On some extensions of Bernstein’s inequality for self-adjoint operators

Stanislav Minsker

Abstract: We present some extensions of Bernstein’s inequality for random self-adjoint operators. The main feature of these results is that they can be applied in the infinite-dimensional setting. In particular, our inequalities refine the previous results of Hsu, Kakade and Zhang.

1. Introduction

Theoretical analysis of many problems, such as low-rank matrix recovery and approximate matrix multiplication, is built upon exponential bounds for \( \mathbb{P}(\left\| \sum_{i} X_i \right\| > t) \) where \( \{X_i\} \) is a finite sequence of self-adjoint random matrices and \( \| \cdot \| \) is the operator norm. Starting with the pioneering work of R. Ahlswede and A. Winter [AW02], the moment-generating function technique was used to produce generalizations of Chernoff, Bernstein and Friedman inequalities to the noncommutative case; see [Tro11b],[Tro11a],[Oli10] for thorough treatment and applications. While being sufficient for most problems, the explicit dependence on the dimension of the matrix does not allow straightforward application of these results in the infinite-dimensional setting.

The main purpose of this note is to provide a dimension-free version of Bernstein inequality for a sequence of independent random matrices as well as for the case of martingale differences. Some results in this direction were previously obtained in [HKZ11], but with a suboptimal tail. The trace quantity appearing in our bounds never exceeds the dimension of the matrix, therefore this result can be seen as a generalization of the finite-dimensional case.

We proceed by stating the main results and giving some applications to estimation of the integral operators.

2. Bernstein’s inequality for independent random matrices

We start with a version of Bernstein’s inequality for the sequence of independent self-adjoint random matrices. Everywhere below, \( \| \cdot \| \) stands for the operator norm \( \| A \| := \max |\lambda_i(A)| \), where \( \lambda_i \) are the eigenvalues of a self-adjoint operator \( A \). Moreover, expectation \( \mathbb{E}X \) is taken elementwise.

Let \( \psi_\sigma(t) := \frac{t^2/2}{\sigma^2 + t^3} \).

Theorem 2.1. Let \( X_1, \ldots, X_n \) be a sequence of \( d \times d \) independent self-adjoint random matrices such that \( \mathbb{E}X_i = 0 \) and \( \| X_i \| \leq 1 \) a.s.
Denote $\sigma^2 := \left\| \sum_{i=1}^{n} \mathbb{E}X_i^2 \right\|$. Then, for any $t > 0$

$$P \left( \left\| \sum_{i=1}^{n} X_i \right\| > t \right) \leq 2 \frac{\text{tr} \left( \sum_{i=1}^{n} \mathbb{E}X_i^2 \right)}{\sigma^2} \exp \left( -\Psi_\sigma(t) \right) \cdot r_\sigma(t)$$

where $r_\sigma(t) = 1 + \frac{6}{\ell^2 \log(1+t/\sigma^2)}$.

Remarks: Note that

1. $\frac{1}{\sigma^2} \text{tr} \left( \sum_{i=1}^{n} \mathbb{E}X_i^2 \right) \leq d$ (in fact, if $\sum_{i=1}^{n} \mathbb{E}X_i^2$ is "approximately low rank", i.e. has many small eigenvalues, $\frac{1}{\sigma^2} \text{tr} \left( \sum_{i=1}^{n} \mathbb{E}X_i^2 \right)$ can be much smaller than $d$).

2. $r_\sigma(t)$ is decreasing, so in the range of $t$ when the inequality becomes nontrivial (e.g., for $t \gtrsim \sqrt{\log \left( (\sigma^2)^{-1} \text{tr} \left( \sum_{i=1}^{n} \mathbb{E}X_i^2 \right) \right)}$), $r_\sigma$ can be replaced by a constant.

Proof. The proof follows the lines of [Tro11b], where the key role is played by Lieb’s concavity theorem [Lie73]:

Theorem (Lieb). Given a fixed self-adjoint matrix $H$, the function

$$A \mapsto \text{tr} \exp(H + \log A)$$

is concave on a positive definite cone.

In [Tro11b], section 4.8, the advantages over the classic method of Ahlswede and Winter based on the Golden-Thompson inequality are discussed.

