Thm 4.1 Subgaussian row

Let A be an $N \times n$ matrix whose rows A_i are independent subgaussian isotropic random vector in $\mathbb{R}^n$. Then for every $t \geq 0$, with probability $\geq 1 - 2\exp(-ct^2)$ one has

$$\sqrt{N} - C\sqrt{n} - t \leq s_{min}(A) \leq s_{max}(A) \leq \sqrt{N} + C\sqrt{n} + t$$

Here $C = C_K$, $c = c_K > 0$ depend only on the subgaussian norm $K = \max \|A_i\|_{\psi_2}$ of the row.

Pf: so far for x fixed,

$$P\left\{ \left| \frac{1}{N} \|Ax\|_2^2 - 1 \right| \geq \frac{\varepsilon}{2} \right\} \leq 2\exp\left(-\frac{c_1}{K^4}(C^2 n + t^2) \right)$$

Step 3 (uniform bound):

$$P\left\{ \max_{x \in \mathcal{N}} \left| \frac{1}{N} \|Ax\|_2^2 - 1 \right| \geq \frac{\varepsilon}{2} \right\}$$

$$\leq 9^n \, 2\exp\Big(\underbrace{-\frac{C_1}{K^4}}_{\text{should} \geq (2\log 9)} \big(C^2 n + t^2 \big) \Big)$$

for $C \geq K^2 \sqrt{\frac{2\log 9}{C_1}}$

$$\leq 2\exp\left(-\frac{C_1 t^2}{K^4} \right)$$

The result follow from the covering argument. (pick x to give s_{min}, s_{max})

Remark: If the A_i are not isotropic, say

$$\Sigma(A) = E A_i A_i^* = \Sigma_i$$

we can define $B_i = \Sigma^{-\frac{1}{2}} A_i$

Then B_i's are isotropic

Hence with probability $\geq 1 - 2\exp(-ct^2)$:

$$\left\| \frac{1}{N} A^* A - \Sigma \right\| = \left\| \Sigma^{\frac{1}{2}} \left(\frac{1}{N} B^* B - I \right) \Sigma^{\frac{1}{2}} \right\|$$

$$\leq \|\Sigma\| \left\| \frac{1}{N} B^* B - I \right\|$$

$$\leq \max(\delta, \delta^2) \|\Sigma\|$$

where $\delta = C\sqrt{\frac{n}{N}} + \frac{t}{\sqrt{N}}$, here C, and c depend on $\max_i \|B_i\|$.

1.30 [illegible]

oral: strategy, main idea

2.11, 18, 25

4.2 Heavy-tailed rows

We saw many distributions are subgaussian, but often their subgaussian norm depends heavily on the dimension.

We hence need a more general theorem

Thm 4.2 (Heavy-tailed rows)

Let A be an $N \times n$ matrix whose rows A_i are independent isotropic random vectors in $\mathbb{R}^n$. Let m be a number s.t. $\|A_i\|_2 \le \sqrt{m}$ a.s. for all i.

($E\|A_i\|_2 = \dim$, m depends on $\dim$)

Then for every $t \geq 0$ one has

$$\sqrt{N} - t\sqrt{m} \le s_{min}(A) \le s_{max}(A) \le \sqrt{N} + t\sqrt{m}$$

with prob $1 - 2n\exp(-c't^2)$.

formal: $\sqrt{N} + C\sqrt{n} \pm t$

now: $\sqrt{N} + \underline{t\sqrt{m}}$

Recall that $\left(E\|A_i\|_2^2\right)^{\frac{1}{2}} = \sqrt{n}$ by lemma 2.25

So the best bound one can hope for is $m = O(n)$.

Then one obtains

$$N - t\sqrt{n} \le s_{min}(A) \le s_{max}(A) \le \sqrt{N} + t\sqrt{n}$$

with prob $\ge 1 - 2\exp(c't^2)$

Note: - t multiplies $\sqrt{n}$

- additional factor of n in probability

} more next time

Pf: We use non-commutative Berstein (Thm 2.28)

Step 1: Reduction to a sum of inde random matrices

First note that $E\|A_i\|_i^2 = n \Rightarrow m \geq n \geq 1$

We again analyse the associated

<u>approximate isometry</u>

We need to show

classical step

$$\left\| \frac{1}{N} A^* A - I \right\| \leq \max(\delta, \delta^2) := \varepsilon$$

where $\delta = t\sqrt{\frac{m}{N}}$.

