University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Michał Strzelecki

Functional and transport inequalities and their applications to concentration of measure

PhD dissertation

Supervisor
dr hab. Radosław Adamczak
Institute of Mathematics
University of Warsaw

October 2018

Author's declaration:
I hereby declare that this dissertation is my own work.

October 31, 2018 ....

Michał Strzelecki

Supervisor's declaration:
The dissertation is ready to be reviewed

October 31, 2018 ....

dr hab. Radosław Adamczak

Abstract

This thesis is devoted to the study of functional and transportation inequalities connected to the concentration of measure phenomenon.

In the first part, we work in the classical setting of smooth functions and are interested in the concentration between the exponential and Gaussian levels. We prove that a probability measure which satisfies a Beckner-type inequality of Latała and Oleszkiewicz, also satisfies a modified log-Sobolev inequality. As a corollary, we obtain improved (dimension-free) two-level concentration for products of such measures.

The second, more extensive, part is concerned with concentration of measure for convex functions. Our main tool, used throughout, is the theory of weak transportation inequalities introduced recently by Gozlan, Roberto, Samson, and Tetali. We start by presenting a characterization of probability measures on the real line which satisfy the convex log-Sobolev inequality. This allows us to give concentration estimates of the lower and upper tails of convex Lipschitz functions (the latter were not known before).

We then prove that a probability measure on $\mathbb{R}^n$ which satisfies the convex Poincaré inequality also satisfies a Bobkov–Ledoux modified log-Sobolev inequality, extending results obtained by other authors for measures on $\mathbb{R}$.

We also present refined concentration of measure inequalities, which are consequences of weak transportation inequalities (or, equivalently, their dual formulations: convex infimum convolution inequalities). This includes applications to concentration for non-Lipschitz convex functions.

Our last result concerns convex infimum convolution inequalities with optimal cost functions for measures with log-concave tails. As a corollary, we obtain comparison of weak and strong moments of random vectors with independent coordinates with log-concave tails.

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

Keywords and phrases. Concentration of measure, convex functions, infimum convolution, log-Sobolev inequality, Poincaré inequality, transportation inequalities.

Streszczenie

Niniejsza rozprawa poświęcona jest nierównościom funkcyjnym i transportowym związanym ze zjawiskiem koncentracji miary.

W pierwszej części zajmujemy się koncentracją dla funkcji gładkich. Dowodzimy, że miara probabilistyczna, która spełnia pochodzącą od Latały i Oleszkiewicza nierówność typu Becknera, spełnia także odpowiednią zmodyfikowaną nierówność logarytmiczną Sobolewa. Jako wniosek dostajemy wzmocnioną dwupoziomową koncentrację dla produktów takich miar.

Druga część jest obszerniejsza i dotyczy koncentracji dla funkcji wypukłych. Naszym głównym narzędziem technicznym jest teoria słabych nierówności transportowych wprowadzonych niedawno przez Gozlana, Roberta, Samsona i Tetaliego. Najpierw przedstawiamy charakteryzację miar probabilistycznych, które spełniają wypukłą nierówność logarytmiczną Sobolewa na prostej. Pozwala nam to wyprowadzić oszacowania koncentracyjne dla górnego i dolnego ogona lipszycowskich funkcji wypukłych (wcześniej znane były jedynie oszacowania dla górnego ogona).

Następnie dowodzimy, że miara probabilistyczna na $\mathbb{R}^n$, spełniająca wypukłą nierówność Poincarégo, spełnia także pochodzącą od Bobkova i Ledoux zmodyfikowaną nierówność logarytmiczną Sobolewa. Wzmacnia to wyniki otrzymane przez innych autorów w przypadku miar na prostej.

Opisujemy także, jakie ogólne nierówności koncentracyjne wynikają ze słabych nierówności transportowych (równoważnie: z ich dualnych sformułowań, czyli wypukłych nierówności splotu infimum). Obejmuje to także wyniki dla nielipszycowskich funkcji wypukłych.

Ostatni wynik dotyczy wypukłych nierówności splotu infimum z optymalnymi funkcjami kosztu dla miar o log-wklęsłych ogonach. Jako wniosek otrzymujemy porównywanie słabych i silnych momentów wektorów losowych o niezależnych współrzędnych z log-wklęsłymi ogonami.

Klasyfikacja tematyczna. 60E15; 26A51, 26B25, 26D10.

Słowa kluczowe. Funkcje wypukłe, koncentracja miary, nierówność logarytmiczna Sobolewa, nierówność Poincarégo, nierówności transportowe, splot infimum.

Contents

Chapter 1

A brief introduction

1.1 The concentration of measure phenomenon on the sphere

The classical isoperimetric problem on the sphere $S^{n-1} \subset \mathbb{R}^n$ (endowed with the uniform, rotation invariant probabilistic measure σ_{n-1}) was solved independently by Lévy [53] and Schmidt [71, 70, 72]. It turns out, that if B is a spherical cap (the geodesic ball) and $A \subset S^{n-1}$ is a measurable set with $\sigma_{n-1}(A) = \sigma_{n-1}(B)$, then

$$\sigma_{n-1}(\{x \in S^{n-1} : d(x, A) < t\}) \geq \sigma_{n-1}(\{x \in S^{n-1} : d(x, B) < t\}),$$

where d is the geodesic metric on S^{n-1}.

In the case when B is a half-sphere it is easy to explicitly compute the measure of the spherical cap $\{x \in S^{n-1} : d(x, B) < t\}$. The above inequality implies that for any set A with measure $1/2$ we have

$$\sigma_{n-1}(\{x \in S^{n-1} : d(x, A) < t\}^c) \leq e^{-(n-2)t^2/2}.$$

Therefore the measure of the set $\{x \in S^{n-1} : d(x, A) < t\}^c$ decays very fast as t grows. Moreover, by taking A to be a half-sphere, we see that—speaking informally—the measure σ_{n-1} is concentrated around the equator (any of the equators) of S^{n-1}.

We can restate this in the following language: if $f \colon S^{n-1} \to \mathbb{R}$ is an L-Lipschitz function, then

$$\sigma_{n-1}(\{x \in S^{n-1} : |f(x) - \mathrm{Med}_{\sigma_{n-1}}(f)| \geq t\}) \leq e^{-(n-2)t^2/(2L^2)},$$

where

$$\mathrm{Med}_{\sigma_{n-1}}(f) := \inf\big\{s \in \mathbb{R} : \ \sigma_{n-1}(\{x \in S^{n-1} : f(x) \leq s\}) \geq 1/2\big\}$$

is the median of f. This means that functions on high-dimensional spheres with small local oscillations are essentially constant—they are concentrated around their median (or mean value).

The importance of this elementary, yet non-trivial, observation was first emphasized by Milman, who used it in his proof of Dvoretzky's theorem [57]. The concentration phenomenon has become one of the main themes of high dimensional probability and geometric analysis, with many applications to, e.g., limit theorems (see Ledoux's and Talagrand's book [52]), non-asymptotic confidence bounds or random constructions of geometric objects with extremal properties (consult the monographs [58] and [10]).

1.2 Gaussian concentration

Another, classical example for the concentration of measure phenomenon is provided by the standard Gaussian measure γ_n on $\mathbb{R}^n$ which has the density $(2\pi)^{-n/2}\exp(-|x|^2/2)$, $x \in \mathbb{R}^n$ (for which the isoperimetric problem is also solved [22, 76]). If $f\colon \mathbb{R}^n \to \mathbb{R}$ is 1-Lipschitz, then

$$\gamma_n\Big(\Big\{x \in \mathbb{R}^n : |f(x) - \int_{\mathbb{R}^n} f d\gamma_n| \geq t\Big\}\Big) \leq 2e^{-t^2/2}, \quad t > 0,$$

or, in probabilistic notation,

$$\mathbb{P}(|f(G) - \mathbb{E}\, f(G)| \geq t) \leq 2e^{-t^2/2}, \quad t > 0,$$

where G is a random vector with law γ_n. Here we restricted our attention to 1-Lipschitz functions, but clearly the general result for L-Lipschitz functions follows just by scaling. Also, we stated the results in terms of concentration around the mean not the median (but it should be intuitively clear, that if a function 'concentrates' around some number, then this number has to be 'close' both to the median and to the mean of this function).

Note that here the concentration is dimension-free, that is, the dimension n does not appear on the right-hand side. The constant $1/2$ in the exponent on the right-hand side is optimal (as standard estimates of the Gaussian tail on the real line show), but—with possibly worse numerical constants—the result can be derived in various ways, some of which we shall discuss below.

1.3 Functional and transport–entropy inequalities

Luckily, concentration inequalities can be obtained without solving the isoperimetric problem for the underlying probability distribution. Let us recall three classical tools which are nicely suited to deriving concentration inequalities: the log-Sobolev inequality, the $\mathbf{T}_2$ inequality, and the Poincaré inequality. The picture sketched below is by no means complete and we refer to, e.g., the monographs [51] and [23] for a thorough introduction to the topic and a vast overview of the field.

Let μ be a Borel probability measure on $\mathbb{R}^n$ and let $|\cdot|$ stand for the standard Euclidean norm on $\mathbb{R}^n$.

We say that μ satisfies the *log-Sobolev inequality* if there exists a constant C_{LS} such that for every smooth function $f\colon \mathbb{R}^n \to \mathbb{R}$,

$$\mathrm{Ent}_\mu(f^2) \le C_{LS} \int_{\mathbb{R}^n} |\nabla f|^2 d\mu. \tag{1.3.1}$$

Here $\mathrm{Ent}_\mu(g)$ denotes the usual entropy of a non-negative function g, i.e.,

$$\mathrm{Ent}_\mu(g) := \int_{\mathbb{R}^n} g \ln(g) d\mu - \int_{\mathbb{R}^n} g d\mu \ln\Big(\int_{\mathbb{R}^n} g d\mu\Big) \tag{1.3.2}$$

if $\int_{\mathbb{R}^n} g \ln(g) d\mu < \infty$ and $\mathrm{Ent}_\mu(g) = \infty$ otherwise. This inequality was first introduced by Gross [41]. By substituting $e^{f/2}$ in place of f it can be rewritten in the form

$$\mathrm{Ent}_\mu(e^f) \le \frac{C_{LS}}{4} \int_{\mathbb{R}^n} |\nabla f|^2 e^f d\mu,$$

which is sometimes more convenient.

We say that μ satisfies *Talagrand's* $\mathbf{T}_2$ *inequality* if there exists a constant C_T such that for every probability measure ν on $\mathbb{R}^n$,

$$\mathcal{T}_2(\mu,\nu) \le C_T H(\nu|\mu), \tag{1.3.3}$$

where

$$\mathcal{T}_2(\mu,\nu) := \inf_\pi \int_{\mathbb{R}^n\times\mathbb{R}^n} |x-y|^2 \pi(dx,dy),$$

is the usual quadratic transport cost considered in the Monge–Kantorovich transport problem (the infimum is taken over all couplings of μ and ν, i.e., over all probability measures π on $(\mathbb{R}^n)^2$ such that $\pi(dx \times \mathbb{R}^n) = \mu(dx)$, $\pi(\mathbb{R}^n \times dy) = \nu(dy)$) and $H(\nu|\mu)$ stands for the relative entropy of ν with respect to μ, given by the formula

$$H(\nu|\mu) = \int_{\mathbb{R}^n} \log\Big(\frac{d\nu}{d\mu}\Big) d\nu \in [0,+\infty] \tag{1.3.4}$$

if ν is absolutely continuous with respect to μ; otherwise one sets $H(\nu|\mu) = +\infty$. This inequality was introduced by Talagrand in [79] and subsequently widely studied (see, e.g., [80] for a complete and detailed presentation). It is also called the transport–entropy inequality, due to the terms appearing in (1.3.3).

We say that μ satisfies the *Poincaré inequality* with constant $C_P > 0$ if for all smooth functions $f\colon \mathbb{R}^n \to \mathbb{R}$ we have

$$\mathrm{Var}_\mu(f) \le C_P \int_{\mathbb{R}^n} |\nabla f|^2 d\mu. \tag{1.3.5}$$

Remark 1.3.1. Suppose that X is a random vector with values in $\mathbb{R}^n$. Slightly abusing the terminology we will say that it satisfies the log-Sobolev, Talagrand's, or the Poincaré inequality whenever its law satisfies the respective inequality. Also, when convenient, we will use probabilistic notation. Thus, e.g., we shall write (1.3.1) and (1.3.5) as

$$\begin{aligned}
\mathrm{Ent}(f(X)) &:= \mathbb{E}\, f(X) \ln f(X) - \mathbb{E}\, f(X) \ln \mathbb{E}\, f(X) \\
&\le C_{LS}\, \mathbb{E}\, |\nabla f(X)|^2, \\
\mathrm{Var}(f(X)) &\le C_P\, \mathbb{E}\, |\nabla f(X)|^2,
\end{aligned}$$

respectively. We shall act similarly with other inequalities.

All three inequalities mentioned above have a dimension-free tensorization property: if a measure satisfies one of the inequalities, then its products also satisfy that inequality and the constant with which the inequality holds does not increase.

The standard Gaussian measure is a flagship example of a measure which satisfies the log-Sobolev and $\mathbf{T}_2$ inequalities. Both of those inequalities imply subgaussian dimension-free concentration: if a measure μ satisfies one of them, then for every positive integer N and every 1-Lipschitz function $f\colon \mathbb{R}^{Nn} \to \mathbb{R}$, one has

$$\mu^{\otimes N}\Big(\big|f - \int_{\mathbb{R}^{Nn}} f d\mu^{\otimes N}\big| \ge t\Big) \le 2\exp(-Kt^2), \tag{1.3.6}$$

where $K = 1/C_{LS}$ in case of the log-Sobolev inequality and $K = 1/C_T$ in case of Talagrand's inequality.

The Poincaré inequality is satisfied, e.g., by the symmetric exponential distribution. If a measure μ satisfies the Poincaré inequality, then for every positive integer N and every 1-Lipschitz function $f\colon \mathbb{R}^{Nn} \to \mathbb{R}$, one has

$$\mu^{\otimes N}\Big(\big|f - \int_{\mathbb{R}^{Nn}} f d\mu^{\otimes N}\big| \ge t\Big) \le 2\exp(-t/\sqrt{C_P}). \tag{1.3.7}$$

There are deep and important connections between the concentration of measure phenomenon and the log-Sobolev inequality, the transport–entropy inequality, and the Poincaré inequality (and infimum convolution inequalities which we shall introduce in Section 1.4 below). Otto and Villani [60] established the implication chain

$$\text{log-Sobolev inequality} \implies \mathbf{T}_2 \text{ inequality} \implies \text{Poincaré inequality.}$$

(A simpler proof, based on the Hamilton-Jacobi partial differential equation, was given by Bobkov, Gentil, and Ledoux in [18], cf. [61].) These implications are strict; the fact that the log-Sobolev inequality can be distinguished from Talagrand's $\mathbf{T}_2$ inequality was proved by Cattiaux and Guillin [24] (see also Section 4.3 of [35] for a nice discussion of other counterexamples in a more general setting).

Moreover, as shown by Gozlan [33], dimension-free subgaussian concentration is in fact equivalent to Talagrand's inequality being satisfied. Another result of Gozlan, Roberto, and Samson [37], asserts that any kind of dimension-free concentration implies dimension-free concentration with—at least—an exponential rate (which was already noticed by Talagrand in [77]) and moreover the Poincaré inequality.

1.4 Infimum convolution inequalities

We shall now recall the definition of another inequality which will play an important role throughout the paper. Let μ be a probability measure on $\mathbb{R}^n$ and let $\varphi : \mathbb{R}^n \to [0, \infty]$ be a measurable function. We say that the pair (μ, φ) satisfies the *infimum convolution inequality* if for every bounded measurable function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\int_{\mathbb{R}^n} e^{f \Box \varphi} d\mu \int_{\mathbb{R}^n} e^{-f} d\mu \leq 1,$$

where $f \Box \varphi$ denotes the infimum convolution of f and φ defined as

$$f \Box \varphi(x) = \inf\{f(y) + \varphi(x - y) : y \in \mathbb{R}^n\} \tag{1.4.1}$$

for $x \in \mathbb{R}^n$. This inequality was introduced by Maurey in [54] and is also known under the name *property* (τ). It behaves nicely under tensorization and it implies the following type of concentration: for every positive integer N, every Borel set $A \subset \mathbb{R}^{nN}$ and any $t > 0$,

$$\mu^{\otimes N}\Big(\big(A + \big\{(x_1, \ldots, x_N) \in (\mathbb{R}^n)^N : \sum_{i=1}^{n} \varphi(x_i) < t\big\}\big)^c\Big) \leq (\mu^{\otimes N}(A))^{-1} e^{-t}$$

(here + stands for the Minkowski addition). Other properties of the infimum convolution inequality will be discussed later on.

We end this introduction be recalling one more, truly seminal result.

1.5 Concentration for convex functions

While initial results on concentration of measure concerned mostly deviation bounds for Lipschitz functions of highly regular random variables, the work by Talagrand [77, 78] has revealed that if one restricts attention to convex Lipschitz functions, dimension-free concentration of measure holds under much weaker conditions. Namely, suppose that $X_1, \ldots, X_n$ are independent random variables, such that $|X_i| \leq 1$ for $i \in \{1, \ldots, n\}$, and $\varphi : \mathbb{R}^n \to \mathbb{R}$ is a convex 1-Lipschitz function. Then, for $t \geq 0$,

$$\mathbb{P}(|\varphi(X_1, \ldots, X_n) - \operatorname{Med} \varphi(X_1, \ldots, X_n)| \geq t) \leq 4 \exp(-t^2/16).$$

We stress that apart from boundedness and independence no further assumptions are placed on the random variables X_i.

Talagrand's approach relied on his celebrated convex distance inequality related to the analysis of isoperimetric problems for product measures. As we shall discuss in detail in Chapter 3 below, 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 convolution inequalities for proving tail inequalities for convex functions. Also, a characterization of convex dimension-free concentration in terms of the convex Poincaré inequality has been established in [37].

This glimpse at the theory of concentration of measure and functional inequalities related to it shall serve us as a basis to introduce other concepts and present our results. In the following chapters we will revisit and explain in more detail some of the results mentioned above.

1.6 Scope of the thesis, overview of the following chapters

Broadly speaking, this thesis is devoted to the study of functional and transportation inequalities connected to the concentration of measure phenomenon. In a higher resolution, the contents of this thesis is the following.

In Chapter 2 we explore connections between some inequalities that give concentration between the exponential and Gaussian levels. This chapter is based on chronologically most recent results obtained while the author was staying in Toulouse, France, and working under the supervision of Franck Barthe. The results of Chapter 2 (of which Franck Barthe should be considered a co-author) are part of a project which at the time of writing is still in progress.

Starting with Chapter 3, we shift our attention to concentration of measure for *convex* functions (and the related weak transport–entropy inequalities). The introductory Chapter 3 contains a historical overview and necessary preliminary results.

In Chapter 4 we investigate (modified) log-Sobolev inequalities for convex functions and provide a characterization on the real line. This chapter is based on joint work with Yan Shu [74].

The next two chapters, are based on joint work with Radosław Adamczak [6]. In Chapter 5 we study the Poincaré inequality for convex functions and extend to $\mathbb{R}^n$ results obtained by other authors for probability measures on the real line. In Chapter 6 we present refined concentration of measure inequalities, which are consequences of weak transportation inequalities and convex infimum convolution inequalities, including applications to concentration for non-Lipschitz convex functions.

In Chapter 7 we discuss the results obtained with Marta Strzelecka and Tomasz Tkocz [75] about the optimal cost function with which, for a given measure on the real line, the convex infimum convolution inequality can hold.

Appendix A gathers some facts connected to Hamilton–Jacobi equations, which are used in the proofs.

1.7 Acknowledgments

I would like to thank my advisor, Radosław Adamczak, for his constant support and guidance, in particular for introducing me to interesting problems and for many hours of conversations on topics related to them.

I also thank Franck Barthe, for his hospitality and kindness during my stay in Toulouse, France, in spring 2018.

I decided against trying to mention here all the other people to whom I am indebted for such important things like great teaching, encouragement, good advice, and interesting discussions, but I would like to end with big thanks to all my co-authors—without them my work would have been far less enjoyable (not speaking about its outcome).

My research was partially supported by National Science Centre, Poland, via the Preludium grant no. 2015/19/N/ST1/00891 and the Etiuda grant no. 2017/24/T/ST1/00323.

Chapter 2

Beckner-type inequalities and modified log-Sobolev inequalities

2.1 Introduction: a tale of two inequalities

Throughout this chapter we always assume that $r \in (1, 2)$ and $q \in (2, \infty)$ satisfy $1/r + 1/q = 1$. By μ_r we denote the probability measure on the real line with density

$$d\mu_r(t) = \frac{e^{-|t|^r} dt}{2\Gamma(1 + 1/r)}, \quad t \in \mathbb{R}.$$

This measure has heavier tails than the standard Gaussian measure and thus cannot satisfy the classical log-Sobolev inequality (1.3.1). In order to prove concentration results for (the products) of the measure μ_r or other measures with similar behavior various variants of the log-Sobolev inequality have been introduced in the literature. To present our motivation and goals we shall recall the definition and some properties of two of those inequalities: Beckner-type inequalities of Latała and Oleszkiewicz from [44] and modified log-Sobolev inequalities introduced by Gentil, Guillin, and Miclo in [31].

We shall say that a probability measure μ on $\mathbb{R}^d$ satisfies the *Latała–Oleszkiewicz inequality* if there exists a constant $C_{LO} < \infty$ such that for every smooth $f : \mathbb{R}^d \to \mathbb{R}$ one has

$$\sup_{p \in (1,2)} \frac{\int_{\mathbb{R}^d} f^2 d\mu - \left(\int_{\mathbb{R}^d} |f|^p d\mu\right)^{2/p}}{(2-p)^{2(1-1/r)}} \leq C_{LO} \int_{\mathbb{R}^d} |\nabla f|^2 d\mu. \tag{2.1.1}$$

(Formally, we should refer to this inequality as, say, the "Latała–Oleszkiewicz inequality *with parameter* r", but for brevity we suppress the dependence on r in the terminology. In other words, the numbers $r \in (1, 2)$ and $q = r/(r-1)$

can be regarded as fixed throughout the chapter. We apply this convention also to other inequalities considered below.) Inequalities of this type, with μ being the standard Gaussian measure and $r = 2$, were first considered by Beckner in [16].

Latała and Oleszkiewicz [44] proved that this inequality has the tensorization property and that whenever μ satisfies the inequality (2.1.1), then for any positive integer n and every 1-Lipschitz function $f\colon \mathbb{R}^{dn} \to \mathbb{R}$, one has

$$\mu^{\otimes n}\Big(\big|f - \int_{\mathbb{R}^{dn}} f d\mu^{\otimes n}\big| \geq t\sqrt{C_{LO}}\Big) \leq 2\exp(-K\min\{t^2, t^r\}) \tag{2.1.2}$$

(their proof yields $K = 1/3$; see also [82] and Section 6 of [12] for an extension to a more general setting). Moreover they showed that the measure μ_r satisfies the inequality (2.1.1) (in this case $d = 1$).

Let us stress here that most of the information is encoded in the speed at which $(2-p)^{2(1-1/r)}$ vanishes as $p \to 2^-$ (by omitting the supremum on the left-hand side of (2.1.1) and only considering a fixed $p \in (1, 2)$ one gets a significantly weaker inequality).

Another approach was suggested by Gentil, Guillin, and Miclo [31]. We say that a probability measure μ on $\mathbb{R}^d$ satisfies the *modified log-Sobolev inequality* if there exists a constant $C_{mLS} < \infty$ such that for every smooth function $f\colon \mathbb{R}^d \to (0,\infty)$ one has

$$\mathrm{Ent}_\mu(f^2) \leq C_{mLS}\int_{\mathbb{R}^d} H_q\Big(\frac{|\nabla f|}{f}\Big) f^2 d\mu, \tag{2.1.3}$$

where $H_q(t) \coloneqq \max\{t^2, |t|^q\}$ for $t \in \mathbb{R}$ (recall that $q = r/(r-1)$). Just for the record, μ_r satisfies (2.1.3).

This inequality also tensorizes: if μ satisfies (2.1.3), then for any positive integer n and every smooth function $f\colon \mathbb{R}^{dn} \to (0,\infty)$ one has

$$\mathrm{Ent}_{\mu^{\otimes n}}(f^2) \leq C_{mLS}\int_{\mathbb{R}^{dn}} \sum_{i=1}^n H_q\Big(\frac{|\nabla_i f|}{f}\Big) f^2 d\mu^{\otimes n},$$

where $\nabla_i f$ denotes the partial gradient with respect to the i-th d-tuple of coordinates of $\mathbb{R}^{dn}$. A modification of Herbst's argument implies that for any $f\colon \mathbb{R}^{dn} \to \mathbb{R}$,

$$\begin{aligned}\mu^{\otimes n}&\Big(\big|f - \int_{\mathbb{R}^{dn}} f d\mu^{\otimes n}\big| \geq t\Big)\\ &\leq 2\exp\Big(-K\min\Big\{\frac{t^2}{C_{mLS}\sup|\nabla f(x)|^2}, \frac{t^r}{C_{mLS}^{r-1}\sup|(|\nabla_i f(x)|)_{i=1}^n|_q^r}\Big\}\Big)\end{aligned} \tag{2.1.4}$$

(one can take $K = 3/8$, see the proof of Corollary 2.2.2 below). Here $|\cdot|_q$ denotes the ℓ^q-norm of a vector in $\mathbb{R}^n$ and the suprema are taken over all $x \in \mathbb{R}^{dn}$.

The inequality (2.1.4) is—up to constants—better than the inequality (2.1.2): t^r is divided by a smaller number, i.e., $\sup_{x\in\mathbb{R}^{dn}} |(|\nabla_i f(x)|)_{i=1}^n|_q^r$ instead of $\sup_{x\in\mathbb{R}^{dn}} |\nabla f(x)|^r$. This difference makes it possible to recover—again up to constants—the tail behavior from the Central Limit Theorem. Namely, if for example $d = 1$ and we consider the function

$$f(x) = f_n(x) = \frac{x_1 + \cdots + x_n}{\sqrt{n}}, \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

then $|\nabla f(x)| = 1$, but $|\nabla f(x)|_q^r = n^{r/2-1}$, and inequality (2.1.4) reads

$$\mu^{\otimes n}\Big(\big|f_n - \int_{\mathbb{R}^{dn}} f_n d\mu^{\otimes n}\big| \geq t\Big) \leq 2\exp\Big(-K \min\Big\{\frac{t^2}{C_{mLS}}, \frac{n^{1-r/2}t^r}{C_{mLS}^{r-1}}\Big\}\Big)$$

(note that $n^{1-r/2} \to \infty$ since $r < 2$).

Recall that Bobkov and Ledoux [17] proved that if a probability measure μ on $\mathbb{R}^d$ satisfies the Poincaré inequality with constant C_P (to which inequality (2.1.1) reduces in the limit case $r = 1$, $q = \infty$), then it satisfies a modified log-Sobolev inequality:

$$\mathrm{Ent}_\mu(e^g) \leq C_{BL} \int_{\mathbb{R}^d} |\nabla g|^2 e^g d\mu$$

for all smooth $g\colon \mathbb{R}^d \to \mathbb{R}$ with $|\nabla g(x)| \leq c < 2/\sqrt{C_P}$ for all $x \in \mathbb{R}^d$ (with C_{BL} depending on C_P and c only). This inequality can be regarded as the limit version of inequality (2.1.3) for $r = 1$, $q = \infty$: formally we replaced H_q by

$$H_{\infty,c}(t) := \lim_{q\to\infty} H_q(t/c) = \begin{cases} t^2/c^2 & \text{if } |t| \leq c, \\ \infty & \text{if } |t| > c \end{cases}$$

and substituted $f^2 = e^g$.

It is natural to conjecture, that similarly the Latała–Oleszkiewicz inequality implies the modified log-Sobolev inequality (and therefore improved two-level concentration), cf. Remark 21 in [15]. Before presenting our main result, which essentially states that this is indeed the case, let us recall two other results.

