The Annals of Probability
2011, Vol. 39, No. 4, 1591–1606
DOI: 10.1214/10-AOP589
© Institute of Mathematical Statistics, 2011

APPROXIMATING THE MOMENTS OF MARGINALS OF HIGH-DIMENSIONAL DISTRIBUTIONS[1]

BY ROMAN VERSHYNIN

University of Michigan

For probability distributions on $\mathbb{R}^n$, we study the optimal sample size $N = N(n, p)$ that suffices to uniformly approximate the pth moments of all one-dimensional marginals. Under the assumption that 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 [*Adv. Math.* **208** (2007) 798–823] which has an extra logarithmic factor in the sample size, and a result of Adamczak et al. [*J. Amer. Math. Soc.* **23** (2010) 535–561] which requires stronger subexponential moment assumptions.

1. Introduction.

1.1. *The estimation problem.* 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 pth 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,

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

Indeed, by the law of large numbers this quantity converges to zero as $N \to \infty$. To understand the 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 [1, 3, 4, 8, 13, 16]. For $p \ne 2$, the problem was also studied in [1, 5–7, 12].

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 $p \ge 2$. Guedon and Rudelson [7] prove the upper bound $N =$

Received December 2009; revised June 2010.
[1]Supported in part by NSF Grants DMS-09-18623 and DMS-10-01829.
MSC2010 subject classifications. Primary 62H12; secondary 46B09, 60B20.
Key words and phrases. High-dimensional distributions, marginals, statistical estimation, heavy-tailed distributions, random matrices.

$O(n^{p/2}\log n)$ for $p \ge 2$ under quite weak moment assumptions[2]

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

The logarithmic term cannot, in general, be removed from the sample complexity; this can be seen by considering a random vector X uniformly distributed in a set of n orthogonal vectors of Euclidean norm $\sqrt{n}$. On the other hand, 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

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

A program aiming at understanding general empirical processes with sub-exponential tails is put forward by Mendelson [11, 12].

1.2. *Distributions with finite moments*: *Main result*. At this moment there is no complete understanding of which distributions on $\mathbb{R}^n$ require logarithmic oversampling and which do not. Clearly there is a gap between the minimal moment assumptions (1.2) of [7] and the subexponential assumptions (1.3) of [1]. The present note makes a step toward closing this gap.

Distributions with tails heavier than exponential frequently arise in statistics, economics, engineering and other exact sciences like geophysics and environmental science. Heavy-tailed distributions are frequently used to model data that exhibit large fluctuations (see, e.g., [2, 9, 10] and the references therein). A very basic theoretical example of a heavy-tailed random vector in $\mathbb{R}^n$ is $X = (\xi_1, \dots, \xi_n)$ where ξ_j are independent random variables with mean zero, unit variance and power-law tails $\mathbb{P}\{|\xi_j| > t\} \sim t^{-q}$ for some fixed exponent $q > 2$ (e.g., normalized Pareto distrbution to mention a specific example). Such random vectors clearly satisfy $\mathbb{E}\|X\|_2^2 = n$, thus $\|X\|_2 = O(\sqrt{n})$ with high probability. Moreover, the marginals have moments $(\mathbb{E}|\langle X, x\rangle|^{q'})^{1/q'} = O(1)$ for all $q' < q$, but the higher moments (for $q' > q$) are infinite.

We shall show that a version of the result of Adamczak et al. [1] holds *under finite moment assumptions* for $p \ne 2$; specifically, the logarithmic oversampling is not needed if we replace p by $4p$ in the minimal moment assumptions (1.2). We shall first consider independent random vectors X_i in $\mathbb{R}^n$ that satisfy

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

[2]The constant implicit in the $O(\cdot)$ notation in the sample complexity N depends only on the constants implicit in the assumptions (1.2); the same convention applies to other results.

THEOREM 1.1 (Approximation of marginals). *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 $N \geq Cn^{p/2}$ where C is a suitably large quantity that depends* (*polynomially*) *only on $K, L, p, \varepsilon, \delta$. Then with probability at least $1 - \delta$ one has*

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

REMARK. 1. A more elaborate version of this result is Theorem 4.3 below. One can get more information on the probability in question using general concentration of measure results as is done in [7]. One can also modify the argument to deduce a version of this result "with high probability" in spirit of [1], that is, 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 also in the range $1 \leq p < 2$. Namely, if the random vectors satisfy (1.4) for some $q \geq 4p$, $q > 4$, then the conclusion (1.5) holds for $N \geq C_{K,L,p,q,\varepsilon,\delta} n$.

