Sharp Matrix Concentration

March Boedihardjo (UC Irvine)
Joint work with Afonso Bandeira and Ramon van Handel

Deep Learning Theory Symposium, 6th December 2021

Supported by NSF DMS-1856221

Problem

Given a random matrix X, how large is the expected spectral norm $\mathbb{E}\|X\|$?

Matrix Khintchine: If X is a $d \times d$ Hermitian matrix with jointly Gaussian entries and $\mathbb{E}X = 0$, then

$$\|\mathbb{E}X^2\|^{\frac{1}{2}} \lesssim \mathbb{E}\|X\| \lesssim \sqrt{\log d}\|\mathbb{E}X^2\|^{\frac{1}{2}}.$$

Matrix Bernstein: If $X = \sum_{i=1}^n Y_i$ and $Y_1, \ldots, Y_n$ are independent $d \times d$ Hermitian matrix with $\mathbb{E}Y_i = 0$, then

$$\mathbb{E}\|X\| \lesssim \sqrt{\log d}\|\mathbb{E}X^2\|^{\frac{1}{2}} + (\log d)(\mathbb{E}\max_i \|Y_i\|^2)^{\frac{1}{2}}.$$

Question: These inequalities are sharp up a log factor, but when can the log factor be removed?

When is the log factor necessary?

(1) If $X = \frac{1}{\sqrt{d}} \begin{bmatrix} g_{1,1} & \cdots & g_{1,d} \\ \vdots & \ddots & \vdots \\ g_{d,1} & \cdots & g_{d,d} \end{bmatrix}$ is symmetric with i.i.d. $N(0,1)$ entries, then $\mathbb{E}X^2 = I$ and $\mathbb{E}\|X\| \approx 2$.

So log factor is not needed.

(2) If $X = \begin{bmatrix} g_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & g_d \end{bmatrix}$ has i.i.d. $N(0,1)$ diagonal entries, then $\mathbb{E}X^2 = I$ and $\mathbb{E}\|X\| \approx \sqrt{2\log d}$.

So log factor is needed.

When is the log factor necessary?

(1) If $X = \frac{1}{\sqrt{d}} \begin{bmatrix} g_{1,1} & \cdots & g_{1,d} \\ \vdots & \ddots & \vdots \\ g_{d,1} & \cdots & g_{d,d} \end{bmatrix}$ is symmetric with i.i.d. $N(0,1)$ entries, then $\mathbb{E}X^2 = I$ and $\mathbb{E}\|X\| \approx 2$.

So log factor is not needed. **(Very noncommutative)**

(2) If $X = \begin{bmatrix} g_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & g_d \end{bmatrix}$ has i.i.d. $N(0,1)$ diagonal entries, then $\mathbb{E}X^2 = I$ and $\mathbb{E}\|X\| \approx \sqrt{2 \log d}$.

So log factor is needed. **(Commutative)**

This suggests that if there is enough noncommutativity then $\mathbb{E}\|X\| \sim \|\mathbb{E}X^2\|^{\frac{1}{2}}$ (i.e., log factor not needed).

How to quantify noncommutativity?

Tropp introduced the following quantity

$$w(X) = \sup_{Q_1, Q_2, Q_3} \|\mathbb{E} X Q_1 Y Q_2 X Q_3 Y\|^{\frac{1}{4}},$$

where the sup is over all unitary Q_1, Q_2, Q_3 and Y is an independent copy of X.

Tropp (2015): If X is a $d \times d$ Hermitian matrix with jointly Gaussian entries and $\mathbb{E}X = 0$, then

$$\mathbb{E}\|X\| \lesssim (\log d)^{\frac{1}{4}} \|\mathbb{E}X^2\|^{\frac{1}{2}} + \sqrt{\log d}\, w(X).$$

- Not sharp due to the $(\log d)^{\frac{1}{4}}$ factor in the first term.
- $w(X)$ is very hard to compute.

How to quantify noncommutativity?

Suppose X is a $d \times d$ Hermitian matrix with jointly Gaussian entries and $\mathbb{E}X = 0$.

Write $X = \sum_{k=1}^{n} g_k A_k$ for some i.i.d. $N(0,1)$ random variables $g_1, \ldots, g_n$ and Hermitian $A_1, \ldots, A_n$.

View X as a Gaussian vector in $\mathbb{R}^{d^2}$. Define $v(X) = \|\mathrm{Cov}(X)\|^{\frac{1}{2}}$.
• $v(X)$ is easy to compute.
• If $A_1, \ldots, A_n$ commute, they can be simultaneously diagonalized so

$$X = \begin{bmatrix} X_{1,1} & & \\ & \ddots & \\ & & X_{d,d} \end{bmatrix}.$$

We have $v(X) \geq \max_i \sqrt{\mathrm{Var}(X_{i,i})} = \|\mathbb{E}(X^2)\|^{\frac{1}{2}}$.
If $v(X) \ll \|\mathbb{E}(X^2)\|^{\frac{1}{2}}$ then $A_1, \ldots, A_n$ are very noncommutative.

Main results

Gaussian case:
If X is a $d \times d$ Hermitian matrix with jointly Gaussian entries and $\mathbb{E}X = 0$, then

$$\mathbb{E}\|X\| \lesssim \|\mathbb{E}X^2\|^{\frac{1}{2}} + (\log d)^{\frac{3}{2}} \|\mathrm{Cov}(X)\|^{\frac{1}{2}}.$$

General:
Let $Z_1, \dots, Z_n$ be independent $d \times d$ matrices with $\mathbb{E}Z_i = 0$. Let $X = \sum_{i=1}^{n} Z_i$. Then

$$\mathbb{E}\|X\| \lesssim \|\mathbb{E}X^*X\|^{\frac{1}{2}} + \|\mathbb{E}XX^*\|^{\frac{1}{2}} + \|\mathrm{Cov}(X)\|^{\frac{1}{2}} (\log d)^{\frac{3}{2}} + (\mathbb{E}[\max_i \|Z_i\|_{\mathrm{F}}^2])^{\frac{1}{2}} (\log d)^2.$$