CS 395T: Sublinear Algorithms	Fall 2016

Lecture 20 — November 3, 2016

Prof. Eric Price | *Scribe: Yitao Chen, Shanshan Wu*

1 Overview

In this lecture, we are going to talk about Fourier RIP matrices, which includes the following

- Fourier Uncertainty Principle
- Proving RIP
  - Symmetrization/Gaussianization
  - Chaining/Dudley's Entropy Integral
  - Maurey's Emprical Method

2 Fourier Uncertainty Principle

For $x \in \mathbb{C}^n$, the Discrete Fourier Transform (DFT) $\hat{x} \in \mathbb{C}^n$ is $\hat{x} = Fx$, where $F_{ij} = \omega^{ij}, \forall i, j$ and $\omega = e^{2\pi\sqrt{-1}/n}$, which implies $\omega^n = 1$,

$$F = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \dots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \dots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \dots & \omega^{(n-1)^2} \end{bmatrix}.$$

For F to be RIP matrix, we just need: $F \cdot F^* = nI_n$ and $|F_{ij}| \le 1, \forall i, j$. Also notice that Hadamard matrix $H_{i,j} = (-1)^{<i,j>}$, where i, j are vectors in $\{0,1\}^{\log n}$ has this property: $HH^T = nI_n$.

Lemma 1. *Let $x \in \mathbb{C}^n$ be k-sparse, then we have supp$(\hat{x}) \ge n/k$.*

Proof. Let $x^{(l)}$ be a "modulation" of x ($x \ne 0$), $x_i^{(l)} = x_i \cdot \omega^{-li}$. Take a look at the jth coordinate of the Fourier transform of the modulated signal $x^{(l)}$, we have

$$(\widehat{x^{(l)}})_j = (Fx^{(l)})_j = \sum_{i=0}^{n-1} \omega^{ij} x_i^{(l)} = \sum_{i=0}^{n-1} x_i \cdot \omega^{i(j-l)} = \hat{x}_{j-l}. \tag{1}$$

Let $X = \text{span}(x^{(0)}, x^{(1)}, \dots, x^{(n-1)})$, Equation (**??**) tells us the Fourier transform of X just rotates the coordinates of X, and modulation does not change the sparseness, so

$$k \ge \dim(X) = \dim(FX). \tag{2}$$

Let $T_l := \cup_{j=0}^{l} \text{supp}(\widehat{x^{(j)}})$ and $S := T_D = \text{supp}(\hat{x})$, we have $|T_n| = n$ and

$$|T_l \setminus T_{l-1}| \leq |\text{supp}(\widehat{x^{(l)}})| = |\text{supp}(\hat{x})| = |S|.$$

We claim

$$\dim(FX) \geq \text{number of times } T_l \neq T_{l-1} \geq \frac{n}{|S|}. \tag{3}$$

Combining Equation (**??**) and (**??**), we have $|S| \geq n/k$. □

3 Proving RIP

Let F be unitary and bounded. Let $\Omega \subseteq [n]$ be a multiset of $m = O(\frac{k}{\epsilon^2} \log^4 n)$ i.i.d. uniform indices. We have the following claim,

Claim 2. *F_Ω = rows of F corresponding to Ω has (k, ϵ) RIP "in expectation". Here "in expectation" means considering*

$$\Delta := \mathop{\mathbb{E}}_{\Omega}\left[\mathop{sup}_{S \subseteq [n], |S| \leq k} \|I - \frac{1}{m} F_{\Omega \times S}^T F_{\Omega \times S}\| \right].$$

Let x_i^T be rows of F, so Δ can be written as

$$\Delta := \mathop{\mathbb{E}}_{\Omega}\left[\sup_{S \subseteq [n], |S| \leq k} \|I - \frac{1}{m} \sum_{i \in \Omega} x_i^S {x_i^S}^T \| \right]. \tag{4}$$

