APPROXIMATING THE MOMENTS OF MARGINALS OF HIGH DIMENSIONAL DISTRIBUTIONS

ROMAN VERSHYNIN

ABSTRACT. For probability distributions on $\mathbb{R}^n$, we study the optimal sample size $N = N(n,p)$ that suffices to uniformly approximate the p-th moments of all one-dimensional marginals. Under the assumption that the support of the distribution lies in the Euclidean ball of radius $O(\sqrt{n})$ and the marginals have bounded $4p$ moments, we obtain the optimal bound $N = O(n^{p/2})$ for $p > 2$. This bound goes in the direction of bridging the two recent results: a theorem of Guedon and Rudelson [6] which has an extra logarithmic factor in the sample size, and a recent result of Adamczak, Litvak, Pajor and Tomczak-Jaegermann [1] which requires stronger subexponential moment assumptions.

1. INTRODUCTION

We study the following problem: how well can one approximate one-dimensional marginals of a distribution on $\mathbb{R}^n$ by sampling? Consider a random vector X in $\mathbb{R}^n$, and suppose we would like to compute the p-th moments of the marginals $\langle X, x\rangle$ for all $x \in \mathbb{R}^n$. To this end, we sample N independent copies $X_1, \dots, X_N$ of X, compute the empirical moment from that sample, and we hope that it gives a good approximation of the actual moment:

$$\sup_{x \in S^{n-1}} \Big| \frac{1}{N} \sum_{i=1}^{N} |\langle X_i, x\rangle|^p - \mathbb{E}|\langle X, x\rangle|^p \Big| \le \varepsilon. \tag{1.1}$$

Indeed, by the law of large numbers this quantity converges to zero as $N \to \infty$. To understand quantitative nature of this convergence one would like to estimate the optimal sample complexity $N = N(n,p,\varepsilon)$ for which (1.1) holds with high probability. For $p = 2$ this problem is equivalent to approximating the covariance matrix of X by a sample covariance matrix, and it was studied in [7, 3, 11, 8, 2, 1]. For $p \neq 2$, the problem was also studied in [4, 5, 6, 9, 1].

A well known *lower* bound for the sample complexity is $N \gtrsim n$ for $1 \le p \le 2$ and $N \gtrsim n^{p/2}$ for $p \ge 2$. O. Guedon and M. Rudelson [6] proved the upper bound $N = O(n^{p/2} \log n)$ for $p \ge 2$ under quite weak and natural moment

Date: November 2, 2009.
Partially supported by NSF grant FRG DMS 0918623.

assumptions:

$$\|X\|_2 = O(\sqrt{n}) \text{ a.s.}, \quad (\mathbb{E}|\langle X, x\rangle|^p)^{1/p} = O(1) \text{ for all } x \in S^{n-1}. \tag{1.2}$$

The logarithmic term can not be in general removed from the sample complexity, as is seen from the example of a random vector X uniformly distributed in a set of n orthogonal vectors of Euclidean norm $\sqrt{n}$. On the other hand, R. Adamczak et al. [1] recently managed to remove the logarithmic term for random vectors X uniformly distributed in an isotropic convex body K in $\mathbb{R}^n$, showing that for such distributions one has $N = O(n)$ for $1 \le p \le 2$ and $N = O(n^{p/2})$ for $p \ge 2$. Their result actually holds for all random vectors X that satisfy the sub-exponential moment assumptions

$$\|X\|_2 = O(\sqrt{n}) \text{ a.s.}, \quad (\mathbb{E}|\langle X, x\rangle|^q)^{1/q} = O(q) \text{ for all } q \ge 1 \text{ and } x \in S^{n-1}. \tag{1.3}$$

At this moment there is no complete understanding which distributions on $\mathbb{R}^n$ require logarithmic oversampling and which do not. There is clearly a gap between the minimal moment assumptions (1.2) of [6] and the subexponential assumptions (1.3) of [1]. The present note makes a step toward closing the gap in this picture. We show that a version of the result of Adamczak et al. [1] for $p \ne 2$ holds *under finite moment assumptions*; for example, the logarithmic oversampling is not needed if we replace p by $4p$ in the minimal moment assumptions (1.2). We shall thus consider independent random vectors X_i in $\mathbb{R}^n$ that satisfy

$$\|X_i\|_2 \le K\sqrt{n} \text{ a.s.}, \quad (\mathbb{E}|\langle X_i, x\rangle|^q)^{1/q} \le L \text{ for all } x \in S^{n-1}. \tag{1.4}$$

