

MOMENT ESTIMATES IMPLIED BY MODIFIED LOG-SOBOLEV INEQUALITIES *

RADOSŁAW ADAMCZAK¹, WITOLD BEDNORZ¹ AND PAWEŁ WOLFF²

Abstract. We study a class of logarithmic Sobolev inequalities with a general form of the energy functional. The class generalizes various examples of modified logarithmic Sobolev inequalities considered previously in the literature. Refining a method of Aida and Stroock for the classical logarithmic Sobolev inequality, we prove that if a measure on \mathbb{R}^n satisfies a modified logarithmic Sobolev inequality then it satisfies a family of L^p -Sobolev-type inequalities with non-Euclidean norms of gradients (and dimension-independent constants). The latter are shown to yield various concentration-type estimates for deviations of smooth (not necessarily Lipschitz) functions and measures of enlargements of sets corresponding to non-Euclidean norms. We also prove a two-level concentration result for functions of bounded Hessian and measures satisfying the classical logarithmic Sobolev inequality.

Mathematics Subject Classification. 60E15, 26D10.

Received May 8, 2016. Accepted December 7, 2016.

1. INTRODUCTION

Concentration of measure inequalities constitute one of the strongest and most widely used tools in modern high dimensional probability, geometry and analysis, crucial in establishing *e.g.* limit theorems or existence proofs by probabilistic method. Their importance was first noted in the 1970s and since then many powerful approaches have been established, which allow to prove concentration results, such as isoperimetric, transportation or functional inequalities (we refer to the monograph [30] by Ledoux for an overview). Among the functional inequalities approaches, the two which have proven particularly useful are those based on the Poincaré and logarithmic Sobolev inequalities. Recall that a Borel probability measure μ on \mathbb{R}^n satisfies the Poincaré inequality with constant D if

$$\mathrm{Var}_\mu f \leq D \mathbb{E}_\mu |\nabla f|^2$$

for all sufficiently smooth functions $f: \mathbb{R}^n \rightarrow \mathbb{R}$, whereas the logarithmic Sobolev inequality holds if for all such f ,

$$\mathrm{Ent}_\mu f^2 \leq D \mathbb{E}_\mu |\nabla f|^2,$$

Keywords and phrases. Concentration of measure, modified logarithmic Sobolev inequalities.

* *Research partially supported by Polish Ministry of Science and Higher Education Inventus Plus Grant no. IP 2011 000171.*

¹ Institute of Mathematics, University of Warsaw, ul. Banacha 2, 02-097 Warszawa, Poland.
R.Adamczak@mimuw.edu.pl; W.Bednorz@mimuw.edu.pl

² Institute of Mathematics, University of Warsaw, ul. Banacha 2, 02-097 Warszawa, Poland; and Institute of Mathematics, Polish Academy of Sciences, ul. Śniadeckich 8, 00-656 Warszawa, Poland. P.Wolff@mimuw.edu.pl

where $\text{Ent}_\mu f^2 = \mathbb{E}_\mu f^2 \log f^2 - \mathbb{E}_\mu f^2 \log \mathbb{E}_\mu f^2$ is the usual entropy of f^2 (throughout the article we use the probabilistic notation, treating f as a random variable on the probability space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu)$, in particular \mathbb{E}_μ denotes integration with respect to μ). Above and in the rest of the paper $|\cdot| = |\cdot|_2$ always denotes the standard Euclidean norm in \mathbb{R}^n .

As is well known, the Poincaré inequality yields subexponential concentration of Lipschitz functions, whereas the logarithmic Sobolev inequality implies sub-Gaussian estimates. There are also other functional inequalities, based either on a modification of the variance functional in the Poincaré inequality [10, 29] or a modification of the right-hand side in the logarithmic Sobolev inequalities [15, 20, 21], which yield concentration estimates with super-Gaussian rates or with rates between subexponential and sub-Gaussian. The general form of such log-Sobolev inequalities is

$$\text{Ent}_\mu f^2 \leq D \mathbb{E}_\mu \Psi \left(\frac{|\nabla f|}{f} \right) f^2 \quad (1.1)$$

for an appropriate function $\Psi: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ (we postpone the introduction of technical conditions on the function Ψ to subsequent sections).

Together with the discovery of two-level concentration inequalities by Talagrand (initially for the exponential distribution [38]), which improve the estimates based on Poincaré inequality and provide sub-Gaussian estimates for relatively small deviations and subexponential bounds for larger ones, a natural question arose whether results of this type could also be obtained *via* functional inequalities. It was soon answered in the affirmative by Bobkov and Ledoux [13] who derived new modified logarithmic Sobolev inequalities, which were subsequently extended by Gentil, Guillin and Miclo [20, 21] to inequalities yielding other (two-level and also more general) types of concentration.

Concentration estimates are usually derived from modified logarithmic Sobolev inequalities *via* differential inequalities on the log-Laplace transform of the function, a method commonly known as the Herbst argument. This method, being very elegant and powerful is however restricted to functions with finite Laplace transform, such as Lipschitz functions. For functions which do not satisfy the Lipschitz condition one can still use a modification of the Herbst approach, proposed by Aida and Stroock [4]. It relies on the analysis of moments and provides L_p -type Sobolev inequalities, which can yield concentration, provided that one controls the gradient of a function. More precisely, Aida and Stroock proved that the logarithmic Sobolev inequality implies inequalities of the form

$$\|f - \mathbb{E}_\mu f\|_p \leq C_D \sqrt{p} \left\| \|\nabla f\| \right\|_p$$

for $p \geq 2$, where $\|g\|_p = (\mathbb{E}_\mu |g|^p)^{1/p}$ is the p th moment of the function g . Clearly, controlling all moments of ∇f allows then to derive concentration results.

The aim of this work is to provide a uniform framework which would allow to treat the aforementioned modified logarithmic inequalities and provide concentration estimates with general profiles of the deviation bound for functions which are not necessarily Lipschitz. Our approach is based on a further refinement of the method by Aida and Stroock, which gives L_p -type Sobolev inequalities with non-Euclidean, p -dependent norms of the gradient on the right-hand side. The form of these norms corresponds to the type of concentration satisfied by the measure. As a particular case we obtain moment inequalities for smooth functions, which generalize moment estimates for linear combinations of independent random variables, derived by Gluskin and Kwapien [22] (see Thm. 3.14 below).

We remark that even in the classical Euclidean framework under defective log-Sobolev inequalities, our results improve certain aspects of the work by Aida and Stroock.

The precise form of the inequalities we obtain depends on the function Ψ in the modified logarithmic Sobolev inequality (1.1) and to describe it we need to introduce some technical notation. For this reason we postpone the precise formulation to subsequent sections and now we just announce that we will obtain inequalities of the form

$$\|f - \mathbb{E}_\mu f\|_p \leq C_D \left\| \|\nabla f|_{\Psi_p}\| \right\|_p,$$

for some norm $|\cdot|_{\Psi_p}$ on \mathbb{R}^n associated with p and the function Ψ . The geometry of the norms $|\cdot|_{\Psi_p}$ will be responsible for the character of concentration of measure valid for the measure μ . In the classical case, considered by Aida and Stroock, we have simply $|x|_{\Psi_p} = \sqrt{p}|x|$.

As corollaries we obtain concentration results for not necessarily Lipschitz functions as well as bounds on the size of enlargement of sets in a setting more general than considered before. We also prove some concentration results for Banach space valued polynomial chaos in the case of not necessarily product measures, extending previous work by Borell [17], Arcones–Giné [6], Łochowski [31] and Adamczak [1]. Additionally we derive comparison principles for real-valued polynomials (or more generally functions with bounded derivatives of higher order) generalizing previous estimates by Adamczak and Wolff [1] and a two-level concentration result for functions with bounded Hessian and measures satisfying the classical logarithmic Sobolev inequality.

Organization of the paper. In Section 2 we introduce the general framework for the inequalities we consider. Section 3 is devoted to the presentation of our results. The proofs are deferred to Section 4.

2. PRELIMINARIES

2.1. Basic notation

We will be working mostly with a fixed probability measure μ , therefore we will denote $\|f\|_p = (\mathbb{E}_\mu |f|^p)^{1/p}$, suppressing the dependence on μ in the notation.

Unless otherwise stated, x_i , $i = 1, \dots, n$, will denote coordinates of a point $x \in \mathbb{R}^n$, i.e. $x = (x_1, \dots, x_n)$. To distinguish norms on \mathbb{R}^n from the notation for moments, we will denote the former with single bars, e.g. for $r \geq 1$, $|\cdot|_r$ will stand for the ℓ_r^n norm, defined by $|x|_r = (\sum_{i=1}^n |x_i|^r)^{1/r}$. Other important norms will be introduced in the sequel. In the case of $r = 2$ we will often suppress the subscript r and write simply $|\cdot|$ for $|\cdot|_2$.

By C, c we will denote universal constants, whereas the notation $C(a)$ or C_a will be used for constants depending only on a parameter a . The values of constants may differ between occurrences.

2.2. Generalized Orlicz functions and modified logarithmic Sobolev inequalities

To formulate our results let us first introduce the general abstract form of the inequalities we will consider. Next we will illustrate it with examples of inequalities known in the literature, which fit our framework.

In what follows we will consider generalized Orlicz functions on \mathbb{R}^n , satisfying some standard technical conditions given in the following

Definition 2.1. We will say that a function $\Psi: \mathbb{R}^n \rightarrow \mathbb{R}_+ \cup \{\infty\}$ satisfies the condition (C) if the following holds

- (C1) $\Psi(0) = 0$ and Ψ is continuous at 0,
- (C2) $\Psi(x) > 0$ for $x \neq 0$,
- (C3) $\lim_{|x| \rightarrow \infty} \Psi(x) = \infty$,
- (C4) for every $x \in \mathbb{R}^n$, the function $t \mapsto \Psi(tx)$ is left-continuous on $(0, \infty)$,
- (C5) for every $x \in \mathbb{R}^n$, the function $t \mapsto \Psi(tx)/t$ is non-decreasing on $(0, \infty)$,
- (C6) Ψ is symmetric, i.e. $\Psi(x) = \Psi(-x)$ for all $x \in \mathbb{R}^n$.

Consider a probability measure μ on \mathbb{R}^n , absolutely continuous with respect to the Lebesgue measure. The general class of functional inequalities we will consider is described in the following definition.

Definition 2.2. Given a function Ψ , satisfying the condition (C) and a positive constant D we will say that μ satisfies the *modified logarithmic Sobolev inequality* $mLSI(\Psi, D)$ if for every bounded locally Lipschitz function $f: \mathbb{R}^n \rightarrow (0, \infty)$,

$$\text{Ent}_\mu f^2 \leq D \mathbb{E}_\mu \Psi \left(\frac{\nabla f}{f} \right) f^2. \quad (2.1)$$

Note that by the Rademacher theorem, ∇f exists μ -a.s. Under additional continuity assumptions on Ψ , by standard arguments one can show that μ satisfies $mLSI(\Psi, D)$ if and only if the above inequality is satisfied by all smooth functions of bounded support.

We will also consider defective versions of the logarithmic Sobolev inequalities.

Definition 2.3. Given a function Ψ , satisfying the condition (C) and constants $D, d \geq 0$ we will say that μ satisfies the *defective modified logarithmic Sobolev inequality* $dmLSI(\Psi, D, d)$, if for every bounded locally Lipschitz function $f: \mathbb{R}^n \rightarrow (0, \infty)$,

$$\text{Ent}_\mu f^2 \leq D \mathbb{E}_\mu \Psi \left(\frac{\nabla f}{f} \right) f^2 + d \mathbb{E}_\mu f^2. \quad (2.2)$$

As already mentioned, inequalities of this form have been considered by many authors starting from the classical work by Stam [37], Federbush [18], Gross [24] on the logarithmic Sobolev inequality, which in our language corresponds to the choice $\Psi(x) = |x|^2$, where $|\cdot|$ is the standard Euclidean norm. Modified versions were first considered by Bobkov–Ledoux [15], then *e.g.* by Bobkov–Zegarliński [16], Gentil–Guillin–Miclo [20, 21], Barthe–Roberto [9] and Barthe–Kolesnikov [8]. The defective versions were investigated *e.g.* by Rothaus [34], Bobkov–Zegarliński [16], Barthe–Kolesnikov [8], who obtained general criteria under which one can infer the non-defective version from the defective one. For instance it is known that under some additional conditions on the function Ψ , the defective inequality implies the non-defective version if one assumes certain Poincaré type inequalities.

Below we present the best known examples of modified log-Sobolev inequalities.

- The inequality (2.1) with $\Psi(x) = \|x\|^q$, where $q \in (1, 2]$ and $\|\cdot\|$ is some norm on \mathbb{R}^n was introduced by Bobkov and Ledoux in [15].
- In [13] Bobkov and Ledoux considered $\Psi(x) = \sum_{i=1}^n H(x_i)$, where

$$H(x) = \begin{cases} x^2 & \text{for } |x| \leq 1/2 \\ \infty & \text{for } |x| > 1/2. \end{cases}$$

This inequality was used to recover Talagrand’s concentration inequality for the exponential distribution [38]. In [20, 21] Gentil, Guillin and Miclo generalized the Bobkov–Ledoux inequality, by considering $H(x) = x^2 \mathbf{1}_{\{|x| \leq 1\}} + \Phi(|x|) \mathbf{1}_{\{|x| > 1\}}$ for a convex function Φ . When $\Phi(x)/x^2$ is non-decreasing on the positive half-line, the characterization of measures on \mathbb{R} which satisfy $mLSI(H, D)$ for some finite D was obtained in [9] (the characterization of the classical case $\Psi(x) = x^2$ was obtained earlier in the seminal paper [14] by Bobkov–Götze).

- The inequalities (2.1) and (2.2) for general Ψ corresponding to a measure $\mu(dx) = e^{-V(x)} dx$ on \mathbb{R}^n , under certain Bakry–Emery type conditions relating Ψ and V were studied *e.g.* by Barthe–Kolesnikov [8], Gentil [19], Shao [36].

