

## Modified log-Sobolev inequalities for convex functions on the real line. Sufficient conditions

by

RADOSŁAW ADAMCZAK and MICHAŁ STRZELECKI (Warszawa)

**Abstract.** We provide a mild sufficient condition for a probability measure on the real line to satisfy a modified log-Sobolev inequality for convex functions, interpolating between the classical log-Sobolev inequality and a Bobkov–Ledoux type inequality. As a consequence we obtain dimension-free two-level concentration results for convex functions of independent random variables with sufficiently regular tail decay.

We also provide a link between modified log-Sobolev inequalities for convex functions and weak transport-entropy inequalities, complementing recent work by Gozlan, Roberto, Samson, and Tetali.

### 1. Introduction and main results

**1.1. Background.** Over the past few decades the concentration of measure phenomenon has become one of the main themes of high dimensional probability and geometric analysis, with applications to limit theorems, non-asymptotic confidence bounds or random constructions of geometric objects with extremal properties. While initial results on concentration of measure concerned mostly deviation bounds for Lipschitz functions of highly regular random variables, the seminal work by Talagrand [22] has revealed that if one restricts attention to convex Lipschitz functions, dimension-free concentration of measure holds under much weaker conditions, e.g. for all product probability measures with bounded support. Talagrand’s approach relied on his celebrated convex distance inequality related to the analysis of isoperimetric problems for product measures. In subsequent papers other authors adapted tools from the classical concentration of measure theory, such as Poincaré and log-Sobolev inequalities or transportation and infimum-

---

2010 *Mathematics Subject Classification*: Primary 60E15; Secondary 26A51, 26B25, 26D10.

*Key words and phrases*: concentration of measure, convex functions, logarithmic Sobolev inequalities.

Received 3 June 2015; revised 15 December 2015.

Published online 21 January 2016.

convolution inequalities [15, 4, 18, 17, 19, 20, 1] to prove tail inequalities for convex functions. More recent work [11, 13] presents deep connections between various approaches.

Talagrand's inequality provides a subgaussian tail estimate for convex Lipschitz functions of independent uniformly bounded random variables (with dimension-independent constants). It is natural to ask whether such a result can hold under weaker assumptions on the tails of the underlying random variables. One can easily see that it is not enough to assume uniform subgaussian tail decay, and some stronger regularity is necessary [14, 1]. In [4] Bobkov and Götze show that for a probability measure on the real line dimension-free concentration for convex functions is equivalent to the convex Poincaré inequality and further to a certain regularity of the tail (see also [11] for a refinement of this result in the setting of abstract metric spaces). Moreover the concentration rate in this case is at least exponential. In [1] a related assumption on the tail is shown to imply the convex log-Sobolev inequality which allows us to obtain dimension-free deviation bounds on the upper tail of convex functions.

The goal of this article is to show that tail conditions interpolating between those considered in [4] and [1] lead to modified log-Sobolev inequalities for convex functions (convex counterparts of inequalities introduced by Bobkov–Ledoux [5] and Gentil–Guillin–Miclo [8, 9]) and thus also two-level deviation bounds interpolating between the subexponential and subgaussian tail decay, in the spirit of Talagrand's inequalities for the exponential or Weibull distribution [21, 18]. Furthermore we demonstrate a connection of modified log-Sobolev inequalities with weak transportation inequalities. In the case of the classical log-Sobolev inequality such a link has been recently discovered by Gozlan, Roberto, Samson, and Tetali [13].

**1.2. Main results.** To formulate our results we need to introduce the following

DEFINITION 1.1. For  $\beta \in [0, 1]$ ,  $m > 0$  and  $\sigma \geq 0$  let  $\mathcal{M}_\beta(m, \sigma^{\beta+1})$  denote the class of probability distributions  $\mu$  on  $\mathbb{R}$  for which

$$\begin{aligned} \forall_{x \geq m} \quad \nu^+([x, \infty)) &\leq \sigma^{\beta+1} \mu([x, \infty)), \\ \forall_{x \leq -m} \quad \nu^-((-\infty, x]) &\leq \sigma^{\beta+1} \mu((-\infty, x]), \end{aligned}$$

where  $\nu^+$  is the measure on  $[m, \infty)$  with density  $x^\beta \mu([x, \infty))$  and  $\nu^-$  is the measure on  $(-\infty, -m]$  with density  $|x|^\beta \mu((-\infty, x])$ .

As one can easily see, the condition defining the class  $\mathcal{M}_\beta(m, \sigma^{\beta+1})$  is equivalent to some type of regular tail decay. More precisely, we have the following proposition, the proof of which is deferred to Section 2.

PROPOSITION 1.2. *Let  $\mu$  be a probability measure on  $\mathbb{R}$ . For  $\beta \in [0, 1]$  and  $m > 0$  the following conditions are equivalent:*

- (i) *There exists  $\sigma \geq 0$  such that  $\mu \in \mathcal{M}_\beta(m, \sigma^{\beta+1})$ .*
- (ii) *For some  $h > 0$ ,  $\alpha < 1$  and all  $x > m$ ,*

$$\begin{aligned}\mu([x + h/x^\beta, \infty)) &\leq \alpha\mu([x, \infty)), \\ \mu((-\infty, -x - h/x^\beta]) &\leq \alpha\mu((-\infty, -x]).\end{aligned}$$

Moreover, the constants in (i) depend only on the constants in (ii) and vice versa. For instance, if (i) holds, we can take  $h = 2\sigma^{\beta+1}$  and  $\alpha = 1/2$ .

Let us recall that a probability measure  $\mu$  on  $\mathbb{R}^n$  satisfies the Poincaré inequality with constant  $C$  for the class  $\mathcal{A}$  of real valued functions on  $\mathbb{R}^n$  if for all  $f \in \mathcal{A}$ ,

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

where  $X$  is a random vector with law  $\mu$ . The measure  $\mu$  satisfies the log-Sobolev inequality with constant  $C$  for the class  $\mathcal{A}$  if for all  $f \in \mathcal{A}$ ,

$$(1.1) \quad \text{Ent } e^{f(X)} \leq C\mathbb{E}|\nabla f(X)|^2 e^{f(X)},$$

where  $\text{Ent } Y$  denotes the usual entropy of a nonnegative random variable  $Y$ , i.e.

$$\text{Ent } Y = \mathbb{E}Y \log Y - \mathbb{E}Y \log \mathbb{E}Y$$

if  $\mathbb{E}Y \log Y < \infty$  and  $\text{Ent } Y = \infty$  otherwise.

We remark that if  $\mathcal{A}$  consists of all smooth functions, one usually states the log-Sobolev inequality in the equivalent form

$$(1.2) \quad \text{Ent } f(X)^2 \leq D\mathbb{E}|\nabla f(X)|^2,$$

where  $D = 4C$ . We have decided to use the form (1.1), since our main motivation is the study of concentration properties for convex functions (note that for  $\mathcal{A}$  being the class of convex functions, (1.1) and (1.2) are no longer equivalent).

Using Proposition 1.2 and the results from [4] one easily finds that the condition  $\mu \in \mathcal{M}_0(m, \sigma)$  for some  $m, \sigma$  is equivalent to the Poincaré inequality for convex functions and further to dimension-free subexponential concentration for convex Lipschitz functions on  $\mathbb{R}^n$  equipped with the measure  $\mu^{\otimes n}$ . On the other hand, as shown in [1] the condition  $\mathcal{M}_1(m, \sigma^2)$  implies the log-Sobolev inequality for convex functions and a dimension-free subgaussian inequality for upper tails of convex Lipschitz functions on  $(\mathbb{R}^n, \mu^{\otimes n})$ —see Corollary 1.7 below.

In what follows,  $\|\cdot\|_p$  denotes the  $\ell_p^n$  norm, i.e. for  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ ,  $\|x\|_p = (|x_1|^p + \dots + |x_n|^p)^{1/p}$ . For  $p = 2$  we will write simply  $|x|$  instead of  $\|x\|_p$ . The symbol  $\mathcal{L}(X)$  stands for the law of a random variable  $X$ .

Our main result is the following theorem.

**THEOREM 1.3.** *Let  $X_1, \dots, X_n$  be independent random variables such that  $\mathcal{L}(X_i) \in \mathcal{M}_\beta(m, \sigma^{\beta+1})$  and let  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$  be a smooth convex function. Then, for  $\beta \in (0, 1]$ ,*

$$(1.3) \quad \text{Ent } e^{\varphi(X_1, \dots, X_n)} \leq C(\beta, m, \sigma) \mathbb{E}(|\nabla\varphi(X_1, \dots, X_n)|^2 \vee \|\nabla\varphi(X_1, \dots, X_n)\|_{(\beta+1)/\beta}^{(\beta+1)/\beta}) e^{\varphi(X_1, \dots, X_n)}.$$

*If  $\beta = 0$  and  $|\partial_i\varphi(x)| \leq 1/(2(m + 3\sigma))$  for  $i \in \{1, \dots, n\}$  and all  $x \in \mathbb{R}^n$ , then*

$$(1.4) \quad \text{Ent } e^{\varphi(X_1, \dots, X_n)} \leq C(\beta = 0, m, \sigma) \mathbb{E}|\nabla\varphi(X_1, \dots, X_n)|^2 e^{\varphi(X_1, \dots, X_n)}.$$

**REMARK 1.4.** Note that in the case  $\beta = 1$ , (1.3) reduces to the classical log-Sobolev inequality

$$\text{Ent } e^{\varphi(X_1, \dots, X_n)} \leq C(m, \sigma) \mathbb{E}|\nabla\varphi(X_1, \dots, X_n)|^2 e^{\varphi(X_1, \dots, X_n)}.$$

The case  $\beta \in (0, 1)$  corresponds to the modified log-Sobolev inequality introduced and studied by Gentil, Guillin and Miclo [8] for the class of smooth (not necessarily convex) functions. The inequality (1.4) for smooth functions was introduced and studied by Bobkov and Ledoux [5].

**REMARK 1.5.** In the above theorem and the following corollaries the assumption about convexity of  $\varphi$  can be relaxed to separate convexity (cf. [15]).

By a version of Herbst's argument we obtain the following corollary, concerning the tail behaviour of smooth convex functions. Below  $C$  denotes the constant  $C(\beta, m, \sigma)$  from Theorem 1.3.

**COROLLARY 1.6.** *Let  $X_1, \dots, X_n$  be independent random variables such that  $\mathcal{L}(X_i) \in \mathcal{M}_\beta(m, \sigma^{\beta+1})$  and let  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$  be a smooth convex Lipschitz function. Then, for  $\beta \in (0, 1]$  and all  $t \geq 0$ ,*

$$\begin{aligned} \mathbb{P}(\varphi(X_1, \dots, X_n) \geq \mathbb{E}\varphi(X_1, \dots, X_n) + t) \\ \leq \exp\left(-\frac{3}{16} \min\left\{\frac{t^2}{C \sup |\nabla\varphi|^2}, \frac{t^{1+\beta}}{C^\beta \sup \|\nabla\varphi\|_{(\beta+1)/\beta}^{1+\beta}}\right\}\right). \end{aligned}$$

*If  $\beta = 0$ , then for all  $t \geq 0$ ,*

$$\begin{aligned} \mathbb{P}(\varphi(X_1, \dots, X_n) \geq \mathbb{E}\varphi(X_1, \dots, X_n) + t) \\ \leq \exp\left(-\frac{1}{4} \min\left\{\frac{t^2}{C \sup |\nabla\varphi|^2}, \frac{t}{(m + 3\sigma) \sup \|\nabla\varphi\|_\infty}\right\}\right). \end{aligned}$$

Using standard smoothing arguments one can also obtain a result for not necessarily smooth functions, expressed in terms of their Lipschitz constants with respect to the  $\ell_2^n$  and  $\ell_{1+\beta}^n$  norms.

COROLLARY 1.7. *Let  $X_1, \dots, X_n$  be independent random variables such that  $\mathcal{L}(X_i) \in \mathcal{M}_\beta(m, \sigma^{\beta+1})$  and let  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$  be a convex function such that*

$$|\varphi(x) - \varphi(y)| \leq L_2 \|x - y\|_2$$

and

$$|\varphi(x) - \varphi(y)| \leq L_{1+\beta} \|x - y\|_{1+\beta}$$

for all  $x, y \in \mathbb{R}^n$  and some  $L_2, L_{1+\beta} < \infty$ . Then, for  $\beta \in (0, 1]$  and all  $t \geq 0$ ,

$$\begin{aligned} \mathbb{P}(\varphi(X_1, \dots, X_n) \geq \mathbb{E}\varphi(X_1, \dots, X_n) + t) \\ \leq \exp\left(-\frac{3}{16} \min\left\{\frac{t^2}{CL_2^2}, \frac{t^{1+\beta}}{C^\beta L_{1+\beta}^{1+\beta}}\right\}\right). \end{aligned}$$

If  $\beta = 0$ , then for all  $t \geq 0$ ,

$$\begin{aligned} \mathbb{P}(\varphi(X_1, \dots, X_n) \geq \mathbb{E}\varphi(X_1, \dots, X_n) + t) \\ \leq \exp\left(-\frac{1}{4} \min\left\{\frac{t^2}{CL_2^2}, \frac{t}{(m + 3\sigma)L_1}\right\}\right). \end{aligned}$$

Finally, we can express concentration in terms of enlargements of convex sets. The proof of the corollary below is a modification of the argument for smooth functions, taking into account some additional difficulties related to the fact that in the convex situation one cannot truncate the function (as this operation destroys convexity).

COROLLARY 1.8. *Let  $X_1, \dots, X_n$  be independent random variables such that  $\mathcal{L}(X_i) \in \mathcal{M}_\beta(m, \sigma^{\beta+1})$  and let  $A \in \mathbb{R}^n$  be a convex set with*

$$\mathbb{P}((X_1, \dots, X_n) \in A) \geq 1/2.$$

Then, for all  $r \geq 0$ ,

$$\mathbb{P}((X_1, \dots, X_n) \notin A + r^{1/2}B_2 + r^{1/(1+\beta)}B_{1+\beta}) \leq e^{-C'r}$$

for some constant  $C'$  depending only on  $\beta$  and  $C(\beta, m, \sigma)$  from Theorem 1.3 (and additionally on  $m + 3\sigma$  in the case  $\beta = 0$ ).

The reader may compare the above corollaries with a general result stated in [2, Proposition 26, Theorem 27] about consequences of inequalities similar to (1.3) in the classical setting of smooth functions (see also [2, Example 29]).

**1.3. Relation to other results, and further questions.** Recently, the classical log-Sobolev inequality for convex functions has been shown to be equivalent to certain transport-entropy inequalities with weak transport cost [13]. In Propositions 3.1 and 3.6 we show that this is also the case for inequalities (1.3) and (1.4). Since the precise formulation of these results requires introducing additional, rather involved notation, we postpone it to Section 3.

Logarithmic Sobolev inequalities for convex functions have one weakness when compared to their counterparts for smooth functions, namely they provide only deviation inequalities for the upper tail of the function (in the classical setting one obtains bounds on the lower tail by applying the inequality to  $-\varphi$ , an approach which is precluded in our situation, since except for trivial cases  $-\varphi$  is not convex).

