HOW CLOSE IS THE SAMPLE COVARIANCE MATRIX TO THE ACTUAL COVARIANCE MATRIX?

ROMAN VERSHYNIN

ABSTRACT. Given a probability distribution in $\mathbb{R}^n$ with general (non-white) covariance, a classical estimator of the covariance matrix is the sample covariance matrix obtained from a sample of N independent points. What is the optimal sample size $N = N(n)$ that guarantees estimation with a fixed accuracy in the operator norm? Suppose the distribution is supported in a centered Euclidean ball of radius $O(\sqrt{n})$. We conjecture that the optimal sample size is $N = O(n)$ for all distributions with finite fourth moment, and we prove this up to an iterated logarithmic factor. This problem is motivated by the optimal theorem of M. Rudelson [23] which states that $N = O(n \log n)$ for distributions with finite second moment, and a recent result of R. Adamczak et al. [1] which guarantees that $N = O(n)$ for sub-exponential distributions.

1. INTRODUCTION

1.1. **Approximation problem for covariance marices.** Estimation of covariance matrices of high dimensional distributions is a basic problem in multivariate statistics. It arises in diverse applications such as signal processing [14], genomics [25], financial mathematics [16], pattern recognition [7], geometric functional analysis [23] and computational geometry [1]. The classical and simplest estimator of a covariance matrix is the sample covariance matrix. Unfortunately, the spectral theory of sample covariance matrices has not been well developed except for product distributions (or affine transformations thereof) where one can rely on random matrix theory for matrices with independent entries. This paper addresses the following basic question: how well does the sample covariance matrix approximate the actual covariance matrix in the operator norm?

We consider a mean zero random vector X in a high dimensional space $\mathbb{R}^n$ and N independent copies $X_1, \dots, X_N$ of X. We would like to approximate the covariance matrix of X

$$\Sigma = \mathbb{E} X \otimes X = \mathbb{E} X X^T$$

Partially supported by NSF grant FRG DMS 0918623.

by the sample covariance matrix

$$\Sigma_N = \frac{1}{N} \sum_{i=1}^N X_i \otimes X_i.$$

Problem. *Determine the minimal sample size $N = N(n, \varepsilon)$ that guarantees with high probability (say, 0.99) that the sample covariance matrix Σ_N approximates the actual covariance matrix Σ with accuracy ε in the operator norm $\ell_2 \to \ell_2$, i.e. so that*

$$\|\Sigma - \Sigma_N\| \le \varepsilon. \tag{1.1}$$

The use of the operator norm in this problem allows one a good grasp of the spectrum of Σ, as each eigenvalue of Σ would lie within ε from the corresponding eigenvalue of Σ_N.

It is common for today's applications to operate with increasingly large number of parameters n, and to require that sample sizes N be moderate compared with n. As we impose no a priori structure on the covariance matrix, we must have $N \ge n$ for dimension reasons. Note that for some structured covariance matrices, such as sparse or having an off diagonal decay, one can sometimes achieve N smaller than n and even comparable to $\log n$, by transforming the sample covariance matrix in order to adhere to the same structure (e.g. by shrinkage of eigenvalues or thresholding of entries). We will not consider structured covariance matrices in this paper; see e.g. [22] and [18].

1.2. **Two examples.** The most extensively studied model in random matrix theory is where X is a random vector with independent coordinates. However, independence of coordinates can not be justified in some important applications, and in this paper we shall consider general random vectors. Let us illustrate this point with two well studied examples.

Consider some non-random vectors $x_1, \dots, x_M$ in $\mathbb{R}^n$ which satisfy Parseval's identity (up to normalization):

$$\frac{1}{M} \sum_{j=1}^M \langle x_j, x \rangle^2 = \|x\|_2^2 \quad \text{for all } x \in \mathbb{R}^n. \tag{1.2}$$

Such generalizations of orthogonal bases (x_j) are called tight frames. They arise in convex geometry via John's theorem on contact points of convex bodies [4] and in signal processing as a convenient mean to introduce redundancy into signal representations [11]. From a probabilistic point of view, we can regard the normalized sum in (1.2) as the expected value of a certain random variable. Indeed, Parseval's identity (1.2) amounts to $\frac{1}{M} \sum_{j=1}^M x_j \otimes x_j = I$. Once we introduce a random vector X uniformly distributed in the set of M points $\{x_1, \dots, x_M\}$, Parseval's identity will read as $\mathbb{E} X \otimes X = I$. In other

words, the covariance matrix of X is identity, $\Sigma = I$. Note that there is no reason to assume that the coordinates of X are independent.

Suppose further that the covariance matrix of X can be approximated by the sample covariance matrix Σ_N for some moderate sample size $N = N(n, \varepsilon)$. Such an approximation $\|\Sigma_N - I\| \le \varepsilon$ means simply that a random subset of N vectors $\{x_{j_1}, \dots, x_{j_N}\}$ taken from the tight frame $\{x_1, \dots, x_M\}$ independently and with replacement is still an approximate tight frame:

$$(1-\varepsilon)\|x\|_2^2 \le \frac{1}{N}\sum_{i=1}^{N} \langle x_{j_i}, x\rangle^2 \le (1+\varepsilon)\|x\|_2^2 \quad \text{for all } x \in \mathbb{R}^n.$$

In other words, a small random subset of a tight frame is still an approximate tight frame; the size of this subset N does not even depend on the frame size M. For applications of this type of results in communications see [28].

Another extensively studied class of examples is the uniform distribution on a convex body K in $\mathbb{R}^n$. A number of algorithms in computational convex geometry (for volume computing and optimization) rely on covariance estimation in order to put K in the isotropic position, see [12, 13]. Note that in this class of examples, the random vector uniformly distributed in K typically does not have independent coordinates.

1.3. **Sub-gaussian and sub-exponential distributions.** Known results on the approximation problem differ depending on the moment assumptions on the distribution. The simplest case is when X is a *sub-gaussian random vector* in $\mathbb{R}^n$, thus satisfying for some L that

$$\mathbb{P}(|\langle X, x\rangle| > t) \le 2e^{-t^2/L^2} \quad \text{for } t > 0 \text{ and } x \in S^{n-1}. \tag{1.3}$$

Examples of sub-gaussian distributions with $L = O(1)$ include the standard Gaussian random distribution in $\mathbb{R}^n$, the uniform distribution on the cube $[-1, 1]^n$, but not the uniform distribution on the unit octahedron $\{x \in \mathbb{R}^n : |x_1| + \cdots + |x_n| \le 1\}$. For sub-gaussian distributions in $\mathbb{R}^n$, the optimal sample size in the approximation problem (1.1) is linear in the dimension, thus $N = O_{L,\varepsilon}(n)$. This known fact follows from a large deviation inequality and an ε-net argument, see Proposition 2.1 below.

Significant difficulties arise when one tries to extend this result to the larger class of *sub-exponential random vectors* X, which only satisfy (1.3) with t^2/L^2 replaced by t/L. This class is important because, as follows from Brunn-Minkowski inequality, the uniform distribution on every convex body K is sub-exponential provided that the covariance matrix is identity (see [10, Section 2.2.(b_3)]). For the uniform distributions on convex bodies, a result of J. Bourgain [6] guaranteed approximation of covariance matrices with sample size slightly larger than linear in the dimension, $N = O_\varepsilon(n \log^3 n)$. Around the

same time, a slightly better bound $N = O_\varepsilon(n \log^2 n)$ was proved by M. Rudelson [23]. It was subsequently improved to $N = O_\varepsilon(n \log n)$ for convex bodies symmetric with respect to the coordinate hyperplanes by A. Giannopoulos et al. [8], and for general convex bodies by G. Paouris [20]. Finally, an optimal estimate $N = O_\varepsilon(n)$ was obtained by G. Aubrun [3] for convex bodies with the symmetry assumption as above, and for general convex bodies by R. Adamczak et al. [1]. The result in [1] is actually valid for all sub-exponential distributions supported in a ball of radius $O(\sqrt{n})$. Thus, if X is a random vector in $\mathbb{R}^n$ that satisfies for some K, L that

$$\|X\|_2 \le K\sqrt{n} \text{ a.s.}, \quad \mathbb{P}(|\langle X, x\rangle| > t) \le 2e^{-t/L} \quad \text{for } t > 0 \text{ and } x \in S^{n-1} \tag{1.4}$$

then the optimal sample size is $N = O_{K,L,\varepsilon}(n)$.

The boundedness assumption $\|X\|_2 = O(\sqrt{n})$ is usually non-restrictive, since many natural distributions satisfy this bound with overwhelming probability. For example, the standard Gaussian random vector in $\mathbb{R}^n$ satisfies this with probability at least $1 - e^{-n}$. It follows by union bound that for any sample size $N \ll e^n$, all independent vectors in the sample $X_1, \dots, X_N$ satisfy this inequality simultaneously with overwhelming probability. Therefore, by truncation one may assume without loss of generality that $\|X\|_2 = O(\sqrt{n})$. A similar reasoning is valid for uniform distributions on convex bodies. In this case one can use the concentration result of G. Paouris [20] which implies that $\|X\|_2 = O(\sqrt{n})$ with probability at least $1 - e^{-\sqrt{n}}$.

1.4. **Distributions with finite moments.** Unfortunately, the class of sub-exponential distributions is too restrictive for many natural applications. For example, discrete distributions in $\mathbb{R}^n$ supported on less than $e^{O(\sqrt{n})}$ points are usually not sub-exponential. Indeed, suppose a random vector X takes values in some set of M vectors of Euclidean length $\sqrt{n}$. Then the unit vector x pointing to the most likely value of X witnesses that $\mathbb{P}(|\langle X, x\rangle| = \sqrt{n}) \ge 1/M$. It follows that in order for the random vector X to be sub-exponential with $L = O(1)$, it must be supported on a set of size $M \ge e^{c\sqrt{n}}$. However, in applications such as (1.2) it is desirable to have a result valid for distributions on sets of moderate sizes M, e.g. polynomial or even linear in dimension n. This may also be desirable in modern statistical applications, which typically operate with large number of parameters n that may not be exponentially smaller than the population size M.

So far, there has been only one approximation result with very weak assumptions on the distribution. M. Rudelson [23] showed that if a random vector X in $\mathbb{R}^n$ satisfies

$$\|X\|_2 \le K\sqrt{n} \text{ a.s.}, \quad \mathbb{E}\langle X, x\rangle^2 \le L^2 \quad \text{for } x \in S^{n-1} \tag{1.5}$$

then the minimal sample size that guarantees approximation (1.1) is $N = O_{K,L,\varepsilon}(n \log n)$. The second moment assumption in (1.5) is very weak; it is equivalent to the boundedness of the covariance matrix, $\|\Sigma\| \le L$. The logarithmic oversampling factor is necessary in this extremely general result, as can be seen from the example of the uniform distribution on the set of n vectors of Euclidean length $\sqrt{n}$. The coupon collector's problem calls for the size $N \gtrsim n \log n$ in order for the sample $\{X_1, \dots, X_N\}$ to contain all these vectors, which is obviously required for a nontrivial covariance approximation.

There is clearly a big gap between the sub-exponential assumption (1.4) where the optimal size is $N \sim n$ and the weakest second moment assumption (1.5) where the optimal size is $N \sim n \log n$. It would be useful to classify the distributions for which the logarithmic oversampling is needed. The picture is far from complete – the uniform distributions on convex bodies in $\mathbb{R}^n$ for which we now know that the logarithmic oversampling is not needed are very far from the uniform distributions on $O(n)$ points for which the logarithmic oversampling is needed. We conjecture that the logarithmic oversampling is not needed for all distributions with q-th moment with appropriate absolute constant q; probably $q = 4$ suffices or even any $q > 2$. We will thus assume that

$$\|X\|_2 \le K\sqrt{n} \text{ a.s.}, \quad \mathbb{E}|\langle X, x\rangle|^q \le L^q \quad \text{for } x \in S^{n-1}. \tag{1.6}$$

Conjecture 1.1. *Let X be a random vector in $\mathbb{R}^n$ that satisfies the moment assumption* (1.6) *for some appropriate absolute constant q and some K, L. Let $\varepsilon > 0$. Then, with high probability, the sample size $N \gtrsim_{K,L,\varepsilon} n$ suffices to approximate the covariance matrix Σ of X by the sample covariance matrix Σ_N in the operator norm: $\|\Sigma - \Sigma_N\| \le \varepsilon$.*

In this paper we prove the Conjecture up to an iterated logarithmic factor.

