Matrix Deviation Inequality for $\ell_p$-Norm

Yuan-Chung Sheu *† Te-Chun Wang ‡§

February 28, 2023

Abstract

Motivated by the general matrix deviation inequality for i.i.d ensemble Gaussian matrix [15, Theorem 11.1.5], we show that this property holds for $\ell_p$-norm with $1 \leq p < \infty$ and i.i.d ensemble sub-Gaussian random matrices, which is a random matrix with i.i.d mean-zero, unit variance, sub-Gaussian entries. As a consequence of our result, we establish the Johnson–Lindenstrauss lemma from $\ell_2^n$-space to $\ell_p^n$-space for all i.i.d ensemble sub-Gaussian random matrices.

Contents

1 Introduction 2
2 $\alpha$-Orlicz Random Variables 4
3 Proofs 5
   3.1 Proof of Theorem 1.2 ................................................................. 5
      3.1.1 Case 1: $x \in \mathbb{R}^n$ and $y = 0$ .................................................. 6
      3.1.2 Case 2: $||x|| = ||y|| = 1$ ............................................................. 9
      3.1.3 Case 3: General Vectors $x, y \in \mathbb{R}^n$ .............................................. 11
   3.2 Proof of Corollary 1.3 ................................................................. 13

*Department of Applied mathematics, National Yang Ming Chiao Tung University, Hsinchu, Taiwan.
†Email: sheu@math.nctu.edu.tw
‡Department of Mathematics and Statistics, University of Victoria, British Columbia, Canada.
§Email: lieb.am07g@gmail.com
1 Introduction

Given an \( m \times n \) random matrix \( A \), the uniform deviation inequality plays an important role in theory of random matrices. Also it has many interesting and important consequences. We first quote a classical result \([15, \text{Theorem 11.1.5}]\) for i.i.d ensemble Gaussian random matrices with respect to positive-homogeneous and subadditive function, which is very useful in asymptotic geometric analysis, and the proof goes back to \([13, \text{Lemma 3}]\), which has a different formulation.

**Theorem 1.1.** Let \( A \in \mathbb{R}^{m \times n} \) be a random matrix with i.i.d \( \mathcal{N}(0,1) \) entries, \( T \subseteq \mathbb{R}^n \), and \( f : \mathbb{R}^m \rightarrow \mathbb{R} \) be a function such that
\[
f(cv) = cf(v) \quad \forall c \geq 0, \ v \in \mathbb{R}^m \quad \text{and} \quad f(u + v) \leq f(u) + f(v) \quad \forall u, v \in \mathbb{R}^m.
\]
Then we have
\[
\mathbb{E}[\sup_{x \in T} |f(Ax) - \mathbb{E}[f(Ax)]|] \leq C \text{Lip}(f) \gamma(T),
\]
where \( \gamma(T) \equiv \mathbb{E}[\sup_{x \in T} |\langle g, x \rangle|], \ g \sim \mathcal{N}(0, I_n), \) and \( \text{Lip}(f) \) is the Lipschitz constant of \( f \). Here \( C \) is an absolute universal constant and \( \gamma(T) \) is called the Gaussian complexity of \( T \).

Note that if \( f(x) \equiv \sup_{y \in S} \langle x, y \rangle \), then \( f(x) \) satisfies the conditions of Theorem 1.1 and \( \text{Lip}(f) = \text{rad}(S) \), so (1) is sharp (see \([15, \text{Theorem 11.2.4}]\) and \([15, \text{Exercise 8.7.2}]\)). An important application of matrix deviation inequality is to establish a Lipschitz embedding between two normed spaces. See \([13, \text{Theorem 7}]\) and \([15, \text{Theorem 11.3.3}]\) for using i.i.d ensemble Gaussian matrix to prove the existence of the embedding from a finite dimensional normed spaces into a low-dimensional Euclidean spaces and the embedding from a high-dimensional Euclidean spaces into a low-dimensional Euclidean spaces. See also \([15, \text{Theorem 8.7.1}]\) and \([8]\) for Chevet inequality and \([2]\) for the \( \ell_p \)-Gaussian-Grothendieck problem.

Still, it is a challenging open problem to study the universality of the matrix deviation inequality (see \([15, \text{Remark 11.1.9}]\)). In other words, whether the general matrix deviation inequality holds for i.i.d ensemble sub-Gaussian random matrix, which is a random matrix with i.i.d mean-zero, unit variance, sub-Gaussian entries. Note that this problem was solved by Liaw et al. \([7, \text{Theorem 1}]\) when \( f(\cdot) = \| \cdot \|_2 \). Namely, if \( A \in \mathbb{R}^{m \times n} \) is an i.i.d ensemble sub-Gaussian random matrix with \( K = \|A_{1,1}\|_{\psi_2} \), where \( \|X\|_{\psi_2} \) is the sub-Gaussian norm of \( X \) (see Definition 2.1), then
\[
\mathbb{E}[\sup_{x \in T} \|Ax\|_2 - m^{1/2}\|x\|_2] \|_{\psi_2} \leq CK^2 \gamma(T) \quad \forall T \subseteq \mathbb{R}^n.
\]
In addition, (2) still holds if \( m^{1/2}\|x\|_2 \) is replaced by \( \mathbb{E}[\|Ax\|_2] \). See \([15, \text{Chapter 9}]\) for the applications.
Main Results. In this paper, we aim to prove the matrix deviation inequality for $\ell_p$-norm, $1 \leq p < \infty$, and i.i.d ensemble sub-Gaussian random matrices.