It turns out that for  $\beta = 0$  the condition  $\mu \in \mathcal{M}_\beta(m, \sigma^{\beta+1})$  is in fact equivalent to a two-sided concentration inequality for convex functions. This is a consequence of a recent result by Feldheim, Marsiglietti, Nayar, and Wang [7] who proved an infimum convolution inequality, which—as follows from our Proposition 3.6—is stronger than (1.4).

**1.3.1. A remark about recent improvements by Gozlan et al.** During the review process of this manuscript a preprint [12] by Gozlan, Roberto, Samson, Shu, and Tetali appeared, in which the authors extended the results by Feldheim et al. and characterized the infimum-convolution inequality (equivalently weak transportation inequality) for probability measures on the real line and a large class of cost functions. Their condition can be shown to be strictly weaker than ours, while (as follows from Proposition 3.1) the inequalities they consider are stronger than modified log-Sobolev inequalities investigated by us. The condition of [12] is expressed in terms of the monotone transport map  $U$  between the symmetric exponential distribution and the measure  $\mu$  and in the case we consider reads

$$\sup_{x \in \mathbb{R}} (U(x+u) - U(x)) \leq C(1+u)^{1/(\beta+1)}$$

for some  $C < \infty$ . Using the same arguments as in Lemma 2.2 below, one can show that this is implied by  $\mu \in \mathcal{M}_\beta(m, \sigma^{\beta+1})$  for some  $m, \sigma^{\beta+1}$ . On the other hand, for measures supported on the positive half-line, the latter condition is equivalent to

$$\sup_{x \in \mathbb{R}} (U(x+u)^{\beta+1} - U(x)^{\beta+1}) \leq D(1+u)$$

for some  $D < \infty$  and so it is not difficult to find a measure which satisfies the condition of [12] but does not belong to  $\mathcal{M}_\beta(m, \sigma^{\beta+1})$  for any  $m, \sigma^{\beta+1}$ .

To the best of our knowledge, except for the case of  $\beta = 0$ , the characterization of measures on the line satisfying the modified log-Sobolev inequality for convex functions is still unknown. For  $\beta = 0$  these are exactly the measures  $\mu$  which belong to the class  $\mathcal{M}_0(m, \sigma)$  for some  $m$  and  $\sigma$  (this follows e.g. from Theorem 1.3 and the characterization of the convex Poincaré inequality due to Bobkov and Götze [4]). We provide a more detailed discussion of the results of [7, 12] and their relation to our results in Section 3 (see Remark 3.8).

**1.4. Organization of the paper.** Section 2 is devoted to the proofs of our main results presented above. In Section 3 we discuss connections with transportation inequalities and in Section 4 we briefly comment on the relation between the class  $\mathcal{M}_\beta(m, \sigma^{\beta+1})$  and the class of functions satisfying the modified log-Sobolev inequality for all smooth functions.

**2. Proofs of the main results**

**2.1. Technical lemmas.** In this section we will prove several technical lemmas which will be used in the proofs of our main results. Of particular importance is Lemma 2.4, which constitutes the core of the proof of Theorem 1.3.

In the proofs we will use the notation  $C(\alpha)$  to denote constants depending only on some parameter  $\alpha$ . The value of such constants may change between occurrences.

*Proof of Proposition 1.2.* Assume (i) holds. For  $x \geq m$  we have

$$\begin{aligned} \sigma^{\beta+1} \mu([x, \infty)) &\geq \nu^+([x, \infty)) \geq \int_x^{x+2\sigma^{\beta+1}/x^\beta} y^\beta \mu([y, \infty)) dy \\ &\geq \frac{2\sigma^{\beta+1}}{x^\beta} \cdot x^\beta \mu([x + 2\sigma^{\beta+1}/x^\beta, \infty)), \end{aligned}$$

which clearly implies the first inequality of (ii). The second inequality follows similarly.

Suppose now that (ii) is satisfied. For  $x \geq m$  define the sequence  $a_0 = x$ ,  $a_{n+1} = a_n + h/a_n^\beta$ . It is easy to see that this sequence is increasing,  $a_n \rightarrow \infty$  and  $a_{n+1}/a_n \rightarrow 1$ . Therefore

$$\begin{aligned} \nu^+([x, \infty)) &= \int_x^\infty y^\beta \mu([y, \infty)) dy \leq \sum_{n=0}^\infty (a_{n+1} - a_n) a_{n+1}^\beta \mu([a_n, \infty)) \\ &\leq K \sum_{n=0}^\infty \alpha^n \mu([a_0, \infty)) = \frac{K}{1 - \alpha} \mu([x, \infty)), \end{aligned}$$

where  $K = \sup (\tilde{a}_{n+1} - a_n) a_{n+1}^\beta = h \sup (a_{n+1}/a_n)^\beta = h \sup (1 + h/a_n^{\beta+1})^\beta \leq h(1 + h/m^{\beta+1})^\beta$ . We can proceed analogously to obtain the condition on the left tail. ■

**LEMMA 2.1.** *Let  $\mu \in \mathcal{M}_\beta(m, \sigma^{\beta+1})$ . Then for all functions  $f: \mathbb{R} \rightarrow [0, \infty)$  which are nonincreasing for  $x \leq x_0$  and nondecreasing for  $x \geq x_0$  (for some  $x_0 \in [-\infty, \infty]$ ) we have*

$$\int_{\tilde{m}}^\infty f(x) x^\beta \mu([x, \infty)) dx \leq \sigma^{\beta+1} \int_{\mathbb{R}} f(x) d\mu(x),$$

where  $\tilde{m} = m \vee ((2\beta)^{1/(\beta+1)} \sigma) + 2\sigma^{\beta+1}/(m \vee ((2\beta)^{1/(\beta+1)} \sigma))^\beta$ .

*Proof.* First notice that the inequalities of Definition 1.1 are also satisfied for sets of the form  $(x, \infty)$ ,  $x \geq m$ . We have

$$(2.1) \quad \int_{\tilde{m}}^{\infty} f(x)x^{\beta}\mu([x, \infty)) dx = \int_{\tilde{m}}^{\infty} \int_0^{\infty} 1_{\{s \leq f(x)\}} x^{\beta} \mu([x, \infty)) ds dx \\ = \int_0^{\infty} \nu^+(\{x \geq \tilde{m} : f(x) \geq s\}) ds.$$

The set  $A = \{x \geq \tilde{m} : f(x) \geq s\}$  is either a (possibly empty) half-line contained in  $[\tilde{m}, \infty)$  or a disjoint union of such a half-line and an interval  $I$  with left end equal to  $\tilde{m}$ . In the former case we have

$$\nu^+(A) \leq \sigma^{\beta+1}\mu(A) \leq \sigma^{\beta+1}\mu(\{x \in \mathbb{R} : f(x) \geq s\}).$$

In the latter case we denote the right end of  $I$  by  $t$ . Let  $a = m \vee ((2\beta)^{1/(\beta+1)}\sigma)$ , so that  $\tilde{m} = a + 2\sigma^{\beta+1}/a^{\beta}$ . Since  $t \geq \tilde{m} \geq a$  we obtain

$$\nu^+(A) \leq \nu^+([\tilde{m}, \infty)) \leq \sigma^{\beta+1}\mu([\tilde{m}, \infty)) \leq \sigma^{\beta+1}\mu([a, \tilde{m})) \\ \leq \sigma^{\beta+1}\mu([a, t]) \leq \sigma^{\beta+1}\mu(\{x \leq t : f(x) \geq s\}),$$

where the second inequality follows from the assumption  $\mu \in \mathcal{M}_{\beta}(m, \sigma^{\beta+1})$ , the third from Proposition 1.2 and the definition of  $\tilde{m}$  (which imply that  $\mu([a, \tilde{m})) = \mu([a, \infty)) - \mu([\tilde{m}, \infty)) \geq \mu([\tilde{m}, \infty))$ ), and the last one from the observation that  $x_0 \geq t$  and thus the function  $f$  is nonincreasing for  $x \leq t$ .

Hence, in both cases we have  $\nu^+(A) \leq \sigma^{\beta+1}\mu(\{x \in \mathbb{R} : f(x) \geq s\})$ . Now we can write

$$\sigma^{\beta+1} \int_0^{\infty} \mu(\{x \in \mathbb{R} : f(x) \geq s\}) ds = \sigma^{\beta+1} \int_{\mathbb{R}} f(x) d\mu(x),$$

which together with (2.1) ends the proof. ■

LEMMA 2.2. *If  $X$  is a random variable such that  $\mathcal{L}(X) \in \mathcal{M}_{\beta}(m, \sigma^{\beta+1})$  then  $\mathbb{P}(|X| \geq t) \leq C_1(\beta, m, \sigma)e^{-t^{\beta+1}/C_2(\beta, m, \sigma)}$  for all  $t \geq 0$ .*

*Proof.* Obviously it is sufficient to prove the inequality for  $t \geq 4m$ . Define continuous functions

$$g(x) = \int_x^{\infty} y^{\beta} \mathbb{P}(X \geq y) dy \quad \text{and} \quad f(z) = \int_z^{\infty} x^{\beta} g(x) dx.$$

By the definition of  $\mathcal{M}_{\beta}(m, \sigma^{\beta+1})$ , for  $x \geq m$  we have

$$g(x) \leq \sigma^{\beta+1} \mathbb{P}(X \geq x).$$

Thus, for  $z \geq m$ ,

$$f(z) = \int_z^{\infty} x^{\beta} g(x) dx \leq \int_z^{\infty} x^{\beta} \sigma^{\beta+1} \mathbb{P}(X \geq x) dx = \sigma^{\beta+1} g(z).$$

We can rewrite this as

$$f'(z) \leq -\frac{1}{\sigma^{\beta+1}} z^\beta f(z), \quad z \geq m,$$

which gives  $f(z) \leq C \exp(-z^{\beta+1}/((\beta+1)\sigma^{\beta+1}))$  with  $C$  depending only on  $\beta$ ,  $m$  and  $\sigma^{\beta+1}$  (note that  $f(m) \leq \sigma^{2\beta+2}$ ). Now, as  $g$  is nonincreasing, we have  $f(z) \geq \int_z^{2z} y^\beta g(y) dy \geq z^{\beta+1} g(2z)$  and hence for  $z \geq m$  we get

$$g(2z) \leq \frac{f(z)}{z^{\beta+1}} \leq \frac{C \exp(-z^{\beta+1}/((\beta+1)\sigma^{\beta+1}))}{z^{\beta+1}}.$$

Similarly

$$\mathbb{P}(X \geq 4x) \leq \frac{g(2x)}{(2x)^{\beta+1}} \leq \frac{C \exp(-x^{\beta+1}/((\beta+1)\sigma^{\beta+1}))}{2^{\beta+1} x^{2\beta+2}}$$

for  $x \geq m$ .

We can analogously deal with the lower tail. ■

The next two lemmas contain the core of our argument and are counterparts of Lemmas 4 and 3 in [1]. While Lemma 2.3 follows quite closely the argument from [1], Lemma 2.4 requires for  $\beta < 1$  a significantly new approach.

LEMMA 2.3. *Let  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  be a smooth convex Lipschitz function, non-increasing on  $(-\infty, 0)$ , and  $X$  a random variable with  $\mathcal{L}(X) \in \mathcal{M}_\beta(m, \sigma^{\beta+1})$ . Then there exists a constant  $C_3(\beta, m, \sigma)$  such that*

$$\int_{\{(x,y) \in \mathbb{R}^2: xy \leq 0\}} \varphi'(x)\varphi'(y)e^{\varphi(y)}\mathbb{P}(X \leq x \wedge y)\mathbb{P}(X \geq x \vee y) dx dy \leq C_3(\beta, m, \sigma)\mathbb{E}\varphi'(X)^2 e^{\varphi(X)}.$$

*Proof.* For  $\beta > 0$  we will assume without loss of generality that  $m \geq 1$ . Let  $\tilde{m}$  be the constant defined in Lemma 2.1 (note that for  $\beta > 0$ ,  $\tilde{m} > 1$ ). For  $x < 0 < y$  we have either  $\varphi'(x)\varphi'(y) \leq 0$  or  $\varphi'(x) \leq \varphi'(y) < 0$  and  $\varphi(x) \geq \varphi(y)$ , so

$$\begin{aligned} & \int_{-\infty}^0 \int_0^\infty \varphi'(x)\varphi'(y)e^{\varphi(y)}\mathbb{P}(X \leq x)\mathbb{P}(X \geq y) dy dx \\ & \leq \int_{-\infty}^0 \int_0^\infty \varphi'(x)^2 e^{\varphi(x)}\mathbb{P}(X \leq x)\mathbb{P}(X \geq y) dy dx \\ & \leq C(\beta, m, \sigma) \int_{-\infty}^0 \varphi'(x)^2 e^{\varphi(x)}\mathbb{P}(X \leq x) dx, \end{aligned}$$

where the last inequality follows from the fact that by Lemma 2.2,

$$\int_0^{\infty} \mathbb{P}(X \geq y) dy = \mathbb{E}X_+ \leq C(\beta, m, \sigma).$$

The integral over the set where  $x > 0 > y$  can be treated similarly, namely we have

$$\begin{aligned} \int_0^{\infty} \int_{-\infty}^0 \varphi'(x)\varphi'(y)e^{\varphi(y)}\mathbb{P}(X \leq y)\mathbb{P}(X \geq x) dy dx \\ \leq \int_0^{\infty} \int_{-\infty}^0 \varphi'(y)^2 e^{\varphi(y)}\mathbb{P}(X \leq y)\mathbb{P}(X \geq x) dy dx \\ \leq C(\beta, m, \sigma) \int_{-\infty}^0 \varphi'(y)^2 e^{\varphi(y)}\mathbb{P}(X \leq y) dy. \end{aligned}$$

Now, by Lemma 2.1 (note that  $-X \in \mathcal{M}_\beta(m, \sigma^{\beta+1})$  and the function  $x \mapsto \varphi'(-x)^2 e^{\varphi(-x)}$  is nonincreasing for  $x \leq x_0$  and nondecreasing for  $x \geq x_0$  for some  $x_0 \in [-\infty, 0]$ ),

$$\begin{aligned} \int_{-\infty}^{-\tilde{m}} \varphi'(x)^2 e^{\varphi(x)}\mathbb{P}(X \leq x) dx &\leq \int_{-\infty}^{-\tilde{m}} \varphi'(x)^2 e^{\varphi(x)}(-x)^\beta \mathbb{P}(X \leq x) dx \\ &\leq \sigma^{\beta+1} \mathbb{E}\varphi'(X)^2 e^{\varphi(X)}. \end{aligned}$$

Moreover

$$\begin{aligned} \int_{-\tilde{m}}^0 \varphi'(x)^2 e^{\varphi(x)}\mathbb{P}(X \leq x) dx &= \mathbb{E} \int_{-\tilde{m}}^0 \varphi'(x)^2 e^{\varphi(x)} 1_{\{X \leq x\}} dx \\ &\leq \tilde{m} \mathbb{E}\varphi'(X)^2 e^{\varphi(X)}, \end{aligned}$$

where the last inequality follows from the fact that  $\varphi'(x)^2 e^{\varphi(x)} \leq \varphi'(X)^2 e^{\varphi(X)}$  for  $X \leq x \leq 0$ . ■

LEMMA 2.4. *Let  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  be a smooth convex Lipschitz function and  $X$  a random variable with  $\mathcal{L}(X) = \mu \in \mathcal{M}_\beta(m, \sigma^{\beta+1})$ . If  $\beta > 0$ , then there exists a constant  $C_4(\beta, m, \sigma)$  such that*

$$(2.2) \quad \int_0^{\infty} \int_0^{\infty} \varphi'(x)\varphi'(y)e^{\varphi(y)}\mathbb{P}(X \leq x \wedge y)\mathbb{P}(X \geq x \vee y) dx dy \leq C_4(\beta, m, \sigma)\mathbb{E}(|\varphi'(X)|^2 \vee |\varphi'(X)|^{(\beta+1)/\beta})e^{\varphi(X)}.$$

If  $\beta = 0$  and  $\varphi' \leq 1/(2(m+3\sigma))$ , then there exists a constant  $C_4(\beta = 0, m, \sigma)$  such that

$$\int_0^\infty \int_0^\infty \varphi'(x)\varphi'(y)e^{\varphi(y)}\mathbb{P}(X \leq x \wedge y)\mathbb{P}(X \geq x \vee y) dx dy \leq C_4(\beta = 0, m, \sigma)\mathbb{E}|\varphi'(X)|^2 e^{\varphi(X)}.$$

*Proof.* Let us first notice that the left-hand side of (2.2) is equal to

$$\int_0^\infty \int_0^y \varphi'(x)\varphi'(y)(e^{\varphi(x)} + e^{\varphi(y)})\mathbb{P}(X \leq x)\mathbb{P}(X \geq y) dx dy$$

(since on  $\{(x, y) : x > y\}$  we can make a change of variables swapping  $x$  and  $y$ ).