In the case when μ is a probability measure on the real line one can prove criteria for the Latała–Oleszkiewicz inequality (2.1.1) and the modified log-Sobolev inequality (2.1.3). Denote by m the median of μ and by n the density of its absolutely continuous part. Barthe and Roberto [14] proved that μ satisfies the Latała–Oleszkiewicz inequality (2.1.1) if and only if $\max\{B^+_{LO}, B^-_{LO}\} < \infty$, where

$$B^+_{LO} \coloneqq \sup_{x>m} \mu([x,\infty)) \log^{2/q}\Big(1 + \frac{1}{2\mu([x,\infty))}\Big) \int_m^x \frac{1}{n(t)} dt \tag{2.1.5}$$

and B^-_{LO} is defined similarly but with $x < m$. Moreover the best possible constant C_{LO} in (2.1.1) is up to numerical constants (not depending on r) comparable to $\max\{B^+_{LO}, B^-_{LO}\}$.

A similar characterization, also due to Barthe and Roberto [15], is likewise available in the case of the modified log-Sobolev inequality (2.1.3) but the picture is more involved. If μ satisfies the Poincaré inequality with constant C_P and $\max\{B^+_{mLS}, B^-_{mLS}\} < \infty$, where

$$B^+_{mLS} \coloneqq \sup_{x>m} \mu([x,\infty)) \log\Big(\frac{1}{\mu([x,\infty))}\Big) \Big(\int_m^x \frac{1}{n(t)^{1/(q-1)}} dt\Big)^{q-1}$$

and B^-_{mLS} is defined similarly but with $x < m$, then μ satisfies the modified log-Sobolev inequality (2.1.3) with constant

$$C_{mLS} \le 235 C_P + 2^{q+1} \max\{B^+_{mLS}, B^-_{mLS}\}.$$

Under some mild technical assumptions on the density n this implication can be reversed: if μ satisfies the modified log-Sobolev inequality (2.1.3), then

$$\max\{B^+_{mLS}, B^-_{mLS}\} < \infty$$

(and this quantity can be estimated in terms of the constant C_{mLS} up to constants depending on q and a number ε which appears in the additional hypotheses about n).

To the best of our knowledge, it is unknown whether in the above criterion for the modified log-Sobolev inequality one can remove the technical assumption about n. It also does not seem to be possible to find a simple argument which would allow us to deduce the modified log-Sobolev inequality from the Latała–Oleszkiewicz inequality (for measures on the real line) at the level of those characterizations. On the other hand, if one assumes for example that $d\mu(x) = \exp(-V(x))dx$, $x \in \mathbb{R}$, where V is symmetric, of class C^2, $\liminf_{x\to\infty} V'(x) > 0$, and

$$\lim_{x\to\infty} \frac{V''(x)}{V'(x)^2} = 0,$$

then the Latała–Oleszkiewicz inequality (2.1.1) is equivalent to the modified log-Sobolev inequality (2.1.3) and furthermore to the condition

$$\limsup_{x\to\infty} \frac{V(x)}{V'(x)^q} < \infty \tag{2.1.6}$$

(by Theorem 12 in [15] and Proposition 15 in [14]).

2.2 Results and organization of the chapter

We are ready to state our main result. We would like to stress that it is not restricted to measures on the real line and that the dimension d does not enter into the dependence of constants.

Theorem 2.2.1. *Let μ be a probability measure on $\mathbb{R}^d$ which satisfies the Latała–Oleszkiewicz inequality* (2.1.1) *with constant C_{LO}. Then μ satisfies the modified log-Sobolev inequality* (2.1.3) *with a constant C_{mLS} depending only on C_{LO} and r.*

For $x = (x_1, \ldots, x_n) \in (\mathbb{R}^d)^n$ and $p \in (1, \infty)$ denote

$$\|x\|_{p,2} := \Big(\sum_{i=1}^{n} |x_i|^p\Big)^{1/p}$$

(here $|\cdot|$ stands for the ℓ_2 norm on $\mathbb{R}^d$; in the notation we suppress the roles of d and n, but they will always be clear from the context).[1]

We immediately obtain the following corollary, which improves upon inequality (2.1.2) in the manner explained above, during the discussion of consequences of inequality (2.1.4).

Corollary 2.2.2. *Let μ be a probability measure on $\mathbb{R}^d$ which satisfies the Latała–Oleszkiewicz inequality* (2.1.1) *with constant C_{LO}. Then there exists a constant $K > 0$, depending only on C_{LO} and r, such that for any positive integer n and any smooth $f \colon \mathbb{R}^{dn} \to \mathbb{R}$,*

$$\mu^{\otimes n}\Big(\Big|f - \int_{\mathbb{R}^{dn}} f d\mu^{\otimes n}\Big| \geq t\Big) \\ \leq 2\exp\Big(-K \min\Big\{\frac{t^2}{\sup_{x\in\mathbb{R}^{dn}} |\nabla f(x)|^2}, \frac{t^r}{\sup_{x\in(\mathbb{R}^d)^n} \|\nabla f(x)\|_{q,2}^r}\Big\}\Big)$$

[1] While $\|\cdot\|_{\ell^p(\ell^2)}$ would be perhaps more self-explanatory (and would not have misleading associations with Lorentz spaces), we favor this more compact notation.

One can take $K = \frac{3}{8}\min\{1/C_{mLS}, 1/C_{mLS}^{r-1}\}$, *where* $C_{mLS} = C_{mLS}(C_{LO}, r)$ *is the constant with which, by Theorem 2.2.1, the modified log-Sobolev inequality holds for* μ.

Using standard smoothing arguments one can also obtain a result for not necessarily smooth functions, expressed in terms of their Lipschitz constants.

Corollary 2.2.3. *Let* μ *be a probability measure on* $\mathbb{R}^d$ *which satisfies the Latała–Oleszkiewicz inequality* (2.1.1) *with constant* C_{LO}. *Then there exists a constant* $K > 0$, *depending only on* C_{LO} *and* r, *such that for any positive integer* n *the following holds: if* $f\colon \mathbb{R}^{dn} \to \mathbb{R}$ *satisfies*

$$\begin{aligned} |f(x) - f(y)| &\leq L_2|x-y|, \\ |f(x) - f(y)| &\leq L_{r,2}\|x-y\|_{r,2}, \end{aligned}$$

for all $x, y \in (\mathbb{R}^d)^n$, *then*

$$\mu^{\otimes n}\Big(\big|f - \int_{\mathbb{R}^{dn}} f d\mu^{\otimes n}\big| \geq t\Big) \leq 2\exp\Big(-K\min\Big\{\frac{t^2}{L_2^2}, \frac{t^r}{L_{r,2}^r}\Big\}\Big).$$

One can take $K = \frac{3}{8}\min\{1/C_{mLS}, 1/C_{mLS}^{r-1}\}$, *where* $C_{mLS} = C_{mLS}(C_{LO}, r)$ *is the constant with which, by Theorem 2.2.1, the modified log-Sobolev inequality holds for* μ.

One can also express concentration in terms of enlargements of sets. Below B_2^{dn} and B_r^{dn} stand for the unit balls in the ℓ_2 and ℓ_r-norms on $\mathbb{R}^{dn}$, respectively. Also, let

$$B_{r,2}^{n,d} := \Big\{(x_1, \ldots, x_n) \in (\mathbb{R}^d)^n : \big(\sum_{i=1}^n |x_i|^r\big)^{1/r} \leq 1\Big\}$$

be the unit ball in the norm $\|\cdot\|_{r,2}$.

Remark 2.2.4. For $r \in (1,2)$,

$$d^{1/2-1/r}B_{r,2}^{n,d} \subset B_r^{dn} \subset B_{r,2}^{n,d} \subset B_2^{dn} \subset n^{1/r-1/2}B_{r,2}^{n,d}.$$

Corollary 2.2.5. *Let* μ *be a probability measure on* $\mathbb{R}^d$ *which satisfies the Latała–Oleszkiewicz inequality* (2.1.1) *with constant* C_{LO}. *Then there exists a constant* $K > 0$, *depending only on* C_{LO} *and* r, *such that for any positive integer* n *and any set* $A \subset \mathbb{R}^{dn}$ *with* $\mu^{\otimes n}(A) \geq 1/2$,

$$\mu^{\otimes n}\Big(A + \Big\{(x_1, \ldots, x_n) \in (\mathbb{R}^d)^n : \sum_{i=1}^n \min\{|x_i|^2, |x_i|^r\} \leq t\Big\}\Big) \geq 1 - e^{-Kt}.$$

In particular,

$$\mu^{\otimes n}\big(A + \sqrt{t}B_2^{dn} + t^{1/r}B_{r,2}^{n,d}\big) \geq 1 - e^{-Kt}. \tag{2.2.1}$$

One can take $K = \frac{3}{128}\min\{1/C_{mLS}, 1/C_{mLS}^{r-1}\}$*, where* $C_{mLS} = C_{mLS}(C_{LO}, r)$ *is the constant with which, by Theorem 2.2.1, the modified log-Sobolev inequality holds for* μ*.*

This corollary should be compared with the results obtained by Gozlan [34], who considered a family of spectral gap inequalities where the length of the gradient is defined with respect to different metrics than the standard Euclidean distance. He proved that if a probability measure μ on $\mathbb{R}^d$ satisfies the Latała–Oleszkiewicz inequality, then it satisfies an appropriate spectral gap inequality with a non-standard length of the gradient (see Corollary 5.17 in [34],), which in turn implies a slightly different type of two-level concentration (see Proposition 2.4 and Proposition 1.2 in [34]). However, unlike in the above two corollaries, the constants which appear in his formulations do depend on the dimension d of the underlying space (even though they do not depend on n). Namely, if we denote $x_i = (x_i^1, \ldots, x_i^d) \in \mathbb{R}^d$ for $i = 1, \ldots, n$, then there exists a constant $K > 0$ (depending only on C_{LO} and r) such that for any positive integer n and any set $A \subset \mathbb{R}^{dn}$ with $\mu^{\otimes n}(A) \geq 1/2$,

$$\mu^{\otimes n}\Big(A + \Big\{(x_1, \ldots, x_n) \in (\mathbb{R}^d)^n : \sum_{i=1}^{n}\sum_{j=1}^{d} \min\Big\{\Big|\frac{x_i^j}{d}\Big|^2, \Big|\frac{x_i^j}{d}\Big|^r\Big\} \leq t\Big\}\Big) \geq 1 - e^{-Kt/d}$$

(the d in the denominator on the left-hand side comes from Corollary 5.17 of [34] and the d on the right-hand side—from Proposition 2.4 therein). In particular, we get

$$\mu^{\otimes n}\big(A + d^{3/2}\sqrt{t}B_2^{dn} + d^{1+1/r}t^{1/r}B_r^{dn}\big) \geq 1 - e^{-Kt}.$$

In terms of the dependence on d this is weaker than (2.2.1), since

$$B_{r,2}^{n,d} \subset d^{1/r-1/2}B_r^{dn} \subset d^{1+1/r}B_r^{dn}$$

(the inclusions are strict for $d \geq 2$).

The organization of the rest of the chapter is the following. In Section 2.3 we introduce some further preliminary results, which we then use in Section 2.4 to prove a proposition essentially equivalent to our main theorem. In Section 2.5 we prove the main result and Corollaries 2.2.2 and 2.2.5. Finally, in Section 2.6 we present some additional remarks on weighted log-Sobolev inequalities connected to (2.1.3).

2.3 Preliminaries: a few more inequalities

We start with the following observation.

Lemma 2.3.1. *Suppose that a probability measure* μ *on* $\mathbb{R}^d$ *satisfies the Latała–Oleszkiewicz inequality* (2.1.1) *with constant* C_{LO}*. Then it satisfies the Poincaré inequality* (1.3.5) *with constant* $C_P = C_{LO}$*.*

Proof. By taking $p \to 1^+$ in (2.1.1) we see that (1.3.5) holds for all positive smooth functions (with constant C_{LO}). Since the variance is translation invariant, we conclude that (1.3.5) holds for all smooth functions bounded from below. The general case follows by approximation. □

Remark 2.3.2. Alternatively, one can deduce the Poincaré inequality from the fact that inequality (2.1.1) implies dimension-free concentration and the results of [37].

Let us now recall another, equivalent form of the Latała–Oleszkiewicz inequality. For $q > 2$ denote

$$F_q(t) = \log^{2/q}(1+t) - \log^{2/q}(2), \quad t \geq 0.$$

We say that a probability measure μ on $\mathbb{R}^d$ satisfies an F_q*-Sobolev inequality* if there exists C such that for every smooth $g \colon \mathbb{R}^d \to \mathbb{R}$,

$$\int_{\mathbb{R}^d} g^2 F_q\Big(\frac{g^2}{\int_{\mathbb{R}^d} g^2 d\mu}\Big) d\mu \leq C \int_{\mathbb{R}^d} |\nabla g|^2 d\mu. \tag{2.3.1}$$

This inequality is *tight*, i.e., we have equality for constant functions (if f is constant and equal to zero on its support, then the expression $0/0$ should be interpreted as 0 here and in inequality (2.3.5) below). We say that μ on $\mathbb{R}^d$ satisfies a *defective* F_q*-Sobolev inequality* if there exists B and C such that for every smooth $g \colon \mathbb{R}^d \to \mathbb{R}$,

$$\int_{\mathbb{R}^d} g^2 F_q\Big(\frac{g^2}{\int_{\mathbb{R}^d} g^2 d\mu}\Big) d\mu \leq B \int_{\mathbb{R}^d} g^2 d\mu + C \int_{\mathbb{R}^d} |\nabla g|^2 d\mu. \tag{2.3.2}$$

In [12] Barthe, Cattiaux, and Roberto provided capacity criteria for, among others, the Latała–Oleszkiewicz and F_q-Sobolev inequalities. We refer to Section 5 of [12] for a thorough overview of the topic. The following theorem is a direct corollary of the results contained therein (and also in Wang's independent paper [82]).

Theorem 2.3.3. *For an absolutely continuous probability measure* μ *on* $\mathbb{R}^d$ *the following conditions are equivalent.*

1. *There exists a constants C_{LO} such that the measure μ satisfies the Latała–Oleszkiewicz inequality* (2.1.1) *(with parameter $r \in (1,2)$).*
2. *There exists a constants C such that the measure μ satisfies the (tight) F_q-Sobolev inequality* (2.3.1) *(with $q = r/(r-1) \in (2,\infty)$).*

In the implications, the constant in the conclusion depends only on the constant in the premise and on the parameter r.

This formulation is sufficient for our purposes, but the above list of equivalent conditions can be extended to include, e.g., the aforementioned capacity criterion or a super Poincaré inequality with an appropriately chosen rate function (see Corollary 1.2 in [82]). Let us only comment that super Poincaré inequalities were initially introduced by Wang in [81].

For the convenience of the reader we provide an outline of the proof of Theorem 2.3.3. We shall use the nice diagram on page 1041 of [12] as our road map.

Sketch of proof of Theorem 2.3.3. We only prove the implication 1. $\implies$ 2., as this is the one we will actually need. Fix $r \in (1,2)$, $q = r/(r-1) \in (2,\infty)$, and denote $T(s) := s^{2(1-1/r)}$. Recall the following definition of capacity: for Borel sets $A \subset \Omega \subset \mathbb{R}^d$, we define

$$\begin{aligned}\mathrm{Cap}_\mu(A,\Omega) &:= \inf\Big\{ \int_{\mathbb{R}^d} |\nabla f|^2 d\mu : f_{|A} \geq 1 \text{ and } f_{|\Omega^c} = 0\Big\}\\ &= \inf\Big\{ \int_{\mathbb{R}^d} |\nabla f|^2 d\mu : \mathbf{1}_A \leq f \leq \mathbf{1}_\Omega\Big\}\end{aligned}$$

(the infimum is taken over all locally Lipschitz functions; the equality follows from an easy truncation). We furthermore set

$$\begin{aligned}\mathrm{Cap}_\mu(A) &:= \inf\Big\{ \mathrm{Cap}_\mu(A,\Omega) : A \subset \Omega \text{ and } \mu(\Omega) \leq 1/2\Big\}\\ &= \inf\Big\{ \mathrm{Cap}_\mu(A,\Omega) : A \subset \Omega \text{ and } \mu(\Omega) = 1/2\Big\}\end{aligned}$$

(the equality follows from absolute continuity of μ and monotonicity of $\mathrm{Cap}_\mu(A,\Omega)$ in Ω).

Theorem 18 and Lemma 19 from [12] imply that if μ satisfies the Latała–Oleszkiewicz inequality (2.1.1) with some constant C_{LO}, then

$$\mu(A)\frac{1}{T\big(\frac{1}{\log(1+1/\mu(A))}\big)} = \mu(A)\log^{2(1-1/r)}\big(1+1/\mu(A)\big) \leq D_1 \,\mathrm{Cap}_\mu(A) \quad (2.3.3)$$

for every $A \subset \mathbb{R}^d$ with $\mu(A) < 1/2$. Here one can take $D_1 = 6C_{LO}$.

Denote $\varphi(x) = \log^{2(1-1/r)}(1+x/2)$ for $x > 0$. This is a concave, non-decreasing function on $(0,\infty)$ with $\varphi(8) = \log^{2(1-1/r)}(5) \geq 1 > 0$. Moreover,

$x\varphi'(x) \le \gamma$ for some $\gamma = \gamma(r)$ as can be checked by a direct calculation. Also, $\varphi(xy) \le M + \varphi(x) + \varphi(y)$ for some M (one can take, e.g., $M = \log^{2(1-1/r)}(4)$, since $\varphi(xy) \le \log^{2(1-1/r)}(4(1+\frac{x}{2})(1+\frac{y}{2})) \le \log^{2(1-1/r)}(4) + \varphi(x) + \varphi(y)$). Since inequality (2.3.3) implies that

$$\mu(A)\varphi(2/\mu(A)) \le D_1 \operatorname{Cap}_\mu(A)$$

for every $A \subset \mathbb{R}^d$ with $\mu(A) < 1/4$, it follows from Theorem 26 of [12] that for every smooth function $f\colon \mathbb{R}^d \to \mathbb{R}$,

$$\int_{\mathbb{R}^d} f^2\varphi(f^2)d\mu - \int_{\mathbb{R}^d} f^2 d\mu\varphi\Big(\int_{\mathbb{R}^d} f^2 d\mu\Big) \le D_2 \int_{\mathbb{R}^d} |\nabla f|^2 d\mu, \qquad (2.3.4)$$

where $D_2 = D_2(C_{LO}, r) = 18\gamma C_P + 24(1 + M/\varphi(8))D_1$.[2]

It remains to substitute $f^2 = 2g^2/\int_{\mathbb{R}^d} g^2 d\mu$ and recall the definition of F_q, to arrive at (2.3.1) (with $C = D_2$). □

Remark 2.3.4. If μ satisfies the Latała–Oleszkiewicz inequality, but is not necessarily absolutely continuous, then it also satisfies the F_q-Sobolev inequality. To see this we shall use an approximation argument. Let γ_ε be a Gaussian measure on $\mathbb{R}^d$ with covariance matrix $\varepsilon\,\mathrm{Id}$. For small enough $\varepsilon > 0$, γ_ε satisfies the Latała–Oleszkiewicz inequality with the same constant as μ and hence, by tensorization, so does $\mu \otimes \gamma_\varepsilon$. Testing the inequality with the function $(x, y) \mapsto f(x - y)$, we conclude that $\mu * \gamma_\varepsilon$ also satisfies the Latała–Oleszkiewicz inequality (with a constant which tends to C_{LO} as $\varepsilon \to 0$). Thus, by Theorem 2.3.3, $\mu * \gamma_\varepsilon$ satisfies the F_q-Sobolev inequality. We fix a bounded smooth Lipschitz function, take $\varepsilon \to 0$, and arrive at the conclusion that μ satisfies the F_q-Sobolev inequality for all bounded smooth Lipschitz functions (we have pointwise convergence and since the function is Lipschitz and bounded we can use the dominated convergence theorem). Now if f is an arbitrary smooth function such that $\int_{\mathbb{R}^d} |\nabla f|^2 d\mu < \infty$, then it suffices to consider functions $f_n = \Psi_n(f)$, where $\Psi_n\colon \mathbb{R} \to \mathbb{R}$ is, say, an odd and non-decreasing function defined by

$$\Psi_n(t) = \begin{cases} \Psi_n(-t) & \text{for } t < 0, \\ t & \text{for } t \in [0, n), \\ \Psi_n(t) = n + \psi(t) & \text{for } t \in [n, n+2], \\ \Psi_n(t) = n + 1 & \text{for } t > n + 2, \end{cases}$$

[2] Let us stress that this is the point where the assumption about absolute continuity of μ comes into play. This is related to the fact that the proof of Theorem 26 in [12] relies on a decomposition of $\mathbb{R}^d$ into level sets $\{f^2 > \rho_k\}$, for some appropriately chosen ρ_k (cf. proof of Theorem 20 in [12]), and one needs to know that the sets $\{f^2 = \rho_k\} \cap \{|\nabla f| \neq 0\}$ are negligible.

and $\psi : [0,2] \to [0,1]$ is smooth and increasing on $(0,2)$, such that $\psi(0) = 0$, $\psi(2) = 1$, $\psi'(0+) = 1$, $\psi'(2-) = 0$, $\psi(t) \le t$ for $t \in [0,2]$. We then use dominated convergence on the right-hand side and monotone convergence on the left-hand side (note that by the Poincaré inequality f is square-integrable).

Finally, we need another inequality introduced by Barthe and Kolesnikov in [13]. We say that a probability measure μ on $\mathbb{R}^d$ satisfies the inequality $I(2/q)$ if there exists B_1 and C_1 such that for every smooth $f\colon \mathbb{R}^d \to \mathbb{R}$,

$$\mathrm{Ent}_\mu(f^2) \le B_1 \int_{\mathbb{R}^d} f^2 d\mu + C_1 \int_{\mathbb{R}^d} |\nabla f|^2 \log^{1-2/q}\Big(e + \frac{f^2}{\int_{\mathbb{R}^d} f^2 d\mu}\Big) d\mu. \quad (2.3.5)$$

Barthe and Kolesnikov proved that if μ satisfies the inequality $I(2/q)$, then it satisfies a defective F_q-Sobolev inequality and a defective modified log-Sobolev inequality. Moreover they developed techniques—extensions of the classical Rothaus lemma [66]—which allow to transfer such defective inequalities into tight ones by means of a Poincaré inequality.

2.4 From F_q-Sobolev to $I(\tau)$ and modified log-Sobolev

The results of this section are our main contribution. We stress that here the assumption of absolute continuity of μ is not needed. Recall that $q \in (2, \infty)$.

Proposition 2.4.1. *Assume that a probability measure μ on $\mathbb{R}^d$ satisfies the defective F_q-Sobolev inequality* (2.3.2) *with constants B and C. Then μ satisfies the $I(2/q)$ inequality* (2.3.5) *with some constants B_1 and C_1 which depend only on B, C, and q.*

Using Theorem 4.1 from [13] we immediately obtain the following corollary.

Corollary 2.4.2. *Assume that a probability measure μ on $\mathbb{R}^d$ satisfies the defective F_q-Sobolev inequality* (2.3.2) *with constants B and C. Then μ satisfies the following defective modified log-Sobolev inequality: there exists B_2, C_2 (depending only on B, C, and q) such that for every smooth $f\colon \mathbb{R}^d \to (0,\infty)$,*

$$\mathrm{Ent}_\mu(f^2) \le B_2 \int_{\mathbb{R}^d} f^2 d\mu + C_2 \int_{\mathbb{R}^d} H_q\Big(\frac{|\nabla f|}{f}\Big) f^2 d\mu,$$

where $H_q(x) = \max\{x^2, |x|^q\}$.

Moreover, if μ satisfies the Poincaré inequality, then one can take $B_2 = 0$ (C_2 will depend on the Poincaré constant).

Proof of Proposition 2.4.1. We will reverse the reasoning from the proof of Theorem 4.1 in [13]. Fix a smooth function f such that the right-hand side of (2.3.5) is finite. We may and do assume that $\int_{\mathbb{R}^d} f^2|\ln(f^2)|d\mu < \infty$.[3] Consider the function

$$\Phi(x) = x^2 \log^{1-2/q}(e+x^2), \quad x \in \mathbb{R},$$

(which is convex since the function $t \mapsto t\log^{1-2/q}(e+t)$ is convex and increasing for $t > 0$). Denote by L the Luxemburg norm of f related to Φ:

$$L = \inf\{\lambda > 0 : \int_{\mathbb{R}^d} \Phi(f/\lambda)d\mu \leq 1\}.$$

Note that $L < \infty$, $\int_{\mathbb{R}^d} \Phi(f/L)d\mu = 1$ (by the definition of L), and $L^2 \geq \int_{\mathbb{R}^d} f^2 d\mu$ (since $\Phi(x) \geq x^2$).

Set $g := \sqrt{\Phi(f/L)}$. We have $\int_{\mathbb{R}^d} g^2 d\mu = 1$ and (2.3.2) reads

$$\int_{\mathbb{R}^d} g^2\big(\log^{2/q}(1+g^2) - \log^{2/q}(2)\big)d\mu \leq B\int_{\mathbb{R}^d} g^2 d\mu + C\int_{\mathbb{R}^d} |\nabla g|^2 d\mu. \quad (2.4.1)$$

Let us first express the right-hand side of this inequality in terms of f. For $x \in \mathbb{R}$ denote $\varphi(x) := x\log^{1/2-1/q}(e+x^2)$. Then

$$\begin{aligned} 0 \leq \varphi'(x) &= \log^{1/2-1/q}(e+x^2) + (1/2-1/q)\frac{2x^2}{e+x^2}\log^{-1/2-1/q}(e+x^2) \\ &\leq 2\log^{1/2-1/q}(e+x^2) \end{aligned}$$

and thus

$$\begin{aligned} |\nabla g|^2 = \frac{|\nabla f|^2}{L^2}\big(\varphi'(f/L)\big)^2 &\leq 4\frac{|\nabla f|^2}{L^2}\log^{1-2/q}(e+f^2/L^2) \\ &\leq 4\frac{|\nabla f|^2}{L^2}\log^{1-2/q}\Big(e+\frac{f^2}{\int_{\mathbb{R}^d} f^2 d\mu}\Big). \end{aligned}$$

[3] Indeed, like above let us define odd and non-decreasing functions $\Psi_n\colon \mathbb{R} \to \mathbb{R}$ by putting $\Psi_n(t) = t$ for $t \in [0,n)$, $\Psi_n(t) = n+1$ for $t > n+2$; for $t \in [n, n+2]$ let us take $\Psi_n(t) = n + \psi(t)$, where $\psi : [0,2] \to [0,1]$ is smooth and increasing on $(0,2)$, and satisfies $\psi(0) = 0$, $\psi(2) = 1$, $\psi'(0+) = 1$, $\psi'(2-) = 0$, $\psi(t) \leq t$. Then the functions $f_n = \Psi_n(f)$ are smooth, bounded (and hence $\int_{\mathbb{R}^d} f_n^2|\ln(f_n^2)|d\mu < \infty$) and converge to f pointwise. After proving that (2.3.5) holds for f_n, we obtain the assertion for f by taking $n \to \infty$ and using monotone convergence on the left-hand side and the Lebesgue dominated convergence theorem on the right-hand side (note that we know that f and f_n are square-integrable, $|\nabla f_n|$ is up to a constant smaller than $|\nabla f|$, $f_n = f$ if $|f| \in [0,n]$, $|f_n| \leq |f|$ if $|f| \in [n, n+2]$, and if $|f| > n+2$, then $\nabla f_n = 0$).