Theorem 1.2. Let $A \in \mathbb{R}^{m \times n}$ be a random matrix with i.i.d, mean-zero, unit variance, sub-Gaussian entries $\{A_{i,j}\}$ and $K = \|A_{i,j}\|_2$. Then we have

$$\left\| \sup_{x \in T} \|Ax\|_p - m^{\frac{1}{p}} \|A_1 x\|_{L^p} \right\|_2 \leq \begin{cases} C_p K^{p} \|A_{1,1}\|_{L^p}^{(p+2)} \text{Lip}(\| \cdot \|_p) \gamma(T), & \text{if } p \in [1, 2] \\ C_p K^{p+2} \text{Lip}(\| \cdot \|_p) \gamma(T), & \text{if } p \in (2, \infty). \end{cases} \tag{3}$$

where $A_i$ is the $i$th row of $A$, $\|A_1 x\|_{L^p} = \mathbb{E}[\|A_1 x\|_p^p]$, and $C_p$ is a positive absolute constant depending only on $p$. In addition, (3) still holds if $m^{1/p}\|A_1 x\|_{L^p}$ is replaced by $\mathbb{E}[\|Ax\|_p]$.

As a consequence of Theorem 1.2, we show that any i.i.d ensemble sub-Gaussian random matrix can be regarded as an embedding from $\ell^2$-space to $\ell^m_p$-space such that the distances between two points do not increase by more than a factor $D_p(1 + \epsilon)$ and do not decrease by more than a factor $d_p(1 - \epsilon)$.

Corollary 1.3. Let $\epsilon \in (0, 1)$, $T$ be a finite subset of $\mathbb{R}^n$ containing $N$ elements, and

$$(d_p, D_p) = \begin{cases} (C_p \|A_{1,1}\|_p, 1), & \text{if } 1 \leq p < 2, \\ (1, 1), & \text{if } p = 2, \\ (1, C_p K), & \text{if } 2 < p < \infty. \end{cases} \tag{4}$$

Then, under assumption of Theorem 1.2, we have

$$\mathbb{P}
\left(
 d_p(1 - \epsilon) \|x - y\|_2 \leq \| \frac{1}{m^{1/p}} A(x - y) \|_p \leq D_p(1 + \epsilon) \|x - y\|_2 \quad \forall x, y \in T
\right)
\geq \begin{cases} 1 - 2 \exp( - C_p K^{2p} (\|A_{1,1}\|_p)^{2(p+3)} \log(N)) & \text{if } 1 \leq p \leq 2 \\ 1 - 2 \exp( - C_p K^{2p+2} \log(N)) & \text{if } 2 < p < \infty. \end{cases} \tag{5}$$

Remark 1.4. For the problem of dimension reduction, Brinkman and Charikar [1] and Ping Li [6] both give overviews of the results in this area. See also [5] for the problem of distortion and [15, Section 11.3] for the random projection. For more general results and similar problems, see [9], [10], and [11].
**Heuristics.** The core of our proof of Theorem 1.2 is to show that

\[ R_x \equiv \|Ax\|_p - m^\frac{1}{p} \|A_1 x\|_{L^p} \quad \forall x \in \mathbb{R}^p \]

has sub-Gaussian increments (see Lemma 3.5). To do this, we will use the approach given by [13, The proof of Lemma 3] and [7, p. 292], which indicates that it suffices to consider some special cases of Lemma 3.5 (see Lemma 3.2 and Lemma 3.4). In order to prove these special cases, we establish the sub-Gaussian concentration inequality with respect to \( \ell_p \)-norm by using [4, Corollary 1.4] to control the tail probability of the sum of i.i.d \( 2/p \)-Orlicz random variables \( |X_j|^p \) (see Lemma 3.3), where \( X_1 \) is a mean-zero sub-Gaussian random variable (note that Lemma 3.3 is a generalization of [7, Proposition 5.1]).

## 2 \( \alpha \)-Orlicz Random Variables

**Definition 2.1.** Let \( \alpha > 0 \) and \( X \) be a random variable. The \( \alpha \)-Orlicz norm of \( X \) is defined by

\[ \|X\|_{\psi_\alpha} \equiv \inf\{t > 0 : \mathbb{E}[\exp(|X|^\alpha/t^\alpha)] \leq 2\} \quad (6) \]

(For convenience, we set \( \inf \emptyset = \infty \)).

We say \( X \) is a \( \alpha \)-Orlicz random variable if \( \|X\|_{\psi_\alpha} < \infty \). In particular, we say \( X \) is a sub-Gaussian random variable if \( \|X\|_{\psi_2} < \infty \). Note that \( \| \cdot \|_{\psi_\alpha} \) is a norm if and only if \( \alpha \geq 1 \). Nevertheless, \( \| \cdot \|_{\psi_\alpha} \) still make sense for any \( \alpha > 0 \).

The following proposition states the equivalent definitions of \( \| \cdot \|_{\psi_\alpha} \), which will be used throughout this paper. Note that the proof of Proposition 2.2 is the same as [15, Proposition 2.5.2].

**Proposition 2.2.** Let \( \alpha > 0 \) and \( X \) be a random variable. Then the following properties are equivalent:

(a) The MGF of \( |X|^\alpha \) is bounded at some point, namely

\[ \mathbb{E}[\exp(|X|^\alpha/K_1^\alpha)] \leq 2. \quad (7) \]

(b) The tails of \( X \) satisfy

\[ \mathbb{P}(|X| \geq t) \leq 2 \exp(-t^\alpha/K_2^\alpha) \quad \forall t \geq 0. \quad (8) \]

(c) The moments of \( X \) satisfy

\[ \|X\|_{L^p} \leq K_3 p^\frac{1}{\alpha} \quad \forall p \geq \alpha \quad (9) \]
Here the parameters $K_i > 0$ appearing in these properties differ from each other by at most a constant that depends on $\alpha$.

