Variations and extensions of the Gaussian concentration inequality, Part I

Daniel J. Fresen*

Abstract

The classical Gaussian concentration inequality for Lipschitz functions is modified and used to prove concentration inequalities in cases where the classical assumptions (i.e. Lipschitz and Gaussian) are not met. The theory is simpler than much of the existing theory designed to handle related generalizations. Applications include concentration of linear combinations of heavy tailed random variables, and a variation of Milman's general Dvoretzky theorem for non-Gaussian random matricies with i.i.d. entries.

1 Introduction

Recall the Gaussian concentration inequality in one of its most classical forms: if $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz, and X is a random vector in $\mathbb{R}^n$ with the standard normal distribution, then for all $t > 0$,

$$\mathbb{P}\left\{\left|f\left(X\right) - \mathbb{M}f\left(X\right)\right| > t\right\} \leq C\exp\left(-c\frac{t^2}{Lip\left(f\right)^2}\right) \tag{1}$$

where $C, c > 0$ are universal constants. $\mathbb{M}f\left(X\right)$ denotes the median of $f\left(X\right)$, and can be replaced with the mean $\mathbb{E}f\left(X\right)$. It follows from the Gaussian isoperimetric inequality of Sudakov and Tsirelson [37] and Borell [4] that this can be improved to

$$\mathbb{P}\left\{f\left(X\right) - \mathbb{M}f\left(X\right) > t\right\} \leq 1 - \Phi\left(\frac{t}{Lip\left(f\right)}\right)$$

where Φ is the standard normal cumulative distribution. Equality clearly holds when f is linear. Assuming for simplicity that f is C^1, it follows from a result of Pisier [33, Theorem 2.2 p176] that if Y is another random vector in $\mathbb{R}^n$ with the standard normal distribution, independent of X, then for any convex function $\varphi : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}\varphi\left(f\left(X\right) - f\left(Y\right)\right) \leq \mathbb{E}\varphi\left(\frac{\pi}{2}\left|\nabla f\left(X\right)\right| Z\right)$$

*University of Pretoria, Department of Mathematics and Applied Mathematics, daniel.fresen@up.ac.za or djfb6b@mail.missouri.edu. 2010 *MSC*: 60E05, 60E15, 52A23. *Keywords*: Gaussian concentration inequality, heavy tailed random variables, random sections of convex bodies.

where Z has the standard normal distribution in $\mathbb{R}$ and is independent of X. This often implies that one can compare the tails of $f(X) - f(Y)$ to that of $\frac{\pi}{2}|\nabla f(X)| Z$, which in turn leads to a bound on the tails of $f(X) - \mathbb{M}f(X)$. Pisier's version of Gaussian concentration has two advantages over the classical version. The first is that the proof (with input from Maurey) is surprisingly simple. The second is that the Lipschitz condition has been removed.

In two papers, 'Variations and extensions of the Gaussian concentration inequality' Part I (here) and Part II (see [15]), we study ways in which to apply the above inequalities in settings where they are usually not applied. The result is that the Gaussian concentration inequality is the mother of a whole family of inequalities that seem to have nothing to do with the normal distribution. This paper, Part I, focuses more on the classical version, while Part II focuses on Pisier's version. The theory is from scratch, in that no background is needed, other than a basic knowledge of elementary mathematics (the one thing that we don't prove is the Gaussian concentration inequality itself). Our opinion is that it is more direct than much of the modern theory of concentration of measure, including for example the theory of Poincaré and log-Sobolev inequalities. We refer the reader to [5] for information on the concentration of measure phenomenon. In Section 1.1 we discuss the methodology and provide a context for the results of Section 2, which is the theoretical backbone of the paper. In Section 3, following up on work of Hitczenko, Montgomery-Smith and Oleszkiewicz [22], we prove a concentration inequality for linear combinations of heavy tailed random variables, and in Section 4 we prove a generalization of Milman's general Dvoretzky theorem [28, 35] for non-uniformly distributed subspaces.

1.1 Methodology: Gaussian concentration

Surgery on f: Consider the setting where X has the standard normal distribution in $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ is, say, C^1, but is not assumed to be Lipschitz. Restrict f to some set E such that $X \in E$ with high probability and $|\nabla f|$ is bounded nicely on E. The set $E = \{x \in \mathbb{R}^n : |\nabla f(x)| \leq R\}$ is often a natural choice. Then bound $Lip(f|_E) \leq \mathfrak{L}(E) \sup_E |\nabla f|$, where $\mathfrak{L}(E)$ is a parameter that measures the connectivity of E. When E is convex $\mathfrak{L}(E) = 1$, in which case we get the obvious bound $Lip(f|_E) \leq \sup_E |\nabla f|$, and whenever E is a certain non-affine deformation of an unconditional convex body $\mathfrak{L}(E) \leq \sqrt{2}$. One can then extend the restriction $f|_E$ to the entire space $\mathbb{R}^n$ so that the extension obeys $Lip(f^*) = Lip(f|_E) \leq \mathfrak{L}(E) \sup_E |\nabla f|$. By applying classical Gaussian concentration of Lipschitz functions to f^* and observing that $\mathbb{P}\{f(X) = f^*(X)\} \geq \mathbb{P}\{X \in E\}$, we may transfer the concentration inequality for $f^*(X)$ about $\mathbb{M}f^*(X)$ to an inequality for $f(X)$ about $\mathbb{M}f(X)$.

The idea of proving concentration inequalities by modifying a function on a set where it behaves badly is not new, see for example Grable [21, Corollary] and Vu [39, Section 3] (where the measure is supported on $[0,1]^n$ and the Lipschitz constant is taken with respect to the Hamming distance, see the bottom of p. 264 there). In [39] and [31, Theorem 5.1], for example, they also make use of the average local Lipschitz constant as opposed to the (global) Lipschitz constant. However this particular observation (restriction based on $|\nabla f(Z)|$, Lipschitz extension, and the parameter $\mathfrak{L}$), goes back to a related observation which was contained in unpublished lecture notes [13] that we prepared and distributed at Yale while teaching graduate classes in 2012 and 2014, where we pointed out that one can prove Lévy's concentration inequality for Lipschitz functions on the sphere S^{n-1} (equivalently on $\sqrt{n}S^{n-1}$) by extending them to $\mathbb{R}^n$ and applying Gaussian concentration to the extension. We acknowledge that a similar observation, without theory surrounding the parameter $\mathfrak{L}(\cdot)$, is also contained in a paper of Bobkov, Nayar, and Tetali [3].

1.2 Methodology: Non-Gaussian concentration

If X is a random vector in $\mathbb{R}^n$ with any distribution μ, then we may always write $X = T(Z)$ for some measurable function $T : \mathbb{R}^n \to \mathbb{R}$, where Z is a random vector with the standard normal distribution on $\mathbb{R}^n$. An example of such a map is the Knöthe-Rosenblatt rearrangement [38], and in some cases one can write down an explicit formula for T. We then write $f(X) = (f \circ T)(Z)$, and under fairly general conditions we may then apply Gaussian concentration to $f \circ T$ to obtain a concentration inequality for $f(X)$. A critical tool here (in the case of product measures) is Proposition 2 (alluded to above) which implies that the inverse image of an unconditional convex body under an appropriate coordinate-wise transformation has parameter $\mathfrak{L}(T^{-1}K) \leq \sqrt{2}$, where $\mathfrak{L}(\cdot)$ is the parameter mentioned above. A related procedure is certainly well known in the one dimensional case, where it is common to write a random variable X as TU, where U is uniformly distributed on $(0, 1)$. In the multivariate case it has also been used, without the parameter $\mathfrak{L}(\cdot)$, see for example the comment in the lower half of p.1046 in Naor's paper [30]. But there it is specifically mentioned in the context of Lipschitz images of the standard Gaussian measure. Note that if we can apply Gaussian concentration to a wider class of functions f as in Section 1.1 above, then we may apply Gaussian concentration to a wider class of non-Gaussian measures (i.e. a wider class of $f \circ T$ in the notation of this section). The observations in Sections 1.1 and those of this section therefore work particularly well *together* and each significantly increases the usefulness of the other. This synergy is at the heart of the paper.

One is thus left with the problem of finding a good choice of T. If μ is an n-fold product measure, then the most natural T acts coordinate-wise in the obvious manner (and is the Knöthe-Rosenblatt rearrangement). If μ is spherically symmetric, then the most natural T acts radially. For most other measures, we expect that the Knöthe-Rosenblatt rearrangement is not a good choice. In the case of log-concave measures, the Brenier map (see for example [24]) may be better, although the Brenier map is best for minimizing a transportation cost, which is not exactly what we want. We focus mainly on the case of product measures.

