ON THE TIGHTNESS OF GAUSSIAN CONCENTRATION FOR CONVEX FUNCTIONS

PETROS VALETTAS

ABSTRACT. The concentration of measure phenomenon in Gauss' space states that every L-Lipschitz map f on $\mathbb{R}^n$ satisfies

$$\gamma_n \left(\{ x : |f(x) - M_f| \geqslant t \} \right) \leqslant 2 e^{-\frac{t^2}{2L^2}}, \quad t > 0,$$

where γ_n is the standard Gaussian measure on $\mathbb{R}^n$ and M_f is a median of f. In this work, we provide necessary and sufficient conditions for when this inequality can be reversed, up to universal constants, in the case when f is additionally assumed to be convex. In particular, we show that if the variance $\mathrm{Var}(f)$ (with respect to γ_n) satisfies $\alpha L \leqslant \sqrt{\mathrm{Var}(f)}$ for some $0 < \alpha \leqslant 1$, then

$$\gamma_n \left(\{ x : |f(x) - M_f| \geqslant t \} \right) \geqslant c e^{-C\frac{t^2}{L^2}}, \quad t > 0,$$

where $c, C > 0$ are constants depending only on α.

1. INTRODUCTION

The concentration of measure phenomenon is by now a fundamental tool in modern probability theory with profound impacts in many research areas. Its significance in the local theory of normed spaces was emphasized by V. Milman in his seminal work [Mil71] on almost spherical sections of high-dimensional convex bodies. Subsequently, applications of concentration have increased remarkably in different fields and concentration techniques have been developed in various contexts. The interested reader may consult the book of Ledoux [Led01], the comprehensive paper of Talagrand [Tal95] and the recent monograph [BLM13] by Boucheron, Lugosi and Massart, for further background and detailed discussions on this very interesting subject.

The prototypical example of concentration is in Gauss' space $(\mathbb{R}^n, \| \cdot \|_2, \gamma_n)$ stating that any L-Lipschitz map f on $\mathbb{R}^n$ satisfies

$$(1.1) \qquad \max \left\{ \gamma_n(\{ x : f(x) \geqslant \mathrm{med}(f) + t \}), \gamma_n(\{ x : f(x) \leqslant \mathrm{med}(f) - t \}) \right\} \leqslant \frac{1}{2} e^{-t^2/2L^2}, \quad t > 0.$$

Formally, the concentration inequality (1.1) follows from the solution to the isoperimetric problem in Gauss' space which was proved independently by Borell [Bor75] and Sudakov and Tsirel'son in [ST74]. The latter asserts that among all Borel sets A with given measure, half-spaces have minimal Gaussian surface area. Equivalently, reformulating the Gaussian isoperimetry using enlargements of sets with respect to the Euclidean ball, one has:

$$(1.2) \qquad \gamma_n(A + t B_2^n) \geqslant \Phi(\Phi^{-1}[\gamma_n(A)] + t), \quad t > 0,$$

for all Borel sets $A \subseteq \mathbb{R}^n$, where Φ denotes the cumulative distribution function of a standard Gaussian.

The concentration inequality (1.1) lies at the center of many important Gaussian inequalities such as the logarithmic Sobolev inequality [Gro75], Nelson's hypercontractive principle [Nel67], the Poincaré inequality [Che82], Ehrhard's inequality [Ehr83] and more.

It is known that (1.1) is sharp for linear functionals. In many cases it provides the correct estimate even if the function is far from being linear, e.g. the ℓ_p^n-norm $z \mapsto \|z\|_p = (\sum_{i \leqslant n} |z_i|^p)^{1/p}$ for $1 \leqslant p \leqslant 2$. The fact that (1.1) is sharp for norms in the large deviation regime $t > \mathrm{med}(f)$ is well known, see e.g. [LT91, Proposition 2.9]. However, there are important examples of Lipschitz maps such as $z \mapsto \max_{j \leqslant n} |z_j|$,

$z \mapsto \|z\|_4$ or an ellipsoidal norm $z \mapsto \|Az\|_2$, $A \in \mathbb{R}^{m \times n}$, in which the classical concentration fails to capture the right behavior in the small deviation regime $0 < t < \mathrm{med}(f)$. Frequently, the functions under consideration are additionally convex, e.g. norms (of vectors or matrices), suprema of linear functionals indexed by sets, among others. In view of the above, several questions arise naturally such as:

(Q1) Do convexity assumptions ensure sharper concentration bounds?

(Q2) Under what conditions is the classical concentration inequality (1.1) tight?

For the first question, and for the deviation below the median, a stronger, variance-sensitive, inequality is available for convex functions, which was established recently in [PV17a]: for any convex map $f \in L_2(\gamma_n)$, one has

$$(1.3) \qquad \gamma_n(\{x : f(x) \leqslant \mathrm{med}(f) - 20t\}) \leqslant \frac{1}{2} e^{-t^2/\mathrm{Var}(f)}, \quad t > 0.$$

The improvement lies in the fact that

$$(1.4) \qquad \mathrm{Var}_{\gamma_n}(f) \leqslant \mathrm{Lip}(f)^2,$$

(which follows by (1.1) or by the Gaussian Poincaré inequality [Che82]). This new type of concentration inequality (1.3) exploits the convexity properties of the Gaussian measure, as opposed to (1.1) which can be explained by isoperimetry. Corresponding estimates can therefore be proved for arbitrary log-concave measures; see [PV17b]. All of these suggest that the left and right distributional behaviors near the median should be treated separately.

In this note, we focus on the second question. To this end let us discuss the aforementioned examples in more detail and review the different reasons that (1.1) and (1.4) can fail to be tight. In particular, for the function $f(z) = \max_{i \leqslant n} |z_i|$ the reason is the *super-concentration phenomenon*, following Chatterjee [Cha14]. Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be ε_n-super-concentrated for some $\varepsilon_n \in (0,1)$, if

$$(1.5) \qquad \mathrm{Var}_{\gamma_n}(f) \leqslant \varepsilon_n \mathbb{E}_{\gamma_n} \|\nabla f\|_2^2.$$

With this terminology we have that $z \mapsto \max_{i \leqslant n} |z_i| = \|z\|_\infty$ is $\frac{C}{\log n}$-super-concentrated[1], since

$$\mathrm{Var}[\|Z\|_\infty] \simeq \frac{1}{\log n} \quad \text{and} \quad \|\nabla\|Z\|_\infty\|_2 = \mathrm{Lip}(\|\cdot\|_\infty) = 1 \quad \text{a.s.},$$

where Z is an n-dimensional standard Gaussian vector. Moreover, the deviation is known to be described by a two-level behavior (see e.g. [Tal91] and [Sch07])

$$ce^{-C\alpha_{n,\infty}(t)} \leqslant \mathbb{P}(|\|Z\|_\infty - \mathbb{E}\|Z\|_\infty| > t) \leqslant Ce^{-c\alpha_{n,\infty}(t)}, \quad \alpha_{n,\infty}(t) = \max\{t^2, t\sqrt{\log n}\}, \ t > 0.$$

In the light of (1.5) we define the *super-concentration constant* of f as follows:[2]

$$\mathbf{s}(f) = \sqrt{\frac{\mathrm{Var}(f)}{\mathbb{E}\|\nabla f\|_2^2}}.$$

With this notation the mapping $z \mapsto \|z\|_\infty$ is super-concentrated with $\mathbf{s}(\|\cdot\|_\infty) \simeq 1/\sqrt{\log n}$.

However, the super-concentration phenomenon is not the only reason for the sub-optimal bounds. In the case of $z \mapsto \|z\|_4$ or $z \mapsto \|Az\|_2$, $A \in \mathbb{R}^{m \times n}$, the reason is that

$$\mathbb{E}\|\nabla f(Z)\|_2^2 \ll \mathrm{Lip}(f)^2.$$

More precisely, we have (see e.g. [Nao07] and [PVZ17])

$$\mathrm{Var}[\|Z\|_4] \simeq \frac{1}{\sqrt{n}} \simeq \mathbb{E}\|\nabla\|Z\|_4\|_2^2, \quad \text{whereas} \quad \mathrm{Lip}(\|\cdot\|_4) = 1,$$

---

[1]Here and everywhere else $C, c, C_1, c_1, \dots$ stand for positive universal constants whose values may change from line to line. For any two quantities A, B depending on dimension, on the parameters of the problem, etc. We write $A \simeq B$ if there exists universal constant $C > 0$ -independent of everything- such that $A \leqslant CB$ and $B \leqslant CA$.