By the definition of $\| \cdot \|_{\psi_\alpha}$, it is clear that we have the following relation.

**Lemma 2.3.** Let $X$ be a random variable such that $\|X\|_{\psi_\alpha \beta} \vee \|X\|_{\psi_\alpha} < \infty$. Then $\|X\|_{\psi_\alpha} = \|X\|_{\psi_\alpha \beta}^\beta$.

### 3 Proofs

#### 3.1 Proof of Theorem 1.2

Consider the norm induced by $A$ as follows and recall the definition of $R_x$:

$$
\|x\| \equiv \|A_1 x\|_{L^p} \quad \text{and} \quad R_x = \|Ax\|_p - \frac{1}{m^p} \|x\| \quad \forall x \in \mathbb{R}^n.
$$

(10)

Recall the Generic chaining bound [15, Theorem 8.5.3]. Note that the proof of [15, Theorem 8.5.3] actually gives the following estimation.

**Proposition 3.1.** Let $T \subseteq \mathbb{R}^n$, $x_0 \in T$, and $\{R_x\}_{x \in T}$ be a random process such that $\|R_x - R_y\|_{\psi_2} \leq \mathcal{K}\|x - y\|_2$ for every $x, y \in T$. Then

$$
\left\| \sup_{x \in T} |R_x - R_{x_0}| \right\|_{\psi_2} \leq C \mathcal{K} \gamma(T),
$$

where $C$ is a positive absolute constant. In particular, if $R_{x_0} = 0$, then $\left\| \sup_{x \in T} |R_x| \right\|_{\psi_2} \leq C \mathcal{K} \gamma(T)$.

To prove Theorem 1.2, it suffices to show that $\{R_x\}_{x \in \mathbb{R}^n}$ has sub-Gaussian increments (i.e., Lemma 3.5). Indeed, since $\gamma(T \cup \{0\}) = \gamma(T)$ and $R_0 = 0$, it follows that (3) is an immediate consequence of Proposition 3.1 and Lemma 3.5. Also, by triangle inequality and $\sup_{x \in T} \|x\|_2 C = \sup_{x \in T} \mathbb{E}[\|g, x\|] = \gamma(T)$, where $C = \mathbb{E}[\|g\|]$ and $g \sim \mathcal{N}(0, 1)$, it is clear that $m^{1/p} \|A_1 x\|_{L^p}$ can be replaced by $\mathbb{E}[\|Ax\|_p]$.

Let us start with some properties that will be used throughout the proof.

(a) Applying Jensen’s inequality shows that

$$
K \geq \inf \{ t > 0 : \exp \left( \frac{\mathbb{E}[\|A_1 x\|_2^2]}{t^2} \right) \leq 2 \} = \sqrt{\frac{1}{\ln 2}} > 1.
$$

(11)
(b) Note that \(\| \cdot \|\) and \(\| \cdot \|_2\) are equivalent. Namely,
\[
C_p \|A_{1,1}\|_{L^p} \|x\|_2 \leq \|x\| \leq \|x\|_2 \quad \forall 1 \leq p \leq 2
\]
and
\[
\|x\|_2 \leq \|x\| \leq C'_p K \|x\|_2 \quad \forall 2 \leq p < \infty,
\]
where \(C_p\) and \(C'_p\) are positive constants that depend on \(p\). The proof of (13) follows from [15, Exercise 2.6.5]. The lower bound of (12) is an immediate consequence of Marcinkiewicz–Zygmund inequality [3, Section 10.3] and Minkowski’s integral inequality [14, Theorem 6.2.7]. Indeed, if \(\|x\|_2 = 1\), then
\[
\|x\| \geq C_p E[(\sum_{j=1}^n (A_{1,j}x_j)^2)^{p/2}]^{1/p} \geq C_p (\sum_{j=1}^n x_j^2 E[|A_{1,j}|^p]^{1/p})^{1/2} = C_p \|A_{1,1}\|_{L^p}.
\]

(c) Applying Hölder inequality gives
\[
\text{Lip}(\| \cdot \|_p) = \begin{cases} \frac{1}{m_p} - \frac{1}{2}, & \text{if } 1 \leq p \leq 2 \\ 1, & \text{if } 2 \leq p < \infty. \end{cases}
\]

3.1.1 Case 1: \(x \in \mathbb{R}^n\) and \(y = 0\)

**Lemma 3.2.** Under assumption of Theorem 1.2, we have
\[
\|R_x\|_{\psi_2} \leq \left\{ \begin{array}{ll} C_p (K/\|A_{1,1}\|_{L^p})^p \text{Lip}(\| \cdot \|_p) \|x\|_2, & \text{if } 1 \leq p \leq 2 \\ C_p K^p \text{Lip}(\| \cdot \|_p) \|x\|_2, & \text{if } 2 \leq p < \infty. \end{array} \right.
\]

To prove Lemma 3.2, it suffices to establish Lemma 3.3. Indeed, by [15, Proposition 2.6.1], we have \(\|A_{1,1}\|_{L^p} \leq CK \|x\|_2\), so applying the following lemma gives
\[
\|R_x\|_{\psi_2} \leq C_p \|x\| K^p (\|x\|_2)^p \leq C_p K^p \|A_{1,1}\|_{L^p} \|x\|_2 \quad \text{if } 1 \leq p \leq 2; \\
\|R_x\|_{\psi_2} \leq C_p \|x\| K^p (\|x\|_2)^p = C_p \|x\|_2 K^p (\|x\|_2)^{p-1} \leq C_p \|x\|_2 K^p \quad \text{if } 2 \leq p < \infty.
\]

