

On the convex Poincaré inequality and weak transportation inequalities

RADOSŁAW ADAMCZAK* and MICHAŁ STRZELECKI**

Institute of Mathematics, University of Warsaw, Banacha 2, 02–097 Warsaw, Poland.

*E-mail: *R.Adamczak@mimuw.edu.pl; **M.Strzelecki@mimuw.edu.pl*

We prove that for a probability measure on \mathbb{R}^n , the Poincaré inequality for convex functions is equivalent to the weak transportation inequality with a quadratic-linear cost. This generalizes recent results by Gozlan, Roberto, Samson, Shu, Tetali and Feldheim, Marsiglietti, Nayar, Wang, concerning probability measures on the real line.

The proof relies on modified logarithmic Sobolev inequalities of Bobkov–Ledoux type for convex and concave functions, which are of independent interest.

We also present refined concentration inequalities for general (not necessarily Lipschitz) convex functions, complementing recent results by Bobkov, Nayar, and Tetali.

Keywords: concentration of measure; convex functions; Poincaré inequality; weak transport-entropy inequalities

1. Introduction

In the last thirty years a substantial body of research has been devoted to the interplay between various functional inequalities, transportation of measure theory, and the concentration of measure phenomenon, showing intimate connection between them. While most of the investigations have been carried out in the setting of general Lipschitz functions, concentration inequalities restricted to the class of convex Lipschitz functions have also been considered by many authors, starting from the seminal work by Talagrand in the 1990s ([30,31], see also [21,24,28,29] and the monograph [23] for subsequent developments). A crucial feature of these results is that they are satisfied under much less restrictive assumptions concerning the regularity of the underlying probability measure when compared to inequalities valid for all Lipschitz functions. Even though the theory of concentration of measure for convex functions to some extent parallels the classical theory, there are some subtle differences related to the fact that convexity is not preserved under general contractions – even under the change of signs – which creates certain difficulties in the proofs and makes many well known arguments, which have been established in the classical context, invalid. As a consequence, the theory of concentration of measure for convex functions has not yet reached a satisfactory level of completeness. Nevertheless, several important results have been obtained in recent years, connecting dimension-free concentration inequalities for convex functions with the convex Poincaré inequality [17] and a new type of weak transportation cost inequalities [18,19]. We will now briefly describe these developments, which will allow us to formulate our main result.

Let $|\cdot|$ stand for the standard Euclidean norm on \mathbb{R}^n . Let μ be a Borel probability measure on \mathbb{R}^n and let X be a random vector with law μ . We say that μ (equivalently X) satisfies the convex Poincaré inequality with constant $\lambda > 0$ if for all convex functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$ we have

$$\text{Var } f(X) \leq \frac{1}{\lambda} \mathbb{E} |\nabla f(X)|^2, \tag{1.1}$$

where by $|\nabla f(x)|$ we mean the length of gradient at x , defined as

$$|\nabla f(x)| = \limsup_{y \rightarrow x} \frac{|f(y) - f(x)|}{|y - x|}. \tag{1.2}$$

Note that this coincides with the length of the “true” gradient provided f is differentiable at x . Also, it is enough to assume that (1.1) holds for convex Lipschitz functions, since an arbitrary convex function can be pointwise approximated by convex Lipschitz functions.

It follows from the results by Gozlan, Roberto, and Samson [17] that μ satisfies the convex Poincaré inequality if and only if there exists a constant $c > 0$ such that for any N , any convex set $A \subseteq (\mathbb{R}^n)^N$ with $\mu^{\otimes N}(A) \geq 1/2$, and any $t > 0$,

$$\mu^{\otimes N}(A + tB_2^{Nn}) \geq 1 - 2\exp(-ct), \tag{1.3}$$

where B_2^k denotes the unit Euclidean ball in \mathbb{R}^k and $+$ stands for the Minkowski addition.

It is not difficult to see that (1.3) is equivalent to the one-sided deviation inequality for convex 1-Lipschitz functions, that is,

$$\mathbb{P}(f(X_1, \dots, X_N) \geq \text{Med } f(X_1, \dots, X_N) + t) \leq 2e^{-ct} \tag{1.4}$$

for all $t \geq 0$, where X_1, \dots, X_N are i.i.d. copies of X , and $\text{Med } Y$ denotes the median of the random variable Y , that is, $\text{Med } Y = \inf\{t \in \mathbb{R} : \mathbb{P}(Y \leq t) \geq 1/2\}$.

Thus, the convex Poincaré inequality is equivalent to a dimension-free deviation inequality for the upper tail of convex Lipschitz functions.

Let us now pass to the connections between the Poincaré inequality and transportation inequalities. Let $\theta : \mathbb{R}^n \rightarrow [0, \infty]$ be a measurable function with $\theta(0) = 0$. Recall that the optimal transport cost between two probability measures μ and ν on \mathbb{R}^n , induced by θ is given by

$$\mathcal{T}_\theta(\nu, \mu) = \inf_{\pi} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \theta(x - y) \pi(dx dy), \tag{1.5}$$

where the infimum is taken over all couplings between μ and ν , that is, over all probability measures on $(\mathbb{R}^n)^2$ such that $\pi(dx \times \mathbb{R}^n) = \mu(dx)$, $\pi(\mathbb{R}^n \times dy) = \nu(dy)$. Recall also that the relative entropy $H(\nu|\mu)$ is defined as

$$H(\nu|\mu) = \int_{\mathbb{R}^n} \log \frac{d\nu}{d\mu} d\nu, \tag{1.6}$$

if ν is absolutely continuous with respect to μ and $H(\nu|\mu) = \infty$ otherwise.

It has been proved in [7] that μ satisfies the Poincaré inequality (1.1) for all smooth functions if and only if there exist constants C, D such that for all probability measures ν ,

$$\mathcal{T}_{\theta_{C,D}}(\nu, \mu) \leq H(\nu|\mu), \tag{1.7}$$

where

$$\theta_{C,D}(x) = \begin{cases} \frac{|x|^2}{2C} & \text{for } |x| \leq CD, \\ D|x| - \frac{CD^2}{2} & \text{for } |x| > CD. \end{cases} \tag{1.8}$$

Recently Gozlan, Roberto, Samson, Shu, and Tetali [18] formulated a similar characterization of the convex Poincaré inequality on the real line. In order to formulate their result we need to introduce the weak transport cost between probability measures and corresponding transportation inequalities as defined in [18,19].

In what follows, by $\mathcal{P}_1(\mathbb{R}^n)$ we denote the class of all probability measures ν on \mathbb{R}^n such that $\int_{\mathbb{R}^n} |x| d\nu(x) < \infty$.

Definition 1.1. Let μ and ν be probability measures on \mathbb{R}^n . Assume that $\nu \in \mathcal{P}_1(\mathbb{R}^n)$. For a convex, lower semicontinuous function $\theta: \mathbb{R}^n \rightarrow [0, \infty]$, such that $\theta(0) = 0$ define the weak transport cost between μ and ν as

$$\overline{\mathcal{T}}_{\theta}(\nu|\mu) = \inf_{\pi} \int_{\mathbb{R}^n} \theta \left(x - \int_{\mathbb{R}^n} y p_x(dy) \right) \mu(dx),$$

where the infimum is taken over all couplings between μ and ν and for $x \in \mathbb{R}^n$, $p_x(\cdot)$ is the conditional measure defined (μ almost surely) by $\pi(dx dy) = p_x(dy)\mu(dx)$.

Note that in the probabilistic notation one can write

$$\overline{\mathcal{T}}_{\theta}(\nu|\mu) = \inf_{(X,Y)} \mathbb{E}\theta(X - \mathbb{E}(Y|X)),$$

where the infimum is taken over all pairs of random vectors (X, Y) with values in $\mathbb{R}^n \times \mathbb{R}^n$, such that X is distributed according to μ and Y according to ν .

Due to the asymmetry between μ and ν , one can now introduce three different inequalities related to the cost $\overline{\mathcal{T}}_{\theta}$.

Definition 1.2. Let $\mu \in \mathcal{P}_1(\mathbb{R}^n)$ and $\theta: \mathbb{R}^n \rightarrow [0, \infty]$ be a convex lower semicontinuous function with $\theta(0) = 0$. We will say that μ satisfies the inequality

- $\overline{\mathbf{T}}_{\theta}^+$ if for every probability measure $\nu \in \mathcal{P}_1(\mathbb{R}^n)$,

$$\overline{\mathcal{T}}_{\theta}(\nu|\mu) \leq H(\nu|\mu),$$

- $\overline{\mathbf{T}}_\theta^-$ if for every probability measure $\nu \in \mathcal{P}_1(\mathbb{R}^n)$,

$$\overline{\mathcal{T}}_\theta(\mu|\nu) \leq H(\nu|\mu),$$

- $\overline{\mathbf{T}}_\theta$ if μ satisfies both $\overline{\mathbf{T}}_\theta^+$ and $\overline{\mathbf{T}}_\theta^-$.

The definition of those inequalities in [19] differs formally from the one presented above (which is taken from [18]). It is not difficult to see that the definitions presented in both articles are equivalent up to universal constants – the version above is more convenient for our purposes.

The authors of [18] proved that a probability measure μ on the real line satisfies the convex Poincaré inequality for some constant $\lambda > 0$ if and only if it satisfies the transportation inequality $\overline{\mathbf{T}}_{\theta C, D}$ for some $C, D > 0$. In a dual formulation (expressed in terms of infimum convolution inequalities), this result has been also obtained in [14].

Our main result is an extension of this equivalence to arbitrary dimension.

Theorem 1.3. *Let μ be a probability measure on \mathbb{R}^n . Then the following conditions are equivalent:*

- (i) *There exists $\lambda > 0$ such that μ satisfies the convex Poincaré inequality (1.1).*
- (ii) *There exist $C, D > 0$ such that μ satisfies the transportation inequality $\overline{\mathbf{T}}_{\theta C, D}$.*

Remark 1.4. The implication (ii) \implies (i) is standard, in this case $\lambda = \frac{1}{C}$. In our proof the constants C, D in the implication (i) \implies (ii) depend not only on λ but also on certain quantiles related to the measure μ (which are always finite but may be of the order of up to \sqrt{n}). This is related to the inequality $\overline{\mathbf{T}}_{\theta C, D}^+$ responsible for the lower tail of convex functions, which is usually more difficult to deal with than the upper tail. We suspect that this is an artefact of our proof and one should be able to obtain $\overline{\mathbf{T}}_{\theta C, D}^+$ with C, D depending only on λ . As for $\overline{\mathbf{T}}_{\theta C, D}^-$ our argument does yield it with C, D depending only on λ (see Corollary 4.3 below for details).

Remark 1.5. Thanks to well known tensorization properties of the inequality $\overline{\mathbf{T}}_{\theta C, D}$, Theorem 1.3 implies that the convex Poincaré inequality is equivalent to improved two-level dimension free concentration inequality for convex functions (see Example 6.9 below for a precise formulation). In the class of Lipschitz functions inequalities of this type have been first obtained by Talagrand [30] in the case of the product exponential distribution (with an alternate proof, using infimum-convolution inequalities, by Maurey [24]). The fact that they are consequences of the Poincaré inequality for smooth functions was established by Bobkov and Ledoux [6]. By results due to Gozlan *et al.* [17], this can be regarded as a self-improvement of dimension-free concentration properties of Lipschitz functions. Our result shows that similar self-improvements are present also in the setting of convex concentration.

Remark 1.6. In [8], Bobkov and Götze provide a simple characterization of measures on \mathbb{R} which satisfy the convex Poincaré inequality for some $\lambda > 0$ (and thus also the inequality $\overline{\mathbf{T}}_{C, D}$) in terms of the probability distribution function. A similar characterization for larger n seems to be a non-trivial open problem.

The organization of the article is as follows. First, in Section 2, we present preliminary properties of measures satisfying the convex Poincaré inequality and weak transportation inequalities, to be used in the proofs. Section 3 contains our most important technical result, that is modified log-Sobolev inequalities for convex and concave functions, which in Section 4 are combined with the Hamilton–Jacobi equations giving the proof of Theorem 1.3.

Next, in Section 5, we briefly discuss operations preserving the convex Poincaré inequality, which may be used to provide new non-trivial examples of measures satisfying it.

In Section 6, we present refined concentration of measure inequalities, which are consequences of weak transportation inequalities. We consider there more general cost functions than the one corresponding to the convex Poincaré inequality and discuss applications both to the Lipschitz and non-Lipschitz setting.

Finally, in Section 7, we state a few open questions. The [Appendix](#) contains basic facts concerning Hamilton–Jacobi equations, which are used in the proof of Theorem 1.3.

2. Preliminaries on the convex Poincaré inequality and weak transportation inequalities

In this section, we present basic concentration of measure properties implied by the convex Poincaré inequality and the dual formulations of weak transportation inequalities. They will be needed in the proof of our main result.

We begin with a simple reformulation of the convex Poincaré inequality.