Theorem 1.2. *Consider a random vector X in $\mathbb{R}^n$ ($n \ge 4$) which satisfies moment assumptions* (1.6) *for some $q > 4$ and some K, L. Let $\delta > 0$. Then, with probability at least $1-\delta$, the covariance matrix Σ of X can be approximated by the sample covariance matrix Σ_N as*

$$\|\Sigma - \Sigma_N\| \lesssim_{q,K,L,\delta} (\log\log n)^2 \Big(\frac{n}{N}\Big)^{\frac{1}{2}-\frac{2}{q}}.$$

Remarks. 1. The notation $a \lesssim_{q,K,L,\delta} b$ means that $a \le C(q, K, L, \delta)b$ where $C(q, K, L, \delta)$ depends only on the parameters q, K, L, δ; see Section 2.3 for more notation. The logarithms are to the base 2. We put the restriction $n \ge 4$ only to ensure that $\log\log n \ge 1$; Theorem 1.2 and other results below clearly hold for dimensions $n = 1, 2, 3$ even without the iterated logarithmic factors.

2. It follows that for every $\varepsilon > 0$, the desired approximation $\|\Sigma - \Sigma_N\| \le \varepsilon$ is guaranteed if the sample has size

$$N \gtrsim_{q,K,L,\delta,\varepsilon} (\log\log n)^p n \quad \text{where } \frac{1}{p} + \frac{1}{q} = \frac{1}{4}.$$

3. A similar result holds for independent random vectors $X_1, \dots, X_N$ that are not necessarily identically distributed; we will prove this general result in Theorem 6.1.

4. The boundedness assumption $\|X\|_2 \le K\sqrt{n}$ in (1.6) can often be weakened or even dropped by a simple modification of Theorem 1.2. This happens, for example, if $\max_{i \le N} \|X_i\| = O(\sqrt{n})$ holds with high probability, as one can apply Theorem 1.2 conditionally on this event. We refer the reader to a thorough discussion of the boundedness assumption in Section 1.3 of [30].

1.5. **Extreme eigenvalues of sample covariance matrices.** Theorem 1.2 can be used to analyze the spectrum of sample covariance matrices Σ_N. The case when the random vector X has i.i.d. coordinates is most studied in random matrix theory. Suppose that both $N, n \to \infty$ while the aspect ratio $n/N \to \beta \in (0, 1]$. If the coordinates of X have unit variance and finite fourth moment, then clearly $\Sigma = I$. The largest eigenvalue $\lambda_1(\Sigma_N)$ then converges a.s. to $(1 + \sqrt{\beta})^2$, and the smallest eigenvalue $\lambda_n(\Sigma_N)$ converges a.s. to $(1 - \sqrt{\beta})^2$, see [5]. For more on the extreme eigenvalues in both asymptotic regime ($N, n \to \infty$) and non-asymptotic regime (N, n fixed), see [24].

Without independence of the coordinates, analyzing the spectrum of sample covariance matrices Σ_N becomes significantly harder. Suppose that $\Sigma = I$. For sub-exponential distributions, i.e. those satisfying (1.4), it was proved in [2] that

$$1 - O(\sqrt{\beta}) \le \lambda_n(\Sigma_N) \le \lambda_1(\Sigma_N) \le 1 + O(\sqrt{\beta}).$$

(A weaker version with extra $\log(1/\beta)$ factors was proved earlier by the same authors in [1].) Under only finite moment assumption (1.6), Theorem 1.2 clearly yields

$$1 - O(\log\log n)\beta^{\frac{1}{2} - \frac{2}{q}} \le \lambda_n(\Sigma_N) \le \lambda_1(\Sigma_N) \le 1 + O(\log\log n)\beta^{\frac{1}{2} - \frac{2}{q}}.$$

Note that for large exponents q, the factor $\beta^{\frac{1}{2} - \frac{2}{q}}$ becomes close to $\sqrt{\beta}$.

1.6. **Norms of random matrices with independent columns.** One can interpret the results of this paper in terms of random matrices with independent columns. Indeed, consider an $n \times N$ random matrix $A = [X_1, \dots, X_N]$ whose columns $X_1, \dots, X_N$ are drawn independently from some distribution on $\mathbb{R}^n$. The sample covariance matrix of this distribution is simply $\Sigma_N = \frac{1}{N} AA^T$, so the eigenvalues of $N^{1/2}\Sigma_N$ are the singular values of A. In particular, under

the same finite moment assumptions as in Theorem 1.2, we obtain the bound on the operator norm

$$\|A\| \lesssim_{q,K,L,\delta} \log\log N \cdot (\sqrt{n} + \sqrt{N}). \tag{1.7}$$

This follows from a result leading to Theorem 1.2, see Corollary 5.2. The bound is optimal up to the $\log\log N$ factor for matrices with i.i.d. entries, because the operator norm is bounded below by the Euclidean norm of any column and any row. For random matrices with independent entries, estimate (1.7) follows (under the fourth moment assumption) from more general bounds by Seginer [26] and Latala [15], and even without the $\log\log N$ factor. Without independence of entries, this bound was proved by the author [29] for products of random matrices with independent entries and deterministic matrices, and also without the $\log\log N$ factor.

1.7. **Organization of the rest of the paper.** In the beginning of Section 2 we outline the heuristics of our argument. We emphasize its two main ingredients – structure of divergent series and a decoupling principle. We finish that section with some preliminary material – notation (Section 2.3), a known argument that solves the approximation problem for sub-gaussian distributions (Section 2.4), and the previous weaker result of the author [30] on the approximation problem in the weak ℓ_2 norm (Section 2.5).

The heart of the paper are Sections 3 and 4. In Section 3 we study the structure of series that diverge faster than the iterated logarithm. This structure is used in Section 4 to deduce a decoupling principle. In Section 5 we apply the decoupling principle to norms of random matrices. Specifically, in Theorem 5.1 we estimate the norm of $\sum_{i\in E} X_i \otimes X_i$ uniformly over subsets E. We interpret this in Corollary 5.2 as a norm estimate for random matrices with independent columns. In Section 6, we deduce the general form of our main result on approximation of covariance matrices, Theorem 6.1.

Acknowledgement. The author is grateful to the referee for useful suggestions.

2. Outline of the method and preliminaries

Let us now outline the two main ingredients of our method, which are a new structure theorem for divergent series and a new decoupling principle. For the sake of simplicity in this discussion, we shall now concentrate on proving the weaker upper bound $\|\Sigma_N\| = O(1)$ in the case $N = n$. Once this simpler case is understood, the full Theorem 1.2 will require a little extra effort using a now standard truncation argument due to J. Bourgain [6]. We thus consider independent copies $X_1, \dots, X_n$ of a random vector X in $\mathbb{R}^n$ satisfying the finite

moment assumptions (1.6). We would like to show with high probability that

$$\|\Sigma_n\| = \sup_{x \in S^{n-1}} \frac{1}{n} \sum_{i=1}^n \langle X_i, x \rangle^2 = O(1).$$

In this expression we may recongize a stochastic process indexed by vectors x on the sphere. For each fixed x, we have to control the sum of independent random variables $\sum_i \langle X_i, x \rangle^2$ with finite moments. Suppose the bad event occurs – for some x, this sum is significantly larger than n. Unfortunately, because of the heavy tails of these random variables, the bad event may occur with polynomial rather than exponential probability $n^{-O(1)}$. This is too weak to control these sums for all x simultaneously on the n-dimensional sphere, where ε-nets have exponential sizes in n. So, instead of working with sums of independent random variables, we try to locate some structure in the summands responsible for the largeness of the sum.

2.1. **Structure of divergent series.** More generally, we shall study the structure of divergent series $\sum_i b_i = \infty$, where $b_i \geq 0$. Let us first suppose that the series diverges faster than logarithmic function, thus

$$\sum_{i=1}^n b_i \gg \log n \quad \text{for some } n \geq 2.$$

Comparing with the harmonic series we see that the non-increasing rearrangement b_i^* of the coefficients at some point must be large:

$$b_{n_1}^* \gg 1/n_1 \quad \text{for some } n_1 \leq n.$$

In other words, one can find n_1 large terms of the sum: there exists an index set $I \subset [n]$ of size $|I| = n_1$ and such that $b_i \gg 1/n_1$ for $i \in I$. This collection of large terms $(b_i)_{i \in I}$ forms a desired structure responsible for the largeness of the series $\sum_i b_i$. Such a structure is well suited to our applications where b_i are independent random variables, $b_i = \langle X_i, x \rangle^2 / n$. Indeed, the events $\{b_i \gg 1/n_1\}$ are independent, and the probability of each such event is easily controlled by finite moment assumptions (2.2) through Markov's inequality. This line was developed in [30], but it clearly leads to a loss of logarithmic factor which we are trying to avoid in the present paper.

We will work on the next level of precision, thus studying the structure of series that diverge slower than the logarithmic function but faster than the iterated logarithm. So let us assume that

$$b_i^* \lesssim 1/i \text{ for all } i; \quad \sum_{i=1}^n b_i \gg \log\log n \text{ for some } n \geq 4.$$

In Proposition 3.1 we will locate almost the same structure as we had for logarithmically divergent series, except up to some factor $\log\log n \ll l \lesssim \log n$,

as follows. For some $n_1 \le n$ there exists an index set $I \subset [n]$ of size $|I| = n_1$, such that

$$b_i \gg \frac{1}{ln_1} \text{ for } i \in I, \quad \text{and moreover } \frac{n}{n_1} \ge 2^{l/2}.$$

2.2. **Decoupling.** The structure that we found is well suited to our application where b_i are independent random variables $b_i = \langle X_i, x\rangle^2/n$. In this case we have

$$\langle X_i, x\rangle^2 \gg \frac{n}{ln_1} \gtrsim \frac{n}{n_1} \Big/ \log\Big(\frac{2n}{n_1}\Big) \gtrsim (n/n_1)^{1-o(1)} \quad \text{for } i \in I. \tag{2.1}$$

The probability that this happens is again easy to control using independence of $\langle X_i, x\rangle$ for fixed x, finite moment assumptions (2.2) and Markov's inequality. Since there are $\binom{n}{n_1}$ number of ways to choose the subset I, the probability of the event in (2.1) is bounded by

$$\binom{n}{n_1}\mathbb{P}\big\{\langle X_i, x\rangle^2 \gg (n/n_1)^{1-o(1)}\big\}^{n_1} \le \binom{n}{n_1}(10n/n_1)^{-(1-o(1))q/2} \ll e^{-n_1}$$

where the last inequality follows because $\binom{n}{n_1} \le (en/n_1)^{n_1}$ and since $q > 2$.

Our next task is to unfix $x \in S^{n-1}$. The exponential probability estimate we obtained allows us to take the union bound over all x in the unit sphere of any fixed n_1-dimensional subspace, since this sphere has an ε-net of size exponential in n_1. We can indeed assume without loss of generality that the vector x in our structural event (2.1) lies in the span of $(X_i)_{i\in I}$ which is n_1-dimensional; this can be done by projecting x onto this span if necessary. Unfortunately, this obviously makes x depend on the random vectors $(X_i)_{i\in I}$ and destroys the independence of random variables $\langle X_i, x\rangle$. This hurdle calls for a decoupling mechanism, which would make x in the structural event (2.1) depend on some small fraction of the vectors $(X_i)_{i\in I}$. One would then condition on this fraction of random vectors and use the structural event (2.1) for the other half, which would quickly lead to completion of the argument.

Our decoupling principle, Proposition 4.1, is a deterministic statement that works for fixed vectors X_i. Loosely speaking, we assume that the structural event (2.1) holds for some x in the span of $(X_i)_{i\in I}$, and we would like to force x to lie in the span of a small fraction of these X_i. We write x as a linear combination $x = \sum_{i\in I} c_i X_i$. The first step of decoupling is to remove the "diagonal" term $c_i X_i$ from this sum, while retaining the largeness of $\langle X_i, x\rangle$. This task turns out to be somewhat difficult, and it will force us to refine our structural result for divergent series by adding a domination ingredient into it. This will be done at the cost of another $\log\log n$ factor. After the diagonal term is removed, the number of terms in the sum for x will be reduced by a probabilistic selection using Maurey's empirical method.