**Lemma 3.3.** Let \(1 \leq p < \infty\) and \(\{X_i\}_{1 \leq i \leq \infty}\) be i.i.d sub-Gaussian random variables such that \(\|X_1\|_{L^p} = 1\) and \(K \equiv \|X_1\|_{\psi_2} < \infty\). Then, for each \(m \geq 1\) and \(X^{(m)} = (X_1, ..., X_m)\), we have
\[
\|\|X^{(m)}\|_p - m^p \| \|_{\psi_2} \leq C_p K^p \text{Lip}(\| \cdot \|_p),
\]
where \(C_p\) is a positive absolute constant depending only on \(p\).
Proof. To prove Lemma 3.3, it suffices to show that
\[
\mathbb{P}\left( \left\| X^{(m)} \right\|_p - m^{\frac{1}{p}} \geq s \right) \leq \begin{cases} 
2 \exp\left(-C_p \frac{s^2}{K^{2p}m^{\frac{1}{p}}-1}\right), & \text{if } 1 \leq p < 2 \\
2 \exp\left(-C_p \frac{s^2}{K^{2p}}\right), & \text{if } 2 \leq p < \infty
\end{cases} \quad \forall s > 0, 
\tag{16}
\]
where \( C_p \) is a positive absolute constant.

Step 1. In this step, we prove (16) when \( 1 \leq p < \infty \) and \( s \leq K^p m^{\frac{1}{p}} \). Note that if \(|z - 1| \geq \delta \) and \( z \geq 0 \), then \(|z^p - 1| \geq \delta\). Then we have
\[
\mathbb{P}\left( \left| \frac{1}{m^p} \left\| X^{(m)} \right\|_p - 1 \right| \geq \delta \right) \leq \mathbb{P}\left( \left| \frac{1}{m} \sum_{i=1}^{m} (\left| X_i \right|^p - 1) \right| \geq \delta \right).
\]
Since \(||X_i||_{\psi_p} \leq c_p ||X_i||_{\psi_2}\) for each \( 1 \leq p < 2 \), we have
\[
\left| \left| X_i \right|^p - 1 \right|_{\psi_1} \leq c_p \left| \left| X_i \right|^p \right|_{\psi_1} \leq c_p \left| \left| X_i \right|^p \right|_{\psi_p} \leq c_p K^p \quad \forall 1 \leq p < 2
\tag{17}
\]
and
\[
\left| \left| X_i \right|^p - 1 \right|_{\psi_2^p} \leq c_p \left| \left| X_i \right|^p \right|_{\psi_2^p} \leq c_p \left| \left| X_i \right|^p \right|_{\psi_2^p} \leq c_p K^p \quad \forall 2 \leq p < \infty
\tag{18}
\]
by [4, Lemma A.3] and Lemma 2.3. Let \( \alpha_i = \frac{1}{m^p} \). Then applying [4, Corollary 1.4] with \( \alpha = 1 \) if \( 1 \leq p < 2 \); \( \alpha = \frac{2}{p} \) if \( 2 \leq p < \infty \) gives
\[
\mathbb{P}\left( \left| \frac{1}{m} \sum_{i=1}^{m} (\left| X_i \right|^p - 1) \right| \geq \delta \right) \leq 2 \exp\left(-C_p \min\{ \frac{\delta^2}{2}, \frac{\delta}{K^{2p}} \} m \right)
\]
\[
= 2 \exp\left(-C_p \frac{\delta^2 m}{K^{2p}} \right) \quad \forall \delta \leq K^p, \quad 1 \leq p < 2
\tag{19}
\]
and
\[
\mathbb{P}\left( \left| \frac{1}{m} \sum_{i=1}^{m} (\left| X_i \right|^p - 1) \right| \geq \delta \right) \leq 2 \exp\left(-C_p \min\{ \frac{\delta^2}{2}, \frac{\delta^\alpha m^\alpha}{K^{2p}} \} \right)
\]
\[
\leq 2 \exp\left(-C_p \min\{ \frac{\delta^2}{2}, \frac{\delta^\alpha m^\alpha}{K^{2p}} \} \right) \quad \forall \delta \leq K^p, \quad 2 \leq p < \infty.
\tag{20}
\]
Therefore, taking \( s = \delta m^{\frac{1}{p}} \) proves (16) when \( 1 \leq p < \infty \) and \( s \leq K^p m^{\frac{1}{p}} \).
Step 2. In this step, we prove (16) when \( 1 \leq p < 2 \) and \( s > Km^{\frac{1}{p}} \). In fact, we only need to prove (16) when \( 1 \leq p < 2 \) and \( s > Km^{1/p}\xi_p \), where \( \xi_p \) is a positive constant that depends on \( p \). Indeed, if \( \xi_p > 1 \) and \( Km^{1/p} < s < Km^{1/p}\xi_p \), then using the result proved in Step 1 gives

\[
\mathbb{P}\left( \left| X^{(m)} \right|_p - m^{\frac{1}{p}} \geq s \right) \leq \mathbb{P}\left( \left| X^{(m)} \right|_p - m^{\frac{1}{p}} \geq Km^{1/p} \right)
\]

\[
\leq 2 \exp\left( -C_p \frac{(Km^{1/p})^2}{s^2} \right) \leq 2 \exp\left( -\frac{C_p}{\xi_p} K^{2p} m^{\frac{2}{p} - 1} \right).
\]

