Geometric Applications of Chernoff-type Estimates

S. Artstein-Avidan[1,2], O. Friedland[3], V. Milman[1,3]

1 Introduction

In this paper we present a probabilistic approach to some geometric problems in asymptotic convex geometry. The aim of this paper is to demonstrate that the well known Chernoff bounds from probability theory can be used in a geometric context for a very broad spectrum of problems, and lead to new and improved results. We begin by briefly describing Chernoff bounds, and the way we will use them.

The following Proposition, which is a version of Chernoff bounds, gives estimates for the probability that at least βN trials out of N succeed, when the probability of success in one trial is p (the proof is standard, see e.g. [HR]).

Proposition 1 (Chernoff). *Let Z_i be independent Bernoulli random variables with mean $0 < p < 1$, that is, Z_i takes value 1 with probability p and value 0 with probability $(1-p)$. Then we have*

1) for every $\beta < p$

$$\mathbb{P}[Z_1 + \cdots + Z_N \geq \beta N] \geq 1 - e^{-NI(\beta,p)},$$

2) for every $\beta > p$

$$\mathbb{P}[Z_1 + \cdots + Z_N > \beta N] \leq e^{-NI(\beta,p)},$$

where $I(\beta,p) = \beta \ln \frac{\beta}{p} + (1-\beta) \ln \frac{1-\beta}{1-p}$.

Assume that X_i is a sequence of independent non-negative random variables. For simplicity assume to begin with, that they are also identically distributed, and even bounded. A good example to consider is $X_i = \|U_i x\|$ where $\|\cdot\|$ is some norm on n-dimensional space $\mathbb{R}^n$, U_i a random orthogonal matrix (with respect to the normalized Haar measure on $O(n)$) and x some fixed point in the space. Define the sequence of partial sums $S_N = \sum_{i=1}^{N} X_i$. The law of large numbers says that $\frac{1}{N} S_N$ converges to the expectation $\mathbb{E}X_i$ as N tends to

[1] partially supported by BSF grant 2002-006

[2] supported by the National Science Foundation under agreement No. DMS-0111298

[3] supported in part by the Israel Science Academy

infinity. In our example the expectation is $|x|M$ where $M = \int_{S^{n-1}} \|u\| d\sigma(u)$ and σ is the rotation invariant probability measure on the sphere S^{n-1}.

To estimate the rate of convergence, one usually turns to large deviation theorems, for example the following well known Bernstein's inequality (see e.g. [BLM]). We say that a centered random variable X is a ψ_2 random variable if there exists some constant A such that $\mathbb{E}e^{\frac{X^2}{A^2}} = 2$, and the minimal A for which this inequality holds we call the ψ_2 norm of X. Below when we say the ψ_2 norm of X we mean the ψ_2 norm of the centered variable $(X - \mathbb{E}X)$.

Proposition 2 (Bernstein). *Let X_i be i.i.d. copies of the random variable X, and assume the X has ψ_2 norm A. Then for any $t > 0$*

$$\mathbb{P}[|\frac{1}{N}S_N - \mathbb{E}X| > t] \leq 2e^{-cNt^2}, \tag{1}$$

where $c = \frac{1}{8A^2}$.

Sometimes it is important to get the probability in (1) to be very small. This is the case in the example of $X_i = \|U_i x\|$, if one wants to have an estimate for all points x in some large net on the sphere (we study this example in more detail in Section 4).

The obvious way to make the probability in (1) smaller is to increase t. However, once t is greater than $\mathbb{E}X$, the estimate in (1) makes sense only as an upper bound for S_N and provides no effective lower bound, since the trivial estimate $0 \leq S_N$ is always true.

Thus, we see that for positive random variables, an estimate of the type (1) does not fully answer our needs, and we actually want an estimate of the type

$$\mathbb{P}[\varepsilon\mathbb{E} \leq \frac{1}{N}S_N \leq t\mathbb{E}] \leq 1 - f(\varepsilon, t, N, X),$$

with f decaying exponentially fast to 0 with N, and moreover, such that the rate of decay will substantially improve as t tends to ∞ and ε tends to 0. This is the aim of our probabilistic method and the subject of the next discussion.

For $\frac{1}{N}S_N$ not to be very small, it is not obligatory that *all* X_is be large, it is enough if some fixed proportion of them are not small. This is the main idea behind our use of Chernoff bounds. The first time this method was applied in our field was in the paper of Milman and Pajor [MP], where in particular a global form of the low M^*-estimate was obtained.

Applying this in our scheme we let $Z_i = 1$ if $X_i \geq \varepsilon$ and $Z_i = 0$ if $X_i < \varepsilon$. Since all X_i are positive, having $\sum_{i=1}^N Z_i \geq \beta N$ means in particular that $\frac{1}{N}S_N \geq \beta\varepsilon$, and this happens with the probability written in Proposition 1, where $p = \mathbb{P}[X_i \geq \varepsilon]$, and β is any number smaller than this p.

Before we proceed let us analyze the estimate. We have

$$I(\beta, p) = u(\beta) - \beta \ln p - (1 - \beta) \ln (1 - p),$$

where we denoted $u(\beta) = [\beta \ln \beta + (1 - \beta) \ln (1 - \beta)]$. The term $u(\beta)$ is a negative, convex function which approaches 0 as $\beta \to 0$ and as $\beta \to 1$, and is symmetric about $1/2$ where it has a minima equal to $-\ln 2$. Thus the whole exponent is of the form

$$e^{-NI(\beta,p)} = p^{\beta N}(1-p)^{(1-\beta)N}e^{-Nu(\beta)} \leq (1-p)^{(1-\beta)N}2^N. \qquad (2)$$

We will usually use the latter, though sometimes we will need the better estimate including $u(\beta)$.

To use the full strength of (2), we will need to have the probability p of success increase rapidly as the parameters in question change. In our example, we will need $\mathbb{P}[X_i \geq \varepsilon]$ to approach 1 fast when $\varepsilon \to 0$. This is not always the case, and additional work is sometimes needed. This will best be demonstrated in Section 3.

In the remainder of this section we outline the main theorems to be proved in this paper and explain the notation to be used throughout.

In Section 2 we give an application to a problem from learning theory, improving a result of Cheang and Barron [CB]. The problem regards the approximation of the n-dimensional euclidean ball by a simpler body, which resembles a polytope but need not be convex, and is described by the set of points satisfying a certain amount of linear inequalities out of a given list of length N. In their paper [CB] Cheang and Barron showed that to ε-approximate the ball one can do with $N = C(n/\varepsilon)^2$ linear inequalities, and we improve this estimate (for fixed ε and $n \to \infty$) to $N = Cn\ln(\frac{1}{\varepsilon})/\varepsilon^2$ (where C is a universal constant). We formulate our theorem (for the proof see [AFM]) and in the remainder of the section we show stability results.

In Section 3 we show three different applications to Khinchine-type inequalities. We reprove, with slightly worse constants, a theorem of Litvak, Pajor, Rudelson, Tomczak-Jaegermann and Vershynin, which is an isomorphic version of Khinchine inequality in the L_1 case, where instead of taking the average of the 2^n terms $|\langle x, \varepsilon\rangle|$ for $\varepsilon \in \{-1,1\}^n$, one averages only over $(1+\delta)n$ of them (for some fixed $\delta > 0$), and the constants in the corresponding inequality depend on δ. Another way to view this result is realizing an n-dimensional euclidean section of $\ell_1^{(1+\delta)n}$ by a matrix of random signs. Schechtman was the first who proved the existence of such an isomorphism for some universal (and large) δ_0, and also together with Johnson proved a non-random version of this fact, see [LPRTV] [S2] [JS]. We remark that an improvement of this result, with a much better dependence on δ will soon be published in [AFMS].

The next application answers a similar question, where instead of random sign vectors, the vectors are random with respect to the volume distribution in an isotropic convex body. We show that when the rows of an $(n \times (1+\delta)n)$ matrix are chosen randomly inside an isotropic convex body, again its image is an n-dimensional euclidean section of $\ell_1^{(1+\delta)n}$. There is a conceptual difference between this result and the preceding one, since now only the *rows* of the matrix are independent, and not *all* entries.

In another application, we reduce the level of randomness, substituting most of it by an explicit sign-matrix. We prove that a Hadamard $(n \times n)$ matrix with extra δn rows of random signs also realizes a euclidean section of $\ell_1^{(1+\delta)n}$, and moreover, the isomorphism constants are polynomially dependent on δ. This is an extension of a result by Schechtman [S1] where he used an $(n \times 2n)$ matrix whose upper half consisted of (a scalar multiple of) the identity matrix and all lower half entries were random signs.

In Section 4 we give a different type of application, proving a Dvoretzky-type theorem in global form. We show that the average of $C(\frac{a}{M^*})^2$ random rotations of a convex body K (with half-mean-width M^* and half-diameter a, see definitions below) is isomorphic to the euclidean ball. This is well known, e.g. [MS]. In the proof we show how the probabilistic method can be adapted to give a new proof of the *upper* bound in this problem. As will be explained below, the main use of the Chernoff method is to provide lower bounds, while upper bounds can usually be obtained straightforwardly with the use of deviation inequalities. However, in the standard proof of the global Dvoretzky Theorem, the upper bound is obtained by using a deep geometric result about concentration on the product of spheres, which itself uses Ricci curvature (see [GrM]). We will show how standard concentration on the sphere, together with our method, provides an alternative proof for the bound. We then show how a reformulation of a conjecture by Vershynin, given by Latała and Oleszkiewicz [LO] about small ball probabilities will imply that the above is true for $(1+\delta)(\frac{a}{M^*})^2$ random rotations, for any δ, with constants of isomorphism depending on δ, a result which will be optimal. In addition we give an alternative parameter that can be used to study these averages, similar to the one introduced by Klartag and Vershynin [KV], which in special cases gives improved results.

The paper includes both new proofs of known result and some new results, and our main goal is to show how the probabilistic method we describe here is applicable in many different situations, and in some sense can be considered as another systematic approach to obtaining lower and upper bounds. In many cases this unifies what were before individual proofs for specific problems.

Notation We work in $\mathbb{R}^n$ which is equipped with the euclidean structure $\langle\cdot,\cdot\rangle$ and write $|\cdot|$ for the euclidean norm. The euclidean unit ball and sphere are denoted by D_n and S^{n-1} respectively. We write σ_n for the rotation invariant probability measure on S^{n-1}, and omit the index n when the dimension is clear from the context. Every symmetric (with respect to 0) convex body K in $\mathbb{R}^n$ induces the norm $\|x\|_K = \inf\{\lambda > 0 : x \in \lambda K\}$. The polar body of K is $K^\circ = \{y \in \mathbb{R}^n : \max_{x\in K} |\langle y, x\rangle| \leq 1\}$ and it induces the dual norm $\|x\|_K^* = \|x\|_{K^\circ} = \max_{y\in K} |\langle y, x\rangle|$. We define $M(K) = \int_{S^{n-1}} \|u\|_K d\sigma_n(u)$ and $M^*(K) = \int_{S^{n-1}} \max_{y\in K} |\langle y, u\rangle| d\sigma_n(u)$. So, $M = M(K)$ is the average of the norm associated to K on the sphere and $M^* = M^*(K) = M(K^\circ)$ is half the mean width of K. We also denote by a and b the least positive constants for which $\frac{1}{a}|x| \leq \|x\|_K \leq b|x|$ holds true for every $x \in \mathbb{R}^n$. Thus, a is half of the diameter of K and $\frac{1}{b}$ is the in-radius of K (so, $\frac{1}{b}D \subseteq K \subseteq aD$). As usual in asymptotic geometric analysis, we will be dealing with finite dimensional normed spaces or convex bodies, and study behavior of some geometric parameters as the dimension grows to infinity. Thus, the dimension n is always assumed large, and the universal constants appearing throughout the paper, denoted usually by c, c_0, c_1, C, do not depend on the dimension and are just numerical constants which can be computed. In addition, throughout, we omit the notation $[\cdot]$ of integer values, and assume the numbers we deal with are integers when needed, to avoid notational inconvenience.

