ROW PRODUCTS OF RANDOM MATRICES

MARK RUDELSON

ABSTRACT. Let $\Delta_1, \dots, \Delta_K$ be $d \times n$ matrices. We define the row product of these matrices as a $d^K \times n$ matrix, whose rows are entry-wise products of rows of $\Delta_1, \dots, \Delta_K$. This construction arises in certain computer science problems. We study the question, to which extent the spectral and geometric properties of the row product of independent random matrices resemble those properties for a $d^K \times n$ matrix with independent random entries. In particular, we show that the largest and the smallest singular values of these matrices are of the same order, as long as $n \ll d^K$.

We also consider a problem of privately releasing the summary information about a database, and use the previous results to obtain a bound for the minimal amount of noise, which has to be added to the released data to avoid a privacy breach.

1. Introduction

This paper discusses spectral and geometric properties of a certain class of random matrices with dependent rows, which are constructed from random matrices with independent entries. Such constructions first appeared in computer science, in the study of privacy protection for contingency tables. The behavior of the extreme singular values of various random matrices with dependent entries has been extensively studied in the recent years [1], [2], [9], [16], [22]. These matrices arise in asymptotic geometric analysis [1], signal processing [2], [16], statistics [22] etc. The row products studied below have also originated in a computer science problem [9].

For two matrices with the same number of rows we define the row product as a matrix whose rows consist of entry-wise product of the rows of original matrices.

Definition 1.1. Let x and y be $1 \times n$ matrices. Denote by $x \otimes_r y$ the $1 \times n$ matrix, whose entries are products of the corresponding entries of x and y: $x \otimes_r y(j) = x(j) \cdot y(j)$. If A is an $N \times n$ matrix, and B is

Key words and phrases. Random matrices, extreme singular values, privacy protection.

Research was supported in part by NSF grants DMS-0907023 and DMS-1161372.

an $M \times n$ matrix, denote by $A \otimes_r B$ an $NM \times n$ matrix, whose rows are entry-wise products of the rows of A and B:

$$(A \otimes_r B)_{(j-1)M+k} = A_j \otimes_r B_k,$$

where $(A \otimes_r B)_l, A_j, B_k$ denote rows of the corresponding matrices.

Row products arise in a number of computer science related problems. They have been introduced in [7] and studied in [24] in the theory of probabilistic automata. They also appeared in compressed sensing, see [3] and [6], as well as in privacy protection problems [9]. These papers use different notation for the row product; we adopt the one from [6].

This paper considers spectral and geometric properties of row products of a finite number of independent random matrices. The definition above assumes a certain order of the rows of the matrix $A \otimes_r B$. This order, however, is not important, since changing the relative positions of rows of a matrix doesn't affect its eigenvalues and singular values. Therefore, to simplify the notation, we will denote the row of the matrix $C = A \otimes_r B$ corresponding to the rows A_j and B_k by $C_{j,k}$. We will use a similar convention for the rows of the row products of more than two matrices.

Recall that the singular values of $N \times n$ random matrix A are the eigenvalues of $(A^*A)^{1/2}$ written in the non-increasing order: $s_1(A) \geq s_2(A) \geq \ldots \geq s_n(A) \geq 0$. The first and the last singular values have a clear geometric meaning: $s_1(A)$ is the norm of A, considered as a linear operator from ℓ_2^n to ℓ_2^N, and if $n \leq N$ and $\operatorname{rank}(A) = n$, then $s_n(A)$ is the reciprocal of the norm of A^{-1} considered as a linear operator from $\ell_2^N \cap A\mathbb{R}^n$ to ℓ_2^n. The quantity $\kappa(A) = s_1(A)/s_n(A)$, called *the condition number* of A, controls the error level and the rate of convergence of many algorithms in numerical linear algebra. The matrices with bounded condition number are "nice" embedding of $\mathbb{R}^n$ into $\mathbb{R}^N$, i.e. they don't significantly distort the Euclidian structure. This property holds, in particular, for random $N \times n$ matrices with independent centered subgaussian entries having unit variance, as long as $N \gg n$.

Obviously, the row product of several matrices is a submatrix of their tensor product. This fact, however, doesn't provide much information about the spectral properties of the row product, since they can be different from those of the tensor product. In particular, for random matrices, the spectra of $A \otimes B$ and $A \otimes_r B$ are, indeed, very different. For example, let $d \leq n \leq d^2$, and consider $d \times n$ matrices A and B with independent ± 1 random values. The spectrum of $A \otimes B$ is the product of spectra of A and B, so the norm of $A \otimes B$ will be of the order

$$O\big((\sqrt{n}+\sqrt{d})^2\big) = O(n),$$

and the last singular value is $O\big((\sqrt{n}-\sqrt{d})^2\big)$ whenever $d < n$, see [17]. From the other side, computer experiments show that the extreme singular values of the row product behave as for the $d^2 \times n$ matrix with independent entries, i.e. the first singular value is

$$O(d+\sqrt{n}) = O(d),$$

and the last one is $O(d-\sqrt{n})$, see [9]. Based on this data, it was conjectured that the extreme singular values of the row product of several random matrices behave like for the matrices with independent entries. This fact was established in [9] up logarithmic terms, whose powers depended on the number of multipliers. We remove these logarithmic terms in Theorems 1.3 and 1.5 for row products of any fixed number of random matrices with independent bounded entries. To formulate these results more precisely, we introduce a class of uniformly bounded random variables, whose variances are uniformly bounded below. To shorten the notation we summarize their properties in the following definition.

Definition 1.2. Let $\delta > 0$. We will call a random variable ξ a δ random variable if $|\xi| \le 1$ a.s., $\mathbb{E}\xi = 0$, and $\mathbb{E}\xi^2 \ge \delta^2$.

We start with an estimate of the norm of the row product of random matrices with independent δ random entries.

Theorem 1.3. *Let $\Delta_1, \dots, \Delta_K$ be $d \times n$ matrices with independent δ random entries. Then the K-times entry-wise product $\Delta_1 \otimes_r \Delta_2 \otimes_r \dots \otimes_r \Delta_K$ is a $d^K \times n$ matrix satisfying*

$$\mathbb{P}\left(\|\Delta_1 \otimes_r \dots \otimes_r \Delta_K\| \ge C'(d^{K/2}+n^{1/2})\right) \le \exp\left(-c\left(d+\frac{n}{d^{K-1}}\right)\right).$$

The constants C', c may depend upon K and δ.

The paper [9] uses an ε-net argument to bound the norm of the row product. This is one of the sources of the logarithmic terms in the bound. To eliminate these terms, we use a different approach. The expectation of the norm is bounded using the moment method, which is one of the standard tools of the random matrix theory. The moment method allows to bound the probability as well. However, the estimate obtained this way would be too weak for our purposes. Instead, we apply the measure concentration inequality for convex functions, which is derived from Talagrand's measure concentration theorem.

The bound for the norm in Theorem 1.3 is the same as for a $d^K \times n$ random matrix with bounded or subgaussian i.i.d. entries, while the probability estimate is significantly weaker than in the independent case. Nevertheless, the estimate of Theorem 1.3 is optimal both in terms of the norm bound and the probability (see Remarks 5.3 and 5.5 for details). In the important for us case $d^K \ge n$ the assertion of Theorem 1.3 reads

$$\mathbb{P}\left(\|\Delta_1 \otimes_r \ldots \otimes_r \Delta_K\| \ge C'\sqrt{d^K}\right) \le \exp\left(-cd\right).$$

It is well-known that with high probability a random $N \times n$ matrix A with independent identically distributed bounded centered random entries has a bounded condition number, whenever $N \gg n$ (see, e.g. [15]). Our next result shows that the same happens for the row products of random matrices as well. For the next theorem we need the iterated logarithmic function.

Definition 1.4. For $q \in \mathbb{N}$ define the function $\log_{(q)} : (0, \infty) \to \mathbb{R}$ by induction.

(1) $\log_{(1)} t = \max\big(\log t, 1\big)$;
(2) $\log_{(q+1)} t = \log_{(1)}\big(\log_{(q)} t\big)$.

Throughout the paper we assume that the constants appearing in various inequalities may depend upon the parameters K, q, δ, but are independent of the size of the matrices, and the nature of random variables.

Theorem 1.5. *Let K, q, n, d be natural numbers. Assume that*

$$n \le \frac{cd^K}{\log_{(q)} d}.$$

Let $\Delta_1, \ldots, \Delta_K$ be $d \times n$ matrices with independent δ random entries. Then the K-times entry-wise product $\Delta_1 \otimes_r \Delta_2 \otimes_r \ldots \otimes_r \Delta_K$ satisfies

$$\mathbb{P}\left(s_n(\Delta_1 \otimes_r \ldots \otimes_r \Delta_K) \le c'\sqrt{d^K}\right) \le C\exp\left(-\bar{c}d\right).$$

This bound, together with the norm estimate above shows that the condition number of the row product of matrices with δ random entries exceeds a constant with probability $O(\exp(-cd))$. While this probability is close to 0, it is much bigger than that for a $d^K \times n$ random matrix with independent random entries, in which case it is of order $\exp(-d^K)$. However, it is easy to show that this estimate is optimal (see Remarks 5.3 and 8.2). This weak probability bound renders standard approaches to singular value estimates unusable. In particular, the size of a (1/2) net on the sphere S^{n-1} is exponential in n, so the union bound in the ε-net argument breaks down.

This weaker bound not only makes the proofs more technically involved, but also leads to qualitative effects which cannot be observed in the context of random matrices with independent entries. One of the main applications of random matrices in asymptotic geometric analysis is to finding roughly Euclidean or almost Euclidean sections of convex bodies. In particular, the classical theorem of Kashin [8] states that a random section of the unit ball of ℓ_1^N by a linear subspace of dimension proportional to N is roughly Euclidean. The original proof of Kashin used a random ± 1 matrix to construct these sections. The optimal bounds were obtained by Gluskin, who used random Gaussian matrices [5].

The particular structure of the ℓ_1 norm plays no role in this result, and it can be extended to a larger class of convex bodies. Let $D \subset \mathbb{R}^N$ be a convex symmetric body such that $B_2^N \subset D$ and define the *volume ratio* [19] of D by

$$\mathrm{vr}(D) = \left(\frac{\mathrm{vol}(D)}{\mathrm{vol}(B_2^N)} \right)^{1/N}.$$

Assume that the volume ratio of D is bounded: $\mathrm{vr}(D) \le V$. Then for a random $N \times n$ matrix A with independent entries satisfying certain conditions,

$$\mathbb{P}\left(\exists x \in \mathbb{R}^n \; \|Ax\|_D \le (cV)^{-\frac{N}{N-n}} N^{1/2} \|x\|_2\right) \le \exp(-cN).$$

This fact was originally established in [18], and extended in [12] to a broad class of random matrices with independent entries. However, the volume ratio theorem doesn't hold for the row product of random matrices. We show in Lemma 3.2 that there exists a convex symmetric body $D \subset \mathbb{R}^{d^K}$ with bounded volume ratio, such that

$$\inf_{x \in S^{n-1}} \left\| \tilde{\Delta} x \right\|_D \le c(Kd)^{1/2}$$

with probability 1. For $K > 1$ this bound is significantly lower than $N = d^{K/2}$, which corresponds to the independent entries case.

Surprisingly, despite the fact that the general volume ratio theorem breaks down, it still holds for the original case of the ℓ_1 ball. The main result of this paper is the following Theorem.

Theorem 1.6. *Let K, q, n, d be natural numbers. Assume that*

$$n \le \frac{cd^K}{\log_{(q)} d}.$$

and let $\Delta_1, \dots, \Delta_K$ *be* $d \times n$ *matrices with independent* δ *random entries. Then the* K*-times entry-wise product* $\tilde{\Delta} = \Delta_1 \otimes_r \Delta_2 \otimes_r \dots \otimes_r \Delta_K$ *is a* $d^K \times n$ *matrix satisfying*

$$\mathbb{P}\left(\exists x \in S^{n-1} \ \left\|\tilde{\Delta}x\right\|_1 \le c' d^K\right) \le C' \exp\left(-\bar{c}d\right).$$

Note that the results similar to Theorems 1.3, 1.5, and 1.6 remain valid if the matrices $\Delta_1, \dots, \Delta_K$ have different numbers of rows, and the proofs require only minor changes.

The rest of the paper is organized as follows. In Section 2 we consider a privacy protection problem from which the study of row products has originated. We derive an estimate on the minimal amount of noise needed to avoid a privacy breach from Theorem 1.5. Section 3 introduces necessary notation. Section 4 contains an outline of the proofs of Theorems 1.3 and 1.6. Theorem 1.3 is proved in the first part of Section 5. The rest of this section and Section 6 develop technical tools needed to prove Theorem 1.6.

In Section 7 we introduce a new technical method for obtaining lower estimates. The minimal norm of Ax over the unit sphere is frequently bounded via an ε-net argument. The implementation of this approach in [9] was one of the main sources of the parasitic logarithmic terms. In Section 7 the lower bound is handled differently. The required bound is written as the infimum of a random process. The most powerful method of controlling the *supremum* of a random process is to use chaining, i.e. to represent the process as a sum of increments, and control the increments separately [21]. Such method, however, cannot be directly applied to control the *infimum* of a positive random process. Indeed, lower estimates for the increments cannot be automatically combined to obtain the lower estimate for the sum. Nevertheless, In Lemma 7.1 we develop a variant of a chaining, which allows to control the *infimum* of a process. This chaining lemma is the major step in proving Theorem 1.6, which is presented in Section 8, where we also derive Theorem 1.5 from it.

Acknowledgement: the author thanks Dick Windecker and Martin Strauss for pointing out to the papers [7, 24, 3, 6], and the referee for correcting numerous typos in the first version of the paper.

2. Minimal noise for attribute non-privacy

Marginal, or contingency tables are the standard way of releasing statistical summaries of data. Consider a database D, which we view as a $d \times n$ matrix with entries from $\{0, 1\}$. The columns of the matrix are n individual records, and the rows correspond to d attributes of each record. Each attribute is binary, so it may be either present, or absent. For any set of $K+1$ different attributes we release the percentage of records having all attributes from this set. The list of these values for all $\binom{d}{K+1}$ sets forms the contingency table. In the row product notation the contingency table is the subset of coordinates of the vector

$$y = \Big(\overbrace{D \otimes_r \ldots \otimes_r D}^{K+1 \text{ times}}\Big) w,$$

which correspond to all sets of $K+1$ different rows of the matrix D. Here $w \in \mathbb{R}^n$ is the vector with coordinates $w = (1, \ldots, 1)$.