Lemma 2.1. *Let X be a random vector in \mathbb{R}^n satisfying the convex Poincaré inequality (1.1). Then for every convex function $f: \mathbb{R}^n \rightarrow \mathbb{R}$,*

$$\mathbb{E}(f(X) - \text{Med } f(X))^2 \leq \frac{2}{\lambda} \mathbb{E}|\nabla f(X)|^2.$$

Proof. Note that for every random variable Z , thanks to the fact that the median minimizes the mean absolute deviation, we have

$$(\mathbb{E}Z - \text{Med } Z)^2 \leq (\mathbb{E}|Z - \text{Med } Z|)^2 \leq (\mathbb{E}|Z - \mathbb{E}Z|)^2 \leq \text{Var } Z.$$

Thus

$$\mathbb{E}(Z - \text{Med } Z)^2 = \text{Var } Z + (\mathbb{E}Z - \text{Med } Z)^2 \leq 2 \text{Var } Z$$

and it is enough to set $Z = f(X)$ and apply (1.1). □

2.1. Concentration inequalities

Let us start with the already mentioned (see (1.4)) upper tail estimate for convex Lipschitz functions implied by the convex Poincaré inequality. The proposition below can be also obtained up to constants by abstract results from [17], but we would like to provide an alternative derivation

based on moments (the possibility of such a proof was suggested in [17]). Our strategy mimics a well known approach from the general Lipschitz case (see e.g. Proposition 2.5 in [25]), however we have to deal with some small difficulties related to the fact that in the convex setting we cannot truncate the function as this operation does not preserve convexity.

Proposition 2.2. *Assume that X is a random vector in \mathbb{R}^n , satisfying the convex Poincaré inequality (1.1). Then for any L -Lipschitz convex function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and any $t > 0$,*

$$\mathbb{P}(f(X) \geq \text{Med } f(X) + t) \leq 8 \exp\left(-\frac{\sqrt{2\lambda}}{eL}t\right).$$

Proof of Proposition 2.2. Consider the random variable $Y = (|X| - a)_+$, where $a \in \mathbb{R}_+$ is arbitrary such that $\mathbb{P}(|X| \leq a) > 1/4$, and let Y' be an independent copy of Y . Since the function $\varphi(x) = (|x| - a)_+$ is convex,

$$\begin{aligned} \frac{1}{\lambda} \mathbb{P}(|X| \geq a) &= \frac{1}{\lambda} \mathbb{E}|\nabla \varphi(X)|^2 \geq \text{Var } Y = \frac{1}{2} \mathbb{E}(Y - Y')^2 \\ &\geq \frac{1}{2} \mathbb{E}(Y - Y')^2 (1_{\{Y>0\}}1_{\{Y'=0\}} + 1_{\{Y=0\}}1_{\{Y'>0\}}) \\ &\geq \frac{1}{4} \mathbb{E}Y^2 1_{\{Y>0\}} \geq \frac{2}{\lambda} \mathbb{P}(|X| > a + 2\sqrt{2/\lambda}) \end{aligned}$$

and so $\mathbb{P}(|X| \geq a + 2\sqrt{2/\lambda}) \leq 2^{-1} \mathbb{P}(|X| \geq a)$, which implies that $|X|$ is exponentially integrable. In particular for every Lipschitz function f and all $p > 0$, $\mathbb{E}|f(X)|^p < \infty$.

Assume now that $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is convex. Then for all $p \geq 2$, applying Lemma 2.1 to the convex function $x \mapsto (f(x) - \text{Med } f(X))_+^{p/2}$ (note that its median is zero and $|\nabla(f(x) - \text{Med } f(X))_+| \leq |\nabla f(x)|$), we obtain

$$\begin{aligned} \mathbb{E}(f(X) - \text{Med } f(X))_+^p &\leq \frac{2}{\lambda} \cdot \frac{p^2}{4} \mathbb{E}(f(X) - \text{Med } f(X))_+^{p-2} |\nabla f(X)|^2 \\ &\leq \frac{p^2}{2\lambda} (\mathbb{E}(f(X) - \text{Med } f(X))_+^p)^{1-2/p} (\mathbb{E}|\nabla f(X)|^p)^{2/p}, \end{aligned}$$

where we used Hölder’s inequality with exponents $p/(p - 2)$, $p/2$. If we additionally assume that f is Lipschitz, so that $\mathbb{E}(f(X) - \text{Med } f(X))_+^p < \infty$, we get

$$(\mathbb{E}(f(X) - \text{Med } f(X))_+^p)^{1/p} \leq \frac{P}{\sqrt{2\lambda}} (\mathbb{E}|\nabla f(X)|^p)^{1/p}, \tag{2.1}$$

which via Chebyshev’s inequality in L_p implies

$$\mathbb{P}\left(f(X) \geq \text{Med } f(X) + e \frac{P}{\sqrt{2\lambda}} (\mathbb{E}|\nabla f(X)|^p)^{1/p}\right) \leq e^{2-p} \tag{2.2}$$

for $p \geq 0$ (the inequality holds trivially for $0 \leq p < 2$ because of the multiplicative constant e^2). Now, if the Lipschitz constant of f equals one, the above inequality yields for $t > 0$,

$$\mathbb{P}(f(X) \geq \text{Med } f(X) + t) \leq \exp\left(2 - \frac{\sqrt{2\lambda}}{e}t\right) \leq 8 \exp\left(-\frac{\sqrt{2\lambda}}{e}t\right). \quad \square$$

Remark 2.3. Another possible approach is based on the Laplace transform: assume without loss of generality that $\mathbb{E}f(X) = 0$ and denote $M(s) = \mathbb{E}e^{sf(X)}$ for $s \geq 0$. Since the function $e^{sf(\cdot)/2}$ is convex, the Poincaré inequality yields

$$M(s) - M(s/2)^2 = \text{Var}(e^{sf(X)/2}) \leq \frac{1}{4\lambda} \mathbb{E}s^2 |\nabla f(X)|^2 e^{sf(X)} \leq \frac{L^2 s^2}{4\lambda} M(s).$$

The idea would be now to regroup the expressions appearing in the above inequality, repeat the procedure (with $s/2$ instead of s), and – after a simple limit argument – obtain a bound on $\mathbb{E}e^{sf(X)}$. After that we could use Markov’s inequality and optimize in s to obtain an estimate of the upper tail of f . However a delicate issue emerges: we have to a priori know that (for reasonable choices of the parameter s) $e^{sf(X)}$ is integrable (in the setting of smooth functions one overcomes this problem simply by truncating f , for convex functions one would need e.g. to repeat the beginning of the proof of Proposition 2.2); cf. the remark following Theorem 6.8 in [17].

We do not know if the convex Poincaré inequality implies similar tail estimates – which depend only on λ and the Lipschitz constant of the function – for the lower tail of convex Lipschitz functions, that is, for $\mathbb{P}(f(X) \leq \text{Med } f(X) - t)$, $t > 0$ (cf. Question 7.3 below).

Nonetheless, we can easily get estimates in terms of λ and certain quantiles of X . They will be crucial in the proof of the implication

$$\text{Convex Poincaré inequality} \implies \overline{\mathbf{T}}_{\theta c, D}^+$$

Lemma 2.4. *Let X be a random vector in \mathbb{R}^n satisfying the convex Poincaré inequality (1.1) and let M be any number such that $\mathbb{P}(|X - \mathbb{E}X| \leq M) \geq 3/4$. Then for every convex $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and for any $t > 32M\mathbb{E}|\nabla f(X)|$,*

$$\mathbb{P}(f(X) \leq \text{Med } f(X) - t) \leq 8 \exp\left(-\frac{\sqrt{2\lambda}}{16e\mathbb{E}|\nabla f(X)|}t\right).$$

Proof. By Proposition 2.2 (note that the function $x \mapsto |x - \mathbb{E}X|$ is convex and 1-Lipschitz),

$$\mathbb{P}(|X - \mathbb{E}X| \geq M + t) \leq 8 \exp\left(-\frac{\sqrt{2\lambda}}{e}t\right), \quad t \geq 0. \quad (2.3)$$

Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex function. Without loss of generality, we may assume $\text{Med } f(X) = 0$. We have

$$\mathbb{P}(f(X) \geq 0) \geq 1/2,$$

$$\begin{aligned} \mathbb{P}(|X - \mathbb{E}X| \leq M) &\geq 3/4, \\ \mathbb{P}(|\nabla f(X)| < 8\mathbb{E}|\nabla f(X)|) &\geq 7/8. \end{aligned}$$

Thus, there exists x_0 such that $f(x_0) \geq 0$, $|x_0 - \mathbb{E}X| \leq M$, and $|\nabla f(x_0)| < 8\mathbb{E}|\nabla f(X)|$. Define

$$\tilde{f}(x) = f(x_0) + \langle u, x - x_0 \rangle, \quad x \in \mathbb{R}^n,$$

where u is any subgradient of f at x_0 , so that $\tilde{f}(x) \leq f(x)$ for all $x \in \mathbb{R}^n$. Taking $x = x_0 + \varepsilon u$ with $\varepsilon \rightarrow 0$ we see that $|u| \leq |\nabla f(x_0)| \leq 8\mathbb{E}|\nabla f(X)|$, and thus we have

$$\begin{aligned} \mathbb{P}(f(X) \leq -t) &\leq \mathbb{P}(\tilde{f}(X) \leq -t) \leq \mathbb{P}(\langle u, X - x_0 \rangle \leq -t) \\ &\leq \mathbb{P}(|u||X - x_0| \geq t) \leq \mathbb{P}(|X - x_0| \geq t/(8\mathbb{E}|\nabla f(X)|)) \\ &\leq \mathbb{P}(|X - \mathbb{E}X| \geq t/(8\mathbb{E}|\nabla f(X)|) - |x_0 - \mathbb{E}X|) \\ &\leq \mathbb{P}(|X - \mathbb{E}X| \geq t/(8\mathbb{E}|\nabla f(X)|) - M). \end{aligned}$$

If now $t/(16\mathbb{E}|\nabla f(X)|) \geq 2M$, we can conclude from (2.3) that

$$\mathbb{P}(f(X) \leq -t) \leq \mathbb{P}(|X - \mathbb{E}X| \geq M + t/(16\mathbb{E}|\nabla f(X)|)) \leq 8 \exp\left(-\frac{\sqrt{2\lambda}}{16e\mathbb{E}|\nabla f(X)|}t\right),$$

which ends the proof. □

2.2. Infimum convolution. Dual formulation of transportation inequalities

We will rely on the following lemma proved in [18] (and in a slightly different version also in [19]). The proof in [18] is presented for the real line, but it is not difficult to see that it generalizes to arbitrary dimension.

Lemma 2.5. *Let $\theta: \mathbb{R}^n \rightarrow \mathbb{R}_+$ be a convex cost function, $\theta(0) = 0$, $\lim_{x \rightarrow \infty} \theta(x) = \infty$. For all functions $f: \mathbb{R}^n \rightarrow \mathbb{R}$ bounded from below, $x \in \mathbb{R}^n$, and $t > 0$ set*

$$Q_t f(x) = Q_t^\theta f(x) = \inf_{y \in \mathbb{R}^n} \left\{ f(y) + t\theta\left(\frac{x - y}{t}\right) \right\}.$$

Then

- (i) μ satisfies $\overline{\mathbf{T}}_\theta^+$ if and only if for all convex $f: \mathbb{R}^n \rightarrow \mathbb{R}$, bounded from below,

$$\exp\left(\int_{\mathbb{R}^n} Q_1 f d\mu\right) \int_{\mathbb{R}^n} e^{-f} d\mu \leq 1; \tag{2.4}$$

- (ii) μ satisfies $\overline{\mathbf{T}}_\theta^-$ if and only if for all convex $f: \mathbb{R}^n \rightarrow \mathbb{R}$, bounded from below,

$$\int_{\mathbb{R}^n} \exp(Q_1 f) d\mu \exp\left(-\int_{\mathbb{R}^n} f d\mu\right) \leq 1; \tag{2.5}$$

(iii) if μ satisfies $\overline{\mathbf{T}}_\theta$, then for all convex $f: \mathbb{R}^n \rightarrow \mathbb{R}$, bounded from below,

$$\int_{\mathbb{R}^n} \exp(Q_t f) d\mu \int_{\mathbb{R}^n} e^{-f} d\mu \leq 1 \tag{2.6}$$

holds with $t = 2$. Conversely, if μ satisfies (2.6) for some $t > 0$, then it satisfies $\overline{\mathbf{T}}_{\tilde{\theta}}$ with $\tilde{\theta}(\cdot) = t\theta(\cdot/t)$.

Moreover, the inequality (2.4) (resp. (2.5)) for all convex, Lipschitz functions bounded from below is a sufficient condition for $\overline{\mathbf{T}}_\theta^+$ (resp. $\overline{\mathbf{T}}_\theta^-$).

The inequality (2.6) was introduced by Maurey in [24] and the relation with transportation cost inequalities was first observed in [9].

3. From convex Poincaré inequality to modified log-Sobolev inequalities for convex and concave functions

In this section, we present modified log-Sobolev inequalities for convex and concave functions which are implied by the convex Poincaré inequality. Our approach builds heavily on the arguments introduced by Bobkov and Ledoux in [6] for arbitrary Lipschitz functions, however some non-trivial modifications will be necessary in order to handle the difficulties imposed by the restriction of the Poincaré inequality to convex functions.

In what follows for a nonnegative random variable Y , we define its entropy as

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

if $\mathbb{E}Y \log Y < \infty$ and $\text{Ent } Y = \infty$ otherwise. We refer for example, to [5,23] for basic properties of entropy and log-Sobolev inequalities.

Throughout this section, we assume that μ is a probability measure on \mathbb{R}^n satisfying the convex Poincaré inequality (1.1) and that X is a random vector with law μ , which will not be explicitly stated in all the theorems.

3.1. Modified log-Sobolev inequalities for convex functions

Theorem 3.1. *Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be convex with $|\nabla f(x)| \leq c < \sqrt{2\lambda}/e$ for all $x \in \mathbb{R}^n$. Then*

$$\text{Ent}(e^{f(X)}) \leq C \mathbb{E}|\nabla f(X)|^2 e^{f(X)}, \tag{3.1}$$

where

$$C = C(\lambda, c) = \frac{1}{3\lambda} \exp(c\sqrt{2/\lambda}) + \frac{1}{3(\sqrt{\lambda/2} - c/2)^2}.$$

Our constants are slightly worse than in [6], basically because we need to work with the median rather than the mean. However the argument (which works also in the classical case) seems to slightly simplify the technicalities of [6]. The proof relies on two propositions.