Theorem 1.1. *Let $p > 2$, $\varepsilon > 0$ and $\delta > 0$. Consider independent random vectors X_i in $\mathbb{R}^n$ which satisfy (1.4) for $q = 4p$. Let $N \ge Cn^{p/2}$ where C is a suitably large quantity that depends only on $K, L, p, \varepsilon, \delta$. Then with probability at least $1 - \delta$ one has*

$$\sup_{x \in S^{n-1}} \Big| \frac{1}{N} \sum_{i=1}^N |\langle X_i, x\rangle|^p - \mathbb{E}|\langle X_i, x\rangle|^p \Big| \le \varepsilon. \tag{1.5}$$

Remark. 1. A more elaborate version of this result is Theorem 4.2 below. One can get more information on the probability in question using general concentration of measure results as is done in [6]. One can also modify the argument to deduce a version of this result "with high probability" in spirit of [1], i.e. with probability converging to 1 (at polynomial rate) as $n \to \infty$.

2. A standard modification of the argument (as in [1]) gives an optimal result in the range $1 \le p < 2$. Namely, if the random vectors satisfy (1.4) for some $q \ge 4p$, $q > 4$, then the conclusion (1.5) holds for $N \ge C_{K,L,p,q,\varepsilon,\delta} n$.

3. The method of the present note does not seem to work for $p = 2$; this important case will hopefully be addressed in a sequel paper.

Theorem 1.1 yields sharp bounds on the norms of random operators $\ell_2 \to \ell_p$. The following corollary is a version of a result of [1] proved there under the stronger sub-exponential moment assumptions (1.3).

Corollary 1.2. *Let $p > 2$ and $\delta > 0$. Consider independent random vectors X_i in $\mathbb{R}^n$ which satisfy (1.4) for $q = 4p$. Then the $N \times n$ random matrix A with rows $X_1, \dots, X_N$ satisfies with probability at least $1 - \delta$ that*

$$\mathbb{E}\|A\|_{\ell_2 \to \ell_p} \le C(n^{1/2} + N^{1/p})$$

where C depends only on K, L, p, δ.

J. Bourgain [3] first demonstrated that proving deviation estimates like (1.5) reduces to bounding the contribution to the sum of the large coefficients – those for which $|\langle X_i, x\rangle| > B$ for a suitably large fixed level B. Such reduction is used in some of the later approaches to the problem [5, 1] as well as in the present note. However, after this reduction we use a different route. Suppose for some vector $x \in S^{n-1}$ there are $s = s(B)$ large coefficients as above. The new ingredient of this note is a decoupling argument which is formalized in Proposition 2.1. It transports the vector x into the linear span of at most $0.01s$ of these X_i, while approximately retaining the largeness of the coefficients, $|\langle X_i, x\rangle| > B/4$. Let us condition on these $0.01s$ random vectors X_i. On the one hand, we have reduced the "complexity" of the problem – our x now lies in a fixed $0.01s$-dimensional subspace, which has an $\frac{1}{2}$-net in the Euclidean metric of cardinality $e^{0.02s}$. On the other hand, the inequality $|\langle X_i, x\rangle| > B$ holds for the remaining $0.99s$ vectors X_i of which x is independent; by (1.4) and Markov's inequality this happens with probability $(L/B)^{qs}$. Choosing the level B suitably large so that $(L/B)^{qs} \ll e^{-0.02s}$ allows us to take the union bound over the net, and therefore to control the contribution of the large coefficients.

In Section 2 we develop the decoupling argument. We use it to control the contribution of the large coefficients in Section 3. This is formalized in Theorem 3.1 where we estimate the norm of a random matrix A with rows X_i in the operator norm $\ell_2 \to \ell_{2,\infty}$, and also in Lemma 4.1. In Section 4, we deduce in a standard way the main results of this note – Theorem 1.1 on approximating the moments of random vectors and Corollary 1.2 on the norms of random matrices $\ell_2 \to \ell_p$.

In what follows, C and c will stand for positive absolute constants (suitably chosen); quantities that depend only on the parameters in question such as K, L, p, q will be denoted $C_{K,L,p,q}$.