1.3 Notation and terminology

The median of a random variable is denoted by the operator $\mathbb{M}$, which may denote any median when not unique. The symbols C, c, C' etc. will usually denote unspecified but fixed positive universal constants that may represent different values at each appearance. Sometimes, upper case letters denote possibly large constants in $[1, \infty)$ while lower case letters denote possibly small constants in $(0, 1]$, but this should not be taken too seriously. Dependence on variables will usually be indicated by subscripts, C_q, c_q etc. The term 'random variable' will be used exclusively for real valued random variables, and not for random vectors in $\mathbb{R}^n$ $(n \geq 2)$, or for complex random variables. In common abuse of terminology we will make statements like 'let μ be a probability measure on $\mathbb{R}$', when in fact μ is defined on Borel subsets of $\mathbb{R}$. The term 'Gaussian concentration' is used to mean concentration of $f(Z)$ about, say, its median, where Z has the standard normal distribution. We use this terminology even if the distribution of $f(Z)$ has super-Gaussian tails, which may happen if f grows more rapidly than linear.

2 General results on which the techniques are based

For any non-empty polygonally connected set $A \subseteq \mathbb{R}^n$, let $\mathfrak{L}(A)$ denote the largest possible ratio of the polygonal distance between points in A and the Euclidean distance (only considering polygonal paths contained in A). In more detail, $\mathfrak{L}(A)$ is the infimum over all values of $\lambda \geq 1$ such that for all $x, y \in A$ there exists a finite sequence $(u_i)_1^N$ in $\mathbb{R}^n$ with $u_1 = x$ and $u_N = y$, such that for all $2 \leq i \leq N$ and all $t \in [0,1]$, $tu_i + (1-t)u_{i-1} \in A$, and such that $\sum_{i=2}^{N} |u_i - u_{i-1}| \leq \lambda |x - y|$. If A is not polygonally connected we set $\mathfrak{L}(A) = \infty$, and set $\mathfrak{L}(\emptyset) = 1$. Whenever A is convex, $\mathfrak{L}(A) = 1$. Conversely if $\mathfrak{L}(A) = 1$ and A is closed, then A is convex. For a function $Q : \mathbb{R}^n \to \mathbb{R}$ define

$$\mathfrak{L}(Q) = \sup_{t \in \mathbb{R}} \mathfrak{L}\left(\{x \in \mathbb{R}^n : Q(x) \leq t\}\right) \tag{2}$$

Let $Lip\,(T, x)$ denote the local Lipschitz constant of a function $T : A \to \mathbb{R}$ around a point x,

$$Lip\,(T, x) = \lim_{\varepsilon \to 0^+} Lip\left(T|_{B(x,\varepsilon)\cap A}\right) \tag{3}$$

Our main reason for defining $\mathfrak{L}(\cdot)$ is the following observation.

Proposition 1 *Let $A \subseteq \mathbb{R}^n$ be any polygonally connected set containing at least two points, with $\mathfrak{L}(A) < \infty$. Then for any function $f : A \to \mathbb{R}$,*

$$Lip\,(f) \leq \mathfrak{L}(A) \sup\left\{Lip\,(T, x) : x \in A\right\}$$

Furthermore, if A is locally convex in the sense that for all $x \in A$ there exists $\varepsilon > 0$ such that $B(x, \varepsilon) \cap A$ is convex, then

$$\mathfrak{L}(A) = \sup_{f : A \to \mathbb{R}} \left\{ \frac{Lip\,(f)}{\sup\left\{Lip\,(T, x) : x \in A\right\}} : 0 < Lip\,(f) < \infty \right\}$$

Proof of Proposition 1. The first part is elementary and reduces to the one dimensional case, and implies a lower bound for $\mathfrak{L}(A)$. We now prove the second part under the assumption that A is locally convex, as defined in the statement of the theorem. For any $x, y \in A$, let $\rho(x, y)$ denote the shortest Euclidean length (infimum) of all polygonal paths between x and y. It is easily seen that ρ is a metric on A. Now consider any $\varepsilon > 0$. It follows (almost immediately) from the definition of $\mathfrak{L}(A)$, that there exist $x, y \in A$ such that $\rho(x, y) > (\mathfrak{L}(A) - \varepsilon)\,|x - y|$. Now consider the function $g : A \to [0, \infty)$ defined as

$g(z) = \rho(x, z)$, for which $g(x) = 0$, $g(y) = \rho(x, y)$, so $Lip(g) > \mathfrak{L}(A) - \varepsilon$. It follows from the definition of ρ, the triangle inequality, and the assumption of local convexity that g is locally 1-Lipschitz. This shows that

$$\mathfrak{L}(A) < \frac{Lip(g)}{\sup\{Lip(g, x) : x \in A\}} + \varepsilon$$

and the result follows by sending $\varepsilon \to 0^+$. ■

We are interested in the case where A is a non-affine deformation of a convex body K, as the inverse image under the action of some map $T : \mathbb{R}^n \to \mathbb{R}^n$. When K is 1-unconditional (i.e. invariant under coordinate reflections) and T acts coordinatewise and monotonically, then $\mathfrak{L}(A) \leq \sqrt{2}$. The actual result can be stated a bit more generally as follows:

Proposition 2 *Let $n \in \mathbb{N}$ and let Υ_n denote the collection of all $K \subseteq \mathbb{R}^n$ with the following property: there exists $a \in [-\infty, \infty]^n$ (depending on K) such that if $x \in K$ and $y \in \mathbb{R}^n$, and each coordinate of y is between the corresponding coordinates of x and a in the non-strict sense (i.e. either $a_i \leq y_i \leq x_i$ or $x_i \leq y_i \leq a_i$), then $y \in K$. For each $1 \leq i \leq n$, let $h_i : \mathbb{R} \to \mathbb{R}$ be a non-decreasing function, and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be defined as $Tx = (h_i(x_i))_1^n$. Then for any $K \in \Upsilon_n$, $\mathfrak{L}(K) \leq \sqrt{2}$ and $T^{-1}K \in \Upsilon_n$, and so $\mathfrak{L}(T^{-1}(K)) \leq \sqrt{2}$.*

Proof of Proposition 2. Consider any $K \in \Upsilon_n$. We first show that $T^{-1}K \in \Upsilon_n$. Since $K \in \Upsilon_n$ $\exists a \in [-\infty, \infty]^n$ as in the statement of the theorem. For each $1 \leq i \leq n$, since h_i is non-decreasing, there exists $b_i \in [-\infty, \infty]$ such that for all $t \in \mathbb{R}$, if $t \leq b_i$, then $h_i(t) \leq a_i$, and if $t \geq b_i$ then $h_i(t) \geq a_i$ (consider three cases: a_i is an upper bound for $range(h_i)$, a_i is a lower bound, or neither). Now consider any $x \in T^{-1}(K)$ and $y \in \mathbb{R}^n$, such that the coordinates of y are between the corresponding coordinates of x and those of b (always meant in the non-strict sense). By the fact that the h_i are non-decreasing, and by construction of b, it follows that the coordinates of Ty are all between the coordinates $Tx \in K$ and those of a. Since $K \in \Upsilon_n$, what we have just shown implies that $Ty \in K$, and therefore $y \in T^{-1}K$. This shows that $T^{-1}K$ satisfies the defining property of Υ_n. We now show that $\mathfrak{L}(K) \leq \sqrt{2}$. If K is empty, or a singleton, then $\mathfrak{L}(K) = 1$, and we may assume without loss of generality that $|K| \geq 2$. Consider any $x, y \in K$ with $x \neq y$. Now define $z \in \mathbb{R}^n$ as follows. If a_i is between x_i and y_i (which is only possible if $a_i \in \mathbb{R}$), then set $z_i = a_i$ (and let the collection of all such i be denoted E), otherwise let z_i be the element of the set $\{x_i, y_i\}$ that is closest to a_i, with the obvious interpretation when $a_i \in \{\pm\infty\}$. For all $\lambda \in [0, 1]$ and all $1 \leq i \leq n$, $\lambda z_i + (1 - \lambda) x_i$ is between a_i and x_i, and therefore $\lambda z + (1 - \lambda) x \in K$. Similarly, $\lambda y + (1 - \lambda) z \in K$, and this defines a polygonal path of length $|x - z| + |y - z|$ in K from x to y. Furthermore,

$$\langle x - z, y - z \rangle = \sum_{i \in E} (x_i - a_i)(y_i - a_i) \leq 0$$

Using this inequality and comparing the ℓ_2^2 and ℓ_1^2 norms,

$$|x - y|^2 = |x - z|^2 + |y - z|^2 - 2 \langle x - z, y - z \rangle \geq \frac{1}{2} (|x - z| + |y - z|)^2$$

and it follows that $\mathfrak{L}(K) \le \sqrt{2}$. ■

Let Φ denote the standard normal cumulative distribution and $\phi = \Phi'$ the standard normal density.