2 A ZigZag Approximation for Balls

2.1 The ZigZag construction and the main theorem

We address the question of approximating the euclidean ball by a simpler set. In many contexts, polytopes are considered to be the simplest sets available, being the intersection of some number of half-spaces, or in other words the set of all points satisfying some list of N linear inequalities. However, it is well known and easy to check that to construct a polytope which is ε-close to the euclidean ball $D_n \subset \mathbb{R}^n$ in the Hausdorff metric one needs to use exponentially many half-spaces, $N \geq e^{Cn\ln(1/\varepsilon)}$ (this can be seen by assuming the polytope is inscribed in D_n, and estimating from above the volume of the cap that each half-space cuts off the sphere S^{n-1}). This is a huge number, and so a different kind of approximation was suggested, first used by Cybenko [C], and by Hornik, Stinchcombe and White [HSW].

The first good bounds in such an approximation result (we describe the approximating set below) were given by Barron [B]. These sets are implemented by what is called single hidden layer neural nets or perception nets, and we will use the simplest version of such sets, for which we suggested the name "ZigZag approximation".

The approximating set is the following, it is no longer convex, but is still described by a list of linear inequalities. Given a set of N inequalities, and a number $k \leq N$, the set consists of all points satisfying no less than k of the N inequalities. We learned of this approximation from a paper by Cheang and Barron [CB], where they showed that there exists a universal constant C such that for any dimension n, one can find $N = C(n/\varepsilon)^2$ linear inequalities, such that the set of points satisfying at least k of the N inequalities is ε-close, in the Hausdorff metric, to D_n (where k is some proportion of N). This is already a huge improvement, from a set described by an exponential number of inequalities to a polynomial number.

Using our approach we improve (in the case of $n \to \infty$) their estimate to $N = Cn\ln(1/\varepsilon)/\varepsilon^2$ linear inequalities, and we use $k = N/2$. The formulation of our result is given in the following Theorem (see [AFM] for its proof).

Theorem 3. *There exists universal constants c, C such that for every dimension n, and every $0 < \varepsilon < 1$, letting $N = [Cn\ln(1/\varepsilon)/\varepsilon^2]$, if $z_1, \dots, z_N$ are random points with respect to Lebesgue measure σ on the sphere S^{n-1}, then* *with probability greater than* $1-e^{-cn}$*, the set*

$$\mathcal{K} = \{x \in \mathbb{R}^n : \exists i_1, \dots, i_{[N/2]} \text{ with } |\langle x, z_{i_j}\rangle| < \frac{c_0}{\sqrt{n}}\}$$

satisfies

$$(1-\varepsilon)D_n \subset \mathcal{K} \subset (1+\varepsilon)D_n,$$

where c_0 *denotes the constant (depending on* n*, but converging to a universal constant as* $n \to \infty$*) for which* $\sigma(u \in S^{n-1} : |\langle\theta, u\rangle| \le \frac{c_0}{\sqrt{n}}) = 1/2$ *for some* $\theta \in S^{n-1}$*.*

2.2 Stability Results

Theorem 3 above is stable, in the following sense, define the body

$$\mathcal{K}(\beta) = \{x \in \mathbb{R}^n : \exists i_1, \dots, i_{[\beta N]} \text{ with } |\langle x, z_{i_j}\rangle| < \frac{c_0}{\sqrt{n}}\}$$

where we have changed the parameter $1/2$ into β. By stability we mean that for N large enough the two bodies $\mathcal{K}_1 = \mathcal{K}(\beta+\delta)$ and $\mathcal{K}_2 = \mathcal{K}(\beta-\delta)$ are close, in the Hausdorff distance, as long as $0 < \delta < \delta_0$, where δ_0 depends only on β. This will readily follow from the fact that both bodies will be close to the euclidean ball of the appropriate radius, depending on β.

We first remark that changing the constant c_0 in the definition of $\mathcal{K}(\beta)$ into c_1 results in multiplication of the body $\mathcal{K}(\beta)$ by the factor $\frac{c_1}{c_0}$. Thus if we denote by c_β the constant so that $\sigma(u \in S^{n-1} : |\langle\theta, u\rangle| \le \frac{c_\beta}{\sqrt{n}}) = \beta$ and define

$$\mathcal{K}'(\beta) = \{x \in \mathbb{R}^n : \exists i_1, \dots, i_{[\beta N]} \text{ with } |\langle x, z_{i_j}\rangle| < \frac{c_\beta}{\sqrt{n}}\}$$

we will have $\mathcal{K}'(\beta) = \frac{c_\beta}{c_0}\mathcal{K}(\beta)$. Notice that the way we defined c_0 at the beginning it actually corresponds in the current notation to $c_{\frac{1}{2}}$.

Now, the fact that these bodies, $\mathcal{K}'(\beta)$, are equivalent to euclidean balls of radius 1 when N is sufficiently large follows in the same way as in Theorem 3. We give the sketch of the proof for $N = C(\beta, \varepsilon)n\log n$ and $\varepsilon > c/\sqrt{n}$. For the proof of the linear dependence on n see complete details in [AFM]. We pick a $1/n$ net of the sphere $(1-\varepsilon)S^{n-1}$. For a point x_0 in the net we check not only $x_0 \in \mathcal{K}'(\beta)$, but more, namely that there exist $i_1, \dots, i_{[\beta N]}$ with $|\langle x_0, z_{i_j}\rangle| < \frac{c_\beta}{\sqrt{n}} - \frac{1}{n}$.

Since the probability of a single event is

$$\sigma(u \in S^{n-1} : |\langle u, \theta\rangle| < (\frac{c_\beta}{\sqrt{n}} - \frac{1}{n})/(1-\varepsilon)) = \beta + p_{\varepsilon,\beta}$$

for some $p_{\varepsilon,\beta} > 0$ (and as long as ε is not too small), we have by Chernoff bounds an exponential probability $1 - e^{-NI(\beta,\beta+p_{\varepsilon,\beta})}$ that x_0 satisfies βN of these inequalities. When N is large enough, greater than $C(\beta,\varepsilon)n\log n$, this probability suffices to take care of the whole net. Then for a point x in $(1-\varepsilon)S^{n-1}$ which is $1/n$-close to a point x_0 in the net, we have that for exactly the same indices, the inequalities $|\langle x, z_{i_j}\rangle| < c_\beta/\sqrt{n}$ are satisfied, which means that $x \in \mathcal{K}'(\beta)$. So we attained $(1-\varepsilon)D_n \subset \mathcal{K}'(\beta)$. The other inclusion is proved similarly.

This implies in particular that if N is large enough

$$(1-\varepsilon)(\frac{c_0}{c_{\beta+\delta}})D_n \subset \mathcal{K}(\beta+\delta) \subset \mathcal{K}(\beta-\delta) \subset (1+\varepsilon)(\frac{c_0}{c_{\beta-\delta}})D_n,$$

as long as $\delta < \delta_0(\beta)$.

The stability is reflected in the rate of change of c_β for β bounded away from 1, which one can estimate by standard volume estimates on the sphere. Thus, $c_{\beta+\delta} < c_{\beta-\delta}(1+C\delta)$. This is what we consider a stability result. We remark that it is not difficult to check that for, say, $\beta > 1/2$ and bounded away from 1, we have $c_0 c\beta < c_\beta < c_0 C\beta$ and thus

$$(\frac{c}{\beta})D_n \subset \mathcal{K}(\beta) \subset (\frac{C}{\beta})D_n.$$

(We mean here, that the constants c and C are universal for, say $1/2 < \beta < 3/4$, and in general depend only on δ_0 when we assume $1/2 < \beta < 1-\delta_0$.) The same is true for $\beta < 1/2$ and bounded away from 0.

The reason that stability results can be important is that sometimes one cannot check *exactly* if a proportion 1/2 of the inequalities is fulfilled, but *can* do the following weaker thing: to have a set so that each point in the set satisfies at least $1/2-\delta$ of the inequalities, and each point outside the set has at least $1/2-\delta$ inequalities which it violates. The stability result implies that we can be sure this set is $C\delta$-isomorphic to the euclidean ball (provided δ is in some bounded range).

Remark 1 The same type of results hold for the following body, where we omit the absolute value,

$$\mathcal{K}(\beta) = \{x \in \mathbb{R}^n : \exists i_1, \ldots, i_{[\beta N]} \text{ with } \langle x, z_{i_j}\rangle < \frac{c_0}{\sqrt{n}}\}.$$

Remark 2 The above discussion implies in particular a probabilistic approach to deciding whether a point is in the ball or not. Indeed, once we have a description of the ball as points satisfying at least 1/2 of the inequalities from a list of N inequalities, we can now for a given point pick randomly say 100 of the inequalities and check what proportion of them is satisfied. Again using Chernoff bounds, we can show that if it satisfies more than 1/2 of them there is a large probability that it is inside $(1+\varepsilon)D_n$ and if it violates more than 1/2 of the inequalities there is a large probability that it is outside $(1-\varepsilon)D_n$. The word "large" here is relative to the choice 100.

3 Khinchine-type Inequalities, or: isomorphic embeddings of ℓ_2^n into ℓ_1^N

3.1 Isomorphic Khinchine-type Inequality

The classical Khinchine inequality states that for any $1 \le p < \infty$ there exist two constants $0 < A_p$ and $B_p < \infty$ such that

$$A_p(\sum_{i=1}^{n} x_i^2)^{1/2} \le (Ave_{\varepsilon_1,\dots,\varepsilon_n} =\pm 1 | \sum_{i=1}^{n} \varepsilon_i x_i |^p)^{\frac{1}{p}} \le B_p(\sum_{i=1}^{n} x_i^2)^{1/2} \tag{3}$$

holds true for every n and arbitrary choice of $x_1, \dots, x_n \in \mathbb{R}$.

In this section we show how Khinchine inequality can be realized without having to go through all 2^n summands in (3). We will insist that instead of going through all 2^n n-vectors of signs we use only N sign-vectors, where $N = (1+\delta)n$ and $0 < \delta < 1$ is any small positive number, and show that we can get inequalities like (3) loosing only in the constants. We know that one cannot do with less than n such vectors since ℓ_p and ℓ_2 are not isomorphic, and this means that the constants of isomorphism will depend on δ and explode as $\delta \to 0$.

Let us rewrite the inequality once again to make this clearer. For simplicity we only deal with the case $p = 1$; the same method works for all other $1 \le p \le 2$ (it is easy to see that $p = 1$ is the hardest case, because of monotonicity). We denote by $\varepsilon(j)$ an n-vector of ± 1, $\varepsilon(j) = (\varepsilon_{i,j})_{i=1}^n$. The average in (3) means summing over *all* possible vectors $\varepsilon(j)$, and there are 2^n of them. We wish to find vectors $\varepsilon(1), \dots, \varepsilon(N)$ such that

$$\frac{1}{N}\sum_{j=1}^{N} |\langle \varepsilon(j), x\rangle| \simeq |x|. \tag{4}$$

Notice that, obviously, this cannot be achieved by $\le n$ vectors since this would give an embedding of ℓ_2^n into ℓ_1^n. However, as we know that $\ell_1^{(1+\delta)n}$ *does* have isomorphic euclidean sections of dimension n (see [K]), it is conceivable that such an embedding can be constructed with a matrix of random signs.

