

Concentration and moment inequalities for heavy-tailed random matrices

Stanislav Minsker^{*1}, Yiqiu Shen² and Martin Wahl³

¹*Department of Mathematics
University of Southern California
Los Angeles, CA 90089
e-mail: minsker@usc.edu*

²*Department of Data Sciences and Operations
University of Southern California
Los Angeles, CA 90089
e-mail: yiqiushe@usc.edu*

³*Department of Mathematics
Bielefeld University
33615 Bielefeld, Germany
e-mail: martin.wahl@math.uni-bielefeld.de*

Abstract: We prove Fuk-Nagaev and Rosenthal-type inequalities for the sums of independent random matrices, focusing on the situation when the norms of the matrices possess finite moments of only low orders. Our bounds depend on the “intrinsic” dimensional characteristics such as the effective rank, as opposed to the dimension of the ambient space. We illustrate the advantages of such results in several applications, including new moment inequalities for the sample covariance operators of heavy-tailed distributions. Moreover, we demonstrate that our techniques yield sharpened versions of the moment inequalities for empirical processes.

MSC 2010 subject classifications: Primary 60E15.

Keywords and phrases: Fuk-Nagaev inequality, Rosenthal’s inequality, random matrix, covariance estimation, heavy tails.

1. Introduction

Fuk-Nagaev inequalities [Fuk and Nagaev, 1971, Nagaev, 1979] generalize exponential deviation inequalities for the sums of independent random variables, such as Bernstein’s, Prokhorov’s and Bennett’s inequalities, to case when the random variables satisfy minimal integrability conditions. For example, a corollary of Fuk and Nagaev’s results is the following bound: for a sequence of independent, centered random variables $X_1, \dots, X_n$ such that $\max_k \mathbb{E}|X_k|^p < \infty$ for some $p \geq 2$,

$$\mathbb{P}\left(\left|\sum_{k=1}^n X_k\right| \geq t\right) \leq 2 \exp\left(-C_1(p) \frac{t^2}{\sum_{k=1}^n \mathbb{E}X_k^2}\right) + \mathbb{P}\left(\max_k |X_k| > t/4\right) + C_2(p) \left(\frac{\sum_{k=1}^n \mathbb{E}|X_k|^p}{t^p}\right)^2.$$

Nagaev [1979] describes the applications of such results to the laws of large numbers and moment inequalities. Later, Einmahl and Li [2008], Rio [2017] and Bakhshizadeh et al. [2023], among others, improved the original estimates by Fuk and Nagaev in several ways: first, the inequalities

*S. Minsker acknowledges support by the National Science Foundation grant DMS CAREER-2045068.

were extended to martingales and Banach-space valued random variables, and second, the constants were sharpened. For example, inequalities due to [Rio \[2017\]](#), [Bakhshizadeh et al. \[2023\]](#) hold with $C_1(p) = 1/2$ and $C_2(p)$ of order p^p . The latter fact is important as the order of growth of these constants translates into the tail behavior of $|\sum_{k=1}^n X_k|$.

The goal of this work is to prove a version of the Fuk-Nagaev inequality for the sums of independent random matrices and use it to sharpen existing moment inequalities. Let $W_1, \dots, W_n \in \mathbb{C}^{d \times d}$ be a sequence of independent self-adjoint¹ random matrices such that $\mathbb{E}W_k = 0_{d \times d}$ for all k and where the expectation is taken element-wise. Assume that for all k , $\|W_k\| \leq U$ with probability 1 where $\|\cdot\|$ stands for the operator (spectral) norm. A line of work by [Ahlsweede and Winter \[2002\]](#), [Oliveira \[2010\]](#), [Tropp \[2012\]](#) culminated in the following version of the “matrix Bernstein’s” inequality: for all $t > 0$,

$$\mathbb{P}\left(\left\|\sum_{k=1}^n W_k\right\| \geq t\right) \leq 2d \exp\left(-\frac{t^2/2}{\sigma^2 + Ut/3}\right)$$

where $\sigma^2 = \|\sum_{k=1}^n \mathbb{E}W_k^2\|$. An attractive feature of this inequality (as opposed to, say, Talagrand’s concentration inequality [[Talagrand, 1996](#)]) is that it yields a bound for $\mathbb{E}\|\sum_{k=1}^n W_k\|$, namely, that $\mathbb{E}\|\sum_{k=1}^n W_k\| \leq K\left(\sigma\sqrt{\log(d)} + U\log(d)\right)$ for some absolute constant $K > 0$. Tropp’s results have been extended in two directions: first, it was shown by [Minsker \[2017\]](#), [Tropp \[2015\]](#) that the dimension factor d can essentially be replaced by the so-called *effective rank* $r(\sum_{k=1}^n \mathbb{E}W_k^2)$ where $r(A) := \frac{\text{trace}(A)}{\|A\|}$ for a positive definite matrix A . In particular, this version of Bernstein’s inequality is applicable in the context of random Hilbert-Schmidt operators acting on Hilbert spaces. Second, the boundedness assumption was relaxed by [Koltchinskii \[2011\]](#) to the requirement that $\max_k \|\|W_k\|\|_{\psi_1} < \infty$ where the ψ_1 norm of a random variable Z is defined via $\|Z\|_{\psi_1} = \inf\left\{r > 0 : \mathbb{E}e^{\left|\frac{Z}{r}\right|} \leq 2\right\}$. Finally, [Klochkov and Zhivotovskiy \[2020\]](#) showed that

$$\mathbb{P}\left(\left\|\sum_{k=1}^n W_k\right\| \geq t\right) \leq c_1 r(V_n^2) \exp\left(-c_2 \frac{t^2}{\|V_n^2\| + Rt}\right)$$

for all $t \geq c_3(\|V_n^2\|^{1/2} + M)$ where V_n^2 is any matrix satisfying $V_n^2 \succeq \sum_{k=1}^n \mathbb{E}W_k^2$ and $R = \|\max_k \|W_k\|\|_{\psi_1}$. Our results allow to relax the integrability assumptions even further and cover the case of heavy-tailed random matrices, namely, random matrices such that $\mathbb{E}\|W\|^p = \infty$ for some $p > 0$ (however, we are still able to recover the known bounds for the “light-tailed” random matrices). For example, Theorem 2.2 below implies that for all $p \geq 1$,

$$\left(\mathbb{E}\left\|\sum_{k=1}^n W_k\right\|^p\right)^{1/p} \leq K\left(\|V_n^2\|^{1/2}\sqrt{q} + q\mathbb{E}\max_k \|W_k\| + \frac{p}{\log(ep)}(\mathbb{E}\max_k \|W_k\|^p)^{1/p}\right)$$

where $q = \log(r(V_n^2)) \vee p$ and K is an absolute constant. This inequality sharpens previously known results of this type by [Junge and Zeng \[2013\]](#), [Dirksen \[2011\]](#), [Chen et al. \[2012\]](#): for example, a version of Rosenthal’s inequality by [Chen et al. \[2012\]](#) states that

$$\left(\mathbb{E}\left\|\sum_{k=1}^n W_k\right\|^p\right)^{1/p} \leq K\left(\|V_n^2\|^{1/2}\sqrt{r} + r\mathbb{E}^{1/p}\max_k \|W_k\|^p\right)$$

¹It is well-known that the case of general rectangular matrices reduces to this one via the so-called “Hermitian dilation,” see [[Tropp, 2015](#), section 2.1.17].

where $r = \log(d) \vee p$. The fact that our bound depends on $r(V_n^2)$ instead of d immediately allows one to extend it to Hilbert-Schmidt operators acting on Hilbert spaces.

Finally, let us remark that the order of the constants in the inequality stated above is optimal: in the scalar case $d = 1$, it is known [Johnson et al., 1985] that without any additional assumptions, the best order of $C(p)$ in the inequality

$$\mathbb{E}^{1/p} \left| \sum_{k=1}^n W_k \right|^p \leq C(p) \left(\left(\sum_{k=1}^n \mathbb{E} W_k^2 \right)^{1/2} + \left(\sum_{k=1}^n \mathbb{E} |W_k|^p \right)^{1/p} \right)$$

is $C(p) = K \frac{p}{\log(p)}$ while Pinelis and Utev [1985] showed that $C_1(p) = K\sqrt{p}$ and $C_2(p) = Kp$ are the best possible in the inequality of the form

$$\mathbb{E}^{1/p} \left| \sum_{k=1}^n W_k \right|^p \leq C_1(p) \left(\sum_{k=1}^n \mathbb{E} W_k^2 \right)^{1/2} + C_2(p) \mathbb{E}^{1/p} \max_k |W_k|^p.$$

It is clear that our results yield a sharper version the inequality (1) for large p whenever $\mathbb{E} \max_k |W_k|$ is much smaller than $\mathbb{E}^{1/p} \max_k |W_k|^p$.

1.1. Organization of the paper.

The rest of the exposition is organized as follows: we present the main results and their proofs in sections 2.1 and 2.2. Applications of the developed techniques to the empirical processes is discussed in section 2.3 while implications for the problems of matrix subsampling and covariance estimation are described in sections 3.1 and 3.2 respectively. Finally, section 4 contains the required background and proofs of the lemmas that were omitted from the main exposition.

2. Main results

In this section we state the new concentration and moment bounds - Theorems 2.1 and 2.2. The required notation will be introduced on demand. Let us remark that throughout the paper, the values of constants K , c , $C(\cdot)$ is often unspecified and can change from line to line; we use K and c to denote absolute constants and $C(\cdot)$ to denote constants whose value depends on the parameters in brackets.