The attribute non-privacy model refers to the situation when $d-1$ rows of the database D are publicly available, or leaked, and one row is sensitive. The analysis of a more general case, where there are more than one sensitive attribute can be easily reduced to this setting. For the comparison of this model with other privacy models see [9], and the references therein. Denote the $(d-1) \times n$ submatrix of D corresponding to non-sensitive attributes by D', and the sensitive vector by x. Then the coordinates of y contain all coordinates of the vector

$$z = \Big(\overbrace{D' \otimes_r \ldots \otimes_r D'}^{K \text{ times}} \otimes_r x^T\Big) w = \Big(\overbrace{D' \otimes_r \ldots \otimes_r D'}^{K \text{ times}}\Big) x,$$

which correspond to K different rows of the matrix D'. Hence, if the database D' is generic, then the sensitive vector y can be reconstructed from D' and the released vector z by solving a linear system. To avoid this privacy breach, the contingency table is released with some random noise. This noise should be sufficient to make the reconstruction impossible, and at the same time, small enough, so that the summary data presented in the contingency table would be reliable. Let z_{noise} be the vector of added noise. Let $\bar{D}'$ be the $\binom{D}{K} \times n$ submatrix of $D' \otimes_r \ldots \otimes_r D'$ corresponding to all K-element subsets of $\{1, \ldots, n\}$. If the last singular value of $\bar{D}'$ is positive, then one can form the left inverse $(\bar{D}')_L^{-1}$ of $\bar{D}'$, and $\left\|(\bar{D}')_L^{-1}\right\| = s_n^{-1}(\bar{D}')$. In this case, knowing the released data $z + z_{\text{noise}}$ we can approximate the sensitive vector x by $x' = (\bar{D}')_L^{-1}(z + z_{\text{noise}})$. Then

$$\|x - x'\|_2 = \left\|(\bar{D}')_L^{-1} z_{\text{noise}}\right\|_2 \le \left\|(\bar{D}')_L^{-1}\right\| \cdot \|z_{\text{noise}}\|_2 .$$

Therefore, if $\|z_{\text{noise}}\|_2 = o(\sqrt{n} \cdot s_n^{-1}(\bar{D}'))$, then $\|x - x'\|_2 = o(\sqrt{n})$. Since the coordinates of x are 0 or 1, we can reconstruct $(1 - o(1))n$ coordinates of x by rounding the coordinates of x'. Thus, the lower estimate of $s_n^{-1}(\bar{D}')$ provides a lower bound for the norm of the noise vector.

We analyze below the case of a random database. Assume that the entries of the database are independent $\{0,1\}$ variables, and the entries in the same column are identically distributed. This means that the distribution of any given attribute is the same for each record, but different attributes can be distributed differently. We exclude almost degenerate attributes, i.e. the attributes having probabilities very close to 0 or 1. In this case bound on the minimal amount of noise follows from

Theorem 2.1. *Let K, q, n, d be natural numbers. Assume that*

$$n \le \frac{cd^K}{\log_{(q)} d}.$$

Let $0 < p' < p'' < 1$, and let $p_1, \dots, p_d$ be any numbers such that $p' < p_j < p''$. Consider a $d \times n$ matrix A with independent Bernoulli entries $a_{j,k}$ satisfying $\mathbb{P}(a_{j,k} = 1) = p_j$ for all $j = 1, \dots, d,\ k = 1, \dots, n$.

Then the K-times entry-wise product $\tilde{A} = A \otimes_r A \otimes_r \dots \otimes_r A$ is a $d^K \times n$ matrix satisfying

$$\mathbb{P}\left(s_n(\tilde{A}) \le c'\sqrt{d^K}\right) \le C' \exp(-\bar{c}d).$$

The constants c, c', C, C' may depend upon the parameters K, q, p', p''.

Proof. This theorem will follow from Theorem 1.5, after we pass to the row product of matrices having independent δ random entries. To this end, notice that if an $m \times n$ matrix U' is formed from the $M \times n$ matrix U by taking a subset of rows, then $s_n(U') \le s_n(U)$.

Let $d = 2Kd' + m$, where $0 \le m < 2K$. For $j = 1, \dots, K$ denote by Δ_j^1 the submatrix of A consisting of rows $(2K(j-1)+1), \dots, (2K(j-1)+K)$, and by Δ_J^0 the submatrix consisting or rows $(2K(j-1)+K+1), \dots, 2Kj$. Let $D_j^1, D_j^0 \in \mathbb{R}^{d'}$ be vectors with coordinates $D_j^1 = (p_{2K(j-1)+1}, \dots, p_{2K(j-1)+K})$ and $D_j^0 = (p_{2K(j-1)+K+1}, \dots, p_{2Kj})$. Set

$$\Delta_j = D_j^0 \otimes_r \Delta_j^1 - D_j^1 \otimes_r \Delta_j^0.$$

Then $\Delta_1, \dots, \Delta_K$ are $d' \times n$ matrices with independent δ random entries for some δ depending on p', p''.

Let $U_s,\ s = 1,2,3$ be $N_s \times n$ matrices, and let $D \in \mathbb{R}^{N_2}$ be a vector with coordinates satisfying $|d_j| \le 1$ for all j. Then for any $x \in \mathbb{R}^n$

$$\left\|(U_1 \otimes_r (D^T \otimes_r U_2) \otimes_r U_3)x\right\|_2 \le \|(U_1 \otimes_r U_2 \otimes_r U_3)x\|_2.$$

Indeed, any coordinate of $(U_1 \otimes_r (D^T \otimes_r U_2) \otimes_r U_3)x$ equals the correspondent coordinate of $(U_1 \otimes_r U_2 \otimes_r U_3)x$ multiplied by some $d_{j,j}$, so the inequality above follows from the bound on $|d_{j,j}|$. This argument shows that for any $(\varepsilon_1, \dots, \varepsilon_K) \in \{0,1\}^K$

$$\left\| \left(\left((D_1^{1-\varepsilon_1})^T \otimes_r \Delta_1^{\varepsilon_1} \right) \otimes_r \dots \otimes_r \left((D_1^{1-\varepsilon_K})^T \otimes_r \Delta_K^{\varepsilon_K} \right) \right) x \right\|_2 \leq \left\| \left(\Delta_1^{\varepsilon_1} \otimes_r \dots \otimes_r \Delta_K^{\varepsilon_K} \right) x \right\|_2 .$$

Therefore,

$$\left\| \left(\Delta_1 \otimes_r \dots \otimes_r \Delta_K \right) x \right\|_2 \leq \sum_{\varepsilon = (\varepsilon_1, \dots, \varepsilon_K) \in \{0,1\}^K} \left\| \left(\Delta_1^{\varepsilon_j} \otimes_r \dots \otimes_r \Delta_K^{\varepsilon_j} \right) x \right\|_2 \leq 2^K \left\| \left(A \otimes_r \dots \otimes_r A \right) x \right\|_2 ,$$

because $\Delta_1^{\varepsilon_j} \otimes_r \dots \Delta_K^{\varepsilon_j}$ is a submatrix of $A \otimes_r \dots \otimes_r A$. Thus, for any $t > 0$

$$\mathbb{P}\left(s_n(A \otimes_r \dots \otimes_r A) < t \right) \leq \mathbb{P}\left(s_n(\Delta_1 \otimes_r \dots \Delta_K) < 2^K t \right).$$

To complete the proof we use Theorem 1.5 with d' in place of d, and note that $d \leq 3Kd'$. □

3. Notation and preliminary results

The coordinates of a vector $x \in \mathbb{R}^n$ are denoted by $(x(1), \dots, x(n))$. Throughout the paper we will intermittently consider x as a vector in $\mathbb{R}^n$ and as an $n \times 1$ matrix. The sequence $e_1, \dots, e_n$ stands for the standard basis in $\mathbb{R}^n$. For $1 \leq p < \infty$ denote by B_p^n the unit ball of the space ℓ_p^n:

$$B_p^n = \left\{ x \in \mathbb{R}^n \mid \|x\|_p = \left(\sum_{j=1}^n |x(j)|^p \right)^{1/p} \leq 1 \right\}.$$

By S^{n-1} we denote the Euclidean unit sphere.

Denote by $\|A\|$ the operator norm of the matrix A, and by $\|A\|_{HS}$ the Hilbert–Schmidt norm:

$$\|A\|_{HS} = \left(\sum_{j,k} |a_{j,k}|^2 \right)^{1/2} .$$

The volume of a convex set $D \subset \mathbb{R}^n$ will be denoted $\mathrm{vol}(D)$, and the cardinality of a finite set J by $|J|$. By $\lfloor x \rfloor$ we denote the integer part of $x \in \mathbb{R}$. Throughout the paper we denote by K the number of terms in the row product, by q the number of iterations of logarithm, and by δ^2 the minimum of the variances of the entries of random matrices. C, c etc. denote constants, which may depend on the parameters K, q, and δ, and whose value may change from line to line.

Let $V \subset \mathbb{R}^n$ be a compact set, and let $\varepsilon > 0$. A set $\mathcal{N} \subset V$ is called an ε-net if for any $x \in K$ there exists $y \in \mathcal{N}$ such that $\|x - y\|_2 \le \varepsilon$. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator, and $\mathcal{N}$ and $\mathcal{N}'$ are ε-nets in B_2^n and B_2^m respectively, then

$$\|T\| \le (1-\varepsilon)^{-1} \sup_{x \in \mathcal{N}} \|Tx\|_2 \le (1-\varepsilon)^{-2} \sup_{x \in \mathcal{N}} \sup_{y \in \mathcal{N}'} \langle Tx, y \rangle.$$

We will use the following *volumetric estimate*. Let $V \subset B_2^n$. Then for any $\varepsilon < 1$ there exists an ε-net $\mathcal{N} \subset V$ such that

$$|\mathcal{N}| \le \left(\frac{3}{\varepsilon} \right)^n .$$

We will repeatedly use Talagrand's measure concentration inequality for convex functions (see [20], Theorem 6.6, or [10], Corollary 4.9).

Theorem (Talagrand)**.** *Let $X_1, \dots, X_n$ be independent random variables with values in $[-1,1]$. Let $f : [-1,1]^n \to \mathbb{R}$ be a convex L-Lipschitz function, i.e.*

$$\forall x, y \in [-1,1]^n \ |f(x) - f(y)| \le L \|x - y\|_2 .$$

Denote by M the median of $f(X_1, \dots, X_n)$. Then for any $t > 0$,

$$\mathbb{P}\Big(|f(X_1, \dots, X_n) - M| \ge t \Big) \le 4 \exp\left(-\frac{t^2}{16L^2} \right).$$

To estimate various norms we will divide the coordinates of a vector $x \in \mathbb{R}^n$ into blocks. Let $\pi : \{1, \dots, n\} \to \{1, \dots, n\}$ be a permutation rearranging the absolute values of the coordinates of x in the non-increasing order: $|x(\pi(1))| \ge \ldots \ge |x(\pi(1))|$. For $l < n$ and $0 \le m$ define

$$N_0 = 0, \ N_m = \sum_{j=0}^{m-1} 4^j l, \text{ and set } I_m = \pi\Big(\{N_m + 1, \dots, N_{m+1}\} \Big).$$

In other words, I_0 contains l largest coordinates of $|z|$, I_1 contains $4l$ next largest, etc. We continue as long as $I_m \neq \emptyset$. The block I_m will be called the m-th block of type l of the coordinates of x. Denote $x|_I$ the restriction of x to the coordinates from the set I. We need the following standard

Lemma 3.1. *Let $b < 1$ and let $x \in B_2^n \cap b B_\infty^n$. For $l \le b^{-2}$ consider blocks $I_0, I_1, \dots$ of type l of the coordinates of x. Then*

$$\sum_{m \ge 0} |I_m| \cdot \|x|_{I_m}\|_\infty^2 \le 5.$$

Proof. Note that the absolute value of any non-zero coordinate of $x_{I_{m-1}}$ is greater or equal $\|x|_{I_m}\|_\infty$. Hence,

$$\sum_{m\geq 0} |I_m| \, \|x|_{I_m}\|_\infty^2 = l \, \|x|_{I_0}\|_\infty^2 + 4 \sum_{m\geq 1} |I_{m-1}| \cdot \|x|_{I_m}\|_\infty^2$$
$$\leq lb^2 + 4 \sum_{m\geq 1} \left\|x|_{I_{m-1}}\right\|_2^2 \leq 5.$$

□

The next lemma shows that Theorem 1.6 cannot be extended from L_1 norm to a general Banach space whose unit ball has a bounded volume ratio.

Lemma 3.2. *There exists a convex symmetric body $D \subset \mathbb{R}^{d^K}$ such that $B_2^{d^K} \subset D$,*

$$\left(\frac{\mathrm{vol}(D)}{\mathrm{vol}(B_2^{d^K})}\right)^{1/d^K} \leq C$$

satisfying

$$\inf_{x\in S^{n-1}} \|(\Delta_1 \otimes_r \ldots \otimes_r \Delta_K)x\|_D \leq c(Kd)^{1/2}$$

for all $d \times n$ matrices $\Delta_1, \ldots, \Delta_K$ with entries 1 or -1.

Proof. Set

$$W = \bigcup_{\varepsilon_1,\ldots,\varepsilon_K \in \{-1,1\}^d} \varepsilon_1 \otimes \ldots \otimes \varepsilon_K$$

and let $D = \mathrm{conv}\left((dK)^{-1/2} W, B_2^{d^K}\right)$. To estimate the volume ratio of D we use Urysohn's inequality [13]:

$$\left(\frac{\mathrm{vol}(D)}{\mathrm{vol}(B_2^{d^K})}\right)^{1/d^K} \leq d^{-K/2} \mathbb{E} \sup_{x\in D} \langle g, x\rangle,$$

where g is a standard Gaussian vector in $\mathbb{R}^{d^K}$. Since

$$D \subset (dK)^{-1/2} \mathrm{conv}(W) + B_2^{d^K},$$

the right hand side of the previous inequality is bounded by

$$1 + d^{-K/2} \cdot (dK)^{-1/2} \mathbb{E} \sup_{x\in W} \langle g, x\rangle \leq 1 + c(dK)^{-1/2} \log^{1/2} |W|,$$

where $|W|$ is the cardinality of W. Since $|W| = 2^{dK}$, the volume ratio of D is bounded by an absolute constant.

Let e_1 be the first basic vector of $\mathbb{R}^n$. The lemma now follows from the equality $(\Delta_1 \otimes_r \ldots \otimes_r \Delta_K)e_1 = \varepsilon_1 \otimes \ldots \otimes \varepsilon_K$, where $\varepsilon_1, \ldots, \varepsilon_K$ are the first columns of the matrices $\Delta_1, \ldots, \Delta_K$. □

4. Outline of the proof

We begin with proving Theorem 1.3. We use the moment method, which is one of the standard random matrix theory tools. To estimate the norm of a rectangular random matrix A with centered entries, one considers the matrix $(A^*A)^p$ for some large $p \in \mathbb{N}$, and evaluates the expectation of its trace using combinatorics. Since $\|A\|^{2p} \le \operatorname{tr}(A^*A)^p$, any estimate of the trace translates into an estimate for the norm. Following a variant of this approach, developed in [4], we obtain an upper bound for the norm of the row product of independent random matrices, which is valid with probability close to 1. However, the moment method alone is insufficient to obtain an exponential bound for the probability. To improve the probability estimate, we combine the bound for the median of the norm, obtained by the moment method, and a measure concentration theorem. To this end we extend Talagrand's measure concentration theorem for convex functions to the functions, which are *polyconvex*, i.e. convex with respect to certain subsets of coordinates.

