Variations and extensions of the Gaussian concentration inequality, Part II

Daniel J. Fresen*

Abstract

Pisier's version of the Gaussian concentration inequality gives birth to a wide class of concentration inequalities for quite general functions sitting on a fairly general class of measures. While the methods here have the potential to be applied more generally, we focus on the case of locally Lipschitz functions defined on product measures, including the case of heavy tailed distributions. This approach is more direct than much of the modern theory of concentration of measure (i.e. Poincaré and log-Sobolev inequalities, estimating moments etc.).

Contents

*University of Pretoria, Department of Mathematics and Applied Mathematics, daniel.fresen@up.ac.za or djfb6b@mail.missouri.edu.

1 Introduction

1.1 Methodology: Tail comparison and Gaussian concentration

Recall Pisier's version of the Gaussian concentration inequality [24] that allows one to remove the assumption that f is Lipschitz (we have added the assumption of differentiability to simplify the wording and considered the special case where $Range(f) \subseteq \mathbb{R}$): if $f : \mathbb{R}^n \to \mathbb{R}$ is C^1, and X and Y are independent random vectors in $\mathbb{R}^n$ each with the standard normal distribution, then for any convex function $\varphi : \mathbb{R} \to \mathbb{R}$,

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

Obviously one may re-word the original conclusion as follows: there exists a random variable Z with the standard normal distribution in $\mathbb{R}$, independent of X, such that for any convex function $\varphi : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}\varphi\left(f(X) - f(Y)\right) \leq \mathbb{E}\varphi\left(\frac{\pi}{2}\left|\nabla f(X)\right| Z\right) \tag{1}$$

Inequality (1) is known as a majorization inequality; we refer the reader to the book [17] for a detailed account of majorization. Taking $\varphi(t) = t$ and then $\varphi(t) = -t$, we see that $\mathbb{E}\left[f(X) - f(Y)\right] = \mathbb{E}\left[\frac{\pi}{2}\left|\nabla f(X)\right| Z\right]$, and about this common mean $f(X) - f(Y)$ is, in a sense *(through the eyes of convex functions)* no more spread out than $\frac{\pi}{2}\left|\nabla f(X)\right| Z$. It is almost as though the quantiles of the distribution of $\left|f(X) - f(Y)\right|$ are bounded above by those of $2^{-1}\pi\left|\nabla f(X)\right| \cdot |Z|$. Indeed, in the most classical case when f is Lipschitz, this is essentially the case (up to a constant, even $o(1)$), and in many particular cases one may use Markov's inequality and an appropriate choice of φ to recover good estimates for $\mathbb{P}\left\{\left|f(X) - f(Y)\right| > t\right\}$. This naturally leads to a bound for $\mathbb{P}\left\{\left|f(X) - \mathbb{M}f(X)\right| > t\right\}$, where $\mathbb{M}$ denotes median.

In Section 2.1, see in particular Proposition 2 and Theorem 3, we prove general results of the type

$$\mathbb{P}\left\{\left|f(X) - f(Y)\right| > t\right\} \leq C\mathbb{P}\left\{2^{-1}\pi\left|\nabla f(X)\right| \cdot |Z| > t\right\}$$

where C can be taken as a universal constant. This estimate is optimal (up to the value of C) in a sense explained later in the paper. This is beyond what is possible using the functions $\varphi(t) = |t|^p$ $(2 \leq p < \infty)$ or $\varphi(t) = \exp(at)$ $(0 < a < \infty)$ in the usual way, which often gives an estimate on $-\log\mathbb{P}\left\{\left|f(X) - \mathbb{M}f(X)\right| > t\right\}$ up to a factor of $(1 + o(1))$ (which is a fairly typical degree of precision in large deviation theory). This is particularly useful when the tails of $2^{-1}\pi\left|\nabla f(X)\right| Z$ are polynomial and the usual use of Markov's inequality fails to recover sub-Gaussian estimates that may hold in a certain range of the distribution. Our results in this direction extend those of Pinelis [22, 23] whose work covers the case where, in the notation here, the function $t \mapsto \mathbb{P}\left\{\frac{\pi}{2}\left\langle\nabla f(X), Y\right\rangle > t\right\}$ can be written in the form $h(t)^{-r}$, where h is convex.

1.2 Methodology: When the distribution of X is not Gaussian

In the more general case where the distribution of X is no longer assumed to be normal but is now arbitrary, there exists a normally distributed random vector Z in $\mathbb{R}^n$ and a measurable function $T : \mathbb{R}^n \to \mathbb{R}^n$ such that the distribution of $T(Z)$ is the same as that of X. Since we are only interested in the distribution of X, we may assume without loss of generality that $X = T(Z)$. An example of such a T is the Knöthe-Rosenblatt rearrangement [30], which is probably not always the best option unless X has independent coordinates. Writing $f(X) = (f \circ T)(Z)$ we are often in a position to apply Gaussian concentration to $f \circ T$ to obtain a concentration inequality for $f(X)$. If the distribution of X has disconnected support (for example the discrete case) we run into problems, since T will no longer be continuous. In general we may have problems when the density of X is small in the central regions of its distribution (the reader will get a better understanding of this in the body of the paper). Whether we can apply Gaussian concentration to $(f \circ T)(Z)$ depends on the regularity of $f \circ T$, which in turn depends on two things: regularity of the distribution of X which translates to regularity of T, and regularity of the function f.

The golden shine of Pisier's version of Gaussian concentration is, besides the very simple proof with contribution from Maurey, the fact that the Lipschitz condition has been removed. This allows for a large class of $f \circ T$ (in the notation above) which in turn allows for a large class of f and of X. For example, if X has i.i.d. coordinates, even heavy tailed, then f takes the form

$$f(x) = (h(x_i))_{i=1}^n$$

where, due to the heavy tails of each X_i, h may grow somewhat rapidly. For example $h(t)$ may grow like t^p when X_i has Weibull type tails, and $\exp(t^2/p)$ when X_i has polynomial tails. Of course this means that the distribution of $|\nabla(f \circ T)(Z)|$ will also have thicker tails. However in many cases, $|\nabla(f \circ T)(Z)|$ is, with high probability, the same order of magnitude as $\mathbb{E}|\nabla(f \circ T)(Z)|$, which ultimately leads to an estmate on $\mathbb{P}\{|f(X) - \mathbb{M}f(X)| > t\}$ with a central subgaussian hump of standard deviation about $\mathbb{E}|\nabla(f \circ T)(Z)|$ and tails that eventually show their thickness for values of t much larger than $\mathbb{E}|\nabla(f \circ T)(Z)|$.

Applying Gaussian concentration when the underlying distribution is not normal has been done before, see for example p.1046 in Naor's paper [19], specifically in the context of Lipschitz images of the normal distribution. A key contribution of this paper is to radically extend this to a much wider class of distributions. Our goal is, colloquially speaking, to milk Pisier's Gaussian concentration inequality for all it is worth. As it turns out, it is worth quite a lot, and it would be only a mild exageration to say that (in the continuous setting) almost any concentration inequality with mild assumptions on f and X can be proved using the Gaussian concentration inequality.

1.3 Overview of what we prove

Let $Lip(T, x)$ denote the local Lipschitz constant of a function around a point x,

$$Lip(T, x) = \lim_{\varepsilon \to 0^+} Lip\left(T|_{B(x,\varepsilon)}\right) \tag{2}$$

For a smooth function $f : \mathbb{R}^n \to \mathbb{R}$, define

$$Lip_s(f) = \sup_{x \in \mathbb{R}^n} \left(\sum_{i=1}^n \left| \frac{\partial f}{\partial x_i}(x) \right|^s \right)^{1/s} \qquad Lip_s^\sharp(f) = \left(\sum_{i=1}^n \sup_{x \in \mathbb{R}^n} \left| \frac{\partial f}{\partial x_i}(x) \right|^s \right)^{1/s}$$

with the usual interpretation when $s = \infty$. Sometimes $Lip_s^\sharp(f)$ is significantly larger than $Lip_s(f)$, but for linear functions or when $s = \infty$, for example, they are equal. We prove concentration inequalities for $f(X)$ about its median involving $Lip_s(f)$ and $Lip_s^\sharp(f)$, where X is a random vector in $\mathbb{R}^n$ with independent coordinates $(X_i)_1^n$, each X_i with distribution μ_i. We assume that the quantile function $F_i^{-1} : (0,1) \to \mathbb{R}$ (the generalized inverse of the CDF of μ_i, see (18)) is locally Lipschitz. There are two main types of estimates that we assume. The first is

$$Lip\left(F_i^{-1} \circ \Phi, s\right) \leq (1 + |s|)^{-1+2/q} : \forall s \in \mathbb{R} \tag{3}$$

where Φ is the standard normal CDF, $Lip\left(F_i^{-1} \circ \Phi, s\right)$ denotes the local Lipschitz constant of $F_i^{-1} \circ \Phi$ at s, see (2) and $0 < q < 1$. This assumption reflects **Weibull type tail behavior** of the form

$$\mathbb{P}\{|X_i| > t\} \leq C \exp(-C_q t^q) \tag{4}$$

(but obviously an assumption on the local Lipschitz constant is stronger than a tail bound of the form 4). Theorem 4 gives deviation inequalities assuming (3), and states that for all $t > 0$,

$$\mathbb{P}\{|f(X) - \mathbb{M}f(X)| > t\} \leq 2\exp\left(-c_q \min\left\{\left(\frac{t}{Lip_2^\sharp(f)}\right)^2, \left(\frac{t}{Lip_\infty^\sharp(f)}\right)^q\right\}\right)$$

and that

$$\begin{aligned} &\mathbb{P}\left\{|f(X) - \mathbb{M}f(X)| > C_q\left(1 + \left(\log\frac{n}{t^{-2+4/q}}\right)^{1/q-1/2}\right)\left(tLip_2(f) + t^{2/q}Lip_\infty(f)\right)\right\} \\ \leq\ & 2\exp\left(-t^2/2\right) \end{aligned}$$

The second type of estimate that is assumed is

$$Lip\left(F_i^{-1}, s\right) \leq \min\{s, 1-s\}^{-1-1/q} : \forall s \in (0,1) \tag{5}$$

where $2 < q < \infty$. This assumption reflects **polynomial tail behavior** of the form

$$\mathbb{P}\{|X_i| > t\} \leq C_q' (|C_q t| + 1)^{-q-1}$$

Theorem 5 gives deviation inequalities assuming (5). Even in the linear case we get estimates that appear to be new, see [9, Theorem 6]. At least they are not found in the work of Hitczenko, Montgomery-Smith and Oleszkiewicz [12], and Latała [15] (see the remarks after Theorem 4 and those in [9, Section 3] for more details). In the case of polynomial tails, our sub-Gaussian estimates go deeper into the tails (up to probability $Cn^{1-q/2}(\log n)^{q/2}$) than does the weighted Berry-Esseen inequality, which implies a sub-Gaussian bound up to probability $Cn^{-1/2}(\log n)^{-r/2}$ (see remarks after Theorem 5 and in Section 3).

Results of Latała and Oleszkiewicz [16], Schechtman and Zinn [26] and Barthe, Cattiaux and Roberto [2] (see Theorems 10, 11 and 12 here) apply to functions Lipschitz with respect to the ℓ_2^n norm, sitting on measures with thicker-than-Gaussian tails (on the appropriate scale). Obviously one cannot hope for sub-Gaussian bounds for all such f. We show in Theorem 6 that one can indeed prove sub-Gaussian bounds for *most* Lipschitz functions (more precisely, most rotations of an arbitrary Lipschitz function).

The methods developed here can also be used to study Gaussian concentration of the ℓ_p^n norm, achieving bounds similar to those of Paouris, Valettas and Zinn [20], and more generally concentration of the ℓ_p^n norm on the ℓ_q^n ball, achieving bounds similar to those of Naor [19]. An earlier version of this paper, still on the arxiv, provides detailed statements and proofs.

Other results and techniques:

• We use a Minkowski functional based on a Poisson point process that effectively estimates the quantiles of linear functionals of probability measures on $\mathbb{R}^n$ and in turn defines a convex body similar to the expected convex hull of a fixed sample size (and to the floating body), but the formula for this functional is more convenient to use. This technique is used in the proof of Theorem 5, see (39).

• Section 4 contains results on order statistics of non-negative random variables, possibly heavy tailed, including a bound for the sum that gives the correct order of magnitude up to a parameter independent of n, and an improvement upon an estimate found in Guédon, Litvak, Pajor and Tomczak-Jaegermann [11], see in particular Corollary 15 and Example 16.

1.4 Work of others

The idea of expressing a probability measure μ in the form $\mu = T\nu$ and showing that μ has a certain property based on a corresponding property of ν, is not at all new. Indeed transportation plays a key role in the modern theory of concentration of measure. Usually one is interested in the distance between x and Tx. Here we are more interested in the distance between Tx and Ty for $x, y \in \mathbb{R}^n$. Among transportation methods most similar to those here, let us mention the following examples (we refer the reader to the papers cited below for more details):

Gozlan [10], see in particular Section 1.4 and (1.15) in his paper, uses the following method of transportation as part of his study of inequalities that he denotes $\mathbb{SG}(\omega, \cdot)$: He considers functions $\omega : \mathbb{R} \to \mathbb{R}$ that are odd, increasing, and such that $\omega(x)/x$ is nondecreasing on $(0, \infty)$, and shows that a measure μ satisfies a Poincaré inequality with respect to the metric

$$d_\omega(x, y) = \left(\sum_{i=1}^{d} |\omega(x_i) - \omega(y_i)|^2 \right)^{1/2} : x, y \in \mathbb{R}^d$$

if and only if the pushforward measure $T\mu$ satisfies the classical Poincaré inequality (with respect to the Euclidean metric), where $Tx = (\omega(x_i))_1^n$.