Since the aforementioned articles introduce many different approaches for proving modified logarithmic Sobolev inequalities and the presentation of all of them is beyond the scope of this paper let us only mention that there is a multitude of examples of measures satisfying the inequalities in question. Currently available tools allow to both find mild sufficient conditions for a measure to satisfy the inequality (2.1) (resp. (2.2)) with a given function Ψ or starting from a measure find an appropriate Ψ so that the inequality (2.1) (resp. (2.2)) holds. Moreover, the usual tensorization and perturbation arguments developed for the classical logarithmic Sobolev inequality [5] work also in the modified setting and allow to construct further examples. In particular, the one-dimensional characterizations of modified logarithmic Sobolev inequalities allow to consider inequalities on \mathbb{R}^n with $\Psi(x) = \sum_{i=1}^n H(x_i)$, first for product measures and then for their bounded perturbations.

2.3. Families of Orlicz norms

Let us introduce another notion we need to formulate our results, namely a family of (quasi-)norms related to the function Ψ . In the Sobolev type inequalities we are about to derive these norms will be applied to the gradient of a function.

For $p > 0$ define

$$\Psi_p(x) = \frac{1}{p}\Psi(px).$$

Assume Ψ satisfies the condition (C). Then so does Ψ_p for all $p > 0$ and we consider a family of (quasi-)norms $|\cdot|_{\Psi_p}$ on \mathbb{R}^n , defined as

$$|x|_{\Psi_p} = \inf\{a > 0: \Psi_p(x/a) \leq 1\} = \inf\{a > 0: \Psi(px/a) \leq p\}.$$

It is easy to see that $|x|_{\Psi_p}$ is indeed a quasi-norm on \mathbb{R}^n , *i.e.* $|x|_{\Psi_p} = 0$ iff $x = 0$, $|tx|_{\Psi_p} = |t||x|_{\Psi_p}$ for $t \in \mathbb{R}$ and $|x + y|_{\Psi_p} \leq K_{\Psi_p}(|x|_{\Psi_p} + |y|_{\Psi_p})$ for some constant K_{Ψ_p} . If Ψ is in addition convex, then $|\cdot|_{\Psi_p}$ is a norm on \mathbb{R}^n . In what follows we will refer to the functional $|\cdot|_{\Psi_p}$ as the Ψ_p -norm or simply the norm, even if the function Ψ_p is not necessarily convex. Note also that from (C5) it follows that the norms $|\cdot|_{\Psi_p}$ are non-decreasing in p , *i.e.* for $0 < p < q$ and any $x \in \mathbb{R}^n$,

$$|x|_{\Psi_p} \leq |x|_{\Psi_q}. \tag{2.3}$$

Example 2.4. To demonstrate the reasons for introducing the above abstract definition of the norms $|\cdot|_{\Psi_p}$ and to show their role in the derivation of concentration of measure results, we will now list some families of norms corresponding to special choices of the function Ψ and present some special cases of known Sobolev type inequalities.

- (1) Clearly, if Ψ is homogeneous, in particular if it is a norm on \mathbb{R}^n , then for any $p > 0$, $|\cdot|_{\Psi_p} = \Psi$. This shows that to obtain interesting families of norms, $|\cdot|_{\Psi_p}$, which can be used to control the behaviour of moments of random variables, one needs to consider functions Ψ which grow faster than linearly.
- (2) If $\Psi(x) = \|x\|^\alpha$ for some norm $\|\cdot\|$ on \mathbb{R}^n and $\alpha > 1$, then $|x|_{\Psi_p} = p^{1/\alpha^*}\|x\|$, where α^* is the Hölder conjugate of α . This simple example can already illustrate the role played by the norms $|\cdot|_{\Psi_p}$ in our estimates. For instance, it is well known [33] that for a standard Gaussian measure γ on \mathbb{R}^n , for every smooth function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and every $p \geq 2$,

$$\|f - \mathbb{E}_\gamma f\|_p \leq C\sqrt{p}\left\|\left|\nabla f\right|\right\|_p \tag{2.4}$$

for some absolute constant C (where the moments $\|\cdot\|_p$ are calculated with respect to γ). Note that if $|\nabla f| \leq L$ on \mathbb{R}^n , then by applying the Chebyshev inequality in L_p and optimizing in p , one recovers (up to constants) the classical Gaussian concentration inequality, *i.e.*

$$\gamma(|f - \mathbb{E}_\gamma f| \geq t) \leq 2\exp(-ct^2/L^2)$$

for some universal constant c . It is easy to see that for $\Psi(x) = |x|^2$, the right hand side of (2.4) can be written as $C\left\|\left|\nabla f\right|_{\Psi_p}\right\|_p$, so the Sobolev inequality (2.4) is equivalent to

$$\|f - \mathbb{E}_\gamma f\|_p \leq C\left\|\left|\nabla f\right|_{\Psi_p}\right\|_p. \tag{2.5}$$

- (3) Let us now consider an example corresponding to a two-level concentration of measure. Let $\Psi(x) = \sum_{i=1}^n(|x_i|^2\mathbf{1}_{\{|x_i| \leq 1\}} + |x_i|^r\mathbf{1}_{\{|x_i| > 1\}})$ for some $r \in [2, \infty)$. Recall that by $|\cdot|_r$ we denote the ℓ_r^n norm, *i.e.* for $x = (x_1, \dots, x_n)$, $|x|_r = (\sum_{i=1}^n |x_i|^r)^{1/r}$. One can show that

$$|x|_{\Psi_p} \simeq \sqrt{p}|x| + p^{1/r^*}|x|_r,$$

where \simeq denotes two-sided estimates matching up to a universal multiplicative constant.

It turns out that the expression given above appears in Sobolev inequalities leading to two-level tail estimates. Namely, in [3] it is proved that if $r \in [2, \infty)$ and a measure μ on \mathbb{R}^n satisfies the $mLSI(\Psi, D)$, then for all smooth functions f and $p \geq 2$,

$$\|f - \mathbb{E}_\mu f\|_p \leq C_{r,D} \left(\sqrt{p} \|\nabla f|_2\|_p + p^{1/r^*} \|\nabla f|_r\|_p \right) \simeq C_{r,D} \left\| |\nabla f|_{\Psi_p} \right\|_p,$$

(where the moments $\|\cdot\|_p$ are calculated with respect to μ), which implies that for any Lipschitz function f and $t > 0$, we have

$$\mu(|f - \mathbb{E}_\mu f| \geq t) \leq 2 \exp \left(-c_{r,D} \min \left(\frac{t^2}{a^2}, \frac{t^{r^*}}{b^{r^*}} \right) \right), \tag{2.6}$$

where $a = \sup_{x \in \mathbb{R}^n} |\nabla f(x)|_2$, $b = \sup_{x \in \mathbb{R}^n} |\nabla f(x)|_r$. This corresponds to Talagrand’s two-level concentration inequality. As we will see, a similar moment bound holds also for $r \in (1, 2)$ and even for more general functions Ψ . We remark that for $r \in [1, 2]$, we have

$$|x|_{\Psi_p} \simeq p^{1/r^*} |(x_i^*)_{i=1}^{\lfloor p \rfloor}|_r + \sqrt{p} |(x_i^*)_{i=\lfloor p \rfloor+1}^n|_2$$

where $x_1^* \geq \dots \geq x_n^*$ is the non-increasing rearrangement of the sequence $|x_1|, \dots, |x_n|$.

- (4) Let us remark that in the special case, when μ is a product of measures on \mathbb{R} with log-concave tails and f is a linear functional, the inequalities of the form

$$\|f - \mathbb{E}_\mu f\|_p \leq \left\| |\nabla f|_{\Psi_p} \right\|_p$$

have been proved by Gluskin and Kwapien in [22]. In Section 3.4.2 we will use their result (which we recall in Thm. 3.14) to give an interpretation of our results in terms of auxiliary i.i.d. sequences.

3. MAIN RESULTS

In this section we will present all our results, deferring their proofs to Section 4.

3.1. Standing assumptions

Let us first state the main assumptions we are going to use throughout the article.

All the measures we will consider are assumed to be absolutely continuous with respect to the Lebesgue measure and this assumption will not be explicitly stated in all the theorems.

Our usual assumption on the function Ψ , beside the condition (C), will be the following growth condition: for some $1 < \alpha \leq 2 \leq \beta < \infty$ and $K \geq 1$,

$$\forall_{x \in \mathbb{R}^n \setminus \{0\}} \forall_{t \geq 1} \quad K^{-1}t^\alpha \leq \frac{\Psi(tx)}{\Psi(x)} \leq Kt^\beta. \tag{G_{K,\alpha,\beta}}$$

Note that the condition $(G_{K,\alpha,\beta})$ is stable under taking max or sum of functions Ψ , and if Ψ satisfies $(G_{K,\alpha,\beta})$ then so does Ψ_p for any $p > 0$.

3.2. Sobolev type inequalities

Let us now present our main results, *i.e.* Sobolev type inequalities, which constitute a basis for all the subsequent corollaries.

3.2.1. The defective case

We will start with moment estimates implied by defective inequalities. Recall the definition of the inequality $dmLSI(\Psi, D, d)$ given in formula (2.2). The proofs of results of this section are provided in Section 4.1.

Theorem 3.1. *Assume that $\Psi: \mathbb{R}^n \rightarrow \mathbb{R}$ satisfies the condition (C) and $(G_{K,\alpha,\beta})$ for some $1 < \alpha \leq 2 \leq \beta < \infty$ and $K \geq 1$. Let μ be a probability measure on \mathbb{R}^n satisfying $dmLSI(\Psi, D, d)$. Then for all locally Lipschitz functions $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and all $p \geq \beta$,*

$$\|f\|_p \leq e^{2d/\beta} \|f\|_\beta + \frac{2e}{\alpha - 1} ((KD)^{1/\alpha} \vee (KD)^{1/\beta}) \left\| |\nabla f|_{\Psi_p} \right\|_p. \tag{3.1}$$

Let us remark that for any $p \geq \beta$ and any $q \in (0, \beta)$ one can actually obtain

$$\|f\|_p \leq 2^{\frac{(p-q)\beta}{(p-\beta)q}} e^{2d\frac{p-q}{(p-\beta)q}} \|f\|_q + \frac{2e}{\alpha - 1} ((KD)^{1/\alpha} \vee (KD)^{1/\beta}) \left\| |\nabla f|_{\Psi_p} \right\|_p. \tag{3.2}$$

This inequality is a simple consequence of the well-known Lemma 4.1 stated in Section 4. Note that the constant $2^{\frac{(p-q)\beta}{(p-\beta)q}} e^{2d\frac{p-q}{(p-\beta)q}}$ obtained with the Lemma 4.1 explodes when $q \rightarrow 0$ or $p \rightarrow \beta$. We do not know if under the assumption of the above theorem, one can prove that for all $p \geq 2$,

$$\|f\|_p \leq C(D, d, \alpha, \beta) \left(\|f\|_2 + \left\| |\nabla f|_{\Psi_p} \right\|_p \right).$$

Fortunately to obtain concentration of measure inequalities, for fixed α, β , it is enough to control the growth of $\|f - \mathbb{E}_\mu f\|_p$ for $p > \beta$. Such a bound will be obtained in the non-defective case. It would be interesting to know if one can obtain meaningful Sobolev inequalities for the case $\beta = \infty$, which corresponds to the Bobkov–Ledoux inequality (satisfied e.g. for the product exponential distribution). In [3] it has been conjectured that in this case (for $d = 0$),

$$\|f - \mathbb{E}_\mu f\|_p \leq C_D \left(\sqrt{p} \left\| |\nabla f|_2 \right\| + p \left\| |\nabla f|_\infty \right\|_p \right)$$

and a weaker inequality was proved, with the second term on the right hand side replaced by $p \left\| |\nabla f|_\infty \right\|_\infty$.

On the other hand, the reason for excluding the case $\alpha = 1$ is made clear by the following proposition and the example below.

Proposition 3.2. *Assume that $\Psi: \mathbb{R}^n \rightarrow \mathbb{R}$ satisfies the condition (C) and for some $\beta \geq 2$ and $K \geq 1$,*

$$\forall x \in \mathbb{R}^n \setminus \{0\} \quad \forall t \geq 1 \quad \frac{\Psi(tx)}{\Psi(x)} \leq Kt^\beta.$$

Let μ be a probability measure on \mathbb{R}^n satisfying $dmLSI(\Psi, D, d)$. Then for all locally Lipschitz functions $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and all $p \geq \beta$,

$$\|f\|_p \leq e^{2d/\beta} \|f\|_\beta + 6 \log(p) (D \vee (KD)^{1/\beta}) \left\| |\nabla f|_{\Psi_p} \right\|_p. \tag{3.3}$$

Example 3.3. Let ν be a probability measure on \mathbb{R} with the distribution function

$$F_\nu(x) = \begin{cases} \frac{1}{2} e^{-e^{-x}+1}, & \text{for } x < 0, \\ 1 - \frac{1}{2} e^{-e^x+1}, & \text{for } x \geq 0. \end{cases} \tag{3.4}$$

Proposition 3.4. *The measure ν defined by (3.4) satisfies $mLSI(\Psi, 2)$ with $\Psi(x) = |x|$. Also, for every locally Lipschitz function $f: \mathbb{R} \rightarrow \mathbb{R}$ and $p > 1$,*

$$\|f\|_{L^p(\nu)} \leq \|f\|_{L^1(\nu)} + \log(p)\|f'\|_{L^p(\nu)}. \tag{3.5}$$

Moreover, if $f(x) = x$, then for $p \geq 1$, $\|f\|_{L^p(\nu)} \geq (2e)^{-1} \log(p)$.

The “moreover” part of the above proposition shows in particular that for $p \rightarrow \infty$, the $\log(p)$ factor in (3.3) or (3.5) cannot be improved.

Further examples:

- (1) If $\Psi(x) = |x|^2$ we are in the setting of the classical defective logarithmic Sobolev inequality. A result by Aida–Stroock [4] says that in this case for $p \geq 2$,

$$\|f\|_p^2 \leq e^{2d/p^*} \left(\|f\|_2^2 + D(p-2) \left\| |\nabla f| \right\|_p^2 \right). \tag{3.6}$$

On the other hand, Theorem 3.1, specialized to this case, asserts that for $p \geq 2$,

$$\|f\|_p \leq e^d \|f\|_2 + 2e\sqrt{D}\sqrt{p} \left\| |\nabla f| \right\|_p. \tag{3.7}$$

Let us note that the constant in front of the term involving ∇f in our inequality does not depend on d , which is not the case in (3.6). Therefore, even though the Aida–Stroock bound may behave in a better way for p close to 2, our estimate (via Chebyshev’s inequality optimized over p) shows that the large deviation behaviour of functions with polynomial growth of moments of ∇f can be controlled independently of d , which does not seem to follow from (3.6). For instance, if $\|f\|_2 < \infty$ and $\left\| |\nabla f| \right\|_p \leq Ap^\gamma$ for some $\gamma \geq 0$, then we obtain

$$\limsup_{t \rightarrow \infty} \frac{1}{t^{2/(1+2\gamma)}} \log \mu(|f| \geq t) \leq -c_{A,D}$$

for some (explicit) constant $c_{A,D} > 0$.