2.3. **Notation and preliminaries.** We will use the following notation throughout this paper. C and c will stand for positive absolute constants; C_p will denote a quantity which only depends on the parameter p, and similar notation will be used with more than one parameter. For positive numbers a and b, the asymptotic inequality $a \lesssim b$ means that $a \le Cb$. Similarly, inequalities of the form $a \lesssim_{p,q} b$ mean that $a \lesssim C_{p,q} b$. Intervals of integers will be denoted by $[n] := \{1, \dots, \lceil n \rceil\}$ for $n \ge 0$. The cardinality of a finite set I is denoted by $|I|$. All logarithms will be to the base 2.

The non-increasing rearrangement of a finite or infinite sequence of numbers $a = (a_i)$ will be denoted by (a_i^*). Recall that the ℓ_p norm is defined as $\|a\|_p = (\sum_i |a_i|^p)^{1/p}$ for $1 \le p < \infty$, and $\|a\|_\infty = \max_i |a_i|$. We will also consider the weak ℓ_p norm for $1 \le p < \infty$, which is defined as the infimum of positive numbers M for which the non-increasing rearrangement $(|a|_i^*)$ of the sequence $(|a_i|)$ satisfies $|a|_i^* \le M i^{-1/p}$ for all i. For sequences of finite length n, it follows from definition that the weak ℓ_p norm is equivalent to the ℓ_p norm up to a $O(\log n)$ factor, thus $\|a\|_{p,\infty} \le \|a\|_p \lesssim \log n \cdot \|a\|_{p,\infty}$ for $a \in \mathbb{R}^n$.

In this paper we deal with the $\ell_2 \to \ell_2$ operator norm of $n \times n$ matrices $\|A\|$, also known as spectral norm. By definition,

$$\|A\| = \sup_{x \in S^{n-1}} \|Ax\|_2$$

where S^{n-1} denotes the unit Euclidean sphere in $\mathbb{R}^n$. Equivalently, $\|A\|$ is the largest singular value of A and the largest eigenvalue of $\sqrt{AA^T}$. We will frequently use that for Hermitian matrices A one has

$$\|A\| = \sup_{x \in S^{n-1}} |\langle Ax, x \rangle|.$$

It will be convenient to work in a slightly more general than in Theorem 1.2, and consider independent random vectors X_i in $\mathbb{R}^n$ that are not necessarily identically distributed. All we need is that moment assumptions (1.6) hold uniformly for all vectors:

$$\|X_i\|_2 \le K\sqrt{n} \text{ a.s.}, \quad (\mathbb{E}|\langle X_i, x \rangle|^q)^{1/q} \le L \text{ for all } x \in S^{n-1}. \tag{2.2}$$

We can view our goal as establishing a *law of large numbers in the operator norm*, and with quantitative estimates on convergence. Thus we would like to show that the approximation error

$$\Big\| \frac{1}{N} \sum_{i=1}^N X_i \otimes X_i - \mathbb{E} X_i \otimes X_i \Big\| = \sup_{x \in S^{n-1}} \Big| \frac{1}{N} \sum_{i=1}^N \langle X_i, x \rangle^2 - \mathbb{E} \langle X_i, x \rangle^2 \Big| \tag{2.3}$$

is small like in Theorem 1.2.

2.4. **Sub-gaussian distributions.** A solution to the approximation problem is well known and easy for sub-gaussian random vectors, those satisfying (1.3). The optimal sample size here is proportional to the dimension, thus $N = O_{L,\varepsilon}(n)$. For the reader's convenience, we recall and prove a general form of this result.

Proposition 2.1 (Sub-gaussian distributions)**.** *Consider independent random vectors $X_1, \dots, X_N$ in $\mathbb{R}^n$, $N \geq n$, which have sub-gaussian distribution as in* (1.3) *for some L. Then for every $\delta > 0$ with probability at least $1 - \delta$ one has*

$$\Big\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - \mathbb{E} X_i \otimes X_i \Big\| \lesssim_{L,\delta} \Big(\frac{n}{N} \Big)^{\frac{1}{2}}.$$

One should compare this with our main result, Theorem 1.2, which yields almost the same conclusion under only finite moment assumptions on the distribution, except for an iterated logarithmic factor and a slight loss of the exponent $1/2$ (the latter may be inevitable when dealing with finite moments).

The well known proof of Proposition 2.1 is based on Bernstein's deviation inequality for independent random variables and an ε-net argument. The latter allows to replace the sphere S^{n-1} in the computation of the norm in (2.3) by a finite ε-net as follows.

Lemma 2.2 (Computing norms on ε-nets)**.** *Let A be a Hermitian $n \times n$ matrix, and let $\mathcal{N}_\varepsilon$ be an ε-net of the unit Euclidean sphere S^{n-1} for some $\varepsilon \in [0, 1)$. Then*

$$\|A\| = \sup_{x \in S^{n-1}} |\langle Ax, x \rangle| \leq (1 - 2\varepsilon)^{-1} \sup_{x \in \mathcal{N}_\varepsilon} |\langle Ax, x \rangle|.$$

Proof. Let us choose $x \in S^{n-1}$ for which $\|A\| = |\langle Ax, x \rangle|$, and choose $y \in \mathcal{N}_\varepsilon$ which approximates x as $\|x - y\|_2 \leq \varepsilon$. It follows by the triangle inequality that

$$\begin{aligned} |\langle Ax, x \rangle - \langle Ay, y \rangle| &= |\langle Ax, x - y \rangle + \langle A(x - y), y \rangle| \\ &\leq \|A\| \|x\|_2 \|x - y\|_2 + \|A\| \|x - y\|_2 \|y\|_2 \leq 2\|A\|\varepsilon. \end{aligned}$$

It follows that

$$|\langle Ay, y \rangle| \geq |\langle Ax, x \rangle| - 2\|A\|\varepsilon = (1 - 2\varepsilon)\|A\|.$$

This completes the proof. □

Proof of Proposition 2.1. Without loss of generality, we can assume that in the sub-gaussian assumption (1.3) we have $L = 1$ by replacing X_i by X_i / L. Identity (2.3) expresses the norm in question as a supremum over the unit sphere S^{n-1}. Next, Lemma 2.2 allows to replace the sphere in (2.3) by its $1/2$-net $\mathcal{N}$ at the cost of an absolute constant factor. Moreover, we can arrange so

that the net has size $|\mathcal{N}| \le 6^n$; this follows by a standard volumetric argument (see [17, Lemma 9.5]).

Let us fix $x \in \mathcal{N}$. The sub-gaussian assumption on X_i implies that the random variables $\langle X_i, x\rangle^2$ are sub-exponential: $\mathbb{P}(\langle X_i, x\rangle^2 > t) \le 2e^{-t}$ for $t > 0$. Bernstein's deviation inequality for independent sub-exponential random variables (see e.g. [27, Section 2.2.2]) yields for all $\varepsilon > 0$ that

$$\mathbb{P}\Big\{\Big|\frac{1}{N}\sum_{i=1}^N \langle X_i, x\rangle^2 - \mathbb{E}\langle X_i, x\rangle^2\Big| > \varepsilon\Big\} \le 2e^{-c\varepsilon^2 N}. \tag{2.4}$$

Now we unfix x. Using (2.4) for each x in the net $\mathcal{N}$, we conclude by the union bound that the event

$$\Big|\frac{1}{N}\sum_{i=1}^N \langle X_i, x\rangle^2 - \mathbb{E}\langle X_i, x\rangle^2\Big| < \varepsilon \quad \text{for all } x \in \mathcal{N}$$

holds with probability at least

$$1 - |\mathcal{N}| \cdot 2e^{-c\varepsilon^2 N} \ge 1 - 2e^{2n - c\varepsilon^2 N}.$$

Now if we choose $\varepsilon^2 = (4/c)\log(2/\delta)\, n/N$, this probability is further bounded below by $1 - \delta$ as required. By the reduction from the sphere to the net mentioned in the beginning of the argument, this completes the proof. □

2.5. **Results in the weak ℓ_2 norm, and almost orthogonality of X_i.** A truncation argument of J. Bourgain [6] reduces the approximation problem to finding an upper bound on

$$\Big\|\sum_{i\in E} X_i \otimes X_i\Big\| = \sup_{x\in S^{n-1}} \sum_{i\in E} \langle X_i, x\rangle^2 = \sup_{x\in S^{n-1}} \|(\langle X_i, x\rangle)_{i\in E}\|_2^2$$

uniformly for all index sets $E \subset [N]$ with given size. A weaker form of this problem, with the weak ℓ_2 norm of the sequence $\langle X_i, x\rangle$ instead of the its ℓ_2 norm, was studied in [30]. The following bound was proved there:

Theorem 2.3 ([30] Theorem 3.1)**.** *Consider random vectors $X_1, \dots, X_N$ which satisfy moment assumptions* (2.2) *for some $q > 4$ and some K, L. Then, for every $t \ge 1$, with probability at least $1 - Ct^{-0.9q}$ one has*

$$\sup_{x\in S^{n-1}} \|(\langle X_i, x\rangle)_{i\in E}\|_{2,\infty}^2 \lesssim_{q,K,L} n + t^2\Big(\frac{N}{|E|}\Big)^{4/q}|E| \quad \textit{for all } E \subseteq [N].$$

For most part of our argument (through decoupling), we treat X_i as fixed non-random vectors. The only property we require from X_i is that they are almost pairwise orthogonal. For random vectors, an almost pairwise orthogonality easily follows from the moment assumptions (2.2):

Lemma 2.4 ([30] Lemma 3.3)**.** *Consider random vectors $X_1, \dots, X_N$ which satisfy moment assumptions* (2.2) *for some $q > 4$ and some K, L. Then, for every $t \geq 1$, with probability at least $1 - Ct^{-q}$ one has*

$$\frac{1}{|E|} \sum_{i \in E,\, i \neq k} \langle X_i, X_k \rangle^2 \lesssim_{q,K,L} t^2 \Big(\frac{N}{|E|} \Big)^{4/q} n \quad \text{for all } E \subseteq [N],\ k \in [N]. \tag{2.5}$$

3. Structure of divergent series

In this section we study the structure of series which diverge slower than the logarithmic function but faster than an iterated logarithm. This is summarized in the following result.

Proposition 3.1 (Structure of divergent series)**.** *Let $\alpha \in (0, 1)$. Consider a vector $b = (b_1, \dots, b_m) \in \mathbb{R}^m$ ($m \geq 4$) that satisfies*

$$\|b\|_{1,\infty} \leq 1, \quad \|b\|_1 \gtrsim_\alpha (\log \log m)^2. \tag{3.1}$$

Then there exist a positive integer $l \leq \log m$ and a subset of indices $I_1 \subseteq [m]$ such that the following holds. Given a vector $\lambda = (\lambda_i)_{i \in I_1}$ such that $\|\lambda\|_1 \leq 1$, one can find a further subset $I_2 \subseteq I_1$ with the following two properties.

(i) (Regularity): the sizes $n_1 := |I_1|$ and $n_2 := |I_2|$ satisfy

$$2^{l/2} \leq \frac{m}{n_1} \leq \frac{m}{n_2} \leq \Big(\frac{m}{n_1} \Big)^{1+\alpha}.$$

(ii) (Largeness of coefficients):

$$|b_i| \geq \frac{1}{l n_1} \text{ for } i \in I_1;$$
$$|b_i| \geq \frac{1}{l n_2} \text{ and } |b_i| \geq 2|\lambda_i| \text{ for } i \in I_2.$$

Furthermore, we can make $l \geq C_\alpha \log \log m$ with arbitrarily large C_α by making the dependence on α implicit in the assumption (3.1) *sufficiently large.*

Remarks. 1. Proposition 3.1 is somewhat nontrivial even if one ignores the vector λ and the further subset I_2. In this simpler form the result was introduced informally in Section 2.1. The structure that we find is located in the coefficients b_i on the index set I_1. Note that the largeness condition (ii) for these coefficients is easy to prove if we disregard the regularity condition (i). Indeed, since $\|b\|_{1,\infty} \gtrsim (\log m)^{-1} \|b\|_1 \gg 1/\log m$, we can choose $l = \log m$ and obtain a set I_1 satisfying (ii) by the definition of the weak ℓ_2 norm. But the regularity condition (i) guarantees the smaller level

$$l \lesssim \log(m/n_1)$$

which will be crucial in our application to decoupling.