Let $\phi(\theta) = e^\theta - \theta - 1$. Note that $\phi$ is nonnegative and increasing on $(0, \infty)$. Denote $S_n := \sum_{i=1}^{n} X_i$ and note that $\sigma^2 = \|\mathbb{E}S_n^2\|$. First, we reduce the bounds on probability to the bounds on moment generating functions through a chain of simple inequalities. Let $\theta > 0$; we have

$$P \left( \lambda_{\max} \left( S_n \right) > t \right) = P \left( \lambda_{\max} \left( \theta S_n \right) > \theta t \right) = P \left( \lambda_{\max} \left( \phi(\theta S_n) \right) > \phi(\theta t) \right) \leq \frac{\text{Etr} \phi(\theta S_n)}{\phi(\theta t)}. \quad (2.1)$$

The following semidefinite relation is straightforward:

$$\log \mathbb{E}e^{\theta X_i} \leq \phi(\theta) \mathbb{E}X_i^2. \quad (2.2)$$
Indeed, writing the series expansion for $e^{\theta X_i}$ and using that $E X_i = 0$, we obtain
\[
E e^{\theta X_i} = I_d + \mathbb{E} \theta^2 X_i^2 \left( \frac{1}{2!} + \ldots + \frac{(\theta X_i)^k}{(k+1)!} + \ldots \right) \leq \]
\[
\leq I_d + \theta^2 \mathbb{E} X_i^2 \left( \frac{1}{2!} + \ldots + \frac{\theta^k \|X_i\|^k}{(k+1)!} + \ldots \right) = \]
\[
= I_d + \theta^2 \mathbb{E} X_i^2 \left( \frac{e^{\theta \|X_i\|} - \theta \|X_i\| - 1}{\theta^2 \|X_i\|^2} \right) \leq \]
\[
\leq I_d + \mathbb{E} X_i^2 \phi(\theta),
\]
where in the last line we used the assumption that $\|X_i\| \leq 1$ and monotonicity of $e^{u} - u - 1/u$. It remains to apply the inequality $I + A \preceq e^A$ which holds for self-adjoint $A := \mathbb{E} X_i^2$.

Next, since $E S_n = 0_d$, Lieb’s concavity theorem and Jensen’s inequality for conditional expectation imply
\[
E \text{tr} \phi(\theta S_n) = E \text{tr} \left( \exp(\theta S_{n-1} + \log e^{\theta X_n}) - I_d \right) = \]
\[
= E \mathbb{E} \left( \text{tr} \left( \exp(\theta S_{n-1} + \log e^{\theta X_n}) - I_d \mid X_1 \ldots X_{n-1} \right) \right) \leq E \left( \text{tr} \left( \exp(\theta S_{n-1} + \log \mathbb{E} e^{\theta X_n}) - I_d \right) \right)
\]

Iterating this argument, we get
\[
E \text{tr} \phi(\theta S_n) \leq \text{tr} \left( \exp \left( \sum_{i=1}^{n} \log \mathbb{E} e^{\theta X_i} \right) - I_d \right),
\]
which together with (2.1) gives
\[
E \text{tr} \phi(\theta S_n) \leq \text{tr} \left( \exp(\phi(\theta) E S_n^2) - I_d \right). \tag{2.3}
\]

Note that
\[
\exp(\phi(\theta) E S_n^2) - I_d = \]
\[
= \phi(\theta) E^{1/2} S_n^2 \left( 1 + \frac{1}{2!} \phi(\theta) E S_n^2 + \ldots + \frac{1}{n!} (\phi(\theta) E S_n^2)^{n-1} + \ldots \right) E^{1/2} S_n^2 \leq \]
\[
\leq \phi(\theta) E S_n^2 \left( 1 + \frac{1}{2!} \phi(\theta) \|E S_n^2\| + \ldots + \frac{1}{n!} (\phi(\theta) \|E S_n^2\|)^{n-1} + \ldots \right) = \]
\[
= E S_n^2 \phi(\theta) \frac{\exp(\phi(\theta) \sigma^2) - 1}{\sigma^2 \phi(\theta)} \leq \frac{E S_n^2}{\sigma^2} \exp(\phi(\theta) \sigma^2). \tag{2.4}
\]