[2]Note that the ratio

$$\frac{1}{\mathbf{s}(f)^2} = \frac{\mathbb{E}\|\nabla f\|_2^2}{\mathrm{Var}(f)} = \frac{\langle -Lf, f \rangle}{\|f\|_{L_2}^2},$$

coincides with the gaussian Rayleigh-Ritz quotient (see e.g. [Ehr84a]) of the operator $-L$ (the generator of the Ornstein-Uhlenbeck semigroup) at f, provided $\int f = 0$, $f \neq 0$.

and the deviation exhibits a three-level behavior:

$$ce^{-C\alpha_{n,4}(t)} \leqslant \mathbb{P}(|\,\|Z\|_4 - \mathbb{E}\|Z\|_4| > t) \leqslant Ce^{-c\alpha_{n,4}(t)}, \quad t > 0,$$

where $\alpha_{n,4}(t) = \max\left\{\min\{t^2 n^{1/2}, t^{1/2} n^{3/8}\}, t^2\right\}$.

For the ellipsoidal norm $z \mapsto Q_A(z) := \|Az\|_2, \ A \in \mathbb{R}^{m \times n}$ one may check that:

$$\mathrm{Var}[Q_A(Z)] \simeq \mathbb{E}\|\nabla Q_A(Z)\|_2^2 \simeq \frac{\|A\|_{S_4}^4}{\|A\|_{\mathrm{HS}}^2}, \quad \text{and} \quad \mathrm{Lip}(Q_A) = \|A\|_{\mathrm{op}},$$

$\|\cdot\|_{S_4}$ is the 4-Schatten norm, $\|\cdot\|_{\mathrm{HS}}$ is the Hilbert-Schmidt norm and $\|\cdot\|_{\mathrm{op}}$ stands for the operator norm of the linear map $A : \ell_2^n \to \ell_2^m$. Again in this case the deviation obeys a multiple-level behavior:

$$ce^{-C\alpha_A(t)} \leqslant \mathbb{P}\left(\left|\|AZ\|_2 - (\mathbb{E}\|AZ\|_2^2)^{1/2}\right| > t\right) \leqslant Ce^{-c\alpha_A(t)}, \quad t > 0,$$

where $\alpha_A(t) = \max\left\{\min\{t^2\|A\|_{\mathrm{HS}}^2/\|A\|_{S_4}^4, t\|A\|_{\mathrm{HS}}/\|A\|_{\mathrm{op}}^2\}, t^2/\|A\|_{\mathrm{op}}^2\right\}$. The right-hand side estimate in the above concentration inequality is due to Hanson and Wright [HW71] and is known to hold for the more general class of sub-gaussian random vectors with independent coordinates, see e.g. [RV13] for a modern exposition and the references therein.[3]

Thus in all of the aforementioned cases, we observe that

$$\mathrm{Var}(f) \ll \mathrm{Lip}(f)^2.$$

In view of the above remarks, and for the purpose of this note, we may intorduce, for any Lipschitz map f, the *over-concentration constant* of f as follows:

$$\mathbf{ov}(f) = \frac{\sqrt{\mathrm{Var}(f)}}{\mathrm{Lip}(f)}.$$

With this terminology the mapping $z \mapsto \|z\|_4$ is over-concentrated (but not super-concentrated) with $\mathbf{ov}(\|\cdot\|_4) \simeq 1/\sqrt[4]{n}$ and the ellipsoidal norm $z \mapsto Q_A(z)$ is over-concentrated when a gap occurs at the top of the spectrum of A.

The main purpose of this note is to show that this parameter quantifies the tightness of the concentration for convex Lipschitz maps. Alternatively, note that if (1.1) can be reversed (up to constants) then it implies a reversal for (1.4), and hence is a necessary condition for the optimality of the concentration. We show that this condition is also sufficient. Namely we prove the following:

Theorem 1.1. *Let $0 < \alpha \leqslant 1$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and L-Lipschitz map. Then, we have:*

$$(1.6) \qquad \mathrm{Var}[f(Z)] \geqslant \alpha^2 L^2 \implies \mathbb{P}(|f(Z) - \mathrm{med}(f)| \geqslant tL) \geqslant c(\alpha)\exp(-C(\alpha)t^2), \quad t > 0,$$

where Z is a standard n-dimensional Gaussian vector and $c(\alpha), C(\alpha) > 0$ depend only on α. Moreover, we can have $c(\alpha) \geqslant c\alpha^8$ and $C(\alpha) \leqslant C\alpha^{-4}\log(e/\alpha)$.

The latter can be viewed as a stability type result in the following sense: since (1.4) holds as equality for the affine maps, and (1.1) is sharp for them, we measure the concentration for functions which are far from linear but now the "distance" is measured in terms of the $\mathbf{ov}(f)$.

In what follows we fix the notation. We use ζ for a standard Gaussian random variable, i.e. $\zeta \sim N(0,1)$ and Z for a standard Gaussian (usually n-dimensional) random vector, i.e. $Z \sim N(\mathbf{0}, I_n)$. We write γ_n for the n-dimensional standard Gaussian measure and simply γ for γ_1. The symbol $\mathbb{E}$ or $\mathbb{E}_{\gamma_n}$ stands for the expectation and Var or Var_{γ_n} stands for the variance. Let $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x e^{-t^2/2}\,dt$ the cumulative distribution function of a standard normal. We write $\mathrm{med}(\xi)$ for a median of a random variable ξ.

---

[3]Usually the Hanson-Wright inequality is stated for quadratic forms, hence for the map $z \mapsto \|Az\|_2^2$. The reason we omit the squares here is because we discuss for Lipschitz functions.

2. Proof of the main result

In this Section we discuss the basic tools in order to establish (1.6) and we finally give the proof in Theorem 2.8. The first ingredient in our approach is Ehrhard's inequality [Ehr83]. Ehrhard proved his inequality in the following form:

Theorem 2.1 (Ehrhard). *Let A, B be convex sets in $\mathbb{R}^n$. Then, for any $\lambda \in (0,1)$ we have:*

$$\Phi^{-1}[\gamma_n((1-\lambda)A + \lambda B)] \geqslant (1-\lambda)\Phi^{-1}[\gamma_n(A)] + \lambda\Phi^{-1}[\gamma_n(B)].$$

The above inequality has been extended to all Borel sets by Borell in [Bor03]. (See [IV15, vH17, NP16] for recent developments and further references). However we will not need these extensions in this work.

An important tool in Ehrhard's work is the notion of the *Gaussian rearrangement* that we recall now. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a measurable function. Following Ehrhard [Ehr84a] we define the Gaussian rearrangement of f as the generalized inverse of the map $t \mapsto \Phi^{-1} \circ \gamma_n(f \leqslant t)$, i.e.

$$f^*(s) = \inf\{t : s \leqslant \Phi^{-1} \circ \gamma_n(f \leqslant t)\}.$$

Note that f^* is non-decreasing and transports the measure γ to the distribution of $f(Z)$ with $Z \sim N(\mathbf{0}, I_n)$. In the following lemma we collect some basic properties of f^* that we will need in the sequel. We sketch the proof of some basic facts for reader's convenience.

Lemma 2.2 (Ehrhard). *Let f be a measurable function on $\mathbb{R}^n$. Then, f^* enjoys the following properties:*

   a. *The map $f^* : \mathbb{R} \to \mathbb{R}$ is convex, if f is convex.*
   b. *The modulus of continuity of f^* satisfies $\omega_{f^*} \leqslant \omega_f$. In particular, if f is Lipschitz, then f^* is Lipschitz with $\|f^*\|_{\mathrm{Lip}} \leqslant \|f\|_{\mathrm{Lip}}$.*
   c. *For all $u \in \mathbb{R}$ we have $\gamma(f^* \leqslant u) = \gamma_n(f \leqslant u)$.*
   d. *For any $1 \leqslant p < \infty$ one has:*

$$\int_{\mathbb{R}} |(f^*)'|^p \, d\gamma \leqslant \int_{\mathbb{R}^n} \|\nabla f\|_2^p \, d\gamma_n.$$

Proof. (a). Applying Ehrhard's inequality (Theorem 2.1) for $A = B = \{f \leqslant t\}$ we obtain that $(f^*)^{-1}$ is concave.