Theorem 3 *Let $n \in \mathbb{N}$, $A > 0$, let μ be a probability measure on $\mathbb{R}^n$, and let $H : \mathbb{R}^n \to \mathbb{R}^n$ and $\psi, Q : \mathbb{R}^n \to \mathbb{R}$ be measurable functions such that $\mu = H\gamma_n$, where γ_n is the standard Gaussian measure on $\mathbb{R}^n$ with density $d\gamma_n/dx = (2\pi)^{-n/2} \exp\left(-|x|^2/2\right)$, and such that $\psi \circ H$ is locally Lipschitz with*

$$Q(x) \ge Lip(\psi \circ H, x)$$

for all $x \in \mathbb{R}^n$. Let X and Z be random vectors in $\mathbb{R}^n$, where the distribution of X is μ, and Z follows the standard normal distribution. Let $R > 0$ and $t > \Phi^{-1}\left(1 - (2A+4)^{-1}\right)$ be such that $\mathbb{P}\{Q(Z) > R\} \le A(1 - \Phi(t))$. Then

$$\mathbb{P}\{|\psi(X) - \mathbb{M}\psi(X)| > 2\mathfrak{L}(Q)Rt\} \le (A+2)(1 - \Phi(t))$$

where $\mathfrak{L}(Q)$ is defined by (2).

Proof of Theorem 3. Since the distribution of the random vector $H(Z)$ is μ, we may assume without loss of generality that $X = H(Z)$. Set $\psi^\sharp = \psi \circ H$, in which case $\psi^\sharp(Z) = \psi(X)$. Let $K = \{x : Q(x) \le R\}$. By assumption, $\mathbb{P}\{Z \in K\} > 1 - A(1 - \Phi(t))$, and for all $x \in K$, $Lip\left(\psi^\sharp|_K, x\right) \le Lip\left(\psi^\sharp, x\right) \le R$. By Proposition 1, $Lip\left(\psi^\sharp|_K\right) \le \mathcal{L}(Q)R$. The function $\psi^\sharp|_K$ may then be extended to a function $\widetilde{\psi} : \mathbb{R}^n \to \mathbb{R}$ such that $Lip\left(\widetilde{\psi}\right) = Lip\left(\psi^\sharp|_K\right)$, (the existence of extensions of Lipschitz functions is a basic result in the theory of metric spaces and Lipschitz functions). By Gaussian concentration of Lipschitz functions and the union bound, it follows that with probability at least $1 - (A+2)(1 - \Phi(t))$, $\psi^\sharp(Z) = \widetilde{\psi}(Z)$, and $\left|\widetilde{\psi}(Z) - \mathbb{M}\widetilde{\psi}(Z)\right| \le \mathcal{L}(Q)Rt$. Since $(A+2)(1 - \Phi(t)) < 1/2$, this implies that greater than 50% of the mass of the distribution of $\psi^\sharp(Z)$ lies in the closed interval from $\mathbb{M}\widetilde{\psi}(Z) - \mathcal{L}(Q)Rt$ to $\mathbb{M}\widetilde{\psi}(Z) + \mathcal{L}(Q)Rt$, and must be a median, so $\left|\mathbb{M}\psi^\sharp(Z) - \mathbb{M}\widetilde{\psi}(Z)\right| \le \mathcal{L}(Q)Rt$. The result now follows by the triangle inequality. ■

The following lemma will be used implicitly.

Lemma 4 *If $f, g : [0,\infty) \to [0,\infty)$ are continuous strictly increasing functions with $f(0) = g(0) = 0$, $t \in [0,\infty)$ and $s = \max\{f(t), g(t)\}$ then $t = \min\{f^{-1}(s), g^{-1}(s)\}$.*

3 Application: Concentration of linear combinations of heavy tailed random variables

Concentration of linear combinations of independent random variables is most classically studied under the assumption of exponential integrability, i.e. $\mathbb{E}\exp(\varepsilon X_i) < \infty$ for some

$\varepsilon > 0$. In this context, the exponential moment method plays an essential role: Using Markov's inequality and independence,

$$\begin{aligned}\mathbb{P}\left\{\sum_{i=1}^{n} a_i X_i > t\right\} &= \mathbb{P}\left\{\exp\left(\lambda \sum_{i=1}^{n} a_i X_i\right) > \exp(\lambda t)\right\} \leq \exp(-\lambda t)\,\mathbb{E}\exp\left(\lambda \sum_{i=1}^{n} a_i X_i\right) \\ &= \exp(-\lambda t) \prod_{i=1}^{n} \mathbb{E}\exp(\lambda a_i X_i)\end{aligned}$$

The resulting estimate is then optimized over $\lambda > 0$ such that $\mathbb{E}\exp(\lambda a_i X_i) < \infty$ for all i. Outside the realm of exponential integrability (still assuming independence), one would estimate power moments and use Markov's inequality,

$$\mathbb{P}\left\{\left|\sum_{i=1}^{n} a_i X_i\right| > t\right\} = \mathbb{P}\left\{\left|\sum_{i=1}^{n} a_i X_i\right|^p > t^p\right\} \leq t^{-p}\mathbb{E}\left|\sum_{i=1}^{n} a_i X_i\right|^p \tag{4}$$

If we assume that each X_i has a symmetric distribution and that $\mathbb{P}\{|X_i| \geq t\} = \exp(-N_i(t))$ for some concave function $N : [0,\infty) \to [0,\infty)$, then it was shown by Hitczenko, Montgomery-Smith and Oleszkiewicz [22, Theorem 1.1] that for all $p \geq 2$,

$$\left(\mathbb{E}\left|\sum_{i=1}^{n} a_i X_i\right|^p\right)^{1/p} \leq C\left(\left(\sum_{i=1}^{n} |a_i|^p\,\mathbb{E}|X_i|^p\right)^{1/p} + \sqrt{p}\left(\sum_{i=1}^{n} |a_i|^2\,\mathbb{E}|X_i|^2\right)^{1/2}\right) \tag{5}$$

where $C > 0$ is a universal constant, with a corresponding lower bound with C replaced by a different constant $c > 0$. In the special case where $(X_i)_1^n$ are i.i.d. symmetric Weibull variables with $\mathbb{P}\{|X_i| \geq t\} = \exp(-t^q)$, for $0 < q \leq 1$, then $c_q p^{1/q} \leq (\mathbb{E}|X_i|^p)^{1/p} \leq C_q p^{1/q}$ and assuming without loss of generality that $|a| = 1$, we show below that $|a|_p$ can be replaced with $|a|_\infty$ and the bound written as

$$c\left((\mathbb{E}|X_1|^p)^{1/p}|a|_\infty + \sqrt{p}\left(\mathbb{E}|X_1|^2\right)^{1/2}\right) \leq \left(\mathbb{E}\left|\sum_{i=1}^{n} a_i X_i\right|^p\right)^{1/p} \leq C_q\left(p^{1/q}|a|_\infty + \sqrt{p}\right) \tag{6}$$

which by Markov's inequality leads to the tail estimate

$$\mathbb{P}\left\{\left|\sum_{i=1}^{n} aX_i\right| > t\right\} \leq 2\exp\left(-c_q \min\left\{\left(\frac{t}{|a|}\right)^2, \left(\frac{t}{|a|_\infty}\right)^q\right\}\right) \tag{7}$$

where $C_q, c_q > 0$ can be taken as universal constants when q is bounded away from 0. This estimate is sharp up to the values of C_q and c_q and ties in naturally with known results in the case $1 \leq q < \infty$ since the dual of ℓ_q^n is isometric to ℓ_∞^n when $0 < q \leq 1$. Surprisingly, we haven't seen this estimate in the literature. Examples of where such a bound might have appeared but doesn't include [1, Eq. (3.6) is specifically for $a_i = n^{-1/2}$], [5, Ex. 2.27 p.50 is specifically for $q \geq 1$], [25, Sec. 4 does not include the sub-Gaussian part], and [22, Th 6.2 (moments estimates, with $|a|_p$ instead of $|a|_\infty$), Cor. 6.5 (they show that $\lim_{t\to\infty} \log_t \ln 1/\mathbb{P}\{|\sum_{i=1}^n aX_i| > t\} = q$)].

