Variations and extensions of the Gaussian concentration inequality, Part I

Daniel J. Fresen[*]

Abstract

The classical Gaussian concentration inequality for Lipschitz functions is adapted to a setting where the classical assumptions (i.e. Lipschitz and Gaussian) are not met. The theory is more direct than much of the existing theory designed to handle related generalizations. An application is presented to linear combinations of heavy tailed random variables.

1 Introduction

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

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

where $C, c > 0$ are universal constants and $Lip(\psi)$ is the Lipschitz constant of ψ. $\mathbb{M}\psi(Z)$ denotes the median of $\psi(Z)$, and can be replaced with the mean $\mathbb{E}\psi(Z)$. It follows from the Gaussian isoperimetric inequality of [22] and [5] that this can be improved to

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

where Φ is the standard normal cumulative distribution. Equality clearly holds when ψ is linear.

Assuming for simplicity that ψ is C^1, it follows from a result of Pisier [21, Theorem 2.2 p176] that if Y is another random vector in $\mathbb{R}^n$ with the standard normal distribution, independent of Z, then for any convex function $\varphi : \mathbb{R} \to \mathbb{R}$,

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

[*]University of Pretoria, Department of Mathematics and Applied Mathematics, daniel.fresen@up.ac.za.

where Z' has the standard normal distribution in $\mathbb{R}$ and is independent of Z. This often implies that one can compare the tails of $\psi(Z) - \psi(Y)$ to that of $\frac{\pi}{2}|\nabla\psi(Z)|\,Z'$, which in turn leads to a bound on the tails of $\psi(Z) - \mathbb{M}\psi(Z)$.

Pisier's version of Gaussian concentration has three advantages over the classical version. The first is that the proof (with input from Maurey) is quite simple. The second is that the Lipschitz condition has been removed. The third is that the resulting concentration inequality is determined by the distribution of $|\nabla\psi(Z)|$ rather than $Lip(\psi) = \sup|\nabla\psi(\cdot)|$. This is really the way things should be, considering that the Lipschitz constant of a function might be determined by behaviour on sets of small measure that have little effect on concentration properties of $\psi(Z)$. This can be seen in the examples

$$u(x) = |x| + \max\left\{0, \varepsilon - \frac{|x|}{\varepsilon}\right\} \qquad\qquad v(x) = |x| + \frac{\varepsilon\sqrt{|x|}}{1 + \sqrt{|x|}}$$

for $x \in \mathbb{R}^n$ and $\varepsilon > 0$.

In two papers, 'Variations and extensions of the Gaussian concentration inequality' Part I (here) and Part II (to appear elsewhere), we study ways in which to apply the above inequalities in settings where they have not previously been applied because they appear to be inapplicable. The result is that the Gaussian concentration inequality gives rise to various 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 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 [6, 16] for information on the concentration of measure phenomenon.

In Section 2 we present the main ideas of the paper. In Section 3, as an illustration of these ideas, we prove a concentration inequality for linear combinations of heavy tailed random variables: if $0 < q \leq 1$ and $(X_i)_1^n$ is a sequence of independent symmetric random variables with $\mathbb{P}\{|X_i| > t\} = \exp(-t^q)$ for all $t > 0$, then

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

(to say that X_i is symmetric means that X_i and $-X_i$ have the same distribution). This 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 $0 < q \leq 1$. Considering that

$$C_1^{1/q}q^{-2/q}|a|^2 \leq \mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) \leq C_2^{1/q}q^{-2/q}|a|^2$$

we see that the coefficient $C^{-1/q}q^{2/q}$ is sharp up to the value of C. By applying symmetrization and contraction techniques it can be applied in the non-symmetric case as

long as $\mathbb{E}X_i = 0$ and $\mathbb{P}\{|X_i| > t\} \le C \exp\left(-t^q\right)$, with $C^{-1/q}q^{2/q}$ replaced by a different quantity. This follows work of Hitczenko, Montgomery-Smith and Oleszkiewicz [12] who estimated the pth moment of $|\sum a_i X_i|$. The authors of [12] presented a significantly weaker result:

$$\lim_{t \to \infty} \log_t \ln \left[\mathbb{P}\left\{\left|\sum_{i=1}^n a_i X_i\right| > t\right\}\right]^{-1} = q.$$

See Cor. 6.5 in their paper. The case of equal coefficients, which avoids the main obstacle, is considered in [3], see their Eq. (3.6). A similar inequality appears in [14] (see their Sec. 4) but does not include the sub-Gaussian part. The classical case $q \ge 1$ is covered in [6] (their Ex. 2.27 p.50) but we did not see any mention of the case $q \in (0, 1)$. Of course in the vast array of techniques available there are undoubtedly several ways to prove (2), we present two, and the result may be folklore. More applications, mostly non-linear, are given in [8, 9] and will appear in future papers.