(b). The isoperimetric inequality (1.2) implies:

$$\Phi^{-1} \circ \gamma_n(A + sB_2^n) \geqslant \Phi^{-1} \circ \gamma_n(A) + s, \quad s > 0.$$

For $A = \{f \leqslant t\}$ says that $(f^*)^{-1}(t + \omega_f(s)) \geqslant (f^*)^{-1}(t) + s$, which yields:

$$f^*\left((f^*)^{-1}(t) + s\right) - t \leqslant \omega_f(s),$$

for all $t \in \mathbb{R}$ and $s > 0$.

(c). We may write:

$$\gamma(\{t \in \mathbb{R} : f^*(t) \leqslant u\}) = \gamma(\{t : F(u) \geqslant t\}) = \Phi(F(u)) = \gamma_n(\{x \in \mathbb{R}^n : f(x) \leqslant u\}),$$

for all $u \in \mathbb{R}$.

(d). For a proof the reader is referred to [Ehr84a] (see also [Ehr84b] for a related application). □

The second ingredient in our approach is a remarkable inequality of Talagrand [Tal94] that improves upon the classical Poincaré inequality [Che82] (see also [Che81]). Before stating his result we need to recall some basic definitions. A function $\psi : [0, \infty) \to \mathbb{R}$ is said to be a Young function if it is convex, increasing and $\psi(0) = 0$. If $(\Omega, \mathcal{A}, \mu)$ is a probability space and ψ is a Young function, the Orlicz norm of an $\mathcal{A}$-measurable function h in $L^\psi(\mu)$ is defined by

$$\|h\|_\psi := \inf\left\{\lambda > 0 : \mathbb{E}_\mu \psi\left(\frac{|h|}{\lambda}\right) \leqslant 1\right\}.$$

With this notation we have the following:

4

Theorem 2.3 (Talagrand)**.** *Let $\varphi(t) := \frac{t^2}{\log(e+t)}$, $t \geqslant 0$. For any smooth function f we have:*

$$\operatorname{Var}(f) \leqslant C \sum_{i=1}^{n} \|\partial_i f\|_{\varphi}^2.$$

Talagrand proved and stated his theorem in the case of the discrete cube (with the normalized counting measure) instead of the Gaussian. The above statement follows by a standard application of the central limit theorem or by mimicking his proof for the cube in Gauss' space (see also [CEL12] for alternative versions).

On the real line Bobkov and Houdré in [BH99] proved that for convex functions the reverse inequality is also true.

Theorem 2.4 (Bobkov-Houdré)**.** *There exist universal constants $c_1, c_2, c_3 > 0$ such that for every convex function $f : \mathbb{R} \to \mathbb{R}$ one has*

$$c_1 \int_{\mathbb{R}} \frac{|f'(t)|^2}{1+t^2} d\gamma(t) \leqslant \operatorname{Var}_{\gamma}(f) \leqslant c_2 \|f'\|_{\varphi}^2 \leqslant c_3 \int_{\mathbb{R}} \frac{|f'(t)|^2}{1+t^2} d\gamma(t).$$

Actually we are going to use only the last inequality which is a consequence of Talagrand's theorem (see [BH99, Lemma 5]). We will also need the following:

Lemma 2.5. *Let g be a non-decreasing and convex map on $\mathbb{R}$. Then, we have:*

(1) *For all $p > 0$,*

$$\|(g - \operatorname{med}(g))_+\|_{L_p(\gamma)}^p \geqslant \sigma_p^p [g'(0+)]^p,$$

   where $\sigma_p^p = \frac{2^{p/2}}{2\sqrt{\pi}} \Gamma(\frac{p+1}{2}) =: \frac{1}{2}\mathbb{E}|\zeta|^p$, $\zeta \sim N(0,1)$.

(2) *Let $s := \sqrt{\operatorname{Var}(g)}/\|g'\|_{L_2(\gamma)}$. Then*

$$\operatorname{Var}[g(\zeta)] \leqslant C_1 \left[g'\left(\frac{C_1}{s}-\right) \right]^2.$$

(3) *For any $t > 0$ we have:*

$$\mathbb{P}\left(g(\zeta) - g(0) \geqslant t\sqrt{\operatorname{Var}[g(\zeta)]} \right) \geqslant 1 - \Phi\left(C_1 \left(\frac{1}{s} + t \right) \right),$$

   where $C_1 > 0$ is a universal constant.

Proof. Note that convexity of g shows that $g(t) - g(0) \geqslant t g'(0+)$ for all $t > 0$. Combining this with the monotonicity we obtain:

$$\|(g - \operatorname{med}(g))_+\|_p^p = \int_0^{\infty} [g(t) - g(0)]^p \, d\gamma(t) \geqslant \frac{(g'(0+))^p}{\sqrt{2\pi}} \int_0^{\infty} t^p e^{-t^2/2} \, dt,$$

for all $p > 0$. This proves (1).

By Theorem 2.4 we have:

$$\frac{1}{c_3}\operatorname{Var}[g(\zeta)] \leqslant \int_{\mathbb{R}} \frac{(g'(t))^2}{1+t^2} \, d\gamma(t) \leqslant g'(\lambda-)^2 \int_{\mathbb{R}} \frac{d\gamma(t)}{1+t^2} + \int_{\lambda}^{\infty} \frac{g'(t)^2}{1+t^2} \, d\gamma(t)$$
$$\leqslant (g'(\lambda-))^2 + \|g'\|_{L_2}^2/\lambda^2,$$

for all $\lambda > 0$. We choose $\lambda = \sqrt{2c_3}/s$. This proves (2).

Let $s = \sqrt{\operatorname{Var}(g)}/\|g'\|_{L_2}$ and $a = C_1/s$, where $C_1 > 0$ is the constant from part (2). Note that for any $x > a$ we have:

(2.1) $$g(x) \geqslant g(a) + (x-a)g'(a+) \geqslant g(0) + \frac{\sqrt{\operatorname{Var}(g)}}{\sqrt{C_1}}(x-a),$$

where we have used part (2) and the monotonicity of g. Hence, for any $t > 0$, we obtain:

$$\left\{ x : x > a + t\sqrt{C_1} \right\} \subseteq \left\{ x : g(x) - g(0) \geqslant t\sqrt{\operatorname{Var}[g(\zeta)]} \right\}$$

where we have used (2.1). Finally,

$$\mathbb{P}\left(g(\zeta) - g(0) \geqslant t\sqrt{\mathrm{Var}[g(\zeta)]}\right) \geqslant 1 - \Phi\left(a + \sqrt{C_1}t\right),$$

for all $t > 0$. This completes the proof. $\qquad\square$

For any Lipschitz map f on $\mathbb{R}^n$, recall the *over-concentration* and the *super-concentration* constant:

$$\mathbf{ov}(f) = \frac{\sqrt{\mathrm{Var}[f(Z)]}}{\mathrm{Lip}(f)} \quad \text{and} \quad \mathbf{s}(f) = \frac{\sqrt{\mathrm{Var}[f(Z)]}}{(\mathbb{E}\|\nabla f(Z)\|_2^2)^{1/2}}.$$

Note that in view of the Gaussian Poincaré inequality [Che82] we have:

(2.2) $$\mathbf{ov}(f) \leqslant \mathbf{s}(f) \leqslant 1.$$

Our first main result is the following inequality:

Proposition 2.6. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a convex function with* $f \in L_2(\gamma_n)$. *Then,*

$$\mathbb{P}\left(f(Z) \geqslant M + t\sqrt{\mathrm{Var}[f(Z)]}\right) \geqslant 1 - \Phi\left(C\left(t + \frac{1}{\mathbf{s}(f)}\right)\right), \quad t > 0,$$

In particular, if f *is not super-concentrated, i.e.* $\mathbf{s}(f) \simeq 1$, *we have*

$$\mathbb{P}\left(|f(Z) - M| \geqslant t\sqrt{\mathrm{Var}[f(Z)]}\right) \geqslant ce^{-Ct^2}, \quad t > 0,$$

where $M = \mathrm{med}(f)$ *and* $C, c > 0$ *are universal constants.*

Proof. Let f^* be the Gaussian rearrangement of f. Then, f^* is convex, non-decreasing, equi-measurable with f (Lemma 2.2). It follows that:

$$\mathbb{P}\left(f(Z) - M \geqslant t\sqrt{\mathrm{Var}(f)}\right) = \mathbb{P}\left(f^*(\zeta) - f^*(0) \geqslant t\sqrt{\mathrm{Var}(f^*)}\right) \geqslant 1 - \Phi\left(C\left(t + \frac{1}{\mathbf{s}(f^*)}\right)\right),$$

for all $t > 0$, where in the last step we have used Lemma 2.5 (3). Again by Lemma 2.2 we have that $\mathbf{s}(f) \leqslant \mathbf{s}(f^*)$, and the result follows. $\qquad\square$

Using the above distributional inequality we may derive lower estimates for the centered moments in a standard fashion.