Proposition 3.2. *Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be convex with $\text{Med } f(X) = 0$ and $|\nabla f(x)| \leq c < \sqrt{2\lambda}/e$ for all $x \in \mathbb{R}^n$. Then*

$$\mathbb{E}f(X)^2 e^{f(X)} \leq C_1 \mathbb{E}|\nabla f(X)|^2 e^{f(X)},$$

where $C_1 = C_1(c, \lambda) = (\sqrt{\lambda/2} - c/2)^{-2}$.

Proof. For $x \in \mathbb{R}$ we define $\Psi(x) = x e^{x/2}$ and

$$\Phi(x) = \begin{cases} x e^{x/2} & \text{for } x \geq -2, \\ -2/e & \text{for } x < -2. \end{cases}$$

One easily checks that $|\Psi(x)| \leq |\Phi(x)|$, $|\Phi'(x)| \leq |\Psi'(x)|$, and Φ is convex nondecreasing.

Denote $a^2 = \mathbb{E}|\Phi(f(X))|^2$ and $b^2 = \mathbb{E}|\nabla f(X)|^2 e^{f(X)}$ (where $a, b \geq 0$). The function $\Phi(f)$ is convex, moreover $\text{Med } \Phi(f(X)) = 0$. Hence, by Lemma 2.1,

$$\begin{aligned} a^2 &\leq \frac{2}{\lambda} \mathbb{E}|\nabla f(X)|^2 (1 + f(X)/2)^2 e^{f(X)} 1_{\{f(X) \geq -2\}} \\ &\leq \frac{2}{\lambda} \left(b^2 + c \mathbb{E}|\nabla f(X)| e^{f(X)/2} \cdot |f(X)| e^{f(X)/2} + \frac{c^2}{4} \mathbb{E}f(X)^2 e^{f(X)} \right) \\ &\leq \frac{2}{\lambda} \left(b^2 + cb \sqrt{\mathbb{E}f(X)^2 e^{f(X)}} + \frac{c^2}{4} a^2 \right) \\ &\leq \frac{2}{\lambda} (b + ca/2)^2. \end{aligned}$$

Note that $a < \infty$ (by Proposition 2.2 and since $c < \sqrt{2\lambda}/e$). Thus $a(\sqrt{\lambda/2} - c/2) \leq b$ and the assertion follows. □

Proposition 3.3. *Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be either convex or concave, with $\text{Med } f(X) = 0$ and $|\nabla f(x)| \leq c$ for all $x \in \mathbb{R}^n$. Then*

$$\mathbb{E}|\nabla f(X)|^2 \leq C_2 \mathbb{E}|\nabla f(X)|^2 e^{f(X)},$$

where $C_2 = C_2(c, \lambda) = \exp(c\sqrt{2/\lambda})$. Consequently,

$$\mathbb{E}f(X)^2 \leq \frac{2}{\lambda} C_2 \mathbb{E}|\nabla f(X)|^2 e^{f(X)}.$$

Proof. If $|\nabla f(X)|$ vanishes with probability one, there is nothing to prove. Otherwise, denote by $\tilde{\mathbb{E}}$ the expectation with respect to the probability measure with density $|\nabla f(X)|^2 / \mathbb{E}|\nabla f(X)|^2$

relative to \mathbb{P} . By Jensen's inequality,

$$\mathbb{E}|\nabla f(X)|^2 e^{-|f(X)|} = \mathbb{E}|\nabla f(X)|^2 \tilde{\mathbb{E}} e^{-|f(X)|} \geq \mathbb{E}|\nabla f(X)|^2 e^{-\tilde{\mathbb{E}}|f(X)|}.$$

Thus, using the trivial inequality $-|f| \leq f$, we conclude that

$$\mathbb{E}|\nabla f(X)|^2 \leq e^{\tilde{\mathbb{E}}|f(X)|} \mathbb{E}|\nabla f(X)|^2 e^{f(X)}.$$

But since by Lemma 2.1 we have

$$\begin{aligned} \mathbb{E}|\nabla f(X)|^2 |f(X)| &\leq c \mathbb{E}|\nabla f(X)| |f(X)| \leq c \sqrt{\mathbb{E}|\nabla f(X)|^2} \sqrt{\mathbb{E}f(X)^2} \\ &\leq c \sqrt{2/\lambda} \mathbb{E}|\nabla f(X)|^2, \end{aligned}$$

we can bound $\tilde{\mathbb{E}}|f(X)|$ by $c\sqrt{2/\lambda}$. This yields the assertion of the proposition. □

Proof of Theorem 3.1. Without loss of generality, assume $\text{Med } f(X) = 0$. Denote $F(t) = \mathbb{E}f(X)^2 e^{tf(X)}$, $t \in [0, 1]$. By the formula $\int_0^1 ta^2 e^{ta} dt = ae^a - e^a + 1$ and the convexity of $t \mapsto F(t)$,

$$\begin{aligned} \text{Ent}(e^{f(X)}) &\leq \mathbb{E}(f(X)e^{f(X)} - e^{f(X)} + 1) = \mathbb{E} \int_0^1 tf(X)^2 e^{tf(X)} dt = \int_0^1 tF(t) dt \\ &\leq \int_0^1 t(1-t)F(0) + t^2F(1) dt = \frac{1}{6}F(0) + \frac{1}{3}F(1) \end{aligned}$$

(note that for this argument to work we do *not* need the expectation of $f(X)$ to vanish). Thus Propositions 3.2 and 3.3 imply the assertion of the theorem. □

3.2. Modified log-Sobolev inequalities for concave functions

Theorem 3.4. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be convex with $|\nabla f(x)| \leq c < \sqrt{2\lambda}/(32e)$ for all $x \in \mathbb{R}^n$. Assume that $M \in \mathbb{R}_+$ satisfies $\mathbb{P}(|X - \mathbb{E}X| \leq M) \geq 3/4$. Then

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

where $C = C(\lambda, c, M)$ is a constant depending only on λ, c, M .

Remark 3.5. If we denote by X_1, \dots, X_n the coordinates of X , then by the Poincaré inequality we have

$$\mathbb{E}|X - \mathbb{E}X|^2 = \sum_{i=1}^n \mathbb{E}|X_i - \mathbb{E}X_i|^2 \leq \frac{n}{\lambda},$$

and hence, by the Chebyshev inequality, $M = 2\sqrt{n/\lambda}$ satisfies $\mathbb{P}(|X - \mathbb{E}X| \leq M) \geq 3/4$. Thus in fixed dimension n and for say $c = \sqrt{2\lambda}/(64e)$, the constant C in Theorem 3.4 can be bounded

uniformly over all probability measures satisfying the convex Poincaré inequality with constant λ .

Proof of Theorem 3.4. We start as in the proof of Theorem 3.1. Denote $g = -f$ (this is a concave function). Without loss of generality, assume $\text{Med } f(X) = 0$. Denote $F(t) = \mathbb{E}g(X)^2 e^{tg(X)}$, $t \in [0, 1]$. By the convexity of $t \mapsto F(t)$,

$$\begin{aligned} \text{Ent}(e^{g(X)}) &\leq \mathbb{E}(g(X)e^{g(X)} - e^{g(X)} + 1) = \mathbb{E} \int_0^1 tg(X)^2 e^{tg(X)} dt = \int_0^1 tF(t) dt \\ &\leq \int_0^1 t(1-t)F(0) + t^2F(1) dt = \frac{1}{6}F(0) + \frac{1}{3}F(1). \end{aligned} \tag{3.2}$$

We have

$$F(1) \leq \mathbb{E}g(X)^2 + \mathbb{E}g_+(X)^2 e^{g_+(X)} = F(0) + \mathbb{E}g_+(X)^2 e^{g_+(X)}. \tag{3.3}$$

By Proposition 3.3, $F(0) \leq \frac{2}{\lambda} \exp(c\sqrt{2/\lambda}) \mathbb{E}|\nabla g(X)|^2 e^{g(X)}$, so it remains to estimate $\mathbb{E}g_+(X)^2 e^{g_+(X)}$.

Integration by parts and Lemma 2.4 yield

$$\begin{aligned} \mathbb{E}e^{2g_+(X)} &= 1 + \int_0^\infty 2e^{2t} \mathbb{P}(g_+(X) \geq t) dt \\ &= 1 + \int_0^{32Mc} 2e^{2t} dt + \int_{32Mc}^\infty 2e^{2t} \mathbb{P}(g_+(X) \geq t) dt \\ &\leq e^{64Mc} + \int_{32Mc}^\infty 16 \exp\left(2t - \frac{\sqrt{2\lambda}}{16ec}t\right) dt < D_1 = D_1(\lambda, c, M) < \infty, \end{aligned}$$

if only $c < \sqrt{2\lambda}/(32e)$. Similarly (using Lemma 2.4 in its full strength),

$$\begin{aligned} \mathbb{E}g_+(X)^4 &= \int_0^\infty 4t^3 \mathbb{P}(g_+(X) \geq t) dt \\ &= \int_0^{32M\mathbb{E}|\nabla f(X)|} 4t^3 dt + \int_{32M\mathbb{E}|\nabla f(X)|}^\infty 4t^3 \mathbb{P}(g_+(X) \geq t) dt \\ &\leq (32M\mathbb{E}|\nabla f(X)|)^4 + 4 \int_{32M\mathbb{E}|\nabla f(X)|}^\infty t^3 \exp\left(-\frac{\sqrt{2\lambda}}{16e\mathbb{E}|\nabla f(X)|}t\right) dt \\ &\leq D_2(\mathbb{E}|\nabla f(X)|)^4 \leq D_2(\mathbb{E}|\nabla f(X)|^2)^2 \end{aligned}$$

for some $D_2 = D_2(\lambda, M)$. Thus, by Proposition 3.3, applied to g ,

$$\begin{aligned} \mathbb{E}g_+(X)^2 e^{g_+(X)} &\leq \sqrt{\mathbb{E}g_+(X)^4} \sqrt{\mathbb{E}e^{2g_+(X)}} \leq \sqrt{D_1 D_2} \mathbb{E}|\nabla f(X)|^2 \\ &\leq \sqrt{D_1 D_2} e^{c\sqrt{2/\lambda}} \mathbb{E}|\nabla f(X)|^2 e^{-f(X)}. \end{aligned}$$

This, together with (3.2) and (3.3), ends the proof:

$$\begin{aligned} \text{Ent}(e^{-f(X)}) &\leq \frac{1}{6}F(0) + \frac{1}{3}F(1) \leq \frac{1}{2}F(0) + \frac{1}{3}\mathbb{E}g_+(X)^2e^{s_+(X)} \\ &\leq \left(\frac{1}{\lambda} + \frac{1}{3}\sqrt{D_1D_2}\right)e^{c\sqrt{2/\lambda}}\mathbb{E}|\nabla f(X)|^2e^{-f(X)}. \end{aligned} \quad \square$$

4. Proof of the main result

We will now present the proof of Theorem 1.3. As already mentioned, the implication (ii) \implies (i) is standard, we provide a sketch of its proof just for the sake of completeness. The proof of the implication (i) \implies (ii) follows the arguments introduced first in [7] and based on the analysis of the Hamilton–Jacobi equations. A crucial element of the proof will be the modified log-Sobolev inequalities obtained in Section 3.

Lemma 4.1. *Let X be a random vector in \mathbb{R}^n . Assume that there exist $C < \infty$ and $L > 0$ such that*

$$\mathbb{E}e^{L|X|} < \infty \tag{4.1}$$

and the inequality

$$\text{Ent}(e^{f(X)}) \leq C\mathbb{E}|\nabla f(X)|^2e^{f(X)} \tag{4.2}$$

holds for every convex (respectively: concave) L -Lipschitz function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Then, for every convex Lipschitz function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ bounded from below,

$$\begin{aligned} \mathbb{E}e^{Q_1^\alpha f(X)}e^{-\mathbb{E}f(X)} &\leq 1, \\ (\text{respectively: } e^{\mathbb{E}Q_1^\alpha f(X)}\mathbb{E}e^{-f(X)} &\leq 1), \end{aligned}$$

where $Q_t^\alpha f(x) = \inf_{y \in \mathbb{R}^n} \{f(x - y) + t\alpha(y/t)\}$, $t > 0$, is the infimum convolution operator with the cost function $\alpha : \mathbb{R}^n \rightarrow \mathbb{R}$ given by the formula

$$\alpha(s) = \begin{cases} \frac{|s|^2}{4C} & \text{for } |s| \leq 2CL, \\ L|s| - L^2C & \text{for } |s| > 2CL. \end{cases} \tag{4.3}$$

Remark 4.2. The condition (4.1) is introduced to exclude heavy-tailed measures for which the only exponentially integrable convex functions are constants. Note that in this case the inequality (4.2) may be trivially satisfied, while the transportation inequality cannot hold (as it implies the existence of exponential moments).

If we recall the dual formulations of the weak transport-entropy inequalities $\overline{\mathbf{T}}^-$ and $\overline{\mathbf{T}}^+$ (see Lemma 2.5), the definition of $\theta_{C,D}$ from (1.8), and the results of the preceding section (namely, Theorems 3.1 and 3.4), we immediately obtain the following corollaries.