3.1 Symmetrization/Gaussianization

Let $\|\cdot\|$ be a norm (it has convexity and triangle inequality holds) and $x_i \sim X$ independently for any i, we claim the following inequality holds,

$$\begin{aligned} \mathop{\mathbb{E}}_{x_1,\ldots,x_m \sim X}\left[\|\frac{1}{m} \sum_{i=1}^m x_i - \mathop{\mathbb{E}}_{x \sim X}[x]\| \right] &\leq 2 \mathop{\mathbb{E}}_{x_1,\ldots,x_m \sim X, s_1,\ldots,s_m \sim \{\pm 1\}}\left[\|\frac{1}{m} \sum_{i=1}^m s_i x_i\| \right] \\ &\leq \sqrt{2\pi} \mathop{\mathbb{E}}_{x_1,\ldots,x_m \sim X, g_1,\ldots,g_m \sim N(0,1)}\left[\|\frac{1}{m} \sum_{i=1}^m g_i x_i\| \right] \end{aligned}$$

Proof. Draw $x_i' \sim X$ independently, let $s_1, \ldots, s_m \sim \{\pm 1\}$, and $g_1, \ldots, g_m \sim N(0,1)$, we have

$$\begin{aligned}
\text{LHS} &= \mathop{\mathbb{E}}_{x}\left[\|\frac{1}{m}\sum_{i=1}^{m} x_i - \mathop{\mathbb{E}}_{x'}\left[\frac{1}{m}\sum_{i=1}^{m} x_i'\|\right]\right] \\
&\leq \mathop{\mathbb{E}}_{x,x'}\left[\|\frac{1}{m}\sum_{i=1}^{m}(x_i - x_i')\|\right] \\
&= \mathop{\mathbb{E}}_{x,x',s}\left[\|\frac{1}{m}\sum_{i=1}^{m} s_i(x_i - x_i')\|\right] \\
&\leq 2\mathop{\mathbb{E}}_{x,s}\left[\|\frac{1}{m}\sum_{i=1}^{m} s_i x_i\|\right] \\
&= 2\mathop{\mathbb{E}}_{x,s}\left[\|\frac{1}{m}\sum_{i=1}^{m} s_i x_i \,\mathbb{E}[|g_i|]\cdot\sqrt{\frac{\pi}{2}}\|\right] \\
&\leq \sqrt{2\pi}\mathop{\mathbb{E}}_{x,s,g}\left[\|\frac{1}{m}\sum_{i=1}^{m} s_i|g_i|x_i\|\right] \\
&= \sqrt{2\pi}\mathop{\mathbb{E}}_{x,g}\left[\|\frac{1}{m}\sum_{i=1}^{m} g_i x_i\|\right].
\end{aligned}$$

This first and the third inequalities follow from the convexity of norm operation. The second inequality follows from triangle inequality. The last equality is true because $s_i|g_i| \sim N(0,1)$. ☐

Next, let $\Sigma_k := \{y \in \mathbb{C}^n : y \text{ is } k\text{-sparse and} \|y\|_2 = 1\}$, we can use the Symmetrization/Gaussianization technique to transform (**??**) as

$$\begin{aligned}
m\Delta &\lesssim \mathop{\mathbb{E}}_{\Omega,g} \sup_{|S|\leq k}\|\sum_{i\in\Omega} g_i x_i^S x_i^{S^T}\| \\
&\lesssim \mathop{\mathbb{E}}_{\Omega,g} \sup_{y\in\Sigma_k}|\sum_{i\in\Omega} g_i \langle x_i, y\rangle^2|. \qquad (5)
\end{aligned}$$

We will now bound the RHS of the (**??**) for every set Ω via a technique called *chaining*. Before introduction to chaining, let us first define a related concept called "Gaussian Process"[1].

