A Bernstein-Chernoff deviation inequality, and geometric properties of random families of operators

Shiri Artstein-Avidan [*]

Mathematics Department, Princeton University

Abstract: In this paper we first describe a new deviation inequality for sums of independent random variables which uses the precise constants appearing in the tails of their distributions, and can reflect in full their concentration properties. In the proof we make use of Chernoff's bounds. We then apply this inequality to prove a global diameter reduction theorem for abstract families of linear operators endowed with a probability measure satisfying some condition. Next we give a local diameter reduction theorem for abstract families of linear operators. We discuss some examples and give one more global result in the reverse direction, and exensions.

Acknowledgement: I would like to thank Prof. Vitali Milman for his support and encouragement, and mainly for his mathematical help and advice.

[*]This research was partially supported by BSF grant 2002-006.

The first theorem in this note is a new Bernstein-type deviation inequality which we prove using Chernoff's bounds. This theorem is different from the classical Bernstein inequality in the following way: whereas the condition in the standard Bernstein inequality is on the global behavior of the random variables in question, for example a condition on the expectation of e^{cX^2}, in Theorem 1 below the condition uses only the constants appearing in the tail of the distribution, and so can reflect concentration. Sometimes one can prove very strong estimates on the tails. In the theorem below these estimates can be then used and are amplified when one averages many i.i.d. copies of the variable. This theorem in a special case was brought forward and used in the paper [AFM] for a specific example. Its proof is straightforward using only Chernoff's bounds, and we find this approach insightful and new.

We first apply the deviation inequality for some geometric question. We present several results regarding the behavior of the diameter of a convex body under some random operations. The first is a global result, namely regarding the Minkowski sums of copies of a convex body acted upon by abstract families of linear operators endowed with a probability measure. The classical global diameter reduction is the well known special case where the family of operators is $O(n)$, the family of orthogonal rotations. This was first observed in [BLM], see also [MiS] for more details. In Section 5 we revisit this case as an example.

The second result we discuss is of a local nature, and is an extension of the now well known diameter reduction phenomenon for random orthogonal projections. This phenomenon was first observed by Milman in his proof for the quotient of a subspace theorem, [Mi2] (and analyzed as a separate proposition in [Mi3], where more references can be found). It can be considered today as a consequence of the classical Dvoretzky-type theorem as proved in [Mi1]. The classical theorem concerns the case where the random operation is intersection with a random subspace or projection onto a random subspace. However, in this paper we consider a more general setting. Instead of working with projections, we deal with an abstract family of linear operators endowed with a probability measure and find a condition on this measure (which is in fact a condition on the probabilistic behavior of the operators on individual elements $x \in \mathbb{R}^n$) which promises that a diameter reduction theorem holds. The proof of the theorem uses Talagrand's Majorizing Measures Theorem, see [Tal].

In Section 4 we describe a global result in the reverse direction, describing in a particular case when does the resulting body contain a euclidean ball. In the classical setting this kind of containment is the only known reason for stabilization of the diameter.

We then discuss some examples. We show how the abstract propositions indeed imply Milman's diameter reduction theorem for usual orthogonal projections and global Dvoretzky's Theorem for unitary transformations (and the diameter reduction which occurs until stabilization). We describe other families of operators for which there is a similar diameter reduction. One of our main goals is to crystalize which properties of the operators are important for diameter reduction results to hold. Finally we give two more variants of the local result.

We remark that the results described in this paper have many similar variants that can be proven in exactly the same way. The choice of conditions in each one depend very much on the applications in mind. Thus as much as we tried to give general and abstract constructions, stating each proposition in full generality would be notationally very inconvenient. We tried to indicate in remarks which main variants are possible for each statement.

Recently I learned that results in the spirit of Proposition 3 below are being studied by the team of A. Litvak, A. Pajor and N. Tomczak-Jaegermann, see [LPT].

Notation: We use $|\cdot|$ to denote the euclidan norm in $\mathbb{R}^n$, and denote by D_n the euclidean unit ball, $D_n = \{x : |x| \le 1\}$. For a centrally symmetric convex body $K \subset \mathbb{R}^n$ we denote by $d = d(K)$ its diameter, so $K \subset d(K)D_n$. We let $M^* = M^*(K)$ denote half its mean width, that is $M^*(K) = \int_{S^{n-1}} \sup_{y\in K}\langle x, y\rangle d\sigma(x)$ where S^{n-1} is the euclidean unit sphere and σ denotes the normalized Lebesgue measure on this sphere. Thus M^* is the average of the dual norm of K, which we denote by $\|x\|^* = \sup_{y\in K}\langle x, y\rangle$.

1 A Deviation Inequality

We first describe our main tool, which is a Bernstein-type deviation Theorem. Its proof follows from Chernoff's bounds, and we provide it below. We wish to point out the main difference between this theorem and the classical Bernstein deviation inequality for, say ψ_2, random variables. The classical theorem, for which we refer the reader to, say, [BLM], gives an upper bound for the probability in (1) below, in the following form: If A is the ψ_2-norm of the random variable X, and X_i are i.i.d. copies of X, then

$$\mathbb{P}[|\frac{1}{N}\sum_{i=1}^{N} X_i - \mathbb{E}X| > t] \le 2e^{-Nt^2/(8A^2)}.$$

The ψ_2-norm of the variable is affected by the constant in the tail estimate, but not only, and for example the expectation or variance may take a part and influence this constant A. The purpose of the deviation inequality in our Theorem 1 is to use the tail estimate itself (and not just the good ψ_p behavior following from it). This type of Proposition was first used, for a special example, in [AFM].

