arXiv:1906.03208v1 [math.FA] 7 Jun 2019

HYPERCONTRACTIVITY, AND LOWER DEVIATION ESTIMATES IN NORMED SPACES

GRIGORIS PAOURIS, KONSTANTIN TIKHOMIROV, AND PETROS VALETTAS

ABSTRACT. We consider the problem of estimating probabilities of lower deviation $\mathbb{P}\{\|G\| \leqslant \delta\mathbb{E}\|G\|\}$ in normed spaces with respect to the Gaussian measure. These estimates occupy central role in the probabilistic study of high-dimensional structures. It has been confirmed in several concrete situations, using ad hoc methods, that lower deviations exhibit very different and more complex behavior than the corresponding upper estimates. A characteristic example of this diverse behavior is $\|G\|_\infty = \max_{i\leqslant n}|g_i|$, where g_i are i.i.d. standard normals. In this work we develop a general method for proving small ball lower deviation estimates for norms. In the case of 1–unconditional norms we show that under certain balancing conditions, which are fulfilled in many classical situations, our bounds are best possible up to numerical constants. We also study the lower small deviation estimates for both 1-unconditional and general norms and we obtain optimal results. In all regimes, $\|G\|_\infty$ arises (at the computational level) as an extremal case in the problem. The proofs exploit the convexity and hypercontractive properties of the Gaussian measure.

Contents

2010 *Mathematics Subject Classification.* Primary 46B07; Secondary 52A21.
Key words and phrases. Talagrand's $L_1 - L_2$ bound, Alon–Milman theorem, Ornstein–Uhlenbeck semigroup, Gaussian convexity, hypercontractivity, superconcentration.
The first author was supported by the NSF grant DMS-1812240.
The second author was partially supported by the Simon foundation.
The third author was supported by the NSF grant DMS-1612936.

1. Introduction

The concentration of measure phenomenon is one of the most important concepts in high-dimensional probability and is an indispensable tool in the study of high-dimensional structures that arise in theoretical and applied fields. In its most simple form it can be stated as follows: functions depending smoothly on many independent variables have small fluctuations. More specifically, in Gauss' space, asserts that any function f on $\mathbb{R}^n$ which is L-Lipschitz is almost constant with overwhelming probability:

$$\max\left(\mathbb{P}\{f(G) \leqslant \mathbb{E}\, f(G) - tL\}, \mathbb{P}\{f(G) \geqslant \mathbb{E}\, f(G) + tL\}\right) \leqslant \exp(-t^2/2), \quad t > 0, \tag{1.1}$$

where G is the standard Gaussian vector in $\mathbb{R}^n$, see e.g. [Pis86, Mau91]. This probabilistic phenomenon is usually addressed as consequence of the solution to the Gaussian isoperimetric problem, which was solved independently by Sudakov and Tsirel'son [ST74] and by Borell [Bor75]. The Gaussian isoperimetric inequality asserts that among all Borel sets in $\mathbb{R}^n$ of a given Gaussian measure, the half-spaces have the smallest Gaussian surface area.

Notably, the first groundbreaking application of the concentration of measure in high-dimensional geometry is the seminal work of V. Milman [Mil71] which provides an almost Euclidean structure of optimal dimension, in any normed space. Applying (1.1) for f being a norm (and for $t = \varepsilon \mathbb{E} f(G)$, $\varepsilon > 0$) we obtain

$$\max\left(\mathbb{P}\{f(G) \leqslant (1-\varepsilon)\mathbb{E}\, f(G)\}, \mathbb{P}\{f(G) \geqslant (1+\varepsilon)\mathbb{E}\, f(G)\}\right) \leqslant \exp(-\tfrac{1}{2}\varepsilon^2 k(f)), \quad \varepsilon > 0,$$

where $k(f) := (\mathbb{E} f(G)/L)^2$ is usually referred to as the *critical dimension* or *Dvoretzky number* of f. It is well known that for norms in the *large deviation regime*, that is for $\varepsilon > 1$, the above estimate is optimal up to universal constants, see e.g. [LT91, Corollary 3.2], [LMS98], [PVZ17, Proposition 2.10]:

$$\mathbb{P}\{f(G) \geqslant (1+\varepsilon)\mathbb{E} f(G)\} \geqslant c \exp(-C\varepsilon^2 k(f)), \quad \varepsilon > 1,$$

where $c, C > 0$ are universal constants.[1] However, depending on the applications, different regimes of the distribution are of interest, and the concentration (for norms) in terms of the Lipschitz constant has a major drawback: it provides suboptimal bounds in several key situations. For example, the behavior in the *small deviation regime*, $\mathbb{P}\{|f(G) - \mathbb{E} f(G)| \geqslant \varepsilon \mathbb{E} f(G)\}, \varepsilon \in (0,1)$ is much more complicated (see e.g. [PVZ17, LT17], for $f(x) = \|x\|_p$). Small changes of the function can considerably affect the estimates in this regime as opposed to the behavior in the large deviation regime which will remain essentially the same, since it is controlled by the Lipschitz constant. To illustrate this instability behavior let us point out that there exist examples of norms $\|\cdot\|, |||\cdot|||$ in $\mathbb{R}^n$, which are isomorphic, i.e. $\|x\| \leqslant |||x||| \leqslant C\|x\|$ for all x, yet $\mathrm{Var}(\|G\|) \asymp 1$ and $\mathrm{Var}(|||G|||) \asymp n^{-c}$; see e.g.[PVZ17], [Val17]. Even more surprising is the case of the ℓ_p norms for p varying with n, where for the variance the transition from polynomially to logarithmically small order happens at an almost isometric scale; see [LT17]. It turns out that the variance estimate

$$\mathrm{Var} f(G) \lesssim L^2, \tag{1.2}$$

provided by (1.1) it can be rough even for nice functions. Conversely, it is proved in [Val17] that (1.1) is tight for convex Lipschitz functions in full regime if and only if $\mathrm{Var} f(G) \asymp L^2$.

Improving upon the variance bounds, one may employ the classical Poincaré inequality [Che82]

$$\mathrm{Var} f(G) \leqslant \mathbb{E}\|\nabla f(G)\|_2^2. \tag{1.3}$$

Although this inequality captures bounds of the right order in the previously discussed case of the ℓ_p-norm (p fixed), it starts being loose when p grows with n. In particular, for the ℓ_∞-norm one has (see e.g. [Cha14, PVZ17, LT17]) that

$$\mathrm{Var}[\|G\|_\infty] \asymp (\log n)^{-1} \lll 1 = \mathbb{E}\|\nabla \|G\|_\infty\|_2^2.$$

This is a typical situation of the so-called *superconcentration* phenomenon. We refer the reader to Chatterjee's monograph [Cha14] for this very interesting topic. The latter phenomenon can also be observed when one

[1] We shall make frequent use of the letters $c, C, c_1, \ldots$ throughout the text for universal constants whose value may change from line to line. We also use the (standard) asymptotic notation: For any two quantities Q_1, Q_2 we write $Q_1 \lesssim Q_2$ if there exists universal constant $c > 0$ such that $Q_1 \leqslant cQ_2$. We also write $Q_2 \gtrsim Q_1$ if $Q_1 \lesssim Q_2$. Finally, we write $Q_1 \asymp Q_2$ if $Q_1 \lesssim Q_2$ and $Q_2 \lesssim Q_1$.

compares the behavior of $\|G\|_\infty$ in the (right) small deviation regime with the one predicted by (1.1). Given that $k(\|\cdot\|_\infty) \asymp \log n$, (1.1) yields

$$\mathbb{P}\{|\|G\|_\infty - \mathbb{E}\|G\|_\infty| \geqslant \varepsilon \mathbb{E}\|G\|_\infty\} \leqslant C\exp(-c\varepsilon^2 \log n), \quad \varepsilon \in (0,1).$$

It is known that in this regime the bound is sub-exponential rather than sub-gaussian, namely

$$\mathbb{P}\{|\|G\|_\infty - \mathbb{E}\|G\|_\infty| \geqslant \varepsilon \mathbb{E}\|G\|_\infty\} \leqslant C\exp(-c\varepsilon \log n), \quad \varepsilon \in (0,1).$$

There exist already several proofs available in the literature for the upper bound, e.g. [Tal91, Sch07, PVZ17] and in [Sch07] it is also computed that this estimate is sharp (up to universal constants),

$$c\exp(-C\varepsilon \log n) \leqslant \mathbb{P}\{\|G\|_\infty > (1+\varepsilon)\mathbb{E}\|G\|_\infty\}, \quad \varepsilon \in (0,1). \tag{1.4}$$

From these approaches, only [Tal91] and [PVZ17] use general concentration tools, hoping for further extension in the general case, while [Sch07] makes explicit use of the stochastic independence of $|g_i|$'s, so that deviation estimates for $\|G\|_\infty$ boil down to tedious computations. Some progress in this direction was achieved in [Tik18] where, exploiting ideas from [PVZ17], it is proved that for any 1-unconditional norm $\|\cdot\|$ in $\mathbb{R}^n$, there exists an invertible linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$\mathbb{P}\left\{\left|\|TG\| - \mathbb{E}\|TG\|\right| > \varepsilon \mathbb{E}\|TG\|\right\} \leqslant C\mathbb{P}\left\{\left|\|G\|_\infty - \mathbb{E}\|G\|_\infty\right| > c\varepsilon \mathbb{E}\|G\|_\infty\right\}, \quad \varepsilon \in (0,1). \tag{1.5}$$

Let us emphasize that if we want to find good tail estimates, the linear structure should be appropriately chosen by a linear map to avoid degeneracies. The case of an arbitrary norm was settled recently in [PV18a], thus establishing that for the concentration of norms in the small deviation regime, the ℓ_∞ is the (approximately) extremal structure.

The latter work indicates that there are roughly two heuristic principles that are responsible for uniformity in high-dimensional normed spaces. Either they have relatively large "typical" parts in which case they are described by the standard concentration, or they contain an extremal geometric structure which equips the system with super-concentration properties (we note that a related dichotomy was earlier observed in [Sch06], [AM83]). This leads to new concentration phenomena that had not been observed in previously studied regimes. Given the importance of the concentration of measure in applications, it is vital to develop new methods for obtaining finer bounds, as well as, to understand the underlying principles and possible limitations in these phenomena. The present work is a step towards this understanding, with a special focus on the class of norms. The main goal is to study systematically the lower deviation regimes, building on ideas developed in our previous works [Tik18, PV19, PV18a] and introduce a new perspective as well.

Let us stress the fact that the situation below the mean is radically different and by far more "fractious". For context and comparison let us consider the diverse distributional behavior of $\|G\|_\infty$. We have already discussed the upper regime, hence we shall focus in the deviation below the mean. A careful study shows that the lower small deviation estimate differs substantially. Namely, it is shown in [Sch07] that

$$c\exp(-n^{C\varepsilon}) < \mathbb{P}\left\{\|G\|_\infty < (1-\varepsilon)\mathbb{E}\|G\|_\infty\right\} < C\exp(-n^{c\varepsilon}), \quad \varepsilon \in (0,1/2), \tag{1.6}$$

while in [KV07] it was noted that in the *small ball regime* one has

$$\exp(-Cn^{1-c\delta^2}) < \mathbb{P}\left\{\|G\|_\infty < \delta \mathbb{E}\|G\|_\infty\right\} < \exp(-cn^{1-C\delta^2}), \quad \delta \in ((\log n)^{-1/2}, 1/2). \tag{1.7}$$

Already these estimates indicate that the concentration below the mean might be much stronger, nonetheless they give no clue what concentration inequalities can be used to obtain such bounds in the general case. Although small ball probabilities occupy central role in high-dimensional probability and have been verified in concrete problems using ad hoc methods, few general estimates are available. Small ball probabilities for norms with respect to log-concave measures have been studied in [Lat99, Gue99, Pao12]. We would like to emphasize that in this note small ball probabilities correspond to deviations inequalities from the *mean* of the random variable. Small ball probabilities of the form $\mathbb{P}(\|X\| < \varepsilon)$ have been studied in [FGG16]. However these results are not comparable with the ones obtained here.

The first small ball estimate, with respect to the Gaussian measure, valid for any norm was given by Latala and Oleszkiewicz in [LO05] asserting that

$$\mathbb{P}\left\{\|G\| < \varepsilon \mathbb{E}\|G\|\right\} < (C\varepsilon)^{ck}, \quad \varepsilon \in (0,1/2), \tag{1.8}$$

where k is the Dvoretzky number of the norm, $k = k(\|\cdot\|) = (\mathbb{E}\|G\|/L)^2$. Their proof makes use of the Gaussian isoperimetric inequality and the B-inequality due to Cordero-Erausquin, Fradelizi and Maurey [CEFM04]. Soon after, Klartag and Vershynin in [KV07] emphasized that stronger small-ball estimates can be obtained. To this end, they associated with any norm f the following parameter

$$d(f,\delta) := -\log \mathbb{P}\left\{f(G) \leqslant \delta \mathrm{med}(f(G))\right\}, \quad \delta \in (0,1), \tag{1.9}$$

and they showed, using the B-inequality, that

$$\mathbb{P}\left\{f(G) \leqslant \varepsilon \mathrm{med}(f(G))\right\} \leqslant \frac{1}{2}\varepsilon^{\frac{d(f,\delta)-\log 2}{\log(1/\delta)}}, \quad \varepsilon \in (0,\delta). \tag{1.10}$$

Note that this formulation recovers Latala–Oleszkiewicz's bound: for $\delta = 1/2$ in the definition of d and by taking into account (1.1) we obtain $d(f,1/2) \gtrsim k(f)$. This general lower bound is insufficient to retrieve the bound in the case of ℓ_∞^n. Note that in the light of (1.7) we have $d(\ell_\infty^n,\delta) \asymp n^{1-c\delta^2}$, whereas $k(\ell_\infty^n) \asymp \log n$. Let us emphasize the fact that the discussed general estimates use mainly the Gaussian isoperimetry and the metric properties of the norm, i.e. the Lipschitz constant.

It turns out that the main reason behind the stronger estimates is different and relies on geometric properties, namely the convexity of both the distribution and the norm. Let us illustrate why one can hope for stronger concentration below the mean under these circumstances. It was observed in [Val17, Proposition 2.12] that the distribution of a convex function of a Gaussian random vector is skew to the right. Although this shows that any convex function deviates less to the left than to the right, this observation doesn't quantify this difference. For this purpose, improved lower deviation inequalities are available for convex functions. As a consequence of the Gaussian isoperimetry and the convexity of f one can obtain an improved version of the lower deviation in (1.1) where the Lipschitz constant of f is replaced by the expected length of the gradient. Further improvements can be achieved if we additionally exploit the Gaussian convexity. This was established in [PV18b], where one replaces the Lipschitz constant now in the lower deviation in (1.1) by the variance. Key ingredient of the proof is Ehrhard's inequality [Ehr83] which can be viewed as the Gaussian counterpart of the classical Brunn–Minkowski inequality for the n-dimensional volume. Although this significant improvement provides a better general bound for $d(f,1/2)$, namely

$$d(f,1/2) \gtrsim 1/\beta(f) := \mathrm{Var}[f(G)]/(\mathbb{E}f(G))^2 \gtrsim k(f), \tag{1.11}$$

it still fails to recover the optimal dependence in the small ball regime in key situations. For example, it is known (see e.g. [PV19]) that $\beta(\|G\|_\infty) \asymp (\log n)^{-2}$, while (1.7) shows that the right bound is of polynomial order. At this point we would like to suggest a plausible explanation of this deficiency. All the aforementioned approaches take into account only the decay of the tail in the microscopic regime $((1-\varepsilon)\mathbb{E}\|G\|, \mathbb{E}\|G\|)$, $\varepsilon \in (0,1/2)$. However, (1.9) and (1.10) indicate that the phenomenon starts becoming more dramatic, in view of the monotonicity of the function $\delta \mapsto d(f,\delta)$, when we transition in the macroscopic regime $(0, \delta\mathbb{E}\|G\|)$, $\delta \in (0,1/2)$. In addition, in this regime, the small deviation estimate is no more elucidative, since it only provides a bound of constant order $e^{-c/\beta}$ for *all* $\delta \in (0,1/2)$. On the other hand it is not difficult to check, based on volumetric arguments, that already for sufficiently small $\delta > 0$ the small ball estimate behaves as δ^{cn}. Therefore, this phenomenon cannot occur from deficiency of the variance-sensitive small deviation inequality, but rather from improper perspective and specifics of the situation. The main challenge that motivates the present work is to overcome this obstacle and upgrade the existing probabilistic tools in order to establish the correct order of magnitude for the probability $\mathbb{P}\{\|G\| \leqslant \delta\mathbb{E}\|G\|\}$, $\delta \in (0,1/2)$.

Our method makes essential use of convexity in various stages. Let us demonstrate some key ingredients of the method. A main feature, as we have already discussed, is that convex functions exhibit stronger one-sided concentration in terms of the variance. Further, smoothening via the Ornstein-Uhlenbeck semigroup P_t endows the function with stronger concentration properties, e.g., it is known that $\mathrm{Var}(P_t f)$ decays exponentially fast with time [BGL14]. In addition, $P_t f$ well approximates f and interpolates between $f = P_0 f$ and $\mathbb{E}f = P_\infty f$ (due to ergodicity), while the expectation is unaltered during the ensuing motion, i.e. $\mathbb{E}P_t f = \mathbb{E}f$, for all $t \geqslant 0$. Introducing the semigroup has a twofold purpose. On the one hand, in view of the above approximation, we may replace f by $P_t f$ and reduce the problem of estimating $\mathbb{P}\{f(G) \leqslant \delta\mathbb{E}f(G)\}$ to a deviation for $P_t f$, which also enjoys better concentration properties. This becomes of particular importance, once it is combined with the simple, but crucial, observation that P_t preserves the convexity of f thanks to

Mehler's formula. Hence, one may apply the stronger variance-sensitive deviation estimate for convex functions to $P_t f$. On the other hand, the latter application of the inequality is not loose in the small deviation regime for $P_t f$, after specific time t. In other words, P_t "lifts" f in order to "slide" the small ball regime of f to the lower small-deviation regime of $P_t f$, where deviation estimate provides bound of the right order of magnitude. Roughly speaking, we have an inequality of the form

$$\mathbb{P}\{f(G) < \delta \mathbb{E} f(G)\} \leqslant \mathbb{P}\{P_t f(G) < \varepsilon(t, \delta) \mathbb{E} P_t f(G)\},$$

for which one would like to choose t as large as possible, when $P_t f$ has almost no deviation. The admissible range for $\varepsilon(t, \delta)$ determines the specific time t that has to be chosen with respect to δ in order to obtain the desired estimate. The goal is to efficiently estimate the decay of the parameters that govern the deviation inequality for $P_t f$ as $t \to \infty$. In the case of the variance, this is done by quantifying the superconcentation of $P_t f$ using hypercontractivity. The superconcentration phenomenon (as presented by Chatterjee [Cha14]) has played central role in the recent developments in the concentration of norms (see [Tik18], [PV18a]).

The described method yields general bounds for $\mathbb{P}\{f(G) < \delta \mathbb{E} f(G)\}$ which involve analytic and statistical (global) parameters of f, e.g. the L_1 norm of the partial derivatives $\partial_i f$ and the variance of f. Later, these bounds are used to obtain sharp small ball estimates for norms that satisfy certain balancing conditions. A norm f in $\mathbb{R}^n$ is said to be in *ℓ-position* if it satisfies

$$\mathbb{E}[\langle G, e_i \rangle \partial_i f(G) f(G)] = \frac{\mathbb{E}\|G\|^2}{n}, \quad i = 1, 2, \ldots, n.$$

We will also consider the case where the norm f satisfies

$$\mathbb{E}|\partial_i f(G)|^p = \mathbb{E}|\partial_j f(G)|^p, \quad i, j = 1, \ldots, n,$$

for $p > 0$ (of particular interest for us is the case $p = 1$). For the purpose of this work, in the latter case, we will say that f satisfies the *$w^{1,p}$–condition* or that the unit ball B_X of $X = (\mathbb{R}^n, \|\cdot\|)$ is in *$w^{1,p}$–position.* Recall that a norm $\|\cdot\|$ in $\mathbb{R}^n$ is said to be 1–unconditional, if it satisfies $\|\sum_{i=1}^n \varepsilon_i \alpha_i e_i\| = \|\sum_{i=1}^n \alpha_i e_i\|$ for all scalars $(\alpha_i) \subset \mathbb{R}$ and any choice of signs $\varepsilon_i = \pm 1$. With this terminology we prove the following:

Theorem 1.1 (Small ball estimates for norms)**.** *Let $\|\cdot\|$ be a norm in $\mathbb{R}^n$.*

(1) *If $\|\cdot\|$ is 1–unconditional, in the ℓ–position or $w^{1,1}$–position, then for any $\delta \in (0, 1/2]$ we have*

$$\mathbb{P}\left\{\|G\| \leqslant \delta \mathbb{E}\, \|G\|\right\} \leqslant 2 \exp\left(-n^{1 - C\delta^2}\right), \tag{1.12}$$

where $C > 0$ is a universal constant.

(2) *If $\|\cdot\|$ is an arbitrary norm in $\mathbb{R}^n$, then there exists a linear invertible map T such that for every $\delta \in (0, 1/2]$ we have*

$$\mathbb{P}\left\{\|TG\| \leqslant \delta \mathbb{E}\, \|TG\|\right\} \leqslant 2 \exp\left(-n^{1/4 - C\delta^2}\right), \tag{1.13}$$

where $C > 0$ is a universal constant.

