Lecture 19 — October 20, 2021

Sebastien Roch, UW-Madison *Scribe: Xufeng Cai*

1 Overview

In the last lecture we have (a) studied the cover number and packing number of a set K and their equivalence, (b) proved the bound of the suprema of Lipschitz sub-gaussian random processes, (c) introduced the definition of ε-separated and it induces ε-nets, and (d) discussed the relation between the cover number and the volume of K in $\mathbb{R}^n$.

In this lecture we first introduce some useful preliminaries on isometries and approximate isometries, and then prove the two-sided bound on sub-gaussian matrices.

2 Isometries and Approximate Isometries

We begin by describing the preliminaries on isometries and approximate isometries. We can refer to the slides `https://people.math.wisc.edu/~roch/hdps/roch-hdps-slides19.pdf` and Section `4.1.5` in [1]. Here we introduce a lemma about approximate isometries that will be useful for proving the bound in Section 3, and its proof can be found in [1].

Lemma 1 (Approximate isometries, Exercise 4.1.4 in [1])**.** *Let A be an $m \times n$ matrix and $\delta > 0$. Suppose that*

$$\|A^\top A - I_n\| \leq \max(\delta, \delta^2).$$

Then

$$(1-\delta)\|\mathbf{x}\|_2 \leq \|A\mathbf{x}\|_2 \leq (1+\delta)\|\mathbf{x}\|_2 \quad \text{for all } \mathbf{x} \in \mathbb{R}^n.$$

Consequently, all singular values of A are between $1-\delta$ and $1+\delta$:

$$1-\delta \leq s_n(A) \leq s_1(A) \leq 1+\delta.$$

3 Bound on Sub-gaussian Matrices

Now we are ready to prove a two sided bound on the entire spectrum of an $m \times n$ sub-gaussian matrix A.

Theorem 2 (Two-sided bound on sub-gaussian matrices, Theorem 4.6.1 in [1])**.** *Let A be an $m\times n$ matrix whose rows A_i are independent, mean zero, sub-gaussian isotropic random vectors in $\mathbb{R}^n$. Then for any $t \geq 0$ we have*

$$\sqrt{m} - CK^2(\sqrt{n}+t) \leq s_n(A) \leq s_1(A) \leq \sqrt{m} + CK^2(\sqrt{n}+t) \tag{1}$$

with probability at least $1-2\exp(-t^2)$. Here $K = \max_i \|A_i\|_{\psi_2}$.

We have the following observations about this bound.

Remark 1. *A tall random matrix A with $m \gg n$ is an approximate isometry.*

Remark 2. *Here the independence of entries is going to be relaxed to just independence of rows. This relaxation is quite important in some data science applications, where the rows of A could be samples from a high-dimensional distribution. The samples are usually independent, but the coordinates of the distribution (the "parameters") are usually not independent. So it is not reasonable to assume the independence of columns of A.*

The main idea to prove this bound is to use an ε-net argument with Bernstein's concentration inequality. Note that by Lemma 1 we can prove a slightly stronger bound instead other than (1), i.e.

$$\|\frac{1}{m}A^\top A - I_n\| \leq K^2 \max(\delta, \delta^2), \tag{2}$$

where $\delta = C\left(\sqrt{\frac{n}{m}} + \frac{t}{\sqrt{m}}\right)$ and C is a large enough constant to be determined later. Before we come to prove Theorem 2, we first introduce two lemmas that will be used in the main proof.

Lemma 3 (Operator norm on a net, Lemma 4.4.1 and Exercise 4.4.3 in [1])**.** *Let A be an $m \times n$ matrix and $\varepsilon \in [0, 1)$. Then for any ε-net $\mathcal{N}$ of the sphere $\mathcal{S}^{n-1}$, we have*

(a) $\|A\| \leq \frac{1}{1-\varepsilon} \cdot \sup_{\mathbf{x}\in\mathcal{N}} \|A\mathbf{x}\|_2$.

(b) If $m = n$ and A is symmetric and $\varepsilon \in [0, 1/2)$, then $\|A\| \leq \frac{1}{1-2\varepsilon} \cdot \sup_{\mathbf{x}\in\mathcal{N}} |\langle A\mathbf{x}, \mathbf{x}\rangle|$.

Proof.