3. The method of the present note does not seem to work for $p = 2$; this important and more difficult case is addressed in [18] with an oversampling by a possibly parasitic $(\log\log n)^{c_{p,q}}$ factor.

The argument of this paper also yields sharp bounds on the norms of random operators $\ell_2 \to \ell_p$. The following result is a version of a result of [1], Corollary 4.12, proved there under the stronger sub-exponential moment assumptions (1.3).

THEOREM 1.2 (Norms or random matrices). *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 $X_1, \ldots, X_N$ satisfies with probability at least $1 - \delta$ that*

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

where C depends (*polynomially*) *only on K, L, p, δ.*

1.3. *On the boundedness assumptions.* Let us take a closer look on our assumptions (1.4) on the distribution. The boundedness assumption $\|X_i\|_2 = O(\sqrt{n})$ a.s. seems to be too strong—even the standard Gaussian distribution in $\mathbb{R}^n$ does not satisfy it. We will observe that, although this assumption cannot be formally dropped, it can be removed by slightly modifying the estimation process—discarding the the sample vectors X_i that do not satisfy it.

First, it is easy to see that the boundedness assumption $\|X_i\|_2 = O(\sqrt{n})$ a.s. cannot be dropped from our results. To this end, one easily constructs a random vector whose Euclidean norm has sufficiently heavy tails[3] so that

[3] For example, one can achieve this by considering a version of a "multidimensional Pareto" distribution [10]—the product of the standard Gaussian random vector in $\mathbb{R}^n$ by an independent scalar random variable ξ with a power-law tail.

$\max_{i\le N}\|X_i\|_2 \gg \sqrt{n}$ with high probability for $N \gg 1$ and, in particular, for the stated number of samples $N \sim n^{p/2}$. Then the approximation inequality (1.5) will fail. Indeed, once we choose x in the direction of the vector X_i with the largest Euclidean norm, we will have with high probability that $|\langle X_i, x\rangle|^p = \|X_i\|_2^p \gg n^{p/2}$, which will force the average of the N terms in (1.5) to be much larger than $n^{p/2}/N \sim \text{const}$ while $\mathbb{E}|\langle X_i, x\rangle|^p = O(1)$.

As a side note, the last observation also shows that the sample size $N \sim n^{p/2}$ in Theorem 1.1 is optimal.

Let us also note that the weaker boundedness assumption

$$(\mathbb{E}\|X_i\|_2^q)^{1/q} \le L\sqrt{n} \tag{1.6}$$

follows automatically from the second (moment) assumption in (1.4). To see this, we represent $\|X_i\|_2^2 = \sum_{j=1}^n Z_j$ where $Z_j = |\langle X_i, e_j\rangle|^2$ and where (e_j) is an orthonormal basis in $\mathbb{R}^n$. Then Minkowski's inequality yields (1.6)

$$\begin{aligned}(\mathbb{E}\|X_i\|_2^q)^{2/q} &= \left[\mathbb{E}\left(\sum_{j=1}^n Z_j\right)^{q/2}\right]^{2/q} \le \sum_{j=1}^n (\mathbb{E}Z_j^{q/2})^{2/q} \\ &= \sum_{j=1}^n (\mathbb{E}|\langle X_i, e_j\rangle|^q)^{2/q} \le L^2 n.\end{aligned}$$

Although, as we noticed before, the strong boundedness assumptions cannot be dropped formally, they can be easily transferred into the estimation process. Instead of using all sample points X_i in the approximation inequality (1.5), one can only use those with moderate norms, $\|X_i\|_2 = O(\sqrt{n})$. This will produce a similar approximation result without any boundedness assumption. Just the previous moment assumption will suffice:

$$(\mathbb{E}|\langle X_i, x\rangle|^q)^{1/q} \le L \qquad \text{for all } x \in S^{n-1}. \tag{1.7}$$