We remark that an improvement of the Aida–Stroock result for Lipschitz functions was obtained by Rothaus in [35].

- (2) Consider now $\Psi(x) = \sum_{i=1}^n (|x_i|^2 \mathbf{1}_{\{|x_i| \leq 1\}} + |x_i|^r \mathbf{1}_{\{|x_i| > 1\}})$ for some $r \in (1, \infty)$, which corresponds to the modified logarithmic Sobolev inequality introduced in [20] and [8] for $r \geq 2$ and in [19] for $r < 2$. In the former case, the inequalities of Theorem 3.1 read as

$$\|f\|_p \leq e^{2d/r} \|f\|_r + 2e(D^{1/2} \vee D^{1/r}) \left(\sqrt{p} \left\| |\nabla f|_2 \right\|_p + p^{1/r^*} \left\| |\nabla f|_r \right\|_p \right)$$

for $p \geq r$ (the assumption $(G_{K,\alpha,\beta})$ is satisfied with $\alpha = 2$, $\beta = r$ and $K = 1$). We remark that if in addition the underlying measure μ satisfies the Poincaré inequality, we can replace $\|f\|_r$ by $\|f\|_2$ and obtain an inequality for any $p \geq 2$ (with altered constants).

If $r \in (1, 2)$, one obtains

$$\|f\|_p \leq e^d \|f\|_2 + \frac{2e}{r-1} (D^{1/r} \vee D^{1/2}) \left(p^{1/r^*} \left\| |(\partial_i^* f)_{i=1}^{\lfloor p \rfloor}|_r \right\|_p + p^{1/2} \left\| |(\partial_i^* f)_{i=\lfloor p \rfloor+1}^n|_2 \right\|_p \right) \tag{3.8}$$

for $p \geq 2$, where $\partial_1^* f(x), \dots, \partial_n^* f(x)$ is the non-increasing rearrangement of the sequence $|\frac{\partial f(x)}{\partial x_1}|, \dots, |\frac{\partial f(x)}{\partial x_n}|$. Note that

$$p^{1/r^*} \left\| |(\partial_i^* f)_{i=1}^{\lfloor p \rfloor}|_r \right\|_p + p^{1/2} \left\| |(\partial_i^* f)_{i=\lfloor p \rfloor+1}^n|_2 \right\|_p \leq Cp^{1/2} \left\| |\nabla f|_2 \right\|_p,$$

so (3.8) is stronger than the bound (3.7) which has been derived from the classical logarithmic Sobolev inequality. Clearly to take advantage of the improvement one needs some additional information about the function f .

3.2.2. *The non-defective case*

Let us now pass to our second result, describing the moment estimates implied by modified LSI without defect, which will be later applied to obtain concentration bounds. Recall the definition of the inequality $mLSI(\Psi, D)$ given in formula (2.1). The proofs of results presented in this section are deferred to Section 4.2.

Denote

$$L(K, D, \alpha, \beta) := \frac{1}{\alpha - 1} (KD)^{1/\beta} + \left(\frac{1}{\alpha - 1} + \beta^{1/\alpha} \right) (KD)^{1/\alpha}.$$

Theorem 3.5. *Assume that $\Psi: \mathbb{R}^n \rightarrow \mathbb{R}$ satisfies the condition (C) and $(G_{K,\alpha,\beta})$ for some $1 < \alpha \leq 2 \leq \beta < \infty$. Let μ be a probability measure on \mathbb{R}^n satisfying $mLSI(\Psi, D)$. Then for all integrable (w.r.t. μ) and locally Lipschitz functions $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and all $p \geq \beta$,*

$$\|f - \mathbb{E}_\mu f\|_p \leq CL(K, D, \alpha, \beta) \left\| |\nabla f|_{\Psi_p} \right\|_p. \tag{3.9}$$

We note that, as is easy to see by truncation arguments, if the right-hand side of (3.9) is finite, then the function f is μ -integrable (in fact the p th moment of f is finite), so the integrability assumption is introduced in the above theorem just for formal reasons.

The advantage of (3.9) with respect to (3.1) is that it provides estimates of central moments of f in terms of norms of the gradient, without further dependence on any norms of f . This allows to derive concentration property for f based on the regularity of the gradient and as a consequence provides also concentration bounds at the level of enlargements of sets.

Theorem 3.5 is derived from Theorem 3.1 by means of Proposition 3.6 below, which allows to handle the central moment of order β .

Proposition 3.6. *Under the assumptions of Theorem 3.5, for every integrable (w.r.t. μ) and locally Lipschitz function $f: \mathbb{R}^n \rightarrow \mathbb{R}$,*

$$\|f - \mathbb{E}_\mu f\|_\beta \leq C \left((KD)^{1/\beta} + (KD\beta)^{1/\alpha} \right) \left\| |\nabla f|_{\Psi_\beta} \right\|_\beta. \tag{3.10}$$

3.3. Corollaries. Concentration: Deviation inequalities and enlargements of sets.

We will now explain how moment estimates of Theorem 3.5 imply concentration results expressed in terms of non-Euclidean norms of the gradient and non-Euclidean enlargements of measurable sets. The proofs of results from this section are presented in Section 4.3.

By Chebyshev’s inequality we obtain the following

Corollary 3.7. *Under the assumptions of Theorem 3.5, for all integrable (w.r.t. μ) and locally Lipschitz functions $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and any $p \geq \beta$,*

$$\mu \left(|f - \mathbb{E}_\mu f| \geq CL(K, D, \alpha, \beta) \left\| |\nabla f|_{\Psi_p} \right\|_p \right) \leq e^{-p}.$$

The above corollary allows to get concentration bounds if one controls the growth of $g(p) := \left\| |\nabla f|_{\Psi_p} \right\|_p$, no Lipschitz-type conditions need to be assumed. However, since in the simplest situation one may control the growth of $g(p)$ via a uniform bound on $|\nabla f|_{\Psi_a}$ for some $a > 0$ we will now specialize to functions which satisfy such a bound. To formulate the next corollary we will need to introduce the function $\omega_\Psi: \mathbb{R}_+ \rightarrow \mathbb{R}_+ \cup \{\infty\}$, defined as

$$\omega_\Psi(t) = \sup_{x \in \mathbb{R}^n \setminus \{0\}, \Psi(x) \neq \infty} \frac{\Psi(tx)}{\Psi(x)}.$$

If Ψ satisfies the condition (C) then ω_Ψ is left-continuous, $\lim_{t \rightarrow 0} \omega_\Psi(t) = 0$, $\lim_{t \rightarrow \infty} \omega_\Psi(t) = \infty$ and $t \mapsto \omega_\Psi(t)/t$ is non-decreasing, so ω_Ψ is strictly increasing on $\{\omega_\Psi < \infty\}$. Consider the inverse $\omega_\Psi^{-1}: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ of ω_Ψ , formally defined as

$$\omega_\Psi^{-1}(s) = \sup\{t > 0: \omega_\Psi(t) \leq s\}.$$

This function is continuous and since $t \mapsto \omega_\Psi(t)/t$ is non-decreasing, the function $s \mapsto s/\omega_\Psi^{-1}(s)$ is also continuous and non-decreasing. If additionally Ψ satisfies $(G_{K,\alpha,\beta})$ with some $K \geq 1$ and $1 < \alpha \leq 2 \leq \beta$ then for all $t > 0$,

$$K^{-1}(t^\alpha \wedge t^\beta) \leq \omega_\Psi(t) \leq K(t^\alpha \vee t^\beta) \tag{3.11}$$

which implies that

$$\lim_{t \rightarrow 0} \frac{\omega_\Psi(t)}{t} = 0, \quad \lim_{t \rightarrow \infty} \frac{\omega_\Psi(t)}{t} = \infty \tag{3.12}$$

and in turn $s/\omega_\Psi^{-1}(s) \rightarrow 0$ as $s \rightarrow 0$ and $s/\omega_\Psi^{-1}(s) \rightarrow \infty$ as $s \rightarrow \infty$. Therefore one can define a function $\omega_\Psi^*: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ to be a right-continuous inverse of $s \mapsto \frac{s}{\omega_\Psi^{-1}(s)}$, i.e.

$$\omega_\Psi^*(t) = \sup \left\{ s > 0: \frac{s}{\omega_\Psi^{-1}(s)} \leq t \right\}.$$

Note that $\omega_\Psi^*(t)$ is strictly increasing. We shall use the following observation (quite standard in the theory of Orlicz functions) which shows that the behaviour of the pair of functions ω_Ψ and ω_Ψ^* is similar to behaviour of conjugate functions:

Lemma 3.8. *Assume that Ψ satisfies the condition (C) and is such that (3.12) holds. Then for any $t > 0$,*

$$\omega_\Psi^*(t) = t \sup \left\{ u > 0: \frac{\omega_\Psi(u)}{u} \leq t \right\}. \tag{3.13}$$

Moreover, if $\lambda(t) = \sup_{y>0}(ty - \omega_\Psi(y))$ is the Legendre transform of ω_Ψ then for all $t > 0$,

$$\lambda(t) \leq \omega_\Psi^*(t) \leq \lambda(2t). \tag{3.14}$$

The role played the function ω_Ψ^* in concentration inequalities is revealed by the following

Corollary 3.9. *Under the assumptions of Theorem 3.5, if a locally Lipschitz function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ satisfies $|\nabla f(x)|_{\Psi_a} \leq b$, μ -a.e. for some $a, b > 0$, then for all $t > 0$,*

$$\mu(|f - \mathbb{E}_\mu f| \geq t) \leq \exp \left(\beta - a\omega_\Psi^* \left(\frac{t}{CL(K, D, \alpha, \beta)b} \right) \right).$$

A version of the above corollary was obtained in ([9], Prop. 26) in the special case when $\Psi(x) = \sum_{i=1}^n H(x_i)$ and H is an even convex function on \mathbb{R} such that $t \mapsto H(t)/t^2$ is non-decreasing on $(0, \infty)$. Our result and the result of [9] are not directly comparable, on the one hand in [9] there is no assumption concerning the parameter β and the constants are explicit, on the other hand the argument there is restricted to functions Ψ of a special form and to $\alpha = 2$, i.e. to the case of super-Gaussian tails. We remark that the proof in [9] is based on the classical Herbst argument with the Laplace transform.

We will now express the concentration property of a measure μ satisfying $mLSI(\Psi, D)$ in the language of enlargements of sets. We will do it under an additional assumption that the function Ψ is convex.

Corollary 3.10. *Assume that Ψ is convex and let Ψ^* be its Legendre transform. Under the assumptions of Theorem 3.5, for every Borel set $A \subseteq \mathbb{R}^n$ such that $\mu(A) \geq 1/2$ and every $u > 0$,*

$$\mu(A + \{x \in \mathbb{R}^n: \Psi^*(x) < u\}) \geq 1 - 2e^{\beta - u/(C(K, D, \alpha, \beta))}.$$

Example 3.11. Consider $\Psi(x) = \sum_{i=1}^n (|x_i|^2 \mathbf{1}_{\{|x_i| \leq 1\}} + |x_i|^r \mathbf{1}_{\{|x_i| > 1\}})$ for some $r \in (1, \infty)$. For $r < 2$, the function Ψ is not convex, but one can easily see that it is equivalent to a convex function, so we can still apply Corollary 3.10 at the cost of adjusting the constants. In this case (after replacing Ψ^* by an equivalent function) if one denotes by B_p^n the unit ball of ℓ_p^n , one obtains that for $r \geq 2$,

$$\mu(A + \sqrt{u}B_2^n + u^{1/r^*} B_{r^*}^n) \geq 1 - C_{r,D} e^{-\frac{u}{C_{r,D}}}$$

whereas for $r \in (1, 2)$,

$$\mu(A + (\sqrt{u}B_2^n) \cap (u^{1/r^*} B_{r^*}^n)) \geq 1 - C_{r,D} e^{-\frac{u}{C_{r,D}}}.$$

In the case of the product distribution with marginal densities proportional to $e^{-|x_i|^{r^*}}$ the above inequalities were first obtained by Talagrand [39] (for $r \geq 1$). The first functional approach was proposed by Bobkov and Ledoux [13] ($r = \infty$) and Barthe and Roberto [9] ($r \in [2, \infty)$) as well as more general concentration rates between subexponential and sub-Gaussian, who used the modified log-Sobolev inequalities introduced by Gentil, Guillin and Miclo. A uniform setting for various types of concentration inequalities, including the ones mentioned above was proposed by Gozlan, who used Poincaré inequalities with modified norms of gradients [23]. There are some subtle differences between the strength of various approaches, for instance Gozlan’s approach works also for $r = \infty$ and his constants do not depend on r . On the other hand in the non-product case his method introduces some dependence on the dimension n (see *e.g.* Prop. 1.2. in [23]).

3.4. Further corollaries. Concentration inequalities for polynomials

In this section we will present corollaries concerning polynomial like functions. First we will consider homogeneous polynomials with coefficients in a Banach space, then arbitrary real valued polynomials or more generally functions with bounded derivatives of order k . The proofs of presented results are deferred to Section 4.4.

To formulate our results in a concise way we will need to introduce some additional notation. Namely for two k -indexed matrices $A = (a_{i_1, \dots, i_k})_{i_1, \dots, i_k=1}^n$ and $B = (b_{i_1, \dots, i_k})_{i_1, \dots, i_k=1}^n$, where $a_{i_1, \dots, i_k} \in E$ for some Banach space E and $b_{i_1, \dots, i_k} \in \mathbb{R}$ we set

$$\langle A, B \rangle = \sum_{i_1, \dots, i_k=1}^n a_{i_1, \dots, i_k} b_{i_1, \dots, i_k}.$$

Moreover, for vectors $x^1, \dots, x^k \in \mathbb{R}^n$ we define $x^1 \otimes \dots \otimes x^k = (x_{i_1}^1 \dots x_{i_k}^k)_{i_1, \dots, i_k=1}^n$. With this convention, the E -valued homogeneous form of degree k , given by matrix A as above, *i.e.*

$$\sum_{i_1, \dots, i_k=1}^n a_{i_1, \dots, i_k} x_{i_1} \dots x_{i_k}$$

can be written simply as $\langle A, x^{\otimes k} \rangle$.