Corollary 2.7. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a convex function with* $f \in L_2(\gamma_n)$ *and* $M = \mathrm{med}(f)$. *Then for every* $p \geqslant C/\mathbf{s}(f)^2$ *we have*

(2.3) $$(\mathbb{E}|f(Z) - M|^p)^{1/p} \geqslant c_1\sqrt{p}\sqrt{\mathrm{Var}[f(Z)]},$$

while for $2 \leqslant p \leqslant C/\mathbf{s}(f)^2$ *we get*

(2.4) $$(\mathbb{E}|f(Z) - M|^p)^{1/p} \geqslant c_2\mathbf{s}(f)\sqrt{p}\sqrt{\mathrm{Var}[f(Z)]}.$$

Moreover, if f *is also Lipschitz we obtain*

$$(\mathbb{E}|f(Z) - M|^p)^{1/p} \geqslant c_2'\mathbf{ov}(f)\mathbf{s}(f)\sqrt{p}\mathrm{Lip}(f), \quad p \geqslant 2,$$

where $C, c_1, c_2, c_2' > 0$ *are universal constants.*

Proof. By Proposition 2.6 we have that for every $t \geqslant 1/\mathbf{s}(f)$,

$$\mathbb{P}\left(f(Z) \geqslant M + t\sqrt{\mathrm{Var}[f(Z)]}\right) \geqslant \Phi\left(-C_1t\right) \geqslant e^{-C_2t^2}.$$

It follows that

$$\mathbb{E}|f(Z) - M|^p = p\left(\mathrm{Var}[f(Z)]\right)^{p/2}\int_0^\infty t^{p-1}\mathbb{P}\left(|f(Z) - M| \geqslant t\sqrt{\mathrm{Var}[f(Z)]}\right)dt$$

$$\geqslant p\left(\mathrm{Var}[f(Z)]\right)^{p/2}\int_{1/\mathbf{s}(f)}^\infty t^{p-1}e^{-C_2t^2}dt$$

$$= \frac{p}{2}\left(C_2^{-1}\mathrm{Var}[f(Z)]\right)^{p/2}\int_{C_2/\mathbf{s}(f)^2}^\infty t^{\frac{p}{2}-1}e^{-t}dt.$$

Using the elementary inequality

$$a > 0, q \geqslant a + 1, \quad q \int_a^\infty t^{q-1} e^{-t}\, dt \geqslant (q/e)^q,$$

we conclude the assertion for $p \geqslant \frac{2C_2}{\mathbf{s}(f)^2}$. For $2 \leqslant p \leqslant \frac{2C_2}{\mathbf{s}(f)^2}$ we clearly have

$$\left(\mathbb{E}|f(Z) - M|^p\right)^{1/p} \geqslant \frac{\sqrt{p}\sqrt{\mathrm{Var}[f(Z)]}}{\sqrt{p}} \geqslant \frac{\mathbf{s}(f)}{\sqrt{2C_2}} \sqrt{p}\sqrt{\mathrm{Var}[f(Z)]}.$$

If f is also Lipschitz the assertion follows immediately by using the definitions. $\qquad\square$

We are ready to prove our second main result. As a consequence we obtain the announced Theorem 1.1 in the Introduction. In fact we prove something slightly more:

Theorem 2.8. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a convex, Lipschitz map with* $L = \mathrm{Lip}(f)$. *Then, we have the following:*

$$(2.5) \qquad\qquad \mathbb{P}\left(|f(Z) - \mathrm{med}(f)| \geqslant tL\right) \geqslant c\tau^4 e^{-C\frac{t^2}{\tau^2}\log(e/\tau)}, \quad t > 0,$$

where $\tau = \mathbf{ov}(f)\mathbf{s}(f)$. *In particular, we get:*

$$(2.6) \qquad\qquad \mathbb{P}\left(|f(Z) - \mathrm{med}(f)| \geqslant tL\right) \geqslant c(\mathbf{ov}(f))^8 e^{-C\frac{t^2}{(\mathbf{ov}(f))^4}\log(e/\mathbf{ov}(f))}, \quad t > 0,$$

where $C, c > 0$ *are universal constants.*

Proof. We prove only the first assertion, since the particular case follows from (2.5) by taking into account (2.2). For this end, set $M = \mathrm{med}(f)$ and note that from Corollary 2.7 we have:

$$(2.7) \qquad\qquad \left(\mathbb{E}|f(Z) - M|^p\right)^{1/p} \geqslant c_1 \tau \sqrt{p} L, \quad p \geqslant 2.$$

Set $t_0 := c_1 \tau / \sqrt{2}$. Then, for any $t > 0$ we may choose $p = p(t) = 2\max\{1, (t/t_0)^2\}$ and take into account (2.7) to write:

$$\mathbb{P}\left(|f(Z) - M| \geqslant tL\right) \geqslant \mathbb{P}\left(|f(Z) - M| \geqslant \frac{1}{2}\|f - M\|_p\right) \geqslant (1 - 2^{-p})^2 \left(\frac{\|f - M\|_p}{\|f - M\|_{2p}}\right)^{2p},$$

where in the last step we have used the Paley-Zygmund inequality, see e.g. [BLM13]. One more application of (2.7) in conjunction with $\|f - M\|_{2p} \leqslant C_1 \sqrt{p} L$, yields:

$$\mathbb{P}\left(|f(Z) - M| \geqslant tL\right) \geqslant \frac{1}{2}(c_2 \tau)^{2p} \geqslant c_3 \tau^4 \exp(-C_3 (t/t_0)^2 \log(e/\tau)),$$

as required. The proof is complete. $\qquad\square$

Summarizing we conclude the following characterization of concentration in terms of the Lipschitz constant.

Corollary 2.9. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be convex and Lipschitz map with* $L = \mathrm{Lip}(f)$. *The following are equivalent:*

   a. *For every* $t > 0$ *we have:*

$$\mathbb{P}(|f(Z) - \mathrm{med}(f)| \geqslant tL) \geqslant a_1 \exp(-t^2/A_1^2).$$

   b. *For all* $p \geqslant 2$ *we have:*

$$\left(\mathbb{E}|f(Z) - \mathrm{med}(f)|^p\right)^{1/p} \geqslant A_2 \sqrt{p} L.$$

   c. *We have:*

$$\mathrm{Var}[f(Z)] \geqslant A_3^2 L^2,$$

where $Z \sim N(\mathbf{0}, I_n)$ *and the constants* $a_1, A_1, A_2, A_3 > 0$ *depend on each other.*

Proof. Note that the implications (a) $\Rightarrow$ (b) $\Rightarrow$ (c) are immediate and they hold for any measurable function. The implication (c) $\Rightarrow$ (a) follows from Theorem 2.8. $\qquad\square$

Remarks 2.10. 1. All previous results can be equivalently stated with the mean, in the light of

$$\frac{1}{2}\left(\mathbb{E}|\xi - \xi'|^p\right)^{1/p} \leqslant \left(\mathbb{E}|\xi - m|^p\right)^{1/p} \leqslant 2\left(\mathbb{E}|\xi - \mathbb{E}\xi|^p\right)^{1/p} \leqslant 2\left(\mathbb{E}|\xi - \xi'|^p\right)^{1/p},$$

for all $1 \leqslant p < \infty$, where ξ is any random variable, m a median of ξ, and ξ' an independent copy of ξ.

2. It might worth mentioning that (2.3), should be compared with the known fact for norms [LMS98]:

$$\left(\mathbb{E}\,|h(Z) - M|^p\right)^{1/p} \geqslant c\sqrt{p}\mathrm{Lip}(h), \quad p \geqslant k(h),$$

for any norm h on $\mathbb{R}^n$, where $k(h) = (\mathbb{E}[h(Z)]/\mathrm{Lip}(h))^2$.

3. In the range $t > 1$, we obtain dependence $A_1 \geqslant cA_3$ which is clearly optimal. Furthermore, if $\mathbf{s}(f) \simeq 1$, the above dependence also holds for the full range of t. However, this may suggest that the restriction on t in Proposition 2.6 (or the restriction on p in Corollary 2.7) is redundant. The following example shows that this is not the case.