Theorem 1 *Assume X is a random variable satisfying*

$$\mathbb{P}[X > t] \leq e^{-Kt^p}$$

for some constant $K > 0$, some $p > 1$, and any $t > K_0$. Let $X_1, \ldots, X_N$ be i.i.d. copies of X. Then for any $s > \max\{C(K,p), K_0\}$,

$$\mathbb{P}[\frac{1}{N}\sum_{i=1}^{N} X_i > 3s] \leq C_0 e^{-N(Ks^p - \ln 2)}, \tag{1}$$

where C_0 is a universal constant for p bounded away from 1, and where $C(K,p) = \frac{(1+\ln 2)}{K^{1/p}}$.

Remark 1. As will be evident from the proof, it is not necessary that the variables be identically distributed, and it is sufficient that they are independent and that each satisfies the tail estimate.

Remark 2. The term $\ln 2$ appearing in the estimate is avoidable, by using the exact form of Chernoff's inequality in the proof, namely using that for i.i.d. p-Bernoulli variables Z_i, and for $\beta < p$,

$$\mathbb{P}[\sum_{i=1}^{N} Z_i \leq \beta N] \leq e^{-N[\beta \ln(\beta/p) + (1-\beta)\ln((1-\beta)/(1-p))]}.$$

For reference on this estimate and on the Chernoff bound used in the proof see for example the survey on geometric applications of Chernoff type estimates

[AFM]. More precisely, if one substitutes the constant 3 by C_1 then instead of $\ln 2$ one can put a constant c_2 such that $c_2 \to 0$ when $C_1 \to \infty$.

Remark 3. In the case $p = 1$ one encounters a problem with the convergence of the probability. However, if one assumes an upper bound d on the random variable X, then the same proof as below will give an upper estimate on the probability in (1) of the form $\approx C_0 \log(d/s) e^{-NKs/\log(d/s)}$, which is sufficient in some cases.

Proof of Theorem 1. We will use the standard Chernoff bound. For $j = \log s + 1, \log s + 2, \dots$ we define

$$A_j = \{2^{j-1} < X \le 2^j\},$$

so that $P[X_i \in A_j] \le e^{-K2^{(j-1)p}}$ (where we have used the assumption $s > K_0$). We set $m_j = Ns2^{-j}/(j - \log s)^2$. We measure the probability of the following event: out of the N variables X_i, for every j, no more than m_j of them are in A_j. This event is included in the event that

$$\frac{1}{N}\sum_{i=1}^{N} X_i \le s(1 + \sum_{j=1}^{\infty} \frac{1}{j^2}) \le 3s$$

We will estimate the probability of the complementary event. It is less than the sum over j over the individual probabilities

$$P_j = \mathbb{P}[\text{ more than } m_j \text{ of the } X_i\text{s are in } A_j].$$

As long as

$$s2^{-j}/(j - \log s)^2 > e^{-K2^{p(j-1)}} \tag{2}$$

(which will give us a condition on s, namely a lower bound on s in terms of K and p), this probability is small, and by Chernoff it is smaller than

$$e^{-N[K2^{p(j-1)}s2^{-j}/(j-\log s)^2-\ln 2]} = e^{-N[Ks2^{(p-1)j-1}/(j-\log s)^2-\ln 2]}.$$

(Here by using the exact form of Chernoff's estimate we may substitute $-\ln 2$ by the term $\beta\ln\beta+(1-\beta)\ln(1-\beta)$ for, say, $\beta = s2^{-j}$, and this will improve the estimate. More precisely, if we sum to begin with for $j = \log(C_1 s)+1, \log(C_1 s)+2, \ldots$ we will have above that $\beta \le 1/(2C_1)$ and so the additional term in the exponent is also small, and tends to 0 when C_1 increases.)

The sum of these probabilities converges (we are using the fact $p > 1$), and is comparable to the first element in the series, which is $e^{-N[s^pK/2-\ln 2]}$ (so, in fact, C_0 in the Theorem depends on p but can be taken universal, and even not large at all, when p is bounded away from 1).

We now have to check condition (2). The left hand side is $2^{-i}/i^2$ and the right hand side is $e^{-K2^{p(i-1)}s^p}$. Taking the natural logarithm of both sides we see that the condition is $(i+2\ln i)/2^{p(i-1)} < Ks^p$. Clearly the left hand side is largest for $i = 1, 2$, so we need only ensure that $s > \frac{(2+2\ln 2)}{2K^{1/p}} = \frac{(1+\ln 2)}{K^{1/p}}$. Thus we have shown that the condition in Chernoff's bound is satisfied and the proof of Theorem 1 is complete. $\square$

2 A Global Proposition

We now state the application of this theorem, which is a global proposition regarding the decrease of diameter of a convex body, which generalizes the well known diameter reduction for averages of random orthogonal rotations of a convex body.

