On the Dimension-Free Concentration of Simple Tensors via Matrix Deviation

Pedro Abdalla Roman Vershynin*

June 12, 2025

Abstract

We provide a simpler proof of a sharp concentration inequality for subgaussian simple tensors obtained recently by Ghattas, Chen, and Alonso. Our approach uses a matrix deviation inequality for ℓ^p norms and a basic chaining argument.

1 Introduction

Let X be a mean-zero random vector in $\mathbb{R}^d$ with covariance $\Sigma = \mathbb{E}XX^\mathsf{T}$, and let $p \geq 2$ be an integer. Can we estimate the expected value of the simple random tensor $X^{\otimes p}$ from a sample $X_1, \ldots, X_N$ of i.i.d. copies of X? The simplest estimator is the empirical mean. The error of this estimator is

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i^{\otimes p} - \mathbb{E}X^{\otimes p} \right\| := \sup_{v \in S^{d-1}} \left| \frac{1}{N} \sum_{i=1}^{N} \langle X_i, v \rangle^p - \mathbb{E}\langle X, v \rangle^p \right|. \tag{1.1}$$

How large is it? We are interested in dimension-free bounds – those that do not depend on the full dimension d, but instead on the "effective dimension" of the distribution, captured by the effective rank of the covariance matrix:

$$r(\Sigma) := \frac{\operatorname{Tr}(\Sigma)}{\|\Sigma\|}.$$

This question is nontrivial already for $p = 2$, where it becomes the covariance estimation problem. The standard ε-net argument gives optimal dimension-dependent bounds (see [12, Chapter 4]). But getting dimension-free bounds even for $p = 2$ is harder. They were first obtained by Lounici and Koltchinskii [5] for subgaussian distributions using generic chaining. A simpler argument for Gaussian distribution was found by Van Handel [10], while Liaw, Mehrabian, Plan and the second author [6] show how to deduce

*Department of Mathematics, UC Irvine

the result for all subgaussian distributions from their matrix deviation inequality (see [12, Chapter 9] for an introduction). That last approach also lets one take the supremum in (1.1) over any bounded set $T \subset \mathbb{R}^d$ – the setting we consider in the present paper, too.

Much less has been known when $p > 2$. Optimal bounds that depend on dimension[1] were proved in [1, 4, 11]. Very recently, Ghattas, Chen and Alonso [2] proved optimal bounds for $p \geq 2$. Their proof relies on a local chaining argument due to Mendelson [7] combined with an intricate analysis of coordinate projections due to Bednorz [3].

The goal of our work is to provide a simpler proof of this fact by exploiting tools from non-asymptotic random matrix theory. Specifically, we can leverage the matrix deviation inequality (see Theorem 2.1) to bypass Bendorz's analysis of coordinate projections, and we can run the chaining argument for the simpler L^2 structure of the process and bypass the delicate chaining argument of Mendelson.

To state the result, let $g \sim N(0, I_d)$ be the standard multivariate Gaussian random vector. For a set $T \subset \mathbb{R}^d$, the radius and the Gaussian complexity are defined by

$$\mathrm{rad}(T) := \sup_{v \in T} \|v\|_2, \quad \gamma(T) := \mathbb{E} \sup_{v \in T} |\langle g, v \rangle|.$$

We say that a random vector X is subgaussian if there is an $K > 0$ such that[2]

$$\|\langle X, u \rangle\|_{\psi_2} \leq K \|\langle X, u \rangle\|_{L^2} \quad \text{for any } u \in S^{d-1}.$$

The following theorem and corollary were proved by Ghattas, Chen and Alonso [2]. We provide a simpler proof of these facts.

Theorem 1.1. *Let $Z_1, \ldots, Z_N$ be i.i.d. copies of a mean-zero subgaussian isotropic random vector Z, and let $T \subset \mathbb{R}^d$ be a bounded set. Then, for any integer $p \geq 2$, we have*

$$\mathbb{E} \sup_{v \in T} \left| \sum_{i=1}^N \langle Z_i, v \rangle^p - N \mathbb{E} \langle Z, v \rangle^p \right| \lesssim_{K,p} \gamma(T)^p + \sqrt{N} \gamma(T)\, \mathrm{rad}(T)^{p-1},$$

where the sign $\lesssim_{K,p}$ hides factors that depend only on K and p.