Example 2.11. Let $\alpha \gg 1$ and let $g_\alpha : \mathbb{R} \to \mathbb{R}$ be the function defined by:

$$g_\alpha(t) = c_\alpha(t - \alpha)_+, \quad c_\alpha = (1 - \Phi(\alpha))^{-1/2}.$$

Then, $\mathbf{s}(g_\alpha) \simeq \alpha^{-1}$ and

$$\mathbb{P}\left(g_\alpha - \mathrm{med}(g_\alpha) > t\sqrt{\mathrm{Var}(g_\alpha)}\right) \leqslant 1 - \Phi\left(\alpha + \frac{ct}{\alpha c_\alpha}\right),$$

for all $t > 0$.

Indeed; we have the asymptotic estimate

(2.8)
$$\int_\alpha^\infty \frac{d\gamma(t)}{1 + t^2} \sim \frac{1 - \Phi(\alpha)}{\alpha^2}, \quad \alpha \to \infty.$$

Thus, we may write:

$$\int \frac{(g_\alpha'(t))^2}{1 + t^2}\,d\gamma(t) = c_\alpha^2 \int_\alpha^\infty \frac{d\gamma(t)}{1 + t^2} \simeq \alpha^{-2}, \quad \alpha \geqslant 2.$$

We may compute that:

$$\mathrm{Var}(g_\alpha) \simeq \int \frac{(g_\alpha'(t))^2}{1 + t^2}\,d\gamma(t) \simeq \alpha^{-2}.$$

In addition we have:

$$\int (g_\alpha'(t))^2\,d\gamma(t) = c_\alpha^2 \int_\alpha^\infty d\gamma(t) = 1.$$

It follows that $s(g_\alpha) \simeq 1/\alpha$ whereas $g_\alpha'(t) = 0$ for $t < \alpha$ and $g_\alpha'(t) = c_\alpha > 0$ for $t > \alpha$. $\qquad\square$

In particular, the above example shows that for $t \simeq \frac{1}{\mathbf{s}(g_\alpha)}$ the estimate in Proposition 2.6 is attained and that one cannot expect super-gaussian behavior for $t \ll \frac{1}{\mathbf{s}(g_\alpha)}$. The discussion shows that there are mainly two reasons for which the classical concentration may fail to give the correct asymptotics. First the super-concentration constant may affect the range of t's and second the over-concentration constant which is apparent on the lower estimate.

2.1. **More comments on the method.** In this subsection we discuss further applications of the methods and techniques used in our result. Mainly, we present several applications of Ehrhard's inequality which we believe are of independent interest.

2.1.1. *On the skewness of Gaussian distribution for convex functions.* Here we show an immediate conse-
quence of Ehrhard's inequality in the spirit of Kwapien's remark from [Kwa94]. The author in [Kwa94]
shows that for any convex function f on $\mathbb{R}^n$ the expectation of $f(Z)$, $Z \sim N(\mathbf{0}, I_n)$ is at least as large as
its median. This fact can be interpreted as the distribution of $f(Z)$ being right-skewed. Another fact which
illustrates this behavior is that the distribution must deviate less below its median than above its median,
which is intuitively clear. Next statement is a rigorous proof of this fact.

Proposition 2.12. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex map. Then, for any $t > 0$ we have:*
$$\mathbb{P}(f(Z) \leqslant \text{med}(f) - t) \leqslant \mathbb{P}(f(Z) > \text{med}(f) + t), \quad Z \sim N(\mathbf{0}, I_n).$$

Proof. Let $m = \text{med}(f)$. We may assume without loss of generality that $\mathbb{P}(f(Z) \leqslant m) = 1/2$ (otherwise
$\inf f = m$ and there is nothing to prove). The map $t \mapsto g(t) = \Phi^{-1} \circ \gamma_n(f \leqslant t)$ is concave and $g(m) = 0$.
Thus,
$$\frac{g(m - t) + g(m + t)}{2} \leqslant g\left(\frac{(m - t) + (m + t)}{2}\right) = g(m) = 0.$$

Finally, recall the property
$$-g(s) = -\Phi^{-1}[\gamma_n(f \leqslant s)] = \Phi^{-1}[1 - \gamma_n(f \leqslant s)] = \Phi^{-1}[\gamma_n(f > s)], \quad s \in \mathbb{R}.$$

Therefore,
$$\Phi^{-1} \circ \mathbb{P}(f(Z) \leqslant m - t) = g(m - t) \leqslant -g(m + t) = \Phi^{-1} \circ \mathbb{P}(f(Z) > m + t).$$

The result follows by the monotonicity of Φ^{-1}. $\qquad\square$

2.1.2. *A small deviation inequality revisited.* The following theorem has been proved in [PV17a] with a worst
(universal) constant. The proof in [PV17a] uses again Ehrhard's inequality but one works with the inverse of
the Gaussian rearrangement. Here we give an alternative short proof using Gaussian rearrangements directly
and we obtain the optimal constant.

Theorem 2.13. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex map with $f \in L_1(\gamma_n)$. Then, we have*
$$\gamma_n\left(\left\{f - \text{med}(f) < -t\|(f - \text{med}(f))_+\|_{L_1(\gamma_n)}\right\}\right) \leqslant \Phi\left(-\frac{t}{\sqrt{2\pi}}\right), \quad t > 0.$$

The equality is attained for affine maps $x \mapsto \langle x, u \rangle + v$, where $u, v \in \mathbb{R}^n$, $u \neq 0$.

Proof. We may assume without loss of generality that $\|(f - \text{med}(f))_+\|_1 > 0$. We introduce the function f^*.
Note that $f^*(0) = \text{med}(f^*) = \text{med}(f) = m$ and $\|(f^* - m)_+\|_{L_1(\gamma)} = \|(f - m)_+\|_{L_1(\gamma_n)}$. Thus, for $u > 0$ we
may write:
$$\gamma_n(x \in \mathbb{R}^n : f(x) - m \leqslant -u) = \gamma(s \in \mathbb{R} : f^*(s) - f^*(0) \leqslant -u) \leqslant \gamma(s : (f^*)'(0-) \cdot s \leqslant -u),$$

where we have used the facts that f^* and f are equi-measurable and $f^*(s) \geqslant f^*(0) + (f^*)'(0-) \cdot s$ for all
$s < 0$, since f^* is convex. Lemma 2.5 shows that:

$$(2.9) \qquad \qquad \|(f^* - m)_+\|_{L_1(\gamma)} \geqslant \frac{(f^*)'(0+)}{\sqrt{2\pi}}.$$

Thus, for $u = t\sqrt{2\pi}\|f - m\|_{L_1(\gamma_n)}$ ($t > 0$ fixed) we may write:
$$\gamma_n\left(x : f(x) - m \leqslant -t\sqrt{2\pi}\|(f - m)_+\|_{L_1}\right) \leqslant \gamma\left(s : (f^*)'(0-) \cdot s \leqslant -t\sqrt{2\pi}\|(f^* - m)_+\|_{L_1}\right) \leqslant \Phi(-t),$$

where in the last step we have used (2.9). $\qquad\square$

It is easy to check that the convexity assumption in the above theorem is essential. Consider the following:

Example 2.14 (The convexity cannot be omitted). Consider the sequence of functions $g_k(t) = t^{2k+1}$, $k =
1, 2, \ldots$ and note that $\text{med}(g_k) = g_k(0) = 0$. Also, $\mathbb{E}(g_k - \text{med}(g_k))_+ = \mathbb{E}(g_k)_+ = \frac{2^k k!}{\sqrt{2\pi}}$. Thus,
$$\mathbb{P}\left(g_k(\zeta) < -t\sqrt{2\pi}\mathbb{E}(g_k(\zeta))_+\right) = \Phi\left(-(t2^k k!)^{\frac{1}{2k+1}}\right).$$

The latter is smaller than $\Phi(-t)$ only when $t < \sqrt{2(k!)^{1/k}} \leqslant \sqrt{2k}$.

2.1.3. *Inequalities for the χ^2 distribution.* Here we present one more application of Ehrhard's inequality. We show that the property that the mapping $t \mapsto \Phi^{-1} \circ \gamma_n(f \leqslant t)$ is concave is shared by other significant distributions at the cost of fairly restricting the class of convex functions. Namely, we have the following:

Proposition 2.15. *Let $k \in \mathbb{N}, k \geqslant 2$ and let $W = (w_1, \ldots, w_n)$ be a random vector with independent coordinates such that $w_j \sim \chi^2(k)$. For any function $f : \mathbb{R}_+^n \to \mathbb{R}$ which is coordinatewise non-decreasing and convex, the mapping*

$$t \mapsto \Phi^{-1} \circ \mathbb{P}(f(W) \leqslant t),$$

is concave.