Before tackling the small ball probability estimate for

$$\min_{x \in S^{n-1}} \|(\Delta_1 \otimes_r \ldots \otimes_r \Delta_K)x\|_1 ,$$

we consider an easier problem of finding a lower bound for $\|(\Delta_1 \otimes_r \ldots \otimes_r \Delta_{K-1} \otimes_r \Delta_K)x\|_1$ for a fixed vector $x \in S^{n-1}$. The entries of the row product are not independent, so to take advantage of independence, we condition on $\Delta_1, \ldots, \Delta_{K-1}$. To use Talagrand's theorem in this context, we have to bound the Lipschitz constant of this norm above, and the median of it below. Such bounds are not available for all matrices $\Delta_1, \ldots, \Delta_{K-1}$, but they can be obtained for "typical" matrices, namely outside of a set of a small probability. Moreover, the bounds will depend on the vector x, so to obtain them, we have to prove these estimates for all submatrices of the row product. This is done in Sections 5.2 and 5.3. Using these results, we bound the small ball probability in Section 6. Actually, we prove a stronger estimate for the *Levy concentration function*, which is the supremum of the small ball probabilities over all balls of a fixed radius.

The final step of the proof is combining the individual small ball probability estimates to obtain an estimate of the minimal ℓ_1-norm over the sphere. This is usually done by introducing an ε-net, and approximating a point on the sphere by its element. Since the small ball probability depends on the direction of the vector x, one ε-net would not be enough. A modification of this method, using several ε-nets was developed in [11]. However, its implementation for the row products lead to appearance of parasitic logarithmic terms, whose degrees rapidly grow with K [9]. To avoid these terms, we develop a new *chaining argument* in Section 7. Unlike standard chaining argument, which is used to bound the supremum of a random process, the method of section 7 applies to the infimum.

In section 8 we combine the chaining lemma with the Levy concentration function bound of Section 6 to complete the proof of Theorem 1.6, and derive Theorem 1.5 from it. We also show that the image of $\mathbb{R}^n$ under the row product of random matrices is a Kashin subspace, i.e. the ℓ_1 and ℓ_2 norms are equivalent on this space.

5. Norm estimates

5.1. **Norm of the matrix.**

We start with a preliminarily estimate of the operator norm of the row product of random matrices. To this end we use the moment method, which is based on bounding the expectation of the trace of high powers of the matrix. This approach, which is standard in the theory of random matrices with independent entries, carries over to the row product setting as well.

Theorem 5.1. *Let $\Delta_1, \dots, \Delta_K$ be $d \times n$ matrices with independent δ random entries. Let $p \in \mathbb{N}$ be a number such that $p \le cn^{1/12K}$. Then the K-times entry-wise product $\tilde{\Delta} = \Delta_1 \otimes_r \Delta_2 \otimes_r \dots \otimes_r \Delta_K$ is a $d^K \times n$ matrix satisfying*

$$\mathbb{E} \left\| \tilde{\Delta} \right\|^{2p} \le p^{2K+1} n \left(d^{1/2} + n^{1/2K} \right)^{2pK}.$$

Proof. The proof of this theorem closely follows [4], so we will only sketch it. Denote the entries of the matrix Δ_l by $\delta_{i,j}^{(l)}$, so the entry of the matrix $\tilde{\Delta}$ corresponding to the product of the entries in the rows $i^{(1)}, i^{(2)} \dots i^{(K)}$ and column j will be denoted $\delta^{(1)}_{i_1^{(1)},j} \cdot \ldots \cdot \delta^{(K)}_{i_1^{(K)},j}$. Then

$$\begin{aligned}
\mathbb{E} \left\| \tilde{\Delta} \right\|^{2p} &\le \mathbb{E} \mathrm{tr} (\tilde{\Delta} \tilde{\Delta}^T)^p \\
&\le \sum_V \mathbb{E} (\delta^{(1)}_{i_1^{(1)},j_1} \cdot \ldots \cdot \delta^{(K)}_{i_1^{(K)},j_1}) \cdot (\delta^{(1)}_{i_2^{(1)},j_1} \cdot \ldots \cdot \delta^{(K)}_{i_2^{(K)},j_1}) \cdot \ldots \\
&\qquad \ldots \cdot (\delta^{(1)}_{i_p^{(1)},j_p} \cdot \ldots \cdot \delta^{(K)}_{i_k^{(K)},j_p}) \cdot (\delta^{(1)}_{i_1^{(1)},j_p} \cdot \ldots \cdot \delta^{(K)}_{i_1^{(K)},j_p}).
\end{aligned}$$

Here V is the set of admissible multi-paths, i.e. a sequence of $2p$ lists $\{(i^{(1)}_{m_1}, j_m), \dots, (i^{(1)}_{m_K}, j_m)\}_{m=1}^{2p}$ such that

1. the column number j_m is the same for all entries of the list m.
2. the first list is arbitrary;

(3) the entries of the second list are in the same column as the entries of the first list, the entries of the third list are in the same rows as the respective entries of the second list, etc.;
(4) the entries of the last list are in the same rows as the respective entries of the first list;
(5) every entry, appearing in each path, appears at list twice.

Since the entries of the matrices $\Delta_1, \dots, \Delta_K$ are uniformly bounded, the expectations are uniformly bounded as well, so

$$\mathbb{E} \left\| \tilde{\Delta} \right\|^{2p} \le |V|.$$

To estimate the cardinality of V denote by $\beta(r_1, \dots, r_K, c)$ the number of admissible multi-paths whose entries are taken from exactly r_1 rows of the matrix Δ_1, exactly r_2 rows of the matrix Δ_2, etc., and exactly from c columns of each matrix. Note that the set of columns through which the path goes is common for the matrices $\Delta_1, \dots, \Delta_K$. An admissible multi-path can be viewed as an ordered K-tuple of closed paths $q_1, \dots, q_K$ of length $2p+1$ in the $d \times n$ bi-partite graph, such that $q_1(2j) = q_2(2j) = \ldots = q_K(2j)$ for $j = 1, \dots, p$, and each edge is traveled at least twice for each path. With this notation we have

$$\mathbb{E} \left\| \tilde{\Delta} \right\|^{2p} \le \sum_J \beta(r_1, \dots, r_K, c), \tag{5.1}$$

where J is the set of sequences of natural numbers $(r_1, \dots, r_K, c)$ satisfying

$$r_l + c \le p + 1 \quad \text{for each } l = 1, \dots, K.$$

The inequality here follows from condition (5) above. Let $\gamma(r_1, \dots, r_K, c)$ be the number of admissible multi-paths, which go through *the first* r_1 rows of the matrix Δ_1, *the first* r_2 rows of the matrix Δ_2, etc., and *the first* c columns. Then

$$\beta(r_1, \dots, r_K, c) \le \binom{n}{c} \cdot \prod_{l=1}^{K} \binom{d}{r_l} \cdot \gamma(r_1, \dots, r_K, c).$$

We call a closed path of length $2p+1$ path in the $d \times n$ bi-partite graph standard if

(1) it starts with the edge $(1, 1)$;
(2) if the path visits a new left (right) vertex, then its number is the minimal among the left (right) vertices, which have not yet been visited by this path;
(3) each edge in the path is traveled at least twice.

Let $m(r,c)$ is the number of the standard paths through r left vertices and c right vertices of the bi-partite graph. Then

$$\gamma(r_1,\dots,r_K,c) \le c! \cdot \prod_{l=1}^{K} r_l! \cdot m(r_l,c).$$

This inequality follows from the fact that all K paths in the admissible multi-path visit a new column vertex at the same time, so the column vertex enumeration defined by different paths of the same multi-path is consistent. Combining two previous estimates, we get

$$\beta(r_1,\dots,r_K,c) \le n^c \cdot \prod_{l=1}^{K} d^{r_l} m(r_l,c).$$

The inequality on page 260 [4] reads

$$m(r,c) \le \binom{p}{r}^2 \cdot p^{12(p-r-c)+14}.$$

Substituting it into the inequality above, we obtain

$$\begin{aligned}
(5.2) \qquad & \sum_J \beta(r_1,\dots,r_K,c) \\
& \le \sum_{c=1}^{p} \sum_{r_1+c\le p+1} \cdots \sum_{r_K+c\le p+1} n^c \cdot \prod_{l=1}^{K} d^{r_l} \cdot \binom{p}{r_l}^2 \cdot p^{12(p-r_l-c)+14} \\
& = \sum_{c=1}^{p} \prod_{l=1}^{K} \sum_{r_l=1}^{p+1-c} n^{c/K} \cdot d^{r_l} \cdot \binom{p}{r_l}^2 \cdot p^{12(p-r_l-c)+14}.
\end{aligned}$$

To estimate the last quantity note that since $p \le \frac{1}{2} n^{1/12K}$,

$$\begin{aligned}
& \sum_{r_l=1}^{p+1-c} n^{c/K} \cdot d^{r_l} \cdot \binom{p}{r_l}^2 \cdot p^{12(p-r_l-c)+14} \\
& = p^2 n^{1/K} \sum_{r_l=1}^{p+1-c} p^{12(p+1-r_l-c)} n^{-(p+1-r_l-c)/K} \binom{p}{r_l}^2 d^{r_l} n^{(p-r_l)/K} \\
& \le p^2 n^{1/K} \left(\sum_{r_l=0}^{p} \binom{p}{r_l} (d^{1/2})^{r_l} (n^{1/2K})^{p-r_l} \right)^2 \\
& = p^2 n^{1/K} \left(d^{1/2} + n^{1/2K} \right)^{2p}.
\end{aligned}$$

Finally, combining this with (5.1) and (5.2), we conclude

$$\mathbb{E} \left\| \tilde{\Delta} \right\|^{2p} \le p^{2K+1} n \left(d^{1/2} + n^{1/2K} \right)^{2pK} \le p^{2K+1} n \cdot \left(d^{1/2} + n^{1/2K} \right)^{2pK}.$$

□

Applying Chebychev's inequality, we can derive a large deviation estimate from the moment estimate of Theorem 5.1.

Corollary 5.2. *Under the conditions of Theorem 5.1,*

$$\mathbb{P}\left(\|\Delta_1 \otimes_r \ldots \otimes_r \Delta_K\| \geq C'(d^{K/2} + n^{1/2})\right) \leq \exp\left(-cn^{\frac{1}{12K}}\right).$$

Remark 5.3. The bound for the norm appearing in Corollary 5.2 matches that for a random matrix with centered i.i.d. entries. This bound is optimal for the row products as well. To see it, assume that the entries of $\Delta_1, \ldots, \Delta_K$ are independent ± 1 random variables. Then $\left\|\tilde{\Delta} e_1\right\|_2 = d^{K/2}$. Also, if $x \in S^{n-1}$ is such that $x(j) = n^{-1/2}\tilde{\delta}_{1,j}$, where $\tilde{\delta}_{1,j}$ is an entry in the first row of the matrix $\tilde{\Delta}$, then $\left\|\tilde{\Delta} x\right\|_2 \geq n^{1/2}$.

More precise versions of the moment method show that the moment bound of the type of Theorem 1.3 is valid for bigger values of p as well, and lead to more precise large deviation bound. We do not pursue this direction here, since these bounds are not powerful enough for our purposes.

Instead, we use the previous corollary to bound the median of the norm of $\Delta_1 \otimes_r \ldots \otimes_r \Delta_K$, and apply measure concentration. The standard tool for deriving measure concentration results for norms of random matrices is Talagrand's measure concentration theorem for convex functions. However, this theorem is not available in our context, since the norm of $\Delta_1 \otimes_r \ldots \otimes_r \Delta_K$ is not a convex function of the entries of $\Delta_1, \ldots, \Delta_K$. We will modify this theorem to apply it to polyconvex functions.

Lemma 5.4. *Consider a function $F : \mathbb{R}^{KM} \to \mathbb{R}$. For $1 \leq k \leq K$ and $x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_K \in \mathbb{R}^M$ define a function $f_{x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_K} : \mathbb{R}^M \to \mathbb{R}$ by*

$$f_{x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_K}(x) = F(x_1, \ldots, x_{k-1}, x, x_{k+1}, \ldots, x_K)$$

Assume that for all $1 \leq k \leq K$ and for all $x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_K \in B_\infty^d$ the functions $f_{x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_K}$ are L-Lipschitz and convex.

Let $(\varepsilon_1, \ldots, \varepsilon_K) = \left((\nu_{1,1}, \ldots, \nu_{1,M}), \ldots, (\nu_{K,1}, \ldots, \nu_{K,M})\right) \in \mathbb{R}^{KM}$ be a set of independent random variables, whose absolute values are uniformly bounded by 1. If

$$\mathbb{P}\left(F(\varepsilon_1, \ldots, \varepsilon_K) \geq \mu\right) \leq 2 \cdot 4^{-K},$$

then for any $t > 0$

$$\mathbb{P}\left(F(\varepsilon_1, \dots, \varepsilon_K) \geq \mu + t\right) \leq 4^K \exp\left(-\frac{ct^2}{K^2L^2}\right).$$

Proof. We prove this lemma by induction on K. In case $K = 1$ the assertion of the lemma follows immediately from Talagrand's measure concentration theorem for convex functions.