2. Decoupling

Proposition 2.1 (Decoupling)**.** *Let $X_1, \dots, X_s$ be vectors in $\mathbb{R}^n$ which satisfy the following conditions for some K_1, K_2:*

$$\|X_k\| \le K_1\sqrt{n}, \qquad \frac{1}{s}\sum_{i\le s,\, i\ne k} \langle X_i, X_k\rangle^2 \le K_2^4 n, \qquad k = 1, \dots, s. \tag{2.1}$$

Let $\delta \in (0,1)$ and let $B \ge C\delta^{-3/2}K_1$, $M \ge C\delta^{-1/2}K_2^2/K_1$. Assume that there exists $x \in S^{n-1}$ such that

$$\langle X_i, x\rangle \ge B\sqrt{n/s} + M, \quad i = 1, \dots, s.$$

Then there exist a subset $I \subseteq \{1, \dots, s\}$, $|I| \ge (1-\delta)s$, and a vector $y \in S^{n-1} \cap \operatorname{span}(X_i)_{i\in I^c}$ such that

$$\langle X_i, y\rangle \ge \frac{1}{4}(B\sqrt{n/s} + M), \quad i \in I.$$

Proof. Without loss of generality we may assume that $\delta > 0$ is smaller than a suitably chosen absolute constant.

Step 1: random selection. Denote $a := B\sqrt{n/s} + M$. Then

$$\langle X_i/a, x\rangle \ge 1, \quad i = 1, \dots, s.$$

The convex hull $K := \operatorname{conv}(X_i/a)_{i\in s}$ is separated in $\mathbb{R}^n$ from the origin by the hyperplane $\{u : \langle u, x\rangle = 1\}$. By a separation argument, one can find a vector $\bar{x} \in \operatorname{conv}(K \cup 0)$, $\|\bar{x}\|_2 = 1$ and such that

$$\langle X_i/a, \bar{x}\rangle \ge 1, \quad i = 1, \dots, s. \tag{2.2}$$

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

$$\bar{x} = \sum_{i=1}^s \lambda_i X_i/a \quad \text{for some } \lambda_i \ge 0, \quad \sum_{i=1}^s \lambda_i \le 1.$$

By Markov's inequality, the set $E := \{i \le s : \ \lambda_i \le 1/\delta s\}$ has cardinality $|E| \ge (1-\delta)s$. We will perform a random selection on E. Let $\delta_1, \dots, \delta_s$ be i.i.d. selectors, i.e. independent $\{0,1\}$ valued random variables with $\mathbb{E}\delta_i = \delta$. We define the random vector

$$\bar{y} := \sum_{i\in E} \delta_i\lambda_i X_i/a + \sum_{i\in E^c} \delta\lambda_i X_i/a; \quad \text{then } \mathbb{E}\bar{y} = \delta\bar{x}.$$

Step 2: control of the norm and inner products. By independence and by definitions of a, E and B we have

$$\mathbb{E}\|\bar{y}-\delta\bar{x}\|_2^2 = \mathbb{E}\Big\| \sum_{i\in E} (\delta_i-\delta)\lambda_i X_i/a \Big\|_2^2 = \sum_{i\in E} \mathbb{E}(\delta_i-\delta)^2 \cdot \lambda_i^2 \frac{\|X_i\|_2^2}{a^2}$$
$$\le s\delta \cdot (1/\delta s)^2 \frac{K_1^2 n}{(B\sqrt{n/s})^2} \le \frac{K_1^2}{\delta B^2} \le 0.1\delta^2.$$

By Chebyshev's inequality, we have with probability at least 0.9 that

$$\|\bar{y}\|_2 \le \|\bar{y}-\delta\bar{x}\|_2 + \|\delta\bar{x}\|_2 \le 2\delta. \tag{2.3}$$

Now fix $k \in E$. By definition of $\bar{y}$ and by (2.2), we have

$$\mathbb{E}\langle X_k/a, \bar{y}\rangle = \delta\langle X_k/a, \bar{x}\rangle \ge \delta. \tag{2.4}$$

We will need a similar bound with high probability rather than in expectation. More accurately, we would like to bound below

$$p_k := \mathbb{P}\big(\langle (1-\delta_k) X_k/a, \bar{y}\rangle \ge \delta/2\big).$$

Consider the random vector $\bar{y}^{(k)}$ obtained by removing from the sum defining $\bar{y}$ the term corresponding to X_k:

$$\bar{y}^{(k)} := \sum_{i\in E,\, i\neq k} \delta_i\lambda_i X_i/a + \sum_{i\in E^c} \delta\lambda_i X_i/a = \bar{y} - \delta_k\lambda_k X_k/a.$$

Then $\bar{y}^{(k)}$ is independent of δ_k, which gives

$$p_k = \mathbb{P}(\delta_k = 0)\cdot \mathbb{P}\big(\langle X_k/a, \bar{y}^{(k)}\rangle \ge \delta/2\big).$$

By definitions of a, E and B we can bound the contribution of the removed term as

$$\langle X_k/a, \lambda_k X_k/a\rangle = \lambda_k \frac{\|X_k\|_2^2}{a^2} \le (1/\delta s)\frac{K_1^2 n}{(B\sqrt{n/s})^2} = \frac{K_1^2}{\delta B^2} \le 0.1\delta^2.$$

