SUMS OF RANDOM HERMITIAN MATRICES AND AN INEQUALITY BY RUDELSON

ROBERTO IMBUZEIRO OLIVEIRA 1

IMPA, Rio de Janeiro, RJ, Brazil, 22460-320.
email: rimfo@impa.br

Submitted January 28, 2010, accepted in final form May 17, 2010

AMS 2000 Subject classification: 60B20
Keywords: Random Hermitian matrices, concentration inequalities, Khintchine Inequalities

Abstract
We give a new, elementary proof of a key inequality used by Rudelson in the derivation of his well-known bound for random sums of rank-one operators. Our approach is based on Ahlswede and Winter’s technique for proving operator Chernoff bounds. We also prove a concentration inequality for sums of random matrices of rank one with explicit constants.

1 Introduction
This note mainly deals with estimates for the operator norm \( \|Z_n\| \) of random sums

\[
Z_n \equiv \sum_{i=1}^{n} \epsilon_i A_i
\]

of deterministic Hermitian matrices \( A_1, \ldots, A_n \) multiplied by random coefficients. Recall that a Rademacher sequence is a sequence \( \{\epsilon_i\}_{i=1}^{n} \) of i.i.d. random variables with \( \epsilon_1 \) uniform over \( \{-1,+1\} \). A standard Gaussian sequence is a sequence i.i.d. standard Gaussian random variables. Our main goal is to prove the following result.

Theorem 1 (proven in Section 3). Given positive integers \( d, n \in \mathbb{N} \), let \( A_1, \ldots, A_n \) be deterministic \( d \times d \) Hermitian matrices and \( \{\epsilon_i\}_{i=1}^{n} \) be either a Rademacher sequence or a standard Gaussian sequence. Define \( Z_n \) as in (1). Then for all \( p \in [1, +\infty) \),

\[
(E \left[\|Z_n\|^p\right])^{1/p} \leq \left(\sqrt{2 \ln(2d)} + C_p\right) \left\|\sum_{i=1}^{n} A_i^2\right\|^{1/2}
\]

where

\[
C_p \equiv \left(p \int_{0}^{+\infty} t^{p-1} e^{-\frac{t^2}{2}} dt\right)^{1/p} \leq c \sqrt{p} \text{ for some universal } c > 0.
\]

1 RESEARCH SUPPORTED BY A “BOLSA DE PRODUTIVIDADE EM PESQUISA” (CNPQ) AND RESEARCH GRANTS FROM CNPQ AND FAPERJ.
For $d = 1$, this result corresponds to the classical Khintchine inequalities, which give sub-Gaussian bounds for the moments of $\sum_{i=1}^{n} c_i a_i$, $(a_1, \ldots, a_n \in \mathbb{R})$. Theorem 1 is implicit in Section 3 of Rudelson’s paper [12], albeit with non-explicit constants. The main Theorem in that paper is the following inequality, which is a simple corollary of Theorem 1 if $Y_1, \ldots, Y_n$ are i.i.d. random (column) vectors in $\mathbb{C}^d$ which are isotropic (i.e. $\mathbb{E} \left[ Y_1 Y_1^* \right] = I$, the $d \times d$ identity matrix), then:

$$\mathbb{E} \left[ \left\| \frac{1}{n} \sum_{i=1}^{n} Y_i Y_i^* - I \right\| \right] \leq C \left( \mathbb{E} \left[ |Y_1|^n \right] \right)^{1/\log n} \sqrt{\frac{\log d}{n}} \tag{2}$$

for some universal $C > 0$, whenever the RHS of the above inequality is at most 1. This important result has been applied to several different problems, such as bringing a convex body to near-isotropic position [12]; the analysis of low-rank approximations of matrices [13] [7] and graph sparsification [14]; estimating of singular values of matrices with independent rows [11]; analysing compressive sensing [4]; and related problems in Harmonic Analysis [17] [16].

The key ingredient of the original proof of Theorem 1 is a non-commutative Khintchine inequality by Lust-Picard and Pisier [10]. This states that there exists a universal $c > 0$ such that for all $Z_n$ as in the Theorem, all $p \geq 1$ and all $d \times d$ matrices $\{B_i, D_i\}_{i=1}^{n}$ with $B_i + D_i = A_i$, $1 \leq i \leq n$,

$$\mathbb{E} \left[ \left\| Z_n \right\|_{Sp}^p \right]^{1/p} \leq c \sqrt{p} \left( \left\| \sum_{i=1}^{n} B_i B_i^* \right\|_{Sp}^{1/2} + \left\| \sum_{i=1}^{n} D_i D_i^* \right\|_{Sp}^{1/2} \right),$$

where $\| \cdot \|_{Sp}$ denotes the $p$-th Schatten norm: $\|A\|_{Sp}^p \equiv \text{Tr}[\text{Tr}(A^*A)^{p/2}]$. Better estimates for $c$, and thus for the constant in Rudelson’s bound, can be obtained from the work of Buchholz [3]. Unfortunately, the proofs of the Lust-Picard/Pisier inequality employs language and tools from non-commutative probability that are rather foreign to most potential users of [2], and Buchholz’s bound additionally relies on delicate combinatorics.