Note that \( |a^r - b^r| \leq |a - b|^r \) if \( 0 < r \leq 1 \) and \( a, b > 0 \). Hence, the tail probability can be estimated as follows:

\[
\mathbb{P}\left( \left| \frac{1}{m^{\frac{1}{p}}} \sum_{i=1}^{m} |X_i|^p - 1 \right| \geq \delta \right) \leq \mathbb{P}\left( \left| \frac{1}{m} \sum_{i=1}^{m} (|X_i|^p - 1) \right| \geq \delta^p \right)
\]

\[
\leq \mathbb{P}\left( \frac{1}{m} \sum_{i=1}^{m} (|X_i|^p - 1) \geq \delta^p \right) + \mathbb{P}\left( \frac{1}{m} \sum_{i=1}^{m} (1 - |X_i|^p) \geq \delta^p \right).
\]

Both of the above terms can be controlled by the same argument. In the following, we only estimate the first term. Note that (18) holds for \( 1 \leq p < 2 \) as well. Hence, applying [12, Proposition 5.2] with random variable \( |X_i|^p - 1 \) and \( \alpha = \frac{2}{p} \) gives

\[
\mathbb{P}\left( \frac{1}{m} \sum_{i=1}^{m} (|X_i|^p - 1) \geq \delta^p \right) \leq \exp\left( -\lambda t + mC\alpha' K^{p\alpha'} \lambda^{\alpha'} \right) \quad \forall \lambda \geq \frac{1}{K^{pC}\alpha},
\]

where \( C\alpha \) is a positive constant that depends on \( \alpha \), \( t = m\delta^p \), \( \alpha' \) is the Hölder conjugates of \( \alpha \). Note that

\[
\left( \frac{t}{mK^{p\alpha'} C\alpha'} \right)^{\frac{1}{\alpha} - 1} \geq \frac{1}{K^{pC}\alpha} \iff \delta \geq K\alpha^{1/p}.
\]

Hence, if \( \delta \geq K\alpha^{1/p} \) and \( \lambda \equiv \left( \frac{t}{mK^{p\alpha'} C\alpha'} \right)^{\frac{1}{\alpha} - 1} \), then

\[
\mathbb{P}\left( \frac{1}{m} \sum_{i=1}^{m} (|X_i|^p - 1) \geq \delta^p \right)
\]

\[
\leq \exp\left( -\left( \frac{t}{mK^{p\alpha'} C\alpha'} \right)^{\frac{1}{\alpha} - 1} t + mK^{p\alpha'} C\alpha' \left( \frac{t}{mK^{p\alpha'} C\alpha'} \right)^{\frac{\alpha'}{\alpha} - 1} \right) = \exp\left( -C_p \frac{\delta^2 m}{K^2} \right),
\]

8
where $C_p$ is a constant that depends on $p$. Therefore, we obtain

$$
P\left( \left| \frac{1}{m^p} ||X^{(m)}||_p - 1 \right| \geq \delta \right) \leq 2 \exp\left(-C_p \frac{\delta^2 m}{K^2}\right) \leq 2 \exp\left(-C_p \frac{\delta^2 m}{2^p}\right) \quad \forall \delta \geq C_{\alpha}^{1/p}K$$

by using (11), so we complete the proof of (16) when $1 \leq p < 2$ and $s > K m^\frac{1}{p}$.

**Step 3.** In this step, we prove (16) when $2 \leq p < \infty$ and $s > m^\frac{1}{p} K^p$. Decompose the tail probability as (21) and apply [4, Corollary 1.4]. Then we have

$$
P\left( \left| \frac{1}{m^p} ||X^{(m)}||_p - 1 \right| \geq \delta \right) \leq 2 \exp\left(-C_p \min\{\frac{\delta^2}{K^{2p}}, \frac{\delta^\alpha}{K^{op}}\} m^\alpha\right)
= 2 \exp(-C_p \frac{m^\alpha \delta^2}{K^2}) \quad \forall \delta > K^p,$$

so, taking $s = \delta m^\frac{1}{p}$, we complete the proof of this step. \hfill \Box

### 3.1.2 Case 2: $||x|| = ||y|| = 1$

**Lemma 3.4.** Under assumption of Theorem 1.2, we have

$$
\left| \left| R_x - R_y \right| \right|_{\psi_2} \leq \begin{cases} 
C_p(K/||A_{1,1}||_{L^p})^p \text{Lip}(|| \cdot ||_p)||x - y||_2, & \text{if } 1 \leq p \leq 2 \\
C_p K^p \text{Lip}(|| \cdot ||_p)||x - y||_2, & \text{if } 2 < p < \infty 
\end{cases} \quad \forall ||x|| = ||y|| = 1. \tag{23}

**Proof.** To prove Lemma (3.4), it suffices to show that

$$
P\left( \left| \frac{||Ax||_p - ||Ay||_p}{||x - y||} \right| \geq s \right) \leq \begin{cases} 
4 \exp(-C_p \frac{s^2}{(K/||A_{1,1}||_p)^2 m^{\frac{1}{p}}}), & \text{if } 1 \leq p \leq 2 \\
4 \exp(-C_p \frac{s^2}{K^{2p}}), & \text{if } 2 < p < \infty, 
\end{cases} \tag{24}

where $C_p$ is a positive constant that depends $p$. Indeed, since $\mathbb{E}||Z||^N = \int_0^\infty N s^{N-1} \mathbb{P}(||Z|| \geq s) ds$, it is clear that (24) implies (23). Note that if $s \geq 2m^\frac{1}{p}$, then (24) is an immediate consequence of Lemma 3.2. Indeed, if $u = \frac{x - y}{||x - y||}$, then

$$
P\left( \left| \frac{||Ax||_p - ||Ay||_p}{||x - y||} \right| \geq s \right) \leq \mathbb{P}(||Au||_p \geq s) = \mathbb{P}(||Au||_p - m^{\frac{1}{p}} \geq s - m^{\frac{1}{p}})
\leq \mathbb{P}(||Au||_p - m^{\frac{1}{p}} \geq \frac{s}{2})$$. 