If X_1 is any random variable and X_1' is an independent copy of X_1, and if $a > 0$ is such that $\mathbb{P}\{|X_1| > a\} \le 1/2$, then

$$\{X_1 > t + a\} \cap \{X_1' \le a\} \subseteq \{X_1 - X_1' > t\} \subseteq \{X_1 > t/2\} \cup \{X_1' < -t/2\}$$

so by independence and identical distributions,

$$\frac{1}{2}\mathbb{P}\{X_1 > t + a\} \le \mathbb{P}\{X_1 - X_1' > t\} \le \mathbb{P}\{|X_1| > t/2\} \tag{8}$$

The significance is that $X_1 - X_1'$ is symmetric. This can be combined with the following contraction principle, see [26, Lemma 4.6] for a more general version:

Lemma 5 *Let $\varphi : [0,\infty) \to [0,\infty)$ be a convex function, $K_1 \ge 1$, $K_2 > 0$, and let $(X_i)_1^n$ and $(Y_i)_1^n$ each be i.i.d. sequences of symmetric random variables with $\mathbb{P}\{|X_i| > t\} \le K_1 \mathbb{P}\{K_2 |Y_i| > t\}$ for all i and all $t > 0$. Then for all $a \in \mathbb{R}^n$,*

$$\mathbb{E}\varphi\left(\left|\sum_{i=1}^n a_i X_i\right|\right) \le \mathbb{E}\varphi\left(K_1 K_2 \left|\sum_{i=1}^n a_i Y_i\right|\right)$$

If we are given an i.i.d. sequence of random variables $(X_i)_1^n$ that satisfy a tail bound such as $\mathbb{P}\{|X_i| > t\} \le h(t)$, then we consider the symmetrized sequence $(X_i - X_i')_1^n$ which obeys a similar tail bound, apply known results for a specific sequence of random variables $(Y_i)_1^n$ with similar tails (e.g. Weibull variables), compare $\mathbb{E}\varphi\left(\left|\sum_{i=1}^n a_i (X_i - X_i')\right|\right)$ to $\mathbb{E}\varphi\left(K_1 K_2 \left|\sum_{i=1}^n a_i Y_i\right|\right)$ using Lemma 5, convert this to a bound on $\mathbb{P}\left\{\left|\sum_{i=1}^n a_i (X_i - X_i')\right| > t\right\}$ (using say Markov's inequality), and then transfer the result for $\sum_{i=1}^n a_i X_i - \sum_{i=1}^n a_i X_i'$ back to a bound on $\mathbb{P}\left\{\left|\sum_{i=1}^n a_i X_i\right| > t\right\}$ using (8). In this way, estimates such as (7) may be extended to the case of tail bounds such as $\mathbb{P}\{|X_i| > t\} \le C\exp(-t^q)$.

Proof of (6) and (7). The case $p = 2$ in (6) follows from the usual estimate of a variance, so assume that $p > 2$. The lower bound in (6) follows trivially from $|a|_\infty \le |a|_p$. By log-convexity of the map $t \mapsto |x|_{1/t}$, $t \in [0,1]$, $1/0 = \infty$, which follows directly from Hölder's inequality, $|a|_p \le |a|_\infty^{1-2/p} |a|^{2/p}$. The replacement of $|a|_p$ with $|a|_\infty$ in the upper bound holds trivially unless $(\mathbb{E}|X_1|^p)^{1/p} |a|_p$ is at least $C_q\sqrt{p}\left(\mathbb{E}|X_1|^2\right)^{1/2}$ (for arbitrary $C_q > 0$), in which case

$$\begin{aligned}
(\mathbb{E}|X_1|^p)^{1/(p-2)} |a|_\infty &\ge (\mathbb{E}|X_1|^p)^{1/(p-2)} |a|_p^{p/(p-2)} \\
&\ge (\mathbb{E}|X_1|^p)^{1/(p-2)} \left\{ (\mathbb{E}|X_1|^p)^{-1/p} \sqrt{p} \left(\mathbb{E}|X_1|^2\right)^{1/2} \right\}^{p/(p-2)} \\
&\ge C\sqrt{p}\left(\mathbb{E}|X_1|^2\right)^{p/[2(p-2)]}
\end{aligned}$$

Now consider $s = s_q > 2$ such that for $(\mathbb{E}|X_1|^s)^{1/s} |a|_s = C_q\sqrt{s}\left(\mathbb{E}|X_1|^2\right)^{1/2}$. $C_q > 0$ can be chosen to ensure that a solution exists in, say, $(3,\infty)$. From what we have observed above, (6) holds for all $p \in [2, s_q]$, and for all $p > s_q$,

$$\frac{p^{1/p} |a|_\infty}{(\mathbb{E}|X_1|^p)^{1/p} |a|_p} \ge c_q \frac{s^{1/s} |a|_\infty}{(\mathbb{E}|X_1|^s)^{1/s} |a|_s} = c_q' \frac{s^{1/s} |a|_\infty}{\sqrt{s}} \ge c_q''$$

so

$$\left(\mathbb{E}\left|\sum_{i=1}^{n}a_iX_i\right|^p\right)^{1/p}\leq C\left(\left(\mathbb{E}\left|X_1\right|^p\right)^{1/p}\left|a\right|_p+\sqrt{p}\left(\mathbb{E}\left|X_1\right|^2\right)^{1/2}\right)\leq C_q\left(p^{1/q}\left|a\right|_\infty+\sqrt{p}\right)$$

(6) in full generality in now established. The probability bound follows by optimizing over p. In Case 1 we assume that $p^{1/q}\left|a\right|_\infty\leq\sqrt{p}$, and then

$$\mathbb{P}\left\{\left|\sum_{i=1}^{n}aX_i\right|>t\right\}\leq t^{-p}\mathbb{E}\left|\sum_{i=1}^{n}a_iX_i\right|^p\leq\left(\frac{C_q\sqrt{p}}{t}\right)^p=\exp\left(-c_qt^2\right)$$

for $p=c_qt^2$. This value of p satisfies the defining inequality of Case 1 if $t\leq c_q\left|a\right|_\infty^{-q/(2-q)}$. In Case 2 we assume that $p^{1/q}\left|a\right|_\infty\geq\sqrt{p}$, and then

$$\mathbb{P}\left\{\left|\sum_{i=1}^{n}aX_i\right|>t\right\}\leq t^{-p}\mathbb{E}\left|\sum_{i=1}^{n}a_iX_i\right|^p\leq\left(\frac{C_qp^{1/q}\left|a\right|_\infty}{t}\right)^p\leq\exp\left(-C_q\frac{t^q}{\left|a\right|_\infty^q}\right)$$

for $p=\left(C_q^{-1}\left|a\right|_\infty^{-1}t\right)^qe^{-1}$. This value of p is allowed in Case 2 provided $t\geq C_q\left|a\right|_\infty^{-q/(2-q)}$. For $c_q\left|a\right|_\infty^{-q/(2-q)}\leq t\leq C_q\left|a\right|_\infty^{-q/(2-q)}$, the result follows by adjusting the values of c_q and C_q and using the fact that the cumulative distribution is non-decreasing. ∎

We now present a direct proof of (7) without using (6) or the results of [22].

Theorem 6 *There exists $C>0$ such that the following is true. Let $n\in\mathbb{N}$, $0<q\leq 1$, $a\in\mathbb{R}^n$, $a\neq 0$, and $(X_i)_1^n$ an i.i.d. sequence of symmetric Weibull random variables with parameter q, i.e. $\mathbb{P}\left\{\left|X_i\right|>t\right\}=\exp\left(-t^q\right)$, $t\geq 0$. Then for all $t>0$,*

$$\mathbb{P}\left\{\left|\sum_{i=1}^{n}aX_i\right|>t\right\}\leq C\exp\left(-\min\left\{C^{-1/q}q^{2/q}\left(\frac{t}{\left|a\right|}\right)^2,C^{-1}\left(\frac{t}{\left|a\right|_\infty}\right)^q\right\}\right)$$

Proof. Write $X=\left(F^{-1}\Phi\left(Z_i\right)\right)_{i=1}^n$, where $F(t)=\mathbb{P}\left\{X_1\leq t\right\}$ and Z is a random vector in $\mathbb{R}^n$ with the standard normal distribution, and define $\psi\left(x\right)=\sum_1^na_iF^{-1}\Phi\left(x_i\right)$, so that $\sum_1^na_iX_i=\psi\left(Z\right)$, and

$$\left|\nabla\psi\left(x\right)\right|=\left(\sum_{i=1}^{n}\left(\frac{a_i\phi\left(\Phi^{-1}\left(\Phi x_i\right)\right)}{f\left(F^{-1}\left(\Phi x_i\right)\right)}\right)^2\right)^{1/2}$$

where $f=F'$. Now, by comparing derivatives and behavior at infinity,

$$\frac{\phi(t)}{1+t}\leq 1-\Phi\left(t\right)\leq\frac{\phi(t)}{t}:t>0$$

which implies that for $1/2\leq t<1$, $\phi\left(\Phi^{-1}\left(t\right)\right)\leq\left(1+\Phi^{-1}\left(t\right)\right)\left(1-t\right)$. Now $1-\Phi\left(t\right)\leq\phi(t)$ for $t\geq 1$, so for $1/2\leq t<1$ we have (taking max $=1$ when second argument undefined),

$$1+\Phi^{-1}\left(t\right)\leq 1+\max\left\{1,\sqrt{2\ln\frac{1}{\sqrt{2\pi}\left(1-t\right)}}\right\}$$