This note presents a more direct proof of Theorem 1. Our argument is based on an improvement of the methodology created by Ahlswede and Winter [21] in order to prove their operator Chernoff bound, which also has many applications e.g. [8] (the improvement is discussed in Section 3.1).

This approach only requires elementary facts from Linear Algebra and Matrix Analysis. The most complicated result that we use is the Golden-Thompson inequality [6] [15]:

$$\forall d \in \mathbb{N}, \forall d \times d \text{ Hermitian matrices } A, B, \text{Tr}(e^{A+B}) \leq \text{Tr}(e^A e^B). \tag{3}$$

The elementary proof of this classical inequality is sketched in Section 5 below.

We have already noted that Rudelson’s bound [2] follows simply from Theorem 1 [12], see Section 3] for details. Here we prove a concentration lemma corresponding to that result under the stronger assumption that $|Y_1|$ is a.s. bounded. While similar results have appeared in other papers [11] [13] [17], our proof is simpler and gives explicit constants.

**Lemma 1** (Proven in Section 4). Let $Y_1, \ldots, Y_n$ be i.i.d. random column vectors in $\mathbb{C}^d$ with $|Y_1| \leq M$ almost surely and $\|E \left[ Y_1 Y_1^* \right]\| \leq 1$. Then:

$$\forall t \geq 0, \mathbb{P} \left( \left\| \frac{1}{n} \sum_{i=1}^{n} Y_i Y_i^* - E \left[ Y_1 Y_1^* \right] \right\| \geq t \right) \leq \left( \frac{d}{2\pi e} \right)^{d} e^{-\frac{t^2}{2n}}.$$


In particular, a calculation shows that, for any \( n, d \in \mathbb{N}, M > 0 \) and \( \delta \in (0, 1) \) such that:

\[
4M \sqrt{\frac{2 \ln(\min\{d, n\}) + 2 \ln 2 + \ln(1/\delta)}{n}} \leq 2,
\]

we have:

\[
P\left( \left\| \frac{1}{n} \sum_{i=1}^{n} Y_i Y_i^* - \mathbb{E} \left[ Y_i Y_i^* \right] \right\| \geq 4M \sqrt{\frac{2 \ln(\min\{d, n\}) + 2 \ln 2 + \ln(1/\delta)}{n}} \right) \geq 1 - \delta.
\]

A key feature of this Lemma is that it gives meaningful results even when the ambient dimension \( d \) is arbitrarily large. In fact, the same result holds (with \( d = \infty \)) for \( Y_i \) taking values in a separable Hilbert space, and this form of the result may be used to simplify the proofs in [11] (especially in the last section of that paper).

To conclude the introduction, we present an open problem: is it possible to improve upon Rudelson’s bound under further assumptions? There is some evidence that the dependence on \( \ln(d) \) in the Theorem, while necessary in general [13, Remark 3.4], can sometimes be removed. For instance, Adamczak et al. [1] have improved upon Rudelson’s original application of Theorem 1 to convex bodies, obtaining exactly what one would expect in the absence of the \( \sqrt{\log(2d)} \) term.

Another setting where our bound is a \( \Theta \left( \sqrt{\ln d} \right) \) factor away from optimality is that of more classical random matrices (cf. the end of Section 3.1 below). It would be interesting if one could sharpen the proof of Theorem 1 in order to reobtain these results. [Related issues are raised by Vershynin [18].]

Acknowledgement: we thank the anonymous referees for providing extra references, pointing out several typos and suggesting a number of small improvements to this note.

## 2 Preliminaries

We let \( \mathbb{C}^{d \times d}_{\text{Herm}} \) denote the set of \( d \times d \) Hermitian matrices, which is a subset of the set \( \mathbb{C}^{d \times d} \) of all \( d \times d \) matrices with complex entries. The spectral theorem states that all \( A \in \mathbb{C}^{d \times d}_{\text{Herm}} \) have \( d \) real eigenvalues (possibly with repetitions) that correspond to an orthonormal set of eigenvectors. \( \lambda_{\text{max}}(A) \) is the largest eigenvalue of \( A \). The spectrum of \( A \), denoted by \( \text{spec}(A) \), is the multiset of all eigenvalues, where each eigenvalue appears a number of times equal to its multiplicity. We let

\[
\|C\| \equiv \max_{v \in \mathbb{C}^d, |v|=1} |Cv|
\]

denote the operator norm of \( C \in \mathbb{C}^{d \times d} \) (\( |\cdot| \) is the Euclidean norm). By the spectral theorem,

\[
\forall A \in \mathbb{C}^{d \times d}_{\text{Herm}}, \|A\| = \max\{\lambda_{\text{max}}(A), \lambda_{\text{max}}(-A)\}.
\]

Moreover, \( \text{Tr}(A) \) (the trace of \( A \)) is the sum of the eigenvalues of \( A \).

