Theorem 4.4

$A \in \mathbb{R}^{N\times n}$ with independent isotropic rows,

$m := \mathbb{E} \max_{i\le N} \|A_i\|_2^2$. Then

$$\mathbb{E} \max_{j\le n} |s_j(A) - \sqrt{N}| \le C\sqrt{m \log \min(m, N)}$$

Lemma 4.5

Let X_i be a finite sequence of independent random vector in some Banach space and (ε_i) be i.i.d. symm Bernoulli, Then

$$\mathbb{E}\left\| \sum (X_i - \mathbb{E}X_i)\right\| \le 2\mathbb{E}\left\| \sum_i \varepsilon_i X_i \right\|$$

Lemma 4.6 Let z be a non-negative r.v.

Then $E|z^2-1| \geqslant \max\left(E|z-1|, (E|z-1|)^2\right)$

Pf of Thm 4.4

Step 1: we need to control

$$E := E\left\|\frac{1}{N}A^*A - I\right\|$$

$$= E\left\|\frac{1}{N}\sum_{i=1}^{N} A_i \otimes A_i - I\right\|$$

$$\underset{\substack{\uparrow \\ \text{lemma 4.5}}}{\leq} \frac{2}{N} E\left\|\sum_{i=1}^{N} \varepsilon_i A_i \otimes A_i\right\|$$

$$\overset{\text{Rudelson conditioning on } A}{\leq} \frac{2}{N} E\left(C\sqrt{\log\min(N,n)} \cdot \underbrace{\max_{i\leq N}\|A_i\|_2}_{E(\cdots)\leq m} \left\|\sum_{i=1}^{N} A_i\otimes A_i\right\|^{\frac{1}{2}}\right)$$

$$\overset{\text{Cauchy ineq}}{\leq} \tilde{C}\underbrace{\sqrt{\frac{m\log(\min(N,n))}{N}}}_{:=a}(E+1)^{\frac{1}{2}}$$

$$\Rightarrow E \leq \sqrt{a^2+\frac{a^4}{4}} + \frac{a^2}{2} \leq \max(3a, 9a^2).$$

the constants are not the best.

Step 2: Diagonalization

$$\left\|\frac{1}{N}A^*A-1\right\| \underset{\text{Diagonalization}}{\overset{\uparrow}{=\!=}} \left\|\frac{1}{N}U^*\begin{pmatrix} s_1^2(A) & & \\ & \ddots & \\ & & s_{\min(N,n)}^2(A)\end{pmatrix}U-\overset{U^*U}{I}\right\|$$

$$\overset{\text{unitary invariance}}{=} \left\|\frac{1}{N}\begin{pmatrix} s_1^2(A)-N & & \\ & \ddots & \\ & & s_{\min(N,n)}^2(A)-N\end{pmatrix}\right\|$$

$$= \max\left(\left|\frac{s_{\min}^2(A)}{N}-1\right|, \left|\frac{s_{\max}^2(A)}{N}-1\right|\right)$$

$$\max(3\delta, 9\delta^2) \geq E\left\|\frac{1}{N}A^*A-I\right\|$$

$$= E\max\left(\left|\frac{s_{\min}^2(A)}{N}-1\right|, \left|\frac{s_{\max}^2(A)}{N}-1\right|\right)$$

$$\geq \max\left\{E\left(\left|\frac{s_{\min}^2(A)}{N}-1\right|\right), E\left(\left|\frac{s_{\max}^2(A)}{N}-1\right|\right)\right\}$$

$$\overset{\text{Lemma 4.6}}{\geq} \max\left\{E\left|\frac{s_{\min}(A)}{\sqrt{N}}-1\right|, \left(E\left|\frac{s_{\min}(A)}{\sqrt{N}}-1\right|\right)^2, E\left|\frac{s_{\max}(A)}{\sqrt{N}}-1\right|, \left(E\left|\frac{s_{\max}(A)}{\sqrt{N}}-1\right|\right)^2\right\}$$

$$\Rightarrow 3\delta \geq \max\left(E\left|\frac{s_{\min}(A)}{\sqrt{N}}-1\right|, E\left|\frac{s_{\max}(A)}{\sqrt{N}}-1\right|\right)$$

Thus

$$\mathbb{E} \max_{j \le n} \left| \frac{s_j(A)}{\sqrt{N}} - 1 \right| = \mathbb{E} \max \left(\left| \frac{s_{min}(A)}{\sqrt{N}} - 1 \right|, \left| \frac{s_{max}(A)}{\sqrt{N}} - 1 \right| \right)$$

$$\le \mathbb{E} \left(\left| \frac{s_{min}(A)}{\sqrt{N}} - 1 \right| + \left| \frac{s_{max}(A)}{\sqrt{N}} - 1 \right| \right)$$

$$\le 2 \max (\dots)$$

$$\le 62$$

Multiplying by $\sqrt{N}$ complete the proof.