Since  $\varphi$  is convex there exists a point  $x_0$  (possibly 0 or infinity) at which  $\varphi$  attains its infimum on  $[0, \infty]$ . Let  $\tilde{m}$  be the number defined in Lemma 2.1. Obviously, for  $\beta > 0$  we can assume that  $\tilde{m} \geq 1$  (note that we do not change the value of  $\tilde{m}$  in the case  $\beta = 0$ ). We split the outer integral into integrals over the intervals  $(0, x_0 \wedge \tilde{m})$ ,  $(x_0 \wedge \tilde{m}, x_0 \vee \tilde{m})$  and  $(x_0 \vee \tilde{m}, \infty)$ .

STEP 1. Since  $\varphi$  is nonincreasing on the interval  $(0, x_0 \wedge \tilde{m})$  we have

$$\varphi'(x)\varphi'(y)(e^{\varphi(x)} + e^{\varphi(y)}) \leq 2\varphi'(x)^2 e^{\varphi(x)}$$

for  $0 \leq x \leq y \leq x_0 \wedge \tilde{m}$ . Therefore

$$\begin{aligned} (2.3) \quad & \int_0^{x_0 \wedge \tilde{m}} \int_0^y \varphi'(x)\varphi'(y)(e^{\varphi(x)} + e^{\varphi(y)})\mathbb{P}(X \leq x)\mathbb{P}(X \geq y) dx dy \\ & \leq 2 \int_0^{x_0 \wedge \tilde{m}} \int_0^y \varphi'(x)^2 e^{\varphi(x)}\mathbb{P}(X \leq x)\mathbb{P}(X \geq y) dx dy \\ & \leq 2\tilde{m}\mathbb{E} \int_0^{x_0 \wedge \tilde{m}} \varphi'(x)^2 e^{\varphi(x)} 1_{\{X \leq x\}} dx \leq 2\tilde{m}^2 \mathbb{E}\varphi'(X)^2 e^{\varphi(X)}, \end{aligned}$$

where the last inequality follows from the fact that  $\varphi'(x)^2 e^{\varphi(x)} \leq \varphi'(X)^2 e^{\varphi(X)}$  for  $X \leq x \leq x_0$ .

STEP 2. To estimate the integral over the interval  $(x_0 \wedge \tilde{m}, x_0 \vee \tilde{m})$  we consider two cases. If  $x_0 < \tilde{m}$ , then

$$\begin{aligned} (2.4) \quad & \int_0^{\tilde{m}} \int_0^y \varphi'(x)\varphi'(y)(e^{\varphi(x)} + e^{\varphi(y)})\mathbb{P}(X \leq x)\mathbb{P}(X \geq y) dx dy \\ & \leq 2 \int_{x_0}^{\tilde{m}} \int_{x_0}^y \varphi'(y)^2 e^{\varphi(y)}\mathbb{P}(X \geq y) dx dy \\ & \leq 2\tilde{m}\mathbb{E} \int_{x_0}^{\tilde{m}} \varphi'(y)^2 e^{\varphi(y)} 1_{\{X \geq y\}} dy \leq 2\tilde{m}^2 \mathbb{E}\varphi'(X)^2 e^{\varphi(X)}. \end{aligned}$$

If  $x_0 > \tilde{m}$ , then as above (the third last passage below follows by Definition 1.1 and the last by Lemma 2.1; for the third passage recall also that  $\tilde{m} \geq 1$  for  $\beta > 0$ )

$$\begin{aligned}
(2.5) \quad & \int_{\tilde{m}}^{x_0} \int_0^y \varphi'(x) \varphi'(y) (e^{\varphi(x)} + e^{\varphi(y)}) \mathbb{P}(X \leq x) \mathbb{P}(X \geq y) dx dy \\
& \leq 2 \int_{\tilde{m}}^{x_0} \int_0^y \varphi'(x)^2 e^{\varphi(x)} \mathbb{P}(X \leq x) \mathbb{P}(X \geq y) dx dy \\
& = 2 \int_0^{x_0} \int_{x \vee \tilde{m}}^{x_0} \varphi'(x)^2 e^{\varphi(x)} \mathbb{P}(X \leq x) \mathbb{P}(X \geq y) dy dx \\
& \leq 2 \int_0^{x_0} \varphi'(x)^2 e^{\varphi(x)} \mathbb{P}(X \leq x) \int_{x \vee \tilde{m}}^{\infty} y^\beta \mathbb{P}(X \geq y) dy dx \\
& \leq 2\sigma^{\beta+1} \int_0^{x_0} \varphi'(x)^2 e^{\varphi(x)} \mathbb{P}(X \leq x) \mathbb{P}(X \geq x) dx \\
& \leq 2\sigma^{\beta+1} \left( \int_0^{\tilde{m}} \varphi'(x)^2 e^{\varphi(x)} \mathbb{P}(X \leq x) dx + \int_{\tilde{m}}^{x_0} \varphi'(x)^2 e^{\varphi(x)} x^\beta \mathbb{P}(X \geq x) dx \right) \\
& \leq 2\sigma^{\beta+1} (\tilde{m} + \sigma^{\beta+1}) \mathbb{E} \varphi'(X)^2 e^{\varphi(X)}.
\end{aligned}$$

STEP 3. It remains to estimate the integral over  $(x_0 \vee \tilde{m}, \infty)$ . In what follows we can assume that  $x_0 < \infty$  and  $\varphi(x_0) = 0$  (since if we subtract  $\varphi(x_0)$  from  $\varphi$ , both sides of (2.2) will change by a factor of  $\exp(-\varphi(x_0))$ ). Fix  $\delta = 1/(2(\tilde{m} + \sigma^{\beta+1}))$  and let  $x_1 = \sup\{x > x_0 : \varphi'(x) \leq \delta\}$  (with the convention that  $x_1 = x_0$  if  $\varphi'(x_0) > \delta$ ). Since  $\varphi'(x) \leq 0 \leq \varphi'(y)$  for  $0 < x \leq x_0 \leq y$  we have

$$\begin{aligned}
(2.6) \quad & \int_{x_0 \vee \tilde{m}}^{\infty} \int_0^y \varphi'(x) \varphi'(y) (e^{\varphi(x)} + e^{\varphi(y)}) \mathbb{P}(X \leq x) \mathbb{P}(X \geq y) dx dy \\
& \leq \int_{x_0 \vee \tilde{m}}^{\infty} \int_{x_0}^y \varphi'(x) \varphi'(y) (e^{\varphi(x)} + e^{\varphi(y)}) \mathbb{P}(X \leq x) \mathbb{P}(X \geq y) dx dy \\
& \leq 2 \int_{x_0 \vee \tilde{m}}^{\infty} \int_{x_0}^y \varphi'(x) \varphi'(y) e^{\varphi(y)} \mathbb{P}(X \geq y) dx dy \\
& = 2 \int_{x_0 \vee \tilde{m}}^{\infty} \varphi(y) \varphi'(y) e^{\varphi(y)} \mathbb{P}(X \geq y) dy
\end{aligned}$$

$$= 2 \int_{x_0 \vee \tilde{m}}^{x_1 \vee \tilde{m}} \varphi(y) \varphi'(y) e^{\varphi(y)} \mathbb{P}(X \geq y) dy + 2 \int_{x_1 \vee \tilde{m}}^{\infty} \varphi(y) \varphi'(y) e^{\varphi(y)} \mathbb{P}(X \geq y) dy.$$

We will estimate the last two integrals separately.

STEP 3a. To estimate the first integral we define a convex function by

$$\tilde{\varphi}(x) = \begin{cases} \varphi(x) & \text{if } x \leq x_1, \\ \varphi(x_1) + (x - x_1)\delta & \text{if } x > x_1. \end{cases}$$

The function  $\tilde{\varphi}$  is  $C^1$ -smooth and  $0 \leq \tilde{\varphi}'(x) \leq \delta$  for  $x \in (x_0, \infty)$ . Moreover  $\varphi(x) = \tilde{\varphi}(x)$  and  $\varphi'(x) = \tilde{\varphi}'(x)$  for  $x \in (x_0 \vee \tilde{m}, x_1 \vee \tilde{m})$  (since this interval is contained in  $(x_0, x_1)$  or is an empty set). Also  $0 \leq \tilde{\varphi}(x) \leq \varphi(x)$  and  $\tilde{\varphi}'(x)^2 \leq \varphi'(x)^2$  for  $x \in \mathbb{R}$ . Hence (the third inequality follows from Lemma 2.1)

$$\begin{aligned} (2.7) \quad & \int_{x_0 \vee \tilde{m}}^{x_1 \vee \tilde{m}} \varphi(y) \varphi'(y) e^{\varphi(y)} \mathbb{P}(X \geq y) dy \\ & \leq \int_{x_0 \vee \tilde{m}}^{\infty} \tilde{\varphi}(y) \tilde{\varphi}'(y) e^{\tilde{\varphi}(y)} \mathbb{P}(X \geq y) dy \\ & \leq \frac{1}{2} \int_{x_0 \vee \tilde{m}}^{\infty} (\tilde{\varphi}(y)^2 + \tilde{\varphi}'(y)^2) e^{\tilde{\varphi}(y)} y^\beta \mathbb{P}(X \geq y) dy \\ & \leq \frac{\sigma^{\beta+1}}{2} \int_{x_0}^{\infty} (\tilde{\varphi}(y)^2 + \tilde{\varphi}'(y)^2) e^{\tilde{\varphi}(y)} d\mu(y) \\ & \leq \frac{\sigma^{\beta+1}}{2} \mathbb{E} \tilde{\varphi}(X)^2 e^{\tilde{\varphi}(X)} 1_{\{X \geq x_0\}} + \frac{\sigma^{\beta+1}}{2} \mathbb{E} \varphi'(X)^2 e^{\varphi(X)}. \end{aligned}$$

In order to complete the estimation of the first integral we have to deal with the expression  $\mathbb{E} \tilde{\varphi}(X)^2 e^{\tilde{\varphi}(X)} 1_{\{X \geq x_0\}}$ . Using integration by parts (recall that  $\tilde{\varphi}(x_0) = 0$ ), the Cauchy–Schwarz inequality and the fact that  $\tilde{\varphi}'$  is bounded by  $\delta$  we can write

$$\begin{aligned} (2.8) \quad & \mathbb{E} \tilde{\varphi}(X)^2 e^{\tilde{\varphi}(X)} 1_{\{X \geq x_0\}} \\ & = 2 \int_{x_0}^{\infty} \tilde{\varphi}'(x) \tilde{\varphi}(x) e^{\tilde{\varphi}(x)} \mathbb{P}(X \geq x) dx + \int_{x_0}^{\infty} \tilde{\varphi}(x)^2 \tilde{\varphi}'(x) e^{\tilde{\varphi}(x)} \mathbb{P}(X \geq x) dx \\ & \leq 2 \left( \int_{x_0}^{\infty} \tilde{\varphi}(x)^2 e^{\tilde{\varphi}(x)} \mathbb{P}(X \geq x) dx \right)^{1/2} \left( \int_{x_0}^{\infty} \tilde{\varphi}'(x)^2 e^{\tilde{\varphi}(x)} \mathbb{P}(X \geq x) dx \right)^{1/2} \\ & \quad + \delta \int_{x_0}^{\infty} \tilde{\varphi}(x)^2 e^{\tilde{\varphi}(x)} \mathbb{P}(X \geq x) dx. \end{aligned}$$

We can now use Lemma 2.1 to estimate the above integrals by expressions of the type  $\mathbb{E}\tilde{\varphi}(X)^2 e^{\tilde{\varphi}(X)} 1_{\{X \geq x_0\}}$  or  $\mathbb{E}\varphi'(X)^2 e^{\varphi(X)}$ . For example

$$\begin{aligned} & \int_{x_0}^{\infty} \tilde{\varphi}(x)^2 e^{\tilde{\varphi}(x)} \mathbb{P}(X \geq x) dx \\ & \leq \mathbb{E} \int_{x_0}^{x_0 \vee \tilde{m}} \tilde{\varphi}(x)^2 e^{\tilde{\varphi}(x)} 1_{\{X \geq x\}} dx + \int_{x_0 \vee \tilde{m}}^{\infty} \tilde{\varphi}(x)^2 e^{\tilde{\varphi}(x)} x^\beta \mathbb{P}(X \geq x) dx \\ & \leq (\tilde{m} + \sigma^{\beta+1}) \mathbb{E}\tilde{\varphi}(X)^2 e^{\tilde{\varphi}(X)} 1_{\{X \geq x_0\}}, \end{aligned}$$

and similarly (at the end we skip the indicator function and replace  $\tilde{\varphi}$  by  $\varphi$ )

$$\int_{x_0}^{\infty} \tilde{\varphi}'(x)^2 e^{\tilde{\varphi}(x)} \mathbb{P}(X \geq x) dx \leq (\tilde{m} + \sigma^{\beta+1}) \mathbb{E}\varphi'(X)^2 e^{\varphi(X)}$$

(note that we use the monotonicity of  $\tilde{\varphi}$ ,  $\tilde{\varphi}'$  to ensure that the assumptions of Lemma 2.1 are satisfied). Plugging this into (2.8) gives us

$$\begin{aligned} (2.9) \quad & \mathbb{E}\tilde{\varphi}(X)^2 e^{\tilde{\varphi}(X)} 1_{\{X \geq x_0\}} \\ & \leq 2(\tilde{m} + \sigma^{\beta+1}) (\mathbb{E}\tilde{\varphi}(X)^2 e^{\tilde{\varphi}(X)} 1_{\{X \geq x_0\}})^{1/2} (\mathbb{E}\varphi'(X)^2 e^{\varphi(X)})^{1/2} \\ & \quad + \delta(\tilde{m} + \sigma^{\beta+1}) \mathbb{E}\tilde{\varphi}(X)^2 e^{\tilde{\varphi}(X)} 1_{\{X \geq x_0\}}, \end{aligned}$$

which we solve with respect to  $\mathbb{E}\tilde{\varphi}(X)^2 e^{\tilde{\varphi}(X)} 1_{\{X \geq x_0\}}$  and (since  $\delta(\tilde{m} + \sigma^{\beta+1}) = 1/2$ ) arrive at

$$\mathbb{E}\tilde{\varphi}(X)^2 e^{\tilde{\varphi}(X)} 1_{\{X \geq x_0\}} \leq C(\beta, m, \sigma) \mathbb{E}\varphi'(X)^2 e^{\varphi(X)}.$$

This allows us to finish the computations of (2.7) and finally conclude that

$$(2.10) \quad \int_{x_0 \vee \tilde{m}}^{x_1 \vee \tilde{m}} \varphi(y) \varphi'(y) e^{\varphi(y)} \mathbb{P}(X \geq y) dy \leq C(\beta, m, \sigma) \mathbb{E}\varphi'(X)^2 e^{\varphi(X)}.$$

This ends the estimation of the first integral from (2.6).