Combining (2.4) with (2.1), we get
\[
\mathbb{P} (\lambda_{\text{max}} (S_n) > t) \leq \text{tr} \left( \frac{E S_n^2}{\sigma^2} \right) \exp (\phi(\theta) \sigma^2) \exp (-\theta t) \frac{\exp(\theta t)}{\phi(\theta)}.
\]
Note that for $y > 0$,
\[
\frac{e^y}{\phi(y)} = 1 + \frac{1 + y}{e^y - y - 1} \leq 1 + \frac{1 + y}{y^2/2 + y^3/6} \leq 1 + \frac{6}{y^2}.
\] (2.5)
Choose $\theta_* := \log \left( 1 + \frac{t}{\sigma^2} \right)$ to minimize $\exp(\phi(\theta)\sigma^2 - \theta t)$. Together with the well-known inequality
\[
(1 + y) \log(1 + y) - y \geq \frac{y^2}{2} + \frac{y^3}{6}, \quad y \geq 0
\] (2.6)
and (2.5), this concludes the proof. It remains to repeat the argument with $X_i$'s replaced by $(-X_i)$'s to obtain a bound for the operator norm.

3. Bernstein’s inequality for the sums of martingale differences

Our next goal is to obtain a concentration inequality for the sums of matrix-valued martingale differences. Although we get a slightly weaker bound compared to the previous inequality, it still improves the multiplicative dimension factor. For $t \in \mathbb{R}$, define $p(t) := \min(-t, 1)$. Note that
1. $p(t)$ is concave;
2. $g(t) := e^t - 1 + p(t)$ is non-negative for all $t$ and increasing for $t > 0$.

Recall the following useful result:

**Proposition 3.1** (Peierls inequality). Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a convex function and $\{u_1, \ldots, u_n\}$ - any orthonormal basis of $\mathbb{C}^n$. For any self-adjoint $A \in \mathbb{C}^{n \times n}$
\[
\sum_{i=1}^{n} f((u_i, Au_i)) \leq \text{tr}(f(A))
\]
An immediate corollary of this fact is that $A \mapsto \text{tr}(f(A))$ is convex for a convex real-valued $f$ and self-adjoint $A$: to show that
\[
\text{tr} \left( f \left( \frac{A + B}{2} \right) \right) \leq \frac{1}{2} \left( \text{tr}(f(A)) + \text{tr}(f(B)) \right),
\]
it is enough to apply Peierls inequality to the orthonormal system given by the eigenvectors of $(A + B)$.
In particular, since $p(t)$ is concave, it follows from Jensen’s inequality that for any random self-adjoint matrix $Y$
\[
\mathbb{E} \text{tr} p(Y) \leq \text{tr} p(\mathbb{E} Y). \quad (3.1)
\]
Everywhere below, $\mathbb{E}_i[\cdot]$ stands for the conditional expectation $\mathbb{E}[\cdot | X_1, \ldots, X_i]$. We are ready to prove the main result of this section:

**Theorem 3.1.** Let $X_1, \ldots, X_n$ be a sequence of martingale differences with values in the set of $d \times d$ self-adjoint matrices and such that $\|X_i\| \leq 1$ a.s.
Denote $W_n := \sum_{i=1}^{n} \mathbb{E}_{i-1} X_i^2$. Then, for any $t > 0$
\[
\mathbb{P} \left( \left\| \sum_{i=1}^{n} X_i \right\| > t, \lambda_{\max}(W_n) \leq \sigma^2 \right) \leq 2 \text{tr} \left[ p \left( -\frac{t}{\sigma^2} \mathbb{E} W_n \right) \right] \exp \left[ -\Psi_\sigma(t) \cdot v_\sigma(t) \right],
\]
where $v_\sigma(t) = 1 + \frac{6}{\Psi_\sigma(t)}$.
Remarks: Note that

1. \( \text{tr} p \left( -\frac{1}{\sigma^2} \mathbb{E} W_n \right) \leq d \) for all \( t > 0 \);
2. \( v_\sigma(t) \) is decreasing, so whenever \( \Psi_\sigma(t) \gtrsim 1 \), \( v_\sigma(t) \) can be replaced by a constant.

Proof. Recall that \( \phi(\theta) = e^\theta - \theta - 1 \). Denote \( S_n := \sum_{i=1}^n X_i \).