Corollary 1.2. *Let X be a mean-zero subgaussian random vector with covariance matrix Σ and $X_1, \ldots, X_N$ be i.i.d. copies of X. Then, for any integer $p \geq 2$, we have*

$$\mathbb{E} \left\| \frac{1}{N} \sum_{i=1}^N X_i^{\otimes p} - \mathbb{E} X^{\otimes p} \right\| \lesssim_{K,p} \|\Sigma\|^{p/2} \left(\frac{r(\Sigma)^{p/2}}{N} + \sqrt{\frac{r(\Sigma)}{N}} \right).$$

[1] Technically, the results are for L^p moments with has the absolute value $|\langle X, v \rangle|^p$, but they directly apply when p is even.

[2] For basics on subgaussian distributions, see [12].

To derive Corollary 1.2, write $X = \Sigma^{1/2} Z$ for some isotropic random vector Z and apply Theorem 1.1 for the ellipsoid $T = \Sigma^{1/2} S^{d-1}$, noting that

$$\mathrm{rad}(\Sigma^{1/2} S^{d-1}) = \|\Sigma\|^{1/2}, \quad \gamma(\Sigma^{1/2} S^{d-1}) \le \mathrm{Tr}(\Sigma)^{1/2} = \|\Sigma\|^{1/2} r(\Sigma)^{1/2}.$$

2 Matrix Deviation Approach

Our approach will also exploit the matrix deviation inequality, as was first done for the case $p = 2$ in [6], see [12, Chapter 9]. We will use the following version for $p \ge 2$ proved by Sheu and Wang [8]:

Theorem 2.1 (ℓ^p matrix deviation inequality). *Let $p \ge 2$. Let $A \in \mathbb{R}^{N \times d}$ be a random matrix whose rows are i.i.d. copies of a mean-zero isotropic subgaussian random vector $Z \in \mathbb{R}^d$. Consider the stochastic process*

$$Z_v := \Big| \|Av\|_p - N^{1/p} \|\langle Z, v\rangle\|_{L^p} \Big|, \quad v \in \mathbb{R}^d.$$

Then Z_v has Lipschitz subgaussian increments:

$$\|Z_v - Z_u\|_{\psi_2} \lesssim_{K,p} \|u - v\|_2 \quad \textit{for every } u, v \in \mathbb{R}^d.$$

Remark 2.2 (A bound on the process). It immediately follows from Theorem 2.1 combined with Talagrand's majorizing measure theorem (see [12, Exercise 8.6.5]) that for any bounded set $T \subset \mathbb{R}^n$ and any $u > 0$, we have

$$\sup_{v \in T} |Z_v| \lesssim_{K,p} \gamma(T) + u \,\mathrm{rad}(T)$$

with probability at least $1 - 2e^{-u^2}$.

Remark 2.3 (Subgaussianity). Theorem 2.1 is stated in [8] under a stronger assumption that the entries A are i.i.d. subgaussian, but the proof actually only relies on the rows being subgaussian.

As we mentioned, for $p = 2$ one can derive Theorem 1.1 directly from matrix deviation inequality. However, for $p > 2$, in addition to matrix deviation, we will need an extra chaining step.

To run the chaining, we need a couple of standard bounds on the order statistic. Given real numbers $X_1, \dots, X_n$, we denote by $X_1^*, \dots, X_n^*$ a nonincreasing rearrangement of the numbers $|X_1|, \dots, |X_n|$, so that $|X_1^*| \ge |X_2^*| \ge \cdots \ge |X_n^*|$.

Lemma 2.4 (Order statistics). *Let $X_1, \dots, X_n$ be independent random variables satisfying $\|X_i\|_{\psi_2} \le 1$ for all i. Let $t > 0$, $q \ge 2$ and $k := t/\ln_+(en/t)$. Then,[3] with probability at least $1 - 2e^{-t}$, we have*

$$\sum_{i \le 3k} (X_i^*)^2 \lesssim t \quad \textit{and} \quad \sum_{i > k} (X_i^*)^q \lesssim_q n. \tag{2.1}$$

[3] By convention, we set $1/0 = \infty$ in defining k, and the sum over an empty set is zero.