COROLLARY 1.3 (Approximation of marginals: no boundedness assumption). *Let $p > 2$, $\varepsilon > 0$, $\delta > 0$ and $K > 0$. Consider independent random vectors X_i in $\mathbb{R}^n$ which satisfy* (1.7) *for $q = 4p$. Let $N \ge Cn^{p/2}$ where C is a suitably large quantity that depends* (*polynomially*) *only on $K, L, p, \varepsilon, \delta$. Denote*

$$I := \{i \le N : \|X_i\|_2 \le K\sqrt{n}\}.$$

Then with probability at least $1 - \delta$ one has

$$\sup_{x\in S^{n-1}} \left| \frac{1}{N} \sum_{i\in I} |\langle X_i, x\rangle|^p - \mathbb{E}|\langle X_i, x\rangle|^p \right| \le \varepsilon + K^{p-q}L^q.$$

PROOF. Consider the events $\mathcal{E}_i = \{\|X_i\|_2 \le K\sqrt{n}\}$. The conclusion then follows by applying Theorem 1.1 to the random vectors $\bar{X}_i = X_i \mathbf{1}_{\mathcal{E}_i}$, which clearly satisfy (1.4). Noting that $|\langle \bar{X}_i, x\rangle|^p = |\langle X_i, x\rangle|^p \mathbf{1}_{\mathcal{E}_i}$, we obtain this way that

$$\sup_{x\in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} |\langle X_i, x\rangle|^p \mathbf{1}_{\mathcal{E}_i} - \mathbb{E}|\langle \bar{X}_i, x\rangle|^p \right| \le \varepsilon. \tag{1.8}$$

To complete the proof, it remains to estimate the error

$$\begin{aligned} \big|\mathbb{E}|\langle X_i, x\rangle|^p - \mathbb{E}|\langle \bar{X}_i, x\rangle|^p\big| &= \mathbb{E}|\langle X_i, x\rangle|^p \mathbf{1}_{\mathcal{E}_i^c} \\ &\le (\mathbb{E}|\langle X_i, x\rangle|^q)^{p/q} (\mathbb{P}(\mathcal{E}_i^c))^{1-p/q}, \end{aligned} \tag{1.9}$$

where we used Hölder's inequality. To estimate the probability of $\mathcal{E}_i^c$ we use (1.6) which follows from our moment assumption (1.7) as we noticed before. By Chebyshev's inequality we obtain

$$\mathbb{P}(\mathcal{E}_i^c) = \{\|X_i\|_2 > K\sqrt{n}\} \le (L/K)^q.$$

Using this and moment assumption (1.7) we conclude that the error (1.9) is bounded by $L^p (L/K)^{q(1-p/q)} = L^q K^{p-q}$. Therefore in (1.8) we can replace $\mathbb{E}|\langle \bar{X}_i, x\rangle|^p$ by $\mathbb{E}|\langle X_i, x\rangle|^p$ by increasing the error bound ε by $L^q K^{p-q}$. This completes the proof. □

REMARKS. 1. Of course one can achieve the approximation error 2ε in Corollary 1.3 by choosing the threshold $K = K(L, \varepsilon)$ sufficiently large.

2. For some distributions one may be able to show that with high probability,

$$\max_{i\le N} \|X_i\|_2 \le K\sqrt{n} \tag{1.10}$$

for some moderate value of K [ideally $K = O(1)$] and for the desired sample size N. In this case, with high probability all events $\mathcal{E}_i$ in Corollary 1.3 hold simultaneously, and therefore they can be dropped from the approximation inequality. One thus obtains the same bound as in Theorem 1.1 except for the extra error term $K^{p-q} L^q$.

This situation occurs, for example, in the estimation result Adamczak et al. [1] mentioned above. For the uniform distribution on an isotropic convex body, the concentration theorem of Paouris [15] implies that $\mathbb{P}(\|X_i\|_2 \ge K\sqrt{n}) \le \exp(-\sqrt{n})$. By union bound this implies that (1.10) holds with probability $1 - N \cdot \exp(-\sqrt{n})$, which is almost 1 for sample sizes N growing linearly or polynomially in n. This is why in the final result of [1] for uniform distributions on convex bodies no boundedness assumption is needed, whereas for general subexponential distributions one needs the boundedness assumption $\|X_i\|_2 = O(\sqrt{n})$ a.s.

1.4. *Heuristics of the proof of Theorem* 1.1. Bourgain [4] 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 [1, 6] 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 Chebyshev'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.