Let \( \theta \) be such that \( \theta t - \phi(\theta)\sigma^2 > 0 \) and define an event \( E \) by

\[
E := \{ \lambda_{\max}(\theta S_n - \phi(\theta)W_n) \geq \theta t - \phi(\theta)\sigma^2 \}.
\]

Note that triangle inequality implies

\[
E \supseteq \{ \lambda_{\max}(S_n) \geq t, \lambda_{\max}(W_n) \leq \sigma^2 \}.
\]

We proceed by bounding \( \mathbb{P}(E) \):

\[
\mathbb{P}(E) = \mathbb{P}(\lambda_{\max}(g(\theta S_n - \phi(\theta)W_n)) \geq \theta t - \phi(\theta)\sigma^2) \leq \text{tr} \mathbb{E}(g(\theta S_n - \phi(\theta)W_n)) \exp(\phi(\theta)\sigma^2 - \theta t) - \frac{\exp(\theta t - \phi(\theta)\sigma^2)}{\exp(\theta t - \phi(\theta)\sigma^2)} \tag{3.2}
\]

The second term in the product, \( \exp(\phi(\theta)\sigma^2 - \theta t) \), is minimized for \( \theta_* := \log(1 + \frac{t}{\sigma^2}) \) and

\[
\exp(\phi(\theta_*)\sigma^2 - \theta_* t) \leq \exp(-\Psi_\sigma(t)) \tag{3.3}
\]

by (2.6). To bound the first term in the product (3.2), note that by Lieb’s theorem

\[
Y_k := \text{tr} \exp(\theta S_k - \phi(\theta)W_k)
\]

is a supermartingale with initial value \( d \) (which can be shown similar to theorem 2.1, or see [Tro11a] for details), so that

\[
\mathbb{E} \text{tr} \exp(\theta S_n - \phi(\theta)W_n) \leq d.
\]

Together with (3.1), this gives

\[
\text{tr} \mathbb{E} g(\theta S_n - \phi(\theta)W_n) = \text{tr} \mathbb{E} (\exp(\theta S_n - \phi(\theta)W_n) - I_d + p(\theta S_n - \phi(\theta)W_n)) \leq \mathbb{E} \text{tr} p(\theta S_n - \phi(\theta)W_n) \leq \text{tr} p(\theta \mathbb{E} S_n - \phi(\theta)W_n) = \text{tr} p(-\phi(\theta)\mathbb{E} W_n). \tag{3.4}
\]

Since \( \mathbb{E} W_n \) is nonnegative definite and due to the obvious estimate

\[
\phi(\theta) \leq e^\theta - 1, \ \theta \geq 0
\]

applied for \( \theta = \theta_* \), bound (3.4) becomes

\[
\text{tr} \mathbb{E} g(\theta S_n - \phi(\theta_*)W_n) \leq \text{tr} p \left( -\frac{t}{\sigma^2} \mathbb{E} W_n \right). \tag{3.5}
\]
Finally, by (2.6)
\[
\theta_+ t - \phi(\theta_+) \sigma^2 = \frac{1}{\sigma^2} \left( (1 + t/\sigma^2) \log(1 + t/\sigma^2) - t/\sigma^2 \right) \geq \frac{t^2/2}{\sigma^2 + t/3} > 0,
\]
and we deduce from (2.5) that
\[
\exp\left( \theta_+ t - \phi(\theta_+) \sigma^2 \right) \leq \frac{1}{g\left( \theta_+ t - \phi(\theta_+) \sigma^2 \right)} \leq 1 + \frac{6}{\Psi_{\sigma}(t)^2},
\]
where \( \Psi_{\sigma}(t) = \frac{t^2/2}{\sigma^2 + t/3} \). Combination of bounds (3.3),(3.5),(3.6) concludes the proof. \( \square \)

The expression \( \text{tr} \left[ p \left( -\frac{t}{\sigma^2} \mathbb{E}W_n \right) \right] \) which replaces the dimension factor in our bound has a very simple meaning: acting on non-negative definite cone, the function \( A \mapsto p(-A) \) just truncates the eigenvalues of \( A \) on the unit level. It is easy to see that if the eigenvalues of \( \mathbb{E}W_n \) decay polynomially, i.e., \( \lambda_i(\mathbb{E}W_n) \lesssim \frac{\sigma^2}{t^p}, \ p > 1 \), then
\[
\text{tr} \left( p \left( -\frac{t}{\sigma^2} \mathbb{E}W_n \right) \right) \leq \min(d, ct^{1/p}).
\]

In particular, this gives an improvement over the bound in [HKZ11].

**Remark.** Clearly, both theorem 2.1 and theorem 3.1 easily extend to the case when \( \{X_i\} \) is a sequence of self-adjoint Hilbert-Schmidt operators \( X_i : \mathbb{H} \to \mathbb{H} \) acting on a separable Hilbert space \( \mathbb{H} \), such that \( \mathbb{E}X_i = 0 \). This can be seen, for example, by showing that Lieb’s theorem holds for this more general case. We provide another direct approach below.

