Matrix Bernstein $\neq$

Recall: (scalar) Bernstein $\neq$ for bounded r.v.'s:

<u>Thm</u> (<u>Bernstein $\neq$</u>) Let $X_1, \dots X_n$ be ind., zero-mean, r.v.'s. Suppose:

$$|X_i| \leq K \text{ a.s.}, \qquad \sum_{i=1}^n \mathbb{E} X_i^2 \leq \sigma^2.$$

Then,

$$P\left(\left|\sum_{i=1}^n X_i\right| > t\right) \leq 2\exp\left(\frac{-\frac{1}{2}t^2}{\sigma^2 + \frac{Kt}{3}}\right)$$

$$\sim 2\exp\left(-c\min\left(\underbrace{\frac{t^2}{\sigma^2}}_{\text{sub-Gauss.}}, \underbrace{\frac{t}{K}}_{\text{sub-exp.}}\right)\right).$$

(We proved a slightly diff. version.)

Similar result holds for sums of random matrices:

<u>Thm</u> (<u>Matrix Bernstein $\neq$</u>) Let $X_1, \dots, X_m$ be ind. self adjoint, mean-zero, $n \times n$ random matrices. Suppose:

$$\max_i \|X_i\| \leq K \text{ a.s.}, \quad \left\|\sum_{i=1}^n \mathbb{E} X_i^2\right\| \leq \sigma^2.$$

Then, $$P\left(\left\|\sum_{i=1}^m X_i\right\| > t\right) \leq 2n\exp\left(\frac{-t^2/2}{\sigma^2 + Kt/3}\right)$$

<u>Remarks:</u>

1. Spectral norm, $\|A\| = \sigma_1(A) = \max_{i=1,\dots,n} |\lambda_i(A)|$

   $=$ largest eigenvalue since A is self-adjoint.

2. Matrix version recovers scalar version when $n=1$.

3. If X_i are not self adjoint, replace them w/ $\begin{bmatrix} 0 & X_i \\ X_i^T & 0 \end{bmatrix}$.

We will prove thm using MGF as before. Set

$S_m = \sum_{i=1}^{m} X_i$.

$$P(\underset{\uparrow\ \text{matrix}}{S_m} > \underset{\text{scalar}}{t}) \overset{?}{\leq} P(\mathbb{E}\, e^{\lambda S_m} > e^{\lambda t})$$

<u>self adjoint matrix relations:</u>

① <u>Partial ordering:</u>

- $A \leq B \overset{\text{def}}{\iff} B - A \succeq 0$ i.e. $B-A$ is PSD

i.e. all eigenvals of $B-A$ are ≥ 0.

② <u>Exponent:</u> $e^A := \sum_{k=0}^{\infty} \frac{A^k}{k!} = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{i=1}^{n} \lambda_i^k u_i u_i^* = \sum_{i=1}^{n} e^{\lambda_i} u_i u_i^*$

eigen-vals   eigen-vecs

(exponentiate eigenvalues)

③ Unfortunately $e^{A+B} \neq e^A e^B$.

But, $\boxed{tr(e^{A+B}) \leq tr(e^A e^B)}$

"Golden Thompson ≠" See T. Tao "What's new" blog post for proof.

<u>Proof</u> (of matrix Bernstein ≠)

$$\|S_m\| \leq t \iff -t I_n \leq S_m \leq t \underset{\text{identity}}{I_n} \quad (\text{All eig-vals} \in [-t,t])$$

$$\iff e^{-\lambda t I_n} \overset{*}{\leq} e^{\lambda S_m} \overset{**}{\leq} e^{\lambda t I_n} \quad (\lambda > 0)$$

We show ** holds w.h.p. (Replace S_m w/ $-S_m$ to control *.)

(**)

$$P(e^{\lambda S_m} \nleq e^{\lambda t I}) \leq P(tr(e^{\lambda S_m}) \geq e^{\lambda t}) \leq \frac{\mathbb{E}\, tr(e^{\lambda S_m})}{e^{\lambda t}}$$

"∃ an eig-val. $> e^{\lambda t}$"   markov ≠

"sum of eig.vals. $> e^{\lambda t}$"

It suffices to bound $\mathbb{E}\, tr(e^{\lambda S_m})$.

$$\mathbb{E}\, tr(e^{\lambda S_m}) = \mathbb{E}\, tr\left(e^{\lambda X_m + \lambda S_{m-1}}\right)$$

$$\leq \mathbb{E}\, tr\left(e^{\lambda X_m} e^{\lambda S_{m-1}}\right) \qquad \text{(By Golden Thompson)}$$

$$= tr\left(\mathbb{E}\, e^{\lambda X_m} \cdot \mathbb{E}\, e^{\lambda S_{m-1}}\right)$$

$$\leq \|\mathbb{E}\, e^{\lambda X_m}\| \cdot tr\left(\mathbb{E}\, e^{\lambda S_{m-1}}\right) \qquad \text{(\underline{Exercize})}$$

$$\leq \dots \leq \qquad \text{(Continue by induction)}$$

$$\prod_{i=1}^{m} \|\mathbb{E}\, e^{\lambda X_i}\| \quad \underbrace{tr\, \underbrace{\mathbb{E}\, e^{0}}_{I_n}}_{n}$$

$$= n \prod_{i=1}^{m} \|\mathbb{E}\, e^{\lambda X_i}\|$$

Now bound $\|\mathbb{E}\, e^{\lambda X_i}\|$. Assume $k=1$, i.e., $\|X_i\| \leq 1$.

Enforce $0 < \lambda \leq 1$. $\Rightarrow \|\lambda X_i\| \leq 1$. Let $Z := \lambda X_i$

Then,

$$\mathbb{E}\, e^{Z} \leq \mathbb{E}\left(I_n + Z + Z^2\right) \qquad \text{Taylor series}$$

$$= I_n + Z^2$$

$$\leq e^{\mathbb{E} Z^2}$$

\underline{Exercize} $\boxed{\text{Justify above steps}}$

$$\Rightarrow \mathbb{E}\, e^{\lambda X_i} \leq e^{\lambda^2 \mathbb{E} X_i^2}$$

$$\Rightarrow \|\mathbb{E}\, e^{\lambda X_i}\| \leq e^{\lambda^2 \|\mathbb{E} X_i^2\|}$$

We have shown that

$$P(S_n \not\preceq t I_n) \leq n e^{-\lambda t} \cdot \exp \lambda^2 \sum_{i=1}^{m} \|\mathbb{E} X_i^2\|$$

Set $\sigma^2 = \sum_{i=1}^{m} \|\mathbb{E} X_i^2\|$. Then,

$$P(S_n \not\preceq t I_n) \leq n \exp\left(-\lambda t + \lambda^2 \sigma^2\right)$$

Optimize over $\lambda \in (0,1]$. QED.

We have proven slightly weaker version w/

$$\sigma^2 = \sum_{i=1}^{n} \| \mathbb{E} X_i^2 \| \quad \text{not} \quad \sigma^2 = \| \sum_{i=1}^{n} \mathbb{E} X_i^2 \|$$

[Ahlswede-Winter] [R. Oliveira], [Tropp]

Literature: [Joel Tropp, "User-friendly..."]