Proposition 2 *There exist universal constants c, C such that: Let $\{A\}$ be some family of operators $A : \mathbb{R}^n \to \mathbb{R}^n$ with some probability measure $\mathbb{P}$. Assume that for some body $K \subset R^n$ and some $\alpha > 0$ the following holds: for every $x \in \mathbb{R}^n$ and for any $s > s_0$*

$$\mathbb{P}[A : \|Ax\|_K > s|x|] \leq e^{-\alpha^2 s^2 n}. \tag{3}$$

Then, if $T < 2(\frac{1}{\alpha s_0})^2$, we have with probability greater than $1 - e^{-cn}$ that for any $x \in \mathbb{R}^n$

$$\frac{1}{T}\sum_{i=1}^{T} \|A_i x\|_K \leq C\frac{1}{\alpha\sqrt{T}}|x|, \tag{4}$$

and if $T \geq 2(\frac{1}{\alpha s_0})^2$ we have with probability greater than $1 - e^{-cn}$ that for any $x \in \mathbb{R}^n$

$$\frac{1}{T}\sum_{i=1}^{T} \|A_i x\|_K \leq C s_0 |x|, \tag{5}$$

where A_i are chosen independently according to the distribution $\mathbb{P}$.

Restating the proposition in geometric form, using duality, gives

Corollary 1 *There exist universal constants c, C such that: Let $\{A\}$ be some family of operators $A : \mathbb{R}^n \to \mathbb{R}^n$ with some probability measure $\mathbb{P}$. Assume that for some body $K \subset R^n$ and some $\alpha > 0$ the following holds: for every $x \in \mathbb{R}^n$ and for any $s > s_0$*

$$\mathbb{P}[A : \|A^* x\|_K^* > s|x|] \leq e^{-\alpha^2 s^2 n}. \tag{6}$$

Then, if $T < 2(\frac{1}{\alpha s_0})^2$, we have with probability greater than $1 - e^{-cn}$ that

$$\frac{A_1K + A_2K + \cdots + A_TK}{T} \subset C\frac{1}{\alpha\sqrt{T}}D_n, \tag{7}$$

and if $T \geq 2(\frac{1}{\alpha s_0})^2$ we have with probability greater than $1 - e^{-cn}$ that

$$\frac{A_1K + A_2K + \cdots + A_TK}{T} \subset Cs_0D_n, \tag{8}$$

where A_i are chosen independently according to the distribution $\mathbb{P}$.

Remark. We are describing the case $p = 2$ because it is the most useful. However, for any $p > 1$ we have a similar result, namely if instead of (6) we have an estimate of the form

$$\mathbb{P}[A : \|A^*x\|_K^* > s|x|] \leq e^{-\alpha^p s^p n},$$

then we get a separation into two cases $T < 2(\frac{1}{\alpha s_0})^p$ and $T \geq 2(\frac{1}{\alpha s_0})^p$. In the first case instead of (7) we get that the average is included in $C\frac{1}{\alpha T^{1/p}}D_n$, and in the second case we get exactly (8). The proof is identical. As for $p = 1$, this is different, since Theorem 1 is different, and one gets an upper estimate on the diameter of the form $\approx \frac{1}{\alpha T}\log d$ if this quantity is greater than s_0.

The meaning of the separation of the two cases inside the proposition seems to be that there is a diameter reduction of order $\sqrt{T}$ when taking an average of T copies of a convex body K operated upon by random operators A_i, until the diameter reaches some critical value on which it stabilizes. In the case of orthogonal rotations we know the reason for stabilization, namely the body becomes a euclidean ball. Of course, the proposition above gives only an upper bound, and by no means implies stabilization. To get any result in the reverse direction (namely, the inclusion of a ball of some radius after a given number of steps, and stabilization) we would need a reverse condition as well, promising that points do not *shrink* very much under the random operation. This is addressed in Proposition 5.

Proof of Proposition 2 We begin with the case of $T < 2(\frac{1}{\alpha s_0})^2$, and we should show that for any $x \in S^{n-1}$

$$\sum_{i=1}^{T} \|A_i x\| \leq C\frac{\sqrt{T}}{\alpha}, \tag{9}$$

under the assumption that for some constant α and for any $s > s_0$

$$\mathbb{P}[A : \|Ax\| > s|x|] \leq e^{-\alpha^2 s^2 n}.$$

Theorem 1 with $p = 2$ tells us that then for $s > \max\{s_0, (1 + \ln 2)/(\sqrt{n}\alpha)\}$ we have

$$\mathbb{P}[\frac{1}{T}\sum_{i=1}^{T} \|A_i x\| > 3s|x|] \leq C_0 e^{-T(\alpha^2 n s^2 - \ln 2)}.$$

We want this true for $3s = \frac{C}{2}\frac{1}{\sqrt{T}\alpha}$, for every x in a 1/2-net on the sphere. Such a net has cardinality less than 5^n. Then successive approximation will guarantee that inequality (9) be true for all $x \in S^{n-1}$. The probability that we get for this is greater than $1 - 5^n e^{-T(nC^2/(36T) - \ln 2)}$, which for large C is high, at least in the case where $T \leq C'n$. If T is much larger than n the term $\ln 2$ interferes, and so we have to use the stronger form of Theorem 1 avoiding this term, which we indicated in the remark following Theorem 1 and also in the proof of the theorem.