By $\mathbf{D}^k f$ we will denote the k th derivative of a function $f: \mathbb{R}^n \rightarrow \mathbb{R}$, which we will identify with the corresponding k -indexed matrix of partial derivatives.

3.4.1. Concentration for Banach space valued chaos

Let $(E, |\cdot|_E)$ be a separable Banach space and $A = (a_{i_1, \dots, i_k})_{i_1, \dots, i_k=1}^n$ a k -indexed E -valued matrix and $X = (X_1, \dots, X_n)$ a random vector in \mathbb{R}^n . We will consider the random variable $Z = |\langle A, X^{\otimes k} \rangle|_E$. Without loss of generality we will assume that A is symmetric, *i.e.* for any permutation σ of the set $\{1, \dots, k\}$, $a_{i_1, \dots, i_k} = a_{i_{\sigma(1)}, \dots, i_{\sigma(k)}}$.

Our main result is the following

Theorem 3.12. *Assume that Ψ is a convex function satisfying the conditions (C) and $(G_{K,\alpha,\beta})$ and let X be a random vector in \mathbb{R}^n , whose law is absolutely continuous and satisfies the $mLSI(\Psi, D)$. For any $p \geq \beta$,*

$$\|Z - \mathbb{E}Z\|_p \leq C_{D,K,\alpha,\beta,k} \sum_{j=1}^k \mathbb{E} \sup_{y^1, \dots, y^j \in A_{\Psi,p}} |\langle A, y^1 \otimes \dots \otimes y^j \otimes X^{\otimes(k-j)} \rangle|_E, \tag{3.15}$$

where $A_{\Psi,p} = \{x \in \mathbb{R}^n : \Psi^*(x) \leq p\}$. As a consequence, for any $p \geq \beta$,

$$\mathbb{P} \left(|Z - \mathbb{E}Z| \geq C_{D,K,\alpha,\beta,k} \sum_{j=1}^k \mathbb{E} \sup_{y^1, \dots, y^j \in A_{\Psi,p}} |\langle A, y^1 \otimes \dots \otimes y^j \otimes X^{\otimes(k-j)} \rangle|_E \right) \leq e^{-p}.$$

Versions of the above theorem were first obtained for Gaussian vectors by Borell [17] and Arcones–Giné [6]. Subsequently they were proved for X with independent coordinates possessing log-concave tails by Lochowski [31] and Adamczak [1]. However, they provided rather estimates of $\|Z\|_p$ and the deviation of Z above $\mathbb{E}Z$, than concentration around $\mathbb{E}Z$.

Let us analyze the quantities appearing on the right-hand side of (3.15). Except for the one corresponding to $j = k$, they are all expectations of suprema of random variables and as such they are difficult to estimate. The exceptional term however is “deterministic” and it is easy to see that for $p \rightarrow \infty$ it dominates the whole sum. Estimates of this form may be therefore used to obtain some large deviation type estimates for $|Z - \mathbb{E}Z|$. Also, in certain situations estimates of the troublesome expectations are available. This is the case *e.g.* if X is a Gaussian vector and E is the real line [28] or more generally E is a Hilbert space (this result is unpublished but it may be recovered from estimates in [28]), and also if $E = \mathbb{R}$, X has independent coordinates with log-concave tails and $k \leq 3$ [2, 26, 27].

Example 3.13. Let us illustrate Theorem 3.12 on a simple example of a real-valued quadratic form $Z = \sum_{i,j=1}^n a_{ij} X_i X_j$ in a centered random vector $X = (X_1, \dots, X_n)$ whose law μ satisfies $mLSI(\Psi, D)$ with $\Psi(x) = |x|_q^q$ for some $q \in (1, 2]$ (the case studied in [15, 16]). We have $\Psi^*(x) = C_q |x|_{q^*}^{q^*}$, therefore we obtain

$$\begin{aligned} \|Z - \mathbb{E}Z\|_p &\leq C_{D,q} \left(p^{1/q^*} \mathbb{E} \sup_{y \in B_{q^*}^n} \left| \sum_{i,j=1}^n a_{ij} y_i X_j \right| + p^{2/q^*} \sup_{x,y \in B_{q^*}^n} \left| \sum_{i,j=1}^n a_{ij} x_i y_j \right| \right) \\ &= C_{D,q} \left(p^{1/q^*} \mathbb{E} \left(\sum_{i=1}^n \left| \sum_{j=1}^n a_{ij} X_j \right|^q \right)^{1/q} + p^{2/q^*} \sup_{x,y \in B_{q^*}^n} \left| \sum_{i,j=1}^n a_{ij} x_i y_j \right| \right) \\ &\leq C_{D,q} \left(p^{1/q^*} \left(\sum_{i=1}^n \mathbb{E} \left| \sum_{j=1}^n a_{ij} X_j \right|^q \right)^{1/q} + p^{2/q^*} \sup_{x,y \in B_{q^*}^n} \left| \sum_{i,j=1}^n a_{ij} x_i y_j \right| \right). \end{aligned}$$

Now, by Proposition 3.6, applied with $\beta = 2$, for each i (note that Ψ satisfies $(G_{K,\alpha,\beta})$ with $K = 1, \alpha = q$ and $\beta = 2$),

$$\mathbb{E} \left| \sum_{j=1}^n a_{ij} X_j \right|^q \leq \left\| \sum_{j=1}^n a_{ij} X_j \right\|_2^q \leq C_{D,q} |(a_{ij})_{j=1}^n|_{\Psi_2}^q = 2^{q/q^*} C_{D,q} |(a_{ij})_{j=1}^n|_q^q.$$

Thus we obtain

$$\|Z - \mathbb{E}Z\|_p \leq C_{D,q} (p^{1/q^*} A + p^{2/q^*} B),$$

where

$$A = \left(\sum_{i,j=1}^n |a_{ij}|^q \right)^{1/q}, \quad B = \sup_{x,y \in B_{q^*}^n} \left| \sum_{i,j=1}^n a_{ij} x_i y_j \right|.$$

As a consequence, for all $t \geq 0$,

$$\mathbb{P}(|Z - \mathbb{E}Z| \geq t) \leq 2 \exp \left(-c_{D,q} \min \left(\frac{t^{q^*}}{A^{q^*}}, \frac{t^{q^*/2}}{B^{q^*/2}} \right) \right). \tag{3.16}$$

Clearly Theorem 3.12 may be applied also to quadratic forms or forms of higher order, with values in Banach spaces, but the resulting estimates will be then given in terms of expectations of suprema, which may not be so easy to estimate. Let us remark that inequalities of the form (3.16) with $q = 2$ are known as Hanson–Wright inequalities. In a slightly weaker form they were first proven in [25] for quadratic forms in independent sub-Gaussian variables.

3.4.2. Comparison principles for real-valued polynomials

We will now restrict to a special choice of the function Ψ related to the study of moments of linear combinations of i.i.d. random variables with logarithmically concave tails. We will start with a brief description of the results by Gluskin–Kwapień [22].

Theorem 3.14. *Let $\Phi: \mathbb{R}_+ \rightarrow \mathbb{R}_+ \cup \{\infty\}$ be a convex non-decreasing function, such that $\Phi(0) = 0$ and $\Phi(1) = 1$. Consider a sequence Z_1, \dots, Z_n of independent symmetric random variables satisfying $\mathbb{P}(|Z_i| \geq t) = e^{-\Phi(t)}$. Define the functions $\tilde{\Phi}: \mathbb{R}_+ \rightarrow \mathbb{R}_+ \cup \{\infty\}$,*

$$\tilde{\Phi}(x) = \begin{cases} |x|^2 & \text{if } |x| \leq 1 \\ \Phi(x) & \text{if } |x| \geq 1 \end{cases}$$

and $\Psi: \mathbb{R}^n \rightarrow \mathbb{R}_+ \cup \{\infty\}$,

$$\Psi(x) = \sum_{i=1}^n \tilde{\Phi}^*(x_i),$$

where $\tilde{\Phi}^*$ is the Legendre transform of $\tilde{\Phi}$.

Then for every sequence x_1, \dots, x_n of real numbers and every $p \geq 2$,

$$\frac{1}{C} |x|_{\Psi_p} \leq \left\| \sum_{i=1}^n x_i Z_i \right\|_p \leq C |x|_{\Psi_p}.$$

We remark that the assumption $\Phi(1) = 1$ is just a normalization condition which allows to obtain two-sided moment estimates with a universal constant C (otherwise one would have to replace C by some (explicit) constant C_Ψ).

As already mentioned in the introduction, modified log-Sobolev inequalities with the function Ψ as in Theorem 3.14 were introduced by Gentil–Guillin–Miclo [20, 21] and further studied by Barthe–Roberto [9] (when $\Phi(x)/x^2$ is non-increasing, which corresponds to super-Gaussian tail behaviour) and Gentil [19] (when $\Phi(x)/x^2$ is non-decreasing, which corresponds to sub-Gaussian tail behaviour).

In view of Theorem 3.14, our Theorem 3.5 can be given an interpretation in terms of independent random variables.

Corollary 3.15. *Under the notation of Theorem 3.14, further assume that for some $K \geq 1$ and $1 < \alpha \leq 2 \leq \beta < \infty$,*

$$K^{-1} t^{\beta^*} \leq \frac{\tilde{\Phi}(tu)}{\tilde{\Phi}(u)} \leq K t^{\alpha^*}, \tag{3.17}$$

for all $t \geq 1$ and $u > 0$. Assume that a measure μ on \mathbb{R}^n satisfies the $mLSI(\Psi, D)$. Let X be a random vector with law μ and a vector $Z = (Z_1, \dots, Z_n)$ be a sequence of i.i.d. symmetric random variables, independent of X , such that $\mathbb{P}(|Z_i| \geq t) = e^{-\Phi(t)}$ for $t \geq 0$. Then for every locally Lipschitz function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and every $p \geq 2$,

$$\|f(X) - \mathbb{E}f(X)\|_p \leq C(K, D, \alpha, \beta) \|\langle \nabla f(X), Z \rangle\|_p.$$

The interest in the above reformulation of moment inequalities stems from the fact that it can be used as a linearization tool, which allows to get estimates for functions with bounded-derivatives of higher order, in particular polynomials.

Theorem 3.16. *In the setting of Corollary 3.15, let Z^1, \dots, Z^k be independent copies of Z , independent of X . Then for every function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ of class \mathcal{C}^k and every $p \geq 2$ we have*

$$\|f(X) - \mathbb{E}f(X)\|_p \leq C_{D,k} \left(\|\langle \mathbf{D}^k f(X), Z^1 \otimes \dots \otimes Z^k \rangle\|_p + \sum_{i=1}^{k-1} \|\langle \mathbb{E}_X \mathbf{D}^i f(X), Z^1 \otimes \dots \otimes Z^i \rangle\|_p \right).$$

Note that all the terms on the right-hand side, except for the first one are moments of polynomials in independent random variables. This is also the case for the first term, provided that f itself is a polynomial of degree k . One can thus think of Theorem 3.16 as a tool which allows to transfer estimates for polynomials in independent random variables to functions with bounded derivatives of higher order of random vectors X , whose law satisfies $mLSI(\Psi, D)$. We remark that there are many results concerning polynomials in independent random variables with log-concave tails, among available results there are hypercontractive estimates, two-sided estimates in terms of expected suprema of certain empirical processes (as in Thm. 3.12) and in some cases (polynomials in Gaussian or exponential variables, polynomials in general variables with log-concave tails of degree at most 3) also precise two-sided inequalities in terms of “deterministic” quantities. We do not present the detailed discussion here, since it would require introducing rather technical notation and would anyway boil down to an application of known estimates. Instead in the example below we work out a simple application, again to a quadratic form.

Example 3.17. Let $\Psi(x) = \sum_{i=1}^n (|x_i|^2 \mathbf{1}_{\{|x_i| \leq 1\}} + |x_i|^r \mathbf{1}_{\{|x_i| > 1\}})$ for some $r \geq 2$ and assume that $X = (X_1, \dots, X_n)$ is a random vector whose law satisfies $mLSI(\Psi, D)$. For simplicity assume further that X is centered. Consider finally a quadratic form $Y = f(X)$ for $f(x) = \sum_{i,j=1}^n a_{ij} x_i x_j$, where we assume without loss of generality that $a_{ij} = a_{ji}$. Thanks to centering, we have $\mathbb{E} \nabla f(X) = 0$. Moreover $\mathbf{D}^2 f = (2a_{ij})_{i,j=1}^n$. Therefore, by Theorem 3.16, if $Z_1, Z'_1, \dots, Z_n, Z'_n$ is a sequence of i.i.d. symmetric random variables, such that $\mathbb{P}(|Z_i| \geq t) = \exp(-t^{r^*})$, we get

$$\|Y - \mathbb{E}Y\|_p \leq C_{D,r} \left\| \sum_{i,j=1}^n a_{ij} Z_i Z'_j \right\|_p$$

for $p \geq 2$.

