LECTURE 22

- For matrices, $e^{A+B} \neq e^A e^B$ in general

Proxy:

- <u>Thm</u> (Lieb's trace inequality)

Let H be a $d \times d$ symmetric matrix. Then the function

$$f(X) := \operatorname{tr}\exp(H + \log X)$$

is <u>concave</u> on $\{X : X \succcurlyeq 0\}$.

- For scalars (1×1 matrices), $f(x) = e^H \cdot x$ is linear.

- If $X \succcurlyeq 0$ is a random matrix, Lieb + Jensen $\Rightarrow$

$$\mathbb{E} f(X) \le f(\mathbb{E}X), \qquad \text{i.e.}$$

$$\mathbb{E}\operatorname{tr}\exp(H + \underbrace{\log X}_{= Z}) \le \operatorname{tr}\exp(H + \log \mathbb{E}\underbrace{X}_{= e^Z}), \qquad \text{i.e.}$$

$$\mathbb{E}\operatorname{tr}\exp(H+Z) \le \operatorname{tr}\exp(H + \log \mathbb{E}e^Z) \qquad \text{for } \forall \text{ random matrix } Z$$

<u>Cor</u> If X_i are independent r.v.'s, then

$$\mathbb{E}\operatorname{tr}\exp\Big(\underbrace{\sum_{i=1}^n X_i}_{= S_n}\Big) \le \operatorname{tr}\exp\Big(\sum_{i=1}^n \log \mathbb{E}e^{X_i}\Big)$$

$\leftarrow$ Proxy of $e^{\Sigma X_i} = \Pi e^{X_i}$

$$\| $$

$$\mathbb{E}_{1,\dots,n-1}\mathbb{E}_n \operatorname{tr}\exp(\underbrace{S_n}_{= H} + \underbrace{X_n}_{= Z}) \le \mathbb{E}_{1,\dots,n-1}\operatorname{tr}\exp(S_n + \log \mathbb{E}_n e^{X_n})$$

$$\dots \le \text{RHS}.$$

keep chopping off one term at a time. $\square$

THM [Matrix Hoeffding Inequality: Tropp'2010] let $\varepsilon_i = \pm 1$ with prob. $\frac{1}{2}$ independently, let A_i be $d \times d$ symmetric random matrices. Then

$$\mathbb{P}\Big\{ \big\| \underbrace{\sum_{i=1}^n \varepsilon_i A_i}_{=S_n} \big\| \geq t \Big\} \leq 2d \cdot \exp\left(-\frac{t^2}{2\sigma^2}\right) \quad \forall t \geq 0$$

where $\sigma^2 = \left\| \sum_{i=1}^n A_i^2 \right\|$

operator norm

the only cost of upgrade to matrices

Proof. • $\|S_n\| \overset{\text{by def. of operator norm}}{=} \max_i |\lambda_i(S)| = \max(\lambda_1(S), \lambda_1(-S))$.

Let $\lambda \geq 0$ be a parameter (its value will be optimized later).

• $$\mathbb{P}\{\lambda_1(S_n) \geq t\} = \mathbb{P}\{e^{\lambda \cdot \lambda_1(S_n)} \geq e^{\lambda t}\} \leq e^{-\lambda t}\, \mathbb{E}\underbrace{e^{\lambda \cdot \lambda_1(S_n)}}_{= \lambda_1(e^{\lambda S_n})} \quad \text{(Markov)}$$