Assume that the lemma holds for $K-1$. Let $F : \mathbb{R}^{KM} \to \mathbb{R}$ be a function satisfying the assumptions of the lemma. Set

$$\Omega = \{(x_1, \dots, x_{K-1}) \in B_\infty^{(K-1)M} \mid \mathbb{P}\left(F(x_1, \dots, x_{K-1}, \varepsilon_K) > \mu\right) \geq 1/2\}.$$

Then Chebychev's inequality yields

$$\mathbb{P}\left((\varepsilon_1, \dots, \varepsilon_{K-1}) \in \Omega\right) \leq 4^{-(K-1)}. \tag{5.3}$$

By Talagrand's theorem, for any $(x_1, \dots, x_{K-1}) \in B_\infty^{(K-1)M} \setminus \Omega$

$$\mathbb{P}\left(F(x_1, \dots, x_{K-1}, \varepsilon_K) \geq \mu + \frac{t}{K}\right) \leq 2\exp\left(-\frac{ct^2}{K^2L^2}\right).$$

Hence,

$$\mathbb{P}\left(F(\varepsilon_1, \dots, \varepsilon_K) \geq \mu + \frac{t}{K} \mid (\varepsilon_1, \dots, \varepsilon_{K-1}) \in B_\infty^{(K-1)M} \setminus \Omega\right) \\ \leq 2\exp\left(-\frac{ct^2}{K^2L^2}\right).$$

Define

$$\Xi = \Big\{x_K \in B_\infty^M \mid \mathbb{P}\Big(F(\varepsilon_1, \dots, \varepsilon_{K-1}, x_K) \geq \mu + \frac{t}{K} \mid \\ (\varepsilon_1, \dots, \varepsilon_{K-1}) \in B_\infty^{(K-1)M} \setminus \Omega\Big) > 4^{-(K-1)}\Big\}.$$

The previous estimate and Chebychev's inequality imply

$$\mathbb{P}\left(\varepsilon_K \in \Xi\right) \leq 2 \cdot 4^{K-1} \exp\left(-\frac{ct^2}{K^2L^2}\right).$$

If $x_K \in \Xi^c$, then combining the conditional probability bound with the estimate (5.3), we obtain

$$\mathbb{P}\left(F(\varepsilon_1, \dots, \varepsilon_{K-1}, x_K) \geq \mu + \frac{t}{K}\right) \\ \leq \mathbb{P}\left(F(\varepsilon_1, \dots, \varepsilon_{K-1}, x_K) \geq \mu + \frac{t}{K} \mid (\varepsilon_1, \dots, \varepsilon_{K-1}) \in B_\infty^{(K-1)M} \setminus \Omega\right) + \mathbb{P}(\Omega) \\ \leq 2 \cdot 4^{-(K-1)}.$$

Hence, applying the induction hypothesis with $\frac{K-1}{K}t$ in place of t, we get

$$\mathbb{P}\left(F(\varepsilon_1,\dots,\varepsilon_{K-1},x_K)\geq \mu+\frac{t}{K}+\frac{K-1}{K}t\right)\leq 4^{K-1}\exp\left(-\frac{ct^2}{K^2L^2}\right).$$

Finally,

$$\begin{aligned}&\mathbb{P}\left(F(\varepsilon_1,\dots,\varepsilon_K)\geq \mu+t\right)\\ &\leq \mathbb{P}\left(F(\varepsilon_1,\dots,\varepsilon_K)\geq \mu+t \mid \varepsilon_K\in B_\infty^M\setminus\Xi\right)+\mathbb{P}\left(\varepsilon_K\in\Xi\right)\\ &\leq 4^K\exp\left(-\frac{ct^2}{K^2L^2}\right),\end{aligned}$$

which completes the proof of the induction step. □

This concentration inequality combined with Corollary 5.2 allows to establish the correct probability bound for large deviations of the norm of the row product of random matrices.

Proof of Theorem 1.3. For $k=1,\dots,K$ let $\varepsilon_k\in\mathbb{R}^{dn}$ be the entries of the matrix Δ_k rewritten as a vector. For any matrices $\Delta_1,\dots,\Delta_{k-1}$, $\Delta_{k+1},\dots,\Delta_K$ the function

$$f_{\Delta_1,\dots,\Delta_{k-1},\Delta_{k+1},\dots,\Delta_K}(\Delta_k)=\left\|\Delta_1\otimes_r\dots\otimes_r\Delta_{k-1}\otimes_r\Delta_k\otimes_r\Delta_{k+1}\otimes_r\dots\otimes_r\Delta_K\right\|$$

is convex. Also, since the absolute values of the entries of the matrices $\Delta_1,\dots,\Delta_{k-1},\Delta_{k+1},\dots,\Delta_K$ do not exceed 1,

$$\begin{aligned}&|f_{\Delta_1,\dots,\Delta_{k-1},\Delta_{k+1},\dots,\Delta_K}(\Delta_k)-f_{\Delta_1,\dots,\Delta_{k-1},\Delta_{k+1},\dots,\Delta_K}(\Delta_k')|\\ &\leq\left\|\Delta_1\otimes_r\dots\otimes_r\Delta_{k-1}\otimes_r(\Delta_k-\Delta_k')\otimes_r\Delta_{k+1}\otimes_r\dots\otimes_r\Delta_K\right\|\\ &\leq\left\|\Delta_1\otimes_r\dots\otimes_r\Delta_{k-1}\otimes_r(\Delta_k-\Delta_k')\otimes_r\Delta_{k+1}\otimes_r\dots\otimes_r\Delta_K\right\|_{HS}\\ &\leq d^{(K-1)/2}\left\|\Delta_k-\Delta_k'\right\|_{HS},\end{aligned}$$

so the Lipschitz constant of this function doesn't exceed $d^{(K-1)/2}$. By Corollary 5.2, we can take $\mu=C'(d^{K/2}+n^{1/2})$. Applying Lemma 5.4 with $t=C''(d^{K/2}+n^{1/2})$ finishes the proof. □

Remark 5.5. The probability bound of Theorem 1.3 is optimal. Indeed, assume first that $d^K\geq n$, and let $\Delta_1,\dots,\Delta_K$ be $d\times n$ matrices with independent random ±1 variables. Choose a number $s\in\mathbb{N}$ such that $\sqrt{s}>C$, where C is the constant in Theorem 1.3, and set $x=(e_1+\dots+e_s)/\sqrt{s}$. With probability $2^{-sK\cdot d}$ all entries in the first s columns of these matrices equal 1, so $\|(\Delta_1\otimes_r\dots\otimes_r\Delta_K)x\|_2=\sqrt{s}\cdot d^{K/2}$.

In the opposite case, $n>d^K$, set $s=C^2n/d^K$, where the constant C is the same as above. Then for x defined above we have

$\|(\Delta_1 \otimes_r \ldots \otimes_r \Delta_K)x\|_2 = \sqrt{s} \cdot d^{K/2} = C\sqrt{n}$ with probability at least $2^{-sK \cdot d} = \exp(C' Kn/d^{K-1})$.

5.2. **Norms of the submatrices.** We start with two deterministic lemmas. The first one is a trivial bound for the norm of the row product of two matrices.

Lemma 5.6. *Let U be an $M \times n$ matrix, and let V be a $d \times n$ matrix. Assume that $|v_{i,j}| \le 1$ for all entries of the matrix V. Then $\|U \otimes_r V\| \le \sqrt{d}\, \|U\|$.*

Proof. The matrix $U \otimes_r V$ consists of d blocks $U \otimes_r v_j,\ j = 1, \ldots, d$, where v_j is a row of V. For any $x \in \mathbb{R}^M$

$$\|(U \otimes_r v_j)x\|_2 = \left\|U(v_j \otimes_r x^T)^T\right\|_2 \le \|U\| \cdot \left\|v_j \otimes_r x^T\right\|_2 \le \|U\| \cdot \|x\|_2 .$$

Hence, $\|U \otimes_r V\|^2 \le \sum_{j=1}^d \|U \otimes_r v_j\|^2 \le d\, \|U\|^2 .$ □

The second lemma is based on the block decomposition of the coordinates of a vector.

Lemma 5.7. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear operator. Set $L = \lceil (1/4) \log_2 n \rceil$ and let $1 \le L_0 < L$. For $l = 1, \ldots, L$ denote*

$$\mathcal{M}_l = \{x \in B_2^n \mid |\mathrm{supp}(x)| \le 4^l, \text{and } x(j) \in \{0, 2^{-l}, -2^{-l}\} \text{ for all } j\}.$$

Let $b \le 2^{-L_0}$. Then

$$\|T : B_2^n \cap bB_\infty^n \to B_2^m\| \le \sqrt{5} \left(\sum_{l=L_0}^{L} \max_{z \in \mathcal{M}_l} \|Tz\|_2^2 \right)^{1/2} .$$

Proof. Let $x \in B_2^n \cap bB_\infty^n$. Let $I_0, I_1, \ldots, I_{L-L_0}$ be blocks of type 4^{L_0} of coordinates of x. Recall that $|I_m| = 4^{L_0+m}$. If $x_m \ne 0$, set

$$y_m = |I_m|^{-1/2} \cdot \frac{x|_{I_m}}{\|x|_{I_m}\|_\infty},$$

otherwise $y_m = 0$. Then $\|y_m\|_\infty \le |I_m|^{-1/2} = 2^{-L_0-m}$, and $\|y_m\|_2 \le 1$, so $y_m \in \mathrm{conv}(\mathcal{M}_{L_0+m})$ for all m. By Cauchy–Schwartz inequality,

$$\begin{aligned} \|Tx\|_2 \le \sum_{m=0}^{L-L_0} \|Tx|_{I_m}\|_2 &\le \left(\sum_{m=0}^{L-L_0} |I_m| \cdot \|x|_{I_m}\|_\infty^2 \right)^{1/2} \cdot \left(\sum_{m=0}^{L-L_0} \|Ty_m\|_2^2 \right)^{1/2} \\ &\le \left(\sum_{m=0}^{L-L_0} |I_m| \cdot \|x|_{I_m}\|_\infty^2 \right)^{1/2} \cdot \left(\sum_{m=0}^{L-L_0} \max_{z \in \mathcal{M}_{L_0+m}} \|Tz\|_2^2 \right)^{1/2} . \end{aligned}$$

The estimate of Lemma 3.1 completes the proof. □

For $k \in \mathbb{N}$ denote by $\mathcal{W}_k$ the set of all $d^k \times n$ matrices V satisfying

$$\|V|_J\| \le C_k \left(d^{k/2} + \sqrt{|J|} \cdot \log^{k/2} \left(\frac{en}{|J|} \right) \right). \tag{5.4}$$

for all non-empty subsets $J \subset \{1, \dots, n\}$. Here $V|_J$ denotes the submatrix of V with columns belonging to J, and C_k is a constant depending on k only. This definition obviously depends on the choice of the constants C_k. These constants will be defined inductively in the proof of Lemma 5.9 and then fixed for the rest of the paper.

We will prove that the row product of random matrices satisfies condition (5.4) with high probability. To this end we need an estimate of the norm of a vector consisting of i.i.d. blocks of coordinates.

Lemma 5.8. *Let W be an $m \times n$ matrix. Let $\theta \in \mathbb{R}^n$ be a vector with independent δ random coordinates. For $l \in \mathbb{N}$ let $Y_1, \dots, Y_l$ be independent copies of the random variable $Y = \|W\theta\|$. Then for any $s > 0$*

$$\mathbb{P}\left(\sum_{j=1}^{l} Y_j^2 \ge 4l \|W\|_{HS}^2 + s \right) \le 2^l \cdot \exp\left(-\frac{cs}{\|W\|^2} \right).$$

Proof. Note that $F : \mathbb{R}^n \to \mathbb{R},\ F(x) = \|Wx\|$ is a Lipschitz convex function with the Lipschitz constant $\|W\|$. By Talagrand's theorem

$$\mathbb{P}\left(|Y - M| \ge t\right) \le 4 \exp\left(-\frac{t^2}{16 \|W\|^2} \right),$$

where $M = \mathbb{M}(Y)$ is the median of Y. For $j = 1, \dots, l$ set $Z_j = |Y_j - M|$. Then the previous inequality means that Z_j is a ψ_2 random variable, i.e.

$$\mathbb{E} \exp\left(\frac{c' Z_j^2}{\|W\|^2} \right) \le 2$$

for some constant $c' > 0$. By the Chebychev inequality and independence of $Z_1, \dots, Z_l$,

$$\mathbb{P}\left(\sum_{j=1}^{l} Z_j^2 > t \right) = \mathbb{P}\left(\frac{c'}{\|W\|^2} \sum_{j=1}^{l} Z_j^2 > \frac{c't}{\|W\|^2} \right) \le 2^l \cdot \exp\left(-\frac{c't}{\|W\|^2} \right).$$

Using the elementary inequality $x^2 \le 2(x-a)^2 + 2a^2$, valid for all $x, a \in \mathbb{R}$, we derive that

$$\mathbb{P}\left(\sum_{j=1}^{l} Y_j^2 > 2lM^2 + 2t \right) \le \mathbb{P}\left(\sum_{j=1}^{l} Z_j^2 > t \right) \le 2^l \cdot \exp\left(-\frac{c't}{\|W\|^2} \right).$$

By Markov's inequality, $M^2 = \mathbb{M}(Y^2) \le 2\mathbb{E}Y^2$. To finish the proof, notice that since the coordinates of θ are independent,

$$\mathbb{E}Y^2 = \sum_{j=1}^{m}\sum_{k=1}^{n} w_{j,k}^2 \cdot \mathbb{E}\theta_k^2 \le \|W\|_{HS}^2 .$$

□

The next lemma shows that a "typical" row product of random matrices satisfies (5.4).

Lemma 5.9. *Let $d, n, k \in \mathbb{N}$ be numbers satisfying $n \ge d^{k+1/2}$. Let $\Delta_1, \dots, \Delta_k$ matrices with independent δ random entries. There exist numbers $C_1, \dots, C_k > 0$ such that*

$$\mathbb{P}\left(\Delta_1 \otimes_r \dots \otimes_r \Delta_k \notin \mathcal{W}_k\right) \le k e^{-cd}.$$

Proof. We use the induction on k.

Step 1. Let $k = 1$. In this case Δ_1 is a matrix with independent δ random entries. For such matrices the result is standard and follows from an easy covering argument. Let $x \in S^{d-1}$, and let $y \in S^{n-1} \cap \mathbb{R}^J$. Then $\langle x, \Delta_1|_J y\rangle$ is a linear combination of independent δ random variables. By Hoeffding's inequality (see e.g. [23]),

$$\mathbb{P}\left(|\langle x, \Delta_1|_J y\rangle| > t\right) \le e^{-ct^2}$$

for any $t \ge 1$. Let $J \subset \{1, \dots, n\}$, $|J| = m$. Let $\mathcal{N}$ be a $(1/2)$-net in S^{d-1}, and let $\mathcal{M}$ be a $(1/2)$-net in $S^{n-1} \cap \mathbb{R}^J$. Then

$$\|\Delta_1|_J\| \le 4 \sup_{x\in\mathcal{N}} \sup_{y\in\mathcal{M}} \langle x, \Delta_1|_J y\rangle.$$

The nets $\mathcal{N}$ and $\mathcal{M}$ can be chosen so that $|\mathcal{N}| \le 6^d$ and $|\mathcal{M}| \le 6^m$. Combining this with the union bound, we get

$$\mathbb{P}\left(\|\Delta_1|_J\| \ge 4t\right) \le |\mathcal{N}|\cdot|\mathcal{M}|\cdot e^{-ct^2} \le \exp\left(-ct^2 + (m+d)\log 6\right) \le e^{-c't^2}$$

provided that $t \ge C(\sqrt{d} + \sqrt{m})$. Let

$$t = t_m = \tau \cdot (\sqrt{d} + \sqrt{m}\sqrt{\log \frac{en}{|J|}}),$$

with $\tau > C$ to be chosen later, and set $C_1 = 4\tau$. Taking the union bound, we get

$$\begin{aligned}\mathbb{P}\left(\Delta_1 \notin \mathcal{W}_1\right) &\le \sum_{m=1}^{n}\sum_{|J|=m} \mathbb{P}\left(\|\Delta_1|_J\| > 4t_m\right) \le \sum_{m=1}^{n}\binom{n}{m} e^{-c't_m^2} \\ &\le \sum_{m=1}^{n} \exp\left[-c'\tau^2 \cdot \left(\sqrt{d} + \sqrt{m}\sqrt{\log\frac{en}{m}}\right)^2 + m\log\frac{en}{m}\right].\end{aligned}$$

We can choose the constant τ so that the last expression doesn't exceed e^{-d}.