Cattiaux, Gozlan, Guillin and Roberto [7, Section 5, Theorem 5.2] show and use the fact that if ν is a spherically symmetric probability measures on $\mathbb{R}^n$ and $\mu = T\nu$ for some radial transformation $Tx = \varphi(|x|)|x|^{-1}x$ with $\varphi : [0, \infty) \to [0, \infty)$ and $\varphi(0) = 0$, and ν satisfies Poincaré's inequality with constant C, then μ satisfies the following weighted

Poincaré inequality,

$$\mathrm{Var}_\mu(f) \le C \int \omega(|x|)^2 |\nabla f|^2 \, d\mu(x) : \forall f$$

with weight defined by

$$\omega(r) = \max\left(\varphi' \circ \varphi^{-1}(r), \frac{r}{\varphi(r)}\right)$$

Tanguy [29] proves bounds for the variance of a function f with respect to a product measure μ^n by writing $\mu^n = T\nu^{(n)}$ for some other product measure ν^n, and applying the classical Poincaré inequality to the function $f \circ T$ with respect to ν^n (assuming that ν satisfies the Poincaré inequality with constant C_ν) to obtain

$$\mathrm{Var}(f(X)) \le C_\nu \sum_{i=1}^{n} \mathbb{E}\left[(\partial_i f)^2 \circ T(Y) \left(\frac{\kappa_\nu(Y_i)}{\kappa_\mu(t(Y_i))}\right)^2\right]$$

where X has i.i.d. coordinates each with distribution μ and Y has i.i.d. coordinates each with distribution ν, $Tx = (t(x_i))_1^n$, and κ_μ, κ_ν denote the hazard functions

$$\kappa_\mu(s) = (\mu(s,\infty))^{-1} \frac{d\mu}{dx}(s) \qquad \kappa_\nu(s) = (\nu(s,\infty))^{-1} \frac{d\nu}{dx}(s)$$

(obviously this can be written in terms of the local Lipschitz constant of the transportation map). In the special case of $f(x) = \max_{1\le i\le n} x_i$, this is used to obtain a deviation inequality (see Theorem 1.2 in his paper).

1.5 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. Dependence on variables will usually be indicated by subscripts, C_q, c_q etc. 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 $\mathbb{R}$.

2 Main results

2.1 Tail comparison inequalities

We consider the problem of estimating the cumulative distribution and quantile function for a random variable X, given that $\mathbb{E}\varphi(X) \le \mathbb{E}\varphi(Y)$ for some random variable Y and all convex functions $\varphi : \mathbb{R} \to \mathbb{R}$ (we can think of X as unknown and Y as given). Motivated by Pisier's version of the Gaussian concentration inequality, this is a special case of the more general setting where we have $\mathbb{E}\varphi(X) \le \Lambda(\varphi)$ for all $\varphi \in E$, where E is some collection of convex functions and $\Lambda : E \to \mathbb{R}$. The most common functions to use are of the form $\varphi(t) = |t|^p$ $(p \geq 2)$ and $\varphi(t) = \exp(at)$ $(a > 0)$. Assuming that φ is non-negative, and strictly increasing on $[t, \infty)$ (and, say, $\varphi(t) \neq 0$), Markov's inequality implies

$$\mathbb{P}\{X > t\} \leq \frac{\mathbb{E}\varphi(X)}{\varphi(t)}$$

We are interested in minimizing this estimate over the class of all such φ. One cannot hope for an estimate better than $\mathbb{P}\{X > t\} \leq \mathbb{P}\{Y > t\}$, since one may take $X = Y$, and we leave it as an exercise to the reader to show that if Z has the standard normal distribution on $\mathbb{R}$, then

$$\liminf_{t\to\infty} \inf_{a>0} \frac{\mathbb{E}e^{aZ}}{e^{at}\mathbb{P}\{Z > t\}} = \liminf_{t\to\infty} \inf_{p>2} \frac{\mathbb{E}|Z|^p}{t^p\mathbb{P}\{|Z| > t\}} = \infty$$

The first expression is easy to evaluate. For the second we recommend more qualitative reasoning related to the techniques of Propositions 1 and 2.

The reader may compare the statement and proof of Proposition 1 below with that of Lemma 2.1 in Pinelis [22] (the special case of $\alpha = 1$ and $\psi = 1_{[t,\infty)}$ in his notation). In his paper he minimizes the RHS of

$$\mathbb{P}\{X > t\} \leq \mathbb{E}\varphi(X) \leq \mathbb{E}\varphi(Y)$$

over the class of all non-decreasing convex functions $\varphi : \mathbb{R} \to \mathbb{R}$ such that $\varphi \geq 1_{[t,\infty)}$. Here we minimize the RHS of

$$\mathbb{P}\{X > t\} \leq \frac{\mathbb{E}\varphi(X)}{\varphi(t)} \leq \frac{\mathbb{E}\varphi(Y)}{\varphi(t)}$$

over the class of all non-negative convex functions $\varphi : \mathbb{R} \to \mathbb{R}$ that are strictly increasing on $[t, \infty)$. While these are different functionals defined on different classes, the minimum is of course the same. The main difference between his work and ours is Proposition 2 below.

Proposition 1 *Let μ be a non-atomic probability measure on $\mathbb{R}$ with $\int_{\mathbb{R}} |x|\, d\mu(x) < \infty$, and let Y be a random variable with distribution μ. Let $t \in \mathbb{R}$ with $t > \mathbb{E}Y$, and let E denote the class of all non-negative convex functions $\varphi : \mathbb{R} \to \mathbb{R}$ that are strictly increasing on $[t, \infty)$ and such that $\varphi(t) \neq 0$. Then the function*

$$\varphi \mapsto \frac{\mathbb{E}\varphi(Y)}{\varphi(t)}$$

defined on E and taking values in $[0, \infty]$ has a global minimum at $\varphi_a(x) = \max\{0, a(x - t) + 1\}$, where $a \in (0, \infty)$ is such that

$$\int_{t-a^{-1}}^{\infty} (x - t)\, d\mu(x) = 0$$

(any such a defines a minimizer, and at least one such a exists), and for this minimizer,

$$\frac{\mathbb{E}\varphi_a(Y)}{\varphi_a(t)} = \mathbb{P}\{Y > t - a^{-1}\}$$

Proof. Replacing φ with $\varphi/\varphi(t)$, we may assume that $\varphi(t) = 1$, and so we add this condition to the constraints on φ. By convexity $\varphi(x) \geq \max\{0, \varphi'(t)(x-t)+1\}$, where $\varphi'(t)$ here denotes $\lim_{h\to 0^+}(\varphi(t+h)-\varphi(t))/h$ (necessarily exists and is finite), and so $\mathbb{E}\varphi(X) \geq \mathbb{E}\max\{0, \varphi'(t)(X-t)+1\}$. This means that we may restrict our attention even further to the collection of functions φ_a, $a \in [0,\infty)$, as defined in the statement of the theorem. Differentiating under the integral sign, for $a \in (0,\infty)$

$$\frac{d}{da}\int_{\mathbb{R}} \max\{0, a(x-t)+1\}\, d\mu(x) = \int_{t-a^{-1}}^{\infty} (x-t)\, d\mu(x)$$

which is a continuous non-decreasing function of a. By the assumption $t > \mathbb{E}Y$, this function is negative for some $a \in (0,\infty)$ and converges to $\int_t^\infty (x-t)\, d\mu(x) \geq 0$ as $a \to \infty$. Therefore there exists $a > 0$ such that $\int_{t-a^{-1}}^{\infty}(x-t)\, d\mu(x) = 0$, and the convex function $a \mapsto \int_{\mathbb{R}} \max\{0, a(x-t)+1\}\, d\mu(x)$ achieves a global minimum over $(0,\infty)$ at this value of a. By continuity, this is also a global minimum over $[0,\infty)$. ■

Consider any $p \in (2,\infty)$ and $b \in (-p^{-1},\infty)\setminus\{0\}$ and define

$$I = \{x \in \mathbb{R} : bx < 1\} = \begin{cases} (-\infty, b^{-1}) & : \ b > 0 \\ (b^{-1},\infty) & : \ b < 0 \end{cases} \tag{6}$$

$$I^+ = \{x \in I : x > 0\} = \begin{cases} (0, b^{-1}) & : \ b > 0 \\ (0,\infty) & : \ b < 0 \end{cases} \tag{7}$$

$$I^- = \{x \in I : x < 0\} = \begin{cases} (-\infty, 0) & : \ b > 0 \\ (b^{-1},0) & : \ b < 0 \end{cases} \tag{8}$$

$$J = \left\{x \in \mathbb{R} : \frac{x}{p} > -1\right\} = (-p,\infty) \tag{9}$$

Since $b > -p^{-1} > -1/2$,

$$0 < \int_{I^+} x(1-bx)^{1/b}\, dx < \infty \qquad 0 < \int_0^\infty x\left(1+\frac{x}{p}\right)^{-p} dx < \infty$$

$$\int_{I^-} -x(1-bx)^{1/b}\, dx = \infty \qquad \int_{-p}^{0} -x\left(1+\frac{x}{p}\right)^{-p} dx = \infty$$

Therefore, there exist $\widehat{a}_{p,b}, a_{p,b} > 0$ such that

$$\left(\int_0^\infty x\left(1+\frac{x}{p}\right)^{-p} dx\right)^{-1} \int_{-a_{p,b}^{-1}}^{0} -x(1-bx)^{1/b}\, dx = 1 \tag{10}$$

$$\left(\int_{I^+} x(1-bx)^{1/b}\, dx\right)^{-1} \int_{-\widehat{a}_{p,b}^{-1}}^{0} -x\left(1+\frac{x}{p}\right)^{-p} dx = 1 \tag{11}$$

Let

$$R^\sharp_{p,b} = \left(\int_{I^+}^{\infty} (1-bx)^{1/b}\, dx\right)^{-1} \int_{-a_{p,b}^{-1}}^{0} \left(1+\frac{x}{p}\right)^{-p} dx \tag{12}$$

$$R^\flat_{p,b} = \left(\int_0^\infty \left(1+\frac{x}{p}\right)^{-p} dx\right)^{-1} \int_{-\widehat{a}_{p,b}^{-1}}^{0} (1-bx)^{1/b}\, dx \tag{13}$$

The quantities $a_{p,b}$, $\widehat{a}_{p,b}$, $R^{\flat}_{p,b}$ and $R^{\sharp}_{p,b}$ are continuous strictly monotonic functions of p and b that are (respectively) inc., dec., inc. and dec. in p and dec., inc., dec. and inc. in b, and satisfy

$$a_{p,b} \leq \widehat{a}_{p,b} \qquad R^{\flat}_{p,b} \leq R^{\sharp}_{p,b} \qquad \lim_{(p,b)\to(\infty,0)} R^{\flat}_{p,b} = \lim_{(p,b)\to(\infty,0)} R^{\sharp}_{p,b} = e$$

$$\widehat{a}_{p,-1/p} = a_{p,-1/p} = \frac{p}{p-1} \qquad R^{\sharp}_{p,-1/p} = R^{\flat}_{p,-1/p}$$

The next proposition applies quite generally. For example, if $f(x)$ is any of the following functions: $f(x) = c\exp\left(-\left(\log\left(1+|x|\right)\right)\log\log\log\left(C+|x|\right)\right)$, $f(x) = c_p \exp\left(-|x|^p\right)$, for $0 < p < \infty$, $f(x) = c\exp\left(-\exp\exp\left(x\right)\right)$, then setting $g = -\log f$, $\lim_{x\to\infty} g''(x)/\left(g'(x)^2\right) = 0$. If $f(x) = (1+|x|)^{-p}$, for $2 < p < \infty$, then $\lim_{x\to\infty} g''(x)/\left(g'(x)^2\right) = -1/p$. The density f in Proposition 2 is allowed to change its behavior back and forth, from ultra rapid decay to polynomial decay.