In the second case where $T \geq 2(\frac{1}{\alpha s_0})^2$ we can no longer take $s = C\frac{1}{\sqrt{T}\alpha}$ but only $s = s_0$. The probability is greater than $1 - 5^n e^{-T(\alpha^2 s_0 n - \ln 2)}$, and from the assumption on T this probability is exponentially close to 1. So, we get inequality (5). We remark that although we wrote a universal constant C, the proof shows that this constant is not large at all and can be chosen to be, say, 5 (and in some cases close to 1). $\square$

3 A Local Proposition

In this section we describe an analogue of Milman's *local* diameter reduction theorem, namely the theorem for orthogonal projections. The proof of Dvoretzky's Theorem in [Mi1] implies that a random projection of a convex body K in $\mathbb{R}^n$ of diameter $d = d(K)$ into a subspace of dimension $k^* = c(\frac{M^*}{d})^2$, is an approximate euclidean ball of radius $M^* = M^*(K)$. The fact that this k^* is indeed the correct formula for the dimension in which a projection is an approximate euclidean ball, and not just a lower estimate, was pointed out in [MiS].

It was then observed by Milman that for any dimension $k > k^*$, when one projects the body K into a k-dimensional subspace, its diameter decreases by a factor around $\sqrt{k/n}$. For a detailed explanation of this fact and more of the history see [Mi4], Section 2.3.1. Thus, there is only one type of behavior of the diameter of a convex body under projections, it decreases like $\sqrt{k/n}$ as long as k is larger than the critical value k^*, and then it stabilizes on the value M^*. (Note that here we know *exact* behavior, not only upper bounds.)

In Proposition 3 below we deal with an abstract family of operators satisfying a condition which has nothing to do with the body K but only describes the way the operators act on individual points. Under this condition, a general reduction of diameter holds for all convex bodies. We discuss some examples for such families of operators in Section 5, including the classical case of orthogonal projections.

Proposition 3 *Let $\{A\}$ be some family of operators $A : \mathbb{R}^n \to \mathbb{R}^n$ with some probability measure $\mathbb{P}$. Assume that for every $x \in \mathbb{R}^n$ and for any $s > s_0$*

$$\mathbb{P}[A : |Ax| > s|x|] \leq e^{-cns^2}.$$

Then there exist universal constants c', C such that given a convex body $K \subset \mathbb{R}^n$ the following holds: If $s_0 > (\frac{M^*(K)}{d(K)})$ then for every $j \geq s_0^2 n$, with probability greater than $1 - e^{-c'j}$ on the choice of A we have

$$AK \subset C\sqrt{\frac{j}{n}} d(K) D_n,$$

and if $s_0 \leq (\frac{M^*(K)}{d(K)})$ then with probability greater than $1 - e^{-c'ns_0^2}$ on the choice of A we have

$$AK \subset CM^*(K)D_n.$$

Moreover, the constant c' appearing in the probability is a function of the constant C, and by increasing C we can have c' as big as desired.

To prove Proposition 3 we will use Talagrand's Majorizing Measures Theorem in the form of Theorem 4 below (see [Tal]); this type of application of the Majorizing Measure Theorem was suggested to me by Prof. Keith Ball and was used in a special case in [Art1].

Theorem 4 (Majorizing Measures) *There exists a universal constant C_0 such that for any dimension n, for every convex body $K \subset R^n$, there exist* *families of points* $B_0 \subset \ldots \subset B_{m-1} \subset B_m \subset \ldots \subset K$ *with cardinality* $|B_m| \leq 2^{2^m}$ *such that for every* $x \in K$

$$\sum_{m=0}^{\infty} d(x, B_m)\sqrt{2^m} \leq C_0 M^*(K)\sqrt{n},$$

where $d(x, B_m)$ *denotes the distance of the point* x *to the* m^{th} *family, i.e.,* $d(x, B_m) = \inf\{d(x, y) : y \in B_m\}$. *(Moreover, these points can be constructed so that if* v_m *is the closest point to* x *in* B_m*, and* v_{m+1} *is the closest point to* x *in* B_{m+1}*, then* v_m *is the closest point to* v_{m+1} *in* B_m*.)*

Proof of Proposition 3 We fix $m_0 = \log c_1 j$, with c_1 universal to be chosen later. Each vector $x \in K$ we write as

$$x = v_{m_0}(x) + \sum_{m=m_0+1}^{\infty} (v_m(x) - v_{m-1}(x))$$

where $v_m(x) \in B_m$ is the member of B_m closest to x. We denote $u_m(x) = v_m(x) - v_{m-1}(x)$, so that $|u_m(x)| \leq 2d(x, B_{m-1})$, and denoting $C_m = \{u_m(x) : x \in K\}$, we have $|C_m| \leq 2^{2^{m+1}}$. We rewrite

$$x = v_{m_0} + \sum_{m=m_0+1}^{\infty} \left(\frac{d(x, B_{m-1})\sqrt{2^{m-1}}}{C_0 M^*(K)\sqrt{n}} \frac{|u_m(x)|}{2d(x, B_{m-1}))} \right) \left(\frac{u_m(x)}{|u_m(x)|} \frac{2C_0 M^*(K)\sqrt{n}}{\sqrt{2^{m-1}}} \right).$$