Now $\frac{1}{N} A^* A - I = \sum_{i=1}^{N} \underbrace{\left(\frac{1}{N}(A_i \otimes A_i - I)\right)}_{X_i}$

Note: X_i are independent and centered

$n \times n$ random matrices

Step 2: Estimating the mean, range and variance

mean: $EX_i = 0$

range: $\|X_i\| \le \frac{1}{N}(\|A_i \otimes A_i\| + 1)$

$$= \frac{1}{N}(\|A_i\|_2^2 + 1)$$

$$\le \frac{1}{N}(m+1) \quad \text{(by assumption)}$$

$$\le \frac{2m}{N} := K$$

Total variance, compute

$$X_i^2 = \frac{1}{N^2}\left((A_i \otimes A_i)^2 - 2(A_i \otimes A_i) + I\right)$$

$$EX_i^2 = \frac{1}{N^2}\left(E(A_i \otimes A_i)^2 - I\right)$$

$$(A_i \otimes A_i)^2 = A_i A_i^T A_i A_i^T = \|A_i\|_2^2 A_i \otimes A_i$$

$$\|EX_i^2\| \le \frac{1}{N^2}\left(\| E\|A_i\|_2^2 A_i \otimes A_i\| + 1\right)$$

$$\underset{\substack{A_i \otimes A_i \succeq 0 \\ \|A_i\|_2^2 \le m}}{\le} \frac{1}{N^2}\left(\|mE(A_i \otimes A_i)\| + 1\right)$$

$$= \frac{1}{N^2}(m+1)$$

$$\le \frac{2m}{N^2}$$

Thus

$$\left\|\sum_{i=1}^{N} EX_i^2\right\| \le N \max_i \|EX_i^2\| \le \frac{2m}{N} := \sigma^2$$

Not too loose
equality for i.i.d.

Step 3: Application of the non-commutative Berstein's inequality

$$P\left\{\left\|\frac{1}{N}A^*A - I\right\| \ge \varepsilon\right\} = P\left\{\left\|\sum_{i=1}^{N} X_i\right\| \ge \varepsilon\right\}$$

$$\le 2n \exp\left(-c\min\left(\frac{\varepsilon^2}{\sigma^2}, \frac{\varepsilon}{K}\right)\right)$$

$$\le 2n \exp\left(-c\min(\varepsilon^2, \varepsilon) \cdot \frac{N}{2m}\right)$$

$$\underset{\delta^2 = \min(\varepsilon^2, \varepsilon)}{\leq} 2n \exp\left(-\frac{c\delta^2 N}{2m}\right) \qquad \begin{cases} \delta > 1 & \varepsilon = \delta^2 \\ \delta < 1 & \varepsilon = \delta \end{cases}$$

$$\underset{\delta = t\sqrt{\frac{m}{N}}}{=} 2n \exp\left(-c\frac{t^2}{2}\right)$$

Difference between the result

- additional factor of n in prob
- bounded ass $\|A_i\| \leq \sqrt{m}$

These aspects are indeed necessary

Boundedness: why do we need the property?

Consider the following example:

Choose $\delta > 0$ arbitrary small

Let ξ be Bernoulli with $P(1) = \delta$, $P(0) = 1 - \delta$. Let $y \in \mathbb{R}^n$ be isotropic independent of ξ.

We study

$$X = \delta^{-\frac{1}{2}} \xi Y$$

$$E X \otimes X = E \delta^{-1} \xi^2 E\, Y \otimes Y$$

$$= I$$

Let $A \in \mathbb{R}^{N \times n}$ be a random matrix with independent copies X_i of X as its rows, where $N < \frac{n}{2\delta}$

Y is not zero and X is not zero

$$P(X \neq 0) \leq \delta$$

$$\Rightarrow P(\sigma_{min}(A) > 0)$$

$$\leq P(\exists \{i_1, \cdots, i_n\} : X_{i_j} \neq 0, \forall j = 1, \cdots, n\}$$

$$\leq \sum_{\substack{I \subseteq \{1, \cdots, N\} \\ |I| = n}} P(X_i \neq 0, i \in I)$$

$$\leq \binom{N}{n} \delta^n \sim \frac{\delta^n N^n}{n^n} < \left(\frac{1}{2}\right)^n$$

No non-trival bounds for σ_{min} are possible (too close to 0)

Additional factor in probability:

means that probability bound is only meaningful when $t \geq C\sqrt{\log n}$

So what Thm 4.2 says is that

with prob 0.9, say,

$$\sqrt{N} - C\sqrt{n\log n} \leq S_{min} \leq S_{max} \leq \sqrt{N} + C\sqrt{n\log n}$$

So there is an additional ~~logarithmic~~ factor, this factor is also necessary

→ next time