

PROBABILITY INEQUALITIES FOR SUMS OF RANDOM MATRICES

JOEL A. TROPP

1. OVERVIEW

Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be independent, self-adjoint random matrices with dimension $d \times d$. Our goal is to provide bounds for the probability

$$\mathbb{P} \left\{ \lambda_{\max} \left(\sum_{k=1}^n \mathbf{X}_k \right) \geq t \right\}. \quad (1.1)$$

The symbol λ_{\max} denotes the (algebraically) maximum eigenvalue of a self-adjoint matrix. We wish to harness properties of the individual summands to obtain information about the behavior of the sum. The approach here leads to simple estimates that are relatively general and easy to use in applied settings. The cost is that the results are not quite sharp for every example.

This research begins with the observation that controlling (1.1) resembles the classical problem of developing tail bounds for a sum of independent real random variables. There are some compelling analogies between self-adjoint matrices and real numbers that suggest it may be possible to extend classical techniques to the matrix setting. Indeed, this dream can be realized.

In a notable paper [AW02], Ahlswede and Winter show that elements from the Laplace transform technique generalize to the matrix setting. Further work in this direction includes [Rec09, Gro09, Oli10a, Oli10b]. These techniques are closely related to noncommutative moment inequalities [LP86, Buc01, JX05] and their applications in random matrix theory [Rud99, RV07].

2. THE MATRIX LAPLACE TRANSFORM METHOD

To begin, we show how Bernstein's Laplace transform technique extends to the matrix setting. The basic idea is due to Ahlswede–Winter [AW02], but we follow Oliveira [Oli10b] in this presentation. Fix a positive number θ . Observe that

$$\begin{aligned} \mathbb{P} \left\{ \lambda_{\max} \left(\sum_k \mathbf{X}_k \right) \geq t \right\} &= \mathbb{P} \left\{ \exp \left\{ \lambda_{\max} \left(\sum_k \theta \mathbf{X}_k \right) \right\} \geq e^{\theta t} \right\} \\ &\leq e^{-\theta t} \cdot \mathbb{E} \exp \left\{ \lambda_{\max} \left(\sum_k \theta \mathbf{X}_k \right) \right\} \\ &= e^{-\theta t} \cdot \mathbb{E} \lambda_{\max} \left(\exp \left\{ \sum_k \theta \mathbf{X}_k \right\} \right) \\ &< e^{-\theta t} \cdot \mathbb{E} \operatorname{tr} \exp \left\{ \sum_k \theta \mathbf{X}_k \right\}. \end{aligned} \quad (2.1)$$

The first identity uses the positive homogeneity of the eigenvalue map; the second relation is Markov's inequality; the third line is the spectral mapping theorem; and the last part holds because the exponential of a self-adjoint matrix is positive definite.

At this point, previous authors interpreted the quantity

$$\mathbb{E} \operatorname{tr} \exp \left\{ \sum_k \theta \mathbf{X}_k \right\}$$

as a matrix extension of the classical moment generating function (mgf). They attempted to generalize the fact that the mgf of an independent sum is the product of the mgfs of the summands.

Date: 2 May 2011.

2010 *Mathematics Subject Classification.* Primary: 60B20.

JAT is with Applied and Computational Mathematics, MC 305-16, California Inst. Technology, Pasadena, CA 91125. E-mail: jtropp@cms.caltech.edu.

Roughly, the hope seemed to be that

$$\langle\langle \mathbb{E} \operatorname{tr} \exp\{\sum_k \theta \mathbf{X}_k\} = \operatorname{tr} \prod_k \mathbb{E} e^{\theta \mathbf{X}_k}. \quad \rangle\rangle$$

This ostensible identity fails completely. In the matrix setting, it is generally not true that $e^{\mathbf{X}+\mathbf{Y}} \neq e^{\mathbf{X}}e^{\mathbf{Y}}$. The Golden–Thompson inequality [Bha97, Ch. IX] can be used as a limited substitute: $\operatorname{tr} e^{\mathbf{X}+\mathbf{Y}} \leq \operatorname{tr} e^{\mathbf{X}}e^{\mathbf{Y}}$. But the obvious extension to three matrices is false: $\operatorname{tr} e^{\mathbf{X}+\mathbf{Y}+\mathbf{Z}} \not\leq \operatorname{tr} e^{\mathbf{X}}e^{\mathbf{Y}}e^{\mathbf{Z}}$. On reflection, it becomes clear that results like this cannot be true because the trace of a product of three positive matrices can be a negative number. In the past, researchers have circumvented this problem using some clever iterative procedures.