### 2.1 Spectral mapping

Let \( f : \mathbb{C} \to \mathbb{C} \) be an entire analytic function with a power-series representation \( f(z) \equiv \sum_{n \geq 0} c_n z^n \) (\( z \in \mathbb{C} \)). If all \( c_n \) are real, the expression:

\[
f(A) \equiv \sum_{n \geq 0} c_n A^n \quad (A \in \mathbb{C}^{d \times d}_{\text{Herm}})
\]
corresponds to a map from \( C^{d \times d}_{\text{Herm}} \) to itself. We will sometimes use the so-called spectral mapping property:

\[
\text{spec}(A) = f(\text{spec}(A)).
\]  

(4)

By this we mean that the eigenvalues of \( f(A) \) are the numbers \( f(\lambda) \) with \( \lambda \in \text{spec}(A) \). Moreover, the multiplicity of \( \xi \in \text{spec}(A) \) is the sum of the multiplicities of all preimages of \( \xi \) under \( f \) that lie in \( \text{spec}(A) \).

2.2  The positive-semidefinite order

We will use the notation \( A \succeq 0 \) to say that \( A \) is positive-semidefinite, i.e. \( A \in C^{d \times d}_{\text{Herm}} \) and its eigenvalues are non-negative. This is equivalent to saying that \( (v, Av) \geq 0 \) for all \( v \in C^{d} \), where \( (\cdot, \cdot) \) is the standard Euclidean inner product.

If \( A, B \in C^{d \times d}_{\text{Herm}} \), we write \( A \succeq B \) or \( B \preceq A \) to say that \( A - B \succeq 0 \). Notice that \( \succeq \) is a partial order and that:

\[
\forall A, B, A', B' \in C^{d \times d}_{\text{Herm}}, (A \preceq A') \land (B \preceq B') \Rightarrow A + A' \preceq B + B'.
\]  

(5)

Moreover, spectral mapping (4) implies that:

\[
\forall A \in C^{d \times d}_{\text{Herm}}, A^{2} \succeq 0.
\]  

(6)

We will also need the following simple fact.

**Proposition 1.** For all \( A, B, C \in C^{d \times d}_{\text{Herm}} \):

\[
(C \succeq 0) \land (A \preceq B) \Rightarrow \text{Tr}(CA) \leq \text{Tr}(CB).
\]  

(7)

**Proof:** To prove this, assume the LHS and observe that the RHS is equivalent to \( \text{Tr}(C\Delta) \geq 0 \) where \( \Delta \equiv B - A \). By assumption, \( \Delta \succeq 0 \), hence it has a Hermitian square root \( \Delta^{1/2} \). The cyclic property of the trace implies:

\[
\text{Tr}(C\Delta) = \text{Tr}(\Delta^{1/2}C\Delta^{1/2}).
\]

Since the trace is the sum of the eigenvalues, we will be done once we show that \( \Delta^{1/2}C\Delta^{1/2} \succeq 0 \). But, since \( \Delta^{1/2} \) is Hermitian and \( C \succeq 0 \),

\[
\forall v \in C^{d}, (v, \Delta^{1/2}C\Delta^{1/2}v) = (\Delta^{1/2}v, C(\Delta^{1/2}v)) = (w, Cw) \geq 0 \text{ (with } w = \Delta^{1/2}v),
\]

which shows that \( \Delta^{1/2}C\Delta^{1/2} \succeq 0 \), as desired. \( \square \)

2.3  Probability with matrices

Assume \( (\Omega, \mathcal{F}, \mathbb{P}) \) is a probability space and \( Z : \Omega \to C^{d \times d}_{\text{Herm}} \) is measurable with respect to \( \mathcal{F} \) and the Borel \( \sigma \)-field on \( C^{d \times d}_{\text{Herm}} \) (this is equivalent to requiring that all entries of \( Z \) be complex-valued random variables). \( C^{d \times d}_{\text{Herm}} \) is a metrically complete vector space and one can naturally define an expected value \( \mathbb{E}[Z] \in C^{d \times d}_{\text{Herm}} \). This turns out to be the matrix \( \mathbb{E}[Z] \in C^{d \times d}_{\text{Herm}} \) whose \((i, j)\)-entry is the expected value of the \((i, j)\)-th entry of \( Z \). [Of course, \( \mathbb{E}[Z] \) is only defined if all entries of \( Z \) are integrable, but this will always be the case in this paper.]