Then the random variable $Z_k := \langle X_k/a, \bar{y}^{(k)}\rangle$ satisfies by (2.4) that

$$\mathbb{E}Z_k = \langle X_k/a, \bar{y}\rangle - \mathbb{E}\langle X_k/a, \delta_k\lambda_k X_k/a\rangle \ge \delta - 0.1\delta^3 \ge 0.9\delta.$$

Similarly to the argument in the beginning of Step 2, we obtain

$$\operatorname{Var} Z_k = \mathbb{E}(Z_k - \mathbb{E}Z_k)^2 = \mathbb{E}\Big\langle X_k/a, \sum_{i\in E,\, i\neq k} (\delta_i-\delta)\lambda_i X_i/a \Big\rangle^2$$
$$= \sum_{i\in E,\, i\neq k} \mathbb{E}(\delta_i-\delta)^2 \cdot \lambda_i^2 \frac{\langle X_k, X_i\rangle^2}{a^4}$$
$$\le \delta\cdot\Big(\frac{1}{\delta s}\Big)^2 \frac{K_2^4 ns}{(B\sqrt{n/s}+M)^4} \le \frac{K_2^4}{\delta B^2 M^2} \le 0.01\delta^3.$$

By Chebyshev's inequality, we conclude that $\mathbb{P}(Z_k \geq \delta/2) \geq 1-\delta$. We have shown that

$$p_k \geq (1-\delta)(1-\delta) \geq 1-2\delta.$$

Step 3: decoupling. Denoting by $\mathcal{E}_k$ the event $\langle (1-\delta_k)X_k/a, \bar{y}\rangle \geq \delta/2$, we have shown that $\mathbb{P}(\mathcal{E}_k) \geq 1-2\delta$ for all $k \in E$. An application of Fubini theorem yields that with probability at least 0.9, at least $(1-20\delta)|E| \geq (1-22\delta)s$ of the events $\mathcal{E}_k$ hold simultaneously. More accurately, with probability at least 0.9 the following event occurs: there exists a subset $I \subset E$, $|I| \geq (1-22\delta)s$, such that $\mathcal{E}_k$ holds for all $k \in I$.

Assume the latter event occurs. By definition of $\mathcal{E}_k$ we clearly have $\delta_k = 0$ whenever $\mathcal{E}_k$ holds. Hence by definition of $\bar{y}$ one has $\bar{y} \in \mathrm{span}(X_i)_{i \in I^c}$. Also, by definition of $\mathcal{E}_k$, one has

$$\langle X_k/a, \bar{y}\rangle \geq \delta/2, \quad k \in I.$$

Once we set $y := \bar{y}/\|\bar{y}\|_2$, this and (2.3) complete the proof. □

3. Norms of random operators $\ell_2 \to \ell_{2,\infty}$

Recall that the weak ℓ_2-norm $\|x\|_{2,\infty}$ of a vector $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$ is defined as the minimal number M for which the non-increasing rearrangement (x_k^*) of the sequence $(|x_k|)$ satisfies $x_k^* \leq M k^{-1/2}$, $k = 1, \ldots, N$. One can easily check that $c_p\|x\|_p \leq \|x\|_{2,\infty} \leq \|x\|_2$ for all $p > 2$.

Although $\|\cdot\|_{2,\infty}$ is not a norm, for linear operators $A : \mathbb{R}^n \to \mathbb{R}^N$ we will be interested in the "norm" $\|A\|_{\ell_2 \to \ell_{2,\infty}}$ defined as the minimal number M such that $\|Ax\|_{2,\infty} \leq M\|x\|_2$ for all $x \in \mathbb{R}^n$.

Theorem 3.1. *Consider random vectors $X_1, \ldots, X_N$ which satisfy* (1.4) *for some $q > 4$. Then, for every $t \geq 1$, the random matrix A whose rows are X_i satisfies the following with probability at least $1 - Ct^{-0.9q}$. For every index set $I \subseteq \{1, \ldots, N\}$, one has*

$$\|P_I A\|_{\ell_2 \to \ell_{2,\infty}} \leq C_{K,L,q}\Big[\sqrt{n} + t\sqrt{|I|}\,\big(N/|I|\big)^{2/q}\Big]$$

where P_I is the coordinate projection in $\mathbb{R}^N$ onto $\mathbb{R}^I$. In particular, one has

$$\|A\|_{\ell_2 \to \ell_{2,\infty}} \leq C_{K,L,q}(\sqrt{n} + t\sqrt{N}).$$

Remarks. 1. This theorem is a finite-moment variant of Corollary 3.7 of [1], where a similar result is proved under the stronger sub-exponential moment assumptions (1.3). The latter is in turn a strengthening of an inequlity of Bourgain [3] that has some unnecessary logarithmic terms.