Hence

$$B\int_{\mathbb{R}^d} g^2 d\mu + C\int_{\mathbb{R}^d} |\nabla g|^2 d\mu \le B + 4C\int_{\mathbb{R}^d} \frac{|\nabla f|^2}{L^2}\log^{1-2/q}\Big(e + \frac{f^2}{\int_{\mathbb{R}^d} f^2 d\mu}\Big)d\mu. \tag{2.4.2}$$

As for the left-hand side of (2.4.1), it is easy to see that there exists $\kappa_1 = \kappa_1(q) > 0$ such that, for $y > 0$,

$$y\log^{1-2/q}(e+y)\Big(\log^{2/q}(1 + y\log^{1-2/q}(e+y)) - \log^{2/q}(2)\Big) \ge y\log(y) - \kappa_1.$$

Applying this inequality with $y = f^2/L^2$, we arrive at

$$\int_{\mathbb{R}^d} g^2\big(\log^{2/q}(1+g^2) - \log^{2/q}(2)\big)d\mu \ge \int_{\mathbb{R}^d} \frac{f^2}{L^2}\log(f^2/L^2)d\mu - \kappa_1.$$

Together with (2.4.2) this yields

$$\int_{\mathbb{R}^d} f^2\log(f^2/L^2)d\mu \le (B+\kappa_1)L^2 + 4C\int_{\mathbb{R}^d} f^2\log^{1-2/q}\Big(e + \frac{f^2}{\int_{\mathbb{R}^d} f^2 d\mu}\Big)d\mu.$$

It remains to replace the expression on the left-hand side by $\mathrm{Ent}_\mu(f^2)$ and estimate L^2.

Since

$$\begin{aligned}\mathrm{Ent}_\mu(f^2) &= \inf_{t>0}\int_{\mathbb{R}^d}\Big(f^2\log(f^2/t) - f^2 + t\Big)d\mu \\ &\le \int_{\mathbb{R}^d}\Big(f^2\log(f^2/L^2) - f^2 + L^2\Big)d\mu,\end{aligned}$$

we conclude that

$$\mathrm{Ent}_\mu(f^2) \le (B+\kappa_1+1)L^2 + 4C\int_{\mathbb{R}^d} f^2\log^{1-2/q}\Big(e + \frac{f^2}{\int_{\mathbb{R}^d} f^2 d\mu}\Big)d\mu. \tag{2.4.3}$$

Finally, it is easy to see that for every $\varepsilon > 0$ there exist $\kappa_2 = \kappa_2(\varepsilon, q)$ such that, for $y > 0$,

$$y\log^{1-2/q}(e+y) \le \varepsilon y\log(y) + \kappa_2$$

Using first the definition of L and the fact that $L^2 \ge \int_{\mathbb{R}^d} f^2 d\mu$, and then the above bound (with $y = f^2/\int_{\mathbb{R}^d} f^2 d\mu$) we can thus estimate

$$\begin{aligned}L^2 = \int_{\mathbb{R}^d} f^2\log^{1-2/q}(e + f^2/L^2)d\mu &\le \int_{\mathbb{R}^d} f^2\log^{1-2/q}\Big(e + \frac{f^2}{\int_{\mathbb{R}^d} f^2 d\mu}\Big)d\mu \\ &\le \varepsilon\,\mathrm{Ent}_\mu(f^2) + \kappa_2\int_{\mathbb{R}^d} f^2 d\mu.\end{aligned}$$

Combining this bound (for ε small enough; recall that by our assumption $\mathrm{Ent}_\mu(f^2) < \infty$) and (2.4.3) yields the assertion. □

2.5 Proof of the main result and its corollaries

Proof of Theorem 2.2.1. Our main results follows immediately by combining Theorem 2.3.3 (and Remark 2.3.4 if μ is not absolutely continuous) and Corollary 2.4.2. $\square$

Remark 2.5.1. Let us comment here that for $d = 1$ it is known that the inequalities (2.1.1) and (2.1.3) hold if and only if they hold with the integration with respect to μ on the right-hand side replaced by integration with respect to μ_{ac}, the absolutely continuous part of μ (cf. [20], Appendix of [55], Appendix of [35]).

For the proofs of the corollaries we need one more technical lemma. We denote by $H_q^*(t) \coloneqq \sup_{s\in\mathbb{R}}\{st - H_q(s)\}$, $t \in \mathbb{R}$, the Legendre transform of H_q (we refer to the book [65] for more information on this topic).

Lemma 2.5.2. *The function H_q^* is given by the formula*

$$H_q^*(t) = \begin{cases} t^2/4 & \text{if } 0 \le |t| \le 2, \\ |t| - 1 & \text{if } 2 \le |t| \le q, \\ \frac{1}{r-1}(\frac{1}{q}|t|)^r & \text{if } |t| \ge q. \end{cases}$$

Moreover, $H_q^*(t) \ge \frac{3}{16}\min\{t^2, |t|^r\}$.

Sketch of the proof. The first part is a straightforward calculation. To prove the second part, first notice that for every $r \in (1, 2)$ there exists exactly one strictly positive number t_0 such that $t_0^2/4 = \frac{1}{r-1}(\frac{1}{q}t_0)^r$. One can check with a direct calculation that $t_0 \in (2, 4)$. Since

$$\inf_{r\in(1,2)} \frac{t_0 - 1}{t_0^2/4} = \inf_{t\in(2,4)} \frac{t-1}{t^2/4} = 3/4$$

and the functions t^2 and $|t|^r$ are convex, we conclude that

$$H_q^*(t) \ge \frac{3}{4}\min\left\{t^2/4, \frac{1}{r-1}\left(\frac{1}{q}|t|\right)^r\right\} \ge \frac{3}{16}\min\{t^2, |t|^r\},$$

where we also used the fact that

$$\inf_{r\in(1,2)} \frac{1}{r-1}\left(\frac{1}{q}\right)^r = 1/4. \qquad \square$$

Proof of Corollary 2.2.2. We follow the classical Herbst argument (see, e.g., [51]) and the calculations from [5]. Take a function $f\colon (\mathbb{R}^d)^n \to \mathbb{R}$ and denote

$$A = \sup_{x\in\mathbb{R}^{dn}} |\nabla f(x)|, \qquad B = \sup_{x\in\mathbb{R}^{dn}} \|\nabla f(x)\|_{q,2}.$$

Moreover, let $F(\lambda) = \int_{\mathbb{R}^{dn}} e^{\lambda f(x)} d\mu^{\otimes n}$. Then

$$\lambda F'(\lambda) = \int_{\mathbb{R}^{dn}} \lambda f(x) e^{\lambda f(x)} d\mu^{\otimes n}$$

and hence, since μ satisfies the modified log-Sobolev inequality with some constant $C = C(C_{LO}, r)$ (by Theorem 2.2.1) and by the tensorization property,

$$\begin{aligned}\lambda F'(\lambda) - F(\lambda)\log F(\lambda) &= \mathrm{Ent}_{\mu^{\otimes n}}(e^{\lambda f}) \\ &\leq C \int_{\mathbb{R}^{dn}} \sum_{i=1}^{n} H_q\Big(\frac{\lambda}{2}|\nabla_i f|\Big) e^{\lambda f} d\mu^{\otimes n} \\ &\leq 2C \max\big\{(A\lambda/2)^2, (B\lambda/2)^q\big\} F(\lambda),\end{aligned}$$

where we used the inequality $\sum_i \max\{a_i^2, b_i^q\} \leq 2\max\{\sum_i a_i^2, \sum_i b_i^q\}$. After dividing both sides by $\lambda^2 F(\lambda)$ we can rewrite this as

$$\Big(\frac{1}{\lambda}\log F(\lambda)\Big)' \leq 2C \max\big\{(A\lambda/2)^2, (B\lambda/2)^q\big\}/\lambda^2.$$

Since the right-hand side is an increasing function of $\lambda > 0$ and

$$\lim_{\lambda \to 0^+} \frac{1}{\lambda} \log F(\lambda) = \int_{\mathbb{R}^{dn}} f d\mu^{\otimes n},$$

we deduce from the last inequality that

$$\frac{1}{\lambda}\log F(\lambda) \leq \int_{\mathbb{R}^{dn}} f d\mu^{\otimes n} + 2C\max\big\{(A\lambda/2)^2, (B\lambda/2)^q\big\}/\lambda,$$

which is equivalent to

$$\int_{\mathbb{R}^{dn}} e^{\lambda f} d\mu^{\otimes n} \leq \exp\Big(\lambda \int_{\mathbb{R}^{dn}} f d\mu^{\otimes n} + 2C\max\big\{(A\lambda/2)^2, (B\lambda/2)^q\big\}\Big).$$

Therefore from Chebyshev's inequality we get, for $t > 0$ and any $\lambda > 0$,

$$\begin{aligned}\mu^{\otimes n}\Big(f \geq \int_{\mathbb{R}^{dn}} f d\mu^{\otimes n} + t\Big) &\leq \frac{\int_{\mathbb{R}^{dn}} e^{2\lambda f} d\mu^{\otimes n}}{\exp(2\lambda \int_{\mathbb{R}^{dn}} f d\mu^{\otimes n} + 2\lambda t)} \\ &\leq \exp\big(-2\lambda t + 2C\max\big\{(A\lambda)^2, (B\lambda)^q\big\}\big).\end{aligned}$$

Now we can optimize the right-hand side with respect to λ. Let U and V be such that $A = U^{1/2}V$, $B = U^{1/q}V$. We have

$$\max\big\{(A\lambda)^2, (B\lambda)^q\big\}\big) = U\max\{(V\lambda)^2, (V\lambda)^q\} = UH_q(V\lambda)$$

and hence

$$\mu^{\otimes n}\Big(f \geq \int_{\mathbb{R}^{dn}} f d\mu^{\otimes n} + t\Big) \leq \exp\big(-2CUH_q^*(\frac{t}{CUV})\big).$$

Using Lemma 2.5.2 and the definitions of U and V we get

$$\mu^{\otimes n}\Big(f \geq \int_{\mathbb{R}^{dn}} f d\mu^{\otimes n} + t\Big) \leq \exp\Big(-\frac{3}{8}\min\Big\{\frac{t^2}{CA^2}, \frac{t^r}{C^{r-1}B^r}\Big\}\Big),$$

which yields the assertion of the corollary. □

Proof of Corollary 2.2.3. Let f_ε be the convolution of f and a Gaussian kernel, i.e., $f_\varepsilon(x) = \mathbb{E}\, f(x + \sqrt{\varepsilon}G)$, where $G \sim \mathcal{N}(0, I)$. This function clearly inherits from f the estimates of the Lipschitz constants. Since it is smooth, the ℓ_2-norm and the norm $\|\cdot\|_{q,2}$ of its gradient can be estimated pointwise by L_2 and $L_{r,2}$, respectively. Therefore we can apply Corollary 2.2.2 to f_ε. Moreover, $|f_\varepsilon(x) - f(x)| \leq L_2\sqrt{\varepsilon}\,\mathbb{E}\,|G|$ and hence f_ε converges uniformly to f as ε tends to zero. This observation ends the proof of the corollary. □

Proof of Corollary 2.2.5. We follow the approach of Bobkov and Ledoux from Section 2 of [17]. Take a set $A \subset \mathbb{R}^{dn}$ with $\mu^{\otimes n}(A) \geq 1/2$. For $x = (x_1, \ldots, x_n) \in (\mathbb{R}^d)^n$, denote

$$F(x) = F_A(x) = \inf_{a \in A} \sum_{i=1}^{n} \min\{|x_i - a_i|^2, |x_i - a_i|^r\}$$

(note that here $|\cdot|$ stands for the ℓ_2-norm on $\mathbb{R}^d$). Take any $t > 0$ and set $f = \min\{F, t\}$. We claim that for all $x, y \in (\mathbb{R}^d)^n$,

$$|f(x) - f(y)| \leq 2\sqrt{t}|x - y|, \qquad |f(x) - f(y)| \leq 2t^{1/q}\|x - y\|_{r,2}. \tag{2.5.1}$$

Suppose that we already know that this holds. Note that $(2t^{1/q})^r = 2^r t^{r-1} \leq 4t^{r-1}$. Also, $\int_{\mathbb{R}^{dn}} f d\mu^{\otimes n} \leq t/2$ since $F = 0$ on A and $\mu^{\otimes n}(A) \geq 1/2$. Consequently, by Corollary 2.2.3 and (2.5.1),

$$\begin{aligned}\mu^{\otimes n}(F_A \geq t) \leq \mu^{\otimes n}(f \geq t) &\leq \mu^{\otimes}\big(f \geq \int_{\mathbb{R}^{dn}} f d\mu^{\otimes n} + t/2\big) \\ &\leq \exp\Big(-\frac{3K_1}{8}\min\Big\{\frac{(t/2)^2}{4t}, \frac{(t/2)^r}{4t^{r-1}}\Big\}\Big) \\ &\leq \exp\Big(-\frac{3K_1}{128}t\Big)\end{aligned}$$

where $K_1 = \min\{1/C_{mLS}^2, 1/C_{mLS}^{r-1}\}$ and C_{mLS} is the constant with which, by Theorem 2.2.1, the modified log-Sobolev inequality holds for μ. Since clearly

$$\{F_A < t\} \subset A + \Big\{(x_1, \ldots, x_n) \in (\mathbb{R}^d)^n : \sum_{i=1}^n \min\{|x_i|^2, |x_i|^r\} \le t\Big\},$$

this yields the first assertion of the corollary. The second part follows by the inclusion

$$\Big\{(x_1, \ldots, x_n) \in (\mathbb{R}^d)^n : \sum_{i=1}^n \min\{|x_i|^2, |x_i|^r\} \le t\Big\} \subset \sqrt{t}B_2^{dn} + t^{1/r} B_{r,2}^{n,d}.$$

It remains to prove the claim (2.5.1). To this end, consider the functions

$$G(x) = \sum_{i=1}^n \min\{|x_i|^2, |x_i|^r\}$$

and $g(x) = \min\{G, t\}$. Since g is locally Lipschitz it suffices to show that, a.e.,

$$\sum_{i=1}^n |\nabla_i g|^2 \le 4t, \qquad \sum_{i=1}^n |\nabla_i g|^q \le 2^q t.$$

Indeed, this will imply that (2.5.1) holds with g in place of f (note that the norm $\|\cdot\|_{r,2}$ is dual to the norm $\|\cdot\|_{q,2}$). Since $f(x) = \inf_{a\in A} g(x-a)$ (and the infimum of Lipschitz functions is Lipschitz with the same constant), the same estimates will be inherited by f.

On the open set $\{G > t\}$ the estimates obviously hold (since g is constant). The set $\{G = t\}$ is Lebesgue negligible. Thus in what follows it suffices to consider the set $\{G < t\}$ on which $g = G$.

If, for some i, $|x_i| < 1$, then

$$|\nabla_i g(x)|^2 = 4|x_i|^2 = 4\min\{|x_i|^2, |x_i|^r\}.$$

Also,

$$|\nabla_i g(x)|^q = 2^q|x_i|^q \le 2^q|x_i|^2 = 2^q \min\{|x_i|^2, |x_i|^r\}.$$

If on the other hand $|x_i| > 1$, then

$$|\nabla_i g(x)|^2 = r^2|x_i|^{2(r-1)} = r^2\big(\min\{|x_i|^2, |x_i|^r\}\big)^{2/q}.$$

In this case,

$$|\nabla_i g(x)|^2 = r^2|x_i|^{2(r-1)} \le 4|x_i|^r = 4\min\{|x_i|^2, |x_i|^r\},$$
$$|\nabla_i g(x)|^q = r^q \min\{|x_i|^2, |x_i|^r\}.$$

Thus, a.e. (the set where $|x_i| = 1$ for some i is negligible),

$$|\nabla_i g(x)|^2 \le 4\min\{|x_i|^2, |x_i|^r\},$$
$$|\nabla_i g(x)|^q \le 2^q \min\{|x_i|^2, |x_i|^r\}.$$

Consequently, on the set $\{G < t\} = \{g < t\}$, we a.e. have

$$\sum_{i=1}^n |\nabla_i g(x)|^2 \le 4g(x) \le 4t,$$
$$\sum_{i=1}^n |\nabla_i g(x)|^q \le 2^q g(x) \le 2^q t.$$

This finishes the proof. □

2.6 Weighted vs. modified log-Sobolev inequality

In the previous sections we showed that the Latała–Oleszkiewicz inequality implies the modified log-Sobolev inequality. To the best of our knowledge the question about the reverse implication is open even for measures on the real line. We would however like to present one somewhat related result.

It is known that if a probability measure μ on $\mathbb{R}^d$ satisfies a certain weighted log-Sobolev inequality (and an integrability condition), then it also satisfies a modified log-Sobolev inequality, see Theorem 3.4 in [25] (in the context of a specific measure on the real line a similar argument appears already in the large entropy case of the proof of Theorem 3.1 from [31]). The goal of this subsection is to show that the converse implication does not hold in general, even for measures on the real line.

We first present a workable criterion for satisfying the weighted log-Sobolev inequality.

Proposition 2.6.1. *Let $d\mu(x) = e^{-V(x)}dx$ be a probability measure on the real line. Assume that $V : \mathbb{R} \to \mathbb{R}$ is symmetric, of class C^2 in some neighborhood of ∞, and that*

(i) $\liminf_{x\to\infty} V'(x) > 0$,

(ii) $\lim_{x\to\infty} \frac{V''(x)}{V'(x)^2} = 0$.

Then, there exists $C < \infty$ such that μ satisfies the following weighted log-Sobolev inequality: for every $f : \mathbb{R} \to \mathbb{R}$,

$$\mathrm{Ent}_\mu(f^2) \le C \int_{\mathbb{R}} f'(x)^2(1 + |x|^{2-r})d\mu(x), \tag{2.6.1}$$

if and only if

$$\limsup_{x\to\infty} \frac{V(x)}{|x|^{2-r}V'(x)^2} < \infty.$$

Remark 2.6.2. The condition (ii) can be weakened to $\limsup_{x\to\infty} \frac{|V''(x)|}{V'(x)^2} < 1$.

Proof of Proposition 2.6.1. Denote $W(x) := V(x) - \log(1 + |x|^{2-r})$, $x \in \mathbb{R}$. By the Bobkov–Götze criterion [20] (cf. Theorems 1 and 3 in [14]), μ satisfies the weighted log-Sobolev inequality if and only if

$$\sup_{x>0} \mu((x,\infty)) \log\Big(\frac{1}{\mu((x,\infty))}\Big) \int_0^x e^{W(t)}dt < \infty. \tag{2.6.2}$$

Of course, it suffices to investigate what happens for $x \to \infty$.

Note that

$$\liminf_{x\to\infty} W'(x) = \liminf_{x\to\infty} \Big(V'(x) - \frac{(2-r)x^{1-r}}{1+x^{2-r}}\Big) > 0$$

(by assumption (i)) and

$$\lim_{x\to\infty} \frac{W''(x)}{W'(x)^2} = 0$$

(by (ii) and the fact that $W''(x) = V''(x) + o(1)$). Thus, as $x \to \infty$,

$$\mu((x,\infty)) = \int_x^\infty e^{-V(t)}dt \sim \frac{e^{-V(x)}}{V'(x)},$$

$$\int_0^x e^{W(t)}dt \sim \frac{e^{W(x)}}{W'(x)} = \frac{e^{V(x)}}{(1+|x|^{2-r})(V'(x)+o(1))}$$

(here by '$\sim$' we mean that the ratio of both sides tends to 1 as $x \to \infty$; to prove that this is indeed the case it suffices to consider the ratio of the derivatives of both sides). Therefore, (2.6.2) holds if and only if

$$\limsup_{x\to\infty} \frac{V(x) + \log V'(x)}{(1+|x|^{2-r})(V'(x)+o(1))V'(x)} < \infty,$$

which, since $V'(x)$ is bounded away from zero as $x \to \infty$, happens if and only if

$$\limsup_{x\to\infty} \frac{V(x)}{|x|^{2-r}V'(x)^2} < \infty.$$

This ends the proof. □

Our example is a modification of the example constructed by Cattiaux and Guillin [24] to prove that the log-Sobolev inequality is strictly stronger than Talagrand's transportation cost inequality.

Proposition 2.6.3. *For* $r \in (1,2)$ *and* $\max\{r/2, r-1/r\} < \beta - 1 < r - 1/2$ *define*

$$V(x) = V_{r,\beta}(x) = |x|^{r+1} + (r+1)|x|^r \sin^2(x) + |x|^\beta, \quad x \in \mathbb{R}.$$

Let $\mu_{r,\beta}$ *be the probability measure with density proportional to* $e^{-V_{r,\beta}(x)}$. *Then* $\mu_{r,\beta}$ *satisfies the modified log-Sobolev inequality* (2.1.3) *and the Latała–Oleszkiewicz inequality* (2.1.1) *(with* $d = 1$*).*

On the other hand, $\mu_{r,\beta}$ *does not satisfy the weighted log-Sobolev inequality* (2.6.1).

Proof. Let us first note that $\beta \in (r, r+1)$. For $x > 0$,

$$V(x) = x^{r+1} + (r+1)x^r \sin^2(x) + x^\beta,$$
$$V'(x) = (r+1)(1 + \sin(2x))x^r + (r+1)rx^{r-1}\sin^2(x) + \beta x^{\beta-1}.$$

Clearly, $V'(x) \geq \beta x^{\beta-1}$; in particular $\liminf_{x\to\infty} V'(x) > 0$. Moreover, for $x > 1$, $|V''(x)|$ can be bounded by Mx^r for some constant $M = M(r,\beta)$. Thus,

$$\lim_{x\to\infty} \frac{|V''(x)|}{V'(x)^2} \leq \lim_{x\to\infty} \frac{Mx^r}{\beta^2 x^{2(\beta-1)}} = 0,$$

since $\beta - 1 > r/2$. We are thus in position to apply workable versions of the criteria for the modified and weighted log-Sobolev inequalities (note that the normalization of $\mu_{r,\beta}$ amounts to adding a constant to the potential V, which does not affect the calculations and reasoning below).

First note that

$$\lim_{x\to\infty} \frac{V(x)}{V'(x)^q} \leq \lim_{x\to\infty} \frac{(r+3)x^{r+1}}{\beta^q x^{(\beta-1)q}} = 0,$$

since $(\beta-1)q > (r-1/r)q = r+1$. Thus, by the Barthe–Roberto criterion (see (2.1.6)), $\mu_{r,\beta}$ satisfies the modified log-Sobolev and the Latała–Oleszkiewicz inequality.

On the other hand, for certain values of $x \to \infty$ (e.g., for $x = k\pi - \pi/4$, $k \in \mathbb{N}$), we have $|V'(x)| \leq ((r+1)r + \beta)x^{\beta-1}$. Hence

$$\limsup_{x\to\infty} \frac{V(x)}{x^{2-r}V'(x)^2} \geq \lim_{x\to\infty} \frac{x^{r+1}}{x^{2-r}((r+1)r+\beta)^2 x^{2(\beta-1)}} = \infty,$$

since $\beta - 1 < r - 1/2$. Thus, by Proposition 2.6.1 above, $\mu_{r,\beta}$ cannot satisfy the weighted log-Sobolev inequality. □

We finish with a short remark.

Remark 2.6.4. The introduction of [73], suggests that the results of our Theorem 2.2.1 are contained in [83], namely that it follows from [83] that the F_q-Sobolev inequality (2.3.1) implies the modified log-Sobolev inequality (2.1.3). We would like to rectify this: Wang's paper [83] deals with measures with *faster* decay than Gaussian. He proves that in that setting an appropriate super Poincaré inequality (or equivalently, an appropriate F-Sobolev inequality) implies a certain weighted log-Sobolev inequality. However, in our setting (measures with tail decay *slower* than Gaussian), we have an example of a measure which satisfies the modified log-Sobolev inequality (2.1.3) and the Latała–Oleszkiewicz inequality (2.1.1) (or equivalently, the F_q-Sobolev inequality (2.3.1)), but does not satisfy the weighted log-Sobolev inequality (2.6.1). Therefore Theorem 2.2.1 cannot be deduced from Wang's paper [83].

Chapter 3

Concentration of measure for convex functions: introduction and preliminaries

3.1 Motivation

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 connections between them (some of which we have discussed in the preceding chapters). 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 1990's ([77, 78], see also [49, 54, 67, 68] and the monograph [51] for subsequent developments). A crucial feature of these results is that they are satisfied under less restrictive assumptions concerning the regularity of the underlying probability measure when compared to inequalities valid for all Lipschitz functions.

The research presented in the next chapters originated from the following question.

Question 3.1.1. *Adamczak [1] gave a sufficient condition for a probability measure on the real line to satisfy the log-Sobolev inequality restricted to the the class of convex functions. Can one give a sufficient and necessary condition? What about the modified versions of the log-Sobolev inequality?*

Remark 3.1.2. For brevity in what follows we shall slightly informally refer to the log-Sobolev inequality for convex functions simply as "the convex log-Sobolev inequality" (with a similar convention for other inequalities).

Around the time when first partial results were obtained by Adamczak and the author [5], a new type of weak transportation cost inequalities has been introduced by Gozlan, Roberto, Samson, and Tetali in [39] and then studied further by those authors and Shu in [38]. As these developments turned out to be crucial for our work, we shall start by describing them. We shall also discuss another important result obtained earlier, connecting dimension-free concentration inequalities for convex functions with the convex Poincaré inequality [37].

3.2 Convex Poincaré inequality and dimension-free concentration for convex sets

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 $f\colon \mathbb{R}^n \to \mathbb{R}$ we have

$$\operatorname{Var} f(X) \leq \frac{1}{\lambda}\, \mathbb{E}\, |\nabla f(X)|^2, \tag{3.2.1}$$

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

$$|\nabla f(x)| = \limsup_{y \to x} \frac{|f(y) - f(x)|}{|y - x|}. \tag{3.2.2}$$

Note that this coincides with the length of the 'true' gradient provided f is differentiable at x. Also, it is enough to assume that (3.2.1) 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 [37] 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 $t > 0$,

$$\mu^{\otimes N}(A + tB_2^{Nn}) \geq 1 - 2\exp(-ct), \tag{3.2.3}$$

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 (3.2.3) is equivalent to the one-sided deviation inequality for convex 1-Lipschitz functions, i.e.

$$\mathbb{P}(f(X_1, \ldots, X_N) \geq \operatorname{Med} f(X_1, \ldots, X_N) + t) \leq 2e^{-ct} \tag{3.2.4}$$

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

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

3.3 Weak transport–entropy inequalities

We turn to weak transport–entropy inequalities. We shall present the necessary definitions and some fundamental results from [39, 38].

3.3.1 Definitions

Following [39] we denote by $\mathcal{P}_1(\mathbb{R}^n)$ the class of all probability measures ν on $\mathbb{R}^n$ such that $\int_{\mathbb{R}^n} |x| d\nu(x) < \infty$.

Definition 3.3.1. Let μ_1 and μ_2 be probability measures on $\mathbb{R}^n$. Assume that $\mu_2 \in \mathcal{P}_1(\mathbb{R}^n)$. For a convex, lower semicontinuous function $\theta \colon \mathbb{R}^n \to [0, \infty]$ with $\theta(0) = 0$ define the weak (barycentric) transport cost between μ and ν as

$$\overline{\mathcal{T}}_\theta(\mu_2|\mu_1) = \inf_\pi \int_{\mathbb{R}^n} \theta\big(x - \int_{\mathbb{R}^n} y p_x(dy)\big)\mu_1(dx),$$

where:

- the infimum is taken over all couplings π between μ_1 and μ_2 (that is probability measures on $\mathbb{R}^n \times \mathbb{R}^n$ such that $\pi(A \times \mathbb{R}^n) = \mu_1(A)$ and $\pi(\mathbb{R}^n \times A) = \mu_2(A)$ for any Borel set $A \subset \mathbb{R}^n$),
- for $x \in \mathbb{R}^n$, $p_x(\cdot)$ is the conditional measure defined μ_1 almost surely by the relation $\pi(dxdy) = p_x(dy)\mu_1(dx)$.

In probabilistic notation one can write

$$\overline{\mathcal{T}}_\theta(\mu_2|\mu_1) = \inf_{(X,Y)} \mathbb{E}\, \theta(X - \mathbb{E}(Y|X)),$$

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 and Y are distributed according to μ_1 and μ_2, respectively. The adjective *weak* stands for the fact that, by Jensen's inequality, $\overline{\mathcal{T}}_\theta(\mu_2|\mu_1)$ is smaller than the classical transport cost,

$$\mathcal{T}_\theta(\mu_2, \mu_1) = \inf_\pi \int_{\mathbb{R}^n \times \mathbb{R}^n} \theta(x - y)\pi(dx, dy). \tag{3.3.1}$$

The word *barycentric* (which we will usually omit) corresponds to the fact that the barycenter $\int_{\mathbb{R}^n} y p_x(dy)$ appears in the definition.