The definition of expectations implies that traces and expectations commute:

\[
\text{Tr}(\mathbb{E}[Z]) = \mathbb{E}[\text{Tr}(Z)].
\]  

(8)
Moreover, one can check that the usual product rule is satisfied:

$$\text{If } Z, W : \Omega \to \mathbb{C}_{\text{Herm}}^{d \times d} \text{ are measurable and independent, } \mathbb{E} [ZW] = \mathbb{E} [Z] \mathbb{E} [W].$$

Finally, the inequality:

$$\text{If } Z : \Omega \to \mathbb{C}_{\text{Herm}}^{d \times d} \text{ satisfies } Z \geq 0 \text{ a.s., } \mathbb{E} [Z] \geq 0$$

is an easy consequence of another easily checked fact: $(v, \mathbb{E} [Z] v) = \mathbb{E} [(v, Zv)]$, $v \in \mathbb{C}^d$.

### 3 Proof of Theorem \[1\]

**Proof:** [of Theorem \[1\]] The usual Bernstein trick implies that for all $t \geq 0$,

$$\forall t \geq 0, \mathbb{P} (\|Z_n\| \geq t) \leq \inf_{\delta > 0} e^{-\delta t} \mathbb{E} [e^{\|Z_n\|}] .$$

Notice that

$$\mathbb{E} [e^{\|Z_n\|}] \leq \mathbb{E} [e^{\lambda_{\text{max}}(Z_n)}] + \mathbb{E} [e^{\lambda_{\text{max}}(-Z_n)}] = 2\mathbb{E} [e^{\lambda_{\text{max}}(Z_n)}]$$

since $\|Z_n\| = \max \{\lambda_{\text{max}}(Z_n), \lambda_{\text{max}}(-Z_n)\}$ and $-Z_n$ has the same law as $Z_n$.

The function “$x \mapsto e^{sx}$” is monotone non-decreasing and positive for all $s \geq 0$. It follows from the spectral mapping property \[4\] that for all $s \geq 0$, the largest eigenvalue of $e^{sZ_n}$ is $e^{\lambda_{\text{max}}(Z_n)}$ and all eigenvalues of $e^{sZ_n}$ are non-negative. Using the equality “trace = sum of eigenvalues” implies that for all $s \geq 0$,

$$\mathbb{E} [e^{\lambda_{\text{max}}(Z_n)}] = \mathbb{E} [\lambda_{\text{max}} (e^{sZ_n})] \leq \mathbb{E} [\text{Tr}(e^{sZ_n})] .$$

As a result, we have the inequality:

$$\forall t \geq 0, \mathbb{P} (\|Z_n\| \geq t) \leq 2 \inf_{\delta > 0} e^{-\delta t} \mathbb{E} [\text{Tr}(e^{sZ_n})] .$$

(12)

Up to now, our proof has followed Ahlswe and Winter’s argument. The next lemma, however, will require new ideas.

**Lemma 2.** For all $s \in \mathbb{R}$,

$$\mathbb{E} [\text{Tr}(e^{sZ_n})] \leq \text{Tr} \left( e^{\frac{s^2 x_n^2 + s^2}{2}} \right) .$$

This lemma is proven below. We will now show how it implies Rudelson’s bound. Let

$$\sigma^2 = \left\| \sum_{i=1}^n A_i^2 \right\| = \lambda_{\text{max}} \left( \sum_{i=1}^n A_i^2 \right) .$$

[The second inequality follows from $\sum_{i=1}^n A_i^2 \geq 0$, which holds because of \(5\) and \(6\).] We note that:

$$\text{Tr} \left( e^{\frac{s^2 x_n^2 + s^2}{2}} \right) \leq \lambda_{\text{max}} \left( e^{\frac{s^2 x_n^2 + s^2}{2}} \right) = d e^\frac{s^2}{2}$$

where the equality is yet another application of spectral mapping \[4\] and the fact that “$x \mapsto e^{s^2 x/2n}$ is monotone non-decreasing. We deduce from the Lemma and \(12\) that:

$$\forall t \geq 0, \mathbb{P} (\|Z_n\| \geq t) \leq 2d \inf_{\delta > 0} e^{-\delta t + \frac{s^2}{2}} = 2d e^{-\frac{t}{2 \sigma^2}} .$$

(13)
This implies that for any $p \geq 1$,
\[
\frac{1}{\sigma^p} \mathbb{E} \left[ (\|Z_n\| - \sqrt{2 \ln(2d)} \sigma)^+ \right] = \int_0^{+\infty} t^{p-1} \mathbb{P} \left( \|Z_n\| \geq \left( \sqrt{2 \ln(2d)} + t \right) \sigma \right) dt
\]