Corollary 4.3. *Let X be a random vector in \mathbb{R}^n satisfying the convex Poincaré inequality (1.1). Then, for any positive $c < \sqrt{2\lambda}/e$, the law of X satisfies the inequality $\overline{\mathbf{T}}_{\theta_{2c,c}}^-$ with*

$$C = C(\lambda, c) = \frac{1}{3\lambda} \exp(c\sqrt{2/\lambda}) + \frac{1}{3(\sqrt{\lambda/2} - c/2)^2}.$$

Corollary 4.4. *Let X be a random vector in \mathbb{R}^n satisfying the convex Poincaré inequality (1.1) and let M be any number such that $\mathbb{P}(|X - \mathbb{E}X| \leq M) \geq 3/4$. Then, for any $c < \sqrt{2\lambda}/(32e)$, the law of X satisfies the inequality $\overline{\mathbf{T}}_{\theta_{2c,c}}^+$ for some constant $C = C(\lambda, c, M)$ depending only on λ , c , and M .*

Proof of Lemma 4.1. Suppose that the log-Sobolev inequality (4.2) holds for all convex and L -Lipschitz functions. We first present a perturbation argument which allows us to work with random vectors with an absolutely continuous law. We then shall follow the approach of [18], Proof of Theorem 1.5.

Let G be a Gaussian random vector in \mathbb{R}^n , independent of X , with the covariance matrix being a sufficiently small multiple of identity, so that it satisfies the usual log-Sobolev inequality with constant C ,

$$\text{Ent} e^{f(G)} \leq C \mathbb{E} |\nabla f(G)|^2 e^{f(G)}$$

for all Lipschitz functions $f: \mathbb{R}^n \rightarrow \mathbb{R}$ (see, e.g., Theorem 5.1 in [23] for an equivalent formulation).

Then, by the tensorization property of entropy (see, e.g., Proposition 5.6 in [23]), the random vector (X, G) on $\mathbb{R}^n \times \mathbb{R}^n$ satisfies the modified log-Sobolev inequality

$$\text{Ent}(e^{F(X,G)}) \leq C \mathbb{E} (|\nabla_X F(X, G)|^2 + |\nabla_G F(X, G)|^2) e^{F(X,G)} \tag{4.4}$$

for all convex functions $F: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ which are L -Lipschitz with respect to the first coordinate (here $|\nabla_X F|$ and $|\nabla_G F|$ denote partial lengths of gradients with respect to the first and second variable, with the other variable fixed).

Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex L -Lipschitz function and consider $\varepsilon > 0$. Applying the inequality (4.4) to the function defined by the formula $F(x, y) = f(x + \varepsilon y)$ for $x, y \in \mathbb{R}^n$ (which is L -Lipschitz with respect to the first variable), we see that the random vector $X_\varepsilon = X + \varepsilon G$ satisfies the modified log-Sobolev inequality

$$\text{Ent}(e^{f(X_\varepsilon)}) \leq C_\varepsilon \mathbb{E} |\nabla f(X_\varepsilon)|^2 e^{f(X_\varepsilon)}, \tag{4.5}$$

where $C_\varepsilon = C(1 + \varepsilon^2)$. Note that the law of X_ε is absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^n , and so almost surely X_ε is a differentiability point of f and $|\nabla f(X_\varepsilon)|$ coincides with the Euclidean length of the “true” gradient $\nabla f(X_\varepsilon)$.

Moreover, (4.5) can be rewritten in the form

$$\text{Ent}(e^{f(X_\varepsilon)}) \leq \mathbb{E} \alpha_\varepsilon^* (\nabla f(X_\varepsilon)) e^{f(X_\varepsilon)}, \tag{4.6}$$

where $\alpha_\varepsilon^* : \mathbb{R}^n \rightarrow \mathbb{R}$ is the Legendre transform of

$$\alpha_\varepsilon(s) = \frac{|s|^2}{4C_\varepsilon} 1_{\{|s| \leq 2C_\varepsilon L\}} + (L|s| - L^2 C_\varepsilon) 1_{\{|s| > 2C_\varepsilon L\}},$$

that is, more explicitly,

$$\alpha_\varepsilon^*(s) = \begin{cases} C_\varepsilon |s|^2 & \text{for } |s| \leq L, \\ +\infty & \text{for } |s| > L. \end{cases}$$

If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex, Lipschitz (with arbitrary Lipschitz constant) and bounded from below, then $Q_t^{\alpha_\varepsilon} f$ is well defined, convex (as an infimum convolution of two convex functions), and L -Lipschitz for $t > 0$ (since $Q_t^{\alpha_\varepsilon} f(x) = \inf_{y \in \mathbb{R}^n} \{f(y) + t\alpha_\varepsilon((y-x)/t)\}$ and the function $x \mapsto t\alpha_\varepsilon((y-x)/t)$ is L -Lipschitz for $t > 0$).

Moreover, the function $u(t, x) = Q_t^{\alpha_\varepsilon} f(x)$ is Lipschitz on $(0, \infty) \times \mathbb{R}^n$ and satisfies the Hamilton–Jacobi equation

$$\frac{d}{dt} u(t, x) + \alpha_\varepsilon^*(\nabla_x u(t, x)) = 0 \quad \text{for Lebesgue almost all } (t, x) \in (0, \infty) \times \mathbb{R}^n$$

(see Proposition A.1 in the Appendix). Set

$$F(t) = \frac{1}{t} \ln(\mathbb{E} e^{t Q_t^{\alpha_\varepsilon} f(X_\varepsilon)}), \quad t \in (0, 1].$$

(Note that $F(t) < \infty$ since $Q_t^{\alpha_\varepsilon} f$ is L -Lipschitz.) Using the integrability properties of X (and as a consequence of X_ε), together with the Lipschitz property of u it is not difficult to see that F is locally Lipschitz and for Lebesgue almost all $t \in (0, 1)$,

$$\begin{aligned} \frac{d}{dt} F(t) &= -\frac{1}{t^2} \ln(\mathbb{E} e^{t Q_t^{\alpha_\varepsilon} f(X_\varepsilon)}) + \frac{1}{t} \frac{\mathbb{E} e^{t Q_t^{\alpha_\varepsilon} f(X_\varepsilon)} (Q_t^{\alpha_\varepsilon} f(X_\varepsilon) + t \frac{d}{dt} Q_t^{\alpha_\varepsilon} f(X_\varepsilon))}{\mathbb{E} e^{t Q_t^{\alpha_\varepsilon} f(X_\varepsilon)}} \\ &= \frac{1}{t^2 \mathbb{E} e^{t Q_t^{\alpha_\varepsilon} f(X_\varepsilon)}} (\text{Ent}(e^{t Q_t^{\alpha_\varepsilon} f(X_\varepsilon)}) - t^2 \mathbb{E} \alpha_\varepsilon^*(\nabla Q_t^{\alpha_\varepsilon} f(X_\varepsilon)) e^{t Q_t^{\alpha_\varepsilon} f(X_\varepsilon)}) \\ &\leq \frac{1}{\mathbb{E} e^{t Q_t^{\alpha_\varepsilon} f(X_\varepsilon)}} C_\varepsilon \mathbb{E} (|\nabla Q_t^{\alpha_\varepsilon} f(X_\varepsilon)|^2 - |Q_t^{\alpha_\varepsilon} f(X_\varepsilon)|^2) e^{t Q_t^{\alpha_\varepsilon} f(X_\varepsilon)} = 0, \end{aligned}$$

where we used (4.6), the definition of α_ε^* , and the fact that $Q_t^{\alpha_\varepsilon} f$ is L -Lipschitz. Thus,

$$F(1) \leq \liminf_{t \rightarrow 0^+} F(t) \leq \lim_{t \rightarrow 0^+} \frac{\ln(\mathbb{E} e^{t f(X_\varepsilon)})}{t} = \mathbb{E} f(X_\varepsilon),$$

or, in other words,

$$\mathbb{E} e^{Q_1^{\alpha_\varepsilon} f(X_\varepsilon)} \leq e^{\mathbb{E} f(X_\varepsilon)}.$$

It is easy to see that by taking $\varepsilon \rightarrow 0$ we arrive at the assertion of the lemma (recall that f and $Q_1^{\alpha_\varepsilon}$ are Lipschitz and $\alpha_\varepsilon \leq \alpha$).

Suppose now that the log-Sobolev inequality (4.2) holds for all *concave* and L -Lipschitz functions. As before, we pass to the random vector X_ε which has an absolutely continuous distribution. Let $g: \mathbb{R}^n \rightarrow \mathbb{R}$ be convex and bounded from below. Then the function $f = -Q_1^{\alpha_\varepsilon} g$ is concave and L -Lipschitz. The same calculation as above yields

$$\mathbb{E}e^{Q_1^{\alpha_\varepsilon} f(X_\varepsilon)} \leq e^{\mathbb{E}f(X_\varepsilon)},$$

or equivalently

$$\mathbb{E}e^{Q_1^{\alpha_\varepsilon} (-Q_1^{\alpha_\varepsilon} g)(X_\varepsilon)} \leq e^{-\mathbb{E}Q_1^{\alpha_\varepsilon} g(X_\varepsilon)}.$$

We stress that now, in order to prove the Hamilton–Jacobi equations via Proposition A.1, we need to use the L -Lipschitz property of f , since in general f is not bounded from below.

Since

$$-g(x) \leq \inf_{y \in \mathbb{R}^n} \sup_{z \in \mathbb{R}^n} \{-g(z) - \alpha_\varepsilon(z - y) + \alpha_\varepsilon(y - x)\} = Q_1^{\alpha_\varepsilon} (-Q_1^{\alpha_\varepsilon} g)(x)$$

(to verify the inequality take $z = x$), a limit argument yields the assertion of the lemma. □

We are now ready for the proof of our main result.

Proof of Theorem 1.3. The implication (i) \implies (ii) follows immediately from Corollaries 4.3 and 4.4, and the definition of $\overline{\mathbf{T}}_{\theta_{2C}, c}$. To obtain the reverse implication one can use a standard Taylor expansion argument. Assume that $\overline{\mathbf{T}}_{\theta_{C,D}}$ holds. Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be convex, Lipschitz, and bounded from below. For $x \in \mathbb{R}^n$ denote

$$f^x(z) = f(x) + \langle u_x, z - x \rangle, \quad z \in \mathbb{R}^n,$$

where u_x is any subgradient of f at x , so that $f^x \leq f$ on \mathbb{R}^n . Taking $z = x + \varepsilon u_x$ with $\varepsilon \rightarrow 0$ we see that $|u_x| \leq |\nabla f(x)|$.

For sufficiently small ε we have $\varepsilon |\nabla f(x)| \leq D$ for all $x \in \mathbb{R}^n$, and hence

$$\begin{aligned} Q_2^{\theta_{C,D}}(\varepsilon f)(x) &\geq \inf_{y \in \mathbb{R}^n} \{\varepsilon f^x(x - y) + 2\theta_{C,D}(y/2)\} \\ &= \varepsilon f(x) + \inf_{y \in \mathbb{R}^n} \{-\varepsilon \langle u_x, y \rangle + 2\theta_{C,D}(y/2)\} \\ &= \varepsilon f(x) - 2\theta_{C,D}^*(\varepsilon u_x) \geq \varepsilon f(x) - \varepsilon^2 C |\nabla f(x)|^2 \end{aligned}$$

(recall that $|u_x| \leq |\nabla f(x)|$). We now substitute εf into the dual formulation (2.6) and use the above estimate. An inspection of the Taylor expansions up to order ε^2 yields

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

This ends the proof. □

5. Operations preserving the convex Poincaré inequality

We will now discuss several tools which allow to construct measures satisfying the convex Poincaré inequality. To shorten the notation, we will denote by \mathbb{E}_μ and Var_μ , respectively, the mean and variance of f seen as a random variable on \mathbb{R}^n equipped with probability measure μ .

Let us start with the well known tensorization property of variance (see, e.g., [5], Proposition 1.4.1), which asserts that whenever μ_i are probability measures on \mathcal{X}_i , $i = 1, \dots, n$, then the product measure $\mu = \mu_1 \otimes \dots \otimes \mu_n$ on $\mathcal{X}_1 \times \dots \times \mathcal{X}_n$, satisfies the inequality

$$\text{Var}_\mu f \leq \sum_{i=1}^n \mathbb{E}_\mu \text{Var}_{\mu_i} f,$$

for every function $f: \mathcal{X}_1 \times \dots \times \mathcal{X}_n \rightarrow \mathbb{R}$, where $\text{Var}_{\mu_i} f$ denotes the variance of f treated as a function on \mathcal{X}_i , with the other coordinates fixed.

This immediately implies the tensorization property for the convex Poincaré inequality, namely if μ_i ($i = 1, \dots, N$) is a probability measure on \mathbb{R}^{n_i} , satisfying the convex Poincaré inequality with constant λ , then the product measure $\mu = \mu_1 \otimes \dots \otimes \mu_N$ on $\mathbb{R}^{n_1 + \dots + n_N}$ satisfies

$$\text{Var}_\mu f \leq \frac{1}{\lambda} \mathbb{E} \sum_{i=1}^N |\nabla_i f|^2, \quad (5.1)$$

for every convex function $f: \mathbb{R}^{n_1 + \dots + n_N} \rightarrow \mathbb{R}$, where $|\nabla_i f|$ denotes the “partial length of gradient” along \mathbb{R}^{n_i} . If the measures μ_i are absolutely continuous with respect to the Lebesgue measure, then by Rademacher’s theorem locally Lipschitz functions are almost everywhere differentiable, in particular the right-hand side of the above inequality coincides with $\lambda^{-1} \mathbb{E} |\nabla f|^2$ and so we obtain that μ satisfies the convex Poincaré inequality with constant λ . The situation is more delicate for measures which are not absolutely continuous, however thanks to results by Gozlan, Roberto and Samson [17], we can obtain the following simple proposition.