Step 2. Let $k > 1$, and assume that $C_1, \dots, C_{k-1}$ are already defined. It is enough to find $C_k > 0$ such that for any $U \in \mathcal{W}_{k-1}$ with $|u_{i,j}| \le 1$ for all i, j

$$\mathbb{P}\left(U \otimes_r \Delta_k \notin \mathcal{W}_k\right) \le e^{-cd}. \tag{5.5}$$

Indeed, in this case

$$\mathbb{P}\left(\Delta_1 \otimes_r \dots \otimes_r \Delta_k \notin \mathcal{W}_k \mid \Delta_1 \otimes_r \dots \otimes_r \Delta_{k-1} \in \mathcal{W}_{k-1}\right) \le e^{-cd}.$$

Hence, the induction hypothesis yields

$$\begin{aligned}
&\mathbb{P}\left(\Delta_1 \otimes_r \dots \otimes_r \Delta_k \notin \mathcal{W}_k\right) \\
&\le \mathbb{P}\left(\Delta_1 \otimes_r \dots \otimes_r \Delta_k \notin \mathcal{W}_k \mid \Delta_1 \otimes_r \dots \otimes_r \Delta_{k-1} \in \mathcal{W}_{k-1}\right) \\
&\quad + \mathbb{P}\left(\Delta_1 \otimes_r \dots \otimes_r \Delta_{k-1} \notin \mathcal{W}_{k-1}\right) \\
&\le k e^{-cd}.
\end{aligned}$$

Fix $U \in \mathcal{W}_{k-1}$. To shorten the notation denote $W = U \otimes_r \Delta_k$. For $j \in \mathbb{N}$ define m_j as the smallest number m satisfying

$$d^j \le m \log^j \left(\frac{en}{m}\right).$$

Our strategy of proving (5.5) will depend on the cardinality of the set $J \subset \{1, \dots, n\}$ appearing in (5.4).

Consider first any set J such that $|J| \le m_{k-1}$. By Lemma 5.6,

$$\begin{aligned}
\|W|_J\| &\le \sqrt{d}\, \|U|_J\| \le \sqrt{d} \cdot C_{k-1}(d^{(k-1)/2} + \sqrt{|J|} \log^{(k-1)/2}(en/|J|)) \\
&\le 2C_{k-1} d^{k/2},
\end{aligned}$$

and so W satisfies the condition $\mathcal{W}_K$ with $C_k = 2C_{k-1}$ for all such J.

Now consider all sets J such that $m_{k-1} < |J| < m_k$. The previous argument shows that any vector $y \in S^{n-1}$ with $|\mathrm{supp}(y)| \le m_{k-1}$ satisfies $\|Wy\| \le 2C_{k-1} d^{k/2}$. Any $x \in S^{n-1}$ can be decomposed as $x = y + z$, where $|\mathrm{supp}(y)| \le m_{k-1}$ and $\|z\|_\infty \le m_{k-1}^{-1/2}$. Therefore, to prove (5.5), it is enough to show that

$$\begin{aligned}
\mathbb{P}\ (&\exists J \subset \{1, \dots, n\}\ m_{k-1} < |J| \le m_k \ \text{ and} \\
&\left\| W|_J : B_2^n \cap m_{k-1}^{-1/2} B_\infty^n \to B_2^{d^k} \right\| > C d^k \Big)) \le e^{-cd}.
\end{aligned}$$

To this end take any $z \in S^{n-1}$ such that $|\mathrm{supp}(z)| \le m_k$ and $\|z\|_\infty \le m_{k-1}^{-1/2}$. We will obtain a uniform bound on $\|Wz\|_2$ over all such z, and use the ε-net argument to derive a bound for $\|W|_J\|$ from it.

Let M be the minimal natural number such that $4^M m_{k-1} \geq m_k$. Let $I_0, \dots, I_M$ be blocks of type m_{k-1} of the coordinates of z. Since $U \in \mathcal{W}_{k-1}$, for any $m \leq M$

$$\|U|_{I_m}\|^2 \leq C_{k-1}(d^{k-1} + |I_m| \log^{k-1}(en/|I_m|)) \leq 2C_{k-1}|I_m| \log^{k-1}(en),$$

because $|I_m| \geq m_{k-1}$.

Let $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ be a row of the matrix Δ_k. Then the coordinates of the vector Wz corresponding to this row form the vector $(U \otimes_r \varepsilon)z = (U \otimes_r z^T)\varepsilon^T$. Let U' be the $d^{k-1} \times |J|$ matrix defined as

$$U' = (U \otimes_r z^T)|_J.$$

The inequality above and Lemma 3.1 imply

$$\begin{aligned} \|U'\|^2 \leq \sum_{m=0}^{M} \|U|_{I_m}\|^2 \cdot \|z|_{I_m}\|_\infty^2 &\leq 2C_{k-1} \log^{k-1}(en) \sum_{m=0}^{M} |I_m| \cdot \|z|_{I_m}\|_\infty^2 \\ &\leq 10C_{k-1} \log^{(k-1)/2}(en) \end{aligned}$$

Also, since all entries of U have absolute value at most 1,

$$\|U'\|_{HS}^2 \leq d^{k-1}.$$

The sequence of coordinates of the vector Wz consists of d independent copies of $U'\varepsilon_I^T$. Therefore, applying Lemma 5.8 with $l = d$ and $s = td^k$, we get

$$\begin{aligned} p(x) := \mathbb{P}\left(\|Wz\|^2 \geq (4+t) \cdot d^k\right) &\leq 2^d \exp\left(-\frac{ctd^k}{\|U'\|^2}\right) \\ &\leq 2^d \exp\left(-\frac{td^k}{c_k' \log^{k-1}(en)}\right), \end{aligned}$$

where $c_k' = 4C_{k-1}^2/c$. By the volumetric estimate, we can construct a $(1/2)$-net $\mathcal{N}$ for the set

$$E_k := \{z \in S^{n-1} \mid |\mathrm{supp}(z)| \leq m_k, \ \|z\|_\infty \leq m_{k-1}^{-1/2}\}$$

in the Euclidean metric, such that

$$|\mathcal{N}| \leq \binom{n}{m_k} 6^{m_k} \leq \exp\left(2m_k \log(en)\right).$$

Since $n \geq d^{k+1/2}$, and $m_k \leq d^k$, we have $\log(en) \leq 2k \log(en/m_k)$, and so

$$m_k \log(en) \leq (2k)^k \frac{d^k}{\log^{k-1}(en)}.$$

Hence, we can chose the constant $t = t_k$ large enough, so that

$$\mathbb{P}\left(\exists z \in \mathcal{N} \mid \|Wz\|^2 \geq C'_k d^k\right) \leq |\mathcal{N}| \cdot 2^d \exp\left(-\frac{t_k d^k}{c'_k \log^{k-1}(en)}\right)$$
$$\leq \exp\left(-\frac{d^k}{\log^{k-1}(en)}\right)$$

with the constant $C'_k = 4 + t_k$. Thus,

$$\mathbb{P}\left(\exists z \in E_k \mid \|Wz\|^2 \geq 4C'_k d^k\right) \leq \exp\left(-\frac{d^k}{\log^{k-1}(en)}\right),$$

which implies condition (5.4) with $C_k = (4C_{k-1}^2 + 4C'_k)^{1/2}$ for all sets J such that $|J| < m_k$.

Finally, consider any set J with $|J| \geq m_k$. As in the previous case, we can split any vector $x \in S^{n-1}$ as $x = y + z$, where $|\mathrm{supp}(y)| \leq m_k$ and $\|z\|_\infty \leq m_k^{-1/2}$. The previous argument shows that with probability greater than $1 - \exp\big(-d^k/\log^{k-1}(en)\big)$,

$$\|Wy\| \leq (4C_{k-1}^2 + 4C'_k)^{1/2} d^k$$

for all such y. Therefore, it is enough to estimate $\max \|W|_J z\|$ over $z \in B_2^n \cap m_k^{-1/2} B_\infty^n$. A (1/2)-net in the set $B_2^n \cap m_k^{-1/2} B_\infty^n$ is too big, so following the argument used in the previous case would lead to the losses that break down the proof. Instead, we will use the sets $\mathcal{M}_l$ defined in Lemma 5.7 and obtain the bounds for $\max \|W|_J z\|$ for each set separately.

To this end, set $b = 1/\sqrt{m_k}$, and let L_0 be the largest number such that $2^{-L_0} \geq b$. Let $l \geq L_0$ and take any $x \in \mathcal{M}_l$. Choose any set $I \supset \mathrm{supp}(x)$ such that $|I| = 4^l$. As in the previous case, let U' be the $d^{k-1} \times 4^l$ matrix defined as

$$U' = (U \otimes_r x^T)|_I.$$

Since all non-zero coordinates of x have absolute value $2^{-l} = 1/\sqrt{|I|}$, the assumption $U \in \mathcal{W}_{k-1}$ implies

$$\|U'\| \leq \frac{1}{\sqrt{|I|}} \|U|_I\| \leq \frac{C_{k-1}}{\sqrt{|I|}} \left(d^{(k-1)/2} + \sqrt{|I|} \cdot \log^{(k-1)/2}\left(\frac{en}{|I|}\right)\right)$$
$$\leq 2C_{k-1} \log^{(k-1)/2}\left(en \cdot 4^{-l}\right)$$

The last inequality holds since for any $m \geq m_k \geq m_{k-1}$

$$d^{(k-1)/2} \leq \sqrt{m} \log^{(k-1)/2}\left(\frac{en}{m}\right).$$

Also, as before, all entries of U have absolute value at most 1, so $\|U'\|_{HS}^2 \leq d^{k-1}$. The sequence of coordinates of the vector Wx consists of d independent copies of $U'\varepsilon_I^T$. Therefore, applying Lemma 5.8, we get

$$\begin{aligned}\mathbb{P}\left(\|Wx\|^2 \geq 4d\cdot d^{k-1}+s\right) &\leq 2^d \exp\left(-\frac{cs}{\|U'\|^2}\right)\\ &\leq 2^d \exp\left(-\frac{s}{c_k' \log^{k-1}\left(en\cdot 4^{-l}\right)}\right)\end{aligned}$$

where $c_k' = 4C_{k-1}^2/c$. Set

$$s = s(l) = 2c_k' \cdot 4^l \log^k\left(en\cdot 4^{-l}\right).$$

Then $s(l) \geq 2c_k' m_k \log^k(en/m_k) \geq 2c_k' d^k$, so the previous inequality can be rewritten as

$$\mathbb{P}\left(\|Wx\|^2 \geq c_k'' s(l)\right) \leq \exp\left(-2\cdot 4^l \log\left(en\cdot 4^{-l}\right)\right).$$

Hence, the union bound implies that there exists a constant C_k satisfying

$$\begin{aligned}&\mathbb{P}\left(\exists l \geq L_0\ \exists x \in \mathcal{M}_l\ \ \|Wx\| > C_k s(l)\right)\\ &\leq \sum_{l=L_0}^{\infty} \binom{n}{4^l} \cdot 3^{4^l} \exp\left(-2\cdot 4^l \log\left(en\cdot 4^{-l}\right)\right) \leq \exp\left(-4^{L_0}\log\left(en\cdot 4^{-L_0}\right)\right)\\ &\leq \exp(-d^k).\end{aligned}$$

Define the event Ω_1 by

$$\Omega_1 = \{\forall l \geq L_0\ \forall x \in \mathcal{M}_l\ \ \ \|Wx\| \leq s(l)\}.$$

The previous inequality means that $\mathbb{P}\left(\Omega_1^c\right) \leq \exp(-d^k)$.

Assume that the event Ω_1 occurs. Let $J \subset \{1,\dots,n\}$ be such that $|J| \geq m_k$, and choose L' so that $4^{L'-1} < |J| \leq 4^{L'}$. Applying Lemma 5.7 to $T = W|_J$ and $b = 1/\sqrt{m_k}$, we obtain

$$\begin{aligned}\left\|W|_J : B_2^n \cap m_k^{-1/2} B_\infty^n \to B_2^{d^k}\right\|^2 &\leq 5c_k'' \sum_{l=L_0}^{L'} s(l) \leq C_k'' 4^{L'} \log^k\left(\frac{en}{4^{L'}}\right)\\ &\leq 4C_k''|J| \log^k\left(\frac{en}{|J|}\right).\end{aligned}$$

This shows that condition (5.4) holds with $C_k = (4C_{k-1}^2 + 4C_k' + 4C_k'')^{1/2}$ for all non-empty sets $J \subset \{1,\dots,n\}$. This completes the induction step and the proof of Lemma 5.9.

$\square$

5.3. **Lower bounds for the Q-norm.** To obtain bounds for the Levy concentration function below, we need a lower estimate for a certain norm of the row product of random matrices.

Definition 5.10. Let $U = (u_{j,k})$ be an $M \times m$ matrix. Denote

$$\|U\|_Q = \sum_{j=1}^{M} \left(\sum_{k=1}^{m} u_{j,k}^2 \right)^{1/2}.$$

In other words, $\|\cdot\|_Q$ is the norm in the Banach space $\ell_1^M(\ell_2^m)$.

If U is an $M \times m$ matrix with independent centered entries of unit variance, then for any $x \in \mathbb{R}^n$,

$$\mathbb{E} \left\| U \otimes_r x^T \right\|_Q \le \sum_{j=1}^{M} \left(\mathbb{E} \sum_{k=1}^{m} u_{j,k}^2 x^2(k) \right)^{1/2} = M \|x\|_2 .$$

Moreover, if the coordinates of x are commensurate, we can expect that a reverse inequality would follow from the Central Limit theorem. This observation leads to the following definition.

Let $\mathcal{V}_L$ be the set of $d^L \times n$ matrices A such that for any $x \in \mathbb{R}^n$

$$\left\| A \otimes_r x^T \right\|_Q \ge \tilde{c} d^L \|x\|_2 . \tag{5.6}$$

We will show below that the row product of L independent $d \times n$ random matrices belongs to $\mathcal{V}_L$ with high probability, provided that the constant $\tilde{c}$ in (5.6) is appropriately chosen. To this end, consider the behavior of $\left\| (\Delta_1 \otimes_r \dots \Delta_L) \otimes_r x^T \right\|_Q$ for a fixed vector $x \in \mathbb{R}^n$.