(a) Take a vector $\mathbf{x} \in \mathcal{S}^{n-1}$ such that $\|A\| = \|A\mathbf{x}\|_2$, and choose a $\mathbf{y} \in \mathcal{N}$ such that $\|\mathbf{x} - \mathbf{y}\|_2 \leq \varepsilon$ to approximate $\mathbf{x}$. By the definition of the operator norm, we have

$$\|Ax - Ay\|_2 = \|A\,(x-y)\|_2 \leq \|A\|\,\|x-y\|_2 \leq \varepsilon\|A\|.$$

Using triangular inequality, we obtain

$$\|Ay\|_2 \geq \|Ax\|_2 - \|Ax - Ay\|_2 \geq \|A\| - \varepsilon\|A\| = (1-\varepsilon)\|A\|.$$

Dividing both sides by $1-\varepsilon$ and $\|Ay\|_2 \leq \sup_{\mathbf{x}\in\mathcal{N}} \|A\mathbf{x}\|_2$, we complete the proof.

(b) By the definition of the operator norm from the quadratic from aspect and since A is symmetric, we take a vector $\mathbf{x} \in \mathcal{S}^{n-1}$ such that $\|A\| = \langle A\mathbf{x}, \mathbf{x}\rangle$. Also, we choose a $\mathbf{y} \in \mathcal{N} \subseteq \mathcal{S}^{n-1}$ such that $\|\mathbf{x} - \mathbf{y}\|_2 \leq \varepsilon$. Then by Cauchy-Schwarz inequality, we have

$$|\langle A\mathbf{x}, \mathbf{x} - \mathbf{y}\rangle| \leq \|A\mathbf{x}\|_2\|\mathbf{x} - \mathbf{y}\|_2 \overset{(i)}{\leq} \varepsilon\|A\|\|\mathbf{x}\|_2 \overset{(ii)}{=} \varepsilon\|A\|,$$

where (i) is by the definition of the operator norm, and (ii) is due to $\mathbf{x} \in \mathcal{S}^{n-1}$. Similarly, we can have

$$|\langle A(\mathbf{x} - \mathbf{y}), \mathbf{y}\rangle| \leq \|A(\mathbf{x} - \mathbf{y})\|_2\|\mathbf{y}\|_2 \leq \|A\|\|\mathbf{x} - \mathbf{y}\|_2 \leq \varepsilon\|A\|.$$

By triangular inequality, we obtain

$$|\langle A\mathbf{y}, \mathbf{y}\rangle| \geq |\langle A\mathbf{x}, \mathbf{x}\rangle| - |\langle A\mathbf{x}, \mathbf{x}\rangle - \langle A\mathbf{y}, \mathbf{y}\rangle| = \|A\| - |\langle A\mathbf{x}, \mathbf{x} - \mathbf{y}\rangle - \langle A(\mathbf{x} - \mathbf{y}), \mathbf{y}\rangle| \geq (1-2\varepsilon)\|A\|.$$

Dividing both sides by $1-2\varepsilon$ and $|\langle A\mathbf{y}, \mathbf{y}\rangle| \leq \sup_{\mathbf{x}\in\mathcal{N}} |\langle A\mathbf{x}, \mathbf{x}\rangle|$, we complete the proof.

$\square$

Lemma 4 (Covering numbers of the Euclidean ball, Corollary 4.2.13 in [1])**.** *The covering number of the unit Euclidean ball B_2^n satisfy the following for any $\varepsilon > 0$:*

$$\mathcal{N}\left(B_2^n, \varepsilon\right) \leq \left(\frac{2}{\varepsilon} + 1\right)^n.$$

The same bound is also true for the unit Euclidean sphere $\mathcal{S}^{n-1}$.

Proof. This is a corollary of Proposition `4.2.12` in [1] (also Lemma 10 in Lecture 18 `https://people.math.wisc.edu/~roch/hdps/roch-hdps-scribe18.pdf`). In fact, since the volume in $\mathbb{R}^n$ scales as $|cB_2^n| = c^n|B_2^n|$ for any $c > 0$, we have

$$\mathcal{N}\left(B_2^n, \varepsilon\right) \leq \frac{|(1+\varepsilon/2)B_2^n|}{|(\varepsilon/2)B_2^n|} = \frac{(1+\varepsilon/2)^n}{(\varepsilon/2)^n} = \left(\frac{2}{\varepsilon} + 1\right)^n.$$

The bound for the sphere $\mathcal{S}^{n-1}$ can be proved in the same way. $\square$

With these two lemmas above, we then come to prove Theorem 2.