2. The freedom to choose λ in Proposition 3.1 ensures that the structure located in the set I_1 is in a sense hereditary; it can pass to subsets I_2. The domination of λ by b on I_2 will be crucial in the removal of the diagonal terms in our application to decoupling.

We now turn to the proof of Proposition 3.1. Heuristically, we will first find many (namely, l) sets I_1 on which the coefficients are large as in (ii), then choose one that satisfies the regularity condition (i). This regularization step will rely on the following elementary lemma.

Lemma 3.2 (Regularization)**.** *Let N be a positive integer. Consider a nonempty subset $J \subset [L]$ with size $l := |J|$. Then, for every $\alpha \in (0,1)$, there exist elements $j_1, j_2 \in J$ that satisfy the following two properties.*

(i) (Regularity):

$$l/2 \le j_1 \le j_2 \le (1+\alpha) j_1.$$

(ii) (Density):

$$|J \cap [j_1, j_2]| \gtrsim_\alpha \frac{l}{\log(2L/l)}.$$

Proof. We will find j_1, j_2 as some consecutive terms of the following geometric progression. Define $j^{(0)} \in J$ to be the (unique) element such that

$$|J \cap [1, j^{(0)}]| = \lceil l/2 \rceil, \quad \text{and let } j^{(k)} := (1+\alpha) j^{(k-1)}, \quad k = 1, 2, \dots.$$

We will only need to consider K terms of this progression, where $K := \min\{k : j^{(k)} \ge L\}$. Since $j^{(0)} \ge \lceil l/2 \rceil \ge l/2$, we have $j^{(k)} \ge (1+\alpha)^k j^{(0)} \ge (1+\alpha)^k l/2$. On the other hand, $j^{(K-1)} \le L$. It follows that $K \lesssim_\alpha \log(2L/l)$.

We claim that there exists a term $1 \le k \le K$ such that

$$|J \cap [j^{(k-1)}, j^{(k)}]| \ge \frac{l}{3K} \gtrsim_\alpha \frac{l}{\log(2L/l)}. \tag{3.2}$$

Indeed, otherwise we would have

$$l = |J| \le |J \cap [1, j^{(0)}]| + \sum_{k=1}^{K} |J \cap [j^{(k-1)}, j^{(k)}]| \le \left\lceil \frac{l}{2} \right\rceil + K \cdot \frac{l}{3K} < l,$$

which is impossible.

The terms $j_1 := j^{(k-1)}$ and $j_2 := j^{(k)}$ for which (3.2) holds clearly satisfy (i) and (ii) of the conclusion. By increasing j_1 and decreasing j_2 if necessary we can assume that $j_1, j_2 \in J$. This completes the proof. □

Proof of Proposition 3.1. We shall prove the following slightly stronger statement. Consider a sufficiently large number

$$K \gtrsim_\alpha \log\log m.$$

Assume that the vector b satisfies

$$\|b\|_{1,\infty} \le 1, \quad \|b\|_1 \gtrsim_\alpha K \log\log m.$$

We shall prove that the conclusion of the Proposition holds with (ii) replaced by:

$$|b_i| \ge \frac{K}{2ln_1} \text{ for } i \in I_1; \tag{3.3}$$

$$|b_i| \ge \frac{K}{2ln_2} \ge 2|\lambda_i| \text{ for } i \in I_2. \tag{3.4}$$

We will construct I_1 and I_2 in the following way. First we decompose the index set $[m]$ into blocks $\Omega_1, \dots, \Omega_L$ on which the coefficients b_i have similar magnitude; this is possible with $L \sim \log m$ blocks. Using the assumption $\|b\|_1 \gtrsim (\log\log m)^2$, one easily checks that many (at least $l \sim \log\log m$) of the blocks Ω_j have large contribution (at least $1/j$) to the sum $\|b\|_1 = \sum_i |b_i|$. We will only focus on such large blocks in the rest of the argument. At this point, the union of these blocks could be declared I_1. We indeed proceed this way, except we first use Regularization Lemma 3.2 on these blocks in order to obtain the required regularity property (ii). Finally, assume we are given coefficients $(\lambda_i)_{i\in I_1}$ with small sum $\sum |\lambda_i| \le 1$ as in the assumption. Since the coefficients b_i are large on I_1 by construction, the pigeonhole principle will yield (loosely speaking) a whole block of coefficients Ω_j where b_i will dominate as required, $|b_i| \ge 2|\lambda_i|$. We declare this block I_2 and complete the proof. Now we pass to the details of the argument.

Step 1: decomposition of $[m]$ into blocks. Without loss of generality,

$$\frac{1}{m} < b_i \le 1, \quad \lambda_i \ge 0 \quad \text{for all } i \in [m].$$

Indeed, we can clearly assume that $b_i \ge 0$ and $\lambda_i \ge 0$. The estimate $b_i \le 1$ follows from the assumption: $\|b\|_\infty \le \|b\|_{1,\infty} \le 1$. Furthermore, the contribution of the small coefficients $b_i \le 1/m$ to the norm $\|b\|_1$ is at most 1, while by the assumption $\|b\|_1 \gtrsim_\alpha K \log\log m \ge 2$. Hence we can ignore these small coefficients by replacing $[m]$ with the subset corresponding to the coefficients $b_i \ge 1/m$.

We decompose $[m]$ into disjoint subsets (which we call blocks) according to the magnitude of b_i, and we consider the contribution of each block Ω_j to the norm $\|b\|_1$:

$$\Omega_j := \left\{ i \in [m] : \ 2^{-j} < b_i \le 2^{-j+1} \right\}; \quad m_j := |\Omega_j|; \quad B_j := \sum_{i\in\Omega_j} b_i.$$

By our assumptions on b, there are at most $\log m$ nonempty blocks Ω_j. As $\|b\|_{1,\infty} \le 1$, Markov's inequality yields for all j that

$$m_j \le \sum_{k \le j} m_k = \big|\big\{i \in [m] : \ b_i > 2^{-j}\big\}\big| \le 2^j; \tag{3.5}$$

$$B_j \le m_j 2^{-j+1} \le 2. \tag{3.6}$$

Only the blocks with large contributions B_j will be of interest to us. Their number is

$$l := \max\big\{j \in [\log m] : \ B_j^* \ge K/j\big\};$$

and we let $l = 0$ if it happens that all $B_j < K/j$. We claim that there are many such blocks:

$$\frac{1}{5} K \log\log m \le l \le \log m. \tag{3.7}$$

Indeed, by the assumption and using (3.6) we can bound

$$K \log\log m \le \|b\|_1 = \sum_{j=1}^{\log m} B_j^* \le 2l + 0.6K \log\log m,$$

which yields (3.7).

Step 2: construction of the set I_1. As we said before, we are only interested in blocks Ω_j with large contributions B_j. We collect the indices of such blocks into the set

$$\bar{J} := \big\{j \in [\log m] : \ B_j \ge K/l\big\}.$$

Since the definition of l implies that $B_l^* \ge K/l$, we have $|\bar{J}| \ge l$. Then we can apply Regularization Lemma 3.2 to the set $\{\log m - j : \ j \in \bar{J}\} \subseteq [\log m]$. Thus we find two elements $j', j'' \in \bar{J}$ satisfying

$$l/2 \le \log m - j' \le \log m - j'' \le (1 + \alpha/2)(\log m - j'), \tag{3.8}$$

and such that the set

$$J := \bar{J} \cap [j'', j']$$

has size $|J| \gtrsim_\alpha l/\log\log m$. Since by our choice of K we can assume that $K \ge 8 \log\log m$, we obtain

$$|J| \ge \frac{8l}{K}. \tag{3.9}$$

We are going to show that the set

$$I_1 := \bigcup_{j \in J} \Omega_j$$

satisfies the conclusion of the Proposition.

Step 3: sizes of the coefficients b_i for $i \in I_1$. Let us fix $j \in J \subseteq \bar{J}$. From the definition of $\bar{J}$ we know that the contribution B_j is large: $B_j \geq K/l$. One consequence of this is a good estimate of the size m_j of the block Ω_j. Indeed, the above bound together with (3.6) this implies

$$\frac{K}{2l} 2^j \leq m_j \leq 2^j \quad \text{for } j \in J. \tag{3.10}$$

Another consequence of the lower bound on B_j is the required lower bound on the individual coefficients b_i. Indeed, by construction of Ω_j the coefficients b_i, $i \in \Omega_j$ are within the factor 2 from each other. It follows that

$$b_i \geq \frac{1}{2|\Omega_j|} \sum_{i \in \Omega_j} b_i = \frac{B_j}{2m_j} \geq \frac{K}{2lm_j} \quad \text{for } i \in \Omega_j,\ j \in J. \tag{3.11}$$

In particuar, since by construciton $\Omega_j \subseteq I_1$, we have $m_j \leq |I_1|$, which implies

$$b_i \geq \frac{K}{2l|I_1|} \quad \text{for } i \in I_1.$$

We have thus proved the required lower bound (3.3).

Step 4: Construction of the set I_2, and sizes of the coefficients b_i for $i \in I_2$. Now suppose we are given a vector $\lambda = (\lambda_i)_{i \in I_1}$ with $\|\lambda\|_1 \leq 1$. We will have to construct a subset $I_2 \subset I_1$ as in the conclusion, and we will do this as follows. Consider the contribution of the block Ω_j to the norm $\|\lambda\|_1$:

$$L_j := \sum_{i \in \Omega_j} \lambda_j, \quad j \in J.$$

On the one hand, the sum of all contributions is bounded as $\sum_{j \in J} L_j = \|\lambda\|_1 \leq 1$. On the other hand, there are many terms in this sum: $|J| \geq 8l/K$ as we know from (3.9). Therefore, by the pigeonhole principle some of the contributions must be small: there exists $j_0 \in J$ such that

$$L_{j_0} \leq K/8l.$$

This in turn implies via Markov's inequality that most of the coefficients λ_i for $i \in \Omega_{j_0}$ are small, and we shall declare these set of indices I_2. Specifically, since $L_{j_0} = \sum_{i \in \Omega_{j_0}} \lambda_j \leq K/8l$ and $|\Omega_{j_0}| = m_{j_0}$, using Markov's inequality we see that the set

$$I_2 := \big\{ i \in \Omega_{j_0} :\ \lambda_i \leq \frac{K}{4lm_{j_0}} \big\}$$

has cardinality

$$\frac{1}{2} m_{j_0} \leq |I_2| \leq m_{j_0}. \tag{3.12}$$

Moreover, using (3.11), we obtain

$$b_i \geq \frac{K}{2lm_{j_0}} \geq 2\lambda_i \quad \text{for } i \in I_2.$$

We have thus proved the required lower bound (3.4).