STEP 3b. To complete the proof we have to estimate the second integral from (2.6). Notice that if  $\beta = 0$ , then  $x_1 = \infty$  since by assumption

$$\varphi' \leq \delta = 1/(2(\tilde{m} + \sigma)) = 1/(2(m + 3\sigma))$$

and in this case that integral disappears. Therefore we can assume that  $\beta > 0$ .

Fix  $\varepsilon > 0$ . Since  $\varphi(x_0) = 0$ , by convexity

$$\varphi(y)^\beta \leq \varphi'(y)^\beta (y - x_0)^\beta \leq \varphi'(y)^\beta y^\beta.$$

Using Lemma 2.1 and Young's inequality with exponents  $1/(1 - \beta)$  and  $1/\beta$  we get

$$\begin{aligned}
(2.11) \quad & \int_{x_1 \vee \tilde{m}}^{\infty} \varphi(y) \varphi'(y) e^{\varphi(y)} \mathbb{P}(X \geq y) dy \\
& \leq \int_{\tilde{m}}^{\infty} \mathbb{1}_{\{y \in (x_1 \vee \tilde{m}, \infty)\}} \varphi(y)^{1-\beta} \varphi'(y)^{1+\beta} e^{\varphi(y)} y^\beta \mathbb{P}(X \geq y) dy \\
& \leq \sigma^{\beta+1} \int_{\mathbb{R}} \mathbb{1}_{\{y \in (x_1 \vee \tilde{m}, \infty)\}} \varphi(y)^{1-\beta} \varphi'(y)^{1+\beta} e^{\varphi(y)} d\mu(y) \\
& \leq \varepsilon(1-\beta) \sigma^{\beta+1} \mathbb{E} \varphi(X) e^{\varphi(X)} \mathbb{1}_{\{X \geq x_1 \vee \tilde{m}\}} \\
& \quad + \sigma^{\beta+1} C(\varepsilon, \beta) \mathbb{E} |\varphi'(X)|^{(1+\beta)/\beta} e^{\varphi(X)}.
\end{aligned}$$

We will estimate  $\mathbb{E} \varphi(X) e^{\varphi(X)} \mathbb{1}_{\{X \geq x_1 \vee \tilde{m}\}}$  by a sum of expressions which appear in the assertion of the lemma and the integral

$$\int_{x_1 \vee \tilde{m}}^{\infty} \varphi(y) \varphi'(y) e^{\varphi(y)} \mathbb{P}(X \geq y) dy.$$

Then we will pick  $\varepsilon = \varepsilon(\beta, m, \sigma)$  small enough to get from (2.11) an estimate of this integral, which will end the estimation of the second integral on the right-hand side of (2.6), and the proof of Lemma 2.4.

Integration by parts gives

$$\begin{aligned}
(2.12) \quad \mathbb{E} \varphi(X) e^{\varphi(X)} \mathbb{1}_{\{X \geq x_1 \vee \tilde{m}\}} &= \varphi(x_1 \vee \tilde{m}) e^{\varphi(x_1 \vee \tilde{m})} \mathbb{P}(X \geq x_1 \vee \tilde{m}) \\
& \quad + \int_{x_1 \vee \tilde{m}}^{\infty} \varphi'(y) e^{\varphi(y)} \mathbb{P}(X \geq y) dy \\
& \quad + \int_{x_1 \vee \tilde{m}}^{\infty} \varphi(y) \varphi'(y) e^{\varphi(y)} \mathbb{P}(X \geq y) dy,
\end{aligned}$$

so it remains to estimate  $\int_{x_1 \vee \tilde{m}}^{\infty} \varphi'(y) e^{\varphi(y)} \mathbb{P}(X \geq y) dy$  and the boundary term  $\varphi(x_1 \vee \tilde{m}) e^{\varphi(x_1 \vee \tilde{m})} \mathbb{P}(X \geq x_1 \vee \tilde{m})$ . Since  $\tilde{m} \geq 1$  and for  $x \geq x_1$  we have  $\varphi'(x) \geq \delta$ , we can write

$$\begin{aligned}
(2.13) \quad \int_{x_1 \vee \tilde{m}}^{\infty} \varphi'(y) e^{\varphi(y)} \mathbb{P}(X \geq y) dy &\leq \frac{1}{\delta} \int_{x_1 \vee \tilde{m}}^{\infty} \varphi'(y)^2 e^{\varphi(y)} y^\beta \mathbb{P}(X \geq y) dy \\
&\leq \frac{\sigma^{\beta+1}}{\delta} \mathbb{E} \varphi'(X)^2 e^{\varphi(X)}.
\end{aligned}$$

As for the boundary term, we write it as

$$(2.14) \quad \mathbb{E} \left( (\varphi(x_1 \vee \tilde{m}) e^{\varphi(x_1 \vee \tilde{m})} - e^{\varphi(x_1 \vee \tilde{m})} + 1) + (e^{\varphi(x_1 \vee \tilde{m})} - 1) \right) \mathbb{1}_{\{X \geq x_1 \vee \tilde{m}\}}.$$

It is easy to deal with the second part, since

$$(2.15) \quad \begin{aligned} \mathbb{E}(e^{\varphi(x_1 \vee \tilde{m})} - 1)1_{\{X \geq x_1 \vee \tilde{m}\}} &\leq \mathbb{E}e^{\varphi(x_1 \vee \tilde{m})}1_{\{X \geq x_1 \vee \tilde{m}\}} \\ &\leq \frac{1}{\delta^2} \mathbb{E}\varphi'(x_1 \vee \tilde{m})^2 e^{\varphi(x_1 \vee \tilde{m})}1_{\{X \geq x_1 \vee \tilde{m}\}} \leq \frac{1}{\delta^2} \mathbb{E}\varphi'(X)^2 e^{\varphi(X)}. \end{aligned}$$

To estimate the first part, we use the fact that  $(\varphi e^\varphi - e^\varphi + 1)' = \varphi' \varphi e^\varphi$  and  $\varphi(x_0)e^{\varphi(x_0)} - e^{\varphi(x_0)} + 1 = 0$ . Therefore

$$(2.16) \quad \begin{aligned} \mathbb{E}(\varphi(x_1 \vee \tilde{m})e^{\varphi(x_1 \vee \tilde{m})} - e^{\varphi(x_1 \vee \tilde{m})} + 1)1_{\{X \geq x_1 \vee \tilde{m}\}} \\ = \mathbb{E} \int_{x_0}^{x_1 \vee \tilde{m}} \varphi'(y)\varphi(y)e^{\varphi(y)} dy 1_{\{X \geq x_1 \vee \tilde{m}\}} \\ \leq \int_{x_0}^{x_1 \vee \tilde{m}} \varphi'(y)\varphi(y)e^{\varphi(y)} \mathbb{P}(X \geq y) dy. \end{aligned}$$

Notice that we already estimated  $\int_{x_0 \vee \tilde{m}}^{x_1 \vee \tilde{m}} \varphi'(y)\varphi(y)e^{\varphi(y)} \mathbb{P}(X \geq y) dy$  in Step 3a (see (2.10)), so it is enough to estimate the integral over  $(x_0, x_0 \vee \tilde{m})$ . By convexity of  $\varphi$  we get

$$(2.17) \quad \begin{aligned} \int_{x_0}^{x_0 \vee \tilde{m}} \varphi'(y)\varphi(y)e^{\varphi(y)} \mathbb{P}(X \geq y) dy \\ \leq \int_{x_0}^{x_0 \vee \tilde{m}} \varphi'(y)^2 (y - x_0) e^{\varphi(y)} \mathbb{P}(X \geq y) dy \\ \leq \tilde{m} \mathbb{E} \int_{x_0}^{x_0 \vee \tilde{m}} \varphi'(y)^2 e^{\varphi(y)} 1_{\{X \geq y\}} dy \leq \tilde{m}^2 \mathbb{E}\varphi'(X)^2 e^{\varphi(X)}, \end{aligned}$$

where the last inequality follows from the fact that  $\varphi'(X)^2 e^{\varphi(X)} \geq \varphi'(y)^2 e^{\varphi(y)}$  for  $X \geq y \geq x_0$ . This (together with (2.14)–(2.16)) completes the estimation of the boundary term from (2.12), i.e. we get

$$\varphi(x_1 \vee \tilde{m})e^{\varphi(x_1 \vee \tilde{m})} \mathbb{P}(X \geq x_1 \vee \tilde{m}) \leq \left( \frac{1}{\delta^2} + \tilde{m}^2 + C(\beta, m, \sigma) \right) \mathbb{E}\varphi'(X)^2 e^{\varphi(X)}.$$

Therefore we get from (2.12), (2.13) the announced estimate

$$\begin{aligned} \mathbb{E}\varphi(X)e^{\varphi(X)}1_{\{X \geq x_1 \vee \tilde{m}\}} &\leq C(\beta, m, \sigma) \mathbb{E}\varphi'(X)^2 e^{\varphi(X)} \\ &\quad + \int_{x_1 \vee \tilde{m}}^{\infty} \varphi(y)\varphi'(y)e^{\varphi(y)} \mathbb{P}(X \geq y) dy. \end{aligned}$$

As explained above, plugging this into (2.11) and taking  $\varepsilon$  small enough, we deduce that

$$(2.18) \quad \int_{x_1 \vee \tilde{m}}^{\infty} \varphi(y) \varphi'(y) e^{\varphi(y)} \mathbb{P}(X \geq y) dy \leq C(\beta, m, \sigma) \mathbb{E}(|\varphi'(X)|^{(1+\beta)/\beta} + \varphi'(X)^2) e^{\varphi(X)}$$

which ends the estimation of the second integral on the right-hand side of (2.6) (and finishes Step 3b and also Step 3).

Bringing together the above results of Step 1 (see (2.3)), Step 2 (see (2.4), (2.5)) and Step 3 (see (2.6), (2.10) and (2.18)) completes the proof of the lemma. ■

## 2.2. Proofs of Theorem 1.3 and Corollaries 1.6–1.8

*Proof of Theorem 1.3.* We will follow Ledoux's approach for bounded random variables. Due to the tensorization property of the entropy (see e.g. [16, Chapter 5]) it is enough to prove the theorem for  $n = 1$ . Also, by a standard approximation argument, we can restrict our attention to convex Lipschitz functions only. Let  $Y$  be an independent copy of  $X$ . By Jensen's inequality we have

$$\begin{aligned} \text{Ent } e^{\varphi(X)} &= \mathbb{E} \varphi(x) e^{\varphi(X)} - \mathbb{E} e^{\varphi(X)} \log \mathbb{E} e^{\varphi(X)} \\ &\leq \frac{1}{2} \mathbb{E} (\varphi(X) - \varphi(Y)) (e^{\varphi(X)} - e^{\varphi(Y)}) \\ &= \mathbb{E} (\varphi(X) - \varphi(Y)) (e^{\varphi(X)} - e^{\varphi(Y)}) 1_{\{X \leq Y\}} \\ &= \mathbb{E} \int_X^Y \varphi'(x) dx \int_X^Y \varphi'(y) e^{\varphi(y)} dy 1_{\{X \leq Y\}} \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi'(x) \varphi'(y) e^{\varphi(y)} \mathbb{P}(X \leq x \wedge y) \mathbb{P}(X \geq x \vee y) dx dy. \end{aligned}$$

We split the double integral into four integrals depending on the signs of  $x$  and  $y$  and use Lemmas 2.4 and 2.3 to obtain the desired inequality—note that  $\mathcal{L}(-X) \in \mathcal{M}_\beta(m, \sigma^{\beta+1})$ , so we can assume that the infimum of  $\varphi$  is attained at some point of  $[0, \infty]$  (possibly at  $\infty$ ), which in particular means that the assumptions of Lemma 2.3 are satisfied. Moreover, we can use Lemma 2.4 to handle the integration over  $(-\infty, 0)^2$  (again by a change of variables and the fact that  $\mathcal{L}(-X) \in \mathcal{M}_\beta(m, \sigma^{\beta+1})$ ). ■

In the proofs of the corollaries we will use some additional notation and observations.

REMARK 2.5. If we substitute  $\varepsilon\varphi$  instead of  $\varphi$  into (1.3) and divide both sides by  $\varepsilon^2$  tending to zero, we recover the Poincaré inequality for convex functions:

$$\text{Var } \varphi(X_1, \dots, X_n) \leq 2C(\beta, m, \sigma) \mathbb{E} |\nabla \varphi(X_1, \dots, X_n)|^2.$$

A standard approximation argument shows that also for  $L$ -Lipschitz (but not necessarily smooth) convex functions we have

$$\text{Var } \varphi(X_1, \dots, X_n) \leq 2C(\beta, m, \sigma)L^2.$$

DEFINITION 2.6. For  $\beta \in [0, 1]$  let  $H_\beta: \mathbb{R} \rightarrow [0, \infty]$  be the function given by

$$H_\beta(t) = \max\{t^2, |t|^{(\beta+1)/\beta}\}$$

and let  $H_\beta^*: \mathbb{R} \rightarrow \mathbb{R}$  be its Legendre transform.

For  $\beta = 0$  the definition of  $H_0$  should be understood as

$$H_0(t) = t^2 1_{\{|t| \leq 1\}} + \infty 1_{\{|t| > 1\}}.$$

Similarly, all indeterminate expressions below should be interpreted as appropriate limits for  $\beta \rightarrow 0^+$ . The next lemma sums up some properties of the functions  $H_\beta$  and  $H_\beta^*$ .

LEMMA 2.7.