By direct computation, for the same range of t,

$$f\left(F^{-1}(t)\right)=q(1-t)\left(\ln\frac{1}{2(1-t)}\right)^{-(-1+1/q)}$$

so

$$\left(\frac{\phi\left(\Phi^{-1}(t)\right)}{f\left(F^{-1}(t)\right)}\right)^2\leq q^{-2}\left(A\Phi^{-1}(t)+B\right)^{-2+4/q}$$

for universal constants $A,B>0$, and

$$\begin{aligned}|\nabla\psi(x)| &= \left(\sum_{i=1}^n\left(\frac{a_i\phi\left(\Phi^{-1}\left(\Phi|x_i|\right)\right)}{f\left(F^{-1}\left(\Phi|x_i|\right)\right)}\right)^2\right)^{1/2}\\ &\leq q^{-1}\left(\sum_{i=1}^n a_i^2\left(A|x_i|+B\right)^{-2+4/q}\right)^{1/2}\end{aligned}$$

The function

$$]x[=\left(\sum_{i=1}^n a_i^2|x_i|^{-2+4/q}\right)^{q/(4-2q)}$$

is a norm, so

$$\begin{aligned}Lip\left(]\cdot[\right) &= \sup\left\{]\theta[:\theta\in S^{n-1}\right\}\\ &= \sup\left\{\left(\sum_{i=1}^n a_i^2|x_i|^{-1+2/q}\right)^{q/(4-2q)}:x_i\geq 0,\sum_{i=1}^n x_i=1\right\}\\ &= |a|_\infty^{q/(2-q)}\end{aligned}$$

The Lipschitz constant of $x\mapsto\,](A|x_i|+B)_1^n[$ is therefore at most $A|a|_\infty^{q/(2-q)}$, and by classical Gaussian concentration applied to this function, with probability at least $1-C\exp\left(-c\lambda^2\right)$,

$$\begin{aligned}&\left(\sum_{i=1}^n a_i^2\left(A|Z_i|+B\right)^{-2+4/q}\right)^{q/(4-2q)}\\ \leq\ &\mathbb{E}\left(\sum_{i=1}^n a_i^2\left(A|Z_i|+B\right)^{-2+4/q}\right)^{q/(4-2q)}+A|a|_\infty^{q/(2-q)}\lambda\\ \leq\ &\left(\sum_{i=1}^n a_i^2\mathbb{E}\left(A|Z_1|+B\right)^{-2+4/q}\right)^{q/(4-2q)}+A|a|_\infty^{q/(2-q)}\lambda\\ =\ &\left(\mathbb{E}\left(A|Z_1|+B\right)^{-2+4/q}\right)^{q/(4-2q)}|a|^{q/(2-q)}+A|a|_\infty^{q/(2-q)}\lambda\end{aligned}$$

i.e.

$$\begin{aligned}
& \left(\sum_{i=1}^{n} a_i^2 \left(A\left|Z_i\right|+B\right)^{-2+4/q}\right)^{1/2} \\
\leq \; & \left[\left(\mathbb{E}\left(A\left|Z_1\right|+B\right)^{-2+4/q}\right)^{q/(4-2q)}|a|^{q/(2-q)}+A\,|a|_{\infty}^{q/(2-q)}\lambda\right]^{-1+2/q}
\end{aligned}$$

The above implies that the convex set

$$K=\left\{x\in\mathbb{R}^n:\left|\nabla\psi\left(x\right)\right|\leq R\right\}$$

where

$$\begin{aligned}
R &= q^{-1}\left[\left(\mathbb{E}\left(A\left|Z_1\right|+B\right)^{-2+4/q}\right)^{q/(4-2q)}|a|^{q/(2-q)}+A\,|a|_{\infty}^{q/(2-q)}\lambda\right]^{-1+2/q} \\
&\leq q^{-1}2^{-1+2/q}\left(\mathbb{E}\left(A\left|Z_1\right|+B\right)^{-2+4/q}\right)^{1/2}|a|+q^{-1}2^{-1+2/q}A\,|a|_{\infty}\lambda^{-1+2/q} \\
&\leq C_2^{1/q}\left(1/q\right)^{1/q}|a|+q^{-1}2^{-1+2/q}A\,|a|_{\infty}\lambda^{-1+2/q}
\end{aligned}$$

satisfies $\gamma_n(K)\geq 1-C\exp\left(-c\lambda^2\right)$. Let $\psi^{\sharp}$ denote a Lipschitz extension of the restriction $\psi|_K$ such that $Lip\left(\psi^{\sharp}\right)\leq R$. Applying Gaussian concentration to $\psi^{\sharp}$,

$$\mathbb{P}\left\{\left|\psi^{\sharp}(Z)-\mathbb{M}\psi^{\sharp}(Z)\right|>\lambda R\right\}\leq C\exp\left(-c\lambda^2\right)$$

yet $\mathbb{P}\left\{\psi^{\sharp}(Z)\neq\psi(Z)\right\}\leq C\exp\left(-c\lambda^2\right)$ so $\mathbb{P}\left\{\left|\psi(Z)-\mathbb{M}\psi^{\sharp}(Z)\right|>\lambda R\right\}\leq 2C\exp\left(-c\lambda^2\right)$. It follows as in the proof of Theorem 3 (and is very standard) that we may replace $\mathbb{M}\psi^{\sharp}(Z)$ with $\mathbb{M}\psi(Z)=0$ by changing the constants involved. Now, remembering that R involves a λ, set $t=\lambda R$ and estimate λ in terms of t using Lemma 4. ■

When the X_i are not symmetric, and the equation $\mathbb{P}\left\{|X_i|>t\right\}=\exp\left(-t^q\right)$ is replaced with $\mathbb{P}\left\{|X_i|>t\right\}\leq C\exp\left(-ct^q\right)$, then one may use symmetrization and contraction arguments to reduce to the case of symmetric Weibull variables, see for example [1, Section 3.2.2].

4 Application: Random sections of convex bodies

We now study random sections of convex bodies generated by random embeddings $W:\mathbb{R}^k\to X$,

$$Wx=\sum_{i=1}^{n}\sum_{j=1}^{k}W_{i,j}x_je_i$$

where $\left(X,|\cdot|_X\right)$ is any real normed space of $\dim(X)\geq n$ and $(e_i)_1^n$ is a linearly independent sequence in X. When $\varepsilon\in(0,1/2)$, $n\geq n_0(k,\varepsilon)$, W is an $n\times k$ standard Gaussian random matrix and $(e_i)_1^n$ are appropriately chosen in X, then it is well known that the corresponding embedding W is, with high probability, close to being an isometry (with respect to the Euclidean norm on $\mathbb{R}^k$), i.e. for all $x \in \mathbb{R}^k$,

$$(1-\varepsilon) M' |x| \leq |Wx|_X \leq (1+\varepsilon) M' |x| \qquad\qquad M' = \mathbb{E}\left|\sum_{i=1}^{n} W_{i,1} e_i\right|_X$$

This is a Gaussian formulation of Dvoretzky's theorem, originally proved by Dvoretzky [9], re-formulated and re-proved using concentration on S^{n-1} by Milman [28], recovering the exact Dvoretzky dimension up to a universal constant (see also [29]), and a Gaussian formulation due to Pisier, see e.g. [33]. We refer the reader to Schechtman's survey article [35] for more information on the subject and more detailed quantitative bounds. We study the case where the matrix W has independent entries that do not necessarily follow the standard Gaussian distribution. Our main result in this direction, Theorem 7, recovers Milman's general Dvoretzky theorem involving the parameter M/b (see Section 4 for more details). When $X = \ell_p^n$ $(1 \leq p < \infty)$ and $(e_i)_1^n$ are the standard basis vectors, then the random section

$$B_X \cap range\,(W)$$

is isomorphic to an ellipsoid but not usually almost isometric, where

$$B_X = \{x \in X : |x|_X \leq 1\}$$

In the case of $X = \ell_\infty^n$ (with the standard basis) we see completely different behavior, and the random section is isomorphic to the floating body, see for example [2, 12, 36], which may be far from ellipsoidal. In the special case of the Gaussian distribution, of course, the floating body is a Euclidean ball.

The spirit here is not to find Euclidean subspaces of largest dimension and smallest distortion, but rather to understand the various subspaces that exist, and to study new populations of subspaces that escape the purview of uniformly distributed subspaces. Like Milman's proof in the case of uniformly distributed subspaces, the proof of Theorem 7 is based on concentration of measure and an epsilon-net argument. What is new is the use of the concentration of measure techniques discussed in Sections 1.1 and 1.2.