1.5. *Organization of the paper.* 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 4.2. In Section 4, we deduce in a standard way the main results of this note—Theorem 1.1 on approximating the moments of marginals and Theorem 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:*

$$\|X_k\|_2 \le K_1\sqrt{n}, \quad \frac{1}{s}\sum_{i\le s, i\ne k} \langle X_i, X_k\rangle^2 \le K_2^4 n, \qquad k = 1, \dots, s. \tag{2.1}$$

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

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

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

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

PROOF. Without loss of generality, we may assume that $\delta > 0$ is smaller than a suitably chosen absolute constant (this can be done by suitably increasing the value of constant C).

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

$$\langle X_i/a, x\rangle \geq 1, \qquad i = 1, \dots, s.$$

The convex hull $K := \operatorname{conv}\{X_i/a, i = 1, \dots, s\}$ 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

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

(Indeed, one chooses $\bar{x} = z/\|z\|_2$ where z is the element of K with the smallest Euclidean norm.) We express $\bar{x}$ as a convex combination

$$\bar{x} = \sum_{i=1}^{s} \lambda_i X_i/a \qquad \text{for some } \lambda_i \geq 0, \qquad \sum_{i=1}^{s} \lambda_i \leq 1.$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $k \in E$. Therefore with probability at least 0.9, at least $(1 - 20\delta)|E|$ of the events $\mathcal{E}_k$ hold simultaneously. Indeed, by linearity of expectation we have

$$\mathbb{E}\sum_{k\in E} \mathbf{1}_{\mathcal{E}_k^c} = \sum_{k\in E} \mathbb{P}(\mathcal{E}_k^c) \leq 2\delta|E|.$$

By Chebyshev's inequality this yields

$$\mathbb{P}\Big\{\sum_{k\in E} \mathbf{1}_{\mathcal{E}_k} \leq (1-20\delta)|E|\Big\} = \mathbb{P}\Big\{\sum_{k\in E} \mathbf{1}_{\mathcal{E}_k^c} \geq 20\delta|E|\Big\} \leq \frac{2\delta|E|}{20\delta|E|} \leq \frac{1}{10}.$$

We have shown that with probability at least 0.9 the following event occurs: there exists a subset $I \subset E$, $|I| \geq (1-22\delta)s \geq (1-22\delta)s$, 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 $\bar{y} \in \operatorname{span}(X_i)_{i \in I^c}$. Also, by definition of $\mathcal{E}_k$, one has

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

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

3. Norms of random operators $\ell_2 \to \ell_{2,\infty}$. Recall that the weak ℓ_2-norm $\|x\|_{2,\infty}$ of a vector $x = (x_1, \dots, x_N) \in \mathbb{R}^N$ is defined as the minimal number M for which the nonincreasing rearrangement (x_k^*) of the sequence $(|x_k|)$ satisfies $x_k^* \leq M k^{-1/2}$, $k = 1, \dots, N$. It is well known that the quasi-norm $\|\cdot\|_{2,\infty}$ is equivalent to a norm on $\mathbb{R}^N$ (see [17]), and 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 independent random vectors $X_1, \dots, X_N$ in $\mathbb{R}^n$ 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 $I \subseteq \{1, \dots, N\}$, one has*

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

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

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

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 inequality of Bourgain [4] 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

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

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 $Z_1, \dots, Z_N \geq 0$ be independent random variables which satisfy $\mathbb{E} Z_i^q \leq B^q$ for some $q > 0$ and some B. Consider the nonincreasing rearrangement (Z_i^*) of (Z_i). Then, for every $t \geq 1$, one has with probability at least $1 - Ct^{-q}/N$ that*

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

In particular, for $q > 4$ (3.1) implies

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

PROOF. By homogeneity, we can assume that $B = 1$. Then by Chebyshev's inequality we have $\mathbb{P}\{Z_j > u\} \leq u^{-q}$ for every $j \leq N$ 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 $i = 1, \dots, N$

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

Choosing $u = t(eN/i)^{2/q}$ we obtain $\mathbb{P}\{Z_i^* > u\} \leq (t^{-q}i/eN)^i$. Then, for $t \geq 1$,

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

This easily implies the first part of the lemma. The second part follows by summation using that $\frac{1}{s}\sum_{i=1}^{s} i^{-r} \leq C_r s^{-r}$ for $0 < r < 1$; here $r = 4/q$. □

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 $t \geq 1$ the following holds with probability at least $1 - Ct^{-q}$. For every subset $E \subseteq \{1, \dots, N\}$ and every $k \leq N$ one has*

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

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

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

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

PROOF OF THEOREM 3.1. By homogeneity, we can assume that $L = 1$. Also, by decomposing I in three sets of roughly equal cardinality we see that it suffices to prove the conclusion for the subsets I of cardinality $|I| \le N/2$.