Comments and notation

$\mathbb{M}$ denotes median, and C, c, C' etc. usually denote positive universal constants that may represent different values at each appearance. Dependence on variables will usually be indicated by subscripts, C_q, c_q etc. The Euclidean norm of a vector $a \in \mathbb{R}^n$ is denoted $|a|$ while $|a|_p = \left(\sum_1^n |a_i|^p\right)^{1/p}$ is its ℓ_p^n-norm ($1 \le p < \infty$), and $|a|_\infty = \max\{|a_i|\}$. Set-builder notation such as $\{x \in \mathbb{R}^n : \psi(x) \le R\}$ will at times be simplified to $\{\psi \le R\}$. The term 'random variable' will be used exclusively for real valued random variables. Let γ_n denote the standard Gaussian measure on $\mathbb{R}^n$ with density

$$\frac{d\gamma_n}{dx} = (2\pi)^{-n/2} \exp\left(-|x|^2/2\right).$$

Let Φ denote the standard normal cumulative distribution and $\phi = \Phi'$ the standard normal density. The expression $Z \sim N(0, I_n)$ means that the random vector Z follows the standard normal distribution on $\mathbb{R}^n$ (with mean 0 and identity covariance matrix I_n). The 'Gaussian' in 'Gaussian concentration' refers to Z, not $\psi(Z)$, and includes the case where $\psi(Z)$ has thicker-than-Gaussian tails, which may occur when ψ grows more rapidly than linear. The C^1 condition that appears from time to time is stronger than necessary; as long as f is locally Lipschitz it will be differentiable almost everywhere, and we will often disregard certain points of non-differentiability without explicitly commenting on the matter. The sigma algebra on $\mathbb{R}^n$ is always assumed to be the Borel sigma algebra, so *measurable* means *Borel measurable*. The following lemma will be used implicitly.

Lemma 1 *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

2　The methodology

2.1　Gaussian concentration for non-Lipschitz functions

In this Section we discuss an alternative technique for handling the case when ψ is non-Lipschitz, or when the Lipschitz constant of ψ is much larger than the average local Lipschitz constant, so (1) either does not apply or becomes crude. The technique is outlined as follows: modify ψ on a set of small probability on which it behaves badly so that the modified function ψ^* is Lipschitz, apply classical Gaussian concentration to ψ^*, and then transfer the result for ψ^* back to ψ. We now describe this in more detail.

Consider the setting where Z has the standard normal distribution in $\mathbb{R}^n$ and $\psi : \mathbb{R}^n \to \mathbb{R}$ is, say, C^1, but is not assumed to be Lipschitz. Restrict ψ to some set E such that $Z \in E$ with high probability and $|\nabla \psi|$ is bounded nicely on E. The set $E = \{x \in \mathbb{R}^n : |\nabla \psi(x)| \leq R\}$ is often a natural choice (with appropriate $R > 0$). Assuming that E is convex (we discuss the non-convex case next), we have $Lip(\psi|_E) \leq \sup_E |\nabla \psi|$. One can then extend the restriction $\psi|_E$ to the entire space $\mathbb{R}^n$ so that the extension ψ^* obeys $Lip(\psi^*) = Lip(\psi|_E) \leq R$. By applying classical Gaussian concentration of Lipschitz functions as in (1) to ψ^* and observing that $\mathbb{P}\{\psi(Z) = \psi^*(Z)\} \geq \mathbb{P}\{Z \in E\}$, we may transfer the concentration inequality for $\psi^*(Z)$ about $\mathbb{M}\psi^*(Z)$ to an inequality for $\psi(Z)$ about $\mathbb{M}\psi(Z)$.

The idea of proving concentration inequalities by modifying a function on a set where it behaves badly is not new. The technique based on restriction involving $|\nabla \psi|$ and Lipschitz extension is found in [1, Lemma 2.2], [2, Section 3], [4, Section 6] and [19, Section 5], and is related to an observation contained in unpublished lecture notes that we prepared and distributed at Yale in 2012 and 2014. In these papers one assumes that E is convex, or at least that $Lip(\psi|_E) \leq \sup_E |\nabla \psi|$ (which is not necessarily the case when E is not convex). In [11, Corollary] and [28, Section 3] a related procedure is also used, but in a slightly different setting 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.

2.2　When $E = \{|\nabla \psi| \leq R\}$ is not convex

Without convexity one cannot apply the bound $Lip(\psi|_E) \leq \sup_E |\nabla \psi|$. This is seen, for example, by considering a set in the shape of the letter C, where the two ends of the C are close together so that the C looks like an O, and a function which increases gradually as we move from one end of the C to the other. We consider a parameter $\mathfrak{L}(E)$, which measures the distance one has to travel inside E to get from a point x to a point y as compared to the Euclidean distance, this ratio maximized over all pairs of distinct points in E. In Proposition 5 we show that $Lip(\psi|_E) \leq \mathfrak{L}(E) \sup_E |\nabla \psi|$ and that $\mathfrak{L}(E)$ is the correct coefficient. But this is not enough: we must provide useful bounds on $\mathfrak{L}(E)$, which we do in Proposition 6. We show that whenever E is a certain non-affine deformation of an unconditional convex body (as the inverse image under a continuous coordinate-wise transformation), then $\mathfrak{L}(E) \leq \sqrt{2}$.

Another approach is to find a convex subset $K \subseteq E$ and then apply the bound $Lip(\psi|_K) \leq \sup_K |\nabla \psi| \leq \sup_E |\nabla \psi|$. Ideally one would want the Gaussian measure

of K to be comparable to the Gaussian measure of E, in the sense that

$$\gamma_n \left(\mathbb{R}^n \setminus K \right) \le C\gamma_n \left(\mathbb{R}^n \setminus C^{-1}E \right)$$

for a universal constant $C > 0$. In Section 4 we present some very basic observations in this direction, motivated by the methodology in Sections 2.1 and 2.3 and of independent interest. We refer the reader to [24, 25] for a deeper discussion of the topic of approximating star bodies by convex sets, including various open problems.