For any probability measure μ on $\mathbb{R}$ and associated cumulative distribution $F : \mathbb{R} \to [0,1]$ defined by $F(t) = \mu(-\infty, t]$, the generalized inverse $F^{-1} : (0,1) \to \mathbb{R}$, known as the quantile function, is defined by

$$F^{-1}(s) = \inf\{t \in \mathbb{R} : F(t) > s\} \tag{9}$$

Theorem 7 *Let $n, k \in \mathbb{N}$ and $T \in [2,\infty)$, and for each $1 \leq i \leq n$ and $1 \leq j \leq k$, let $\mu_{i,j}$ be a probability measure on $\mathbb{R}$ with cumulative distribution $F_{i,j}$. Let $F_{i,j}^{-1}$ be the generalized inverse of $F_{i,j}$ as defined by (9), and assume that each $F_{i,j}^{-1}$ is locally Lipschitz on $(0,1)$, and that each $\mu_{i,j}$ has finite first moment,*

$$\int_{-\infty}^{\infty} |x|\, d\mu_{i,j}(x) < \infty$$

Let W be an $n \times k$ random matrix with independent entries $(W_{i,j})$, where the distribution of $W_{i,j}$ is $\mu_{i,j}$, and assume that there is no single hyperplane in which the distribution of *every row* $(W_{i,j})_{j=1}^k$ *is supported. Let* G *be an* $n \times k$ *standard Gaussian random matrix. Let* $K \subset \mathbb{R}^n$ *be a convex body with* $0 \in int(K)$ *and set* $b = \sup\{|\theta|_K : \theta \in S^{n-1}\}$*. Consider the following norm on* $\mathbb{R}^{n\times k}$*,*

$$|A|_\sharp = \max_{1\le i\le n} \sup_{0\ne y\in\mathbb{R}^k} \frac{\left(\sum_{j=1}^k A_{i,j}^2 y_j^2\right)^{1/2}}{\mathbb{E}\,|Wy|_K}$$

Let $Q : \mathbb{R}^{n\times k} \to \mathbb{R}$ *be a measurable function such that for all* $A \in \mathbb{R}^{n\times k}$*,*

$$Q(A) \ge b \left|\left(Lip\left(F_{i,j}^{-1}\circ\Phi, A_{i,j}\right)\right)_{i,j}\right|_\sharp$$

where $Lip(\cdot,\cdot)$ *is defined in (3). Let* $\xi : [0,\infty) \to [0,\infty)$ *be a non-decreasing function such that for all* $t \ge 0$, $\mathbb{P}\{Q(G) > \xi(t)\} \le 2(1-\Phi(t))$*. Set*

$$\varepsilon = 8\mathfrak{L}(Q)T\xi(T) + 28\mathfrak{L}(Q)\sqrt{\int_2^\infty \xi(t)^2 t^3 \exp\left(-t^2/2\right) dt} \tag{10}$$

where $\mathfrak{L}(\cdot)$ *is defined in (2), and assume that both of the following conditions hold*

$$0 < \varepsilon \le 1/2 \qquad\qquad k \le (1/19)\left(\log \varepsilon^{-1}\right)^{-1} T^2 \tag{11}$$

Then with probability at least $1-\exp\left(-T^2/4\right)$ *the following event occurs: for all* $x \in \mathbb{R}^k$*,*

$$(1-\varepsilon)\,\mathbb{E}\,|Wx|_K \le |Wx|_K \le (1+\varepsilon)\,\mathbb{E}\,|Wx|_K$$

Comments for Theorem 7:

• **Interpretation:** The body $K^\flat = \left\{x \in \mathbb{R}^k : \mathbb{E}\,|Wx|_K \le 1\right\}$ is a compact convex set with $0 \in int(K^\flat)$, and the theorem can be interpreted geometrically as a deviation inequality for the random body $K \cap Range(W)$ (with appropriate coordinates) about the deterministic body $K^\flat$. When the entries of W are i.i.d. then the body $K^\flat$ is invariant under coordinate permutations. When the distribution of each entry is even (symmetric about 0), then $K^\flat$ is invariant under reflections about the coordinate axes (i.e. unconditional).

• **Bounding $\mathfrak{L}(Q)$:** Since $|\cdot|_\sharp$ is unconditional, Lemma 2 implies that typically (as with our other results) we may choose Q so that $\mathfrak{L}(Q) \le \sqrt{2}$.

• **$k \le n$:** While this is not explicitly assumed, the conclusion guarantees that $null(W) = \{0\}$. The case $k > n$ is indeed impossible because, since $\varepsilon \le 1/2$ (10) implies an upper bound for T and (11) then implies a bound on k.

• **Milman's general Dvoretzky theorem as a special case:** Suppose that K is centrally symmetric (i.e. $-K = K$) in which case $|\cdot|_K$ is a norm, and that W is a standard Gaussian random matrix. It follows that $\mathbb{E}\,|Wx|_K = (1-\delta_n)\sqrt{n}M\,|x|$, where $1-\delta_n = n^{-1/2}\mathbb{E}\,|Y_1|$ depends only on n (Y_1 denotes the first column of W), $\delta_n \to 0$ as $n \to \infty$,

$$M = \int_{S^{n-1}} |x|_K \, d\sigma_n(x)$$

and σ_n is normalized Haar measure on S^{n-1}. In particular

$$|A|_{\sharp} = (1-\delta_n)^{-1} n^{-1/2} M^{-1} \max_{i,j} |A_{i,j}|$$

and we may take $Q(A) = \xi(t) = 2n^{-1/2}M^{-1}b$ in which case $\mathfrak{L}(Q) = 1$. Assuming $n^{1/2}Mb^{-1} \geq C'$ (which we may without loss of generality by considering the probability bound below), choosing any $\varepsilon \in (0, 1/2)$ as an independent variable and setting

$$T \approx C\varepsilon n^{1/2} M b^{-1}$$

this recovers Milman's general Dvoretzky theorem with the sufficient condition $k \leq c\left(\log \varepsilon^{-1}\right)^{-1} \varepsilon^2 n M^2 b^{-2}$, probability $1 - C\exp\left(-c\varepsilon^2 n M^2 b^{-2}\right)$, and estimate

$$(1-\varepsilon) M |x| \leq \frac{1}{\sqrt{n}} |Wx|_K \leq (1+\varepsilon) M |x|$$

Here we have used the fact that

$$0 \leq \delta_n \leq Cn^{-1/2} \leq CbM^{-1}n^{-1/2} \leq \varepsilon/10$$

otherwise the probability bound becomes trivial, so we may ignore δ_n. Such a bound for k recovers the correct 'Dvoretzky dimension' nM^2b^{-2} and is optimal in a certain sense for fixed ε, say $\varepsilon = 1/4$, up to a universal constant, see [29]. It also gives the original dependence on ε until it was improved by Schechtman [34] to $k \leq c\varepsilon^2 nM^2b^{-2}$, see also the paper by Gordon [19, Theorem 7] that gives Euclidean subspaces of dimension $c\varepsilon^2 \log n$. It is known that using an affine map we may always place K in John's position, where the ellipsoid of maximal volume in K is B_2^k, and in this case $M \geq c\sqrt{n^{-1}\log n}$ and $b = 1$. In the language of functional analysis, this implies that real Hilbert space is finitely representable in any infinite dimensional real Banach space (also known over $\mathbb{C}$).

• **Non-Gaussian sections of B_p^n $(1 \leq p < \infty)$ are isomorphic but not usually almost-isometric to $n^{-1/p}B_2^k$:** When $K = B_p^n$ $(1 \leq p < \infty)$ and each $\mu_{i,j} = \mu$ for some probability measure μ with mean zero, variance one and log-concave density, and we denote the rows of W by $(X_i)_1^n$, then for each $\theta \in S^{n-1}$, $\langle \theta, X_i \rangle$ has a log-concave distribution with mean zero and variance one, and by Jensen's inequality

$$\mathbb{E}\,|Wx|_p = \mathbb{E}\left(\sum_{i=1}^{n} |\langle x, X_i\rangle|^p\right)^{1/p} \leq Cpn^{1/p}|x|$$

on the other hand

$$\mathbb{E}\,|Wx|_p \geq \frac{1}{2}\mathbb{M}\left(\sum_{i=1}^{n} |\langle x, X_i\rangle|^p\right)^{1/p} \geq Cn^{1/p}|x|$$

Therefore a random section of B_p^n using a matrix with i.i.d. mean zero variance one log-concave entries is isomorphic to $n^{-1/p}B_2^k$ with distortion depending on p (but usually not almost isometric).

In the case where each $W_{i,j}$ has a two sided exponential distribution with density $2^{-1}e^{-|t|}$, for example, $Lip\left(F_{i,j}^{-1}\Phi,t\right) \leq C|t|$, $|A|_\sharp \leq C_p n^{-1/p}\max_{i,j}|A_{i,j}|$, set $Q(A) = C_p b n^{-1/p}\max_{i,j}|A_{i,j}|$, $\varepsilon = 1/4$, and $\xi(t) = C_p b n^{-1/p}\left((\log n)^{1/2}+t\right)$. Setting $T \approx c_p b^{-1/2}n^{1/(2p)}$, we may take

$$k = c_p b^{-1} n^{1/p} = \begin{cases} c_p n^{1/2} & : \quad 1 \leq p \leq 2 \\ c_p n^{1/p} & : \quad 2 \leq p < \infty \end{cases}$$

and the corresponding section of B_p^n is (with high probability) isomorphic to B_2^k with distortion C_p.