The Majorizing Measures Theorem tells us that the sum of the coefficients in the left brackets of the infinite sum is less than 1. Therefore, for any linear operator A

$$|Ax| \leq |A(v_{m_0}(x))| + \sup_{m_0 < m < \infty} \frac{|A(u_m(x))|}{|u_m(x)|} \frac{2C_0 M^*(K)\sqrt{n}}{\sqrt{2^{m-1}}}.$$

In other words, AK is contained in a ball of radius

$$\max_{v \in B_{m_0}} |Av| + \sup_{m_0 < m < \infty, u_m \in C_m} \frac{|Au_m|}{|u_m|} \frac{2C_0 M^*(K)\sqrt{n}}{\sqrt{2^{m-1}}}. \tag{10}$$

We first discuss the case $s_0 > (\frac{M^*(K)}{d(K)})$. We use the assumption in the statement of the Proposition,

$$\mathbb{P}[A : |Ax| > s|x|] \leq e^{-cns^2},$$

to show that with proability $1 - e^{-c'j}$ we have that

(a) For every $v \in B_{m_0}$,

$$|Av| \leq C_1 \sqrt{\frac{j}{n}} |v|,$$

and (b) For every $m > m_0$, for every $u_m \in C_m$,

$$|Au_m| \leq C_1 \frac{d\sqrt{2^{m-1}}}{M^*(K)\sqrt{n}} \sqrt{\frac{j}{n}} |u_m|.$$

This will complete the proof of the case $s_0 > (\frac{M^*(K)}{d})$ with $C = C_1 + 2C_0C_1$.

For (a), since the probability of the event happening for a specific v is at least $1 - e^{-cC_1^2 j}$, we see that if we have in B_{m_0} no more than $e^{c_1 j}$ elements, for $c_1 = cC_1^2/2$, the probability that for all of them we have this property is greater than $1 - e^{-c_1 j}$, which is precisely why we chose $m_0 = \log c_1 j$ (and we have thus specified c_1). In estimating the probability, we have used the assumption that $C_1\sqrt{j/n} > s_0$, which is clearly satisfied if $C_1 > 1$.

For (b) to happen we take care of each m separately. For each m the probability is bounded by

$$1 - 2^{2^{m+1}} P[|Ax| > t\sqrt{\frac{j}{n}}|x|],$$

for $t = \frac{d}{M^*} C_1 \frac{\sqrt{2^{m-1}}}{\sqrt{n}}$. Since we consider $m > m_0$, we have that $2^{m-1} \geq c_1 j$, and hence we can apply the estimate for the probability, as long as, say, $C_1 > 1/c_1$, getting

$$1 - 2^{2^{m+1}} e^{-cC_1^2 2^{m-1} \frac{j}{n} (\frac{d}{M^*})^2}.$$

We are assuming that $s_0 > (\frac{M^*(K)}{d(K)})$, so $\frac{j}{n} > (\frac{M^*(K)}{d})^2$. Therefore (if, say, $C_1 \geq 4/\sqrt{c}$) we can bound this probability from below by $1 - 2^{-c_2 2^m}$, and c_2 can be large provided that C_1 is chosen large. Adding up for all m the probability of failure in (b), and adding also the probability of failure in (a) as the first summand, we see that

$$e^{-c_1 j} + \sum_{m=m_0+1}^{\infty} 2^{-c_2 2^m} \leq 2^{-c' j},$$

which completes the proof in the first case. Clearly by increasing C_1 we can increase c' as much as required.

In the second case, where $s_0 < (\frac{M^*}{d})$, we take $m_0 = \log(c_1 n s_0^2)$, again c_1 to be chosen later. We use the assumption

$$\mathbb{P}[A : |Ax| > s|x|] \leq e^{-cns^2}$$

for $s = C_1 s_0$ to show that with probability $1 - e^{-cC_1^2 n s_0^2}$

(a) For every $v \in B_{m_0}$,

$$|Av| \leq C_1 s_0 |v| \leq C_1 M^*(K),$$

and for $s = C_1 \frac{\sqrt{2^{m-1}}}{\sqrt{n}}$ to show

(b) For every $m > m_0$, for every $u_m \in C_m$,

$$|Au_m| \leq C_1 \frac{\sqrt{2^{m-1}}}{\sqrt{n}} |u_m|,$$

and following (10) this will complete the proof in the second case $s_0 \leq (\frac{M^*}{d})$.