(a) The function  $H_\beta^*$  is given by the formula

$$H_\beta^*(t) = \begin{cases} t^2/4 & \text{if } 0 \leq |t| \leq 2, \\ |t| - 1 & \text{if } 2 \leq |t| \leq (\beta + 1)/\beta, \\ \frac{1}{\beta} \left( \frac{\beta}{\beta + 1} |t| \right)^{1+\beta} & \text{if } |t| \geq (\beta + 1)/\beta. \end{cases}$$

(b) We have

$$H_\beta^*(t) \geq \frac{3}{16} \min\{t^2, |t|^{1+\beta}\}.$$

(c) The derivative of the function  $H_\beta^*$  is given by the formula

$$\frac{d}{dt} H_\beta^*(t) = \begin{cases} |t| \operatorname{sgn}(t)/2 & \text{if } 0 \leq |t| \leq 2, \\ \operatorname{sgn}(t) & \text{if } 2 \leq |t| \leq (\beta + 1)/\beta, \\ \left( \frac{\beta}{\beta + 1} |t| \right)^\beta \operatorname{sgn}(t) & \text{if } |t| \geq (\beta + 1)/\beta. \end{cases}$$

(d) We have

$$\frac{d}{dt} H_\beta^*(t) \leq \left| \frac{d}{dt} H_\beta^*(t) \right| \leq \min\{|t|, |t|^\beta\}.$$

*Sketch of the proof.* A straightforward calculation gives (a); (c) and (d) follow directly from it. To prove (b), first notice that for every  $\beta \in [0, 1]$  there exists exactly one  $t_0 > 0$  such that  $t_0^2/4 = \frac{1}{\beta} \left( \frac{\beta}{\beta+1} t_0 \right)^{1+\beta}$ . Since

$$\inf_{\beta \in [0, 1]} \frac{t_0 - 1}{t_0^2/4} = \inf_{t \in [2, 4]} \frac{t - 1}{t^2/4} = 3/4$$

and the functions  $t^2$  and  $|t|^{1+\beta}$  are convex, we conclude that

$$H_\beta^*(t) \geq \frac{3}{4} \min \left\{ t^2/4, \frac{1}{\beta} \left( \frac{\beta}{\beta+1} |t| \right)^{1+\beta} \right\} \geq \frac{3}{16} \min \{ t^2, |t|^{1+\beta} \},$$

where we have also used the fact that

$$\inf_{\beta \in [0,1]} \frac{1}{\beta} \left( \frac{\beta}{\beta+1} \right)^{1+\beta} = 1/4. \blacksquare$$

*Proof of Corollary 1.6 for  $\beta > 0$ .* We follow the classical Herbst argument (see e.g. [16]). Denote

$$A = \sup \{ \|\nabla \varphi(x)\| : x \in \mathbb{R}^n \}, \quad B = \sup \{ \|\nabla \varphi(x)\|_{(\beta+1)/\beta} : x \in \mathbb{R}^n \}$$

and for  $\lambda > 0$  define

$$\Phi(\lambda) = \mathbb{E} e^{\lambda \varphi(X_1, \dots, X_n)}.$$

Then

$$\lambda \Phi'(\lambda) = \mathbb{E} \lambda \varphi(X_1, \dots, X_n) e^{\lambda \varphi(X_1, \dots, X_n)}$$

and hence

$$\begin{aligned} \lambda \Phi'(\lambda) - \Phi(\lambda) \log \Phi(\lambda) &= \text{Ent} e^{\lambda \varphi(X_1, \dots, X_n)} \\ &\leq C \mathbb{E} (|\lambda \nabla \varphi(X_1, \dots, X_n)|)^2 \\ &\quad \vee (\lambda \|\nabla \varphi(X_1, \dots, X_n)\|_{(\beta+1)/\beta})^{(\beta+1)/\beta} e^{\lambda \varphi(X_1, \dots, X_n)} \\ &\leq C((A\lambda)^2 \vee (B\lambda)^{(\beta+1)/\beta}) \Phi(\lambda). \end{aligned}$$

After dividing both sides by  $\lambda^2 \Phi(\lambda)$  we can rewrite this as

$$\left( \frac{1}{\lambda} \log \Phi(\lambda) \right)' \leq C((A\lambda)^2 \vee (B\lambda)^{(\beta+1)/\beta}) / \lambda^2.$$

Since the right-hand side is an increasing function of  $\lambda$  and

$$\lim_{\lambda \rightarrow 0^+} \frac{1}{\lambda} \log \Phi(\lambda) = \mathbb{E} \varphi(X_1, \dots, X_n),$$

we deduce from the last inequality that

$$\frac{1}{\lambda} \log \Phi(\lambda) \leq \mathbb{E} \varphi(X_1, \dots, X_n) + C((A\lambda)^2 \vee (B\lambda)^{(\beta+1)/\beta}) / \lambda,$$

which is equivalent to

$$\mathbb{E} e^{\lambda \varphi(X_1, \dots, X_n)} \leq \exp(\lambda \mathbb{E} \varphi(X_1, \dots, X_n) + C((A\lambda)^2 \vee (B\lambda)^{(\beta+1)/\beta})).$$

Therefore from Chebyshev's inequality we get

$$\begin{aligned} \mathbb{P}(\varphi(X_1, \dots, X_n) \geq t + \mathbb{E} \varphi(X_1, \dots, X_n)) \\ &\leq \frac{\mathbb{E} e^{\lambda \varphi(X_1, \dots, X_n)}}{\exp(\lambda \mathbb{E} \varphi(X_1, \dots, X_n) + \lambda t)} \\ &\leq \exp(-\lambda t + C((A\lambda)^2 \vee (B\lambda)^{(\beta+1)/\beta})). \end{aligned}$$

Now we can optimize the right-hand side with respect to  $\lambda$ . Let  $K$  and  $L$  be such that  $A = K^{1/2}L$ ,  $B = K^{\beta/(\beta+1)}L$ . We have

$$(A\lambda)^2 \vee (B\lambda)^{(\beta+1)/\beta} = K((L\lambda)^2 \vee (L\lambda)^{(\beta+1)/\beta}) = KH_\beta(L\lambda)$$

and hence

$$\mathbb{P}(\varphi(X_1, \dots, X_n) \geq t + \mathbb{E}\varphi(X_1, \dots, X_n)) \leq \exp(-CKH_\beta^*(t/CKL)).$$

Using Lemma 2.7 and the definitions of  $K$  and  $L$  we get

$$\begin{aligned} \mathbb{P}(\varphi(X_1, \dots, X_n) \geq t + \mathbb{E}\varphi(X_1, \dots, X_n)) \\ \leq \exp\left(-\frac{3}{16} \min\left\{\frac{t^2}{CA^2}, \frac{t^{1+\beta}}{C^\beta B^{1+\beta}}\right\}\right), \end{aligned}$$

which is the assertion of the lemma. ■

*Proof of Corollary 1.6 for  $\beta = 0$ .* We proceed exactly as in the case  $\beta > 0$ , with the only difference that we have the additional constraint

$$\lambda \leq 1/(2B(m + 3\sigma))$$

since the gradient of the function  $\lambda\varphi$  has to be small. We arrive at

$$\mathbb{P}(\varphi(X_1, \dots, X_n) \geq t + \mathbb{E}\varphi(X_1, \dots, X_n)) \leq \exp(-\lambda t + CA^2\lambda^2).$$

Now we optimize the right-hand side with respect to  $\lambda \in [0, 1/(2B(m+3\sigma))]$ . For  $t/(2CA^2) > 1/(2B(m+3\sigma))$  it is best to take  $\lambda = 1/(2B(m+3\sigma))$ , for which the polynomial becomes

$$\begin{aligned} -\lambda t + CA^2\lambda^2 &= -t/(2B(m+3\sigma)) + CA^2/(2B(m+3\sigma))^2 \\ &\leq -t/(4B(m+3\sigma)) = -\min\{t/(4B(m+3\sigma)), t^2/(4CA^2)\}. \end{aligned}$$

For  $t/(2CA^2) \leq 1/(2B(m+3\sigma))$  we set  $\lambda = t/(2CA^2)$ , which gives

$$\begin{aligned} -\lambda t + CA^2\lambda^2 &= -t^2/(2CA^2) + t^2/(4CA^2) \\ &\leq -t^2/(4CA^2) = -\min\{t/(4B(m+3\sigma)), t^2/(4CA^2)\}. \end{aligned}$$

The assertion follows from those two inequalities. ■

*Proof of Corollary 1.7.* We will consider only the case  $\beta > 0$ ; the case  $\beta = 0$  is similar. The function  $\varphi$  is differentiable almost everywhere (with respect to the Lebesgue measure) and

$$\langle \nabla\varphi(x), h \rangle \leq \varphi(x+h) - \varphi(x) \quad \text{a.e.}$$

After plugging  $h = \nabla\varphi(x)$  and  $h = (|\partial_i\varphi(x)|^{1/\beta} \text{sgn}(\partial_i\varphi(x)))_{i=1}^n$ , and using the Lipschitz conditions, we arrive at

$$\|\nabla\varphi(x)\|_2^2 \leq L_2^2, \quad \|\nabla\varphi(x)\|_{(\beta+1)/\beta}^{1+\beta} \leq L_{1+\beta}^{1+\beta} \quad \text{a.e.}$$

Let  $\varphi_\varepsilon$  be the convolution of  $\varphi$  and a Gaussian kernel, i.e.

$$\varphi_\varepsilon(x) = \mathbb{E}\varphi(x + \sqrt{\varepsilon}G),$$

where  $G \sim \mathcal{N}(0, I)$ . The function  $\varphi_\varepsilon$  is smooth and convex, and inherits

from  $\varphi$  the estimates of the gradient (which are now satisfied for all  $x \in \mathbb{R}^n$ ). Therefore from Corollary 1.6 we get

$$\begin{aligned} \mathbb{P}(\varphi_\varepsilon(X_1, \dots, X_n) \geq t + \mathbb{E}\varphi_\varepsilon(X_1, \dots, X_n)) \\ \leq \exp\left(-\frac{3}{16} \min\left\{\frac{t^2}{CL_2^2}, \frac{t^{1+\beta}}{C^\beta L_{1+\beta}^{1+\beta}}\right\}\right). \end{aligned}$$

Moreover  $|\varphi_\varepsilon(x) - \varphi(x)| \leq L_2\sqrt{\varepsilon}\mathbb{E}\|G\|_2$  and hence  $\varphi_\varepsilon$  converges uniformly to  $\varphi$  as  $\varepsilon$  tends to zero. This observation ends the proof of the corollary. ■

*Proof of Corollary 1.8.* We give a detailed proof only for the case  $\beta > 0$ ; the case  $\beta = 0$  is similar. Throughout the proof we denote by  $C(\beta, m, \sigma)$  the constant from Theorem 1.3 and by  $x_i, i = 1, \dots, n$ , the coordinates of the vector  $x \in \mathbb{R}^n$ . The function  $\Phi(x) = \inf_{a \in A} \sum_{i=1}^n H_\beta^*(x_i - a_i)$  is convex. Therefore, the set

$$\tilde{A} = A + \left\{x \in \mathbb{R}^n : \sum_{i=1}^n H_\beta^*(x_i) < 3r/16\right\} = \{x \in \mathbb{R}^n : \Phi(x) < 3r/16\}$$

is open and convex.

By Rademacher's theorem, the function  $\Phi$  is differentiable almost everywhere (with respect to the Lebesgue measure) and hence we can choose a dense set  $D \subset \tilde{A}$  consisting of points for which  $\nabla\Phi$  exists. We define

$$\varphi(x) = \sup_{y \in D} \{\Phi(y) + \langle \nabla\Phi(y), x - y \rangle\}.$$

This function is convex as a supremum of linear functions. Moreover  $\varphi \leq \Phi$  and  $\varphi = \Phi$  on the set  $\text{cl } \tilde{A} = \{x \in \mathbb{R}^n : \Phi(x) \leq 3r/16\}$ .

Note that the set  $\{x \in \mathbb{R}^n : \varphi(x) < 3r/16\}$  is open and convex, contains  $\tilde{A}$  (on  $\tilde{A}$  we have  $\varphi = \Phi < 3r/16$ ), and is disjoint from  $\text{cl } \tilde{A} \setminus \tilde{A} = \{x \in \mathbb{R}^n : \Phi(x) = 3r/16\}$  (on  $\text{cl } \tilde{A} \setminus \tilde{A}$  we have  $\varphi = \Phi = 3r/16$ ). Therefore we have  $\{x \in \mathbb{R}^n : \varphi(x) < 3r/16\} = \tilde{A}$  and hence

$$\begin{aligned} (2.19) \quad \{x \in \mathbb{R}^n : \varphi(x) < 3r/16\} &= A + \left\{x \in \mathbb{R}^n : \sum_{i=1}^n H_\beta^*(x_i) < 3r/16\right\} \\ &\subset A + \left\{x \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 \wedge |x_i|^{1+\beta} < r\right\} \subset A + r^{1/2}B_2 + r^{1/(1+\beta)}B_{1+\beta}, \end{aligned}$$

where we have used Lemma 2.7 and the observation that if  $\sum_{i=1}^n x_i^2 \wedge |x_i|^{1+\beta} < r$ , then  $x$  can be written as  $y + z$  with  $y \in r^{1/2}B_2$  and  $z \in r^{1/(1+\beta)}B_{1+\beta}$  (since we can set  $y_i = x_i$  if  $x_i^2 < |x_i|^{1+\beta}$ ,  $y_i = 0$  otherwise, and  $z_i = x_i$  if  $x_i^2 \geq |x_i|^{1+\beta}$ ,  $z_i = 0$  otherwise).

We will show that

$$\{x \in \mathbb{R}^n : \varphi(x) < r/16 + \mathbb{E}\varphi(X_1, \dots, X_n)\} \subset \{x \in \mathbb{R}^n : \varphi(x) < 3r/16\}$$

for  $r$  large enough (greater than some constant depending only on  $C(\beta, m, \sigma)$ ), and that

$$|\varphi(x) - \varphi(y)| \leq L_2 \|x - y\|_2, \quad |\varphi(x) - \varphi(y)| \leq L_{1+\beta} \|x - y\|_{1+\beta}$$

with

$$L_2^2 = \sup_{y \in D} \|\nabla \Phi(y)\|_2^2 \leq r, \quad L_{1+\beta}^{1+\beta} = \sup_{y \in D} \|\nabla \Phi(y)\|_{(\beta+1)/\beta}^{1+\beta} \leq r^\beta.$$

By Corollary 1.7 the above claims together with (2.19) will give

$$\mathbb{P}((X_1, \dots, X_n) \notin A + r^{1/2} B_2 + r^{1/(1+\beta)} B_{1+\beta}) \leq \exp(-C'r)$$

for  $r$  large enough and some constant  $C'$  depending only on  $C(\beta, m, \sigma)$  and  $\beta$ . By decreasing  $C'$  we will be able to guarantee that this inequality holds for all  $r > 0$  (recall that  $\mathbb{P}((X_1, \dots, X_n) \in A) \geq 1/2$ ).