This problem has a history. It was first shown by Schechtman in [S2] that the above is possible with a random selection of $N = Cn$ vectors, where C is a universal constant, and then repeated in [BLM] in a more general context including Kahane type generalization. Schechtman showed that for this quantity of vectors, if chosen randomly, (4) holds with universal constants, with exponentially large probability. The question then remained whether the constant C can be reduced to be close to 1. This was resolved by Johnson and Schechtman, and follows from their paper [JS]. However, they showed the *existence* of such vectors, and not that it is satisfied for *random* $N = (1+\delta)n$ sign-vectors. Very recently in a paper by Litvak, Pajor, Rudelson, Tomczak-Jaegermann and Vershynin [LPRTV] this was demonstrated. We reprove this result, using our method, getting slightly weaker dependence of the constants on δ. In a recent paper, joint with S. Sodin, we were able to significantly improve the dependence, from exponential in $(1/\delta)$ to polynomial, loosing only a little in the probability, see [AFMS].

One final remark is that even if we take an L_2 average instead of an L_1 average in formula (4) above, it is not correct that we can do with n random vectors alone. This is because, although the norm defined in (4) would be euclidean, it will correspond to some ellipsoid rather than to the standard ball D_n. This leads to the question of finding the smallest eigenvalue of an $(n \times n)$ matrix of random signs, which is itself an interesting question. Even the fact that with probability going to 1 exponentially fast such a matrix is *invertible* is a non trivial theorem due to Kahn, Komlós, and Szemerédy [KKS] (for a new improvement by Tao and Vu see [TV], see also [R]). The same question remains when one asks for smallest singular values of an $((1+\delta)n \times n)$ matrix of signs (where now the expectation of the smallest singular value is a constant depending on δ). This is also addressed in [LPRT], and follows also from our methods in the same way replacing $p = 1$ by $p = 2$. See also [AFMS] for better dependence on δ.

Our goal is to prove that with large probability on the choice of $N = (1+\delta)n$ vectors $\varepsilon(1), \varepsilon(2), \ldots, \varepsilon(N)$, where $\varepsilon(j) = (\varepsilon_{i,j})_{i=1}^n \in \{-1,1\}^n$, we have for every x the estimate (4) where the isomorphism constants depend only on $\delta > 0$. Throughout this section we demonstrate our method by proving the following Theorem.

Theorem 4. *For any $0 < \delta < 1$ there exists a constant $0 < c(\delta)$, depending only on δ and universal constants $0 < c', C < \infty$, such that for large enough n, for $N = (1+\delta)n$ random sign vectors $\varepsilon(1), \ldots, \varepsilon(N) \in \{-1,1\}^n$, with probability greater than $1 - e^{-c'n}$, one has for every $x \in \mathbb{R}^n$*

$$c(\delta)|x| \leq \frac{1}{N}\sum_{j=1}^{N} |\langle \varepsilon(j), x \rangle| \leq C|x|.$$

Remark 1 The constant $c(\delta)$ which our proof provides is $c(\delta) = (c_1\delta)^{1+2/\delta}$, where c_1 is an absolute constant. The constant in [LPRTV] is better: $c_1^{1/\delta}$. In [AFMS] we get a polynomial dependence on δ, but with a slightly worse exponent in the probability: $1 - e^{-c'\delta n^{1/6}}$.
Remark 2 It is easy to see that once you learn the theorem for small δ, it holds for large δ as well. This applies also to Theorem 7 and Theorem 11. Thus we may always assume that $\delta < \delta_0$ for some universal δ_0.

Before beginning the proof we want to remark on one more point. The technique we show below works for the $\ell_2^n \to \ell_2^N$ case as well, that is, to estimating the smallest singular number of an almost-square matrix. We present the proof for the $\ell_2^n \to \ell_1^N$ case, which is, even formally, more difficult. It is important to emphasize however that in the proof we do not *use* any known fact about the smallest singular number of the matrix (differently from what we do in [AFMS]). Thus, in fact, although proving $\ell_2^n \to \ell_1^N$ is formally more difficult, the main difficulty, and the reason for the exponentially bad bound that we get, lies primarily in the euclidean case. This section gives in particular another way to get lower bounds on smallest singular value of a random sign matrix using Chernoff bounds.

Proof of Theorem 4 We will denote $|||x||| = \frac{1}{N}\sum_{j=1}^{N}|\langle\varepsilon(j),x\rangle|$. This is a *random* norm depending on the choice of N sign vectors.

We need to estimate $\mathbb{P}[\varepsilon(1),\dots,\varepsilon(N) : \forall x \in S^{n-1}\ c \le |||x||| \le C]$. The following step is standard: this probability is greater than

$$1-\mathbb{P}[\exists x, |||x||| > C|x|] - \mathbb{P}[(\forall y, |||y||| \le C|y|) \textit{ and } (\exists x, |||x||| < c|x|)]. \qquad (5)$$

We begin by estimating $\mathbb{P}[\exists x \in S^{n-1}, |||x||| > C]$. This is relatively easy, and does not require a new method; we do it in a similar way to the one in [BLM]: Let $\mathcal{N} = \{x_i\}_{i=1}^{m}$ be a $\frac{1}{2}$-net of S^{n-1}, with $m \le 5^n$. For each $i = 1,\dots,m$ define the random variables $\{X_{i,j}\}_{j=1}^{N}$ by

$$X_{i,j} = |\langle\varepsilon(j),x_i\rangle|,$$

and denote $r = \mathbb{E}|\langle\varepsilon,x\rangle|$. It is obvious that $r \le |x| = 1$.

We use Proposition 2 from Section 1. It is well known that $X_{i,j}$ are ψ_2 random variables and $\|X_{i,j}\|_{\psi_2} \le c_3$ for some absolute constant $c_3 > 0$ (it follows from Khinchine inequality and the basic facts about ψ_2 random variables). Proposition 2 then implies that for every $t > 0$, and a fixed i, we have

$$\mathbb{P}[\varepsilon(1),\dots,\varepsilon(N) : \frac{1}{N}\sum_{j=1}^{N}X_{i,j} > r+t] \le 2e^{-t^2N/8c_3^2},$$

which in turn implies that (using that $r \le 1$) for a fixed point $x_i \in \mathcal{N}$ and any $t > 1$ we have

$$\mathbb{P}[\varepsilon(1),\dots,\varepsilon(N) : \frac{1}{N}\sum_{j=1}^{N}|\langle\varepsilon(j),x_i\rangle| > t] \le 2e^{-(t-1)^2N/8c_3^2}, \qquad (6)$$

We choose t so that $2e^{(-(t-1)^2N/8c_3^2)}5^n \le e^{-n}$, for example $t = 6c_3+1$. Then, with probability at least $1-e^{-n}$, for every $i = 1,\dots,m$,

$$\frac{1}{N}\sum_{j=1}^{N}|\langle\varepsilon(j),x_i\rangle| \le t.$$

We thus have an upper estimate for a net on the sphere. It is standard to transform this to an upper estimate on *all* the sphere (an important difference between lower and upper estimates). One uses consecutive approximation of a point on the sphere by points from the net to get that $|||x||| \le 2t = 12c_3+2$ for every $x \in S^{n-1}$. This completes the proof of the upper bound, where $C = 12c_3+2$ is our universal constant.

We now turn to the second term to be estimated in (5). Notice that when estimating this term we know *in advance* that the (random) norm $|||\cdot|||$ is bounded from above on the sphere. This is crucial in order to transform a lower bound on a net on the sphere to a lower bound on the whole sphere. For the lower bound we use our method, as described in Section 1, to estimate the following probability

$$\mathbb{P}[\big(\forall y \in S^{n-1}, |||y||| \leq C\big) \ \textit{and} \ \big(\exists x \in S^{n-1}, |||x||| < c\big)]. \tag{7}$$

Let us denote by $p_{x,\alpha}$ the probability that for a random $\varepsilon \in \{-1,1\}^n$ we have $|\langle \varepsilon, x\rangle| \geq \alpha$, where $\alpha > 0$ and x is some point on S^{n-1}:

$$p_{x,\alpha} := \mathbb{P}[|\langle \varepsilon, x\rangle| \geq \alpha]. \tag{8}$$

If "doing an experiment" means checking whether $|\langle \varepsilon, x\rangle| \geq \alpha$ (with ε a random sign vector) then for $|||x|||$ to be greater than some c it is enough that c/α of the experiments succeed.

Of course, we will eventually not want to do this on all points x on the sphere, but just some dense enough set. This set turns out to be slightly more complicated than usual nets, because of the estimates we get for $p_{x,\alpha}$, but the underlying idea is still the usual simple one.

We first estimate $p_{x,\alpha}$. In estimating this probability we will consider two cases. Notice that in the simple example of $x = (\frac{1}{2}, \frac{1}{2}, 0, \ldots, 0)$, for every $0 < \alpha < 1$ we have $p_{x,\alpha} = \frac{1}{2}$. This is not a very high probability, and if we look again at the estimate (2) we see that we cannot make use of the parameters (in this case, decreasing α) to increase the rate of decay. This is a bad situation, however this is the worst that can happen, as shown in Lemma 5. Moreover, for *most* points x (these will be points x with 'many' small coordinates), a much better estimate holds, which we present in Lemma 6. The proof of the following lemma is not difficult, and we include it for the convenience of the reader.

Lemma 5. *There exists a universal constant $\alpha_0 > 0$ such that for every $x \in S^{n-1}$ we have*

$$\mathbb{P}[|\langle \varepsilon, x\rangle| \geq \alpha_0] \geq 1/2 \tag{9}$$

where $\varepsilon \in \{-1,1\}^n$ is chosen uniformly.

Proof We prove this Lemma in two stages. First, assume that one of the coordinates of x is greater than or equal to α_0 (we later choose α_0, and it will be universal). Without loss of generality we may assume $x_1 \geq \alpha_0$. Then, using conditional probability

$$\mathbb{P}[|\sum_{i=1}^n \varepsilon_i x_i| \geq \alpha_0] \geq \frac{1}{2}\mathbb{P}[\sum_{i=2}^n \varepsilon_i x_i \geq 0] + \frac{1}{2}\mathbb{P}[\sum_{i=2}^n \varepsilon_i x_i \leq 0] = \frac{1}{2}.$$

This proves the statement in the case where one of the coordinates is greater than α_0. In the case where *all* the coordinates of x are smaller than α_0 we use the Berry-Esséen Theorem (see [Ha]), which will promise us that the distribution of the sum is close to gaussian, for which we can estimate the probability exactly. The theorem of Berry-Esséen states that for $X_1, X_2, \dots$ independent random variables with mean zero and finite third moments, setting $S_n = \sum_{j=1}^n X_j$ and $s_n^2 = \mathbb{E}(S_n^2)$ one has

$$\sup_t |\mathbb{P}[S_n \le s_n t] - \Phi(t)| \le C' s_n^{-3} \sum_{j=1}^n \mathbb{E}(|X_j|)^3 \tag{10}$$

for all $n \ge 1$, where C' is a universal constant and where $\Phi(t)$ is the gaussian distribution function, i.e., $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-s^2/2} ds$.