9
so applying Lemma 3.2 gives (24). Thus, it remains to prove (24) when \( s < 2m^{\frac{1}{p}} \). Since \( a^{p-1}|a-b| \leq |a^p - b^p| \) if \( 1 \leq p < \infty \) and \( a, b > 0 \), it follows that
\[
P\left( \frac{||Ax||_p - ||Ay||_p}{||x - y||} \geq s \right) \leq P\left( \frac{||Ax||_p^p - ||Ay||_p^p}{||x - y||} \geq s||Ax||_p^{p-1} \right)
\]
\[
\leq P\left( \frac{||Ax||_p^p - ||Ay||_p^p}{||x - y||} \geq s||Ax||_p^{p-1}, ||Ax||_p \geq \frac{m^{\frac{1}{p}}}{2} \right) + P\left( ||Ax||_p < \frac{m^{\frac{1}{p}}}{2} \right)
\]
\[
\leq P\left( \frac{||Ax||_p^p - ||Ay||_p^p}{||x - y||} \geq \frac{sm^{\frac{1}{p}}}{2p-1} \right) + P\left( ||Ax||_p < \frac{m^{\frac{1}{p}}}{2} \right) \equiv \mathcal{A}_1 + \mathcal{A}_2.
\]
Since \( s < 2m^{\frac{1}{p}} \) and \( ||x|| = 1 \), applying Lemma (3.2) gives
\[
\mathcal{A}_2 \leq P\left( ||Ax||_p - m^{\frac{1}{p}}||x|| \geq \frac{m^{\frac{1}{p}}}{2} \right) \leq P\left( ||Ax||_p - m^{\frac{1}{p}}||x|| \geq \frac{s}{4} \right)
\]
\[
\leq \begin{cases} 
2\exp(-C_p\frac{s^2}{(K/||A_{1,1}||_{L_p})^{2p}m^{\frac{1}{p}}}), & \text{if } 1 \leq p \leq 2 \\
2\exp(-C_p\frac{s^2}{K^{2p}}), & \text{if } 2 < p < \infty.
\end{cases} \tag{25}
\]
To estimate \( \mathcal{A}_1 \), we write \( \mathcal{A}_1 \) as
\[
\mathcal{A}_1 = P\left( \frac{1}{m} \sum_{i=1}^{m} \frac{|A_i x|^p - |A_i y|^p}{||x - y||} \geq \delta \right), \text{ where } \delta = \frac{s}{2p-1m^{\frac{1}{p}}},
\]
so it suffices to show that
\[
\frac{||A_i x|^p - |A_i y|^p}{||x - y||} \leq \begin{cases} 
C_p(K/||A_{1,1}||_{L_p})^p, & \text{if } 1 \leq p \leq 2 \\
C_pK^p, & \text{if } 2 < p < \infty.
\end{cases} \tag{26}
\]
Indeed, since \( \delta = \frac{s}{2p-1m^{\frac{1}{p}}} \leq 2^{2-p} \leq 2K^p \) for every \( 1 \leq p < \infty \), applying [4, Corollary 1.4] similar to (19) and (20) yields
\[
\mathcal{A}_1 \leq \begin{cases} 
2\exp(-C_p\frac{(\delta/2)^2m^{\frac{1}{2}}}{K^{2p}}) = 2\exp(-C_p\frac{s^2}{(K/||A_{1,1}||_{L_p})^{2p}m^{\frac{1}{p}}}), & \text{if } 1 \leq p \leq 2 \\
2\exp(-C_p\frac{(\delta/2)^2m^{2p}}{K^{2p}}) = 2\exp(-C_p\frac{s^2}{K^{2p}}), & \text{if } 2 < p < \infty.
\end{cases} \tag{27}
\]
Hence, it remains to prove (26). Note that \( |a^p - b^p| \leq p|a - b|\sqrt{a^{2p-2} + b^{2p-2}} \) if \( 1 \leq p < \infty \) and \( a, b > 0 \). Thus, we have
\[
\frac{|||A_i x|^p - |A_i y|^p||_{\psi^\frac{1}{p}}}{\psi^\frac{1}{p}} \leq p \left( |||A_i x| - |A_i y||\sqrt{|A_i x|^{2p-2} + |A_i y|^{2p-2}}\right)_{\psi^\frac{1}{p}}.
\]
Also, by Hölder’s inequality, we get \(||XY||_{\psi^2_p} \leq ||X||_{\psi^2_{\frac{p}{r}}} ||Y||_{\psi^2_{\frac{p}{s}}} \) if \(\frac{1}{r} + \frac{1}{s} = 1\), so it follows that

\[
\|\|A_1x| - |A_1y|\|_{\psi^2_p} \leq \|\|A_1x\|^{2p-2} + |A_1y|^{2p-2}\|_{\psi^2_{\frac{p}{r}}} \leq \|\|A_1x| - |A_1y|\|_{\psi^2_p}
\]

Applying \([15, Proposition 2.6.1]\) and Lemma 2.3 gives

\[
\|\|A_1x| - |A_1y|\|_{\psi^2_p} \leq \|\|A_1(x - y)\|_{\psi^2_p} \leq \begin{cases} C_p(\|A_1\|_{L^p})\|x - y\|, & \text{if } 1 \leq p \leq 2 \\ C_p\|x - y\|, & \text{if } 2 < p < \infty \end{cases}
\]

and

\[
\|\|\sqrt{|A_1x|^{2p-2} + |A_1y|^{2p-2}}\|_{\psi^2_{\frac{p}{r}}} \leq \left( C_p K^{2(p-1)}\|x\|^{2p-2} + C_p K^{2(p-1)}\|y\|^{2p-2} \right)^{\frac{1}{2}}
\]

\[
\leq \begin{cases} C_p(\|A_1\|_{L^p})^{p-1}, & \text{if } 1 \leq p \leq 2 \\ C_p K^{p-1}, & \text{if } 2 < p < \infty. \end{cases}
\]

Thus, combining (28) and (29) yields (26). Therefore, by (27) and (25), we establish (24) when \(s < 2m^{\frac{1}{r}}\), which completes the proof of Lemma 3.4.

