Sketching Algorithms for Big Data Fall 2017

Lecture 15 — October 24th, 2017

Prof. Piotr Indyk *Scribe: Yueqi Sheng*

1 Overview

In the past two lectures, we covered two algorithms for compressed sensing, Basis Pursuit (BP) and Iterative hard thresholding (ITH). Recall that in compressed sensing, we want to recover $x \in \mathbb{R}^n$ from Ax for some sketch matrix $A \in \mathbb{R}^{m\times n}$. We showed that if A is (ϵ, Ck) RIP, then one can recover $\hat{x}$ s.t.

$$\|\hat{\mathrm{x}}\text{-x}\|_1 \leq c(k) \min_{y \text{ is } k-sparse} \|\text{y-x}\|_1$$

Our goal is to have a good compression while be able to recover x.

One question to ask is which kind of sketch matrix works the best. Intuitively, the trade off between dense vs sparse matrices are as follows: While dense matrix gives shorter sketches (smaller m), sparse matrix is more computationally efficient. In this lecture we focus on construction of sketch matrices that balance between the two cases.

2 RIP_1 matrix

Examples of dense matrix that satisfies $(c, k)-RIP$ property are Gaussian/Bernoulli matrices with $m = O(k \log(\frac{n}{k}))$ and random Fourier matrices with $m = O(k \log^{O(1)n})$.

For sparse matrices: It was shown in [2] that all sparse binary matrix that satisfies RIP must has $m = \Omega(k^2)$.

To get around this, we look at RIP property w.r.t. l_1 norm instead of l_2 norm. Turns out that sketch matrices satisfies RIP_1 is enough for BP to approximate the original input x.

Definition 1 (RIP_1)**.** *A matrix A is (ϵ, k)-RIP_1 if for all k sparse vector v,*

$$(1-\epsilon)\|\text{v}\|_1 \leq \|\text{Av}\|_1 \leq (1+\epsilon)\|\text{v}\|_1$$

3 Construction of RIP_1 matrix

In this section we give a construction of sparse binary matrix satisfies RIP_1 property. The idea is to view A as the (bi)adjancy matrix of a bipartite graph. If the underlying graph is an unbalanced expander, then A satisfies RIP_1.

Definition 2 (Expander)**.** A (l, ϵ)-*unbalanced expander* is a bipartite simple graph $G = (U, V, E)$, $|U| = n, |V| = m$, with left degree d such that for any $X \subset U$ with $|X| \leq l$, the set of neighbors $N(X)$ of X has size $|N(X)| \geq (1-\epsilon)d|X|$.

From $A \in \{0,1\}^{m\times n}$, one can construct a $G = (U, V, E)$ as follows: let $U = [n]$, $V = [m]$. $E = \{(i,j) : i \in U, j \in V, A_{j,i} = 1\}$.

Here we assume each column of A has d 1s or $\forall i \in U$, $deg(i) = d$.

Theorem 3 ([1]). *For $A \in \{0,1\}^{m\times n}$, if the underlying bipartite graph is a $(k, d(1-\frac{\epsilon}{2}))$ expander, then for all k sparse vector v*

$$d(1-\epsilon)\left\|\mathrm{v}\right\|_1 \le \left\|\mathrm{Av}\right\|_1 \le d\left\|\mathrm{v}\right\|_1$$

(The other direction is also true: Given a binary sparse matrix satisfies RIP_1, the underlying graph is an expander.)

Proof. $\left\|\mathrm{Av}\right\|_1 \le d\left\|\mathrm{v}\right\|_1$: for any $v \in \mathbb{R}^n$, think of Av as for each $i \in U$, send $v_i \to j$ if $(i,j) \in E$, each v_i is seen d times from V.

$$\left\|\mathrm{Av}\right\|_1 = \sum_i |(Av)_i| \le \sum_i |\sum_{j:(i,j)\in E} v_j| \le \sum_{(i,j)\in E} |v_i| = d\left\|\mathrm{v}\right\|_1$$

$d(1-\epsilon)\left\|\mathrm{v}\right\|_1 \le \left\|\mathrm{Av}\right\|_1$: Let v be some k sparse vector. WLOG, sort the corrdinates of v s.t. $v_1 \ge v_2 \ge \cdots \ge v_k > v_{k+1} = \cdots = v_n = 0$.

Sort $e = (i,j)$ in lexicographic order. Let $r(e) = 1$ if e is not seen before and $r(e) = -1$ if $\exists i' < i$ s.t. $(i', j) \in E$. The inequality follows from the following two claims.

Claim 4. $\left\|\mathrm{Av}\right\|_1 \ge \sum_{(i,j)=e\in E} r(e)|v_i|$

Claim 5. $\sum_{(i,j)=e\in E} r(e)|v_i| \ge (1-\epsilon)d\left\|\mathrm{v}\right\|_1$

Combine claim 4, 5 completes the proof. □

proof of claim 4. For $j \in U$, if $|N(j)| = 1$, then $|(Av)_j| = |v_{N(j)}|$. Otherwise, let $a_j = argmax\{i : i \in N(j)\}$. By construction, $|v_{a_j}| \ge |v_i|$ for all other $i \in N(j)$. We also get $|(Av)_j| \ge |v_{a_j}| - \sum_{i\in N(j), i\ne a_j} |v_i| = \sum_{i\in N(j)} |v_i| r(e)$. This completes the proof. □

proof of claim 5. Since the underlying graph is $(k, d(1-\frac{\epsilon}{2}))$ expander, then for any $i, i' \in \{1, \cdots, k\}$, $|N(i) \cap N(i')| \le \epsilon d$. By definition, for $e = (i,j)$, $r(e) = -1$ iff $\exists i' < i$ s.t. $(i', j) \in E$, let $r'(e) = -1$ for the top ϵd neighbors of i for $i > 1$, observe that

$$\sum_{(i,j)=e\in E} r(e)|v_i| \ge \sum_{(i,j)=e\in E} r'(e)|v_i| \ge d\left\|\mathrm{v}\right\|_1 - \epsilon d\left\|\mathrm{v}\right\|_1$$

□

4 RIP_1 + Expander allows l_1 minimization

Recall that in Lecture 13, we showed if all vector in the null space of A doesn't have mass concentrated on some small subset of coordinate (**null space condition**), then l_1 minimization gives good error guarantee. In this section, we will show that matrix satisfies RIP_1 property also satisfies the null space condition.