Proof. By Lemma 1, it is sufficient to prove Inequality (2), i.e.

$$\|\frac{1}{m}A^\top A - I_n\| \leq K^2 \max(\delta, \delta^2),$$

where $\delta = C\left(\sqrt{\frac{n}{m}} + \frac{t}{\sqrt{m}}\right)$ and C is a constant to be determined later. By the definition of the operator norm, we know that

$$\left\|\frac{1}{m}A^\top A - I_n\right\| = \sup_{\mathbf{x}\in\mathcal{S}^{n-1}} \left|\left\langle \left(\frac{1}{m}A^\top A - I_n\right)\mathbf{x}, \mathbf{x}\right\rangle\right|.$$

Here we consider choosing an appropriate ε-net $\mathcal{N}$ of the unit sphere $\mathcal{S}^{n-1}$ and evaluate the operator norm on the $\mathcal{N}$. By Lemma 4, we can find a $\frac{1}{4}$-net $\mathcal{N}$ of $\mathcal{S}^{n-1}$ with cardinality $|\mathcal{N}| \leq 9^n$. Then by Lemma 3, with setting $\varepsilon = \frac{1}{4}$, we can obtain that

$$\left\|\frac{1}{m}A^\top A - I_n\right\| \leq 2\sup_{\mathbf{x}\in\mathcal{N}} \left|\left\langle \left(\frac{1}{m}A^\top A - I_n\right)\mathbf{x}, \mathbf{x}\right\rangle\right| = 2\sup_{\mathbf{x}\in\mathcal{N}}\left|\frac{1}{m}\|A\mathbf{x}\|_2^2 - 1\right|. \tag{3}$$

Note that for any $\mathbf{x} \in \mathcal{S}^{n-1}$, we can express $\|A\mathbf{x}\|_2^2$ as a sum of random variables:

$$\|A\mathbf{x}\|_2^2 = \sum_{i=1}^m \langle A_i, \mathbf{x}\rangle^2 =: \sum_{i=1}^m X_i^2.$$

Since A_i are independent, isotropic, and sub-gaussian random vectors with $\|A_i\|_{\psi_2} \leq K$, $X_i = \langle A_i, x\rangle$ here are also independent sub-gaussian random variables with $\mathbb{E}X_i^2 = 1$ and $\|X_i\|_{\psi_2} \leq K$. Therefore, we can construct independent, mean zero, and sub-exponential random variables $X_i^2 - 1$ such that

$$\left\|X_i^2 - 1\right\|_{\psi_1} \leq CK^2$$

for some constant $C > 0$, which can be verified as in the proof of Theorem `3.1.1` in [1] (also Theorem 3 in Lecture 8 `https://people.math.wisc.edu/~roch/hdps/roch-hdps-scribe8.pdf`).

So we can use Bernstein's inequality, and obtain that for $\varepsilon = K^2 \max\left(\delta, \delta^2\right) > 0$,

$$\begin{aligned}
\mathbb{P}\left\{\left|\frac{1}{m}\|A\mathbf{x}\|_2^2 - 1\right| \geq \frac{\varepsilon}{2}\right\} &= \mathbb{P}\left\{\left|\frac{1}{m}\sum_{i=1}^{m} X_i^2 - 1\right| \geq \frac{\varepsilon}{2}\right\} \\
&\leq 2\exp\left[-c\min\left(\frac{\varepsilon^2}{K^4}, \frac{\varepsilon}{K^2}\right) m\right] \\
&\overset{(i)}{=} 2\exp\left[-c\delta^2 m\right] \\
&\overset{(ii)}{\leq} 2\exp\left[-cC^2\left(n + t^2\right)\right],
\end{aligned} \tag{4}$$

where (i) is due to $\frac{\varepsilon}{K^2} = \max\left(\delta, \delta^2\right)$ and (ii) can be derived by using the fact that $(a+b)^2 \geq a^2 + b^2$ for $a, b \geq 0$.

Recalling that $\mathcal{N}$ has cardinality bounded by 9^n, then by Inequality (4) we have

$$\mathbb{P}\left\{\sup_{\mathbf{x}\in\mathcal{N}}\left|\frac{1}{m}\|A\mathbf{x}\|_2^2 - 1\right| \geq \frac{\varepsilon}{2}\right\} \leq 9^n \cdot 2\exp\left[-cC^2\left(n + t^2\right)\right] \overset{(i)}{\leq} 2\exp\left(-t^2\right), \tag{5}$$

where (i) can be obtained by choosing large enough $C > 0$. With combining Inequalities (5) and (3), we complete the proof. □

References

[1] Roman Vershynin, *High-dimensional probability: An introduction with applications in data science*, Cambridge University Press, 2018.