2. The conclusion of Theorem 3.1 can be equivalently stated as follows. For every subset $I \subseteq \{1,\dots,N\}$, one has

$$\Big\| \sum_{i \in I} X_i \Big\|_2 \le C_{K,L,q} \Big[\sqrt{n|I|} + t|I| \, \big(N/|I|\big)^{2/q} \Big].$$

3. It seems possible that Theorem 3.1 holds for the spectral norm $\|A\|_{\ell_2 \to \ell_2}$. This would imply that Theorem 3.1 holds in the important case $p = 2$.

The proof of Theorem 3.1 is based on the Decoupling Proposition 2.1. So we will first need to verify the assumptions on the vectors (2.1).

Lemma 3.2. *Let $Z_1, \dots Z_N \ge 0$ be independent random vectors which satisfy $\mathbb{E} Z_i^q \le B^q$ for some $q > 4$ and some B. Consider the non-increasing rearrangement (Z_i^*) of (Z_i). Then, for every $t \ge 1$, one has with probability at least $1 - Ct^{-q}/N$ that*

$$Z_i^* \le 2tB \, (N/i)^{2/q}, \quad i = 1, \dots, N. \tag{3.1}$$

In particular, (3.1) *implies*

$$\frac{1}{s} \sum_{i=1}^{s} (Z_i^*)^2 \le C_q t^2 B^2 \, (N/s)^{4/q}, \quad s = 1, \dots, N.$$

Proof. By homogeneity, we can assume that $B = 1$. Then by Markov's inequality we have $\mathbb{P}(Z_j > u) \le u^{-q}$ for every $j \le N$ and $u > 0$. Now, if $Z_i^* > u$ then there exists a set $J \subseteq \{1, \dots, N\}$, $|J| = i$ such that $|Z_j| > u$ for all $j \in J$. Taking union bound over possible choices of the subsets J, using independence and Stirling's approximation, we obtain for all $i \le N/2$:

$$\mathbb{P}(Z_i^* > u) \le \binom{N}{i} \big[\max_{j \le N} \mathbb{P}(Z_j > u) \big]^i \le \binom{N}{i} u^{-qi} \le (eu^{-q}N/i)^i.$$

Choosing $u = t(eN/i)^{2/q}$ with a sufficiently large absolute constant C, we conclude that $\mathbb{P}(Z_i^* > t) \le (t^{-q} i/eN)^i$. Then

$$\mathbb{P}(\exists i \le N/2 : \; Z_i^* > t) \le \sum_{i=1}^{N/2} (t^{-q} i/eN)^i \le Ct^{-q}/N.$$

This easily implies the first part of the lemma. The second part follows by summation. □

Lemma 3.3. *Consider independent random vectors $X_1, \dots, X_N$ in $\mathbb{R}^n$ which satisfy* (1.4) *for some $q > 4$. Then for every $t \ge 1$ the following holds with probability at least $1 - Ct^{-q}$. For every subset $E \subseteq \{1, \dots, N\}$ and every $k \le N$ one has*

$$\frac{1}{|E|} \sum_{i \in E, \, i \ne k} \langle X_i, X_k \rangle^2 \le C_q t^2 K^2 L^2 \, \big(N/|E|\big)^{4/q} n.$$

Proof. We fix $k \le N$ and apply Lemma 3.2 to the random variables $Z_i^{(k)} := |\langle X_i, X_k\rangle|$, $i \le N$, $i \ne k$. By assumptions (1.4), we have $\mathbb{E} Z_i^q \le (KL\sqrt{n})^q$. Then with probability at least $1 - Ct^{-q}/N$, we have

$$\frac{1}{s}\sum_{i=1}^{s}((Z^{(k)})_i^*)^2 \le C_q t^2 K^2 L^2\,(N/s)^{4/q} n, \quad s = 1, \dots, N.$$

Taking union bound over $k \le N$ completes the proof. □

Proof of Theorem 3.1. By homogeneity, we can assume that $L = 1$. Denote by $\mathcal{E}$ the event in the conclusion of Lemma 3.3. If $\mathcal{E}$ holds, then the assumptions (2.1) of Decoupling Proposition 2.1 are satisfied for every s and every subset $(X_i)_{i\in E}$, $E \subseteq \{1,\dots,N\}$, $|E| = s$, and with parameters $K_1 = K$, $K_2^4 = C_q t^2 K^2 (N/s)^{4/q}$. So, in view of application of Decoupling Proposition 2.1, we consider $B = B(K,\delta)$ and $M_1 = M_1(q,\delta,t)$ defined as

$$B := C\delta^{-3/2}K_1, \qquad M = C\delta^{-1}K_2^2/K_1 = C_q'\delta^{-1}t\,(N/s)^{2/q} =: M_1\,(N/s)^{2/q}.$$

Note that we can assume that $C_q' \ge 12$, which we will use later.