• **Non-Gaussian sections of B_∞^n are not necessarily isomorphic to B_2^k, but instead are almost-isometric to the floating body:** The case $p = \infty$ is quite different to the case $1 \leq p < \infty$. Note that the body $\left\{x \in \mathbb{R}^k : \mathbb{E}|Wx|_\infty \leq 1\right\}$ is the polar (dual) of the (symmetrized) expected convex hull

$$\mathbb{E}conv\left\{\pm X_i\right\}_1^n = \left\{x \in \mathbb{R}^k : \forall\theta \in S^{k-1}, \langle x,\theta\rangle \leq \mathbb{E}\max_{1\leq i\leq n}|\langle X_i,\theta\rangle|\right\}$$

The bound

$$(1-\varepsilon)\,\mathbb{E}|Wx|_\infty \leq |Wx|_\infty \leq (1+\varepsilon)\,\mathbb{E}|Wx|_\infty$$

valid for all $x \in \mathbb{R}^k$, implies that

$$(1-\varepsilon)\mathbb{E}conv\,(X_i)_1^n \subseteq conv\,(X_i)_1^n \subseteq (1+\varepsilon)\mathbb{E}conv\,(X_i)_1^n$$

Results of this type (i.e. concentration of the convex hull within the space of convex bodies) are multivariate generalizations of Gnedenko's law of large numbers on the maximum and minimum of a random sample, since for a compact set $E \subset \mathbb{R}^1$, $conv\,(E) = [\inf E, \sup E]$. Such multivariate extensions were studied in [12, 14] (under the assumption that the rows are i.i.d. and have a log-concave distribution, but individual coordinates need not be independent), and other variations in [8, 10, 11, 17, 18, 23, 27] (we refer the reader to the introduction in [14] for a more detailed discussion). When all rows share a common distribution (say ν) that is rotationally invariant (in our case this corresponds to the standard Gaussian matrix), then $\mathbb{E}conv\,(X_i)_1^n$ is a Euclidean ball, but for most other distributions $\mathbb{E}conv\,(X_i)_1^n$ is far from Euclidean. When ν is log-concave and not contained in any affine hyperplane, the expected convex hull is similar to the floating body $F_{1/n}$ (which is what remains of $\mathbb{R}^k$ after all half-spaces of ν-measure less than $1/n$ have been deleted, see e.g. [2, 12, 36]) and in this case it was shown in [14] that for $n \geq 4$,

$$\left(1-\frac{C}{\log n}\right)F_{1/n} \subseteq \mathbb{E}conv\,(X_i)_1^n \subseteq \left(1+\frac{C}{\log n}\right)F_{1/n}$$

Non-Gaussian embeddings into ℓ_∞^n were also studied by Gordon, Litvak, Pajor, and Tomczak-Jaegermann [20]: given a symmetric convex body $K \subset \mathbb{R}^k$, one chooses the rows of W to be i.i.d. from the dual unit ball, and then the corresponding embedding W embeds $\left(\mathbb{R}^k, |\cdot|_K\right) \hookrightarrow_{1+\varepsilon} \ell_\infty^n$. In the case $K = B_p^n$ $(1 \leq p \leq \infty)$, the entries of such a matrix are almost i.i.d. by the volume representation of B_q^n.

Proof of Theorem 7. By a smoothing argument we may assume without loss of generality that each $\mu_{i,j}$ has a C^∞ density function $f_{i,j} = d\mu_{i,j}/dx$. Since this is a standard procedure in analysis and we wish to simplify notation, we won't work with the smoothed measure and then take a limit at the end of the proof, but will rather just assume that $f_{i,j}$ exists and is C^∞ from the start. The Knöthe-Rosenblatt map $T : \mathbb{R}^{n\times k} \to \mathbb{R}^{n\times k}$ takes the form $(T(A))_{i,j} = F_{i,j}^{-1}(\Phi(A_{i,j}))$. By the triangle inequality it follows that $|\cdot|_K$ is b-Lipschitz on $\mathbb{R}^n$. Let $\sigma > 0$ and let $g^{(\sigma)} = |\cdot|_K * \varphi_\sigma$, where $\varphi_\sigma(x) = (2\pi\sigma)^{-n/2}\exp\left(-|\sigma^{-1}x|^2/2\right)$, in which case $g^{(\sigma)}$ is C^∞ and $\left|\nabla g^{(\sigma)}(x)\right| \leq b$ for all $x \in \mathbb{R}^n$. The set $K^\flat = \left\{x \in \mathbb{R}^k : \mathbb{E}|Wx|_K \leq 1\right\}$ is seen to be compact and convex, with $0 \in int(K)$. Consider any (momentarily fixed) $\theta \in \partial K^\flat$, and define $\psi_{\sigma,\theta} : \mathbb{R}^{n\times k} \to \mathbb{R}$ by

$$\psi_{\sigma,\theta}(A) = g^{(\sigma)}((TA)\theta)$$

where $y \longmapsto (TA)y$ denotes the standard action of $\mathbb{R}^{n\times k}$ on $\mathbb{R}^k$. A direct calculation shows that

$$\begin{aligned}
\left|\nabla\psi_{\sigma,\theta}(A)\right| &\leq \left(\sum_{i=1}^n\sum_{j=1}^k\left[\frac{\phi(A_{i,j})\theta_j}{f_{i,j}F_{i,j}^{-1}\Phi(A_{i,j})}g_i^{(\sigma)}\left(\left(\sum_{w=1}^k F_{u,w}^{-1}\Phi(A_{u,w})\theta_w\right)_{u=1}^n\right)\right]^2\right)^{1/2} \\
&= \left(\sum_{i=1}^n\left[g_i^{(\sigma)}\left(\left(\sum_{w=1}^k F_{u,w}^{-1}\Phi(A_{u,w})\theta_w\right)_{u=1}^n\right)^2\sum_{j=1}^k\left(\frac{\phi(A_{i,j})\theta_j}{f_{i,j}F_{i,j}^{-1}\Phi(A_{i,j})}\right)^2\right]\right)^{1/2} \\
&\leq b\max_{1\leq i\leq n}\left(\sum_{j=1}^k\left(\frac{\phi(A_{i,j})\theta_j}{f_{i,j}F_{i,j}^{-1}\Phi(A_{i,j})}\right)^2\right)^{1/2} \qquad (12)
\end{aligned}$$

By definition of $|\cdot|_\sharp$ and Q, this is bounded above by

$$b\left|\left(Lip\left(F_{i,j}^{-1}\circ\Phi, A_{i,j}\right)\right)_{i,j}\right|_\sharp \mathbb{E}|W\theta|_K \leq Q(A)$$

By Theorem 3 with $A = 2$ it follows that for all $t \geq 2 > \Phi^{-1}(7/8)$, with probability at least $1 - \exp(-t^2/2)$, $\left|\psi_{\theta,\sigma}(G) - \mathbb{M}\psi_{\theta,\sigma}(G)\right| \leq 2\mathfrak{L}(Q)t\xi(t)$. By sending $\sigma \to 0$, we get (with the same probability), $||W\theta|_K - \mathbb{M}|W\theta|_K| \leq 2\mathfrak{L}(Q)t\xi(t)$. Comparing the mean and median,

$$|\mathbb{E}|W\theta|_K - \mathbb{M}|W\theta|_K| \leq \sqrt{2\mathrm{Var}|W\theta|_K} \leq \sqrt{2\mathbb{E}||W\theta|_K - \mathbb{M}|W\theta|_K|^2} \qquad (13)$$

By our bound on the distribution of $||W\theta|_K - \mathbb{M}|W\theta|_K|$,

$$\mathbb{E}||W\theta|_K - \mathbb{M}|W\theta|_K|^2 \leq 16\mathfrak{L}(Q)^2\xi(2)^2 + \int_2^\infty 4\mathfrak{L}(Q)^2t^2\xi(t)^2t\exp\left(-t^2/2\right)dt$$

Since ξ is non-decreasing,

$$\begin{aligned}
\xi(2)^2 &\leq \left(\int_2^\infty t^3\exp\left(-t^2/2\right)dt\right)^{-1}\int_2^\infty t^3\xi(t)^2\exp\left(-t^2/2\right)dt \\
&\leq (1.236)\int_2^\infty t^3\xi(t)^2\exp\left(-t^2/2\right)dt
\end{aligned}$$

and

$$\mathbb{E}\left||W\theta|_K - \mathbb{M}\,|W\theta|_K\right|^2 \leq 24\mathfrak{L}(Q)^2 \int_2^\infty t^3 \xi(t)^2 \exp\left(-t^2/2\right) dt$$