Proposition 5.1. *Assume that μ_i are probability measures on \mathbb{R}^{n_i} , $i = 1, \dots, N$, satisfying the convex Poincaré inequality with constant λ . Then the measure $\mu = \mu_1 \otimes \dots \otimes \mu_N$ on $\mathbb{R}^{n_1 + \dots + n_N}$ satisfies the convex Poincaré inequality with constant λ/C for some universal constant C .*

Proof. We provide only a sketch of the proof, leaving some computational details to the Reader. Denote $n = n_1 + \dots + n_N$ and consider an arbitrary convex smooth 1-Lipschitz function f on \mathbb{R}^n , $k \geq 1$. By (5.1) we have $\text{Var}_{\mu^{\otimes k}} f \leq \lambda^{-1} \mathbb{E}_{\mu^{\otimes k}} |\nabla f|^2 \leq 1/\lambda$. Using an analogous argument as in the proof of Proposition 2.2 (for $p > 2$, to remain in the smooth setting) we arrive at

$$\mu^{\otimes k}(f \geq \text{Med}_{\mu^{\otimes k}} f + t) \leq 8e^{-\sqrt{\lambda}t/2} \quad (5.2)$$

for all 1-Lipschitz smooth convex functions. We can extend this inequality to arbitrary 1-Lipschitz convex function (approximating them with 1-Lipschitz smooth convex functions, for

example, by convolving them with Gaussian densities, see [28], p. 429), so in particular we get that for any convex set $A \subseteq \mathbb{R}^{nk}$, with $\mu^{\otimes k}(A) \geq 1/2$, and all $t > 0$,

$$\mu^{\otimes k}(A + tB_2^{nk}) \geq 1 - 8e^{-\sqrt{\lambda}t/2},$$

where B_2^{nk} is the unit Euclidean ball in \mathbb{R}^{nk} . Recall the notation

$$|\nabla^- f(x)| = \limsup_{y \rightarrow x} \frac{(f(y) - f(x))_-}{|x - y|}.$$

By [17], Theorem 6.7, the dimension-free subexponential concentration for convex sets of the form (5.2) implies that μ satisfies the Poincaré inequality

$$\text{Var}_\mu f \leq \frac{1}{\lambda'} \mathbb{E} |\nabla^- f|^2 \leq \frac{1}{\lambda'} \mathbb{E} |\nabla f|^2 \tag{5.3}$$

for all convex functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$, where

$$\sqrt{\lambda'} = \sup \left\{ \frac{\bar{\Phi}^{-1}(8 \exp(-\sqrt{\lambda}r/2))}{r} : r \geq \frac{2 \log(16)}{\sqrt{\lambda}} \right\},$$

where $\bar{\Phi}$ is the Gaussian tail function. Using the estimate $\bar{\Phi}(x) \geq \frac{1}{2}e^{-x^2}$ and performing some elementary calculations, we arrive at the assertion of the proposition. \square

Remark 5.2. The above argument shows that if μ satisfies the Poincaré inequality (1.1) then it also satisfies the formally stronger inequality (5.3) with $\lambda' = \lambda/C$. We remark that in the category of all Lipschitz functions it is known that the Poincaré inequalities with the length of gradients $|\nabla^- f|$ and $|\nabla f|$ are equivalent and the involved constants do not change (cf. [17], Remark 1.1).

Tensorization allows in particular to pass from one-dimensional measures satisfying the convex Poincaré inequality (characterized in [8]) to product measures in higher dimensions. Another standard tool for producing new examples is perturbation: if μ satisfies the convex Poincaré inequality with constant λ and ν is a measure with density e^U with respect to μ , then ν satisfies the convex Poincaré inequality with constant $\lambda \exp(\inf U - \sup U)$. For the proof see, for example, [5], Chapter 3.4, (the proof therein is written in the context of Markov processes and Dirichlet forms but it is based only on the elementary observation that $\text{Var} f = \inf_{a \in \mathbb{R}} \mathbb{E} |f - a|^2$ and works in exactly the same way in the convex setting).

Perturbation and tensorization are tools that appeared for the first time in the “classical” theory of Poincaré and log-Sobolev inequalities for smooth (or locally Lipschitz) functions. The next proposition does not have a counterpart in the classical setting and significantly extends the set of tools for creating new examples. Namely, we will show that the convex Poincaré inequality passes to mixtures of measures. This cannot be the case for the classical Poincaré inequality since it clearly cannot hold for measures with disconnected support. We note however that the preservation of the Poincaré and log-Sobolev inequalities by mixtures of measures with overlapping

supports has been investigated by Chafaï and Malrieu in [11]. In particular, the Proposition 5.3 below has been inspired by calculations in Section 4.1 therein.

Let $\mathcal{T}_2(\mu_0, \mu_1)$ stand for the usual Kantorovich transport cost between μ_1 and μ_0 (defined by taking $\theta(x) = |x|^2$ in (1.5)), in other words the square of the Kantorovich–Wasserstein distance W_2 .

Proposition 5.3. *Let $\{\mu_\theta\}_{\theta \in \Theta}$ be a family of probability measures on \mathbb{R}^n which satisfy the convex Poincaré inequality (1.1) with constants $\{\lambda_\theta\}_{\theta \in \Theta}$ respectively. Let ν be a probability measure on Θ and assume that for each Borel set $A \subseteq \mathbb{R}^n$, the map $\theta \rightarrow \mu_\theta(A)$ is measurable. Then the measure $\mu = \int \mu_\theta d\nu(\theta)$, satisfies the convex Poincaré inequality (1.1) with constant*

$$\lambda = \left(\sup_{\theta \in \Theta} \{1/\lambda_\theta\} + 2 \operatorname{diam}(\{\mu_\theta\}_{\theta \in \Theta})^2 \right)^{-1},$$

where $\operatorname{diam}(\{\mu_\theta\}_{\theta \in \Theta})^2 = \sup_{\theta_1, \theta_2 \in \Theta} \mathcal{T}_2(\mu_{\theta_1}, \mu_{\theta_2})$.

Proof. If $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex function, then as one can easily check,

$$\begin{aligned} \operatorname{Var}_\mu(f) &= \int \operatorname{Var}_{\mu_\theta}(f) d\nu(\theta) + \frac{1}{2} \iint (\mathbb{E}_{\mu_{\theta_1}} f - \mathbb{E}_{\mu_{\theta_2}} f)^2 d\nu(\theta_1) d\nu(\theta_2) \\ &\leq \sup_{\theta \in \Theta} \{1/\lambda_\theta\} \mathbb{E}_\mu |\nabla f|^2 + \frac{1}{2} \iint (\mathbb{E}_{\mu_{\theta_1}} f - \mathbb{E}_{\mu_{\theta_2}} f)^2 d\nu(\theta_1) d\nu(\theta_2) \end{aligned}$$

and it suffices to estimate the last term.

For fixed $\theta_1, \theta_2 \in \Theta$ let X and Y be random vectors in \mathbb{R}^n with laws μ_{θ_1} and μ_{θ_2} respectively. By convexity of f ,

$$\begin{aligned} |\mathbb{E}f(X) - \mathbb{E}f(Y)| &\leq \mathbb{E}(|\nabla f(X)| + |\nabla f(Y)|) |X - Y| \\ &\leq \left(\sqrt{\mathbb{E}|\nabla f(X)|^2} + \sqrt{\mathbb{E}|\nabla f(Y)|^2} \right) \sqrt{\mathbb{E}|X - Y|^2}. \end{aligned}$$

Taking the infimum over all realizations of X and Y , we conclude that

$$(\mathbb{E}_{\mu_{\theta_1}} f - \mathbb{E}_{\mu_{\theta_2}} f)^2 \leq 2(\mathbb{E}_{\mu_{\theta_1}} |\nabla f|^2 + \mathbb{E}_{\mu_{\theta_2}} |\nabla f|^2) \operatorname{diam}(\{\mu_\theta\}_{\theta \in \Theta})^2.$$

Thus,

$$\begin{aligned} &\frac{1}{2} \iint (\mathbb{E}_{\mu_{\theta_1}} f - \mathbb{E}_{\mu_{\theta_2}} f)^2 d\nu(\theta_1) d\nu(\theta_2) \\ &\leq \iint (\mathbb{E}_{\mu_{\theta_1}} |\nabla f|^2 + \mathbb{E}_{\mu_{\theta_2}} |\nabla f|^2) d\nu(\theta_1) d\nu(\theta_2) \operatorname{diam}(\{\mu_\theta\}_{\theta \in \Theta})^2 \\ &= 2 \operatorname{diam}(\{\mu_\theta\}_{\theta \in \Theta})^2 \mathbb{E}_\mu |\nabla f|^2. \end{aligned}$$

This implies the assertion of the proposition. □

Example 5.4. Having the above results concerning the preservation of the convex Poincaré inequality under appropriate transformations of measures, as well as the characterization of one-dimensional measures satisfying it (obtained by Bobkov and Götze in [8]), one can create many examples of measures in high dimensions, satisfying the convex Poincaré inequalities but not satisfying the usual Poincaré inequality for smooth functions or any stronger functional inequalities for convex functions.

To illustrate this, we will provide a specific class of examples, built from perturbation, products and mixtures of one dimensional measures. Fix $h > 0$ and consider a bounded set Θ in \mathbb{R}^d and a probability measure ν supported on Θ . For each $x \in \Theta$ let μ_x be a measure on \mathbb{R} such that for all $t > \text{Med } \mu_x$, $\mu_x([t + h, \infty)) \leq \frac{1}{2}\mu_x([t, \infty))$ and a symmetric condition is satisfied by the lower tail of the measure. Assume also that the set $\{\text{Med } \mu_x\}_{x \in \Theta}$ is bounded by some constant l . By the results in [8], the measures μ_x satisfy (1.1) with constant λ_x uniformly bounded from below by some $c_h > 0$. It is easy to see that by Proposition 5.3, the measure μ on $\mathbb{R}^{d+1} = \mathbb{R}^d \times \mathbb{R}$ defined as $\mu(dx dt) = \mu_x(dt)\nu(dx)$ satisfies (1.1) with $\lambda > 0$, depending only on h, l and the diameter of Θ . Note that if $\text{supp } \nu \subseteq \Theta$ is not connected, then μ cannot satisfy the usual Poincaré inequality for smooth functions. Moreover, if for x from a set of positive ν -measure, the measures μ_x have tails bounded from below by e^{-cx} for some $c < \infty$ (which may happen under the requirements on μ_x introduced above), then at least one marginal of μ also has exponential tails, which means that μ cannot satisfy any weak transportation inequality or convex modified log-Sobolev inequality substantially stronger than the convex Poincaré inequality. Now, thanks to stability of the convex Poincaré inequality under bounded perturbations and tensorization, one can pass to products and perturbations of measures μ (for various choices of the driving measure ν) and build more complicated examples in higher dimension, still satisfying the convex Poincaré inequality with constants depending only on h, l and the diameter of Θ .

6. Refined concentration of measure derived from infimum convolution inequalities

In this section, we explain what concentration inequalities for convex functions can be obtained from general infimum convolution inequalities of the form (2.6). While some parts of our derivation are well known and are included only for the sake of completeness, we also provide new inequalities valid beyond the setting of Lipschitz functions. Their proofs are elementary but to our best knowledge they have not been noted in the literature before.

Throughout this section, $\theta: \mathbb{R}^n \rightarrow [0, \infty)$ is a convex function. We also assume the following conditions:

- $\theta(x) = \theta(-x)$ for all $x \in \mathbb{R}^n$,
- $\theta(x) = 0$ if and only if $x = 0$ (in particular, by convexity, $\lim_{x \rightarrow \infty} \theta(x) = \infty$).

We remark that at the cost of some technical work one can obtain the results we present below for more general cost functions (e.g., taking the value ∞ or not satisfying the symmetry condition). We restrict to a smaller class to simplify the presentation.

In what follows, for a function $f: \mathbb{R}^n \rightarrow \mathbb{R}$, bounded from below, we set

$$Qf(x) = Q_1^\theta f(x) = \inf_{y \in \mathbb{R}^n} \{f(y) + \theta(x - y)\}.$$

We also denote

$$B_\theta(r) = \{x \in \mathbb{R}^n : \theta(x) < r\}, \quad r > 0.$$

6.1. Enlargements of sets and concentration for Lipschitz functions

Let us start with the classical description of concentration of measure in terms of enlargements of sets. The following proposition goes back to [24].

Proposition 6.1. *Assume that μ is a probability measure on \mathbb{R}^n , satisfying*

$$\int_{\mathbb{R}^n} e^{Qf} d\mu \int_{\mathbb{R}^n} e^{-f} d\mu \leq 1 \tag{6.1}$$

for all convex functions $f: \mathbb{R}^n \rightarrow \mathbb{R}$, bounded from below. Then for all convex subsets $A \subseteq \mathbb{R}^n$ and $r > 0$, we have

$$\mu((A + B_\theta(r))^c) \mu(A) \leq e^{-r}.$$

Proof. Consider $f = \infty 1_{(cA)^c}$ and note that $Qf(x) < r$ if and only if there exists $y \in A$ such that $\theta(x - y) < r$. Applying the inequality (6.1) to f (which can be justified by monotone approximation), we obtain

$$e^r \mu((A + B_\theta(r))^c) \mu(A) \leq \int_{\mathbb{R}^n} e^{Qf} d\mu \int_{\mathbb{R}^n} e^{-f} d\mu \leq 1. \quad \square$$

To formulate corollaries to the above proposition we need to introduce new notation, which at first may seem rather abstract. However, as the examples presented in the subsequent parts of this section will show, it will prove useful in providing a uniform framework for concentration inequalities, especially in the non-Lipschitz case.