We will now need a convenient interpretation the conclusion of the theorem. Given $x \in S^{n-1}$, we denote by $|\langle X_{\pi(i)}, x\rangle|$ a non-increasing rearrangement of the sequence $|\langle X_i, x\rangle|$, $i = 1,\dots,N$. Denote by D the minimal number such that for every $x \in S^{n-1}$ and every $s \le N/2$ one has

$$|\langle X_{\pi(s)}, x\rangle| \le R_s := D\big[B\sqrt{n/s} + M_1\,(N/s)^{2/q}\big].$$

Since $q \ge 4$, the quantity $\sqrt{s}(N/s)^{2/q}$ is non-decreasing in s. Therefore one has for every $s \le m \le N/2$:

$$|\langle X_{\pi(s)}, x\rangle| \le D\big[B\sqrt{n/s} + M_1\sqrt{m/s}\,(N/m)^{2/q}\big].$$

It follows that for every $x \in S^{n-1}$, every $m \le N/2$, and every index set $I \subseteq \{1,\dots,N\}$, $|I| = m$, one has

$$\|(\langle X_i, x\rangle)_{i\in I}\|_{2,\infty} \le D\Big[B\sqrt{n} + M_1\sqrt{m}\,(N/m)^{2/q}\Big].$$

If we are able to show that $D \le 1$ with the high probability as required in Theorem 3.1, this would clearly complete the proof.

Since the event $\mathcal{E}$ holds with probability at least $1 - Ct^{-q}$, it suffices to show that the event $\{\mathcal{E}$ holds and $D > 1\}$ occurs with probability at most $Ct^{-0.99q}$. Let us assume that the latter event does occur. By definition of D, one can find a number $s \le N$, a subset $E \subseteq \{1,\dots,N\}$, $|E| = s$, and a vector $x \in S^{n-1}$ such that

$$|\langle X_i, x\rangle| \ge R_s, \quad i \in E.$$

By the definition of R_s, B, M above, Decoupling Proposition 2.1 can be applied for $(X_i)_{i\in E}$, and it yields the following. There exists a decomposition $E = I \cup J$

into disjoint sets I and J such that $|I| \geq (1-\delta)s$, $|J| \leq \delta s$, and there exists a vector $y \in \mathrm{span}(X_j)_{j \in J}$, $\|y\|_2 = 1$, such that

$$|\langle X_i, y\rangle| \geq R_s/4, \quad i \in I. \tag{3.2}$$

Let $\beta = \beta(\delta) \geq 0$ be a sufficiently small constant to be determined later. Consider a β-net $\mathcal{N}_\beta$ of the sphere $S^{n-1} \cap \mathrm{span}(X_j)_{j\in J}$. As in known by volumetric argument (see e.g. [10] Lemma 2.6), one can choose such a net with cardinality

$$|\mathcal{N}_\beta| \leq (3/\beta)^{|J|}.$$

We can assume that the random set $\mathcal{N}_\beta$ depends only on β and the random variables $(X_j)_{j\in J}$. There exists $y \in \mathcal{N}_\beta$ such that $\|y - y_0\|_2 \leq \beta$. By definition of D, this implies that

$$|\langle X_{\pi(\delta s)}, y - y_0\rangle| \leq R_{\delta s} \cdot \beta \leq (R_s/\sqrt{\delta})\beta = R_s/8$$

if we choose $\beta = \sqrt{\delta}/8$. This means that all but at most δs coefficients i in I satisfy the inequality $|\langle X_i, y - y_0\rangle| \leq R_s/8$, and therefore (by (3.2)) also the inequality $|\langle X_i, y_0\rangle| \geq R_s/8$. Let us denote the set of these coefficients by I_0.