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 $C'_q \ge 8$, which we will use later.

We will now need a convenient interpretation of the conclusion of the theorem. Given $x \in S^{n-1}$, we denote by $|\langle X_{\pi(i)}, x\rangle|$ a nonincreasing 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

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

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

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

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 $\{\mathcal{E}$ 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 an integer $s \le N/2$, a subset $E \subseteq \{1, \dots, N\}$, $|E| = s$ and a vector $x \in S^{n-1}$ such that

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

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 $|I| \ge (1 - \delta)s$, $|J| \le \delta s$, and there exists a vector $y \in \operatorname{span}(X_j)_{j \in J}$, $\|y\|_2 = 1$, such that

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

Let $\beta = \beta(\delta) \geq 0$ be a sufficiently small quantity to be determined later. Consider a β-net $\mathcal{N}_J$ of the sphere $S^{n-1} \cap \mathrm{span}(X_j)_{j \in J}$. As in known by volumetric argument (see, e.g., [14], Lemma 2.6), one can choose such a net with cardinality

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

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

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

if we choose $\beta = \sqrt{\delta}/8$. This means that all but at most δs indices 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. Note that

$$\begin{aligned} R_s/8 &\geq \frac{1}{8} M_1 (N/s)^{2/q} \qquad \text{(by definition of } R_s \text{ and since } D > 1) \\ &\geq \frac{C_q'}{8} (t/\delta)(N/s)^{2/q} \qquad \text{(by definition of } M_1) \\ &\geq (t/\delta)(N/s)^{2/q} \qquad (\text{since } C_q' \geq 8). \end{aligned}$$

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

$$|\langle X_i, y_0\rangle| \geq (t/\delta)(N/s)^{2/q}, \qquad i \in I_0.$$

Note that by Chebyshev'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

$$\begin{aligned} \mathbb{P}\{|\langle X_i, y_0\rangle| \geq (t/\delta)(N/s)^{2/q}, i \in I_0\} &\leq \big((t/\delta)(N/s)^{2/q}\big)^{-q|I_0|} \\ &= \big((\delta/t)^q (s/N)^2\big)^{|I_0|}. \end{aligned} \tag{3.3}$$

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

$$\mathbb{P}(\mathcal{E}_0) \leq \sum_{s=1}^{N/2} \binom{N}{|I_0|} \binom{N}{|J|} |\mathcal{N}_J| \big((\delta/t)^q (s/N)^2\big)^{|I_0|}$$

(recall that I_0 and J in this sum may depend on s). Also recall that with our choice $\beta = \sqrt{\delta}/8$, we have $|\mathcal{N}_J| \leq (24/\sqrt{\delta})^{|J|}$. Further, by our choice of M_1 we have

$R_s/8 \ge \delta^{-1} t(N/s)^{2/q}$. Using Stirling's approximation, we obtain

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

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

$$\mathbb{P}(\mathcal{E}_0) \le \sum_{s=1}^{N/2} \left(C \left(\frac{\delta}{t}\right)^q \frac{s}{N} \right)^{(1-2\delta)s} \left(\frac{CN}{\delta^{3/2} s} \right)^{\delta s} \le \sum_{s=1}^{N/2} \left(\frac{t^{-q} s}{10N} \right)^{(1-3\delta)s} \le t^{-0.9q} N^{-0.9}.$$

This completes the proof of Theorem 3.1. □

4. Approximation of marginals and the $\ell_2 \to \ell_p$ norms of random operators. In this section we deduce from Theorem 3.1 the main results of this paper, Theorems 1.1 and 1.2. The method of this deduction is by now standard; it was used in particular in [1]. It consists of an application of symmetrization, truncation, and contraction principle, and it reduces the problem to estimating the contribution to the sum of large coefficients.