We postpone the proof to Appendix A and prove the main result.

Proof of Theorem 1.1. Step 1: Symmetrization. Denote by $\mathcal{F}$ be the class of functions $f : \mathbb{R}^d \to \mathbb{R}$ of the form $f(Z) = \langle Z, v\rangle$, where $v \in T$. Then, by the classical Gine-Zinn symmetrization (see for example [12, Chapter 6]), we have

$$\mathbb{E}\sup_{v\in T}\Big|\sum_{i=1}^{N}\langle Z_i, v\rangle^p - N\mathbb{E}\langle Z, v\rangle^p\Big| = \mathbb{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^{N}\big(f^p(Z_i) - \mathbb{E}f^p(Z_i)\big)\Big| \lesssim \mathbb{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^{N}\varepsilon_i f^p(Z_i)\Big|, \tag{2.2}$$

where $\varepsilon_1, \dots, \varepsilon_N$ are i.i.d. Rademacher random variables, independent of (Z_i).

Step 2: Generic chaining. Given an admissible sequence[4] $(T_s)_{s\ge0}$ in T, we consider the collection of maps $\pi_s : T \to T_s$ that map $v \in T$ to a nearest point $\pi_s v \in T_s$. Denote by $\mathcal{F}_s$ the class of functions $(\pi_s f)(Z) := \langle Z, \pi_s v\rangle$, where $v \in T$.

By Talagrand's majorizing measure theorem (see [12, Chapter 8] or [9, Chapter 2]), there exists an admissible sequence such that

$$\sup_{f\in\mathcal{F}}\sum_{s\ge0}2^{s/2}\|\Delta_s f\|_{L^2(Z)} = \sup_{v\in T}\sum_{s\ge0}2^{s/2}\|\Delta_s v\|_2 \lesssim \gamma(T), \tag{2.3}$$

where $\Delta_s f := \pi_{s+1}f - \pi_s f$ and $\Delta_s v := \pi_{s+1}v - \pi_s v$. Fix this admissible sequence. We can decompose any $f \in \mathcal{F}$ along this chain:

$$f^p(Z) = \sum_{s\ge0}\big((\pi_{s+1}f)^p - (\pi_s f)^p\big)(Z) + (\pi_0 f)^p(Z),$$

where the series converges in $L^2(Z)$. Decomposing $f^p(Z_i)$ this way gives

$$\sum_{i=1}^{N}\varepsilon_i f^p(Z_i) = \sum_{s\ge0}\sum_{i=1}^{N}\varepsilon_i\big((\pi_{s+1}f)^p - (\pi_s f)^p\big)(Z_i) + \sum_{i=1}^{N}\varepsilon_i(\pi_0 f)^p(Z_i). \tag{2.4}$$

Now, $\pi_0 f \in \mathcal{F}_0$ and $\mathcal{F}_0$ is a singleton (as is T_0), so $\pi_0 f$ does not depend on the choice of f. Thus, if we take supremum over $f \in \mathcal{F}$ and then take expectation, we can bound (2.2) as follows:

$$\mathbb{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^{N}\varepsilon_i f^p(Z_i)\Big| \lesssim \mathbb{E}\sup_{f\in\mathcal{F}}\Big|\sum_{s\ge0}\sum_{i=1}^{N}\varepsilon_i\big((\pi_{s+1}f)^p - (\pi_s f)^p\big)(Z_i)\Big| + \mathbb{E}\Big|\sum_{i=1}^{N}\varepsilon_i(\pi_0 f)^p(Z_i)\Big|. \tag{2.5}$$

[4] An admissible sequence in T is a collection of sets $(T_s)_{s\ge0} \subset T$ satisfying $|T_s| \le 2^{2^s}$ and $|T_0| = 1$.