We estimate the Lipschitz constants first. For  $y \in D \subset \tilde{A}$  we can choose  $a^y \in \text{cl } A$  such that

$$\Phi(y) = \inf_{a \in A} \sum_{i=1}^n H_\beta^*(y_i - a_i) = \sum_{i=1}^n H_\beta^*(y_i - a_i^y).$$

Note that then  $y - a^y \in \{x \in \mathbb{R}^n : \sum_{i=1}^n H_\beta^*(x_i) \leq 3r/16\}$ . Let  $e^i$  be the  $i$ th vector of the canonical basis in  $\mathbb{R}^n$ . We have

$$\begin{aligned} \Phi(y + te^i) - \Phi(y) &= \inf_{a \in A} \sum_{j=1}^n H_\beta^*(y_j + t1_{\{i=j\}} - a_j) - \sum_{j=1}^n H_\beta^*(y_j - a_j^y) \\ &\leq H_\beta^*(y_i + t - a_i^y) - H_\beta^*(y_i - a_i^y). \end{aligned}$$

The function  $t \mapsto H_\beta^*(y_i + t - a_i^y) - H_\beta^*(y_i - a_i^y) - \Phi(y + te^i) + \Phi(y)$  attains its minimum for  $t = 0$  and hence  $\partial_i \Phi(y) = (H_\beta^*)'(y_i - a_i^y)$ . Using Lemma 2.7 we arrive at  $|\partial_i \Phi(y)| \leq |y_i - a_i^y| \wedge |y_i - a_i^y|^\beta$ . Therefore,

$$\begin{aligned} \|\nabla \Phi(y)\|_2^2 &\leq \sum_{i=1}^n (|y_i - a_i^y| \wedge |y_i - a_i^y|^\beta)^2 \\ &\leq \sum_{i=1}^n |y_i - a_i^y|^2 \wedge |y_i - a_i^y|^{1+\beta} \leq 16 \sum_{i=1}^n H_\beta^*(y_i - a_i^y)/3 \leq r, \end{aligned}$$

and similarly

$$\begin{aligned}
\|\nabla\Phi(y)\|_{(\beta+1)/\beta}^{1+\beta} &\leq \left(\sum_{i=1}^n |y_i - a_i^y|^{(\beta+1)/\beta} \wedge |y_i - a_i^y|^{1+\beta}\right)^\beta \\
&\leq \left(\sum_{i=1}^n |y_i - a_i^y|^2 \wedge |y_i - a_i^y|^{1+\beta}\right)^\beta \\
&\leq \left(16 \sum_{i=1}^n H_\beta^*(y_i - a_i^y)/3\right)^\beta \leq r^\beta.
\end{aligned}$$

This finishes the estimation of the Lipschitz constants.

In order to show that for  $r$  large enough we have

$$\{x \in \mathbb{R}^n : \varphi(x) < r/16 + \mathbb{E}\varphi(X_1, \dots, X_n)\} \subset \{x \in \mathbb{R}^n : \varphi(x) < 3r/16\},$$

we claim that  $\mathbb{E}\varphi(X_1, \dots, X_n) \leq r/8$  for  $r > 2^8 C(\beta, m, \sigma)$ . Indeed, if this were not the case, then we could write

$$\begin{aligned}
1/2 &\leq \mathbb{P}(\varphi(X_1, \dots, X_n) \leq 0) \\
&\leq \mathbb{P}(\varphi(X_1, \dots, X_n) \leq \mathbb{E}\varphi(X_1, \dots, X_n) - r/8) \\
&\leq \mathbb{P}(|\varphi(X_1, \dots, X_n) - \mathbb{E}\varphi(X_1, \dots, X_n)| \geq r/8) \\
&\leq 64 \operatorname{Var} \varphi(X_1, \dots, X_n)/r^2 \leq 64 \cdot 2C(\beta, m, \sigma)/r,
\end{aligned}$$

where the first inequality follows from the fact that on  $A$  we have  $\varphi = \Phi = 0$ , and the last from the fact that  $\mu$  satisfies the Poincaré inequality (see Remark 2.5) and  $\varphi$  is  $\sqrt{r}$ -Lipschitz. This yields a contradiction. ■

**3. A link to transport inequalities.** In this section we will present several equivalent formulations of modified log-Sobolev inequalities for convex functions in terms of transportation inequalities or  $\tau$ -log-Sobolev inequalities, in the spirit of [13, Theorem 6.8]. We emphasize that our arguments in this part of the article follow quite closely the approach in [13] (in fact the authors of [13], while working with the classical convex log-Sobolev inequality, i.e.  $\beta = 1$ , mention the possibility of extending their results to more general cost functions). In particular we do not claim any novelty of the methods we use.

Our main motivation for providing an equivalent formulation of modified log-Sobolev inequalities for all  $\beta \in [0, 1]$  is the hope that it may help find a characterization of measures satisfying the inequalities in question. In fact, one interesting observation which follows from our considerations is that modified log-Sobolev inequalities are equivalent to formally stronger results in which the entropy is replaced by a larger functional (see Remark 3.4).

Since we will rely heavily on techniques developed by other authors, we decided to present only the main steps and ideas of the proofs. We include

numerous references to appropriate sources for the reader interested in all technical details.

Before stating the actual results we have to introduce some notation which we will use throughout this section.

For  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$  and  $\beta \in (0, 1]$  (we will treat the case  $\beta = 0$  later) we define

$$H_\beta^{(n)}(x) = \sum_{i=1}^n H_\beta(x_i), \quad H_\beta^{(n)*}(x) = \sum_{i=1}^n H_\beta^*(x_i),$$

where  $H_\beta^*$  is the Legendre transform of the function  $H_\beta(s) = s^2 \vee |s|^{(\beta+1)/\beta}$  (see Definition 2.6 and Lemma 2.7). For  $t > 0$ ,  $\beta \in (0, 1]$ , and a function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  we define the infimum-convolution operator

$$Q_t f(x) = \inf_{y \in \mathbb{R}^n} \{f(y) + tH_\beta^{(n)*}((x - y)/t)\} \in [-\infty, \infty).$$

Let  $P_1(\mathbb{R}^n)$  denote the class of all probability measures on  $\mathbb{R}^n$  with finite first moment. If  $\mu \in P_1(\mathbb{R}^n)$ , we define the transport cost

$$\bar{\mathcal{T}}(\mu|\nu) = \inf_{\pi} \left\{ \int_{\mathbb{R}^n} H_\beta^{(n)*} \left( x - \int_{\mathbb{R}^n} y dp_x(y) \right) d\nu(x) \right\},$$

where the infimum is taken over all probability measures  $\pi$  on  $\mathbb{R}^n \times \mathbb{R}^n$  with the first marginal  $\nu$  and the second  $\mu$ ; here  $p_x$  denotes the probability kernel such that  $\pi(dx dy) = \nu(dx)p_x(dy)$ .

If  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  is bounded below by an affine function, we define for  $\lambda > 0$ ,

$$R^\lambda f(x) = \inf_p \left\{ \int_{\mathbb{R}^n} f(y) dp(y) + \lambda H_\beta^{(n)*} \left( x - \int_{\mathbb{R}^n} y dp(y) \right) \right\} \in [-\infty, \infty)$$

where the infimum is taken over all probability measures  $p \in P_1(\mathbb{R}^n)$ . Note that  $R^\lambda f \leq f$ .

Recall also that for two probability measures  $\mu, \nu$  on  $\mathbb{R}^n$  the relative entropy of  $\nu$  with respect to  $\mu$  is given by the formula

$$H(\nu|\mu) = \int_{\mathbb{R}^n} \log \left( \frac{d\nu}{d\mu} \right) d\nu$$

if  $\nu$  is absolutely continuous with respect to  $\mu$ ; otherwise we set  $H(\nu|\mu) = +\infty$ .

The main result of this section provides a characterization of the modified log-Sobolev inequalities with  $\beta \in (0, 1]$  for convex functions in terms of transportation inequalities.

**PROPOSITION 3.1.** *Let  $X$  be an integrable random vector with values in  $\mathbb{R}^n$  and law  $\mu$ . For  $\beta \in (0, 1]$  the following conditions are equivalent:*

(i) There exists  $b > 0$  such that for all probability measures  $\nu$  on  $\mathbb{R}^n$ ,

$$\overline{\mathcal{T}}(\mu|\nu) \leq bH(\nu|\mu).$$

(ii) For all  $s > 0$  we have  $\mathbb{E}e^{s|X|} < \infty$  and there exist  $\lambda, D > 0$  such that

$$(3.1) \quad \text{Ent } e^{f(X)} \leq D\mathbb{E}(f(X) - R^\lambda f(X))e^{f(X)}$$

for all functions  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  such that  $\mathbb{E}f(X)e^{f(X)} < \infty$  and  $f$  is bounded below by an affine function.

(iii) For all  $s > 0$  we have  $\mathbb{E}e^{s|X|} < \infty$  and there exists  $C > 0$  such that

$$(3.2) \quad \text{Ent } e^{\varphi(X)} \leq C\mathbb{E}H_\beta^{(n)}(\nabla\varphi(X))e^{\varphi(X)}$$

for all smooth convex Lipschitz functions  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$ .

(iv) For all  $s > 0$  we have  $\mathbb{E}e^{s|X|} < \infty$  and there exists  $B > 0$  such that for every  $t_0 \in [0, B]$ , every  $t > 0$ , and every convex Lipschitz function  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$  bounded from below,

$$(3.3) \quad \|e^{Q_t\varphi(X)}\|_{k(t)} \leq \|e^{\varphi(X)}\|_{k(0)},$$

where  $k(t) = (1 + B^{-1}(t - t_0))1_{\{t \leq t_0\}} + (1 + B^{-1}\beta^{-1}(t - t_0))^\beta 1_{\{t > t_0\}}$ . Here, for a positive random variable  $Z$ , we use the notation  $\|Z\|_0 = \exp(\mathbb{E} \log Z)$  and  $\|Z\|_k = (\mathbb{E}Z^k)^{1/k}$  if  $k \neq 0$ .

Moreover, in each of the above implications the constants in the conclusion depend only on the constants in the premise (in particular they do not depend on the dimension).

REMARK 3.2. Using the terminology and notation introduced in [13], the above proposition can be stated more briefly: (i) means that the transport inequality  $\overline{\mathbf{T}}_c^-(b)$  holds (for the cost function  $c(x, p) = H_\beta^{(n)*}(x - \int_{\mathbb{R}^n} y dp(y))$ ; see [13, Definition 5.1]), and (ii) is the  $\tau$ -log-Sobolev inequality, denoted  $(\tau)$ -**LSI** $_c(\lambda, D)$  (for the same cost function; see [13, Definition 6.1]).

REMARK 3.3. The exponential integrability of the norm in items (ii)–(iv) is introduced to exclude heavy tailed measures for which the only exponentially integrable convex functions are constants. Using the observation that  $R^\lambda f(x) = \inf_y \{\varphi(y) + \lambda H_\beta^{(n)*}(x - y)\}$ , where  $\varphi$  is the greatest convex minorant of  $f$  (see [13, Corollary 3.11(2)]), one can easily see that such measures trivially satisfy (3.1) with  $D = 1$  and arbitrary  $\lambda > 0$ , while they cannot satisfy the inequality of point (i), as it implies subexponential concentration.

REMARK 3.4. As we will see from the proof (see (3.5)), the conditions of the proposition remain equivalent if we replace the entropy on the left-hand side of the inequalities in (ii) and (iii) by the (greater) expression

$$\int_{\mathbb{R}^n} \varphi e^\varphi d\mu - \int_{\mathbb{R}^n} \varphi d\mu \int_{\mathbb{R}^n} e^\varphi d\mu = \frac{1}{2}\mathbb{E}(\varphi(X) - \varphi(Y))(e^{\varphi(X)} - e^{\varphi(Y)})$$

(where  $X, Y$  are independent with law  $\mu$ ). Therefore the first estimate from the proof of Theorem 1.3 is actually quite sharp. Note that there does not exist a finite constant  $C$  such that  $C \text{Ent} e^X \geq \mathbb{E}(X - Y)(e^X - e^Y)$  for all bounded i.i.d. random variables  $X$  and  $Y$  (as one can see by considering e.g.  $X$  with the distribution  $\frac{1}{2}\delta_{-a} + \frac{1}{2}\delta_a$  for  $a$  sufficiently large).

REMARK 3.5. In the proofs we will use a dual formulation of (i): the transport cost inequality holds if and only if

$$(3.4) \quad \left( \int_{\mathbb{R}^n} e^{b^{-1}Q_1\varphi(x)} d\mu(x) \right)^b e^{-\int_{\mathbb{R}^n} \varphi(x) d\mu(x)} \leq 1$$

for all convex Lipschitz functions  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$  bounded from below (see [13, Proposition 5.5(ii'') and Corollary 3.11(2)]; note also that  $Q_1\varphi = R^1\varphi$  since  $\varphi$  is convex).

For  $\beta = 0$  a result similar to Proposition 3.1 holds, but we have to introduce a slightly different notation which will additionally depend on a number  $\delta > 0$ . Namely, for  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$  we define

$$H_{0,\delta}^{(n)}(x) = \sum_{i=1}^n H_{0,\delta}(x_i), \quad H_{0,\delta}^{(n)*}(x) = \sum_{i=1}^n H_{0,\delta}^*(x_i),$$

where  $H_{0,\delta}(s) = H_0(s/\delta) = \delta^{-2}s^2 1_{\{|s| \leq \delta\}} + \infty 1_{\{|s| > \delta\}}$  and  $H_{0,\delta}^*(s) = H_0^*(s\delta) = \frac{1}{4}\delta^2 s^2 1_{\{|\delta|s| \leq 2\}} + (\delta|s| - 1) 1_{\{|\delta|s| > 2\}}$  is its Legendre transform. We define  $Q_{t,\delta}, \overline{T}_\delta, R_\delta^\lambda$  as  $Q_t, \overline{T}, R^\lambda$  above, but with  $H_{0,\delta}^{(n)}$  and  $H_{0,\delta}^{(n)*}$  in place of  $H_\beta^{(n)}$  and  $H_\beta^{(n)*}$  respectively, i.e. we set

$$\begin{aligned} Q_{t,\delta} f(x) &= \inf_{y \in \mathbb{R}^n} \{f(y) + tH_{0,\delta}^{(n)*}((x-y)/t)\}, \\ \overline{T}_\delta(\mu|\nu) &= \inf_{\pi} \left\{ \int_{\mathbb{R}^n} H_{0,\delta}^{(n)*} \left( x - \int_{\mathbb{R}^n} y dp_x(y) \right) d\nu(x) \right\}, \\ R_\delta^\lambda f(x) &= \inf_{p \in P_1(\mathbb{R}^n)} \left\{ \int_{\mathbb{R}^n} f(y) dp(y) + \lambda H_{0,\delta}^{(n)*} \left( x - \int_{\mathbb{R}^n} y dp(y) \right) \right\}. \end{aligned}$$

We then have the following

PROPOSITION 3.6. *Let  $X$  be an integrable random vector with values in  $\mathbb{R}^n$  and law  $\mu$ . The following conditions are equivalent:*

- (i) *There exist  $\delta_1$  and  $b > 0$  such that for all probability measures  $\nu$  on  $\mathbb{R}^n$ ,*

$$\overline{T}_{\delta_1}(\mu|\nu) \leq bH(\nu|\mu).$$

- (ii) *There exist  $\delta_2 > 0, \lambda, D > 0$  such that  $\mathbb{E}e^{D^{-1}H_{0,\delta_2}^{(n)*}(X)} < \infty$  and*

$$\text{Ent} e^{f(X)} \leq D\mathbb{E}(f(X) - R_{\delta_2}^\lambda f(X))e^{f(X)}$$

for all functions  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  such that  $\mathbb{E}f(X)e^{f(X)} < \infty$  and  $f$  is bounded below by an affine function.