Let us emphasize that for finding the optimal small ball behavior, one has to choose appropriately the linear structure. Generally, for any semi-norm $f(\cdot)$ and any log-concave probability measure μ one has $\mu\{x : f(x) \leqslant \varepsilon \mathbb{E}_\mu f(x)) \leqslant c\varepsilon$ (see [Lat99] and [Gue99]). This estimate is sharp in the case where $f(\cdot) = |\langle \cdot, \theta \rangle|$ for some $\theta \in S^{n-1}$. The linear image that appears in the statement of Theorem 1.1.(2) is different (in our approach) than the commonly used positions in the geometry of normed spaces, e.g. the ℓ-position. Moreover, the existence of the linear map T is based on several results; some of topological nature (the Borsuk–Ulam theorem), structural (the Alon–Milman theorem) and analytic ones. This technique has been developed in [PV18a] and plays significant role in the present work, too. Let us point out that the existence of a linear invertible map T for which $d(\|TG\|, 1/2)$ grows faster than any polylogarithmic (but slower than any polynomial) can be derived by classical means. However, to the extend of our knowledge, it has not appeared in the literature before. We outline this argument in §3.2.

Now we turn in discussing small deviations below the mean, which are by far a more delicate matter. Note that (recall (1.4), (1.6)) in the case of the ℓ_∞ norm one essentially has

$$-\log \mathbb{P}\{\|G\|_\infty < (1 - \varepsilon)\mathbb{E}\|G\|_\infty\} \geqslant 1/\mathbb{P}\{\|G\|_\infty > (1 + \varepsilon)\mathbb{E}\|G\|_\infty\},$$

reflecting that $\|G\|_\infty$ is skewed left in a strong sense. The method we described for small ball probabilities can immediately yield a double exponential bound in this regime by approximating the function f and apply the variance-sensitive small deviation estimate for $P_t f$. However, since an approximation of f by $P_t f$ satisfies an a priori error bound $f - P_t f = O(\sqrt{t})$ (see e.g. [Pis86] Chapter 1), this approach would only yield a tail estimate of $\exp(-n^{c\varepsilon^2})$. To obtain the matching bound $\exp(-n^{c\varepsilon})$ requires extra effort. The smoothening via the Ornstein-Uhlenbeck-semigroup is still involved but the main difference now is that instead of applying P_t directly to f we apply it to an appropriately chosen norming set of f. We perform a step-by-step procedure (algorithm) based on smoothening and elimination of spiky parts of the norm, which requires at most polynomial number of steps, and allows to deform the original norm to one which almost preserves the mean and enjoys better variance bounds. The final stage employs the variance-sensitive deviation inequality once more. Let us emphasize that our method heavily depends on the fact that the variance governs the lower deviation estimate for obtaining both the right dependence on ε and the full regime $\varepsilon \in ((\log n)^{-1}, 1)$. This is evident at the computational stage of the argument which apparently wouldn't have produced the optimal form if a small-deviation inequality with different variance proxy has been used.

As before, we establish this result for 1-unconditional norms in canonical positions, and in the case of general norms, we have to apply once more the "non-standard" linear transformation. Namely, we obtain the following:

Theorem 1.2 (Small deviations for norms). *Let $\|\cdot\|$ be a norm in $\mathbb{R}^n$.*

(1) *If $\|\cdot\|$ is 1-unconditional in the ℓ–position or $w^{1,1}$–position, then for any $\varepsilon \in (0, 1/2)$ we have*

$$\mathbb{P}\left\{\|G\| \leqslant (1-\varepsilon)\mathbb{E}\,\|G\|\right\} \leqslant 2\exp\left(-n^{c\varepsilon}\right), \tag{1.14}$$

where $c > 0$ is a universal constant.

(2) *In the general case, there exists an invertible linear map T such that for any $\varepsilon \in (0, 1/2)$ we have*

$$\mathbb{P}\{\|TG\| \leqslant (1-\varepsilon)\mathbb{E}\|G\|\} \leqslant 2\exp\left(-n^{c\varepsilon}\right), \tag{1.15}$$

where $c > 0$ is a universal constant.

We have chosen to present the results in this introductory part according to the regime under study because we believe it is more instructive. Nonetheless, this presentation may raise questions, such as what is the relation between the linear transformation of Theorem 1.1 with Theorem 1.2? The reader of the subsequent material will convince himself that the argument in the case of an arbitrary norm can be arranged in order to obtain an invertible linear transformation T for which $\|TG\|$ satisfies the same estimates as $\|G\|_\infty$ in (almost) all regimes. More precisely, there is a $T \in GL(n)$ such that for all $\varepsilon > 0$ we have

$$\mathbb{P}\{\|TG\| > (1+\varepsilon)\mathbb{E}\|TG\|\} \leqslant 2\exp(-c\max\{\varepsilon^2, \varepsilon\}\log n), \tag{1.16}$$

while for all $\varepsilon \in (0, 1/2)$ we have

$$\mathbb{P}\{\|TG\| < (1-\varepsilon)\mathbb{E}\|TG\|\} \leqslant 2\exp(-n^{c\varepsilon}), \tag{1.17}$$

and for any $\delta \in (0, 1/2)$ we have

$$\mathbb{P}\{\|TG\| < \delta\mathbb{E}\|TG\|\} \leqslant 2\exp(-cn^{1/4-C\delta^2}), \tag{1.18}$$

where $C, c > 0$ are universal constants. Estimate (1.16) has been previously established in [PV18a].

We believe that the method we develop for establishing small ball probabilities can be applied in a much more general setting than the one we pursue here. However, our motivation to work with norms is initiated by Dvoretzky's theorem. It is the seminal work of V. Milman [Mil71] that introduced the concentration of measure in geometry and led to a new, probabilistic, understanding of the structure of convex bodies in high dimensions. The Dvoretzky theorem was also the motivation in the work of Klartag and Vershynin [KV07]. They used their small ball probability (1.10) to show that for all $m \lesssim d(f)$ the random m-dimensional subspace (with respect to the Haar measure on the Grassmannian) satisfies a lower ℓ_2-estimate, that is

$$\|x\| \geqslant c\frac{\mathbb{E}f(G)}{\sqrt{n}}\|x\|_2, \quad x \in F,$$

with overwhelming probability. In geometric language the latter asserts that any high-dimensional convex body admits (random) sections of dimension proportional to d, which have relative small diameter. In view of the new small ball estimates this phenomenon is established in polynomial dimension for all norms. We provide the details in the last section.

The rest of the paper is organized as follows: In Section 2 we recall the main tools for Gaussian measure and lay them out carefully, in order to assemble them into the proof of the general lower deviation estimates. We provide two variants of the method. In Section 3 we discuss applications of the aforementioned deviations in the framework of normed spaces, establishing Theorem 1.1 (Corollary 3.4, Theorem 3.6, and Theorem 3.10). In Section 4 we describe the algorithmic construction, based on smoothing and truncation of peaks, which allows to deform any given norm to a smaller one which almost preserves the Gaussian expectation and admits improved fluctuations. Employing this construction we show that any norm exhibits double exponential decay in the lower small deviation regime, similar to the case of ℓ_∞-norm. This establishes Theorem 1.2 (Theorem 4.1 and Theorem 4.2). Finally, in Section 5 we conclude with applications in asymptotic geometric analysis and questions that arise from our work.

Acknowledgment. Part of this work was conducted while the authors were in residence at the Mathematical Sciences Research Institute in Berkeley, California, supported by NSF grant DMS-1440140. The hospitality of MSRI and of the organizers of the program on Geometric Functional Analysis is gratefully acknowledged.

2. Main tools

In this section we introduce a general method (see §2.4) for proving small ball probability estimates for convex, positively homogeneous functions. We provide two variants of the method. The first uses the hypercontractivity property of the Gaussian measure; the second is based on the superconcentration phenomenon, whereas both of them depend heavily on the convexity properties of the underlying function. The convexity ensures that the function deviates less below the median and this becomes even more drastic when the function is proportionally smaller than its mean. We start by providing the background material on Gaussian tools that we will need for proving the aforementioned results.

2.1. **Gaussian deviation inequalities for convex functions.** The fact that for any convex function the deviation below the median is smaller than the deviation above the median, with respect to the Gaussian measure, it is perhaps more or less intuitively clear:

$$\mathbb{P}\{f(G) \leqslant \mathrm{med}(f) - t\} \leqslant \mathbb{P}\{f(G) \geqslant \mathrm{med}(f) + t\}, \quad t > 0, \quad G \sim N(\mathbf{0}, I_n). \tag{2.1}$$

Another fact which illustrates this skew behavior is due to Kwapien [Kwa94] and states that the expectation of a convex function is at least as large as the median

$$\mathbb{E}[f(G)] \geqslant \mathrm{med}(f(G)), \quad G \sim N(\mathbf{0}, I_n). \tag{2.2}$$

A rigorous proof of these facts can be given by employing Ehrhard's inequality [Ehr83] (see [Val17] for the details). However, the inequality (2.1) does not give any information how this skew behavior can be quantified. To this end, there exist one-sided deviation inequalities which improve upon the classical deviation inequality (1.1), when the function under consideration is additionally convex.

An inequality of this type can be traced back to the works of Samson [Sam03] and Bobkov-Götze [BG99] (see also [PV19, §5.2]). It essentially states that one may replace the Lipschitz constant of f by the $(\mathbb{E}\|\nabla f(G)\|_2^2)^{1/2}$. Note that for any smooth Lipschitz function one has

$$\mathbb{E}\|\nabla f(G)\|_2^2 \leqslant \mathrm{Lip}(f)^2. \tag{2.3}$$

The precise statement is the following:

Theorem 2.1 (isoperimetry+convexity). *For any smooth, convex map $f : \mathbb{R}^n \to \mathbb{R}$ with $\|\nabla f\|_2 \in L_2(\gamma_n)$, we have*

$$\mathbb{P}\left\{f(G) \leqslant \mathbb{E}f(G) - t\left(\mathbb{E}\|\nabla f(G)\|_2^2\right)^{1/2}\right\} \leqslant e^{-t^2}, \quad t > 0, \tag{2.4}$$

where G is the standard Gaussian vector in $\mathbb{R}^n$.

Let us mention that this inequality is obtained by melting together the Gaussian isoperimetry and the convexity properties of the function. In fact, it holds true for a much larger class of measures, e.g. for all measures satisfying a logarithmic-Sobolev inequality. We refer the reader to [Led01, Chapter 6] and the references therein for related notions.

Another inequality of this type was recently established in [PV18b]. For convex functions, one may replace the L_2-norm of the gradient by the variance of the function. More precisely, we have the following:

Theorem 2.2 (convexity+convexity)**.** *Let f be a convex map in $L_2(\gamma_n)$. Then, we have*

$$\mathbb{P}\left\{f(G) \leqslant \mathbb{E}f(G) - t\sqrt{\mathrm{Var}[f(G)]}\right\} \leqslant e^{-t^2/100}, \quad t > 0, \tag{2.5}$$

where G is the standard Gaussian vector in $\mathbb{R}^n$.

Note that, in the light of the Gaussian Poincaré inequality [Che82],

$$\mathrm{Var}[f(G)] \leqslant \mathbb{E}\|\nabla f(G)\|_2^2, \tag{2.6}$$

the estimate (2.5) improves considerably upon (2.4). Moreover, there exist classical examples which indicate that the variance can be dramatically smaller than $\mathbb{E}\|\nabla f(G)\|_2^2$ or $\mathrm{Lip}(f)^2$ (see [PVZ17], [LT17], [Val17], [Cha14]). Let us emphasize the fact that, the aforementioned variance-sensitive inequality is strongly connected with the Gaussian convexity, via Ehrhard's inequality, and it is not known if holds true for other than Gaussian-like distributions, see [PV18b, Theorem 2.2], [PV19, Theorem 5.6] and [Val17, §2.1.3].

In many problems, one is interested in studying the small deviation from the mean (or the median) of a positive convex function f

$$\mathbb{P}\{f(G) \leqslant (1-\varepsilon)\mathbb{E}[f(G)]\}, \quad 0 < \varepsilon < 1.$$

This probability can be estimated by using either (2.4) or (2.5) to obtain

$$\mathbb{P}\{f(G) \leqslant (1-\varepsilon)\mathbb{E}[f(G)]\} \leqslant \exp\left(-\varepsilon^2/\tilde{\beta}(f)\right), \quad 0 < \varepsilon < 1, \quad \tilde{\beta}(f) := \frac{\mathbb{E}\|\nabla f(G)\|_2^2}{(\mathbb{E}[f(G)])^2}, \tag{2.7}$$

in the first case, and

$$\mathbb{P}\{f(G) \leqslant (1-\varepsilon)\mathbb{E}[f(G)]\} \leqslant 2\exp\left(-c\varepsilon^2/\beta(f)\right), \quad 0 < \varepsilon < 1, \quad \beta(f) := \frac{\mathrm{Var}[f(G)]}{(\mathbb{E}[f(G)])^2}, \tag{2.8}$$

in the latter case. If f is smooth and Lipschitz, by taking into account (2.6) and (2.3) we infer that

$$k(f) \leqslant \frac{1}{\tilde{\beta}(f)} \leqslant \frac{1}{\beta(f)},$$

where $k(f)$ is the Dvoretzky number of f and $\beta(f)$ is referred to as the normalized variance (see e.g. [PV18b]). Furthermore, if f is additionally a norm one may easily check (see e.g. [PV19]) that

$$n \asymp k(\|\cdot\|_2) \geqslant k(f) \geqslant k(|\langle\cdot,\theta\rangle|) = 2/\pi, \quad \beta(f) \geqslant \beta(\|\cdot\|_2) \asymp 1/n,$$

where θ is some (any) unit vector.

2.2. **Talagrand's $L_1 - L_2$ bound.** We have already reviewed that there exist mainly two reasons for having suboptimal concentration, but the fact that $\mathrm{Var}[f(G)] \ll \mathbb{E}\|\nabla f(G)\|_2^2$ is by far more delicate. Following Chatterjee [Cha14, Definition 3.1], we will say that a smooth function $f : \mathbb{R}^n \to \mathbb{R}$ is *ε_n-superconcentrated*, for some $\varepsilon_n \in (0,1)$, if we have

$$\mathrm{Var}[f(G)] \leqslant \varepsilon_n \mathbb{E}\|\nabla f(G)\|_2^2. \tag{2.9}$$

In view of the above, we may define the superconcentration constant of f (see also [Val17]) by

$$s(f) := \frac{\mathrm{Var}[f(G)]}{\mathbb{E}\|\nabla f(G)\|_2^2}. \tag{2.10}$$

Although, superconcentration occurs quite often, unfortunately, not many methods for establishing it are available. Hence, it is of great importance to develop general approaches for quantifying this phenomenon efficiently. A way for proving superconcentration is via Talagrand's inequality [Tal94], which improves upon the classical Poincaré inequality (2.6).

Theorem 2.3 (Talagrand)**.** *For any absolutely continuous function $f : \mathbb{R}^n \to \mathbb{R}$, we have*

$$\mathrm{Var}_{\gamma_n}(f) \leqslant C \sum_{i=1}^{n} \frac{\|\partial_i f\|_{L_2}^2}{1 + \log\left(\|\partial_i f\|_{L_2}/\|\partial_i f\|_{L_1}\right)}, \tag{2.11}$$

where $\partial_i f$ stands for the i-th partial derivative of f.

In the study of the asymptotic theory of finite-dimensional normed spaces, Talagrand's inequality and the superconcentration phenomenon were put forward in [PVZ17]. The authors there, invoke (2.11) to prove that the ℓ_p-norms are superconcentrated for $p > \log n$ and to establish sharp concentration inequalities for this range of p. Soon after, the inequality was used in [Tik18] for proving that every 1-unconditional norm, in ℓ-position, has concentration at least as good as the ℓ_∞-norm. Subsequently, it was exploited in [PV18a], to show that every norm admits an invertible linear image with concentration at least as good as the ℓ_∞-norm.

In the present work we will use Talagrand's inequality, to give an upper bound for the super-concentration constant, in the following form:

Corollary 2.4. *Let f be a smooth function on $\mathbb{R}^n$. For the following parameters:*

$$s(f) = \frac{\mathrm{Var}[f(G)]}{\mathbb{E}\|\nabla f(G)\|_2^2}, \quad R(f) = \frac{\mathbb{E}\|\nabla f(G)\|_2^2}{\sum_{j=1}^n \|\partial_j f\|_{L_1}^2}, \tag{2.12}$$

we have

$$\frac{2}{s(f)} \geqslant \alpha \log R(f), \tag{2.13}$$

where $\alpha > 0$ is a universal constant.

This estimate follows from Talagrand's inequality with an appropriate application of Jensen's inequality (see e.g. [Cha14, Chapter 5] for the details). Since an explicit constant is required for our approach, we provide a proof in §2.4.2 which yields $\alpha = 1$. The proof is nothing more than a repetition of the existing argument in [Cha14, Chapter 5], or [BKS03], with a careful bookkeeping of the constants in each computational stage.

Let us mention that Talagrand [Tal94] proved (2.11) for the uniform probability measure on the Hamming cube. An alternative approach was presented by Benjamini, Kalai and Schramm in [BKS03]. Both approaches rest on the Bonami-Beckner hypercontractive inequality and therefore (2.11) holds true for any hypercontractive measure, see e.g. [CEL12]. An explicit proof of the Gaussian version (Theorem 2.3) can be found in [CEL12] or [Cha14, Chapter 5]. Our approach also depends on the hypercontractive property of the Gaussian measure. We recall some basic facts in the next paragraph.

2.3. **Ornstein-Uhlenbeck semigroup.** The hypercontractive property of the Gaussian measure can be expressed in terms of the associated Ornstein-Uhlenbeck semigroup (OU-semigroup). Let us first recall the definition. For any $f \in L_1(\gamma_n)$ we define

$$P_t f(x) = \mathbb{E} f\left(e^{-t}x + \sqrt{1-e^{-2t}}G\right), \quad x \in \mathbb{R}^n, \quad t \geqslant 0, \tag{2.14}$$

where G is the standard Gaussian vector in $\mathbb{R}^n$. It is known that $P_t f$ is solution of the following heat equation:

$$\begin{cases} \partial_t u = Lu \\ u(x,0) = f \end{cases}, \tag{2.15}$$

where L stands for the generator of the semigroup, that is $Lu = \Delta u - \langle x, \nabla u\rangle$ for sufficiently smooth function u. For background material on the OU-semigroup we refer the reader to [BGL14]. In the next lemma we collect several properties of the semigroup $(P_t)_{t\geqslant 0}$ that will be useful for our approach.

Lemma 2.5. *The Onrnstein-Uhlenbeck semigroup $(P_t)_{t\geqslant 0}$ enjoys the following properties:*

1. $P_0 f(x) = f(x)$ *and* $P_\infty f(x) = \lim_{t\to\infty} P_t f(x) = \mathbb{E}[f(G)]$, $x \in \mathbb{R}^n$.
2. *For each* $i \leqslant n$, *we have* $\partial_i(P_t f) = e^{-t} P_t(\partial_i f)$.
3. $\mathbb{E}P_t f(G) = \mathbb{E}f(G)$.
4. *If* $f \geqslant 0$, *then* $P_t f \geqslant 0$.

(5) *For every $1 \leqslant p \leqslant \infty$ we have that $P_t : L_p \to L_p$ is a linear contraction.*
(6) *If f is convex, then $P_t f$ is also convex.*

These properties can be easily verified from the definition of the OU-semigroup. Alternatively, proofs of these facts can be found, e.g., in [Led01, Chapter 5] or [BGL14]. Let us emphasize the fact that properties (1)-(5) are satisfied by any Markov semigroup, while property (6) depends crucially on the fact that the OU-semigroup admits the integral representation (2.14) with respect to the Mehler kernel. Indeed; for any $\lambda \in [0,1]$ and $x, y \in \mathbb{R}^n$ we may write

$$\begin{aligned} P_t f((1-\lambda)x+\lambda y) &= \mathbb{E}\left[f\left(e^{-t}\left((1-\lambda)x+\lambda y\right)+\sqrt{1-e^{-2t}}G\right)\right] \\ &= \mathbb{E}\left[f\left((1-\lambda)\left(e^{-t}x+\sqrt{1-e^{-2t}}G\right)+\lambda\left(e^{-t}y+\sqrt{1-e^{-2t}}G\right)\right)\right] \\ &\leqslant \mathbb{E}\left[(1-\lambda)f\left(e^{-t}x+\sqrt{1-e^{-2t}}G\right)+\lambda f\left(e^{-t}y+\sqrt{1-e^{-2t}}G\right)\right] \\ &= (1-\lambda)P_t f(x)+\lambda P_t f(y), \end{aligned}$$

where we have used the convexity of f pointwise.

The fact that P_t is a linear contraction admits an improvement by, roughly speaking, relaxing the integrability assumption in the domain. This important property of the Ornstein-Uhlenbeck semigroup called hypercontractivity is due to Nelson [Nel67]. We have the following:

Theorem 2.6 (Nelson)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function and let $1 < p \leqslant q$. Then, for any $t \geqslant 0$ with $q \leqslant 1 + e^{2t}(p-1)$, we have*

$$\|P_t f\|_{L_q(\gamma_n)} \leqslant \|f\|_{L_p(\gamma_n)}. \tag{2.16}$$

Gross [Gro75] showed that the hypercontractive property is intimately connected to the logarithmic Sobolev inequality. In particular, he proved that a probability measure is hypercontractive if and only if satisfies a logarithmic Sobolev inequality. See also [BGL14, Chapter 5].