2.3 Non-Gaussian concentration

If X is a random vector in $\mathbb{R}^n$ with any distribution μ, then X has the same distribution as $T(Z)$ for some measurable function $T : \mathbb{R}^n \to \mathbb{R}^n$, where Z is a random vector with the standard normal distribution on $\mathbb{R}^n$, denoted γ_n. One such map is the Knöthe-Rosenblatt rearrangement, see for example [27], and in some cases one can write down an explicit formula for T. For our purposes we may assume that $X = T(Z)$, so $f(X) = (f \circ T)(Z)$. Under fairly general conditions we may then apply Gaussian concentration to $\psi = f \circ T$ to obtain a concentration inequality for $f(X)$. Let us reiterate this in words: *any function of any random vector has the same distribution as a function of a Gaussian random vector.*

Transportation methods, and a variety of them, are quite standard in the theory of concentration of measure, see for example [6, 7, 10, 16, 23, 26]. The above procedure in particular is alluded to in [20, p. 1046] in the context where μ is a Lipschitz image of the standard Gaussian measure, in anticipation of applying the classical Gaussian concentration inequality for Lipschitz functions. Note that if one can apply Gaussian concentration to a wider class of functions ψ as in Section 2.1 above, then one may apply Gaussian concentration to a wider class of non-Gaussian measures. The observations in Section 2.1 and those of this section therefore work particularly well *together* and each 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 [13]) may be better, although the Brenier map is best for minimizing a transportation cost, which is not exactly what we want.

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. $\mathbb{E}\exp\left(\varepsilon X_i\right) < \infty$ for some $\varepsilon > 0$. In this context, the exponential moment method plays an essential role: Using

Markov's inequality and independence,

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

$$
= \exp(-\lambda t)\prod_{i=1}^{n} \mathbb{E}\exp(\lambda a_i X_i).
$$

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 and linearity), one would estimate power moments and use Markov's inequality,

$$
\mathbb{P}\left\{\left|\sum_{i=1}^{n} a_i X_i\right| > t\right\} = \mathbb{P}\left\{\left|\sum_{i=1}^{n} a_i X_i\right|^p > t^p\right\} \leq t^{-p}\,\mathbb{E}\left|\sum_{i=1}^{n} a_i X_i\right|^p. \tag{3}
$$

Power moments of sums are more difficult to compute than exponential moments because power functions lack the critical property of the exponential function as used in the exponential moment method. 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 in [12, Theorem 1.1] that for all $p \geq 2$,

$$
\left(\mathbb{E}\left|\sum_{i=1}^{n} a_i X_i\right|^p\right)^{1/p} \leq C\left(\left(\sum_{i=1}^{n} |a_i|^p\, \mathbb{E}\,|X_i|^p\right)^{1/p} + \sqrt{p}\left(\sum_{i=1}^{n} |a_i|^2\, \mathbb{E}\,|X_i|^2\right)^{1/2}\right) \tag{4}
$$

where $C > 0$ is a universal constant, with a corresponding lower bound with C replaced by a different constant $c > 0$. General upper and lower bounds for the pth moment of a sum of independent random variables were later presented in [15], reducing the problem of computing these moments to that of evaluating a certain Orlicz norm type expression.

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 $0 < q \leq 1$, we show below that in (4), $|a|_p$ can be replaced with $|a|_\infty$ at the cost of $C^{1/q}$, i.e.

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

which by Markov's inequality leads to the tail estimate

$$
\mathbb{P}\left\{\left|\sum_{i=1}^{n} a_i X_i\right| > t\right\} \leq 2\exp\left(-c_q \min\left\{\left(\frac{t}{|a|}\right)^2, \left(\frac{t}{|a|_\infty}\right)^q\right\}\right) \tag{6}
$$

where $c_q > 0$ is made explicit in Theorem 3. This is more subtle than observing that the ℓ_p^n norm is equivalent to the ℓ_∞^n norm for $p > c\ln n$ and comes down to the geometry of Minkowski sums of the form $uB_{\ell_1} + vB_{\ell_2}$ and the question of when they contain B_{ℓ_r} for any given $1 < r < 2$ (n does not play a role here).

6

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

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

so by independence and identical distributions,

$$\frac{1}{2}\mathbb{P}\{X_1 > t + a\} \leq \mathbb{P}\{X_1 - X_1' > t\} \leq \mathbb{P}\{|X_1| > t/2\}. \tag{7}$$

The significance is that $X_1 - X_1'$ is symmetric. This can be combined with the following contraction principle, see [17, Lemma 4.6] for a more general version:

Lemma 2 *Let $\varphi : [0, \infty) \to [0, \infty)$ be a convex function, $K_1 \geq 1$, $K_2 > 0$, and let $(X_i)_1^n$ and $(Y_i)_1^n$ each be i.i.d. sequences of symmetric random variables with $\mathbb{P}\{|X_i| > t\} \leq K_1\mathbb{P}\{K_2|Y_i| > t\}$ for all i and all $t > 0$. Then for all $a \in \mathbb{R}^n$,*

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

If we are given an i.i.d. sequence of random variables $(X_i)_1^n$ that satisfy a tail bound such as $\mathbb{P}\{|X_i| > t\} \leq h(t)$, 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 2, convert this to a bound on $\mathbb{P}\{\left|\sum_{i=1}^n a_i(X_i - X_i')\right| > t\}$ (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 $\mathbb{P}\{\left|\sum_{i=1}^n a_i X_i\right| > t\}$ using (7). In this way, estimates such as (6) may be extended to the case of tail bounds such as $\mathbb{P}\{|X_i| > t\} \leq C\exp(-t^q)$.