Lemma 5.11. *Let $\Delta_1, \dots, \Delta_L$ be $d \times m$ random matrices with independent δ random entries. Then for any $x \in \mathbb{R}^m$*

$$\mathbb{P} \left(\left\| (\Delta_1 \otimes_r \dots \Delta_L) \otimes_r x^T \right\|_Q \le c d^L \|x\|_2 \right) \le \exp \left(- \frac{c d^L \|x\|_2^2}{\|x\|_\infty^2} \right).$$

Proof. Without loss of generality, assume that $\|x\|_\infty = 1$, so $\|x\|_2 \ge 1$. Let $\alpha > 0$, and let $\nu_1, \dots, \nu_m \in [0, 1]$ be independent random variables satisfying $\mathbb{E}\nu_j \ge \alpha$ for all $j = 1, \dots, m$. The standard symmetrization and Bernstein's inequality [23] yield

$$\mathbb{P} \left(\left| \sum_{j=1}^{m} x^2(j) \nu_j - \mathbb{E} \sum_{j=1}^{m} x^2(j) \nu_j \right| > t \right) \le 2 \exp \left(- \frac{t^2}{2(\sum_{j=1}^{m} x^4(j) + t/3)} \right).$$

Setting $t = (\alpha/2) \|x\|_2^2$, and using $\|x\|_\infty \le 1$, we get

$$\mathbb{P} \left(\sum_{j=1}^{m} x^2(j) \nu_j < \frac{\alpha}{2} \|x\|_2^2 \right) \le 2 \exp \left(- \frac{\alpha^2}{16} \|x\|_2^2 \right).$$

Applying the previous inequality to the random variable $Y_i,\ i = 1, \dots, d^L$, which is the ℓ_2-norm of a row of the matrix $(\Delta_1 \otimes_r \dots \Delta_L) \otimes_r x^T$, we obtain $\mathbb{P}(Y_i < c\|x\|_2) \le 2\exp(-c'\|x\|_2^2)$. Let $0 < \theta < 1$. If

$$\left\|(\Delta_1 \otimes_r \dots \Delta_L) \otimes_r x^T\right\|_Q = \sum_{i=1}^{d^L} Y_i \le \theta \cdot d^L \|x\|_2,$$

then $Y_i < c\|x\|_2$ for at least $(1-\theta)d^L$ numbers i. Hence,

$$\begin{aligned}
&\mathbb{P}\left(\|(\Delta_1 \otimes_r \dots \Delta_L) \otimes_r x\|_Q \le \theta c d^L \|x\|_2\right) \\
&\le \binom{d^L}{\lfloor (1-\theta)d^L \rfloor} \exp(-c(1-\theta)d^L\|x\|_2^2) \\
&\le \exp\left(-d^L\left(c(1-\theta)\|x\|_2^2 - \theta \log\frac{e}{\theta}\right)\right) \le \exp(-(c/2)d^L\|x\|_2^2),
\end{aligned}$$

if θ is small enough. □

We will use Lemma 5.11 to show that the row product $\Delta_1 \otimes_r \dots \otimes_r \Delta_{K-1}$ satisfies condition (5.6) with high probability.

Lemma 5.12. *There exists a constant $\tilde{c} > 0$ for which the following holds. Let $K > 1$, and let $n \le d^K$. For $d \times n$ matrices $\Delta_1, \dots, \Delta_{K-1}$ be matrices with independent δ random entries*

$$\mathbb{P}(\Delta_1 \otimes_r \dots \otimes_r \Delta_{K-1} \notin \mathcal{V}_{K-1}) \le \exp(-cd^{K-1}).$$

Proof. Denote for shortness $\bar{\Delta} = \Delta_1 \otimes_r \dots \otimes_r \Delta_{K-1}$. To conclude that $\bar{\Delta} \in \mathcal{V}_{K-1}$, it is enough to show that condition (5.6) holds for any $x \in S^{n-1}$.

For $x \in S^{n-1}$ denote by $\Omega(x)$ the set of matrices A such that $\|A \otimes_r x\|_Q \le cd^{K-1}$. For $L = K-1$ Lemma 5.11 yields

$$\mathbb{P}(\bar{\Delta} \in \Omega^c(x)) \le \exp\left(-\frac{cd^{K-1}}{2\|x\|_\infty^2}\right). \tag{5.7}$$

As the first step in proving the lemma, we will show that for $A = \bar{\Delta}$ condition (5.6) holds for all x from some subset of the sphere. More precisely, we will prove the following claim.

Claim. Let $a > 0$ and $m \le n$. Denote

$$S(a,m) = \{x \in S^{n-1} \mid \|x\|_\infty \le a,\ |\mathrm{supp}(x)| \le m\}.$$

If $a^2 m \log d < Cd^{K-1}$, then

$$\mathbb{P}(\bar{\Delta} \notin \bigcap_{x \in S(a,m)} \Omega(x)) \le \exp\left(-\frac{c'd^{K-1}}{a^2}\right).$$

It is enough to prove the claim for $0 < a \leq 1$. Note that if $0 \leq |y(j)| \leq |x(j)|$ for any $j = 1, \dots, k$, then $\left\| \bar{\Delta} \otimes_r y^T \right\|_Q \leq \left\| \bar{\Delta} \otimes_r x^T \right\|_Q$. Hence, to prove the claim, it is enough to construct a set $\mathcal{N}$ of vectors $y \in B_2^n \setminus (1/2) B_2^n$ such that for any $x \in S(a,m)$ there is $y \in \mathcal{N}$ with $|y(j)| \leq |x(j)|$ for all j and

$$\mathbb{P}\,(\bar{\Delta} \notin \bigcap_{y \in \mathcal{N}} \Omega(y)) \leq \exp\left(-\frac{c' d^{K-1}}{a^2}\right).$$

Set

$$\mathcal{N} = \left\{ y \in \left(\frac{1}{2\sqrt{m}}\right) \mathbb{Z}^n \mid |\mathrm{supp}(y)| \leq m, \ \|y\|_\infty \leq a \text{ and } \frac{1}{2} \leq \|y\|_2 \leq 1 \right\}.$$

By the volumetric considerations

$$|\mathcal{N}| \leq \binom{n}{m} C^m \leq \exp(cm \log n) \leq \exp(C' m \log d),$$

since $n \leq d^K$. For $x \in S(a,m)$ consider the vector y with coordinates $y(j) = (1/2\sqrt{m}) \cdot \lfloor 2\sqrt{m} |x(j)| \rfloor$. Then $|y(j)| \leq |x(j)|$, and $\|y\|_2 \geq 1 - \|x - y\|_2 \geq 1/2$, so $y \in \mathcal{N}$. By the union bound and (5.7),

$$\mathbb{P}\,(\bar{\Delta} \notin \bigcap_{y \in \mathcal{N}} \Omega(y)) \leq |\mathcal{N}| \exp\left(-\frac{c d^{K-1}}{2a^2}\right).$$

The claim now follows from the assumption $a^2 m \log d \leq C d^{K-1}$ for a suitable constant C.

The lemma can be easily derived from the claim. For a and m as above denote $\Omega(a,m) = \bigcap_{x \in S(a,m)} \Omega(x)$. Set

$$a_i = 3 d^{(1-i)K/6}, \quad m_i = \min\left(d^{iK/3}, n\right), \quad i = 1,2,3.$$

Then $m_3 = n$, and the condition $a_i^2 m_i \log d \leq C d^{K-1}$, $i = 1,2,3$ is satisfied. Set

$$\mathcal{V} = \bigcap_{i=1}^{3} \Omega(a_i, m_i).$$

By the claim, $\mathbb{P}\,(\mathcal{V}^c) \leq \exp(-c d^{K-1})$.

Assume now that $\bar{\Delta} \in \mathcal{V}$. Using the non-increasing rearrangement of $|x(j)|$, we can decompose any $x \in S^{n-1}$ as $x = x_1 + x_2 + x_3$,where x_1, x_2, x_3 have disjoint supports, $|\mathrm{supp}(x_i)| \leq m_i$, $\|x_i\|_\infty \leq a_i/3$. By the triangle inequality, $\|x_i\|_2 \geq 1/3$ for some i. Thus,

$$\left\| \bar{\Delta} \otimes_r x^T \right\|_Q \geq \left\| \bar{\Delta} \otimes_r x_i^T \right\|_Q \geq \left\| \bar{\Delta} \otimes_r \frac{x_i^T}{\|x_i\|_2} \right\|_Q \cdot \frac{1}{3} \geq \frac{c}{3} d^{K-1},$$

since $x_i / \|x_i\|_2 \in S(a_i, m_i)$. This proves the Lemma with $\tilde{c} = c/3$. $\square$

6. Bounds for the Levy concentration function

Definition 6.1. Let $\rho > 0$. Define the Levy concentration function of a random vector $X \in \mathbb{R}^n$ by

$$\mathcal{L}_1(X, \rho) = \sup_{x \in \mathbb{R}^n} \mathbb{P}\left(\|X - x\|_1 \leq \rho\right).$$

Unlike the standard definition of the Levy concentration function, we use the ℓ_1-norm instead of the ℓ_2-norm. We need the following standard

Lemma 6.2. *Let $X \in \mathbb{R}^n$ be a random vector, and let X' be an independent copy of X. Then for any $\rho > 0$*

$$\mathcal{L}_1(X, \rho) \leq \mathbb{P}^{1/2}(\|X - X'\|_1 \leq 2\rho).$$

Proof. Let $y \in \mathbb{R}^n$ be any vector. Then

$$\begin{aligned} \mathbb{P}^2(\|X - y\|_1 \leq \rho) &= \mathbb{P}\left(\|X - y\|_1 \leq \rho \text{ and } \|X' - y\|_1 \leq \rho\right) \\ &\leq \mathbb{P}\left(\|X - X'\|_1 \leq 2\rho\right). \end{aligned}$$

Taking the supremum over $y \in \mathbb{R}^n$ proves the Lemma. □

In the next lemma, we bound the Levy concentration function using Talagrand's inequality, in the same way it was done in the proof of Lemma 5.8.

Lemma 6.3. *Let $U = (u_{i,j})$ be any $N \times n$ matrix, and let $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ be a vector with independent δ random coordinates. Then for any $x \in \mathbb{R}^n$*

$$\mathcal{L}_1\left((U \otimes_r \varepsilon^T)x, c\left\|U \otimes_r x^T\right\|_Q\right) \leq 2\exp\left(-c' \frac{\left\|U \otimes_r x^T\right\|_Q^2}{N\left\|U \otimes_r x^T\right\|^2}\right). \tag{6.1}$$

Proof. Note that $(U \otimes_r \varepsilon^T)x = (U \otimes_r x^T)\varepsilon$. Let $\varepsilon'_1, \dots, \varepsilon'_n$ be independent copies of $\varepsilon_1, \dots, \varepsilon_n$. Applying Lemma 6.2, we obtain for any $\rho > 0$

$$\mathcal{L}_1\left((U \otimes_r x^T)\varepsilon, \rho\right) \leq \mathbb{P}^{1/2}\left(\left\|(U \otimes_r x^T)(\varepsilon - \varepsilon')\right\|_1 \leq 2\rho\right). \tag{6.2}$$

Consider a function $F : \mathbb{R}^n \to \mathbb{R}$, defined by

$$F(y) = \left\|(U \otimes_r x^T)y\right\|_1,$$

where $y \in \mathbb{R}^n$. Then F is a convex function with the Lipschitz constant $L \leq \left\|U \otimes_r x^T : B_2^n \to B_1^N\right\| \leq \sqrt{N}\left\|U \otimes_r x^T\right\|$.

By Talagrand's measure concentration theorem

$$\mathbb{P}\left(|F(\varepsilon - \varepsilon') - \mathbb{M}(F)| > s\right) \leq 4\exp\left(-\frac{cs^2}{L^2}\right),$$

where $\mathbb{M}(F)$ is a median of F, considered as a function on $\mathbb{R}^n$ equipped with the probability measure defined by the vector $\varepsilon - \varepsilon'$. This tail estimate implies

$$|\mathbb{M}(F) - \mathbb{E}F| \le c_1 L \le c_1 \sqrt{N} \left\| U \otimes_r x^T \right\|.$$

By Lemma 2.6 [15] we have

$$\begin{aligned}\mathbb{E}F = \mathbb{E} \sum_{i=1}^N \left| \sum_{j=1}^n u_{i,j} x(j) \cdot (\varepsilon(j) - \varepsilon'(j)) \right| &\ge c_2 \sum_{i=1}^N \left(\sum_{j=1}^n u_{i,j}^2 x^2(j) \right)^{1/2} \\ &= c_2 \left\| (U \otimes_r x^T) \right\|_Q .\end{aligned}$$

Note that if the constant c' in the formulation of the lemma is chosen small enough, we may assume that $2c_1\sqrt{N} \left\| U \otimes_r x^T \right\| \le c_2 \left\| (U \otimes_r x^T) \right\|_Q$. Indeed, if this inequality does not hold, the right-hand side of (6.1) would be greater than 1. Combining the previous estimates yields $\mathbb{M}(F) \ge (c_2/2) \left\| (U \otimes_r x^T) \right\|_Q$. Hence,

$$\begin{aligned}&\mathbb{P} \left(\left\| (U \otimes_r x^T)(\varepsilon - \varepsilon') \right\|_1 \le \frac{c_2}{4} \left\| U \otimes_r x^T \right\|_Q \right) \\ &\le \mathbb{P} \left(|F(\varepsilon - \varepsilon') - \mathbb{M}(F)| \ge \frac{1}{4} \mathbb{M}(F) \right) \\ &\le 4 \exp \left(- \frac{c \mathbb{M}^2(F)}{L^2} \right) \le 4 \exp \left(-c' \frac{\left\| U \otimes_r x^T \right\|_Q^2}{N \left\| U \otimes_r x^T \right\|^2} \right).\end{aligned}$$

This inequality and (6.2), applied with $\rho = \frac{c_2}{8} \left\| (U \otimes_r x^T) \right\|_Q$, finish the proof. □

For the next result we need the following standard Lemma.

Lemma 6.4. *Let $s_1, \dots, s_d$ be independent non-negative random variables such that $\mathbb{P}(s_j \le R) \le p$ for all j. Then*

$$\mathbb{P} \left(\sum_{j=1}^d s_j \le \frac{1}{2} R d \right) \le (4p)^{d/2}.$$

Proof. If $\sum_{j=1}^d s_j \le \frac{1}{2} R d$, then $s_j \le R$ for at least $d/2$ numbers j. □

Combining Lemma 6.3 with this inequality, we obtain the tensorized version of Lemma 6.3.

Corollary 6.5. *Let $U = (u_{i,j})$ be any $N \times n$ matrix, and let V be a $d \times n$ matrix with independent δ random coordinates. Then for any*

$x \in \mathbb{R}^n$
(6.3)

$$\mathcal{L}_1\left((U \otimes_r V)x, cd \|U \otimes_r x\|_Q\right) \le C2^d \exp\left(-c' \frac{d \left\|U \otimes_r x^T\right\|_Q^2}{N \left\|U \otimes_r x^T\right\|^2}\right).$$

Proof. The coordinates of the vector $(U \otimes_r V)x \in \mathbb{R}^{Nd}$ consist of d independent blocks $(U \otimes_r \varepsilon_1)x, \dots, (U \otimes_r \varepsilon_d)x$, where $\varepsilon_1, \dots, \varepsilon_d$ are the rows of V. The corollary follows from Lemma 6.4, applied to the random variables $s_j = \|(U \otimes_r \varepsilon_j)x - y_j\|_1$, where $y_1, \dots, y_d \in \mathbb{R}^N$ are any fixed vectors. □

To prove Theorem 1.6 we have to bound the probability that the matrix $\Delta_1 \otimes_r \dots \otimes_r \Delta_K$ maps some vector from the unit sphere into a small ℓ_1 ball. Before doing that, we consider an easier problem of estimating the probability that this matrix maps a *fixed* vector into a small ℓ_1 ball. We phrase this estimate in terms of the Levy concentration function.