2.1. Fuk-Nagaev-type inequality

The following proposition is the key technical result that will serve as the starting point for the derivation of the main results. Therefore, we state it separately. Everywhere below, $M = \max_{k=1, \dots, n} \|W_k\|$, $Q_{1/2}(Z)$ stands for the median of a real-valued random variable Z , and $\varepsilon_1, \dots, \varepsilon_n$ denote independent symmetric Bernoulli random variables that are independent from $W_1, \dots, W_n$. Finally, $\|\cdot\|_2$ is the Euclidean norm.

Proposition 2.1. *Let $W_1, \dots, W_n \in \mathbb{C}^{d \times d}$ be a sequence of centered, independent, self-adjoint random matrices. Let $U > 0$, and assume that V_n^2 satisfies $V_n^2 \succeq \sum_k \mathbb{E} W_k^2 I\{\|W_k\| \leq U\}$. Finally, set $\sigma_U^2 = \|V_n^2\|$ where $r(\cdot)$ represents the effective rank. Then, whenever*

$$\frac{t}{2} \geq \sigma_U \vee U/3 \vee \sup_{\|v\|_2=1} Q_{1/2} \left(\left\langle \left(\sum_{k=1}^n W_k \right) v, v \right\rangle \right),$$

the following inequality holds:

$$\begin{aligned} \mathbb{P} \left(\left\| \sum_{k=1}^n W_k \right\| > 12t \right) &\leq 64r(V_n^2) \exp \left[-\frac{(t/2)^2}{\sigma_U^2 + tU/6} \right] \\ &\quad + 16\mathbb{P} \left(\left\| \sum_{k=1}^n \varepsilon_k W_k \mathbf{1}_{\{\|W_k\| > U\}} \right\| > t/2 \right) \mathbb{P} \left(\left\| \sum_{k=1}^n \varepsilon_k W_k \right\| > t \right) + 4\mathbb{P}(M > t). \end{aligned}$$

If the random matrices are symmetrically distributed (that is, W_j and $-W_j$ are equidistributed for all j), then

$$\begin{aligned} \mathbb{P} \left(\left\| \sum_{k=1}^n W_k \right\| > 3t \right) &\leq 16r(V_n^2) \exp \left[-\frac{(t/2)^2}{\sigma^2 + tU/6} \right] \\ &\quad + 4\mathbb{P} \left(\left\| \sum_{k=1}^n W_k \mathbf{1}_{\{\|W_k\| > U\}} \right\| > t/2 \right) \mathbb{P} \left(\left\| \sum_{k=1}^n W_k \right\| > t \right) + \mathbb{P}(M > t) \end{aligned}$$

under the assumption that $\frac{t}{2} \geq \sigma_U \vee U/3$.

Remark 2.1. If $W_1, \dots, W_n$ have symmetric distribution, then $\mathbb{P}(\|\sum_{k=1}^n W_k\| > t) \geq \frac{1}{2}\mathbb{P}(M > t)$ in view of Lévy's inequality [Ledoux and Talagrand, 2013, Proposition 2.3]. This shows that the quantile U is necessary in the lower bound for t . The term σ_U is also known to be necessary – for instance, take $W_j = \xi_j A_j$ where $\xi_1, \dots, \xi_n$ are i.i.d. $N(0, 1)$ random variables and $A_1, \dots, A_n$ are fixed self-adjoint matrices Tropp [2012].

Proof. Let us reduce the general case to the situation when $W_1, \dots, W_n$ are symmetric. To this end, it suffices to apply Lemma 2.3.7 in van der Vaart and Wellner [1996]: it implies that whenever $6t \geq \sup_{\|v\|_2=1} Q_{1/2}(\langle (\sum_{k=1}^n W_k) v, v \rangle)$,

$$\mathbb{P} \left(\left\| \sum_{k=1}^n W_k \right\| > 12t \right) \leq 4\mathbb{P} \left(\left\| \sum_{k=1}^n \varepsilon_k W_k \right\| > 3t \right).$$

Obviously, this inequality holds without any assumptions if $W_1, \dots, W_k$ are symmetrically distributed. Next, in view of Hoffmann-Jørgensen inequality (Proposition 4.1),

$$\mathbb{P} \left(\left\| \sum_{k=1}^n \varepsilon_k W_k \right\| > 3t \right) \leq 4\mathbb{P} \left(\left\| \sum_{k=1}^n \varepsilon_k W_k \right\| > t \right)^2 + \mathbb{P}(M > t).$$

Given $U > 0$, we define, for each $k = 1, \dots, n$,

$$\widetilde{W}_k := \varepsilon_k W_k \mathbf{1}_{\{\|W_k\| \leq U\}} \text{ and } \Delta_k := \varepsilon_k W_k \mathbf{1}_{\{\|W_k\| > U\}}.$$

Clearly, $\|\sum_{k=1}^n \varepsilon_k W_k\| \leq \|\sum_{k=1}^n \widetilde{W}_k\| + \|\sum_{k=1}^n \Delta_k\|$, all the random matrices $\widetilde{W}_k$, Δ_k are symmetric, and

$$\mathbb{P} \left(\left\| \sum_{k=1}^n \varepsilon_k W_k \right\| > t \right) \leq \mathbb{P} \left(\left\| \sum_{k=1}^n \widetilde{W}_k \right\| > t/2 \right) + \mathbb{P} \left(\left\| \sum_{k=1}^n \Delta_k \right\| > t/2 \right) := A_1 + A_2.$$

Therefore,

$$\mathbb{P} \left(\left\| \sum_{k=1}^n \varepsilon_k W_k \right\| > 3t \right) \leq 4A_1 + 4A_2 \mathbb{P} \left(\left\| \sum_{k=1}^n \varepsilon_k W_k \right\| > t \right) + \mathbb{P}(M > t).$$

The first term on the right-hand side of inequality (2.1) can be bounded directly via Theorem 4.1. Let V_n^2 satisfy $V_n^2 \succeq \sum_k \mathbb{E} W_k^2 I\{\|W_k\| \leq U\}$. Then for $\sigma_U^2 = \|V_n^2\|$ and t such that $\frac{t}{2} \geq \sigma_U + U/3$,

$$4A_1 \leq 16r(V_n^2) \exp \left[-\frac{(t/2)^2}{\sigma_U^2 + tU/6} \right].$$

The result follows. $\square$

We are now ready to deduce the first main result of the paper.

Theorem 2.1. *Let $W_1, \dots, W_n \in \mathbb{C}^{d \times d}$ be a sequence of centered, independent, self-adjoint random matrices, and assume that $\mathbb{E} M^p < \infty$ for some $p > 1$. Moreover, suppose that V_n^2 satisfies $V_n^2 \succeq \sum_k \mathbb{E} W_k^2$, and set $\sigma^2 = \|V_n^2\|$. Then*

$$\begin{aligned} \mathbb{P} \left(\left\| \sum_{k=1}^n W_k \right\| > 12t \right) &\leq K \left(r(V_n^2) \exp \left[-\frac{(t/2)^2}{\sigma^2 + 4t\mathbb{E} M} \right] + \mathbb{P}(M \geq t) \right. \\ &\quad \left. + \left(\frac{p}{\log(ep)} \right)^{2p} \left(\frac{\mathbb{E} M^p}{t^p} \right)^2 \right) \end{aligned}$$

whenever $t \geq 2(\sigma \vee \mathbb{E} M/3)$ and where K is an absolute constant.

Remark 2.2. *In principle, the term $\left(\frac{p}{\log(ep)} \right)^{2p} \left(\frac{\mathbb{E} M^p}{t^p} \right)^2$ can be replaced by $\left(\left(\frac{p}{\log(ep)} \right)^p \frac{\mathbb{E} M^p}{t^p} \right)^{2^k}$ for any integer $k \geq 1$, at the cost of increasing the constants appearing in the bound.*

Proof. We will continue using the notation introduced in the proof of Proposition 2.1. First of all, note that $\sup_{\|v\|_2=1} Q_{1/2} \left(\left\langle \left(\sum_{j=1}^n W_j \right) v, v \right\rangle \right) \leq \sigma\sqrt{2}$ for the choice of σ stated above. Next, plugging the inequality

$$\mathbb{P} \left(\left\| \sum_{k=1}^n \varepsilon_k W_k \right\| > t \right) \leq \mathbb{P} \left(\left\| \sum_{k=1}^n \widetilde{W}_k \right\| > t/2 \right) + \mathbb{P} \left(\left\| \sum_{k=1}^n \Delta_k \right\| > t/2 \right)$$

into relation (2.1) with V_n^2 and σ^2 specified in the conditions of the theorem, we deduce that

$$\begin{aligned} \mathbb{P} \left(\left\| \sum_{k=1}^n W_k \right\| > 12t \right) &\leq 128r(V_n^2) \exp \left[-\frac{(t/2)^2}{\sigma^2 + tU/6} \right] + 4\mathbb{P}(M > t) \\ &\quad + 16 \left(\mathbb{P} \left(\left\| \sum_{k=1}^n \Delta_k \right\| > t/2 \right) \right)^2. \end{aligned}$$

Next, we apply Markov's inequality to get the bound

$$\mathbb{P} \left(\left\| \sum_{k=1}^n \Delta_k \right\| > t/2 \right) \leq \frac{\mathbb{E} \left\| \sum_{k=1}^n \Delta_k \right\|^p}{(t/2)^p}.$$