In our case, we let $X_j = \varepsilon_j x_j$, where ε_j's are independent ± 1 valued Bernoulli random variables. We are assuming that $\sum_{j=1}^n x_j^2 = 1$, and thus $s_n = 1$. Also, $\sum_{j=1}^n \mathbb{E}(|X_j|)^3 = \sum_{j=1}^n x_j^3$. Since we are in the case that for all j, $x_j < \alpha_0$, we have that $\sum_{j=1}^n \mathbb{E}(|X_j|)^3 \le \alpha_0$. Inequality (10) tells us that

$$\sup_t |\mathbb{P}[\langle \varepsilon, x \rangle \le t] - \Phi(t)| \le C' \alpha_0.$$

We choose once $t = \alpha_0$ and once $t = -\alpha_0$, and get

$$\begin{aligned} \mathbb{P}[|\langle \varepsilon, x \rangle| \le \alpha_0] &= \\ \mathbb{P}[\langle \varepsilon, x \rangle \le \alpha_0] - \mathbb{P}[\langle \varepsilon, x \rangle < -\alpha_0] &\le \\ \Phi(\alpha_0) - \Phi(-\alpha_0) + 2C'\alpha_0 &\le \frac{2\alpha_0}{\sqrt{2\pi}} + 2C'\alpha_0. \end{aligned}$$

We choose $\alpha_0 = \frac{1}{4(\frac{1}{\sqrt{2\pi}} + C')}$, then the sum is less than or equal to 1/2 and this completes the proof of Lemma 5. □

Looking above, one sees that in the case where the coordinates of x are small we can very much improve the estimate $\frac{1}{2}$ in the lemma, by decreasing α_0. In the next lemma we push further this point of view. We estimate (8) when not necessarily *all* the coordinates are small (smaller than a), but a significant "weight" of them, γ^2, is. We can interplay with these two parameters a and γ, where for a given x, the parameter a determines γ, however it is the ratio that enters the estimate.

This has recently been done independently by the group Litvak, Pajor, Rudelson and Tomczak-Jaegermann in [LPRT], and the reader can either adapt the proof above or refer to Proposition 3.2 in [LPRT] for the proof of the following Lemma.

Lemma 6. *Let $x \in S^{n-1}$ and assume that for $j = 1, \dots, j_0$ we have $|x_j| < a$, and that $\sum_{j=1}^{j_0} x_j^2 > \gamma^2$. Then for any $\alpha > 0$ one has*

$$\mathbb{P}[|\langle \varepsilon, x \rangle| > \alpha] \ge 1 - (\frac{2\alpha}{\sqrt{2\pi}} + 2C'a)/\gamma$$

where C' is the universal constant from (10).

We return now to the proof of Theorem 4; we need to estimate the probability in (7). Note that we can bound it in the following way for any choice of a and γ (both x and y below are assumed to be in S^{n-1}):

$$\begin{aligned}&\mathbb{P}[(\forall y, |||y||| \le C) \textit{ and } (\exists x \textit{ s.t.} |||x||| < c)] \le \\ &\quad \mathbb{P}[(\forall y, |||y||| \le C) \textit{ and } \left(\exists x \textit{ s.t.} \sum_{\{i:|x_i|\le a\}} x_i^2 > \gamma^2 \textit{ and } |||x||| < c\right)] + \\ &\quad \mathbb{P}[(\forall y, |||y||| \le C) \textit{ and } \left(\exists x \textit{ s.t.} \sum_{\{i:|x_i|\le a\}} x_i^2 \le \gamma^2 \textit{ and } |||x||| < c\right)].\end{aligned}$$

This type of decomposition is by now considered standard, we were introduced to it by Schechtman, who used a similar decomposition in his paper [S1]. It is also used in [LPRT]. We need to estimate these two probabilities, choosing a and γ in the right way. We start by estimating the easy part, which is the second probability (again, in (11) both x and y belong to S^{n-1}):

$$\mathbb{P}[(\forall y \ |||y||| \le C) \textit{ and } \left(\exists x \textit{ s.t.} \sum_{\{i:|x_i|\le a\}} x_i^2 \le \gamma^2 \textit{ and } |||x||| < c\right)] \tag{11}$$

If there exists $x \in S^{n-1}$ with $|||x||| < c$ and $\sum_{\{i:|x_i|\le a\}} x_i^2 \le \gamma^2$, then it is close to a vector with small support, let us denote it by $y = y(x)$. The vector $y(x)$ is defined as $y_i = 0$ when $|x_i| \le a$ and $y_i = x_i$ when $|x_i| > a$. Thus $|x - y| < \gamma$. Since $|y| \le |x| = 1$, it is clear that the support of y, the number of coordinates where y is non zero, cannot be larger than $[1/a^2]$. We prefer to use a normalized version, namely $y' = y/|y|$, which also has support no larger than $[1/a^2]$, is on the sphere, and satisfies

$$|y' - x| \le |y' - y| + |y - x| \le 1 - (1-\gamma^2)^{1/2} + \gamma \le 2\gamma.$$

In addition we know that $|||y'||| \le |||x||| + |||y' - x||| \le c + C|x - y'| \le c + 2C\gamma$.

We let $\mathcal{N}$ be a subset of S^{n-1} such that for every y' with $|y'| = 1$ and which is supported on no more than $[1/a^2]$ coordinates, there is a vector $v \in \mathcal{N}$ with $|y' - v| \le \theta_1$. (The parameter θ_1 will be chosen later.) For this we take a θ_1-net on each $[1/a^2]$-dimensional coordinate sub-sphere of S^{n-1}, and let $\mathcal{N}$ be the union of all these nets. We thus have $|\mathcal{N}| \le \binom{n}{[1/a^2]}(\frac{3}{\theta_1})^{[1/a^2]}$. If there exists x as above, and correspondingly y and y', then there exists $v \in \mathcal{N}$ with $|||v||| \le |||y'||| + |||v - y'||| \le c + 2C\gamma + C\theta_1$. Hence we can estimate probability (11) by

$$\mathbb{P}[\exists v \in \mathcal{N} : |||v||| \le c + 2C\gamma + C\theta_1]. \tag{12}$$

By Lemma 5, for a given $v \in \mathcal{N}$ (for any unit vector, for that matter) $p_{v,\alpha_0} = \mathbb{P}[|\langle \varepsilon, v\rangle| \ge \alpha_0] \ge \frac{1}{2}$. In order to estimate the probability in (12), we choose in our scheme $\beta = 1/4$ (so, it is smaller than p_{v,α_0}) to be the proportion of "trials" $\{|\langle \varepsilon, v\rangle| \ge \alpha_0\}$ we want to succeed. We want $\beta\alpha_0 \ge c + 2C\gamma + C\theta_1$, so we have to make sure that γ, θ_1 and c are small enough, each say less than $\alpha_0/20C$. At this point we *choose* both γ and θ_1 to be equal $\alpha_0/20C$. The choice of c is postponed to later on since in the second part of the proof we have some more conditions on it.

Proposition 1 gives that for a given v

$$\mathbb{P}[|||v||| \leq c + 2C\gamma + C\theta_1] \leq e^{-NI(\frac{1}{4},\frac{1}{2})}.$$

Combining this with the size of $\mathcal{N}$, and the trivial calculation for $I(\frac{1}{4},\frac{1}{2})$, we get that

$$\mathbb{P}[\exists v \in \mathcal{N} : |||v||| \leq c + 2C\gamma + C\theta_1] \leq \binom{n}{[1/a^2]}(\frac{3}{\theta_1})^{[1/a^2]}e^{-c''n} \tag{13}$$

for $c'' = \ln(3^{3/4}/2)$.

We want this probability to be very small, less than $\frac{1}{2}e^{-c'n}$. Thus we get a restriction on a which is very mild (θ_1 has already been chosen), which we keep in mind for the time when we choose the constants. (The parameter a will later be chosen to be a small constant depending only on δ, and since n is assumed to be large, this condition will automatically be satisfied.)

We turn now to the more difficult task of estimating (again, x and y are assumed to be in S^{n-1}):

$$\mathbb{P}[(\forall y \ |||y||| \leq C) \ and \ \left(\exists x \ s.t. \sum_{\{i:|x_i|\leq a\}} x_i^2 > \gamma^2 \ and \ |||x||| < c\right)]. \tag{14}$$

Let $\mathcal{N}$ be this time a θ-net on S^{n-1}, θ is yet another parameter we will choose later on. We can find one with cardinality $\leq (\frac{3}{\theta})^n$. We bound (14) by

$$\mathbb{P}[\exists v \in \mathcal{N}' \ s.t. \ |||v||| < c + C\theta] \tag{15}$$

where $\mathcal{N}' = \{v \in \mathcal{N} : \sum_{\{i:|v_i|\leq a+\theta\}} v_i^2 \geq (\gamma - \theta)^2\}$. Indeed, if there exists $x \in S^{n-1}$ such that $\sum_{\{i:|x_i|\leq a\}} x_i^2 > \gamma^2$ and $|||x||| < c$ then there is a vector $v \in \mathcal{N}$ such that $|x - v| \leq \theta$ and we have $|||v||| \leq |||x||| + |||x - v||| < c + C\theta$. Also, all the coordinates i for which $|x_i| \leq a$ satisfy of course $|v_i| \leq a + \theta$, and the square root of the sum of squares of these coordinates for v cannot differ by more than θ from the square root of the sum of squares of these coordinates for x. Therefore when taking squares the difference is at most $(\gamma - \theta)^2$. Hence if for the norm $|||\cdot|||$ there exist an $x \in S^{n-1}$ for (14), then there exists also some $v \in \mathcal{N}'$ for (15). By Lemma 6, for a given $v \in \mathcal{N}'$ we have for any $\alpha > 0$ that

$$p_{v,\alpha} = \mathbb{P}[|\langle \varepsilon, v\rangle| \geq \alpha] \geq 1 - (\frac{2\alpha}{\sqrt{2\pi}} + 2C'(a + \theta))/(\gamma - \theta).$$

We return to our scheme, in order to estimate the probability in (15). Assume $\beta\alpha \geq c + C\theta$ (where β will be the portion of good trials out of N according to our scheme, and α another constant we later choose); Proposition 1 together with the estimate (2) gives that for a given v

$$\mathbb{P}[|||v||| \le c + C\theta] \le 2^N (1 - p_{v,\alpha})^{(1-\beta)N},$$

and so for a given $v \in \mathcal{N}'$ we can estimate

$$\mathbb{P}[|||v||| \le c + C\theta] \le (2(\frac{(2\alpha/\sqrt{2\pi}) + 2C'(a+\theta)}{(\gamma-\theta)})^{(1-\beta)})^N.$$

Combining this with the size of $\mathcal{N}'$ (which is at most the size of $\mathcal{N}$) we get that

$$\mathbb{P}[\exists v \in \mathcal{N}' : |||v||| \le c + C\theta] \le (\frac{3}{\theta})^n (2(\frac{(2\alpha/\sqrt{2\pi}) + 2C'(a+\theta)}{(\gamma-\theta)})^{(1-\beta)})^N.$$

We choose β such that $(1-\beta)(1+\delta)n = (1+\frac{\delta}{2})n$, (so, $\beta = \frac{\delta}{2(1+\delta)}$) thus we have (remembering that $N = (1+\delta)n$) that

$$\mathbb{P}[\exists v \in \mathcal{N}' \ s.t. \ |||v||| \le c + C\theta] \le$$
$$[(\frac{3}{\theta})2^{(1+\delta)} \frac{(2\alpha/\sqrt{2\pi}) + 2C'(a+\theta)}{(\gamma-\theta)}]^n \cdot [\frac{(2\alpha/\sqrt{2\pi}) + 2C'(a+\theta)}{(\gamma-\theta)}]^{\frac{\delta}{2}n}$$

We are now in the place to choose all the various constants. We let $a = c = \theta$. As θ will soon be chosen very small, smaller than $\gamma/2$ (which was already specified in the first part) we have that $\gamma - \theta$ is bounded from below by a universal constant $\alpha_0/40C$. We need to make sure that $\beta\alpha \ge c + C\theta$, so we let $\alpha = 12C\theta/\delta$. What we get, so far, without choosing θ yet, is that

$$\mathbb{P}[\exists v \in \mathcal{N}' \ s.t. \ |||v||| \le c + C\theta] \le (\frac{C_1}{\delta})^n \cdot (\frac{C_2\theta}{\delta})^{\frac{\delta}{2}n}$$

for universal constants C_1 and C_2. To make this probability less than $\frac{1}{2}e^{-c'n}$ we choose $\theta \le (\frac{1}{2}e^{-c'}\frac{\delta}{C_1})^{2/\delta}\frac{\delta}{C_2}$ and the proof of the estimate for the probability (14), and of the whole of Theorem 4, is complete. □

3.2 Euclidean sections of ℓ_1^N generated by isotropic convex bodies

The second application we present also deals with Khinchine-type inequalities, this time when the matrix elements are chosen differently. The conceptual difference is that they are not all independent anymore.