2.4. **Lower deviations with respect to global parameters.** We are now ready to present the proofs of the new small ball estimates governed by the parameters $\tilde{\beta}$ and β.

2.4.1. *From hypercontractivity to small ball estimates.* Recall that for any smooth function f with $\mathbb{E}[f(G)] \neq 0$ we define

$$\tilde{\beta}(f) = \frac{\mathbb{E}\|\nabla f(G)\|_2^2}{(\mathbb{E}[f(G)])^2}. \tag{2.17}$$

Our goal is to prove the following small ball estimate:

Theorem 2.7. *Let $f : \mathbb{R}^n \to [0,\infty)$ be a positively homogeneous[2], convex map with $\|\nabla f\|_2 \in L_2(\gamma_n)$ and let $\tilde{\beta} = \tilde{\beta}(f(G))$. Suppose that for some $L > 0$ we have*

$$\sum_{j=1}^n \|\partial_j f\|_{L_1(\gamma_n)}^2 \leqslant L(\mathbb{E}f(G))^2. \tag{2.18}$$

Then, for any $\delta \in (0,1/2)$ we have

$$\mathbb{P}\left\{f(G) \leqslant \delta \mathbb{E}[f(G)]\right\} \leqslant \exp\left\{-c\delta^2 \left(\frac{1}{\tilde{\beta}}\right)^{\tau(\delta)} \left(\frac{1}{L}\right)^{1-\tau(\delta)}\right\},$$

where $\tau(\delta) \asymp \delta^2$.

Next proposition is the key observation.

[2]A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be positively homogeneous if $f(\lambda x) = \lambda f(x)$ for all $\lambda > 0$ and $x \in \mathbb{R}^n$.

Proposition 2.8. *Let f be a smooth function with $\|\nabla f\|_2 \in L_2$ and let $L > 0$ be such that*

$$\sum_{j=1}^{n} \|\partial_j f\|^2_{L_1(\gamma_n)} \leqslant L(\mathbb{E}[f(G)])^2.$$

Then, for all $t \geqslant 0$ we have

$$\frac{1}{\tilde{\beta}(t)} \geqslant e^{2t} \left(\frac{1}{\tilde{\beta}}\right)^{\frac{2e^{-2t}}{1+e^{-2t}}} \left(\frac{1}{L}\right)^{\frac{1-e^{-2t}}{1+e^{-2t}}},$$

where $\tilde{\beta}(t) := \tilde{\beta}(P_t f)$.

For the proof we will need the following lemma which is a consequence of Nelson's hypercontractivity.

Lemma 2.9. *Let h be a function in L_2. Then, for all $t \geqslant 0$ one has*

$$\|P_t h\|_{L_2} \leqslant \|h\|_{L_2} \left(\frac{\|h\|_{L_1}}{\|h\|_{L_2}}\right)^{\frac{2}{p(t)}-1}, \quad p(t) = 1 + e^{-2t}. \tag{2.19}$$

Proof. Using the hypercontractivity we get

$$\|P_t h\|_2 \leqslant \|h\|_p, \quad p = 1 + e^{-2t}.$$

By Hölder's inequality we derive

$$\|h\|_p^p \leqslant \|h\|_1^{2-p} \|h\|_2^{2(p-1)}.$$

The result follows. □

Now we turn to proving the key proposition. The argument we follow can be traced back to [BKS03] (see also [Cha14, Theorem 5.1]):

Proof of Proposition 2.8. Let $A_i = \|\partial_i f\|_{L_2}$ and $a_i = \|\partial_i f\|_{L_1}$. We may write

$$\mathbb{E}\|\nabla(P_t f)\|_2^2 = e^{-2t} \sum_{i=1}^{n} \|P_t(\partial_i f)\|^2_{L_2} \leqslant e^{-2t} \sum_{i=1}^{n} A_i^2 \left(\frac{a_i^2}{A_i^2}\right)^{\tanh t}, \tag{2.20}$$

where we have also used Lemma 2.9. Note that the function $u \mapsto u^{\tanh t}$, $u > 0$ is concave, thus Jensen's inequality implies

$$\sum_{i=1}^{n} A_i^2 \left(\frac{a_i^2}{A_i^2}\right)^{\tanh t} \leqslant \mathbb{E}\|\nabla f\|_2^2 \left(\frac{\sum_{i=1}^n a_i^2}{\mathbb{E}\|\nabla f\|_2^2}\right)^{\tanh t}. \tag{2.21}$$

Combining (2.20) with (2.21) we arrive at

$$\mathbb{E}\|\nabla(P_t f)\|_2^2 \leqslant e^{-2t} \mathbb{E}\|\nabla f\|_2^2 R^{-\tanh t}. \tag{2.22}$$

Dividing both sides with $(\mathbb{E}[f(G)])^2$, using the assumption and the definition of $\tilde{\beta}$, the above estimate yields

$$\tilde{\beta}(t) \leqslant e^{-2t} \tilde{\beta} \left(\frac{L}{\tilde{\beta}}\right)^{\tanh t},$$

as required. □

Now we turn to proving our first main result.

Proof of Theorem 2.7. Since f is positively homogeneous and convex it is also sub-additive, hence

$$P_t f(x) \leqslant e^{-t} f(x) + \sqrt{1 - e^{-2t}} \mathbb{E} f(G), \quad x \in \mathbb{R}^n, \quad t \geqslant 0.$$

We fix $\delta \in (0, 1/2)$. For any $t > 0$, using Theorem 2.1, we may write

$$\mathbb{P}\left(f(G) \leqslant \delta \mathbb{E} f(G)\right) \leqslant \mathbb{P}\left(P_t f(G) \leqslant \left(\delta e^{-t} + \sqrt{1 - e^{-2t}}\right) \mathbb{E}[P_t f(G)]\right) \leqslant \exp\left(-c\varepsilon(t)^2 / \tilde{\beta}(t)\right),$$

provided that

$$\varepsilon(t) := 1 - \delta e^{-t} - \sqrt{1-e^{-2t}} = e^{-t}\left(\frac{e^{-t}}{1+\sqrt{1-e^{-2t}}} - \delta\right) > 0.$$

We select time $t_0 = t_0(\delta) > 0$ such that

$$\frac{e^{-t_0}}{1+\sqrt{1-e^{-2t_0}}} = 2\delta.$$

For this choice we have $e^{-t_0} \asymp \delta$ and by taking into account Proposition 2.8 we conclude with

$$\tau(\delta) = \frac{e^{-2t_0}}{1+e^{-2t_0}} \asymp \delta^2.$$

The proof is complete. □

2.4.2. *From superconcentration to small ball estimates.* Recall the definition of the parameter β. For any smooth function f with $\mathbb{E}f \neq 0$ we have

$$\beta(f) = \frac{\mathrm{Var}[f(G)]}{(\mathbb{E}[f(G)])^2}. \tag{2.23}$$

The main result of this section is the following variant of Theorem 2.7:

Theorem 2.10. *Let $f : \mathbb{R}^n \to [0,\infty)$ be a positively homogeneous, convex function in $L_2(\gamma_n)$ and let $\beta = \beta(f(G))$. Suppose that*

$$\sum_{j=1}^n (\mathbb{E}|\partial_j f(G)|)^2 \leqslant L(\mathbb{E}f(G))^2, \quad L > 0.$$

Then, for any $\delta \in (0, 1/2)$ we have

$$\mathbb{P}\left\{f(G) \leqslant \delta\mathbb{E}f(G)\right\} \leqslant \exp\left\{-c\delta^2\left(\frac{1}{\beta}\right)^{\omega(\delta)}\left(\frac{1}{L}\right)^{1-\omega(\delta)}\right\},$$

with $\omega(\delta) \asymp \delta$.

The proof follows the same lines as in the previous paragraph with the appropriate modifications. Unlike to the previous one, this approach exploits the exponential decay of the variance along the flow P_t as was previously discussed in the Introduction. More precisely, for any Markov semigroup P_t with invariant measure μ one has

$$\mathrm{Var}(P_t f) \leqslant e^{-\lambda_1 t}\mathrm{Var}(f), \quad t \geqslant 0, \tag{2.24}$$

where λ_1 is the spectral gap of μ. This well known property can be found in any classical text concerned with semigroup tools, see e.g. [Led00, BGL14]. However, this dimension free estimate cannot later provide dimension dependent bounds of the right order of magnitude, which is the central theme of this work due to the conjectured extreme case of ℓ_∞-norm. Note that the latter is genuinely high-dimensional functional whereas (2.24) takes into account the extremals in Poincaré inequality, i.e. linear functionals. In other words, the bound in terms of λ_1 is the worst case scenario amongst the ones we have to encounter in the light of

$$\lambda_1(\mu) = \inf\left\{\frac{1}{s_\mu(f)} : f \textit{ smooth, nonconstant}\right\}.$$

By carefully revisiting the proof of (2.24) we see that something more is true, which fixes this sub-optimality issue. Therefore, the proof of the next lemma is standard and can be found implicitly in [Led00, BGL14]. Nonetheless, this formulation is not explicitly stated there, hence we shall include a (sketch of) proof for reader's convenience.

Lemma 2.11. *Let f be a smooth function on $\mathbb{R}^n$. We define $v(t) = \mathrm{Var}[P_t f(G)]$ and $s(t) = s(P_t f)$. Then, we have the following properties:*

i. $v'(t) = -2\mathbb{E}\|\nabla(P_t f)\|_2^2$.

ii. $v'(t)s(t) = -2v(t)$, *hence for all* $t \geqslant 0$ *we have*

$$v(t) = v(0)\exp\left(-2\int_0^t \frac{dz}{s(z)}\right).$$

iii. $v(t)$ *is log-convex and hence,* $s(t)$ *is nondecreasing.*

Proof. Since all assertions include shift invariant quantities we may assume without loss of generality that $\mathbb{E}f = 0$. Thus, $v(t) = \mathbb{E}(P_t f)^2$ and differentiation in terms of t, under the integral sign, yields

$$\frac{dv}{dt} = \frac{d}{dt}\left(\mathbb{E}(P_t f)^2\right) = 2\mathbb{E}[P_t f L P_t f] = -2\mathbb{E}\langle \nabla P_t f, \nabla P_t f\rangle = -2\mathbb{E}\|\nabla(P_t f)\|_2^2,$$

where we have used that P_t solves the heat equation (2.15) and the generator L satisfies the integration by parts formula, i.e. $\mathbb{E}(uLv) = -\mathbb{E}\langle\nabla u, \nabla v\rangle$ for u, v smooth functions. The rest of the assertions easily follow. □

Remark 2.12. The above lemma provides a link between the superconcentration phenomenon and the variance decay, during the ensuing motion. This can be viewed as an alternative definition to the superconcentration. Alternative (equivalent) definitions (via the gap in Poincaré inequality or in connection with chaos) can be found in [Cha14, Chapter 3 & 4]. Note also that Lemma 2.11 is true for any Markov semigroup.

Proof of Corollary 2.4 (with $\alpha = 1$). Recall estimate (2.22),

$$\mathbb{E}\|\nabla(P_t f)\|_2^2 \leqslant e^{-2t}\mathbb{E}\|\nabla f\|_2^2 R^{-\tanh t}, \quad t \geqslant 0.$$

Integrating the latter and taking into account the fact that

$$\mathrm{Var}[f(G)] = v(0) = -\int_0^\infty v'(t)\,dt = 2\int_0^\infty \mathbb{E}\|\nabla(P_t f)\|_2^2\,dt,$$

which follows from Lemma 2.5, we get the assertion. □

We will need the next proposition, which is the analogue of Proposition 2.8 in terms of the parameter β.

Proposition 2.13. *Let* $f \in L_2(\gamma_n)$ *be a smooth function and let* $L > 0$ *such that*

$$\sum_{j=1}^n (\mathbb{E}|\partial_j f(G)|)^2 \leqslant L(\mathbb{E}f(G))^2.$$

Then, for all $t \geqslant 0$ *we have*

$$\frac{1}{\beta(t)} \geqslant e^{2t-2}\left(\frac{1}{\beta}\right)^{e^{-t}}\left(\frac{1}{L}\right)^{1-e^{-t}},$$

where $\beta(t) := \beta(P_t f)$.

Proof. First, note that Lemma 2.11 yields

$$\frac{1}{\beta(t)} = \frac{1}{\beta}\exp(\psi(t)), \quad \psi(t) := 2\int_0^t \frac{dz}{s(z)}. \tag{2.25}$$

Next, we employ Corollary 2.4 to link $s(t)$ with $\beta(t)$. To this end, it suffices to bound from below the parameter $R(t) = R(P_t f)$. Indeed; we have

$$R(t) = \frac{\mathbb{E}\|\nabla(P_t f)(G)\|_2^2}{\sum_j \|\partial_j(P_t f)\|_{L_1}^2} \geqslant \frac{\mathbb{E}\|\nabla(P_t f)(G)\|_2^2}{e^{-2t}\sum_j \|\partial_j f\|_{L_1}^2} \geqslant \frac{e^{2t}\mathrm{Var}[P_t f(G)]}{L(\mathbb{E}f(G))^2} = e^{2t}\beta(t)/L,$$

where we have also used Lemma 2.5 and the Poincaré inequality (2.6). Whence, we obtain

$$\psi'(t) = \frac{2}{s(t)} \geqslant \alpha\log(e^{2t}\beta(t)/L). \tag{2.26}$$

Inserting (2.25) into (2.26) we obtain

$$\psi'(t) \geqslant \alpha\log\left(\frac{\beta}{L}e^{2t-\psi(t)}\right) \Longrightarrow \psi'(t) + \alpha\psi(t) \geqslant 2\alpha t + \alpha\log(\beta/L).$$

Integrating the above in $[0,t]$ we find

$$
\begin{aligned}
e^{\alpha t}\psi(t) = \int_0^t e^{\alpha z}\left(\psi'(z)+\alpha\psi(z)\right)\,dz \geqslant \int_0^t e^{\alpha z}\left(2\alpha z+\alpha\log(\beta/L)\right)\,dz \\
= 2te^{\alpha t}-\frac{2}{\alpha}(e^{\alpha t}-1)+(e^{\alpha t}-1)\log(\beta/L) \\
= 2te^{\alpha t}+(e^{\alpha t}-1)\log(e^{-2/\alpha}\beta/L).
\end{aligned}
$$

It follows that

$$
\psi(t) \geqslant 2t+(1-e^{-\alpha t})\log(e^{-2/\alpha}\beta/L), \quad t\geqslant 0.
$$

Plug the latter in (2.25) we arrive at

$$
\frac{1}{\beta(t)} \geqslant \frac{e^{2t}}{\beta}\left(\frac{\beta}{e^{2/\alpha}L}\right)^{1-e^{-\alpha t}} \Longrightarrow \frac{1}{\beta(t)} \geqslant e^{2t}\left(\frac{1}{\beta}\right)^{e^{\alpha t}}\left(\frac{\beta}{e^{2/\alpha}L}\right)^{1-e^{-\alpha t}},
$$

as required. It remains to notice that we have established Corollary 2.4 with $\alpha = 1$. □

Remark 2.14. Note that the above estimate holds true for any hypercontractive semigroup; see [CEL12] for a proof of the fact that hypercontractive measures satisfy Talagrand's inequality (2.3).

Now we turn to proving our second main result.

Proof of Theorem 2.10. The proof follows the same lines as before: For $\delta\in(0,1/2)$ and $t>0$ we have

$$
\mathbb{P}\left\{f(G)\leqslant \delta\mathbb{E}f(G)\right\} \leqslant \exp\left(-c\varepsilon(t)^2/\beta(t)\right),
$$

provided that

$$
\varepsilon(t) := 1-\delta e^{-t}-\sqrt{1-e^{-2t}} = e^{-t}\left(\frac{e^{-t}}{1+\sqrt{1-e^{-2t}}}-\delta\right) > 0.
$$

Selecting time $t_0 = t_0(\delta) > 0$ such that

$$
\frac{e^{-t_0}}{1+\sqrt{1-e^{-2t_0}}} = 2\delta,
$$

and taking into account Proposition 2.13, we conclude with $\omega(\delta)=\exp(-t_0)\asymp\delta$. □

3. Small ball probabilities for norms

In this section we apply Theorem 2.7 to derive optimal small ball estimates in normed spaces. Since our study takes into account the unconditional structure of the norm both explicitly and implicitly, we begin with some auxiliary results in this context.

3.1. **Unconditional structure.** Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ and let (b_i) be a basis. Following [FJ80], we define the unconditional constant of the norm with respect to the basis, denoted by $\mathrm{unc}(\|\cdot\|,\{b_i\})$, to be the least $r>0$ such that

$$
\left\|\sum_{i=1}^n \varepsilon_i\alpha_i b_i\right\| \leqslant r\left\|\sum_{i=1}^n \alpha_i b_i\right\|, \tag{3.1}
$$

for all choices of signs $\varepsilon_i=\pm1$ and all scalars $(\alpha_i)\subset\mathbb{R}$. Next, we define

$$
\mathrm{unc}(\mathbb{R}^n,\|\cdot\|) = \inf\left\{\mathrm{unc}(\|\cdot\|,\{b_i\}) : (b_i) \text{ basis}\right\}. \tag{3.2}
$$

We denote by (e_i) the standard (orthonormal) basis in $\mathbb{R}^n$.

The following lemma can be viewed as a Lozanovski type result (see e.g. [Sza80] or [Pis89]):

Lemma 3.1. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$, let $a_i=\mathbb{E}|\partial_i\|G\||$ for $i=1,\dots,n$ and let $r=\mathrm{unc}(\|\cdot\|,\{e_i\})$.*

(1) *For every $x \in \mathbb{R}^n$, we have*

$$\frac{1}{r}\sum_{i=1}^n a_i|x_i| \leqslant \|x\| \leqslant r\sqrt{\frac{\pi}{2}}\mathbb{E}\|G\| \cdot \|x\|_\infty. \tag{3.3}$$

(2) *The following estimate holds:*

$$c\frac{\mathbb{E}\|G\|}{\sqrt{\log n}} \leqslant \sum_{i=1}^n a_i \leqslant r\sqrt{\frac{\pi}{2}}\mathbb{E}\|G\|. \tag{3.4}$$

In particular,

$$\frac{c'}{n\log n} \leqslant \frac{\sum_{i=1}^n a_i^2}{(\mathbb{E}\|G\|)^2}.$$

(3) *Assuming that $a_i = a_j$ for all $i, j = 1, 2, \ldots, n$, we have the following estimate:*

$$|B_X|^{1/n}\mathbb{E}\|G\| \leqslant Cr\sqrt{\log n}. \tag{3.5}$$

Proof. (1). Note that we have

$$\forall x, y \in \mathbb{R}^n, \quad y \neq 0, \quad \sum_{i=1}^n x_i \cdot \partial_i\|y\| = \langle x, \nabla\|y\|\rangle \leqslant \|x\|.$$

Applying the latter for $x_i \to x_i\mathrm{sgn}\{x_i\partial_i\|y\|\} \equiv \varepsilon_i x_i$, we obtain

$$\sum_{i=1}^n |x_i| \cdot |\partial_i\|y\|| \leqslant \left\|\sum_{i=1}^n \varepsilon_i x_i e_i\right\| \leqslant r\|x\|.$$

Integration with respect to y yields the lower estimate. For the upper estimate we may argue as follows:

$$\|x\| \leqslant r\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i x_i e_i\right\| \leqslant r\|x\|_\infty \mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i e_i\right\|,$$

where we have used the contraction principle [LT91, Theorem 4.4]. Jensen's inequality and the fact that $(\varepsilon_i|g_i|)$ have the same distribution as (g_i) (see e.g. [Pis74, Proposition 1] or [LT91, Lemma 4.5]) yields

$$\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i e_i\right\| \leqslant (\mathbb{E}|g_1|)^{-1}\mathbb{E}\left\|\sum_{i=1}^n g_i e_i\right\|.$$

The assertion follows.

(2). The rightmost inequality follows by integrating the left-hand side of (3.3) and using the fact that $\mathbb{E}|g| = \sqrt{2/\pi}$. For the leftmost we argue as before. Set $A = \{\|G\|_\infty > 100\sqrt{\log n}\}$ and note that $\mathbb{P}(A) \leqslant n^{-2}$. Then, we may write

$$\mathbb{E}\|G\| = \mathbb{E}\left[\sum_{i=1}^n g_i\partial_i\|G\|\right] \leqslant \mathbb{E}\left[\|G\|_\infty \cdot \|\nabla\|G\|\|_1\right] \leqslant 100\sqrt{\log n}\mathbb{E}\left\|\nabla\|G\|\right\|_1 + nC\mathbb{E}\|G\| \cdot \mathbb{P}(A),$$

where we have used the fact that

$$|\partial_i\|G\|| \leqslant b \leqslant C\mathbb{E}\|G\|, \quad a.s.$$

The in particular part follows from the leftmost estimate and the Cauchy-Schwarz inequality.