Let \( L_1 \subset L_2 \subset \ldots \) be a nested sequence of finite dimensional subspaces of \( \mathbb{H} \) such that \( \bigcup L_j = \mathbb{H} \) and let \( P_{L_j} \) be an orthogonal projector on \( L_j \). For any fixed \( j \), we will apply theorems 2.1 (similarly, theorem 3.1) to a sequence of finite dimensional operators \( \{P_{L_j}X_iP_{L_j}\}_i \) mapping \( L_j \) into itself.

Note that \( \left\| \sum_{i=1}^{n} (X_i - P_{L_j}X_iP_{L_j}) \right\| \xrightarrow{j \to \infty} 0 \) almost surely, hence
\[
\mathbb{P}\left( \left\| \sum_{i=1}^{n} X_i \right\| > t \right) \leq \liminf_{j \to \infty} \mathbb{P}\left( \left\| \sum_{i=1}^{n} P_{L_j}X_iP_{L_j} \right\| > t \right).
\]

Note that, since \( A \preceq B \) implies \( \text{SAS}^* \preceq \text{SBS}^* \), taking \( A = P_{L_j}, \ B = I \) and \( S = P_{L_j}X \) gives
\[
(P_{L_j}X P_{L_j})^2 \preceq P_{L_j}X^2 P_{L_j}
\]
thus
\[
\liminf_{j \to \infty} \frac{\text{tr} \left( \sum_{i=1}^{n} \mathbb{E}(P_{L_j}X_iP_{L_j})^2 \right)}{\lambda_{\max} \left( \sum_{i=1}^{n} \mathbb{E}(P_{L_j}X_iP_{L_j})^2 \right)} \leq \liminf_{j \to \infty} \frac{\text{tr} \left( \sum_{i=1}^{n} \mathbb{E}(P_{L_j}X_iP_{L_j}) \right)}{\lambda_{\max} \left( \sum_{i=1}^{n} \mathbb{E}(P_{L_j}X_iP_{L_j})^2 \right)} \leq \frac{\text{tr} \left( \sum_{i=1}^{n} \mathbb{E}X_i^2 \right)}{\limsup_{j \to \infty} \lambda_{\max} \left( \sum_{i=1}^{n} \mathbb{E}(P_{L_j}X_iP_{L_j})^2 \right)} = \frac{\text{tr} \left( \sum_{i=1}^{n} \mathbb{E}X_i^2 \right)}{\lambda_{\max} \left( \sum_{i=1}^{n} \mathbb{E}X_i^2 \right)},
\]
where in the last step we used a simple bound
\[ \|X^2 - (P_{L_j} XP_{L_j})^2\| = \|X(X - P_{L_j} XP_{L_j}) + (X - P_{L_j} XP_{L_j})P_{L_j} XP_{L_j}\| \leq 2\|X\|\|X - P_{L_j} XP_{L_j}\| \to 0, \text{ almost surely.} \]

4. Application: estimation of the integral operators.

Let \((S, \Pi)\) be a measurable space, with \(\Pi\) being a probability measure. Let \(K(\cdot, \cdot)\) be a symmetric continuous positive definite kernel with \(\kappa := \sup_{x \in S} |K(x, x)| < \infty\) and let \(H_K\) be the corresponding reproducing kernel Hilbert space. For \(x \in S\), let \(K_x(\cdot) := K(\cdot, x)\).