Proof of (5) and (6). By Hölder's inequality and optimization, for all $p > 2$ and $s \in (0, \infty)$,

$$|a|_p \leq |a|_\infty^{1-2/p}|a|^{2/p} \leq \omega(s)|a|_\infty + s|a|$$

where

$$\omega(s) = \begin{cases} (2/p)^{2/(p-2)}(1 - 2/p)s^{-2/(p-2)} & : \quad 0 < s \leq 2/p \\ 1 - s & : \quad 2/p < s \leq 1 \\ 0 & : \quad s > 1 \end{cases}.$$

To see this, assume momentarily and without loss of generaility that $|a| = 1$ and optimize $(r^{1-2/p} - s)/r$ over $r \in [0, 1]$. Setting $s = 2p^{1/2-1/q}$,

$$|a|_p \leq \begin{cases} (1 - 2/p)p^{-(3q-2)/[(p-2)q]}|a|_\infty + 2p^{1/2-1/q}|a| & : \quad 0 < q < 2/3 \\ |a|_\infty + 2p^{1/2-1/q}|a| & : \quad 2/3 \leq q \leq 1 \end{cases}.$$

We use this unless $0 < q < 2/3$ and $2 \leq p \leq 3$ (3 is arbitrary), in which case we use $|a|_p \leq |a|$. The result is that $|a|_p \leq C^{1/q}\left(|a|_\infty + p^{1/2-1/q}|a|\right)$, and this clearly holds for $p = 2$ as well. (5) now follows from (4) using $\mathbb{E}|X_1|^r = (r/q)\Gamma(r/q)$ and $cs \leq \Gamma(s)^{1/s} \leq Cs$

for all $s \geq 2$. The probability bound follows by optimizing over p. In Case 1 we assume that $p^{1/q} |a|_\infty \leq \sqrt{p}$, and then

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

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

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

for $p = \left(C_q^{-1} |a|_\infty^{-1} t\right)^q e^{-1}$. This value of p is allowed in Case 2 provided $t \geq C_q |a|_\infty^{-q/(2-q)}$. For $c_q |a|_\infty^{-q/(2-q)} \leq t \leq C_q |a|_\infty^{-q/(2-q)}$, 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 (6) without using (5) or the results of [12]. The methodology used is that outlined in Section 2 and does not make use of linearity. The result is dimension independent and applies in the infinite dimensional setting with $a \in \ell_2$ (the partial sums are Cauchy in L^2).

Theorem 3 *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. $\mathbb{P}\{|X_i| > t\} = \exp(-t^q)$, $t \geq 0$. Then for all $t > 0$,*

$$\mathbb{P}\left\{\left|\sum_{i=1}^n a_i X_i\right| > t\right\} \leq C \exp\left(-\min\left\{C^{-1/q} q^{2/q}\left(\frac{t}{|a|}\right)^2, C^{-1}\left(\frac{t}{|a|_\infty}\right)^q\right\}\right). \quad (8)$$

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

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

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

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

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

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

8

By direct computation, for the same range of t,

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

so

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

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

$$
\begin{aligned}
\left|\nabla\psi\left(x\right)\right| &= \left(\sum_{i=1}^{n}\left(\frac{a_{i}\phi\left(\Phi^{-1}\left(\Phi\left|x_{i}\right|\right)\right)}{f\left(F^{-1}\left(\Phi\left|x_{i}\right|\right)\right)}\right)^{2}\right)^{1/2} \\
&\leq q^{-1}\left(\sum_{i=1}^{n}a_{i}^{2}\left(A\left|x_{i}\right|+B\right)^{-2+4/q}\right)^{1/2}.
\end{aligned}
\tag{9}
$$

The function

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

is a norm, so

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

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

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

i.e.

$$\left(\sum_{i=1}^{n} a_i^2 \left(A\left|Z_i\right| + B\right)^{-2+4/q}\right)^{1/2}$$

$$\leq \quad \left[\left(\mathbb{E}\left(A\left|Z_1\right| + B\right)^{-2+4/q}\right)^{q/(4-2q)} |a|^{q/(2-q)} + A\left|a\right|_{\infty}^{q/(2-q)} \lambda\right]^{-1+2/q}. \tag{10}$$

Define

$$K \quad = \quad \left\{x \in \mathbb{R}^n : q^{-1}\left(\sum_{i=1}^{n} a_i^2 \left(A\left|x_i\right| + B\right)^{-2+4/q}\right)^{1/2} \leq R\right\}$$

where

$$R \quad = \quad q^{-1}\left[\left(\mathbb{E}\left(A\left|Z_1\right| + B\right)^{-2+4/q}\right)^{q/(4-2q)} |a|^{q/(2-q)} + A\left|a\right|_{\infty}^{q/(2-q)} \lambda\right]^{-1+2/q}$$

$$\leq \quad q^{-1}2^{-1+2/q}\left(\mathbb{E}\left(A\left|Z_1\right| + B\right)^{-2+4/q}\right)^{1/2} |a| + q^{-1}(2A)^{-1+2/q} |a|_{\infty} \lambda^{-1+2/q}$$

$$\leq \quad C_2^{1/q}\left(1/q\right)^{1/q} |a| + q^{-1}(2A)^{-1+2/q} |a|_{\infty} \lambda^{-1+2/q}. \tag{11}$$