Step 5: the sizes of the sets I_1 and I_2. It remains to check the regularity property (i) of the conclusion of the Proposition. We bound

$$\begin{aligned} |I_1| &= \sum_{j \in J} m_j \quad \text{(by definition of } I_1\text{)} \\ &\leq \sum_{j \leq j'} m_j \quad \text{(by definition of } J\text{)} \\ &\leq 2^j. \quad \text{(by (3.5))} \end{aligned}$$

Therefore, using (3.8) we conclude that

$$\frac{m}{|I_1|} \geq 2^{\log m - j'} \geq 2^{l/2}. \tag{3.13}$$

We have thus proved the first inequality in (i) of the conclusion of the Proposition. Similarly, we bound

$$\begin{aligned} |I_2| &\geq \frac{1}{2} m_{j_0} \quad \text{(by (3.12))} \\ &\geq \frac{K}{4l} 2^{j_0} \quad \text{(by (3.10), and since } j_0 \in J\text{)} \\ &\geq \frac{K}{4l} 2^{j''}. \quad \text{(by definition of } J\text{, and since } j_0 \in J\text{)} \end{aligned}$$

Therefore

$$\begin{aligned} \frac{m}{|I_2|} &\leq \frac{4l}{K} 2^{\log m - j''} \\ &\leq \frac{8}{K} (\log m - j') 2^{(1+\alpha/2)(\log m - j')} \quad \text{(by (3.8))} \\ &\leq \frac{\alpha}{2} (\log m - j') 2^{(1+\alpha/2)(\log m - j')} \quad \text{(by the assumption on } K\text{)} \\ &\leq 2^{(1+\alpha)(\log m - j')} \\ &\leq \Big(\frac{m}{|I_1|} \Big)^{1+\alpha}. \quad \text{(by (3.13))} \end{aligned}$$

This completes the proof of (i) of the conclusion, and of the whole Proposition 3.1. □

4. Decoupling

In this section we develop a decoupling principle, which was informally introduced in Section 2.2. In contrast to other decoupling results used in probabilistic contexts, our decoupling principle is non-random. It is valid for arbitrary fixed vectors X_i which are almost pairwise orthogonal as in (2.5). An example of such vectors are random vectors, as we observed earlier in Lemma 2.4. Thus in this section we will consider vectors $X_1, \dots, X_m \in \mathbb{R}^n$ that satisfy the following almost pairwise orthogonality assumptions for some $r' \geq 1, K_1, K_2$:

$$\begin{aligned} &\|X_i\|_2 \leq K_1 \sqrt{n}; \\ &\frac{1}{|E|} \sum_{i \in E,\, i \neq k} \langle X_i, X_k \rangle^2 \leq K_2^4 \Big(\frac{N}{|E|} \Big)^{1/r'} n \quad \text{for all } E \subseteq [m],\ k \in [m]. \end{aligned} \tag{4.1}$$

In the earlier work [30] we developed a weaker decoupling principle, which was valid for the weak ℓ_2 norm instead of ℓ_2 norm. Let us recall this result first. Assume that for vectors X_i satisfying (4.1) with $r' = r$ one has

$$\sup_{x \in S^{n-1}} \|(\langle X_i, x \rangle)_{i=1}^m\|_{2,\infty}^2 \gtrsim_{r, K_1, K_2} n + \Big(\frac{N}{m} \Big)^{1/r} m.$$

Then the Decoupling Proposition 2.1 of [30] implies that there exist disjoint sets of indices $I, J \subseteq [m]$ such that $|J| \leq \delta |I|$, and there exists a vector $y \in S^{n-1} \cap \operatorname{span}(X_j)_{j \in J}$, such that

$$\langle X_i, y \rangle^2 \geq \Big(\frac{N}{|I|} \Big)^{1/r} \quad \text{for } i \in I.$$

Results of this type are best suited for applications to random independent vectors X_i. Indeed, the events that $\langle X_i, y \rangle^2$ is large are independent for $i \in I$ because y does not depend on $(X_i)_{i \in I}$. The probability of each such event is easy to bound using the moment assumptions (2.2).

In our new decoupling principle, we replace the weak ℓ_2 norm by the ℓ_2 norm at the cost of an iterated logarithmic factor and a slight loss of the exponent. Our result will thus operate in the regime where the weak ℓ_2 norm is small while ℓ_2 norm is large. We summarize this in the following proposition.

Proposition 4.1 (Decoupling)**.** *Let $n \geq 1$ and $4 \leq m \leq N$ be integers, and let $1 \leq r < \min(r', r'')$ and $\delta \in (0, 1)$. Consider vectors $X_1, \dots, X_m \in \mathbb{R}^n$ which satisfy the weak orthonormality conditions* (4.1) *for some K_1, K_2. Assume that*

for some $K_3 \geq \max(K_1, K_2)$ one has

$$\sup_{x \in S^{n-1}} \|(\langle X_i, x\rangle)_{i=1}^m\|_{2,\infty}^2 \leq K_3^2 \Big[n + \Big(\frac{N}{m}\Big)^{1/r} m \Big], \tag{4.2}$$

$$\Big\| \sum_{i=1}^m X_i \otimes X_i \Big\| = \sup_{x \in S^{n-1}} \sum_{i=1}^m \langle X_i, x\rangle^2 \gtrsim_{r,r',r'',\delta} K_3^2 (\log\log m)^2 \Big[n + \Big(\frac{N}{m}\Big)^{1/r} m \Big]. \tag{4.3}$$

Then there exist nonempty disjoint sets of indices $I, J \subseteq [m]$ such that $|J| \leq \delta |I|$, and there exists a vector $y \in S^{n-1} \cap \operatorname{span}(X_j)_{j \in J}$, such that

$$\langle X_i, y\rangle^2 \geq K_3^2 \Big(\frac{N}{|I|}\Big)^{1/r''} \quad \text{for } i \in I.$$

The proof of the Decoupling Proposition 4.1 will use Proposition 3.1 in order to locate the structure of the large coefficients $\langle X_i, x\rangle$. The following elementary lemma will be used in the argument.

Lemma 4.2. *Consider a vector $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$ which satisfies*

$$\|\lambda\|_1 \leq 1, \quad \|\lambda\|_\infty \leq 1/K$$

for some integer K. Then, for every real numbers $(a_1, \dots, a_n) \in \mathbb{R}^n$ one has

$$\sum_{i=1}^n \lambda_i a_i \leq \frac{1}{K} \sum_{i=1}^K a_i^*.$$

Proof. It is easy to check that each extreme point of the convex set

$$\Lambda := \{\lambda \in \mathbb{R}^n : \ \|\lambda\|_1 \leq 1, \ \|\lambda\|_\infty \leq 1/K\}$$

has exactly K nonzero coefficients which are equal to $\pm 1/K$. Evaluating the linear form $\sum \lambda_i a_i$ on these extreme points, we obtain

$$\sup_{\lambda \in \Lambda} \sum_{i=1}^n \lambda_i a_i = \sup_{\lambda \in \operatorname{ext}(\Lambda)} \sum_{i=1}^n \lambda_i a_i = \frac{1}{K} \sum_{i=1}^K a_i^*.$$

The proof is complete. □

Proof of Decoupling Proposition 4.1. By replacing X_i with X_i/K_3 we can assume without loss of generality that $K_1 = K_2 = K_3 = 1$. By perturbing the vectors X_i slightly we may also assume that X_i are all different.

Step 1: separation and the structure of coefficients. Suppose the assumptions of the Proposition hold, and let us choose a vector $x \in S^{n-1}$ which attains the supremum in (4.3). We denote

$$a_i := \langle X_i, x\rangle \quad \text{for } i \in [m],$$

and without loss of generality we may assume that $a_i \neq 0$. We also denote

$$\bar{n} := n + \Big(\frac{N}{m}\Big)^{1/r} m.$$

We choose parameter $\alpha = \alpha(r, r', r'', \delta) \in (0,1)$ sufficiently small; its choice will become clear later on in the argument. At this point, we may assume that

$$\|a\|_{2,\infty}^2 \leq \bar{n}, \quad \|a\|_2^2 \gtrsim_\alpha (\log\log m)^2 \bar{n}.$$

We can use Structure Proposition 3.1 to locate the structure in the coefficients a_i. To this end, we apply this result for $b_i = a_i^2/\bar{n}$ and obtain a number $l \leq \log m$ and a subset of indices $I_1 \subseteq [m]$. We can also assume that l is sufficiently large – larger than an arbitrary quantity which depends on α.

Since a vector $x \in S^{n-1}$ satisfies $\langle X_i/a_i, x\rangle = 1$ for all $i \in I_1$ (in fact for all $i \in [m]$), a separation argument for the convex hull $K := \mathrm{conv}(X_i/a_i)_{i\in I_1}$ yields the existence of a vector $\bar{x} \in \mathrm{conv}(K \cup 0)$ that satisfies

$$\|\bar{x}\|_2 = 1, \quad \langle X_i/a_i, \bar{x}\rangle \geq 1 \quad \text{for } i \in I_1. \tag{4.4}$$

We express $\bar{x}$ as a convex combination

$$\bar{x} = \sum_{i\in I_1} \lambda_i X_i/a_i \quad \text{for some } \lambda_i \geq 0, \quad \sum_{i\in I_1} \lambda_i \leq 1. \tag{4.5}$$

We then read the conclusion of Structure Proposition 3.1 as follows. There exists a futher subset of indices $I_2 \subseteq I_1$ such that the sizes $n_1 := |I_1|$ and $n_2 := |I_2|$ are regular in the sense that

$$2^{l/2} \leq \frac{m}{n_1} \leq \frac{m}{n_2} \leq \Big(\frac{m}{n_1}\Big)^{1+\alpha}, \tag{4.6}$$

and the coefficients on I_1 and I_2 are large:

$$a_i^2 \geq \frac{\bar{n}}{l n_1} \qquad \text{for } i \in I_1, \tag{4.7}$$

$$a_i^2 \geq \frac{\bar{n}}{l n_2} \text{ and } a_i^2 \geq 2\lambda_i \bar{n} \qquad \text{for } i \in I_2. \tag{4.8}$$

Furthermore, we can make l sufficiently large depending on α, say $l \geq 100/\alpha^2$.

Step 2: random selection. We will reduce the number of terms n_1 in the sum (4.5) defining $\bar{x}$ using random selection, trying to bring this number down to about n_2. As is usual in dealing with sums of independent random variables, we will need to ensure that all summands $\lambda_i X_i/a_i$ have controlled magnitudes. To this end, we have $\|X_i\|_2 \leq \sqrt{n}$ by the assumption, and we can bound $1/a_i$ through (4.7). Finally, we have an a priori bound $\lambda_i \leq 1$ on the coefficients of the convex combination. However, the latter bound will turn out to be too weak, and we will need $\lambda_i \lesssim_\alpha 1/n_2$ instead. To make this happen,

instead of the sets I_1 and I_2 we will be working on their large subsets I_1' and I_2' defines as

$$I_1' := \big\{ i \in I_1 : \ \lambda_i \le \frac{C_\alpha}{n_2} \big\}, \quad I_2' := \big\{ i \in I_2 : \ \lambda_i \le \frac{C_\alpha}{n_2} \big\}$$

where C_α is a sufficiently large quantity whose value we will choose later. By Markov's inequality, this incurs almost no loss of coefficients:

$$|I_1 \setminus I_1'| \le \frac{n_2}{C_\alpha}, \quad |I_2 \setminus I_2'| \le \frac{n_2}{C_\alpha}. \tag{4.9}$$

We will perform a random selection on I_1 using B. Maurey's empirical method [21]. Guided by the representation (4.5) of $\bar{x}$ as a convex combination, we will treat λ_i as probabilities, thus introducing a random vector V with distribution

$$\mathbb{P}\{V = X_i/a_i\} = \lambda_i, \quad i \in I_1'.$$

On the remainder of the probability space, we assign V zero value: $\mathbb{P}\{V = 0\} = 1 - \sum_{i \in I_1'} \lambda_i$. Consider independent copies $V_1, V_2, \dots$ of V. We are not going to do a random selection on the set $I_1 \setminus I_1'$ where the coefficients λ_i may be out of control, so we just add its the contribution by defining independent random vectors

$$Y_j := V_j + \sum_{i \in I_1 \setminus I_1'} \lambda_i X_i a_i, \quad j = 1, 2, \dots$$

Finally, for $C_\alpha' := C_\alpha/\alpha$, we consider the average of about n_2 such vectors:

$$\bar{y} := \frac{C_\alpha'}{n_2} \sum_{j=1}^{n_2/C_\alpha'} Y_j. \tag{4.10}$$

We would like to think of $\bar{y}$ as a random version of the vector $\bar{x}$. This is certainly true in expectation:

$$\mathbb{E}\bar{y} = \mathbb{E}Y_1 = \sum_{i \in I_1} \lambda_i X_i/a_i = \bar{x}.$$

Also, like $\bar{x}$, the random vector $\bar{y}$ is a convex combination of terms X_i/a_i (now even with equal weights). The advantage of $\bar{y}$ over $\bar{x}$ is that it is a convex combination of much fewer terms, as $n_2/C_\alpha' \ll n_2 \le n_1$. In the next two steps, we will check that $\bar{y}$ is similar to $\bar{x}$ in the sense that its norm is also well bounded above, and at least $\sim n_2$ of the inner products $\langle X_i/a_i, \bar{y} \rangle$ are still nicely bounded below.

Step 3: control of the norm. By independence, we have

$$\begin{aligned}\mathbb{E}\|\bar{y}-\bar{x}\|_2^2 &= \mathbb{E}\Big\|\frac{C'_\alpha}{n_2}\sum_{j=1}^{n_2/C'_\alpha}(Y_j-\mathbb{E}Y_j)\Big\|_2^2 = \Big(\frac{C'_\alpha}{n_2}\Big)^2\sum_{j=1}^{n_2/C'_\alpha}\mathbb{E}\|Y_j-\mathbb{E}Y_j\|_2^2\\ &\lesssim_\alpha \frac{1}{n_2}\mathbb{E}\|Y_1-\mathbb{E}Y_1\|_2^2 = \frac{1}{n_2}\mathbb{E}\|V-\mathbb{E}V\|_2^2 \le \frac{4}{n_2}\mathbb{E}\|V\|_2^2\\ &= \frac{4}{n_2}\sum_{i\in I'_1}\lambda_i\|X_i\|_2^2/a_i^2 \le \frac{4}{n_2}\max_{i\in I_1}\|X_i\|_2^2/a_i^2,\end{aligned}$$

where the last inequality follows because $I'_1 \subseteq I_1$ and $\sum_{i\in I_1}\lambda_i \le 1$.

