodruff, D. P. (2010, January). Lower Bounds for Sparse Recovery. In SODA (Vol. 10, pp. 1190-1197).