By (10) and its probability bound, $\gamma_n\left(K\right) \geq 1 - C\exp\left(-c\lambda^2\right)$. Since $-2 + 4/q \geq 2$ the function $x \mapsto \sum_{i=1}^{n} a_i^2 \left(A\left|x_i\right| + B\right)^{-2+4/q}$ is the sum of convex functions and is therefore itself convex, which implies that K is convex. Since K is convex and, by (9), $\left|\nabla\psi\left(x\right)\right| \leq R$ on K, the restricted function $\psi|_K$ is Lipschitz, with corresponding Lipschitz constant at most R. Let $\psi^\sharp$ denote a Lipschitz extension of the restriction $\psi|_K$ such that $Lip\left(\psi^\sharp\right) \leq R$. The existence of extensions of real valued Lipschitz functions preserving the Lipschitz constant is a basic result in the theory of metric spaces, see e.g. [18]. Applying Gaussian concentration to $\psi^\sharp$,

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

yet $\mathbb{P}\left\{\psi^\sharp\left(Z\right) \neq \psi\left(Z\right)\right\} \leq C\exp\left(-c\lambda^2\right)$ so $\mathbb{P}\left\{\left|\psi\left(Z\right) - \mathbb{M}\psi^\sharp\left(Z\right)\right| > \lambda R\right\} \leq 2C\exp\left(-c\lambda^2\right)$. Assuming that this probability bound is less than $1/2$,

$$\mathbb{P}\left\{\psi\left(Z\right) < \mathbb{M}\psi^\sharp\left(Z\right) - \lambda R\right\} < 1/2$$
$$\mathbb{P}\left\{\psi\left(Z\right) > \mathbb{M}\psi^\sharp\left(Z\right) + \lambda R\right\} < 1/2$$

which implies that $\left|\mathbb{M}\psi\left(Z\right) - \mathbb{M}\psi^\sharp\left(Z\right)\right| \leq \lambda R$. The bound for concentration of $\psi\left(Z\right)$ around $\mathbb{M}\psi^\sharp\left(Z\right)$ can now be written as a concentration inequality around $\mathbb{M}\psi\left(Z\right)$, i.e.

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

By symmetry, $\mathbb{M}\psi\left(Z\right) = 0$. Now, remembering that R is bounded in terms of λ, see (11), set $t = \lambda R$ and estimate λ in terms of t using Lemma 1. By increasing the value of C in the probability bound as stated in the theorem, this probability bound becomes trivial unless the probability bound above is less than $1/2$ as required. ∎

4 Beyond convexity

Here we elaborate on the discussion in Section 2.2.

4.1 The parameter $\mathfrak{L}(\cdot)$

For any set $A \subseteq \mathbb{R}^n$ containing at least two points, let

$$\mathfrak{L}(A) = \sup \left\{ \frac{\inf \{\operatorname{Var}(f) : f \in C([0,1], A), f(0) = x, f(1) = y\}}{|x-y|} : x, y \in A, x \neq y \right\}$$

where $C([0,1], A)$ is the collection of all continuous functions $f : [0,1] \to A$ and $\operatorname{Var}(f)$ is the total variation of f,

$$\operatorname{Var}(f) = \sup \left\{ \sum_{i=1}^{N} |f(t_i) - f(t_{i-1})| : N \in \mathbb{N}, 0 = t_0 < t_1 < \dots < t_N = 1 \right\}.$$

Note that the numerator in the definition of $\mathfrak{L}(A)$ is a sort of geodesic distance. We assume the usual convention that $\inf \emptyset = \infty$, so unless A is path connected, $\mathfrak{L}(A) = \infty$. If A contains fewer than two points, set $\mathfrak{L}(A) = 1$.

Whenever A is convex, $\mathfrak{L}(A) = 1$. Conversely,

Proposition 4 *If $A \subseteq \mathbb{R}^n$ is closed and $\mathfrak{L}(A) = 1$ then A is convex.*

Proof. Consider any $x, y \in A$ with $x \neq y$, any $\lambda \in (0, 1)$ and any $f \in C([0,1], A)$ such that $f(0) = x$, $f(1) = y$. By continuity there exists $t \in (0, 1)$ and $\theta \perp (y - x)$ such that $f(t) = \lambda x + (1 - \lambda) y + \theta$. Now $f(t) \in A \cap B(\lambda x + (1 - \lambda) y, |\theta|)$ and

$$\operatorname{Var}(f) \geq \sqrt{(1-\lambda)^2 |x-y|^2 + |\theta|^2} + \sqrt{\lambda^2 |x-y|^2 + |\theta|^2}$$

and the result follows by taking $\operatorname{Var}(f) \to |x - y|$ and using the fact that A is closed. ∎

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

$$Lip(\psi, x) = \lim_{\varepsilon \to 0^+} Lip\left(\psi|_{B(x,\varepsilon) \cap A}\right) \tag{12}$$

where $B(x, \varepsilon)$ denotes the Euclidean ball centred at x of radius ε. When A contains a neighbourhood of $x \in \mathbb{R}^n$ and $\nabla \psi$ is continuous at x, then $Lip(\psi, x) = |\nabla \psi(x)|$. Our main reason for defining $\mathfrak{L}(\cdot)$ is the following observation.