Remark 3.3.2. More often than not, we will assume that θ is symmetric: $\theta(x) = \theta(-x)$ for $x \in \mathbb{R}^n$.

Due to the asymmetry between μ_1 and μ_2, one can now introduce three different inequalities related to the cost $\overline{\mathcal{T}}_\theta$. Recall that the relative entropy $H(\nu|\mu)$ has been defined above in (1.3.4).

Definition 3.3.3. Let $\mu \in \mathcal{P}_1(\mathbb{R}^n)$ and $\theta\colon \mathbb{R}^n \to [0,\infty]$ 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 $\nu \in \mathcal{P}_1(\mathbb{R}^n)$,
$$\overline{\mathcal{T}}_\theta(\nu|\mu) \le H(\nu|\mu),$$
- $\overline{\mathbf{T}}_\theta^-$ if for every probability measure $\nu \in \mathcal{P}_1(\mathbb{R}^n)$,
$$\overline{\mathcal{T}}_\theta(\mu|\nu) \le H(\nu|\mu),$$
- $\overline{\mathbf{T}}_\theta$ if μ satisfies both $\overline{\mathbf{T}}_\theta^+$ and $\overline{\mathbf{T}}_\theta^-$.

The definition of those inequalities given by Gozlan, Roberto, Samson, and Tetali in [39] differs formally from the one presented above (which is taken from the follow-up paper [38]). 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.

3.3.2 Dual formulations

We will rely on the following Bobkov–Götze type criterion for the weak-transport inequality proved in Lemma 4.1 in [38] (and in a slightly different version also in Proposition 4.5 in [39]). The proof in [38] is presented for the real line, but it is not difficult to see that it generalizes to arbitrary dimension.

Proposition 3.3.4. *Let $\theta\colon \mathbb{R}^n \to \mathbb{R}_+$ be a convex cost function, $\theta(0) = 0$, $\lim_{x\to\infty} \theta(x) = \infty$. For all functions $f\colon \mathbb{R}^n \to \mathbb{R}$ bounded from below, $x \in \mathbb{R}^n$, and $t > 0$ set*
$$Q_t f(x) = Q_t^\theta f(x) = \inf_{y\in\mathbb{R}^n} \big\{ f(y) + t\theta\Big(\frac{x-y}{t}\Big)\big\}.$$
Then

(i) μ satisfies $\overline{\mathbf{T}}_\theta^+$ if and only if for all convex $f\colon \mathbb{R}^n \to \mathbb{R}$, bounded from below,
$$\exp\Big(\int_{\mathbb{R}^n} Q_1 f d\mu\Big) \int_{\mathbb{R}^n} e^{-f} d\mu \le 1; \qquad (3.3.2)$$

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

$$\int_{\mathbb{R}^n} \exp(Q_1 f)d\mu \ \exp\Big(-\int_{\mathbb{R}^n} f d\mu\Big) \le 1; \tag{3.3.3}$$

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

$$\int_{\mathbb{R}^n} \exp(Q_t f)d\mu \int_{\mathbb{R}^n} e^{-f} d\mu \le 1 \tag{3.3.4}$$

holds with $t = 2$. Conversely, if μ satisfies (3.3.4) *for some $t > 0$, then it satisfies $\overline{\mathbf{T}}_{\tilde\theta}$ with $\tilde\theta(\cdot) = t\theta(\cdot/t)$.*

Moreover, the inequality (3.3.2) *(resp.* (3.3.3)*) for all convex, Lipschitz functions bounded from below is a sufficient condition for $\overline{\mathbf{T}}_\theta^+$ (resp. $\overline{\mathbf{T}}_\theta^-$).*

We will refer to Q_t^θ as the infimum convolution operator and to the inequality (3.3.4) as the convex infimum convolution inequality. From our point of view, the importance of infimum convolution inequalities (whether for convex or all smooth functions) lies also in the fact that they imply concentration inequalities (cf. Subsection 1.4 above). To this end, the inequality (3.3.4) was introduced by Maurey [54] (who considered both the classical setting of smooth functions and the convex setting). The relation with transportation cost inequalities was first observed by Bobkov and Götze [20].

Remark 3.3.5. The infimum convolution operator appears in the Hopf–Lax formula for the solution of the Hamilton–Jacobi partial differential equation (see, e.g., Chapter 3 of [29]). We will use this fact repeatedly in the proofs.

3.3.3 Relation to convex log-Sobolev inequalities

Let us now briefly explain why weak transportation inequalities can be helpful when looking for an answer to Question 3.1.1.

Recall that in the classical setting there are strong links between the log-Sobolev inequality, transport–entropy inequalities, and the infimum convolution inequality (see Section 1.3 above). Similar connections have been observed by Gozlan, Roberto, Samson, and Tetali in [39] for *convex* log-Sobolev inequalities and *weak* transport–entropy inequalities. More specifically, the following is proved in Theorem 8.8 of [39].

Proposition 3.3.6. *Let $\theta(x) = |x|^2$ for $x \in \mathbb{R}^n$. For a probability measure μ on $\mathbb{R}^n$ the following conditions are equivalent.*

(i) *There exists $a > 0$ such that the measure μ satisfies $\overline{\mathbf{T}}^-_{\theta(a\cdot)}$.*

(ii) For every $s > 0$ we have $\int_{\mathbb{R}} e^{s|x|} d\mu(x) < \infty$ and there exists $c > 0$ such that the convex log-Sobolev inequality holds, i.e.,

$$\mathrm{Ent}_\mu(e^\varphi) \le c \int_{\mathbb{R}} |\nabla\varphi|^2 e^\varphi d\mu$$

for every smooth convex Lipschitz function $\varphi\colon \mathbb{R} \to \mathbb{R}$.

We shall come back to this equivalence later on, in Chapter 4, in the more general case of modified log-Sobolev inequalities for convex functions and non-quadratic cost functions θ, as well as in Chapter 5, in the case of transportation inequalities with a quadratic-linear cost function θ. Let us only mention, that similar connections hold also for inequalities for concave functions and inequality $\overline{\mathbf{T}}_\theta^+$ (with some additional technical restrictions), cf. Sections 4.4.3 and 5.4.

3.3.4 Characterization on the real line

Let τ be the symmetric exponential measure on $\mathbb{R}$ with density $\frac{1}{2}e^{-|x|}$. For a Borel probability measure μ on $\mathbb{R}$ we denote by U_μ the unique left-continuous and non-decreasing map transporting τ onto the reference measure μ. More precisely, let

$$U_\mu(x) := F_\mu^{-1} \circ F_\tau(x) = \begin{cases} F_\mu^{-1}(\frac{1}{2}e^{-|x|}) & \text{if } x < 0, \\ F_\mu^{-1}(1 - \frac{1}{2}e^{-|x|}) & \text{if } x \ge 0, \end{cases}$$

where

$$F_\mu^{-1}(t) := \inf\{y \in \mathbb{R} : F_\mu(y) \ge t\} \in \mathbb{R} \cup \{\pm\infty\}, \quad t \in [0,1],$$

is the generalized inverse of the cumulative distribution function defined as

$$F_\mu(x) = \mu((-\infty, x]), \qquad x \in \mathbb{R}.$$

Denote moreover

$$\Delta_\mu(h) := \sup_{x\in\mathbb{R}}\{U_\mu(x+h) - U_\mu(x)\}, \quad h > 0. \tag{3.3.5}$$

Recall the following result obtained by Gozlan, Roberto, Samson, Shu, and Tetali (see Theorem 1.3 of [38]).

Proposition 3.3.7. *Let $\theta\colon \mathbb{R} \to [0,\infty)$ be a symmetric convex cost function. Assume that $\theta(t) = t^2$ for $|t| \le t_0$ (for some $t_0 > 0$). For a probability measure $\mu \in \mathcal{P}_1(\mathbb{R})$ the following conditions are equivalent.*

(i) There exists $a > 0$ such that μ satisfies $\overline{\mathbf{T}}_{\theta(a\cdot)}$.

(ii) There exists $b > 0$ such that for all $h > 0$ we have

$$\Delta_\mu(h) \leq \frac{1}{b}\theta^{-1}(t_0^2 + h),$$

where $\theta^{-1}\colon [0,\infty) \to [0,\infty)$ denotes the inverse of the (increasing) function θ restricted to $[0,\infty)$.

The dependence of the constants is the following: (i) implies (ii) with $b = \kappa_1 a$ and (ii) implies (i) with $a = \kappa_2 b$, where

$$\kappa_1 = \frac{t_0}{8\theta^{-1}(\log(3) + t_0^2)}, \qquad \kappa_2 = \frac{\min(1, t_0)}{210\theta^{-1}(2 + t_0^2)}.$$

3.4 Organization of the rest of the thesis

In the following chapters we study functional inequalities for convex functions, using extensively and extending the results of [39] and [38]. Apart from the above short introduction, those chapters are essentially self-contained. They cover respectively the following topics:

- the convex log-Sobolev inequality on the real line (in particular, we answer Question 3.1.1);
- the convex Poincaré inequality in $\mathbb{R}^n$;
- general concentration inequalities which can be derived from infimum convolution inequalities for convex functions of the form (3.3.4);
- application to convex infimum convolution inequalities with optimal cost functions and comparison of moments.

Finally, Appendix A contains basic facts concerning Hamilton–Jacobi equations, which are used in the proof of Theorem 5.1.1.

Chapter 4

Convex log-Sobolev inequality: characterization on the real line

4.1 Introduction and main results

Let μ be a Borel probability measure on $\mathbb{R}^n$. We say that μ satisfies the *(modified) log-Sobolev inequality for a class of functions* $\mathcal{F}$ (with cost function $H\colon \mathbb{R}^n \to [0,\infty)$ and constant $c < \infty$) if for every $f \in \mathcal{F}$ we have

$$\mathrm{Ent}_\mu(e^f) \le \int_{\mathbb{R}^n} H(c\nabla f) e^f d\mu. \tag{4.1.1}$$

In the most classical setting where $H(x) = |x|^2$ and $\mathcal{F}$ is the class of C^1 functions this inequality was first introduced by Gross in [41]. In this case it can be rewritten in the form

$$\mathrm{Ent}_\mu(g^2) \le 4C \int_{\mathbb{R}^n} |\nabla g|^2 d\mu, \tag{4.1.2}$$

where $C = c^2$, or in yet another form which states that the entropy of a positive function g is bounded by its Fisher information:

$$\mathrm{Ent}_\mu(g) \le C \int_{\mathbb{R}^n} \frac{|\nabla g|^2}{g} d\mu.$$

Remark 4.1.1. In our definition of the log-Sobolev inequality the constant c is introduced as a scaling of the argument of the function H rather than as a multiplicative constant outside of the integral. We decided to use this form because it simplifies some of the calculations in Section 4.2. Clearly, in the most common cases, e.g., in the case of the functions H_p from Example 4.1.7 below, the two formulations are equivalent up to numerical constants (for the functions H_p those constants depend on p).

Due to its tensorization property the log-Sobolev inequality is a powerful tool and can be used to obtain dimension-free concentration bounds (via the so-called Herbst argument, see Proof of Corollary 2.2.2 above for a version of it). It has been investigated also in more general settings of Riemannian manifolds and in the context of applications to the study of Markov chains (see, e.g., the monographs [9, 11] and the expository article [28]).

In Chapter 2 we have discussed some results concerning the inequality (4.1.1) with $\mathcal{F} = C^1$ and non-quadratic functions H. Now we shall be interested in the case when $\mathcal{F}$ is the class of *convex* functions.[1] The restriction of the class of functions allows us to work with measures which satisfy much weaker regularity conditions. Most importantly, their supports do not need to be connected. On the other hand, a disturbing issue arises: the log-Sobolev inequality for convex functions yields via standard reasonings only deviation inequalities for the upper tail of functions, i.e.

$$\mu^{\otimes N}\Big(\Big\{x \in (\mathbb{R}^n)^N : f(x) \geq \int_{(\mathbb{R}^n)^N} f \, d\mu^{\otimes N} + t\Big\}\Big), \quad t \geq 0$$

(see, e.g., [49]; in the classical setting of smooth functions one obtains bounds on the lower tail simply by working with $-f$ instead of f, but this is precluded in our situation because $-f$ is usually not convex).

Our goal is to give an intrinsic characterization of probability measures on the real line for which the convex log-Sobolev inequality holds. As a corollary we will obtain dimension-free concentration bounds for upper *and* lower tails of convex functions of independent random variables satisfying the convex log-Sobolev inequality. Before stating our main result let us outline what has been known in the convex setting.

In [1] Adamczak found a sufficient condition for a probability measure on the real line to satisfy the convex log-Sobolev inequality with $H(x) = x^2$, $x \in \mathbb{R}$. This has been extended to functions of the form $H(x) = \max\{x^2, x^q\}$, where $q > 2$, by Adamczak and the author in [5].

Recall from Section 3.3 that Gozlan, Roberto, Samson, Shu, and Tetali [38] established a condition equivalent to the weak transport inequality $\overline{\mathbf{T}}_\theta$ on the real line. By the very definition, it is therefore sufficient for the formally weaker inequality $\overline{\mathbf{T}}_\theta^-$, which in turn is equivalent to a certain convex modified log-Sobolev inequality on the real line (as already hinted in Section 3.3.3; see Proposition 4.2.1 below for a more precise statement). Their condition, expressed in terms of the unique left-continuous and non-decreasing map

[1] Note that in this case the dimension-free tensorization property still holds, but the alternative formulations (1.3.2) and (4.1.2)—with g being convex, respectively convex and non-negative—are no longer equivalent to (4.1.1).

transporting the symmetric exponential measure onto μ (see Proposition 3.3.7), is in fact weaker than the condition considered in [5].

On the other hand, it follows from [38] and the independent work of Feldheim, Marsiglietti, Nayar, and Wang in [30] that in the case when H is quadratic on an interval near zero and then infinite the following are equivalent:

- the condition on the tail of the measure μ from [5] in the case $\beta = 0$,
- the condition on monotone transport map obtained in [38],
- the log-Sobolev inequality for convex functions

(and further: the convex Poincaré inequality, the convex infimum convolution inequality with a quadratic-linear cost function). In what follows we extend this result to more general choices of the function H.

Before formulating our main result, we recall for the reader's convenience the notation introduced in Section 3.3.4 above:

- U_μ is the unique left-continuous and non-decreasing map transporting the symmetric exponential measure onto the reference measure μ,

$$U_\mu(x) = F_\mu^{-1} \circ F_\tau(x) = \begin{cases} F_\mu^{-1}(\frac{1}{2}e^{-|x|}) & \text{if } x < 0, \\ F_\mu^{-1}(1 - \frac{1}{2}e^{-|x|}) & \text{if } x \geq 0, \end{cases}$$

  where the generalized inverse of the cumulative distribution function F_μ is defined as

$$F_\mu^{-1}(t) = \inf\{y \in \mathbb{R} : F_\mu(y) \geq t\} \in \mathbb{R}, \quad t \in (0,1);$$

- the modulus of continuity of U_μ is denoted

$$\Delta_\mu(h) = \sup_{x \in \mathbb{R}}\{U_\mu(x+h) - U_\mu(x)\}, \quad h > 0.$$

We stress that by all means we can have $\lim_{h \to 0^+} \Delta_\mu(h) > 0$. This corresponds to the fact that the support of μ does not need to be connected.

Recall also that $H^* \colon \mathbb{R} \to \mathbb{R}$ stands for the Fenchel-Legendre transform of $H \colon \mathbb{R} \to \mathbb{R}$, given by the formula

$$H^*(x) = \sup_{y \in \mathbb{R}}\{xy - H(y)\}, \quad x \in \mathbb{R}.$$

Remark 4.1.2. Note that if H is a symmetric convex function, such that $H(x) = \frac{1}{4}x^2$ for $x \in [-2t_0, 2t_0]$ for some $t_0 > 0$, then H^* is also quadratic near zero (namely, $H^*(x) = x^2$ for $x \in [-t_0, t_0]$, since for such x the supremum in the definition of H^* is attained at $y = 2x$).

Our main result is the following. The results are new already in the case of the quadratic function $H(x) = \frac{1}{4}x^2$.

Theorem 4.1.3. *Let $H\colon \mathbb{R} \to [0,\infty)$ be a symmetric convex function, such that $H(x) = \frac{1}{4}x^2$ for $x \in [-2t_0, 2t_0]$ for some $t_0 > 0$. Suppose moreover that there exist $A \in [1,\infty)$ and $\alpha \in (1,2]$ such that*

$$\forall_{x\in\mathbb{R}}\ \forall_{s\in[0,1]} \quad H(sx) \le A s^\alpha H(x). \tag{4.1.3}$$

Denote $\theta(t) = H^*(t)$. For a probability measure μ on the real line the following conditions are equivalent.

(i) For every $s > 0$ we have $\int_{\mathbb{R}} e^{s|x|}d\mu(x) < \infty$ and there exists $c > 0$ such that

$$\mathrm{Ent}_\mu(e^\varphi) \le \int_{\mathbb{R}} H(c\varphi')e^\varphi d\mu$$

for every smooth convex Lipschitz function $\varphi\colon \mathbb{R} \to \mathbb{R}$.

(ii) There exists $b > 0$ such that for all $h > 0$,

$$\Delta_\mu(h) \le \frac{1}{b}\theta^{-1}(t_0^2 + h),$$

where $\theta^{-1}\colon [0,\infty) \to [0,\infty)$ denotes the inverse of the (increasing) function θ restricted to $[0,\infty)$.

Let us state three technical remarks in this place.

Remark 4.1.4. The dependence of constants is explicit but complicated and hence we shall only specify it throughout the proof of the theorem. However, in the case when $H(x) = \frac{x^2}{4}$ the dependence of constants can be simplified: (ii) implies (i) with $c = \frac{2}{\kappa b}$ where

$$\kappa = \max_{t_0>0}\left\{\frac{\min(1,t_0)}{210\sqrt{2+t_0^2}}\right\} = \frac{1}{210\sqrt{3}};$$

(i) implies that $\Delta_\mu(h) \le 16c(\frac{2}{3} + \sqrt{h/2})$.

Remark 4.1.5. In (i) the assumption about exponential integrability is added in order to exclude very heavy-tailed measures for which the only exponentially integrable convex Lipschitz functions are constants and hence the convex log-Sobolev inequality is trivially satisfied, whereas (ii) cannot hold.

Remark 4.1.6. Suppose for simplicity that μ is symmetric and has a nowhere vanishing density. By the definition of U_μ we have

$$\mu([U_\mu(x+h),\infty)) = \tau([x+h,\infty)) = e^{-h}\tau([x,\infty)) = e^{-h}\mu([U_\mu(x),\infty))$$

for $x, h \ge 0$. This easy computation shows that (ii) implies:

(ii') There exists $b > 0$ such that for every $h > 0$,

$$\mu([x + g(h), \infty)) \leq e^{-h}\mu([x, \infty)) \quad \forall x \geq 0,$$

where $g(h) = \frac{1}{b}\theta^{-1}(t_0^2 + h)$.

This observation can be used to compare the condition on Δ_μ with the sufficient conditions obtained in [1, 5] (see definition of the class $\mathcal{M}_\beta$; see also Subsection 1.3.1 in [5]).

The following example shows that our theorem covers among others the modified log-Sobolev inequalities from Chapter 2 restricted to the class of convex functions. Hence, it can be viewed as a convex counterpart of the results of Barthe and Roberto from [15].

Example 4.1.7. The condition (4.1.3) is stable under taking convex combinations or maxima of functions and, for $1 < p < \infty$, the function

$$H(x) = H_p(x) = \begin{cases} \frac{1}{4}x^2 & \text{if } |x| \leq 2, \\ \frac{2}{p}(|x/2|^p - 1) + 1 & \text{if } |x| > 2, \end{cases}$$

satisfies (4.1.3) with $\alpha = \min\{p, 2\}$ and $A = 1$.[2] Indeed, if $x > 0$ and $s \in (0, 1)$, then by Cauchy's mean value theorem

$$\frac{H(sx)}{H(x)} = \frac{sH'(s\xi)}{H'(\xi)} = \begin{cases} s^2 & \text{if } 0 \leq s\xi \leq \xi \leq 2, \\ s^2(\xi/2)^{2-p} = s^p(s\xi/2)^{2-p} & \text{if } 0 \leq s\xi \leq 2 < \xi, \\ s^p & \text{if } 2 < s\xi \leq \xi \end{cases}$$

for some $\xi \in (0, x)$. In either case,

$$H(sx) \leq \max\{s^2, s^p\}H(x) = s^{\min\{p,2\}}H(x).$$

Taking into account the results from [38], we can give a handful of conditions equivalent to the convex log-Sobolev inequality on the real line. For simplicity we state the result only for the quadratic cost.

Theorem 4.1.8. *Let $\theta(t) = t^2$ for $t \in \mathbb{R}$. For a probability measure μ on the real line the following conditions are equivalent.*

(i) For every $s > 0$ we have $\int_{\mathbb{R}} e^{s|x|}d\mu(x) < \infty$ and there exists $C > 0$ such that

$$\mathrm{Ent}_\mu(e^\varphi) \leq C\int_{\mathbb{R}} |\varphi'|^2 e^\varphi d\mu$$

for every smooth convex Lipschitz function $\varphi\colon \mathbb{R} \to \mathbb{R}$.

[2] There is a conflict in notation with the function H_q introduced in Chapter 2, but for $p = q$ both functions are comparable up to multiplicative constants.

(ii) There exist $a, b > 0$ such that for all $h > 0$,

$$\Delta_\mu(h) \le \sqrt{a + bh}.$$

(iii) There exists $a_1 > 0$ such that μ satisfies the inequality $\overline{\mathbf{T}}^-_{\theta(a_1\cdot)}$.

(iv) There exists $a_2 > 0$ such that μ satisfies the inequality $\overline{\mathbf{T}}_{\theta(a_2\cdot)}$.

(v) There exists $t > 0$ such that μ satisfies the infimum convolution inequality

$$\int_{\mathbb{R}} \exp(Q_t^\theta f) d\mu \int_{\mathbb{R}} \exp(-f) \le 1$$

for every convex function $f\colon \mathbb{R} \to \mathbb{R}$ bounded from below.

In each of the implications the constants in the conclusion depend only on the constants in the premise.

As a consequence, we immediately obtain concentration bounds for both the upper and lower tails of convex Lipschitz functions. Indeed, they follow at once from the convex infimum convolution inequality. Let us stress here once again that for the lower tail such estimates were previously unknown (since working directly with the convex log-Sobolev inequality leads only to bounds for the upper tail, see [49]). For simplicity we also state the result in the case $H(x) = \frac{1}{4}x^2$, but one can obtain appropriate bounds also in the case when H is not quadratic[3]. We refer to, e.g., Corollary 3 from [30] for a similar statement (with H quadratic on an interval near zero and then infinite); see also Corollary 5.11 from [39] and Chapter 6 below for a general overview of concentration properties implied by weak transportation inequalities and convex infimum convolution inequalities.

Corollary 4.1.9. *Let μ be a probability measure on $\mathbb{R}$ such that for every $s > 0$ we have $\int_{\mathbb{R}} e^{s|x|} d\mu(x) < \infty$ and*

$$\mathrm{Ent}_\mu(e^\varphi) \le C \int_{\mathbb{R}} |\varphi'|^2 e^\varphi d\mu$$

for every smooth convex Lipschitz function $\varphi\colon \mathbb{R} \to \mathbb{R}$. Then there exist $A, B < \infty$ (depending only on C), such that for any convex function $\varphi\colon \mathbb{R}^N \to \mathbb{R}$ which is 1-Lipschitz (with respect to the Euclidean norm on $\mathbb{R}^N$) we have

$$\mu^{\otimes N}\big(\{x \in \mathbb{R}^N : |\varphi(x) - \mathrm{Med}_{\mu^{\otimes N}}(\varphi)| \ge t\}\big) \le Be^{-t^2/A}, \quad t \ge 0,$$

[3] Using simply the fact that condition (ii) from Theorem 4.1.3 is, by Proposition 3.3.7 and Proposition 3.3.4 above, equivalent to the infimum convolution inequality for convex functions; for the quadratic cost these results are summarized in Theorem 4.1.8

where

$$\mathrm{Med}_{\mu^{\otimes N}}(\varphi) = \inf\big\{s \in \mathbb{R} : \ \mu^{\otimes N}(\{x \in \mathbb{R}^N : \varphi(x) \le s\}) \ge 1/2\big\}$$

is the median of φ.

The rest of the chapter is organized as follows. In Sections 4.2 and 4.3 we prove that the conditions (i) and (ii) are both equivalent to a weak transport–entropy inequality. Finally, in Section 4.4 we summarize the results of the previous sections and give the proof of Theorem 4.1.3 (and Corollary 4.1.9). We also recapitulate all conditions equivalent to the convex log-Sobolev inequality in the quadratic case, present a corollary connecting the convex log-Sobolev inequality to Talagrand's $\mathbf{T}_2$ inequality, and pose some open questions.

4.2 Equivalence of the convex log-Sobolev inequality and the weak transportation inequality

In this section we establish the equivalence of the convex log-Sobolev inequality and the weak transport–entropy inequality. In the case of the quadratic cost this was done in [39] (see also [5] for related results for other cost functions). Using the techniques developed therein, especially in the dual formulation (3.3.3), we extend this result to a wider class of cost functions. We work with measures on the real line, but in contrast to Section 4.3 there are no problems with extending the results of this section to a higher dimensional setting (cf. [39, 5] and Lemma 5.4.1 below).

Let $H\colon \mathbb{R} \to [0,\infty)$ be a symmetric convex function, such that $H(x) = \frac{1}{4}x^2$ for $x \in [-2t_0, 2t_0]$ for some $t_0 > 0$. Recall that H^* is is also quadratic near zero (see Remark 4.1.2). Moreover, we assume that there exist $A \in [1,\infty)$ and $\alpha \in (1,2]$ such that

$$\forall_{x\in\mathbb{R}}\ \forall_{s\in[0,1]} \quad H(sx) \le A s^\alpha H(x). \tag{4.2.1}$$

Note that we have

$$H(x) \ge \frac{1}{4A} t_0^{2-\alpha} x^\alpha \quad \text{for } x \ge t_0 \tag{4.2.2}$$

(this follows immediately by taking $s = t_0/x$ in condition (4.2.1)). Also, $\lim_{x\to\infty} H^*(x)/x = \infty$, since otherwise there would exist $M > 0$ such that

$H^*(x) \le M|x|$ for all sufficiently large x, which would lead to a contradiction with the assumption that H takes finite values. In particular, H^* is not Lipschitz.

The main technical result of this section is the following.

Proposition 4.2.1. *For a probability measure μ on the real line the following conditions are equivalent.*

(i) There exists $a > 0$ such that the measure μ satisfies $\overline{\mathbf{T}}^{-}_{H^*(a\cdot)}$.

(ii) For every $s > 0$ we have $\int_{\mathbb{R}} e^{s|x|} d\mu(x) < \infty$ and there exists $c > 0$ such that

$$\mathrm{Ent}_\mu(e^\varphi) \le \int_{\mathbb{R}} H(c\varphi')e^\varphi d\mu$$

for every smooth convex Lipschitz function $\varphi\colon \mathbb{R} \to \mathbb{R}$.

The dependence of the constants is the following: (i) implies (ii) with $c = 2/a$; (ii) implies (i) with $a = ((\alpha - 1)/A)^{1/\alpha} c^{-1}$.

The implication (i) $\implies$ (ii) is a general fact and no special assumptions are used in the proof. For the sake of completeness we sketch the main argument here.

Proof of Proposition 4.2.1, (i) $\implies$ (ii). The exponential integrability follows from the dual formulation (3.3.3) tested with the function $x \mapsto s|x|$ (cf. [5, p. 86]; note that we use the fact that H^* is not Lipschitz).