(We will omit heavy-tail columns, do sub-gau/exp columns)

5. Random matrices with independent columns

Results of section 4 can be applied to matrices with independent columns. (consider the transpose)

One obtains that with high probability

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

Lower bound is no longer meaningful for $n < N$.

We seek similar inequality with ____ roles of n and N.

Why does the previous approach fail?

The previous approach amounts to analyzing the spectrum of $AA^* \in \mathbb{R}^{N\times N}$.

If $N>n$, this matrix must be singular, so no sharp bounds can be expected.

Instead, one should again studying

$$\mathbb{R}^{n\times n} \ni \underbrace{A^*A}_{G} = \underbrace{\left(\langle A_j, A_k\rangle\right)_{j,k=1}}_{\text{Gram Mat}}$$

It follows from Lemma 2.33 that it sufficies to show that G is an approximate isometry.

For G, neither the row, nor the column are independent. However, in each of the entries two of the column enter in a (bi)linear ~~fashion~~ pattern.

11:0_ , 1.27 watch video

This can be seen an analogy to the chaos studied in Thm 2.11

→ Try a variant of decoupling (Lemma 2.11)

different for subgaussian and heavy-tailed.

5.1 Subgaussian columns

Theorem 5.1 (Subgaussian columns)

Let A be an $N \times n$ matrix ($N \geq n$) whose column A_i are independent subgaussian isotropic random vectors in $\mathbb{R}^N$ with $\|A_j\|_2 = \sqrt{N}$ a.s.

we need a more strict assumption here, in formal one we only need a bound

Then for every $t \geq 0$, one has

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

with probability $\geq 1 - 2\exp(-ct^2)$, where $C = C_K'$, $c = c_K' > 0$ depend only on the subgaussian norm bound $K := \max_j \|A_j\|_{\psi_2}$ of the columns.

Difference to Thm 4.1

Normalization of the columns (necessarity as we will see)

Remark: By Lemma 2.33, the conclusion of the Theorem is equiv to $\|\frac{1}{N}A^*A - I\| \leq C\sqrt{\frac{n}{N}} + \frac{t}{\sqrt{N}}$.

Hence $\|\frac{1}{N}G - I\| = O(\sqrt{\frac{n}{N}})$ w. high prob, the normalized Gram matrix is an approximate isometry.

For the proof, we decoupling for random subsets.

Definition 5.2 (Random Subsets)

can be arbitrary

Let $n \in \mathbb{N}$, $m \in \{0, \cdots, n\} := [n]$.

Consider independent $\{0,1\}$-valued random variables $\delta_1, \cdots, \delta_n$ with $E(\delta_i) = \frac{m}{n}$, we call that these are indepent selection.

Then we define a random subset of $[n]$ of average size m to the set-valued r.v. given by $T = \{i \in [n] : \delta_i = 1\}$.

Note that $E|T| = E\sum_{j=1}^{n} \delta_j = m$.

Lemma 5.3 (Decoupling with random subsets)

Consider a double array of real numbers $(a_{ij})_{i=1}^{n}$ s.t. $a_{ii} = 0$ for all i. Then

$$\sum_{i,j \in [n]} a_{ij} = 4E \sum_{i \in T, j \in T^c} a_{ij},$$

where T is a random subset of $\{1, \cdots, n\}$ with average size $\frac{n}{2}$.

In particular,

$$4 \min_{T\subseteq[n]} \sum_{\substack{i\in T,\\ j\in T^c}} a_{ij} \leq \sum_{i,j\in[n]} \leq 4 \max_{T\subseteq[n]} \sum_{i\in T} \sum_{j\in T^c} a_{ij}$$

where the min and max are taken over all $T\subseteq[n]$.

Pf: Using the definition of a random subset, we write

$T=\{i\in[n], \delta_i=1\}$ where the δ_i are independent, sections with

$E\,\delta_i = \frac{1}{2}$.

We obtain

$$E\sum_{i\in T}\sum_{j\in T^c} a_{ij} = E\sum_{i,j\in\{1,\dots,n\}} \delta_i(1-\delta_j)\,a_{ij} = \frac{1}{4}\sum_{i,j\in[n]} a_{ij}$$

The second part follows by estimating E by max and min.

Proof of Thm 5.1:

Step 1: w.l.o.g. assume that the columns A_i have zero mean.

Indeed, multipling all of the columns A_i by independent symmetric Bernoulli random variables ε_i keeps zero mean while not changing the singular of the matrix.