Summarizing, we have shown that the event $\{\mathcal{E}$ holds and $D > 1\}$ implies the following event that we call $\mathcal{E}_0$: there exist a number $s \leq N$, disjoint index subsets $I_0, J \subseteq \{1,\dots,N\}$ with cardinalities $|I_0| \geq (1-2\delta)s$, $|J| \leq \delta s$, and a vector $y_0 \in \mathcal{N}_J$ such that

$$|\langle X_i, y_0\rangle| \geq R_s/8, \quad i \in I_0.$$

Without loss of generality, we can assume that $|I_0| = \lceil (1-2\delta)s\rceil$, $|J| = \lfloor \delta s\rfloor$. Note that by Markov's inequality and independence, for a fixed $y_0 \in S^{n-1}$ and a fixed set $I_0 \subset \{1,\dots,N\}$ as above, one has

$$\mathbb{P}\big(|\langle X_i, y_0\rangle| \geq R_s/8,\ i \in I_0\big) \leq (R_s/8)^{-q|I_0|}. \tag{3.3}$$

Then we can bound the probability of $\mathcal{E}_0$ 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 net $\mathcal{N}_J$), taking the union bound over $y_0 \in \mathcal{N}_J$, and finally evaluating the probability using (3.3). This yields

$$\mathbb{P}(\mathcal{E}_0) \leq \sum_{s=1}^{N/2} \binom{N}{|I_0|}\binom{N}{|J|} |\mathcal{N}_\beta|\,(R_s/8)^{-q|I_0|}.$$

Recall that with our choice $\beta = \sqrt{\delta}/12$, we have $|\mathcal{N}_\beta| \leq (24/\sqrt{\delta})^{|J|}$. Further, by our choice of M_1 we have $R_s/8 \geq \delta^{-1} t\,(N/s)^{2/q}$. Using Stirling's approximation, we obtain

$$\mathbb{P}(\mathcal{E}_0) \leq \sum_{s=1}^{N/2} \Big(\frac{eN}{|I_0|}\Big(\frac{\delta}{t}\Big)^q \Big(\frac{s}{N}\Big)^2\Big)^{|I_0|} \Big(\frac{eN}{|J|}\cdot\frac{24}{\sqrt{\delta}}\Big)^{|J|}.$$

Replacing s in the summand by $2|I_0|$ and using the inequalities $|I_0| \geq (1-2\delta)s$ and $|J| \leq \delta s$ along with monotonicity, we conclude for a sufficiently small δ that

$$\mathbb{P}(\mathcal{E}_0) \leq \sum_{s=1}^{N/2} \Big(C\Big(\frac{\delta}{t}\Big)^q \frac{s}{N}\Big)^{(1-2\delta)s} \Big(\frac{CN}{\sqrt{\delta}\, s}\Big)^{\delta s} \leq \sum_{s=1}^{N/2} \Big(\frac{t^{-q}s}{8N}\Big)^{(1-3\delta)s} \leq t^{-0.9q} N^{-0.9}.$$

This completes the proof of Theorem 3.1. □

4. Sampling and norms of random operators $\ell_2 \to \ell_p$

In this section we consider independent random vectors $X_1, \dots, X_N$ in $\mathbb{R}^n$ which satisfy (1.4) for some $q > 4$. To facilitate the notation, we will write $a \lesssim b$ if $a \leq C_{K,L,p,\delta} b$. For $B \geq 1$ and $x \in S^{n-1}$, consider the index set of large coefficients

$$E_B = E_B(x) = \{ i \leq N : \ |\langle X_i, x\rangle| \geq B\}.$$

Lemma 4.1 (Large coefficients)**.** *Let $t \geq 1$, $\varepsilon \in (0,1)$ and $B \geq t(\varepsilon N/n)^{2/(q-4)}$. Then with probability at least $1 - Ct^{-0.9q}$, one has for every $x \in S^{n-1}$:*

$$|E_B| \lesssim t^2 n/\varepsilon B^2, \qquad \|(\langle X_i, x\rangle)_{i \in E_B}\|_{2,\infty} \lesssim t\sqrt{n/\varepsilon}.$$

Proof. By definition of the set E_B and the norm $\|\cdot\|_{2,\infty}$ and using Theorem 3.1, we obtain with the required probability:

$$B^2 |E_B| \leq \|(\langle X_i, x\rangle)_{i\in E_B}\|_{2,\infty}^2 \lesssim n + t^2 |E_B| \left(N/|E_B|\right)^{4/q}. \tag{4.1}$$

It follows that $|E_B| \lesssim n/B^2 + N(t/B)^{q/2}$. This and the assumption on B implies that $|E_B| \lesssim t^2 n/\varepsilon B^2$ as required. Substituting this estimate into the second inequality in (4.1), we complete the proof. □

Proposition 4.2 (Deviation)**.** *Let $p > 2$, $\varepsilon \in (0,1)$ and $\delta > 0$. Consider independent random vectors X_i in $\mathbb{R}^n$ which satisfy* (1.4) *for $q = 4p$. Then with probability at least $1 - \delta$ one has*

$$\sup_{x \in S^{n-1}} \Big| \frac{1}{N} \sum_{i=1}^N |\langle X_i, x\rangle|^p - \mathbb{E}|\langle X_i, x\rangle|^p \Big| \lesssim \varepsilon^{1/2} + \frac{(n/\varepsilon)^{p/2}}{N} + \Big(\frac{n}{\varepsilon^2 N}\Big)^{3/2}. \tag{4.2}$$

Remarks. 1. Theorem 1.1 follows immediately from this result.