4 Gaussian Process

Suppose we are given a bounded (but may be infinite) set of vectors $S \subset \mathbb{R}^n$ (e.g., S can be all k-sparse vectors with $\|\cdot\|_2 \leq 1$). For each vector $x \in S$, we associate it with a random (usually zero-mean) Gaussian variable G_x. This is called a Gaussian Process. One property about $\{G_x\}_{x\in S}$ is that we can derive concentration bounds on $G_x - G_y$ in terms of a distance measure $d(x,y)$. More specifically, define $d(x,y) = \mathbb{E}[(G_x - G_y)^2]^{1/2}$, then $G_x - G_y$ is distributed as $N(0, d(x,y)^2)$.

[1] We found this blog post [**?**] by Jelani Nelson very useful for understanding chaining methods.

As an example, let $g \in \mathbb{R}^n$ be a random vector with entries drawn i.i.d. from standard normal distribution. We can define $G_x := \langle g, x \rangle$. Then

$$d(x,y)^2 = \mathbb{E}[(G_x - G_y)^2] = \mathbb{E}[(\langle g, x-y \rangle)^2] = \|x-y\|_2^2, \quad \forall x, y \in S.$$

Since $G_x - G_y$ is distributed as $N(0, d(x,y)^2)$, we can bound it as

$$\mathbb{P}(|G_x - G_y| \geq t) \lesssim e^{-\frac{t^2}{2\|x-y\|_2^2}}.$$

One nice property about Gaussian Process is that we can use chaining technique (which we will show in the next section) to bound

$$\mathbb{E}[\sup_{x \in S} G_x]. \tag{6}$$

Many problems can be transformed into this form of (**??**). For example, consider the problem of computing $\mathbb{E}[\|A\|_2]$ for a random matrix $A \in \mathbb{R}^{m \times n}$ with entries being i.i.d. $N(0,1)$. This problem can be formulated as

$$\mathbb{E}[\sup_{x \in \mathbb{S}^{n-1},\, y \in \mathbb{S}^{m-1}} y^T A x],$$

where $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ is the n-dimensional unit sphere. In this case, we define a Gaussian Process on the set $S = \{(x,y) : x \in \mathbb{S}^{n-1},\ y \in \mathbb{S}^{m-1}\}$, with distance metric given by

$$\begin{aligned} d((x,y),(x',y'))^2 &= \mathbb{E}[(y^T A x - y'^T A x')^2] = \mathbb{E}[(\mathrm{tr}(y^T A x) - \mathrm{tr}(y'^T A x'))^2] \\ &= \mathbb{E}[(\mathrm{tr}(A x y^T) - \mathrm{tr}(A x' y'^T))^2] = \mathbb{E}[(\mathrm{tr}(A x y^T) - \mathrm{tr}(A x' y'^T))^2] \\ &= \mathbb{E}[(\mathrm{tr}(A(x y^T - x' y'^T))^2] = \|x y^T - x' y'^T\|_F^2, \end{aligned}$$

where the last equality follows from the fact each entry of A is i.i.d. standard Gaussian.

5 Chaining

Given a set S and a Gaussian process $\{G_x\}_{x \in S}$, we are interested in bounding $\mathbb{E}[\sup_{x \in S} G_x]$. We show how chaining techniques can be used to obtain a non-trivial bound.

5.1 Naive bound

In Homework 1, we have shown that the maximum value of n i.i.d. N(0,1) random variables scales as $\sqrt{\log n}$. The same technique (for upper bound) can be used here to derive a bound when S is a finite set. Suppose $0 \in S$, $G_0 = 0$, then we have[2]

$$\mathbb{E}[\sup_{x \in S} G_x] \lesssim \max_{x \in S} d(x,0) \sqrt{\log |S|}, \tag{7}$$

where $d(x,0) = \mathbb{E}[G_x^2]^{1/2}$ captures the radius of S under the given Gaussian process. The above bound depends on $|S|$, and hence is not applicable if S is a infinite set.

[2]A detailed proof can be found in [**?**].