Step 3: Hoeffding's inequality. The main work goes into bounding the first term on the right hand side of (2.5). To do so, let's first fix $s \geq 0$ and $f \in \mathcal{F}$. Split the sum $\sum_{i=1}^N$ into $\sum_{i\in I_s}$ and $\sum_{i\notin I_s}$, where the set $I_s \subset [N]$ will be chosen later independently of the signs ε_i. To bound $\sum_{i\in I_s}$, use the triangle inequality. To bound $\sum_{i\notin I_s}$, use Hoeffding's inequality conditionally on (Z_i).

We get that, with probability at least $1 - 2e^{-2u^2 2^s}$,

$$\begin{aligned}&\Big|\sum_{i=1}^N \varepsilon_i\big((\pi_{s+1}f)^p - (\pi_s f)^p\big)(Z_i)\Big| \\ &\lesssim \underbrace{\sum_{i\in I_s}|(\pi_{s+1}f)^p - (\pi_s f)^p|(Z_i)}_{:=A_s(f)} + \underbrace{u2^{s/2}\Big(\sum_{i\notin I_s}\big((\pi_{s+1}f)^p - (\pi_s f)^p\big)^2(Z_i)\Big)^{1/2}}_{:=B_s(f)}.\end{aligned} \tag{2.6}$$

By the numeric inequality $|a^p - b^p| \leq |a-b|(|a|^{p-1} + |b|^{p-1})$, we have

$$\begin{aligned}A_s(f) &\lesssim_p \sum_{i\in I_s}|(\Delta_s f)(Z_i)|\Big(|(\pi_{s+1}f)(Z_i)|^{p-1} + |(\pi_s f)(Z_i)|^{p-1}\Big),\\ B_s(f) &\lesssim_p u2^{s/2}\Big(\sum_{i\notin I_s}(\Delta_s f)(Z_i)^2\Big(|(\pi_{s+1}f)(Z_i)|^{2p-2} + |(\pi_s f)(Z_i)|^{2p-2}\Big)\Big)^{1/2}.\end{aligned}$$

Step 4: Order statistics. Toward applying Lemma 2.4, let $t := 2u^2 2^s$ and $k_s := t/\ln_+(en/t)$, and choose $I_s \subset [N]$ as a set with cardinality $\lfloor 3k_s \rfloor$ which contains $\lfloor k_s \rfloor$ largest coordinates of $|(\Delta_s f)(Z_i)|$, $\lfloor k_s \rfloor$ largest coordinates of $|(\pi_{s+1}f)(Z_i)|$ and $\lfloor k_s \rfloor$ largest coordinates of $|(\pi_s f)(Z_i)|$ (and any other indices in $[N]$ to make the total cardinality $\lfloor 3k_s \rfloor$). Thus, applying Cauchy-Schwarz inequality and recalling that $\pi_{s+1}f, \pi_s f \in \mathcal{F}$, we get

$$A_s(f) \leq 2\underbrace{\Big(\sum_{i\leq 3k_s}\big((\Delta_s f)(Z_i)^*\big)^2\Big)^{1/2}}_{:=A'_s(f)}\underbrace{\sup_{\phi\in\mathcal{F}}\Big(\sum_{i=1}^N|\phi(Z_i)|^{2(p-1)}\Big)^{1/2}}_{:=A''}, \tag{2.7}$$

where the star indicates a nonincreasing rearrangement of N numbers (introduced just above Lemma 2.4).

To bound $A'_s(f)$, we apply Lemma 2.4 for $k = k_s$ and for the random variables $X_i = (\Delta_s f)(Z_i) = \langle \Delta_s v, Z_i\rangle$, which satisfy $\|X_i\|_{\psi_2} \lesssim_K \|\Delta_s v\|_2$ by assumption. After rescaling, we get with probability at least $1 - 2e^{-2u^2 2^s}$ that

$$A'_s(f) \lesssim_K u2^{s/2}\|\Delta_s v\|_2. \tag{2.8}$$

Step 5: Matrix deviation inequality. To bound A'', recall from the definition of $\mathcal{F}$ that

$$A'' = \sup_{v\in T}\Big(\sum_{i=1}^{N} |\langle Z_i, v\rangle|^{2(p-1)}\Big)^{1/2} = \sup_{v\in T}\|Av\|_{2(p-1)}^{p-1}. \tag{2.9}$$