(3). The left-hand side of (3.3) implies that $B_X \subseteq \frac{r}{a}B_1^n$. Hence, taking volumes on both sides, we get

$$|B_X|^{1/n} \leqslant \frac{r}{a}\frac{2}{(n!)^{1/n}} \leqslant \frac{2er}{an} \leqslant \frac{2er}{c\mathbb{E}\|G\|}\sqrt{\log n},$$

where in the last step we have used the leftmost estimate from (3.4).

Remark 3.2. (1) Both lower and upper estimates are sharp (up to constants) in the case of ℓ_∞^n and ℓ_1^n-norm, respectively.

(2) Note that Hölder's inequality implies that $|B_X|^{1/n}\mathbb{E}\|G\| \geqslant c$. Thus, (3.5) should be viewed as a reverse Hölder estimate.

Theorem 3.3. *Let $\|\cdot\|$ be a norm in $\mathbb{R}^n$ which satisfies $\mathbb{E}|\partial_i\|G\|| = \mathbb{E}|\partial_j\|G\||$ for all $i, j = 1, 2, \ldots, n$. If $r = \mathrm{unc}(\|\cdot\|, \{e_i\})$, then for any $\delta \in (0, 1/2)$ we have*

$$\mathbb{P}\left\{\|G\| \leqslant \delta\mathbb{E}\|G\|\right\} < \exp\left(-c\left(\frac{1}{\tilde{\beta}}\right)^{\tau(\delta)}\left(\frac{n}{r^2}\right)^{1-\tau(\delta)}\right), \tag{3.6}$$

with $\tau(\delta) \asymp \delta^2$.

Proof. By Theorem 2.7 we get

$$\mathbb{P}\left\{\|G\| \leqslant \delta\mathbb{E}\|G\|\right\} < \exp\left(-c\delta^2(1/\tilde{\beta})^{\tau(\delta)}(1/L)^{1-\tau(\delta)}\right),$$

where L is given by

$$L = \frac{\sum_{i=1}^n (\mathbb{E}|\partial_i\|G\||)^2}{(\mathbb{E}\|G\|)^2}.$$

We distinguish two cases.

- $\delta \geqslant \frac{1}{10r}\sqrt{Ln}$. Then, $1/L \geqslant \frac{n}{100r^2\delta^2}$ and thus,
$$\delta^2(1/L)^{1-\tau(\delta)} \geqslant c'(n/r^2)^{1-\tau(\delta)}.$$
- $\delta < \frac{1}{10r}\sqrt{Ln}$. Note that by Lemma 3.1 we have
$$a\|x\|_1 \leqslant r\|x\|, \quad a = \mathbb{E}|\partial_i\|G\|| = \mathbb{E}\|G\|\sqrt{\frac{L}{n}},$$
by the definition of L and the equality for the partial derivatives of the norm. Therefore, we obtain
$$\{x : \|x\| \leqslant \delta\mathbb{E}\|G\|\} \subset \left\{x : \|x\|_1 \leqslant \frac{\delta r}{a}\mathbb{E}\|G\|\right\} \subset \left\{x : \|x\|_1 \leqslant \frac{n}{10}\right\},$$
where for the last inclusion we have used the assumption on δ.

In each case we get the desired result. □

The following corollary is immediate:

Corollary 3.4. *Let $\|\cdot\|$ be a 1-unconditional norm on $\mathbb{R}^n$, which satisfies $\mathbb{E}|\partial_i\|G\|| = \mathbb{E}|\partial_j\|G\||$ for all $i, j = 1, 2, \ldots, n$. Then, for any $\delta \in (0, 1/2)$ we have*

$$\mathbb{P}\left\{\|G\| \leqslant \delta\mathbb{E}\|G\|\right\} < \exp\left(-cn^{1-C\delta^2}\right), \tag{3.7}$$

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

3.2. **The general case.** Here we show that every centrally-symmetric convex body A in $\mathbb{R}^n$ has an invertible linear image A_1 which satisfies a small-ball estimate with $d(A_1, 1/2)$ of power type. A linear image $\widetilde{A}$ of A which satisfies $d(\widetilde{A}, 1/2) \gtrsim e^{c\sqrt{\log n}}$ can be derived by using a dichotomy based on the Alon–Milman theorem [AM83]. Let us mention that a refined version of the dichotomy in [AM83, p.278-279] was also used by Schechtman [Sch06] in connection with the problem of the dependence on ε in the existential form of Dvoretzky's theorem. Since our method is also rooted in a structural dichotomy which employs the Alon–Milman theorem we find it instructive to outline the former argument and comment on the resemblance and the difference between the two. A (slightly reformulated) form of Alon–Milman theorem reads as follows.

Proposition 3.5 (Alon–Milman). *Let $X = (\mathbb{R}^n, \|\cdot\|)$ be a normed space whose unit ball B_X is in John's position, that is B_2^n is the maximal volume ellipsoid in B_X. Then,*

1. *Either $k(X) \geqslant ce^{c\sqrt{\log n}}$, or*
2. *There exists a subspace F with $\dim F \geqslant ce^{c\sqrt{\log n}}$ such that*
$$c\|x\|_\infty \leqslant \|x\| \leqslant C\|x\|_\infty, \quad x \in F.$$

We may argue as follows. In view of Proposition 3.5 we may assume that we are in the second case (otherwise we employ the general estimate $d \gtrsim k$, see e.g. (1.11), to conclude), hence there exists a subspace F with $\dim F = m \geqslant e^{c\sqrt{\log n}}$ such that $\|x\|_\infty \asymp \|x\|$ for all $x \in F$. Therefore,

$$\mathbb{P}(\|W\| \leqslant \delta \mathbb{E}\|W\|) \leqslant \mathbb{P}(\|W\|_\infty \leqslant C\delta \mathbb{E}\|W\|_\infty) \leqslant e^{-cm^{1-C\delta^2}}, \quad \delta \in (0, 1/(2C)),$$

where W is the standard Gaussian vector on F. Finally, we argue as in [PV18a] in order to "lift" the identity map on F up to a linear invertible map on $\mathbb{R}^n$ by adding a sufficiently small multiple of the identity on $F^\perp$. That is, we choose $a > 0$ such that $\mathbb{E}\|TG\| \leqslant 2\mathbb{E}\|W\|$, where $T = I_F \oplus (aI_{F^\perp})$ and G is the standard Gaussian vector in $\mathbb{R}^n$. Since,

$$\mathbb{P}\left(\|TG\| \leqslant \frac{\delta}{2}\mathbb{E}\|TG\|\right) \leqslant \mathbb{P}(\|W\| \leqslant \delta\mathbb{E}\|W\|), \quad \delta > 0,$$

the assertion follows.

Our method shares common points with the described proof. The lifting argument and the dichotomy based on the Alon–Milman theorem are still involved. The main difference is that instead of using the Alon–Milman theorem (as a black box) in order to locate ℓ_∞-structure, we utilize it to preclude extremal unconditional structure in some lower dimensional subspace F. This permits us to construct a linear map $S : F \to F$ which equips the norm with extra properties. Moreover, the argument is integrated with the new small ball estimate (Theorem 2.7) and allows to boost the lower bound from $e^{\sqrt{\log n}}$ to polynomial. Namely, we have the following:

Theorem 3.6. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Then, there exists $T \in GL(n)$ with the following property: for any $0 < \delta < \frac{1}{4}$ one has*

$$\mathbb{P}\left\{\|TG\| \leqslant \delta\mathbb{E}\|TG\|\right\} \leqslant \exp\left(-cn^{\frac{1}{4}-C\delta^2}\right), \quad G \sim N(\mathbf{0}, I_n), \tag{3.8}$$

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

Let us demonstrate the steps of the proof in more detail. First we show that an estimate of polynomial type holds true conditionally, i.e. if the underlying norm has moderate unconditional structure. In view of Theorem 3.3, this can be accomplished if the norm also satisfies the $w^{1,1}$–condition. This is promised by the following lemma from [PV18a].

Lemma 3.7. *Let f be a smooth norm on $\mathbb{R}^m$. Then, there exist $\lambda_1, \ldots, \lambda_m > 0$ such that*

$$\|\partial_i(f \circ \Lambda)\|_{L_1} = \|\partial_j(f \circ \Lambda)\|_{L_1}, \quad i, j = 1, \ldots, m,$$

where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_m)$.

Combining this with Theorem 3.3 we derive the following:

Theorem 3.8. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^m$. There exists $S \in GL(m)$ such that, for any $\delta \in (0, 1/2)$ we have*

$$\mathbb{P}\left\{\|SG\| \leqslant \delta\mathbb{E}\|SG\|\right\} < \exp\left(-c\left(m/r^2\right)^{1-C\delta^2}\right), \tag{3.9}$$

where $r = \operatorname{unc}(\mathbb{R}^m, \|\cdot\|)$ and $c, C > 0$ are absolute constants.

This result almost reaches the final goal, except from the parameter $r = \mathrm{unc}X$. Note that John's theorem [Joh48] readily implies that $\mathrm{unc}X \leqslant \sqrt{\dim X}$ for any finite dimensional normed space. On the other hand, it is known (see e.g. [FKPc77] and [FJ80]) that a "typical" subspace F of ℓ_∞^n of proportional dimension satisfies $\mathrm{unc}F \gtrsim \sqrt{\dim F}$, which marks the end of usefulness of Theorem 3.8 in those cases. Hence, the next crucial step is to locate moderate unconditional structure in *every* normed space. This is possible by using the Alon–Milman theorem, and its sharper form due to Talagrand [Tal95]. Below we state it as a lemma in a customized form that will be of immediate use for us.

Lemma 3.9. *Let $X = (\mathbb{R}^n, \|\cdot\|)$ be a normed space and let B_2^n be the maximal volume ellipsoid inscribed in B_X. Then, there exists $\sigma \subset [n]$ with $|\sigma| \geqslant c\sqrt{n}$ such that*

$$\frac{1}{8}\max_{i\in\sigma}|\alpha_i| \leqslant \left\|\sum_{i\in\sigma}\alpha_i e_i\right\| \leqslant 4\sqrt{k(X)}\max_{i\in\sigma}|\alpha_i|,$$

for all scalars $(\alpha_i)_{i\in\sigma} \subset \mathbb{R}$.

The reader is referred to [PV18a, Theorem 4.4] for the detailed proof of this formulation which uses the classical Dvoretzky-Rogers lemma [DR50]. Combining Lemma 3.9 with Theorem 3.8, we can prove the announced small ball estimate for any norm.

Proof of Theorem 3.6. We may assume that $A = \{x : \|x\| \leqslant 1\}$ is in John's position, and $k(A) = (\mathbb{E}\|G\|)^2 \leqslant n^{1/4}$, since $d(A, 1/2) \gtrsim k(A)$. Then, Lemma 3.9 yields a $\sigma \subset [n]$ with $|\sigma| \asymp \sqrt{n}$ such that

$$c_1 \max_{i\in\sigma} |\alpha_i| \leqslant \left\| \sum_{i\in\sigma} \alpha_i e_i \right\| \leqslant C_1 \sqrt{k(A)} \max_{i\in\sigma} |\alpha_i|, \tag{3.10}$$

for all scalars $(\alpha_i) \subset \mathbb{R}$. Set $E_\sigma = \mathrm{span}\{e_i : i \in \sigma\}$ and note that the above estimate shows that

$$r := \mathrm{unc}(E_\sigma, \|\cdot\|) \leqslant C_2 \sqrt{k(A)} \leqslant C_2 n^{1/8}.$$

Therefore, Theorem 3.8 yields the existence of $S \in GL(E_\sigma)$ with the property that for all $\delta \in (0, 1/2)$ one has

$$\mathbb{P}\left\{\|SZ\| \leqslant \delta \mathbb{E}\|SZ\|\right\} \leqslant \exp\left(-(|\sigma|/r^2)^{1-C\delta^2}\right) \leqslant \exp\left(-cn^{\frac{1}{4}-C'\delta^2}\right), \quad Z \sim N(\mathbf{0}, I_{E_\sigma}).$$

Now we apply the "lifting" argument as before: We consider the linear map $T = S \oplus (aI_{E_\sigma^\perp})$ to get

$$\mathbb{E}\|TG\| \leqslant 2\mathbb{E}\|SZ\|,$$

provided $a > 0$ is sufficiently small and, moreover,

$$\mathbb{P}\left(\|TG\| \leqslant t\right) \leqslant \mathbb{P}\left(\|SZ\| \leqslant t\right), \quad t > 0,$$

The result readily follows. $\square$

3.3. **1–unconditional norms in the ℓ-position and position of minimal M.** Theorem 2.7 allowed to prove optimal small-ball estimates for 1–unconditional norms in the position where absolute first gaussian moments of partial derivatives of the norm are all equal (Corollary 3.4). In the classical ℓ or minimal M positions, a direct adaptation of the argument will lead to a weaker estimate for small δ "close" to $\log^{-1/2} n$. The reason is that partial derivatives of the norm in the ℓ– (or in minimal M–) position are not necessarily equal and can vary up to the factor $O(\log^{1/2} n)$. Indeed, in that setting we rely on the identities $\mathbb{E}\left(G_i\, \partial_i\|G\|\right) = \mathbb{E}\left(G_j\, \partial_j\|G\|\right)$ for all $i \neq j$ (in the position of minimal M) and $\mathbb{E}\left(G_i\, \partial_i\|G\|\,\|G\|\right) = \mathbb{E}\left(G_j\, \partial_j\|G\|\,\|G\|\right)$ (in the ℓ–position) which, as can be easily checked, imply $\mathbb{E}\,|\partial_i\|G\|| \leqslant C\sqrt{\log n}\,\mathbb{E}\,|\partial_j\|G\||$ for all $i \neq j$, and considering a direct sum of ℓ_2 and ℓ_∞ (of appropriate dimensions), one can check that the inequalities are in general optimal up to the constant multiple C.

In this subsection, we prove the small-ball inequality for 1–unconditional norms in position of minimal M, which matches the estimate in Corollary 3.4. For this, we augment the above argument, based on Theorem 2.7, with specially constructed norm replacement. The proof can be repeated for the ℓ–position with minor modifications, so we will state the main result for both.

Theorem 3.10. *Let $\|\cdot\|$ be a 1–unconditional norm in the position of minimal M or ℓ–position. Then for any $\delta \in (0, 1/2]$ we have*

$$\mathbb{P}\big\{\|G\| \leqslant \delta \mathbb{E}\,\|G\|\big\} \leqslant 2\exp(-cn^{1-C\delta^2}),$$

where c, C are universal constants.

The main technical step of the proof is to construct a seminorm $\mathcal{T}(\cdot)$ satisfying three conditions:

- $\mathbb{E}\,\mathcal{T}(G) \geqslant c\,\mathbb{E}\,\|G\|$,
- $\mathcal{T}(x) \leqslant C\|x\|$ for all $x \in \mathbb{R}^n$, and
- $\frac{\sum_{i=1}^n (\mathbb{E}\,|\partial_i \mathcal{T}(G)|)^2}{(\mathbb{E}\,\mathcal{T}(G))^2} \leqslant C\delta^2/n$.

The first two conditions immediately yield

$$\mathbb{P}\big\{\|G\| \leqslant \delta \mathbb{E}\,\|G\|\big\} \leqslant \mathbb{P}\big\{\mathcal{T}(G) \leqslant C'\delta\, \mathbb{E}\,\mathcal{T}(G)\big\},$$

while the third condition, together with Theorem 2.7, implies the desired deviation estimate. The seminorm is constructed in Proposition 3.18 as a composition of three mappings: a diagonal contraction $\widetilde{D}$, an auxiliary seminorm $\mathcal{U}(\cdot)$ and a mapping $F_{\cdot,\cdot}(\cdot)$ defined below. Each of the mappings from the composition is responsible for particular structural properties of $\mathcal{T}(\cdot)$: $\widetilde{D}$ is needed to "balance" the quantities $\mathbb{E}\, G_i \partial_i \mathcal{T}(G)$, the auxiliary seminorm $\mathcal{U}(\cdot)$ provides an upper bound on the ℓ_1^n–norm of norming functionals, and $F_{\cdot,\cdot}(\cdot)$ controls the size of the support of the norming functionals, which, together with the estimate for $\mathbb{E}\, G_i \partial_i \mathcal{T}(G)$, implies the upper bound for $\sum_{i=1}^n (\mathbb{E}\, |\partial_i \mathcal{T}(G)|)^2$.

Everywhere in this subsection, by a *norming functional* for a seminorm $\|\cdot\|$ in $\mathbb{R}^n$ we mean any vector $x \in \mathbb{R}^n$ such that $\sup\limits_{\|y\|\leqslant 1} \langle y, x\rangle = 1$ and $\langle y, x\rangle = 0$ for all y with $\|y\| = 0$.

As the first step of the proof, we define a mapping F on the class of all 1–unconditional seminorms in $\mathbb{R}^n$.

Definition 3.11. Let $\|\cdot\|$ be a 1–unconditional seminorm, and fix a parameter $\tau \geqslant 1$. We will define a new 1–unconditional seminorm $F_{\|\cdot\|,\tau}(\cdot)$ as follows. Take any norming functional x for the seminorm $\|\cdot\|$. We associate with x a collection of vectors $\{v(x,I)\}_I \subset \mathbb{R}^n$ indexed over all subsets $I \subset [n]$, and with each $v(x,I)$ defined by

$$v(x,I) := \Big(1 + \frac{\|x\,\mathbf{1}_I\|_{\ell_1^n}}{\tau \|x\|_{\ell_1^n}}\Big) x\,\mathbf{1}_{[n]\setminus I},$$

where $\mathbf{1}_J$ denotes the indicator of a subset of indices J. In a sense, we truncate and rescale the original functional x. Now, set

$$F_{\|\cdot\|,\tau}(y) := \sup_x \sup_{I\subset[n]} \langle v(x,I), y\rangle, \quad y \in \mathbb{R}^n,$$

where the supremum is taken over all norming functionals x for the seminorm $\|\cdot\|$.

It is immediately clear that $F_{\|\cdot\|,\tau}(\cdot)$ is a 1–unconditional seminorm. Another elementary observation is

Lemma 3.12. *For any* 1*–unconditional seminorm* $\|\cdot\|$ *and* $\tau \geqslant 1$*, the seminorm* $F_{\|\cdot\|,\tau}(\cdot)$ *satisfies*

$$\|y\| \leqslant F_{\|\cdot\|,\tau}(y) \leqslant \Big(1 + \frac{1}{\tau}\Big)\|y\|, \quad \forall\, y \in \mathbb{R}^n.$$

The following is a crucial property that says that norming functionals for $F_{\|\cdot\|,\tau}(\cdot)$ are supported on the sets of large coordinates of corresponding vectors.