(see (13)) \leq 2pd \int_0^{+\infty} t^{p-1} e^{-\frac{(1+\sqrt{2\ln(2d)})^2}{2}} dt
\leq 2pd \int_0^{+\infty} t^{p-1} e^{-\frac{t^2}{2}} dt = C_p^p
\]

Since $0 \leq \|Z_n\| \leq \sqrt{2 \ln(2d)} \sigma + (\|Z_n\| - \sqrt{2 \ln(2d)} \sigma)_+$, this implies the $L^p$ estimate in the Theorem. The bound $C_p \leq c \sqrt{p}$ is standard and we omit its proof. \[\square\]

To finish, we now prove Lemma 2.

Proof: [of Lemma 2] Define $D_0 = \sum_{i=1}^n \frac{s^2 A_i^2}{2}$ and
\[
D_j \equiv D_0 + \sum_{i=1}^n \left( se_i A_j - \frac{s^2 A_i^2}{2} \right) (1 \leq j \leq n).
\]

We will prove that for all $1 \leq j \leq n$:
\[
\mathbb{E} \left[ \text{Tr} \left( \exp \left( D_j \right) \right) \right] \leq \mathbb{E} \left[ \text{Tr} \left( \exp \left( D_{j-1} \right) \right) \right]. \tag{14}
\]

Notice that this implies $\mathbb{E} \left[ \text{Tr}(e^{D_0}) \right] \leq \mathbb{E} \left[ \text{Tr}(e^{D_n}) \right]$, which is the precisely the Lemma. To prove (14), fix $1 \leq j \leq n$. Notice that $D_{j-1}$ is independent from $se_i A_j - \frac{s^2 A_i^2}{2}$ since the $\{\epsilon_i\}_{i=1}^n$ are independent. This implies that:
\[
\mathbb{E} \left[ \text{Tr} \left( \exp \left( D_j \right) \right) \right] = \mathbb{E} \left[ \text{Tr} \left( \exp \left( D_{j-1} + se_i A_j - \frac{s^2 A_i^2}{2} \right) \right) \right]
\]

(see Golden-Thompson (9)) \leq \mathbb{E} \left[ \text{Tr} \left( \exp \left( D_{j-1} \right) \exp \left( se_i A_j - \frac{s^2 A_i^2}{2} \right) \right) \right]

(Tr(\cdot) and $\mathbb{E}[\cdot]$ commute, (8)) = \text{Tr} \left( \mathbb{E} \left[ \exp \left( D_{j-1} \right) \exp \left( se_i A_j - \frac{s^2 A_i^2}{2} \right) \right] \right).

(see product rule, (9)) = \text{Tr} \left( \mathbb{E} \left[ \exp \left( D_{j-1} \right) \right] \mathbb{E} \left[ \exp \left( se_i A_j - \frac{s^2 A_i^2}{2} \right) \right] \right).

By the monotonicity of the trace (7) and the fact that $\exp(D_{j-1}) \succeq 0$ (cf. (4)) implies $\mathbb{E} \left[ \exp \left( D_{j-1} \right) \right] \succeq 0$ (cf. (10)), we will be done once we show that:
\[
\mathbb{E} \left[ \exp \left( se_i A_j - \frac{s^2 A_i^2}{2} \right) \right] \succeq 1. \tag{15}
\]

The key fact is that $se_i A_j$ and $-s^2 A_i^2/2$ always commute, hence the exponential of the sum is the product of the exponentials. Applying (9) and noting that $e^{-s^2 A_i^2/2}$ is constant, we see that:
\[
\mathbb{E} \left[ \exp \left( se_i A_j - \frac{s^2 A_i^2}{2} \right) \right] = \mathbb{E} \left[ \exp \left( se_i A_j \right) \right] e^{-\frac{s^2 A_i^2}{2}}.
\]
In the Gaussian case, an explicit calculation shows that
\[ \mathbb{E} \left[ \exp \left( \sigma \epsilon_{ij} A_{ij} \right) \right] = e^{\sigma^2 A_{ij}^2 / 2}, \]
hence (15) holds. In the Rademacher case, we have:
\[ \mathbb{E} \left[ \exp \left( \sigma \epsilon_{ij} A_{ij} \right) \right] e^{-\sigma^2 A_{ij}^2 / 2} = f(A_{ij}) \]
where \( f(z) = \cosh(sz)e^{-sz^2 / 2} \). It is a classical fact that \( 0 \leq \cosh(x) \leq e^{x^2 / 2} \) for all \( x \in \mathbb{R} \) (just compare the Taylor expansions); this implies that \( 0 \leq f(\lambda) \leq 1 \) for all eigenvalues of \( A_{ij} \). Using spectral mapping (4), we see that:
\[ \text{spec} f(A_{ij}) = f(\text{spec}(A_{ij})) \subseteq [0, 1], \]
which implies that \( f(A_{ij}) \leq I \). This proves (15) in this case and finishes the proof of (14) and of the Lemma.