Lemma 6.6. *Let $U \in \mathcal{W}_{K-1} \cap \mathcal{V}_{K-1}$ be a $d^{K-1} \times n$ matrix, and let Δ_K be a $d \times n$ random matrix with independent δ random entries. For any $x \in \mathbb{R}^n$*

$$\begin{aligned}&\mathcal{L}_1\left((U \otimes_r \Delta_K)x, \tilde{c} d^K \|x\|_2\right)\\ &\le \exp\left(-\frac{c'' d \|x\|_2^2}{\|x\|_\infty^2}\right) + \exp\left(-\frac{c'' d^K}{\log^{K-1}\left(\frac{en\|x\|_\infty^2}{\|x\|_2^2}\right)}\right).\end{aligned}$$

Proof. To use Corollary 6.5, we have to estimate the Q-norm and the operator norms of $U \otimes_r x^T$. The estimate of the Q-norm is given by (5.6).

To estimate the operator norm, assume that $\|x\|_2 = 1$, and set $s = \left\lfloor \|x\|_\infty^{-2} \right\rfloor$. Let L be the maximal number l such that $2^l s \le n$, and let $I_0, \dots, I_L$ be the blocks of coordinates of x of type s. Then $\|x|_{J_l}\|_\infty \le 2^{-l} \|x\|_\infty$, and by Lemma 3.1

$$\sum_{j=0}^{L} |J_l| \cdot \|x|_{J_l}\|_\infty^2 \le 5.$$

Let $y \in \mathbb{R}^n$. By Cauchy–Schwartz inequality, we have

$$\begin{aligned}\left\|(U \otimes_r x^T)y\right\|_2^2 &= \left\|\sum_{l=0}^{L}(U|_{J_l} \otimes_r x^T|_{J_l})y|_{J_l}\right\|_2^2 \\ &\le \left(\sum_{l=0}^{L}\left\|U|_{J_l} \otimes_r x^T|_{J_l}\right\|_2^2\right)\cdot\left(\sum_{l=0}^{L}\|y|_{J_l}\|_2^2\right) \\ &\le \left(\sum_{l=0}^{L}\left\|U|_{J_l} \otimes_r x^T|_{J_l}\right\|_2^2\right)\cdot\|y\|_2^2,\end{aligned}$$

which means

$$\left\|U \otimes_r x^T\right\|^2 \le \sum_{l=0}^{L}\left\|U|_{J_l} \otimes_r x^T|_{J_l}\right\|^2 \le \sum_{l=0}^{L}\|U|_{J_l}\|^2 \cdot \|x|_{J_l}\|_\infty^2 .$$

Since $U \in \mathcal{W}_{K-1}$, and $|J_l| \ge |J_1| = s$ for all $l \le L$, the previous inequality implies

$$\begin{aligned}\left\|U \otimes_r x^T\right\|^2 &\le C\sum_{l=0}^{L}\left(d^{K-1} + |J_l|\cdot\log^{K-1}\left(\frac{en}{|J_l|}\right)\right)\cdot\|x|_{J_l}\|_\infty^2 \\ &\le C\left(d^{K-1}\|x\|_\infty^2 + \log^{K-1}\left(\frac{en}{s}\right)\right).\end{aligned}$$

Therefore, by Corollary 6.5 and condition (5.6),

$$\begin{aligned}&\mathcal{L}_1\left((U \otimes_r \Delta_K)x, cd^K\right) \\ &\le \exp\left(-\frac{Cd^K}{d^{K-1}\|x\|_\infty^2 + \log^{K-1}\left(\frac{en}{s}\right)}\right).\end{aligned}$$

The lemma follows from an elementary inequality $\exp(-\frac{a}{b+c}) \le \exp(-\frac{a}{2b})+\exp(-\frac{a}{2c})$. □

7. Lower bounds via the chaining argument

To get a global bound for the Levy concentration function using the bounds for each fixed vector, we prove a chaining-type estimate. Chaining argument is one of the main approaches to obtaining bounds for the *supremum* of a random process [21]. Let $\{X_t \mid t \in T\}$ be a random process indexed by a set T. The chaining method is based on representing X_t as a sum of increments and proving an upper estimate for each increment separately, and combining these estimates using the union bound.

A similar approach, based on passing from a random variable to increments can be applied to estimating the *infimum* of a random process as well. In this case we isolate one "big" increment, whose position in the chain depends on t. The rest of the increments is divided in two groups. In one group the increments are small, and we can bound their absolute values above, and use the triangle inequality. The increments from the other group may be big, but they belong to a small set of random variables. In such situation, we can condition on these increments, and obtain a lower bound on the conditional probability using the Levy concentration function of the "big" increment. Then we sum up these conditional probabilities over the small set. As usual for the chaining method, this step requires a balance between the estimate of the Levy concentration function, and the size of the set.

Lemma 7.1. *Let $R > 0$, $\alpha \in (0, 1/2)$ and let $\{l_j\}_{j=0}^L$ be a sequence of natural numbers such that $l_0 = 1$ and $l_{j+1} \ge 2l_j$ for all $j = 0, \dots, L$. Set $n = \sum_{j=1}^L l_j$. Let $A : \mathbb{R}^n \to \mathbb{R}^N$ be a random matrix with independent columns. Assume that for any $j = 1, \dots, L$ there exists $p_j > 0$ such that for any $x \in S^{n-1}$ with $|\mathrm{supp}(x)| \le l_j$, $\|x\|_\infty \le l_{j-1}^{-1/2}$*

$$\mathcal{L}_1(Ax, R) \le p_j \le \left(\frac{6en}{l_j \alpha^j}\right)^{-8l_j}. \tag{7.1}$$

Then for any $y \in \mathbb{R}^N$

$$\mathbb{P}\left(\exists x \in S^{n-1} \ \ \|Ax - y\|_1 \le \frac{\alpha^{L-1}R}{4}\right) \le p_1^{1/2} + \mathbb{P}\left(\|A\| > \frac{R}{8\alpha\sqrt{N}}\right).$$

Proof. Denote $\|A\|_{2\to 1} = \left\|A : B_2^n \to B_1^N\right\|$. Let $j \in \{1, \dots, L\}$ and let J be a l_j-element subset of $\{1, \dots, L\}$. Denote

$$S_J = \{x \in S^{n-1} \mid |\mathrm{supp}(x) \subset J, \ \|x\|_\infty \le l_{j-1}^{-1/2}\}.$$

Set $m_j = \sum_{i=1}^{j-1} l_i$. Since the sequence $\{l_j\}_{j=1}^L$ increases exponentially, $m_j \le l_j$. We will need the following

Claim. Let $y \in \mathbb{R}^N$. Let

$$Q_J = \{w \mid, \|w\|_2 \le 2\alpha^{1-j}, \ \mathrm{supp}(w) \cap J = \emptyset, \ |\mathrm{supp}(w)| \le m_j\}.$$

Then

$$\mathbb{P}\left(\exists z \in Q_J + S_J \ \ \|Az - y\|_1 \le R - \alpha \|A\|_{2\to 1}\right) \le p_j^{3/4}.$$

By the volumetric estimate we can choose an $(\alpha/2)$-net $\mathcal{M}_J$ in S_J such that $|\mathcal{M}_J| \le (6/\alpha)^{l_j}$.

Take any $x \in \mathcal{M}_J$ and $w \in Q_J$. Denote $y' = y - Aw$. Then the vectors Ax and y' are independent. Conditioning on the columns of A with indexes from $\mathrm{supp}(w)$, and using (7.1), we get

$$\mathbb{P}\left(\|A(w + x) - y\|_1 < R \mid A|_{J^c}\right) \le \mathbb{P}\left(\|Ax - y'\|_1 < R \mid A|_{J^c}\right) \le p_j.$$

Taking the expectation with respect to $A|_{J^c}$ yields

$$\mathbb{P}\left(\|A(w+x)-y\|_1 < R\right) \le p_j.$$

The volumetric estimate guarantees the existence of a $(\alpha/2)$-net $\mathcal{N}_J$ in Q_J such that

$$|\mathcal{N}_J| \le \binom{n}{m_j}\left(6\alpha^{-j}\right)^{m_j} \le \left(\frac{6en}{\alpha^j m_j}\right)^{m_j}.$$

Since $m_j \le l_j$, the last quantity does not exceed $\left(\frac{6en}{\alpha^j l_j}\right)^{l_j}$. By the union bound and assumption (7.1),

$$\begin{aligned}&\mathbb{P}\left(\exists x \in \mathcal{M}_J\ \exists w \in \mathcal{N}_J\ \ \|A(w+x)-y\|_1 < R\right)\\ &\le |\mathcal{N}_J|\cdot|\mathcal{M}_J|\cdot p_j \le \left(\frac{6en}{\alpha^j l_j}\right)^{l_j}\cdot\left(\frac{6}{\alpha}\right)^{l_j}\cdot p_j \le p_j^{3/4}.\end{aligned}$$

Assume that a point $x'+w' \in S_J+Q_J$ satisfies $\|A(w'+x')-y'\|_1 < R-\alpha\|A\|_{2\to1}$. Then, approximating it by a point $x+w \in \mathcal{M}_J+\mathcal{N}_J$, such that $\|x'+w'-x-w\|_2 < \alpha$, we get $\|A(w+x)-y\|_1 < R$. This, in combination with the probability estimate above, proves the claim.

Applying the union bound again, we see that the event

$$\begin{aligned}\Omega = \{&\exists j \le n\ \exists J \subset \{1,\dots,n\}\ |J| = l_j\ \exists x \in S_J\ \exists w \in Q_J\\ &\|A(w+x)-y\|_1 < R-\alpha\|A\|_{2\to1}\}\end{aligned}$$

satisfies

$$\mathbb{P}(\Omega) \le \sum_{j=1}^{L}\sum_{|J|=l_j} p_j^{3/4} \le \max_{j=1,\dots,L}\binom{n}{l_j}p_j^{1/4}\cdot\sum_{j=1}^{L} p_j^{1/2}.$$

By condition (7.1), $\binom{n}{l_j}p_j^{1/4} \le p_j^{1/8} \le 1/2$. The same condition and the exponential growth of l_j show also that the sequence $\{p_j^{1/2}\}_{j=1}^{L}$ decays exponentially, and $\sum_{j=1}^{L} p_j^{1/2} \le 2p_1^{1/2}$. This implies

$$\mathbb{P}(\Omega) \le p_1^{1/2}. \tag{7.2}$$

Now let $x \in S^{n-1}$ be any point. Let $\pi : \{1,\dots,n\} \to \{1,\dots,n\}$ be a permutation rearranging the absolute values of the coordinates of x in the non-increasing order: $|x_{\pi(1)}| \ge |x_{\pi(2)}| \ge \dots \ge |x_{\pi(n)}|$. Let $I_1 \cup I_2 \cap \dots \cap I_L = \{1,\dots,n\}$ be the decomposition of $\{1,\dots,n\}$ into a disjoint union of consecutive intervals such that $|I_j| = l_j$. Set $J_j = \pi^{-1}(I_j)$. In other words, the set J_1 contains l_1 largest coordinates of x, J_2 contains l_2 next largest etc. Let x_j be the coordinate projection of x to J_j, i.e. $x_j(i) = x(i)\cdot\mathbf{1}_{J_j}(i)$. Since the largest coordinate of x_j

has the position $\sum_{i=1}^{j-1} l_i + 1$ in the non-increasing rearrangement, and $\|x\|_2 = 1$, we conclude that

$$\|x_j\|_\infty \le \left(\sum_{i=1}^{j-1} l_i + 1\right)^{-1/2} \le l_{j-1}^{-1/2}.$$

If for all $j = 1, \dots, L$ $\|x_j\|_2 \le \alpha^{j-1}/2$, then

$$\|x\|_2 \le \sum_{j=1}^{L} \|x_j\|_2 \le \frac{1}{2} \cdot \frac{1}{1-\alpha} < 1.$$

Hence, there exists a j such that $\|x_j\|_2 > \alpha^{j-1}/2$. Let j be the largest number satisfying this inequality. Then the vector $u = \sum_{i=j+1}^{L} x_i$ satisfies $\|u\|_2 \le \sum_{i=j+1}^{L} \|x_i\|_2 \le \alpha^j$.

Assume that $\|Ax - y\|_1 \le \alpha^{L-1}(R/2 - 2\alpha \|A\|_{2\to 1})$. Then

$$\begin{aligned}\left\| A(\sum_{i=1}^{j} x_i) - y \right\|_1 &\le \alpha^{j-1}(R/2 - 2\alpha \|A\|_{2\to 1}) + \|A\|_{2\to 1} \cdot \|u\|_2 \\ &\le \alpha^{j-1}(R/2 - \alpha \|A\|_{2\to 1}).\end{aligned}$$

Set $J = \mathrm{supp}(x_j)$, $z = x_j / \|x_j\|_2$, and $w = (\sum_{i=1}^{j-1} x_i)/\|x_j\|_2$. Since $\|x_j\|_2 > \alpha^{j-1}/2$, $w \in Q_J$ and the inequality above implies $\left\|A(w+z) - y/\|x_j\|_2\right\|_1 \le R - 2\alpha \|A\|_{2\to 1}$. Hence, the assumption above implies that the event Ω occurs. Therefore,

$$\begin{aligned}&\mathbb{P}\left(\exists x \in S^{n-1} \ \ \|Ax - y\|_1 \le \frac{\alpha^{L-1} R}{4}\right) \\ &\le \mathbb{P}\left(\exists x \in S^{n-1} \ \ \|Ax - y\|_1 \le \alpha^{L-1}\Big(\frac{R}{2} - 2\alpha \|A\|_{2\to 1}\Big) \text{ and } \ \|A\|_{2\to 1} \le \frac{R}{8\alpha}\right) \\ &\quad + \mathbb{P}\left(\|A\|_{2\to 1} > \frac{R}{8\alpha}\right) \le \mathbb{P}(\Omega) + \mathbb{P}\left(\|A\|_{2\to 1} > \frac{R}{8\alpha}\right).\end{aligned}$$

Since $\|A\|_{2\to 1} \le \sqrt{N}\,\|A\|$, the lemma is proved. □

8. Lower bounds for ℓ_1 and ℓ_2 norms

In this section we use the chaining lemma 7.1 to prove Theorems 1.6 and 1.5. Actually we will prove a statement, which is stronger than Theorem 1.6.