Proposition 5 *For any non-empty set $A \subseteq \mathbb{R}^n$ with $\mathfrak{L}(A) < \infty$ and any function $\psi : A \to \mathbb{R}$,*

$$Lip(\psi) \leq \mathfrak{L}(A) \sup \{Lip(\psi, x) : x \in A\}. \tag{13}$$

If, furthermore, A contains at least two points and 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

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

Proof. We start with (13). Let $L = \sup\{Lip(\psi, x) : x \in A\}$. Consider any $x, y \in A$, any $\varepsilon > 0$, and any continuous path $f : [0, 1] \to A$ between x and y. It follows from the definition of L and of the local Lipschitz constant that for each $z \in f([0, 1])$ there exists a nonempty open ball B_z centered at z with $Lip(\psi|_{B_z}) < L + \varepsilon$. Let $\mathcal{I}$ be the collection of all connected components of inverse images of these balls (each will be an interval that is open in the subspace topology of $[0, 1]$). The elements of $\mathcal{I}$ cover $[0, 1]$ so by compactness there is a finite subcover $\mathcal{I}^* \subseteq \mathcal{I}$. After removing unnecessary elements of $\mathcal{I}^*$ one by one, we arrive at a minimal subcover $\mathcal{I}^{**}$. The elements of $\mathcal{I}^{**}$ can then be ordered as follows: $I_1 < I_2$ if there exists $t_1 \in I_1$ such that for all $t_2 \in I_2$, $t_1 < t_2$, and $I_1 \leq I_2$ if either $I_1 < I_2$ or $I_1 = I_2$. By minimality this is a linear ordering. The elements of $\mathcal{I}^{**}$ can then be labelled $I_1, I_2...I_N$, and we write $I_1 = [a_1, b_1)$, $I_i = (a_i, b_i)$ for $2 \leq i \leq N - 1$, and $I_N = (a_N, b_N]$, with $a_1 = 0$ and $b_N = 1$. We assume that $N \geq 3$ and leave the simpler case $N \in \{1, 2\}$ to the reader. Now $a_{i+1} < b_i$ for all $1 \leq i \leq N - 1$, otherwise $b_i \notin \cup I_j$, so there exists $c_i \in (a_{i+1}, b_i)$. By minimality, $b_{i-1} < a_{i+1}$ for $2 \leq i \leq N - 1$ (otherwise $I_i \subseteq I_{i-1} \cup I_{i+1}$). So, setting $c_0 = 0$ and $c_N = 1$, the sequence $(c_i)_0^N$ is strictly increasing. It also follows that for all $1 \leq i \leq N$, $[c_{i-1}, c_i] \subseteq I_i$, and from the construction of $\mathcal{I}$ and $\mathcal{I}^{**}$, that ψ is $(L+\varepsilon)$-Lipschitz on each $f(I_i)$. This implies that $|\psi(f(c_i)) - \psi(f(c_{i-1}))| \leq (L + \varepsilon)|f(c_i) - f(c_{i-1})|$, and by the triangle inequality $|\psi(f(0)) - \psi(f(1))| \leq (L + \varepsilon)\sum_{i=1}^{N}|f(c_{i-1}) - f(c_i)| \leq (L + \varepsilon)\mathrm{Var}(f)$. We now choose f such that $\mathrm{Var}(f) \leq (\mathfrak{L}(A) + \varepsilon)|x - y|$ and we let $\varepsilon \to 0$. This implies that $|\psi(x) - \psi(y)| \leq \mathfrak{L}(A)L|x - y|$, which implies (13). This also 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

$$\rho(x, y) = \inf\{\mathrm{Var}(f) : f \in C([0, 1], A), f(0) = x, f(1) = y\}$$

It is easily seen that ρ is a metric on A. Now consider any $\varepsilon > 0$. It follows 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 $\varepsilon \to 0$. ∎

Consider the case where A is a non-affine deformation of a convex body K, as the inverse image under the action of some continuous 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 $\mathfrak{L}(A) \leq \sqrt{2}$. So, for example, $\mathfrak{L}(B_p^n) \leq \sqrt{2}$ for all $p \in (0, 1)$, where

$$B_p^n = \left\{x \in \mathbb{R}^n : \sum_{i=1}^{n}|x_i|^p \leq 1\right\}.$$

The general case is covered by the following proposition.

Proposition 6 *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 continuous 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. 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 of $Tx \in K$ 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 $\mathfrak{L}(K) \leq \sqrt{2}$. $\blacksquare$

For a function $\psi : \mathbb{R}^n \to \mathbb{R}$ define

$$\mathfrak{L}(\psi) = \sup_{t \in \mathbb{R}} \mathfrak{L}(\{\psi \leq t\}). \tag{14}$$

The following result summarizes the methodology outlined in Section 2.

Theorem 7 *Let $n \in \mathbb{N}$, $A > 0$, let μ be a probability measure on $\mathbb{R}^n$, and let $T : \mathbb{R}^n \to \mathbb{R}^n$ and $f, Q : \mathbb{R}^n \to \mathbb{R}$ be measurable functions such that $\mu = T\gamma_n$, and such that $f \circ T$ is locally Lipschitz with*

$$Q(x) \geq Lip(f \circ T, x)$$

13

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 $\mathbb{P}\{Q(Z) > R\} \leq A(1 - \Phi(t))$. Then

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

where $\mathfrak{L}(Q)$ is defined by (14).