3.1 Remarks on the original AW approach

A direct adaptation of the original argument of Ahlswede and Winter [2] would lead to an inequality of the form:
\[ \mathbb{E} \left[ \text{Tr} \left( e^{\epsilon Z_n} \right) \right] \leq \text{Tr} \left( \mathbb{E} \left[ e^{\epsilon_{ij} A_{ij}} \right] \mathbb{E} \left[ e^{\epsilon Z_n} \right] \right). \]
One sees that:
\[ \mathbb{E} \left[ e^{\epsilon_{ij} A_{ij}} \right] \leq e^{\sigma^2 A_{ij}^2 / 2} \leq e^{\sigma^2 |A_{ij}|^2 / 2} I. \]
However, only the second equality seems to be useful, as there is no obvious relationship between
\[ \text{Tr} \left( e^{\sigma^2 A_{ij}^2 / 2} \mathbb{E} \left[ e^{\epsilon Z_n} \right] \right) \]
and
\[ \text{Tr} \left( \mathbb{E} \left[ e^{\epsilon_{ij} A_{ij} - (m-1)} \right] \mathbb{E} \left[ e^{\epsilon Z_n + \sigma^2 |A_{ij}|^2 / 2} \right] \right), \]
which is what we would need to proceed with induction. [Note that Golden-Thompson (3) cannot be undone and fails for three summands, [15].] The best one can do with the second inequality is:
\[ \mathbb{E} \left[ \text{Tr} \left( e^{\epsilon Z_n} \right) \right] \leq d e^{\frac{\sigma^2 \sum_{i,j} |A_{ij}|^2}{2}}. \]
This would give a version of Theorem 1 with \( \sum_{i=1}^{m} |A_{ij}|^2 \) replacing \( \| \sum_{i=1}^{m} A_{ij}^2 \| \). This modified result is always worse than the actual Theorem, and can be dramatically so. For instance, consider the case of a Wigner matrix where:
\[ Z_n = \sum_{1 \leq i < j \leq m} \epsilon_{ij} A_{ij} \]
with the \( \epsilon_{ij} \) i.i.d. standard Gaussian and each \( A_{ij} \) has ones at positions \((i, j)\) and \((j, i)\) and zeros elsewhere (we take \( d = m \) and \( n = \binom{m}{2} \) in this case). Direct calculation reveals:
\[ \left\| \sum_{ij} A_{ij}^2 \right\| = \|(m-1)I\| = m - 1 \leq \binom{m}{2} = \sum_{ij} |A_{ij}|^2. \]
We note in passing that neither approach is sharp in this case, as \( \| \sum_{ij} \epsilon_{ij} A_{ij} \| \) concentrates around \( 2\sqrt{m} \). The same holds when the \( \epsilon_{ij} \) are Rademacher [5].
4 Concentration for rank-one operators

In this section we prove Lemma 1.

Proof: [of Lemma 1] Let

\[ \phi(s) \equiv \mathbb{E} \left[ \exp \left( s \left\| \frac{1}{n} \sum_{i=1}^{n} Y_i Y_i^* - \mathbb{E} [Y_1 Y_1^*] \right\| \right) \right]. \]

We will show below that:

\[ \forall s \geq 0, \phi(s) \leq 2 \min \{d, n\} e^{2M^2s^2/n} \phi(2M^2s^2/n). \]  \hspace{1cm} (16)

By Jensen’s inequality, \( \phi(2M^2s^2/n) \leq \phi(s)^{2M^2/n} \) whenever \( 2M^2s/n \leq 1 \), hence (16) implies:

\[ \forall 0 \leq s \leq n/2M^2, \phi(s) \leq (2 \min \{d, n\})^{1/2M^2s^2/n} e^{2M^2s^2/n}. \]

Since

\[ \forall s \geq 0, \mathbb{P} \left( \left\| \frac{1}{n} \sum_{i=1}^{n} Y_i Y_i^* - \mathbb{E} [Y_1 Y_1^*] \right\| \geq t \right) \leq e^{-st} \phi(s), \]

the Lemma then follows from the choice

\[ s = \frac{n}{8M^2} \min \{2, t\} \]

and a few simple calculations. [Notice that \( 2M^2s \leq n/2 \) with this choice, hence \( 1/(1 - 2M^2s/n) \leq 2 \) and \( 2M^2s^2/(n - 2M^2s) \leq 4M^2s^2/n \).]