According to Proposition 8.3 from [39], (i) implies that the so-called (τ)−log-Sobolev inequality holds: for all functions $f\colon \mathbb{R} \to \mathbb{R}$ such that $\int_{\mathbb{R}} f e^f d\mu < \infty$ we have

$$\mathrm{Ent}_\mu(e^f) \le \frac{1}{1-\lambda} \int_{\mathbb{R}} (f - R^\lambda f)e^f d\mu$$

for every $\lambda \in (0, 1)$. Here

$$R^\lambda f(x) := \inf_p \Big\{ \int_{\mathbb{R}} f(y)p(dy) + \lambda H^*\big(a(x - \int_{\mathbb{R}} yp(dy))\big) \Big\},$$

where the infimum is taken over all probability measures p on $\mathbb{R}$ (note that we skip the dependence on H^* in the notation). For convex functions f the infimum above is achieved at some Dirac measure:

$$R^\lambda f(x) = \inf_{y\in\mathbb{R}} \{f(y) + \lambda H^*(a(x - y))\}$$

(indeed, if we replace the measure p by a Dirac mass at the point $y_p = \int_{\mathbb{R}} yp(dy)$, then, by Jensen's inequality, the expression under the infimum in the definition of R^λ will not increase; cf. Theorem 2.11 and Subsection 8.2 in [39]). Now, by convexity of f,

$$\begin{aligned} f(x) - R^\lambda f(x) &= f(x) + \sup_{y\in\mathbb{R}}\{-f(y) - \lambda H^*(a(x-y))\} \\ &\le f(x) + \sup_{y\in\mathbb{R}}\{-f(x) - f'(x)(y-x) - \lambda H^*(a(x-y))\} \\ &= \lambda H\Big(\frac{f'(x)}{a\lambda}\Big), \end{aligned}$$

where we have used the fact that $H^{**} = H$. Thus, after taking $\lambda = 1/2$, we arrive at the assertion (with $c = 2/a$). □

For the proof of the second implication we need the following simple lemma. It is based on an argument of Maurey (cf. Proof of Theorem 3 in [54]), but takes into account the observation that for compactly supported measures it doesn't matter whether for large arguments the cost function is quadratic or equal to $+\infty$. Recall that $|\cdot|$ stands for the Euclidean norm.

Lemma 4.2.2. *Let μ be a probability measure on $\mathbb{R}^n$ such that the diameter of the support of μ is not greater than D and denote*

$$\theta_D(x) = \begin{cases} \frac{1}{4D^2}|x|^2 & \text{if } |x| \le D, \\ +\infty & \text{if } |x| > D. \end{cases}$$

Then for any convex function $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ bounded from below

$$\int_{\mathbb{R}^n} e^{Q_1^{\theta_D}\varphi} d\mu \int_{\mathbb{R}^n} e^{-\varphi} d\mu \le 1, \tag{4.2.3}$$

where $Q_1^{\theta_D}\varphi(x) = \inf\{\varphi(y) + \theta_D(x-y) : y \in \mathbb{R}^n\}$, $x \in \mathbb{R}^n$, stands for the infimum convolution.

Conversely, if inequality (4.2.3) holds for some D, then the support of μ is bounded: if $x, y \in \operatorname{Supp}\mu$, then $|x-y| \le D$.

Proof. Assume that the diameter of the support of μ is bounded by D. Take a convex function φ, bounded from below. By adding a constant to φ, we may assume that $\inf_{\operatorname{Supp}\mu}\varphi = 0$. Take any $\varepsilon > 0$, any $x \in \operatorname{Supp}\mu$, and let $z \in \operatorname{Supp}\mu$ be such that $\varphi(z) < \varepsilon$. Moreover, define $y = (1-\lambda)x + \lambda z$, where $\lambda \in [0,1]$. Then $|x-y| \le D$ and hence

$$\begin{aligned} Q_1^{\theta_D}\varphi(x) &\le \varphi(y) + \frac{1}{4D^2}|x-y|^2 \le (1-\lambda)\varphi(x) + \lambda\varphi(z) + \frac{\lambda^2}{4D^2}|x-z|^2 \\ &\le (1-\lambda)\varphi(x) + \lambda\varepsilon + \frac{\lambda^2}{4}. \end{aligned}$$

We now let $\varepsilon \to 0^+$, and then optimize with respect to $\lambda \in [0,1]$: if $\varphi(x) \geq 1/2$ we take $\lambda = 1$, and if $0 \leq \varphi(x) \leq 1/2$ we take $\lambda = 2\varphi(x)$. This gives $Q_1^{\theta_D}\varphi(x) \leq k(\varphi(x))$, where

$$k(u) = (u - u^2) \cdot 1_{\{u \in [0,1/2)\}} + \tfrac{1}{4} \cdot 1_{\{u \geq 1/2\}}.$$

Note that we have $e^{k(u)} \leq 2 - e^{-u}$. Indeed, for $u = 1/2$ (or larger) the inequality holds, and for $u \in [0, 1/2)$ we have

$$(e^{u-u^2} + e^{-u})/2 \leq e^{-u^2/2} \cosh(u - u^2/2) \leq e^{-u^2/2} \cosh(u) \leq 1.$$

Hence

$$\int e^{Q_1^{\theta_D}\varphi} \mu \leq \int e^{k(\varphi)} d\mu \leq 2 - \int e^{-\varphi} d\mu \leq \Big(\int e^{-\varphi} d\mu \Big)^{-1}.$$

Conversely, assume that inequality (4.2.3) holds, but there exist $x_0, y_0 \in \operatorname{Supp} \mu$ such that $|x_0 - y_0| > D$. Then there exist $\varepsilon, \delta > 0$, such that $\mu(B(x_0, \varepsilon)) > \delta$ and $\mu(\mathbb{R}^n \setminus B(x_0, D + 2\varepsilon)) > \delta$. Consider now $\varphi_a : \mathbb{R}^n \to \mathbb{R}$ defined by the formula $\varphi_a(x) = a \operatorname{dist}\big(x, B(x_0, \varepsilon)\big)$ for $a > 0$. For $x \in \mathbb{R}^n \setminus B(x_0, D + 2\varepsilon)$ we have

$$Q_1^{\theta_D} \varphi_a(x) = \inf_{y \in \mathbb{R}^n : |x-y| \leq D} \big\{ a \operatorname{dist}\big(y, B(x_0, \varepsilon)\big) + \frac{1}{4D^2} |x - y|^2 \big\} \geq a\varepsilon.$$

Moreover, $\varphi_a = 0$ on $B(x_0, \varepsilon)$. Thus for sufficiently large $a > 0$,

$$\begin{aligned} \int_{\mathbb{R}^n} e^{Q_1^{\theta_D}\varphi_a} d\mu \int_{\mathbb{R}^n} e^{-\varphi_a} d\mu &\geq \int_{\mathbb{R}^n \setminus B(x_0, D+2\varepsilon)} e^{Q_1^{\theta_D}\varphi_a} d\mu \int_{B(x_0,\varepsilon)} e^{-\varphi_a} d\mu \\ &\geq \delta^2 \exp(a\varepsilon) > 1, \end{aligned}$$

which contradicts the inequality (4.2.3). □

Proof of Proposition 4.2.1, (ii) $\implies$ (i). Assume that (ii) holds. Without loss of generality we can assume that μ is absolutely continuous with respect to the Lebesgue measure. Indeed, if γ is a uniform probability distribution on $[0, \delta]$, then by Lemma 4.2.2, Proposition 3.3.4 (iii), and the already proved implication (i) $\implies$ (ii) of Proposition 4.2.1 it satisfies the convex log-Sobolev inequality with a quadratic-linear function

$$H_0(x) = \delta^2 x^2 \mathbf{1}_{\{|x| \leq 1/(2\delta)\}} + (\delta|x| - 1/4)\mathbf{1}_{\{|x| > 1/(2\delta)\}}$$

(and constant $c = 2$). Hence by (4.2.2) the product measure $\mu \otimes \gamma$ on $\mathbb{R}^2$ satisfies (for sufficiently small $\delta > 0$)

$$\operatorname{Ent}_{\mu\otimes\gamma}(e^{\phi}) \leq \int_{\mathbb{R}^2} \big(H(c\phi'_x) + H(c\phi'_y) \big) e^{\phi} d\mu \otimes \gamma \tag{4.2.4}$$

for all smooth convex Lipschitz functions $\phi\colon \mathbb{R}^2 \to \mathbb{R}$. Let $\varphi\colon \mathbb{R} \to \mathbb{R}$ be a smooth convex Lipschitz function and let $\phi\colon \mathbb{R}^2 \to \mathbb{R}$ be defined by the formula $\phi(x, y) = \varphi(x + \varepsilon y)$, $x, y \in \mathbb{R}$. Applying (4.2.4) to the function ϕ and using the assumption (4.2.1), we see that the convolution $\mu * \gamma_\varepsilon$, where $\gamma_\varepsilon(\cdot) = \gamma(\cdot/\varepsilon)$, satisfies, up to a multiplicative constant which tends to 1 as $\varepsilon \to 0^+$, the same modified log-Sobolev inequality as μ:

$$\mathrm{Ent}_{\mu*\gamma_\varepsilon}(e^\varphi) \leq (1 + A\varepsilon^\alpha) \int_{\mathbb{R}} H(c\varphi')e^\varphi d\mu{*}\gamma_\varepsilon.$$

The reader will easily check that the proof below shows that $\mu * \gamma_\varepsilon$ satisfies $\overline{\mathbf{T}}^-_{H^*(a_\varepsilon \cdot)}$ with $a_\varepsilon = ((\alpha - 1)/A_\varepsilon)^{1/\alpha} c^{-1}$, where $A_\varepsilon = A \cdot (1 + A\varepsilon^\alpha)$ (the multiplicative constant $1{+}A\varepsilon^\alpha$ will appear in one place in the estimate of $F'(t)$). By the Lebesgue dominated convergence theorem applied to the dual formulation this implies that μ satisfies $\overline{\mathbf{T}}^-_{H^*(a\cdot)}$ with $a = ((\alpha - 1)/A)^{1/\alpha} c^{-1}$. Indeed, in the dual formulation it suffices to consider convex Lipschitz functions only. If f is Lipschitz, then for every $x \in \mathbb{R}$, there exists a compact set K_x such that for all $\varepsilon \in [0, 1]$,

$$Q_1^{H^*(a_\varepsilon \cdot)} f(x) = \inf_{y \in K_x} \big\{ f(x - y) + H^*(a_\varepsilon y) \big\}$$

(the infimum cannot be attained for big values of y, since f is Lipschitz and H^* grows faster than a linear function when its argument tends to $+\infty$). Since H^* is uniformly continuous on K_x, we have pointwise convergence $Q_1^{H^*(a_\varepsilon \cdot)} f(x) \to Q_1^{H^*(a \cdot)} f(x)$ as $\varepsilon \to 0^+$. We can use the Lebesgue dominated convergence theorem, because

$$Q_1^{H^*(a_\varepsilon \cdot)} f(x) \leq f(x) \leq f(0) + \mathrm{Lip}(f)|x|$$

and by assumption $\int_{\mathbb{R}} e^{s|x|} d\mu(x) < \infty$ for every $s > 0$.

Note that if μ is absolutely continuous, then standard approximation shows that (ii) holds for all convex Lipschitz functions (by the Rademacher theorem the gradient is then almost surely well defined).

Take a convex Lipschitz function $\varphi\colon \mathbb{R} \to \mathbb{R}$, denote for brevity

$$Q_t\varphi(x) = Q_t^{H^*(\cdot)}\varphi(x) = \inf_{y \in \mathbb{R}} \Big\{ \varphi(y) + tH^*\Big(\frac{x - y}{t}\Big) \Big\},$$

and set $F(t) = \int_{\mathbb{R}} e^{k(t)Q_t\varphi(x)} d\mu(x)$ for $t > 0$ (for some non-decreasing function k yet to be determined). The Hamilton–Jacobi partial differential equation,

$$\partial_t Q_t\varphi + H(\partial_x Q_t\varphi) = 0,$$

holds almost surely on $(0,\infty)\times\mathbb{R}$ (see, e.g., Chapter 3 of [29]). Using first this fact, then applying the log-Sobolev inequality, and finally the estimate $H(ck(t)\cdot)\le Ac^\alpha k(t)^\alpha H(\cdot)$ which follows from the assumption (4.2.1) if only $ck(t)\le 1$, we arrive at

$$\begin{aligned}
k(t)F'(t) &= k(t)\int_{\mathbb{R}} e^{k(t)Q_t\varphi(x)}\Big(k'(t)Q_t\varphi(x)+k(t)\partial_t Q_t\varphi(x)\Big)d\mu(x)\\
&= k(t)\int_{\mathbb{R}} e^{k(t)Q_t\varphi(x)}\Big(k'(t)Q_t\varphi(x)-k(t)H\big(\partial_x Q_t\varphi(x)\big)\Big)d\mu(x)\\
&= k'(t)F(t)\log F(t)+k'(t)\operatorname{Ent}_\mu\big(e^{k(t)Q_t\varphi}\big)\\
&\quad - k^2(t)\int_{\mathbb{R}} e^{k(t)Q_t\varphi(x)}H\big(\partial_x Q_t\varphi(x)\big)d\mu(x)\\
&\le k'(t)F(t)\log F(t)+k'(t)\int_{\mathbb{R}} e^{k(t)Q_t\varphi(x)}H\big(ck(t)\partial_x Q_t\varphi(x)\big)d\mu(x)\\
&\quad - k^2(t)\int_{\mathbb{R}} e^{k(t)Q_t\varphi(x)}H\big(\partial_x Q_t\varphi(x)\big)d\mu(x)\\
&\le k'(t)F(t)\log F(t)\\
&\quad + \big[Ac^\alpha k'(t)k(t)^\alpha - k^2(t)\big]\cdot\int_{\mathbb{R}} e^{k(t)Q_t\varphi(x)}H\big(\partial_x Q_t\varphi(x)\big)d\mu(x).
\end{aligned}$$

Denote $\widetilde{A}=A^{1/\alpha}$ and take

$$k(t)=(\widetilde{A}c)^{-\alpha/(\alpha-1)}\big((\alpha-1)t\big)^{1/(\alpha-1)}.$$

Then $k(0)=0$, $ck(t)\le 1$ for $t\in[0,\widetilde{A}^\alpha c/(\alpha-1)]$, and

$$Ac^\alpha k'(t)k(t)^\alpha - k^2(t)=0.$$

For such a choice of k, the differential inequality obtained above is equivalent to $(\log(F(t))/k(t))'\le 0$ for almost all $t\in(0,\widetilde{A}^\alpha c/(\alpha-1))$ (note that—by the fact that $Q_t\varphi$ is Lipschitz and the integrability properties of μ—the function $\log(F(t))/k(t)$ is absolutely continuous on bounded closed intervals contained in $(0,\infty)$). Since $Q_t\varphi\le\varphi$, this yields

$$\frac{\log F(t)}{k(t)}\le\liminf_{s\to 0^+}\frac{\log F(s)}{k(s)}\le\lim_{s\to 0^+}\frac{\log(\int_{\mathbb{R}} e^{k(s)\varphi(x)}d\mu(x))}{k(s)}=\int_{\mathbb{R}}\varphi d\mu$$

for $t\in(0,\widetilde{A}^\alpha c/(\alpha-1)]$. This is exactly the dual formulation of $\overline{\mathbf{T}}^-_{tk(t)H^*(\cdot/t)}$ (see Proposition 3.3.4 (ii)). Taking $t=t_*=\widetilde{A}c/(\alpha-1)^{1/\alpha}$ we see that $t_*k(t_*)=1$, $t_*\le\widetilde{A}^\alpha c/(\alpha-1)$ (recall that $A\ge 1$ and $1<\alpha\le 2$), and thus also $ck(t_*)\le 1$. We conclude that μ satisfies $\overline{\mathbf{T}}^-_{H^*(a\cdot)}$ with $a=1/t_*=((\alpha-1)/A)^{1/\alpha}c^{-1}$. □

4.3 From the weak transportation inequality to the condition on U_μ

In the previous section we showed the equivalence of the convex log-Sobolev inequality and the weak transport–entropy inequality. In this section, working towards the proof of Theorem 4.1.3, we deal with weak transport–entropy inequalities.

Throughout this section let μ be a measure on the real line (which is not a Dirac mass) with median $m = F_\mu^{-1}(1/2)$. Denote

$$s_\mu \coloneqq \inf \operatorname{Supp}(\mu) \in [-\infty, \infty), \quad t_\mu \coloneqq \sup \operatorname{Supp}(\mu) \in (-\infty, \infty].$$

Let $\theta\colon \mathbb{R} \to [0, \infty)$ be a symmetric convex cost function such that $\theta(t) = t^2$ for $t \in [-t_0, t_0]$ for some $t_0 > 0$. Note that by convexity θ is increasing. We moreover assume that

$$\int_0^\infty \theta(x) e^{-\lambda x} dx < \infty \quad \text{for any } \lambda > 0. \tag{4.3.1}$$

The goal of this section is to provide a proof of the following stronger version of Proposition 3.3.7, where $\overline{\mathbf{T}}_{\theta(a\cdot)}$ is replaced by the formally weaker inequality $\overline{\mathbf{T}}^-_{\theta(a\cdot)}$. Note that this in particular means that the inequalities $\overline{\mathbf{T}}_\theta$ and $\overline{\mathbf{T}}^-_\theta$ are equivalent.

Proposition 4.3.1. *The following conditions are equivalent.*

(i) There exists $a > 0$ such that μ satisfies $\overline{\mathbf{T}}^-_{\theta(a\cdot)}$.

(ii) There exists $b > 0$ such that for all $h > 0$ we have

$$\Delta_\mu(h) \le \frac{1}{b} \theta^{-1}(t_0^2 + h).$$

The dependence of the constants is the following: (i) implies (ii) with

$$b = \frac{\min(a, 1)}{16} \Big(1 + \frac{1}{a t_0} \theta^{-1}\Big(\frac{\log(2e^{C_\theta/2} - 1)}{2}\Big)\Big)^{-1},$$

where $C_\theta = \int_0^\infty \theta(2 + \frac{1}{\log 2} t) e^{-t} dt$, and (ii) implies (i) with $a = \kappa b$, where $\kappa = \frac{\min(1, t_0)}{210 \theta^{-1}(2 + t_0^2)}$.

For the proof we need the following lemma which explains the connection between the condition (ii) satisfied by the map U_μ and transport–entropy inequalities connected to costs which are equal to zero in a neighborhood of zero. The lemma is an immediate consequence of Theorem 2.2 from [35] (cf. Theorem 6.1 in [38]). In what follows the symbol $\int_{t_1}^{t_2}$ always denotes an integral over the *open* interval (t_1, t_2).

Lemma 4.3.2. *Let $\beta\colon [0,\infty) \to [0,\infty)$ be a function which is equal to zero on the interval $[0,t_0]$ and then strictly increasing; denote its inverse by $\beta^{-1}\colon [0,\infty) \to [t_0,\infty)$. The following conditions are equivalent.*

(i) There exists $d > 0$ such that for all $h > 0$ and $x \in \mathbb{R}$,

$$\Delta_\mu(h) \le \tfrac{1}{d}\beta^{-1}(h).$$

(ii) There exist $k > 0$, $K < \infty$ such that

$$\sup_{x\in[m,t_\mu)} \frac{1}{\mu((x,\infty))}\int_x^\infty \exp\bigl(\beta\bigl(k(u-x)\bigr)\bigr)\mu(du) \le K,$$
$$\sup_{x\in(s_\mu,m]} \frac{1}{\mu((-\infty,x))}\int_{-\infty}^x \exp\bigl(\beta\bigl(k(x-u)\bigr)\bigr)\mu(du) \le K.$$

The dependence of the constants is the following: (i) implies (ii) with $K = 3$ and $k = d\frac{t_0}{18\beta^{-1}(2)}$; (ii) implies (i) with $d = k\frac{t_0}{4\beta^{-1}(\log K)}$.

The above lemma will be our main tool in the proof of the implication (i) $\Longrightarrow$ (ii) of Proposition (4.3.1), but first we need a couple of preparatory results. They concern the consequences of the convex Poincaré inequality, which as we recall below is satisfied in our setting.

Lemma 4.3.3. *If μ satisfies $\overline{\mathbf{T}}^-_{\theta(a\cdot)}$, then μ satisfies the following convex Poincaré inequality: for any convex Lipschitz $f : \mathbb{R} \to \mathbb{R}$ we have*

$$\mathrm{Var}_\mu(f) \le \frac{1}{2a^2}\int_{\mathbb{R}} |\nabla f|^2 d\mu, \tag{4.3.2}$$

where $|\nabla f(x)|$ is the length of the gradient of f at x.

Proof. This follows by a standard Taylor expansion argument from the dual formulation (3.3.3) of the transport entropy inequality: we plug in εf instead of f, use the estimate

$$\begin{aligned} Q_1^\theta(\varepsilon f)(x) &= \inf_{y\in\mathbb{R}}\{\varepsilon f(x-y) + \theta(y)\} \ge \varepsilon f(x) + \inf_{y\in\mathbb{R}}\{-\varepsilon|\nabla f(x)|y + \theta(y)\} \\ &\ge \varepsilon f(x) - \theta^*(\varepsilon|\nabla f(x)|) \end{aligned}$$

(valid for convex functions; note that $|\nabla f(x)|$ is just the maximum of the one-sided derivatives which exist at every point), and take $\varepsilon \to 0^+$ (recall that $\theta^*(t) = t^2/4$; alternatively, one could use Proposition 4.2.1 and deduce the Poincaré inequality with a slightly worse constant from the log-Sobolev inequality by a similar argument). □

By the characterization of Bobkov and Götze from Theorem 4.2 in [19] (cf. Theorem 1.5 in [38]), if μ satisfies the convex Poincaré inequality, then there exist $D_1, D_2 > 0$ such that $\Delta_\mu(h) \le D_1 + D_2 h$ for all $h \ge 0$. Following the proof of Theorem 1 from [30] we get a version with explicit constants.

Lemma 4.3.4. *If μ satisfies the convex Poincaré inequality* (4.3.2)*, then we have $U_\mu(x+h) - U_\mu(x) \le \frac{4}{a} + \frac{1}{a\log(2)}h$. However, the constant $4/a$ may be replaced by $2/a$ if we know that x and $x+h$ are of the same sign.*

Proof. Let X, X' be two independent random variables with distribution μ. Fix $u \ge m$ and plug the function $f(x) = \max\{x-u, 0\}$ into (4.3.2):

$$\begin{aligned}
\frac{1}{a^2}\mu([u,\infty)) &\ge 2\operatorname{Var}_\mu(f) = \mathbb{E}(f(X) - f(X'))^2 \\
&\ge \mathbb{E}(f(X) - f(X'))^2\big[\mathbf{1}_{\{X'\le m,\ X\ge u+2/a\}} + \mathbf{1}_{\{X\le m,\ X'\ge u+2/a\}}\big] \\
&\ge \mathbb{E}\, f(X)^2 \mathbf{1}_{\{X\ge u+2/a\}} \ge \frac{4}{a^2}\mu([u+2/a,\infty)),
\end{aligned}$$

Thus,

$$\mu([u+2/a,\infty)) \le \frac{1}{4}\mu([u,\infty)), \quad u \ge m,$$

and similarly

$$\mu((-\infty, u-2/a]) \le \frac{1}{4}\mu((-\infty,u]), \quad u \le m.$$

By the definition of U_μ (and the fact that F_μ is right-continuous) for $x \in \mathbb{R}$ and all $\varepsilon > 0$,

$$F_\mu(U_\mu(x) - \varepsilon) < F_\tau(x) \le F_\mu(U_\mu(x)).$$

Denote now $h_0 = 2\ln(2)$ and let $x, x+h_0 \le 0$. Then, for $\varepsilon > 0$,

$$\begin{aligned}
\mu\big((-\infty, U_\mu(x+h_0) - 2/a - \varepsilon]\big) &\le \frac{1}{4}\mu\big((-\infty, U_\mu(x+h_0) - \varepsilon]\big) \\
&< e^{-h_0}\tau\big((-\infty, x+h_0]\big) = \tau\big((-\infty,x]\big) \\
&\le \mu\big((-\infty, U_\mu(x)]\big).
\end{aligned}$$

The inequality is strict and thus $U_\mu(x+h_0) - U_\mu(x) \le 2/a + \varepsilon$. Since $\varepsilon > 0$ was arbitrary, $U_\mu(x+h_0) - U_\mu(x) \le 2/a$ for $x, x+h_0 \le 0$.

Similarly $U_\mu(x+h_0) - U_\mu(x) \le 2/a$ for $x, x+h_0 \ge 0$, since

$$\begin{aligned}
\mu\big([U_\mu(x) + 2/a + \varepsilon, \infty)\big) &\le \frac{1}{4}\mu\big([U_\mu(x) + \varepsilon, \infty)\big) \le \frac{1}{4}\big(1 - F_\mu(U_\mu(x))\big) \\
&\le e^{-h_0}\big(1 - F_\tau(x)\big) = 1 - F_\tau(x+h_0) \\
&< \mu\big([U_\mu(x+h_0) - \varepsilon, \infty)\big).
\end{aligned}$$

Using these inequalities in a telescoping manner at most $\lceil h/h_0 \rceil \le 1 + h/h_0$ times we conclude that

$$U_\mu(x+h) - U_\mu(x) \le \frac{2}{a} + \frac{1}{a\log(2)}h$$

for any $x \in \mathbb{R}$ and $h \ge 0$ such that x and $x+h$ are of the same sign. If $x < 0 < x+h$, then the additive constant $2/a$ in the last estimate has to be replaced by $4/a$. □

With this result in hand, we can prove our last preparatory lemma.

Lemma 4.3.5. *If μ satisfies the convex Poincaré inequality* (4.3.2)*, then*

$$\frac{1}{\mu((x,\infty))}\int_x^\infty \theta(a(u-x))\mu(du) \le C_\theta \quad \textit{for } \ x \in [m, t_\mu),$$
$$\frac{1}{\mu((-\infty,x))}\int_{-\infty}^x \theta(a(x-u))\mu(du) \le C_\theta \quad \textit{for } x \in (s_\mu, m],$$

where $C_\theta = \int_0^\infty \theta(2 + \frac{1}{\log 2}t)e^{-t}dt$. Moreover, if $\theta(x) = x^2$ then one can choose $C_\theta = 1$.

Proof of Lemma 4.3.5. We only need to prove the first inequality (where $x \in [m, t_\mu)$), the second one can be taken care of in a similar way.

First, we deal with the case $\theta(x) := x^2$. Inequality (4.3.2) implies that

$$A - B \le \frac{1}{2}\mu((x,\infty)),$$

where $A = \int_x^\infty (a(u-x))^2 \mu(du)$ and $B = \left(\int_x^\infty a(u-x)\mu(du)\right)^2$. (This is again obtained by testing (4.3.2) with $u \mapsto a\max\{u-x, 0\}$; a minuscule limit argument is needed to obtain the version with the open interval on the right hand-side: it suffices to consider $x + 1/n$, $n \to \infty$, instead of x.) By the Cauchy-Schwarz inequality $B \le A\mu((x,\infty))$ and thus

$$A \le \frac{1}{2}\mu((x,\infty)) + A\mu((x,\infty)),$$

which, since $x \ge m$, leads to

$$\frac{1}{\mu((x,\infty))}\int_x^\infty \theta(a(u-x))\mu(du) \le \frac{1}{2(1-\mu((x,\infty)))} \le 1.$$

Now we turn to the general θ. By the characterization of Bobkov and Götze [19], there exist $D_1, D_2 > 0$ such that $\Delta_\mu(h) \le D_1 + D_2 h$ for all

$h \geq 0$. By Lemma 4.3.4, we see that in our case one can choose $D_1 = \frac{2}{a}$ and $D_2 = \frac{1}{a \log 2}$.