By Theorem 2.1 (see Remark 2.2), with probability at least $1-2e^{-u^2}$, the following event holds:

$$\sup_{v\in T}\Big|\|Av\|_{2(p-1)} - N^{1/2(p-1)}\|\langle Z, v\rangle\|_{L^{2(p-1)}}\Big| \lesssim_{K,p} \gamma(T) + u\,\mathrm{rad}(T). \tag{2.10}$$

Suppose this event occurs. By the triangle inequality together with the bound $\|\langle Z, v\rangle\|_{L^p} \lesssim_{K,p} \|\langle Z, v\rangle\|_{L^2} = \|v\|_2 \le \mathrm{rad}(T)$, we have

$$\sup_{v\in T}\|Av\|_{2(p-1)} \lesssim_{K,p} \gamma(T) + u\,\mathrm{rad}(T) + N^{1/2(p-1)}\,\mathrm{rad}(T).$$

Then, using the numeric inequality $(a+b+c)^{p-1} \lesssim_p a^{p-1}+b^{p-1}+c^{p-1}$ and recalling (2.9), we conclude that the event

$$\mathcal{E} := \Big\{A'' \lesssim_{K,p} \gamma(T)^{p-1} + (\sqrt{N} + u^{p-1})\,\mathrm{rad}(T)^{p-1}\Big\}$$

is likely:

$$\mathbb{P}(\mathcal{E}) \ge 1 - 2e^{-u^2}. \tag{2.11}$$

Substituting this and (2.8) into (2.7), we conclude that, for any $s \ge 0$ and $f \in \mathcal{F}$, the following bound holds with probability at least $1-2e^{-2u^2 2^s}$:

$$\mathbf{1}_{\mathcal{E}} A_s(f) \lesssim_{K,p} u2^{s/2}\|\Delta_s v\|_2 \underbrace{\Big(\gamma(T)^{p-1} + (\sqrt{N} + u^{p-1})\,\mathrm{rad}(T)^{p-1}\Big)}_{:=\lambda(T)}. \tag{2.12}$$

For the term $B_s(f)$, it is enough to bound

$$\tilde{B}_s(f) := u2^{s/2}\Big(\sum_{i\notin I_s}(\Delta_s f)(Z_i)^2|(\pi_s f)(Z_i)|^{2p-2}\Big)^{1/2},$$

as the bound with $\pi_{s+1}f$ would follows similarly. Applying first the Cauchy-Schwarz inequality and then Lemma 2.4, we obtain with probability at least $1-2e^{-2u^2 2^s}$ that

$$\begin{aligned}
\tilde{B}_s(f) &\le u2^{s/2}\Big(\sum_{i>k_s}((\Delta_s f)(Z_i)^*)^4\Big)^{1/4}\Big(\sum_{i>k_s}|(\pi_s f)(Z_i)^*|^{4p-4}\Big)^{1/4}\\
&u2^{s/2}\Big(\sum_{i>k_s}(\Delta_s^* f)^4(Z_i)\Big)^{1/4}\Big(\sum_{i>k_s}|\pi_s f(Z_i)^*|^{4p-4}\Big)^{1/4}\\
&\lesssim_{K,p} u2^{s/2}\Big(N\|\Delta_s v\|_2^4\Big)^{1/4}\Big(N\|\pi_s v\|_2^{4p-4}\Big)^{1/4}\\
&\lesssim_{K,p} u2^{s/2}\sqrt{N}\|\Delta_s v\|_2\,\mathrm{rad}(T)^{p-1}.
\end{aligned}$$

The same bound holds replacing π_{s+1} by π_s. Therefore, with probability at least $1-4e^{-2u^2 2^s}$, we get

$$B_s(f) \lesssim_{K,p} u2^{s/2}\sqrt{N}\|\Delta_s v\|_2 \operatorname{rad}(T)^{p-1}. \tag{2.13}$$