Definition 6.2. Define the norm $|\cdot|_{\frac{1}{p}\theta}$ on \mathbb{R}^n , as the Orlicz norm corresponding to the function $x \mapsto \frac{1}{p}\theta(x)$, i.e.

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

Define also the norm $|\cdot|_{\theta,p}$ on \mathbb{R}^n as the dual to $|\cdot|_{\frac{1}{p}\theta}$, that is,

$$|x|_{\theta,p} = \sup \left\{ \sum_{i=1}^n x_i y_i : \theta(y) \leq p \right\}.$$

The norm $|x|_{\theta,p}$ is equivalent (up to universal constants) to the Orlicz norm $|\cdot|_{\theta_p^*}$ related to the function $\theta_p^*(x) = \frac{1}{p}\theta^*(px)$, explicitly given by

$$|\cdot|_{\theta_p^*} = \inf\{a > 0 : \theta_p^*(x/a) \leq 1\} = \inf\{a > 0 : \theta^*(px/a) \leq p\}.$$

It was observed by Gluskin and Kwapien in [15] that norms of this type play an important role in moment estimates for sums of independent random variables. Recently it has been noted [1,3] that they also appear in moment estimates for smooth functions of random vectors satisfying modified log-Sobolev inequalities. Since in the context of transportation or infimum convolution inequalities one starts from the function θ and not from θ^* (which is the case in the corresponding log-Sobolev setting) it is more convenient to work with $|\cdot|_{\theta,p}$ rather than with the equivalent norm $|\cdot|_{\theta^*}$ used in [1,3].

In what follows we will need a simple inequality which can be easily derived from convexity of θ and the assumption $\theta(0) = 0$. For $x \in \mathbb{R}^n$, $p > 0$, and $t \geq 1$,

$$|x|_{\theta,tp} \leq t|x|_{\theta,p}. \tag{6.2}$$

The following corollary to Proposition 6.1 is again based on by now standard arguments, written however in the language of the norms $|\cdot|_{\theta,p}$.

Corollary 6.3. *Let X be a random vector with law μ , satisfying (6.1) for all convex functions $f: \mathbb{R}^n \rightarrow \mathbb{R}$ bounded from below. Then for any smooth convex Lipschitz function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and $p \geq 0$,*

$$\mathbb{P}\left(|f(X) - \text{Med } f(X)| > \sup_{x \in \mathbb{R}^n} |\nabla f(x)|_{\theta,p}\right) \leq 4e^{-p}. \tag{6.3}$$

Remark 6.4. It is easy to see that if the inequality (6.3) holds for all smooth convex Lipschitz functions, then one can apply it to arbitrary convex Lipschitz function, replacing $\sup_{x \in \mathbb{R}^n} |\nabla f(x)|_{\theta,p}$ by the Lipschitz constant of f with respect to the norm $|\cdot|_{\frac{1}{p}\theta}$. To verify this it is enough to consider convolutions of f with a sequence of Gaussian densities converging to Dirac’s mass at zero – they are smooth, have the same Lipschitz constant as f and converge to f uniformly (see, e.g., [28], p. 429).

Proof of Corollary 6.3. Let $A = \{y \in \mathbb{R}^n : f(y) \leq \text{Med } f(X)\}$, so that $\mathbb{P}(X \in A) \geq 1/2$. Then by convexity, for any $y \in A$,

$$f(X) \leq f(y) + \langle \nabla f(X), X - y \rangle \leq \text{Med } f(X) + |\nabla f(X)|_{\theta,p} \cdot |X - y|_{\frac{1}{p}\theta}. \tag{6.4}$$

Thus,

$$\begin{aligned} & \mathbb{P}\left(f(X) > \text{Med } f(X) + \sup_{x \in \mathbb{R}^n} |\nabla f(x)|_{\theta,p}\right) \\ & \leq \mathbb{P}\left(\inf_{y \in A} |X - y|_{\frac{1}{p}\theta} > 1\right) \\ & = \mathbb{P}(X \notin A + \text{cl } B_\theta(p)) \leq \frac{e^{-p}}{\mathbb{P}(X \in A)} \leq 2e^{-p}, \end{aligned} \tag{6.5}$$

where in the second inequality we used Proposition 6.1.

Let now $A = \{y \in \mathbb{R}^n : f(y) < \text{Med } f(X) - \sup_{x \in \mathbb{R}^n} |\nabla f(x)|_{\theta,p}\}$. Similarly as above, we obtain

$$\begin{aligned} 1/2 &\leq \mathbb{P}(f(X) \geq \text{Med } f(X)) \leq \mathbb{P}\left(\inf_{y \in A} |X - y|_{\frac{1}{p}\theta} \geq 1\right) \\ &\leq \mathbb{P}(X \notin A + B_{\theta}(p)) \leq \frac{e^{-p}}{\mathbb{P}(X \in A)}, \end{aligned}$$

which shows that

$$\mathbb{P}\left(f(X) < \text{Med } f(X) - \sup_{x \in \mathbb{R}^n} |\nabla f(x)|_{\theta,p}\right) \leq 2e^{-p}.$$

Combining the last inequality with (6.5) proves the corollary. □

6.2. Concentration inequalities for general convex functions

We are now ready to state the main result of this section, contained in the following theorem, dealing with general (not necessarily Lipschitz) convex functions. In its formulation, we adopt the convention $\frac{0}{0} = 0$. The proof of the theorem as well as of its corollary is postponed to Section 6.3

We would like to emphasize, that in the theorem we assume only (6.3), which is strictly weaker than the infimum-convolution inequality (6.1).

Theorem 6.5. *Let X be a random vector satisfying (6.3) for all smooth convex Lipschitz functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Then for any smooth convex function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, the following properties hold.*

(i) For any $p \geq 1$,

$$\left\| \frac{(f(X) - \text{Med } f(X))_+}{|\nabla f(X)|_{\theta,p}} \right\|_p \leq 3^{1/p}. \tag{6.6}$$

(ii) Let $p > 0$, $q \in (1/2, 1]$ and let $M_{p,q} \in \mathbb{R}$ satisfy $\mathbb{P}(|\nabla f(X)|_{\theta,p} \leq M_{p,q}) \geq q$. Then

$$\mathbb{P}(f(X) < \text{Med } f(X) - M_{p,q}(1 + \log(8/(2q - 1)))) \leq 4e^{-p}.$$

In particular for $p \geq 0$,

$$\mathbb{P}(f(X) < \text{Med } f(X) - 16\mathbb{E}|\nabla f(X)|_{\theta,p}) \leq 4e^{-p}. \tag{6.7}$$

(iii) For all $p \geq 1$,

$$\|(f - \text{Med } f(X))_-\|_p \leq 48\mathbb{E}|\nabla f(X)|_{\theta,p}.$$

Remark 6.6. As will become clear in the proof, the part (i) of the above theorem holds in fact under one-sided concentration, that is, it is enough to assume that

$$\mathbb{P}\left(f(X) - \text{Med } f(X) > \sup_{x \in \mathbb{R}^n} |\nabla f(x)|_{\theta,p}\right) \leq 4e^{-p}. \tag{6.8}$$

Let us now illustrate the above theorem with a few concrete examples and a corollary. In particular, we will show what the norms $|\cdot|_{\theta,p}$ look like for various choices of the cost function θ .

Example 6.7. If $\theta(x) = c|x|^r$ for some $r \geq 1$ and $c > 0$, then $|x|_{\theta,p} = c^{-1/r} p^{1/r} |x|$ and (6.3) is equivalent to

$$\mathbb{P}(|f(X) - \text{Med } f(X)| \geq t) \leq 4 \exp(-ct^r) \tag{6.9}$$

for all 1-Lipschitz convex functions (in particular for $r = 2$ we get the sub-Gaussian concentration). The first part of Theorem 6.5 gives then the following inequality for all (not necessarily Lipschitz) convex functions and $p \geq 1$,

$$\left\| \frac{(f(X) - \text{Med } f(X))_+}{|\nabla f(X)|} \right\|_p \leq 3^{1/p} c^{-1/r} p^{1/r}.$$

Thus by the L^p -Chebyshev inequality, with $p = ct^r/(3e)^r$, we obtain for $t \geq 0$,

$$\mathbb{P}\left(\frac{f(X) - \text{Med } f(X)}{|\nabla f(X)|} \geq t\right) \leq e \exp\left(-\frac{ct^r}{(3e)^r}\right) \tag{6.10}$$

(the additional factor e on the right-hand side is introduced artificially to encompass all $t \geq 0$, also those for which $p < 1$; note that in this case the right-hand side exceeds one). We remark that similar self-normalized inequalities are known for example, in the theory of empirical processes (see [12]).

The lower tail inequalities give

$$\mathbb{P}(f(X) \leq \text{Med } f(X) - t) \leq 4 \exp\left(-c \frac{t^r}{16^r (\mathbb{E}|\nabla f(X)|)^r}\right). \tag{6.11}$$

Moreover, using the full strength of part (ii) of Theorem 6.5, one can replace $\mathbb{E}|\nabla f(X)|$ by $4^{-1}M_{3/4}$, where $M_{3/4}$ is the 3/4 quantile of $|\nabla f(X)|$. Thus, no integrability conditions on the gradient are in fact required.

Remark 6.8. Let us note that inequalities similar to (6.11) were previously known with the quantity $(\mathbb{E}|\nabla f(X)|^2)^{1/2}$ instead of the quantile or $\mathbb{E}|\nabla f(X)|$ (see [28] or [22], Chapter 3.3). Very recently, Paouris and Valettas [26] have proved that the standard Gaussian vector in \mathbb{R}^n satisfies a similar inequality (for $r = 2$) with $\mathbb{E}|f(X) - \text{Med } f(X)|$ in place of $\mathbb{E}|\nabla f(X)|$. Their proof uses in a crucial way isoperimetric properties of Gaussian measures. The version with $\mathbb{E}|\nabla f(X)|$ follows simply by an application of the (1, 1)-Poincaré inequality for the Gaussian measure, that is, $\mathbb{E}|f(X) - \text{Med } f(X)| \leq C\mathbb{E}|\nabla f(X)|$ (see, e.g., [25,27]). In fact, the proof in

[26] gives also inequalities in terms of quantiles of $|f(X) - M|$. We do not know if they are comparable to our estimates (specialized to the standard Gaussian measure) in terms of quantiles of $|\nabla f(X)|$.

Note also that (6.9) for $r = 1$ is a consequence of the convex Poincaré inequality (however, we do not know if (1.1) implies (6.9) with c depending only on λ and not on the dimension n , see Question 7.3 below).

Example 6.9. Let us now consider a measure μ on \mathbb{R}^n satisfying the convex Poincaré inequality with constant λ . Then, by Theorem 3.1 it satisfies the convex Bobkov–Ledoux inequality (3.1) with constants C and c depending only on λ . By the classical Herbst argument, it follows (see, e.g., [2,6]) that for each $N \geq 1$, if X is an Nn -dimensional random vector with law $\mu^{\otimes N}$, then for any smooth convex function $f : \mathbb{R}^{Nn} \rightarrow \mathbb{R}$ and any $t > 0$,

$$\begin{aligned} & \mathbb{P}(f(X) \geq \mathbb{E}f(X) + t) \\ & \leq 2 \exp\left(-c'(\lambda) \min\left\{\frac{t^2}{\sup_{x \in \mathbb{R}^{Nn}} |\nabla f(x)|^2}, \frac{t}{\sup_{x \in \mathbb{R}^{Nn}} \max_{i \leq N} |\nabla_i f(x)|}\right\}\right), \end{aligned}$$

where for $x = (x_1, \dots, x_N) \in (\mathbb{R}^n)^N = \mathbb{R}^{Nn}$, $\nabla_i f(x)$ denotes the partial gradient with respect to x_i .

Moreover, by the Poincaré inequality

$$|\mathbb{E}f(X) - \text{Med } f(X)| \leq \frac{1}{\sqrt{\lambda}} \sup_{x \in \mathbb{R}^{Nn}} |\nabla f(x)|,$$

which at the cost of changing the constant allows to replace the mean by the median in the above inequality. Thus, we obtain that for some constant $c''(\lambda)$ and $p > 0$,

$$\mathbb{P}\left(f(X) \geq \text{Med } f(X) + c''(\lambda) \sup_{x \in \mathbb{R}^{Nn}} \left(\sqrt{p} |\nabla f(x)| + p \max_{i \leq N} |\nabla_i f(x)|\right)\right) \leq 2e^{-p}.$$

It is easy to see that up to universal constants $c''(\lambda)(\sqrt{p}|x| + p \max_{i \leq N} |x_i|)$ is equivalent to $|x|_{\theta, p}$, where

$$\theta(x) = \sum_{i=1}^N \left(\left| \frac{x_i}{c''(\lambda)} \right|^2 1_{\left\{ \left| \frac{x_i}{c''(\lambda)} \right| \leq 1 \right\}} + \left(2 \left| \frac{x_i}{c''(\lambda)} \right| - 1 \right) 1_{\left\{ \left| \frac{x_i}{c''(\lambda)} \right| > 1 \right\}} \right)$$

(e.g., by comparing θ with $\theta_1(x) = \sum_{i=1}^N \min(|x_i/c''(\lambda)|^2, |x_i/c''(\lambda)|)$, which is not convex but is comparable to θ up to multiplicative constants). More precisely,

$$|x|_{\theta, p} \leq c''(\lambda) \left(\sqrt{p}|x| + p \max_{i \leq N} |x_i| \right) \leq 4|x|_{\theta, p}.$$

Thus, the first part of Theorem 6.5 together with Remark 6.6 gives for arbitrary smooth convex function f on \mathbb{R}^{Nn} , the inequality

$$\left\| \frac{(f(X) - \text{Med } f(X))_+}{\sqrt{p}|\nabla f(X)| + p \max_{i \leq N} |\nabla_i f(X)|} \right\|_p \leq c'''(\lambda),$$

for $p \geq 1$, where $c'''(\lambda)$ depends only on λ . By Chebyshev's inequality, this implies that

$$\mathbb{P}\left(\frac{(f(X) - \text{Med } f(X))_+}{\sqrt{t}|\nabla f(X)| + t \max_{i \leq N} |\nabla_i f(X)|} \geq ec'''(\lambda)\right) \leq e^{-t}$$

for $t \geq 1$ (note that contrary to (6.10) this time t cannot be removed from the denominator).