Lemma 3.13. *Let* $K > 0$ *be a parameter, and let* $\|\cdot\|$ *be a* 1*–unconditional seminorm such that every norming functional for* $\|\cdot\|$ *has* ℓ_1^n*–norm at most* K*. Further, let* $\tau \geqslant 1$ *and let* $F_{\|\cdot\|,\tau}(\cdot)$ *be as above. Take a vector* $y \in \mathbb{R}^n$ *with* $\|y\| \neq 0$ *and let* $\widetilde{x}$ *be a norming functional with respect to the seminorm* $F_{\|\cdot\|,\tau}(\cdot)$*, such that* $\langle y, \widetilde{x}\rangle = F_{\|\cdot\|,\tau}(y)$*. Then necessarily*

$$\operatorname{supp}\, \widetilde{x} \subset \Big\{ i \leqslant n:\ |y_i| \geqslant \frac{\tau\|y\|}{(\tau+1)^2 K} \Big\}.$$

Proof. Let x be the "original" norming functional with respect to $\widetilde{x}$, i.e. let x be a norming functional with respect to the seminorm $\|\cdot\|$, and let $I \subset \operatorname{supp}(\widetilde{x})^c$ be the set such that

$$\widetilde{x} = \Big(1 + \frac{\|x\,\mathbf{1}_I\|_{\ell_1^n}}{\tau \|x\|_{\ell_1^n}}\Big) x\,\mathbf{1}_{[n]\setminus I}.$$

Since our seminorms are 1–unconditional, we may assume without loss of generality that all components of y and x are non-negative. Fix any $i \in \operatorname{supp} \widetilde{x}$, define $J := I \cup \{i\}$, and set

$$x' := \Big(1 + \frac{\|x\,\mathbf{1}_J\|_{\ell_1^n}}{\tau \|x\|_{\ell_1^n}}\Big) x\,\mathbf{1}_{[n]\setminus J}.$$

Then x' belongs to the collection of functionals $\{v(x,I)\}_I$ from the definition of $F_{\|\cdot\|,\tau}(\cdot)$. Hence, $\langle \widetilde{x}, y\rangle = F_{\|\cdot\|,\tau}(y) \geqslant \langle x', y\rangle$. On the other hand,

$$\frac{1}{\widetilde{\alpha}}\langle \widetilde{x}, y\rangle = \frac{1}{\alpha'}\langle x', y\rangle + x_i y_i,$$

with

$$\widetilde{\alpha} := \Big(1 + \frac{\|x\,\mathbf{1}_I\|_{\ell_1^n}}{\tau \|x\|_{\ell_1^n}}\Big); \quad \alpha' := \Big(1 + \frac{\|x\,\mathbf{1}_J\|_{\ell_1^n}}{\tau \|x\|_{\ell_1^n}}\Big).$$

Note that

$$\frac{\alpha'}{\widetilde{\alpha}} = \frac{\tau\|x\|_{\ell_1^n} + \|x\,\mathbf{1}_J\|_{\ell_1^n}}{\tau\|x\|_{\ell_1^n} + \|x\,\mathbf{1}_I\|_{\ell_1^n}} = \frac{\tau\|x\|_{\ell_1^n} + \|x\,\mathbf{1}_I\|_{\ell_1^n} + x_i}{\tau\|x\|_{\ell_1^n} + \|x\,\mathbf{1}_I\|_{\ell_1^n}}.$$

Therefore, from the above we get

$$\langle \widetilde{x}, y\rangle \geqslant \frac{\tau\|x\|_{\ell_1^n} + \|x\,\mathbf{1}_I\|_{\ell_1^n} + x_i}{\tau\|x\|_{\ell_1^n} + \|x\,\mathbf{1}_I\|_{\ell_1^n}}\langle \widetilde{x}, y\rangle - \alpha' x_i y_i, \quad \text{so that} \quad \frac{\langle \widetilde{x}, y\rangle}{\tau\|x\|_{\ell_1^n} + \|x\,\mathbf{1}_I\|_{\ell_1^n}} \leqslant \alpha' y_i.$$

This inequality, together with the relation $\|y\| \leqslant F_{\|\cdot\|,\tau}(y) = \langle \widetilde{x}, y\rangle$ from Lemma 3.12, implies that

$$\frac{\|y\|}{(\tau+1)\|x\|_{\ell_1^n}} \leqslant \frac{\|y\|}{\tau\|x\|_{\ell_1^n} + \|x\,\mathbf{1}_I\|_{\ell_1^n}} \leqslant \frac{\langle \widetilde{x}, y\rangle}{\tau\|x\|_{\ell_1^n} + \|x\,\mathbf{1}_I\|_{\ell_1^n}} \leqslant \alpha' y_i \leqslant \Big(1 + \frac{1}{\tau}\Big) y_i,$$

and the result follows. □

Proposition 3.14. *For any $\varepsilon \in (0,1]$ there are $C, c > 0$ depending only on ε with the following property. Let $n \geqslant C$, let $\|\cdot\|$ be a 1–unconditional norm in the position of minimal M, with $\mathbb{E}\,\|G\| = 1$, and let $\delta > 0$. Then at least one of the following two assertions is true:*

- *Either $\mathbb{P}\{\|G\| \leqslant \delta\mathbb{E}\,\|G\|\} \leqslant \exp(-cn)$, or*
- *There is a 1–unconditional seminorm $\mathcal{U}(\cdot)$ in $\mathbb{R}^n$ such that $\mathcal{U}(y) \leqslant \|y\|$ for all $y \in \mathbb{R}^n$; all norming functionals for $\mathcal{U}(\cdot)$ have ℓ_1^n–norm at most $C\delta$; and for any subset $J \subset [n]$ of size $|J| \leqslant cn$ we have $\mathbb{E}\,\mathcal{U}(G\,\mathbf{1}_{[n]\setminus J}) \geqslant (1-\varepsilon)\mathbb{E}\,\|G\|$.*

Proof. Let I be the subset of all indices i such that $\mathbb{E}\,|\partial_i\|G\|| \geqslant C'\delta/n$, where $C' = C'(\varepsilon)$ will be chosen later. Note that for any vector $y \in \mathbb{R}^n$ we have

$$\|y\,\mathbf{1}_I\| \geqslant \mathbb{E}\sum_{i\in I} |y_i\,\partial_i\|G\|| \geqslant \frac{C'\delta}{n}\|y\,\mathbf{1}_I\|_1. \tag{3.11}$$

We will consider two possibilities:

- $|I| \geqslant \frac{2n}{C'}$. Then, in view of (3.11) and applying standard concentration estimates for Lipschitz functions, we get

$$\begin{aligned}\mathbb{P}\big\{\|G\| \leqslant \delta\mathbb{E}\,\|G\|\big\} &\leqslant \mathbb{P}\big\{\|G\,\mathbf{1}_I\| \leqslant \delta\big\}\\ &\leqslant \mathbb{P}\big\{\|G\,\mathbf{1}_I\|_1 \leqslant n/C'\big\}\\ &\leqslant \mathbb{P}\big\{\|G\,\mathbf{1}_I\|_1 \leqslant |I|/2\big\} \leqslant \exp(-cn),\end{aligned}$$

  for some $c > 0$ depending on ε.
- $|I| < \frac{2n}{C'}$. First, let

$$B^* := \big\{x \in \mathbb{R}^n : \max_{y:\,\|y\|\leqslant 1}\langle x, y\rangle \leqslant 1 \text{ and } \langle x, y\rangle = 0 \text{ for all } y \in \mathbb{R}^n \text{ with } \|y\| = 0\big\},$$

  and for any $t \geqslant 1$ define a collection of functionals

$$S(t) := \big\{x\,\mathbf{1}_{[n]\setminus I} : \ x \in B^* \quad \text{and} \quad \|x\,\mathbf{1}_{[n]\setminus I}\|_1 \leqslant C'\delta\,t\big\}.$$

  Choose any subset $J \subset [n]$. By our assumption and by Markov's inequality, we have

$$\mathbb{P}\Big\{\sum_{i\in[n]\setminus(I\cup J)} |\partial_i\|G\|| \leqslant C'\delta\,t\Big\} \geqslant 1 - \frac{1}{t}, \quad t \geqslant 1.$$

  On the other hand, again by Markov's inequality and the assumption that $\|\cdot\|$ is in the position of minimal M, we have

$$\mathbb{P}\Big\{\sum_{i\in I\cup J} G_i\,\partial_i\|G\| \leqslant t(|I| + |J|)/n\Big\} \geqslant 1 - \frac{1}{t}, \quad t \geqslant 1,$$

whence

$$\mathbb{P}\Big\{\sum_{i\in[n]\setminus(I\cup J)} G_i\,\partial_i\|G\| \geqslant \big(\|G\| - t(|I|+|J|)/n\big) \text{ and }$$
$$\sum_{i\in[n]\setminus(I\cup J)} |\partial_i\|G\|| \leqslant C'\delta\, t\Big\} \geqslant 1-\frac{2}{t}, \quad t\geqslant 1.$$

Together with the definition of $S(t)$ this gives

$$\mathbb{P}\Big\{\sup_{w\in S(t)} \langle G\,\mathbf{1}_{[n]\setminus J}, w\rangle \geqslant \big(\|G\| - t(|I|+|J|)/n\big)\Big\} \geqslant 1-\frac{2}{t}, \quad t\geqslant 1, \quad J\subset[n].$$

Finally, we define $\mathcal{U}(\cdot)$ as

$$\mathcal{U}(y) := \sup\big\{\langle y, w\rangle :\ w\in S(8/\varepsilon)\big\}, \quad y\in\mathbb{R}^n.$$

Clearly, $\mathcal{U}(\cdot)$ is a 1–unconditional seminorm with $\mathcal{U}(\cdot)\leqslant\|\cdot\|$. Further, it is known that in minimal M position one has that $k(\|\cdot\|)$ is at least of order $\log n$ [Proposition ? []]. So, since we have assume that $\mathbb{E}\|G\| = 1$ we have that (for large enough n) $\|G\|\geqslant 1-\frac{\varepsilon}{8}$ with probability at least $1-\frac{\varepsilon}{8}$. Then the above relations imply that for any $J\subset[n]$ with $|J|\leqslant 2n/C'$ we have

$$\mathbb{P}\Big\{\mathcal{U}(G\,\mathbf{1}_{[n]\setminus J}) \geqslant (1-\varepsilon/8 - 32/(C'\varepsilon))\Big\} \geqslant 1-\frac{\varepsilon}{8}-\frac{\varepsilon}{4}.$$

Finally, choose $C' := \frac{512}{\varepsilon^2}$. It is then easy to see that the last relation implies that $\mathbb{E}\,\mathcal{U}(G\,\mathbf{1}_{[n]\setminus J})\geqslant 1-\varepsilon$, completing the proof.

□

The next technical lemma describes a continuous diagonal contraction, which will be used later to "balance" coordinates of a seminorm.

Lemma 3.15. *Let $\mathcal{V}(\cdot)$ be any 1–unconditional seminorm in $\mathbb{R}^n$. Denote by $\mathbf{D}$ the set of diagonal $n\times n$ matrices with diagonal entries in $[0,1]$. Then for any $L>0$ there is a differentiable function $D:[0,\infty)\to\mathbf{D}$ having the following properties:*

- *$D(t)$ is a solution to initial value problem*
$$\begin{cases} D(0) := I_n; \\ \frac{\partial}{\partial t} d_{ii}(t) = H(D(t))_i, \quad i\leqslant n, \quad t\in[0,\infty), \end{cases}$$
where for each $i\leqslant n$, and $A=(a_{kj})\in\mathbf{D}$
$$H(A)_i := -\max\big(0, \mathbb{E}\,(G_i\,\partial_i\mathcal{V}(A\cdot)(G)) - L\big) - \max\Big(0, \frac{1}{4} - \Big|a_{ii}-\frac{1}{4}\Big|\Big),$$
and $\partial_i\mathcal{V}(A\cdot)(G)$ denotes i-th partial derivative of the seminorm
$$y \longrightarrow \mathcal{V}\Big(\sum_{i=1}^n a_{ii}y_i e_i\Big), \quad y\in\mathbb{R}^n,$$
at point G;
- *Setting $\widetilde{D} = (\widetilde{d}_{ii})\in\mathbf{D}$ to be the entry-wise limit of $D(t)$ when $t\to\infty$, we have $\widetilde{d}_{ii}\in\{0\}\cup[1/2,1]$ for all $i\leqslant n$,*
$$\mathbb{E}\,(G_i\,\partial_i\mathcal{V}(\widetilde{D}\cdot)(G)) \leqslant L \quad \text{for all } i \text{ with } \widetilde{d}_{ii}\neq 0,$$
and
$$\big|\{i\leqslant n:\ \widetilde{d}_{ii}=0\}\big| \leqslant \frac{2\mathbb{E}\,\mathcal{V}(G)}{nL}.$$

Proof. For any $A\in\mathbf{D}$ we have $H(A)\in(-\infty,0]^n$, and the function H is continuous everywhere on $\mathbf{D}$. Further, the relation

$$\mathbb{E}\,(G_i\,\partial_i\mathcal{V}(A\cdot)(G)) \leqslant \mathbb{E}\,\mathcal{V}\big(A(G_ie_i)\big) < a_{ii}\mathcal{V}(e_i), \quad i\leqslant n,$$

implies that $|H(A)_i|\leqslant a_{ii}$ whenever $a_{ii}\leqslant\frac{L}{\mathcal{V}(e_i)}$. Hence, the initial value problem stated above has a solution, by the Cauchy–Peano theorem.

Further, if $d_{ii}(t_0) < 1/2$ for some $t_0 > 0$ then the term $-\max\big(0, \frac{1}{4} - \big|d_{ii}(t) - \frac{1}{4}\big|\big)$ in the definition of the function H forces $d_{ii}(t)$ to converge to zero when $t \to \infty$. This implies $\widetilde{d}_{ii} \in \{0\} \cup [1/2, 1]$, $i \leqslant n$.

For the next property, note that for any two matrices $A, A + \Delta A \in \mathbf{D}$, and any vector $y \in \mathbb{R}^n$,

$$\mathcal{V}(Ay) \geqslant \mathcal{V}((A + \Delta A)y) + \langle x, -\Delta A y\rangle,$$

where x is a norming functional for $(A+\Delta A)y$ with respect to the seminorm $\mathcal{V}(\cdot)$. It is not difficult to check that the functional x can be represented in the form

$$x = \big((a_{ii} + \Delta a_{ii})^{-1}\partial_i \mathcal{V}((A + \Delta A)\cdot)(y)\big)_{i=1}^n,$$

where $\partial_i(\mathcal{V}((A+\Delta A)\cdot))(y)$ denotes the partial derivative of the seminorm $\mathcal{V}((A+\Delta A)\cdot)$, evaluated at point y (as long as the gradient of $\mathcal{V}((A + \Delta A)\cdot)$ at point y is well defined). Hence,

$$\mathcal{V}(Ay) \geqslant \mathcal{V}((A + \Delta A)y) + \sum_{i=1}^n \frac{-\Delta a_{ii}\, y_i}{a_{ii} + \Delta a_{ii}} \partial_i \mathcal{V}((A + \Delta A)\cdot)(y).$$

Applying this last identity to $D(t)$ and integrating (we are interested only in the intervals corresponding to $d_{ii}(t) \geqslant 1/2$), we obtain

$$\mathcal{V}(y) \geqslant \mathcal{V}(\widetilde{D}y) + \sum_{i=1}^n \int_0^\infty \mathbf{1}_{\{d_{ii}(t)\geqslant 1/2\}} \frac{-\frac{\partial}{\partial t} d_{ii}(t)\, y_i}{d_{ii}(t)} \partial_i \mathcal{V}(D(t)\cdot)(y)\, dt$$

for almost all $y \in \mathbb{R}^n$. Now, our definition of $D(t)$ implies that $\frac{\partial}{\partial t} d_{ii}(t) < 0$ and $d_{ii}(t) \geqslant 1/2$ only if $\mathbb{E}\,(G_i\, \partial_i \mathcal{V}(D(t)\cdot)(G)) \geqslant L$. Using this inequality and integrating the above relation with respect to the Gaussian measure, we get

$$\mathbb{E}\,\mathcal{V}(G) \geqslant \mathbb{E}\,\mathcal{V}(\widetilde{D}G) + \sum_{i=1}^n \int_0^\infty (-L)\mathbf{1}_{\{d_{ii}(t)\geqslant 1/2\}} \frac{\partial}{\partial t} d_{ii}(t)\, dt \geqslant \mathbb{E}\,\mathcal{V}(\widetilde{D}G) + \frac{nL}{2}\big|\big\{i \leqslant n :\ \widetilde{d}_{ii} = 0\big\}\big|,$$

implying the bound on the cardinality of $\big|\big\{i \leqslant n :\ \widetilde{d}_{ii} = 0\big\}\big|$. Finally, for any $i \leqslant n$ such that $\widetilde{d}_{ii} \neq 0$, we necessarily have $\mathbb{E}\,(G_i\, \partial_i \mathcal{V}(\widetilde{D}\cdot)(G)) \leqslant L$.

□

Remark 3.16. On the conceptual level, the continous contraction $D(t)$ constructed above is designed to act on the coordinates which give the main input to the expectation $\mathbb{E}\,\mathcal{V}(G)$. This way, we balance coordinates by making their input approximately equal. When "reasonable" balancing does not work for some coordinates, that is, when after being rescaled by 1/2 they still produce a large input to the norm, we zero them out.

Definition 3.17. Let $\|\cdot\|$ be a 1–unconditional seminorm in $\mathbb{R}^n$ and let $\tau \geqslant 1$ be a parameter. We will say that $\|\cdot\|$ is in *M_τ–position* if for any $i \leqslant n$ we have

$$\mathbb{E}\,(G_i\, \partial_i \|G\|) \leqslant \frac{\tau}{n} \mathbb{E}\,\|G\|,$$

where G is the standard Gaussian vector in $\mathbb{R}^n$.

Proposition 3.18. *There are universal constants $C, c > 0$ with the following property. Let $n \geqslant C$, let $\|\cdot\|$ be a 1–unconditional norm in the position of minimal M, with $\mathbb{E}\,\|G\| = 1$, and let $\delta > 0$. Then at least one of the following is true:*

- *Either $\mathbb{P}\{\|G\| \leqslant \delta \mathbb{E}\,\|G\|\} \leqslant \exp(-cn)$, or*
- *There is a 1–unconditional seminorm $\mathcal{T}(\cdot)$ in $\mathbb{R}^n$ in the M_C–position such that $\mathcal{T}(y) \leqslant 2\|y\|$ for all $y \in \mathbb{R}^n$; $\mathbb{E}\,\mathcal{T}(G) \geqslant \frac{1}{4}\mathbb{E}\,\|G\|$; and $\mathbb{E}\,|\partial_i \mathcal{T}(G)| \leqslant \frac{C\delta}{n}\mathbb{E}\,\|G\|$ for all $i \leqslant n$.*

Proof. We will apply Proposition 3.14 with $\varepsilon := 1/2$. Assuming that the constant $c > 0$ is sufficiently small and that the first assertion of the statement does not hold, there is a 1–unconditional seminorm $\mathcal{U}(\cdot)$ with the properties stated in Proposition 3.14. Define a seminorm $\mathcal{W}(\cdot)$ in $\mathbb{R}^n$ by setting

$$\mathcal{W}(y) := F_{\mathcal{U}(\cdot),1}(\mathcal{U}(y)), \quad y \in \mathbb{R}^n,$$

where the transformation $F_{\mathcal{U}(\cdot),1}$ was constructed earlier. Note that, in view of Lemma 3.12 and properties of $\mathcal{U}(\cdot)$,

$$\mathbb{E}\,\mathcal{W}(G) \leqslant 2. \tag{3.12}$$

Next, we apply Lemma 3.15 with $\mathcal{V} := \mathcal{W}$ and $L := \frac{\tau}{n}\mathbb{E}\,\|G\|$ (τ to be chosen a bit later) to obtain a diagonal contraction operator $\widetilde{D}$ with $\widetilde{d}_{ii} \in \{0\} \cup [1/2, 1]$, $i \leqslant n$, and with the set

$$J := \big\{i \leqslant n :\ \widetilde{d}_{ii} = 0\big\}$$

of cardinality at most $\frac{2n\mathbb{E}\,\mathcal{W}(G)}{\tau}$, which is less than $\frac{4n}{\tau}$, in view of (3.12). Set

$$\mathcal{T}(y) := \mathcal{W}(\widetilde{D}y), \quad y \in \mathbb{R}^n.$$

We claim that $\mathcal{T}(\cdot)$ is a 1–unconditional seminorm in the M_C–position, for an appropriate $C > 0$. By the properties of $\widetilde{D}$, for all $i \in [n] \setminus J$ we have $\widetilde{d}_{ii} \geqslant 1/2$, whence $\mathcal{T}(y) \geqslant \frac{1}{2}\mathcal{W}(y)$ for any $y \in \mathbb{R}^n$ with $\operatorname{supp} y \subset [n] \setminus J$. Take $\tau := 4/c_\varepsilon$, where the constant c_ε is taken from Proposiion 3.14. Then, by that proposition, we have

$$\mathbb{E}\,\mathcal{T}(G) \geqslant \mathbb{E}\,\mathcal{T}(G\,\mathbf{1}_{[n]\setminus J}) \geqslant \frac{1}{2}\mathbb{E}\,\mathcal{W}(G\,\mathbf{1}_{[n]\setminus J}) \geqslant \frac{1}{4}\mathbb{E}\,\|G\|.$$

Hence, by Lemma 3.15 and the definition of L,

$$\mathbb{E}\,(G_i \partial_i \mathcal{T}(G)) \leqslant \frac{\tau}{n}\mathbb{E}\,\|G\| \leqslant \frac{4\tau}{n}\mathbb{E}\,\mathcal{T}(G), \quad i \leqslant n,$$

so indeed $\mathcal{T}(\cdot)$ is in M_C–position for $C := 4\tau$.

It only remains to check the assertion about the partial derivatives, i.e. that

$$\mathbb{E}\,|\partial_i \mathcal{T}(G)| \leqslant \frac{C''\delta}{n}\mathbb{E}\,\|G\|, \quad i \leqslant n,$$

for some $C'' > 0$. For that, we will apply Lemma 3.13. Pick any vector $y \in \mathbb{R}^n$ with $\operatorname{supp} y \subset [n] \setminus J$ and $\mathcal{T}(y) \neq 0$. Observe that $\widetilde{D}x$ is a norming functional for y with respect to the seminorm $\mathcal{T}(\cdot)$ if and only if x is a norming functional for $\widetilde{D}y$ with respect to the seminorm $\mathcal{W}(\cdot)$. On the other hand, according to Lemma 3.13, we have

$$\operatorname{supp}\, x \subset \Big\{i \leqslant n :\ |\widetilde{d}_{ii} y_i| \geqslant \frac{\mathcal{U}(y)}{2^2 K}\Big\},$$

where K is the maximal ℓ_1^n–norm of a norming functional for $\mathcal{U}(\cdot)$. The definition of $\mathcal{U}(\cdot)$ implies that $K \leqslant C'\delta$. Since all diagonal entries of $\widetilde{D}$ indexed over $[n] \setminus J$, belong to the interval $[1/2, 1]$, we get from the above

$$\operatorname{supp}\, \widetilde{D}x \subset \Big\{i \leqslant n :\ |y_i| \geqslant \frac{\mathcal{T}(y)}{16C'\delta}\Big\},$$

or, equivalently,

$$\partial_i \mathcal{T}(z) \neq 0 \quad \text{only if } |z_i| \geqslant \frac{\mathcal{T}(z)}{16C'\delta} \quad \text{for almost all } z \in \mathbb{R}^n.$$

But this immediately implies that for any $i \leqslant n$ we have

$$|\partial_i \mathcal{T}(G)| \leqslant \frac{16C'\delta}{\mathcal{T}(G)} G_i \partial_i \mathcal{T}(G)$$

almost everywhere on the probability space, whence, together with the above relations,

$$\mathbb{E}\,|\partial_i \mathcal{T}(G)| \leqslant \frac{C_1 \tau \delta}{n}$$

for a universal constant $C_1 > 0$. The result follows. □

Proof of Theorem 3.10. We need to show that

$$\mathbb{P}\,\{\|G\| \leqslant \delta\mathbb{E}\,\|G\|\} < 2\exp\big(-cn^{1-C\delta^2}\big), \quad \delta \in (0, 1/2]. \tag{3.13}$$

The proof essentially follows by combining Proposition 3.18 and Theorem 2.7. If the first assertion of Proposition 3.18 holds then we are done. Otherwise, let $\mathcal{T}(\cdot)$ be the seminorm defined within the second assertion of that proposition. We set

$$L := \frac{\sum_{i=1}^n (\mathbb{E}\,|\partial_i \mathcal{T}(G)|)^2}{(\mathbb{E}\,\mathcal{T}(G))^2}.$$

Then, by the properties of $\mathcal{T}(\cdot)$, we have $L \leqslant C\delta^2/n$ for some $C > 0$. On the other hand, again by the properties of $\mathcal{T}(\cdot)$,

$$\mathbb{P}\left\{\|G\| \leqslant \delta\mathbb{E}\,\|G\|\right\} \leqslant \mathbb{P}\left\{\mathcal{T}(G) \leqslant 8\delta\mathbb{E}\,\mathcal{T}(G)\right\}.$$

It remains to apply Theorem 2.7. □

4. Small deviations for norms

The goal of this section is to obtain lower deviation estimates $\mathbb{P}\big\{\|G\| < (1-\varepsilon)\mathbb{E}\,\|G\|\big\}$, $\varepsilon \in (0, 1/2]$, for norms $\|\cdot\|$ satisfying certain "balancing" conditions. The following two theorems are consequences of the main technical result of the section (Theorem 4.16):

Theorem 4.1. *For any norm $\|\cdot\|$ in $\mathbb{R}^n$ there is an invertible linear transformation T such that*

$$\mathbb{P}\big\{\|T(G)\| < (1-\varepsilon)\mathbb{E}\,\|T(G)\|\big\} \leqslant 2\exp(-n^{c\varepsilon}), \quad \varepsilon \in (0, 1/2],$$

where $c > 0$ is a universal constant.