Proof. Let $F : \mathbb{R}^{kn} \to \mathbb{R}$ be the function defined by:

$$F(x_{11}, \ldots, x_{1n}, \ldots, x_{k1}, \ldots, x_{kn}) = f\left(\sum_{i=1}^{k} x_{i1}^2, \ldots, \sum_{i=1}^{k} x_{in}^2\right).$$

We may check that F is convex. Hence, if $(\zeta_{ij})_{i,j=1}^{k,n}$ are independent with $\zeta_{ij} \sim N(0,1)$, by Ehrhard's inequality we have that the mapping $t \mapsto \Phi^{-1} \circ \mathbb{P}(F(Z) \leqslant t)$ is concave, where $Z = (\zeta_{11}, \ldots, \zeta_{1n}, \ldots, \zeta_{k1}, \ldots, \zeta_{kn}) \sim N(\mathbf{0}, I_{kn})$. Finally, the observation that for $j = 1, 2, \ldots, n$ the random variables $\sum_{i=1}^{k} \zeta_{ij}^2$ are independent and $\chi^2(k)$, yields that

$$\mathbb{P}(F(Z) \leqslant t) = \mathbb{P}(f(W) \leqslant t),$$

for all $t \in \mathbb{R}$. □

As an immediate consequence of Proposition 2.15 we have the following:

Corollary 2.16. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be 1-unconditional and convex function. Then, the mapping*

$$t \mapsto \Phi^{-1} \circ \nu_1^n(f \leqslant t),$$

is concave.

3. An application to finite-dimensional normed spaces

The purpose of this section is to provide an application of the tightness result on the concentration proved in Theorem 2.8 in the context of norms. Namely, we show that for any given norm on $\mathbb{R}^n$ there exists a 5-equivalent norm say, which exhibits optimal Gaussian concentration in terms of its Lipschitz constant. In turn this implies an instability result for the dependence on ε in the almost isometric version of the randomized Dvoretzky's theorem. In order to give the precise statements we have to recall some definitions.

Let $\|\cdot\|$ be an arbitrary norm on $\mathbb{R}^n$ and let $X = (\mathbb{R}^n, \|\cdot\|)$. We define the global parameters of X:

$$b(X) := \max\{\|\theta\| : \|\theta\|_2 = 1\}, \quad k(X) := (\mathbb{E}\|Z\|/b(X))^2, \quad Z \sim N(\mathbf{0}, I_n).$$

The parameter $k(X)$ is usually referred to as the *critical dimension* of the normed space X.

First we show the following instability result for the concentration of norms:

Theorem 3.1. *There exists an universal constant $C \geqslant 1$ with the following property: for any $n \geqslant 1$ and for any norm $\|\cdot\|$ on $\mathbb{R}^n$, there exists a 5-equivalent norm $\|\!|\cdot|\!\|$ such that*

$$\mathbb{P}\left(\left|\|\!|Z|\!\| - \mathbb{E}\|\!|Z|\!\|\right| > \varepsilon\right) \geqslant \frac{1}{C}\exp(-C\varepsilon^2/b(Y)^2), \quad Z \sim N(\mathbf{0}, I_n),$$

for all $\varepsilon > 0$, where $Y = (\mathbb{R}^n, \|\!|\cdot|\!\|)$.

At this point we should mention that, although for any given norm there exists a 5-equivalent norm for which the classical concentration is optimal, there are several examples (established recently) which show that choosing appropriately the position of the norm (via a linear map) one can exhibit better tail estimates; see [PV15] and [Tik17].

3.1. Tilted norms and instability of the concentration. Given any norm $\| \cdot \|$ on $\mathbb{R}^n$, let $x_0^* \in S_{X^*}$ such that $\|x_0^*\|_2 = b(X)$, where $X = (\mathbb{R}^n, \| \cdot \|)$. For any $t > 0$ we define the norm

$$(3.1) \qquad f_t(x) = \|x\| + t|\langle x, x_0^* \rangle|, \quad x \in \mathbb{R}^n.$$

Let also $X_t = (\mathbb{R}^n, f_t)$ be the induced normed space. In the next easily verified lemma we collect some of the basic properties of the norms f_t:

Lemma 3.2. *Let $X = (\mathbb{R}^n, \| \cdot \|)$ be a normed space and let $(f_t)_{t>0}$ be the family of norms defined above.*
   a. *For all $x \in \mathbb{R}^n$ and $t > 0$ we have: $\|x\| \leqslant f_t(x) \leqslant (1+t)\|x\|$.*
   b. *For all $t > 0$ we have: $b_t \equiv \sup\{f_t(\theta) : \|\theta\|_2 = 1\} = (1+t)b = \mathrm{Lip}(f_t)$ and $k_t \equiv k(X_t) = (1+t)^{-2}(\sqrt{k(X)} + t\sqrt{2/\pi})^2$.*
   c. *For $t \geqslant 4$, we have $\sqrt{\mathrm{Var}[f_t(Z)]} \geqslant \frac{1}{8}\mathrm{Lip}(f_t)$.*

Proof. a. Note that for all x, $\|x\| \leqslant f_t(x) \leqslant \|x\| + t\|x\| \cdot \|x_0^*\|_* \leqslant (1+t)\|x\|$.
b. Let $\|\theta_0\|_2 = 1$ with $\langle x_0^*, \theta_0 \rangle = b$. Then, $f_t(\theta_0) = b + tb = (1+t)b$. In addition, we have:

$$(3.2) \qquad \mathbb{E}f_t(Z) = \mathbb{E}\|Z\| + tb\mathbb{E}|\zeta| = b\sqrt{k} + tb\sqrt{2/\pi}, \quad \zeta \sim N(0,1).$$

c. We use the inequality $\sqrt{\mathrm{Var}(\xi_1)} - \sqrt{\mathrm{Var}(\xi_2)} \leqslant \sqrt{\mathrm{Var}(\xi_1 + \xi_2)}$ to write:

$$\sqrt{\mathrm{Var}[f_t(Z)]} \geqslant \sqrt{\mathrm{Var}(t|\langle Z, x_0^* \rangle|)} - \sqrt{\mathrm{Var}\|Z\|} = t\|x_0^*\|_2\sqrt{\mathrm{Var}|\zeta|} - \sqrt{\mathrm{Var}\|Z\|} \geqslant tb\sqrt{\mathrm{Var}|\zeta|} - b.$$

We choose $t \geqslant 4 > 2/\sqrt{\mathrm{Var}|\zeta|}$ to conclude.
$\square$

Theorem 3.1 immediately follows from the next result:

Proposition 3.3. *Let $\| \cdot \|$ be a norm on $\mathbb{R}^n$ and let $t \geqslant 4$. The tilted norms (f_t) defined in (3.1) satisfy:*

$$\mathbb{P}(|f_t(Z) - \mathbb{E}f_t(Z)| \geqslant \varepsilon) \geqslant c_2 e^{-C_2 \varepsilon^2 / b_t^2},$$

for all $\varepsilon > 0$, where $C_2, c_2 > 0$ are universal constants.

Proof. From Lemma 3.2, note that $\sqrt{\mathrm{Var}[f_t(Z)]} \geqslant \frac{1}{8}\mathrm{Lip}(f_t)$ for all $t \geqslant 4$. Thus, by (2.6) we obtain:

$$\mathbb{P}\left(|f_t(Z) - \mathbb{E}[f_t(Z)]| \geqslant \varepsilon\mathrm{Lip}(f_t)\right) \geqslant c_2 e^{-C_2 \varepsilon^2},$$

for all $\varepsilon > 0$, as required.
$\square$

3.2. Random almost spherical sections of convex bodies. The critical dimension $k(X)$ of a normed space was introduced by V. Milman in his work [Mil71] on the random version of Dvoretzky's theorem [Dvo61]:

Theorem 3.4 (Dvoretzky 1961, V. Milman 1971). *For any $\varepsilon \in (0,1)$ there exists $\eta(\varepsilon) > 0$ with the following property: For any $n \geqslant 1$ and any norm $\| \cdot \|$ on $\mathbb{R}^n$, the random k-dimensional subspace F (with respect to the Haar measure $\nu_{n,k}$ on the Grassmannian $G_{n,k}$) satisfies with high probability*

$$(3.3) \qquad (1-\varepsilon)M \leqslant \|y\| \leqslant (1+\varepsilon)M\|y\|_2, \quad y \in F,$$

as long as $k \leqslant \eta(\varepsilon)k(X)$, where $M = \int_{S^{n-1}} \|\theta\| \, d\sigma(\theta)$ and σ is the uniform probability measure on S^{n-1}.