Theorem 4.2 implies that for all $p \geq 2$,

$$\left(\mathbb{E} \left\| \sum_{k=1}^n \Delta_k \right\|^p \right)^{1/p} \leq K \frac{p}{\log(p)} \left(\mathbb{E} \left\| \sum_{k=1}^n \Delta_k \right\| + \left(\mathbb{E} \max_{k \leq n} \|\Delta_k\|^p \right)^{1/p} \right).$$

Moreover, if we set $U := 24\mathbb{E}M$, then Proposition 4.2 applies with $q = 1$ and $t_0 = 0$, the latter due to the inequality

$$\mathbb{P}\left(\left\|\sum_{k=1}^n \Delta_k\right\| > 0\right) \leq \mathbb{P}(M > U) \leq \frac{\mathbb{E}M}{U} \leq 1/24.$$

Therefore,

$$\mathbb{E}\left\|\sum_{k=1}^n \Delta_k\right\| \leq 6\mathbb{E}\max_{k \leq n} \|\Delta_k\| \leq 6\mathbb{E}\max_{k \leq n} \|W_k\| \leq 6\left(\mathbb{E}\max_{k \leq n} \|W_k\|^p\right)^{1/p},$$

implying that

$$\mathbb{P}\left(\left\|\sum_{k=1}^n \Delta_k\right\| > t/2\right) \leq \left(K \frac{p}{\log(ep)}\right)^p \frac{\mathbb{E}M^p}{t^p}.$$

Note that whenever $0 < p < 2$, the same result still holds since

$$\left(\mathbb{E}\left\|\sum_{k=1}^n \Delta_k^{(1)}\right\|^p\right)^{1/p} \leq K\mathbb{E}^{1/p} \max_{k \leq n} \|W_k\|^p$$

in view of Proposition 4.2. The conclusion follows. $\square$

Let us mention that using Markov's inequality with the exponential instead of polynomial moments leads to the version of Bernstein's inequality due to [Klochkov and Zhivotovskiy \[2020\]](#) (their argument relies on similar tools but does not require symmetrization and the Hoffmann-Jørgensen inequality).

2.2. Moment inequalities

Now we will establish moment inequalities by integrating the tail estimates of Proposition 2.1. As before, assume that V_n^2 satisfies $V_n^2 \succeq \sum_k \mathbb{E}W_k^2$ and let $\sigma^2 = \|V_n^2\|$.

Theorem 2.2. *Let $W_1, \dots, W_n \in \mathbb{C}^{d \times d}$ be a sequence of centered, independent, self-adjoint random matrices, and let $Q_p = \inf\{s > 0 : \mathbb{P}(\|\sum_{k=1}^n \Delta_k\| > s/2) \leq \frac{1}{8}3^{-p}\}$. Then for all $p \geq 1$,*

$$\left(\mathbb{E}\left\|\sum_{k=1}^n W_k\right\|^p\right)^{1/p} \leq K\left(\sigma\sqrt{q} + qQ_1 + Q_p + \mathbb{E}^{1/p}M^p\right),$$

where $q = \log(r(V_n^2)) \vee p$ and $K > 0$ is an absolute constant. In particular, we have the following “closed-form” Rosenthal-type moment inequalities:

$$\begin{aligned} \left(\mathbb{E}\left\|\sum_{k=1}^n W_k\right\|^p\right)^{1/p} &\leq K\left(\sigma\sqrt{q} + q\mathbb{E}M + \frac{p}{\log(ep)}(\mathbb{E}M^p)^{1/p}\right) \text{ and} \\ \left(\mathbb{E}\left\|\sum_{k=1}^n W_k\right\|^p\right)^{1/p} &\leq K\left(\sigma\sqrt{q} + \log(r(V_n^2))\mathbb{E}M + p\|M\|_{\psi_1}\right). \end{aligned}$$

Remark 2.3. *The well-known relation $\|M\|_{\psi_1} \leq K \log(n) \max_{k \leq n} \|X_k\|_{\psi_1}$ [[van der Vaart and Wellner, 1996](#)] could be useful when combined with the inequality (2.2).*

Proof. Observe that

$$\begin{aligned} \mathbb{E} \left\| \sum_{k=1}^n W_k \right\|^p &\leq 2^p \mathbb{E} \left\| \sum_{k=1}^n \varepsilon_k W_k \right\|^p \\ &= 2^p p \int_0^\infty t^{p-1} \mathbb{P} \left(\left\| \sum_{k=1}^n \varepsilon_k W_k \right\| \geq t \right) dt = 6^p \cdot p \int_0^\infty t^{p-1} \mathbb{P} \left(\left\| \sum_{k=1}^n \varepsilon_k W_k \right\| \geq 3t \right) dt, \end{aligned}$$

where we used the symmetrization inequality [van der Vaart and Wellner, 1996, Lemma 2.3.6] on the first step and the integration by parts formula on the second step, and the linear change of variables on the third step. To estimate the last integral, we choose $U = Q_1/2$ and apply inequality (2.1) in the range $t \geq t_0 := 2(\sigma \vee U/3)$ to deduce that

$$\begin{aligned} \mathbb{E} \left\| \sum_{k=1}^n W_k \right\|^p &\leq 12^p (\sigma \vee U/3)^p \\ &\quad + 6^p p \left(\int_0^\infty t^{p-1} \left(16r(V_n^2) \exp \left[-\frac{(t/2)^2}{\sigma^2 + tU/6} \right] \wedge 1 \right) dt \right. \\ &\quad \left. + \int_0^\infty t^{p-1} \mathbb{P}(M \geq t) dt + \int_0^\infty 4t^{p-1} \mathbb{P} \left(\left\| \sum_{k=1}^n \Delta_k \right\| > t/2 \right) \mathbb{P} \left(\left\| \sum_{k=1}^n \varepsilon_k W_k \right\| > t \right) dt \right). \end{aligned}$$

Recalling the definition of Q_p , one easily checks that

$$\begin{aligned} 6^p p \int_0^\infty 4t^{p-1} \mathbb{P} \left(\left\| \sum_{k=1}^n \Delta_k \right\| > t/2 \right) \mathbb{P} \left(\left\| \sum_{k=1}^n \varepsilon_k W_k \right\| > t \right) dt \\ \leq 4 \cdot 6^p Q_p^p + 2^{p-1} \mathbb{E} \left\| \sum_{k=1}^n \varepsilon_k W_k \right\|^p. \end{aligned}$$

Combined with the first line of display (2.2), this inequality implies that

$$\begin{aligned} 2^{p-1} \mathbb{E} \left\| \sum_{k=1}^n \varepsilon_k W_k \right\|^p &\leq 12^p (\sigma \vee U/3)^p + 6^p \mathbb{E} M^p + 4 \cdot 6^p Q_p^p \\ &\quad + 6^p p \int_0^\infty t^{p-1} \left(16r(V_n^2) \exp \left[-\frac{(t/2)^2}{\sigma^2 + tU/6} \right] \wedge 1 \right) dt. \end{aligned}$$

Application of Corollary 4.1 to the last integral yields the inequality (2.2). Finally, let us prove the inequalities (2.2) and (2.2). To this end, we need to obtain the upper bounds for the quantities Q_1 and Q_p . To estimate Q_1 , recall the inequality (2.1) which implies that $Q_1/2 \leq 24\mathbb{E}M$. Similarly, by (2.1) and Markov's inequality, we deduce that $Q_p \leq K \frac{p}{\log(ep)} \mathbb{E}^{1/p} M^p$. If, on the other hand, $\|M\|_{\psi_1} < \infty$, then the second inequality of Theorem 4.2 combined with the bound (2.1) and the well-known estimate $\mathbb{E}M \leq K\|M\|_{\psi_1}$ imply that $\|\sum_{k=1}^n \Delta_k\|_{\psi_1} \leq \|M\|_{\psi_1}$, whence

$$\mathbb{P} \left(\left\| \sum_{k=1}^n \Delta_k \right\| > t \right) \leq e^{-C \frac{t}{\|M\|_{\psi_1}}}$$

and $Q_p \leq C'p\|M\|_{\psi_1}$. Let us remark that the inequality (2.2) can be also obtained by integrating the tail bound (2.1) directly. $\square$

Next, we deduce a version of the previous result that holds for sums of nonnegative definite random matrices.

Theorem 2.3. *Let $W_1, \dots, W_n \in \mathbb{C}^{d \times d}$ be a sequence of independent, nonnegative definite matrices and let $M = \max_{j=1, \dots, n} \|W_j\|$. Moreover, let $A_n := \sum_{j=1}^n \mathbb{E}W_j$. Then for all $p \geq 1$,*

$$\begin{aligned} \left(\mathbb{E} \left\| \sum_{k=1}^n W_k \right\|^p \right)^{1/p} &\leq K \left(\|A_n\| + q \mathbb{E}M + \frac{p}{\log(ep)} (\mathbb{E}M^p)^{1/p} \right) \text{ and} \\ \left(\mathbb{E} \left\| \sum_{k=1}^n W_k \right\|^p \right)^{1/p} &\leq K (\|A_n\| + \log(r(A_n)) \mathbb{E}M + p \|M\|_{\psi_1}), \end{aligned}$$

where $q = \log(r(A_n)) \vee p$ and $K > 0$ is an absolute constant.