Proposition 2 *Let $\mu = \mu_1 + \mu_2$ be a probability measure on $\mathbb{R}$ such that $\int_{\mathbb{R}} |x|\, d\mu(x) < \infty$, with discrete and continuous components $\mu_1 \perp m$ and $\mu_2 \ll m$ respectively (with respect to Lebesgue measure m). Let $f = d\mu_2/dx$ and define $g : \mathbb{R} \to (-\infty, \infty]$ by $g = -\log f$. Let Y be a random variable with distribution μ. Let $t \in \mathbb{R}$ such that $t > \mathbb{E}Y$, and let $p > 2$, $b \in (-p^{-1}, \infty) \setminus \{0\}$. Let $a_{p,b}, \widehat{a}_{p,b}, R^{\sharp}_{p,b}, R^{\flat}_{p,b} > 0$ be defined as in (10)-(13). Suppose that $g(t) < \infty$, g is differentiable at t with $g'(t) > 0$, $\mu_1 \left[t - \widehat{a}_{p,b}/g'(t), \infty\right) = 0$, and for all $x \in \left[t - \widehat{a}_{p,b}/g'(t), \infty\right)$, $g''(x)$ exists and obeys*

$$-\frac{1}{p} \leq \frac{g''(x)}{\left(g'(x)\right)^2} \leq b \tag{14}$$

If φ_a is a minimizer as in Proposition 1, then $a_{p,b} \leq a \leq \widehat{a}_{p,b}$ and

$$R^{\flat}_{p,b}\mathbb{P}\left\{Y > t\right\} \leq \frac{\mathbb{E}\varphi_a(Y)}{\varphi_a(t)} \leq R^{\sharp}_{p,b}\mathbb{P}\left\{Y > t\right\}$$

Proof. Set $r = 1/g'(t)$ and consider the function $h(x) = g(t+rx) - g(t)$, in which case $h(0) = 0$, $h'(0) = 1$ and

$$\frac{h''(x)}{\left(h'(x)\right)^2} = \frac{g''(t+rx)}{\left(g'(t+rx)\right)^2}$$

Recall the definitions in (6)-(13). The function $q(x) = -\lambda^{-1}\log(1-\lambda x)$ satisfies

$$q''(x) = \lambda\left(q'(x)\right)^2 \qquad q(0) = 0 \qquad q'(0) = 1$$

and therefore by (14), setting $\lambda = b$ and then $\lambda = -1/p$,

$$\begin{aligned} h(x) &\leq -b^{-1}\log(1-bx) : x \in I \\ h(x) &\geq p\log\left(1+p^{-1}x\right) : x \in J \end{aligned}$$

This follows from monotonicity properties of autonomous differential equations, used to compare h' and q'. This implies

$$
\begin{aligned}
& \left(\int_0^\infty x\exp\left(-h(x)\right)dx\right)^{-1}\int_{-a_{p,b}^{-1}}^0 -x\exp\left(-h(x)\right)dx \\
\geq\ & \left(\int_0^\infty x\left(1+\frac{x}{p}\right)^{-p}dx\right)^{-1}\int_{-a_{p,b}^{-1}}^0 -x\left(1-bx\right)^{1/b}dx = 1
\end{aligned}
$$

and

$$
\begin{aligned}
& \left(\int_{I^+} x\exp\left(-h(x)\right)dx\right)^{-1}\int_{-\widehat{a}_{p,b}^{-1}}^0 -x\exp\left(-h(x)\right)dx \\
\leq\ & \left(\int_{I^+} x\left(1-bx\right)^{1/b}dx\right)^{-1}\int_{-\widehat{a}_{p,b}^{-1}}^0 -x\left(1+\frac{x}{p}\right)^{-p}dx = 1
\end{aligned}
$$

By Proposition 1, $\int_{t-a^{-1}}^\infty (x-t)\exp\left(-g(x)\right)dx = 0$, which can be written as

$$
\left(\int_0^\infty x\exp\left(-h(x)\right)dx\right)^{-1}\int_{-r^{-1}a^{-1}}^0 -x\exp\left(-h(x)\right)dx = 1
$$

Consider the function

$$
\zeta(s) = \left(\int_0^\infty x\exp\left(-h(x)\right)dx\right)^{-1}\int_{-r^{-1}s}^0 -x\exp\left(-h(x)\right)dx
$$

This function is continuous and strictly increasing in s, and by what we have shown above,

$$
\zeta\left(r\widehat{a}_{p,b}^{-1}\right)\leq 1 \qquad \zeta\left(a^{-1}\right) = 1 \qquad \zeta\left(ra_{p,b}^{-1}\right)\geq 1
$$

which implies that $r^{-1}a_{p,b}\leq a\leq r^{-1}\widehat{a}_{p,b}$. By Proposition 1 and the definition of $R_{p,b}^\sharp$,

$$
\begin{aligned}
\frac{\mathbb{E}\varphi_a\left(Y\right)}{\varphi_a(t)} &= \frac{\mathbb{P}\left\{Y>t-a^{-1}\right\}}{\mathbb{P}\left\{Y>t\right\}}\mathbb{P}\left\{Y>t\right\} \\
&= \mathbb{P}\left\{Y>t\right\}\left(\int_0^\infty \exp\left(-h(x)\right)dx\right)^{-1}\int_{-r^{-1}a^{-1}}^0 \exp\left(-h(x)\right)dx \\
&\leq R_{p,b}^\sharp\mathbb{P}\left\{Y>t\right\}
\end{aligned}
$$

and the lower bound follows similarly. ■

Going back to the setting where $\mathbb{E}\varphi\left(X\right)\leq\mathbb{E}\varphi\left(Y\right)$ for all (non-decreasing) convex φ, what we have shown is that provided Y has a sufficiently regular density, then the estimate

$$
\mathbb{P}\left\{X>t\right\}\leq\frac{\mathbb{E}\varphi_a\left(Y\right)}{\varphi_a(t)} \tag{15}
$$

comes close to the best possible estimate of $\mathbb{P}\{Y > t\}$. If $\lim_{x\to\infty} g''(x)/(g'(x))^2 = q$, for some $q \in (-1/2, 0]$, then replacing the estimator $\mathbb{E}\varphi_a(Y)/\varphi_a(t)$ with

$$\frac{\mathbb{E}\varphi_a(Y)}{R^\flat_{p(t),-1/p(t)}\varphi_a(t)}$$

where $p : (\mathbb{E}Y, \infty) \to (-1/2, 0)$ and $\lim_{t\to\infty} p(t) = q$, we recover $\mathbb{P}\{Y > t\}$ up to a factor of $(1 + o(1))$ as $t \to \infty$.

If Y does not have a sufficiently regular density and (14) does not hold, we may replace Y with another variable Y' such that $\mathbb{P}\{Y' > t\} \geq \mathbb{P}\{Y > t\}$ for all $t \in \mathbb{R}$ and then apply Proposition 2 to Y'. For example if $\mathbb{P}\{|Y| > t\} \leq \exp(-t^p)$ for all $t > 0$ then define Y' so that $\mathbb{P}\{Y' > t\} = \exp(-t^p)$ for all $t > 0$. Then $\mathbb{E}\varphi(X) \leq \mathbb{E}\varphi(Y) \leq \mathbb{E}\varphi(Y')$ (assuming φ non-decreasing) and we may apply Proposition 2 and (15) to X and Y'.

The following result is an example of how Pisier's version of Gaussian concentration (Theorem 9) may be combined with Theorem 2. The C^1 condition is not essential and can usually be weakened to f being locally Lipschitz (or even weaker) by applying standard smoothing techniques.

Theorem 3 *Let $n \in \mathbb{N}$, $T_0, A, b \in (0, \infty)$, $p \in (2, \infty)$, let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^1 function, and let X be a standard normal random vector. Let $\xi : [0, \infty) \to [0, \infty)$ be a continuous non-decreasing function, twice differentiable on $(0, \infty)$, and assume that*

$$-\frac{1}{p} \leq \frac{g''(x)}{(g'(x))^2} \leq b : \forall x \geq T_0 \qquad g(x) := -\log\left[-\frac{d}{dx}(A + 1/2)\exp\left(-\eta(x)^2/2\right)\right] \tag{16}$$

$$(A + 1/2)\exp\left(-\eta(T_0)^2/2\right) < 1/2$$

where $\eta : [0, \infty) \to [0, \infty)$ is the inverse of $t \mapsto 2^{-1}\pi t\xi(t)$, and assume that for all $t > 0$,

$$\mathbb{P}\{|\nabla f(X)| \geq \xi(t)\} \leq A\exp\left(-t^2/2\right)$$

Then for all $t > \max\{\mathbb{E}W, T_0\}$ (see 17) such that $t - \widehat{a}_{p,b}/g'(t) \geq T_0$, where $\widehat{a}_{p,b}$ is defined in (11),

$$\mathbb{P}\{|f(X) - \mathbb{M}f(X)| > t\} \leq 4(A + 1/2) R^\sharp_{p,b}\exp\left(-\eta(t)^2/2\right)$$

Proof. Let Y be an independent copy of X. By Theorem 1 there exists a standard normal random variable Z (in $\mathbb{R}$) such that

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

where φ_a is the minimizer from Proposition 1. Note that $\mathbb{P}\{2^{-1}\pi|\nabla f(X)| Z \geq 2^{-1}\pi s\xi(s)\} \leq (A + 1/2)\exp(-s^2/2)$. Let W be a random variable with

$$\mathbb{P}\{W \geq t\} = \left\{\begin{array}{ccc} \min\left\{1/2, (A + 1/2)\exp\left(-\eta(t)^2/2\right)\right\} & : & t > 0 \\ \mathbb{P}\{2^{-1}\pi|\nabla f(X)| Z \geq t\} & : & t \leq 0 \end{array}\right\} \geq \mathbb{P}\left\{2^{-1}\pi|\nabla f(X)| Z \geq t\right\} \tag{17}$$

W has a density given by $\exp(-g(t))$ for all $t \geq 0$ such that $(A+1/2)\exp\left(-\eta(t)^2/2\right) < 1/2$. Now consider any fixed $t > 0$ such that $t - \widehat{a}_{p,b}/g'(t) \geq T_0$. By the condition imposed on t, if $x \geq t - \widehat{a}_{p,b}/g'(t)$ then $x \geq T_0$ and the bounds in (16) apply. By Markov's inequality and Proposition 2,

$$\begin{aligned}
\mathbb{P}\{f(X) - f(Y) > t\} &\leq \frac{\mathbb{E}\varphi_a(f(X) - f(Y))}{\varphi_a(t)} \leq \frac{1}{\varphi_a(t)}\mathbb{E}\varphi_a\left(\frac{\pi}{2}|\nabla f(X)|Z\right) \\
&\leq \frac{1}{\varphi_a(t)}\mathbb{E}\varphi_a(W) \leq R^{\sharp}_{p,b}\mathbb{P}\{W > t\} \\
&\leq (A+1/2)R^{\sharp}_{p,b}\exp\left(-\eta(t)^2/2\right)
\end{aligned}$$

The result now follows because concentration about the median is equivalent to concentration about an independent random value,

$$\begin{aligned}
\frac{1}{2}\mathbb{P}\{f(X) - \mathbb{M}f(X) > t\} &\leq \mathbb{P}\{f(Y) \leq \mathbb{M}f(X)\}\mathbb{P}\{f(X) - \mathbb{M}f(X) > t\} \\
&\leq \mathbb{P}\{f(X) - f(Y) > t\}
\end{aligned}$$

■

2.2 Concentration inequalities

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

$$F^{-1}(s) = \inf\{t \in \mathbb{R} : F(t) > s\} \tag{18}$$

If U is a random variable uniformly distributed in $(0,1)$, then $F^{-1}(U)$ has distribution μ. If μ has a continuous density $h = d\mu/dx$ then for all points $s \in (0,1)$ and $t \in \mathbb{R}$,

$$Lip\left(F^{-1}, s\right) = \frac{1}{h(F^{-1}(s))} \qquad Lip\left(F^{-1} \circ \Phi, t\right) = \frac{\phi(t)}{h(F^{-1}(\Phi(t)))}$$

where $Lip(\cdot,\cdot)$ is the local Lipschitz constant as defined in (2). For a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ and $1 \leq s \leq \infty$ set

$$Lip_s(f) = \sup_{x \in \mathbb{R}^n} |\nabla f(x)|_s \qquad Lip^{\sharp}_s(f) = \left(\sum_{i=1}^n \sup_{x \in \mathbb{R}^n}\left|\frac{\partial f}{\partial x_i}(x)\right|^s\right)^{1/s} \tag{19}$$

with the usual interpretation (involving max) when $s = \infty$. In general, $Lip_s(f) \leq Lip^{\sharp}_s(f)$. There are two special cases when $Lip_s(f) = Lip^{\sharp}_s(f)$. The first is when f is linear. The second is when $s = \infty$. By considering a path integral, and a local linear approximation, it follows that $Lip_s(f)$ is the Lipschitz constant of f with respect to the $\ell^n_{s^*}$ norm on $\mathbb{R}^n$, where $s^* = s/(s-1)$ when $1 < s \leq \infty$ (setting $\infty/(\infty - 1) := 1$). We will also use $s^* = \infty$ when $0 < s < 1$. In Theorem 4 we study probability distributions

satisfying condition (20) below. This condition is satisfied (for example) in the following cases:
• $d\mu_i/dt = f_i(t) = (2\Gamma(1+1/q))^{-1} C_q \exp(-|C_q t|^q)$.
• X_i has a symmetric Weibull distribution with $\mathbb{P}\{|X_i| \geq t\} = \exp(-|C_q t|^q)$
• $X_i = C_q Z_i |Z_i|^{-1+2/q}$ or $X_i = C_q |Z_i|^{2/q}$ for a standard normal random vector $Z = (Z_i)_1^n$.