Since $\bar{n} \ge n$, (4.7) gives us the lower bound

$$a_i^2 \ge \frac{n}{ln_1} \quad \text{for } i \in I_1.$$

Together with the assumption $\|X_i\|_2^2 \le n$, this implies that

$$\mathbb{E}\|\bar{y}-\bar{x}\|_2^2 \lesssim_\alpha \frac{1}{n_2}\cdot\frac{n}{n/ln_1} \le \frac{ln_1}{n_2}.$$

Since $\|\bar{x}\|_2^2 = 1 \le ln_1/n_2$, we conclude that with probability at least 0.9, one has

$$\|\bar{y}\|_2^2 \lesssim_\alpha \frac{ln_1}{n_2}. \tag{4.11}$$

Step 4: removal of the diagonal term. We know from (4.4) that $\langle X_i/a_i, \bar{x}\rangle \ge 1$ for many terms X_i. We would like to replace $\bar{x}$ by its random version $\bar{y}$, establishing a lower bound $\langle X_k/a_k, \bar{y}\rangle \ge 1$ for many terms X_k. But at the same time, our main goal is decoupling, in which we would need to make the random vector $\bar{y}$ independent of those terms X_k. To make this possible, we will first remove from the sum (4.10) defining $\bar{y}$ the "diagonal" term containing X_k, and we call the resulting vector $\bar{y}^{(k)}$.

To make this precise, let us fix $k \in I'_2 \subseteq I'_1 \subseteq I_1$. We consider independent random vectors

$$V_j^{(k)} := V_j\mathbf{1}_{\{V_j\ne X_k/a_k\}}, \quad Y_j^{(k)} := V_j^{(k)} + \sum_{i\in I_1\setminus I'_1}\lambda_i X_i a_i, \quad j=1,2,\dots$$

Note that

$$\mathbb{P}\{Y_j^{(k)} \ne Y_j\} = \mathbb{P}\{V_j^{(k)} \ne V_j\} = \mathbb{P}\{V_j = X_k/a_k\} = \lambda_k. \tag{4.12}$$

Similarly to the definition (4.10) of $\bar{y}$, we define

$$\bar{y}^{(k)} := \frac{C'_\alpha}{n_2}\sum_{j=1}^{n_2/C'_\alpha} Y_j^{(k)}.$$

Then

$$\mathbb{E}\bar{y}^{(k)} = \mathbb{E}Y_1^{(k)} = \mathbb{E}Y_1 - \lambda_k X_k/a_k = \bar{x} - \lambda_k X_k/a_k. \tag{4.13}$$

As we said before, we would like to show that the random variable

$$Z_k := \langle X_k/a_k, \bar{y}^{(k)} \rangle$$

is bounded below by a constant with high probability. First, we will estimate its mean

$$\mathbb{E}Z_k = \langle X_k/a_k, \bar{x} \rangle - \lambda_k \|X_k\|_2^2/a_k^2.$$

To estimate the terms in the right hand side, note that $\langle X_i/a_i, \bar{x} \rangle \geq 1$ by (4.4) and $\|X_k\|_2^2 \leq n$ by the assumption. Now is the crucial point when we use that a_i^2 dominate λ_i as in the second inequality in (4.8). This allows us to bound the "diagonal" term as

$$\lambda_k \|X_k\|_2^2/a_k^2 \leq n/2\bar{n} \leq n/2n = 1/2.$$

As a result, we have

$$\mathbb{E}Z_k \geq 1 - 1/2 = 1/2. \tag{4.14}$$

Step 5: control of the inner products. We would need a stronger statement than (4.14) – that Z_k is bounded below not only in expectation but also with high probability. We will get this immediately by Chebyshev's inequality if we can upper bound the variance of Z_k. In a way similar to Step 3, we estimate

$$\begin{aligned} \operatorname{Var} Z_k = \mathbb{E}(Z_k - \mathbb{E}Z_k)^2 &= \mathbb{E}\Big\langle X_k/a_k, \frac{C'_\alpha}{n_2} \sum_{j=1}^{n_2/C'_\alpha} (Y_j^{(k)} - \mathbb{E}Y_j^{(k)}) \Big\rangle^2 \\ &\lesssim_\alpha \frac{1}{n_2} \mathbb{E}\langle X_k/a_k, V_1^{(k)} \rangle^2 = \frac{1}{n_2} \sum_{i \in I'_1,\, i \neq k} \lambda_i \langle X_k/a_k, X_i/a_i \rangle^2. \end{aligned} \tag{4.15}$$

Now we need to estimate the various terms in the right hand side of (4.15).

We start with the estimate on the inner products, collecting them into

$$S := \sum_{i \in I'_1,\, i \neq k} \lambda_i \langle X_k, X_i \rangle^2 = \sum_{i \in I'_1} \lambda_i p_{ki} \quad \text{where } p_{ki} = \langle X_k, X_i \rangle^2 \mathbf{1}_{\{k \neq i\}}.$$

Recall that, by the construction of λ_i and of $I'_1 \subseteq I$, we have $\sum_{i \in I'_1} \lambda_i \leq 1$ and $\lambda_i \leq C_\alpha/n_2$ for $i \in I'_1$. We use Lemma 4.2 on order statistics to obtain the bound

$$S \leq \frac{C_\alpha}{n_2} \sum_{i=1}^{n_2/C_\alpha} (p_k)_i^* = \frac{C_\alpha}{n_2} \max_{\substack{E \subseteq [m] \\ |E| = n_2/C_\alpha}} \sum_{i \in E,\, i \neq k} \langle X_i, X_k \rangle^2.$$

Finally, we use our weak orthonormality assumption (4.1) to conclude that

$$S \lesssim_\alpha \Big(\frac{N}{n_2}\Big)^{1/r'} n.$$

To complete the bound on the variance of Z_k in (4.15) it remains to obtain some good lower bounds on a_k and a_i. Since $k \in I_2' \subseteq I_2$, (4.8) yields

$$a_k^2 \geq \frac{\bar{n}}{l n_2} \geq \frac{n}{l n_2}.$$

Similarly we can bound the coefficients a_i in (4.15): using (4.7) we have $a_i^2 \geq \bar{n}/l n_1$ since since $i \in I_1' \subseteq I$. But here we will not simply replace $\bar{n}$ by n, as we shall try to use a_i^2 to offset the term $(N/n_2)^{1/r'}$ in the estimate on S. To this end, we note that $\bar{n} \geq (N/m)^{1/r} m \geq (N/n_1)^{1/r} n_1$ because $m \geq n_1$. Therefore, using the last inequality in (4.6) and that $N \geq m$, we have

$$\frac{\bar{n}}{n_1} \geq \Big(\frac{N}{n_1}\Big)^{1/r} \geq \Big(\frac{N}{n_2}\Big)^{\frac{1}{(1+\alpha)r}}. \tag{4.16}$$

Using this, we obtain a good lower bound

$$a_i^2 \geq \frac{\bar{n}}{l n_1} \geq \frac{1}{l}\Big(\frac{N}{n_2}\Big)^{\frac{1}{(1+\alpha)r}}, \quad \text{for } i \in I_1'.$$

Combining the estimates on S, a_k and a_i, we conclude our lower bound (4.15) on the variance of Z_k as follows:

$$\begin{aligned}
\operatorname{Var} Z_k &\lesssim_\alpha \frac{1}{n_2} \cdot \frac{l n_2}{n} \cdot l\Big(\frac{n_2}{N}\Big)^{\frac{1}{(1+\alpha)r}} \cdot \Big(\frac{N}{n_2}\Big)^{1/r'} n \\
&\leq l^2 \Big(\frac{n_2}{N}\Big)^{\alpha} \quad \text{(by choosing } \alpha \text{ small enough depending on } r,\, r') \\
&\leq l^2 \Big(\frac{n_2}{m}\Big)^{\alpha} \leq l^2 2^{-\alpha l/2} \quad \text{(by (4.6))} \\
&\leq \alpha/16. \quad \text{(since } l \text{ is large enough depending on } \alpha)
\end{aligned}$$

Combining this with the lower bound (4.14) on the expectation, we conclude by Chebyshev's inequality the desired estimate

$$\mathbb{P}\big\{ Z_k = \langle X_k/a_k, \bar{y}^{(k)} \rangle \geq \frac{1}{4} \big\} \geq 1 - \alpha \quad \text{for } k \in I_2'. \tag{4.17}$$

Step 6: decoupling. We are nearing the completion of the proof. Let us consider the good events

$$\mathcal{E}_k := \big\{ \langle X_k/a_k, \bar{y} \rangle \geq \frac{1}{4} \text{ and } \bar{y} = \bar{y}^{(k)} \big\} \quad \text{for } k \in I_2'.$$

To show that each $\mathcal{E}_k$ occurs with high probability, we note that by definition of $\bar{y}$ and $\bar{y}^{(k)}$ one has

$$\begin{aligned}\mathbb{P}\{\bar{y} \neq \bar{y}^{(k)}\} &\le \mathbb{P}\{Y_j \neq Y_j^{(k)} \text{ for some } j \in [n_2/C'_\alpha]\} \\ &\le \sum_{j=1}^{n_2/C'_\alpha} \mathbb{P}\{Y_j \neq Y_j^{(k)}\} = \frac{n_2}{C'_\alpha} \cdot \lambda_k \quad \text{(by (4.12))} \\ &\le \frac{n_2}{C'_\alpha} \cdot \frac{C_\alpha}{n_2} \quad \text{(by definition of } I'_2) \\ &= \alpha. \quad \text{(as we chose } C'_\alpha = C_\alpha/\alpha)\end{aligned}$$

From this and using (4.17) we conclude that

$$\mathbb{P}\{\mathcal{E}_k\} \ge 1 - \mathbb{P}\Big\{\langle X_k/a_k, \bar{y}^{(k)}\rangle < \frac{1}{4}\Big\} - \mathbb{P}\{\bar{y} \neq \bar{y}^{(k)}\} \ge 1 - 2\alpha \quad \text{for } k \in I'_2.$$

An application of Fubini theorem yields that with probability at least 0.9, at least $(1-20\alpha)|I'_2|$ of the events $\mathcal{E}_k$ hold simultaneously. More accurately, with probability at least 0.9 the following event occurs, which we denote by $\mathcal{E}$. There exists a subset $I \subseteq I'_2$ of size $|I| \ge (1-20\alpha)|I'_2|$ such that $\mathcal{E}_k$ holds for all $k \in I$. Note that using (4.9) and choosing C_α sufficiently large we have

$$(1-21\alpha)n_2 \le |I| \le n_2. \tag{4.18}$$

Recall that the norm bound (4.11) also holds with high probability 0.9. Hence with probability at least 0.8, both $\mathcal{E}$ and this norm bound holds. Let us fix a realization of our random variables for which this happens. Then, first of all, by definition of $\mathcal{E}_k$ we have

$$\langle X_k/a_k, \bar{y}\rangle \ge \frac{1}{4} \quad \text{for } k \in I. \tag{4.19}$$

Next, we are going to observe that $\bar{y}$ lies in the span of few vectors X_i. Indeed, by construction $\bar{y}^{(k)}$ lies in the span of the vectors $Y_j^{(k)}$ for $j \in [n_2/C'_\alpha]$. Each such $Y_j^{(k)}$ by construction lies in the span of the vectors X_i, $i \in I_1 \setminus I'_1$ and of one vector $V_j^{(k)}$. Finally, each such vector $V_j^{(k)}$, again by construction, is either equal zero or V_j, which in turn equals X_{i_0} for some $i_0 \neq k$. Since $\mathcal{E}$ holds, we have $\bar{y} = \bar{y}^{(k)}$ for all $k \in I$. This implies that there exists a subset $I_0 \subseteq [m]$ (consisting of the indices i_0 as above) with the following properties. Firstly, I_0 does not contain any of indices $k \in I$; in other words I_0 is disjoint from I. Secondly, this set is small: $|I_0| \le n_2/C'_\alpha$. Thirdly, $\bar{y}$ lies in the span of X_i, $i \in I_0 \cup (I_1 \setminus I'_1)$. We claim that this set of indices,

$$J := I_0 \cup (I_1 \setminus I'_1)$$

satisfies the conclusion of the Proposition.