Proof of Theorem 7. Since the distribution of the random vector $T(Z)$ is μ, we may assume without loss of generality that $X = T(Z)$. Set $\psi = f \circ T$, in which case $\psi(Z) = f(X)$. Let $K = \{x : Q(x) \leq R\}$. By assumption, $\mathbb{P}\{Z \in K\} \geq 1 - A(1 - \Phi(t))$, and for all $x \in K$, $Lip(\psi|_K, x) \leq Lip(\psi, x) \leq R$. By Proposition 5, $Lip(\psi|_K) \leq \mathcal{L}(Q)R$. As in the proof of Theorem 3, the function $\psi|_K$ may be extended to a function $\widetilde{\psi} : \mathbb{R}^n \to \mathbb{R}$ such that $Lip\left(\widetilde{\psi}\right) = Lip(\psi|_K)$. By Gaussian concentration of Lipschitz functions and the union bound, it follows that with probability at least $1 - (A+2)(1 - \Phi(t))$, $\psi(Z) = \widetilde{\psi}(Z)$, and $\left|\widetilde{\psi}(Z) - \mathbb{M}\widetilde{\psi}(Z)\right| \leq \mathcal{L}(Q)Rt$. Since $(A+2)(1 - \Phi(t)) < 1/2$, this implies that greater than 50% of the mass of the distribution of $\psi(Z)$ lies in the closed interval from $\mathbb{M}\widetilde{\psi}(Z) - \mathcal{L}(Q)Rt$ to $\mathbb{M}\widetilde{\psi}(Z) + \mathcal{L}(Q)Rt$, so $\left|\mathbb{M}\psi(Z) - \mathbb{M}\widetilde{\psi}(Z)\right| \leq \mathcal{L}(Q)Rt$. The result now follows by the triangle inequality. $\blacksquare$

4.2   Convex subsets with large Gaussian measure

The relevance of the following result is that, in the notation there, $\lambda\sqrt{n}B_2^n$ is an approximating ellipsoid for the possibly non-convex Orlicz ball λB_F, in the sense of the Gaussian measure of the complement. The result applies more generally with essentially the same proof; we state it as is for simplicity.

Proposition 8 *Let $F : [0, \infty) \to [0, \infty)$ be strictly increasing and concave, such that $F(0) = 0$, $F(1) = 1$ and $\lim_{x \to \infty} F(x) = \infty$. Let $n \in \mathbb{N}$ and define*

$$B_F = \left\{x \in \mathbb{R}^n : \sum_{i=1}^n F(|x_i|) \leq n\right\}.$$

Then for all $\lambda > 1$ we have $\lambda\sqrt{n}B_2^n \subseteq \lambda B_F$ and

$$\gamma_n\left(\mathbb{R}^n \setminus C\lambda\sqrt{n}B_2^n\right) \leq \gamma_n\left(\mathbb{R}^n \setminus \lambda B_F\right)$$

where $C > 1$ is a universal constant.

Proof. By convexity there exists $m \in (0, 1)$ such that $F(s) \leq m(s - 1) + 1$ for all $s \in [0, \infty)$. Therefore, for any $x \in \sqrt{n}B_2^n$,

$$\sum_{i=1}^n F(|x_i|) \leq \sum_{i=1}^n (m(|x_i| - 1) + 1) \leq m\sqrt{n}|x| + n(1 - m) \leq n$$

so $\lambda\sqrt{n}B_2^n \subseteq \lambda B_F$ as claimed. By polar integration and using the tangent line approximation to a concave function,

$$\gamma_n\left(\mathbb{R}^n \setminus C\lambda\sqrt{n}B_2^n\right) = (2\pi)^{-n/2}n \cdot \mathrm{vol}\left(B_2^n\right) \int_{C\lambda\sqrt{n}}^{\infty} e^{-r^2/2+(n-1)\ln r}dr$$

$$\leq \left(\frac{n}{2\pi}\right)^{n/2} \mathrm{vol}\left(B_2^n\right) C^{n-2}\lambda^{n-2} \exp\left(\frac{-C^2\lambda^2 n}{2}\right).$$

On the other hand, since F is strictly increasing, $\lambda B_F \subseteq \{x \in \mathbb{R}^n : \exists i, |x_i| \leq \lambda\}$ so using the fact that the tails of the standard normal density are convex and therefore lie above their tangent lines,

$$\gamma_n\left(\mathbb{R}^n \setminus \lambda B_F\right) \geq \left(2\int_{\lambda}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt\right)^n \geq \left(\frac{1}{\sqrt{2\pi}\lambda}e^{-\lambda^2/2}\right)^n$$

and the result follows by taking C large enough. ∎

For general star bodies it may not be as easy to find approximating convex subsets as in Proposition 8, although as $\lambda \to \infty$ this is straightforward. Let σ_{n-1} denote Haar measure on S^{n-1}, normalized so that $\sigma_{n-1}\left(S^{n-1}\right) = 1$.

Proposition 9 *Let $n \in \mathbb{N}$ with $n \geq 2$, let $\|\cdot\| : S^{n-1} \to (0,\infty)$ be any bounded Borel measurable function, let $b = \sup(\|\cdot\|)$, and let K be the unit ball of the homogeneous extension of $\|\cdot\|$, i.e.*

$$K = \{0\} \cup \left\{x \in \mathbb{R}^n \setminus \{0\} : |x|\left\|\frac{x}{|x|}\right\| \leq 1\right\}.$$