Theorem 4 (Weibull type tails $\sim \exp(-|t|^q)$, $0 < q < 1$) *Let $n \in \mathbb{N}$, $0 < q < 1$, and let $(\mu_i)_1^n$ be a sequence of probability measures on $\mathbb{R}$, each with corresponding cumulative distribution F_i and quantile function F_i^{-1} as in (18) such that for all $s \in \mathbb{R}$,*

$$Lip\left(F_i^{-1} \circ \Phi, s\right) \leq (1+|s|)^{-1+2/q} \tag{20}$$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable and let $(X_i)_1^n$ be a sequence of independent random variables, each with corresponding distribution μ_i. Then for all $t > 0$,

$$\mathbb{P}\{|f(X) - \mathbb{M}f(X)| > t\} \leq 2\exp\left(-c_q \min\left\{\left(\frac{t}{Lip_2^\sharp(f)}\right)^2, \left(\frac{t}{Lip_\infty^\sharp(f)}\right)^q\right\}\right) \tag{21}$$

and

$$\begin{aligned} & \mathbb{P}\left\{|f(X) - \mathbb{M}f(X)| > C_q\left(1 + \left(\log\frac{n}{t^{-2+4/q}}\right)^{1/q-1/2}\right)\left(tLip_2(f) + t^{2/q}Lip_\infty(f)\right)\right\} \\ \leq \;& 2\exp\left(-t^2/2\right) \end{aligned} \tag{22}$$

Comments for Theorem 4:
• For $q \in [1,2]$ related estimates can be proved using results of Talagrand [28, Theorem 2.4] and Gozlan [10, Proposition 1.2], which give bounds of the form

$$\mu^n\left(A + tB_2^n + t^{2/q}B_q^n\right) \geq 1 - \exp\left(-Dt^2\right) : t \geq 0$$

assuming that μ satisfies a certain Poincaré inequality on $\mathbb{R}$ with constant C, μ^n is the n-fold product of μ, and D is a constant depending only on C. This includes the case $d\mu/dx = (2\Gamma(1+1/q))^{-1}\exp(-|x|^q)$. Taking $A = \{x : f(x) \leq \mathbb{M}f(x)\}$, if $x \in A + tB_2^n + t^{2/q}B_q^n$ then there exists $a \in A$, $u \in B_2^n$, $v \in B_q^n$ such that $x = a + tu + t^{2/q}v$, so $|f(x) - f(a+tu)| \leq t^{2/q}Lip_{q^*}(f)$ and $|f(a+tu) - f(a)| \leq tLip_2(f)$, which implies $f(x) \leq \mathbb{M}f(x) + tLip_2(f) + t^{2/q}Lip_{q^*}(f)$. Together with a similar lower bound,

$$\mathbb{P}\{|f(X) - \mathbb{M}f(X)| > t\} \leq 2\exp\left(-c_q \min\left\{\left(\frac{t}{Lip_2(f)}\right)^2, \left(\frac{t}{Lip_{q^*}(f)}\right)^q\right\}\right)$$

which is the perfect nonlinear version of (28).
• In the case where f is linear, $f(x) = \sum_1^n a_i x_i$ where we may assume without loss of generality that $a \in S^{n-1}$. In this case it follows from symmetrization and the contraction principle (see for example the discussions in [1, 9]) that the basic assumption (20) may be weakened to $\mathbb{P}\{|X_i| \geq t\} \leq \exp(-|C_q t|^q)$, and the bound given by I recovers (29) which is optimal (up to the value of c_q) for all $a \in S^{n-1}$. As far as we are aware, even when f

is linear such an explicit probability bound (for *all* linear functionals) has not appeared before. We refer the reader to Section 3, see in particular (29), and [9], where we discuss a result of Hitczenko, Montgomery-Smith and Oleszkiewicz [12], and how to modify their result and prove related tail bounds in the linear case.

• The quantity

$$\left(\log \frac{n}{t^{-2+4/q}}\right)^{1/q-1/2}$$

cannot be completely erased from (22), otherwise this could be written as

$$\mathbb{P}\left\{|f(X) - \mathbb{M}f(X)| > t\right\} \leq 2\exp\left(-c_q \min\left\{\left(\frac{t}{Lip_2(f)}\right)^2, \left(\frac{t}{Lip_\infty(f)}\right)^q\right\}\right)$$

which would imply that $\mathrm{Var}\left(|X|_\infty\right) < C_q$, when in fact $\mathrm{Var}\left(|X|_\infty\right) \approx (\log n)^{-2+2/q}$. The current bound gives the estimate $\mathrm{Var}\left(|X|_\infty\right) \leq C_q (\log n)^{-1+2/q}$ which is off by a factor of $\log n$, exactly the same factor by which the classical Gaussian concentration inequality is off by in the case when $q = 2$ (specifically for the ℓ_∞^n norm). For certain functions it may be possible to decrease the exponent of the logarithmic term, say from $1/q - 1/2$ to $1/q - 1$, using Talagrand's L_1-L_2 inequality (the Gaussian version as in [8]), or by applying the methodology presented here to $\eta \circ f$ for an appropriate choice of η to achieve a superconcentrated estimate, which is a trick we have exploited in connection with Gaussian concentration of the ℓ_p^n norm (in an earlier version of this paper still available on arXiv).

• In case $|\nabla f(X)|_s$ is with high probability much smaller than $Lip_s(f)$, including the case when $Lip_s(f) = \infty$ and/or when one has a high probability bound on $\partial_i f(X)$, $1 \leq i \leq n$, one can prove variations of Theorem 4 (using a similar proof) that take the distribution of these into account.

Before stating a result under the assumption of polynomial tails, we need to discuss the following norm on $\mathbb{R}^n$. For any $r \in [1, \infty)$ and $q \in (1, \infty)$, let $E_{r,q} \subset \mathbb{R}^n$ be defined by

$$E_{r,q} = conv\left\{\max\left\{|u|_1, r|u|_q\right\}^{-1} u : u \in \{0, \pm 1\}^n, u \neq 0\right\}$$

The Minkowski functional of $E_{r,q}$ is the norm $|x|_{r,q} = \inf\{\lambda > 0 : x \in \lambda E_{r,q}\}$. It follows from the definition of $E_{r,q}$ that

• $|x|_{r,q} = \max\left\{|x|_1, r|x|_q\right\} = |x|_{B_1^n \cap r^{-1}B_q^n}$ for all $x \in \{0, \pm 1\}^n$.

• If $\|\cdot\|$ is any norm on $\mathbb{R}^n$ and $\|x\| \leq |x|_{r,q}$ for all $x \in \{0, \pm 1\}^n$, then $\|x\| \leq |x|_{r,q}$ for all $x \in \mathbb{R}^n$.

The dual Minkowski functional of $E_{r,q}$ is

$$\begin{aligned} |y|_{r,q}^\circ &= \sup\left\{\sum_{i=1}^n x_i y_i : x \in E_{r,q}\right\} \\ &\leq 2\sup\left\{r^{-1}k^{-1/q}\sum_{i=1}^k y^{(i)} : 1 \leq k \leq \min\left\{r^{q/(q-1)}, n\right\}\right\} \leq 2|y|_{r,q}^\circ \end{aligned}$$

where $y^{(1)} \geq y^{(2)} \ldots \geq y^{(n)} \geq 0$ is the non-increasing rearrangement of $(|y_i|)_1^n$. Since the canonical embedding of a normed space into its bidual is an isometry,

$$|x|_{r,q} = \sup \left\{ \sum_{i=1}^{n} x_i y_i : |y|_{r,q}^{\circ} \leq 1 \right\} \leq C_q \left(|x|_1 + r \sum_{i=1}^{n} i^{-1+1/q} x^{(i)} \right) \leq C_q' |x|_{r,q}$$

(we leave it to the reader to find y with $|y|_{r,q}^{\circ} \leq 1$ that maximizes $\sum x_i y_i$). $\mathbb{R}^n$ endowed with $|\cdot|_{r,q}$ is therefore isomorphic to a Lorentz space with constant independent of n. Using $E_{r,q} \subseteq B_1^n \cap r^{-1} B_q^n$ for the lower bound and Hölder's inequality for the upper bound (for the coordinates where $r i^{-1+1/q} > 1$), for all $x \in \mathbb{R}^n$,

$$\max \left\{ |x|_1, r|x|_q \right\} \leq |x|_{r,q} \leq C_q \left(1 + \log \min \left\{ r^{q/(q-1)}, n \right\}\right)^{(q-1)/q} \max \left\{ |x|_1, r|x|_q \right\}$$

Furthermore, these norms are equivalent with

$$\max \left\{ |x|_1, r|x|_q \right\} \leq |x|_{r,q} \leq A \max \left\{ |x|_1, r|x|_q \right\} \tag{23}$$

in all of the following cases:

• There exists $p > 1/q$ and $C_{p,q} > 0$ such that $x^{(i)} \leq C_{p,q} x^{(1)} i^{-p}$ for all i, with $A = C_{p,q}'$,

• There exists $0 < p < 1/q$ and $C_{p,q} \geq c_{p,q} > 0$ such that $c_{p,q} x^{(1)} i^{-p} \leq x^{(i)} \leq C_{p,q} x^{(1)} i^{-p}$ for all i, with $A = C_{p,q}'$,

• The empirical distribution of the coordinates of x (or a scalar multiple thereof) loosely approximates some given probability measure on $[0, \infty)$ with finite $(q+\varepsilon)^{th}$ moment: more generally and precisely, if $H_1(i/(n+1)) \leq x^{(n-i+1)} \leq H_2(i/(n+1))$ for all $1 \leq i \leq n$ and two non-decreasing functions $H_1, H_2 : (0,1) \to [0, \infty)$ such that

$$\int_0^1 H_1(t)\, dt > 0 \qquad \int_0^1 (1-t)^{-1+1/q} H_2(t)\, dt < \infty$$

then (23) holds with $A = C(q, H_1, H_2)$,

• $x \in \Theta_n$, with $A = C_q'$ (where $\Theta_n \subset S^{n-1}$ is a particular set, not depending on q, with $\sigma_n(\Theta_n) > 1 - 2\exp(-cn)$, and σ_n denotes normalized Haar measure on S^{n-1}). This can be shown by generating a random $\theta \in S^{n-1}$ as $\theta = Z/|Z|$, where Z is a standard normal random vector, and using the crude estimate $Z^{(i)} = \Phi^{-1}\left(1 - \exp\left(-X^{(i)}\right)\right) \leq C X^{(i)}$, where $(X_i)_1^n$ is an i.i.d. sample from the standard exponential distribution, and then using the Renyi representation of $\left(X_{(i)}\right)_1^n$, see the proof of Lemma 14, and estimating $\sum_1^n i^{-1+1/q} X^{(i)}$ by changing the order of a double sum and using (28).

In Theorem 5 we consider probability distributions satisfying condition (24). This condition is satisfied, for example, when μ_i has a density $f_i(t) = 2^{-1} q C_q (|C_q t| + 1)^{-q-1}$.

Theorem 5 (polynomial tails) *There exists a universal constant $C > 0$ such that the following is true. Let $n \in \mathbb{N}$, $2 < q < \infty$, $2q(q-2)^{-1} < p < \infty$, and let $(\mu_i)_1^n$ be a sequence of probability measures on $\mathbb{R}$, each with corresponding cumulative distribution F_i and quantile function F_i^{-1} as in (18) such that for all $s \in (0,1)$,*

$$Lip\left(F_i^{-1}, s\right) \leq \min\{s, 1-s\}^{-1-1/q} \tag{24}$$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable and let $(X_i)_1^n$ be a sequence of independent random variables, each with corresponding distribution μ_i. Let $Lip_s(f)$ be defined as in (19). Then for all $t > 0$,

$$\mathbb{P}\left\{|f(X) - \mathbb{M}f(X)| > C_q t \left| \left(\sup_{x \in \mathbb{R}^n} \left| \frac{\partial f}{\partial x_i}(x) \right|^2 \right)_{i=1}^n \right|_{r,q/2}^{1/2} \right\} \leq C \exp\left(-t^2/2\right) \tag{25}$$

where $r = Ct^2 \exp(t^2/q)$, $|\cdot|_{r,q/2}$ is the norm discussed above, $C_q > 0$ is a function of q, and

$$\mathbb{P}\left\{|f(X) - \mathbb{M}f(X)| > C_{p,q} Lip_p(f) \left(n^{1/2-1/p} t + n^{1/q} t \exp\left(\frac{t^2}{2q} \right) \right) \right\} \leq C \exp\left(-t^2/2\right) \tag{26}$$

where $C_{p,q} > 0$ is a function of (p, q).

Consider the special case of Theorem 5 where $f(x) = n^{-1/2} \sum_{i=1}^n x_i$, and (25) implies

$$\mathbb{P}\left\{|f(X) - \mathbb{M}f(X)| > C_q t \left(1 + n^{1/q-1/2} t \exp\left(\frac{t^2}{2q}\right)\right)\right\} \leq C \exp\left(-t^2/2\right)$$

which is sub-Gaussian up to probability $Cn^{1-q/2}(\log n)^{q/2}$ and matches the variance up to the factor C_q. In this case (and whenever f is linear), we may replace the basic assumption (24) with $\mathbb{P}\{|X_i| > t\} \leq c_q (1+t)^{-q}$ (see the discussion in Section 3). Compare this with the non-uniform version of the Berry-Esseen bound, Theorem 8, that gives a sub-Gaussian bound up to probability $C_r n^{-1/2} (\log n)^{-r/2}$ under the assumption that $\mathbb{E}|X_i|^r < C_r$ $(3 \leq r < \infty)$.

In Section 3 (see in particular Theorems 10, 11 and 12) we state concentration inequalities of Latała and Oleszkiewicz [16], Schechtman and Zinn [26], and Barthe, Cattiaux and Roberto for functions $f : \mathbb{R}^n \to \mathbb{R}$, Lipschitz with respect to the ℓ_2^n norm, sitting on measures with thicker-than-Gaussian tails (on the appropriate scale). By taking f to be a coordinate functional, we see that these inequalities cannot be improved to give sub-Gaussian bounds. However, for 'most' Lipschitz functions one can in fact obtain sub-Gaussian bounds. If there is such a thing as a typical Lipschitz function in a coordinate free sense, it is certainly covered by the case of a typical rotation of an arbitrary Lipschitz function.