Proof. In view of Minkowski's inequality followed by the symmetrization inequality,

$$\begin{aligned} \left(\mathbb{E} \left\| \sum_{k=1}^n W_k \right\|^p \right)^{1/p} &\leq \|A_n\| + \left(\mathbb{E} \left\| \sum_{k=1}^n W_k - \mathbb{E}W_k \right\|^p \right)^{1/p} \\ &\leq \|A_n\| + 2 \mathbb{E} \left(\mathbb{E} \left\| \sum_{k=1}^n \varepsilon_k W_k \right\|^p \right)^{1/p}, \end{aligned}$$

where $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d random signs independent from $W_1, \dots, W_n$. To estimate the second term in the sum above, we will apply the inequality (2.2) together with the following choice of V_n^2 : recall that in (2.2), we set $U = 24\mathbb{E}M$ and note that for all j ,

$$\mathbb{E} [W_j^2 I\{\|W_j\| \leq 24\mathbb{E}M\}] \preceq 24\mathbb{E}M \cdot \mathbb{E}W_j$$

since $W_j \succeq 0$ with probability 1. This relation implies that we can set $V_n^2 = 24\mathbb{E}M \cdot A_n$, whence $r(V_n^2) = r(A_n)$. Moreover, $\sigma\sqrt{q} = \sqrt{24q\|A_n\|\mathbb{E}M} \leq \|A_n\| + 6q\mathbb{E}M$, hence (2.2) yields the bound

$$\left(\mathbb{E} \left\| \sum_{k=1}^n \varepsilon_k W_k \right\|^p \right)^{1/p} \leq K' \left(\|A_n\| + q\mathbb{E}M + \frac{p}{\log(ep)} (\mathbb{E}M^p)^{1/p} \right),$$

implying the claim. The second inequality is obtained in a similar manner where the inequality (2.2) is used in place of (2.2). $\square$

2.3. Inequalities for the empirical processes

The only part of the previous arguments that exploits the “non-commutative” nature of the random variables is the application of Matrix Bernstein's inequality. In this section, we state the results produced by our method for general empirical processes. The only required modification is the application of Bousquet's version of Talagrand's concentration inequality 4.3 in place of Bernstein's inequality. We state only the versions of Theorems 2.1 and 2.2 and remark on the key differences. The required changes to the proofs are minimal hence we avoid the details.

Let $\mathcal{F}$ be a set of measurable real-valued functions and let $X_1, \dots, X_n$ be i.i.d copies of X . Assume that $\mathbb{E}f(X) = 0$ for all $f \in \mathcal{F}$. Let us set $F(x) := \sup_{f \in \mathcal{F}} |f(x)|$, $M = \max_{k=1, \dots, n} F(X_k)$, and suppose that $\mathbb{E}M^p < \infty$ for some $p \geq 2$. Denote $Z = \sup_{f \in \mathcal{F}} \sum_{k=1}^n f(X_k)$; for simplicity, we

will assume that Z is measurable. Finally, let σ_* satisfy $\sigma_*^2 \geq n \sup_{f \in \mathcal{F}} \mathbb{E} f^2(X)$. For example, in the main case of interest of this paper, $\|\sum_{k=1}^n W_k\| = \sup_{\|v\|_2=1} \text{tr}((\sum_{k=1}^n W_k)vv^T)$ corresponding to $\mathcal{F} = \{f_v(\cdot) = \text{tr}((\cdot)vv^T), \|v\|_2 = 1\}$. The following result can be viewed as an extension of Adamczak's inequality [Adamczak, 2008] to the heavy-tailed case.

Theorem 2.4. *For all $t \geq \sqrt{2}\sigma_*$,*

$$\mathbb{P}(Z > 24(\mathbb{E}Z + t)) \leq K \left(\exp\left(-\frac{t^2}{2\sigma_*^2 + 64t\mathbb{E}M}\right) + \mathbb{P}(M \geq t) + \left(\frac{p}{\log(ep)}\right)^{2p} \left(\frac{\mathbb{E}M^p}{t^p}\right)^2 \right)$$

where K is an absolute constant.

Integrating this tail bound, we obtain the following moment inequalities.

Theorem 2.5. *For all $p \geq 1$,*

$$(\mathbb{E}Z^p)^{1/p} \leq K \left(\mathbb{E}Z + \sigma_*\sqrt{p} + p\mathbb{E}M + \frac{p}{\log(ep)}(\mathbb{E}M^p)^{1/p} \right).$$

Let us compare this result with the bound of Theorem 2.2. When applied to the sums of random matrices, we get the inequality

$$\left(\mathbb{E} \left\| \sum_{k=1}^n W_k \right\|^p \right)^{1/p} \leq K \left(\mathbb{E} \left\| \sum_{k=1}^n W_k \right\| + \sigma_*\sqrt{p} + p\mathbb{E}M + \frac{p}{\log(ep)}(\mathbb{E}M^p)^{1/p} \right)$$

The main difference is that this bounds includes $\mathbb{E} \|\sum_{k=1}^n W_k\|$ on the right-hand side. However, for large values of p , it is better than (2.2) since σ_*^2 can be much smaller than $\|\sum_k \mathbb{E}W_k^2\|$. Moreover, Theorem 2.5 improves upon the inequality proved by [Boucheron et al., 2005, Theorem 12]: the latter states that for all $p \geq 2$,

$$(\mathbb{E}Z^p)^{1/p} \leq K \left(\mathbb{E}Z + \sigma_*\sqrt{p} + p(\mathbb{E}M^p)^{1/p} \right).$$

The estimate provided by (2.5) is better for large values of p if $\mathbb{E}M$ is much smaller than $\mathbb{E}^{1/p}M^p$.

3. Applications

In this section, we apply the inequalities to get improved bounds to two classical problems - matrix subsampling and covariance estimation.

3.1. Norms of random submatrices.

Let B be a self-adjoint matrix, and let $\delta_1, \dots, \delta_d$ be i.i.d. Bernoulli random variables with $\mathbb{E}\delta_1 = \delta \in (0, 1)$. Define $R = \text{diag}(\delta_1, \dots, \delta_d)$. We are interested in the spectral norm of the matrix BR formed by the columns B_i , $i \in I$ of B with indices corresponding to the random set $I = \{1 \leq$

$i \leq d : \delta_i = 1\}$. This problem has previously been studied by Rudelson and Vershynin [2007] and Tropp [2015] who showed that

$$\mathbb{E}\|BR\|^2 \leq K \left(\delta\|B\|^2 + \frac{\log(n\delta)}{\lfloor \delta^{-1} \rfloor} \sum_{k=1}^{\lfloor \delta^{-1} \rfloor} \|B_{(k)}\|_2^2 \right)$$

and

$$\mathbb{E}\|BR\|^2 \leq 1.72 \left(\delta\|B\|^2 + \log \left(2 \frac{\|B\|_F^2}{\|B\|^2} \right) \|B_{(1)}\|_2^2 \right)$$

respectively, where $B_{(j)}$ denotes the column with the j -th largest norm and K is a numerical constant. Note that $\|B_{(1)}\|_2^2 \geq \frac{1}{\lfloor \delta^{-1} \rfloor} \sum_{k=1}^{\lfloor \delta^{-1} \rfloor} \|B_{(k)}\|_2^2$ but it is possible that $\log(n\delta) > \log \left(2 \frac{\|B\|_F^2}{\|B\|^2} \right)$ when the matrix B has small “stable rank” $\text{srnk}(B) := \frac{\|B\|_F^2}{\|B\|^2}$. Next, we will show that Tropp’s bound can be improved, and that (3.1) holds with $\log(n\delta)$ replaced with $\log(n\delta) \wedge \log(\text{srnk}(B))$. To this end, let $e_1, \dots, e_d$ denote the standard Euclidean basis, and observe that

$$\|BR\|^2 = \left\| \sum_{k=1}^d \delta_k B_k e_k^T \right\|^2 = \left\| \sum_{k=1}^d \delta_k B_k B_k^T \right\|.$$

We will apply Theorem 2.3 to the last expression with $W_k = \delta_k B_k B_k^T$. Note that $A_n = \delta B B^T$ so that $\|A_n\| = \delta\|B\|^2$, and that $\mathbb{E}M = \mathbb{E}(\max_{k=1, \dots, d} \delta_k \|B_k\|_2^2)$. According to Lemma 5.1 in [Rudelson and Vershynin, 2007] or Proposition 2.3 in [Tropp, 2008],

$$\mathbb{E} \max_{k=1, \dots, d} \delta_k \|B_k\|_2^2 \leq \frac{2}{\lfloor \delta^{-1} \rfloor} \sum_{k=1}^{\lfloor \delta^{-1} \rfloor} \|B_{(k)}\|_2^2.$$

Finally, $r(A_n) = \frac{\|B\|_F^2}{\|B\|^2}$ is the stable rank of B . We record the following bound.

Corollary 3.1. *The inequalities*

$$\mathbb{E}\|BR\|^2 \leq K \left(\delta\|B\|^2 + \log \left(\frac{\|B\|_F^2}{\|B\|^2} \right) \frac{1}{\lfloor \delta^{-1} \rfloor} \sum_{k=1}^{\lfloor \delta^{-1} \rfloor} \|B_{(k)}\|_2^2 \right)$$

and

$$\mathbb{E}\|BR - \delta B\|^2 \leq K(1 - \delta) \left(\delta\|B\|^2 + \log \left(\frac{\|B\|_F^2}{\|B\|^2} \right) \frac{1}{\lfloor \delta^{-1} \rfloor} \sum_{k=1}^{\lfloor \delta^{-1} \rfloor} \|B_{(k)}\|_2^2 \right)$$

hold for all $\delta \in (0, 1)$ and a numerical constant $K > 0$.

