LECTURE 29

Last class: Pretend that matrices are scalars

	Scalars	Matrices
1.	$a \in \mathbb{R}$	A : $d \times d$ symmetric
2.	$+, -$	$+, -$
3.	$\times, \div$	$\times$ (matrix product: NON-commutative), $A/B = B^{-1/2} A B^{-1/2}$
4.	$0, 1$	$0, I_d$
5.	$\|a\|$	$\|A\|$ (operator norm)
6.	Functions $f(a)$, $f: \mathbb{R} \to \mathbb{R}$	• If $A = \sum_{i=1}^{d} \lambda_i u_i u_i^T \Rightarrow f(A) \stackrel{def}{=} \sum_{i=1}^{d} f(\lambda_i) u_i u_i^T$ • Thus, $\boxed{\lambda_i(f(A)) = f(\lambda_i(A))}$.
7.	Order $\leq$	• Löwner order: $A \succeq 0$ if A is PSD. $A \preceq B$ if $B - A \succeq 0$.

$f(x) \leq g(x)\ \forall x \in \mathbb{R} \Longrightarrow f(X) \preceq g(X)\ \forall X$ matrix

8.	Monotonicity of f, $f: \mathbb{R} \to \mathbb{R}$ increasing $\not\Rightarrow$	Matrix monotonicity: "$X \preceq Y \Rightarrow f(X) \preceq f(Y)$" true for $f(x) = \log x$ and for $\mathrm{tr} f(X)$.
9.	Random var's	Random matrices. ☺

• **Matrix** versions of Hoeffding, Chernoff, Bernstein?

Yes: [Tropp '2010]

Recall scalar H.I from lecture 6:

THM (Hoeffding ineq.)

Let $\varepsilon_i = \pm 1$ with probability $1/2$ independently; $a_i \in \mathbb{R}$. Then

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

subgaussian tail

where $\sigma^2 = \operatorname{Var}(S_n) = \sum_{i=1}^n a_i^2$

Recall proof, based on MGF:

$$\mathbb{P}\{S_n \geq t\} = \mathbb{P}\{e^{\lambda S_n} \geq e^{\lambda t}\} \qquad \forall \lambda \geq 0$$

$$\leq \frac{\mathbb{E} e^{\lambda S_n}}{e^{\lambda t}} \quad \text{(Markov's inequality).} \qquad e^{\lambda S_n} = \prod_{i=1}^n e^{\lambda \varepsilon_i a_i} \Rightarrow$$

$$\leq e^{-\lambda t} \prod_{i=1}^n \underbrace{\mathbb{E} e^{\lambda \varepsilon_i a_i}}_{\displaystyle \| \atop \displaystyle \frac{1}{2}e^{\lambda a_i} + \frac{1}{2}e^{-\lambda a_i} \leq e^{\lambda^2 a_i^2/2}} \quad \text{(independence)}$$

since $\boxed{\dfrac{e^x + e^{-x}}{2} \leq e^{x^2/2} \quad \forall x}$

$$\leq \exp\left(-\lambda t + \sum_{i=1}^n \lambda^2 a_i^2/2\right) = \exp\left(-\lambda t + \lambda^2\sigma^2/2\right)$$

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

$$\leq \exp(-t^2/2\sigma^2).$$

QED

(7) p.1

$$\boxed{\frac{e^X + e^{-X}}{2} \preceq e^{X^2/2}} \quad (*)$$

- Matrix version? $\leq \mapsto \preceq$; $a_i \longmapsto A_i$?
- $e^{A+B} \neq e^A e^B$:(

Instead:

- **Thm** (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 concave on $\{X : X \succeq 0\}$.

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

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

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

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

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

Cor If X_i are independent r.v.'s, then

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

$$\|$$

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

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

keep chopping off one term at a time.

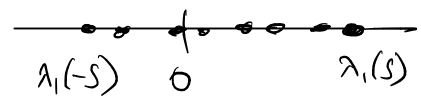
THM [Matrix Hoeffding Inequality: Tropp'2010] Let $\varepsilon_i = \pm 1$ with prob. $\frac{1}{2}$ independently, let A_i be $d \times d$ symmetric random matrices. Then

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

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

operator norm

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



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

• $$\mathbb{P}\{\lambda_1(S_n) \geq t\} = \mathbb{P}\{ e^{\lambda \cdot \lambda_1(S_n)} \geq e^{\lambda t} \} \leq e^{-\lambda t}\, \mathbb{E} \underbrace{e^{\lambda \cdot \lambda_1(S_n)}}_{= \lambda_1(e^{\lambda S_n})\ (6\,\text{p.}1)} \quad \text{(Markov)}$$

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

$$\leq e^{-\lambda t}\, \mathbb{E}\operatorname{tr} e^{\lambda S_n} \leq e^{-\lambda t} \operatorname{tr}\exp\Big(\sum_{i=1}^{n} \log \underbrace{\mathbb{E} e^{\lambda \varepsilon_i A_i}}_{= \frac{e^{\lambda A_i} + e^{-\lambda A_i}}{2} \preceq e^{\lambda^2 A_i^2/2}\ (*\text{p.}2)} \Big)$$

• Take $\log()$ on both sides (matrix monotone, 8 p.1)

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

• 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. Matrix monotone (8 p.1) $\Rightarrow$

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

$$\overset{(6\,\text{p.}1)}{=} \sum_{i=1}^{d} \exp\left(\frac{\lambda^2 \cdot \lambda_i(Z)}{2}\right) \leq d \cdot \exp\left(\frac{\lambda^2 \sigma^2}{2}\right). \qquad \lambda_i(Z) \leq \|Z\| = \sigma^2$$

• Substitute into (*) p. 3 $\Rightarrow$

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

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

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

Q.E.D.