Since I and I_0 are disjoint and $I \subseteq I_2' \subseteq I_1'$, it follows that I and J are disjoint as required. Moreover, by (4.9) and by choosing C_α, C_α' sufficiently large we have

$$|J| \le |I_0| + |I_1 \setminus I_1'| \le \frac{n_2}{C_\alpha'} + \frac{n_2}{C_\alpha} \le \alpha n_2.$$

When we combine this with (4.18) and choose α sufficiently small depending on δ, we achieve

$$|J| \le \delta |I|$$

as required. Finally, we claim that the normalized vector

$$y := \frac{\bar{y}}{\|\bar{y}\|_2}$$

satisfies the conclusion of the Proposition. Indeed, we already noted that $\bar{y} \in \mathrm{span}(X_j)_{j\in J}$, as required. Next, for each $k \in I \subseteq I_2' \subseteq I_2$ we have

$$\begin{aligned}
\langle X_k, y\rangle^2 &\ge \frac{a_k^2}{16\|\bar{y}\|_2^2} \quad \text{(by (4.19))} \\
&\gtrsim_\alpha \frac{\bar{n}}{l n_2} \cdot \frac{n_2}{l n_1} \quad \text{(by (4.8) and (4.11))} \\
&= \frac{\bar{n}}{l^2 n_1} \ge \frac{1}{l^2}\Big(\frac{N}{n_2}\Big)^{\frac{1}{(1+\alpha)r}}. \quad \text{(by (4.16))}
\end{aligned}$$

We can get rid of l^2 in this estimate using the bound

$$\begin{aligned}
\Big(\frac{N}{n_2}\Big)^{\frac{\alpha}{(1+\alpha)r}} &\ge \Big(\frac{m}{n_2}\Big)^{\frac{\alpha}{(1+\alpha)r}} \ge 2^{\frac{\alpha l}{2(1+\alpha)r}} \quad \text{(by (4.6))} \\
&\ge 2^{\frac{\alpha l}{3r}} \ge 2^{\alpha^2 l} \quad \text{(choosing } \alpha \text{ small enough depending on } r) \\
&\ge l^2. \quad \text{(since } l \text{ is large enough depending on } \alpha)
\end{aligned}$$

Therefore

$$\langle X_k, y\rangle^2 \ge \Big(\frac{N}{n_2}\Big)^{\frac{1-\alpha}{(1+\alpha)r}} \ge \Big(\frac{N}{n_2}\Big)^{1/r''} \quad \text{for } k \in I$$

where the last inequality follows by choosing α sufficiently small depending on r, r''. This completes the proof of Decoupling Proposition 4.1. □

5. Norms of random matrices with independent columns

In this section we apply our decoupling principle, Proposition 4.1, to estimate norms of random matrices with independent columns. As we said, a simple truncation argument of J. Bourgain [6] reduces the approximation problem for covariance matrices to bounding the norm of the random matrix $\sum_{i\in E} X_i \otimes X_i$ uniformly over index sets E. The following result gives such an estimate for random vectors X_i with finite moment assumptions.

Theorem 5.1. *Let $1 \le n \le N$ be integers, and let $4 < p < q$ and $t \ge 1$. Consider independent random vectors $X_1, \dots, X_N$ in $\mathbb{R}^n$ ($1 \le n \le N$) which satisfy the moment assumptions* (2.2)*. Then with probability at least $1 - Ct^{-0.9q}$, for every index set $E \subseteq [N]$, $|E| \ge 4$, one has*

$$\Big\| \sum_{i \in E} X_i \otimes X_i \Big\| \lesssim_{p,q,K,L} t^2 (\log\log |E|)^2 \Big[n + \Big(\frac{N}{|E|} \Big)^{4/p} |E| \Big].$$

We can state Theorem 5.1 in terms of random matrices with independent columns.

Corollary 5.2. *Let $1 \le n \le N$ be integers, and let $4 < p < q$ and $t \ge 1$. Consider the $n \times N$ random matrix A whose columns are independent random vectors $X_1, \dots, X_N$ in $\mathbb{R}^n$ which satisfy* (2.2)*. Then with probability at least $1 - Ct^{-0.9q}$ one has*

$$\|A\| \lesssim_{p,q,K,L} t \log\log N \cdot (\sqrt{n} + \sqrt{N}).$$

Moreover, with the same probability all $n \times m$ submatrices B of A simultaneously satisfy the following for all $4 \le m \le N$:

$$\|B\| \lesssim_{p,q,K,L} t \log\log m \cdot \Big[\sqrt{n} + \Big(\frac{N}{m} \Big)^{2/p} \sqrt{m} \Big].$$

Proof of Theorem 5.1. By replacing X_i with $X_i / \max(K, L)$ we can assume without loss of generality that $K = L = 1$. As we said, the argument will be based on Decoupling Proposition 4.1. Its assumptions follow from known results. Indeed, the pairwise almost orthogonality of the vectors X_i follows from Lemma 2.4, which yields (2.5) with probability at least $1 - Ct^{-q}$. Also, the required bound on the weak ℓ_2 norm follows from Theorem 2.3, which gives with probability at least $1 - Ct^{-0.9q}$ that

$$\sup_{x \in S^{n-1}} \|(\langle X_i, x \rangle)_{i \in I}\|_{2,\infty}^2 \lesssim_q t^2 \Big[n + \Big(\frac{N}{|I|} \Big)^{4/q} |I| \Big] \quad \text{for } I \subseteq [N]. \tag{5.1}$$

Consider the event $\mathcal{E}$ that both required bounds (2.5) and (5.1) hold.

Let $\mathcal{E}_0$ denote the event in the conclusion of the Theorem. It remains to prove that $\mathbb{P}(\mathcal{E}_0^c \text{ and } \mathcal{E})$ is small. To this end, assume that $\mathcal{E}$ holds but $\mathcal{E}_0$ does not. Then there exists an index set $E \subset [N]$ whose size we denote by $m := |E|$, and which satisfies

$$\Big\| \sum_{i \in E} X_i \otimes X_i \Big\| = \sup_{x \in S^{n-1}} \sum_{i \in E} \langle X_i, x \rangle^2 \gtrsim_{p,q} t^2 (\log\log m)^2 \Big[n + \Big(\frac{N}{m} \Big)^{4/p} m \Big].$$

Recalling (2.5) and (5.1) we see that the assumptions of Decoupling Proposition 4.1 hold for $1/r = 4/p$, $1/r' = 4/q$, $r'' = r'$, $K_1 = K = 1$, $K_2 = C_q \sqrt{t}$

for suitably large C_q, $K_3 = \max(K_1, K_2, 100t^2)$, and for $\delta = \delta(p,q) > 0$ sufficiently small (to be chosen later). Applying Decoupling Proposition 4.1 we obtain disjoint index sets $I, J \subseteq E \subseteq [N]$ with sizes

$$|I| =: s, \quad |J| \le \delta s,$$

and a vector $y \in S^{n-1} \cap \operatorname{span}(X_j)_{j\in J}$ such that

$$\langle X_i, y\rangle^2 \ge 100t^2 \Big(\frac{N}{s}\Big)^{1/r''} \quad \text{for } i \in I. \tag{5.2}$$

We will need to discretize the set of possible vectors y. Let

$$\varepsilon := \Big(\frac{\delta s}{N}\Big)^5$$

and consider an ε-net $\mathcal{N}_J$ of the sphere $S^{n-1} \cap \operatorname{span}(X_j)_{j\in J}$. As in known by a volumetric argument (see e.g. [19] Lemma 2.6), one can choose such a net with cardinality

$$|\mathcal{N}_J| \le (3/\varepsilon)^{|J|} \le \Big(\frac{2N}{\delta s}\Big)^{5\delta s}.$$

We can assume that the random set $\mathcal{N}_J$ depends only on the number ε, the set J and the random variables $(X_j)_{j\in J}$. Given a vector y as we have found above, we can approximate it with some vector $y_0 \in \mathcal{N}_J$ so that $\|y - y_0\|_2 \le \varepsilon$. By (5.1) we have

$$\|(\langle X_i, y - y_0\rangle)_{i\in I}\|_{2,\infty}^2 \lesssim_q \varepsilon^2 t^2 \Big[n + \Big(\frac{N}{s}\Big)^{1/r'} s\Big].$$

This implies that all but at most δs indices i in I satisfy the inequality

$$\langle X_i, y - y_0\rangle^2 \lesssim_q \frac{\varepsilon^2 t^2}{\delta s}\Big[n + \Big(\frac{N}{s}\Big)^{1/r'} s\Big]. \tag{5.3}$$

Let us denote the set of these indices by $I_0 \subseteq I$. The bound in (5.3) can be simplified as

$$\frac{\varepsilon^2}{\delta s}\Big[n + \Big(\frac{N}{s}\Big)^{1/r'} s\Big] \le 2\delta.$$

Indeed, this estimate follows from the two bounds

$$\frac{\varepsilon^2}{\delta s} \cdot n \le \Big(\frac{\delta s}{N}\Big)^{10} \cdot \frac{n}{\delta s} \le \delta \quad (\text{because } n \le N);$$

$$\frac{\varepsilon^2}{\delta s} \cdot \Big(\frac{N}{s}\Big)^{1/r'} s \le \frac{1}{\delta}\Big(\frac{\delta s}{N}\Big)^{10}\Big(\frac{N}{s}\Big)^{1/r'} \le \delta. \quad (\text{because } \delta \le 1,\ r' \ge 1)$$

In particular, by choosing $\delta = \delta(q) > 0$ sufficiently small, (5.3) implies

$$|\langle X_i, y - y_0\rangle| \le t \quad \text{for } i \in I_0.$$

Together with (5.2) this yields by triangle inequality that

$$|\langle X_i, y_0\rangle| \ge 10t\Big(\frac{N}{s}\Big)^{1/2r''} - t \ge 9t\Big(\frac{N}{s}\Big)^{1/2r''} \quad \text{for } i \in I_0.$$

Summarizing, we have shown that the event $\{\mathcal{E}_0^c \text{ and } \mathcal{E}\}$ implies the following event: there exists a number $s \le N$, disjoint index subsets $I_0, J \subseteq [N]$ with sizes $|I_0| \ge (1-\delta)s$, $|J| \le \delta s$, and a vector $y_0 \in N_J$ such that

$$|\langle X_i, y_0\rangle| \ge 9t\Big(\frac{N}{s}\Big)^{1/2r''} \quad \text{for } i \in I_0.$$

It will now be easy to estimate the probability of this event. First of all, for each fixed vector $y_0 \in S^{n-1}$ and each index i, the moment assumptions (2.2) imply via Markov's inequality that

$$\mathbb{P}\big\{|\langle X_i, y_0\rangle| \ge 9t\Big(\frac{N}{s}\Big)^{1/2r''}\big\} \le \frac{1}{(9t)^q}\Big(\frac{N}{s}\Big)^{-q/2r''} \le \frac{1}{9t^q}\Big(\frac{N}{s}\Big)^{-2}$$

where the last line follows from our choice of q and r''. By independence, for each fixed vector $y_0 \in S^{n-1}$ and a fixed index set $I_0 \subseteq [N]$ of size $|I_0| \ge (1-\delta)s$ we have

$$\begin{aligned}\mathbb{P}\big\{|\langle X_i, y_0\rangle| \ge 9\Big(\frac{N}{s}\Big)^{1/2r''} \text{ for } i \in I_0\big\} &\le \Big[\frac{1}{9t^q}\Big(\frac{N}{s}\Big)^{-2}\Big]^{|I_0|} \\ &\le 9^{-(1-\delta)s} t^{-(1-\delta)q}\Big(\frac{N}{s}\Big)^{-2(1-\delta)s}. \qquad (5.4)\end{aligned}$$

Then we bound the probability of event $\{\mathcal{E}_0^c \text{ and } \mathcal{E}\}$ by taking the union bound over all s, I_0, J as above, conditioning on the random variables $(X_j)_{j\in J}$ (which fixes the ε-net $\mathcal{N}_J$), taking the union bound over the choice of $y_0 \in \mathcal{N}_J$, and finally evaluating the probability for using (5.4). This way we obtain via Stirling's approximation of the binomial coefficients that

$$\begin{aligned}\mathbb{P}\{\mathcal{E}_0^c \text{ and } \mathcal{E}\} &\le \sum_{s=1}^{N} \binom{N}{|I_0|}\binom{N}{|J|} |\mathcal{N}_J|\; 9^{-(1-\delta)s} t^{-(1-\delta)q}\Big(\frac{N}{s}\Big)^{-2(1-\delta)s} \\ &\le t^{-(1-\delta)q} \sum_{s=1}^{N} \Big(\frac{eN}{(1-\delta)s}\Big)^{(1-\delta)s}\Big(\frac{eN}{\delta s}\Big)^{\delta s}\Big(\frac{2N}{\delta s}\Big)^{5\delta s}\; 9^{-(1-\delta)s}\Big(\frac{N}{s}\Big)^{-2(1-\delta)s} \\ &\le t^{-0.9q} \sum_{s=1}^{N} \Big(\frac{N}{2s}\Big)^{s} \quad \text{(by choosing } \delta > 0 \text{ small enough)} \\ &\le t^{-0.9q}.\end{aligned}$$

It follows that

$$\mathbb{P}\{\mathcal{E}_0^c\} \le \mathbb{P}\{\mathcal{E}_0^c \text{ and } \mathcal{E}\} + \mathbb{P}\{\mathcal{E}^c\} \le t^{-0.9q} + Ct^{-q} + Ct^{-0.9q} \lesssim t^{-0.9q}.$$

This completes the proof of Theorem 5.1. □

6. Approximating covariance matrices

In this final section, we deduce our main result on the approximation of covariance matrices for random vectors with finite moments.