2. One could of course optimize the right hand side in ε; we did not do this in order to make clear where the three terms come from.

Proof. Without loss of generality, one can assume that $K = 1$. Choose $t = t(\delta)$ to be sufficiently large so that the probability in Lemma 4.1 satisfies $1 - Ct^{-0.9q} \geq 1 - \delta/2$. Let also $B := t(\varepsilon N/n)^{2/(q-4)}$.

The argument described in [1] in the beginning of proof of Proposition 4.3, which consists of an application of symmetrization, truncation, and contraction principle, reduces the problem to estimating the contribution to the sum of large coefficients. Denoting by E the left hand side of (4.2), one obtains this way with probability at least $1-\delta/2$ that

$$E \lesssim B^{p-1}\sqrt{\frac{n}{N}} + \sup_{x\in S^{n-1}} \frac{1}{N}\sum_{i\in E_B} |\langle X_i, x\rangle|^p + \sup_{x\in S^{n-1}} \mathbb{E}\frac{1}{N}\sum_{i\in E_B} |\langle X_i, x\rangle|^p. \tag{4.3}$$

(In fact one obtains a similar estimate on the expectation on E, from which (4.3) follows by Markov's inequality).

By our choice of B, the first term in the right hand side of (4.3) is $\lesssim \varepsilon^{1/2}$ as required. The second term can be bound using Lemma 4.1. Since $\|\cdot\|_p \lesssim \|\cdot\|_{2,\infty}$ for $p>2$, we obtain with probability at least $1-\delta/2$ that

$$\sup_{x\in S^{n-1}} \frac{1}{N}\sum_{i\in E_B} |\langle X_i, x\rangle|^p \lesssim \frac{1}{N}\|(\langle X_i, x\rangle)_{i\in E_B}\|_{2,\infty}^p \lesssim \frac{(n/\varepsilon)^{p/2}}{N}$$

as required. To compute the third term in the right hand side of (4.3), consider for a fixed x the random variable $Z_i = |\langle X_i, x\rangle|$. Since $\mathbb{E}Z_i^q \le L^q$, an application of Hölder and Markov inequalities yield $\mathbb{E}Z_i^p \mathbf{1}_{\{Z\ge B\}} \le (\mathbb{E}Z^q)^{p/q}(\mathbb{P}(Z \ge B))^{1-p/q} \le L^p B^{p-q}$. Therefore, by our choice of B, we have

$$\sup_{x\in S^{n-1}} \mathbb{E}\frac{1}{N}\sum_{i\in E_B} |\langle X_i, x\rangle|^p = \sup_{x\in S^{n-1}} \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}Z_i^p \mathbf{1}_{\{Z\ge B\}} \le L^p B^{p-q} \lesssim \Big(\frac{n}{\varepsilon^2 N}\Big)^{3/2}.$$

Combining these estimates, we complete the proof. □

Remark. Corollary 1.2 now follows easily. Indeed, for $N \le n$ this result follows from Theorem 3.1 since $\|A\|_{\ell_2\to\ell_p} \lesssim \|A\|_{\ell_2\to\ell_{2,\infty}}$. For $N \ge n$, the result follows from Proposition 4.2.

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*, Journal of AMS, to appear.

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

[3] 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.

[4] J. Bourgain, J. Lindenstrauss, V. D. Milman, *Approximation of zonoids by zonotopes*, Acta Math. 162 (1989), 73–141.

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

[6] O. Guedon, M. Rudelson, *L_p-moments of random vectors via majorizing measures*, Adv. Math. 208 (2007), 798–823.

[7] R. Kannan, L. Lovasz, M. Simonovits, *Random walks and an $O^*(n^5)$ volume algorithm for convex bodies*, Random Structures Algorithms 11 (1997), 1–50.

[8] S. Mendelson, A. Pajor, *On singular values of matrices with independent rows*, Bernoulli 12 (2006), 761-773.

[9] S. Mendelson, *On weakly bounded empirical processes*, Math. Annalen 340 (2008), 293-314.

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

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

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

E-mail address: romanv@umich.edu