Fix $x \geq m$ and define $v := \sup\{u : U_\mu(u) \leq x\}$. Since the map U_μ is left-continuous we have $U_\mu(v) \leq x < U_\mu(v + \varepsilon)$ for any $\varepsilon > 0$; also $v \geq 0$ since $x \geq m$. Recall that τ denotes the exponential measure and that

$$\int_{\mathbb{R}} f d\mu = \int_{\mathbb{R}} f \circ U_\mu d\tau$$

for any measurable function f. Thus (note that $\mu((U_\mu(v), x]) = 0$ if we have $U_\mu(v) < x$)

$$\begin{aligned}
\frac{1}{\mu((x,\infty))} \int_x^\infty \theta(a(u-x))\mu(du) &= \frac{1}{\tau((v,\infty))} \int_v^\infty \theta\big(a(U_\mu(u) - x)\big)\tau(du) \\
&\leq e^v \int_v^\infty \theta\big(a(U_\mu(u) - U_\mu(v))\big)e^{-u} du \\
&\leq \int_v^\infty \theta\big(a(D_1 + D_2(u-v))\big)e^{-(u-v)} du \\
&= \int_0^\infty \theta(a(D_1 + D_2 t))e^{-t} dt \\
&= \int_0^\infty \theta\big(2 + \frac{1}{\log 2} t\big)e^{-t} dt = C_\theta < \infty,
\end{aligned}$$

where the last inequality follows from the integrability condition (4.3.1) placed on θ. □

Now we are ready to prove Proposition 4.3.1.

Proof of Proposition 4.3.1. Due to Proposition 3.3.7 we only need to check that if μ satisfies the inequality $\overline{\mathbf{T}}^{-}_{\theta(a\cdot)}$, then the condition from (ii) is satisfied by U_μ.

Fix $x > m$ and consider the function $f(t) = \theta(a[t-x]_+)$. Then clearly $Q_1^{\theta(a\cdot)} f(t) = 0$ if $t \leq x$. For $t > x$,

$$\begin{aligned}
Q_1^{\theta(a\cdot)} f(t) &= \inf_{y \in \mathbb{R}} \{\theta(a[y-x]_+) + \theta(a(t-y))\} \\
&= \inf_{y \in [x,t]} \{\theta(a(y-x)) + \theta(a(t-y))\} = 2\theta(a(t-x)/2),
\end{aligned}$$

where the last equality follows from the fact that we have an inequality due to the convexity of θ and on the other hand the infimum is attained at $y = (x+t)/2$. Hence the dual formulation (3.3.3) of the weak transport inequality implies that

$$\mu((-\infty, x]) + \int_x^\infty \exp\big(2\theta(a(t-x)/2)\big) d\mu(t) \leq \exp\Big(\int_x^\infty \theta(a(t-x))\mu(dt)\Big).$$

Denote $k = 1/2$ and $\beta(u) = 2\theta(a[u - t_0]_+)$ for $u > 0$. Since θ is increasing on $(0, \infty)$ we have $\beta(ku) \le 2\theta(au/2)$. Therefore,

$$\int_x^\infty \exp\big(\beta(k(t-x))\big)\mu(dt) \le \exp\Big(\int_x^\infty \theta(a(t-x))\mu(dt)\Big) - 1 + \mu((x,\infty)). \tag{4.3.3}$$

By Lemmas 4.3.3 and 4.3.5 there exists $C_\theta < \infty$ such that $\int_x^\infty \theta(a(t - x))\mu(dt) \le C_\theta \mu((x, \infty))$. Hence, since $\mu((x, \infty)) \in [0, 1/2]$,

$$\begin{aligned}\frac{\int_x^\infty \exp\big(\beta(k(t-x))\big)\mu(dt)}{\mu((x,\infty))} &\le \frac{\exp\big(C_\theta \mu((x,\infty))\big) - 1 + \mu((x,\infty))}{\mu((x,\infty))} \\ &\le 2e^{C_\theta/2} - 1.\end{aligned}$$

One can deal with $x \le m$ similarly. Since

$$\beta^{-1}(h) = \big(t_0 + \tfrac{1}{a}\theta^{-1}(h/2)\big),$$

Lemma 4.3.2 implies that

$$\begin{aligned}\Delta_\mu(h) &\le \tfrac{1}{d}\beta^{-1}(h) = \tfrac{1}{d}\big(t_0 + \tfrac{1}{a}\theta^{-1}(h/2)\big) \le \frac{1}{d\min(a,1)}\big(t_0 + \theta^{-1}(h/2)\big) \\ &\le \frac{2}{d\min(a,1)}\theta^{-1}(t_0^2 + h),\end{aligned} \tag{4.3.4}$$

where

$$d = \frac{t_0}{8\beta^{-1}(\log(2e^{C_\theta/2} - 1))} = \frac{t_0}{8(t_0 + \frac{1}{a}\theta^{-1}(\frac{1}{2}\log(2e^{C_\theta/2} - 1)))}$$

(recall that $k = 1/2$). This finishes the proof. □

4.4 Summary

4.4.1 Proof of the main results and dependence of constants for $H(x) = \frac{1}{4}x^2$

The results of the two preceding sections allow us to prove our main result.

Proof of Theorem 4.1.3. The implication (ii) $\Longrightarrow$ (i) has been proved in [38] (cf. Proposition 3.3.7 above). The implication (i) $\Longrightarrow$ (ii) follows immediately by combining Propositions 4.2.1 and 4.3.1. The only assumption we need to check is that $\int_0^\infty \theta(x)e^{-\lambda x}dx < \infty$ for any $\lambda > 0$ (i.e., (4.3.1)), but this follows from the scaling condition placed on H. Indeed, for $s \in (0,1]$,

$$H^*(y/s) = \big(H(s\cdot)\big)^*(y) \geq \big(As^\alpha H(\cdot)\big)^*(y) = As^\alpha H^*(y/(As^\alpha)).$$

Taking $z \geq 1$ and substituting into the above inequality $s = z^{-1/(\alpha-1)}$ and $y = As = Az^{-1/(\alpha-1)}$ we arrive at

$$\theta(z) = H^*(z) \leq H^*(A)A^{-1}z^{\alpha/(\alpha-1)},$$

which implies the claim. □

As for the dependence of constants, in the case $H(x) = \frac{1}{4}x^2$ one can take $A = 1$ and $\alpha = 2$ in (4.1.3). Let us consider the implication (i) $\implies$ (ii) from Theorem 4.1.3. In Proposition 4.2.1 we have $a = 1/c$ and moreover we can take $C_\theta = 1$ in Lemma 4.3.5. Therefore, inequality (4.3.4) reads

$$\Delta_\mu(h) \leq \tfrac{1}{d}\big(t_0 + c\sqrt{h/2}\big) \leq 8\frac{t_0 + \frac{2}{3}c}{t_0}\left(t_0 + c\sqrt{h/2}\right), \tag{4.4.1}$$

since

$$d = \frac{t_0}{8\big(t_0 + \frac{1}{a}\sqrt{\frac{1}{2}\log(2e^{1/2}-1)}\big)} \geq \frac{t_0}{8(t_0 + \frac{2}{3}c)}.$$

Taking $t_0 = \frac{2}{3}c$ we obtain the result announced in Remark 4.1.4 (the dependence of constants for the implication (ii) $\implies$ (i) follows directly from Proposition 4.2.1 and Proposition 3.3.7). In fact, we can take $t_0 = c\sqrt[4]{2h/9}$ (which minimizes the right-hand side of (4.4.1)) to obtain a slightly better estimate

$$\Delta_\mu(h) \leq 8c\Big(2/3 + \sqrt{h/2} + 2\sqrt[4]{2h/9}\Big).$$

Proof of Theorem 4.1.8. The assertion follows immediately by combining Theorem 4.1.3, Proposition 3.3.7, and Proposition 3.3.4. □

Finally, the proof of Corollary 4.1.9 reduces to standard techniques. We postpone a more detailed discussion of concentration inequalities which can be derived from the convex infimum convolution inequality to Chapter 6.

Proof of Corollary 4.1.9. By Theorem 4.1.8 the measure μ satisfies the inequality $\overline{\mathbf{T}}_\theta$ for the quadratic cost. An application of, e.g., Corollary 5.11 of [39] completes the proof. Alternatively, for a more self-contained reasoning, one can use item (v) of Theorem 4.1.8, that is the convex infimum convolution inequality, and adapt the approach of [54]. □

4.4.2 Relation to Talagrand's inequality

Assume that $\theta(t) = t^2$ for $t \in \mathbb{R}$ (one can formulate a similar result also for costs other than the quadratic cost). Recall from Chapter 1 that we say that a probability measure μ on the real line satisfies Talagrand's inequality (with constant C) if

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

for every probability measure ν ($\mathcal{T}_\theta$ was defined in (3.3.1)). In the classical setting of smooth functions we have the implication chain

$$\begin{aligned} \text{log-Sobolev inequality} &\implies \text{Talagrand's inequality} \\ &\implies \text{Poincaré inequality} \end{aligned}$$

and these implications are strict (see Section 4.3 of [35] for a nice discussion). From [38] we also know that Talagrand's inequality is strictly stronger than the convex log-Sobolev inequality. The following corollary explains what additional information is carried by it. It is an immediate consequence of Theorem 4.1.3 above and Theorem 1.1 from [35].

Corollary 4.4.1. *A probability measure μ on the real line satisfies Talagrand's inequality if and only if it satisfies the Poincaré inequality for smooth functions and the log-Sobolev inequality for convex functions.*

4.4.3 Further questions

We conclude with three open questions, which to the best of our knowledge are open even in the case $\theta(t) = t^2$.

Question 4.4.2. *Suppose that a probability measure μ on $\mathbb{R}^n$, $n \geq 2$, satisfies the inequality $\overline{\mathbf{T}}^-_{\theta(a\cdot)}$ for some $a > 0$. Does it satisfy the inequality $\overline{\mathbf{T}}_{\theta(a'\cdot)}$ for some $a' > 0$?*

Question 4.4.3. *Suppose that a probability measure μ on the real line satisfies the inequality $\overline{\mathbf{T}}^+_{\theta(a\cdot)}$ for some $a > 0$. Does it satisfy the inequality $\overline{\mathbf{T}}^-_{\theta(a'\cdot)}$, and thus $\overline{\mathbf{T}}_{\theta(\min\{a',a\}\cdot)}$, for some $a' > 0$?*

While these questions are stated in terms of the weak transport–entropy inequalities, they can be equivalently expressed with the use of log-Sobolev inequalities for convex and concave functions. We refer to Theorem 8.15 from [39] and Remark 8.8 from [39] for details and subtleties concerning the log-Sobolev inequality for concave functions and their relation to inequality $\overline{\mathbf{T}}^+_\theta$.

As for the case of the quadratic-linear cost function, we present a partial answer to the above two questions in the next chapter.

The last question is deliberately somewhat vague. It would be nice to have a more dimensional version of Theorem 4.1.3, that is a condition equivalent to the convex log-Sobolev inequality which is not expressed in terms of quantifiers ranging over families of functions or measures (like, e.g., the conditions from Propositions 3.3.4 and 3.3.6), but rather in terms of the measure μ itself.

Question 4.4.4. *Find an intrinsic characterization of probability measures on $\mathbb{R}^n$, $n \geq 2$, which satisfy the convex log-Sobolev inequality.*

Unfortunately, the results of this chapter, and even the very statement of Theorem 4.1.3, relied on techniques specific to the real line (and, moreover, the proofs were rather indirect and based on many other characterizations). This leads to the suspicion that Question 4.4.4 may be very hard to answer.

Chapter 5

Convex Poincaré inequality

5.1 Introduction

Recall from Chapter 3 that we say that a probability measure μ (or, equivalently, a random variable X with law μ) satisfies the *convex Poincaré inequality* with constant $\lambda > 0$ if for all convex functions $f : \mathbb{R}^n \to \mathbb{R}$ we have

$$\operatorname{Var} f(X) \leq \frac{1}{\lambda} \mathbb{E} |\nabla f(X)|^2, \tag{5.1.1}$$

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

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

We already explained in Section 3.2, that—by a result of Gozlan, Roberto, and Samson [37]—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 $\theta : \mathbb{R}^n \to [0, \infty]$ 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

$$\mathcal{T}_\theta(\nu, \mu) = \inf_\pi \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \theta(x - y) \pi(dxdy), \tag{5.1.2}$$

where the infimum is taken over all couplings between μ and ν.

It has been proved in [18] that μ satisfies the Poincaré inequality for *all smooth* functions if and only if there exist constants C, D such that for all probability measures ν,

$$\mathcal{T}_{\theta_{C,D}}(\nu, \mu) \leq H(\nu|\mu), \tag{5.1.3}$$

where

$$\theta_{C,D}(x) = \begin{cases} \frac{|x|^2}{2C} & \text{for } |x| \le CD, \\ D|x| - \frac{CD^2}{2} & \text{for } |x| > CD. \end{cases} \tag{5.1.4}$$

Recently Gozlan, Roberto, Samson, Shu, and Tetali [38] formulated a similar characterization of the *convex* Poincaré inequality *on the real line.* They 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 weak 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 independently in [30].

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

Theorem 5.1.1. *For a probability measure μ on $\mathbb{R}^n$ the following conditions are equivalent.*

(i) There exists $\lambda > 0$ such that μ satisfies the convex Poincaré inequality (5.1.1)*.*

(ii) There exist $C, D > 0$ such that μ satisfies the transportation inequality $\overline{\mathbf{T}}_{\theta_{C,D}}$*.*

Remark 5.1.2. The implication (ii) $\implies$ (i) is standard, in this case $\lambda = \frac{1}{C}$. Unfortunately, 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. We suspect that this is an artifact 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 5.4.3 below for details).

Remark 5.1.3. Thanks to well known tensorization properties of the inequality $\overline{\mathbf{T}}_{\theta_{C,D}}$, Theorem 5.1.1 implies that the convex Poincaré inequality is equivalent to improved two-level dimension-free concentration inequality for convex functions (see Example 6.3.5 below for a precise formulation). Recall from the previous chapters, that in the class of Lipschitz functions such a fact was established by Bobkov and Ledoux [17]; by results due to Gozlan et al. [37] 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 5.1.4. Recall from Chapter 4 that Bobkov and Götze [19] provided a simple characterization of measures on $\mathbb{R}$ which satisfy the convex Poincaré inequality for some $\lambda > 0$ (and thus also the inequality $\overline{\mathbf{T}}_{C,D}$). A similar characterization for $n \geq 2$ seems to be a non-trivial open problem.

The organization of the chapter is as follows. First, in Section 5.2, we present preliminary properties of measures satisfying the convex Poincaré inequality, to be used in the proofs (we refer to Section 3.3 for the definitions of the weak transport–entropy inequalities and preliminary results about them). Section 5.3 contains our most important technical result, i.e., modified log-Sobolev inequalities for convex and concave functions, which in Section 5.4 are combined with the Hamilton–Jacobi equations giving the proof of Theorem 5.1.1.

Next, in Section 5.5 we briefly discuss operations preserving the convex Poincaré inequality, which may be used to provide new non-trivial examples of measures satisfying it.

Finally, in Section 5.6 we state a few open questions.

5.2 Preliminaries on the convex Poincaré inequality

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 5.2.1. *Suppose that X is a random vector in $\mathbb{R}^n$ satisfying the convex Poincaré inequality* (5.1.1)*. Then for every convex function $f\colon \mathbb{R}^n \to \mathbb{R}$,*

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

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

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

Thus

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

and it is enough to set $Z = f(X)$ and apply (5.1.1). □

Let us start with the already mentioned (see (3.2.4)) 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 [37], but we would like to provide an alternative derivation based on moments (the possibility of such a proof was suggested in [37]). Our strategy mimics a well known approach from the general Lipschitz case (see, e.g., Proposition 2.5 in [56]), 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 5.2.2. *Assume that X is a random vector in $\mathbb{R}^n$, satisfying the convex Poincaré inequality* (5.1.1). *Then for any L-Lipschitz convex function $f\colon \mathbb{R}^n \to \mathbb{R}$ and any $t > 0$,*

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

Proof of Proposition 5.2.2. 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,

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

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 $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex. Then for all $p \geq 2$, applying Lemma 5.2.1 to the convex function $x \mapsto (f(x) - \operatorname{Med} f(X))_+^{p/2}$ (note that its median is zero and $|\nabla(f(x) - \operatorname{Med} f(X))_+| \leq |\nabla f(x)|$), we obtain

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

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

$$\big(\mathbb{E}(f(X) - \operatorname{Med} f(X))_+^p\big)^{1/p} \leq \frac{p}{\sqrt{2\lambda}}\big(\mathbb{E}\,|\nabla f(X)|^p\big)^{1/p}, \qquad (5.2.1)$$

which via Chebyshev's inequality in L_p implies

$$\mathbb{P}\Big(f(X) \geq \operatorname{Med} f(X) + e\frac{p}{\sqrt{2\lambda}}\big(\mathbb{E}\,|\nabla f(X)|^p\big)^{1/p}\Big) \leq e^{2-p} \tag{5.2.2}$$

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 $t > 0$,

$$\mathbb{P}(f(X) \geq \operatorname{Med} f(X) + t) \leq \exp\Big(2 - \frac{\sqrt{2\lambda}}{e}t\Big) \leq 8\exp\Big(-\frac{\sqrt{2\lambda}}{e}t\Big). \qquad \square$$

Remark 5.2.3. 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

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

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 5.2.2); cf. the remark following Theorem 6.8 in [37].

We do not know whether 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, i.e., for $\mathbb{P}(f(X) \leq \operatorname{Med} f(X) - t)$, $t > 0$ (cf. Question 5.6.3 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 5.2.4. *Let X be a random vector in $\mathbb{R}^n$ which satisfies the convex Poincaré inequality* (5.1.1) *and let M be any number such that $\mathbb{P}(|X - \mathbb{E}\,X| \leq M) \geq 3/4$. Then for every convex $f\colon \mathbb{R}^n \to \mathbb{R}$ and for any $t > 32M\,\mathbb{E}\,|\nabla f(X)|$,*

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

Proof. By Proposition 5.2.2 (note that the function $x \mapsto |x - \mathbb{E}\,X|$ is convex and 1-Lipschitz),

$$\mathbb{P}(|X - \mathbb{E}\,X| \geq M + t) \leq 8 \exp\Big(-\frac{\sqrt{2\lambda}}{e} t\Big), \quad t \geq 0. \tag{5.2.3}$$

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function. Without loss of generality we may assume $\operatorname{Med} f(X) = 0$. We have $\mathbb{P}(f(X) \geq 0) \geq 1/2$,

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

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

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

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 $\varepsilon \to 0$ we see that $|u| \leq |\nabla f(x_0)| \leq 8\,\mathbb{E}\,|\nabla f(X)|$, and thus we have

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

If now $t/(16\,\mathbb{E}\,|\nabla f(X)|) \geq 2M$, we can conclude from (5.2.3) that

$$\begin{aligned}
\mathbb{P}(f \leq -t) &\leq \mathbb{P}\big(|X - \mathbb{E}\,X| \geq M + t/(16\,\mathbb{E}\,|\nabla f(X)|)\big) \\
&\leq 8 \exp\Big(-\frac{\sqrt{2\lambda}}{16e\,\mathbb{E}\,|\nabla f(X)|} t\Big),
\end{aligned}$$

which ends the proof. □

5.3 From convex Poincaré to convex and concave modified log-Sobolev inequalities

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 [17] 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.

Recall that for a nonnegative random variable Y we define its entropy as

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

if $\mathbb{E}\, Y \log Y < \infty$ and $\operatorname{Ent} Y = \infty$ otherwise. We refer to, e.g., [9, 51] 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 (5.1.1) and that X is a random vector with law μ, which will not be explicitly stated in all the theorems.

5.3.1 Convex modified log-Sobolev inequalities

Theorem 5.3.1. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex with $|\nabla f(x)| \le c < \sqrt{2\lambda}/e$ for all $x \in \mathbb{R}^n$. Then*

$$\operatorname{Ent}(e^{f(X)}) \le C\, \mathbb{E}\, |\nabla f(X)|^2 e^{f(X)}, \tag{5.3.1}$$

where

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

Our constants are slightly worse than in [17], 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 [17]. The proof relies on two propositions.

Proposition 5.3.2. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex with* $\operatorname{Med} f(X) = 0$ *and $|\nabla f(x)| \le c < \sqrt{2\lambda}/e$ for all $x \in \mathbb{R}^n$. Then*

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

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

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

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

One easily checks that $|\Psi(x)| \le |\Phi(x)|$, $|\Phi'(x)| \le |\Psi'(x)|$, 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 $\operatorname{Med}\Phi(f(X)) = 0$. Hence, by Lemma 5.2.1,

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

Note that $a < \infty$ (by Proposition 5.2.2 and since $c < \sqrt{2\lambda}/e$). Thus $a(\sqrt{\lambda/2} - c/2) \leq b$ and the assertion follows. □

Proposition 5.3.3. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be either convex or concave, with* $\operatorname{Med} f(X) = 0$ *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 $\widetilde{\mathbb{E}}$ 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,

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

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

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

But since by Lemma 5.2.1 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 $\widetilde{\mathbb{E}}|f(X)|$ by $c\sqrt{2/\lambda}$. This yields the assertion of the proposition. □

Proof of Theorem 5.3.1. Without loss of generality assume $\operatorname{Med} f(X) = 0$. Denote $F(t) = \mathbb{E}\, f(X)^2 e^{tf(X)}$, $t \in [0,1]$. By the formula $\int_0^1 ta^2e^{ta}dt = ae^a - e^a + 1$ and the convexity of $t \mapsto F(t)$,

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

(note that for this argument to work we do *not* need the expectation of $f(X)$ to vanish). Thus Propositions 5.3.2 and 5.3.3 imply the assertion of the theorem. □

5.3.2 Concave modified log-Sobolev inequalities

Theorem 5.3.4. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex with $|\nabla f(x)| \le c < \sqrt{2\lambda}/(32e)$ for all $x \in \mathbb{R}^n$. Assume that $M \in \mathbb{R}_+$ satisfies $\mathbb{P}(|X - \mathbb{E}\,X| \le M) \ge 3/4$. Then*

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

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

Remark 5.3.5. If we denote by $X_1, \ldots, X_n$ the coordinates of X, then by the Poincaré inequality we have

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

and hence, by the Chebyshev inequality, $M = 2\sqrt{n/\lambda}$ satisfies $\mathbb{P}(|X - \mathbb{E}\,X| \le M) \ge 3/4$. Thus in fixed dimension n and for say $c = \sqrt{2\lambda}/(64e)$, the constant C in Theorem 5.3.4 can be bounded uniformly over all probability measures satisfying the convex Poincaré inequality with constant λ.

Proof of Theorem 5.3.4. We start as in the proof of Theorem 5.3.1. Denote $g = -f$ (this is a *concave* function). Without loss of generality assume $\operatorname{Med} g(X) = 0$. Denote $F(t) = \mathbb{E}\, g(X)^2e^{tg(X)}$, $t \in [0,1]$. By the convexity of $t \mapsto F(t)$,

$$\begin{aligned}\operatorname{Ent}(e^{g(X)}) &\le \mathbb{E}(g(X)e^{g(X)} - e^{g(X)} + 1) = \mathbb{E}\int_0^1 tg(X)^2e^{tg(X)}dt = \int_0^1 tF(t)dt \\ &\le \int_0^1 t(1-t)F(0) + t^2F(1)dt = \frac{1}{6}F(0) + \frac{1}{3}F(1). \qquad (5.3.2)\end{aligned}$$

We have

$$F(1) \le \mathbb{E}\, g(X)^2 + \mathbb{E}\, g_+(X)^2 e^{g_+(X)} = F(0) + \mathbb{E}\, g_+(X)^2 e^{g_+(X)} \tag{5.3.3}$$

By Proposition 5.3.3,

$$F(0) \le \frac{2}{\lambda} \exp(c\sqrt{2/\lambda})\, \mathbb{E}\, |\nabla g(X)|^2 e^{g(X)}, \tag{5.3.4}$$

so it remains to estimate $\mathbb{E}\, g_+(X)^2 e^{g_+(X)}$.

Integration by parts and Lemma 5.2.4 yield

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

if only $c < \sqrt{2\lambda}/(32e)$. Similarly (using Lemma 5.2.4 in its full strength),

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

for some $D_2 = D_2(\lambda, M)$. Thus, by Proposition 5.3.3, applied to g,

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

This, together with (5.3.2), (5.3.3), and (5.3.4) ends the proof:

$$\begin{aligned}
\mathrm{Ent}(e^{-f(X)}) &\le \frac{1}{6}F(0) + \frac{1}{3}F(1) \le \frac{1}{2}F(0) + \frac{1}{3}\,\mathbb{E}\, g_+(X)^2 e^{g_+(X)} \\
&\le \Big(\frac{1}{\lambda} + \frac{1}{3}\sqrt{D_1 D_2}\Big) e^{c\sqrt{2/\lambda}}\, \mathbb{E}\,|\nabla f(X)|^2 e^{-f(X)}. \qquad \square
\end{aligned}$$

5.4 Proof of the main result

We are ready to present the proof of Theorem 5.1.1. 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 [18] and based on the analysis of the Hamilton–Jacobi equations. The modified log-Sobolev inequalities obtained in Section 5.3 are a crucial element of the proof.

Lemma 5.4.1. *Let X be a random vector in $\mathbb{R}^n$. Assume that there exist $C < \infty$ and $L > 0$ such that*

$$\mathbb{E}\, e^{L|X|} < \infty \tag{5.4.1}$$

and the inequality

$$\mathrm{Ent}(e^{f(X)}) \le C\,\mathbb{E}\,|\nabla f(X)|^2 e^{f(X)} \tag{5.4.2}$$

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

$$\mathbb{E}\, e^{Q_1^\alpha f(X)} e^{-\mathbb{E}\, f(X)} \le 1$$

$$\big(\textit{respectively:} \quad e^{\mathbb{E}\, Q_1^\alpha f(X)}\,\mathbb{E}\, e^{-f(X)} \le 1\big),$$

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 $\alpha\colon \mathbb{R}^n \to \mathbb{R}$ given by the formula

$$\alpha(s) = \begin{cases} \frac{|s|^2}{4C} & \textit{for } |s| \le 2CL, \\ L|s| - L^2 C & \textit{for } |s| > 2CL. \end{cases} \tag{5.4.3}$$

Remark 5.4.2. Similarly as in Chapter 4, the condition (5.4.1) is introduced to exclude heavy-tailed measures for which the only exponentially integrable convex functions are constants, cf. Remark 4.1.5.

If we recall the dual formulations of the weak transport–entropy inequalities $\overline{\mathbf{T}}^-$ and $\overline{\mathbf{T}}^+$ (see Proposition 3.3.4), the definition of $\theta_{C,D}$ from (5.1.4), and the results of the preceding section (namely, Theorems 5.3.1 and 5.3.4), we immediately obtain the following corollaries.

Corollary 5.4.3. *Let X be a random vector in $\mathbb{R}^n$ satisfying the convex Poincaré inequality* (5.1.1)*. Then, for any $c < \sqrt{2\lambda}/e$, the law of X satisfies the inequality $\overline{\mathbf{T}}^-_{\theta_{2C,c}}$ with*

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

Corollary 5.4.4. *Let X be a random vector in $\mathbb{R}^n$ satisfying the convex Poincaré inequality* (5.1.1) *and let M be any number such that $\mathbb{P}(|X - \mathbb{E}\,X| \le M) \ge 3/4$. 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 5.4.1. Suppose that the log-Sobolev inequality (5.4.2) 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 [38, 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,

$$\mathrm{Ent}\, e^{f(G)} \le C\,\mathbb{E}\,|\nabla f(G)|^2 e^{f(G)}$$

for all Lipschitz functions $f\colon \mathbb{R}^n \to \mathbb{R}$ (see, e.g., Theorem 5.1 in [51] for an equivalent formulation).