Definition 6 (null space condition)**.** *A satisfies the null space property of order k if for any η s.t. $A\eta = 0$ and $S \subset [n]$ s.t. $|S| \le k$,*

$$\|\eta_{\mathrm{S}}\|_1 \le C(\epsilon)\, \|\eta\|_1$$

4.1 Original notes from lecture

We also define $E(X : Y) = E \cap (X \times Y)$ to be the set of edges between the sets X and Y.

The following well-known proposition can be shown using Chernoff bounds.

Claim 7. For any $n/2 \ge l \ge 1$, $\epsilon > 0$, there exists a (l, ϵ)-unbalanced expander with left degree $d = O(\log(n/l)/\epsilon)$ and right set size $O(ld/\epsilon) = O(l \log(n/l)/\epsilon^2)$.

Now we show that the expander matrices have the null-space property. Let A be an $m \times n$ adjacency matrix of an unbalanced $(2k, \epsilon)$-expander G with left degree d. Let $\alpha(\epsilon) = (2\epsilon)/(1 - 2\epsilon)$.

Lemma 8. *Consider any $\eta \in \mathbb{R}^n$ such that $A\eta = 0$, and let S be any set of k coordinates of η. Then we have*

$$\|\eta_S\|_1 \le \alpha(\epsilon)\|\eta\|_1$$

Proof. Without loss of generality, we can assume that S consists of the largest (in magnitude) coefficients of η. We partition coordinates into sets $S_0, S_1, S_2, \ldots S_t$, such that (i) the coordinates in the set S_l are not-larger (in magnitude) than the coordinates in the set S_{l-1}, $l \ge 1$, and (ii) all sets but S_t have size k. Therefore, $S_0 = S$. Let A' be a submatrix of A containing rows from $N(S)$.

The basic idea of the proof is as follows. Assume (by contradiction) that $\|\eta_S\|_1$ is "large" compared to $\|\eta\|_1$, which (by RIP1) implies that $\|A'\eta_S\|_1$ is "large". Since $0 = \|A'\eta\|_1 = \|A'\eta_S + A'\eta_{-S}\|_1$, it follows that $\|A'\eta_{-S}\|_1$ must be "large", to cancel the contribution of $A'\eta_S$. The only way for this to happen though is if there are many edges in G from $-S$ to $N(S)$. This however would mean that the neighborhoods of S and blocks S_i have large overlaps, which cannot happen since the graph is an expander.

The formal proof follows.

From the RIP-1 property we know that $\|A'\eta_S\|_1 = \|A\eta_S\|_1 \ge d(1 - 2\epsilon)\|\eta_S\|_1$. At the same time, we know that $\|A'\eta\|_1 = 0$. Therefore

$$\begin{aligned} 0 = \|A'\eta\|_1 &\ge \|A'\eta_S\|_1 - \sum_{l\ge1} \sum_{(i,j)\in E, i\in S_l, j\in N(S)} |\eta_i| \\ &\ge d(1-2\epsilon)\|\eta_S\|_1 - \sum_{l\ge1} |E(S_l : N(S))| \min_{i\in S_{l-1}} |\eta_i| \\ &\ge d(1-2\epsilon)\|\eta_S\|_1 - \sum_{l\ge1} |E(S_l : N(S))| \cdot \|\eta_{S_{l-1}}\|_1/k \end{aligned}$$

From the expansion properties of G it follows that, for $l \ge 1$, we have $|N(S \cup S_l)| \ge d(1-\epsilon)|S \cup S_l|$.

It follows that at most $d\epsilon 2k$ edges can cross from S_l to $N(S)$, and therefore

$$\begin{aligned} 0 &\geq d(1-2\epsilon)\|\eta_S\|_1 - \sum_{l\geq 1} |E(S_l : N(S))| \cdot \|\eta_{S_{l-1}}\|_1/k \\ &\geq d(1-2\epsilon)\|\eta_S\|_1 - d\epsilon 2k \sum_{l\geq 1} \|\eta_{S_{l-1}}\|_1/k \\ &\geq d(1-2\epsilon)\|\eta_S\|_1 - 2d\epsilon\|\eta\|_1 \end{aligned}$$

It follows that $d(1-2\epsilon)\|\eta_S\|_1 \leq 2d\epsilon\|\eta\|_1$, and thus $\|\eta_S\|_1 \leq (2\epsilon)/(1-2\epsilon)\|\eta\|_1$. □

References

[1] R. Berinde and A. C. Gilbert and P. Indyk and H. Karloff and M. J. Strauss.

Combining geometry and combinatorics: A unified approach to sparse signal recovery *2008 46th Annual Allerton Con-ference on Communication, Control, and Computing (Urbana-Champaign, 2008), 798805.*

[2] Venkat Chandar, A negative result concerning explicit matrices with the restricted isometry property.

2008