39.1 Matrix Bernstein's Inequality

h) Lieb's concavity theorem. Let $H = H^T \in \mathbb{R}^{d\times d}$. Then

$$f(A) = \mathrm{tr}(e^{H+\log A})$$

is a concave function on the set of positive definite matrices, sometimes called the noncommutative Bernstein's inequality.

h') Golden-Thompson inequality. Let A, B be symmetric $d \times d$ matrices. Then $\mathrm{tr} e^{A+B} \leq \mathrm{tr}(e^A e^B)$.

Let $X_1, \dots, X_n$ be independent random matrices such that $||X_j|| \leq M$ with probability 1, and let $\sigma^2 = \left|\left|\sum_{j=1}^n \mathrm{Var}(X_j)\right|\right|$. Then

$$P\left(\left|\left|\sum_{j=1}^n X_j - \mathbb{E}(\sum X_j)\right|\right| > t\right) \leq 2d\exp\left(\frac{-t^2/2}{\sigma^2 + Mt/3}\right).$$

Assume WLOG that $\mathbb{E}X_j = 0$. Then

$$\begin{aligned}\left|\left|\sum X_j\right|\right| &= \sup_{||u||=1} \left\langle \left(\sum X_j\right) u, u\right\rangle \\ &= \sup_{||u||=1} \langle \sum X_i, uu^T\rangle =: \sup_{||u||=1} |Y_n(u)|\end{aligned}$$

Classical concentration inequalities describe the fluctuations of $\sup Y_n(u)$ about its mean. In other words, we would be able to control

$$P\left(|\left|\left|\sum(X_j - \mathbb{E}X_j)\right|\right| - \mathbb{E}\left|\left|\sum(X_j - \mathbb{E}X_j)\right|\right| | \geq t\right).$$

Given $A = A^T \in \mathbb{R}^{d\times d}$, the singular values are the absolute values of the eigenvalues. $-\lambda_{min}(A) = \lambda_{max}(-A)$. Let $\theta > 0$ (to be fixed later). Then

$$\begin{aligned}P(\lambda_{max}(\sum X_j) \geq t) &= P(\lambda_{max}(\theta \sum X_j) \geq \theta t) \\ &= P(e^{\lambda_{max}\theta \sum X_j} \geq e^{\theta t}) = P(\lambda_{max} e^{\theta \sum X_j} \geq e^{\theta t}).\end{aligned}$$

Hence $\lambda_{max} e^{\theta \sum X_j} \leq tr e^{\theta \sum X_j}$, and

$$P(\lambda_{max}(\sum X_j) \geq t) \leq P(\mathrm{tr}\, e^{\theta \sum X_j} \geq e^{\theta t}) \leq \frac{\mathbb{E}\,\mathrm{tr}\, e^{\theta \sum X_j}}{e^{-\theta t}}.$$

True for all θ, so we can take infimum over θ. We can define

$$F(\theta) = \operatorname{tr} e^{\theta \sum X_j}$$

the trace MGF. Note that

$$\mathbb{E} \operatorname{tr} e^{\theta \sum X_j} = \mathbb{E}\mathbb{E} \left(\operatorname{tr} e^{\theta \sum^{n-1} X_j + \theta X_n} | X_1, \ldots, X_{n-1} \right).$$

Let $H = \theta \sum^{n-1} X_j$. By Lieb's and Jensen's for concave functions, we get

$$\begin{aligned} \mathbb{E}(\operatorname{tr} e^{\theta \sum X_j + \theta X_n} | X_1, \ldots, X_n) &= \mathbb{E}(\operatorname{tr} e^{\theta \sum X_j + \log e^{\theta X_n}}) \\ &\leq \mathbb{E} \operatorname{tr} e^{\theta X_j + \theta \log \mathbb{E} e^{\theta X_n}} \ldots \\ &\leq \operatorname{tr} e^{\sum \log \mathbb{E} e^{\theta X_j}} \\ &\leq d \left|\left| e^{\sum \log \mathbb{E} e^{\theta X_j}} \right|\right|. \end{aligned}$$

Now we bound $\log \mathbb{E} e^{\theta X_i}$.

$$\mathbb{E} e^{\theta X_1} = I + \theta^2 \frac{\mathbb{E} X_1^2}{2} + \sum_{j \geq 2} \frac{\theta^j X_1^j}{j!}.$$

Because X_1 is symmetric,

$$X_1^k = X_1 X_1^{k-2} X_1 \leq X_1 M^{k-2} X_1 = M^{k-2} X_1^2.$$

Taking expectations,

$$\mathbb{E} e^{\theta X_1} \leq I + \frac{\theta^2}{2} \mathbb{E} X_1^2 + \frac{\theta^2 E X_1^2}{M^2} \sum \frac{M^j \theta^j}{j!} = I + \frac{\theta^2}{2} \mathbb{E} X_1^2 + \frac{\mathbb{E} X_1^2}{M^2} \left(e^{\theta M} - 1 - \theta M \right).$$