Combining this with (2.12), we conclude that the following bound holds with probability at least $1-6e^{-2u^2 2^s}$:

$$\mathbf{1}_{\mathcal{E}}\big(A_s(f)+B_s(f)\big) \lesssim_{K,p} u2^{s/2}\|\Delta_s v\|_2 \lambda(T).$$

Substituting into (2.6), we obtain that, with probability at least $1-7e^{-2u^2 2^s}$,

$$\mathbf{1}_{\mathcal{E}}\Big|\sum_{i=1}^{N}\varepsilon_i\big((\pi_{s+1}f)^p-(\pi_s f)^p\big)(Z_i)\Big| \lesssim_{K,p} u2^{s/2}\|\Delta_s v\|_2 \lambda(T). \tag{2.14}$$

Step 6: Union bound. Now take a union bound over all $f\in\mathcal{F}$, or equivalently over all $v\in T$. By definition of an admisible sequence, the number of different pairs $(\pi_s(f),\pi_{s+1}(f))$ is at most $2^{2^s}\cdot 2^{2^{s+1}} \le 2^{3\cdot 2^s}$. So, taking the union bound over all these, we conclude that, with probability at least $1-2^{3\cdot 2^s}\cdot 7e^{-u^2 2^s} \ge 1-7e^{-u^2 2^s}$, the bound (2.14) holds for all $f\in\mathcal{F}$ simultaneously.

Next, take a union bound over all $s\ge 0$. We obtain that, with probability at least

$$1-\sum_{s\ge 0} 7e^{-u^2 2^s} \ge 1-10e^{-u^2},$$

the bound (2.14) holds for all $f\in\mathcal{F}$ and $s\ge 0$ simultaneously. Furthermore, recalling from (2.11) that the event $\mathcal{E}$ is likely, we conclude that, with probability at least $1-12e^{-u^2}$, the bound

$$\Big|\sum_{i=1}^{N}\varepsilon_i\big((\pi_{s+1}f)^p-(\pi_s f)^p\big)(Z_i)\Big| \lesssim_{K,p} u2^{s/2}\|\Delta_s v\|_2 \lambda(T)$$

holds for all $f\in\mathcal{F}$ and $s\ge 0$ simultaneously.

Sum up these inequalities over all $s\ge 0$ and take supremum over $f\in\mathcal{F}$ (or equivalently, over $v\in T$) on both sides. We get with probability at least $1-12e^{-u^2}$:

$$\begin{aligned}
&\sup_{f\in\mathcal{F}}\Big|\sum_{s\ge 0}\sum_{i=1}^{N}\varepsilon_i\big((\pi_{s+1}f)^p-(\pi_s f)^p\big)(Z_i)\Big| \\
&\quad\lesssim_{K,p} u\lambda(T)\sup_{v\in T}\sum_{s\ge 0}2^{s/2}\|\Delta_s v\|_2 \\
&\quad\lesssim u\lambda(T)\gamma(T) \quad \text{(by the choice of admissible sequence in (2.3))} \\
&\quad\lesssim u^p\big(\gamma(T)^p+\sqrt{N}\gamma(T)\operatorname{rad}(T)^{p-1}\big) \quad \text{(by definition of } \lambda(T) \text{ in (2.12))}.
\end{aligned}$$

Integrating the tail, we conclude that

$$\mathbb{E}\sup_{f\in\mathcal{F}}\left|\sum_{s\geq 0}\sum_{i=1}^{N}\varepsilon_i\big((\pi_{s+1}f)^p-(\pi_s f)^p\big)(Z_i)\right| \lesssim_{K,p} \gamma(T)^p+\sqrt{N}\gamma(T)\,\mathrm{rad}(T)^{p-1}. \tag{2.15}$$