Instead of considering the norm of the form given in (4), with N random sign vectors, we do the same but with vectors distributed uniformly in some isotropic convex body K (just as in (4) they were distributed uniformly in the discrete cube). By isotropic we mean that K satisfies $Vol(K) = 1$, $\int_K x = 0$ and, most importantly, for every $\theta \in S^{n-1}$ the integral $\int_K \langle x, \theta\rangle^2$ is a constant independent of θ, depending only on K, which is called the (square of the) isotropic constant of K and denoted L_K^2. It is easy to check that every body has a linear image which is isotropic. In other words, saying that the body is in isotropic position only means that we identify the right euclidean structure with which to work.

We want to check, as in Section 3.1, how close the randomly defined norm $\frac{1}{N}\sum_{j=1}^N |\langle z_j, x\rangle|$ is to being euclidean, when the points z_j are chosen randomly with respect to the volume distribution in K. We prove the following theorem.

Theorem 7. *For any $0 < \delta < 1$ there exist a constants $0 < c(\delta)$, depending only on δ and universal constants $0 < c', C < \infty$ such that for large enough n, for any convex body $K \subset \mathbb{R}^n$ in isotropic position, with probability greater than $1 - e^{-c'n}$ we have that*

$$c(\delta)L_K|x| \le \frac{1}{N}\sum_{j=1}^N |\langle z_j, x\rangle| \le CL_K|x|,$$

where $N = (1+\delta)n$ and z_j are chosen independently and uniformly inside the body K.

Proof We begin with the upper estimate. As explained before in this paper, upper bounds usually present less difficulties, and the use of Chernoff bounds is not needed. When a point z is chosen uniformly inside a convex body, the distribution of the random variable $\langle x, z\rangle$ (where x is some fixed point) is *not necessarily* a ψ_2 distribution. For example for the unit ball of ℓ_1^n and the point $x = (1, 0, 0, \ldots, 0)$, the decay of the distribution function is only exponential and not gaussian. This is the worst that can happen though. We say that a random variable X has ψ_1 behavior if there exists a constant λ such that $\mathbb{E}e^{\frac{X}{\lambda}} \le 2$. The smallest λ for which this inequality holds is what we call the ψ_1 norm of X. The following Lemma (resulting from the work of C. Borell) shows that our random variables are always ψ_1 (for proof see [MS] Appendix III and [GiM2] Section 1.3 and Lemma 2.1)

Lemma 8. *There exists a universal constant C' such that for any isotropic convex body K, and any direction $\theta \in S^{n-1}$ the random variable $X = |\langle \theta, z\rangle|$ where z is chosen uniformly in K has ψ_1 distribution and its ψ_1 norm is equivalent to L_K and to its expectation, namely*

$$L_K \le \|X\|_{\psi_1} \le C'\mathbb{E}X \le {C'}^2 L_K.$$

We thus need a proposition of the like of Proposition 2 but for ψ_1 distributions and it is the following, the proof of which is standard, in the same lines of the inequality in Proposition 2.

Proposition 9. *Let $\{X_j\}_{j=1}^N$ be i.i.d. copies of the random variable X. Assume that X is ψ_1 and that the ψ_1 norm of X is smaller than some constant A. Then for any t,*

$$\mathbb{P}[|\frac{1}{N}\sum_{j=1}^N X_j - \mathbb{E}X| > t] \le 2e^{-Nt/(3A)}. \tag{16}$$

Thus, for $t = C''L_K$ with C'' large enough, this probability is enough to take care of a $1/2$ net of the sphere, and then by successive approximation one has an upper bound for the whole sphere.

We turn to the lower bound, where we will use our method. We need, as usual, to estimate the probability $\mathbb{P}[z \in K : |\langle x, z\rangle| < L_K\alpha]$. This is done in the following Proposition:

Proposition 10. *There exists a universal constant C_1 such that for any $\alpha > 0$ and for any symmetric isotropic convex body K, for every direction $u \in S^{n-1}$*

$$\mathbb{P}[x \in K : |\langle x, u\rangle| < L_K\alpha] < C_1\alpha.$$

Proof We use two well known facts from Asymptotic Geometric Analysis. First, all central sections of an isotropic convex body have volume $\approx \frac{1}{L_K}$. Second, for a centrally symmetric convex body K and a direction u, of all sections of K by hyperplanes orthogonal to u, the one with the largest volume is the central section (for proofs see e.g. the survey [GiM1]). In particular the two facts imply that there exists some universal constant C_1 such that for any direction u, any section of K orthogonal to u has $(n-1)$-dimensional volume $\leq \frac{C_1}{2L_K}$. Now use Fubini Theorem to get that $\mathbb{P}[x \in K : |\langle x, u\rangle| < L_K\alpha] < C_1\alpha$. □

Notice that, differently from what was going on in Section 3.1, here for *any* point x, we can make the probability as small as we want by reducing α. This allows us to use just one simple net: take a θ-net $\mathcal{N}$ in S^{n-1}, with less than $(\frac{3}{\theta})^n$ points x_i. Define the random variables $X_{i,j} = |\langle z_j, x_i\rangle|$. We know that for $\beta < 1 - C_1\alpha$ (which is hardly a restriction, α will be very small and so will β) we have

$$\mathbb{P}[\frac{1}{N}\sum_{j=1}^{N} X_{i,j} > \beta L_K\alpha] \geq 1 - e^{-NI(\beta, 1-C_1\alpha)}.$$

We choose β so that $(1+\delta)(1-\beta) = (1+\delta/2)$, hence $\beta = \frac{\delta}{2(1+\delta)}$. We choose $\theta = \beta\alpha/2C$, where C comes from the upper bound (which is CL_K). To make sure that the probability that the above holds for all points in the net we ask that

$$(\frac{3}{\theta})^n 2^N (C_1\alpha)^{(1+\frac{\delta}{2})n} \leq \frac{1}{2}e^{-c'n}.$$

For this we choose $\alpha = (C_2\delta)^{\frac{2}{\delta}}$ for some universal C_2, and get the lower bound for each point of the θ-net of S^{n-1}. Now using the upper bound, for every $x \in S^{n-1}$ we have for some i that (denoting $|||x||| = \frac{1}{N}\sum_1^N |\langle x, z_i\rangle|$)

$$|||x||| \geq |||x_i||| - |||x - x_i||| \geq \beta L_K\alpha - \theta C L_K.$$

Thus the proof of the lower bound, and of Theorem 7, is complete. □

3.3 Reducing the level of randomness

Another variant of the question answered in Section 3.1 which we discuss in this section is related to a more "explicit" construction of n-dimensional euclidean sections of $\ell_1^{(1+\delta)n}$. In Section 3.1 we described Schechtman's question about realizing such a euclidean section by the image of a random sign matrix. In a different paper, [S1], Schechtman showed that for $\delta = 1$, that is, a $2n \times n$ matrix, one can take the upper half to be the identity matrix, and the lower to be n random sign vectors, and this gives an isomorphic euclidean section of ℓ_1^{2n}. Using this method we can also take only the identity with only δn additional random sign vectors (so, get a section of $\ell_1^{(1+\delta)n}$), and the isomorphism constant will depend on δ. Below we present a similar construction, in which we use our method to show that when the upper half (that is, the first n vectors) is a Hadamard matrix, namely a matrix of signs whose rows are orthogonal, and add to it δn random sign vectors below, you also get an isomorphic euclidean section of $\ell_1^{(1+\delta)n}$.

Remark While it is not known precisely for which n a Hadamard matrix exists (the Hadamard conjecture is that they exist for $n = 1, 2$ and all multiples of 4), it is known that the orders of Hadamard matrices are dense in the sense that for all ε if n is sufficiently large there will exist a Hadamard matrix of order between n and $n(1-\varepsilon)$. However, we only use the fact that the first n rows are an orthonormal basis of $\mathbb{R}^n$ and Theorem 11 below holds if we replace the Hadamard matrix by any other orthonormal matrix (normalized properly). For more information on Hadamard matrices we refer the reader to [H]. We chose Hadamard matrices since this way the section we get is generated by a sign matrix.

Denote the rows of the $n \times n$ Hadamard matrix W_n by $\frac{1}{\sqrt{n}}\varepsilon(j)$ for $j = 1, \dots, n$. They form an orthonormal basis of $\mathbb{R}^n$. We prove below that by adding the random sign vectors $\varepsilon(n+1), \dots, \varepsilon(n+\delta n)$ we get a matrix which gives an isomorphic euclidean section of $\ell_1^{(1+\delta)n}$. We prove

Theorem 11. *Let $0 < \delta < 1$, and denote $N = (1+\delta)n$. There exists a constant $c(\delta)$ depending only on δ, and universal constants c', C, such that for large enough n, with probability $1 - e^{-c'\delta n}$, for δn random sign-vectors $\varepsilon(j) \in \{-1,1\}^n$, with $j = n+1, \dots, n+\delta n$, one has for every $x \in \mathbb{R}^n$*

$$c(\delta)|x| \le \frac{1}{N}\sum_{j=1}^{N} |\langle x, \varepsilon(j)\rangle| \le (1 + \sqrt{\delta}C)|x|, \tag{17}$$

where one may take $c(\delta) = c_1\delta^{3/2}/(1+\ln(1/\delta))$ for a universal c_1.