Calculating the probabilities, using that now $\sqrt{2^m/n} > \sqrt{c_1} s_0$, we get that, when $C_1 > 1/\sqrt{c_1}$, the probability is greater than

$$1 - e^{-cC_1^2 s_0^2 n} - \sum_{m=m_0+1}^{\infty} 2^{2^{m+1}} e^{-cC_1^2 2^{m-1}}$$

and for C_1 sufficiently large the probability is greater than $1 - e^{-c' n s_0^2}$. $\square$

4 A Global Proposition in the reverse direction

To have a complete global Dvoretzky type statement, Proposition 2 provides just one direction, namely it shows that the average of a certain amount of copies of K is contained inside a euclidean ball of an appropriate size. Actually, under these conditions nothing stronger can be stated. Proposition 5 below gives the reverse side, namely the containment of a ball. Naturally, it involves a condition which promises that individual points are not shrunk "too much" by the operators. It also includes an a-priori assumption of diameter reduction, which can be obtained for example by using Proposition 2. We remark that the condition in the proposition is about a specific value ε_0 which is (four times) the radius of the ball we want inside our body. If one knows a more global condition satisfied for different values of ε, for example a small ball probability estimate such as in [LO], one can sometimes get an "inner diameter increase" lemma by applying the condition each time to a different pair (ε, T). In this sense the proof below is very simple, and so can be adapted to various initial conditions.

Proposition 5 *Let $\{A\}$ be some family of operators $A : \mathbb{R}^n \to \mathbb{R}^n$ with some probability measure $\mathbb{P}$. Assume that for some body K and some $\varepsilon_0 > 0$ the following holds: for every $x \in \mathbb{R}^n$*

$$\mathbb{P}[A : \|Ax\|_K < \varepsilon_0 |x|] \le (F(\varepsilon_0))^n,$$

where $(F(\varepsilon_0))^n < 1/4$. Assume, further, that for some R and some T, with *probability p we have*

$$\frac{1}{T}\sum_{i=1}^{T}\|A_i x\|_K \le R|x|.$$

Then, if T satisfies $T > \ln(2+16R/\varepsilon_0)/\ln(1/(2F(\varepsilon_0)))$ we have with probability greater than $1-2^{-n}-p$ that

$$\frac{\varepsilon_0}{4}|x| \le \frac{1}{T}\sum_{i=1}^{T}\|A_i x\|_K,$$

where A_i are chosen independently according to the distribution $\mathbb{P}$.

Again we may reformulate the above in geometric form, namely assume that the family of operators $A:\mathbb{R}^n \to \mathbb{R}^n$ satisfies for some body K and some $\varepsilon_0 > 0$ and $F(\varepsilon_0) < (1/4)^{1/n}$, that for every $x \in \mathbb{R}^n$ one has $\mathbb{P}[A : \|A^* x\|_K^* < \varepsilon_0|x|] \le (F(\varepsilon_0))^n$, and that for some R and some T, with probability p one has $\frac{1}{T}\sum_{i=1}^{T} A_i K \subset RD_n$. Then, if T satisfies $T > \ln(2+16R/\varepsilon_0)/\ln(1/(2F(\varepsilon_0)))$, one has with probability greater than $1-2^{-n}-p$ that $\frac{\varepsilon_0}{4}D_n \le \frac{1}{T}\sum_{i=1}^{T} A_i K$.

Proof of Proposition 5 We will show that for an $\frac{\varepsilon_0}{4R}$ net on the sphere, where R is the upper bound we are assuming, the inequality holds with $\varepsilon_0/2$. Then by the triangle inequality we will have for every $x \in S^{n-1}$

$$\frac{1}{T}\sum_{i=1}^{T}\|A_i x\|_K \ge \varepsilon_0/4.$$

The net has cardinality less than $(1+8R/\varepsilon_0)^n$. From Chernoff, the probability that more than half of the numbers $\|A_i x\|_K$ will be greater than ε_0 is greater than $1-(2(F(\varepsilon_0))^n)^T$. The probability that this is true for every point in the net is greater than

$$1-(1+8R/\varepsilon_0)^n(2F(\varepsilon_0)^n)^T.$$

We now write the condition on T which promises that this quantity is greater than $1 - 2^{-n}$ and the proof is complete. $\square$

5 Some Examples

1. Orthogonal Projections For an integer $1 \le k \le n$ let $\{P_k\}$ be the family of orthogonal projections onto k-dimensional subspaces of $\mathbb{R}^n$, endowed with the normalized Haar measure. It is well known, and was shown for example (with precise estimates on c_1, c_2 below) in [Art1], that for $s > c_1\sqrt{\frac{k}{n}}$

$$\mathbb{P}[P_k : |P_k x| > s|x|] \le e^{-c_2 n s^2}.$$

Proposition 3 then implies Milman's diameter reduction theorem for random projections. (And by duality, one sided estimates for Dvoretzky's Theorem for sections). Namely, for projection onto a random subspace of dimension k greater than $k^* = c(\frac{M^*}{d})^2 n$ the diameter decreases like $\sqrt{\frac{k}{n}}$, and for dimension lower than this the diameter is around M^*. The other side of Dvoretzky's Theorem, namely that a projection onto a subspace of dimension k^* is already with high probability isomorphic to a euclidean ball of radius M^* is not included in this statement.