Step 7: The first term of chaining. We successfully bounded the first term in (2.5). The second term is much easier, since there is no supremum:

$$\begin{aligned}\mathbb{E}\left|\sum_{i=1}^{N}\varepsilon_i(\pi_0 f)^p(Z_i)\right| &= \mathbb{E}\left|\sum_{i=1}^{N}\varepsilon_i\langle Z_i,\pi_0 v\rangle^p\right| \leq \mathbb{E}\left(\sum_{i=1}^{N}\langle Z_i,\pi_0 v\rangle^{2p}\right)^{1/2}\\ &\lesssim_{K,p} \sqrt{N}\,\mathrm{rad}(T)^p \lesssim \sqrt{N}\gamma(T)\,\mathrm{rad}(T)^{p-1}. \end{aligned}\tag{2.16}$$

The first inequality follows by first taking expectation with respect to the random signs (ε_i) and then with respect to (Z_i). To get the second inequality, follow the steps we made to bound A'' in (2.11), using matrix deviation inequality (2.10) with $2p$ instead of $2(p-1)$. The third inequality follows since we always have $\mathrm{rad}(T)\lesssim\gamma(T)$.

Substitute (2.15) and (2.16) into (2.5) and then into (2.2), and Theorem 1.1 is proved. □

3 Remarks

The proof does not rely on p being an integer. The same bound (up to an absolute constant) holds for the process

$$\sup_{v\in T}\left|\frac{1}{N}\sum_{i=1}^{N}|\langle Z_i,v\rangle|^p-\mathbb{E}|\langle Z,v\rangle|^p\right|,$$

where $p\geq 2$ may not be an integer. We assumed that p is an integer in the case of simple tensors just because the function x^p might not be well-defined for non-integer p.

Similarly as in the proof of Ghattas, Chen and Alonso, our proof extends to any star-shaped class of functions $\mathcal{F}$ such that for any $f\in\mathcal{F}$, it holds that $\|f(Z)-g(Z)\|_{\psi_2}\lesssim\|f(Z)-g(Z)\|_{L^2}$.

A Proof of Lemma 2.4

Case 1: $t\leq n$. In this regime, we have

$$k\leq n \quad\text{and}\quad t\asymp k\ln(en/k). \tag{A.1}$$

The first bound follows if we replace t with n the definition of k. The second bound follows from $k=t/\ln(en/t)$.

To prove the first bound in (2.1), note that if $\sum_{i\le k}(X_i^*)^2 > a$, then there exists a subset $I \subset [n]$ with $|I| \le 3k$ for which $\sum_{i\in I} X_i^2 > a$. There are at most $(en/k)^{3k}$ subsets of cardinality bounded by $3k$, so the union bound gives

$$\mathbb{P}\Big\{\sum_{i\le k}(X_i^*)^2 > C(k+t)\Big\} \le \Big(\frac{en}{k}\Big)^{3k} \max_{|I|=k} \mathbb{P}\Big\{\sum_{i\in I} X_i^2 > C(k+t)\Big\}. \tag{A.2}$$

Each X_i satisfies $\mathbb{E}X_i^2 \le C\|X_i\|_{\psi_2}^2 \le C$, so

$$\mathbb{P}\Big\{\sum_{i\in I} X_i^2 > C(k+t)\Big\} = \mathbb{P}\Big\{\sum_{i\in I}(X_i^2 - \mathbb{E}X_i^2) > Ct\Big\}.$$

The random variables $X_i^2 - \mathbb{E}X_i^2$ are independent, mean zero, and subexponential: $\|X_i^2 - \mathbb{E}X_i^2\|_{\psi_1} \lesssim \|X_i^2\|_{\psi_1} = \|X_i\|_{\psi_2}^2 \le 1$, so Bernstein's inequality [12, Theorem 2.9.1] gives the following bound on the probability above:

$$\le 2\exp\Big(-c\min((Ct)^2/k, Ct)\Big) \le 2\exp(-cCt),$$

where we used that $t \gtrsim k$ from (A.1) and chose C sufficiently large. Hence the probability in (A.2) is bounded by

$$2(en/k)^{3k}\exp(-cCt) \le 2\exp(-t),$$

where we again used (A.1) and chose C sufficiently large. Thus, with probability at least $1-2e^{-t}$, we have

$$\sum_{i\le k}(X_i^*)^2 \lesssim k+t \lesssim t,$$

where the last bound follows from (A.1). The first bound in (2.1) is proved.