For 1–unconditional norms, we can choose certain classical positions to get these deviation estimates:

Theorem 4.2. *For any 1–unconditional norm $\|\cdot\|$ in $\mathbb{R}^n$ in the position of minimal M, or ℓ–position, or $w^{1,1}$–position, we have*

$$\mathbb{P}\big\{\|G\| < (1-\varepsilon)\mathbb{E}\,\|G\|\big\} \leqslant 2\exp(-n^{c\varepsilon}), \quad \varepsilon \in (0, 1/2],$$

where $c > 0$ is a universal constant.

For the standard ℓ_∞^n–norm in $\mathbb{R}^n$, a reverse estimate is known [Sch07]:

$$c\exp(-n^{C\varepsilon}) < \mathbb{P}\big\{\|G\|_\infty < (1-\varepsilon)\mathbb{E}\,\|G\|_\infty\big\}, \quad \varepsilon \in (0, 1/2],$$

which implies that the above result is optimal (in an appropriate sense).

The proof of the main result follows a different strategy compared to the results of Section 2. Although the operator P_t still plays a fundamental role, instead of "replacing" the original norm with $P_t\|G\|$ (as is done in the proofs of Theorems 2.7 and 2.10), we will replace it with an auxiliary seminorm by carefully choosing a subset of norming functionals for $\|\cdot\|$, in such a way that the seminorm satisfies strong lower deviation estimates. In that respect, the proof is similar to the approach from Subsection 3.3, although the actual construction is completely different.

4.1. **Construction of seminorms.** Let $H \geqslant 1$, $\tau \in [\log^{-1} n, 1/2]$ and $\delta \in (0, 1]$ be parameters. Given a norm $f(\cdot) = \|\cdot\|$ in $\mathbb{R}^n$ with $H \geqslant \mathrm{unc}(\|\cdot\|, \{e_i\}_{i=1}^n)$, and

$$n^{-\delta} \geqslant \frac{1}{R(f)} = \frac{\sum_{i=1}^n (\mathbb{E}\,|\partial_i f(G)|)^2}{\mathbb{E}\,\|\nabla f(G)\|_2^2}, \tag{4.1}$$

and a collection of Borel subsets $(\mathcal{E}_t)_{t\geqslant\tau}$ of $\mathbb{R}^n$, define the seminorm $\Upsilon(\cdot)$ as follows: For every $t \geqslant \tau$, set

$$F_t := \big\{x \in \mathbb{R}^n :\ \langle x, P_t\nabla f(x)\rangle \geqslant (1-4t)\mathbb{E}\,f(G) \quad \text{and} \quad \|P_t\nabla f(x)\|_2 \leqslant \|\nabla f\|_{L_2(\gamma_n)}\, n^{-\delta t/8}\big\}, \tag{4.2}$$

and

$$w(x) := \sup\big\{\langle x, P_t\nabla f(y)\rangle :\ y \in F_t \setminus \mathcal{E}_t,\ t \geqslant \tau\big\}, \quad x \in \mathbb{R}^n$$

(if $F_t \setminus \mathcal{E}_t = \emptyset$ for all $t \geqslant \tau$ then we set $w \equiv 0$). Then for every vector $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ we define

$$\Upsilon(x) := \max\Big(w(x), \frac{1}{H}\max_{\epsilon_1,\ldots,\epsilon_n\in\{-1,1\}} w\Big(\sum_{i=1}^n \epsilon_i x_i e_i\Big)\Big).$$

The next statement immediately follows from the definition:

Lemma 4.3. *For any parameters H, τ, δ and Borel subsets $(\mathcal{E}_t)_{t \geqslant \tau}$, the seminorm $\Upsilon(\cdot)$ satisfies*

- $\mathrm{unc}(\Upsilon(\cdot), \{e_i\}_{i=1}^n) \leqslant H$;
- $\Upsilon(x) \leqslant f(x)$ *for all* $x \in \mathbb{R}^n$;
- $\mathrm{Lip}(\Upsilon) \leqslant \mathrm{Lip}(f)\, n^{-\delta\tau/8}$.

We will view the above procedure as a construction of a family $\mathcal{F}$ of seminorms parameterized by H, τ, δ, f and Borel sets $(\mathcal{E}_t)_{t \geqslant \tau}$. In what follows, by $\Upsilon(\cdot)$ we will understand a particular seminorm from the family $\mathcal{F}$ for a given realization of parameters (which are clear from the context and are either given explicitly or required to take values from a specific range).

4.2. **Lower bound for the expectation $\mathbb{E}\,\Upsilon(G)$.** In what follows, G and Z denote independent standard Gaussian vectors in $\mathbb{R}^n$, and for any $t > 0$, by G_t we denote the random vector $e^{-t}G + \sqrt{1-e^{-2t}}Z$. The goal of this subsection is to prove

Proposition 4.4. *Let the norm $f(\cdot)$ satisfy* (4.1), *as well as the conditions $\delta\sqrt{\log n}\,\mathrm{Lip}(f) \leqslant \mathbb{E}\, f(G)$ and*

$$\mathbb{P}\big\{|f(G) - \mathbb{E}\, f(G)| \geqslant s\,\mathbb{E}\, f(G)\big\} \leqslant 2\exp(-\delta s \log n), \quad s \geqslant 0,$$

for some $\delta \in (0,1]$. Let $H > 0$, $\tau \in [\log^{-1} n, 1/2]$ and assume that the Borel sets $\mathcal{E}_t$, $t \in [\tau, 1/2]$, satisfy $\mathbb{P}(\mathcal{E}_t) \leqslant 2n^{-\delta t/4}$. Then the seminorm $\Upsilon(\cdot)$ defined above with parameters H, τ, δ, $(\mathcal{E}_t)_{t\geqslant\tau}$, satisfies

$$\mathbb{E}\,\Upsilon(G) \geqslant (1 - C_{4.4}\delta^{-2}\tau)\mathbb{E}\, f(G)$$

for some universal constant $C_{4.4} > 0$.

Remark 4.5. We note that the proposition, combined with the lower deviation Theorem 2.2, allows to obtain the lower deviation inequality for the norm $f(\cdot)$ in the regime $\varepsilon \in [C' \log\log n/\log n, 1/2]$ (taking $\mathcal{E}_t = \emptyset$ for all admissible t), namely

$$\mathbb{P}\big\{f(G) \leqslant (1-\varepsilon)\mathbb{E}\, f(G)\big\} \leqslant 2\exp(-n^{c'\varepsilon}),$$

for some $c', C' > 0$ depending only on δ. Indeed, setting $\tau := \delta^2\varepsilon/(2C)$, we get

$$\mathbb{P}\big\{f(G) \leqslant (1-\varepsilon)\mathbb{E}\, f(G)\big\} \leqslant \mathbb{P}\big\{\Upsilon(G) \leqslant (1-\varepsilon)(1-C\delta^{-2}\tau)^{-1}\mathbb{E}\,\Upsilon(G)\big\} \leqslant 2\exp\big(-c\varepsilon^2(\mathbb{E}\,\Upsilon(G))^2/\mathrm{Var}\Upsilon(G)\big),$$

where, in view of Lemma 4.3,

$$\mathrm{Var}\Upsilon(G) \leqslant \mathrm{Lip}(\Upsilon)^2 \leqslant \mathrm{Lip}(f)^2\, n^{-\delta\tau/4} \leqslant \delta^{-2}(\log n)^{-1}(\mathbb{E}\, f(G))^2\, n^{-\delta\tau/4},$$

implying the bound. In order to extend the range to all $\varepsilon \in (0, 1/2]$, we will need to carefully choose the events $\mathcal{E}_t$ in the definition of $\Upsilon(\cdot)$ in order to obtain stronger bounds on the variance $\mathrm{Var}\Upsilon(G)$. That will be done in the next subsections.

Lemma 4.6. *Assume that a norm $f(\cdot)$ in $\mathbb{R}^n$ satisfies $\delta\sqrt{\log n}\,\mathrm{Lip}(f) \leqslant \mathbb{E}\, f(G)$ for some parameter $\delta > 0$. Then for any $t \in (0, 1/2]$ and $x \in \mathbb{R}^n \setminus \{0\}$ such that the gradient of f at x is well defined, we have*

$$\mathbb{P}\big\{\langle e^{-t}x + \sqrt{1-e^{-2t}}Z, \nabla f(x)\rangle \leqslant e^{-t}f(x) - \delta^{-1}\sqrt{t/\log n}\,u\,\mathbb{E}\, f(G)\big\} \leqslant 2\exp(-cu^2), \quad u \geqslant 1,$$

where $c > 0$ is a universal constant.

Proof. Obviously,

$$\langle e^{-t}x + \sqrt{1-e^{-2t}}Z, \nabla f(x)\rangle = e^{-t}f(x) + \sqrt{1-e^{-2t}}\langle Z, \nabla f(x)\rangle.$$

The inner product $\langle Z, \nabla f(x)\rangle$ is equidistributed with $g\|\nabla f(x)\|_2$, where g is a standard Gaussian variable. Further, $\|\nabla f(x)\|_2 \leqslant \mathrm{Lip}(f) \leqslant \mathbb{E}\, f(G)/(\delta\sqrt{\log n})$, by the assumptions of the lemma. The result follows. □

Lemma 4.7. *Let the norm $f(\cdot)$ satisfy $\delta\sqrt{\log n}\,\mathrm{Lip}(f) \leqslant \mathbb{E}\, f(G)$ and*

$$\mathbb{P}\big\{|f(G) - \mathbb{E}\, f(G)| \geqslant s\mathbb{E}\, f(G)\big\} \leqslant 2\exp(-\delta s \log n), \quad s \geqslant 0,$$

for some parameter $\delta \in (0,1]$. Then for any $u \geqslant 1$ and $t \in [\log^{-1} n, 1/2]$ we have

$$\mathbb{P}\big\{\langle G, P_t\nabla f(G)\rangle \leqslant e^{-t}f(G) - \delta^{-1}\sqrt{t/\log n}\,u\,\mathbb{E}\, f(G)\big\} \leqslant 2\exp(-cu),$$

for some universal constant $c > 0$.

Proof. Fix any $u \geqslant 1$. Define a subset S of $\mathbb{R}^n$ by setting

$$S := \big\{x \in \mathbb{R}^n :\ \langle x, P_t \nabla f(x)\rangle \leqslant e^{-t} f(x) - \delta^{-1}\sqrt{t/\log n}\, u\, \mathbb{E}\, f(G)\big\}.$$

Fix for a moment any $x \in S$. By definition,

$$e^{-t} f(x) - \delta^{-1}\sqrt{t/\log n}\, u\, \mathbb{E}\, f(G) \geqslant \langle x, P_t \nabla f(x)\rangle = \mathbb{E}_Z \langle x, \nabla f(e^{-t}x + \sqrt{1-e^{-2t}}Z)\rangle.$$

The upper bound on the expectation then implies that there is an integer $i = i(x) \geqslant 1$ with

$$\mathbb{P}_Z\big\{\langle x, \nabla f(e^{-t}x + \sqrt{1-e^{-2t}}Z)\rangle \leqslant e^{-t} f(x) - ci\delta^{-1}\sqrt{t/\log n}\, u\, \mathbb{E}\, f(G)\big\} \geqslant \frac{1}{i^2}, \tag{4.3}$$

for a sufficiently small universal constant $c > 0$ (instead of $\frac{1}{i^2}$ on the right hand side, we can take the i-th element of arbitrary convergent series). Indeed, this immediately follows from the obvious relation

$$\begin{aligned}\mathbb{E}\,\xi &\geqslant (a-b)\mathbb{P}\{\xi > a-b\} + \sum_{i=1}^{\infty} \big(a-(i+1)b\big)\mathbb{P}\{a-(i+1)b < \xi \leqslant a - ib\}\\ &\geqslant a - b - 2b\sum_{i=1}^{\infty}\mathbb{P}\{\xi \leqslant a - ib\},\end{aligned}$$

which is valid for any variable ξ and any parameters $a \in \mathbb{R}$ and $b > 0$.

For a given $i \geqslant 1$, denote the collection of all $x \in S$ satisfying (4.3) by S_i. Since $\bigcup_{i=1}^{\infty} S_i = S$, we get that there is $i_0 \geqslant 1$ with $\gamma_n(S_{i_0}) \geqslant c'\gamma_n(S)/i_0^2$. Thus,

$$\mathbb{P}\big\{\langle G, \nabla f(G_t)\rangle \leqslant e^{-t} f(G) - ci_0\delta^{-1}\sqrt{t/\log n}\, u\, \mathbb{E}\, f(G)\big\} \geqslant \frac{c'}{i_0^4}\gamma_n(S).$$

This implies

$$\begin{aligned}&\mathbb{P}\big\{\langle G, \nabla f(G_t)\rangle \leqslant e^{-t} f(G_t) - \tfrac{c}{2} i_0\delta^{-1}\sqrt{t/\log n}\, u\, \mathbb{E}\, f(G)\big\}\\ &\quad\geqslant \frac{c'}{i_0^4}\gamma_n(S) - \mathbb{P}\big\{|f(G) - f(G_t)| \geqslant \tfrac{c}{2} i_0\delta^{-1}\sqrt{t/\log n}\, u\, \mathbb{E}\, f(G)\big\}.\end{aligned}$$

Observe that the probability on the left hand side of the inequality is equal to

$$\mathbb{P}\big\{\langle G_t, \nabla f(G)\rangle \leqslant e^{-t} f(G) - \tfrac{c}{2} i_0\delta^{-1}\sqrt{t/\log n}\, u\, \mathbb{E}\, f(G)\big\},$$

and in this form can be estimated with help of Lemma 4.6. This gives

$$\frac{c'}{i_0^4}\gamma_n(S) \leqslant \mathbb{P}\big\{|f(G) - f(G_t)| \geqslant \tfrac{c}{2} i_0\delta^{-1}\sqrt{t/\log n}\, u\, \mathbb{E}\, f(G)\big\} + 2\exp(-\tilde{c} i_0^2 u^2).$$

It remains to note that the condition on the concentration of $f(G)$ implies

$$\mathbb{P}\big\{|f(G) - f(G_t)| \geqslant \tfrac{c}{2} i_0\delta^{-1}\sqrt{t/\log n}\, u\, \mathbb{E}\, f(G)\big\} \leqslant 2e^{-c'' i_0 u\sqrt{t\log n}},$$

and solve the inequality for $\gamma_n(S)$. $\square$

Lemma 4.8. *Let $f(\cdot)$ be a norm in $\mathbb{R}^n$ satisfying* (4.1). *Then for any $t \in (0, 1/2]$ we have*

$$\mathbb{P}\big\{\|P_t \nabla f(G)\|_2 \leqslant \|\nabla f\|_{L_2(\gamma_n)}\, n^{-\delta t/8}\big\} \geqslant 1 - n^{-c\delta t},$$

where $c > 0$ is a universal constant.

Proof. Repeating the argument from the proof of Proposition 2.8, or employing (2.22), we get

$$\begin{aligned}\mathbb{E}\,\|P_t \nabla f(G)\|_2^2 \leqslant \mathbb{E}\,\|\nabla f(G)\|_2^2 R^{-\tanh t} &\leqslant \mathbb{E}\,\|\nabla f(G)\|_2^2\, n^{-\delta \tanh t}\\ &\leqslant \mathbb{E}\,\|\nabla f(G)\|_2^2\, n^{-\delta t/e}.\end{aligned}$$

The result follows by a standard application of Markov's inequality. $\square$

Lemma 4.9. *Let the norm $f(\cdot)$ satisfy $\delta\sqrt{\log n}\,\mathrm{Lip}(f) \leqslant \mathbb{E}\, f(G)$ and*

$$\mathbb{P}\big\{|f(G) - \mathbb{E}\, f(G)| \geqslant s\mathbb{E}\, f(G)\big\} \leqslant 2\exp(-\delta s \log n), \quad s \geqslant 0,$$

for some parameter $\delta \in (0,1]$. Then for any $t \in [\log^{-1} n, 1/2]$ we have

$$\mathbb{P}\big\{\langle G, P_t \nabla f(G)\rangle < (1-4t)\,\mathbb{E}\, f(G)\big\} \leqslant 2\exp(-c\delta\sqrt{t\log n})$$

for a universal constant $c > 0$.

Proof. We have

$$\begin{aligned}
&\mathbb{P}\big\{\langle G, P_t \nabla f(G)\rangle < (1-4t)\,\mathbb{E}\, f(G)\big\} \\
&\quad\leqslant \mathbb{P}\big\{\langle G, P_t \nabla f(G)\rangle < (1-t) f(G) - 2t\,\mathbb{E}\, f(G)\big\} \\
&\qquad\qquad + \mathbb{P}\big\{(1-t) f(G) - 2t\,\mathbb{E}\, f(G) < (1-4t)\,\mathbb{E}\, f(G)\big\}.
\end{aligned}$$

Applying Lemma 4.7, we get

$$\mathbb{P}\big\{\langle G, P_t \nabla f(G)\rangle < (1-t) f(G) - 2t\,\mathbb{E}\, f(G)\big\} \leqslant 2\exp\big(-c\delta\sqrt{t\log n}\big)$$

for some universal constant $c > 0$. On the other hand, the assumptions of the lemma imply

$$\mathbb{P}\big\{(1-t) f(G) - 2t\,\mathbb{E}\, f(G) < (1-4t)\,\mathbb{E}\, f(G)\big\} \leqslant 2\exp(-c'\delta t\log n)$$

for a universal constant $c' > 0$. The result follows. $\square$

Proof of Proposition 4.4. Without loss of generality, $\tau \leqslant 1/4$. We have

$$\begin{aligned}
\mathbb{E}\,\Upsilon(G) &= \int_0^\infty \mathbb{P}\big\{\Upsilon(G) \geqslant s\big\}\, ds \\
&\geqslant 4\int_\tau^{1/4} \mathbb{P}\big\{\Upsilon(G) \geqslant (1-4s)\mathbb{E}\, f(G)\big\}\,\mathbb{E}\, f(G)\, ds \\
&= (1-4\tau)\mathbb{E}\, f(G) - 4\int_\tau^{1/4} \mathbb{P}\big\{\Upsilon(G) < (1-4s)\mathbb{E}\, f(G)\big\}\,\mathbb{E}\, f(G)\, ds.
\end{aligned}$$

Take any $s \in [\tau, 1/4]$. Combining Lemmas 4.8 and 4.9 and the conditions on events $\mathcal{E}_s$, we get that the event $\{G \in F_s \setminus \mathcal{E}_s\}$ (where the sets $F_\cdot$ are defined in (4.2)) has probability at least $1 - 2\exp(-c'\delta\sqrt{s\log n})$, for some constant $c' > 0$. By the definition of the seminorm $\Upsilon(\cdot)$, we have $\Upsilon(G) \geqslant (1-4s)\mathbb{E}\, f(G)$ whenever $G \in F_s \setminus \mathcal{E}_s$. Hence,

$$\int_\tau^{1/4} \mathbb{P}\big\{\Upsilon(G) < (1-4s)\mathbb{E}\, f(G)\big\}\, ds \leqslant 2\int_\tau^{1/4} \exp(-c'\delta\sqrt{s\log n})\, ds \leqslant C\delta^{-2}\tau,$$

for some constant $C > 0$. The result follows. $\square$

4.3. **Seminorm deformation.** Let T be non-empty closed bounded origin-symmetric subset of $\mathbb{R}^n \setminus \{0\}$, and let $w(\cdot)$ be a seminorm in $\mathbb{R}^n$ given by

$$w(x) := \max_{y \in T}\langle x, y\rangle, \quad x \in \mathbb{R}^n.$$

Further, assume that for some $i \leqslant n$ and $\alpha \in (0,1)$ we have

$$\mathbb{E}\,|\partial_i w(G)| \geqslant \alpha\sqrt{\mathbb{E}\,\|\nabla w(G)\|_2^2}.$$

We are interested in properties of a new seminorm $\widetilde{w}(\cdot)$ obtained from $w(\cdot)$ by "removing" functionals with large scalar products with the i-th coordinate vector:

$$\widetilde{w}(x) := \max_{y \in T_i}\langle x, y\rangle, \quad x \in \mathbb{R}^n,$$

where

$$T_i := \big\{y \in T : \ |y_i| \leqslant \tfrac{\alpha}{4}\sqrt{\mathbb{E}\,\|\nabla w(G)\|_2^2}\big\}.$$

The main statement of the section is the following:

Proposition 4.10. *Let $w(\cdot)$, $\widetilde{w}(\cdot)$, $\alpha \in (0,1)$ and $i \leqslant n$ be as above. Then, we have*

$$\mathbb{E}\,\widetilde{w}(G) \leqslant \mathbb{E}\,w(G) - c\alpha^4\sqrt{\mathbb{E}\,\|\nabla w(G)\|_2^2},$$

where $c > 0$ is a universal constant.