To prove (16), we begin with symmetrization (see e.g. Lemma 6.3 in Chapter 6 of [9]):

\[ \phi(s) \leq \mathbb{E} \left[ \exp \left( 2s \left\| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i Y_i Y_i^* \right\| \right) \right], \]

where \( \{\epsilon_i\}_{i=1}^{n} \) is a Rademacher sequence independent of \( Y_1, \ldots, Y_n \). Let \( \mathcal{S} \) be the (random) span of \( Y_1, \ldots, Y_n \) and \( \text{Tr}_{\mathcal{S}} \) denote the trace operation on linear operators mapping \( \mathcal{S} \) to itself. Using the same argument as in (11), we notice that:

\[ \mathbb{E} \left[ \exp \left( 2s \left\| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i Y_i Y_i^* \right\| \right) \left| Y_1, \ldots, Y_n \right. \right] \leq 2 \mathbb{E} \left[ \text{Tr}_{\mathcal{S}} \left( \exp \left( 2s \frac{1}{n} \sum_{i=1}^{n} \epsilon_i Y_i Y_i^* \right) \right) \right] \left| Y_1, \ldots, Y_n \right. \].

Lemma 2 implies:

\[ \mathbb{E} \left[ \text{Tr}_{\mathcal{S}} \left( \exp \left( 2s \frac{1}{n} \sum_{i=1}^{n} \epsilon_i Y_i Y_i^* \right) \right) \right] \left| Y_1, \ldots, Y_n \right. \] \hspace{1cm} \leq \hspace{1cm} 2 \text{Tr}_{\mathcal{S}} \left( \exp \left( 2s^2 \frac{1}{n^2} \sum_{i=1}^{n} (Y_i Y_i^*)^2 \right) \right) \left| Y_1, \ldots, Y_n \right. \hspace{1cm} \leq \hspace{1cm} 2 \min \{d, n\} \exp \left( \frac{2s^2}{n^2} \sum_{i=1}^{n} (Y_i Y_i^*)^2 \right) \text{ a.s.,} \]

using spectral mapping [4], the equality “trace = sum of eigenvalues” and the fact that \( \mathcal{S} \) has dimension \( \leq \min \{d, n\} \). A quick calculation shows that \( 0 \geq (Y_i Y_i^*)^2 = |Y_i|^2 Y_i Y_i^* \leq M^2 Y_i Y_i^* \), hence (5) implies:

\[ 0 \geq \frac{2s^2}{n^2} \sum_{i=1}^{n} (Y_i Y_i^*)^2 \leq \frac{2M^2s^2}{n} \left( \frac{1}{n} \sum_{i=1}^{n} Y_i Y_i^* \right). \]
Therefore:
\[
\left\| \frac{2s^2}{n} \sum_{i=1}^{n} (Y_i Y_i^*) \right\|^2 \leq \frac{2M^2s^2}{n} \left\| \frac{1}{n} \sum_{i=1}^{n} Y_i Y_i^* \right\|^2 \leq \frac{2M^2s^2}{n} \left\| \frac{1}{n} \sum_{i=1}^{n} Y_i Y_i^* - \mathbb{E} [Y_1 Y_1^*] \right\| + \frac{2M^2s^2}{n}.
\]

[We used \( \| \mathbb{E} [Y_1 Y_1^*] \| \leq 1 \) in the last inequality.] Plugging this into the conditional expectation above and integrating, we obtain (16):
\[
\phi(s) \leq 2 \min\{d, n\} \mathbb{E} \left[ \exp \left( \frac{2M^2s^2}{n} \left\| \frac{1}{n} \sum_{i=1}^{n} Y_i Y_i^* - \mathbb{E} [Y_1 Y_1^*] \right\| + \frac{2M^2s^2}{n} \right) \right] = 2 \min\{d, n\} e^{\frac{2M^2s^2}{n}} \phi(\frac{2M^2s^2}{n}).
\]

\section{Proof sketch for Golden-Thompson inequality}

As promised in the Introduction, we sketch an elementary proof of inequality (3). We will need the Trotter-Lie formula, a simple consequence of the Taylor formula for \( e^X \):
\[
\forall A, B \in \mathbb{C}^{d \times d}_{\text{Herm}}, \lim_{n \to +\infty} (e^{A/n}e^{B/n})^n = e^{A+B}. \tag{17}
\]

The second ingredient is the inequality:
\[
\forall k \in \mathbb{N}, \forall X, Y \in \mathbb{C}^{d \times d}_{\text{Herm}} : X, Y \succeq 0 \Rightarrow \text{Tr}((X Y)^{2^k+1}) \leq \text{Tr}((X^2 Y^2)^{2^k}). \tag{18}
\]