Using the triangle inequality, (13) and the assumptions of the theorem, we see that with probability at least $1 - \exp(-T^2/2)$, $\left||W\theta|_K - \mathbb{E}\,|W\theta|_K\right| \leq \varepsilon/4$, where

$$\varepsilon = 8\mathfrak{L}(Q)T\xi(T) + 28\mathfrak{L}(Q)\sqrt{\int_2^\infty \xi(t)^2 t^3 \exp(-t^2/2)\,dt}$$

Recall that $\theta \in \partial K^\flat$ was momentarily fixed (but arbitrary). We now apply the standard epsilon-net argument to achieve a uniform bound over all $\partial K^\flat$. Let $\mathcal{N} \subset \partial K^\flat$ be an $\varepsilon/4$-net with respect to the distance function $\rho(x, y) = |y - x|_{K^\flat}$, where $|x|_{K^\flat} = \mathbb{E}\,|Wx|_K$. By Lemma 5.2 in [14], which is a modification of Lemmas 4.10 and 4.11 in [32] for centrally symmetric bodies, one can choose $\mathcal{N}$ so that $|\mathcal{N}| \leq (12/\varepsilon)^k$, and for each $x \in \partial K^\flat$ one has the series expansion $x = \omega_0 + \sum_1^\infty \varepsilon_i \omega_i$, with $0 \leq \varepsilon_i < (\varepsilon/4)^i$ and $(\omega_i)_0^\infty \subseteq \mathcal{N}$. With probability at least $1 - (12/\varepsilon)^k\, 4\exp(-T^2/2)$, for all $\omega \in \mathcal{N}$, $\left||W\omega|_K - \mathbb{E}\,|W\omega|_K\right| \leq (\varepsilon/4)\,\mathbb{E}\,|W\omega|_K$ and by the series expansion and the triangle inequality it follows that for all $x \in \partial K^\flat$,

$$|Wx|_K \leq |W\omega_0|_K + \sum_{i=1}^\infty \varepsilon_i\,|W\omega_i|_K \leq (1+\varepsilon)\,\mathbb{E}\,|W\omega_0|_K$$

and

$$|Wx|_K \geq |W\omega_0|_K - \sum_{i=1}^\infty \varepsilon_i\,|W\omega_i|_K \geq (1-\varepsilon)\,\mathbb{E}\,|W\omega_0|_K$$

The result now follows by positive homogeneity of $|x|_{K^\flat}$. ∎

Acknowledgements

Thanks to Gusti van Zyl for various suggestions and for pointing out a typing error. Part of this work was done while the author was a postdoctoral fellow at the Weizmann Institute of Science.

References

[1] Adamczak, R., Litvak, A. E., Pajor, A., Tomczak-Jaegermann, N.: Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling. Constr. Approx. 34 (1), 61-88 (2011)

[2] Bobkov, S. G.: Convex bodies and norms associated to convex measures. Probab. Theory Related Fields 147 (1-2), 303-332 (2010)

[3] Bobkov, S. G., Nayar, P., Tetali, P.: Concentration properties of restricted measures with applications to non-Lipschitz functions. Geometric aspects of functional analysis, 25-53, Lecture Notes in Math. 2169, Springer, Cham, 2017.

[4] Borell, C.: The Brunn-Minkowski inequality in Gauss space. Invent Math. 30 (2), 207-216 (1975)

[5] Boucheron, S., Lugosi, G., Massart, P.: Concentration Inequalities, A nonasymptotic theory of independence. Clarendon press, Oxford 2012.

[6] Chatterjee, S.: Superconcentration and related topics. Springer Monographs in Mathematics. (2014)

[7] Cordero-Erausquin, D., Ledoux, M.: Hypercontractive measures, Talagrand's inequality, and influences. Geometric aspects of functional analysis, 169-189, Lecture Notes in Math., 2050, Springer (2012)

[8] Davis, R., Mulrow, E., Resnick, S.: Almost sure limit sets of random samples in $\mathbb{R}^d$. Adv. in Appl. Probab. 20 (3), 573-599 (1988)

[9] Dvoretzky, A.: Some results on convex bodies and Banach spaces. 1961 Proc. Internat. Sympos. Linear Spaces (Jerusalem, 1960) pp. 123–160 Jerusalem Academic Press, Jerusalem; Pergamon, Oxford.

[10] Fisher, L. D.: The convex hull of a sample. Bull. Amer. Math. Soc. 72, 555-558 (1966)

[11] Fisher, L. D.: Limiting sets and convex hulls of samples from product measures. Ann. Math. Statist. 40, 1824-1832 (1969)

[12] Fresen, D.: A multivariate Gnedenko law of large numbers. Ann. Probab. 41 (5), 3051-3080 (2013)

[13] Fresen, D.: Dvoretzky's theorem, from the ground up. Unpublished lecture notes distributed at Yale in 2012 and 2014.

[14] Fresen, D., Vitale, R.: Concentration of random polytopes around the expected convex hull. Electron. Commun. Probab. 19 (59), 1-8 (2014)

[15] Fresen, D.: Variations and extensions of the Gaussian concentration inequality, Part II.

[16] Fresen, D.: Random Euclidean embeddings in finite dimensional Lorentz sequence spaces. Preprint.

[17] Geffroy, J.: Localisation asymptotique du polyèdre d'appui d'une échantillon Laplacien à k dimensions. Publ. Inst. Statist. Univ. Paris 10, 213-228 (1961)

[18] Goodman, V.: Characteristics of normal samples. Ann. Probab. 16, 1281-1290 (1988)

[19] Gordon, Y.: Gaussian processes and almost spherical sections of convex bodies. Ann. Probab. 16 (1), 180-188 (1988)

[20] Gordon, Y., Litvak, A. E., Pajor, A., Tomczak-Jaegermann, N.: Random ε-nets and embeddings in ℓ_∞^N. Studia Math. 178 (1), 91-98 (2007)

[21] Grable, D. A.: A large deviation inequality for functions of independent , multi-way choices. Combin. Probab. Comput. 7 (1), 57-63 (1998)

[22] Hitczenko, P., Montgomery-Smith, S. J., Oleszkiewicz, K.: Moment inequalities for sums of certain independent symmetric random variables. Studia Math. 123 (1), 15-42 (1997)

[23] Kinoshita, K., Resnick, S.: Convergence of scaled random samples in $\mathbb{R}^d$. Ann. Probab. 19 (4), 1640-1663 (1991)

[24] Klartag, B., Kolesnikov, A.: Eigenvalue distribution of optimal transportation. Anal. PDE 8 (1), 33-55 (2015)

[25] Klartag, B., Sodin, S.: Variations on the Berry-Esseen theorem. Theory Probab. Appl. 56 (3), 403-419 (2012)

[26] Ledoux, M., Talagrand, M.: Probability in Banach Spaces. Springer-Verlag 1991.

[27] McBeth, D., Resnick, S.: Stability of random sets generated by multivariate samples. Comm. Statist. Stochastic Models 10 (3), 549-574 (1994)

[28] Milman, V. D.: A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies. (Russian) Funkcional. Anal. i Priložen. 5 (4), 28-37 (1971)

[29] Milman, V. D., Schechtman, G.: Global versus local asymptotic theories of finite-dimensional normed spaces, Duke Math. Journal 90 (1), 73-93 (1997)

[30] Naor, A.: The surface measure and cone measure on the sphere of ℓ_p^n. Trans. Amer. Math. Soc. 359 (3), 1045-1079 (2007)

[31] Paouris, G., Valettas, P., Zinn, J.: Random Version of Dvoretzky's theorem in ℓ_p^n. Stochastic Process. Appl. 127 (10), 3187-3227 (2017)

[32] Pisier, G.: The Volume of Convex Bodies and Banach Space Geometry. Cambridge Tracts in Mathematics 94, Cambridge University Press 1989.

[33] Pisier, G.: Probabilistic methods in the geometry of Banach spaces. CIME, Varenna, 1985. Lecture Notes in Mathematics 1206, 167-241 (1986)

[34] Schechtman, G.: A remark concerning the dependence on ϵ in Dvoretzky's theorem. Geometric aspects of functional analysis (1987-88), 274-277, Lecture Notes in Math., 1376, Springer, Berlin, 1989.

[35] Schechtman, G.: Euclidean sections of convex bodies. Asymptotic geometric analysis, 271-288, Fields Inst. Commun., 68, Springer, New York, 2013.

[36] Schütt, C., Werner, E.: The convex floating body. Math. Scand. 66, 275-290 (1990)

[37] Sudakov, V. N., Cirel B. S. (Tsirelson): Extremal properties of half-spaces for spherically invariant measures. (Russian) Problems in the theory of probability distributions, II. Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) 41, 14-24 (1974)

[38] Villani, C.: Optimal transport Old and new. Grundlehren der Mathematischen Wissenschaften 338. Springer-Verlag, Berlin 2009.

[39] Vu, V. H.: Concentration of non-Lipschitz functions and applications. Probabilistic methods in combinatorial optimization. Random Structures Algorithms. 20 (3), 262-316 (2002)