To prove the second bound in (2.1), choose a sufficiently large absolute constant C, fix any $i \in [n]$ and argue like above to get

$$\begin{aligned}\mathbb{P}\Big\{X_i^* > C\sqrt{\ln\frac{en}{i}}\Big\} &\le \binom{n}{i}\max_{i\in[n]}\Big(\mathbb{P}\Big\{X_i > C\sqrt{\ln\frac{en}{i}}\Big\}\Big)^i \\ &\le \Big(\frac{en}{i}\Big)^i\Big(\exp\big(-2\ln\frac{en}{i}\big)\Big)^i = \Big(\frac{en}{i}\Big)^{-i}.\end{aligned}$$

Thus

$$\mathbb{P}\Big\{\exists i > k:\ X_i^* > C\sqrt{\ln\frac{en}{i}}\Big\} \le \sum_{i>k}\Big(\frac{en}{i}\Big)^{-i} \lesssim \Big(\frac{en}{k}\Big)^{-k} \le e^{-t}.$$

Therefore, with probability at least $1-e^{-t}$, we have

$$\sum_{i>k}(X_i^*)^q \lesssim \sum_{i=1}^{n}\Big(\ln\frac{en}{i}\Big)^{q/2} \lesssim n,$$

proving the second bound in (2.1).

Case 2: $t > n$. In this regime, we have $k > n$, which can be seen by replacing t with n in the definition of k. So, only the first bound in (2.1) needs to be shown. We have

$$\mathbb{P}\left\{\sum_{i \le 3k} (X_i^*)^2 > C(n+t)\right\} = \mathbb{P}\left\{\sum_{i=1}^{n} X_i^2 > C(n+t)\right\} \le 2\exp(-t),$$

where the last bound follows in a way similar to Case 1 (but with n instead of k). Thus, with probability at least $1 - 2e^{-t}$, we have

$$\sum_{i \le 3k} (X_i^*)^2 \lesssim n + t \lesssim t.$$

The first bound in (2.1) is proved, finishing the proof of Lemma 2.4.

Acknowledgments

PA and RV are supported by NSF and Simons Research Collaborations on the Mathematical and Scientific Foundations of Deep Learning. RV is also supported by NSF Grant DMS 1954233.

References

[1] Radosław Adamczak, Alexander Litvak, Alain Pajor, and Nicole Tomczak-Jaegermann. Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *Journal of the American Mathematical Society*, 23(2):535–561, 2010.

[2] Omar Al-Ghattas, Jiaheng Chen, and Daniel Sanz-Alonso. Sharp concentration of simple random tensors. *arXiv preprint arXiv:2502.16916*, 2025.

[3] Witold Bednorz. Concentration via chaining method and its applications. *arXiv preprint arXiv:1405.0676*, 2014.

[4] Olivier Guédon and Mark Rudelson. Lp-moments of random vectors via majorizing measures. *Advances in Mathematics*, 208(2):798–823, 2007.

[5] Vladimir Koltchinskii and Karim Lounici. Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*, 23(1):110 – 133, 2017.

[6] Christopher Liaw, Abbas Mehrabian, Yaniv Plan, and Roman Vershynin. A simple tool for bounding the deviation of random matrices on geometric sets. In *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2014–2016*, pages 277–299. Springer, 2017.

[7] Shahar Mendelson. Upper bounds on product and multiplier empirical processes. *Stochastic Processes and their Applications*, 126(12):3652–3680, 2016.

[8] Yuan-Chung Sheu and Te-Chun Wang. Matrix deviation inequality for ℓ_p-norm. *Random Matrices: Theory and Applications*, 12(04):2350007, 2023.

[9] Michel Talagrand. *Upper and lower bounds for stochastic processes: decomposition theorems*, volume 60. Springer Nature, 2022.

[10] Ramon van Handel. Structured random matrices. In *Convexity and concentration*, volume 161 of *IMA Vol. Math. Appl.*, pages 107–156. Springer, New York, 2017.

[11] Roman Vershynin. Approximating the moments of marginals of high-dimensional distributions. *The Annals of Probability*, 39(4):1591 – 1606, 2011.

[12] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.