Remark 4.11. The proposition can be interpreted as follows: if the partial derivative of the i–th coordinate of a seminorm is large then removing the "spikes" in the definition of the set of norming functionals produces a seminorm with expectation significantly less than the expectation of the original seminorm. In a sense, this implies that the total number of such coordinates cannot be large. This will be very important in the next subsection when we choose the events $(\mathcal{E}_t)$ for the parametric family of seminorms $\Upsilon(\cdot)$ and prove the main technical statement of the section.

Lemma 4.12. *Let $x \in \mathbb{R}^n$ be a vector such that the gradient $\partial_i w(x)$ is well defined and let us assume that $\partial_i w(x) \geqslant \frac{\alpha}{2}\sqrt{\mathbb{E}\,\|\nabla w(G)\|_2^2}$. Then, for any $r > 0$ one has*

$$\widetilde{w}(x + re_i) \leqslant w(x + re_i) - \frac{r\alpha}{4}\sqrt{\mathbb{E}\|\nabla w(G)\|_2^2}.$$

Proof. Let $r > 0$ and let $y \in T_i$ such that $\widetilde{w}(x + re_i) = \langle x + re_i, y\rangle$. It follows that

$$\widetilde{w}(x + re_i) = \langle x, y\rangle + ry_i \leqslant w(x) + \frac{r\alpha}{4}\sqrt{\mathbb{E}\|\nabla w(G)\|_2^2}, \tag{4.4}$$

where we have used the definition of T_i and the fact that $T_i \subseteq T$. On the other hand, the mapping $x_i \mapsto w(x_1, \dots, x_i, \dots x_n)$ is convex (when the other coordinates remain fixed), thus

$$w(x) \leqslant w(x + re_i) - r\partial_i w(x) \leqslant w(x + re_i) - \frac{r\alpha}{2}\sqrt{\mathbb{E}\|\nabla w(G)\|_2^2}.$$

Inserting the latter into (4.4) we get the result. □

Proof of Proposition 4.10. We fix $1 \leqslant i \leqslant n$ such that $\mathbb{E}|\partial_i w(G)| \geqslant \alpha\sqrt{\mathbb{E}\|\nabla w(G)\|_2^2}$. The trivial bound $\mathbb{E}|\partial_i w(G)|^2 \leqslant \mathbb{E}\|\nabla w(G)\|_2^2$ and the Paley-Zygmund inequality [BLM13, p.47] imply

$$\mathbb{P}\left\{|\partial_i w(G)| \geqslant \frac{\alpha}{2}\sqrt{\mathbb{E}\|\nabla w(G)\|_2^2}\right\} \geqslant \mathbb{P}\left\{|\partial_i w(G)| \geqslant \frac{1}{2}\mathbb{E}|\partial_i w(G)|\right\} \geqslant \frac{1}{4}\frac{(\mathbb{E}|\partial_i w(G)|)^2}{\mathbb{E}|\partial_i w(G)|^2} \geqslant \frac{\alpha^2}{4}.$$

It follows that[3] the set

$$Q := \left\{x : \partial_i w(x) \geqslant \frac{\alpha}{2}\sqrt{\mathbb{E}\,\|\nabla w(G)\|_2^2}\right\},$$

satisfies $\gamma_n(Q) \geqslant \alpha^2/8$. By a result of Kuelbs and Li [LK98, Theorem 1] we have for all $r > 0$ that

$$\gamma_n(Q + re_i) \geqslant \Phi(\Phi^{-1}(\gamma_n(Q)) - r).$$

The choice $r_0 := \frac{1}{8}\left(-\log(\gamma_n(Q)/2)\right)^{-1/2}$ yields

$$\gamma_n(Q + r_0 e_i) \geqslant \frac{\gamma_n(Q)}{4}.$$

Indeed; setting $p = \gamma_n(Q)/2 \in (0, 1/2)$ we may write

$$\Phi^{-1}(p) - \Phi^{-1}(p/2) = \int_{1/2}^{1}\frac{p}{I(tp)}\,dt \geqslant \frac{p}{2I(p)} \geqslant \frac{1}{8\sqrt{\log(1/p)}} = r_0,$$

[3]Note that if $Q = \left\{\partial_i w \geqslant \frac{\alpha}{2}\sqrt{\mathbb{E}\|\nabla w(G)\|_2^2}\right\}$, then the evenness of w implies $-Q = \left\{-\partial_i w \geqslant \frac{\alpha}{2}\sqrt{\mathbb{E}\|\nabla w(G)\|_2^2}\right\}$, and hence $\left\{|\partial_i w| \geqslant \frac{\alpha}{2}\sqrt{\mathbb{E}\|\nabla w(G)\|_2^2}\right\} \subseteq Q \cup (-Q)$.

where $I = \Phi' \circ \Phi^{-1}$ stands for the Gaussian isoperimetric function and we have also used the facts that $p \mapsto I(p)$ is nondecreasing in $(0, 1/2)$ and satisfies $I(p) \leqslant 4p\sqrt{\log(1/p)}$ for all $p \in (0, 1/2)$, see e.g. [BLM13, Lemma 10.3]. By Lemma 4.12, for any $z \in Q + r_0 e_i$ we have

$$\widetilde{w}(z) \leqslant w(z) - \frac{r_0 \alpha}{4} \sqrt{\mathbb{E}\,\|\nabla w(G)\|_2^2}.$$

This, together with the obvious relation $\widetilde{w}(y) \leqslant w(y)$, $y \in \mathbb{R}^n$, gives

$$\begin{aligned}\mathbb{E}\,w(G) - \mathbb{E}\,\widetilde{w}(G) \geqslant \mathbb{E}[(w(G) - \widetilde{w}(G))\mathbf{1}_{Q+r_0e_i}] &\geqslant \frac{r_0 \alpha \gamma_n(Q)}{16} \sqrt{\mathbb{E}\,\|\nabla w(G)\|_2^2} \\ &\geqslant \frac{c\alpha\gamma_n(Q)}{\sqrt{-\log(\gamma_n(Q)/2)}} \sqrt{\mathbb{E}\,\|\nabla w(G)\|_2^2}.\end{aligned}$$

The bound $\gamma_n(Q) \geqslant \alpha^2/8$ yields the estimate

$$\mathbb{E}w(G) - \mathbb{E}\widetilde{w}(G) \geqslant \frac{c\alpha^3}{\sqrt{\log(1/\alpha)}} \sqrt{\mathbb{E}\|\nabla w(G)\|_2^2},$$

as required. □

4.4. **Balancing the seminorm: improving bounds on the variance.** Our bound on the Lipschitz constant of a seminorm $\Upsilon(\cdot)$ from Proposition 4.4 holds under very general (or no) assumptions on parameters δ, H, the norm $f(\cdot)$ and events $(\mathcal{E}_t)_{t\geqslant\tau}$, but is not sufficiently strong to imply the main results of the section. However, the bound on the variance of the seminorm can be improved provided that H is "sufficiently small" and that the events $(\mathcal{E}_t)_{t\geqslant\tau}$ are carefully chosen to guarantee that the seminorm is "well balanced". It will be convenient to revise our notation a little. Let $\mathcal{F}$ be the collection of all seminorms defined as in the Subsection 4.1. We will inductively construct a finite sequence of seminorms $(\Upsilon_k)_{k=0}^m \subset \mathcal{F}$, with each $\Upsilon_k(\cdot)$ sharing the same set of parameters with the other seminorms of the sequence, except for the collection of events $(\mathcal{E}_t^k)_{t\geqslant\tau}$, which will be defined inductively, starting with $\mathcal{E}_t^0 := \emptyset$, $t \geqslant \tau$ for the initial seminorm $\Upsilon_0(\cdot)$.

Let $\theta > 0$ be a parameter (later, we will connect it with paramters H and δ), and assume that for some $k \geqslant 1$, the seminorm $\Upsilon_{k-1}(\cdot)$ has been constructed. If for every $i \leqslant n$ we have $\mathbb{E}\,|\partial_i \Upsilon_{k-1}(G)| < \theta \mathbb{E}\,\Upsilon_0(G)$, then we set $m := k-1$ and stop the construction. Otherwise, choose an index $i \leqslant n$ with $\mathbb{E}\,|\partial_i \Upsilon_{k-1}(G)| \geqslant \theta \mathbb{E}\,\Upsilon_0(G)$, and for any $t \geqslant \tau$ let

$$\mathcal{E}_t^k := \mathcal{E}_t^{k-1} \cup \big\{y \in \mathbb{R}^n : \ |\langle e_i, P_t \nabla f(y)\rangle| \geqslant \tfrac{\theta}{4} \mathbb{E}\,\Upsilon_0(G)\big\}.$$

Then we let $\Upsilon_k(\cdot)$ to be the seminorm from $\mathcal{F}$ defined with parameters τ, δ, H, f and $(\mathcal{E}_t^k)_{t\geqslant\tau}$, and proceed to the next step.

The following is a consequence of results of the previous subsection:

Lemma 4.13. *The sequence of seminorms constructed above is finite, with $m \leqslant C\theta^{-4}$, and for any $k \leqslant m$ we have*

$$\mathbb{E}\,\Upsilon_k(G) \leqslant \mathbb{E}\,\Upsilon_{k-1}(G) - c\theta^4\, \mathbb{E}\,\Upsilon_0(G),$$

for universal constants $c, C > 0$.

Proof. Take any two adjacent seminorms $\Upsilon_{k-1}(\cdot)$ and $\Upsilon_k(\cdot)$ from the sequence. According to the construction, there is an index $i \leqslant n$ such that $\sqrt{\mathbb{E}\,\|\nabla \Upsilon_{k-1}(G)\|_2^2} \geqslant \mathbb{E}\,|\partial_i \Upsilon_{k-1}(G)| \geqslant \theta \mathbb{E}\,\Upsilon_0(G)$. Define $\alpha > 0$ by $\alpha := \theta \mathbb{E}\,\Upsilon_0(G)/\sqrt{\mathbb{E}\,\|\nabla \Upsilon_{k-1}(G)\|_2^2}$. Note that the definition of the seminorm $\Upsilon_k(\cdot)$ implies that every norming functional of $\Upsilon_k(\cdot)$ is (a coordinate projection of) a vector $P_t \nabla f(y)$ for some $y \in F_t \setminus \mathcal{E}_t^k$ or a vector $\frac{1}{H}(\epsilon_j P_t \partial_j f(y))_{j=1}^n$ for some $y \in F_t \setminus \mathcal{E}_t^k$ and $\epsilon_j \in \{-1, 1\}$, $j \leqslant n$. Together with the definition of the events $\mathcal{E}_t^k$, $t \geqslant \tau$, this implies that every norming functional $v = (v_1, \ldots, v_n)$ of $\Upsilon_k(\cdot)$ satisfies $|v_i| \leqslant \frac{\theta}{4} \mathbb{E}\,\Upsilon_0(G) = \frac{\alpha}{4}\sqrt{\mathbb{E}\,\|\nabla \Upsilon_{k-1}(G)\|_2^2}$.

Then, by the definition of $\Upsilon_k(\cdot)$ and by Proposition 4.10, we get

$$\mathbb{E}\,\Upsilon_k(G) \leqslant \mathbb{E}\,\Upsilon_{k-1}(G) - c'\alpha^4 \sqrt{\mathbb{E}\,\|\nabla \Upsilon_{k-1}(G)\|_2^2} \leqslant \mathbb{E}\,\Upsilon_{k-1}(G) - \widetilde{c}\theta^4\, \mathbb{E}\,\Upsilon_0(G)$$

for a universal constant $\widetilde{c} > 0$. Recursive application of this bound gives

$$\mathbb{E}\,\Upsilon_k(G) \leqslant (1 - \widetilde{c}\theta^4 k)\mathbb{E}\,\Upsilon_0(G), \quad k \geqslant 0,$$

whence $m \leqslant \frac{1}{\tilde{c}\theta^4}$. $\square$

Next, we will estimate the Gaussian measure of events $\mathcal{E}_t^k$ constructed above.

Lemma 4.14. *Let the norm $f(\cdot)$ satisfy condition (4.1), and assume that $\delta\sqrt{\log n}\,\mathrm{Lip}(f) \leqslant \mathbb{E}\, f(G)$ and*

$$\mathbb{P}\big\{|f(G) - \mathbb{E}\, f(G)| \geqslant s\,\mathbb{E}\, f(G)\big\} \leqslant 2\exp(-\delta s \log n), \quad s \geqslant 0.$$

Further, assume that $\log^{-1} n \leqslant \tau \leqslant \delta^2/(2C_{4.4})$, where the constant $C_{4.4}$ is taken from Proposition 4.4. Then for any $k \leqslant m$ and $t \geqslant \tau$ we have

$$\gamma_n(\mathcal{E}_t^k) \leqslant \frac{Ck}{\theta} n^{-\delta/2},$$

where $C > 0$ is a universal constant.

Proof. It is sufficient to show that for any given $i \leqslant n$, we have

$$q := \gamma_n\Big\{y \in \mathbb{R}^n :\ |P_t \partial_i f(y)| \geqslant \frac{\theta}{4}\mathbb{E}\,\Upsilon_0(G)\Big\} \leqslant \frac{C}{\theta} n^{-\delta/2}.$$

By our conditions on parameters and in view of Proposition 4.4,

$$\mathbb{E}\,\Upsilon_0(G) \geqslant \frac{1}{2}\mathbb{E}\, f(G).$$

Hence,

$$\mathbb{E}\,|\partial_i f(G)| \geqslant \mathbb{E}\,|P_t\partial_i f(G)| \geqslant \frac{\theta q}{8}\mathbb{E}\, f(G) \geqslant c\theta q\sqrt{\mathbb{E}\,\|\nabla f(G)\|_2^2}.$$

This, together with condition (4.1), implies the bound. $\square$

As a combination of Lemma 4.3, Proposition 4.4 and the above Lemmas 4.13 and 4.14, we obtain

Proposition 4.15. *There is a universal constant $C_{4.15} > 0$ with the following property. Let $n \geqslant C_{4.15}$; let the norm $f(\cdot)$ and parameter δ be as in Lemma 4.14. Assume additionally that $\log^{-1} n \leqslant \tau \leqslant \delta^2/(2C_{4.15})$ and that $\theta^5 \geqslant n^{-\delta/4}$. Then the seminorm $\Upsilon_m(\cdot)$ constructed above satisfies*

- *$\Upsilon_m(x) \leqslant f(x)$ for all $x \in \mathbb{R}^n$;*
- *$\mathbb{E}\,\Upsilon_m(G) \geqslant (1 - C_{4.15}\delta^{-2}\tau)\mathbb{E}\, f(G)$;*
- *$\mathrm{Lip}(\Upsilon_m) \leqslant \mathrm{Lip}(f)\, n^{-\delta\tau/8}$;*
- *$\mathbb{E}\,|\partial_i(\Upsilon_m(G))| \leqslant 2\theta\mathbb{E}\,\Upsilon_m(G)$ for all $i \leqslant n$.*

Finally, we can prove the main technical result of the section. For convenience, we will explicitly restate all conditions on the parameters involved.

Theorem 4.16. *There are universal constants $C_{4.16}, c_{4.16} > 0$ with the following property. Let $\delta \in (0,1]$, and let $f(\cdot)$ be a norm in $\mathbb{R}^n$ such that $n^{\delta/64} \geqslant H := \mathrm{unc}(f, \{e_i\}_{i=1}^n)$. Further, take τ satisfying $\log^{-1} n \leqslant \tau \leqslant \delta^2/(2C_{4.16})$. Assume that (4.1) is satisfied, that $\delta\sqrt{\log n}\,\mathrm{Lip}(f) \leqslant \mathbb{E}\, f(G)$ and*

$$\mathbb{P}\big\{|f(G) - \mathbb{E}\, f(G)| \geqslant s\,\mathbb{E}\, f(G)\big\} \leqslant 2\exp(-\delta s \log n), \quad s \geqslant 0.$$

Then

$$\mathbb{P}\big\{f(G) \leqslant (1 - C_{4.16}\delta^{-2}\tau)\mathbb{E}\, f(G)\big\} \leqslant 2\exp\big(-n^{c_{4.16}\delta\tau}\big).$$

Proof. We will assume that $C_{4.16} \geqslant 2C_{4.15}$. Let $\theta := n^{-\delta/32}$, and let $\Upsilon_m(\cdot)$ be the seminorm constructed above. We have $\Upsilon_m(\cdot) \leqslant f(\cdot)$ and

$$\mathbb{E}\,\Upsilon_m(G) \geqslant (1 - C_{4.15}\delta^{-2}\tau)\mathbb{E}\, f(G),$$

by Proposition 4.15. This implies

$$\mathbb{P}\big\{f(G) \leqslant (1 - C_{4.16}\delta^{-2}\tau)\mathbb{E}\, f(G)\big\} \leqslant \mathbb{P}\big\{\Upsilon_m(G) \leqslant (1 - C_{4.15}\delta^{-2}\tau)\mathbb{E}\,\Upsilon_m(G)\big\}.$$

Next, we will estimate the variance $\mathrm{Var}\Upsilon_m(G)$. As a corollary of Talagrand's $L_1 - L_2$ bound (see Corollary 2.4), we obtain

$$\mathrm{Var}\Upsilon_m(G) \leqslant \frac{\widetilde{C}\mathbb{E}\,\|\nabla\Upsilon_m(G)\|_2^2}{1 + \log\frac{\mathbb{E}\,\|\nabla\Upsilon_m(G)\|_2^2}{\sum_{i=1}^n \|\partial_i\Upsilon_m\|_{L_1(\gamma_n)}^2}}, \tag{4.5}$$

for a universal constant $\widetilde{C} > 0$. Clearly, $\mathbb{E}\,\|\nabla\Upsilon_m(G)\|_2^2 \leqslant \mathrm{Lip}(\Upsilon_m)^2 \leqslant n^{-\delta\tau/4}\mathrm{Lip}(f)^2$, in view of Proposition 4.15. Further, by the same proposition we have for any $i \leqslant n$:

$$\|\partial_i \Upsilon_m\|_{L_1(\gamma_n)} \leqslant 2\theta \mathbb{E}\,\Upsilon_m(G).$$

This implies

$$\sum_{i=1}^n \|\partial_i \Upsilon_m\|_{L_1(\gamma_n)}^2 \leqslant \max_{j\leqslant n} \|\partial_j \Upsilon_m\|_{L_1(\gamma_n)} \Big(\sum_{i=1}^n \|\partial_i \Upsilon_m\|_{L_1(\gamma_n)}\Big) \leqslant C\theta H(\mathbb{E}\,\Upsilon_m(G))^2 \leqslant Cn^{-\delta/64}(\mathbb{E}\,\Upsilon_m(G))^2,$$

where we have used H–unconditionality of the standard vector basis with respect to $\Upsilon_m(\cdot)$. Now, if $(\mathbb{E}\,\Upsilon_m(G))^2 < n^{\delta/128}\mathbb{E}\,\|\nabla\Upsilon_m(G)\|_2^2$ then the last estimate gives

$$\sum_{i=1}^n \|\partial_i \Upsilon_m\|_{L_1(\gamma_n)}^2 \leqslant Cn^{-\delta/128}\mathbb{E}\,\|\nabla\Upsilon_m(G)\|_2^2,$$

so, (4.5) implies

$$\mathrm{Var}\Upsilon_m(G) \leqslant \frac{C'n^{-\delta\tau/4}\mathrm{Lip}(f)^2}{1+\delta\log n} \leqslant \frac{C'n^{-\delta\tau/4}\delta^{-2}(\mathbb{E}\,f(G))^2}{\log n + \delta\log^2 n}.$$

On the other hand, if $(\mathbb{E}\,\Upsilon_m(G))^2 \geqslant n^{\delta/128}\mathbb{E}\,\|\nabla\Upsilon_m(G)\|_2^2$ then the Poincaré inequality immediately implies a bound

$$\mathrm{Var}\Upsilon_m(G) \leqslant 2n^{-c\delta}(\mathbb{E}\,\Upsilon_m(G))^2 \leqslant 2n^{-c\delta}(\mathbb{E}\,f(G))^2.$$

