MTH 995-001: Intro to CS and Big Data	Spring 2015
Lecture 21 — April 4th, 2015	
Inst. Mark Iwen	*Scribe: Minh Pham*

1 Overview

In the last lecture we proved the JohnsonLindenstrauss lemma (J-L lemma) implies the RIP property

In this lecture we will discuss the lower bound on number of rows of a RIP matrices and compare this to the matrix the J-L lemma gives us.

2 Lecture

Lemma 1. *Given* $K, N \in \mathbb{N}, K < N$.
Then $\exists n \geq (\frac{N}{4K})^{K/2}$ *subsets* $S_1, S_2, ..., S_n \subseteq [N]$ *such that*

$$\begin{cases} |S_j| = K, \forall j \in [n] \\ |S_i \cap S_j| < \frac{K}{2}, \forall i \neq j \end{cases}$$

Proof: Assume $K < \frac{N}{4}$
Let $C = \{S \subseteq [N] | K = |S|\}$
Pick $S_1 \in C$
Let $C_1 \subseteq C : C_1 = \{S \in C | \frac{K}{2} \leq |S \cap S_1|\}$ i.e. things that looks like S_1
$|C_1| = \sum_{S=\lceil K/2 \rceil}^{K} \binom{K}{S}\binom{N-K}{K-S}$
$\leq 2^K \max_{\lceil \frac{K}{2} \rceil \leq S \leq K} \binom{N-K}{K-S}$, by sum of binomial $\leq 2^K$
$= 2^K \binom{N-K}{\lfloor K/2 \rfloor}$ because N-K is small, so it maximize when K-S maximize

We will pick $S_2, S_3, ..., S_n$ using the following algorithm:

$Pick : S_1, C_1, C$
$n \leftarrow 1$
$While : |C \setminus \bigcup_{l=1}^{n} C_l| > 0$
$- \quad Choose : S_{n+1} \in C \setminus \bigcup_{l=1}^{n} C_l$
$- \quad Set : C_{n+1} = \{S \in C \setminus \bigcup_{l=1}^{n} C_l | \frac{K}{2} \leq |S \cap S_{n+1}|\}$
$- \quad n \leftarrow n + 1$

Note that by construction: $|S_i \cap S_j| < \frac{K}{2}, \forall i \neq j$
Algorithm stop when

$n \geq \frac{|C|}{\max_{1\leq i\leq n}|C_i|} \geq \frac{\binom{N}{K}}{2^K\binom{N-K}{\lfloor K/2\rfloor}} \geq (\frac{N}{4K})^{\frac{K}{2}}$ by expanding binomial coefficient

□

Theorem 1. *Given* $A \in \mathbb{R}^{m\times N}$
Condition for Basis Pursuit (BP), i.e. :
$\forall x \in \mathbb{R}^N$ *with* $||x||_0 \leq 2K, \forall z \in \mathbb{R}^N$ *with* $Az = Ax$ *then* $||x||_1 \leq ||z||_1$
Then
$m \geq \frac{K}{ln9} ln(\frac{N}{4K})$

Proof: Consider $[x] = x + Ker(A)$
associated with a norm: $||[x]|| := \inf_{v\in Ker(A)} ||x - v||_1$
Indentify (i.e., note the existence of the bijection) $y \in Ker(A)^{\perp}$ with $[y]$
This bijection induces a norm in Ker(A): $||y||_S = ||[y]||$
Suppose $||x||_0 \leq 2K$
Project x on $Ker(A)^{\perp}$:

$$\begin{aligned} ||\prod_{Ker(A)^{\perp}} x||_S &= ||[\prod_{Ker(A)^{\perp}} x]|| \\ &= \inf_{v\in Ker(A)} ||\prod_{Ker(A)^{\perp}} x - v||_1 \\ &= \inf_{v\in Ker(A)} ||x - v||_1 \\ &= ||x||_1 \end{aligned}$$

The last equality is due to Basis Pursuit condition (note that $A(x - v) = A(x)$)

Let S_j be the subsets from lemma 1 :

$$\begin{cases} |S_j| = K, \forall j \in [n] \\ |S_i \cap S_j| < \frac{K}{2}, \forall i \neq j \end{cases}$$

Let y_j be vector in R^N such that:

$$(y_j)_l = \begin{cases} \frac{1}{K}\ if\ l \in S_j \\ 0\ otherwise \end{cases}$$

Note that $||y_j||_0 = k$ and $||y_j||_1 = 1$
Define $x_j = \prod_{Ker(A)^{\perp}} y_j$
Note:
$||x_j||_S = ||\prod_{Ker(A)^{\perp}} y_j||_S = ||y_j||_1 = 1$
$||x_j - x_l||_S = ||y_j - y_l||_1 > 1$ since $K \leq ||y_j - y_l||_0 \leq 2K$
Hence:

$$\begin{aligned}
(\frac{N}{4K})^{\frac{K}{2}} &\le n \quad by\ lemma\ 1 \\
&\le P_1(||.||_S - ball\ in\ Ker(A)^{\perp}) \quad since\ x_j\ can\ be\ a\ packing \\
&\le (1+\frac{2}{1})^{rank(A)} \\
&= 3^m
\end{aligned}$$

Take the log of both side and we are done. □

This theorem give us a lower bound on the number of rows, m, of matrices with RIP and NSP (null space property), which is in order of k. This means our random matrices from J-L lemma are pretty close to optimal as they are also in order of k.