V. Milman's proof provides $\eta(\varepsilon) \simeq \varepsilon^2 / \log(1/\varepsilon)$ while Gordon [Gor85] and Schechtman [Sch89] proved that one can always have $\eta(\varepsilon) \simeq \varepsilon^2$.

For any given normed space $X = (\mathbb{R}^n, \| \cdot \|)$ and $\varepsilon \in (0,1)$ we define $k(X, \varepsilon)$ the maximal positive integer $k \leqslant n$ for which the random k-dimensional subspace F of X is $(1+\varepsilon)$-*spherical*, i.e.

$$\max_{z \in S_F} \|z\| / \min_{z \in S_F} \|z\| < 1 + \varepsilon,$$

with probability at least $2/3$.

With this terminology Theorem 3.4 implies that $k(X, \varepsilon) \geqslant c\varepsilon^2 k(X)$, for any normed space $X = (\mathbb{R}^n, \| \cdot \|)$. Note that there are spaces for which the dependence on ε in the above asymptotic formula is much better, e.g. $k(\ell_\infty^n, \varepsilon) \simeq \frac{\varepsilon}{\log(1/\varepsilon)} k(\ell_\infty^n)$; see [Sch07] and [Tik14]. The reader may consult [MS86] for further background on the local theory of normed spaces.

Using the construction introduced in previous paragraph we may show that for any norm $\|\cdot\|$ on $\mathbb{R}^n$ there exists a t-equivalent norm $f_t(\cdot)$ such that $k(X_t, \varepsilon) \simeq \varepsilon^2 k(X_t)$ for all, not so large, $t > 0$. In geometric language this can be interpreted as follows: in the space of centrally symmetric n-dimensional convex bodies, the ones which admit random almost spherical sections, with high probability, in dimension at most $C\varepsilon^2 k(X)$ form a C_0-net with respect to the geometric distance. More precisely we prove the following:

Theorem 3.5 (instability)**.** *There exists an universal constant $C > 1$ with the following property: For any normed space $X = (\mathbb{R}^n, \|\cdot\|)$ with $C \leqslant t \leqslant \sqrt{k(X)}$, the normed spaces $X_t = (\mathbb{R}^n, f_t)$ defined in (3.1) satisfy:*

    a. *For all $x \in \mathbb{R}^n$ we have $\|x\| \leqslant f_t(x) \leqslant 2t\|x\|$.*

    b. *For every $\varepsilon \in (0, 1/3)$ one has $k(X_t, \varepsilon) \simeq \varepsilon^2 k(X_t)$.*

Proof. The argument follows the same lines as in [Sch07] (see also [PVZ17, Section 5]), hence we roughly sketch the details for reader's convenience. Let $\varepsilon \in (0, \frac{1}{3})$ and let the set of k-dimensional subspaces of X_t,

$$\mathcal{F}_\varepsilon := \left\{ F \in G_{n,k} \mid (1+\varepsilon)^{-1} M_F \leqslant f_t(\theta) \leqslant (1+\varepsilon) M_F \ \forall \theta \in S_F \right\},$$

where $M_F = \int_{S_F} f_t(u)\, d\sigma_F(u)$ and σ_F denotes the uniform probability measure on the sphere $S_F = S^{n-1} \cap F$. Define further,

$$\mathcal{B}_\varepsilon := \left\{ F \in \mathcal{F}_\varepsilon \mid (1 - 2\varepsilon)\frac{\mathbb{E}[f_t(Z)]}{\mathbb{E}\|Z\|_2} \leqslant M_F \leqslant (1 + 2\varepsilon)\frac{\mathbb{E}[f_t(Z)]}{\mathbb{E}\|Z\|_2} \right\}.$$

An application of [Sch07, Lemma 1] yields:

$$\nu_{n,k}(\mathcal{F}_\varepsilon) = \nu_{n,k}(\mathcal{F}_\varepsilon \setminus \mathcal{B}_\varepsilon) + \nu_{n,k}(\mathcal{B}_\varepsilon)$$

$$\leqslant \left[\mathbb{P}\left(\left\{ f_t(Z) \geqslant \frac{1 + 2\varepsilon}{1 + \varepsilon}\frac{\mathbb{E}[f_t(Z)]}{\mathbb{E}\|Z\|_2}\|Z\|_2 \text{ or } f_t(Z) \leqslant (1 + \varepsilon)(1 - 2\varepsilon)\frac{\mathbb{E}[f_t(Z)]}{\mathbb{E}\|Z\|_2}\|Z\|_2 \right\} \right) \right]^k +$$

$$\left[\mathbb{P}\left(\left\{ \frac{1 - 2\varepsilon}{1 + \varepsilon}\|Z\|_2\frac{\mathbb{E}[f_t(Z)]}{\mathbb{E}\|Z\|_2} \leqslant f_t(Z) \leqslant (1 + \varepsilon)(1 + 2\varepsilon)\frac{\mathbb{E}[f_t(Z)]}{\mathbb{E}\|Z\|_2}\|Z\|_2 \right\} \right) \right]^k.$$

In order to proceed we will need the following estimate:

Lemma 3.6. *For any $C_4 \leqslant t \leqslant \sqrt{k(X)}$ and for every $0 < \delta < 1/3$ we have:*

$$c_4 e^{-C_4 \delta^2 k_t} \leqslant \mathbb{P}\left(f_t(Z) \leqslant \frac{(1 - \delta)\mathbb{E}[f_t(Z)]}{\mathbb{E}\|Z\|_2}\|Z\|_2 \text{ or } f_t(Z) \geqslant \frac{(1 + \delta)\mathbb{E}[f_t(Z)]}{\mathbb{E}\|Z\|_2}\|Z\|_2 \right) \leqslant C_4 e^{-c_4 \delta^2 k_t}.$$

Taking Lemma 3.6 for granted it suffices to consider $1/\sqrt{k_t} < \varepsilon < 1/3$, hence we obtain:

$$\nu_{n,k}(\mathcal{F}_\varepsilon) \leqslant C_5^k e^{-k\varepsilon^2 k_t / C_5} + (1 - C_5^{-1} e^{-C_5 \varepsilon^2 k_t})^k$$

$$\leqslant e^{-\frac{1}{2C_5} k\varepsilon^2 k_t} + 1 - C_5^{-1} e^{-C_5 \varepsilon^2 k_t},$$

provided that $\varepsilon^2 k_t > 2C_5 \log C_5$. Now assuming that $\nu_{n,k}(\mathcal{F}_\varepsilon) \geqslant 1 - e^{-\beta k} \geqslant 2/3$ for some universal constant $\beta > 0$ and restricting further $\max\{2\beta C_5, 2C_5 \log C_5\} < \varepsilon^2 k_t$, we obtain:

$$1 - e^{-2C_5 \varepsilon^2 k_t} \geqslant 1 - C_5^{-1} e^{-C_5 \varepsilon^2 k_t} \geqslant 1 - e^{-\beta k} - e^{-\frac{1}{2C_5} k\varepsilon^2 k_t} \geqslant 1 - 2e^{-\beta k} \geqslant 1 - e^{-c_0 \beta k},$$

which implies $k \leqslant \frac{2C_5}{c_0 \beta}\varepsilon^2 k_t$, as required. $\qquad\square$

Proof of Lemma 3.6. Let $\xi = f_t(Z)/\mathbb{E}[f_t(Z)]$ and $\eta = \|Z\|_2/\mathbb{E}\|Z\|_2$. For any $s \in (0,1)$ we define the sets:

$$A_s = \{|\xi - \eta| > s\eta\}, \quad B_s = \{|\eta - 1| > s\}, \quad \Gamma_s = \{|\xi - 1| > s\}.$$

Our aim is to show that:

$$c_4 e^{-C_4 \delta^2 k_t} \leqslant \mathbb{P}(A_\delta) \leqslant C_4 e^{-c_4 \delta^2 k_t}.$$

Note that for any $0 < s < 1/2$ we have

$$\mathbb{P}(A_s) \leqslant \mathbb{P}(\Gamma_{s/4}) + \mathbb{P}(B_{s/2}).$$

Using the Gaussian concentration for $f_t(\cdot)$ and $\|\cdot\|_2$ we infer:

$$\mathbb{P}(A_\delta) \leqslant C_0 e^{-c_0 \delta^2 k_t} + C_0 e^{-c_0 \delta^2 n} \leqslant 2C_0 e^{-c_0 \delta^2 k_t}, \quad 0 < \delta < 1/3,$$

where we have also used the general fact that $k(X) \leqslant n$. This proves the rightmost inequality.