Nevertheless, we need a new idea if we want to find the natural extension the classical approach. The key observation is that we should try to **extend the additivity rule for cumulants**. To do so, we need more tools. The following result is one of the crown jewels of matrix analysis.

Theorem 2.1 (Lieb [Lie73]). *Let \mathbf{H} be a self-adjoint matrix. Then the map*

$$\mathbf{A} \mapsto \operatorname{tr} \exp\{\mathbf{H} + \log \mathbf{A}\}$$

is concave on the positive-definite cone.

We apply Lieb’s theorem through the following simple corollary.

Corollary 2.2 (Tropp 2010). *Let \mathbf{H} be a fixed self-adjoint matrix, and let \mathbf{X} be a random self-adjoint matrix. Then*

$$\mathbb{E} \operatorname{tr} \exp\{\mathbf{H} + \mathbf{X}\} \leq \operatorname{tr} \exp\{\mathbf{H} + \log \mathbb{E} e^{\mathbf{X}}\}.$$

When we apply the corollary iteratively, we obtain the following inequality in our setting.

$$\operatorname{tr} \exp\{\log \mathbb{E} \exp\{\sum_k \theta \mathbf{X}_k\}\} = \mathbb{E} \operatorname{tr} \exp\{\sum_k \theta \mathbf{X}_k\} \leq \operatorname{tr} \exp\{\sum_k \log \mathbb{E} e^{\theta \mathbf{X}_k}\}. \quad (2.2)$$

The bound (2.2) states that the cumulant generating function (cgf) of a sum of independent random matrices is controlled by the sum of the cgfs of the individual matrices. Introducing (2.2) into (2.1), we reach

$$\mathbb{P}\{\lambda_{\max}(\sum_k \mathbf{X}_k) \geq t\} \leq \inf_{\theta > 0} [e^{-\theta t} \cdot \operatorname{tr} \exp\{\sum_k \log \mathbb{E} e^{\theta \mathbf{X}_k}\}]. \quad (2.3)$$

The latter inequality is the natural matrix extension of the classical Laplace transform approach.

3. EXAMPLE: MATRIX RADEMACHER SERIES

The simplest application of (2.3) concerns Rademacher series with matrix coefficients. Let $\{\mathbf{A}_k\}$ be a finite sequence of fixed, self-adjoint matrices with dimension d . Let $\{\varepsilon_k\}$ be a sequence of independent Rademacher random variables. We claim that

$$\mathbb{P}\{\lambda_{\max}(\sum_k \varepsilon_k \mathbf{A}_k) \geq t\} \leq d \cdot e^{-t^2/2\sigma^2} \quad \text{where} \quad \sigma^2 = \|\sum_k \mathbf{A}_k^2\|. \quad (3.1)$$

The symbol $\|\cdot\|$ denotes the spectral norm, or Hilbert space operator norm, of a matrix. A related calculation, which we omit, yields

$$\mathbb{E} \lambda_{\max}(\sum_k \varepsilon_k \mathbf{A}_k) \leq \sigma \cdot \sqrt{2 \log d}.$$

For every example, this bound on the expectation is sharp up to the square-root log factor.

The inequality (3.1) has some interesting relations to earlier results. An alternative proof uses sharp noncommutative Khintchine inequalities [Buc01] to bound the matrix mgf. In comparison, the approach described by Ahlswede and Winter [AW02] leads to the weaker inequality

$$\mathbb{P}\{\lambda_{\max}(\sum_k \varepsilon_k \mathbf{A}_k) \geq t\} \leq d \cdot e^{-t^2/2\rho^2} \quad \text{where} \quad \rho^2 = \sum_k \|\mathbf{A}_k\|^2.$$

The latter estimate also follows from Tomczak-Jaegermann’s moment bounds [TJ74] for Rademacher series in the Schatten classes.