(iii) There exist  $\delta_3, C > 0$  such that  $\mathbb{E}e^{C^{-1}H_{0,\delta_3}^{(n)*}(X)} < \infty$  and

$$\text{Ent} e^{\varphi(X)} \leq C \mathbb{E} H_{0,\delta_3}^{(n)}(\nabla\varphi(X)) e^{\varphi(X)}$$

for all smooth convex Lipschitz functions  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$  with  $\|\nabla\varphi\|_\infty \leq \delta_3$ .

(iv) There exist  $\delta_4, B > 0$  such that

$$\mathbb{E}e^{B^{-1}H_{0,\delta_4}^{(n)*}(X)} < \infty$$

and for every  $t_0 \in [0, B]$ , every  $t > 0$ , and every convex Lipschitz function  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$  bounded from below,

$$\|e^{Q_{t,\delta_4}\varphi(X)}\|_{k(t)} \leq \|e^{\varphi(X)}\|_{k(0)},$$

where  $k(t) = B^{-1} \min\{1 + B^{-1}(t - t_0), 1\}$ .

Moreover, in each of the above implications the constants in the conclusion depend only on the constants in the premise (in particular they do not depend on the dimension).

REMARK 3.7. The dependence of the constants is the following: (i) implies (ii) with  $\lambda \in (0, 1/b)$ ,  $D = 1/(1 - \lambda b)$ , and  $\delta_2 = \delta_1$ ; (ii) implies (iii) with  $C = \lambda D$  and  $\delta_3 = \delta_2(\lambda \wedge (2D)^{-1})$ ; (iii) implies (iv) with  $B = C \vee 1$  and  $\delta_4 = \delta_3$ ; (iv) implies (i) with  $b = B$  and  $\delta_1 = \delta_4$ . Note also that for  $a \geq 1$  we have  $H_{0,a\delta} \leq H_{0,\delta}$  and  $H_{0,\delta}^* \leq H_{0,a\delta}^* \leq a^2 H_{0,\delta}^*$ , and hence if the inequalities in (i)–(iv) hold for some  $\delta > 0$ , then they also hold for every  $\delta' \in (0, \delta)$ .

*Sketch of the proof of Proposition 3.1.* For simplicity we assume  $n = 1$ . To prove that (i) implies (ii) we follow the proof of [13, Proposition 6.3]. Fix  $\lambda \in (0, 1/b)$  and let  $f: \mathbb{R} \rightarrow \mathbb{R}$  be a function with  $\int_{\mathbb{R}} f e^f d\mu < \infty$ . We define  $\nu_f$  to be the measure with density  $e^f d\mu / (\int_{\mathbb{R}} e^f d\mu)$ . Let  $\pi$  be a measure on  $\mathbb{R}^2$  with the first marginal  $\nu_f$  and the second  $\mu$ . We have  $\pi(dxdy) = \nu_f(dx)p_x(dy)$  for some probability kernel  $p_x$  and hence

$$\begin{aligned} \int_{\mathbb{R}} f d\nu_f - \int_{\mathbb{R}} f d\mu &= \int_{\mathbb{R}^2} (f(x) - f(y)) \pi(dxdy) \\ &= \int_{\mathbb{R}} \left( f(x) - \int_{\mathbb{R}} f(y) dp_x(y) \right) d\nu_f(x) \\ &\leq \int_{\mathbb{R}} (f(x) - R^\lambda f(x)) d\nu_f(x) \\ &\quad + \lambda \int_{\mathbb{R}} H_\beta^* \left( x - \int_{\mathbb{R}} y dp_x(y) \right) d\nu_f(x), \end{aligned}$$

where the last inequality follows from the definition of  $R^\lambda$ . Optimizing over all measures  $\pi$  as above gives us

$$\begin{aligned} \int_{\mathbb{R}} f d\nu_f - \int_{\mathbb{R}} f d\mu &\leq \int_{\mathbb{R}} (f(x) - R^\lambda f(x)) d\nu_f(x) + \lambda \overline{T}(\mu|\nu_f) \\ &\leq \int_{\mathbb{R}} (f(x) - R^\lambda f(x)) d\nu_f(x) + \lambda b H(\nu_f|\mu). \end{aligned}$$

It follows from Jensen's inequality that

$$H(\nu_f|\mu) = \int_{\mathbb{R}} \log\left(\frac{e^f}{\int_{\mathbb{R}} e^f d\mu}\right) d\nu_f \leq \int_{\mathbb{R}} f d\nu_f - \int_{\mathbb{R}} f d\mu$$

and hence we conclude from the preceding inequality that

$$\int_{\mathbb{R}} f d\nu_f - \int_{\mathbb{R}} f d\mu \leq \frac{1}{1 - \lambda b} \int_{\mathbb{R}} (f(x) - R^\lambda f(x)) d\nu_f(x).$$

By the definition of  $\nu_f$  this means that

$$(3.5) \quad \int_{\mathbb{R}} f e^f d\mu - \int_{\mathbb{R}} f d\mu \int_{\mathbb{R}} e^f d\mu \leq \frac{1}{1 - \lambda b} \int_{\mathbb{R}} (f(x) - R^\lambda f(x)) e^f d\mu(x).$$

Therefore (3.1) is satisfied with  $D = 1/(1 - \lambda b)$  (the above inequality is formally stronger, since its left-hand side is by Jensen's inequality greater than  $\text{Ent } e^f$ ; see Remark 3.4).

It remains to prove the exponential integrability of the norm (we present the argument on the real line, however it can be easily adapted to  $\mathbb{R}^n$ , e.g. if we decide to work with the  $\ell_1^n$  norm all the calculations are performed on each coordinate separately; clearly the choice of the norm for this problem is irrelevant).

Consider the function  $\varphi(x) = r|x|$ . By the dual formulation of the transport cost inequality (see Remark 3.5) and the assumption that  $X$  is integrable, we get  $\mathbb{E}e^{b^{-1}Q_1\varphi(X)} < \infty$ . It is easy to see that for  $|x|$  large enough (depending on  $r$ ) the infimum in the definition of  $Q_1\varphi(x)$  is attained for  $y = x - \text{sgn}(x)((H_\beta^*)')^{-1}(r)$ , where  $((H_\beta^*)')^{-1}$  is the inverse of  $(H_\beta^*)'$  restricted to  $[0, \infty)$ . Thus for  $|x|$  large enough, we have  $Q_1\varphi(x) \geq r|x|/2$ . Since  $r$  is arbitrary, we conclude that  $\mathbb{E}e^{s|X|} < \infty$  for all  $s$ .

To prove that (ii) implies (iii) note that if  $\varphi$  is convex, then

$$\begin{aligned} \varphi(x) - R^\lambda\varphi(x) &= \sup_p \left\{ \int_{\mathbb{R}} (\varphi(x) - \varphi(y)) dp(y) - \lambda H_\beta^* \left( x - \int_{\mathbb{R}} y dp(y) \right) \right\} \\ &\leq \sup_p \left\{ \int_{\mathbb{R}} \varphi'(x)(x - y) dp(y) - \lambda H_\beta^* \left( x - \int_{\mathbb{R}} y dp(y) \right) \right\} \\ &\leq \sup_p \left\{ \varphi'(x) \cdot \left( x - \int_{\mathbb{R}} y dp(y) \right) - \lambda H_\beta^* \left( x - \int_{\mathbb{R}} y dp(y) \right) \right\}. \end{aligned}$$

We have

$$uv \leq CH_\beta(u) + \lambda H_\beta^*(v)$$

for some constant  $C$  depending only on  $\lambda$  and  $\beta$ . Therefore

$$\varphi(x) - R^\lambda \varphi(x) \leq CH_\beta(\varphi'(x))$$

and hence (iii) holds.

To see that (iii) implies (iv) (with  $B = C$ ) we follow the reasoning from the proof of [10, Theorem 1.11]. By a perturbation argument we can assume that  $\mu$  is absolutely continuous. Indeed, if  $\gamma$  is a Gaussian measure on  $\mathbb{R}^n$  whose covariance matrix is a sufficiently small multiple of identity, then the product measure  $\mu \otimes \gamma$  on  $\mathbb{R}^n \times \mathbb{R}^n$  satisfies the modified log-Sobolev inequality with constant  $C$ . Let  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a smooth convex Lipschitz function and let  $\Phi: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  be defined by  $\Phi(x, y) = \varphi(x + \varepsilon y)$ ,  $x, y \in \mathbb{R}$ . Applying the modified log-Sobolev inequality to the measure  $\mu \otimes \gamma$  and the function  $\Phi$ , we see that the convolution  $\mu * \gamma_\varepsilon$ , where  $\gamma_\varepsilon(A) = \gamma(A/\varepsilon)$ , satisfies the modified log-Sobolev inequality on  $\mathbb{R}^n$  with constant  $C_\varepsilon \rightarrow C$  as  $\varepsilon \rightarrow 0$ . If one can prove that (3.3) is satisfied by  $\mu * \gamma_\varepsilon$  with  $B_\varepsilon = C_\varepsilon$ , then one can obtain it for  $\mu$  with  $B = C$  by the Lebesgue dominated convergence theorem. Note that if  $\mu$  is absolutely continuous, then another standard approximation shows that (iii) holds for all convex Lipschitz functions (by the Rademacher theorem the gradient is then almost surely well defined).

If  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  is a convex function, then so is  $Q_t \varphi$  and it satisfies the Hamilton–Jacobi equation

$$(3.6) \quad \partial_t Q_t \varphi + H_\beta(\partial_x Q_t \varphi) = 0$$

(see [6, Chapter 3.3.2]). More precisely, the function  $(t, x) \mapsto Q_t \varphi(x)$  is Lipschitz and the equation is satisfied a.e. with respect to the Lebesgue measure on  $(0, \infty) \times \mathbb{R}^n$ . We fix  $t_0 \in [0, C]$ , define

$$k(t) = (1 + C^{-1}(t - t_0))1_{\{t \leq t_0\}} + (1 + C^{-1}\beta^{-1}(t - t_0))^\beta 1_{\{t > t_0\}},$$

and set

$$F(t) = \int_{\mathbb{R}} e^{k(t)Q_t \varphi(x)} d\mu(x) \quad \text{for } t > 0.$$

Using the absolute continuity of  $\mu$  together with integrability properties of Lipschitz functions, one can show that  $F$  is locally Lipschitz,  $F'(t)$  exists for almost all  $t > 0$ , and

$$\begin{aligned}
k(t)F'(t) &= k(t) \int_{\mathbb{R}} e^{k(t)Q_t\varphi(x)} (k'(t)Q_t\varphi(x) + k(t)\partial_t Q_t\varphi(x)) d\mu(x) \\
&= k(t) \int_{\mathbb{R}} e^{k(t)Q_t\varphi(x)} (k'(t)Q_t\varphi(x) - k(t)H_\beta(\partial_x Q_t\varphi(x))) d\mu(x) \\
&= k'(t)F(t) \log F(t) + k'(t) \text{Ent} e^{k(t)Q_t\varphi(X)} \\
&\quad - k(t)^2 \int_{\mathbb{R}} e^{k(t)Q_t\varphi(x)} H_\beta(\partial_x Q_t\varphi(x)) d\mu(x) \\
&\leq k'(t)F(t) \log F(t) + Ck'(t) \int_{\mathbb{R}} e^{k(t)Q_t\varphi(x)} H_\beta(k(t)\partial_x Q_t\varphi(x)) d\mu(x) \\
&\quad - k(t)^2 \int_{\mathbb{R}} e^{k(t)Q_t\varphi(x)} H_\beta(\partial_x Q_t\varphi(x)) d\mu(x) \\
&\leq k'(t)F(t) \log F(t) \\
&\quad + \int_{\mathbb{R}} e^{k(t)Q_t\varphi(x)} H_\beta(\partial_x Q_t\varphi(x)) d\mu(x) \\
&\quad \cdot (Ck'(t) \max\{|k(t)|^2, |k(t)|^{(\beta+1)/\beta}\} - k(t)^2) \\
&= k'(t)F(t) \log F(t),
\end{aligned}$$

where we have used (3.6), the inequality (3.2), the fact that

$$H_\beta(ax) \leq \max\{|a|^2, |a|^{(\beta+1)/\beta}\} H_\beta(x),$$

and the definition of  $k(t)$  in the last equality. The above differential inequality is equivalent to  $(\log(F(t))/k(t))' \leq 0$  for almost all  $t > 0$ . Since  $Q_t\varphi \leq \varphi$ , we arrive at

$$\frac{\log F(t)}{k(t)} \leq \liminf_{s \rightarrow 0^+} \frac{\log F(s)}{k(s)} \leq \lim_{s \rightarrow 0^+} \frac{\log(\int_{\mathbb{R}} e^{k(s)\varphi(x)} d\mu(x))}{k(s)} = \log \|e^{\varphi(X)}\|_{k(0)},$$

which is equivalent to the assertion.

Finally, (iv) implies (i) (with  $b = B \vee 1$ ). Indeed, we plug  $t_0 = B$  and  $t = B \wedge 1 \leq 1$  (so that  $k(t) = (B \vee 1)^{-1}$  and  $k(0) = 0$ ) into inequality (3.3), and using the fact that  $Q_t\varphi$  decreases with  $t$  we arrive at the dual formulation of the transport cost inequality (see Remark 3.5). This ends the proof of the last implication. ■

The proof in the case  $\beta = 0$  follows roughly the same lines. Below we sum up the most important changes one has to make.

*Sketch of the proof of Proposition 3.6.* For simplicity we assume  $n = 1$ . To prove that (i) implies the entropy bound of (ii) (with  $\delta_2 = \delta_1$ ) we follow exactly the same reasoning as in the proof of Proposition 3.1. As for the exponential integrability of  $H_{0,\delta_1}^{(n)*}$ , we again consider  $\varphi(x) = r|x|$  and note that by the dual formulation of (i),  $\mathbb{E}e^{b^{-1}Q_{1,\delta_1}\varphi(X)} < \infty$ . But, as one can easily

see, the boundedness of the derivative of  $H_{0,\delta_1}^*$  implies that for  $r$  sufficiently large,

$$Q_{1,\delta_1}\varphi(x) = \inf_y \{r|y| + H_{0,\delta_1}^*(x-y)\} = H_{0,\delta_1}^*(x),$$

which shows that  $\mathbb{E}e^{b^{-1}H_{0,\delta_1}^*(X)} < \infty$ .