Proof. The first inequality has already been established. The proof of the second bound is quite similar: it suffices to note that

$$\mathbb{E}\|BR - \delta B\|^2 = \mathbb{E} \left\| \sum_{k=1}^d (\delta_k - \delta)^2 B_k B_k^T \right\|,$$

and that $\mathbb{E} \sum_{k=1}^d (\delta_k - \delta)^2 B_k B_k^T = \delta(1 - \delta)\|B\|^2$ and $\mathbb{E} \max_{k=1, \dots, d} (\delta_k - \delta)^2 \|B_k\|_2^2 \leq \frac{2(1 - \delta)}{\lfloor \delta^{-1} \rfloor} \sum_{k=1}^{\lfloor \delta^{-1} \rfloor} \|B_{(k)}\|_2^2$. $\square$

3.2. Covariance estimation.

In this section, we consider applications of our results to the covariance estimation problem. Let $X \in \mathbb{R}^d$ be a random vector such that $\mathbb{E}X = 0$ and $\mathbb{E}[XX^\top] = \Sigma$. Given a sequence $X_1, \dots, X_n \in \mathbb{R}^d$ of i.i.d. copies of X , what is an upper bound for the error of the sample covariance matrix? In other words, we would like to estimate $E_n := \left\| \frac{1}{n} \sum_{j=1}^n X_j X_j^\top - \Sigma \right\|$. One of the long-standing open questions asks for the minimal assumptions on the distribution of X such that $n = C(\varepsilon)d$ suffices to guarantee that $\mathbb{E}E_n \leq \varepsilon \|\Sigma\|$, or that $E_n \leq \varepsilon \|\Sigma\|$ with high probability. Results of this type are often referred to as the “quantitative versions of the Bai-Yin theorem,” after [Bai and Yin, 2008]. Let us give an (incomplete) overview of the long history of the problem. It has long been known that sub-Gaussian distributions satisfy the required conditions [Vershynin, 2012]. Moreover, very general and precise characterization of the behavior of the sample covariance of Banach-space valued random vectors has been found by Koltchinskii and Lounici [2017] and, very recently, further sharpened by Han [2022] in the finite-dimensional case. For the log-concave and the sub-exponential distributions, the problem was first considered by Kannan et al. [1997], and the bounds were significantly improved and refined by Bourgain [1996], Rudelson [1999]. It took much longer to remove the logarithmic factors, until the problem was finally solved by Adamczak et al. [2010, 2011]. Finally, the case of heavy-tailed distributions was investigated by Vershynin [2011], Srivastava and Vershynin [2013], Mendelson and Paouris [2014], Guédon et al. [2017], until Tikhomirov [2018] showed that $4 + \varepsilon$ moments are essentially sufficient to get the desired bound. Specifically, his results imply that if $\Sigma = I_d$ and $\sup_{\|v\|_2=1} \mathbb{E}|\langle X, v \rangle|^p = T < \infty$ for some $p > 4$, then

$$E_n \leq C(p) \left(T^{2/p} \sqrt{\frac{d}{n}} + \frac{\max_j \|X_j\|_2^2}{n} \right)$$

with probability at least $1 - 1/n$. Abdalla and Zhivotovskiy [2024] refined Tikhomirov’s estimates and essentially showed that $n = C(\varepsilon)r(\Sigma)$ samples suffices to get the desired guarantees in expectation, although they considered the sample covariance based on properly truncated random vectors. Next, we show that the results by Abdalla and Zhivotovskiy [2024] can be combined with the moment inequalities developed in this paper to get sharp moment inequalities for E_n .

Theorem 3.1. *Let $X \in \mathbb{R}^d$ be a random vector such that $\mathbb{E}X = 0$ and $\mathbb{E}[XX^\top] = \Sigma$. Let $X_1, \dots, X_n \in \mathbb{R}^d$ be i.i.d. copies of X . Assume that $\frac{r(\Sigma)}{n} \leq c$ for a sufficiently small positive constant c , and that for some $p > 4$*

$$\sup_{\|v\|_2=1} \frac{\mathbb{E}\langle X, v \rangle^p}{(\mathbb{E}\langle X, v \rangle^2)^{p/2}} = \kappa < \infty.$$

Then

$$\left(\mathbb{E} \left\| \frac{1}{n} \sum_{j=1}^n X_j X_j^\top - \Sigma \right\|^2 \right)^{1/2} \leq C(\kappa, p) \left(\|\Sigma\| \sqrt{\frac{r(\Sigma)}{n}} + \frac{\mathbb{E}^{2/p} \max_{j \leq n} \|X_j\|_2^p}{n} \right).$$

Remark 3.1. In the course of the proof, we obtain a slightly stronger inequality

$$\left(\mathbb{E} \left\| \frac{1}{n} \sum_{j=1}^n X_j X_j^\top - \Sigma \right\|^2 \right)^{1/2} \leq C(\kappa, p) \left(\|\Sigma\| \sqrt{\frac{r(\Sigma)}{n}} + \frac{\mathbb{E}^{1/2} \max_{j \leq n} \|X_j\|_2^4}{n} \wedge \frac{\mathbb{E}^{2/p} \max_{j \leq n} \|X_j\|_2^p}{n} \left(\frac{1}{n} + e^{-r(\Sigma)} \right)^{\frac{p-4}{2p}} \right).$$

Proof. Our proof builds on the results by [Abdalla and Zhivotovskiy \[2024\]](#) which in turn sharpen the inequality due to [Tikhomirov \[2018\]](#). Before we dive into the details, let us mention that the “hypercontractivity” condition (3.1) implies in particular that $\|\mathbb{E}\|X\|^2 X X^\top\| \leq \kappa^2 \text{tr}(\Sigma) \|\Sigma\|$; the proof of this fact can be found in [\[Wei and Minsker, 2017, Lemma 2.3\]](#). Note that

$$\left\| \frac{1}{n} \sum_{j=1}^n X_j X_j^\top - \Sigma \right\| = \sup_{\|v\|_2=1} \left| \frac{1}{n} \sum_{j=1}^n \langle X_j, v \rangle^2 - \mathbb{E} \langle X, v \rangle^2 \right|.$$

Let us state the following decomposition of the error into “peaky” and “spread” parts [\[Tikhomirov, 2018\]](#) that holds for arbitrary $\lambda > 0$:

$$\begin{aligned} \sup_{\|v\|_2=1} \left| \frac{1}{n} \sum_{j=1}^n \langle X_j, v \rangle^2 - \mathbb{E} \langle X, v \rangle^2 \right| &\leq \underbrace{\sup_{\|v\|_2 \leq 1} \frac{1}{n} \sum_{j=1}^n \langle X_j, v \rangle^2 \mathbf{1}_{\{\lambda \langle X_j, v \rangle^2 > 1\}}}_{\text{Peak part}} \\ &\quad + \underbrace{\sup_{\|v\|_2 \leq 1} \left| \frac{1}{\lambda n} \sum_{j=1}^n \psi(\lambda \langle X_j, v \rangle^2) - \mathbb{E} \langle X, v \rangle^2 \right|}_{\text{Spread part}}, \end{aligned}$$

where

$$\psi(x) = \begin{cases} x, & \text{for } x \in [-1, 1]; \\ \text{sign}(x) & \text{for } |x| > 1. \end{cases}$$

We will estimate the two terms separately, starting with the “spread” part. To this end, we will apply Proposition 4.1 in [\[Abdalla and Zhivotovskiy, 2024\]](#) which implies that for $\lambda = \frac{1}{\kappa^2 \|\Sigma\|} \sqrt{\frac{r(\Sigma)}{n}}$,

$$\sup_{\|v\|_2 \leq 1} \left| \frac{1}{\lambda n} \sum_{j=1}^n \psi(\lambda \langle X_j, v \rangle^2) - \mathbb{E} \langle X, v \rangle^2 \right| \leq C \kappa^2 \|\Sigma\| \left(\sqrt{\frac{r(\Sigma)}{n}} + \frac{t}{\sqrt{r(\Sigma)n}} \right)$$

with probability at least $1 - e^{-t}$. For $t = r(\Sigma)$, we get in particular that

$$\sup_{\|v\|_2 \leq 1} \left| \frac{1}{\lambda n} \sum_{j=1}^n \psi(\lambda \langle X_j, v \rangle^2) - \mathbb{E} \langle X, v \rangle^2 \right| \leq C \kappa^2 \|\Sigma\| \sqrt{\frac{r(\Sigma)}{n}}$$

with probability at least $1 - e^{-r(\Sigma)}$. Next, we will estimate the “peaky” term in the inequality (3.2). Equation (5) in the work by [Abdalla and Zhivotovskiy \[2024\]](#) states that for all subsets

$\mathcal{J} \subseteq [n]$ of cardinality at most k ,

$$\left\| \frac{1}{n} \sum_{j \in \mathcal{J}} X_j X_j^\top \right\| \leq \frac{f(k, [n])}{n}$$

for the function $f(k, [n])$ defined via

$$f(k, [n]) = \sup_{\|y\|_2=1, \|y\|_0 \leq k, \text{supp}(y) \subseteq [n]} \left\| \sum_{j=1}^n y_j X_j \right\|_2^2$$

and the bounds holds uniformly over all such subsets J . The following result provides a bound for $f(k, [n])$.

Theorem 3.2. [*Abdalla and Zhivotovskiy, 2024, Theorem 3*] Assume that $\frac{r(\Sigma)}{n} \leq c'$ for a sufficiently small positive constant c' . Then

$$f(k, [n]) \leq C(p, \kappa) \left(\max_{j \leq n} \|X_j\|_2^2 + \|\Sigma\| k \left(\frac{n}{k} \right)^{4/(4+p)} \log^4 \frac{n}{k} \right),$$

with probability at least $1 - \frac{c(p)}{n}$, and the bound holds simultaneously for all integers k satisfying $r(\Sigma) \leq k \leq c'n$.