Theorem 6 (typical Lipschitz function) *Let $n \in \mathbb{N}$, $p \in (1, \infty)$ and $q \in (4, \infty)$, such that $q/(q-2) < p < q/2$, and let $g : \mathbb{R}^n \to \mathbb{R}$ be a 1-Lipschitz function with respect to the standard Euclidean metric on $\mathbb{R}^n$, i.e. $|g(x) - g(y)| \leq |x - y|$ for all $x, y \in \mathbb{R}^n$. Let $(\mu_i)_1^n$ be a sequence of probability measures on $\mathbb{R}$, each with center of mass at 0, cumulative distribution F_i and quantile function F_i^{-1} as in (18) such that for all $t \in (0, 1)$,*

$$Lip\left(F_i^{-1}, t\right) \leq \min\{t, 1-t\}^{-1-1/q} \tag{27}$$

Let σ_n denote Haar measure on $SO(n)$, normalized so that $\sigma_n(SO(n)) = 1$, and for any $U \in SO(n)$ let $g_U = g \circ U$. Then there exists a Borel subset $E \subseteq SO(n)$ with

$\sigma_n(E) \geq 0.99$ *such that any* $U \in E$ *has the following property: Let* $X = (X_i)_1^n$ *be a sequence of independent random variables, where the distribution of* X_i *is* μ_i*. Then for all* $\lambda \geq 0$ *such that*

$$\exp\left(\frac{\lambda^2}{2q}\right) \leq \min\left\{\lambda^{-1} n^{1/2-1/(2p)-1/q}, n^{1/(2p)-1/q}\right\}$$

we have

$$\mathbb{P}\left\{|g_U(X) - \mathbb{M}g_U(X)| \leq C_{p,q}\lambda\right\} \geq 1 - C\exp\left(-\lambda^2/2\right)(\log(2+\lambda))(\log\log(3+\lambda))^2$$

where $C > 0$ *is a universal constant and* $C_{p,q} > 0$ *is a continuous function of* (p,q)*.*

Comments for Theorem 6:

• The condition in (27) holds when (for example) $d\mu_i/dx = 2^{-1}qC_q(1+|C_q t|)^{-q-1}$. One can prove versions of this result under more rapid tail decay, say $d\mu_i/dx = c_q \exp(-|t|^q)$ for $0 < q < \infty$, and the sub-Gaussian bounds then go deeper into the tails.

• Setting $p = 2(\log n)/(\log n - \log\log n)$, this sub-Gaussian bound goes up to probability

$$C_q n^{-q/4+1}(\log n)^{q/4}(\log\log n)(\log\log\log n)^2$$

The scale of this sub-Gaussian bound matches the known variance of linear functionals $g(x) = \langle x, \theta\rangle$, $\theta \in S^{n-1}$, up to the value of $C_{p,q}$.

• This probability measures randomness due to X, and U is considered fixed. $\mathbb{M}g_U(X)$ can be replaced with a single quantity that does not depend on U. One can also prove a version (which is actually part of the proof) where U is not considered fixed and $\mathbb{P}\{\cdot\}$ measures randomness due to X and U jointly. One can also prove a version where the quantile is fixed and the set $E \subseteq SO(n)$ is chosen depending on this quantile. These different versions will all have slightly different bounds, and we refer the reader to the existing proof for these.

• The value 0.99 is arbitrary and can be replaced with any constant in $(0,1)$.

3 Background

Let $0 < q < \infty$, let $a \in \mathbb{R}^n$ be nonzero, and let $(X_i)_1^n$ be a sequence of independent random variables with $\mathbb{E}X_i = 0$ and $\mathbb{P}\{|X_i| \geq t\} \leq 2\exp(-t^q)$ for all $t > 0$. If $1 \leq q < \infty$ then by the exponential moment method (see for example [5, Exercise 2.27 p.50]),

$$\mathbb{P}\left\{\left|\sum_{i=1}^n aX_i\right| > t\right\} \leq 2\exp\left(-c_q \min\left\{\left(\frac{t}{|a|}\right)^2, \left(\frac{t}{|a|_{q/(q-1)}}\right)^q\right\}\right) \tag{28}$$

where in the case $q = 1$ we define $1/0 = \infty$. For $2 < q < \infty$ there are some directions in which this can be improved slightly, but this is not the case of main interest here. For $0 < q < 1$ the exponential moment method does not apply and it follows from a recent result in [9] that

$$\mathbb{P}\left\{\left|\sum_{i=1}^n aX_i\right| > t\right\} \leq 2\exp\left(-c_q \min\left\{\left(\frac{t}{|a|}\right)^2, \left(\frac{t}{|a|_\infty}\right)^q\right\}\right) \tag{29}$$

For certain directions, e.g. $a_i = 1$, 29 was known earlier, and follows from moment estimates of Hitczenko, Montgomery-Smith and Oleszkiewicz [12, Theorem 1.1] (see the discussion in [9]). The general formula for arbitrary nonzero $a \in \mathbb{R}^n$ seems to be new. Note that (29) is a very natural extension of (28) considering that the dual of ℓ_q^n is $\ell_{q/(q-1)}^n$ for $1 \leq q \leq \infty$ (with appropriate interpretations at 1 and ∞), while the dual is ℓ_∞^n for $0 < q < 1$.

Consider the Berry-Esseen and non-uniform Berry-Esseen bounds; see [21, Chapter V §2-4 Theorems 3 and 13] for these and numerous other variations.

Theorem 7 *Let $a \in S^{n-1}$ and let $(X_i)_1^n$ be a sequence of independent random variables such that for all $1 \leq i \leq n$, $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = 1$, and $\mathbb{E}|X_i|^3 < \infty$. Then*

$$\sup_{t \in \mathbb{R}} \left| \Phi(t) - \mathbb{P}\left\{ \sum_{i=1}^n a_i X_i \leq t \right\} \right| \leq C \sum_{i=1}^n |a_i|^3 \, \mathbb{E}|X_i|^3$$

Theorem 8 *Let $r \in [3, \infty)$, and let $(X_i)_1^n$ be an i.i.d. sequence with $\mathbb{E}X_1 = 0$, $\mathbb{E}X_1^2 = 1$, and $\mathbb{E}|X_1|^r < \infty$. Then for all $x \in \mathbb{R}$,*

$$\left| \Phi(x) - \mathbb{P}\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \leq x \right\} \right| \leq C_r \left(1 + |x|\right)^{-r} \left(n^{-1/2} \mathbb{E}|X_1|^3 + n^{-(r-2)/2} \mathbb{E}|X_1|^r \right)$$

Assuming that $\mathbb{E}|X_1|^r < C_r$ and $n > C_r$, this gives a sub-Gaussian bound on $n^{-1/2} \sum X_i$ up to probability $C_r (\log n)^{-r/2} n^{-1/2}$.

The following result is Pisier's version of the Gaussian concentration inequality, see [24, Theorem 2.2 p176]. We have taken some liberty in the wording, and (31) follows from the Gaussian isoperimetric inequality of Sudakov and Tsirelson [27] and Borell [4] (see also [3]). The C^1 condition is of course not essential.

Theorem 9 *There exist universal constants $C, c > 0$ such that the following is true. Let $n \in \mathbb{N}$, let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^1 function, and let X and Y be independent standard normal random vectors. Then there exists a random variable Z with the standard normal distribution on $\mathbb{R}$, independent of X, such that for any convex function $\varphi : \mathbb{R} \to \mathbb{R}$,*

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

In the special case where f is Lipschitz, this implies that for all $t > 0$

$$\mathbb{P}\left\{ \left| f(X) - \mathbb{M}f(X) \right| > t \right\} \leq C \exp\left(-\frac{ct^2}{Lip(f)^2} \right) \tag{30}$$

This can be sharpened to

$$\left. \begin{array}{l} \mathbb{P}\left\{ f(X) > \mathbb{M}f(X) + Lip(f)\, t \right\} \\ \mathbb{P}\left\{ f(X) < \mathbb{M}f(X) - Lip(f)\, t \right\} \end{array} \right\} \leq 1 - \Phi(t) \tag{31}$$

The following two results are due to Latała and Oleszkiewicz [16, Theorems 1 and 2] (as formulated in [26, Theorem 4.2]), and Schechtman and Zinn [26, Theorems 3.1 and 4.1] (respectively).

Theorem 10 *There exist universal constants $C, c > 0$ such that if $1 \leq p \leq 2$, $f : \mathbb{R}^n \to \mathbb{R}$ satisfies $|f(x) - f(y)| \leq |x - y|$ for all $x, y \in \mathbb{R}^n$, and X is a random vector in $\mathbb{R}^n$ with probability density (with respect to Lebesgue measure) $d\mu/dx = c_p^n \exp\left(-|x|_p^p\right)$, then for all $t > 0$*

$$\mathbb{P}\{|f(X) - \mathbb{E}f(X)| > t\} \leq C\exp(-ct^p)$$

Theorem 11 *There exist universal constants $C, c > 0$ such that if $1 \leq p \leq 2$, $f : \partial B_p^n \to \mathbb{R}$ satisfies $|f(x) - f(y)| \leq |x - y|$ for all $x, y \in \partial B_p^n$, and X is a random vector on ∂B_p^n distributed according to normalized cone measure, i.e. for all Borel sets $E \subseteq \partial B_p^n$,*

$$\mathbb{P}\{X \in E\} = \frac{\mathrm{vol}_n\{r\theta : 0 \leq r \leq 1, \theta \in E\}}{\mathrm{vol}_n\left(B_p^n\right)}$$

then for all $t > 0$,

$$\mathbb{P}\{|f(X) - \mathbb{E}f(X)| > t\} \leq C\exp(-cnt^p)$$

The following result is due to Barthe, Cattiaux and Roberto [2, Example 5.3] (see Theorem 5.1 in their paper for a more general result).

Theorem 12 *Let X be a random vector in $\mathbb{R}^n$ with probability density (with respect to Lebesgue measure) $d\mu/dx = \alpha^n 2^{-n} \prod_{i=1}^n (1 + |x_i|)^{-1-\alpha}$, and let $f : \mathbb{R}^n \to \mathbb{R}$ satisfy $|f(x) - f(y)| \leq |x - y|$ for all $x, y \in \mathbb{R}^n$. Then there exists $t_0(\alpha) > e$ and $C(\alpha) > 0$ such that for all $t \geq t_0(\alpha)$,*

$$\mathbb{P}\left\{|f(X) - \mathbb{M}f(X)| > tn^{1/\alpha}\right\} \leq C(\alpha)\left(\frac{\log t}{t}\right)^{\alpha}$$

4 Order statistics

An i.i.d. random sample $(X_i)_1^n$ from any probability measure μ in $\mathbb{R}$ with CDF $F(t) = \mu(-\infty, t]$ and quantile function F^{-1} as defined in (18) can be represented as $X_i = F^{-1}(\gamma_i)$, where $(\gamma_i)_1^n$ is an i.i.d. random sample from the uniform distribution on $(0, 1)$. Since F^{-1} is non-decreasing, the order statistics (non-decreasing rearrangement) $\left(X_{(i)}\right)_1^n$ may be written as $X_{(i)} = F^{-1}\left(\gamma_{(i)}\right)$.

Lemma 13 *If we define $\xi_1 : [0, 1] \to [0, 1]$ and $\xi_2 : [0, \infty) \to (0, 1]$ by*

$$\xi_1(t) = e^t(1 - t) \qquad \xi_2(t) = e^{-t}(1 + t) \tag{32}$$

then

$$\begin{aligned}
\xi_1^{-1}(t) &\leq \min\left\{\sqrt{2(1-t)}, 1 - e^{-1}t\right\} : 0 \leq t \leq 1 \\
\xi_2^{-1}(t) &\leq \begin{cases} \log t^{-1} + \log\left(1 + 4\log t^{-1}\right) & : \quad 0 < t \leq 2e^{-1} \\ \sqrt{2\log t^{-1} + 10\left(\log t^{-1}\right)^{3/2}} & : \quad 2e^{-1} \leq t \leq 1 \end{cases}
\end{aligned}$$

Proof. The estimates for ξ_1^{-1} follow since $\xi_1(t) \leq \min\{1-t^2/2, e(1-t)\}$. To estimate ξ_2^{-1} we re-write $y = e^{-t}(1+t)$ as $z = t - \log(1+t)$, where $z = \log y^{-1}$. If $z < 1 - \log 2$ then $t < 1$, since $t \mapsto t - \log(1+t)$ is strictly increasing. Since $\log(1+t) = \sum_1^\infty (-1)^{j+1} j^{-1} t^j$ is alternating, with terms that decrease in absolute value, $z = t - \log(1+t) \geq t - (t - t^2/2 + t^3/3) \geq t^2/6$. But then $z = t - \log(1+t) \geq t - (t - t^2/2 + t^3/3)$ and so $t^2/2 \leq z + t^3/3 \leq z + 2\sqrt{6} z^{3/2}$. If $z \geq 1 - \log 2$ then $t \geq 1$ and $\log(1+t) \leq t \log(2)$ so $z = t - \log(1+t) \geq (1 - \log(2))\, t$ and $t \leq (1-\log(2))^{-1} z$. But then $t = z + \log(1+t) \leq z + \log\left(1 + (1-\log(2))^{-1} z\right)$. ■

The bounds in the following lemma are natural, in light of the central limit theorem and the large deviation behavior of a sum of exponential random variables (which is governed by ξ_1 and ξ_2), and the fact that the uniform empirical process on $[0,1]$ converges to a Brownian bridge. A different result with log replaced by log log can also be proved, and the logs can be removed entirely if one only wants estimates for individual order statistics.