Using results from [27], one can find a deterministic expression equivalent to the p th moment on the right-hand side above. It is expressed in terms of certain norms of the matrix $A = (a_{ij})_{i,j=1}^n$, treated as a multi-linear functional on products of certain ℓ_2 and ℓ_{r^*} spaces. More precisely,

$$\begin{aligned} \left\| \sum_{i,j=1}^n a_{ij} Z_i Z'_j \right\|_p &\simeq p^{1/2} \|A\|_{\{1,2\}|\emptyset} + p \|A\|_{\{1\}\{2\}|\emptyset} + p^{1/r^*} \|A\|_{\emptyset|\{1,2\}} \\ &\quad + p^{1/2+1/r^*} \|A\|_{\{1\}\{2\}} + p^{2/r^*} \|A\|_{\emptyset|\{1\}\{2\}}, \end{aligned}$$

where

$$\begin{aligned} \|A\|_{\{1,2\}|\emptyset} &= \sup \left\{ \sum_{i,j=1}^n a_{ij}x_{ij} : \sum_{i,j} x_{ij}^2 \leq 1 \right\} = \left(\sum_{i,j=1}^n a_{ij}^2 \right)^{1/2}, \\ \|A\|_{\{1\}\{2\}|\emptyset} &= \sup \left\{ \sum_{i,j=1}^n a_{ij}x_i y_j : \sum_{i=1}^n x_i^2 \leq 1, \sum_{j=1}^n y_j^2 \leq 1 \right\}, \\ \|A\|_{\{1\}\{2\}} &= \sup \left\{ \sum_{i,j=1}^n a_{ij}x_i y_j : \sum_{i=1}^n x_i^2 \leq 1, \sum_{j=1}^n |y_j|^{r^*} \leq 1 \right\}, \\ \|A\|_{\emptyset|\{1,2\}} &= \sup \left\{ \sum_{i,j=1}^n a_{ij}x_{ij} : \sum_{i=1}^n \left(\sum_{j=1}^n |x_{ij}|^2 \right)^{r^*/2} \leq 1 \right\} = \left(\sum_{i=1}^n \left(\sum_{j=1}^n |a_{ij}|^2 \right)^{r/2} \right)^{1/r}, \\ \|A\|_{\emptyset|\{1\}\{2\}} &= \sup \left\{ \sum_{i,j=1}^n a_{ij}x_i y_j : \sum_{i=1}^n |x_i|^{r^*} \leq 1, \sum_{j=1}^n |y_j|^{r^*} \leq 1 \right\}. \end{aligned}$$

As a consequence we obtain that for $p \geq 2$,

$$\|Y - \mathbb{E}Y\|_p \leq C_{D,r} \left(p^{1/2} \|A\|_{\{1,2\}|\emptyset} + p \|A\|_{\{1\}\{2\}|\emptyset} + p^{1/r^*} \|A\|_{\emptyset|\{1,2\}} + p^{1/2+1/r^*} \|A\|_{\{1\}\{2\}} + p^{2/r^*} \|A\|_{\emptyset|\{1\}\{2\}} \right)$$

and so for $t \geq 0$,

$$\mathbb{P}(|Y - \mathbb{E}Y| \geq t) \leq 2 \exp \left(-c_{D,r} \min \left(\frac{t^2}{\|A\|_{\{1,2\}|\emptyset}^2}, \frac{t}{\|A\|_{\{1\}\{2\}|\emptyset}}, \frac{t^{r^*}}{\|A\|_{\emptyset|\{1,2\}}^{r^*}}, \frac{t^{\frac{2r^*}{r^*+2}}}{\|A\|_{\{1\}\{2\}}^{\frac{2r^*}{r^*+2}}}, \frac{t^{r^*/2}}{\|A\|_{\emptyset|\{1\}\{2\}}^{r^*/2}} \right) \right).$$

In the class of random vectors satisfying $mLSI(\Psi, D)$ this estimate is optimal up to constants (as it can be reversed for the vector $Y = (Z_1, \dots, Z_n)$). A similar derivation may be also carried out for cubic forms as two-sided estimates of their moments are known [2], however it would involve 10 different norms of the corresponding 3-indexed matrix (under the assumption that X is isotropic). As for forms of higher order, they can also be reduced to forms in variables Z_1, \dots, Z_n , by means of Theorem 3.16. However finding two-sided estimates for moments of the latter forms remains open.

3.5. Concentration results for functions with bounded Hessian under the logarithmic Sobolev inequality

In this section we will consider the setting of the classical logarithmic Sobolev inequality and we will prove a two-level concentration estimate for functions with bounded derivatives of second order, which slightly improves on the special \mathcal{C}^2 case of Theorem 1.2. in [3] and Theorem 3.16. Our approach is inspired by a very recent development by Bobkov, Chistyakov and Götze [12] who considered second order concentration on the sphere S^{n-1} . While the authors of [12] were interested mostly in subexponential concentration, it turns out that using their approach one can also obtain two-level bounds. The goal of this section is to describe this derivation. Actually, for consistency with previous sections, we will consider a slightly more general setting and obtain inequalities in terms of moments, which allows to obtain concentration for functions with unbounded but controlled Hessian. It will be at a cost of deteriorating constants with respect to what can be obtained by working with Laplace transforms in the bounded Hessian case (as in [12]).

Recall that for a matrix $A = (A_{i,j})_{i,j=1}^n$, by $|A|_{\text{HS}}$ we denote the Hilbert–Schmidt norm of A , whereas $|A|_{\text{op}}$ stands for the operator norm of A , i.e. $|A|_{\text{HS}} = \sqrt{\sum_{i,j=1}^n a_{ij}^2}$, $|A|_{\text{op}} = \sup_{|x|,|y|\leq 1} \sum_{i,j=1}^n a_{ij}x_iy_j$.

The main result of this section is the following

Theorem 3.18. *Let μ be a probability measure on \mathbb{R}^n , such that for every $p \geq 2$ and every locally Lipschitz function $f: \mathbb{R}^n \rightarrow \mathbb{R}$,*

$$\|f - \mathbb{E}_\mu f\|_p \leq L\sqrt{p} \|\nabla f\|_2. \tag{3.18}$$

Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be a function of class \mathcal{C}^2 , such that the operator norm of $\mathbf{D}^2 f$ is uniformly bounded on \mathbb{R}^n . Then for every $t > 0$,

$$\mu(|f - \mathbb{E}_\mu f| \geq t) \leq e^2 \exp\left(-\min\left(\frac{t^2}{a^2}, \frac{t}{b}\right)\right),$$

where

$$a^2 = 4e^2 \left(\sqrt{2}L^2 \|\mathbf{D}^2 f\|_{\text{HS}} + L\mathbb{E}_\mu \|\nabla f\|_2\right)^2, \quad b = 2eL^2 \|\mathbf{D}^2 f\|_{\text{op}}.$$

In fact we shall prove a more general result, from which the above theorem easily follows.

Theorem 3.19. *Let μ be as in Theorem 3.18. Then for every $k \geq 2$, every $f: \mathbb{R}^n \rightarrow \mathbb{R}$ of class \mathcal{C}^k and for every $p \geq 2$,*

$$\begin{aligned} \|f - \mathbb{E}_\mu f\|_p &\leq L\sqrt{p}\mathbb{E}_\mu \|\nabla f\|_2 + L^2 p \|\mathbf{D}^2 f\|_{\text{op}} \\ &\leq \sqrt{p} \left(2^{(k-1)/2} L^k \|\mathbf{D}^k f\|_2 + \sum_{m=1}^{k-1} 2^{(m-1)/2} L^m \mathbb{E}_\mu \|\mathbf{D}^m f\|_2\right) + L^2 p \|\mathbf{D}^2 f\|_{\text{op}}, \end{aligned} \tag{3.19}$$

where for an m -indexed matrix $A = (a_{i_1, \dots, i_m})_{i_1, \dots, i_m=1}^n$ we denote $|A|_2 = \sqrt{\sum_{i_1, \dots, i_m=1}^n a_{i_1, \dots, i_m}^2}$.

The advantage of Theorem 3.18 over the \mathcal{C}^2 case of Theorem 1.2. in [3] stems from the fact that the tail bound obtained in [3] uses $\|\mathbf{D}^2 f\|_{\text{HS}}$ instead of $\|\mathbf{D}^2 f\|_{\text{op}}$. On the other hand it is not clear to us whether Theorem 3.19 could lead to similar improvements of the results in [3] in the case of functions with bounded derivatives of order higher than 2, since instead of the term $\|\mathbf{D}^2 f\|_{\text{op}}$ the bounds in [3] involve $\mathbb{E}_\mu \|\mathbf{D}^2 f\|_{\text{op}}$ (at the cost of introducing some additional norms of higher order derivatives). We refer the Reader to [3] for the details.

4. PROOFS

In the proofs we will drop the subscript μ and write simply \mathbb{E}, Ent for $\mathbb{E}_\mu, \text{Ent}_\mu$.

4.1. Proofs of results from Section 3.2.1

Let us first state without proof the following well-known lemma, which follows from the convexity of the function $p \mapsto \log \|X\|_{1/p}$.

Lemma 4.1. *If X is a random variable, such that for some $p > q > 0$ and $A \geq 1$, $\|X\|_p \leq A\|X\|_q$, then for all $0 < r < q$,*

$$\|X\|_p \leq A^{\frac{(p-r)q}{(p-q)r}} \|X\|_r.$$

We are also going to use the following observation on the norms $|\cdot|_{\Psi_p}$.

Lemma 4.2. *If Ψ satisfies (C) and $(G_{K,\alpha,\beta})$ then for any $x \in \mathbb{R}^n$,*

$$\Psi(x) \leq K \left(|x|_{\Psi}^{\alpha} + |x|_{\Psi}^{\beta} \right).$$

Proof. First note that (C) implies that $\Psi(x/|x|_{\Psi}) \leq 1$. If $|x|_{\Psi} \geq 1$, then

$$\Psi(x) = \Psi \left(|x|_{\Psi} \frac{x}{|x|_{\Psi}} \right) \leq K |x|_{\Psi}^{\beta},$$

and if $|x|_{\Psi} \leq 1$, then

$$1 \geq \Psi \left(\frac{x}{|x|_{\Psi}} \right) \geq K^{-1} |x|_{\Psi}^{-\alpha} \Psi(x). \quad \square$$

We are now ready to prove Theorem 3.1.

Proof of Theorem 3.1. Consider an arbitrary locally Lipschitz bounded positive function $f: \mathbb{R}^n \rightarrow \mathbb{R}$. Arguing as in the proof of Theorem 3.4. in [3], we get

$$\frac{d}{dt} (\mathbb{E} f^t)^{2/t} = \frac{2}{t^2} (\mathbb{E} f^t)^{\frac{2}{t}-1} \text{Ent} f^t.$$

Thus by $dmLSI(\Psi, D, d)$ applied to the function $f^{t/2}$ and by Lemma 4.2 applied to Ψ_p we have

$$\begin{aligned} \frac{d}{dt} (\mathbb{E} f^t)^{2/t} &\leq \frac{2D}{t^2} (\mathbb{E} f^t)^{\frac{2}{t}-1} \mathbb{E} f^t \Psi \left(\frac{t \nabla f}{2f} \right) + \frac{2d}{t^2} (\mathbb{E} f^t)^{2/t} \\ &= \frac{2Dp}{t^2} \|f\|_t^{2-t} \mathbb{E} f^t \Psi_p \left(\frac{t \nabla f}{2pf} \right) + \frac{2d}{t^2} (\mathbb{E} f^t)^{2/t} \\ &\leq \frac{2^{1-\alpha} KD}{t^{2-\alpha} p^{\alpha-1}} \|f\|_t^{2-t} \mathbb{E} f^{t-\alpha} |\nabla f|_{\Psi_p}^{\alpha} + \frac{2^{1-\beta} KD}{t^{2-\beta} p^{\beta-1}} \|f\|_t^{2-t} \mathbb{E} f^{t-\beta} |\nabla f|_{\Psi_p}^{\beta} + \frac{2d}{t^2} \|f\|_t^2. \end{aligned}$$

Further denote $M = KD$. Using Hölder's inequality with the pairs of exponents $\frac{t}{t-\alpha}, \frac{t}{\alpha}$ and $\frac{t}{t-\beta}, \frac{t}{\beta}$, for $t \in (\beta, p)$ we have

$$\begin{aligned} \frac{d}{dt} \|f\|_t^2 &\leq \frac{2^{1-\alpha} M}{t^{2-\alpha} p^{\alpha-1}} \|f\|_t^{2-t} \|f\|_t^{t-\alpha} \|\nabla f|_{\Psi_p}\|_t^{\alpha} + \frac{2^{1-\beta} M}{t^{2-\beta} p^{\beta-1}} \|f\|_t^{2-t} \|f\|_t^{t-\beta} \|\nabla f|_{\Psi_p}\|_t^{\beta} + \frac{2d}{t^2} \|f\|_t^2 \\ &\leq \frac{2^{1-\alpha} M}{t^{2-\alpha} p^{\alpha-1}} \|f\|_t^{2-\alpha} \|\nabla f|_{\Psi_p}\|_p^{\alpha} + \frac{2^{1-\beta} M}{t^{2-\beta} p^{\beta-1}} \|f\|_t^{2-\beta} \|\nabla f|_{\Psi_p}\|_p^{\beta} + \frac{2d}{t^2} \|f\|_t^2. \end{aligned}$$

For $t \in [\beta, p]$ define

$$x(t) = \|f\|_t^2 / \left\| \nabla f|_{\Psi_p} \right\|_p^2.$$

Note that by Condition (C), if the denominator above vanishes, then ∇f vanishes μ -a.s. and so we obtain

$$\frac{d}{dt} \|f\|_t^2 \leq \frac{2d}{t^2} \|f\|_t^2,$$

which *via* Gronwall's lemma gives

$$\|f\|_p^2 \leq \|f\|_{\beta}^2 \exp \left(\int_{\beta}^p \frac{2d}{t^2} dt \right) \leq e^{2d/\beta} \|f\|_{\beta}^2,$$

which is stronger than the assertion of the theorem. In what follows we will thus assume that $\|\nabla f|_{\Psi_p}\|_p \neq 0$ and so the function x is well defined.