Since $p > 4$, this result implies that

$$f(k, [n]) \leq C'(p, \kappa) \left(\max_{j \leq n} \|X_j\|_2^2 + \|\Sigma\| \sqrt{nk} \right).$$

Next, let $\lambda = \frac{1}{\kappa^2 \|\Sigma\|} \sqrt{\frac{r(\Sigma)}{n}}$. Following [Tikhomirov \[2018\]](#), [Abdalla and Zhivotovskiy \[2024\]](#), let us define the random set

$$I_v = \{j \in [n] : \langle X_j, v \rangle^2 > 1/\lambda\}$$

and $m = \sup_{\|v\|_2 \leq 1} |I_v|$. Then, in view of the inequality (3.2), we see that

$$\frac{m}{n\lambda} \leq \sup_{\|v\|_2 \leq 1} \frac{1}{n} \sum_{j=1}^n \langle X_j, v \rangle^2 \mathbf{1}_{\{\lambda \langle X_j, v \rangle^2 > 1\}} \leq \frac{f(m, [n])}{n}.$$

Now, if $r(\Sigma) \leq m \leq c'n$, then we can employ Theorem 3.2 to derive the following bound that holds with probability at least $1 - \frac{c(p)}{n}$:

$$\begin{aligned} m \leq \lambda f(m, [n]) &\leq C'(p) \left(\max_{j \leq n} \|X_j\|_2^2 + \|\Sigma\| \sqrt{nm} \right) \cdot \frac{1}{\kappa^2 \|\Sigma\|} \sqrt{\frac{r(\Sigma)}{n}} \\ &\leq \frac{C'(p)}{\kappa^2} \left(\frac{\max_{j \leq n} \|X_j\|_2^2}{\|\Sigma\|} \sqrt{\frac{r(\Sigma)}{n}} + \sqrt{m} \sqrt{r(\Sigma)} \right). \end{aligned}$$

Solutions to the inequality $x \leq a\sqrt{x} + b$ satisfy $x \leq 2 \max(a^2, b)$. Therefore, with probability at least $1 - \frac{c(p)}{n}$,

$$m \leq C_1(p) \max \left(r(\Sigma), \frac{\max_{j \leq n} \|X_j\|_2^2}{\kappa^2 \|\Sigma\|} \sqrt{\frac{r(\Sigma)}{n}} \right).$$

It remains to show that $m \leq c'n$ if $r(\Sigma) < cn$ for c small enough (clearly, if $m < r(\Sigma)$, then (3.2) holds). By the definition of I_v , for any $v \in \mathbb{R}^d$,

$$|I_v| = \sum_{j=1}^n \mathbf{1} \left\{ \langle X_j, v \rangle^2 > \kappa^2 \|\Sigma\| \sqrt{\frac{n}{r(\Sigma)}} \right\} = \sum_{j=1}^n \mathbf{1} \left\{ \frac{|\langle X_j, v \rangle|}{\left(\kappa^2 \|\Sigma\| \sqrt{\frac{n}{r(\Sigma)}} \right)^{1/2}} > 1 \right\} \\ \leq \sum_{j=1}^n \rho \left(\frac{|\langle X_j, v \rangle|}{\left(\kappa^2 \|\Sigma\| \sqrt{\frac{n}{r(\Sigma)}} \right)^{1/2}} \right),$$

where $\rho(x) = \begin{cases} 0 & x \leq 1/2 \\ 2x - 1 & x \in (1/2, 1] \\ 1 & x > 1 \end{cases}$ is such that $\mathbf{1}\{x \geq 1/2\} \geq \rho(x) \geq \mathbf{1}\{x \geq 1\}$. For brevity,

set $Z_j(v) := \frac{|\langle X_j, v \rangle|}{\left(\kappa^2 \|\Sigma\| \sqrt{\frac{n}{r(\Sigma)}} \right)^{1/2}}$. In view of Markov's inequality and assumption (3.1),

$$\mathbb{E} \rho(Z_j) \leq \mathbb{P} \left(\langle X, v \rangle^2 > \frac{\kappa^2}{2} \|\Sigma\| \sqrt{\frac{n}{r(\Sigma)}} \right) \leq \frac{4\mathbb{E} \langle X, v \rangle^4}{\kappa^4 \|\Sigma\|} \cdot \frac{r(\Sigma)}{n} \leq \frac{4r(\Sigma)}{n}.$$

Denote $S = \sup_{\|v\|_2 \leq 1} \left(\sum_{j=1}^n \rho(Z_j(v)) - \mathbb{E} \rho(Z_j(v)) \right)$ and $\sigma^2 = \sup_{\|v\|_2 \leq 1} \text{Var}(\rho(Z_1(v)))$. We deduce that

$$\mathbb{P} \left(\sup_{\|v\|_2 \leq 1} |I_v| \geq c'n \right) \leq \mathbb{P} \left(\sup_{\|v\|_2 \leq 1} \sum_{j=1}^n \rho(Z_j) > c'n \right) \leq \mathbb{P}(S > c'n - 4r(\Sigma)).$$

We will apply Theorem 4.3 to estimate the right hand side in the display above. Specifically, $\mathbb{P}(S > c'n - 4r(\Sigma)) \leq e^{-t}$ whenever

$$c'n - 4r(\Sigma) - 2\mathbb{E}S - \sigma\sqrt{2tn} - 4t/3 \geq 0.$$

To prove that this relation holds for a suitable choices of parameters c and t (where $r(\Sigma)/n \leq c$), first observe that

$$\sigma^2 \leq \mathbb{E}(\rho(Z)^2) \leq \mathbb{E}\rho(Z) \leq \frac{4r(\Sigma)}{n},$$

where the last inequality follows from the bound (3.2). Next we will estimate $\mathbb{E}S$. Let $\{\varepsilon_j\}$, $j = 1, \dots, n$ be a sequence of independent i.i.d. random signs. The standard argument based on the symmetrization and contraction [Ledoux and Talagrand, 2013, Theorem 4.4] inequalities, together with the fact that $\rho(x)$ is Lipschitz continuous with Lipschitz constant equal to 2, yields

that

$$\begin{aligned}
\mathbb{E}S &= \mathbb{E} \sup_{\|v\|_2 \leq 1} \left(\sum_{j=1}^n \rho(Z_j) - \mathbb{E}\rho(Z_j) \right) \\
&\leq 2\mathbb{E} \sup_{\|v\|_2 \leq 1} \left| \sum_{j=1}^n \varepsilon_j \rho(Z_j) \right| \leq 4\mathbb{E} \sup_{\|v\|_2 \leq 1} \left| \sum_{j=1}^n \varepsilon_j \frac{|\langle X_j, v \rangle|}{\left(\kappa^2 \|\Sigma\| \sqrt{\frac{n}{r(\Sigma)}} \right)^{1/2}} \right| \\
&\leq \frac{8}{\left(\kappa^2 \|\Sigma\| \sqrt{\frac{n}{r(\Sigma)}} \right)^{1/2}} \left(\mathbb{E} \sup_{\|v\|_2 \leq 1} \left(\sum_{j=1}^n \langle X_j, v \rangle \right)^2 \right)^{1/2}.
\end{aligned}$$

Since $\mathbb{E} \sup_{\|v\|_2 \leq 1} \left(\sum_{j=1}^n \langle X_j, v \rangle \right)^2 = \mathbb{E} \left\| \sum_{j=1}^n X_j \right\|^2 \leq \kappa^2 n \|\Sigma\| r(\Sigma)$ by assumption (3.1), we conclude that

$$\mathbb{E} \sup_{\|v\|_2 \leq 1} \left(\sum_{j=1}^n \rho(Z_j) - \mathbb{E}\rho(Z_j) \right) \leq 8n^{1/4} (r(\Sigma))^{3/4}.$$

Elementary algebraic computation show that it suffices to choose parameters c and t so that $(4/3 + \sqrt{2})t \leq c'n/2$ and $16c^{3/4} + (4 + \sqrt{2})c \leq c'/2$, where c' is a constant defined in Theorem 3.2. If these relations hold, we conclude that $m \leq c'n$ with probability at least $1 - e^{-t}$, and that the inequality (3.2) holds with probability at least $1 - 1 - c'(p)/n$. Combining this result with the estimates (3.2), (3.2), and Theorem 3.2, we conclude that with probability at least $1 - c'(p)/n$, the “peaky” admits an upper bound of the form

$$\begin{aligned}
\sup_{\|v\|_2 \leq 1} \frac{1}{n} \sum_{j=1}^n \langle X_j, v \rangle^2 \mathbf{1} \left\{ \langle X_i, v \rangle^2 > \kappa^2 \|\Sigma\| \sqrt{\frac{n}{r(\Sigma)}} \right\} \\
\leq \frac{f(m, [n])}{n} \leq C_2(p, \kappa) \left(\frac{\max_{j \leq n} \|X_j\|_2^2}{n} + \|\Sigma\| \sqrt{\frac{r(\Sigma)}{n}} \right).
\end{aligned}$$

Combining the estimates (3.2) and (3.2) with the decomposition (3.2), we conclude that with probability at least $1 - e^{-r(\Sigma)} - c'(p)/n$,

$$\left\| \frac{1}{n} \sum_{j=1}^n X_j X_j^\top - \Sigma \right\| \leq C(p, \kappa) \left(\frac{\max_{j \leq n} \|X_j\|_2^2}{n} + \|\Sigma\| \sqrt{\frac{r(\Sigma)}{n}} \right).$$