(def of function of a matrix)

one eigenvalue $\lambda_1 \leq$ sum of all eigenvalues

$$\leq e^{-\lambda t}\, \mathbb{E}\operatorname{tr} e^{\lambda S_n} \underset{\text{Lieb}}{\leq} e^{-\lambda t} \operatorname{tr}\exp\Big(\sum_{i=1}^n \log \underbrace{\mathbb{E} e^{\lambda \varepsilon_i A_i}}_{\| \; \frac{e^{\lambda A_i} + e^{-\lambda A_i}}{2}}\Big)$$

scalar inequality in 1 var extends to matrices (last class)

$$\frac{e^{\lambda A_i} + e^{-\lambda A_i}}{2} \preceq e^{\lambda^2 A_i^2/2}$$

• Take $\log(\cdot)$ on both sides (matrix monotone, last class)

$$\Rightarrow \quad \log \mathbb{E} e^{\lambda \varepsilon_i A_i} \preceq \frac{\lambda^2}{2} A_i^2$$

• Sum up $\Rightarrow$

$$\sum_{i=1}^n \log \mathbb{E} e^{\lambda \varepsilon_i A_i} \preceq \frac{\lambda^2}{2} \boxed{\sum_{i=1}^n A_i^2} =: Z$$

• Take $\operatorname{tr}\exp(\cdot)$ on both sides. Trace monotonicity (last class) $\Rightarrow$

$$\operatorname{tr}\exp\Big(\sum_1^n \log \mathbb{E} e^{\lambda \varepsilon_i A_i}\Big) \leq \operatorname{tr}\exp\left(\frac{\lambda^2 Z}{2}\right) = \sum_{i=1}^d \lambda_i\left(\exp\left(\frac{\lambda^2 Z}{2}\right)\right)$$

(def of e^x)

$$= \sum_{i=1}^d \exp\left(\frac{\lambda^2 \cdot \lambda_i(Z)}{2}\right) \leq d \cdot \exp\left(\frac{\lambda^2 \sigma^2}{2}\right).$$

$\lambda_i(Z) \leq \|Z\| = \sigma^2$

• $\Rightarrow$

$$\mathbb{P}\{\lambda_1(S_n) \geq t\} \leq e^{-\lambda t} \cdot d \cdot \exp\left(\frac{\lambda^2\sigma^2}{2}\right) = d \cdot \exp\left(-\lambda t + \frac{\lambda^2 \sigma^2}{2}\right)$$

Optimize λ (set $\lambda := t/\sigma^2$)

$$\leq d \cdot \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

Q.E.D.

- The cost of the upgrade $= d$. Is it big? Curse of H.D.?

$$\mathbb{P}\{\|S_n\| \geq t\} \leq 2\exp\left(\log d - \frac{t^2}{2\sigma^2}\right) \leq 0.01$$

if we choose $t = 10\sigma\sqrt{\log d}$

$$\Rightarrow \quad \boxed{\|S_n\| \leq 10\,\sigma\sqrt{\log d} \quad \text{with probability} \geq 0.99}$$

the cost is logarithmic ☺

- Similarly, we can upgrade Bernstein's inequality. **(HW)**

Matrix Bernstein Inequality (MBI)

Let X_i be independent, mean zero, symmetric $d \times d$ random matrices. Assume $\|X_i\| \leq 1$ with prob. 1. Then $\forall t \geq 0$:

$$\mathbb{P}\left\{\left\|\sum_{i=1}^{n} X_i\right\| \geq t\right\} \leq 2d \cdot \exp\left[-c\left(\frac{t^2}{\sigma^2} \wedge \frac{t}{K}\right)\right]$$

minimum

where $\sigma^2 = \left\|\sum_{i=1}^{n} \mathbb{E}X_i^2\right\|$.

"matrix variance"

absolute constant > 0

- Cost of upgrade? $\quad \text{RHS} = 2\exp\left[\log d - \left(\frac{ct^2}{\sigma^2} \wedge \frac{ct}{K}\right)\right] \leq 0.01$

if we choose $t = C\sigma\sqrt{\log d} \vee CK\log d$, with a sufficiently large absolute constant C.

$$\Rightarrow \quad \boxed{\left\|\sum_{1}^{n} X_i\right\| \leq C\sigma\sqrt{\log d} + CK\log d \quad \text{with high probability}}$$ ☺