To establish the claim (3.1), we need to study the cgf of a fixed matrix modulated by a Rademacher variable. Note that

$$\log \mathbb{E} e^{\varepsilon \theta \mathbf{A}} = \log \cosh(\theta \mathbf{A}) \leq \frac{\theta^2}{2} \mathbf{A}^2.$$

The semidefinite relation follows from the scalar inequality $\log \cosh(x) \leq x^2/2$. Introduce this estimate (with appropriate justifications!) into the tail bound (2.3) to reach

$$\begin{aligned} \mathbb{P} \left\{ \lambda_{\max} \left(\sum_k \varepsilon_k \mathbf{A}_k \right) \geq t \right\} &\leq \inf_{\theta > 0} e^{-\theta t} \cdot \text{tr} \exp \left\{ \frac{\theta^2}{2} \sum_k \mathbf{A}_k^2 \right\} \\ &\leq \inf_{\theta > 0} e^{-\theta t} \cdot \exp \left\{ \frac{\theta^2}{2} \cdot \lambda_{\max} \left(\sum_k \mathbf{A}_k^2 \right) \right\} \\ &= \inf_{\theta > 0} e^{-\theta t} \cdot e^{\theta^2 \sigma^2 / 2}. \end{aligned}$$

Optimize with respect to θ to complete the proof of (3.1).

Finally, let us mention that these ideas can be extended to study rectangular matrices. Consider a finite sequence $\{\mathbf{B}_k\}$ of fixed $d_1 \times d_2$ matrices. Then

$$\mathbb{P} \left\{ \left\| \sum_k \varepsilon_k \mathbf{B}_k \right\| \geq t \right\} \leq (d_1 + d_2) \cdot e^{-t^2 / 2\sigma^2} \quad \text{where} \quad \sigma^2 = \left\| \sum_k \mathbf{B}_k \mathbf{B}_k^* \right\| \vee \left\| \sum_k \mathbf{B}_k^* \mathbf{B}_k \right\|.$$

Remarkably, this estimate follows immediately from (3.1) by applying that result to the self-adjoint matrices

$$\mathbf{A}_k = \begin{bmatrix} \mathbf{0} & \mathbf{B}_k \\ \mathbf{B}_k^* & \mathbf{0} \end{bmatrix}.$$

We omit the details.

ACKNOWLEDGMENTS

This work has been supported in part by ONR awards N00014-08-1-0883 and N00014-11-1-0025, AFOSR award FA9550-09-1-0643, and a Sloan Fellowship.

REFERENCES

- [AW02] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Trans. Inform. Theory*, 48(3):569–579, Mar. 2002.
- [Bha97] R. Bhatia. *Matrix Analysis*. Number 169 in Graduate Texts in Mathematics. Springer, Berlin, 1997.
- [Buc01] A. Buchholz. Operator Khintchine inequality in non-commutative probability. *Math. Ann.*, 319:1–16, 2001.
- [Gro09] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inform. Theory*, Oct. 2009. To appear. Available at [arXiv:0910.1879](https://arxiv.org/abs/0910.1879).
- [JX05] M. Junge and Q. Xu. On the best constants in some non-commutative martingale inequalities. *Bull. London Math. Soc.*, 37:243–253, 2005.
- [Lie73] E. H. Lieb. Convex trace functions and the Wigner–Yanase–Dyson conjecture. *Adv. Math.*, 11:267–288, 1973.
- [LP86] F. Lust-Piquard. Inégalités de Khintchine dans C_p ($1 < p < \infty$). *C. R. Math. Acad. Sci. Paris*, 303(7):289–292, 1986.
- [Oli10a] R. I. Oliveira. Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. Available at [arXiv:0911.0600](https://arxiv.org/abs/0911.0600), Feb. 2010.
- [Oli10b] R. I. Oliveira. Sums of random Hermitian matrices and an inequality by Rudelson. *Elect. Comm. Probab.*, 15:203–212, 2010.
- [Rec09] B. Recht. Simpler approach to matrix completion. *J. Mach. Learn. Res.*, Oct. 2009. To appear. Available at <http://pages.cs.wisc.edu/~brecht/papers/09.Recht.ImprovedMC.pdf>.
- [Rud99] M. Rudelson. Random vectors in the isotropic position. *J. Funct. Anal.*, 164:60–72, 1999.
- [RV07] M. Rudelson and R. Vershynin. Sampling from large matrices: An approach through geometric functional analysis. *J. Assoc. Comput. Mach.*, 54(4):Article 21, 19 pp., Jul. 2007. (electronic).
- [TJ74] N. Tomczak-Jaegermann. The moduli of smoothness and convexity and the Rademacher averages of trace classes S_p ($1 \leq p < \infty$). *Studia Math.*, 50:163–182, 1974.