Define the integral operator \(L_K : H_K \to H_K\) by
\[
(L_K f)(x) := \int_S K(x, y) f(y) d\Pi(y) = \int_S K(x, y) \langle K_y, f \rangle_{H_K} d\Pi(y)
\]
where the second equality follows from the reproducing property. Note that \(L_K\) is self-adjoint and trace-class, with \(\text{tr} \, L_K = \mathbb{E} K(X, X)\). In many problems, \(\Pi\) is unknown and \(L_K\) is approximated by its empirical version \(L_{K,n}\):
\[
(L_{K,n} f)(x) := \frac{1}{n} \sum_{i=1}^{n} \langle f, K_{X_i} \rangle K_{X_i}(x)
\]
where \(X_1, \ldots, X_n\) is an iid sample from \(\Pi\). The natural question to ask is: what is the degree of approximation provided by \(L_{K,n}\), measured in the operator norm? Theorem 2.1 gives an answer to this question. To apply the theorem, define the operator-valued random variables
\[
\xi_i := \langle \cdot, K_{X_i} \rangle K_{X_i} - L_K.
\]
Note that \(\xi_i\)'s are iid with mean zero. Setting \(u_i = \frac{K_{X_i}}{\|K_{X_i}\|_{H_K}}\), we have
\[
\|\langle \cdot, K_{X_i} \rangle K_{X_i}\| = \|K_{X_i}\|_{H_K}^2 \|\langle \cdot, u_i \rangle u_i\| \leq \|K_{X_i}\|_{H_K}^2 = K(X_i, X_i) \leq \kappa,
\]
hence \(\|\xi_i\| \leq 2\kappa\). At the same time, since \(U_i := \langle \cdot, u_i \rangle u_i\) is a projector, it satisfies \(U_i^2 = U_i\) and
\[
\|\mathbb{E} \xi_i^2\| \leq \mathbb{E} \langle \cdot, K_{X_i} \rangle K_{X_i} \leq \mathbb{E} \|K_{X_i}\|_{H_K}^2 \|\langle \cdot, K_{X_i} \rangle K_{X_i}\| \leq \kappa \mathbb{E} \|K_{X_i}\|_{H_K}^2 = \kappa \mathbb{E} K(X, X).
\]

Note that in many cases \(\mathbb{E} K(X, X)\) is much smaller than \(\kappa\). Applying theorem 2.1, we get

**Corollary 4.1.** **Under our assumptions on the kernel,**
\[
\mathbb{P} (\|L_{K,n} - L_K\| > t) \leq 2 \frac{\text{tr} \mathbb{E} \xi_i^2}{\|\mathbb{E} \xi_i^2\|} \exp \left( -\frac{nt^2}{2\kappa (\mathbb{E} K(X, X) + 2t/3)} \right) (1 + \gamma(t)),
\]
where \(\gamma(t) = \left( \frac{6\kappa^2}{n^2 t^2 \log^2 \left( 1 + \frac{t}{\|\mathbb{E} \xi_i^2\|} \right)} \right)^{1/2} \).

This can be used together with the fact that
\[
\|L_{K,n} - L_K\| \geq \sup_j |\lambda_j(L_K) - \lambda_j(L_{K,n})|
\]
where the eigenvalues of \(L_K\) and \(L_{K,n}\) are ordered increasingly. In particular, in many cases our bound improves upon the estimate of Proposition 1 in [SZ09].
References

[AW02] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Trans. Inform. Theory*, 48(3):569–579, 2002.

[Car09] E. Carlen. Trace inequalities and quantum entropy: an introductory course. Lecture notes, 2009. Available at: http://www.mathphys.org/AZschool/material/AZ09-carlen.pdf.

[HJ85] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1985.

[HKZ11] D. Hsu, S. M. Kakade, and T. Zhang. Dimension-free tail inequalities for sums of random matrices. Preprint, 2011. Available at: http://arxiv.org/abs/1104.1672.

[KG00] V. Koltchinskii and E. Giné. Random matrix approximation of spectra of integral operators. *Bernoulli*, 6(1):113–167, 2000.

[Kol11] V. Koltchinskii. *Oracle inequalities in empirical risk minimization and sparse recovery problems*. Springer, 2011. Lectures from the 38th Probability Summer School held in Saint-Flour, 2008, École d’Été de Probabilités de Saint-Flour.

[Lie73] E. H. Lieb. Convex trace functions and the Wigner-Yanase-Dyson conjecture. *Advances in Math.*, 11:267–288, 1973.

[MP06] S. Mendelson and A. Pajor. On singular values of matrices with independent rows. *Bernoulli*, 12(5):761–773, 2006.

[Oli10] R. I. Oliveira. Concentration of the adjacency matrix and of the laplacian in random graphs with independent edges. preprint, 2010. http://arxiv.org/abs/arXiv:0911.0600.

[SZ09] S. Smale and D.-X. Zhou. Geometry on probability spaces. *Constr. Approx.*, 30(3):311–323, 2009.

[Tro11a] J. Tropp. Freedman’s inequality for matrix martingales. *Electron. Commun. Probab.*, 16:262–270, 2011.

[Tro11b] J. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, pages 1–46, 2011.