To obtain the desired bound in expectation, let us define the event

$$A := \left\{ \left\| \frac{1}{n} \sum_{j=1}^n X_j X_j^\top - \Sigma \right\| \leq C(p, \kappa) \left(\frac{\max_{j \leq n} \|X_j\|_2^2}{n} + \|\Sigma\| \sqrt{\frac{r(\Sigma)}{n}} \right) \right\}.$$

Then $P(A^c) \leq e^{-r(\Sigma)} + c'(p)/n$, and

$$\begin{aligned} \mathbb{E}^{1/2} \left\| \frac{1}{n} \sum_{j=1}^n X_j X_j^\top - \Sigma \right\|^2 &\leq C_1(p, \kappa) \left(\frac{\mathbb{E}^{1/2} \max_{j \leq n} \|X_j\|_2^4}{n} + \|\Sigma\| \sqrt{\frac{r(\Sigma)}{n}} \right) \\ &\quad + \mathbb{E}^{1/2} \left[\left\| \frac{1}{n} \sum_{j=1}^n X_j X_j^\top - \Sigma \right\|^2 \mathbf{1}(A^c) \right]. \end{aligned}$$

Hölder's inequality implies that

$$\mathbb{E}^{1/2} \left[\left\| \frac{1}{n} \sum_{j=1}^n X_j X_j^\top - \Sigma \right\|^2 \mathbf{1}(A^c) \right] \leq \mathbb{E}^{2/p} \left\| \frac{1}{n} \sum_{j=1}^n X_j X_j^\top - \Sigma \right\|^{p/2} \left(\frac{c'(p)}{n} \vee e^{-r(\Sigma)} \right)^{\frac{p-4}{2p}}.$$

Finally, we invoke Rosenthal's inequality (2.2) to deduce that

$$\begin{aligned} \mathbb{E}^{2/p} \left\| \frac{1}{n} \sum_{j=1}^n X_j X_j^\top - \Sigma \right\|^{p/2} &\leq C(p) \left[\sqrt{\frac{r(\Sigma)}{n}} \|\Sigma\| \sqrt{\log(er(\Sigma))} \right. \\ &\quad \left. + \frac{\log(er(\Sigma))}{n} \mathbb{E} \max_{j \leq n} \|X_j\|_2^2 + \frac{1}{n} \mathbb{E}^{2/p} \max_{j \leq n} \|X_j\|_2^p \right]. \end{aligned}$$

Since $r(\Sigma) < cn$ by assumption, $\log(er(\Sigma)) < c(p) (n \vee e^{r(\Sigma)})^{(p-4)/2p}$, implying the final form of the bound. $\square$

4. Auxiliary results

In this section, we collect the background material and technical results that our arguments rely on.

Theorem 4.1. [*Minsker, 2017, Theorem 3.1*] and [*Tropp, 2015, Theorem 7.7.1*] Let $W_1, \dots, W_n \in \mathbb{C}^{d \times d}$ be a sequence of independent, centered, self-adjoint random matrices such that $\|W_k\| \leq U$, $k = 1, \dots, n$ almost surely. Assume that $V_n^2 \succeq \sum_k \mathbb{E} W_k^2$ and let $\sigma^2 = \|V_n^2\|$. Then for any $t \geq \sigma + U/3$,

$$\mathbb{P} \left(\left\| \sum_{k=1}^n W_k \right\| > t \right) \leq 4r(V_n^2) \exp \left[-\frac{t^2/2}{\sigma^2 + tU/3} \right].$$

Corollary 4.1. Let $p \geq 1$. Under the assumptions of Theorem 4.1,

$$\mathbb{E}^{1/p} \left\| \sum_{k=1}^n W_k \right\|^p \leq K (\sigma \sqrt{q} + Uq)$$

where $q = \log(er(V_n^2)) \vee p$ and $K > 0$ is a numerical constant.

Proof. Fix $t_0 > 0$ and note that

$$\begin{aligned}\mathbb{E}(\|\mathbf{Y}\|^p) &= \int_0^\infty pt^{p-1}\mathbb{P}(\|\mathbf{Y}\| > t)dt \\ &= \int_0^{t_0} pt^{p-1}\mathbb{P}(\|\mathbf{Y}\| > t)dt + \int_{t_0}^\infty pt^{p-1} \cdot 4r \exp\left[-\frac{t^2/2}{\sigma^2 + tU/3}\right] \\ &\leq t_0^p + \int_{t_0}^\infty 4r(V_n^2)pt^{p-1}e^{-t^2/4\sigma^2}dt + \int_{t_0}^\infty 4r(V_n^2)pt^{p-1}e^{-3t/4U}dt.\end{aligned}$$

Next, take $t_0 = 2\sigma\sqrt{\log(er(V_n^2))} + \frac{4}{3}U\log(er(V_n^2))$. The first integral on the right-hand side of (4) admits the following bound due to Lemma 4.1:

$$\begin{aligned}\int_{t_0}^\infty 4r(V_n^2)pt^{p-1}e^{-t^2/2\sigma^2}dt &\leq 4r(V_n^2)p(2\sigma)^{p/2} \int_{\sqrt{\log(er(V_n^2))}}^\infty u^{p-1}e^{-u^2}du \\ &\leq 4rp2^{p/2}\sigma^p2^{(p-1)/2}\Gamma((p+1)/2) \int_{\mu_1/\sqrt{2\sigma^2}}^\infty e^{-u^2/2}e^{-u^2}du \\ &\leq 4rp2^p\sigma^p((p+1)/4)^{(p-1)/2}\frac{\sigma}{\mu_1}e^{-\mu_1^2/4\sigma^2} \\ &\leq 4rp2^p\sigma^p((p+1)/4)^{(p-1)/2}e^{-(2\sigma(\sqrt{\log(1+r)}))^2/4\sigma^2} \\ &\leq 16 \cdot 2^p\sigma^p((p+1)/4)^{(p+1)/2}\end{aligned}$$

where $\Gamma(x) \leq (x/2)^{x-1}$ for $x \geq 2$. Similarly, for the second integral on the last line of (4), introducing $u = \frac{3t}{2U}$ and applying Lemma 4.1, we deduce that

$$\begin{aligned}\int_{\mu}^\infty 4rpt^{p-1}e^{-3t/2U}dt &\leq \int_{\mu_2}^\infty 4rpt^{p-1}e^{-3t/2U}dt \leq 4rp\left(\frac{3\mu_2}{2U}\right)^p \int_{3\mu_2/2U}^\infty u^{p-1}e^{-u}du \\ &\leq 4rp\left(\frac{3\mu_2}{2U}\right)^p 2^{p-1}\Gamma(p) \int_{3\mu_2/2U}^\infty e^{-u/2}e^{-u}du \leq 4rp\left(\frac{3\mu_2}{2U}\right)^p 2^p(p/2)^{p-1}e^{-3(\frac{4}{3}U\log(1+r))/4U} \\ &\leq 8(2\log(1+r))^p(2p)^p.\end{aligned}$$

□

Proposition 4.1. [Hoffmann-Jørgensen inequality: [Ledoux and Talagrand, 2013](#), Proposition 6.7] Let $X_1, \dots, X_n$ be independent, symmetrically distributed random variables with values in a separable Banach space with norm $\|\cdot\|_B$. Set $S_k = \sum_{i=1}^k X_i$, $k \leq N$. Then for any $s, t > 0$,

$$\mathbb{P}(\|S_N\|_B > 2t + s) \leq 4(\mathbb{P}(\|S_N\|_B > t))^2 + \mathbb{P}\left(\max_{i \leq N} \|X_i\|_B > s\right).$$

Proposition 4.2. [[Ledoux and Talagrand, 2013](#), Proposition 6.8] Let $0 < q < \infty$ and let $X_1, \dots, X_n$ be independent, symmetrically distributed random variables with values in a separable Banach space with norm $\|\cdot\|_B$. Set $S_k = \sum_{i=1}^k X_i$, $k \leq N$. Then for $t_0 = \inf\{t > 0 : \mathbb{P}(\|S_N\|_B > t) \leq (2 \cdot 3^p)^{-1}\}$,

$$\mathbb{E}\|S_N\|_B^p \leq 2 \cdot 3^p \mathbb{E} \max_{i \leq N} \|X_i\|_B^p + 2(3t_0)^p.$$

Theorem 4.2. [[Ledoux and Talagrand, 2013](#), Theorems 6.20 and 6.21] Let $X_1, \dots, X_n$ be independent random variables with values in a separable Banach space with norm $\|\cdot\|_B$. There exists

a numerical constant K such that for all $p > 1$,

$$\mathbb{E}^{1/p} \left\| \sum_{k=1}^n X_k \right\|_B^p \leq K \frac{p}{\log(ep)} \left(\mathbb{E} \left\| \sum_{k=1}^n X_k \right\|_B + \mathbb{E}^{1/p} \max_k \|X_k\|_B^p \right) \text{ and}$$

$$\left\| \left\| \sum_{k=1}^n X_k \right\|_B \right\|_{\psi_1} \leq K \left(\mathbb{E} \left\| \sum_{k=1}^n X_k \right\|_B + \left\| \max_k \|X_k\|_B \right\|_{\psi_1} \right).$$

Theorem 4.3. [Bousquet, 2003, Theorem 7.3] Let $\mathcal{F}$ be a countable set of measurable real-valued functions and let $X_1, \dots, X_n$ be i.i.d. Assume that $\mathbb{E}f(X_1) = 0$ for all $f \in \mathcal{F}$ and that $\sup_{f \in \mathcal{F}} |f(X_1)| \leq U$ with probability 1. Denote $Z = \sup_{f \in \mathcal{F}} \sum_{k=1}^n f(X_k)$. Assume that $\sigma_*^2 \geq n \sup_{f \in \mathcal{F}} \mathbb{E}f^2(X_1)$ and set $v = \sigma_*^2 + 2\mathbb{E}[Z]$. Then for all $t \geq 0$,

$$\mathbb{P} \left(Z \geq \mathbb{E}Z + \sqrt{2tv} + \frac{tU}{3} \right) \leq e^{-t}.$$

The inequality (4.3) immediately implies that with probability at least $1 - e^{-t}$, $Z \leq 2\mathbb{E}Z + \sigma_* \sqrt{2t} + 4tU/3$. Alternatively, the bound can be stated as

$$\mathbb{P}(Z \geq 2\mathbb{E}Z + t) \leq \exp \left(-\frac{t^2/2}{\sigma_*^2 + 4Ut/3} \right).$$

Finally, we state two elementary integral estimates required in the proofs.