Proof Since $\frac{1}{\sqrt{n}}\varepsilon(1), \dots, \frac{1}{\sqrt{n}}\varepsilon(n)$ is an orthonormal basis of $\mathbb{R}^n$, every $x \in S^{n-1}$ can be uniquely written as $x = \frac{1}{\sqrt{n}}\sum a_i\varepsilon(i)$, and $a_i = \frac{1}{\sqrt{n}}\langle x, \varepsilon(i)\rangle$. So, $\sum a_i^2 = 1$, and $a = (a_i)_{i=1}^n \in S^{n-1}$ depends on x. Our aim is to show that inequality (17) holds. We can rewrite it as

$$c(\delta)|x| \le \frac{1}{(1+\delta)}\frac{1}{n}\sum_{j=1}^{n}|\langle x,\varepsilon(j)\rangle| + \frac{1}{N}\sum_{j=1}^{\delta n}|\langle x,\varepsilon(n+j)\rangle| \le (1+\sqrt{\delta}C)|x|. \quad (18)$$

Fix $x \in S^{n-1}$. To prove the upper bound, first notice that the first summand satisfies

$$\frac{1}{(1+\delta)}\frac{1}{\sqrt{n}}\sum|a_i| \le \frac{1}{(1+\delta)}\sqrt{\sum a_i^2} = \frac{1}{(1+\delta)}.$$

As for the second one, we can use a standard upper bound approach as in Section 3.1. Notice that the second term is in fact

$$[\frac{\delta}{1+\delta}]\frac{1}{\delta n}\sum_{j=1}^{\delta n}|\langle x,\varepsilon(n+j)\rangle|,$$

so the upper bound we would expect for this part is $\simeq \delta C$. However, this is not true, since if we would go ahead trying to prove this, the probability we would get for an individual x to satisfy this would be $1-e^{-c\delta n}$ and this is not enough to take care of say a 1/2-net of the sphere. Thus, we need to take a larger deviation in order to increase the probability. Take a 1/2-net $\mathcal{N}$ of S^{n-1}, then taking $t = C/(2\sqrt{\delta})$ in inequality (6) with $N = \delta n$ we get that for a fixed $x \in \mathcal{N}$

$$\mathbb{P}[\frac{1}{\delta n}\sum_{j=1}^{\delta n}|\langle x,\varepsilon(n+j)\rangle| \le C/(2\sqrt{\delta})|x|] \ge 1 - e^{-\delta n(\frac{C}{2\sqrt{\delta}}-1)^2/c_4}$$

for some universal c_4. For large enough C, this probability is enough to take care of the whole 1/2-net, and by successive approximation we get that with high probability $1-e^{-cn}$ we have for every x

$$[\frac{\delta}{1+\delta}]\frac{1}{\delta n}\sum_{j=1}^{\delta n}|\langle x,\varepsilon(n+j)\rangle| \le \sqrt{\delta}C|x|. \quad (19)$$

The bound for the whole expression is thus as wanted, and in fact we will later use the bound $\sqrt{\delta}C$ for the second term separately.

For the lower bound, denote

$$A_\gamma = \{x \in S^{n-1} : \frac{1}{\sqrt{n}}\sum_{i=1}^{n}|a_i| \le \gamma\}.$$

If $x \notin A_\gamma$ then in inequality (17) we have

$$\frac{1}{N}\sum_{j=1}^{N}|\langle x,\varepsilon(j)\rangle| \ge \frac{1}{(1+\delta)}\frac{1}{n}\sum_{j=1}^{n}|\langle x,\varepsilon(j)\rangle| \ge \gamma/(1+\delta)$$

and so a lower bound of the order $\gamma/(1+\delta)$ holds. We want to choose γ so that all $x \in A_\gamma$ are taken care of by the δn random sign vectors, that is, by the right hand side term in equation (18).

We need the following observation: Let $\alpha < 1$ be some proportion. If $\frac{1}{\sqrt{n}}\sum_{i=1}^n |a_i| \le \gamma$, denote by a_{i_0} the term a_i which is in absolute value the (αn)'st largest one. Then

$$\gamma \ge \frac{1}{\sqrt{n}}\sum_{i=1}^n |a_i| \ge \frac{1}{\sqrt{n}}\sum_{\alpha n\ biggest} |a_i| \ge \frac{\alpha n}{\sqrt{n}} \cdot |a_{i_0}| \ge \alpha\sqrt{n}\left(\frac{1}{(1-\alpha)n}\sum_{i\in I}|a_i|^2\right)^{1/2}$$

where I is the set of the $(1-\alpha)n$ coordinates a_i which are smallest in absolute value. Thus for some set I of coordinates, with $|I| = (1-\alpha)n$, we have $(\sum_{i\in I}|a_i|^2)^{1/2} \le \gamma\frac{\sqrt{1-\alpha}}{\alpha}$.

We let E stand for a subspace spanned by αn of the (normalized) Hadamard basis row vectors $\varepsilon(1),\ldots,\varepsilon(n)$. The observation tells us that every $x \in A_\gamma$ can be written as $x = y + z$ with y in some such E, and $|z| < \gamma\frac{\sqrt{1-\alpha}}{\alpha}$. We will choose α so that the δn additional random vectors take care of *all* vectors in *all* the E's, with a lower bound c''. We will then choose γ such that $\gamma\frac{\sqrt{1-\alpha}}{\alpha}C\sqrt{\delta} \le c''/2$ (where C is from the upper bound in (19)) and by this we will finish, since then

$$\frac{1}{N}\sum_{j=1}^{\delta n}|\langle x, \varepsilon(n+j)\rangle| \ge \frac{1}{N}\sum_{j=1}^{\delta n}|\langle y, \varepsilon(n+j)\rangle| - C\sqrt{\delta}|z| \ge c''/2.$$

(So, we will have a lower bound $c(\delta) = \min(\gamma, c''/2)$.) We make sure that $\gamma\frac{\sqrt{1-\alpha}}{\alpha} < 1/2$, so that $|y| > 1/2$.

We thus have to find α and c'' such that for a set of $\binom{n}{\alpha n}$ subspaces E of dimension αn we have for all $y \in E \cap S^{n-1}$ that

$$\frac{1}{N}\sum_{j=1}^{\delta n}|\langle y, \varepsilon(n+j)\rangle| \ge c''.$$

We take a θ-net on this set (the value of θ will be chosen later). Its cardinality is less than $\binom{n}{\alpha n}(\frac{3}{\theta})^{\alpha n}$, this is $\le (\frac{e}{\alpha})^{\alpha n}(\frac{3}{\theta})^{\alpha n}$. For a single y in the net we estimate the probability that

$$\frac{1}{N}\sum_{j=1}^{\delta n}|\langle y, \varepsilon(n+j)\rangle| \ge c''$$

by our usual method. The probability for a single experiment $|\langle y, \varepsilon\rangle| \ge \alpha_0|y|$ is bounded below, for a suitably chosen α_0, by $1/2$, from Lemma 5. Choose, say $\beta = 1/4$, and just as in previous sections

$$\mathbb{P}[\frac{1}{N}\sum_{j=1}^{\delta n}|\langle y, \varepsilon(n+j)\rangle| > \frac{\delta}{1+\delta}\beta\alpha_0/2] \ge 1 - e^{-2c'\delta n}$$

(this is our *definition* of c'). We choose, say, $\theta = \frac{\alpha_0\sqrt{\delta}}{100C}$ (since the upper bound for the second part in (18) is $\sqrt{\delta}C$ and so we are able to transfer the bound from the net to the whole set) and then we choose α such that

$$(\frac{3e}{\alpha\theta})^{\alpha n} e^{-2\delta n c'} \leq e^{-c'\delta n}.$$

This holds if $\alpha \leq c_2\delta/(1 + \ln\frac{1}{\delta})$ for some universal c_2. This finally gives, say, $c'' = \delta\alpha_0/16$. We still have to return to γ, which we can choose to be $\gamma = \alpha c''/(4C\sqrt{\delta})$, and this is the order of the lower bound we achieve, $c(\delta) = c_1\delta^{3/2}/(1 + \log(1/\delta))$. □

4 Dvoretzky-type theorems

In this Section we deal with a different question, namely with a global Dvoretzky-type theorem. We will first illustrate yet another application where our method works, reproving a well known version of the global Dvoretzky Theorem. First, we will show how the upper bound can be obtained using Chernoff's inequalities and standard concentration on the sphere. This is different from the standard way of proof for global Dvoretzky Theorem (which we also indicate below), where usually the upper bound is obtained by a deep geometric argument about concentration on the product of spheres, (inequality (20) below). The lower bound we then obtain by using our Chernoff scheme.

We will then state, as a conjecture, a natural strengthening of the global Dvoretzky Theorem (which would be optimal), the local analogue of which is known to hold. We show how this strengthened theorem would be implied by the validity of a small ball probability conjecture of Latała and Oleszkiewicz [LO].

In the last part of the section we discuss an alternative parameter that is of interest, and is similar to a parameter introduced by Klartag and Vershynin in [KV], and which clarifies some other cases where the global Dvoretzky-type theorem holds in an improved form.

4.1 About Global Dvoretzky's Theorem

The global analogue of Dvoretzky's Theorem first appeared in [BLM], in a non explicit form, and explicitly in [MS2], and is the following Theorem, which, by duality, means that the Minkowski average of $C'(\frac{a}{M^*})^2$ random rotations of a convex body K with radius a and mean width M^* is isomorphic to a euclidean ball of radius M^*.

Theorem 12. *There exist universal constants c, c', C and C' such that for every symmetric convex body $K \subset \mathbb{R}^n$ satisfying $\frac{1}{b}D \subseteq K$, letting $M = M(K)$, we have with probability $1 - e^{-c'n}$, that the $N = C'(\frac{b}{M})^2$ random orthogonal* *transformations $U_1, \ldots, U_N \in O(n)$ satisfy for every $x \in \mathbb{R}^n$ that*

$$cM|x| \leq \frac{1}{N}\sum_{i=1}^{N} \|U_i x\|_K \leq CM|x|.$$

Remark We later show that in fact the constant C' above can be chosen to be $(4+\delta)$ for any $\delta > 0$, and then all other constants depend on δ. We also later conjecture that in fact $(1+\delta)$ for any $\delta > 0$ should be the optimal constant.

It is clear that we are dealing with a lower and an upper bound for a sum of random variables. It is also clear what our experiments will be: for a random orthogonal transformation U_j, there is some fixed probability (for a given $x \in S^{n-1}$) that $\|U_j x\| \geq \alpha M$. We say that the experiment is a success if this happens. In fact, taking $\alpha = 1$ and taking M to be the median of the norm instead of its expectation (they are very close, see [MS]), this probability is exactly $1/2$. If at least $1/4$ of the trials succeed, we get the average above to be at least $M/4$. This can be thought of as the main idea, however, we need something stronger in order to get that $N \simeq (b/M)^2$ rotations are enough, and this naive approach will only give $N \simeq n$. which is typically much larger.

4.1.1 The upper bound, using concentration on the product of spheres

We start with the upper bound. The upper bound is usually handled with the estimate (see 6.5.2 in [MS]): Fix $x \in S^{n-1}$, then for random $U_j \in O(n)$, $j = 1, \ldots, N$, and $t > 0$ we have

$$\mathbb{P}[(U_1, \ldots, U_N) : |\frac{1}{N}\sum_{j=1}^{N} \|U_j x_i\| - M| > tM] \leq \sqrt{\frac{\pi}{2}} e^{-t^2 N (\frac{M}{b})^2 \frac{n-2}{2}}, \qquad (20)$$

which is a concentration result on the product of N spheres.