Specifically, given a threshold $B \ge 0$ and a vector $x \in S^n$, we define the set of large coefficients with respect to random vectors $X_1, \dots, X_N$ as

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

The truncation argument in the beginning of proof of Proposition 4.4 in [1] yields the following bound:

LEMMA 4.1 (Reduction to the few large coefficients). *Let $p \ge 2$, $B \ge 0$, $t \ge 1$. Consider independent random vectors X_i in $\mathbb{R}^n$ which satisfy (1.7) for $q = 2p$. Then for every positive integer N, with probability at least*

$$1 - \exp\big(-c \min\big(t^2 n B^{2p-2}, t\sqrt{Nn}/B\big)\big)$$

one has

$$\begin{aligned} \sup_{x \in S^{n-1}} & \left| \frac{1}{N} \sum_{i=1}^N |\langle X_i, x\rangle|^p - \mathbb{E}|\langle X_i, x\rangle|^p \right| \\ & \le 16 t B^{p-1} \sqrt{\frac{n}{N}} + \sup_{x \in S^{n-1}} \frac{1}{N} \sum_{i \in E_B(x)} |\langle X_i, x\rangle|^p \\ & \quad + \sup_{x \in S^{n-1}} \mathbb{E} \frac{1}{N} \sum_{i \in E_B(x)} |\langle X_i, x\rangle|^p, \end{aligned} \tag{4.1}$$

where $c = c_{p,K} > 0$ depends only on p and the parameter K in the moment assumption (1.7).

This lemma reduces the approximation problem in Theorem 1.1 to finding an upper bound on the contribution of the large coefficients $\frac{1}{N}\sum_{i\in E_B(x)} |\langle X_i, x\rangle|^p$. In the following lemma, we observe that a slightly stronger bound (for the $\|\cdot\|_{2,\infty}$ norm rather than $\|\cdot\|_p$ norm) follows from Theorem 3.1. To facilitate the notation, throughout the end of this section we will write $a \lesssim b$ if $a \le C_{K,L,p,q,\delta} b$.

LEMMA 4.2 (Large coefficients). *Let $q > 4$, $t \ge 1$, $\varepsilon \in (0,1)$ and $B \ge t(\varepsilon N/n)^{2/(q-4)}$. Consider independent random vectors $X_1, \dots, X_N$ in $\mathbb{R}^n$ which satisfy (1.4). 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

$$B^2|E_B| \le \|(\langle X_i, x\rangle)_{i\in E_B}\|_{2,\infty}^2 \lesssim n + t^2|E_B|(N/|E_B|)^{4/q}. \tag{4.2}$$

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 (4.2), we complete the proof. □

PROPOSITION 4.3 (Deviation). *Let $p > 2$, $\varepsilon \in (0,1)$, $\delta > 0$ and $N \ge n/\varepsilon + C$ where $C = C_{p,K,\delta}$ is suitably large. 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*

$$\sup_{x\in S^{n-1}} \left| \frac{1}{N}\sum_{i=1}^N |\langle X_i, x\rangle|^p - \mathbb{E}|\langle X_i, x\rangle|^p \right| \lesssim \varepsilon^{1/2} + \frac{(n/\varepsilon)^{p/2}}{N} + \left(\frac{n}{\varepsilon N}\right)^{3/2}. \tag{4.3}$$

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 OF PROPOSITION 4.3. We choose $B := t(\varepsilon N/n)^{2/(q-4)}$ so that Lemma 4.2 holds.

Next, we choose $t = t(\delta, K)$ and $C = C_{p,K,\delta}$ sufficiently large so that the probabilities in Lemmas 4.1 and 4.2 are at least $1 - \delta/2$ each. This is indeed possible for the probability in Lemma 4.1 as one can check that $t^2 n B^{2p-2} = t^{2p}\varepsilon N \ge t^{2p}$ and $t\sqrt{Nn}/B \ge N^{1/2-2/(q-4)} \ge C^{(p-2)/2(p-1)}$; for the probability in Lemma 4.2 this is straightforward.

Let us assume that the conclusions of both these lemmas hold; as we now know this holds with probability at least $1 - \delta$. Our goal is to estimate the three terms in the right-hand side of (4.1).