Clearly x is non-decreasing and in the view of the above it satisfies

$$\frac{dx}{dt} \leq Ma(t)x^{1-\alpha/2} + Mb(t)x^{1-\beta/2} + dc(t)x \tag{4.1}$$

for $t \in (\beta, p)$, where

$$a(t) = 2^{1-\alpha}t^{\alpha-2}p^{1-\alpha}, \quad b(t) = 2^{1-\beta}t^{\beta-2}p^{1-\beta}, \quad c(t) = \frac{2}{t^2}.$$

Now, consider three cases:

Case 1. $x(p) \leq \frac{1}{(\alpha-1)^2}(M^{2/\alpha} \vee M^{2/\beta})$. In this case we simply have

$$\|f\|_p \leq \frac{1}{\alpha-1}(M^{1/\alpha} \vee M^{1/\beta}) \left\| |\nabla f|_{\Psi_p} \right\|_p.$$

Case 2. $x(\beta) \geq \frac{1}{(\alpha-1)^2}(M^{2/\alpha} \vee M^{2/\beta})$. It is easy to check that for $t = \beta$ we have

$$Mx(t)^{1-\alpha/2} \leq M^{1/\alpha}x(t)^{1/2}, \quad Mx(t)^{1-\beta/2} \leq M^{1/\beta}x(t)^{1/2}, \tag{4.2}$$

and since $x(t)$ is non-decreasing, we clearly have the above for all $t \in (\beta, p)$. Combining (4.1) with (4.2) yields

$$\frac{dx}{dt} \leq (M^{1/\alpha}a(t) + M^{1/\beta}b(t))x^{1/2} + dc(t)x.$$

Substituting $y = x^{1/2}$ we get

$$\frac{dy}{dt} \leq \frac{1}{2}(M^{1/\alpha}a(t) + M^{1/\beta}b(t)) + \frac{1}{2}dc(t)y,$$

from which we easily obtain

$$\begin{aligned} y(p) &\leq y(\beta)e^{\frac{d}{2} \int_{\beta}^p c} + \frac{1}{2} \left(\int_{\beta}^p (M^{1/\alpha}a(t) + M^{1/\beta}b(t))e^{-\frac{d}{2} \int_{\beta}^t c} dt \right) e^{\frac{d}{2} \int_{\beta}^p c} \\ &\leq y(\beta)e^{\frac{d}{2}} + \frac{1}{2} \left(\frac{M^{1/\alpha}}{\alpha-1} + \frac{M^{1/\beta}}{2^{\beta-1}(\beta-1)} \right) e^{\frac{d}{2}} \\ &\leq y(\beta)e^{\frac{d}{2}} + \frac{1}{\alpha-1} (M^{1/\alpha} \vee M^{1/\beta}) e^{\frac{d}{2}}. \end{aligned}$$

If $d \leq \beta$ then the above yields

$$y(p) \leq y(\beta)e^{d/\beta} + \frac{e}{\alpha-1} (M^{1/\alpha} \vee M^{1/\beta}),$$

which means

$$\|f\|_p \leq e^{d/\beta} \|f\|_{\beta} + \frac{e}{\alpha-1} (M^{1/\alpha} \vee M^{1/\beta}) \left\| |\nabla f|_{\Psi_p} \right\|_p,$$

and if $d > \beta$ then using $y(\beta) \geq \frac{1}{\alpha-1}(M^{1/\alpha} \vee M^{1/\beta})$,

$$y(p) \leq 2y(\beta)e^{d/\beta} \leq e^{2d/\beta} y(\beta)$$

hence

$$\|f\|_p \leq e^{2d/\beta} \|f\|_{\beta}.$$

Case 3. $x(t_0) = \frac{1}{(\alpha-1)^2} (M^{2/\alpha} \vee M^{2/\beta})$ for some $t_0 \in (\beta, p)$. Arguing as in Case 2, for $y = x^{1/2}$ we have

$$\begin{aligned} y(p) &\leq y(t_0)e^{\frac{d}{2} \int_{t_0}^p c} + \frac{1}{2} \left(\int_{t_0}^p (M^{1/\alpha} a(t) + M^{1/\beta} b(t)) e^{-\frac{d}{2} \int_{t_0}^t c} dt \right) e^{\frac{d}{2} \int_{t_0}^p c} \\ &\leq y(t_0)e^{d(\frac{1}{t_0} - \frac{1}{p})} + \frac{1}{\alpha-1} (M^{1/\alpha} \vee M^{1/\beta}) e^{d(\frac{1}{t_0} - \frac{1}{p})}. \end{aligned} \tag{4.3}$$

Case 3a. $t_0 \geq \frac{dp}{d+p}$. Then $\frac{1}{t_0} - \frac{1}{p} \leq \frac{1}{d}$ and using the fact that $y(t_0) = \frac{1}{\alpha-1} (M^{1/\alpha} \vee M^{1/\beta})$, the inequality (4.3) gives

$$y(p) \leq 2e \frac{1}{\alpha-1} (M^{1/\alpha} \vee M^{1/\beta}),$$

hence

$$\|f\|_p \leq 2e \frac{1}{\alpha-1} (M^{1/\alpha} \vee M^{1/\beta}) \left\| |\nabla f|_{\Psi_p} \right\|_p.$$

Case 3b. $t_0 < \frac{dp}{d+p}$. Again, using the fact that $y(t_0) = \frac{1}{\alpha-1} (M^{1/\alpha} \vee M^{1/\beta})$ the inequality (4.3) implies

$$\begin{aligned} y(p) &\leq y(t_0)e^{d(\frac{1}{t_0} - \frac{1}{p})} + \frac{1}{\alpha-1} (M^{1/\alpha} \vee M^{1/\beta}) e^{d(\frac{1}{t_0} - \frac{1}{p})} \\ &= 2y(t_0)e^{d(\frac{1}{t_0} - \frac{1}{p})} \leq y(t_0)e^{2d(\frac{1}{t_0} - \frac{1}{p})}, \end{aligned}$$

which means

$$\|f\|_p \leq A \|f\|_{t_0},$$

with $A = e^{2d(\frac{1}{t_0} - \frac{1}{p})}$. Using Lemma 4.1 with $q = t_0$ and $r = \beta$ we obtain

$$\begin{aligned} \|f\|_p &\leq A \frac{(p-\beta)t_0}{(p-t_0)^\beta} \|f\|_\beta = e^{2d \frac{p-t_0}{t_0 p} \frac{(p-\beta)t_0}{(p-t_0)^\beta}} \|f\|_\beta \\ &\leq e^{2d/\beta} \|f\|_\beta. \end{aligned}$$

This ends the proof for bounded positive functions. Let us now assume that $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a bounded locally Lipschitz function. Set $g_m = |f| + 1/m$ for $m = 1, 2, \dots$ and note that almost everywhere with respect to the Lebesgue measure, f and all the functions g_m are differentiable, moreover $\nabla g_m(x) \neq 0$ implies that $\nabla f(x) = \nabla g_m(x)$. Thus the inequality for f follows by a limiting argument. Removing the boundedness assumption is straightforward by a truncation argument. \square

Let us now pass to the sketch of the proof of Proposition 3.2.

Proof of Proposition 3.2. It is enough to follow the steps of Theorem 3.1 and to replace the splitting value $\frac{1}{(\alpha-1)^2} (M^{2/\alpha} \vee M^{2/\beta})$ for $x(t)$ with $(D^2 \vee (KD)^{2/\beta})(\log p)^2$. \square

We will now provide the proof of Proposition 3.4, which shows that the lower bound in our assumption $(G_{K,\alpha,\beta})$ cannot be avoided.

Proof of Proposition 3.4. Let us note that for $\Psi(x) = |x|$, $mLSI(\Psi, 2)$ is equivalent to that for all bounded, locally Lipschitz functions $f \geq 0$,

$$\text{Ent}_\nu f \leq \mathbb{E}_\nu |f'|. \tag{4.4}$$

For a Borel set $A \subseteq \mathbb{R}$, denote $\nu^+(A) = \liminf_{\varepsilon \rightarrow 0^+} \frac{\nu(A + (-\varepsilon, \varepsilon)) - \nu(A)}{\varepsilon}$. It is a general fact, based on the co-area formula, that the isoperimetric inequality

$$\nu(A) \log \frac{1}{\nu(A)} \vee (1 - \nu(A)) \log \frac{1}{1 - \nu(A)} \leq \nu^+(A) \tag{4.5}$$

valid for all Borel sets $A \subseteq \mathbb{R}$ implies (4.4). Indeed, on the one hand, using the variational formula for the entropy,

$$\text{Ent}_\nu f = \sup \left\{ \int fg \, d\nu : g : \mathbb{R} \rightarrow \mathbb{R} \text{ is measurable, bounded and } \int e^g \, d\nu \leq 1 \right\}, \tag{4.6}$$

and the Fubini theorem, for any measurable and bounded $g : \mathbb{R} \rightarrow \mathbb{R}$ with $\int e^g \, d\nu \leq 1$ and ρ defined as a finite (signed) measure on \mathbb{R} such that $d\rho = g d\nu$ we obtain

$$\begin{aligned} \int fg \, d\nu &= \int f \, d\rho = \int_0^\infty \rho(\{f > t\}) \, dt = \int_0^\infty \int_{\mathbb{R}} \mathbf{1}_{\{f(x) > t\}} g(x) \, \nu(dx) \, dt \\ &\leq \int_0^\infty \text{Ent}_\nu(\mathbf{1}_{\{f > t\}}) \, dt = \int_0^\infty \nu(\{f > t\}) \log \frac{1}{\nu(\{f > t\})} \, dt, \end{aligned}$$

where the inequality follows from (4.6). Hence,

$$\text{Ent}_\nu f \leq \int_0^\infty \nu(\{f > t\}) \log \frac{1}{\nu(\{f > t\})} \, dt.$$

On the other hand, by the co-area formula (see *e.g.* Thm. 8.5.1. in [7]),

$$\mathbb{E}_\nu |f'| \geq \int_0^\infty \nu^+(\{f > t\}) \, dt,$$

which combined with the previous formula shows the implication (4.5) \implies (4.4).

For the isoperimetric inequality (4.5) itself, since the density of ν w.r.t. the Lebesgue measure

$$f_\nu(x) = F'_\nu(x) = \begin{cases} \frac{1}{2} e^{-(e^{-x} - (1-x))}, & \text{for } x < 0, \\ \frac{1}{2} e^{-(e^x - (1+x))}, & \text{for } x \geq 0 \end{cases}$$

is log-concave, the result of Bobkov [11] asserts that it is enough to check (4.5) for half-lines, and in fact, by symmetry of ν , for $A = [x, \infty)$ with $x \geq 0$.

Let $t \in (0, 1/2]$. Then $F_\nu^{-1}(1-t) = \log(1 + \log \frac{1}{2t})$ and thus

$$f_\nu(F_\nu^{-1}(1-t)) = t \left(1 + \log \frac{1}{2t} \right) \geq t \log \frac{1}{t} = t \log \frac{1}{t} \vee (1-t) \log \frac{1}{1-t},$$

which proves the isoperimetric inequality (4.5).

For the moment estimate (3.5) one can repeat the argument from the proof of Theorem 3.1 to get that the function $x(t)$ as defined therein satisfies

$$\frac{dx}{dt} \leq \frac{2}{t} x^{1/2}(t),$$

for $t \in (1, p)$, *i.e.* $(x^{1/2})' \leq \frac{1}{t}$ and thus

$$x^{1/2}(p) - x^{1/2}(1) \leq \log p,$$

which implies (3.5). In order to show the “moreover” part of the proposition note that

$$\begin{aligned} \|x\|_{L^p(\nu)} &= \left(e \int_0^\infty p x^{p-1} e^{-e^x} \, dx \right)^{1/p} \geq \left(ep \int_{\frac{1}{2} \log p}^{\log p} \left(\frac{1}{2} \log p \right)^{p-1} e^{-e^{\log p}} \, dx \right)^{1/p} \\ &= \left(ep \left(\frac{1}{2} \log p \right)^p e^{-p} \right)^{1/p} \geq \frac{\log p}{2e}. \end{aligned} \quad \square$$

4.2. Proofs of results from Section 3.2.2

Let us first prove the Poincaré type inequality given in Proposition 3.6.

Proof of Proposition 3.6. The proof follows the approach by Bobkov and Zegarliński [16] who considered the function $\Psi(x) = |x|^q$ for $q \in [1, 2]$. Let Mf be a median of f under μ . We have $\|f - \mathbb{E}f\|_\beta \leq 2\|f - Mf\|_\beta$ so it is enough to prove (3.10) with the mean replaced by the median. In what follows without loss of generality we will assume that $Mf = 0$.

Note that for any bounded, locally Lipschitz function $g: \mathbb{R}^n \rightarrow (0, \infty)$, the inequality $mLSI(\Psi, D)$ applied to $g^{\beta/2}$ and Lemma 4.2 yield

$$\begin{aligned} \text{Ent}g^\beta &\leq D\mathbb{E}g^\beta\Psi\left(\frac{\beta\nabla g}{2g}\right) = D\beta\mathbb{E}g^\beta\Psi_\beta\left(\frac{\nabla g}{2g}\right) \\ &\leq KD\beta\left(2^{-\beta}\mathbb{E}|\nabla g|_{\Psi_\beta}^\beta + 2^{-\alpha}\mathbb{E}g^{\beta-\alpha}|\nabla g|_{\Psi_\beta}^\alpha\right). \end{aligned}$$

If $\alpha < \beta$ we apply the Young inequality to the last expectation to get

$$\mathbb{E}g^{\beta-\alpha}|\nabla g|_{\Psi_\beta}^\alpha \leq A^{\beta/\alpha}\frac{\alpha}{\beta}\mathbb{E}|\nabla g|_{\Psi_\beta}^\beta + A^{-\beta/(\beta-\alpha)}\frac{\beta-\alpha}{\beta}\mathbb{E}g^\beta,$$

and the choice $A = (2^{1-\alpha}KD(\beta-\alpha))^{(\beta-\alpha)/\beta}$ yields

$$\begin{aligned} \text{Ent}g^\beta &\leq KD\beta\left(2^{-\beta} + 2^{-\alpha}\frac{\alpha}{\beta}(2^{1-\alpha}KD(\beta-\alpha))^{(\beta-\alpha)/\alpha}\right)\mathbb{E}|\nabla g|_{\Psi_\beta}^\beta + \frac{1}{2}\mathbb{E}g^\beta \\ &\leq C\left((KD)^{1/\beta} + (\beta KD)^{1/\alpha}\right)^\beta\mathbb{E}|\nabla g|_{\Psi_\beta}^\beta + \frac{1}{2}\mathbb{E}g^\beta. \end{aligned} \tag{4.7}$$

Note that (4.7) is obviously valid also in the case $\alpha = \beta$. Also, this inequality can be extended to arbitrary non-negative locally Lipschitz function $g: \mathbb{R}^n \rightarrow \mathbb{R}$. Moreover, by Lemma 2.2 in [16], for any non-negative $h: \mathbb{R}^n \rightarrow \mathbb{R}$,

$$\text{Ent}h \geq \left(\log\frac{1}{\mu(\{h > 0\})}\right)\mathbb{E}h,$$

which used for $h = g^\beta$ and combined with (4.7) gives that for any non-negative locally Lipschitz function $g: \mathbb{R}^n \rightarrow \mathbb{R}$,

$$\left(\log\frac{1}{\mu(\{g > 0\})} - \frac{1}{2}\right)\mathbb{E}g^\beta \leq C\left((KD)^{1/\beta} + (\beta KD)^{1/\alpha}\right)^\beta\mathbb{E}|\nabla g|_{\Psi_\beta}^\beta. \tag{4.8}$$

Now, applying (4.8) to the functions $g = f_+$ and $g = f_-$ and using the assumption $Mf = 0$ and the implication $\nabla f_\pm \neq 0 \implies \nabla f_\pm = \pm\nabla f$, we obtain

$$\begin{aligned} (\log 2 - 1/2)\mathbb{E}f_+^\beta &\leq C\left((KD)^{1/\beta} + (\beta KD)^{1/\alpha}\right)^\beta\mathbb{E}|\nabla f_+|_{\Psi_\beta}^\beta \\ &\leq C\left((KD)^{1/\beta} + (\beta KD)^{1/\alpha}\right)^\beta\mathbb{E}|\nabla f|_{\Psi_\beta}^\beta \end{aligned}$$

and

$$\begin{aligned} (\log 2 - 1/2)\mathbb{E}f_-^\beta &\leq C\left((KD)^{1/\beta} + (\beta KD)^{1/\alpha}\right)^\beta\mathbb{E}|\nabla f_-|_{\Psi_\beta}^\beta \\ &\leq C\left((KD)^{1/\beta} + (\beta KD)^{1/\alpha}\right)^\beta\mathbb{E}|\nabla f|_{\Psi_\beta}^\beta. \end{aligned}$$

Summing the above inequalities we get

$$\frac{1}{10}\mathbb{E}|f|^\beta \leq 2C\left((KD)^{1/\beta} + (\beta KD)^{1/\alpha}\right)^\beta \mathbb{E}|\nabla f|_{\Psi_\beta}^\beta,$$

which ends the proof. □

Having proven Proposition 3.6 we can reduce Theorem 3.5 to Theorem 3.1.