Concentration on the product of n spheres is quite a strong tool, and at first glance this seems appropriate since we are searching for a strong result: not a sum of $N \simeq n$ variables, but much less (typically), $N \simeq (b/M)^2$. In what follows we will several times use the well known and easily provable fact that $b \leq \sqrt{n}M$. To complete the upper bound using (20) we simply take a 1/2-net on the sphere, with at most 5^n points x_i. For each i we use (20) with, say, $t = 4$, and get that with probability at least $1 - 5^n \sqrt{\frac{\pi}{2}} e^{-8N(\frac{M}{b})^2(n-2)}$ we have for every x_i in the net that

$$\frac{1}{N}\sum_{j=1}^{N} \|U_j x_i\| < 5M.$$

We clearly see that if $N \geq (\frac{b}{M})^2$, the probability above is exponentially close to 1. Passing from a net on the sphere to the whole sphere, in an upper bound, is standard, and may be done by successive approximation, which gives us that for every $x \in \mathbb{R}^n$

$$\frac{1}{N}\sum_{j=1}^{N}\|U_j x_i\| < 10M.$$

Before moving to the lower bound, we would like to offer an alternative proof for the upper bound, which does not use (20) directly, but gives a proof of a slightly weaker estimate (which is sufficient for our needs) by using only Chernoff's bounds. We remark that (20), which is concentration on the product of n spheres, is a much deeper fact than the concentration estimate on the sphere, see [GrM].

4.1.2 The upper bound, avoiding concentration on the product of spheres

In this paper, up till now, we have mostly shown how the use of Chernoff bounds is useful in obtaining lower bounds, where the standard large deviation technology was not enough. Below we will show how standard concentration on the sphere, together with Chernoff bounds, provides an alternative proof for the upper bound. This approach was pursued further in the paper [Ar], and one of its merits is that it is quite robust.

We will use concentration on the sphere which states that

Lemma 13. *For $t > 0$*

$$\sigma\left(x \in S^{n-1} : |\|x\| - M| \geq tM\right) \leq \sqrt{\frac{\pi}{2}}e^{-t^2(\frac{M}{b})^2\frac{n-2}{2}}, \tag{21}$$

and is simply the case $N = 1$ of inequality (20).

Fix $x \in S^{n-1}$, and denote

$$A_j = \{U \in O(n) : 2^j M < \|Ux\| \leq 2^{j+1}M\},$$

where $j = t, t+1, \ldots, \log(\frac{b}{M})$, for an integer $t \geq 2$. By Lemma 13 we thus have $\mathbb{P}(U \in A_j) \leq \sqrt{\frac{\pi}{2}}e^{-(2^j-1)^2(\frac{M}{b})^2(n-2)/2}$. We also denote $m_j = N2^{-j}/j^2$. If out of the N transformations $U_1, \ldots, U_N$, for every $j \geq t$, less than m_j of them belong to A_j then

$$\frac{1}{N}\sum_{i=1}^{N}\|U_i x\| \leq [2^t + \sum_{j=t}^{\log(\frac{b}{M})} 2^{j+1}\frac{m_j}{N}]M \leq (2^t + 2)M.$$

Fix some $j \geq t$. We use Chernoff's Proposition 1 to bound from above the probability that more than m_j of the N transformations are in A_j. For us now $p = \sqrt{\frac{\pi}{2}}e^{-(2^j-1)^2(\frac{M}{b})^2(n-2)/2}$ and $\beta = 2^{-j}/j^2$, and in particular $\beta > p$ since $j \geq 2$. Our scheme implies that this probability is bounded by $(2\sqrt{\frac{\pi}{2}})^N e^{-N(\frac{M}{b})^2(n-2)(2^j-2)/2j^2}$.

Adding these expressions up for $j = t, t+1, \ldots$ we get that

$$\mathbb{P}[(U_1, \ldots, U_N) : \frac{1}{N}\sum_{i=1}^{N}\|U_i x\| > (2^t + 1)M] \leq e^{-c'N(\frac{M}{b})^2(n-2)}$$

for some c' depending only on t, as long as t is bigger than some universal constant. (Where we use, as usual, that always $(b/M) \leq \sqrt{n}$.) Thus by taking N above to be say $2\ln 5(\frac{M}{b})^{-2}/c'$ we get the upper bound $(2^t+2)M$ for a whole (1/2)-net. Successive approximation then gives the upper bound CM with $C = 2^{t+1}+4$. Recall that t can be chosen to be anything above some universal constant C_0. Enlarging t will make the probability better, which means we can take N as any constant proportion, even < 1, of $(\frac{M}{b})^{-2}$, and have an upper bound depending on this proportion. We have thus proved the upper bound, using only standard concentration, and Chernoff.

4.1.3 The lower bound

One crucial point in the lower bound's proof is estimating the probability of a success in a specific experiment, that is, $\mathbb{P}[U \in O(n) : \|Ux\| \geq \alpha M]$. What we usually need, is to be able to decrease this probability significantly by sending α to 0. The standard concentration argument on the sphere, such as Lemma 13, gives that for $\alpha < 1$ and a fixed $x \in S^{n-1}$

$$\mathbb{P}[U \in O(n) : \|Ux\| \geq \alpha M] \geq 1 - \sqrt{\frac{\pi}{8}} e^{-(1-\alpha)^2(\frac{M}{b})^2\frac{n-2}{2}}. \qquad (22)$$

This is enough for proving Theorem 12 with a universal C', but sending α to 0 does not help to change the rate of decrease of the probability in (22), and this is the reason for not getting the conjectured (below) optimal constant.

To complete the proof of the lower bound we take again a net on the sphere, this time an $(1/4C)$-net where CM is the upper bound which we already know from either one of the two previous subsections. We use (22) and our scheme with $\alpha = 1/2$ and $\beta < \alpha$ (small) to be specified later, and have that for a given x in the net, with high probability, more than βN of the operators U_j satisfy $\|U_j x\| \geq M/2$, more precisely

$$\mathbb{P}[U_1 \ldots, U_N \in O(n) : \frac{1}{N}\sum_{j=1}^{N} \|U_j x\| \geq \alpha\beta M] \geq$$
$$1 - 2^{-u(\beta)N}\left(\sqrt{\frac{\pi}{8}} e^{-\frac{1}{8}(\frac{M}{b})^2(n-2)}\right)^{(1-\beta)N}.$$

We see that if $N = C'(\frac{b}{M})^2$ this probability is greater than

$$1 - 2^{-u(\beta)C'(b/M)^2} e^{-(1-\beta)(n-2)/8},$$

and so for $\beta = \beta_0$ for some universal β_0, and for large enough C' (which, notice, depends on the bound C we have achieved before), we can have this probability so big that it happens simultaneously for the whole net (and even this we can make sure happens with exponentially high probability). Now with use of the upper bound and the inverse triangle inequality we transfer the estimate to the whole sphere, and the proof is complete. □

4.2 With conjectured small-ball probability estimate

In this section we discuss the following conjecture which is different from Theorem 12 by specifying the constant C' in that theorem.

Conjecture 14. *For every δ there exists a constant $c(\delta)$ depending only on δ, and universal constants c' and C, such that for every symmetric convex body $K \subset \mathbb{R}^n$ satisfying $\frac{1}{b}D \subseteq K$, and with $M = M(K)$ there exist $N = (1+\delta)(\frac{b}{M})^2$ orthogonal transformations $U_1, \dots, U_N \in O(n)$ such that*

$$c(\delta)M|x| \leq \frac{1}{N}\sum_{i=1}^{N} \|U_i x\|_K \leq CM|x|. \tag{23}$$

Remark For large $\delta > 0$, this is Theorem 12. Also, we may assume $\delta < \delta_0$ for some universal δ_0 since if we prove Conjecture 14 for such δ it will then follow from standard arguments that the same holds for all $\delta > 0$. Finally, this is the best we can hope for in the general case, in the sense that using less than $N = (\frac{b}{M})^2$ transformations U_i, we cannot expect the average always to be isomorphic to euclidean, as is implied by the example of a cylinder with basis of dimension $(M/b)^2 n$ (see a parallel local version in [GMT]; the local version also follows from an earlier result by Gordon [Go]).

To achieve such an estimate for N we need a stronger estimate than (22). Such an estimate was conjectured (for different applications) by Vershynin, and reformulated by Latała and Oleszkiewicz with an extra non-degeneration condition, as follows (see [LO], Conjecture 1 and its Corollaries). Below we formulate a variant of their conjecture, which was originally formulated in the Gaussian context, but the translation is straightforward. Notice that we formulate a variant with M being the mean of the norm whereas in [LO] the median is used;

Conjecture 15. *For every constant $\kappa < 1$ there exists universal constants $C' = C'(\kappa)$, $c_0 = C_0(\kappa)$ and $w_0 = w_0(\kappa) > c_0(\kappa)$ such that if for some norm we have $(b/M)^2 \leq n/w_0$ then for any $\alpha < 1$,*

$$\sigma(x \in S^{n-1} : \|x\| < \alpha M) < (C'\alpha)^{\kappa(\frac{M}{b}\sqrt{n}-c_0)^2_+}. \tag{24}$$

Notice that this estimates precisely the same quantity as in (22). Here we see that as $\alpha \to 0$, the estimates improve significantly.

Proof of the implication Conjecture 15 → Conjecture 14 We will first prove the implication in the non-degenerate case. Assume that Conjecture 15 is true. We start with a given $\delta > 0$, and first prove that the statement of Conjecture 14 must hold for bodies K with $b(K)/M(K) \leq \sqrt{n/w_0}$, where $w_0 = \min(w_0(1 - \delta/10), c_0/(10\delta))$ comes from the constants in Conjecture 15. We then show why knowing the conjecture in these cases implies all other cases.

So, let $\delta > 0$ and define $\kappa = (1 - \delta/10)$. By Conjecture 15, for every body satisfying $b(K)/M(K) \leq \sqrt{n}/w_0$ we have that

$$\sigma(x \in S^{n-1} : \|x\| < \alpha M) < (C\alpha)^{(1-\delta/10)(\frac{M}{b}\sqrt{n}-c_0)^2}$$

(where C, c_0 and κ, now depend δ). We choose β small enough so that $(1-\beta)N(\frac{M}{b})^2(1-\delta/10)^2 \geq (1+\delta/2)$, so for example $\beta = \delta/10$ is small enough.

We now repeat the proof from Section 4.1, using the new estimate on the probability. Since β is very small (having assumed $\delta < \delta_0$), the term $u(\beta)$ from (2) has hardly any effect. We thus have an estimate $2^{u(\beta)N}(C'\alpha)^{(1+\delta/2)n}$ for the probability that for a single x a lower estimate in (23) of order $\beta\alpha M$ does not hold. We work the same way as in Section 3.1, having some choice of $\alpha = \alpha(\delta)$ for which this probability is small enough to take care of a whole $\alpha\beta/2C$-net of the sphere, where C is from the upper bound CM which we have already shown. This α will, as usual, be exponentially bad in δ, even before taking into account that $C' = C'(1-\delta/10)$ can itself have a bad dependency on δ. Notice that in this case we get not only the existence of operators U_i satisfying the inequality (23), but that (23) is true with high probability on the choice of operators (the probability coming from Chernoff, so, at least $1 - e^{-cn}$).