5.2 An ϵ-cover

Consider an ϵ-cover of S under the distance metric d, i.e., for every $x \in S$, there exists a vector $c(x)$ such that $d(x, c(x)) \leq \epsilon$. Let $\mathcal{N}(S, d, \epsilon)$ be the covering number. Since $G_x = G_x - G_{c(x)} + G_{c(x)}$, we have

$$\begin{aligned}\mathbb{E}[\sup_{x\in S} G_x] &\leq \mathbb{E}[\sup_{x\in S} G_x - G_{c(x)}] + \mathbb{E}[\sup_{x\in S} G_{c(x)}] \\ &\lesssim \epsilon\sqrt{\log|S|} + \max_{x\in S} d(c(x), 0)\sqrt{\log \mathcal{N}(S, d, \epsilon)}, \end{aligned} \tag{8}$$

where the last inequality follows from (**??**) and $d(x - c(x), 0) = d(x, c(x)) \leq \epsilon$. The above bound is better than (**??**), however, it still depends on $|S|$. To remove this dependence, we can use the same idea repeatedly by constructing a sequence of ϵ-covers, a technique called chaining.

5.3 Chaining

The idea of chaining is to partition S into a sequence of covers at different resolutions, and then use (**??**) to bound their difference. Let $R = \max_{x\in S} d(x, 0)$ be the radius S under distance metric d. Let $S_i \subset S$ be a $R/2^i$-cover of S. For every $x \in S$, let $x_i \in S_i$ be the closest point to x, then $d(x, x_i) \leq R/2^i$. Suppose $S_0 = \{0\}$, then $x_0 = 0$. We have

$$G_x = G_{x_0} + G_{x_1} - G_{x_0} + G_{x_2} - G_{x_1} + \cdots = \sum_{r=1}^{\infty} G_{x_r} - G_{x_{r-1}} = \sum_{r=1}^{\infty} G_{x_r - x_{r-1}}.$$

Let $\mathcal{N}(S, d, R/2^i)$ be the corresponding covering number, and suppose that $|S_i| = \mathcal{N}(S, d, R/2^i)$. Then

$$\mathbb{E}[\sup_{x\in S} G_x] \leq \sum_{r=1}^{\infty} \mathbb{E}[\sup_{x\in S} G_{x_r - x_{r-1}}] \lesssim \sum_{r=1}^{\infty} \frac{R}{2^r}\sqrt{\log \mathcal{N}(S, d, \frac{R}{2^r})^2} \lesssim \int_0^{\infty} \sqrt{\log \mathcal{N}(S, d, u)}\mathrm{d}u, \tag{9}$$

where the second inequality uses (**??**), with triangle inequality

$$d(x_r, x_{r-1}) \leq d(x, x_r) + d(x, x_{r-1}) \leq \frac{R}{2^r} + \frac{R}{2^{r-1}} = \frac{3R}{2^r},$$

being used to bound $\max_{x\in S} d(x_r, x_{r-1})$. Note that $|\{x_r - x_{r-1} : x \in S\}| \leq \mathcal{N}(S, d, \frac{R}{2^r})^2$.

The bound given in (**??**) is called *Dudley's Entropy Integral*. How good is this bound? It can lose a log factor. An even tighter bound can be obtained using the generic chaining method [**?**, **?**].

References

[JN16] Jelani Nelson. Chaining methods continued. Blog post. URL: `https://windowsontheory.org/2016/02/29/chaining-methods-continued-guest-post-by-jelani-nelson/`, February, 2016.

[Dud67] Richard M. Dudley. The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Functional Analysis*, 1:29–330, 1967.

[Fer75] Xavier Fernique. Regularit des trajectoires des fonctions alatoires gaussiennes. *Lecture Notes in Math.*, 480:1–96, 1975.

[Tal14] Michel Talagrand. Upper and lower bounds for stochastic processes: modern methods and classical problems. *Springer*, 2014.