As the final step of the proof, we observe that, according to Theorem 2.2, we have

$$\mathbb{P}\big\{\Upsilon_m(G) \leqslant (1 - C_{4.15}\delta^{-2}\tau)\mathbb{E}\,\Upsilon_m(G)\big\} \leqslant 2\exp\big(-c\delta^{-4}\tau^2(\mathbb{E}\,\Upsilon_m(G))^2/\mathrm{Var}\Upsilon_m(G)\big).$$

The result follows from the upper bound on the variance obtained above. □

4.5. **Proof of Theorems 4.1 and 4.2.**

Proof of Theorem 4.1. Apart from applying Theorem 4.16, the proof essentially reproduces the ideas from Subsection 3.2, so we only sketch the argument. The starting point is the dichotomy employed earlier in paper [PV18a]: given any norm $f(\cdot)$ in $\mathbb{R}^n$ with the unit ball in John's position, either the critical dimension $k(f)$ is greater than n^c, or, as a corollary of Alon–Milman's theorem [AM83], there is a subspace $X := (E, f(\cdot))$ of $(\mathbb{R}^n, f(\cdot))$ of dimension $c'n^{1/2}$ which has a $Cn^{c/2}$–unconditional basis. Here, $c > 0$ is a sufficiently small universal constant, which will be chosen later. In the former case, the required result immediately follows from standard concentration estimates for Lipschitz functions in the Gauss space. In the latter case, applying an appropriate linear transformation to X (whose existence is guaranteed by the Borsuk–Ulam theorem, see [PV18a, Lemma 3.1]), we obtain a normed space $(\mathbb{R}^{c'n^{1/2}}, \widetilde{f}(\cdot))$ such that the standard basis $\{e_i\}_{i=1}^{c'n^{1/2}}$ is $Cn^{c/2}$–unconditional, and, moreover, $\mathbb{E}\,|\partial_i \widetilde{f}(G')| = \mathbb{E}\,|\partial_j \widetilde{f}(G')|$ for all $i \neq j$, where G' is the standard Gaussian vector in $\mathbb{R}^{c'n^{1/2}}$. The $Cn^{c/2}$–unconditionality of the standard basis then implies

$$\begin{aligned}\sum_{i=1}^{c'n^{1/2}} \big(\mathbb{E}\,|\partial_i \widetilde{f}(G')|\big)^2 &\leqslant C'n^{c/2}\mathbb{E}\,\widetilde{f}(G')\mathbb{E}\,|\partial_1 \widetilde{f}(G')| \\ &= C''n^{c/2-1/2}\mathbb{E}\,\widetilde{f}(G') \sum_{i=1}^{c'n^{1/2}} \mathbb{E}\,|\partial_i \widetilde{f}(G')| \\ &\leqslant C'''n^{c-1/2}(\mathbb{E}\,\widetilde{f}(G'))^2.\end{aligned}$$

If $(\mathbb{E}\,\widetilde{f}(G'))^2 \leqslant n^c\mathbb{E}\,\|\nabla\widetilde{f}(G')\|_2^2$ and assuming that the constant $c > 0$ is sufficiently small, the norm $\widetilde{f}(\cdot)$ satisfies assumptions of Theorem 4.16 with $c'n^{1/2}$ in place of n, and with $\delta := 256c$, implying the lower deviation estimates for $\widetilde{f}(\cdot)$ (the conditions $\delta\sqrt{\log n}\,\mathrm{Lip}(\widetilde{f}) \leqslant \mathbb{E}\,\widetilde{f}(G')$ and $\mathbb{P}\big\{|\widetilde{f}(G') - \mathbb{E}\,\widetilde{f}(G')| \geqslant s\,\mathbb{E}\,\widetilde{f}(G')\big\} \leqslant 2\exp(-\delta s\log n)$, $\quad s \geqslant 0$, follow from the argument in the proof of Theorem 3.4 of [PV18a]). On the other hand, if $(\mathbb{E}\,\widetilde{f}(G'))^2 \geqslant n^c\,\mathbb{E}\,\|\nabla\widetilde{f}(G')\|_2^2$ then the deviation estimates follow immediately from Theorem 2.2 and the Poincaré inequality. Finally, we apply the argument from [PV18a] (see the proof of Theorem 4.5 there) to "lift" the lower deviation estimates for $\widetilde{f}(\cdot)$ back to an appropriate linear transformation of $(\mathbb{R}^n, f(\cdot))$. □

Proof of Theorem 4.2. Clearly, it is sufficient to show that any 1–unconditional norm $f(\cdot)$ in the position of minimal M, or ℓ–position, or $w^{1,1}$–position satisfies the conditions of Theorem 4.16 with δ being a universal constant. We consider the position of minimal M for concreteness. The conditions $\delta\sqrt{\log n}\,\mathrm{Lip}(f) \leqslant \mathbb{E}\, f(G)$ and $\mathbb{P}\big\{|f(G) - \mathbb{E}\, f(G)| \geqslant s\, \mathbb{E}\, f(G)\big\} \leqslant 2\exp(-\delta s \log n)$, $\quad s \geqslant 0$ (for constant δ) are known; see, in particular, [Tik18, PV18a]. If, additionally, (4.1) holds then a direct application of Theorem 4.16 gives the result. On the other hand, if (4.1) does not hold, say, for $\delta = 1/10$, then it can be easily checked, using the balancing conditions $\mathbb{E}\,(G_i \partial_i f(G)) = \frac{1}{n}\mathbb{E}\, f(G)$, $i \leqslant n$ for the position of minimal M, and the Poincaré inequality, that $\mathrm{Var} f(G) \leqslant n^{-c}(\mathbb{E}\, f(G))^2$ for a universal constant $c > 0$. Then Theorem 2.2 implies the required bound. □

5. Further Remarks and Questions

We conclude this paper with an application of Theorem 3.6 in the asymptotic geometric analysis. As we have already mentioned in the Introduction the motivation to study small-ball probabilities for norms stems for the need to better understand the (random) local Euclidean structure.

5.1. **Dvoretzky-type results.**

Although the connection of local almost Euclidean structure with the Gaussian concentration has its origins in Milman's work [Mil71], the link to the Gaussian small ball estimate had not been considered until the work of Klartag and Vershynin in [KV07]. The authors there establish the remarkable phenomenon that any centrally symmetric convex body in $\mathbb{R}^n$, admits random sections of dimension larger than the Dvoretzky number which are well bounded in terms of the expected radius.

Theorem 5.1 (Klartag–Vershynin). *Let A be a centrally symmetric convex body in $\mathbb{R}^n$. Then, for any $1 \leqslant m \leqslant cd(A)$, the random m-dimensional subspace F (with respect to the Haar probability measure on the Grassmannian $G_{n,m}$) satisfies*

$$A \cap F \subseteq \frac{C}{M(A)} B_F, \quad M(A) = \int_{S^{n-1}} \|\theta\|_A \, d\sigma(\theta),$$

with probability greater than $1 - e^{-cd(A)}$.

In order to make the connection of the small ball estimate with the one-sided random version of Dvoretzky's theorem more transparent, let us first recall the definition of the lower Dvoretzky dimension d due to Klartag and Vershynin: For a centrally symmetric convex body A in $\mathbb{R}^n$ and $\delta \in (0,1)$, one defines

$$d(A,\delta) = \min\left\{n, -\log \mathbb{P}\left(\|G\|_A \leqslant \delta \mathrm{med}(\|\cdot\|_A)\right)\right\},$$

where $\mathrm{med}(\|G\|_A)$ is the median of $\|G\|_A$. For $\delta = 1/2$ we simply write $d(A) = d(A, 1/2)$. Recall (1.10)

$$\mathbb{P}\left\{\|G\|_A \leqslant c\varepsilon \mathbb{E}\|G\|_A\right\} \leqslant \varepsilon^{\frac{d(A,\delta)-\log 2}{\log(1/\delta)}}, \quad \varepsilon \in (0,\delta). \tag{5.1}$$

Let us point out that the original formulation uses the uniform probability measure σ on the sphere S^{n-1}. However, one can equivalently work with the Gaussian measure for this problem in the light of the following (standard) estimates that compare the two measures:

$$\gamma_n(tB_2^n)\sigma(A \cap S^{n-1}) \leqslant \gamma_n(tA), \quad \gamma_n(sA) \leqslant \gamma_n(sB_2^n) + \sigma(A \cap S^{n-1}), \quad t, s > 0,$$

where A is any centrally symmetric convex body in $\mathbb{R}^n$; see in particular [KV07] and [LO05].

The next step is to express small ball estimates in terms of reverse Hölder inequalities. More precisely, the estimate (5.1) implies the following:

$$\left(\mathbb{E}\|G\|_A^{-q}\right)^{-1/q} \geqslant c\delta \mathbb{E}\|G\|_A, \quad q \leqslant cd(A,\delta)/\log(1/\delta).$$

The final step requires to bound the diameter of random sections in terms of the negative moments of $\|G\|_A$. Note that this cannot be achieved now by the standard net argument. The reason is that the latter works once we have first established an upper bound for the norm in the subspace, equivalently that the section contains relatively large ball. This is not the case in this regime. To overcome this obstacle Klartag and Vershynin devise a dimension lift technique based on the low M-inequality from [Kla04].

Lemma 5.2 (Dimension lift). *Let A be a centrally symmetric convex body in $\mathbb{R}^n$. Then, for any $1 \leqslant m \leqslant q \leqslant n/8$ we have*

$$\left(\int_{G_{n,m}} \operatorname{diam}(A \cap F)^m \, d\nu_{n,m}(F) \right)^{1/m} \leqslant \frac{C}{M(A)} \left(\frac{\mathbb{E}\|G\|_A}{\left(\mathbb{E}\|G\|_A^{-q}\right)^{-1/q}} \right)^2 . \tag{5.2}$$

The latter shows that the diameter of a random section is controlled by the tightness of the reverse Hölder inequality for the negative moments. In view of the above and the findings of this paper one gets the following:

Theorem 5.3. *Let A be a centrally symmetric convex body in $\mathbb{R}^n$ and let $\|\cdot\|_A$ be the corresponding norm.*

(1) *If $\|\cdot\|_A$ is 1-unconditional in the position of minimal M, or ℓ-position, or $w^{1,1}$–position, then for any $\delta \in (0,1/2)$ and $k \leqslant cn^{1-C\delta^2}$, the random k-dimensional subspace E of $\mathbb{R}^n$ satisfies*
$$A \cap E \subseteq \frac{C\delta^{-2}}{M(A)} B_E,$$
with probability greater than $1 - e^{-cn^{1-C\delta^2}}$.

(2) *In the general case, there exists an invertible linear map T with the following property: for every $1 \leqslant m \leqslant n^{1/5}$, the random m-dimensional subspace F satisfies*
$$TA \cap F \subseteq \frac{C}{M(TA)} B_2^n \cap F,$$
with probability greater than $1 - e^{-cn^{1/5}}$, where $C, c > 0$ are universal constants.

Quermassintegrals. Similar Dvoretzky-type results can be proved for other geometric quantities associated with the convex body than the negative moments of the norm. For example, as an application of the dimension lift (Lemma 5.2) and the Gaussian deviation inequality (Theorem 2.2) it is proved in [PPV19] a quantitive reversal of the classical Alexandrov inequality. The latter says that the sequence of quermassintegrals of a convex body A, appropriately normalized, is monotone. Recall that the k-th quermassintegral of A is the average of volume of the random k-dimensional projection of A, that is

$$Q_k(A) := \left(\frac{1}{|B_2^k|} \int_{G_{n,k}} |P_F A| \, d\nu_{n,k}(F) \right)^{1/k},$$

where $\nu_{n,k}$ is the (unique) Haar probability measure invariant under the orthogonal group action on the Grassmannian $G_{n,k}$. With this normalization Alexandrov's inequality reads as follows:

$$Q_n(A) \leqslant Q_{n-1}(A) \leqslant \ldots \leqslant Q_1(A).$$

Note that $Q_n(A)$ is the volume radius of A, $Q_{n-1}(A)$ is (an appropriate multiple of) the surface area of A, while $Q_1(A)$ stands for the mean width. We refer the reader to Schneider's monograph [Sch14] for related background material and an excellent exposition in convex geometry. The main result of [PPV19] asserts that a long initial segment of the above sequence is essentially constant, where the length and the almost constant behavior is quantified in terms of the parameter $1/\beta(A^\circ)$, namely

$$Q_k(A) \geqslant \left(1 - \sqrt{k\beta(A^\circ) \log\left(\frac{c}{k\beta(A^\circ)} \right)} \right) Q_1(A),$$

where A° is the polar body of A.

The strong small-ball probabilities established in the present paper yield an isomorphic reversal (up to constant) which extends to a polynomial order length, regardless the order of magnitude of $\beta(A)$. More precisely, we have the following:

Corollary 5.4. *Let A be a centrally symmetric convex body in $\mathbb{R}^n$ and let A° its polar.*

(1) *If A is 1–unconditional, and A° is in the ℓ-position, or position of minimal M, or $w^{1,1}$–position, then for any $\delta \in (0,1/2)$ and for any $k \leqslant cn^{1-C\delta^2}$ we have*
$$Q_k(A) \geqslant c\delta^2 Q_1(A).$$

(2) *In the general case there exists a linear image $\widetilde{A}$ of A such that for all $k \leqslant cn^{1/5}$ we have*

$$Q_k(\widetilde{A}) \geqslant cQ_1(\widetilde{A}).$$

The corollary follows by applying Theorem 5.3 in the dual setting to the polar body A°. We omit the details.

5.2. **Open questions.** 1. *Optimality of Corollary 2.4.* We do not know what is the optimal constant α in Corollary 2.4 for (even functions). If Corollary 2.4 can be established with $\alpha = 2$, then the argument described in §2.4.1. through the parameter $\widetilde{\beta}$ is redundant and then one has only Theorem 2.10 with β and $\omega(\delta) \asymp \delta^2$. Note that a posteriori the sharp constant α in Corollary 2.4 cannot exceed 2.

2. *Optimality of Theorem 3.6.* The exponent 1/4 we derive in the small ball estimate in the general case is product of the optimization in the dichotomy argument based on dependences from the Alon–Milman theorem. It seems plausible that a better organization of the argument and the tools might provide a slightly better exponent. However, we are not aware if one should expect that the exponent can be pushed arbitrary close to 1, or how this can be achieved.

3. *Further research direction.* Additional results on small ball probabilities for norms requires further investigation in other classical positions, say John's position or minimal surface area position. This can be achieved either by estimating (if possible) in those positions the global parameters $R(f)$ or $L(f)$ that govern the probabilistic estimates, or by rebuilding the existing methodology and tailoring it for geometric properties of the aforementioned positions. We do not pursue these research directions here.

References

[AM83] N. Alon and V. D. Milman. Embedding of l^k_∞ in finite-dimensional Banach spaces. *Israel J. Math.*, 45(4):265–280, 1983.

[BG99] S. G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163(1):1–28, 1999.

[BGL14] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.

[BKS03] I. Benjamini, G. Kalai, and O. Schramm. First passage percolation has sublinear distance variance. *Ann. Probab.*, 31(4):1970–1978, 2003.

[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.

[CEFM04] D. Cordero-Erausquin, M. Fradelizi, and B. Maurey. The (B) conjecture for the Gaussian measure of dilates of symmetric convex sets and related proble. *J. Funct. Anal.*, 214(2):410–427, 2004.

[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.

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

[DR50] A. Dvoretzky and C. A. Rogers. Absolute and unconditional convergence in normed linear spaces. *Proc. Nat. Acad. Sci. U. S. A.*, 36:192–197, 1950.

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

[FGG16] O. Friedland, O. Giladi, and O. Guédon. Small ball estimates for quasi-norms. *J. Theoret. Probab.*, 29(4):1624–1643, 2016.

[FJ80] T. Figiel and W. B. Johnson. Large subspaces of l^n_∞ and estimates of the Gordon-Lewis constant. *Israel J. Math.*, 37(1-2):92–112, 1980.

[FKPc77] T. Figiel, S. Kwapień, and A. Peł czyński. Sharp estimates for the constants of local unconditional structure of Minkowski spaces. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.*, 25(12):1221–1226, 1977.

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

[Gue99] O. Guedon. Kahane-Khinchine type inequalities for negative exponent. *Mathematika*, 46(1):165–173, 1999.

[Joh48] F. John. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, pages 187–204. Interscience Publishers, Inc., New York, N. Y., 1948.

[Kla04] B. Klartag. A geometric inequality and a low M-estimate. *Proc. Amer. Math. Soc.*, 132(9):2619–2628, 2004.

[KV07] B. Klartag and R. Vershynin. Small ball probability and Dvoretzky's theorem. *Israel J. Math.*, 157:193–207, 2007.

[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.

[Lat99] R. Latała. On the equivalence between geometric and arithmetic means for log-concave measures. In *Convex geometric analysis (Berkeley, CA, 1996)*, volume 34 of *Math. Sci. Res. Inst. Publ.*, pages 123–127. Cambridge Univ. Press, Cambridge, 1999.

[Led00] M. Ledoux. The geometry of Markov diffusion generators. *Ann. Fac. Sci. Toulouse Math. (6)*, 9(2):305–366, 2000. Probability theory.

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

[LK98] W. V. Li and J. Kuelbs. Some shift inequalities for Gaussian measures. In *High dimensional probability (Oberwolfach, 1996)*, volume 43 of *Progr. Probab.*, pages 233–243. Birkhäuser, Basel, 1998.

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

[LO05] R. Latała and K. Oleszkiewicz. Small ball probability estimates in terms of widths. *Studia Math.*, 169(3):305–314, 2005.

[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.

[LT17] A. Lytova and K. E. Tikhomirov. The variance of the ℓ_p^n-norm of the gaussian vector, and Dvoretzky's theorem, (2017). preprint.

[Mau91] B. Maurey. Some deviation inequalities. *Geom. Funct. Anal.*, 1(2):188–197, 1991.

[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.

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

[Pao12] G. Paouris. Small ball probability estimates for log-concave measures. *Trans. Amer. Math. Soc.*, 364(1):287–308, 2012.

[Pis74] G. Pisier. "Type" des espaces normés. pages 12 pp. (errata, p. E.1), 1974.

[Pis86] G. Pisier. Probabilistic methods in the geometry of Banach spaces. In *Probability and analysis (Varenna, 1985)*, volume 1206 of *Lecture Notes in Math.*, pages 167–241. Springer, Berlin, 1986.

[Pis89] G. Pisier. *The volume of convex bodies and Banach space geometry*, volume 94 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1989.

[PPV19] G. Paouris, P. Pivovarov, and P. Valettas. On a quantitative reversal of Alexandrov's inequality. *Trans. Amer. Math. Soc.*, 371(5):3309–3324, 2019.

[PV18a] G. Paouris and P. Valettas. Dichotomies, structure, and concentration in normed spaces. *Adv. Math.*, 332:438–464, 2018.

[PV18b] G. Paouris and P. Valettas. A Gaussian small deviation inequality for convex functions. *Ann. Probab.*, 46(3):1441–1454, 2018.

[PV19] G. Paouris and P. Valettas. Variance estimates and almost Euclidean structure. *Adv. Geom.*, 19(2):165–189, 2019.

[PVZ17] G. Paouris, P. Valettas, and J. Zinn. Random version of Dvoretzky's theorem in ℓ_p^n. *Stochastic Process. Appl.*, 127(10):3187–3227, 2017.

[Sam03] P. M. Samson. Concentration inequalities for convex functions on product spaces. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 33–52. Birkhäuser, Basel, 2003.

[Sch06] G. Schechtman. Two observations regarding embedding subsets of Euclidean spaces in normed spaces. *Adv. Math.*, 200(1):125–135, 2006.

[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.

[Sch14] R. Schneider. *Convex bodies: the Brunn-Minkowski theory*, volume 151 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, expanded edition, 2014.

[ST74] V. N. Sudakov and B. S. Tsirelson. 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.

[Sza80] S. J. Szarek. Volume estimates and nearly Euclidean decompositions for normed spaces. In *Seminar on Functional Analysis, 1979–1980 (French)*, pages Exp. No. 25, 8. École Polytech., Palaiseau, 1980.

[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. Embedding of l_k^∞ and a theorem of Alon and Milman. In *Geometric aspects of functional analysis (Israel, 1992–1994)*, volume 77 of *Oper. Theory Adv. Appl.*, pages 289–293. Birkhäuser, Basel, 1995.

[Tik18] K. E. Tikhomirov. Superconcentration, and randomized Dvoretzky's theorem for spaces with 1-unconditional bases. *J. Funct. Anal.*, 274(1):121–151, 2018.

[Val17] P. Valettas. On the tightness of Gaussian concentration for convex functions. *Journal d' Analyse Mathématique*, (to appear), 2017. https://arxiv.org/abs/1706.09446.

Department of Mathematics, Mailstop 3368, Texas A&M University, College Station, TX, 77843-3368.
E-mail address: grigorios.paouris@gmail.com

School of Mathematics, GeorgiaTech, Atlanta, GA, 30332.
E-mail address: ktikhomirov6@gatech.edu

Mathematics Department, University of Missouri, Columbia, MO, 65211.
E-mail address: valettasp@missouri.edu