2. Sign-Projections For an integer $1 \le k \le n$ let $\{S_k\}$ be the family of k-dimensional sign-projections, defined as follows: for k sign-vectors $\varepsilon_i \in \{-1, 1\}^n$, $i = 1, \ldots, k$ let $S_k(x) = \sum_{i=1}^k \langle x, \varepsilon_i/\sqrt{n}\rangle e_i \in \mathbb{R}^k$. We consider the uniform measure on this set, namely each ε_i is chosen with respect to the uniform measure on the n-dimensional discrete cube. It is not difficult to show that for $s > c_1\sqrt{\frac{k}{n}}$

$$\mathbb{P}[S_k : |S_k x| > s|x|] \leq e^{-c_2 n s^2}.$$

This is explained for example in [Art2], where in the same spirit as of the proof of Proposition 3 it was shown that also for this family of operators a diameter reduction statement holds. In this case, however, the statement is not sharp, namely for certain bodies, such as $B(\ell_1^n)$, the decrease of diameter continues beyond the value k^*. For a more detailed discussion see [Art2].

3. Orthogonal Rotations Consider the family $O(n)$ of orthogonal rotations in $\mathbb{R}^n$, endowed with the normalized Haar measure. Proposition 2 and Proposition 5 together give us the famous global version of Dvoretzky's Theorem, namely that the average of $(\frac{d}{M^*})^2$ random rotations of a body K is isomorphic to a euclidean ball of radius M^*. This theorem first appeared in [BLM]. Indeed, the estimates that we use are the famous concentration of measure estimate for $t > 0$

$$\mathbb{P}[x : |\|x\|_K^* - M^*| > tM^*|x|] \leq \sqrt{\pi/2}e^{-(M^*/d)^2 t^2 n/2}.$$

(Where for one side we use this with one specific $t = 1/2$ say, and apply Proposition 5, and for the other side we use the tail estimate for $t > 1$ and Proposition 2).

4. ψ_2-bodies

A convex body T of volume 1 is called a "ψ_2-body" if there exists a constant A such that for each $\theta \in S^{n-1}$ we have that the random variable $X_\theta = \langle y, \theta \rangle$ (where y is random in T with respect to the volume distribution) is ψ_2 with constant less than A, that is, $\mathbb{E}(e^{(X_\theta/A)^2}) \leq 2$. (When the expectation is with respect to the volume distribution in T)

We can define another projection-type operator as follows: Consider k random points y_i inside this convex body T (random with respect to the volume distribution). Define $G_k(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{k} \langle x, y_i \rangle e_i$. That is, a projection-type operator from $\mathbb{R}^n$ into $\mathbb{R}^k$.

To apply Proposition 3, we need to find s_0 such that for every $x \in \mathbb{R}^n$ and for any $s > s_0$

$$\mathbb{P}[G_k : |G_k x| > s|x|] \leq e^{-cns^2}.$$

To do this we use the ψ_2 behavior in the standard Markov inequality scheme

$$\begin{aligned} \mathbb{P}[|G_k x| > s|x|] = \mathbb{P}[\sum_{i=1}^{k} |\langle x, y_i \rangle|^2 > s^2 n|x|] &\leq \\ \mathbb{E}(e^{(X/A)^2})^k e^{-s^2 n/A^2} \leq e^{-n(s^2/A^2 - (k/n)\ln 2)} &\leq \quad e^{-k(s^2/2A^2)} \end{aligned}$$

for $s > A\sqrt{\ln 2/2}\sqrt{k/n}$.

Following Proposition 3 we get exactly the same behavior as in the case of orthogonal projections and of sign-projections. Notice that we have two *different* convex bodies involved, one which helps us define a random operator, and another whose diameter is reduced by applying this operator. The first is a ψ_2-body, the second is arbitrary.

The case of the same definition of an operator but when the body T with which we define the random operator is general and not necessarily ψ_2 is different, and is discussed in the next section.

6 Some further continuations

We want to describe two extensions of the above propositions, which, joint together, are relevant to the example indicated at the end of the above section.

The first extension of the local Proposition 3 is to the case of ψ_p behavior, or more precisely to the case where the tails are not as good, subgaussian, as in the statement of the proposition. To describe the result we need to introduce the parameter $\gamma_p(K)$, $1 \le p \le 2$, associated to a convex body K. This parameter was introduced by M. Talagrand in his Majorizing Measures theory, and is by now a widely used geometric parameter. Define

$$\gamma_p(K) = \inf \sup_{x \in K} \sum_{m \ge 0} d(x, B_m) 2^{m/p}$$

where the infimum is taken over all families $B_m \subset K$, $m = 0, 1, \ldots$ with $|B_m| \le 2^{2^m}$. For $p = 2$ we have in effect already used the parameter $\gamma_2(K)$ because $\gamma_2(K) \approx \sqrt{n} M^*(K)$ (this is exactly Talagrand's Majorizing Measures Theorem, one side of which is theorem 4). For motivations for this definition, computation of γ_p for certain bodies, and many applications see [Tal].