Theorem 6.1. *Consider independent random vectors $X_1, \dots, X_N$ in $\mathbb{R}^n$, $4 \le n \le N$, which satisfy moment assumptions (2.2) for some $q > 4$ and some K, L. Then for every $\delta > 0$ with probability at least $1 - \delta$ one has*

$$\Big\| \frac{1}{N} \sum_{i=1}^N X_i \otimes X_i - \mathbb{E} X_i \otimes X_i \Big\| \lesssim_{q,K,L,\delta} (\log\log n)^2 \Big(\frac{n}{N}\Big)^{\frac{1}{2}-\frac{2}{q}}. \tag{6.1}$$

In our proof of Theorem 6.1, we can clearly assume that $K = L = 1$ in the moment assumptions (2.2) by rescaling the vectors X_i. So in the rest of this section we suppose X_i are such random vectors.

For a level $B > 0$ and a vector $x \in S^{n-1}$, we consider the (random) index set of large coefficients

$$E_B = E_B(x) := \{i \in [N] : \ |\langle X_i, x\rangle| \ge B\}.$$

Lemma 6.2 (Large coefficients)**.** *Let $t \ge 1$. With probability at least $1 - Ct^{0.9q}$, one has*

$$|E_B| \lesssim_q n/B^2 + N(t/B)^{q/2} \quad \text{for } B > 0.$$

Proof. This estimate follows from Theorem 2.3. By definition of the set E_B and the weak ℓ_2 norm, we obtain with the required probability that

$$B^2 |E_B| \le \|(\langle X_i, x\rangle)_{i \in E_B}\|_{2,\infty}^2 \lesssim_q n + t^2 \Big(\frac{N}{|E_B|}\Big)^{4/q} |E_B|.$$

Solving for $|E_B|$ we obtain the bound as in the conclusion. □

Proof of Theorem 6.1. The truncation argument described in [1] in the beginning of proof of Proposition 4.3 reduces the problem to estimating the contribution to the sum of large coefficients. Denote

$$E = \Big\| \frac{1}{N} \sum_{i=1}^N X_i \otimes X_i - \mathbb{E} X_i \otimes X_i \Big\| = \sup_{x \in S^{n-1}} \Big| \frac{1}{N} \sum_{i=1}^N \langle X_i, x\rangle^2 - \mathbb{E}\langle X_i, x\rangle^2 \Big|.$$

The truncation argument yields that for every $B \ge 1$, one has with probability at least $1 - \delta/3$ that

$$\begin{aligned} E &\lesssim_{q,\delta} B\sqrt{\frac{n}{N}} + \sup_{x \in S^{n-1}} \frac{1}{N} \sum_{i \in E_B} \langle X_i, x\rangle^2 + \sup_{x \in S^{n-1}} \frac{1}{N} \mathbb{E} \sum_{i \in E_B} \langle X_i, x\rangle^2 \\ &=: I_1 + I_2 + I_3. \end{aligned} \tag{6.2}$$

We choose the value of the level

$$B = \Big(\frac{N}{n}\Big)^{2/q}$$

so that, using Lemma 6.2, with probability at least $1 - \delta/3$ we have

$$|E_B| \lesssim_{q,\delta} n. \tag{6.3}$$

It remains to estimate the right hand side of (6.2) using (6.3).

First, we clearly have

$$I_1 = \Big(\frac{n}{N}\Big)^{\frac{1}{2}-\frac{2}{q}}.$$

An estimate of I_2 follows from Theorem 5.1 for some $p = p(q) \in (4, q)$ to be determined later. Note that enlarging E_B can only make I_2 and I_3 larger. So without loss of generality we can assume that $|E_B| \geq 4$ as required in Theorem 5.1. This way, we obtain with probability at least $1 - \delta/3$ that

$$\begin{aligned} I_2 &\lesssim_{q,\delta} \frac{1}{N}(\log\log|E_B|)^2\Big[n + \Big(\frac{N}{|E_B|}\Big)^{4/p}|E_B|\Big] \\ &\lesssim_{q,\delta} (\log\log n)^2\Big[\frac{n}{N} + \Big(\frac{n}{N}\Big)^{1-\frac{4}{p}}\Big]. \qquad \text{(by (6.3))} \end{aligned}$$

Finally, to estimate I_3 let us fix x and consider the random variable $Z_i = |\langle X_i, x\rangle|$. Since $\mathbb{E}Z_i^q \leq 1$, an application of Hölder's and Markov's inequalities yield

$$\mathbb{E}Z_i^2\mathbf{1}_{\{Z_i \geq B\}} \leq (\mathbb{E}Z_i^q)^{2/q}(\mathbb{P}(Z_i \geq B))^{1-2/q} \leq B^{2-q} \lesssim_{q,\delta} \Big(\frac{n}{N}\Big)^{2-\frac{4}{q}}.$$

Therefore

$$I_3 = \sup_{x\in S^{n-1}} \frac{1}{N}\sum_{i=1}^N \mathbb{E}Z_i^2\mathbf{1}_{\{Z_i\geq B\}} \lesssim_{q,\delta} \Big(\frac{n}{N}\Big)^{2-\frac{4}{q}}.$$

Since we are free to choose $p = p(q)$ in the interval $(4, q)$, we choose the middle of the interval, $p = (q+4)/2$. Returning to (6.2) we conclude that

$$E \lesssim_{q,\delta} (\log\log n)^2\Big(\frac{n}{N}\Big)^{\frac{1}{2}-\frac{2}{q}}.$$

This completes the proof of Theorem 6.1. □

References

[1] R. Adamczak, A. Litvak, A. Pajor, N. Tomczak-Jaegermann, *Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles*, J. Amer. Math. Soc. 234 (2010), 535–561.

[2] R. Adamczak, A. Litvak, A. Pajor, N. Tomczak-Jaegermann, *Sharp bounds on the rate of convergence of the empirical covariance matrix*, preprint.

[3] G. Aubrun, *Sampling convex bodies: a random matrix approach*, Proc. Amer. Math. Soc. 135 (2007), 1293–1303

[4] K. Ball, *An elementary introduction to modern convex geometry*. Flavors of geometry, 1–58, Math. Sci. Res. Inst. Publ., 31, Cambridge Univ. Press, Cambridge, 1997

[5] Z. D. Bai and Y. Q. Yin, *Limit of the smallest eigenvalue of a large dimensional sample covariance matrix*, Ann. Probab. 21 (1993), 1275–1294

[6] J. Bourgain, *Random points in isotropic convex sets*. Convex geometric analysis (Berkeley, CA, 1996), 53–58, Math. Sci. Res. Inst. Publ., 34, Cambridge Univ. Press, Cambridge, 1999

[7] J. Dahmen, D. Keysers, M. Pitz, H. Ney, *Structured Covariance Matrices for Statistical Image Object Recognition*. In: 22nd Symposium of the German Association for Pattern Recognition. Springer, 2000, pp. 99–106

[8] A. Giannopoulos, M. Hartzoulaki, A. Tsolomitis, *Random points in isotropic unconditional convex bodies*, J. London Math. Soc. 72 (2005), 779–798.

[9] A. Giannopoulos, V. D. Milman, *Concentration property on probability spaces*, Adv. Math. 156 (2000), 77–106.

[10] A. Giannopoulos, V. D. Milman, *Euclidean structure in finite dimensional normed spaces*. Handbook of the geometry of Banach spaces, Vol. I, 707–779, North-Holland, Amsterdam, 2001.

[11] E. Kovačević, A. Chebira, *An Introduction to Frames*, Foundations and Trends in Signal Processing, Now Publishers, 2008

[12] R. Kannan, L. Lovász, M. Simonovits, *Random walks and $O^*(n^5)$ volume algorithm for convex bodies*, Random Structures and Algorithms 2 (1997), 1–50

[13] R. Kannan, L. Rademacher, *Optimization of a convex program with a polynomial perturbation*, Operations Research Letters 37(2009), 384–386

[14] H. Krim and M. Viberg, *Two decades of array signal processing research: The parametric approach*, IEEE Signal Process. Mag. 13 (1996), 67–94

[15] R. Latala, *Some estimates of norms of random matrices*, Proc. Amer. Math. Soc. 133 (2005), 1273–1282

[16] O. Ledoit, M. Wolf, *Improved estimation of the covariance matrix of stock returns with an application to portfolio selection*, Journal of Empirical Finance 10 (2003), 603–621

[17] M. Ledoux, M. Talagrand, *Probability in Banach spaces. Isoperimetry and processes*. Ergebnisse der Mathematik und ihrer Grenzgebiete (3), 23 Springer-Verlag, Berlin, 1991

[18] E. Levina, R. Vershynin, *Partial estimation of covariance matrices*, preprint

[19] V. Milman, G. Schechtman, *Asymptotic theory of finite dimensional normed spaces*. Lecture Notes in Mathematics, 1200. Springer-Verlag, Berlin, 1986

[20] G. Paouris, *Concentration of mass on convex bodies*, Geom. Funct. Anal. 16 (2006), 1021–1049

[21] G. Pisier, *Remarques sur un resultat non publie de B. Maurey*, Seminar on Functional Analysis, 1980-1981, Exp. no V, 13 pp., Ecole Polytech, Palaiseau, 1981.

[22] A. J. Rothman, E. Levina, J. Zhu, *Generalized Thresholding of Large Covariance Matrices*, J. Amer. Statist. Assoc. (Theory and Methods) 104 (2009), 177–186

[23] M. Rudelson, *Random vectors in the isotropic position*, J. Funct. Anal. 164 (1999), 60–72

[24] M. Rudelson, R. Vershynin, *Non-asymptotic theory of random matrices: extreme singular values*, Proceedings of the International Congress of Mathematicians, Hyderabad, India, 2010, to appear.

[25] J. Schäfer, K. Strimmer, *A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics*, Statistical Applications in Genetics and Molecular Biology 4 (2005), Article 32

[26] Y. Seginer, *The expected norm of random matrices*, Combin. Probab. Comput. 9 (2000), 149–166

[27] A. W. Van der Vaart, J. A. Wellner, *Weak convergence and Empirical Processes.* Springer-Verlag, 1996.

[28] R. Vershynin, *Frame expansions with erasures: an approach through the non-commutative operator theory*, Applied and Computational Harmonic Analysis 18 (2005), 167–176

[29] R. Vershynin, *Spectral norm of products of random and deterministic matrices*, Probability Theory and Related Fields, to appear.

[30] R. Vershynin, *Approximating the moments of marginals of high dimensional distributions*, Annals of Probability, to appear.

Department of Mathematics, University of Michigan, 530 Church Street, Ann Arbor, MI 48109, U.S.A.

E-mail address: romanv@umich.edu