For the lower estimate note that for any $0 < s < 1$ one has

$$\mathbb{P}(A_{s/3}) \geqslant \mathbb{P}(\Gamma_s) - \mathbb{P}(B_{s/3}).$$

Using the Gaussian concentration for $\|\cdot\|_2$ and Proposition 3.3 we obtain:

$$\mathbb{P}(A_\delta) \geqslant c_2 e^{-9C_2\delta^2 k_t} - C_0 e^{-c_0\delta^2 n},$$

provided that $t \geqslant 4$, where $C_2 > 0$ is the constant from Proposition 3.3. The latter is larger than $\frac{c_2}{2} e^{-9C_2\delta^2 k_t}$ provided that $(c_0 n - 9C_2 k_t)\delta^2 \geqslant \log(2C_0/c_0)$. Note that since $4 \leqslant t \leqslant \sqrt{k(X)}$, Lemma 3.2.b yields $k_t \leqslant C_3 t^{-2} k(X) \leqslant C_3 n/t^2$, thus it suffices to have $(c_0 n - 9C_2 C_3 t^{-2} n)\delta^2 \geqslant \log(2C_0/c_0)$. The last one holds if $t \geqslant \max\{4, \sqrt{18C_2 C_3/c_0}\}$ and $\delta \geqslant \sqrt{\frac{2}{c_0 n} \log(\frac{2C_0}{c_0})}$. The assertion of the lemma follows. $\square$

References

[BH99]   S. G. Bobkov and C. Houdré. A converse Gaussian Poincaré-type inequality for convex functions. *Statist. Probab. Lett.*, 44(3):281–290, 1999.

[BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities.* Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.

[Bor75]   C. Borell. The Brunn-Minkowski inequality in Gauss space. *Invent. Math.*, 30(2):207–216, 1975.

[Bor03]   C. Borell. The Ehrhard inequality. *C. R. Math. Acad. Sci. Paris*, 337(10):663–666, 2003.

[CEL12] D. Cordero-Erausquin and M. Ledoux. Hypercontractive measures, Talagrand's inequality, and influences. In *Geometric aspects of functional analysis*, volume 2050 of *Lecture Notes in Math.*, pages 169–189. Springer, Heidelberg, 2012.

[Cha14] S. Chatterjee. *Superconcentration and related topics.* Springer Monographs in Mathematics. Springer, Cham, 2014.

[Che81] H. Chernoff. A note on an inequality involving the normal distribution. *Ann. Probab.*, 9(3):533–535, 1981.

[Che82] L. H. Y. Chen. An inequality for the multivariate normal distribution. *J. Multivariate Anal.*, 12(2):306–315, 1982.

[Dvo61] A. Dvoretzky. Some results on convex bodies and Banach spaces. In *Proc. Internat. Sympos. Linear Spaces (Jerusalem, 1960)*, pages 123–160. Jerusalem Academic Press, Jerusalem; Pergamon, Oxford, 1961.

[Ehr83] A. Ehrhard. Symétrisation dans l'espace de Gauss. *Math. Scand.*, 53(2):281–301, 1983.

[Ehr84a] A. Ehrhard. Inégalités isopérimétriques et intégrales de Dirichlet gaussiennes. *Ann. Sci. École Norm. Sup. (4)*, 17(2):317–332, 1984.

[Ehr84b] A. Ehrhard. Sur l'inégalité de Sobolev logarithmique de Gross. In *Seminar on probability, XVIII*, volume 1059 of *Lecture Notes in Math.*, pages 194–196. Springer, Berlin, 1984.

[Gor85] Y. Gordon. Some inequalities for Gaussian processes and applications. *Israel J. Math.*, 50(4):265–289, 1985.

[Gro75] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4):1061–1083, 1975.

[HW71] D. L. Hanson and F. T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Statist.*, 42:1079–1083, 1971.

[IV15]   P. Ivanisvili and A. Volberg. Bellman partial differential equation and the hill property for classical isoperimetric problems. `https://arxiv.org/abs/1506.03409`, (2015). preprint.

[Kwa94] S. Kwapień. A remark on the median and the expectation of convex functions of Gaussian vectors. In *Probability in Banach spaces, 9 (Sandjberg, 1993)*, volume 35 of *Progr. Probab.*, pages 271–272. Birkhäuser Boston, Boston, MA, 1994.

[Led01] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI, 2001.

[LMS98] A. E. Litvak, V. D. Milman, and G. Schechtman. Averages of norms and quasi-norms. *Math. Ann.*, 312(1):95–124, 1998.

[LT91]   M. Ledoux and M. Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)].* Springer-Verlag, Berlin, 1991. Isoperimetry and processes.

[Mil71]   V. D. Milman. A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies. *Funkcional. Anal. i Priložen.*, 5(4):28–37, 1971.

[MS86]   V. D. Milman and G. Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov.

[Nao07] A. Naor. The surface measure and cone measure on the sphere of l_p^n. *Trans. Amer. Math. Soc.*, 359(3):1045–1079, 2007.

[Nel67]   E. Nelson. *Dynamical theories of Brownian motion.* Princeton University Press, Princeton, N.J., 1967.

[NP16]   J. Neeman and G. Paouris. An interpolation proof of Ehrhard's inequality. `https://arxiv.org/abs/1605.07233`, (2016). preprint.

[PV15]   G. Paouris and P. Valettas. On Dvoretzky's theorem for subspaces of L_p. `https://arxiv.org/abs/1510.07289`, (2015). preprint.

[PV17a] G. Paouris and P. Valettas. A Gaussian small deviation inequality for convex functions. *The Annals of Probability*, 2017. (to appear).

[PV17b] G. Paouris and P. Valettas. Variance estimates and almost Euclidean structure. `https://arxiv.org/abs/1703.10244`, (2017). preprint.

[PVZ17] G. Paouris, P. Valettas, and J. Zinn. Random version of Dvoretzky's theorem in ℓ_p^n. *Stochastic Processes and their Applications*, (2017). (to appear).

[RV13]   M. Rudelson and R. Vershynin. Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.*, 18:no. 82, 9, 2013.

[Sch89] G. Schechtman. A remark concerning the dependence on ϵ in Dvoretzky's theorem. In *Geometric aspects of functional analysis (1987–88)*, volume 1376 of *Lecture Notes in Math.*, pages 274–277. Springer, Berlin, 1989.

[Sch07] G. Schechtman. The random version of Dvoretzky's theorem in ℓ_∞^n. In *Geometric aspects of functional analysis*, volume 1910 of *Lecture Notes in Math.*, pages 265–270. Springer, Berlin, 2007.

[ST74]   V. N. Sudakov and B. S. Tsirel'son. Extremal properties of half-spaces for spherically invariant measures. *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 41:14–24, 165, 1974. Problems in the theory of probability distributions, II.

[Tal91] M. Talagrand. A new isoperimetric inequality and the concentration of measure phenomenon. In *Geometric aspects of functional analysis (1989–90)*, volume 1469 of *Lecture Notes in Math.*, pages 94–124. Springer, Berlin, 1991.

[Tal94] M. Talagrand. On Russo's approximate zero-one law. *Ann. Probab.*, 22(3):1576–1587, 1994.

[Tal95] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.*, (81):73–205, 1995.

[Tik14] K. E. Tikhomirov. The randomized Dvoretzky's theorem in l_∞^n and the χ-distribution. In *Geometric aspects of functional analysis*, volume 2116 of *Lecture Notes in Math.*, pages 455–463. Springer, Cham, 2014.

[Tik17] K. E. Tikhomirov. Superconcentration, and randomized Dvoretzky's theorem for spaces with 1-unconditional bases. `https://arxiv.org/abs/1702.00859`, (2017). preprint.

[vH17]   R. van Handel. The Borell-Ehrhard game. *Probability Theory and Related Fields*, 2017. (to appear).

MATHEMATICS DEPARTMENT, UNIVERSITY OF MISSOURI, COLUMBIA, MO, 65211.

E-mail address: `valettasp@missouri.edu`