Proof of Theorem 3.5. We apply Theorem 3.1 to the function $|f - \mathbb{E}f|$ to get for $p \geq \beta$,

$$\|f - \mathbb{E}f\|_p \leq \|f - \mathbb{E}f\|_\beta + \frac{2e}{\alpha - 1}\left((KD)^{1/\alpha} + (KD)^{1/\beta}\right)\left\|\nabla f\right\|_{\Psi_p}.$$

Thus by Proposition 3.6, (2.3) and Hölder’s inequality we obtain

$$\begin{aligned} \|f - \mathbb{E}f\|_p &\leq C\left((KD)^{1/\beta} + (KD\beta)^{1/\alpha}\right)\left\|\nabla f\right\|_\beta + \frac{C}{\alpha - 1}\left((KD)^{1/\alpha} + (KD)^{1/\beta}\right)\left\|\nabla f\right\|_{\Psi_p} \\ &\leq C\left(\frac{1}{\alpha - 1}(KD)^{1/\beta} + \left(\frac{1}{\alpha - 1} + \beta^{1/\alpha}\right)(KD)^{1/\alpha}\right)\left\|\nabla f\right\|_{\Psi_p}. \end{aligned}$$
□

4.3. Proofs of results from Section 3.3

Let us start with the proof of Lemma 3.8.

Proof of Lemma 3.8. The definition of ω_Ψ^* can be written equivalently as

$$\omega_\Psi^*(t) = t \sup\{u > 0: \omega_\Psi^{-1}(tu) \geq u\}.$$

Now recall that if $f: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is left-continuous, non-decreasing and satisfies $\lim_{x \rightarrow 0} f(x) = 0$, $\lim_{x \rightarrow \infty} f(x) = \infty$ and $g: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a right-continuous inverse of f , i.e.

$$g(y) = \sup\{x > 0: f(x) \leq y\},$$

then for all $x, y > 0$,

$$g(y) \geq x \iff f(x) \leq y. \tag{4.9}$$

Applying (4.9) with ω_Ψ as f we obtain that $\omega_\Psi^{-1}(tu) \geq u$ if and only if $\omega_\Psi(u) \leq tu$, which proves (3.13).

For the first inequality in (3.14), fix $t > 0$ and take $u > 0$ for which the supremum in (3.13) is attained (such u exists due to (3.12) and left-continuity of ω_Ψ). Then for all $y > u$,

$$t < \frac{\omega_\Psi(y)}{y} \leq \frac{\omega_\Psi(y)}{y - u},$$

hence for all $y > u$,

$$ty - \omega_\Psi(y) \leq tu.$$

Moreover the above inequality holds trivially for $y \in (0, u]$. Therefore $\lambda(t) \leq tu = \omega_\Psi^*(t)$.

To prove the second inequality of (3.14) fix any $u > 0$ satisfying $\omega_\Psi(u)/u \leq t$ and note that

$$tu \leq 2tu - \omega_\Psi(u) \leq \sup_{y>0}(2ty - \omega_\Psi(y)) = \lambda(2t). \tag{□}$$

Next, we will prove Corollary 3.9.

Proof of Corollary 3.9. Denote $L := CL(K, D, \alpha, \beta)$. In view of Corollary 3.7 it is enough to show that for all $t, p > 0$,

$$p \leq a\omega_{\Psi}^*(t/(Lb)) \implies \|\nabla f(x)\|_{\Psi_p} \leq t/L, \mu\text{-a.e.} \tag{4.10}$$

Then for a given $t > 0$ take $p = a\omega_{\Psi}^*(t/(Lb))$ and use Corollary 3.7 and (4.10) to obtain

$$\begin{aligned} \mu(|f - \mathbb{E}f| \geq t) &\leq \mu\left(|f - \mathbb{E}f| \geq L\left\|\nabla f\right\|_{\Psi_p}\right) \\ &\leq e^{-p}\mathbf{1}_{\{p \geq \beta\}} + \mathbf{1}_{\{p < \beta\}} \leq e^{\beta-p} = e^{\beta - a\omega_{\Psi}^*(t/(Lb))}. \end{aligned}$$

Note that the hypothesis $\|\nabla f(x)\|_{\Psi_a} \leq b, \mu\text{-a.e.}$ implies that for any $t, p > 0$,

$$\begin{aligned} \Psi_p\left(\frac{L\nabla f(x)}{t}\right) &= \frac{1}{p}\Psi\left(\frac{pL\nabla f(x)}{t}\right) \leq \frac{a}{p}\frac{1}{a}\Psi\left(\frac{a\nabla f(x)}{b}\right)\omega_{\Psi}\left(\frac{Lpb}{at}\right) \\ &\leq \frac{a}{p}\omega_{\Psi}\left(\frac{Lpb}{at}\right). \end{aligned}$$

From (3.13) in Lemma 3.8 we have

$$\omega_{\Psi}\left(\frac{Lb}{t}\frac{p}{a}\right) \leq \frac{p}{a} \iff \omega_{\Psi}^*\left(\frac{t}{Lb}\right) \geq \frac{p}{a},$$

thus (4.10) follows. □

Before we prove Corollary 3.10, let us formulate a simple lemma.

Lemma 4.3. *Under the assumptions of Corollary 3.9 we have*

$$\mu(|f - M_{\mu}f| \geq t) \leq 2 \exp\left(\beta - a\omega_{\Psi}^*\left(\frac{t}{2Lb}\right)\right),$$

where $L = CL(K, D, \alpha, \beta)$ and C is the constant from Corollary 3.9.

Proof. Since ω_{Ψ}^* is increasing and right-continuous, we can take t_0 to be the smallest positive real satisfying

$$\omega_{\Psi}^*\left(\frac{t_0}{Lb}\right) \geq \frac{\beta + \log 2}{a},$$

or equivalently, $\exp(\beta - a\omega_{\Psi}^*(t/(Lb))) \leq 1/2$. Then by Corollary 3.9, for all $\varepsilon > 0$, $\mu(|f - \mathbb{E}f| \geq t_0 + \varepsilon) < 1/2$ and thus $|Mf - \mathbb{E}f| \leq t_0$. Therefore, using Corollary 3.9 for $t \geq 2t_0$,

$$\mu(|f - Mf| \geq t) \leq \mu\left(|f - \mathbb{E}f| \geq \frac{t}{2}\right) \leq \exp\left(\beta - a\omega_{\Psi}^*\left(\frac{t}{2Lb}\right)\right).$$

On the other hand if $t < 2t_0$ then by the definition of t_0 ,

$$2 \exp\left(\beta - a\omega_{\Psi}^*\left(\frac{t}{2Lb}\right)\right) > 1,$$

so the inequality of the lemma holds trivially. □

Proof of Corollary 3.10. We recall that if Φ is a Young function on \mathbb{R}^n , Φ^* is the Legendre transform of Φ and $|\cdot|_{\Phi}^*$ denotes the norm on \mathbb{R}^n , dual to $|\cdot|_{\Phi}$, then for all $x \in \mathbb{R}^n$,

$$|x|_{\Phi^*} \leq |x|_{\Phi}^* \leq 2|x|_{\Phi^*}. \tag{4.11}$$

Using the lower bound from $(G_{K,\alpha,\beta})$ together with condition (C), one can show that $\Psi^*(x) = 0$ iff $x = 0$ and $\Psi^*(x) < \infty$ at all x . Further, notice that $\{x: \Psi^*(x) < u\} = B(|\cdot|_{u^{-1}\Psi^*}, 1) := \{x: |x|_{u^{-1}\Psi^*} < 1\}$. Define the function

$$f(x) = \inf_{y \in A} |x - y|_{u^{-1}\Psi^*}.$$

The function f is 1-Lipschitz with respect to the norm $|\cdot|_{u^{-1}\Psi^*}$, which implies that ∇f exists almost everywhere and $|\nabla f|_{u^{-1}\Psi^*}^* \leq 1$. Since $u^{-1}\Psi^* = (\Psi_u)^*$, (4.11) implies that $|\nabla f|_{\Psi_u} \leq 1$.

Therefore, by Lemma 4.3, applied with $a = u$, $b = 1$, $t = 1$,

$$\mu(f \geq Mf + 1) \leq 2 \exp(\beta - u\omega_{\Psi}^*(1/L)),$$

where $L = CL(K, D, \alpha, \beta)$. By the assumption $\mu(A) \geq 1/2$, we have $Mf = 0$ and so

$$A + \{\Psi^*(x) < u\} = \{f < 1\} = \{f < Mf + 1\},$$

therefore the above inequality yields

$$\mu(A + \{\Psi^*(x) < u\}) \geq 1 - 2 \exp(\beta - u\omega_{\Psi}^*(1/L)).$$

It remains to bound $\omega_{\Psi}^*(1/L)$ from below. Note that by the convexity of Ψ and (3.11), the function ω_{Ψ} is convex and everywhere finite, hence continuous. Thus by Lemma 3.8,

$$\omega_{\Psi}^*(1/L) = w/L, \tag{4.12}$$

where w is such that $\omega_{\Psi}(w)/w = 1/L$. On the other hand, the upper bound in (3.11) yields

$$\frac{1}{L} = \frac{\omega_{\Psi}(w)}{w} \leq K(w^{\alpha-1} \vee w^{\beta-1}),$$

hence

$$w \geq \frac{1}{(KL)^{1/(\alpha-1)}} \wedge \frac{1}{(KL)^{1/(\beta-1)}} \geq \frac{1}{(KL)^{1/(\alpha-1)}} \wedge 1.$$

Combining the above lower bound on w with (4.12) we obtain

$$\omega_{\Psi}^*(1/L) \geq (K^{-1/(\alpha-1)}L^{-\alpha/(\alpha-1)}) \wedge L^{-1}. \quad \square$$

4.4. Proofs of results from Section 3.4

Let us start with the proofs of results concerning Banach space valued homogeneous polynomials.

Proof of Theorem 3.12. We will proceed by induction on k . Note that the norm $|\cdot|_E$ can be expressed as a supremum of countably many functionals, therefore it is enough to prove the theorem for $(E, |\cdot|) = (\ell_{\infty}^N, |\cdot|_{\infty})$ with arbitrarily large, finite N . We will denote the index related to the ℓ_{∞}^N structure in the superscript, i.e. $a_{i_1, \dots, i_k} = (a_{i_1, \dots, i_k}^1, \dots, a_{i_1, \dots, i_k}^N)$ and $A^r = (a_{i_1, \dots, i_k}^r)_{i_1, \dots, i_k=1}^n$.

For $k = 1$, the inequality in question is then just a dual formulation of Theorem 3.5 for $f(x) = \max_{r \leq N} |\sum_{i=1}^n a_i^r x_i|$. Indeed, it is enough to note that $A_{\Psi, p}$ is the unit ball for the Orlicz norm corresponding to the function $\frac{1}{p}\Psi^*$, which up to universal multiplicative constants is equivalent to the dual norm for $|\cdot|_{\Psi_p}$.