Lemma 14 *Let $(\gamma_i)_1^n$ be an i.i.d. sample from $(0,1)$ with corresponding order statistics $\left(\gamma_{(i)}\right)_1^n$ and let $t > 0$. With probability at least $1 - 3^{-1}\pi^2 \exp(-t^2/2)$, the following event occurs: for all $1 \leq k \leq n$, $\gamma_{(k)}$ is bounded above by both of the following quantities*

$$\frac{k}{n+1}\left(1 + \xi_2^{-1}\left(\exp\left(\frac{-t^2 - 4\log k}{2k}\right)\right)\right) \tag{33}$$

$$1 - \frac{n-k+1}{n+1}\left(1 - \xi_1^{-1}\left(\exp\left(\frac{-t^2 - 4\log(n-k+1)}{2(n-k+1)}\right)\right)\right) \tag{34}$$

and with probability at least $1 - C\exp(-t^2/2)$ the following event occurs: for all $1 \leq k \leq n$,

$$\begin{aligned}
\gamma_{(k)} &\leq 1 - \frac{n-k}{n}\exp\left(-c\max\left\{\frac{\left(t+\sqrt{\log k}\right)\sqrt{k}}{\sqrt{n(n-k+1)}}, \frac{t^2 + \log k}{n-k+1}\right\}\right) && (35)\\
&\leq \frac{k}{n} + c\frac{n-k}{n}\max\left\{\frac{\left(t+\sqrt{\log k}\right)\sqrt{k}}{\sqrt{n(n-k+1)}}, \frac{t^2+\log k}{n-k+1}\right\}
\end{aligned}$$

For $k \leq n/2$, (33) gives a typical deviation about the mean at most $C\sqrt{k\log k}/n$ but breaks down as $t \to \infty$ and n, k are fixed. For $k \geq n/2$ (34) gives a typical deviation at most $C\sqrt{(n-k+1)\log(n-k+1)}/n$, and remains non-trivial (i.e. < 1) for all $1 \leq k \leq n$ as $t \to \infty$. For $k \leq n/2$ (35) also gives a typical deviation of $C\sqrt{k\log k}/n$: it is not quite as precise as (33) (which includes the exact function ξ_2) for $0 < t < t_{n,k}$ but eventually improves upon (33) and remains non-trivial as $t \to \infty$.

Proof. If B has a binomial distribution with parameters (n, p), and $np \leq s < n$, then using the exponential moment method,

$$\mathbb{P}\{B \geq s\} = \mathbb{P}\left\{e^{\lambda B} \geq e^{\lambda s}\right\} \leq e^{-\lambda s}\left(1 - p + pe^{\lambda}\right)^n = \left(\frac{np}{s}\right)^s \left(\frac{n-np}{n-s}\right)^{n-s}$$

See e.g. [5, Ex. 2.11 p48]. Let $\#(E)$ denote the number of $1 \le i \le n$ such that $\gamma_i \in E$. Then (recycling the variable s),

$$\begin{aligned}
\mathbb{P}\left\{\gamma_{(k)} \ge \frac{k+s\sqrt{k}}{n+1}\right\} &= \mathbb{P}\left\{\#\left(\frac{k+s\sqrt{k}}{n+1},1\right) \ge n-k+1\right\} \\
&\le \left(1-\frac{s\sqrt{k}}{n-k+1}\right)^{n-k+1}\left(1+\frac{s}{\sqrt{k}}\right)^{k-1}\left(\frac{k}{k-1}\right)^{k-1} \div \left(\frac{n+1}{n}\right)^{n} \\
&\le \left(\xi_2\left(\frac{s}{\sqrt{k}}\right)\right)^{k} \le \frac{\exp\left(-t^2/2\right)}{k^2}
\end{aligned}$$

provided

$$s \ge \sqrt{k}\xi_2^{-1}\left(k^{-2/k}\exp\left(\frac{-t^2}{2k}\right)\right)$$

We then apply the union bound over all $1 \le k \le n$. (34) follows the same lines:

$$\mathbb{P}\left\{\gamma_{(k)} \ge \frac{k+s\sqrt{n-k+1}}{n+1}\right\} = \mathbb{P}\left\{\#\left(\frac{k+s\sqrt{n-k+1}}{n+1},1\right) \ge n-k+1\right\}$$

To prove (35), we make use of the Rényi representation of order statistics from the exponential distribution (which we heard of from [6, Theorem 2.5]): there exist i.i.d. standard exponential random variables $(Z_j)_1^n$ such that

$$\log\left(1-\gamma_{(k)}\right)^{-1} = \sum_{j=1}^{k} \frac{Z_j}{n-j+1}$$

(this is an easy consequence of the fact that for all $1 \le k \le n$, the order statistics $\left(\gamma_{(j)}\right)_{k+1}^{n}$ are (after being re-scaled to fill $(0,1)$) independent of $\left(\gamma_{(j)}\right)_1^k$ and distributed as the order statistics from a sample of size $n-k$. Thus we may write

$$1-\gamma_{(k)} = \left(1-\gamma_{(1)}\right)\prod_{j=2}^{k}\left(1-\gamma_{(j)}\right)\left(1-\gamma_{(j-1)}\right)^{-1}$$

which is the product of k independent variables). Concentration of $\log\left(1-\gamma_{(k)}\right)^{-1}$ about its mean (with probability $1-Ck^{-2}\exp\left(-t^2/2\right)$) can now be studied using the basic estimate (28) with $q=1$, and the result transferred back to $\gamma_{(k)}$ using the transformation $t \mapsto 1-\exp(-t)$. ■

Corollary 15 *Let $n \in \mathbb{N}$, $1 \le k < (n+1)/2$, $\lambda \in [2,\infty)$, and let $(Y_i)_1^n$ be an i.i.d. sequence of non-negative random variables, each with cumulative distribution F and quantile function F^{-1} as defined in (18), and corresponding order statistics $\left(Y_{(i)}\right)_1^n$. With probability at least $1-3^{-1}\pi^2\exp\left(-\lambda^2/2\right)$, the following inequalities hold:*

$$Y_{(\lfloor (n+1)/2 \rfloor)} \le F^{-1}\left(1-\frac{1}{12}\exp\left(-\frac{\lambda^2}{n+1}\right)\right)$$

$$\begin{aligned}\sum_{i=1+\lfloor (n+1)/2 \rfloor}^{n-k} Y_{(i)} &\leq (n+1) \int_{k/(n+1)}^{1/2} F^{-1}\left(1 - t\left(1 - \xi_1^{-1}\left(\exp\left(\frac{-\lambda^2 - 4\log((n+1)t)}{2(n+1)t}\right)\right)\right)\right) dt \\ &\leq (n+1) \int_{k/(n+1)}^{1/2} F^{-1}\left(1 - e^{-1-2/e} t \exp\left(\frac{-\lambda^2}{2(n+1)t}\right)\right) dt\end{aligned}$$

$$Y_{(n)} \leq F^{-1}\left(1 - \frac{1}{e(n+1)} \exp\left(-\frac{\lambda^2}{2}\right)\right)$$

Proof. This follows from Lemmas 14 and 13 applied to the random vector $\left(\gamma_{(i)}\right)_1^n$, where $(\gamma_i)_1^n$ is an i.i.d. sample from the uniform distribution on $(0,1)$. We write $Y_{(i)} = F^{-1}\left(\gamma_{(i)}\right)$ and compare the sum to an integral using monotonicity of the integrand. The condition $\lambda \geq 2$ is included because $x \mapsto \left(\lambda^2 + 4\log x\right)/x$ is decreasing provided $\log x \geq 1 - \lambda^2/4$. We also use the fact that $m^{1/m} \leq 3^{1/3}$ for $m = 1, 2, 3 \ldots$, and $2e3^{2/3} \leq 12$. ■

Example 16 *If in Corollary 15 we assume that the tail probabilities satisfy $1 - F_i(t) \leq t^{-p}$ with $p \in (1, \infty)$ (we do not need identically distributed variables), now assuming $1 \leq k \leq (n+1)/4$ (a relatively superficial condition), then for all $\lambda > 0$,*

$$\mathbb{P}\left\{\sum_{i=1}^{n-k+1} Y_{(i)} < \frac{Cpn}{p-1} + Cn^{1/p}k^{-1/p}\left(1 + pk^2\lambda^{-2}\right)\exp\left(\frac{\lambda^2}{2pk}\right)\right\} > 1 - C\exp\left(-\lambda^2/2\right) \tag{36}$$

We write $\sum_{i=1}^{n-k+1} Y_{(i)} = Y_{(n-k+1)} + \sum_{i=1}^{n-k} Y_{(i)}$, estimating $Y_{(n-k+1)}$ using Lemmas 14 and 13, and estimating $\sum_{i=1}^{n-k} Y_{(i)}$ using Corollary 15. This covers the case where $\lambda \gg k$ and $Y_{(n-k+1)}$ dominates (and the integral approximation of $\sum_{i=1}^{n-k+1} Y_{(i)}$ is not optimal, but the integral approximation of $\sum_{i=1}^{n-k} Y_{(i)}$ is absorbed into $Y_{(n-k+1)}$), as well as the case $\lambda \ll k$ where the mass of $\sum_{i=1}^{n-k+1} Y_{(i)}$ is not concentrated on any single term and the integral approximation of $\sum_{i=1}^{n-k} Y_{(i)}$ is good). Setting $k = 1$,

$$\mathbb{P}\left\{\sum_{i=1}^{n} Y_i < \frac{Cpn}{p-1} + Cn^{1/p}\exp\left(\frac{\lambda^2}{2p}\right)\left(1 + p\lambda^{-2}\right)\right\} > 1 - C\exp\left(-\lambda^2/2\right)$$

This gives the correct order of magnitude for $\sum_1^n Y_i$ in the i.i.d. case up to a factor of $C\left(1 + p\lambda^{-2}\right)$, since the same bound describes the order of magnitude of $n\mathbb{E}Y_1 + \max_{1 \leq i \leq n} Y_i$. Setting $s^k = C\exp\left(\lambda^2/2\right)$, (36) can be written as

$$\mathbb{P}\left\{\sum_{i=1}^{n-k+1} Y_{(i)} < \frac{Cpn}{p-1} + Cn^{1/p}k^{-1/p}\left(1 + \frac{pk}{\log s}\right)s^{1/p}\right\} > 1 - s^{-k}$$

Compare this to the following bound [11, Lemma 4.4]: for all $s \in (1, \infty)$,

$$\mathbb{P}\left\{\sum_{i=1}^{n-k+1} Y_{(i)} \leq \frac{12p\,(es)^{1/p}}{p-1} n\right\} > 1 - s^{-k}$$

Similar computations can be carried out when $p \in (0, 1]$.

5 Proofs for Section 2.2

The following lemma will be used implicitly several times.

Lemma 17 *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 $t = \min\{f^{-1}(s), g^{-1}(s)\}$.*

Proof of Theorem 4. Without loss of generality we may assume that each F_i^{-1} is differentiable. Let $Tx = \left(\left(F_i^{-1}\Phi(x_i)\right)_{i=1}^n\right)$, and let Z be a random vector in $\mathbb{R}^n$ with the standard normal distribution. Then

$$\begin{aligned}
|\nabla(f \circ T)(Z)| &= \left(\sum_{i=1}^n f_i(TZ)^2 \, Lip\left(F_i^{-1}\Phi, Z_i\right)^2\right)^{1/2} \leq \left(\sum_{i=1}^n f_i(TZ)^2 (1+|Z_i|)^{-2+4/q}\right)^{1/2} \\
&\leq \left(\sum_{i=1}^n \sup_{x\in\mathbb{R}^n} |f_i(x)|^2 \, \mathbb{E}(1+|Z_i|)^{-2+4/q}\right)^{1/2} \\
&\quad + \left|\sum_{i=1}^n \sup_{x\in\mathbb{R}^n} |f_i(x)|^2 \left\{(1+|Z_i|)^{-2+4/q} - \mathbb{E}(1+|Z_i|)^{-2+4/q}\right\}\right|^{1/2}
\end{aligned}$$

Since $0 < q < 1$, for all $t > 0$,

$$\mathbb{P}\left\{\left|(1+|Z_i|)^{-2+4/q} - \mathbb{E}(1+|Z_i|)^{-2+4/q}\right| > t\right\} \leq 2\exp\left(-c_q t^{q/(2-q)}\right)$$

Noting that $0 < 2/(2-q) < 1$ and using (29), with probability at least $1 - 2\exp(-t^2/2)$,

$$\begin{aligned}
&\left|\sum_{i=1}^n \sup_{x\in\mathbb{R}^n} |f_i(x)|^2 \left\{(1+|Z_i|)^{-2+4/q} - \mathbb{E}(1+|Z_i|)^{-2+4/q}\right\}\right|^{1/2} \\
\leq\; & C_q\left(t^{1/2} Lip_4^\sharp(f) + t^{(2-q)/q} Lip_\infty^\sharp(f)\right)
\end{aligned}$$

When adding this to the remaining term of $C_q Lip_2^\sharp(f)$, the term involving $Lip_4^\sharp(f)$ can be erased since by Hölder's inequality $|\cdot|_4 \leq |\cdot|_2^{1/2} |\cdot|_\infty^{1/2}$ and since we may assume that $t \geq 1$, $t^{1/2} \leq (1)^{1/2}\left(t^{(2-q)/q}\right)^{1/2}$. What we have shown is that with probability at least $1 - 2\exp(-t^2/2)$,

$$|\nabla(f \circ T)(Z)| \leq C_q Lip_2^\sharp(f) + C_q t^{(2-q)/q} Lip_\infty^\sharp(f)$$