### 3.1.3 Case 3: General Vectors \(x, y \in \mathbb{R}^n\)

**Lemma 3.5.** Under assumption of Theorem 1.2, we have

\[
\|\|R_x - R_y\|_{\psi^2_p} \leq \begin{cases} C_p K^p\|A_1\|_{L^p}^{(p+2)} Lip(||\cdot||_{L^p})\|x - y\|_2, & \text{if } 1 \leq p \leq 2 \\ C_p K^{p+2} Lip(||\cdot||_{L^p})\|x - y\|_2, & \text{if } 2 < p < \infty \end{cases} \forall x, y \in \mathbb{R}^n.
\]

**Proof.** Without loss of generality, we may suppose that \(\|x\| = 1\) and \(\|y\| > 1\). Set \(\overline{y} \equiv \frac{y}{\|y\|}\).

Observe that

\[
\|\|R_x - R_y\|_{\psi^2_p} \leq \|\|R_x - R\overline{y}\|_{\psi^2_p} + \|R\overline{y} - R_y\|_{\psi^2_p}
\]

\[
\leq \begin{cases} \|\|R_x - R\overline{y}\|_{\psi^2_p} + \|R\overline{y}\|_{\psi^2_p} \|y - \overline{y}\|_2, & \text{if } p \in [1, 2] \\ \|\|R_x - R\overline{y}\|_{\psi^2_p} + \|R\overline{y}\|_{\psi^2_p} C_p K \|y - \overline{y}\|_2, & \text{if } p \in (2, \infty). \end{cases}
\]
Hence, it suffices to show the reverse triangle inequality:

\[
||x - \overline{y}||_2 + ||y - \overline{y}||_2 \leq \begin{cases} 
C_p||A_{1,1}||_{L^p}||x - y||_2, & \text{if } 1 \leq p \leq 2 \\
C_pK||x - y||_2, & \text{if } 2 < p < \infty \end{cases} \forall ||x|| = 1, ||y|| > 1 \tag{32}
\]

since applying Lemma 3.2, Lemma 3.4, and (32) to (31) yields Lemma 3.5. Let \( \theta \) be the angle between \( x - \overline{y} \) and \( y - \overline{y} \) such that \( 0 \leq \theta \leq \pi \), i.e., \( \cos \theta = \frac{(x - \overline{y})(y - \overline{y})}{||x - \overline{y}||_2||y - \overline{y}||_2} \). It is easy to see that (32) holds if \( \frac{\pi}{2} \leq \theta \leq \pi \). Indeed, since \( \cos \theta \leq 0 \), applying the law of cosines gives

\[
(||x - \overline{y}||_2 + ||y - \overline{y}||_2)^2 \leq 2(||x - \overline{y}||_2^2 + ||y - \overline{y}||_2^2) - 4\cos(\theta)||x - \overline{y}||_2||y - \overline{y}||_2
\]

\[
= 2||x - y||_2^2.
\]

In addition, if \( \theta = 0 \), then \( \overline{y} = x \) and so there is nothing to prove.

Now, it remains to consider the case of \( 0 < \theta < \frac{\pi}{2} \). Note that there are two possible positions for \( y \) (as shown in Figure 1):

1. If \( y = y_1 \) (see the left of Figure 1), then

\[
||\overline{y} - y||_2 + ||x - \overline{y}||_2 \leq \frac{\cos \theta}{\sin \theta} \sin \tilde{\theta}||x - y||_2 + \frac{1}{\sin \theta} \sin \tilde{\theta}||x - y||_2 \leq \frac{2}{\sin \theta}||x - y||_2; \tag{33}
\]

2. If \( y = y_2 \) (see the right of Figure 1), then

\[
||\overline{y} - y||_2 + ||x - \overline{y}||_2 = \frac{\cos \theta}{\sin \theta} \sin \tilde{\theta}||x - y||_2 + \cos \tilde{\theta}||x - y||_2
\]

\[
+ \frac{1}{\sin \theta} \sin \tilde{\theta}||x - y||_2 \leq \frac{3}{\sin \theta}||x - y||_2. \tag{34}
\]

Thus, it suffices to show that

\[
\sin \theta \geq \begin{cases} 
C_p||A_{1,1}||_{L^p}, & \text{if } 1 \leq p \leq 2 \\
C_pK^{-1}, & \text{if } 2 < p < \infty \end{cases} \forall ||x|| = 1, ||y|| > 1 \text{ such that } 0 < \theta \leq \frac{\pi}{2}. \tag{35}
\]