Let us thus assume that the theorem is true for all homogeneous forms of degree strictly smaller than k . We can assume that the real-valued matrices $\pm(a_{i_1, \dots, i_k}^r)$ are pairwise distinct and nonzero for $1 \leq r \leq N$. Since the set of zeros of a non-trivial polynomial is Lebesgue null, there exist open sets B_r^ε , $r = 1, \dots, N$, $\varepsilon = \pm 1$, such that $\mathbb{R}^n \setminus (\bigcup_{r=1}^N \bigcup_{\varepsilon=\pm 1} B_r^\varepsilon)$ is Lebesgue null and on B_r^ε ,

$$|\langle A, x^{\otimes k} \rangle|_E = \varepsilon \langle A^r, x^{\otimes k} \rangle.$$

Denoting $[k] = \{1, \dots, k\}$, we get that almost everywhere on B_r^ε , the function $f(x) = |\langle A, x^{\otimes k} \rangle|_E$ satisfies

$$\begin{aligned} \nabla f(x) &= \varepsilon \left(\sum_{i_1, \dots, i_k=1}^n \left(a_{i_1, \dots, i_k}^r \sum_{s=1}^k \mathbf{1}_{\{i_s=j\}} \prod_{\substack{1 \leq u \leq k \\ u \neq s}} x_{i_u} \right) \right)_{j=1}^n \\ &= \varepsilon \sum_{s=1}^k \left(\sum_{(i_m)_{m \in [k] \setminus \{s\}} \in [n]^{k-1}} \left(a_{i_1, \dots, i_{s-1}, j, i_{s+1}, \dots, i_k}^r \prod_{\substack{1 \leq u \leq k \\ u \neq s}} x_{i_u} \right) \right)_{j=1}^n \\ &= k\varepsilon \left(\sum_{i_2, \dots, i_k=1}^n a_{j, i_2, \dots, i_k}^r x_{i_2} \cdots x_{i_k} \right)_{j=1}^n, \end{aligned}$$

where in the last equality we used the symmetry of the coefficients a_{i_1, \dots, i_k}^r . Denoting $B_r = B_r^1 \cup B_r^{-1}$ we see that on B_r , we have

$$\begin{aligned} |\nabla f(x)|_{\Psi_p} &\leq Ck \sup_{y \in A_{\Psi, p}} \left| \sum_{i_1, \dots, i_k=1}^n a_{i_1, \dots, i_k}^r y_{i_1} x_{i_2} \cdots x_{i_k} \right| \\ &= Ck \sup_{y \in A_{\Psi, p}} |\langle A_r, y \otimes x^{\otimes(k-1)} \rangle|. \end{aligned}$$

Thus almost surely

$$|\nabla f(X)|_{\Psi_p} \leq Ck \max_{1 \leq r \leq N} \sup_{y \in A_{\Psi, p}} |\langle A^r, y \otimes X^{\otimes(k-1)} \rangle|.$$

The right-hand side above is a supremum of homogeneous forms of degree $k - 1$ in X , moreover by separability of $A_{\Psi, p}$ it can be clearly approximated by suprema of a finite number of such forms. Therefore, by the induction assumption,

$$\begin{aligned} \left\| |\nabla f(X)|_{\Psi_p} \right\|_p &\leq Ck \mathbb{E} \max_{1 \leq r \leq N} \sup_{y \in A_{\Psi, p}} |\langle A^r, y \otimes X^{\otimes(k-1)} \rangle| \\ &\quad + C_{D, K, \alpha, \beta, k-1} \sum_{j=2}^k \mathbb{E} \max_{1 \leq r \leq N} \sup_{y \in A_{\Psi, p}} \sup_{y^2, \dots, y^j \in A_{\Psi, p}} |\langle A^r, y \otimes y^2 \otimes \cdots \otimes y^j \otimes X^{\otimes(k-j)} \rangle| \\ &= C_{D, K, \alpha, \beta, k} \sum_{j=1}^k \mathbb{E} \sup_{y^1, y^2, \dots, y^j \in A_{\Psi, p}} \max_{1 \leq r \leq N} |\langle A^r, y^1 \otimes y^2 \otimes \cdots \otimes y^j \otimes X^{\otimes(k-j)} \rangle|. \end{aligned}$$

This ends the proof of (3.15), since by Theorem 3.5,

$$\left\| Z - \mathbb{E}Z \right\|_p \leq C_{D, K, \alpha, \beta} \left\| |\nabla f(X)|_{\Psi_p} \right\|_p.$$

The second estimate of Theorem 3.12 follows now by the Chebyshev inequality. □

We will now pass to the proof of Theorem 3.16. Let us start with the main tool, which is Corollary 3.15.

Proof of Corollary 3.15. Relating the Legendre transform of $\tilde{\Phi}$ to the conjugation in a sense of Lemma 3.8, one can deduce that the growth condition (3.17) on $\tilde{\Phi}$ implies that $\tilde{\Phi}^*$ satisfies

$$K'^{-1}t^\alpha \leq \frac{\tilde{\Phi}^*(tu)}{\tilde{\Phi}^*(u)} \leq K't^\beta$$

for all $t \geq 1$ and $u > 0$, where $K' = C(K, \alpha, \beta)$. As a consequence, Ψ satisfies $(G_{K', \alpha, \beta})$ and therefore, for $p \geq \beta$ the corollary is a direct consequence of Theorem 3.5 and Theorem 3.14.

For $2 \leq p < \beta$ we use the fact that if μ satisfies $mLSI(\Psi, D)$, then it also satisfies the Poincaré inequality

$$\text{Var}(f) \leq C_D \mathbb{E}|\nabla f|^2$$

(see [21]), which as is well known (see e.g. [16] or [32]) implies that

$$\|f - \mathbb{E}f\|_p \leq C'_D p \left\| |\nabla f| \right\|_p.$$

Moreover, due to the normalization $\Phi(1) = 1$ one can easily get $|x| \leq C|x|_{\Psi_p}$ for $p \geq 1$, which allows to deduce the corollary for $2 \leq p < \beta$. □

Proof of Theorem 3.16. Given Corollary 3.15, the proof follows with just formal changes the proof of Proposition 3.2. in [3]. □

4.5. Proofs of results from Section 3.5

In this section we will present the proof of Theorem 3.19. Theorem 3.18 will then follow by specializing to $k = 2$ and using Chebyshev's inequality optimized in p .

Let us start with the following simple lemma.

Lemma 4.4. *In the setting of Theorem 3.18, for every $m < k$,*

$$\left\| |\mathbf{D}^m f - \mathbb{E} \mathbf{D}^m f|_2 \right\|_2 \leq \sqrt{2}L \left\| |\mathbf{D}^{m+1} f|_2 \right\|_2$$

Proof. We will regard $\mathbf{D}^m f$ as a vector in $(\mathbb{R}^n)^{\otimes m} \simeq \mathbb{R}^{n^m}$. Let X be a random vector distributed according to μ and G a standard Gaussian vector in \mathbb{R}^{n^m} , independent of X . Then

$$\begin{aligned} \left\| |\mathbf{D}^m f - \mathbb{E} \mathbf{D}^m f|_2 \right\|_2^2 &= \mathbb{E}_X |\mathbf{D}^m f(X) - \mathbb{E}_X \mathbf{D}^m f(X)|_2^2 = \mathbb{E}_G \mathbb{E}_X \langle \mathbf{D}^m f(X) - \mathbb{E}_X \mathbf{D}^m f(X), G \rangle^2 \\ &\leq 2L^2 \mathbb{E}_G \mathbb{E}_X |\mathbf{D} \langle \mathbf{D}^m f(X), G \rangle|_2^2, \end{aligned}$$

where the second equality follows from the Fubini theorem and the inequality from the assumption (3.18), applied conditionally on G to the function $x \mapsto \langle \mathbf{D}^m f(x), G \rangle$. Now, it is easy to see that

$$\mathbb{E}_G |\mathbf{D} \langle \mathbf{D}^m f(X), G \rangle|_2^2 = |\mathbf{D}^{m+1} f(X)|_2^2,$$

which ends the proof. □

Corollary 4.5. *In the setting of Theorem 3.18, for all $k \geq 2$,*

$$\left\| |\nabla f|_2 \right\|_2 \leq (\sqrt{2}L)^{k-1} \left\| |\mathbf{D}^k f|_2 \right\|_2 + \sum_{m=1}^{k-1} (\sqrt{2}L)^{m-1} \mathbb{E}_\mu |\mathbf{D}^m f|_2.$$

Proof. An induction on k , using Lemma 4.4. □

Proof of Theorem 3.19. By (3.18), we have

$$\|f - \mathbb{E}_\mu f\|_p \leq L\sqrt{p} \left\| |\nabla f|_2 \right\|_p. \quad (4.13)$$

It is easy to prove that $|\nabla f|_2$ is locally Lipschitz and $|\nabla|\nabla f|_2|_2 \leq |\mathbf{D}^2 f|_{\text{op}}$ μ -a.s. Indeed, we have for $x \in \mathbb{R}^n$ and $|h| \rightarrow 0$, by the triangle inequality and Taylor's expansion,

$$\begin{aligned} \frac{\left| |\nabla f(x+h)|_2 - |\nabla f(x)|_2 \right|}{|h|} &\leq \frac{\left| \left(\sum_{j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(x) h_j + o(|h|) \right)_{i=1}^n \right|_2}{|h|} \\ &\leq \frac{|\mathbf{D}^2 f(x)h|_2}{|h|} + o(1) \leq |\mathbf{D}^2 f(x)|_{\text{op}} + o(1), \end{aligned}$$

which *via* standard compactness arguments yields that $|\nabla f|_2$ is locally Lipschitz and that if $\nabla|\nabla f(x)|_2$ exists (which happens μ -a.s.), then its Euclidean norm does not exceed $|\mathbf{D}^2 f(x)|_{\text{op}}$.

Thus, another application of (3.18) gives

$$\left\| |\nabla f|_2 \right\|_p \leq \mathbb{E}_\mu |\nabla f|_2 + \left\| |\nabla f|_2 - \mathbb{E}_\mu |\nabla f|_2 \right\|_p \leq \mathbb{E}_\mu |\nabla f|_2 + L\sqrt{p} \left\| |\mathbf{D}^2 f|_{\text{op}} \right\|_p,$$

which together with (4.13) gives the first inequality of (3.19). The second inequality follows now by Corollary 4.5. □

REFERENCES

- [1] R. Adamczak, Logarithmic Sobolev inequalities and concentration of measure for convex functions and polynomial chaoses. *Bull. Pol. Acad. Sci. Math.* **53** (2005) 221–238.
- [2] R. Adamczak and R. Latała, Tail and moment estimates for chaoses generated by symmetric random variables with logarithmically concave tails. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** (2012) 1103–1136.
- [3] R. Adamczak and P. Wolff, Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. *Probab. Theory Related Fields* **162** (2015) 531–586.
- [4] S. Aida and D. Stroock, Moment estimates derived from Poincaré and logarithmic Sobolev inequalities. *Math. Res. Lett.* **1** (1994) 75–86.
- [5] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto and G. Scheffer, Sur les inégalités de Sobolev logarithmiques. Vol. 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris (2000).
- [6] M.A. Arcones and E. Giné, On decoupling, series expansions and tail behavior of chaos processes. *J. Theor. Probab.* **6** (1993) 101–122.
- [7] D. Bakry, I. Gentil and M. Ledoux, Analysis and geometry of Markov diffusion operators. Springer, Cham (2014).
- [8] F. Barthe and A.V. Kolesnikov, Mass transport and variants of the logarithmic Sobolev inequality. *J. Geom. Anal.* **18** (2008) 921–979.
- [9] F. Barthe and C. Roberto, Modified logarithmic Sobolev inequalities on \mathbb{R} . *Potential Anal.* **29** (2008) 167–193.
- [10] W. Beckner, A generalized Poincaré inequality for Gaussian measures. *Proc. Amer. Math. Soc.* **105** (1989) 397–400.
- [11] S. Bobkov, Extremal properties of half-spaces for log-concave distributions. *Ann. Probab.* **24** (1996) 35–48.
- [12] S. Bobkov, G. Chistyakov and F. Götze, Second order concentration on the sphere. *Commun. Contemp. Math.* **19** (2017) 1650058.
- [13] S. Bobkov and M. Ledoux, Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields* **107** (1997) 383–400.
- [14] S.G. Bobkov and F. Götze, Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** (1999) 1–28.
- [15] S.G. Bobkov and M. Ledoux, From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.* **10** (2000) 1028–1052.
- [16] S.G. Bobkov and B. Zegarlinski, Entropy bounds and isoperimetry. *Mem. Amer. Math. Soc.* **176** (2005) (829) .
- [17] C. Borell, On the Taylor series of a Wiener polynomial. Seminar Notes on multiple stochastic integration, polynomial chaos and their integration. Case Western Reserve Univ., Cleveland (1984).
- [18] P. Federbush, Partially alternate derivation of a result of Nelson. *J. Math. Phys.* **10** (1969) 5052.

- [19] I. Gentil, From the Prékopa-Leindler inequality to modified logarithmic Sobolev inequality. *Ann. Fac. Sci. Toulouse Math.* **17** (2008) 291–308.
- [20] I. Gentil, A. Guillin and L. Miclo, Modified logarithmic Sobolev inequalities and transportation inequalities. *Probab. Theory Related Fields* **133** (2005) 409–436.
- [21] I. Gentil, A. Guillin and L. Miclo, Modified logarithmic Sobolev inequalities in null curvature. *Rev. Mat. Iberoam.* **23** (2007) 235–258.
- [22] E.D. Gluskin and S. Kwapien, Tail and moment estimates for sums of independent random variables with logarithmically concave tails. *Studia Math.* **114** (1995) 303–309.
- [23] N. Gozlan, Poincaré inequalities and dimension free concentration of measure. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** (2010) 708–739.
- [24] L. Gross, Logarithmic Sobolev inequalities. *Amer. J. Math.* **97** (1975) 1061–1083.
- [25] D.L. Hanson and F.T. Wright, A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Statist.* **42** (1971) 1079–1083.
- [26] R. Latała, Tail and moment estimates for sums of independent random vectors with logarithmically concave tails. *Studia Math.* **118** (1996) 301–304.
- [27] R. Latała, Tail and moment estimates for some types of chaos. *Studia Math.* **135** (1999) 39–53.
- [28] R. Latała, Estimates of moments and tails of Gaussian chaoses. *Ann. Probab.* **34** (2006) 2315–2331.
- [29] R. Latała and K. Oleszkiewicz, Between Sobolev and Poincaré. In Geometric aspects of functional analysis. Vol. 1745 of *Lect. Notes Math.* Springer, Berlin (2000) 147–168
- [30] M. Ledoux, The concentration of measure phenomenon. Vol. 89 of *Mathematical Surveys and Monographs.* Amer. Math. Society, Providence, RI (2001).
- [31] R. Łochowski, Moment and tail estimates for multidimensional chaoses generated by symmetric random variables with logarithmically concave tails. In Approximation and probability. Vol. 72 of Banach Center Publ. Polish Acad. Sci., Warsaw (2006) 161–176.
- [32] E. Milman, On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.* **177** (2009) 1–43.
- [33] G. Pisier, Probabilistic methods in the geometry of Banach spaces. In Probability and analysis (Varenna, 1985). Vol. 1206 of *Lect. Notes Math.* Springer, Berlin (1986) 167–241.
- [34] O.S. Rothaus, Analytic inequalities, isoperimetric inequalities and logarithmic Sobolev inequalities. *J. Funct. Anal.* **64** (1985) 296–313.
- [35] O.S. Rothaus, Logarithmic Sobolev inequalities and the growth of L^p norms. *Proc. Amer. Math. Soc.* **126** (1998) 2309–2314.
- [36] J. Shao, Modified logarithmic Sobolev inequalities and transportation cost inequalities in \mathbb{R}^n . *Potential Anal.* **31** (2009) 183–202.
- [37] A.J. Stam, Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inform. Control* **2** (1959) 101–112.
- [38] M. Talagrand, A new isoperimetric inequality and the concentration of measure phenomenon. In Geometric aspects of functional analysis (1989–90). Vol. 1469 of *Lect. Notes Math.* Springer, Berlin (1991) 94–124.
- [39] M. Talagrand, The supremum of some canonical processes. *Amer. J. Math.* **116** (1994) 283–325.