Replacing f with $f/Lip_2^\sharp(f)$ we may assume that $Lip_2^\sharp(f) = 1$. Since $q < 2$, $Lip_\infty^\sharp(f) \leq Lip_2^\sharp(f)$. We will now apply Theorem 3. Defining

$$g(t) = C_{q,f} + \frac{1}{2}\left\{t\left(\frac{t}{Lip_\infty^\sharp(f)}\right)^{q/2}\right\}^2 \left(t + \left(\frac{t}{Lip_\infty^\sharp(f)}\right)^{q/2}\right)^{-2} \approx \min\left\{t, \left(\frac{t}{Lip_\infty^\sharp(f)}\right)^{q/2}\right\}^2$$

where $\approx$ means the same order of magnitude up to a constant factor and in abuse of notation only applies to the two terms on either side (i.e. does not include $C_{q,f}$), and $C_{q,f}$ (which actually depends only on q and $Lip_\infty^\sharp(f) \in (0,1]$) is chosen so that

$$\frac{5}{2}\int_0^\infty \exp\left(-g\left(s\right)\right) ds = 1$$

We will use Theorem 3 with $A = 2$, where 2 comes from the probability bound above, so $A + 1/2 = 5/2$. This is chosen so that (37) below matches (16) of Theorem 3. As $Lip_\infty^\sharp(f) \to 0$,

$$\frac{1}{2}\left\{t\left(\frac{t}{Lip_\infty^\sharp(f)}\right)^{q/2}\right\}^2\left(t+\left(\frac{t}{Lip_\infty^\sharp(f)}\right)^{q/2}\right)^{-2} \to \frac{t^2}{2}$$

and therefore $c_q \le C_{q,f} \le C_q$. By direct differentiation, if $0 < r < 1$ and $h(x) = \left(Bx^{1+r}/\left(x+Bx^r\right)\right)^2/2$, and for all $x > 1$ (say),

$$\begin{aligned}\left|\frac{h''(x)}{\left(h'(x)\right)^2}\right| &= \left|\frac{\left(1+B^{-1}x^{1-r}\right)^2\left[1-B^{-1}x^{1-r}\left(r^2-5r+2\right)+B^{-2}x^{2-2r}r\left(2r-1\right)\right]}{x^2\left(1+rB^{-1}x^{1-r}\right)^2}\right| \\ &\le \left|\frac{1-B^{-1}x^{1-r}\left(r^2-5r+2\right)+B^{-2}x^{2-2r}r\left(2r-1\right)}{x^2}\right|\end{aligned}$$

This implies that there exists $C_q' > 0$, independent of $Lip_\infty^\sharp(f)$, such that $\left|g''(x)/g'(x)^2\right| \le 1/4$ for all $x \ge C_q'$. Defining $\eta : [0,\infty) \to [0,\infty)$ by

$$\eta(x) = \left(-2\log\left(1-\frac{5}{2}\int_0^x \exp\left(-g(s)\right) ds\right)\right)^{1/2}$$

we see that η is continuous, strictly increasing, and satisfies

$$-\frac{d}{dx}\frac{5}{2}\exp\left(-\frac{1}{2}\eta(x)^2\right) = \exp\left(-g(x)\right) \qquad \eta(0) = 0 \qquad c_q g(x)^{1/2} \le \eta(x) \le C_q g(x)^{1/2} \tag{37}$$

uniformly over all possible values of $Lip_\infty^\sharp(f)$. The equation holds for all $x > 0$ while the inequalities hold for all $x > 1$. Defining $\xi : [0,\infty) \to [0,\infty)$ so that $t \mapsto 2^{-1}\pi t\xi(t)$ is the inverse of η, we see that

$$\xi(x) \ge C_q Lip_2^\sharp(f) + C_q t^{(2-q)/q} Lip_\infty^\sharp(f)$$

and therefore

$$\mathbb{P}\left\{\left|\nabla(f\circ T)(Z)\right| \ge \xi(t)\right\} < 2\exp\left(-t^2/2\right)$$

Ineq. (21) now follows from Theorem 3. For $r \ge 1$ consider the following norm on $\mathbb{R}^n$,

$$|x|_{B_1^n \cap r^{-1}B_\infty^n} = \max\left\{|x|_1 + r\,|x|_\infty\right\}$$

After a brief consideration we see that the dual norm is given by

$$|y|_{conv\left(B_\infty^n, rB_1^n\right)} = \max\left\{\sum_{i=1}^n x_i y_i : |x|_1 \le 1, |x|_\infty \le r^{-1}\right\} \le r^{-1} \sum_{i=1}^{\min\{\lceil r\rceil, n\}} y^{(i)} \le 2\,|y|_{conv\left(B_\infty^n, rB_1^n\right)}$$

where $y^{(1)} \ge y^{(2)} \ldots$ is the non-increasing rearrangement of $(|y_i|)_1^n$. Setting $r = t^{-2+4/q}$ and $k = \lceil r\rceil$, by a variation of Corollary 15, with probability at least $1 - C\exp\left(-t^2/2\right)$, (assuming first that $t^{-2+4/q} \le c_q n$),

$$\left|\left((1+|Z_i|)^{-2+4/q}\right)_1^n\right|_{conv\left(B_\infty^n, rB_1^n\right)} \le C_q + C_q\left(\log\frac{n+1}{k}\right)^{-1+2/q}$$

For $t^{-2+4/q} > c_q n$, it follows by classical Gaussian concentration of the $\ell_{-2+4/q}^n$ norm (which is 1-Lipschitz since $-2+4/q \ge 2$) that with probability at least $1 - C\exp\left(-t^2/2\right)$,

$$\left|\left((1+|Z_i|)^{-2+4/q}\right)_1^n\right|_{conv\left(B_\infty^n, rB_1^n\right)} \le C_q n^{-1}\sum_{i=1}^n (1+|Z_i|)^{-2+4/q} \le C_q + C_q t^{-2+4/q} n^{-1} \le C_q$$

In either case,

$$|\nabla(f\circ T)(Z)|^2 \le \left|\left(f_i(TZ)^2\right)_1^n\right|_{B_1^n \cap r^{-1}B_\infty^n}\left|\left((1+|Z_i|)^{-2+4/q}\right)_1^n\right|_{conv\left(B_\infty^n, rB_1^n\right)}$$

The bound on the gradient then translates to a bound on the distribution of $|f(X) - \mathbb{M}f(X)|$ by Pisier's version of the Gaussian concentration inequality (Theorem 9) and, for example, the results of Section 2.1. ■

Proof of Theorem 5. Without loss of generality we may assume that each F_i^{-1} is differentiable, and that $Lip_2^\sharp(f) = 1$. Let $Tx = \left(F_i^{-1}\Phi(x_i)\right)_{i=1}^n$ and let Z be a random vector in $\mathbb{R}^n$ with the standard normal distribution. Then

$$|\nabla(f\circ T)(Z)| = \left(\sum_{i=1}^n f_i(TZ)^2\, Lip\left(F_i^{-1}\Phi, Z_i\right)^2\right)^{1/2} \le \left(\sum_{i=1}^n f_{i,\sharp}^2 U_i\right)^{1/2}$$

where

$$f_{i,\sharp} = \sup_{x\in\mathbb{R}^n} |f_i(x)| \qquad U_i = C\min\{\Phi(Z_i), 1-\Phi(Z_i)\}^{-2/q}\log\min\{\Phi(Z_i), 1-\Phi(Z_i)\}^{-1} \tag{38}$$

Let $U^{(j)} = \left(U_i^{(j)}\right)_{i=1}^n$, $j\in\mathbb{N}$, be an i.i.d. sequence of random vectors in $\mathbb{R}^n$ (independent of Z), where each $U^{(j)}$ has the same distribution as $(U_i)_1^n$ (let this distribution be denoted as μ_*), let $U^{(0)} = 0$, and for some $\delta\in(0,1/2)$ let $N \sim Pois\left(\delta^{-1}\right)$ be independent of $\left(U^{(j)}\right)_1^\infty$ and Z (jointly). Then the empirical measure

$$\sum_{j=1}^N \delta\left(U^{(j)}\right)$$

where $\delta(x)$ denotes the Dirac point mass at x (not to be confused with $\delta \in (0,1/2)$), is a Poisson point process on $\mathbb{R}^n$ with intensity $\delta^{-1}\mu_*$. Consider the norm

$$[x]_\delta = \mathbb{E} \max_{0\le j\le N} \sum_{i=1}^{n} U_i^{(j)} |x_i| \tag{39}$$

From the definition of a Poisson process, for all $t > 0$,

$$\mathbb{P}\left\{ \max_{1\le j\le N} \left\langle \left(f_{i,\sharp}^2\right)_1^n, U^{(j)} \right\rangle < t \right\} = \exp\left(-\delta^{-1}\mathbb{P}\left\{ \sum_{i=1}^{n} f_{i,\sharp}^2 U_i > t \right\}\right)$$

Comparing the median and expected value, noting that $\left\langle \left(f_{i,\sharp}^2\right)_1^n, U^{(0)} \right\rangle = 0$, and re-writing the above equation,

$$\mathbb{P}\left\{ \sum_{i=1}^{n} f_{i,\sharp}^2 U_i > 2\left[\left(f_{i,\sharp}^2\right)_1^n\right]_\delta \right\} \le \delta \log 2$$

It follows from the definition that each U_i has quantile function

$$G^{-1}(t) = \left(\frac{1-t}{2}\right)^{-2/q} \log\left(\frac{1-t}{2}\right)^{-1} : 0 < t < 1 \tag{40}$$

and therefore by Corollary 15, for all $1 \le k \le n$, with probability at least $1 - C\exp\left(-t^2/2\right)$, $\sum_{i=1}^{k} U_i \le C_q\left(k + k^{2/q} t^2 \exp\left(t^2/q\right)\right)$. The details here are tedious, but the impatient reader may compare the bound to $Ck\mathbb{E}U_i + C\max_{1\le i\le k} U_i$. Again, from the definition of a Poisson point process,

$$\mathbb{P}\left\{ \max_{1\le j\le N} \sum_{i=1}^{k} U_i^{(j)} < t \right\} = \exp\left(-\delta^{-1}\mathbb{P}\left\{ \sum_{i=1}^{k} U_i > t \right\}\right) \ge 1 - \delta^{-1}\mathbb{P}\left\{ \sum_{i=1}^{k} U_i > t \right\}$$

The expected value may then be written in terms of the quantile function we have just calculated,

$$\mathbb{E} \max_{1\le j\le N} \sum_{i=1}^{k} U_i^{(j)} \le \int_0^1 C_q\left(k + k^{2/q}(\delta t)^{-2/q}\log(\delta t)^{-1}\right) dt \le C_q\left(k + k^{2/q}\delta^{-2/q}\log\delta^{-1}\right)$$

which implies that $[x]_\delta \le C_q\left(|x|_1 + \delta^{-2/q}\log\delta^{-1}|x|_{q/2}\right)$ for all $x \in \{0, \pm 1\}^n$. It follows that $[x]_\delta \le C_q |x|_{r,q/2}$ for all $x \in \mathbb{R}^n$, where $r = \delta^{-2/q}\log\delta^{-1}$ and $|x|_{r,q/2}$ is defined as in the discussion preceding the statement of Theorem 5. Putting this together,

$$\mathbb{P}\left\{ \sum_{i=1}^{n} f_{i,\sharp}^2 U_i > C_q \left|\left(f_{i,\sharp}^2\right)_1^n\right|_{r,q/2} \right\} \le \delta \log 2$$

Setting $\delta = C\exp\left(-t^2/2\right)$, one can express this as follows: with probability at least $C\exp\left(-t^2/2\right)$,

$$|\nabla (f \circ T)(Z)| \le C_q \left(Lip_2^\sharp(f) + t\exp\left(\frac{t^2}{2q}\right)\left(\sum_{i=1}^{n} i^{-1+1/q} f_{i,\sharp}^2\right)^{1/2} \right)$$

Eq. (25) now follows from Pisier's version of the Gaussian concentration inequality (Theorem 9) and the results of Section 2.1. We now consider Eq. (26). No longer assuming that $Lip_2^\sharp(f) = 1$, using Hölder's inequality for $\ell_{p/2}$ and $\ell_{p/(p-2)}$,

$$|\nabla (f \circ T)(Z)| \leq Lip_p(f) \left(\sum_{i=1}^n U_i^{p/(p-2)} \right)^{(p-2)/(2p)}$$

By Corollary 15, with probability at least $1 - C\exp(-t^2/2)$,

$$\left(\sum_{i=1}^n U_i^{p/(p-2)} \right)^{(p-2)/(2p)} \leq C_{p,q} \left(n^{1/2-1/p} + n^{1/q} t \exp\left(\frac{t^2}{2q} \right) \right)$$

The result now follows as before by Pisier's version of the Gaussian concentration inequality (Theorem 9) and the results of Section 2.1. ■