To prove that (ii) implies (iii) with  $\delta_3 = \delta_2(\lambda \wedge (2D))^{-1}$  and  $C = \lambda D$ , notice that if  $\varphi$  is convex and  $|\varphi'| \leq \delta_3$ , then  $\mathbb{E}\varphi(X)e^{\varphi(X)} < \infty$  and

$$\varphi(x) - R_{\delta_2}^\lambda \varphi(x) \leq \sup_p \left\{ \varphi'(x) \cdot \left( x - \int_{\mathbb{R}} y dp(y) \right) - \lambda H_{0,\delta_2}^* \left( x - \int_{\mathbb{R}} y dp(y) \right) \right\}.$$

We have

$$\begin{aligned} uv &\leq (\lambda H_{0,\delta_2}^*)^*(u) + \lambda H_{0,\delta_2}^*(v) = \lambda H_{0,\delta_2}(u/\lambda) + \lambda H_{0,\delta_2}^*(v) \\ &= \lambda H_{0,\lambda\delta_2}(u) + \lambda H_{0,\delta_2}^*(v) \leq \lambda H_{0,\delta_3}(u) + \lambda H_{0,\delta_2}^*(v). \end{aligned}$$

Therefore  $\varphi(x) - R_{\delta_2}^\lambda \varphi(x) \leq \lambda H_{0,\delta_3}(\varphi'(x))$  and hence (iii) holds.

As in the case  $\beta > 0$ , for the proof that (iii) implies (iv) (with  $\delta_4 = \delta_3$  and  $B = C \vee 1$ ) we can assume that  $\mu$  is absolutely continuous. Thus, by standard approximation arguments, the log-Sobolev inequality of (iii) is satisfied for all convex Lipschitz functions (the gradient is then almost surely well defined). In what follows, without loss of generality, we will also assume that  $C \geq 1$ .

Note that

$$|\partial_x t H_{0,\delta_3}^*((x-y)/t)| = |(H_{0,\delta_3}^*)'((x-y)/t)| \leq \delta_3,$$

so for  $t > 0$ ,  $x \mapsto Q_{t,\delta_3}\varphi(x)$  is  $\delta_3$ -Lipschitz. An adaptation of the reasoning from [6, Chapter 3.3.2] shows that  $(t, x) \mapsto Q_{t,\delta_3}\varphi(x)$  is Lipschitz and

$$\partial_t Q_{t,\delta_3}\varphi + H_{0,\delta_3}(\partial_x Q_{t,\delta_3}\varphi) = \partial_t Q_{t,\delta_3}\varphi + \delta_3^{-2}(\partial_x Q_{t,\delta_3}\varphi)^2 = 0$$

Lebesgue a.e. on  $(0, \infty) \times \mathbb{R}$  (we remark that for  $n > 1$  one needs to work with the  $\ell_1^n$  rather than the Euclidean norm to deduce that  $t \mapsto Q_{t,\delta_3}\varphi(x)$  is Lipschitz).

Next, one shows that for any  $\tilde{C} \geq C$ , the function

$$F(t) = \int_{\mathbb{R}} e^{\tilde{k}(t)Q_{t,\delta_3}\varphi(x)} d\mu(x), \quad t > 0,$$

is well defined, where we have denoted

$$\tilde{k}(t) = \tilde{C}^{-1} \min\{1 + \tilde{C}^{-1}(t - t_0), 1\},$$

and  $t_0 \in [0, \tilde{C}]$  is a fixed number. Indeed,

$$\tilde{k}(t)Q_{t,\delta_3}\varphi(x) \leq \tilde{k}(t)tH_{0,\delta_3}^*(x/t) + \tilde{k}(t)\varphi(0) \leq \tilde{k}(t)\delta_3|x| + \tilde{k}(t)\varphi(0).$$

For  $|x| > 2/\delta_3$  the right-hand side above is bounded by  $\tilde{C}^{-1}(H_{0,\delta_3}^*(x) + 1 + \varphi(0))$ , which together with the assumed exponential integrability of  $H_{0,\delta_3}^*$  shows that the integral defining  $F$  is finite.

Using absolute continuity of  $F$  and again exponential integrability of  $H_{0,\delta_3}^*$  one proves that for  $\tilde{C} > C$ ,  $F$  is differentiable a.e. Moreover, the same argument as for  $\beta > 0$  shows that  $\tilde{k}(t)F'(t) \leq \tilde{k}'(t)F(t) \log F(t)$  for almost all  $t > 0$ . Let us list the additional observations needed to carry out the calculations. First,

$$|\partial_x \tilde{k}(t) Q_{t,\delta_3} \varphi| \leq |\partial_x Q_{t,\delta_3} \varphi| \leq \delta_3,$$

which allows us to use the modified log-Sobolev inequality. Second,

$$H_{0,\delta_3}(\tilde{k}(t) \partial_x Q_{t,\delta_3} \varphi(x)) = \tilde{k}(t)^2 H_{0,\delta_3}(\partial_x Q_{t,\delta_3} \varphi(x)).$$

Finally,

$$\tilde{C} \tilde{k}'(t) \tilde{k}(t)^2 \leq \tilde{k}(t)^2.$$

We remark that to verify the above properties one uses the assumption that  $C \geq 1$ .

Integration of the differential inequality and taking the limit  $\tilde{C} \rightarrow C$  ends the proof of the implication.

Finally, (iv) implies (i) (with  $b = B$  and  $\delta_1 = \delta_4$ ) as in the case  $\beta > 0$ : it is enough to check that

$$\left( \int_{\mathbb{R}} e^{B^{-1} Q_{1,\delta_3} \varphi(x)} d\mu(x) \right)^B e^{-\int_{\mathbb{R}} \varphi(x) d\mu(x)} \leq 1$$

for all convex Lipschitz functions  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  bounded from below (see Remark 3.5). This is easy to recover from the hypercontractive inequality in (iv). If  $B < 1$  one sets  $t_0 = t = B$  and uses the inequality  $Q_{1,\delta_3} \varphi \leq Q_{t,\delta_3} \varphi$ , otherwise one sets  $t = t_0 = 1$ . ■

REMARK 3.8. As mentioned in the Introduction, very recently Feldheim et al. [7] proved that in the case of symmetric measures on  $\mathbb{R}$ , the condition  $\mu \in \mathcal{M}_0(m, \sigma)$  for some  $m, \sigma$  is equivalent to the infimum-convolution inequality

$$\int_{\mathbb{R}} e^{Q_{1,\delta} \varphi(x)} d\mu(x) \int_{\mathbb{R}} e^{-\varphi(x)} d\mu(x) \leq 1$$

for some  $\delta$ , which is a formally stronger property than the inequality (i) of Proposition 3.6 (see Remark 3.5). Thus for  $\beta = 0$  the modified log-Sobolev inequality is in fact equivalent to a stronger concentration property. Gozlan et al. [12] extended the characterization of [7] to much more general transport costs, including all the cases considered by us. It would be interesting to provide an intrinsic characterization also for measures on the line satisfying general modified log-Sobolev inequalities. One expects that for  $\beta \in (0, 1]$  such measures form a strictly larger class than the class of measures satisfying the inequality of [12].

**4. A remark about the nonconvex setting.** As noted in Remark 2.5, a Borel probability measures satisfying inequality (1.4) for all smooth convex functions (with bounded derivatives) also satisfies the Poincaré inequality for all smooth convex functions. It follows from [4, Theorem 1.4] that in this case  $\mu \in \mathcal{M}_0(m, \sigma)$  for some  $m$  and  $\sigma$ .

We will now show that sufficiently nice Borel probability measures  $\mu$  which satisfy the modified logarithmic Sobolev inequality (1.3) for all (not necessarily log-convex) smooth functions belong to the class  $\mathcal{M}_\beta(m, \sigma^{\beta+1})$  for some  $m$  and  $\sigma$ . More precisely, for  $\beta > 0$  we have the following result.

**PROPOSITION 4.1.** *Let  $\mu$  be a Borel probability measure on  $\mathbb{R}$ , absolutely continuous with respect to the Lebesgue measure. Suppose there exist constants  $C < \infty$  and  $\beta \in (0, 1]$  such that for all smooth functions  $f$ ,*

$$\text{Ent } f(X)^2 \leq C \mathbb{E}(f'(X)^2 \vee |f'(X)|^{(\beta+1)/\beta}),$$

where  $X$  is a random variable with law  $\mu$ . Denote by  $M$  the median and by  $n$  the density of  $\mu$ , and assume also that

$$\frac{1}{\beta} n(x)^{-\beta} \geq \varepsilon \int_{x \wedge M}^{x \vee M} n(t)^{-\beta} dt$$

for some  $\varepsilon > 0$  and all  $x \neq M$ . Then there exist constants  $m, \sigma < \infty$  such that  $\mu \in \mathcal{M}_\beta(m, \sigma^{\beta+1})$ .

*Proof.* It follows from the Barthe–Roberto criterion [2, Theorem 10] that for some constant  $K$  and all  $y \geq M$ ,

$$(4.1) \quad \mu([y, \infty)) \log \frac{1}{\mu([y, \infty))} \left( \int_M^y n(t)^{-\beta} dt \right)^{1/\beta} \leq K.$$

Hölder's inequality (with exponents  $\beta + 1$  and  $(\beta + 1)/\beta$ ) gives

$$\begin{aligned} \frac{1}{x^\beta} &= \int_x^{x+1/x^\beta} n(t)^{-\beta/(\beta+1)} n(t)^{\beta/(\beta+1)} dt \\ &\leq \left( \int_x^{x+1/x^\beta} n(t)^{-\beta} dt \right)^{1/(\beta+1)} \left( \int_x^{x+1/x^\beta} n(t) dt \right)^{\beta/(\beta+1)}, \end{aligned}$$

and therefore

$$\frac{1}{x^{1+\beta}} \leq \left( \int_x^{x+1/x^\beta} n(t)^{-\beta} dt \right)^{1/\beta} \mu([x, x + 1/x^\beta]).$$

Since  $\mu$  has the concentration property of order  $1 + \beta$ , using (4.1) for  $y = x + 1/x^\beta$  we get

$$\begin{aligned} \frac{D}{K} x^{1+\beta} \mu([x + 1/x^\beta, \infty)) &\leq \left( \int_M^{x+1/x^\beta} n(t)^{-\beta} dt \right)^{-1/\beta} \\ &\leq x^{1+\beta} \mu([x, x + 1/x^\beta)) \end{aligned}$$

for some constant  $D$  and  $x$  large enough. Hence, for  $x$  large enough,

$$\mu([x + 1/x^\beta, \infty)) \leq \frac{1}{1 + D/K} \mu([x + 1/x^\beta, \infty)).$$

Since we can deal analogously with the left tail, the claim follows from Proposition 1.2. ■

REMARK 4.2. As proved in [1], in the case of  $\beta = 1$  we do not need the assumption of absolute continuity, as in this case the Barthe–Roberto criterion can be replaced by a more general Bobkov–Götze criterion [3], in which one does not require the existence of a density.

**Acknowledgements.** We would like to thank Ivan Gentil and Arnaud Guillin who back in 2006 during a trimester Phenomena in High Dimensions held at the IHP in Paris introduced the first-named author to their theory of modified log-Sobolev inequalities and raised a question concerning their convex counterparts. Some ideas used in the proof of the  $\beta = 0$  case of our results originated in those conversations. We also thank Piotr Nayar for communicating to us the recent results of [7]. Finally, we thank the anonymous referee for a very careful reading of the first version of this article, for pointing out that it is possible to obtain hypercontractive estimates in Section 3, and for many helpful comments.

Research partially supported by the NCN grant no. 2012/05/B/ST1/00412 (R.A.).

## References

- [1] R. Adamczak, *Logarithmic Sobolev inequalities and concentration of measure for convex functions and polynomial chaoses*, Bull. Polish Acad. Sci. Math. 53 (2005), 221–238.
- [2] F. Barthe and C. Roberto, *Modified logarithmic Sobolev inequalities on  $\mathbb{R}$* , Potential Anal. 29 (2008), 167–193.
- [3] S. G. Bobkov and F. Götze, *Exponential integrability and transportation cost related to logarithmic Sobolev inequalities*, J. Funct. Anal. 163 (1999), 1–28.
- [4] S. G. Bobkov and F. Götze, *Discrete isoperimetric and Poincaré-type inequalities*, Probab. Theory Related Fields 114 (1999), 245–277.
- [5] 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.
- [6] L. C. Evans, *Partial Differential Equations*, 2nd ed., Grad. Stud. Math. 19, Amer. Math. Soc., Providence, RI, 2010.

- [7] N. Feldheim, A. Marsiglietti, P. Nayar, and J. Wang, *A note on the convex infimum convolution inequality*, arXiv:1505.00240 (2015).
- [8] I. Gentil, A. Guillin, and L. Miclo, *Modified logarithmic Sobolev inequalities and transportation inequalities*, Probab. Theory Related Fields 133 (2005), 409–436.
- [9] I. Gentil, A. Guillin, and L. Miclo, *Modified logarithmic Sobolev inequalities in null curvature*, Rev. Mat. Iberoamer. 23 (2007), 235–258.
- [10] N. Gozlan, C. Roberto, and P.-M. Samson, *Hamilton Jacobi equations on metric spaces and transport entropy inequalities*, Rev. Mat. Iberoamer. 30 (2014), 133–163.
- [11] N. Gozlan, C. Roberto, and P.-M. Samson, *From dimension free concentration to the Poincaré inequality*, Calc. Var. Partial Differential Equations 52 (2015), 899–925.
- [12] N. Gozlan, C. Roberto, P.-M. Samson, Y. Shu, and P. Tetali, *Characterization of a class of weak transport-entropy inequalities on the line*, arXiv:1509.04202v1 (2015).
- [13] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali, *Kantorovich duality for general transport costs and applications*, arXiv:1412.7480v1 (2014).
- [14] P. Hitczenko, S. Kwapien, W. V. Li, G. Schechtman, T. Schlumprecht, and J. Zinn, *Hypercontractivity and comparison of moments of iterated maxima and minima of independent random variables*, Electron. J. Probab. 3 (1998), no. 2, 26 pp.
- [15] M. Ledoux, *On Talagrand’s deviation inequalities for product measures*, ESAIM Probab. Statist. 1 (1995/97), 63–87.
- [16] M. Ledoux, *The Concentration of Measure Phenomenon*, Math. Surveys Monogr. 89, Amer. Math. Soc., Providence, RI, 2001.
- [17] K. Marton, *A measure concentration inequality for contracting Markov chains*, Geom. Funct. Anal. 6 (1996), 556–571.
- [18] B. Maurey, *Some deviation inequalities*, Geom. Funct. Anal. 1 (1991), 188–197.
- [19] P.-M. Samson, *Concentration of measure inequalities for Markov chains and  $\Phi$ -mixing processes*, Ann. Probab. 28 (2000), 416–461.
- [20] P.-M. Samson, *Concentration inequalities for convex functions on product spaces*, in: Stochastic Inequalities and Applications, Progr. Probab. 56, Birkhäuser, Basel, 2003, 33–52.
- [21] M. Talagrand, *A new isoperimetric inequality and the concentration of measure phenomenon*, in: Geometric Aspects of Functional Analysis (1989–90), Lecture Notes in Math. 1469, Springer, Berlin, 1991, 94–124.
- [22] M. Talagrand, *Concentration of measure and isoperimetric inequalities in product spaces*, Inst. Hautes Études Sci. Publ. Math. 81 (1995), 73–205.

Radosław Adamczak, Michał Strzelecki  
Institute of Mathematics  
University of Warsaw  
Banacha 2  
02-097 Warszawa, Poland  
E-mail: R.Adamczak@mimuw.edu.pl  
M.Strzelecki@mimuw.edu.pl