As for the lower tail, by Theorem 1.3, Remark 1.4, Lemma 2.5 and tensorization properties of infimum convolution inequalities (see Lemma 5 in [24]) we obtain that X satisfies (6.1) and thus also (6.3) with $\theta(x) = K(\lambda, n) \sum_{i=1}^N (|x_i|^2 1_{\{|x_i| \leq 1\}} + (2|x_i| - 1) 1_{\{|x_i| > 1\}})$, where $K(\lambda, n)$ depends only on λ and the dimension n . Thus, by the second part of Theorem 6.5,

$$\mathbb{P}\left(f(X) \leq \text{Med } f(X) - K'(\lambda, n)^{-1} \left[\sqrt{p} \mathbb{E}|\nabla f(X)| + p \mathbb{E} \max_{i \leq N} |\nabla_i f(X)| \right]\right) \leq 4e^{-p},$$

or equivalently (up to constants depending only on λ, n),

$$\mathbb{P}(f(X) \leq \text{Med } f(X) - t) \leq 4 \exp\left(-K''(\lambda, n) \min\left\{\frac{t^2}{(\mathbb{E}|\nabla f(X)|)^2}, \frac{t}{\mathbb{E} \max_{i \leq N} |\nabla_i f(X)|}\right\}\right).$$

We stress that all the above inequalities are dimension-free in the sense that the constants do not depend on the number N but just on the initial dimension n (cf. Remark 1.5).

Example 6.10. Finally, we remark that general cost functions θ lead to other concentration profiles, which have been studied in the literature. One can for instance, consider products of measures on \mathbb{R} , satisfying (6.1) with

$$\theta(x) \simeq |x|^2 1_{\{|x| \leq 1\}} + |x|^r 1_{\{|x| > 1\}}$$

for $r \geq 1$ (such measures are characterized thanks to results in [18]). If we denote for $x \in \mathbb{R}^n$, $|x|_r = (|x_1|^r + \dots + |x_n|^r)^{1/r}$ and let r^* be the Hölder conjugate of r , then such costs correspond for $r \in [1, 2]$ to norms of the form $|x|_{\theta, p} \simeq \sqrt{p}|x| + p^{1/r}|x|_{r^*}$ (the case $r = 1$ has been discussed above), while for $r > 2$ to

$$|x|_{\theta, p} \simeq p^{1/r} |(x_i^*)_{i=1}^p|_{r^*} + \sqrt{p} |(x_i^*)_{i=p+1}^n|,$$

where $(x_i^*)_{i=1}^n$ is the non-increasing rearrangement of the sequence $(|x_i|)_{i=1}^n$.

We will now present a corollary to Theorem 6.5, providing concentration inequalities for non-Lipschitz convex functions, in the spirit of recent results due to Bobkov, Nayar, and Tetali [10].

Corollary 6.11. *Under the assumptions of Theorem 6.5 for all convex functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $t \geq 0$,*

$$\mathbb{P}(f(X) - \text{Med } f(X) \geq t) \leq \inf_{p \geq 1} \{e^{-p} + \mathbb{P}(|\nabla f(X)|_{\theta,p} \geq t/(3e))\}.$$

Moreover, for any $p \geq 1$,

$$\mathbb{P}(|f(X) - \text{Med } f(X)| \geq 3e^2 \|\nabla f(X)\|_{\theta,p}) \leq 6e^{-p}. \tag{6.12}$$

Let us note that inequalities of the form (6.12) have been obtained in [1] for all smooth functions of random vectors satisfying modified log-Sobolev inequalities (assumed to hold for all smooth functions). Therein, the function θ had to satisfy some appropriate growth condition.

Example 6.12. In particular for $\theta(x) = c|x|^2$, the above corollary gives

$$\mathbb{P}(f(X) - \text{Med } f(X) \geq t) \leq \inf_{p \geq 1} \{e^{-p} + \mathbb{P}(\sqrt{p/c} |\nabla f(X)| \geq t/(3e))\}.$$

By substituting $p = \frac{ct^2}{(3e)^2 L^2}$ and adjusting the constant, we obtain

$$\mathbb{P}(f(X) - \text{Med } f(X) \geq t) \leq \inf_{L > 0} \{2e^{-c' \frac{t^2}{L^2}} + \mathbb{P}(|\nabla f(X)| \geq L)\}, \tag{6.13}$$

where c' is positive and depends only on c . The factor 2 in the above inequality is introduced for notational simplicity to allow the whole range of $L > 0$ in the infimum (note that for large L we have $p < 1$ and we cannot apply Corollary 6.11, on the other hand in this case for an appropriate choice of c' the value under the infimum above exceeds one, so we can indeed extend the range of L to the whole positive half-line).

Recall also the second part of Theorem 6.5 which for $q = 3/4$ gives in this case

$$\mathbb{P}(f(X) \leq \text{Med } f(X) - t) \leq 4 \exp\left(-c'' \frac{t^2}{M_{3/4}^2}\right), \tag{6.14}$$

where $M_{3/4} = \inf\{x \in \mathbb{R}^n : \mathbb{P}(|\nabla f(X)| \leq x) \geq 3/4\}$ and c'' again depends only on c .

The above inequalities should be compared with a recent result in [10], which asserts that for some positive constant c''' depending only on c ,

$$\mathbb{P}(|f(X) - f(Y)| \geq t) \leq 2 \inf_{L \geq \text{Med } |\nabla f(X)|} \{e^{-c''' \frac{t^2}{L^2}} + \mathbb{P}(|\nabla f(X)| \geq L)\}, \tag{6.15}$$

where Y is an independent copy of X .

It is not difficult to see that in the regime of t for which the above inequalities are of interest, that is, the right-hand sides are small, (6.13) gives estimates on the upper tail which (up to numerical constants) are comparable to those implied by (6.15), whereas for the lower tail, the inequality (6.14) improves over (6.15).

Example 6.13. Consider the function $\theta(x) = \sum_{i=1}^N (|x_i/c|^2 \mathbf{1}_{\{|x_i/c| \leq 1\}} + (2|x_i/c| - 1) \mathbf{1}_{\{|x_i/c| > 1\}})$, which we have already discussed in Example 6.9. From Corollary 6.11, we get

$$\mathbb{P}(f(X) - \text{Med } f(X) \geq t) \leq \inf_{p \geq 1} \left\{ e^{-p} + \mathbb{P}\left(\sqrt{p} |\nabla f(X)| + p \max_{i \leq N} |\nabla_i f(X)| \geq t/c' \right) \right\}.$$

By substituting $p = \min\{\frac{t^2}{(2c')^2 L^2}, \frac{t}{2c'M}\}$ and using the union bound, we obtain

$$\begin{aligned} &\mathbb{P}(f(X) - \text{Med } f(X) \geq t) \\ &\leq \inf_{L, M > 0} \left\{ 2 \exp\left(-c'' \min\left\{\frac{t^2}{L^2}, \frac{t}{M}\right\}\right) + \mathbb{P}(|\nabla f(X)| \geq L) + \mathbb{P}\left(\max_{i \leq N} |\nabla_i f(X)| \geq M\right) \right\}, \end{aligned}$$

with c'' depending only on c . As in the preceding example, the factor 2 is introduced to allow for all positive values of L, M .

Remark 6.14. Let us note that another way of obtaining estimates on the upper tail of non-Lipschitz convex functions under the convex Poincaré inequality is to use the estimates (2.1) and (2.2). By approximating arbitrary convex functions with Lipschitz ones, we can easily see that they hold in fact for all convex functions. Thus, if one controls the moments of $|\nabla f(X)|$, one can obtain tail estimates beyond the Lipschitz case. Such inequalities are however different from those of the above example as they are of exponential type and not of mixed exponential/Gaussian type. On the other hand, the weak transportation inequality with the cost function of Example 6.9 arises usually as a consequence of tensorization, so in order to apply it we need some additional product structure of the measure.

6.3. Proofs of Theorem 6.5 and Corollary 6.11

Proof of Theorem 6.5. Let us start with (i), the proof of which is quite similar to the proof of Corollary 6.3. Let us again define $A = \{x \in \mathbb{R}^n : f(x) \leq \text{Med } f(X)\}$. Using (6.2) and (6.4), we can write for $t \geq 1$,

$$\frac{f(X) - \text{Med } f(X)}{t |\nabla f(X)|_{\theta, p}} \leq \frac{f(X) - \text{Med } f(X)}{|\nabla f(X)|_{\theta, tp}} \leq \inf_{y \in A} |X - y|_{\frac{1}{tp} \theta}.$$

Hence for $t \geq 1$,

$$\mathbb{P}\left(\frac{f(X) - \text{Med } f(X)}{|\nabla f(X)|_{\theta, p}} > t\right) \leq \mathbb{P}\left(\inf_{y \in A} |X - y|_{\frac{1}{tp} \theta} > 1\right) \leq 4e^{-pt},$$

where we used the fact that the function $g(x) = \inf_{y \in A} |x - y|_{\frac{1}{tp} \theta}$ is convex, 1-Lipschitz with respect to $|\cdot|_{\frac{1}{tp} \theta}$ and $\text{Med } g(X) = 0$, together with Corollary 6.3 and Remark 6.4. We can now

integrate by parts and get

$$\mathbb{E} \left| \frac{(f(X) - \text{Med } f(X))_+}{|\nabla f(X)|_{\theta,p}} \right|^p \leq 1 + 4 \int_1^\infty pt^{p-1} e^{-pt} dt \leq 1 + 4 \int_1^\infty e^{-t} dt \leq 3$$

(the integrand is pointwise non-increasing with respect to $p \geq 1$, as the computation of the derivative with respect to p reveals), which proves the first part of the theorem.

Let us now pass to the second part. Assume without loss of generality that $\text{Med } f(X) = 0$ and $p \geq 1$. Consider the set $B = \{x \in \mathbb{R}^n : |\nabla f(x)|_{\theta,p} \leq M_{p,q}\}$. By the definition of $M_{p,q}$, we have $\mathbb{P}(X \in B) \geq q$. Let $\tilde{f} : \mathbb{R}^n \rightarrow \mathbb{R}$ be defined as

$$\tilde{f}(x) = \sup_{y \in B} \{f(y) + \langle \nabla f(y), x - y \rangle\}.$$

Then \tilde{f} is convex, moreover by convexity of f we have $\tilde{f} \leq f$ pointwise and $\tilde{f} = f$ on B . By the definition of the set B and inequality (6.2), for any $t \geq 1$ all linear functionals $x \mapsto \langle \nabla f(y), x \rangle$, $y \in B$, are $(tM_{p,q})$ -Lipschitz with respect to $|\cdot|_{\frac{1}{t} \theta}$ and therefore so is \tilde{f} . By Corollary 6.3 and Remark 6.4, this implies that for any $t \geq 1$,

$$\mathbb{P}(|\tilde{f}(X) - \text{Med } \tilde{f}(X)| > tM_{p,q}) \leq 4e^{-tp}. \tag{6.16}$$

We also have $\mathbb{P}(\tilde{f}(X) \geq 0) \geq \mathbb{P}(f(X) \geq 0 \text{ and } X \in B) \geq q - 1/2$. Therefore, the above inequality applied with $t \searrow \log(8/(2q - 1)) > 1$ gives (recall that $p \geq 1$)

$$\text{Med } \tilde{f}(X) + M_{p,q} \log(8/(2q - 1)) \geq 0,$$

which by another application of (6.16) implies

$$\mathbb{P}(f(X) < -M_{p,q}(1 + \log(8/(2q - 1)))) \leq \mathbb{P}(\tilde{f}(X) < \text{Med } \tilde{f}(X) - M_{p,q}) \leq 4e^{-p}.$$

This proves the first inequality of part (ii).

The second inequality of part (ii) follows from the first one by specializing to $q = 3/4$, $M_{p,q} = 4\mathbb{E}|\nabla f(X)|_{\theta,p}$ and some elementary calculations.

As for part (iii), using again (6.2) and (6.7), we get for $t \geq 16\mathbb{E}|\nabla f(X)|_{\theta,p}$

$$\mathbb{P}(f(X) - \text{Med } f(X) \leq -t) \leq 4 \exp\left(-\frac{pt}{16\mathbb{E}|\nabla f(X)|_{\theta,p}}\right).$$

Now, again by integration by parts,

$$\begin{aligned} & \mathbb{E}(f(X) - \text{Med } f(X))_-^p \\ & \leq (16\mathbb{E}|\nabla f(X)|_{\theta,p})^p + 4p \int_{16\mathbb{E}|\nabla f(X)|_{\theta,p}}^\infty t^{p-1} \exp\left(-\frac{pt}{16\mathbb{E}|\nabla f(X)|_{\theta,p}}\right) dt \\ & \leq 3(16\mathbb{E}|\nabla f(X)|_{\theta,p})^p, \end{aligned}$$

which ends the proof. □

Proof of Corollary 6.11. To prove the first inequality, it is enough to note that if $|\nabla f(X)|_{\theta,p} \leq t/(3e)$ and $f(X) - \text{Med } f(X) \geq t$, then

$$Z := \frac{(f(X) - \text{Med } f(X))_+}{|\nabla f(X)|_{\theta,p}} \geq 3e \geq e\|Z\|_p,$$

where the last inequality follows from (6.6). The assertion follows thus from Chebyshev’s inequality: $\mathbb{P}(Z \geq e\|Z\|_p) \leq e^{-p}$.