Proof of Theorem 6. We start by including the randomness of U, uniformly distributed on $SO(n)$, and then show how to remove this randomness. By approximation we may assume that each μ_i has a continuous nonvanishing density f_i and that g is differentiable. Consider the function $h(x) = g(UTx)$, where $Tx = \left(F_i^{-1}\Phi(x_i)\right)_1^n$. Then $\nabla h(x) = \nabla g(UTx)\, U(DT(x))$, where $DT(x)$ has $\phi(x_i)/f_i\left(F_i^{-1}\Phi(x_i)\right)$ along its diagonal and zeros elsewhere, and we view $\nabla g(Ux)$ as a $1 \times n$ matrix. We may write $X = TZ$, where Z is a random vector in $\mathbb{R}^n$ with the standard normal distribution. For any two non-antipodal vectors $x, y \in S^{n-1}$, there exists a unique $U_{y,x} \in SO(n)$ such that $U_{y,x}x = y$ and for all $z \in \{x, y\}^\perp$, $U_{y,x}z = z$. We don't have to worry about the case when x and y are antipodal since we will be dealing with random vectors and this happens with probability zero. Let

$$W = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \widetilde{W} & \\ 0 & & & \end{bmatrix}$$

where $\widetilde{W}$ is a random matrix uniformly distributed in $SO(n-1)$ and independent of X and Z, and let Y be a random vector on S^{n-1}, also uniformly distributed, and independent of X, Z and $\widetilde{W}$ (jointly). Let $\widehat{X} = X/|X|$. Then

$$U = U_{Y,\widehat{X}} U_{\widehat{X},e_1} W U_{e_1,\widehat{X}}$$

is uniformly distributed in $SO(n)$ and $UTZ = |X|\,Y$. Considering the action of a diagonal matrix and applying Hölder's inequality to the pairing of ℓ_p^n and $\ell_{p/(p-1)}^n$,

$$|\nabla h(x)| \leq |\nabla g(UTx)\, U|_{2p} \left(\sum_{i=1}^n Lip\left(F_i^{-1}\Phi, x_i\right)^{2p/(p-1)} \right)^{(p-1)/2p}$$

Then

$$\nabla g(UTZ)\, U = \nabla g(|X|\,Y)\, U_{Y,\widehat{X}} U_{\widehat{X},e_1} W U_{e_1,\widehat{X}}$$

Note that everything to the left of W is independent of W. Writing $\nabla g\left(|X|\,Y\right) = aY + bH$, where $H \in S^{n-1} \cap Y^{\perp}$ and $|a|, |b| \leq 1$, it follows from elementary properties of orthogonal matrices and properties of the transpose that

$$\nabla g\left(|X|\,Y\right) U_{Y,\widehat{X}} U_{\widehat{X},e_1} W U_{e_1,\widehat{X}} = a\widehat{X} + bAU_{e_1,\widehat{X}}$$

where A is uniformly distributed on $\{\theta \in S^{n-1} : \theta_1 = 0\}$ (here we don't make too much of a distinction between row and column vectors and matrices with either one row or one column, considering all of these as vectors in $\mathbb{R}^n$ under the natural isomorphism). Let $G = \chi A + G_1 e_1$ where χ follows the chi distribution with $n-1$ degrees of freedom and $G_1 \sim N(0,1)$, jointly independent of everything else, and each other. Then G has the standard normal distribution on $\mathbb{R}^n$ and $A = \left(\sum_2^n G_i^2\right)^{-1/2}(0, G_2, G_3 \ldots, G_n)$. Then

$$AU_{e_1,\widehat{X}} = \left(\sum_{i=2}^n G_i^2\right)^{-1/2} \left(GU_{e_1,\widehat{X}} - G_1\widehat{X}\right)$$

Putting this together

$$\left|\nabla g\left(UTZ\right)U\right|_{2p} \leq \left(1 + |G_1| \left(\sum_{i=2}^n G_i^2\right)^{-1/2}\right) |X|_{2p} / |X| + \left(\sum_{i=2}^n G_i^2\right)^{-1/2} \left|GU_{e_1,\widehat{X}}\right|_{2p}$$

We are now in a position to estimate the quantiles of $|\nabla h(Z)|$. By definition of G, $\mathbb{P}\{|G_1| > \lambda\} \leq C\exp\left(-\lambda^2/2\right)$. Using $\mathrm{vol}_n\left(rB_2^n\right) \leq \left(Crn^{-1/2}\right)^n$ and a bound on the density (of product measures), the following is bounded above by $C\exp\left(-\lambda^2/2\right)$,

$$\mathbb{P}\left\{|X| < c_q n^{1/2}\exp\left(-\lambda^2 n^{-1}/2\right)\right\} + \mathbb{P}\left\{\left(\sum_{i=2}^n G_i^2\right)^{1/2} < cn^{1/2}\exp\left(-\lambda^2\left(n-1\right)^{-1}/2\right)\right\}$$

Even though the definition of G can be traced back to include $\widehat{X}$, because W is uniformly distributed and independent of $\widehat{X}$, information about $\widehat{X}$ is lost and G is independent of $\widehat{X}$. This implies that $GU_{e_1,\widehat{X}}$ has the standard normal distribution on $\mathbb{R}^n$, and by Gaussian concentration of the ℓ_{2p}^n norm

$$\mathbb{P}\left\{\left|GU_{e_1,\widehat{X}}\right|_{2p} > C_p n^{1/(2p)} + \lambda\right\} \leq C\exp\left(-\lambda^2/2\right)$$

By Example 16 (which is based on Corollary 15),

$$|X|_{2p} \leq C_{p,q}\left(n^{1/(2p)} + n^{1/q}\exp\left(\frac{\lambda^2}{2q}\right)\right)$$

By the standard estimates for Φ and the assumed bound (27) on $Lip\left(F_i^{-1}, \cdot\right)$, see also the similar estimates in the proof of Theorem 5, in particular (38) and (40),

$$\mathbb{P}\left\{Lip\left(F_i^{-1}\Phi, Z_i\right)^{2p/(p-1)} \geq C^{p/(p-1)} s^{-2p/[q(p-1)]}\left(\log s^{-1}\right)^{p/(p-1)}\right\} \leq s$$

and therefore, since $2p/\left[q(p-1)\right] < 1$, it follows from Corollary 15 that with probability at least $1 - C\exp\left(-\lambda^2/2\right)$,

$$\sum_{i=1}^{n} Lip\left(F_i^{-1}\Phi, Z_i\right)^{2p/(p-1)} \leq C_{p,q}\left(n + n^{2p/[q(p-1)]}\lambda^{2p/(p-1)}\exp\left(\frac{\lambda^2 p}{q(p-1)}\right)\right)$$

This is actually the same bound as for $\sum U_i^{p/(p-2)}$ at the end of the proof of Theorem 5, with $p/(p-2) < q/2$ there replaced with $p/(p-1) < q/2$ here. So after dropping terms dominated by others and changing an n to an $n-1$, with probability at least $1 - C\exp\left(-\lambda^2/2\right)$,

$$\begin{aligned}|\nabla h(Z)| \;\leq\; & C_{p,q}n^{-1/2}\exp\left((n-1)^{-1}\lambda^2/2\right)\left(1 + \lambda n^{-1/2}\exp\left((n-1)^{-1}\lambda^2/2\right)\right) \\ & \times\left(n^{1/(2p)} + n^{1/q}\exp\left(\frac{\lambda^2}{2q}\right)\right)\left(n^{(p-1)/(2p)} + n^{1/q}\lambda\exp\left(\frac{\lambda^2}{2q}\right)\right)\end{aligned}$$

By Pisier's version of the Gaussian concentration inequality and the results of Section 2.1, see for example Theorem 3 and its use in the proof of Theorem 4), (for any $\lambda > 0$) with probability at least $1 - C\exp\left(-\lambda^2/2\right)$,

$$\begin{aligned}|h(Z) - \mathbb{M}h(Z)| \;\leq\; & C_{p,q}\lambda n^{-1/2}\exp\left((n-1)^{-1}\lambda^2/2\right)\left(1 + \lambda n^{-1/2}\exp\left((n-1)^{-1}\lambda^2/2\right)\right) \\ & \times\left(n^{1/(2p)} + n^{1/q}\exp\left(\frac{\lambda^2}{2q}\right)\right)\left(n^{(p-1)/(2p)} + n^{1/q}\lambda\exp\left(\frac{\lambda^2}{2q}\right)\right)\end{aligned}$$

Restricting the values of λ as in the statement of the theorem,

$$n^{1/q}\lambda\exp\left(\frac{\lambda^2}{2q}\right) \leq n^{(p-1)/(2p)} \qquad\qquad n^{1/q}\exp\left(\frac{\lambda^2}{2q}\right) \leq n^{1/(2p)}$$

so

$$\mathbb{P}\left\{|h(Z) - \mathbb{M}h(Z)| > C_{p,q}\lambda\right\} \geq C\exp\left(-\lambda^2/2\right)$$

We now show how to get rid of the randomness due to U. In slight abuse of notation, $\mathbb{M}g(UX) = \mathbb{M}h(Z)$ denotes the median of the single random variable $g(UX)$ where both U and X are random (i.e. not conditioned on U), even when followed by ': U'. Let $E(t)$ be the event $\{|g(UX) - \mathbb{M}g(UX)| > t\}$ and define $\mathbb{P}\{|g(UX) - \mathbb{M}g(UX)| > t : U\}$ to mean $\mathbb{E}\left(1_{E(t)} : U\right)$. By Markov's inequality,

$$\begin{aligned} & \mathbb{P}\left\{\mathbb{P}\left\{|g(UX) - \mathbb{M}g(UX)| > C_q n^{1/2}t : U\right\} > C\exp\left(-t^2/2\right)(\log(2+t))(\log\log(3+t))^2\right\} \\ \leq\; & c\exp\left(t^2/2\right)(\log(2+t))^{-1}(\log\log(3+t))^{-2}\,\mathbb{E}\mathbb{P}\left\{|g(UX) - \mathbb{M}g(UX)| > C_q n^{1/2}t : U\right\} \\ =\; & c\exp\left(t^2/2\right)(\log(2+t))^{-1}(\log\log(3+t))^{-2}\,\mathbb{P}\left\{|g(UX) - \mathbb{M}g(UX)| > C_q n^{1/2}t\right\} \\ =\; & c\exp\left(t^2/2\right)(\log(2+t))^{-1}(\log\log(3+t))^{-2}\,\mathbb{P}\left\{|h(Z) - \mathbb{M}h(Z)| > C_q n^{1/2}t\right\} \\ \leq\; & c(\log(2+t))^{-1}(\log\log(3+t))^{-2}\end{aligned}$$

We now apply this with $t = C^m$, $m \in \{1, 2, 3\ldots\}$, where we allow m to grow until C^m is the same order of magnitude as the maximum value of λ allowed. By the union bound, with probability at least 0.99, the following holds for all such values of t,

$$\mathbb{P}\left\{|g(UX) - \mathbb{M}g(UX)| > C_q n^{1/2}t : U\right\} < C\exp\left(-t^2/2\right)(\log(2+t))(\log\log(3+t))^2$$

This gives estimates for the quantiles of $|g(UX) - \mathbb{M}g(UX)|$. Since the quantile function is non-decreasing, we can pass from these estimates to the full (continuous) range of t (or λ) and the theorem is proved. ■

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] Barthe, F., Cattiaux, P., Roberto C.: Concentration for independent random variables with heavy tails. AMRX, Appl. Math. Res. Express (2), 39-60 (2005)

[3] Bobkov, S. G.: An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space. Ann. Probab. 25 (1), 206-214 (1997)

[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] Boucheron, S., Thomas, M.: Concentration inequalities for order statistics. Electron. Commun. Probab. 17 (51), 1-12 (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] 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)

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

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

[11] Guédon, O., Litvak, A. E., Pajor, A., Tomczak-Jaegermann, N.: On the interval of fluctuations of the singular values of random matrices. J. Eur. Math. Soc. (JEMS) 19 (5), 1469-1505 (2017)

[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] Latała, R.: Estimates of moments of sums of independent real random variables. Ann. Probab. 25 (3), 1502-1513 (1997)

[16] Latała, R., Oleszkiewicz, K., Between Sobolev and Poincaré. Geometric aspects of functional analysis, 147-168, Lecture Notes in Math., 1745, Springer, Berlin, 2000.

[17] Marshall, A. W., Olkin, I., Arnold, B. C.: Inequalities: theory of majorization and its applications. Second edition. Springer Series in Statistics. Springer, New York, 2011. xxviii+909 pp. ISBN: 978-0-387-40087-7

[18] Mendelson, S., Pajor, A., Tomczak-Jaegermann, N.: Uniform uncertainty principle for Bernoulli and subgaussian ensembles. Constr. Approx. 28 (3), 277-289 (2008)

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

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

[21] Petrov, V. V.: Sums of independent random variables. Springer-Verlag 1975 (translated from Russian into English)

[22] Pinelis, I.: Optimal tail comparison based on comparison of moments, Proceedings of the Conference on High Dimensional Probability (Oberwolfach, Germany, 1996), Progress in Probability, 43, Birkhauser, Basel, Switzerland, (1998), 297-314.

[23] Pinelis, I.: Fractional sums and integrals of r-concave tails and applications to comparison probability inequalities. Advances in stochastic inequalities (Atlanta, GA, 1997), 149-168, Contemp. Math., 234, Amer. Math. Soc., Providence, RI, 1999

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

[25] Schechtman, G., Zinn, J.: On the volume of the intersection of two L_p^n balls. Proc. Amer. Math. Soc. 110 (1), 217-224 (1990)

[26] Schechtman, G., Zinn, J.: Concentration on the ℓ_p^n ball. Geometric aspects of functional analysis, 245-256, Lecture Notes in Math., 1745, Springer, Berlin, 2000.

[27] 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)

[28] Talagrand, M.: The supremum of some canonical processes. Amer. J. Math. 116 (2), 283-325 (1994)

[29] Tanguy, K.: Non asymptotic variance bounds and deviation inequalities by optimal transport. arXiv: 1708.08620.

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