This is proven in \cite{6} via an argument using the existence of positive-semidefinite square-roots for positive-semidefinite matrices, and the Cauchy-Schwartz inequality for the standard inner product over \( \mathbb{C}^{d \times d} \). Iterating (18) implies:
\[
\forall X, Y \in \mathbb{C}^{d \times d}_{\text{Herm}} : X, Y \succeq 0 \Rightarrow \text{Tr}((X Y)^{2^k}) \leq \text{Tr}(X^{2^k} Y^{2^k}).
\]

Apply this to \( X = e^{A/2^k} \) and \( Y = e^{B/2^k} \) with \( A, B \in \mathbb{C}^{d \times d}_{\text{Herm}} \). Spectral mapping (4) implies \( X, Y \succeq 0 \) and we deduce:
\[
\text{Tr}(e^{A/2^k} e^{B/2^k}) \leq \text{Tr}(e^A e^B).
\]

Inequality (3) follows from letting \( k \to +\infty \), using (17) and noticing that \( \text{Tr}(\cdot) \) is continuous.

\section*{References}

\begin{enumerate}
\item Radoslaw Adamczak, Alexander E. Litvak, Alain Pajor and Nicole Tomczak-Jaegermann. “Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles.” \textit{Journal of the American Mathematical Society} 23: 535–561, 2010.
\item Rudolf Ahlswede and Andreas Winter. “Strong converse for identification via quantum channels.” \textit{IEEE Transactions on Information Theory} 48(3): 569–579, 2002. MR1889969
\item Artur Buchholz. “Operator Khintchine inequality in non commutative probability.” \textit{Mathematische Annalen} 319, 1-Ü16, 2001.
\end{enumerate}
[4] Emmanuel Candès and Justin Romberg. “Sparsity and incoherence in compressive sampling.” *Inverse Problems* 23:969–985, 2007.

[5] Zoltan Füredi and János Komlós. “The eigenvalues of random symmetric matrices.” *Combinatorica* 1(3): 233–241, 1981. [MR0637828](https://www.ams.org/mathscinet-getitem?mr=0637828)

[6] Sidney Golden. “Lower Bounds for the Helmholtz Function.” *Physical Review* 137: B1127–B1128, 1965.

[7] Nathan Halko, Per-Gunnar Martinsson and Joel A. Tropp. “Finding structure with randomness: Stochastic algorithms for constructing approximate matrix decompositions.” Preprint arXiv:0909.4061.

[8] Zeph Landau and Alexander Russell. “Random Cayley graphs are expanders: a simplified proof of the Alon-Roichman theorem.” *Electronic Journal of Combinatorics* Research Paper 62, 6 pp. (electronic), 2004. [MR2097328](https://www.ams.org/mathscinet-getitem?mr=2097328)

[9] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces.* Springer (1991). [MR1102015](https://www.ams.org/mathscinet-getitem?mr=1102015)

[10] Françoise Lust-Piquard and Gilles Pisier. “Non commutative Khintchine and Paley inequalities.” *Arkiv för Matematik*, 29(2): 241–260, 1991. [MR1150376](https://www.ams.org/mathscinet-getitem?mr=1150376)

[11] Shahar Mendelson and Alain Pajor. “On singular values of matrices with independent rows”. *Bernoulli* 12(5): 761–773, 2006. [MR2265341](https://www.ams.org/mathscinet-getitem?mr=2265341)

[12] Mark Rudelson. “Random vectors in the isotropic position.” *Journal of Functional Analysis*, 164 (1): 60–72, 1999. [MR1694526](https://www.ams.org/mathscinet-getitem?mr=1694526)

[13] Mark Rudelson and Roman Vershynin. “Sampling from large matrices: an approach through geometric functional analysis.” *Journal of the ACM* 54(4): Article 21, 2007. [MR2351844](https://www.ams.org/mathscinet-getitem?mr=2351844)

[14] Daniel Spielman and Nikhil Srivastava. “Graph sparsification by effective resistances.” In Proceedings of the 40th annual ACM symposium on Theory of Computing (STOC 2008).

[15] Colin J. Thompson. “Inequality with applications in statistical mechanics.” *Journal of Mathematical Physics*, 6: 1812–1813, 1965. [MR0189688](https://www.ams.org/mathscinet-getitem?mr=0189688)

[16] Joel A. Tropp. “On the conditioning of random subdictionaries.” *Applied and Computational Harmonic Analysis* 25(1): 1–24, 2008. [MR2419702](https://www.ams.org/mathscinet-getitem?mr=2419702)

[17] Roman Vershynin. “Frame expansions with erasures: an approach through the non-commutative operator theory.” *Applied and Computational Harmonic Analysis* 18: 167–176, 2005. [MR2121508](https://www.ams.org/mathscinet-getitem?mr=2121508)

[18] Roman Vershynin. “Spectral norm of products of random and deterministic matrices." To appear in *Probability Theory and Related Fields*. 