A tail-bound for sums of independent positive semidefinite random matrices

Andreas Maurer
am@andreas-maurer.eu

July 8, 2011

In this note I prove the following one-sided Berstein-type inequality for sums of independent positive semidefinite $N \times N$-matrices:

Theorem 1 *Let $X_1, ..., X_n$ be independent random $N \times N$-matrices with $X_i \succeq 0$ and $t > 0$. Then*

$$\Pr\left\{\lambda_{\max}\left(\sum_i \left(E\left[X_i\right] - X_i\right)\right) > t\right\} \leq N \ \exp\left(\frac{-t^2}{2\lambda_{\max}\left(\sum_i E\left[X_i^2\right]\right)}\right).$$

This is the noncommutative version of an inequality in [1], where it is argued that it improves over Bernstein's inequality for very heterogeneous summands. Given the machinery introduced in [2] the proof is surprisingly simple, easier than that of Bernstein's inequality. The method of Ahlswede and Winter might possibly also be used to arrive at the same result, but I have not checked the details.

Before giving the proof I state the necessary auxiliary results (which can all be found in [2]), a trivial corollary and a simple lemma of my own. The word "matrix" will always refer to a real $N \times N$ matrix.

Lemma 2 *(Proposition 3.1 in [2], from Ahlswede and Winter, Oliveira). Let Y be random symmetric matrix and $t \in \mathbb{R}$. Then $\forall \beta > 0$*

$$\Pr\left\{\lambda_{\max} Y \geq t\right\} \leq e^{-\beta t} E \ tr e^{\beta Y}$$

The following beautiful trick is derived in [2] from Lieb's work on convex trace functions.

Lemma 3 *(Lemma 3.4 in [2]) : Let $\mathbf{X} = (X_1, ..., X_n)$ be a vector of independent random symmetric matrices. Then*

$$E \ tr \exp\left(\sum X_i\right) \leq tr \exp\left(\sum \ln E e^{X_i}\right).$$

With deterministic $X_0 = A$ we immediately obtain

Corollary 4 *Let A be a symmetric matrix and let $\mathbf{X} = (X_1, ..., X_n)$ be a vector of independent random symmetric matrices. Then*

$$E\ tr \exp\left(A + \sum X_i\right) \leq tr \exp\left(A + \sum \ln E e^{X_i}\right).$$

Lemma 5 *For a matrix $X \succeq 0$*

$$\ln E e^{-X} \preceq -E[X] + \frac{1}{2} E\left[X^2\right]$$

Proof. For $x \geq 0$ calculus shows that $e^{-x} \leq 1 - x + x^2/2$. Thus, by the transfer rule ((2.2) in [2]), $e^{-X} \preceq I - X + X^2/2$ and

$$E e^{-X} \preceq I - E[X] + \frac{1}{2} E\left[X^2\right] = I + T \preceq e^T = \exp\left(-E[X] + \frac{1}{2} E\left[X^2\right]\right),$$

where we used $1 + t \leq e^t$ and again the transfer rule with $T = -E[X] + E\left[X^2\right]/2$. Taking the logarithm completes the proof. ∎

We also use the following monotonicity property of the trace exponential (also stated in [2]): For symmetric matrices A and B

$$A \preceq B \implies tr\ e^A \leq tr\ e^B. \tag{1}$$

Proof of Theorem 1. For $\beta > 0$

$$\begin{aligned}
& \Pr\left\{\lambda_{\max}\left(\sum_i (E[X_i] - X_i)\right) > t\right\} \\
\leq\ & e^{-\beta t} E\ tr \exp\left(\left(\beta \sum E[X_i]\right) + \sum(-\beta X_i)\right) \text{ by Lemma 2} \\
\leq\ & e^{-\beta t} tr \exp\left(\beta \sum E[X_i] + \sum \ln E e^{-\beta X_i}\right) \text{ by Corollary 4} \\
\leq\ & e^{-\beta t} tr \exp\left(\beta \sum E[X_i] + \sum\left(-\beta E[X_i] + \frac{\beta^2}{2} E\left[X_i^2\right]\right)\right) \text{ Lemma 5 and (1)} \\
=\ & e^{-\beta t} tr \exp\left(\frac{\beta^2}{2} \sum E\left[X_i^2\right]\right) \\
\leq\ & N\ e^{-\beta t} \lambda_{\max}\left(\exp\left(\frac{\beta^2}{2} \sum E\left[X_i^2\right]\right)\right) \text{ since } tr(A) \leq N\ \lambda_{\max}(A) \text{ for } A \succeq 0 \\
=\ & N\ \exp\left(\frac{\beta^2}{2} \lambda_{\max}\left(\sum E\left[X_i^2\right]\right) - \beta t\right) \text{ by spectral mapping.}
\end{aligned}$$

Using calculus to minimize in β gives the result. ∎

References

[1] A. Maurer, A bound on the deviation probability for sums of nonnegative random variables, *J. Inequal. Pure. Appl. Math.*, 4(1) , Art.15, 2003

[2] Joel Tropp, user friendly tail bounds for sums of random matrices, arxive.