As for the second inequality, we apply the first one with $t = 3e^2\|\nabla f(X)\|_{\theta,p}$ and combine it with the estimate (6.7). □

7. Further questions

Let us conclude with some open questions, which seem natural in view of our results.

As already mentioned in the [Introduction](#), in our proof of the implication

μ satisfies the convex Poincaré inequality with constant λ

$$\implies \mu \text{ satisfies the inequality } \overline{\mathbf{T}}_{\theta_{C,D}}$$

the constants C, D do not depend just on λ , but also on certain quantiles of the measure μ . In fact, the issue comes from the inequality $\overline{\mathbf{T}}^+$, since the constants in $\overline{\mathbf{T}}^-$ do depend only on λ (see Corollary 4.3). This gives rise to our first question.

Question 7.1. *Does the convex Poincaré inequality with constant λ imply the weak transportation inequality $\overline{\mathbf{T}}_{\theta_{C,D}}$ with constants C, D depending only on λ ?*

The inspection of our proof shows that in order to answer the above question in the affirmative, it is enough to remove the restriction on t in Lemma 2.4. An improved version of this lemma, valid for all $t > 0$ would follow by part (ii) of Theorem 6.5 provided that one can show that the convex Poincaré inequality with constant λ implies subexponential concentration for convex 1-Lipschitz functions, with constants depending only on λ . The problem lies in the lower-tail (as the upper one is handled by Proposition 2.2). More precisely, we have the following result.

Theorem 7.2. *Assume that μ is a probability measure on \mathbb{R}^n , satisfying the convex Poincaré inequality (1.1) with constant λ and c is a positive constant, such that for all 1-Lipschitz convex functions $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and all $t > 0$,*

$$\mu(\{x \in \mathbb{R}^n: f(x) \leq \text{Med}_\mu f - t\}) \leq 2 \exp(-ct).$$

Then μ satisfies the inequality $\overline{\mathbf{T}}_{\theta_{C,D}}$ with C, D depending only on λ and c .

This motivates the following question, which is clearly of interest also in its own right.

Question 7.3. *Does the convex Poincaré inequality (1.1) with constant λ imply subexponential estimates for the lower-tail of convex 1-Lipschitz functions, with constants depending only on λ ? Specifically, is it true that whenever μ is a probability measure on \mathbb{R}^n satisfying (1.1), then for every convex 1-Lipschitz function $f : \mathbb{R}^n \rightarrow \mathbb{R}$,*

$$\mu(\{x \in \mathbb{R}^n : f(x) \leq \text{Med}_\mu f - t\}) \leq 2 \exp(-c(\lambda)t),$$

where the constant $c(\lambda)$ depends only on λ ?

The inequality provided by Lemma 2.4 introduces an additional dependence on n , which carries over to the dependence of constants in Theorem 1.3. Let us point out that all the proofs of lower-tail estimates based on the Poincaré inequality and available for the category of all smooth functions, which we have been able to find in the literature, seem to break down in the convex setting (see e.g. the arguments in [4,17,20]).

Appendix A: Facts related to Hamilton–Jacobi equations

We will now present some basic properties of Hamilton–Jacobi equations related to infimum convolution operators with the cost $\theta(x) = \alpha(x)$, where α is given by (4.3), which have been exploited in the proof of Lemma 4.1. We remark that all the facts we will rely on are quite standard, however in the literature they are usually considered under slightly different sets of assumptions, which makes it difficult to find an off the shelf result applicable to our situation. We will briefly indicate how the reasonings from [13], Chapter 3, can be modified to yield the properties we need. Alternatively, as in [18], one could rely on a modification of the results from [16], where the theory of Hamilton–Jacobi equations is extended to the setting of metric spaces.

Proposition A.1. *Let C, L be positive constants and let α be defined by (4.3). Assume that $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is either bounded from below or L -Lipschitz and let $u : (0, \infty) \times \mathbb{R}^n \rightarrow \mathbb{R}$ be given by $u(t, x) = Q_t^\alpha f(x)$, where*

$$Q_t^\alpha f(x) = \inf_{y \in \mathbb{R}^n} \{f(y) + t\alpha((x - y)/t)\}, \quad t > 0.$$

Then the following conditions hold.

- (a) For every $s, t > 0$ and every $x \in \mathbb{R}^n$, $Q_t Q_s f(x) = Q_{t+s} f(x)$.
- (b) The function u is Lipschitz on $(0, \infty) \times \mathbb{R}^n$,
- (c) At every point $(t, x) \in (0, \infty) \times \mathbb{R}^n$ of differentiability of u , one has

$$\frac{d}{dt}u(t, x) + \alpha^*(\nabla_x u(t, x)) = 0,$$

where $\alpha^* : \mathbb{R}^n \rightarrow \mathbb{R}$ is the Legendre transform of α , given explicitly by the formula

$$\alpha^*(s) = \begin{cases} C|s|^2 & \text{for } |s| \leq L, \\ +\infty & \text{for } |s| > L. \end{cases}$$

Sketch of proof. Let us note that if f is bounded from below or L -Lipschitz, then $Q_t f$ is well defined.

Ad (a). To show the semigroup property, one can repeat the argument from the proof of [13], Chapter 3.3.2, Lemma 1, however in our setting one needs to work with infima rather than minima.

Ad (b). For fixed t , u is L -Lipschitz in x , as an infimum of L -Lipschitz functions. Indeed for each y , the function $x \mapsto t\alpha((x-y)/t)$ is L -Lipschitz. As for the Lipschitz property with respect to t , the argument in the proof of [13], Chapter 3.3.2, Lemma 2, shows that if f is L -Lipschitz, then for any x ,

$$|u(t, x) - f(x)| \leq Mt,$$

where $M = \max_{|x| \leq L} \alpha^*(x) = CL^2$. Now the Lipschitz condition with respect to $t > 0$ (for general f , which may not be L -Lipschitz) follows from the semigroup property and the fact that $Q_t f$ is an L -Lipschitz function of x .

Ad (c). Using again the fact that $Q_t f$ is L -Lipschitz, it is enough to consider the case when so is f . One can then repeat the proof of [13], Chapter 3.3.2, Theorem 5, provided that one can prove that the infimum in the definition of $Q_t f$ is in fact achieved. To this end, it is enough to note that whenever $|y-x| > 2CLt$ we have, denoting $z = x + 2CLt(y-x)/|x-y|$,

$$\begin{aligned} & f(y) + t\alpha((x-y)/t) \\ &= f(z) + t\alpha((x-z)/t) + (f(y) - f(z)) + t\alpha((x-y)/t) - t\alpha((x-z)/t) \\ &\geq f(z) + t\alpha((x-z)/t) - L|z-y| + t\alpha((x-y)/t) - t\alpha((x-z)/t) \\ &= f(z) + t\alpha((x-z)/t), \end{aligned}$$

where the inequality holds by the Lipschitz property of f and the last equality follows from the definition of α (and the fact that z lies on the interval with endpoints x and y). Thus $Q_t f(x) = \inf_{|y-x| \leq 2CLt} \{f(y) + t\alpha((y-x)/t)\}$ and the existence of the minimizer follows from compactness and continuity of f and α . \square

Acknowledgements

Research partially supported by the National Science Centre, Poland, grants no. 2015/18/E/ST1/00214 (R. Adamczak) and 2015/19/N/ST1/00891 (M. Strzelecki).

References

- [1] Adamczak, R., Bednorz, W. and Wolff, P. (2017). Moment estimates implied by modified log-Sobolev inequalities. *ESAIM Probab. Stat.* **21** 467–494.
- [2] Adamczak, R. and Strzelecki, M. (2015). Modified log-Sobolev inequalities for convex functions on the real line. Sufficient conditions. *Studia Math.* **230** 59–93. [MR3456588](#)
- [3] Adamczak, R. and Wolff, P. (2015). Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. *Probab. Theory Related Fields* **162** 531–586. [MR3383337](#)

- [4] Aida, S. and Stroock, D. (1994). Moment estimates derived from Poincaré and logarithmic Sobolev inequalities. *Math. Res. Lett.* **1** 75–86.
- [5] Ané, C., Blachère, S., Chafaï, D., Fougères, P., Gentil, I., Malrieu, F., Roberto, C. and Scheffer, G. (2000). *Sur les inégalités de Sobolev logarithmiques. Panoramas et Synthèses [Panoramas and Syntheses]* **10**. Paris: Société Mathématique de France. [MR1845806](#)
- [6] Bobkov, S. and Ledoux, M. (1997). Poincaré’s inequalities and Talagrand’s concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields* **107** 383–400. [MR1440138](#)
- [7] Bobkov, S.G., Gentil, I. and Ledoux, M. (2001). Hypercontractivity of Hamilton–Jacobi equations. *J. Math. Pures Appl.* (9) **80** 669–696.
- [8] Bobkov, S.G. and Götze, F. (1999). Discrete isoperimetric and Poincaré-type inequalities. *Probab. Theory Related Fields* **114** 245–277.
- [9] Bobkov, S.G. and Götze, F. (1999). Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** 1–28. [MR1682772](#)
- [10] Bobkov, S.G., Nayar, P. and Tetali, P. (2017). Concentration properties of restricted measures with applications to non-Lipschitz functions. In *Geometric Aspects of Functional Analysis. Lecture Notes in Math.* **2169** 25–53. Cham: Springer.
- [11] Chafaï, D. and Malrieu, F. (2010). On fine properties of mixtures with respect to concentration of measure and Sobolev type inequalities. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** 72–96.
- [12] de la Peña, V.H., Klass, M.J. and Lai, T.L. (2004). Self-normalized processes: Exponential inequalities, moment bounds and iterated logarithm laws. *Ann. Probab.* **32** 1902–1933.
- [13] Evans, L.C. (2010). *Partial Differential Equations*, 2nd ed. *Graduate Studies in Mathematics* **19**. Providence, RI: Amer. Math. Soc.
- [14] Feldheim, N., Marsiglietti, A., Nayar, P. and Wang, J. (2018). A note on the convex infimum convolution inequality. *Bernoulli* **24** 257–270.
- [15] Gluskin, E.D. and Kwapien, S. (1995). Tail and moment estimates for sums of independent random variables with logarithmically concave tails. *Studia Math.* **114** 303–309.
- [16] Gozlan, N., Roberto, C. and Samson, P.-M. (2014). Hamilton Jacobi equations on metric spaces and transport entropy inequalities. *Rev. Mat. Iberoam.* **30** 133–163. [MR3186934](#)
- [17] Gozlan, N., Roberto, C. and Samson, P.M. (2015). From dimension free concentration to the Poincaré inequality. *Calc. Var. Partial Differential Equations* **52** 899–925.
- [18] Gozlan, N., Roberto, C., Samson, P.M., Shu, Y. and Tetali, P. (2017). Characterization of a class of weak transport-entropy inequalities on the line. *Ann. Inst. Henri Poincaré Probab. Stat.* To appear. Available at [arXiv:1509.04202v2](#).
- [19] Gozlan, N., Roberto, C., Samson, P.-M. and Tetali, P. (2017). Kantorovich duality for general transport costs and applications. *J. Funct. Anal.* **273** 3327–3405. [MR3706606](#)
- [20] Gromov, M. and Milman, V.D. (1983). A topological application of the isoperimetric inequality. *Amer. J. Math.* **105** 843–854.
- [21] Ledoux, M. (1995/97). On Talagrand’s deviation inequalities for product measures. *ESAIM Probab. Stat.* **1** 63–87 (electronic).
- [22] Ledoux, M. (1999). Concentration of measure and logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXIII. Lecture Notes in Math.* **1709** 120–216. Berlin: Springer.
- [23] Ledoux, M. (2001). *The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs* **89**. Providence, RI: Amer. Math. Soc.
- [24] Maurey, B. (1991). Some deviation inequalities. *Geom. Funct. Anal.* **1** 188–197.
- [25] Milman, E. (2009). On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.* **177** 1–43.
- [26] Paouris, G. and Valettas, P. (2017). A small deviation inequality for convex functions. *Ann. Probab.* To appear. Available at [arXiv:1611.01723](#).

- [27] Pisier, G. (1986). Probabilistic methods in the geometry of Banach spaces. In *Probability and Analysis (Varenna, 1985)*. *Lecture Notes in Math.* **1206** 167–241. Berlin: Springer. [MR0864714](#)
- [28] Samson, P.M. (2000). Concentration of measure inequalities for Markov chains and Φ -mixing processes. *Ann. Probab.* **28** 416–461.
- [29] Samson, P.M. (2003). Concentration inequalities for convex functions on product spaces. In *Stochastic Inequalities and Applications*. *Progress in Probability* **56** 33–52. Basel: Birkhäuser.
- [30] Talagrand, M. (1991). A new isoperimetric inequality and the concentration of measure phenomenon. In *Geometric Aspects of Functional Analysis (1989–90)*. *Lecture Notes in Math.* **1469** 94–124. Berlin: Springer.
- [31] Talagrand, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.* **81** 73–205.

Received March 2017 and revised July 2017