Lemma 4.1. Fix $p \geq 2$. Then for any $t > 0$,

$$\int_t^\infty u^{p-1} e^{-u^2/2} du \leq e^{-t^2/2} \cdot 2^{p-2} (t^{p-2} + (\sqrt{p}/t)^p).$$

Proof. Making the change of variables $v = u - t$ and using the relation $e^{-u^2/2} \leq 1$ that holds for all u gives that

$$\begin{aligned} \int_t^\infty u^{p-1} e^{-u^2/2} du &\leq \int_0^\infty (t+v)^{p-1} e^{-(t+v)^2/2} dv \\ &\leq e^{-t^2/2} \int_0^\infty 2^{p-2} (t^{p-1} + v^{p-1}) e^{-tv} e^{-v^2/2} dv \\ &\leq e^{-t^2/2} \cdot 2^{p-2} (t^{p-2} + t^{-p} \cdot \Gamma(p)) \leq e^{-t^2/2} \cdot 2^{p-2} (t^{p-2} + (\sqrt{p}/t)^p) \end{aligned}$$

where $\Gamma(\cdot)$ denotes Euler's gamma function and where we used the inequality $\Gamma(p) \leq p^{p/2}$. $\square$

Lemma 4.2. For any positive real numbers $p > q$ and any $t > 0$,

$$t^p \leq 2^{p/q} \Gamma(p/q + 1) e^{t^q/2},$$

where $\Gamma(\cdot)$ is the Euler's gamma function.

Proof. Start with the identity $t^p = \frac{(t^q/2)^{p/q}}{\Gamma(p/q+1)} 2^{p/q} \Gamma(p/q+1)$ and observe that it suffices to show that $\frac{(t^q/2)^{p/q}}{\Gamma(p/q+1)} \leq e^{t^q/2}$. The latter is equivalent to the inequality $\frac{a^x}{\Gamma(x+1)} \leq e^a$ for all $x > 1$, $a > 0$. This inequality follows from the following estimate:

$$\frac{\Gamma(x+1)}{a^x} = \frac{\int_0^\infty t^x e^{-t} dt}{a^x} \geq \int_a^\infty \left(\frac{t}{a} \right)^x e^{-t} dt \geq \int_a^\infty e^{-t} dt = e^{-a}.$$

$\square$

References

- Abdalla, P. and Zhivotovskiy, N. (2024). Covariance estimation: Optimal dimension-free guarantees for adversarial corruption and heavy tails. to appear in the Journal of the European Mathematical Society.
- Adamczak, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to markov chains. Electronic Journal of Probability [electronic only], 13:1000–1034.
- Adamczak, R., Litvak, A., Pajor, A., and Tomczak-Jaegermann, N. (2010). Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. Journal of the American Mathematical Society, 23(2):535–561.
- Adamczak, R., Litvak, A. E., Pajor, A., and Tomczak-Jaegermann, N. (2011). Sharp bounds on the rate of convergence of the empirical covariance matrix. Comptes Rendus. Mathématique, 349(3-4):195–200.
- Ahlsweide, R. and Winter, A. (2002). Strong converse for identification via quantum channels. IEEE Transactions on Information Theory, 48(3):569–579.
- Bai, Z.-D. and Yin, Y.-Q. (2008). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. In Advances In Statistics, pages 108–127. World Scientific.
- Bakhshizadeh, M., Maleki, A., and De La Pena, V. H. (2023). Sharp concentration results for heavy-tailed distributions. Information and Inference: A Journal of the IMA, 12(3):1655–1685.
- Boucheron, S., Bousquet, O., Lugosi, G., and Massart, P. (2005). Moment inequalities for functions of independent random variables. Annals of Probability, pages 514–560.
- Bourgain, J. (1996). Random points in isotropic convex sets. Convex geometric analysis, 34:53–58.
- Bousquet, O. (2003). Concentration inequalities for sub-additive functions using the entropy method. In Stochastic inequalities and applications, pages 213–247. Springer.
- Chen, R. Y., Gittens, A., and Tropp, J. A. (2012). The masked sample covariance estimator: an analysis using matrix concentration inequalities. Information and Inference: A Journal of the IMA, 1(1):2–20.
- Dirksen, S. (2011). Noncommutative and vector-valued rosenthal inequalities. Ph.D. thesis.
- Einmahl, U. and Li, D. (2008). Characterization of LIL behavior in Banach space. Transactions of the American Mathematical Society, 360(12):6677–6693.
- Fuk, D. K. and Nagaev, S. V. (1971). Probability inequalities for sums of independent random variables. Theory of Probability & Its Applications, 16(4):643–660.
- Guédon, O., Litvak, A. E., Pajor, A., and Tomczak-Jaegermann, N. (2017). On the interval of fluctuation of the singular values of random matrices. Journal of the European Mathematical Society, 19(5).
- Han, Q. (2022). Exact spectral norm error of sample covariance. arXiv preprint arXiv:2207.13594.
- Johnson, W. B., Schechtman, G., and Zinn, J. (1985). Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. The Annals of Probability, pages 234–253.
- Junge, M. and Zeng, Q. (2013). Noncommutative Bennett and Rosenthal inequalities. Annals of probability: An official journal of the Institute of Mathematical Statistics, 41(6):4287–4316.
- Kannan, R., Lovász, L., and Simonovits, M. (1997). Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. Random Structures & Algorithms, 11(1):1–50.
- Klochkov, Y. and Zhivotovskiy, N. (2020). Uniform Hanson-Wright type concentration inequalities for unbounded entries via the entropy method. Electron. J. Probab, 25(22):1–30.
- Koltchinskii, V. (2011). Von Neumann entropy penalization and low-rank matrix estimation. The Annals of Statistics, 39(6):2936–2973.

- Koltchinskii, V. and Lounici, K. (2017). Concentration inequalities and moment bounds for sample covariance operators. Bernoulli, pages 110–133.
- Ledoux, M. and Talagrand, M. (2013). Probability in Banach Spaces: isoperimetry and processes. Springer Science & Business Media.
- Mendelson, S. and Paouris, G. (2014). On the singular values of random matrices. Journal of the European Mathematical Society, 16(4).
- Minsker, S. (2017). On some extensions of Bernstein’s inequality for self-adjoint operators. Statistics & Probability Letters, 127:111–119.
- Nagaev, S. V. (1979). Large deviations of sums of independent random variables. The Annals of Probability, pages 745–789.
- Oliveira, R. (2010). Sums of random Hermitian matrices and an inequality by Rudelson.
- Pinelis, I. and Utev, S. (1985). Estimates of the moments of sums of independent random variables. Theory of Probability & Its Applications, 29(3):574–577.
- Rio, E. (2017). About the constants in the Fuk-Nagaev inequalities. Electronic Communications in Probability, 22(28):12p.
- Rudelson, M. (1999). Random vectors in the isotropic position. Journal of Functional Analysis, 164(1):60–72.
- Rudelson, M. and Vershynin, R. (2007). Sampling from large matrices: An approach through geometric functional analysis. Journal of the ACM (JACM), 54(4):21–es.
- Srivastava, N. and Vershynin, R. (2013). Covariance estimation for distributions with $2 + \varepsilon$ moments. The Annals of Probability, 41(5):3081–3111.
- Talagrand, M. (1996). New concentration inequalities in product spaces. Inventiones mathematicae, 126(3):505–563.
- Tikhomirov, K. (2018). Sample covariance matrices of heavy-tailed distributions. International Mathematics Research Notices, 2018(20):6254–6289.
- Tropp, J. A. (2008). Norms of random submatrices and sparse approximation. Comptes Rendus. Mathématique, 346(23-24):1271–1274.
- Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. Foundations of computational mathematics, 12:389–434.
- Tropp, J. A. (2015). An introduction to matrix concentration inequalities. Foundations and Trends® in Machine Learning, 8(1-2):1–230.
- van der Vaart, A. W. and Wellner, J. A. (1996). Weak convergence and empirical processes. Springer Series in Statistics. Springer-Verlag, New York.
- Vershynin, R. (2011). Approximating the moments of marginals of high-dimensional distributions. The Annals of Probability, 39(4):1591–1606.
- Vershynin, R. (2012). How close is the sample covariance matrix to the actual covariance matrix? Journal of Theoretical Probability, 25(3):655–686.
- Wei, X. and Minsker, S. (2017). Estimation of the covariance structure of heavy-tailed distributions. Advances in neural information processing systems, 30.