Define \( B \equiv \{ z \in \mathbb{R}^n : ||z|| \leq 1 \} \) and \( B_2(a, r) \equiv \{ z \in \mathbb{R}^n : ||z - a||_2 \leq r \} \). Applying (12) and (13) yields that \( ||z|| \leq \frac{1}{R_p}||z||_2 \) for every \( z \in \mathbb{R}^n \), where \( R_p = 1 \) if \( 1 \leq p \leq 2 \); \( R_p = \frac{1}{C_pK} \) if \( 2 < p < \infty \). Thus, it follows that \( B_2(0, R_p) \subseteq B, ||x||_2 > R_p/2, \) and \( ||\overline{y}||_2 > R_p/2 \). Hence, there exists an unique \( w \in \partial B_2(0, R_p/2) \) such that \( \overrightarrow{ow} \perp \overrightarrow{yy} \). Let \( \theta' \) be the angle between \( \overrightarrow{x\overline{y}} \) and \( \overrightarrow{yy} \) on the left of Figure 2. Observe that \( 0 < \theta' \leq \theta \). Indeed, if \( \theta < \theta' \) (see the right of Figure 2), there exists \( z \in B \) such
that $z = r\overline{y}$ for some $r > 1$ since $B$ is convex. However, since $w \in B$, we have $||z|| \leq 1$, so we get a contradiction. Therefore applying (12) and (13) implies
\[
\sin \theta \geq \sin \theta' = \frac{Ow}{Oy} = \frac{R_p/2}{||y||_2} \geq \begin{cases} 
C_p||A_{1,1}||_{L^p}, & \text{if } 1 \leq p \leq 2 \\
C_pK^{-1}, & \text{if } 2 < p < \infty
\end{cases}
\]
Similarly, if $n \geq 3$, we consider the two dimensional space spanned by $x, \overline{y}$. Hence, (35) still holds, so we complete the proof of Lemma 3.5.

3.2 Proof of Corollary 1.3

Let $S = \{\frac{x-y}{||x-y||} : x, y \in T \text{ and } x \neq y\}$ and $\tilde{S} = \{\frac{x-y}{||x-y||_2} : x, y \in T \text{ and } x \neq y\}$. Recall that $d_p||z||_2 \leq ||z|| \leq D_p||z||_2$ for every $z \in \mathbb{R}^n$, where $d_p$ and $D_p$ are defined in (4). Thus, we have $\gamma(S) \leq \frac{1}{d_p}\gamma(\tilde{S})$, so it follows that $\gamma(S) \leq \frac{1}{d_p}\sqrt{\log N}$ by using [15, (9.13)]. Therefore applying
Theorem 1.2 and (14) gives

$$
\left\| \sup_{x,y \in T, x \neq y} \frac{1}{m^p} \left\| A(x-y) \right\|_p - 1 \right\|_{\psi_2} = \frac{1}{m^p} \left\| \sup_{z \in S} |R_z| \right\|_{\psi_2} \leq \begin{cases}
C_p K^p \| A_{1,1} \|_p \left( \frac{p+3}{p} \right) \sqrt{\log(N)} \frac{1}{m^{1/p}}, & \text{if } 1 \leq p \leq 2 \\
C_p K^{p+2} \sqrt{\log(N)} \frac{1}{m^{1/p}}, & \text{if } 2 < p < \infty,
\end{cases}
$$

which implies (5).

References

[1] Bo Brinkman and Moses Charikar. On the impossibility of dimension reduction in $\ell_1$. *J. ACM*, 52(5):766–788, 2005. doi:10.1145/1089023.1089026.

[2] Wei-Kuo Chen and Arnab Sen. On $\ell_p$-Gaussian–Grothendieck Problem. *International Mathematics Research Notices*, 11 2021. doi:10.1093/imrn/rnab311.

[3] Y.S. Chow, H. Teicher, G. Casella, S. Fienberg, and I. Olkin. *Probability Theory: Independence, Interchangeability, Martingales*. Springer Texts in Statistics. Springer New York, 1997. doi:10.1007/978-1-4612-1950-7.

[4] Friedrich Götze, Holger Sambale, and Arthur Sinulis. Concentration inequalities for polynomials in $\alpha$-sub-exponential random variables. *Electronic Journal of Probability*, 26:1 – 22, 2021. doi:10.1214/21-EJP606.

[5] James R. Lee and Assaf Naor. Embedding the diamond graph in $L_p$ and dimension reduction in $L_1$. *Geometric and Functional Analysis*, 14(4):745–747, 2004. doi:10.1007/s00039-004-0473-8.

[6] Ping Li. Stable random projections and conditional random sampling, two sampling techniques for modern massive datasets. 2007.

[7] 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*, pages 277–299. Springer International Publishing, 2017. doi:10.1007/978-3-319-45282-1_18.
[8] Michael B. Marcus and Michel Talagrand. Chevet's theorem for stable processes II. *Journal of Theoretical Probability*, 1(1):65–92, January 1988. doi:10.1007/bf01076288.

[9] Shahar Mendelson. On weakly bounded empirical processes. *Mathematische Annalen*, 340(1):293–314, 2008. doi:10.1007/s00208-007-0152-9.

[10] Shahar Mendelson, Alain Pajor, and Nicole Tomczak-Jaegermann. Uniform uncertainty principle for bernoulli and subgaussian ensembles. *Constructive Approximation*, 28:277–289, 12 2008. doi:10.1007/s00365-007-9005-8.

[11] Shahar Mendelson and Nicole Tomczak-Jaegermann. A subgaussian embedding theorem. *Israel Journal of Mathematics*, 164:349–364, 2008. doi:10.1007/s11856-008-0034-1.

[12] Holger Sambale. Some notes on concentration for $\alpha$-subexponential random variables. 2020. doi:10.48550/ARXIV.2002.10761.

[13] Gideon Schechtman. Two observations regarding embedding subsets of euclidean spaces in normed spaces. *Advances in Mathematics*, 200(1):125–135, 2006. doi:10.1016/j.aim.2004.11.003.

[14] D.W. Stroock. *Essentials of Integration Theory for Analysis*. Graduate Texts in Mathematics. Springer New York, 2011. doi:10.1007/978-1-4614-1135-2.

[15] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018. doi:10.1017/9781108231596.