Using this definition and the method as in the proof of Proposition 3 one gets the following

Proposition 6 *Let $\{A\}$ be some family of operators $A : \mathbb{R}^n \to \mathbb{R}^n$ with some probability measure $\mathbb{P}$. Assume that for some $1 \le p \le 2$, for every $x \in \mathbb{R}^n$ and for any $s > s_0$*

$$\mathbb{P}[A : |Ax| > s|x|] \le e^{-cns^p}.$$

Then there exist universal constants c', C such that given a convex body $K \subset \mathbb{R}^n$ the following holds: If $s_0 > (\frac{\gamma_p(K)}{d(K)n^{1/p}})$ then for every $j \ge s_0^p n$, with probability greater than $1 - e^{-c'j}$ on the choice of A we have

$$AK \subset C(\frac{j}{n})^{1/p} d(K) D_n,$$

and if $s_0 \leq (\frac{\gamma_p(K)}{d(K)n^{1/p}})$ *then with probability greater than* $1 - e^{-c'ns_0^p}$ *on the choice of* A *we have*

$$AK \subset C\frac{\gamma_p(K)}{n^{1/p}}D_n.$$

Moreover, the constant c' *appearing in the probability is a function of the constant* C*, and by increasing* C *we can have* c' *as big as desired.*

The second extension we describe is to the case where instead of estimating the ℓ_2-norm of Ax, we consider its size with respect to some other norm. The most important case to consider is the ℓ_p-norm, because if we build a projection-type operator where the different coordinates are ψ_p and not ψ_2 (for a general body this will be ψ_1), then to get an estimate for their sum we have to sum p-powers of them and not squares (which may have terrible tails). We discuss the general case of measuring the size of Ax by an arbitrary norm $\|\cdot\|$ with unit ball say B. If our assumption is that the images shrink in the norm $\|\cdot\|$, we get that the image of K is inside appropriate copies of B. More precisely

Proposition 7 *Let* $\{A\}$ *be some family of operators* $A : \mathbb{R}^n \to \mathbb{R}^n$ *with some probability measure* $\mathbb{P}$*. Let* $\|\cdot\|$ *be some abstract norm, with unit ball* B*. Assume that for every* $x \in \mathbb{R}^n$ *and for any* $s > s_0$

$$\mathbb{P}[A : \|Ax\| > s|x|] \leq e^{-cns^2}.$$

Then there exist universal constants c', C *such that given a convex body* $K \subset \mathbb{R}^n$ *the following holds: If* $s_0 > (\frac{M^*(K)}{d(K)})$ *then for every* $j \geq s_0^2 n$*, with probability greater than* $1 - e^{-c'j}$ *on the choice of* A *we have*

$$AK \subset C\sqrt{\frac{j}{n}}d(K)B,$$

and if $s_0 \leq (\frac{M^*(K)}{d(K)})$ *then with probability greater than* $1-e^{-c'ns_0^2}$ *on the choice of* A *we have*

$$AK \subset CM^*(K)B.$$

Moreover, the constant c' *appearing in the probability is a function of the constant* C*, and by increasing* C *we can have* c' *as big as desired.*

One can similarly combine the two above propositions. The proofs of these two propositions are similar to the proofs given in this note. Not to overload this paper technically, we will publish them, together with some more applications, such as example 4 of Section 5 with T a general convex body, elsewhere.

References

[Art1] S. Artstein *Proportional Concentration Phenomena on the Sphere*, Israel J. Math., Vol 132 (2002), 337-358.

[Art2] S. Artstein *The change in the diameter of a convex body under a random sign-projection*, Geometric Aspects of Functional Analysis, Israel Seminar Notes (2002-2003) (Eds. Milman-Schechtman), Springer Lect. Notes series 1850, 31–40.

[AFM] S. Artstein, O. Friedland and V. Milman *Geometric Applications of Chernoff Type Estimates* To appear in GAFA Seminar notes.

[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.

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

[LPT] A. Litvak, A. Pajor and N. Tomczak-Jaegermann, *Diameters of sections, and coverings of convex bodies*, J. Func. Anal. 231 (2006) 483-457.

[Mi1] V. D. Milman, *A new proof of the theorem of A. Dvoretzky on sections of convex bodies*, Funct. Anal. Appl. 5 (1971), 28-37 (translated from Russian).

[Mi2] V. D. Milman, *Almost Euclidean quotient spaces of subspaces of a finite dimensional normed space*, Proceedings AMS 94, no. 3 (1985), 445-449.

[Mi3] V. D. Milman, *A note on a low M^*-estimate* Geometry of Banach spaces (1989), London Math. Soc. Lecture Note Ser., 158, Cambridge Univ. Press, Cambridge, 1990, 219-229.

[Mi4] V. D. Milman, *Topics in Asymptotic Geometric Analysis*, Geom. Funct. Anal., Special Volume GAFA2000 (2000), 792-815.

[MiS] V. D. Milman and G. Schechtman, *Global vs. Local asymptotic theories of finite dimensional normed spaces*, Duke Math. J. 90 (1997) 73-93.

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

Shiri Artstein-Avidan,
Department of Mathematics,
Princeton University,
Fine Hall, Washington Road,
Princeton NJ 08544-1000 USA
Email: artstein@princeton.edu