Theorem 8.1. *Let K, q, n, d be natural numbers. Assume that*

$$n \le \frac{c d^K}{\log_{(q)} d}.$$

and let $\Delta_1, \dots, \Delta_K$ be $d \times n$ matrices with independent δ random entries. Then for any $y \in \mathbb{R}^{d^K}$

$$\mathbb{P}\left(\exists x \in S^{n-1} \ \ \|(\Delta_1 \otimes_r \dots \otimes_r \Delta_K)x - y\|_1 \le c' d^K\right) \le C' \exp\left(-\bar{c} d\right).$$

Proof. Assume first, that $d^K \ge n \ge d^{K-1/2}$ so the condition of Lemma 5.9 holds for $k = K-1$. Set $R = \tilde{c} d^K$, where $\tilde{c}$ is the constant from Lemma 6.6 Set $\alpha = 8\tilde{c}/C'$, where C' is the constant from Theorem 1.3. By this Corollary,

$$\mathbb{P}\left(\|\Delta_1 \otimes_r \dots \otimes_r \Delta_K\| > \frac{R}{8\alpha\sqrt{d^K}}\right) \le \exp\left(-cd\right).$$

Denote $U = \Delta_1 \otimes_r \dots \otimes_r \Delta_{K-1}$, and let $\mathcal{U}$ be the set of all $d^{K-1} \times n$ matrices A satisfying

$$\mathbb{P}\left(\|A \otimes_r \Delta_K\| > \frac{R}{8\alpha\sqrt{d^K}}\right) \le \exp\left(-c' d\right),$$

where $c' = c/2$. By the Chebychev's inequality $\mathbb{P}\left(U \in \mathcal{U}^c\right) \le \exp\left(-c'd\right)$. Let $y \in \mathbb{R}^{d^K}$. By Lemmata 5.9 and 5.12,

$$\begin{aligned}
&\mathbb{P}\left(\exists x \in S^{n-1} \mid \|(\Delta_1 \otimes_r \dots \otimes_r \Delta_K)x - y\|_1 < c d^K\right) \\
&\le \mathbb{P}\left(\exists x \in S^{n-1} \mid \|(U \otimes_r \Delta_K)x - y\|_1 < c d^K \text{ and } U \in \mathcal{W}_{K-1} \cap \mathcal{V} \cap \mathcal{U}\right) \\
&\quad + c e^{-c'' d}.
\end{aligned}$$

This estimate shows that it is enough to bound the conditional probability

$$\mathbb{P}\left(\exists x \in S^{n-1} \mid \|(U \otimes_r \Delta_K)x - y\|_1 < c d^K \mid U\right)$$

for all matrices $U \in \mathcal{W}_{K-1} \cap \mathcal{V} \cap \mathcal{U}$. This bound is based on Lemma 7.1. Fix a matrix $U \in \mathcal{W}_{K-1} \cap \mathcal{V} \cap \mathcal{U}$ for the rest of the proof. Let $L = K + q$. It is enough to define numbers $l_1, \dots, l_L \in \mathbb{N}$, and $p_1, \dots, p_L \in (0,1)$ which satisfy the conditions of Lemma 7.1. These numbers will be constructed differently for $j \le K$ and $j > K$. The difference between these cases stems from the different behavior of the bound in Lemma 6.6. For relatively small l_j the ℓ_∞ norm of a vector x is large, and the second term in Lemma 6.6 is negligible, compare to the first one. However, for $l_j \ge c d^K / \log^K n$ the picture is opposite, and the second term is dominating.

We consider the case $1 \le j \le K$ first. Set $l_0 = 1$ and $c_0 = 1$. For $1 \le j \le K$ set

$$l_j = \left\lfloor \frac{c_j d^j}{\log^j d} \right\rfloor, \tag{8.1}$$

where the constants $c_1, \dots, c_K$ will be defined inductively. Assume that $c_1, \dots, c_{j-1}$ are already defined. Applying Lemma 6.6 to any vector $x \in S^{n-1}$ with $\|x\|_\infty \le l_{j-1}^{-1/2}$, we get

$$\mathbb{P}\left(\|(U \otimes_r \Delta_K)x - y\|_1 \le cd^K\right) \le \exp\left(-cdl_{j-1}\right) + \exp\left(-\frac{cd^K}{\log^{K-1} n}\right)$$
$$\le \exp\left(-\frac{c'_{j-1}d^j}{\log^{j-1} d}\right) =: p_j,$$

where we can take $c'_{j-1} = c \cdot c_{j-1}/2$. Inequality (7.1) reads

$$\frac{c'_{j-1}d^j}{\log^{j-1} d} \ge 8\frac{c_j d^j}{\log^j d} \cdot \log\left(\frac{6en \cdot \log^j d}{c_j d^j \alpha^j}\right).$$

Since $n \le d^K$, this inequality follows from

$$c'_{j-1} \ge \frac{8c_j}{\log d} \cdot \log\left(\frac{6ed^K}{c_j \alpha^K}\right).$$

Therefore, we can choose c_j independently of d, so that the inequality above is satisfied. Thus, the sequence $l_1, \dots, l_K$ satisfies condition (7.1). Also, if $d \ge d_0$ for some d_0 depending only on K and δ, then $l_{j+1} \ge 2l_j$ for all $j = 1, \dots, K-1$.

Let us now define the numbers l_{K+s} for $s = 1, \dots, q+1$. To this end define the sequence $\{\beta_s\}_{s=0}^q$ by induction. Set

$$\beta_0 = \frac{\log^K d}{c_K} \quad \text{and}$$
$$\beta_s = \tilde{c}\log_{(1)}^K\left(6e\beta_{s-1}\right) \quad \text{for } 1 \le s \le q,$$

where the number $\tilde{c} \ge 1$ will be chosen below. For $0 \le s \le q$ set

$$l_{K+s} = \lfloor d^K/\beta_s \rfloor.$$

Note that for $s = 0$ this formula agrees with (8.1). Let $1 \le s \le q$. By Lemma 6.6, any vector $x \in S^{n-1}$ with $\|x\|_\infty \le l_{K+s-1}^{-1/2}$ satisfies

$$\mathbb{P}\left(\|(U \otimes_r \Delta_K)x - y\|_1 \le cd^K\right)$$
$$\le \exp\left(-cdl_{K+s-1}\right) + \exp\left(-\frac{cd^K}{\log^{K-1}\left(\frac{en}{l_{K+s-1}}\right)}\right)$$
$$\le 2\exp\left(-\frac{cd^K}{\log^{K-1}\left(\frac{en}{l_{K+s-1}}\right)}\right) =: p_{K+s}.$$

The last inequality follows from $dl_{K+s-1} > d^K$. In this case, condition (7.1) reads

$$\frac{cd^K}{\log^{K-1}\left(\frac{en}{l_{K+s-1}}\right)} \geq 8l_{K+s} \cdot \log\left(\frac{6en}{l_{K+s}\alpha^{K+s}}\right),$$

which can be rewritten as

$$\frac{c}{\log^{K-1}\left(\frac{en\beta_{s-1}}{d^K}\right)} \geq \frac{8}{\beta_s} \cdot \log\left(\frac{6en\beta_s}{\alpha^{K+s}d^K}\right).$$

Since the sequence $\{\beta_s\}_{s=0}^q$ is decreasing, and $n \leq d^K$, the previous inequality holds, provided

$$\beta_s \geq \frac{8}{c}\log^{K-1}(e\beta_{s-1}) \cdot \left[\log(6e\beta_{s-1}) + (K+q)\log\frac{1}{\alpha}\right].$$

Since by the definition of β_s, $\log(e\beta_{s-1}) \geq 1$, we can choose

$$\tilde{c} = \frac{8}{c} \cdot (K+q)\log\frac{1}{\alpha}.$$

The inductive definition of the numbers $\beta_1, \dots, \beta_q$ is complete, and the sequences $l_1, \dots, l_{K+q}$, $p_1, \dots, p_{K+q}$ satisfy condition (7.1). Also, if $d \geq d_1$ for some d_1 depending only K, q, and δ, then $\beta_{s+1} \leq \beta_s/2$, and so $l_{K+s+1} \geq 2l_{K+s}$ for $s = 0, \dots, q-1$.

Set $\tilde{n} = \sum_{j=1}^{K+q} l_j$. Then $l_{K+q} \leq \tilde{n} \leq 2l_{K+q}$. From the definition of β_s and induction follows that $1 \leq \beta_s \leq c' \log_{(s)} d$ for all $s = 1, \dots, q$. Hence, there exists $c > 0$ depending only on K, q, δ such that

$$\frac{cd^K}{\log_{(q)} d} \leq \tilde{n} \leq d^K.$$

Thus, for $d \geq \max(d_0, d_1)$, and $n = \tilde{n}$, the assertion of Theorem 8.1 follows from Lemma 7.1. It automatically extends to all $n \leq \tilde{n}$, since for any $y \in \mathbb{R}^{d^K}$ the quantity

$$\min_{x \in S^{\tilde{n}-1}} \|(\Delta_1 \otimes_r \dots \otimes_r \Delta_K)x - y\|_1$$

can only increase, if we take the minimum over $S^{\tilde{n}-1} \cap \mathbb{R}^n$, instead of the whole sphere, in other words, if we consider a submatrix of $\Delta_1 \otimes_r \dots \otimes_r \Delta_K$ consisting of n first columns. It can be also automatically extended to the case $d < \max(d_0, d_1)$ by choosing a large constant C' in the formulation of the Theorem. The proof is now complete. $\square$

Remark 8.2. The probability estimate of Theorem 8.1 is actually optimal. Indeed, let $y = 0$, and assume that the entries of the matrices

$\Delta_1, \dots, \Delta_K$ are i.i.d. random variables taking values $0, 1, -1$ with probability $1/3$ each. Then with probability $(1/3)^d$, the first column of Δ_1 is 0, and so the first column of $\Delta_1 \otimes_r \dots \otimes_r \Delta_K$ is 0 as well.

We conclude with the proof of Theorem 1.5. Set $\tilde{\Delta} = \Delta_1 \otimes_r \dots \otimes_r \Delta_K$. By Theorem 1.6, with probability at least $1 - \exp(-cd)$,

$$\left\| \tilde{\Delta} \right\| \le C' d^{K/2} \quad \text{ and } \forall x \in S^{n-1} \ \left\| \tilde{\Delta} x \right\|_1 \ge c' d^K.$$

Then for any $x \in S^{n-1}$

$$c' d^K \le \left\| \tilde{\Delta} x \right\|_1 \le d^{K/2} \left\| \tilde{\Delta} x \right\|_2 \le d^{K/2} \left\| \tilde{\Delta} \right\|_2 \cdot \|x\|_2 \le C' d^K,$$

so all these norms are equivalent. Comparison between the first and the third term of this inequality implies Theorem 1.5. Moreover, as in [14], we can conclude that $\tilde{\Delta}\mathbb{R}^n$ is a Kashin subspace of $\mathbb{R}^{d^K}$, i.e. the L_1 and L_2 norms are equivalent on it. More precisely, this establishes the following corollary.

Corollary 8.3. *Under the conditions of Theorem 1.6*

$$\mathbb{P}\left(\forall y \in \tilde{\Delta}\mathbb{R}^n \ \ \|y\|_1 \le d^{K/2} \|y\|_2 \le C'' \|y\|_1\right) \ge 1 - \exp\left(-cd\right).$$

References

[1] R. Adamczak, A. E. Litvak, A. Pajor, N. Tomczak-Jaegermann, *Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles*, J. Amer. Math. Soc. 23 (2010), no. 2, 535561.

[2] R. Adamczak, A. E. Litvak, A. Pajor, N. Tomczak-Jaegermann, *Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling*, Constructive Approximation, 34 (2011), 61–88.

[3] G. Cormode, S. Muthukrishnan,*Combinatorial algorithms for Compressed Sensing*, Proc. of SIROCCO, pages 280-294, 2006.

[4] S. Geman, *A limit theorem for the norm of random matrices*, Ann. of Probability 8, no. 2 (1980), 252–261.

[5] Y. A. Garnaev, E. D. Gluskin, *The widths of a Euclidean ball*, (Russian) Dokl. Akad. Nauk SSSR 277 (1984), no. 5, 10481052.

[6] A. Gilbert, M. Strauss, J. Tropp, R. Vershynin, *One sketch for all: fast algorithms for compressed sensing*, STOC'07–Proceedings of the 39th Annual ACM Symposium on Theory of Computing, 237–246, ACM, New York, 2007.

[7] G. S. Glinski and W.-K. Chung, *A Markov model for reliability of probabilistic sequential circuit with Bernoulli inputs*, Proc. 6th Annual Allerton Conf. on Circuit and System Theov, 1968, pp. 568–517.

[8] B. Kashin, *The widths of certain finite-dimensional sets and classes of smooth functions*, (Russian), Izv. Akad. Nauk SSSR Ser. Mat., 41 (1977), 334–351.

[9] S. P. Kasiviswanathan, M. Rudelson, A. Smith, J. Ullman, *The price of privately releasing contingency tables and the spectra of random matrices with correlated rows*, STOC 2010, 775–784.

[10] M. Ledoux, The concentration of measure phenomenon, Mathematical Surveys and Monographs, vol 89, AMS 2001.

[11] A. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes*, Adv. Math. 195 (2005), 491–523.

[12] A. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, R.Vershynin, *Euclidean embeddings in spaces of finite volume ratio via random matrices*, J. Reine Angew. Math. 589 (2005), 1–19.

[13] G. Pisier, Pisier, The volume of convex bodies and Banach space geometry. Cambridge Tracts in Mathematics, 94. Cambridge University Press, Cambridge, 1989.

[14] M. Rudelson, *Lower estimates for the singular values of random matrices*, C. R. Math. Acad. Sci. Paris 342 (2006), no. 4, 247–252.

[15] M.Rudelson, R.Vershynin, *The Littlewood-Offord Problem and invertibility of random matrices*, Adv. Math. 218 (2008), no. 2, 600–633.

[16] M.Rudelson, R.Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*, Communications on Pure and Applied Mathematics 61 (2008), 1025-1045.

[17] M.Rudelson, R.Vershynin, *Smallest singular value of a random rectangular matrix*, Communications on Pure and Applied Mathematics 62 (2009), 1707–1739.

[18] S. J. Szarek, *On Kashin's almost Euclidean orthogonal decomposition of* l_n^1, Bull. Acad. Polon. Sci., Sér. Sci. Math. Astronom. Phys. 26 (1978), 691–694.

[19] S. J. Szarek, N. Tomczak-Jaegermann, *On nearly Euclidean decomposition for some classes of Banach spaces*, Compositio Math., 40 (1980), 367–385.

[20] Talagrand, *A new look at independence*, Ann. Probab. 24 (1996), no. 1, 134.

[21] M. Talagrand, The generic chaining. Upper and lower bounds of stochastic processes. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2005.

[22] R. Vershynin, *How close is the sample covariance matrix to the actual covariance matrix?* Journal of Theoretical Probability, to appear.

[23] A. W. van der Vaart, J. A. Wellner, Weak convergence and empirical processes. With applications to statistics. Springer Series in Statistics. Springer-Verlag, New York, 1996.

[24] R. C. Windecker, *Stochastic combinatorial networks*, Information Sciences 16 (1978), 185–234.

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109.

E-mail address: `rudelson@umich.edu`