Then, by the tensorization property of entropy (see, e.g., Proposition 5.6 in [51]), the random vector (X, G) on $\mathbb{R}^n \times \mathbb{R}^n$ satisfies the modified log-Sobolev inequality

$$\mathrm{Ent}(e^{F(X,G)}) \le C\,\mathbb{E}(|\nabla_X F(X,G)|^2 + |\nabla_G F(X,G)|^2)e^{F(X,G)} \tag{5.4.4}$$

for all convex functions $F\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ 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 $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex L-Lipschitz function and consider $\varepsilon > 0$. Applying the inequality (5.4.4) 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

$$\mathrm{Ent}(e^{f(X_\varepsilon)}) \le C_\varepsilon\,\mathbb{E}\,|\nabla f(X_\varepsilon)|^2 e^{f(X_\varepsilon)}, \tag{5.4.5}$$

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, (5.4.5) can be rewritten in the form

$$\mathrm{Ent}(e^{f(X_\varepsilon)}) \le \mathbb{E}\,\alpha^*_\varepsilon(\nabla f(X_\varepsilon))e^{f(X_\varepsilon)}, \tag{5.4.6}$$

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

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

i.e., more explicitly,

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

If $f\colon \mathbb{R}^n \to \mathbb{R}$ 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 $Q_t^{\alpha_\varepsilon} f(x) = \inf_{y\in\mathbb{R}^n}\{f(y)+t\alpha_\varepsilon((x-y)/t)\}$ and the function $x \mapsto t\alpha_\varepsilon((x-y)/t)$ 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.0.1 in Appendix A). Set

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

(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 $t \in (0,1)$,

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

where we used (5.4.6), the definition of α_ε^*, and the fact that $Q_t^{\alpha_\varepsilon} f$ is L-Lipschitz. Thus

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

or, in other words,

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

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

Suppose now that the log-Sobolev inequality (5.4.2) 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 $g\colon \mathbb{R}^n \to \mathbb{R}$ 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

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

or equivalently

$$\mathbb{E}\, e^{Q_1^{\alpha_\varepsilon}(-Q_1^{\alpha_\varepsilon} g)(X_\varepsilon)} \le e^{-\,\mathbb{E}\, Q_1^{\alpha_\varepsilon} g(X_\varepsilon)}. \tag{5.4.7}$$

We stress that now, in order to prove the Hamilton–Jacobi equations via Proposition A.0.1, we need to use the L-Lipschitz property of f, since in general f is not bounded from below.

Since

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

(to verify the inequality take $z = x$), we deduce from (5.4.7) that

$$e^{\mathbb{E}\, Q_1^{\alpha_\varepsilon} g(X_\varepsilon)}\, \mathbb{E}\, e^{-g(X_\varepsilon)} \le 1.$$

A limit argument yields the assertion. □

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

Proof of Theorem 5.1.1. The implication (i) $\Longrightarrow$ (ii) follows immediately from Corollaries 5.4.3 and 5.4.4, 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 $f\colon \mathbb{R}^n \to \mathbb{R}$ 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 $f^x \le f$ on $\mathbb{R}^n$. Taking $z = x + \varepsilon u_x$ with $\varepsilon \to 0$ we see that $|u_x| \le |\nabla f(x)|$.

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

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

(recall that $|u_x| \leq |\nabla f(x)|$). We now substitute εf into the dual formulation (3.3.4) and use the above estimate. An inspection of the Taylor expansions up to order ε^2 yields

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

This ends the proof. □

5.5 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 $\mathbb{E}_\mu$ 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., [9, Proposition 1.4.1]), which asserts that whenever μ_i are probability measures on $\mathcal{X}_i$, $i = 1, \ldots, n$, then the product measure $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ on $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$, satisfies the inequality

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

for every function $f\colon \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$, where $\mathrm{Var}_{\mu_i} f$ 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 $(i = 1, \ldots, N)$ is a probability measure on $\mathbb{R}^{n_i}$, satisfying the convex Poincaré inequality with constant λ, then the product measure $\mu = \mu_1 \otimes \cdots \otimes \mu_N$ on $\mathbb{R}^{n_1+\cdots+n_N}$ satisfies

$$\mathrm{Var}_\mu f \leq \frac{1}{\lambda}\,\mathbb{E}\sum_{i=1}^N |\nabla_i f|^2, \tag{5.5.1}$$

for every convex function $f\colon \mathbb{R}^{n_1+\cdots+n_N} \to \mathbb{R}$, 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 [37], we can obtain the following simple proposition.

Proposition 5.5.1. *Assume that μ_i are probability measures on $\mathbb{R}^{n_i}$, $i = 1, \ldots, N$, satisfying the convex Poincaré inequality with constant λ. Then the measure $\mu = \mu_1 \otimes \cdots \otimes \mu_N$ on $\mathbb{R}^{n_1+\cdots+n_N}$ 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 $n = n_1 + \cdots + n_N$ and consider an arbitrary convex smooth 1-Lipschitz function f on $\mathbb{R}^{nk}$, $k \geq 1$. By (5.5.1) we have $\mathrm{Var}_{\mu^{\otimes k}} f \leq \lambda^{-1}\,\mathbb{E}_{\mu^{\otimes k}}\,|\nabla f|^2 \leq 1/\lambda$. Using an analogous argument as in the proof of Proposition 5.2.2 (for $p > 2$, to remain in the smooth setting) we arrive at

$$\mu^{\otimes k}(f \geq \mathrm{Med}_{\mu^{\otimes k}} f + t) \leq 8e^{-\sqrt{\lambda}t/2} \tag{5.5.2}$$

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, e.g., by convolving them with Gaussian densities, see [67, 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

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

By [37, Theorem 6.7], the dimension-free subexponential concentration for convex sets of the form (5.5.2) implies that μ satisfies the Poincaré inequality

$$\mathrm{Var}_\mu f \leq \frac{1}{\lambda'}\,\mathbb{E}\,|\nabla^- f|^2 \leq \frac{1}{\lambda'}\,\mathbb{E}\,|\nabla f|^2 \tag{5.5.3}$$

for all convex functions f, 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. □

Remark 5.5.2. The above argument shows that if μ satisfies the Poincaré inequality (5.1.1) then it also satisfies the formally stronger inequality (5.5.3) 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. [37, Remark 1.1]).

Tensorization allows in particular to pass from one-dimensional measures satisfying the convex Poincaré inequality (characterized in [19]) 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, e.g., [9, 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 $\operatorname{Var} f = \inf_{a\in\mathbb{R}} \mathbb{E}\,|f-a|^2$ 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. Note that 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 [26]. In particular, the Proposition 5.5.3 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 (5.1.2)), in other words the square of the Kantorovich-Wasserstein distance W_2.

Proposition 5.5.3. *Let* $\{\mu_\vartheta\}_{\vartheta\in\Theta}$ *be a family of probability measures on* $\mathbb{R}^n$ *which satisfy the convex Poincaré inequality* (5.1.1) *with constants* $\{\lambda_\vartheta\}_{\vartheta\in\Theta}$

respectively. Let ν *be a probability measure on* Θ *and assume that for each Borel set* $A \subseteq \mathbb{R}^n$*, the map* $\vartheta \to \mu_\vartheta(A)$ *is measurable. Then the measure* $\mu = \int \mu_\vartheta d\nu(\vartheta)$*, satisfies the convex Poincaré inequality* (5.1.1) *with constant*

$$\lambda = \Big(\sup_{\vartheta\in\Theta}\{1/\lambda_\vartheta\} + 2\,\mathrm{diam}\big(\{\mu_\vartheta\}_{\vartheta\in\Theta}\big)^2\Big)^{-1},$$

where $\mathrm{diam}\big(\{\mu_\vartheta\}_{\vartheta\in\Theta}\big)^2 = \sup_{\vartheta_1,\vartheta_2\in\Theta} \mathcal{T}_2(\mu_{\vartheta_1},\mu_{\vartheta_2})$.

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

$$\begin{aligned}\mathrm{Var}_\mu(f) &= \int \mathrm{Var}_{\mu_\vartheta}(f)d\nu(\vartheta) + \frac{1}{2}\iint (\mathbb{E}_{\mu_{\vartheta_1}} f - \mathbb{E}_{\mu_{\vartheta_2}} f)^2 d\nu(\vartheta_1)d\nu(\vartheta_2)\\ &\le \sup_{\vartheta\in\Theta}\{1/\lambda_\vartheta\}\,\mathbb{E}_\mu\,|\nabla f|^2 + \frac{1}{2}\iint (\mathbb{E}_{\mu_{\vartheta_1}} f - \mathbb{E}_{\mu_{\vartheta_2}} f)^2 d\nu(\vartheta_1)d\nu(\vartheta_2)\end{aligned}$$

and it suffices to estimate the last term.

For fixed $\vartheta_1,\vartheta_2 \in \Theta$ 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,

$$\begin{aligned}\big|\mathbb{E}\,f(X) - \mathbb{E}\,f(Y)\big| &\le \mathbb{E}(|\nabla f(X)| + |\nabla f(Y)|)|X-Y|\\ &\le (\sqrt{\mathbb{E}\,|\nabla f(X)|^2} + \sqrt{\mathbb{E}\,|\nabla f(Y)|^2})\sqrt{\mathbb{E}\,|X-Y|^2}.\end{aligned}$$

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

$$(\mathbb{E}_{\mu_{\vartheta_1}} f - \mathbb{E}_{\mu_{\vartheta_2}} f)^2 \le 2(\mathbb{E}_{\mu_{\vartheta_1}}|\nabla f|^2 + \mathbb{E}_{\mu_{\vartheta_2}}|\nabla f|^2)\,\mathrm{diam}\big(\{\mu_\vartheta\}_{\vartheta\in\Theta}\big)^2.$$

Thus,

$$\begin{aligned}&\frac{1}{2}\iint (\mathbb{E}_{\mu_{\vartheta_1}} f - \mathbb{E}_{\mu_{\vartheta_2}} f)^2 d\nu(\vartheta_1)d\nu(\vartheta_2)\\ &\quad\le \iint \big(\mathbb{E}_{\mu_{\vartheta_1}}|\nabla f|^2 + \mathbb{E}_{\mu_{\vartheta_2}}|\nabla f|^2\big)d\nu(\vartheta_1)d\nu(\vartheta_2)\,\mathrm{diam}\big(\{\mu_\vartheta\}_{\vartheta\in\Theta}\big)^2\\ &\quad= 2\,\mathrm{diam}\big(\{\mu_\vartheta\}_{\vartheta\in\Theta}\big)^2\,\mathbb{E}_\mu\,|\nabla f|^2.\end{aligned}$$

This implies the assertion. □

Example 5.5.4. 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 [19]), 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 ech $x \in \Theta$ let μ_x be a measure on $\mathbb{R}$ such that for all $t > \operatorname{Med}\mu_x$, $\mu_x([t+h, \infty)) \le \frac{1}{2}\mu([t, \infty))$ and a symmetric condition is satisfied by the lower tail of the measure. Assume also that the set $\{\operatorname{Med}\mu_x\}_{x\in\Theta}$ is bounded by some constant l. By the results in [19], the measures μ_x satisfy the convex Poincaré inequality with constant λ_x uniformly bounded from below by some $c_h > 0$. It is easy to see that by Proposition 5.5.3, the measure μ on $\mathbb{R}^{d+1} = \mathbb{R}^d \times \mathbb{R}$ defined as $\mu(dxdt) = \mu_x(dt)\nu(dx)$ satisfies (5.1.1) with $\lambda > 0$, depending only on h, l and the diameter of Θ. Note that if 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.

5.6 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

$$\begin{aligned}&\mu \text{ satisfies the convex Poincaré inequality with constant } \lambda\\ &\implies \mu \text{ satisfies the inequality } \overline{\mathbf{T}}_{\theta_{C,D}} \text{ for some } C, D,\end{aligned}$$

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 5.4.3). This gives rise to our first question.

Question 5.6.1. *Does the 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 5.2.4. Actually, by (ii) of Theorem 6.3.1 below, it would be sufficient to 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 only in the lower-tail (as the upper one is handled by Proposition 5.2.2). More precisely, we have the following result.

Theorem 5.6.2. *Assume that μ is a probability measure on $\mathbb{R}^n$, satisfying the convex Poincaré inequality* (5.1.1) *with constant λ and c is a positive constant, such that for all 1-Lipschitz convex functions $f\colon \mathbb{R}^n \to \mathbb{R}$ and all $t > 0$,*

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

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 5.6.3. *Does the convex Poincaré inequality* (5.1.1) *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* (5.1.1)*, then for every convex 1-Lipschitz function $f\colon \mathbb{R}^n \to \mathbb{R}$,*

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

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

The inequality provided by Lemma 5.2.4 introduces an additional dependence on n, which carries over to the dependence of constants in Theorem 5.1.1. 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 [40, 8, 37]).

Chapter 6

Refined concentration of measure derived from convex infimum convolution inequalities

6.1 Goals and notation

In this chapter we present refined concentration of measure inequalities, which are consequences of weak transportation inequalities. We consider more general cost functions than the one corresponding to, e.g., the convex Poincaré inequality considered in the previous chapter and discuss applications both to the Lipschitz and non-Lipschitz setting. Some of the ideas presented below have been used recently by Adamczak, Kotowski, Polaczyk, and the author in [3] to obtain concentration bounds for quadratic forms in bounded, dependent random variables (under some assumptions, which are fulfilled, e.g., in the Ising model satisfying the Dobrushin condition).

First, we shall explain what concentration inequalities for convex functions can be obtained from general convex infimum convolution inequalities of the form (3.3.4). 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 $\theta\colon \mathbb{R}^n \to [0,\infty)$ 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, $\lim_{x\to\infty} \theta(x) = \infty$).

We remark that at the cost of some technical work one can obtain the results we present below for more general cost functions (say 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 $f\colon \mathbb{R}^n \to \mathbb{R}$, bounded from below, we set

$$f\Box\theta(x) \coloneqq Q_1^\theta f(x) = \inf_{y\in\mathbb{R}^n}\big\{f(y) + \theta(x-y)\big\}. \tag{6.1.1}$$

We also denote

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

6.2 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 [54].

Proposition 6.2.1. *Assume that μ is a probability measure on $\mathbb{R}^n$, satisfying*

$$\int_{\mathbb{R}^n} e^{f\Box\theta}d\mu \int_{\mathbb{R}^n} e^{-f}d\mu \le 1 \tag{6.2.1}$$

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

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

Proof. Consider $f = \infty\mathbf{1}_{(\mathrm{cl}\,A)^c}$ and note that $f\Box\theta(x) < r$ if and only if there exists $y \in A$ such that $\theta(x-y) < r$. Applying the inequality (6.2.1) to f (which can be justified by monotone approximation), we obtain

$$e^r\mu\big((A + B_\theta(r))^c\big)\mu(A) \le \int_{\mathbb{R}^n} e^{f\Box\theta}d\mu \int_{\mathbb{R}^n} e^{-f}d\mu \le 1. \qquad \Box$$

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 6.2.2. 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.,

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

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

$$|x|_{\theta,p} = \sup\Big\{\sum_{i=1}^{n} x_i y_i \colon \theta(y) \le p\Big\}.$$

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

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

It was observed by Gluskin and Kwapień in [32] that norms of this type play an important role in moment estimates for sums of independent random variables. Recently it has been noted [7, 2] 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 $|\cdot|_{\theta_p^*}$ used in [7, 2].

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 $t \ge 1$,

$$|x|_{\theta,tp} \le t|x|_{\theta,p}. \tag{6.2.2}$$

The following corollary to Proposition 6.2.1 is again based on by now standard arguments, written however in the language of the norms $|\cdot|_{\theta,p}$.

Corollary 6.2.3. *Let X be a random vector with law μ, satisfying* (6.2.1) *for all convex functions $f\colon \mathbb{R}^n \to \mathbb{R}$ bounded from below. Then for any smooth convex Lipschitz function $f\colon \mathbb{R}^n \to \mathbb{R}$ and $p \ge 0$,*

$$\mathbb{P}(|f(X) - \operatorname{Med} f(X)| > \sup_{x\in\mathbb{R}^n} |\nabla f(x)|_{\theta,p}) \le 4e^{-p}. \tag{6.2.3}$$

Remark 6.2.4. It is easy to see that if the inequality (6.2.3) 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., [67, p. 429]).

Proof of Corollary 6.2.3. Let $A = \{y \in \mathbb{R}^n \colon f(y) \le \operatorname{Med} f(X)\}$, so that $\mathbb{P}(X \in A) \ge 1/2$. Then by convexity, for any $y \in A$,

$$f(X) \le f(y) + \langle \nabla f(X), X - y\rangle \le \operatorname{Med} f(X) + |\nabla f(X)|_{\theta,p} \cdot |X - y|_{\frac{1}{p}\theta}. \tag{6.2.4}$$

Thus

$$\begin{aligned}\mathbb{P}(f(X) > \operatorname{Med} f(X) + \sup_{x\in\mathbb{R}^n} |\nabla f(x)|_{\theta,p}) &\le \mathbb{P}(\inf_{y\in A} |X - y|_{\frac{1}{p}\theta} > 1)\\ &= \mathbb{P}(X \notin A + \operatorname{cl} B_\theta(p)) \le \frac{e^{-p}}{\mathbb{P}(X \in A)} \le 2e^{-p}, \end{aligned} \tag{6.2.5}$$

where in the second inequality we used Proposition 6.2.1.

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

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

which shows that

$$\mathbb{P}(f(X) < \operatorname{Med} f(X) - \sup_{x\in\mathbb{R}^N} |\nabla f(x)|_{\theta,p}) \le 2e^{-p}.$$

Combining the last inequality with (6.2.5) proves the corollary. □

6.3 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 6.4

We would like to emphasize, that in the theorem we assume only (6.2.3), which is strictly weaker than the infimum-convolution inequality (6.2.1).

Theorem 6.3.1. *Let X be a random vector satisfying* (6.2.3) *for all smooth convex Lipschitz functions $f \colon \mathbb{R}^n \to \mathbb{R}$. Then for any smooth convex function $f \colon \mathbb{R}^n \to \mathbb{R}$, the following properties hold.*

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

$$\Big\| \frac{(f(X) - \operatorname{Med} f(X))_+}{|\nabla f(X)|_{\theta,p}} \Big\|_p \leq 3^{1/p}. \tag{6.3.1}$$

(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*

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

In particular for $p \geq 0$,

$$\mathbb{P}(f(X) < \operatorname{Med} f(X) - 16\,\mathbb{E}\,|\nabla f(X)|_{\theta,p}) \leq 4e^{-p}. \tag{6.3.2}$$

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

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

Remark 6.3.2. As will become clear in the proof, the part (i) of the above theorem holds in fact under one-sided concentration, i.e., it is enough to assume that

$$\mathbb{P}(f(X) - \operatorname{Med} f(X) > \sup_{x \in \mathbb{R}^n} |\nabla f(x)|_{\theta,p}) \leq 4e^{-p}. \tag{6.3.3}$$

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 different choices of the cost function θ.

Example 6.3.3. 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 (6.2.3) is equivalent to

$$\mathbb{P}(|f(X) - \operatorname{Med} f(X)| \geq t) \leq 4\exp(-ct^r) \tag{6.3.4}$$

for all 1-Lipschitz convex functions (in particular for $r = 2$ we get the subgaussian concentration). The first part of Theorem 6.3.1 gives then the following inequality for all (not necessarily Lipschitz) convex functions and $p \geq 1$,

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

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

$$\mathbb{P}\Big(\frac{f(X) - \operatorname{Med} f(X)}{|\nabla f(X)|} \geq t \Big) \leq e \exp\Big(- \frac{ct^r}{(3e)^r} \Big) \tag{6.3.5}$$

(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, e.g., in the theory of empirical processes (see [27]).

The lower tail inequalities give

$$\mathbb{P}(f(X) \leq \operatorname{Med} f(X) - t) \leq 4 \exp\Big(-c \frac{t^r}{16^r (\mathbb{E}\,|\nabla f(X)|)^r} \Big). \tag{6.3.6}$$

Moreover, using the full strength of part (ii) of Theorem 6.3.1, 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 6.3.4. Let us note that inequalities similar to (6.3.6) 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 [67] or [50, Chapter 3.3]). Very recently, Paouris and Valettas [63] have proved that the standard Gaussian vector in $\mathbb{R}^n$ satisfies a similar inequality (for $r = 2$) with $\mathbb{E}\,|f(X) - \operatorname{Med} f(X)|$ 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, i.e., $\mathbb{E}\,|f(X) - \operatorname{Med} f(X)| \leq C\,\mathbb{E}\,|\nabla f(X)|$ (see, e.g., [64, 56]). In fact the proof in [63] 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 (6.3.4) for $r = 1$ is a consequence of the convex Poincaré inequality (however we do not know if (5.1.1) implies (6.3.4) with c depending only on λ and not on the dimension n, see Question 5.6.3 below).

Example 6.3.5. Let us now consider a measure μ on $\mathbb{R}^n$ satisfying the convex Poincaré inequality with constant λ. Then, by Theorem 5.3.1 it satisfies the convex Bobkov–Ledoux inequality (5.3.1) with constants C and c depending only on λ. By the classical Herbst argument it follows (see, e.g., [17, 5]) 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 $f\colon \mathbb{R}^{Nn} \to \mathbb{R}$ and any $t > 0$,

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

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

Moreover, by the Poincaré inequality

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

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 $p > 0$,

$$\mathbb{P}\big(f(X) \ge \operatorname{Med} f(X) + c''(\lambda) \sup_{x\in\mathbb{R}^{Nn}} (\sqrt{p}|\nabla f(x)| + p \max_{i\le N} |\nabla_i f(x)|)\big) \le 2e^{-p}.$$

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

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

(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

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

Thus, the first part of Theorem 6.3.1 together with Remark 6.3.2 gives for arbitrary smooth convex function f on $\mathbb{R}^{Nn}$, the inequality

$$\Big\|\frac{(f(X) - \operatorname{Med} f(X))_+}{\sqrt{p}|\nabla f(X)| + p\max_{i\le N}|\nabla_i f(X)|}\Big\|_p \le c'''(\lambda),$$

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

$$\mathbb{P}\Big(\frac{(f(X) - \operatorname{Med} f(X))_+}{\sqrt{t}|\nabla f(X)| + t\max_{i\le N}|\nabla_i f(X)|} \ge ec'''(\lambda)\Big) \le e^{-t}$$

for $t \ge 1$ (note that contrary to (6.3.5) this time t cannot be removed from the denominator).

As for the lower tail, by Theorem 5.1.1, Remark 5.1.2, Proposition 3.3.4 and tensorization properties of infimum convolution inequalities (see Lemma 5 in [54]) we obtain that X satisfies (6.2.1) and thus also (6.2.3) with $\theta(x) = K(\lambda, n)\sum_{i=1}^N(|x_i|^2\mathbf{1}_{\{|x_i|\le 1\}} + (2|x_i| - 1)\mathbf{1}_{\{|x_i|>1\}})$, where $K(\lambda, n)$ depends only on λ and the dimension n (note that here $K(\lambda, n)$ denotes a constant which is perhaps small). Thus, by the second part of Theorem 6.3.1,

$$\mathbb{P}\big(f(X) \le \operatorname{Med} f(X) - K'(\lambda, n)^{-1}\big[\sqrt{p}\,\mathbb{E}\,|\nabla f(X)| + p\,\mathbb{E}\max_{i\le N}|\nabla_i f(X)|\big]\big) \le 4e^{-p}, \quad (6.3.7)$$

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

$$\begin{aligned}&\mathbb{P}(f(X) \le \mathrm{Med}\, f(X) - t)\\&\qquad \le 4\exp\Big(-K''(\lambda,n)\min\Big\{\frac{t^2}{(\mathbb{E}\,|\nabla f(X)|)^2}, \frac{t}{\mathbb{E}\max_{i\le N}|\nabla_i f(X)|}\Big\}\Big).\end{aligned}$$

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 5.1.3).

Example 6.3.6. 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 (6.2.1) with

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

for $r \ge 1$ (such measures are characterized thanks to results in [38]). If we denote for $x \in \mathbb{R}^n$, $|x|_r = (|x_1|^r + \cdots + |x_n|^r)^{1/r}$ 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 6.3.1, providing concentration inequalities for non-Lipschitz convex functions, in the spirit of recent results due to Bobkov, Nayar, and Tetali [21].

Corollary 6.3.7. *Under the assumptions of Theorem 6.3.1 for all convex functions $f\colon \mathbb{R}^n \to \mathbb{R}$ and $t \ge 0$,*

$$\mathbb{P}(f(X) - \mathrm{Med}\, f(X) \ge t) \le \inf_{p\ge 1}\big\{e^{-p} + \mathbb{P}\big(|\nabla f(X)|_{\theta,p} \ge t/(3e)\big)\big\}.$$

Moreover, for any $p \ge 1$,

$$\mathbb{P}\big(|f(X) - \mathrm{Med}\, f(X)| \ge 3e^2\big\||\nabla f(X)|_{\theta,p}\big\|_p\big) \le 6e^{-p} \tag{6.3.8}$$

Let us note that inequalities of the form (6.3.8) have been obtained in [2] 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 6.3.8. In particular for $\theta(x) = c|x|^2$, the above corollary gives

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

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

$$\mathbb{P}(f(X) - \operatorname{Med} f(X) \geq t) \leq \inf_{L>0}\big\{2e^{-c'\frac{t^2}{L^2}} + \mathbb{P}(|\nabla f(X)| \geq L)\big\}, \qquad (6.3.9)$$

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 6.3.7, on the other hand the above inequality becomes then trivial, as the right-hand side exceeds one).

Recall also the second part of Theorem 6.3.1 which for $q = 3/4$ gives in this case

$$\mathbb{P}(f(X) \leq \operatorname{Med} f(X) - t) \leq 4\exp\big(-c''\frac{t^2}{M_{3/4}^2}\big), \qquad (6.3.10)$$

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

The above inequalities should be compared with a recent result in [21], which asserts that for some constant positive c''' depending only on c,

$$\mathbb{P}(|f(X) - f(Y)| \geq t) \leq 2\inf_{L\geq \operatorname{Med}|\nabla f(X)|}\big\{e^{-c'''\frac{t^2}{L^2}} + \mathbb{P}(|\nabla f(X)| \geq L)\big\}, \qquad (6.3.11)$$

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, i.e., the right-hand sides are small, (6.3.9) gives estimates on the upper tail which (up to numerical constants) are comparable to those implied by (6.3.11), whereas for the lower tail, the inequality (6.3.10) improves over (6.3.11).

Example 6.3.9. 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 6.3.5. From Corollary 6.3.7 we get

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

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

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

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 6.3.10. Let us note that another way of obtaining estimates on the upper tail of non-Lipschitz functions under the convex Poincaré inequality is to use the estimates (5.2.1) and (5.2.2). 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 than 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 6.3.5 arises usually as a consequence of tensorization, so in order to apply it we need some additional structure of the measure.

6.4 Proofs of Theorem 6.3.1, Corollary 6.3.7

Proof of Theorem 6.3.1. Let us start with (i), the proof of which is quite similar to the proof of Corollary 6.2.3. Let us again define $A = \{x \in \mathbb{R}^n \colon f(x) \leq \mathrm{Med}\, f(X)\}$. Using (6.2.2) and (6.2.4), we can write for $t \geq 1$,

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

Hence for $t \geq 1$,

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

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 $\mathrm{Med}\, g(X) = 0$, together with Corollary 6.2.3 and Remark 6.2.4. We can now integrate by parts and get

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

(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 $\operatorname{Med} f(X) = 0$ and $p \geq 1$. Consider the set $B = \{x \in \mathbb{R}^n \colon |\nabla f(x)|_{\theta,p} \leq M_{p,q}\}$. By the definition of $M_{p,q}$, we have $\mathbb{P}(X \in B) \geq q$. Let $\tilde{f} \colon \mathbb{R}^n \to \mathbb{R}$ be defined as

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

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 (6.2.2), 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 $|\cdot|_{\frac{1}{tp}\theta}$ and therefore so is $\tilde{f}$. By Corollary 6.2.3 and Remark 6.2.4 this implies that for any $t \geq 1$,

$$\mathbb{P}(|\tilde{f}(X) - \operatorname{Med} \tilde{f}(X)| > tM_{p,q}) \leq 4e^{-tp}. \tag{6.4.1}$$

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

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

which by another application of (6.4.1) implies

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

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 inequalities (6.2.2) and (6.3.2), we get for $t \geq 16\,\mathbb{E}\,|\nabla f(X)|_{\theta,p}$

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

Now, again by integration by parts,

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

which ends the proof. □

Proof of Corollary 6.3.7. To prove the first inequality it is enough to note that if $|\nabla f(X)|_{\theta,p} \le t/(3e)$ and $f(X) - \operatorname{Med} f(X) \ge t$, then

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

where the last inequality follows from (6.3.1). The assertion follows thus from Chebyshev's inequality: $\mathbb{P}(Z \ge e\|Z\|_p) \le e^{-p}$.