For simplicity, assume that the L^∞ norm of $\|\cdot\|$ (i.e. its essential supremum) is also b. Then for all $\lambda \geq \sqrt{2(n-1)}b$ and all $\delta \in (0,1)$,

$$\gamma_n\left(\mathbb{R}^n \setminus RB_2^n\right) \leq \gamma_n\left(\mathbb{R}^n \setminus \lambda K\right)$$

where

$$R = \frac{\lambda}{(1-\delta)b} + \frac{2(1-\delta)b}{\lambda}\ln\left[\frac{4}{\sigma_{n-1}\left(\{\|\theta\| \geq (1-\delta)b\}\right)}\right].$$

Proof. By polar integration, and using $t^{n-1} \geq r^{n-1}$ for $t \geq r \geq 0$ and the standard normal tail bound from the proof of Proposition 8, and since $\lambda > b$, $\gamma_n\left(\mathbb{R}^n \setminus \lambda K\right)$ is bounded below by

$$\frac{n \cdot \mathrm{vol}\left(B_2^n\right)}{2(2\pi)^{n/2}} \int_{S^{n-1}} \left(\frac{\lambda}{\|\theta\|}\right)^{n-2} \exp\left(-\frac{1}{2}\left(\frac{\lambda}{\|\theta\|}\right)^2\right) d\sigma_{n-1}(\theta)$$

$$\geq \sigma_{n-1}\left(\{\|\theta\| \geq (1-\delta)b\}\right)\frac{n \cdot \mathrm{vol}\left(B_2^n\right)}{2(2\pi)^{n/2}}r^{n-2}e^{-r^2/2} \tag{15}$$

where $r = (1-\delta)^{-1}b^{-1}\lambda$. The inequality on the second line holds since $\lambda > b\sqrt{n-2}$ and $t^{n-2}e^{-t^2/2}$ is decreasing for $t \geq \sqrt{n-2}$. Since $R > r$, by log-concavity,

$$R^{n-2}e^{-R^2/2} \leq r^{n-2}e^{-r^2/2}\exp\left[-(R-r)\left(r - \frac{n-2}{r}\right)\right] \tag{16}$$

15

and it follows from the definition of R that

$$\exp\left[-(R-r)\left(r-\frac{n-2}{r}\right)\right] \leq \frac{1}{4}\sigma_{n-1}\left(\{\|\theta\| \geq (1-\delta)b\}\right). \qquad (17)$$

It follows from (16) and (17) that the quantity in (15) is bounded below by

$$\frac{2n \cdot \mathrm{vol}\left(B_2^n\right)}{(2\pi)^{n/2}}R^{n-2}e^{-R^2/2}.$$

Since $R \geq \sqrt{2(n-1)}$, it follows by polar integration and the upper bound for a standard normal tail based on log-concavity that this is bounded below by $\gamma_n\left(\mathbb{R}^n \setminus RB_2^n\right)$. $\blacksquare$

Acknowledgements

References

[1] Aubrun, G., Szarek, S. J., and Ye, D.: Entanglement thresholds for random induced states. Comm. Pure Appl. Math., 67 (1), 129-171 (2014)

[2] Adamczak, R.: A note on the Hanson-Wright inequality for random vectors with dependencies. Electron. Commun. Probab., 20 (72), 13pp (2015)

[3] 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)

[4] 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.

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

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

[7] Cattiaux, P., Gozlan, N., Guillin A., Roberto, C.: Functional inequalities for heavy tailed distributions and application to isoperimetry. Electon. J. Probab. 15 (13), 346-385 (2010)

[8] Fresen, D. J.: Variations and extensions of the Gaussian concentration inequality, Part II. Expected to appear in Electron. J. Probab. (pending revision)

[9] Fresen, D. J.: Random Euclidean embeddings in finite dimensional Lorentz spaces. arXiv: 2104.11974. Submitted.

[10] Gozlan, N.: Poincaré inequalities and dimension free concentration of measure. Ann. Inst. Henri Poincaré Probab. Stat. 46 (3), 708-739 (2010)

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

[12] 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)

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

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

[15] Latala, R.: Estimation of moments of sums of independent real random variables. Ann. Probab. 25 (3), 1502–1513 (1997).

[16] Ledoux, M.: The Concentration of Measure Phenomenon, Volume 89 of Mathematical Surveys and Mono-graphs. American Mathematical Society, Providence, RI, 2001.

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

[18] Mattila, P.: Geometry of Sets and Measures in Euclidean Spaces: Fractals and Rectifiability. Cambridge University Press, Cambridge, 1995.

[19] Meckes, M., Szarek, S., J.: Concentration for noncommutative polynomials in random matrices. Proc. Amer. Math. Soc. 140 (5), 1803-1813 (2012)

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

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

[22] Sudakov, V. N., Tsirel'son B. S.: 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)

[23] Talagrand, M.: A new isoperimetric inequality and the concentration of measure phenomenon. Geometric aspects of functional analysis (1989–90), 94–124, Lecture Notes in Math., 1469, Springer, Berlin, 1991.

[24] Talagrand, M.: Are all sets of positive measure essentially convex? Geometric aspects of functional analysis (Israel, 1992–1994), 295–310, Oper. Theory Adv. Appl., 77, Birkhauser, Basel, 1995.

[25] Talagrand, M.: Are many small sets explicitly small? STOC'10—Proceedings of the 2010 ACM International Symposium on Theory of Computing, 13–35, ACM, New York, 2010.

[26] Tanguy, K.: Non asymptotic variance bounds and deviation inequalities by optimal transport. Electron. J. Probab. 24, 18pp (2019)

[27] Villani, C.: Optimal Transport Old and New. Grundlehren der Mathematischen Wissenschaften 338. Springer-Verlag, Berlin 2009.

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