By our choice of B, the first term in the right-hand side of (4.1) is $\lesssim \varepsilon^{1/2}$ as required. The second term can be bounded using Lemma 4.2. Since $\|\cdot\|_p \lesssim \|\cdot\|_{2,\infty}$ for $p > 2$, we obtain 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 (4.1), consider for a fixed x the random variable $Z_i = |\langle X_i, x\rangle|$. Since $\mathbb{E}Z_i^q \le L^q$, we have

$$\mathbb{E}Z_i^p \mathbf{1}_{\{Z_i\ge B\}} \le \mathbb{E}Z_i^p (Z_i/B)^{q-p}\mathbf{1}_{\{Z_i\ge B\}} \le \mathbb{E}Z_i^q/B^{q-p} \le L^q B^{p-q}.$$

Therefore, by our choice of B, we have

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

Combining these estimates, we complete the proof. □

REMARK. Theorem 1.2 now follows easily. We can assume that $N \ge C$ where $C = C_{p,K,\delta}$ is suitably large. Now, 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 4.3 with $\varepsilon = 1$, noting that $(\mathbb{E}|\langle X_i, x\rangle|^p)^{1/p} \le L$ as $p \le q$.

Acknowledgment. The author is grateful to the referees for their thorough reading of the first two versions of this manuscript and for many suggestions, which greatly improved the presentation of this paper.

REFERENCES

[1] ADAMCZAK, R., LITVAK, A. E., PAJOR, A. and TOMCZAK-JAEGERMANN, N. (2010). Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. Amer. Math. Soc.* **23** 535–561. MR2601042

[2] ADLER, R. J., FELDMAN, R. E. and TAQQU, M. S., eds. (1998). *A Practical Guide to Heavy Tails. Statistical Techniques and Applications*. Birkhäuser, Boston, MA. MR1652283

[3] AUBRUN, G. (2007). Sampling convex bodies: A random matrix approach. *Proc. Amer. Math. Soc.* **135** 1293–1303 (electronic). MR2276637

[4] BOURGAIN, J. (1999). Random points in isotropic convex sets. In *Convex Geometric Analysis* (*Berkeley*, *CA*, 1996). *Math. Sci. Res. Inst. Publ.* **34** 53–58. Cambridge Univ. Press, Cambridge. MR1665576

[5] BOURGAIN, J., LINDENSTRAUSS, J. and MILMAN, V. (1989). Approximation of zonoids by zonotopes. *Acta Math.* **162** 73–141. MR0981200

[6] GIANNOPOULOS, A. A. and MILMAN, V. D. (2000). Concentration property on probability spaces. *Adv. Math*. **156** 77–106. MR1800254

[7] GUÉDON, O. and RUDELSON, M. (2007). L_p-moments of random vectors via majorizing measures. *Adv. Math*. **208** 798–823. MR2304336

[8] KANNAN, R., LOVÁSZ, L. and SIMONOVITS, M. (1997). Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures Algorithms* **11** 1–50. MR1608200

[9] KLÜPPELBERG, C. and MIKOSCH, T. (1997). Large deviations of heavy-tailed random sums with applications in insurance and finance. *J. Appl. Probab*. **34** 293–308. MR1447336

[10] MEERSCHAERT, M. M. and SCHEFFLER, H.-P. (2001). *Limit Distributions for Sums of Independent Random Vectors. Heavy Tails in Theory and Practice*. Wiley, New York. MR1840531

[11] MENDELSON, S. Empirical processes with a bounded ψ_1 diameter. Unpublished manuscript.

[12] MENDELSON, S. (2008). On weakly bounded empirical processes. *Math. Ann*. **340** 293–314. MR2368981

[13] MENDELSON, S. and PAJOR, A. (2006). On singular values of matrices with independent rows. *Bernoulli* **12** 761–773. MR2265341

[14] MILMAN, V. D. and SCHECHTMAN, G. (1986). *Asymptotic Theory of Finite-dimensional Normed Spaces*. *Lecture Notes in Math*. **1200**. Springer, Berlin. MR0856576

[15] PAOURIS, G. (2006). Concentration of mass on convex bodies. *Geom. Funct. Anal*. **16** 1021–1049. MR2276533

[16] RUDELSON, M. (1999). Random vectors in the isotropic position. *J. Funct. Anal*. **164** 60–72. MR1694526

[17] STEIN, E. M. and WEISS, G. (1971). *Introduction to Fourier Analysis on Euclidean Spaces*. *Princeton Mathematical Series* **32**. Princeton Univ. Press, Princeton, NJ. MR0304972

[18] VERSHYNIN, R. (2010). How close is the sample covariance matrix to the actual covariance matrix? Unpublished manuscript. Available at http://arxiv.org/abs/1004.3484.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109
USA
E-MAIL: romanv@umich.edu