We turn to the case where, after specifying δ, we have a body for which $b(K)/M(K) > \sqrt{n}/w_0$ where w_0 was indicated above and depends only on δ. This means in effect that K is very degenerate. We then do a preliminary "regulating" procedure. We pick randomly k operators U_i, $i = 1, \dots, k$, for $k = 2(b/M)^2 w_0/n$. This is a small number depending on δ, since $k < 2w_0^2$. We now define K' to be the unit ball of the norm

$$\|x\|_{K'} = \frac{1}{k}\sum_{i=1}^{k} \|U_i x\|_K.$$

Of course, since M is simply the average of the norm on the sphere, we have that $M(K) = M(K')$. However, it is well known that the diameter of the average of k random rotates of a body is smaller by a factor about $\frac{1}{\sqrt{k}}$ than the diameter of the body. Since we are speaking about norms, this means that $b(K') \simeq b(K)/\sqrt{k}$. We will need the more precise result, namely that the diameter decreases almost isometrically by $1/\sqrt{k}$, provided k is not large compared to b/M, which is our case since K is degenerate. We formulate the lemma we need in its more familiar, dual form:

Lemma 16. *For any $0 < \varepsilon < 1$ there exist constants c_ε and $c(\varepsilon)$ such that for a symmetric body T, if $k < (c(\varepsilon)(diam(T)/M^*(T))^{1/2}$ and $k \ln k \leq n\varepsilon^2/8$ then for random $U_1, \dots, U_k \in O(n)$ we have with probability greater than $1 - e^{-c_\varepsilon n/k^2}$ that*

$$diam\left(\frac{1}{k}\sum_{i=1}^{k} U_i T\right) \leq \frac{(1+\varepsilon)}{\sqrt{k}} diam(T).$$

The proof of this fact follows from standard considerations, see [AM] for the case $k = 2$ which generalizes directly. In fact $c(\varepsilon)$ can be taken linear in ε and c_ε to be linear in ε^2.

Applying Lemma 16 to K° we get that for $k < \min(C(\varepsilon)(b/M)^2, n\varepsilon^2/8 \ln n)$ we have $b(K') \leq (1+\varepsilon)b(K)/\sqrt{k}$. For our choice of k (and for a fixed $\delta > 0$) clearly the condition holds for n large enough, since w_0 doesn't depend on n.

This means that $(b(K')/M(K')) \leq (1+\varepsilon)\sqrt{n/2w_0}$, so as long as $\varepsilon \leq \sqrt{2}-1$ we may apply the proof of the first part to get that there exist rotations $V_1, \dots, V_{N'}$, for $N' = (b(K')/M(K'))^2(1+\delta)$ so that

$$\frac{1}{N'}\sum_{j=1}^{N'} \|V_j x\|_{K'} \simeq |x|,$$

with constants of isomorphism depending only on δ. Taking the $N = N'k$ rotations U_iV_j we have that

$$\frac{1}{N'k}\sum_{j=1}^{N'}\sum_{i=1}^{k} \|V_jU_i x\|_K \simeq |x|.$$

From our choice of k and the estimate on N' we see that

$$N = N'k = k(b(K')/M(K'))^2(1+\delta) \leq (1+\varepsilon)(1+\delta)(b(K)/M(K))^2.$$

For ε of the same order as δ, we have the desired result.

We remark that although we have proved the existence of a set of rotations, we provided rotations with a certain structure and did not show that for random N rotations (23) is satisfied.

Remark A weakening of Conjecture 15 was proved by Latała and Oleszkiewicz, see Theorem 3 in [LO]. It states that for every symmetric K one has

$$\sigma(x \in S^{n-1} : \|x\| < \alpha M') < (12\alpha)^{(\frac{M'}{b}\sqrt{n}-6)_+^2/4}, \tag{25}$$

where M' is the median of the norm (which, in the non degenerate case, is known to be close to the mean of the norm, M). This estimate can be used instead of (22), in the same way that (24) was used in the proof of the implication above, to prove Conjecture 14 with instead of constant $(1+\delta)$, constant $(4+\delta)$ for any $\delta > 0$. We omit the details.

4.3 Improvements in some special cases

From our method of proof it is obvious that the parameter which plays the leading role in the lower bound is not $(M/b)^2$ but rather

$$\frac{1}{n}\log\sigma(x \in S^{n-1} : \|x\| > \alpha M).$$

(This parameter, for $\alpha = 1/2$, is very similar to the one introduced in [KV] to study local Dvoretzky type theorems.) To be precise, let us denote

$$f(\alpha) = \frac{1}{n}\log(1/\sigma(x \in S^{n-1} : \|x\| > \alpha M)).$$

Then for any proportion $\beta < \sigma(x \in S^{n-1} : \|x\| > \alpha M)$, that is, $\beta < 1 - e^{-nf(\alpha)})$ we have by Chernoff (2) that for a single x

$$\mathbb{P}[U_1 \ldots, U_N \in O(n) : \frac{1}{N}\sum_{i=1}^{N} \|U_i x\| \geq \beta\alpha M] \geq 1 - e^{-nf(\alpha)N(1-\beta)} e^{-Nu(\beta)}.$$

If we want this probability to suffice for a $\beta\alpha/2C$-net of the sphere (where C is from the upper bound), we need to have

$$e^{-Nu(\beta)} e^{-nf(\alpha)N(1-\beta)} e^{n \log(1+4C/(\alpha\beta))} < 1.$$

This gives us the bound on N, namely that for every $\alpha < 1$ and $0 < \beta < 1 - e^{-nf(\alpha)}$, we may choose

$$N = 2\frac{\log(1 + 4C/(\alpha\beta))}{f(\alpha)(1-\beta) + u(\beta)/n}$$

and have a lower bound $(\alpha\beta/2)M$ on the norm defined in (23), where CM is the *upper* bound we have on this norm defined in (23). In other words, we can take N as close as we want to

$$\inf_{\alpha<1, 0<\beta<1-e^{-nf(\alpha)}} \frac{\log(1 + 4C/(\alpha\beta))}{f(\alpha)(1-\beta) + u(\beta)/n},$$

getting that for the average of this number of rotations, assuming an upper bound CM, is isomorphic to euclidean, paying only with the isomorphism constants.

In many special cases the estimates for $f(\alpha)$ are better than what is given above, see examples in [KV]. The question remains whether one can give a general condition under which there are estimates for $f(\alpha)$ significantly better than (22) and (25).

Notice, however, that this is just the N for the lower bound, *assuming* an upper bound. It is well known that one always need to take at least $N = \lambda(b/M)^2$ for some constant $\lambda > 0$, to get the right order upper bound in (23). In particular, we need the upper bound so that we can transform the bounds on the net to bounds on the whole sphere. Thus, the improvement in the special cases where one computes $f(\alpha)$ and sees that it is larger than expected, i.e., that the infimum above is $o((b/M)^2)$, will be that averaging over $N = \lambda(b/M)^2$ rotations for a proportion $0 < \lambda$ is enough to get a norm isomorphic to euclidean.

References

[Ar] S. Artstein-Avidan, *A Bernstein-Chernoff deviation inequality, and geometric properties of random families of operators*, to appear in Israel J. of Math.

[AFM] S. Artstein-Avidan, O. Friedland and V. Milman, *Geometric applications of Chernoff-type estimates and a ZigZag approximation for balls*, Proc. Amer. Math. Soc., in press.

[AFMS] S. Artstein-Avidan, O. Friedland, V. Milman and S. Sodin, *Better bounds for large random Bernoulli sections of ℓ_1^N*, submitted.

[AM] S. Artstein-Avidan and V. Milman, *Logarithmic reduction of the level of randomness in some probabilistic geometric constructions*, to appear in J. Funct. Anal.

[B] A. Barron, *Universal approximation bounds for superpositions of a sigmoidal function*, IEEE Trans. Inform. Theory 39 (1993), 930-945

[BLM] J. Bourgain, J. Lindenstrauss and V. Milman, *Minkowski sums and symmetrizations*, Geometric aspects of functional analysis (1986/87), Lecture Notes in Math., 1317, Springer, Berlin, 1988, 44-66.

[C] G. Cybenko, *Approximation by superpositions of sigmoidal functions*, Proc. of the 1994 IEEE-IMS Workshop on Info. Theory and Stat., (1994).

[CB] G. Cheang and A. Barron *A better approximation for balls*, J. Approx. Theory 104 (2000) no. 2, 183-302.

[GiM1] A. Giannopoulos and V. Milman, *Euclidean structure in finite dimensional normed spaces*, Handbook of the geometry of Banach spaces, Vol. I, 707-779.

[GiM2] A. Giannopoulos and V. Milman, *Concentration property on probability spaces.* Adv. Math. 156 (2000), no. 1, 77-106.

[GMT] A. Giannopoulos, V. Milman and A. Tsolomitis, *Asymptotic formulas for the diameter of sections of symmetric convex bodies*, J. Funct. Anal. 223 (2005), no. 1, 86–108.

[Go] Y. Gordon, *On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$*, Geometric aspects of functional analysis (1986/87), Lecture Notes in Math. 1317, Springer, Berlin-New York, 1988, 84-106.

[GrM] M. Gromov and V. Milman, *A topological application of the isoperimetric inequality*, Amer. J. Math. 105 (1983), no. 4., 843-854.

[H] M. Hall, *Combinatorial Theory*, Second edition, Wiley, NY. 0-273-08565-4.

[Ha] P. Hall, *Rates of convergence in the central limit theorem*, Research Notes in Mathematics 62, Pitman (Advanced Publishing Program), Boston, Mass.-London, 1982.

[HR] T. Hagerup and C. Rüb, *A guided tour of Chernoff bounds*, Info Proc Lett 33 (1990) no. 6, 305-308.

[HSW] K. Hornik, M.B. Stinchcombe and H. White, *Multi-layer feedforward networks are universal approximators*, Neural Networks 2 (1988), 359-336.

[JS] W. Johnson and G. Schechtman, *Very tight embeddings of subspaces of L_p $1 \leq p < 2$ into ℓ_p^n*, Geom. Funct. Anal. 13 (2003), no. 4, 845–851.

[K] B. Kashin, *Section of some finite-dimensional sets and classes of smooth functions (in Russian)*, Izv. Acad. Nauk. SSSR 41 (1977) 334-351.

[KKS] J. Kahn, J. Komlós and E.Szemerédy, *On the Probability that a random* ±1-*Matrix is Singular*, Journal of the AMS, Vol 8 Number 1, (1995) 223-240.

[KV] B. Klartag and R. Vershynin, *Small ball probability and Dvoretzky Theorem*, Israel Journal of Mathematics, to appear.

[LO] R. Latała and K. Oleszkiewicz, *Small ball probability estimates in terms of width*, Studia Math. 169 (2005) no. 3, 305-314.

[LPRT] A. Litvak, A. Pajor, M. Rudelson and N. Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes*, Adv. Math. 195 (2005), no. 2, 491–523.

[LPRTV] A.E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann and R. Vershynin, *Random Euclidean embeddings in spaces of bounded volume ratio*, C. R. Math. Acad. Sci. Paris 339 (2004), no. 1, 33–38.

[MP] V. Milman and A. Pajor, *Regularization of start bodies by random hyperplane cut off*, Studia Mathematica 159 (2), 2003, 247-261.

[MS] V. Milman and G. Schechtman, *Asymptotic theory of finite-dimensional normed spaces. With an appendix by M. Gromov*, Lecture Notes in Mathematics, 1200. Springer-Verlag, Berlin, 1986.

[MS2] V. Milman and G. Schechtman, *Global versus local asymptotic theories of finite-dimensional normed spaces*, Duke Math. J. 90 (1997), no. 1, 73–93.

[R] M. Rudelson, *Invertibility of random matrices: norm of the inverse*, submitted.

[S1] G. Schechtman, *Special orthogonal splittings of L_1^{2k}* (English. English summary), Israel J. Math. 139 (2004), 337-347.

[S2] G. Schechtman, *Random embeddings of euclidean spaces in sequence spaces*, Israel Journal of Mathematics, Vol. 40, No. 2, 1981, pp. 187-192.

[TV] T. Tao and V. Vu, *On the singularity probability of random Bernoulli matrices*, preprint.