As for the second inequality, we it suffices to apply the first one with $t = 3e^2\| |\nabla f(X)|_{\theta,p}\|_p$ and combine it with the estimate (6.3.2). $\square$

Chapter 7

Convex infimum convolution inequalities with optimal cost functions

7.1 Introduction

In the previous chapters we have seen that the convex infimum convolution inequality (3.3.4) plays a crucial role in theory of weak transport–entropy and convex log-Sobolev inequalities, and also what general concentration properties follow from it. Now we shall take a look at the results from [39] and [38] from a new perspective. To this end, we introduce slightly different notation.

Let X be a random vector with values in $\mathbb{R}^n$ and let $\varphi\colon \mathbb{R}^n \to [0,\infty]$ be a measurable function. Recall that we say that the pair (X,φ) satisfies the *infimum convolution inequality* if for every bounded measurable function $f\colon \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}\, e^{f\Box\varphi(X)}\, \mathbb{E}\, e^{-f(X)} \leq 1, \tag{7.1.1}$$

where $f\Box\varphi$ denotes the infimum convolution of f and φ (see (1.4.1) or (6.1.1) above for the definition). We also say that the pair (X,φ) satisfies the *convex infimum convolution inequality* if (7.1.1) holds for every convex function $f\colon \mathbb{R}^n \to \mathbb{R}$ bounded from below.

Maurey [54] showed that Gaussian and exponential random variables satisfy the infimum convolution inequality with a quadratic and quadratic-linear cost function respectively. Thanks to the tensorization property of the infimum convolution inequality, he recovered—up to constants—the Gaussian concentration inequality as well as the so-called Talagrand two-level concentration inequality for the exponential product measure. Moreover, Maurey proved that bounded random variables satisfy the convex infimum convolution inequality with a quadratic cost function (see equation (3) in [69] for an improvement and consult Lemma 2.3 in [68] for results for non-symmetric Bernoulli measures; see also Lemma 4.2.2 above).

Later on, Maurey's idea was developed further by Latała and Wojtaszczyk who studied comprehensively the infimum convolution inequality in [48]. By testing with linear functions, they observed that the optimal cost function is given by the Legendre transform of the cumulant-generating function (here optimal means largest possible, up to a scaling constant, because the inequality (7.1.1) improves when we increase the cost function). They introduced the notion of optimal infimum convolution inequalities, established them for log-concave product measures and uniform measures on ℓ_p-balls, and put forward important, challenging and far-reaching conjectures (see also [43]).

In this chapter we go along Latała and Wojtaszczyk's line of research and study the optimal convex infimum convolution inequality. Using the characterization of weak transport cost inequalities on the real line (see Proposition 3.3.7), we show that product measures with symmetric marginals having log-concave tails satisfy the optimal convex infimum convolution inequality, which complements Latała and Wojtaszczyk's result about log-concave product measures. This has applications to concentration and moment comparison of any norm of such vectors in the spirit of celebrated Paouris' inequality (see [62] and [4]) and addresses some questions posed lately in [46]. We also offer an example showing that the assumption of log-concave tails cannot be weakened substantially.

7.2 Main results

For a random vector X in $\mathbb{R}^n$ we define

$$\Lambda_X^*(x) := \mathcal{L}\Lambda_X(x) := \sup_{y\in\mathbb{R}^n}\{\langle x,y\rangle - \ln\mathbb{E}\, e^{\langle y,X\rangle}\},$$

which is the Legendre transform of the cumulant-generating function

$$\Lambda_X(x) := \ln\mathbb{E}\, e^{\langle x,X\rangle}, \qquad x\in\mathbb{R}^n.$$

If X is symmetric and the pair (X,φ) satisfies the infimum convolution inequality, then $\varphi(x)\le\Lambda_X^*(x)$ for every $x\in\mathbb{R}^n$ (see Remark 2.12 in [48]). In other words, Λ_X^* is the optimal cost function φ for which the infimum convolution inequality can hold. Since this conclusion is obtained by testing (7.1.1) with linear functions, the same holds for the convex infimum convolution inequality. Following [48] we shall say that X satisfies (convex) IC(β) if the pair $(X, \Lambda_X^*(\cdot/\beta))$ satisfies the (convex) infimum convolution inequality.

We are ready to present the first result of this chapter.

Theorem 7.2.1. *Let X be a symmetric random variable with log-concave tails, i.e., such that the function*

$$t \mapsto N(t) := -\ln \mathbb{P}(|X| \geq t), \quad t \geq 0,$$

is convex. Then there exists a universal constant $\beta \leq 1680e$ such that X satisfies the convex IC(β).

The (convex) infimum convolution inequality tensorizes and, consequently, the property (convex) IC tensorizes: if independent random vectors X_i satisfy (convex) IC(β_i), $i = 1, \ldots, n$, then the vector $(X_1, \ldots, X_n)$ satisfies (convex) IC($\max \beta_i$) (see Lemma 5 in [54] and Proposition 2.14 in [48]). Therefore we have the following corollary.

Corollary 7.2.2. *Let X be a symmetric random vector with values in $\mathbb{R}^n$ and independent coordinates with log-concave tails. Then X satisfies the convex IC(β) with a universal constant $\beta \leq 1680e$.*

Note that the class of distributions from Corollary 7.2.2 is wider than the class of symmetric log-concave product distributions considered by Latała and Wojtaszczyk in [48] (since, by the Prékopa–Leindler inequality, log-concave measures have log-concave tails). Among others, it contains measures which do not have a connected support, e.g., a symmetric Bernoulli random variable.

In order to comment on the relevance of the assumptions of Theorem 7.2.1 and present applications to comparison of weak and strong moments, we need the following definition. Let X be a random vector with values in $\mathbb{R}^n$. We say that the moments of X *grow α-regularly* if for every $p \geq q \geq 2$ and every $\theta \in S^{n-1}$ we have

$$\|\langle X, \theta\rangle\|_p \leq \alpha \frac{p}{q} \|\langle X, \theta\rangle\|_q$$

Clearly, if the moments of X grow α-regularly, then α has to be at least 1 (unless $X = 0$ a.s.).

Remark 7.2.3. If X is a symmetric random variable with log-concave tails, then its moments grow 1-regularly (this classical fact follows for instance from Proposition 5.5 from [42] and the proof of Proposition 3.8 from [48]).

The assumption of log-concave tails in Theorem 7.2.1 cannot be replaced by a weaker one of α-regularity of moments: if X is a symmetric random variable defined by

$$\mathbb{P}(|X| > t) = 1_{[0,2)}(t) + \sum_{k=1}^{\infty} e^{-2^k} 1_{[2^k, 2^{k+1})}(t), \quad t \geq 0, \tag{7.2.1}$$

then the moments of X grow α-regularly (for some $\alpha < \infty$), but—as we shall prove in Section 7.5 below—there does not exist $C > 0$ such that the pair $(X, x \mapsto \max\{(Cx)^2, C|x|\})$ satisfies the convex infimum convolution inequality. All the more, X cannot satisfy convex IC(β) with any $\beta < \infty$. Thus it seems that the assumptions of Theorem 7.2.1 are not far from necessary conditions for the convex infimum convolution inequality to hold with an optimal cost function (random variables with moments growing regularly are akin to random variables with log-concave tails as the former can essentially be sandwiched between the latter, see (4.6) in [47]).

Our second main result is an application of Theorem 7.2.1 to moment comparison. Recall that for a random vector X its p-th weak moment associated with a norm $\|\cdot\|$ on $\mathbb{R}^n$ is the quantity defined as

$$\sigma_{\|\cdot\|,X}(p) := \sup_{\|t\|_* \leq 1} \|\langle t, X\rangle\|_p,$$

where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$. The following version of Proposition 3.15 from [48] holds (some non-trivial modifications of the proof are necessary in order to deal with the fact that the inequality (7.1.1) only holds for convex functions).

Theorem 7.2.4. *Let X be a symmetric random vector with values in $\mathbb{R}^n$ the moments of which grow α-regularly. Suppose moreover that X satisfies the convex IC(β). Then for every norm $\|\cdot\|$ on $\mathbb{R}^n$ and every $p \geq 2$ we have*

$$\Big(\mathbb{E}\big|\|X\| - E\|X\|\big|^p\Big)^{1/p} \leq C\alpha\beta\sigma_{\|\cdot\|,X}(p),$$

where C is a universal constant (one can take $C = 4\sqrt{2}e < 16$).

Immediately, in view of Corollary 7.2.2 and Remark 7.2.3, we obtain the following corollary in the spirit of the results from [62, 4, 46, 45]. Similar inequalities for Rademacher sums with the emphasis on exact values of constants have also been studied by Oleszkiewicz (see Theorem 2.1 in [59]).

Corollary 7.2.5. *Let X be a symmetric random vector with values in $\mathbb{R}^n$ and with independent coordinates which have log-concave tails. Then for every norm $\|\cdot\|$ on $\mathbb{R}^n$ and every $p \geq 2$ we have*

$$\big(\mathbb{E}\,\|X\|^p\big)^{1/p} \leq \mathbb{E}\,\|X\| + D\sigma_{\|\cdot\|,X}(p), \tag{7.2.2}$$

where D is a universal constant (one can take $D = 6720\sqrt{2}e^2 < 70223$).

Note that each of the terms on the right-hand side of (7.2.2) is, up to a constant, dominated by the left-hand side of (7.2.2), so (7.2.2) yields the comparison of weak and strong moments of the norms of X.

Note also that the constant standing at $\mathbb{E}\,\|X\|$ is equal to 1. If we only assume that the coordinates of X are independent and their moments grow α-regularly, then (7.2.2) does not always hold (the example here is a vector with independent coordinates distributed like in (7.2.1); see Section 7.5 for details), although by Theorem 1.1 from [46] it holds if we allow the constant in front of $\mathbb{E}\,\|X\|$ to be greater than 1 and to depend on α. Hence Corollary 7.2.5 and example (7.2.1) partially answer the following question raised by Latała and Strzelecka in [46]: "For which vectors does the comparison of weak and strong moments hold with constant 1 at the first strong moment?"

The organization of the chapter is the following. In Section 7.3 and 7.4 we present the proofs of Theorem 7.2.1 and Theorem 7.2.4 respectively. In Section 7.5 we discuss example (7.2.1) in details.

7.3 Proof of Theorem 7.2.1

Our approach is based on a characterization provided in [38] of measures on the real line which satisfy a weak transport–entropy inequality (see Proposition 3.3.7 above). We emphasize that our optimal cost functions need not be quadratic near the origin, therefore we cannot apply the result as is, but have to first fine-tune the cost functions a bit. We shall also need the following simple lemma.

Lemma 7.3.1. *If X is a symmetric random variable and $\mathbb{E}\,X^2 = \beta_1^{-2}$, then*

$$\Lambda_X^*(x/\beta_1) \le x^2 \quad \text{for } |x| \le 1.$$

Proof. Since X is symmetric, we have

$$\mathbb{E}\,e^{tX} = 1 + \sum_{k=1}^{\infty} \frac{\|X\|_{2k}^{2k} t^{2k}}{(2k)!} \ge 1 + \sum_{k=1}^{\infty} \frac{\|X\|_2^{2k} t^{2k}}{(2k)!} = 1 + \sum_{k=1}^{\infty} \frac{\beta_1^{-2k} t^{2k}}{(2k)!} = \cosh(\beta_1^{-1}|t|).$$

Moreover, $\mathcal{L}\big(\ln\cosh(\cdot)\big)(|u|) \le |u|^2$ for $|u| \le 1$ (see for example the proof of Proposition 3.3 in [48]). Therefore

$$\Lambda_X^*(x/\beta_1) = \mathcal{L}(\Lambda_X(\beta_1 \cdot))(x) \le \mathcal{L}(\ln\cosh(\cdot))(x) \le x^2 \qquad \text{for } |x| \le 1. \quad \square$$

Throughout the proof g^{-1} stands for the generalized inverse of a function g defined as

$$g^{-1}(y) := \inf\{x : g(x) \geq y\}.$$

Proof of Theorem 7.2.1. Note that $N(0) = 0$ and the function N is non-decreasing. First we tweak the assumptions and change the assertion to a more straightforward one.

Step 1 (first reduction). We claim that it suffices to prove the assertion for random variables for which the function N is strictly increasing on the set where it is finite (or, in other words, $N(t) = 0$ only for $t = 0$). Indeed, suppose we have done this and let now X be any random variable satisfying the assumptions of the theorem. Let X_ε be a symmetric random variable such that $\mathbb{P}(|X_\varepsilon| \geq t) = \exp(-N_\varepsilon(t))$, where $N_\varepsilon(t) = N(t) \vee \varepsilon t$ (note that this function is convex). Consider versions of X and X_ε on the probability space $(0, 1)$ (equipped with Lebesgue measure) constructed as the (generalized) inverses of their cumulative distribution functions. Then $|X_\varepsilon| \leq |X|$ almost surely. Hence $\Lambda_{X_\varepsilon} \leq \Lambda_X$ and therefore also $\Lambda^*_{X_\varepsilon} \geq \Lambda^*_X$.

The theorem applied to the random variable X_ε and the above inequality imply that the pair $(X_\varepsilon, \Lambda^*_X(\cdot/\beta))$ satisfies the convex infimum convolution inequality. Since by construction $X_\varepsilon \to X$ a.s. for $\varepsilon \to 0^+$, we get the assertion for X (in the second integral we just use the fact that the test function f is bounded from below and thus e^{-f} is bounded from above; for the first integral it suffices to prove the convergence of integrals on any interval $[-M, M]$, and on such an interval we have $f \square \Lambda^*_X(x/\beta) \leq f(x) + \Lambda^*_X(0) = f(x)$, and thus $\exp(\max_{[-M,M]} f)$ is a good majorant).

Step 2 (second reduction). We claim that it suffices to prove the assertion for random variables such that $\Lambda_X < \infty$. Indeed, suppose we have done this and let X be any random variable satisfying the assumptions of the theorem. Let $N_\varepsilon(t) = N(t) \vee \varepsilon^2 t^2$ and let X_ε be a symmetric random variable such that $\mathbb{P}(|X_\varepsilon| \geq t) = \exp(-N_\varepsilon(t))$. Then, similarly as in Step 1., $\Lambda_{X_\varepsilon} \leq \Lambda_Y < \infty$, where Y is symmetric and $\mathbb{P}(|Y| \geq t) = \exp(-\varepsilon^2 t^2)$. Thus we can apply the proposition to X_ε and we continue as in Step 1.

Step 3 (scaling). Due to the scaling properties of the Legendre transform, we can fix the value of $\mathbb{E}\, X^2 = \beta_1^{-2}$. We choose $\beta_1 := 2e$ (the case where $X \equiv 0$ is trivial), so that, by Markov's inequality, $e^{-N(1/2)} = \mathbb{P}(|X| \geq \frac{1}{2}) \leq 4\,\mathbb{E}\, X^2 = e^{-2}$ and equivalently

$$N(1/2) \geq 2. \tag{7.3.1}$$

Step 4 (reformulation). For $x \in \mathbb{R}$ let

$$\varphi(x) := \big(x^2 1_{\{|x|<1\}} + (2|x| - 1)1_{\{|x|\geq 1\}}\big) \vee \Lambda^*_X(x/(2\beta_1)).$$

We claim that there exists a universal constant $\widetilde{b} \leq 1/420$, such that the pair $(X, \varphi(\widetilde{b}\,\cdot))$ satisfies the convex infimum convolution inequality. Of course the assertion of the theorem follows immediately from that.

Note that φ is convex, increasing on $[0,\infty)$ (because it is convex and $\varphi(x) = 0$ only for $x = 0$). Crucially, we have $\varphi(x) = x^2$ for $x \in [0,1]$ (by Lemma 7.3.1), so the cost function φ is quadratic near zero. Moreover, by Lemma 7.3.1, $\varphi^{-1}(3) = 2$.

Let $U = F^{-1} \circ F_\nu$, where F, F_ν are the distribution functions of X and the symmetric exponential measure ν on $\mathbb{R}$, respectively. By Propositions 3.3.4 and 3.3.7 we know that if there exists $b > 0$ such that for every $x, y \in \mathbb{R}$ we have

$$\big|U(x) - U(y)\big| \leq \frac{1}{b}\varphi^{-1}\big(1 + |x-y|\big), \tag{7.3.2}$$

then the pair $(X, \varphi(\widetilde{b}\,\cdot))$, where $\widetilde{b} = \frac{b}{210\varphi^{-1}(2+1^2)} = \frac{b}{420}$, satisfies the convex infimum convolution inequality. We will show that (7.3.2) holds with $b = 1$.

Step 5 (further reformulation). Let $a = \inf\{t > 0 : N(t) = \infty\}$. We have three possibilities (recall that N is left-continuous):

- $a = \infty$. Then N is continuous, increasing, and transforms $[0,\infty]$ onto $[0,\infty]$. Also, F is increasing and therefore F^{-1} is the usual inverse of F.
- $a < \infty$ and $N(a) < \infty$. Then X has an atom at a. Moreover, $N(a) = \lim_{t\to a^-} N(t)$.
- $a < \infty$ and $N(a) = \infty = \lim_{t\to a^-} N(t)$.

Of course, in the first case one can extend N by putting $N(a) = \infty$, so that all formulas below make sense.

Note that

$$F(t) = \begin{cases} \frac{1}{2}\exp(-N(|t|)) & \text{if } t < 0, \\ 1 - \frac{1}{2}\exp(-N_+(t)) & \text{if } t \geq 0, \end{cases}$$

where $N_+(t)$ denotes the right-sided limit of N at t (which is different from $N(t)$ only if $t = a$ and X has an atom at a). Hence, F is continuous on the interval $(-a, a)$, the image of $(-a,a)$ under F is the interval $\big(\frac{1}{2}\exp(-N(a)), 1 - \frac{1}{2}\exp(-N(a))\big)$, and we have $F(-a) = \frac{1}{2}\exp(-N(a))$ and $F(a) = 1$. Since the image of $\mathbb{R}$ under U is equal to the image of $(0,1)$ under F^{-1}, we conclude that $U(\mathbb{R}) = (-a,a)$ if $N(a) = \infty$ and $U(\mathbb{R}) = [-a,a]$ if $N(a) < \infty$. Denote $A := U(\mathbb{R})$.

When $N(a) < \infty$, it suffices to check the condition (7.3.2) for $x, y \in [-a,a]$ (otherwise one can change x, y and decrease the right-hand side while not changing the value of the left-hand side of (7.3.2)). For $x \in [-a,a]$ we can write $U^{-1}(x) = N(|x|)\operatorname{sgn}(x)$ and $U^{-1}(x) \in \mathbb{R}$. When $N(a) = \infty$, U

is a bijection (on its image), so we can obviously write again $U^{-1}(x) = N(|x|)\,\mathrm{sgn}(x)$ for any $x \in \mathbb{R}$.

Therefore, in order to verify (7.3.2) with $b = 1$ we need to check that

$$|x - y| \le \varphi^{-1}\big(1 + \big|N(|x|)\,\mathrm{sgn}(x) - N(|y|)\,\mathrm{sgn}(y)\big|\big) \qquad \text{for } x, y \in A. \quad (7.3.3)$$

Since we consider the case when $\Lambda_X(t)$ is finite for every $t \in \mathbb{R}$, the Chernoff inequality applies, so for $t \ge \mathbb{E}\,X = 0$ we have

$$\frac{1}{2}e^{-N(t)} = \mathbb{P}(X \ge t) \le e^{-\Lambda_X^*(t)},$$

so

$$N(t) \ge \Lambda_X^*(t) - \ln 2. \quad (7.3.4)$$

Note that $\varphi(|x-y|) < \infty$ for $x, y \in A$, since $\varphi(|x-y|) = \infty$ would imply $\Lambda_X^*(|x-y|/(2\beta_1)) = \infty$, and hence $\Lambda_X^*(|x-y|/2) = \infty$, and – by (7.3.4) – also $N(|x-y|/2) = \infty$, but for $x, y \in A$ we have $|x-y|/2 \in [0, a)$ when $N(a) = \infty$ or $|x-y|/2 \in [0, a]$ when $N(a) < \infty$ and in either case $N(|x-y|/2)$ is finite. Therefore for every $x, y \in A$ we have $\varphi(|x-y|) < \infty$. Since $\varphi^{-1}(\varphi(z)) = z$ for z such that $\varphi(z) < \infty$ (because φ is then continuous and increasing on $[0, z]$), the condition (7.3.3) is implied by

$$\varphi\big(|x-y|\big) \le 1 + \big|N(|x|)\,\mathrm{sgn}\,x - N(|y|)\,\mathrm{sgn}\,y\big| \qquad \text{for } x, y \in A. \quad (7.3.5)$$

In the next step we check that this is indeed satisfied.

Step 6 (checking the condition). Let $x_0 = \inf\{x \ge 1 : 2x - 1 = \Lambda_X^*(\frac{x}{2\beta_1})\}$ (if $x_0 = \infty$ we simply do not have to consider Case 2 below). We consider three cases. We repeatedly use the fact that $uN(t) \ge N(ut)$ for $u \le 1$, $t \ge 0$, which follows by the convexity of N and the property $N(0) = 0$.

Case 1. $|x-y| \le 1$. Then $\varphi\big(|x-y|\big) = (x-y)^2 \le 1$, so (7.3.5) is trivially satisfied.

Case 2. $|x-y| \ge x_0$. Then $\varphi\big(|x-y|\big) = \Lambda_X^*(\frac{1}{2\beta_1}|x-y|) \le \Lambda_X^*(|x-y|/2)$. Inequality (7.3.4) implies that in order to prove (7.3.5) it suffices to show that if x, y are of the same sign, say $x, y \ge 0$, then $N\big(|x-y|/2\big) \le |N(x) - N(y)|$ and if x, y have different signs, we have $N\big((|x| + |y|)/2\big) \le N(|x|) + N(|y|)$.

By the convexity of N, for $s, t \ge 0$ we have

$$N(s/2) + N(t) \le N(s) + N(t) \le \frac{s}{s+t}N(s+t) + \frac{t}{s+t}N(s+t) = N(s+t)$$

and

$$N\big((s+t)/2\big) \le \frac{1}{2}N(s) + \frac{1}{2}N(t) \le N(s) + N(t),$$

so N satisfies the aforementioned two inequalities. This finishes the proof of (7.3.5) in Case 2.

Case 3. $1 \le |x-y| \le x_0$. Then $\varphi\big(|x-y|\big) = 2|x-y|-1$. Consider two sub-cases:

(i) x, y have different signs. Without loss of generality we may assume $x \ge |y| \ge 0 \ge y$. Thus in order to obtain (7.3.5) it suffices to show that $N(x) \ge 2x + 2|y|$. Note that $1 \le x + |y| \le 2x$, so $x \ge \frac{1}{2}$. Thus

$$N(x) \ge N(1/2)2x \overset{(7.3.1)}{\ge} 4x \ge 2x + 2|y|,$$

which finishes the proof in case (i).

(ii) x, y have the same sign. Without loss of generality we may assume $x \ge y \ge 0$. Thus it suffices to show that $2(x-y) \le N(x) - N(y)$. Note that due to the assumption of Case 3 we have $x \ge x - y \ge 1 \ge \frac{1}{2}$, so by the convexity of N we have

$$\frac{N(x)-N(y)}{x-y} \ge \frac{N(\frac{1}{2}) - N(0)}{\frac{1}{2} - 0} \overset{(7.3.1)}{\ge} 4 \ge 2$$

This ends the examination of case (ii) and the proof of the theorem. □

7.4 Comparison of weak and strong moments

The goal of this section is to establish the comparison of weak and strong moments with respect to any norm $\|\cdot\|$ for random vectors X with independent coordinates having log-concave tails (Corollary 7.2.5). In view of Theorem 7.2.1 and Remark 7.2.3, it is enough to show Theorem 7.2.4.

Our proof of Theorem 7.2.4 comprises three steps: first we exploit α-regularity of moments of X to control the size of its cumulant-generating function Λ_X, second we bound from below the infimum convolution of the optimal cost function with the convex test function being the norm $\|\cdot\|$ properly rescaled, and finally by the property convex IC(β) we obtain exponential tail bounds which integrated out give the desired moment inequality.

We start with two lemmas corresponding to the first two steps described above and then we put everything together.

Lemma 7.4.1. *Let $p \ge 2$ and suppose that the moments of a random vector X in $\mathbb{R}^n$ grow α-regularly. If for a vector $u \in \mathbb{R}^n$ we have $\|\langle u, X\rangle\|_p \le 1$, then*

$$\Lambda_X((2e\alpha)^{-1}pu) \le p.$$

Proof. Let k_0 be the smallest integer larger than p. If $\alpha e\|\langle u, X\rangle\|_p \le 1/2$, then by α-regularity we have

$$
\begin{aligned}
\Lambda_X(pu) &\le \ln\Big(\sum_{k\ge 0} \frac{\mathbb{E}\,|\langle pu, X\rangle|^k}{k!}\Big) \\
&\le \ln\Big(\sum_{0\le k\le p} p^k \frac{\|\langle u, X\rangle\|_p^k}{k!} + \sum_{k>p} (\alpha k)^k \frac{\|\langle u, X\rangle\|_p^k}{k!}\Big) \\
&\le \ln\Big(\sum_{0\le k\le p} \frac{p^k\|\langle u, X\rangle\|_p^k}{k!} + \sum_{k>p} \big(\alpha e\|\langle u, X\rangle\|_p\big)^k\Big) \\
&\le \ln\Big(\sum_{0\le k\le p} \frac{p^k\|\langle u, X\rangle\|_p^k}{k!} + 2(\alpha e\|\langle u, X\rangle\|_p)^{k_0}\Big) \\
&\le \ln\Big(\sum_{0\le k\le p} \frac{p^k\|\langle u, X\rangle\|_p^k}{k!} + \frac{(2\alpha e p\|\langle u, X\rangle\|_p)^{k_0}}{k_0!}\Big) \\
&\le \ln\Big(\sum_{0\le k\le k_0} \frac{(2\alpha e p\|\langle u, X\rangle\|_p)^k}{k!}\Big) \le 2\alpha e p\|\langle u, X\rangle\|_p \le p.
\end{aligned}
$$

Replace u with $(2e\alpha)^{-1}u$ to get the assertion. □

Lemma 7.4.2. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ and let X be a random vector with values in $\mathbb{R}^n$ and moments growing α-regularly. For $\beta > 0$, $p \ge 2$, and $x \in \mathbb{R}^n$ we have*

$$\big(\Lambda_X^*(\cdot/\beta)\,\Box a\|\cdot\|\big)(x) \ge a\|x\| - p,$$

where $a = p(2e\alpha\beta\sigma_{\|\cdot\|,X}(p))^{-1}$.

Proof. For $f(x) = a\|x\|$ with positive a being arbitrary for now we bound the infimum convolution as follows

$$
\begin{aligned}
\big(\Lambda_X^*(\cdot/\beta)\Box f\big)(x) &= \inf_y \sup_z \big\{\beta^{-1}\langle y, z\rangle - \Lambda_X(z) + a\|x-y\|\big\} \\
&= \inf_y \sup_u \big\{(2e\alpha\beta)^{-1}p\langle y, u\rangle - \Lambda_X((2e\alpha)^{-1}pu) + a\|x-y\|\big\} \\
&\ge \inf_y \sup_{u:\|\langle u,X\rangle\|_p\le 1} \big\{(2e\alpha\beta)^{-1}p\langle y, u\rangle - p + a\|x-y\|\big\},
\end{aligned}
$$

where in the last inequality we have used Lemma 7.4.1. Let us choose $u = \sigma_{\|\cdot\|,X}(p)^{-1}v$ with $\|v\|_* \le 1$ such that $\langle y, v\rangle = \|y\|$. Then clearly $\|\langle u, X\rangle\|_p \le 1$ and thus

$$\Lambda_X^*(\cdot/\beta)\Box f(x) \ge \inf_y \big\{(2e\alpha\beta\sigma_{\|\cdot\|,X}(p))^{-1}p\|y\| - p + a\|x-y\|\big\}.$$

If we now set $a = p(2e\alpha\beta\sigma_{\|\cdot\|,X}(p))^{-1}$, then by the triangle inequality we obtain the desired lower bound

$$\left(\Lambda_X^*(\cdot/\beta)\,\Box a\|\cdot\|\right)(x) \geq a\|x\| - p. \qquad \Box$$

Proof of Theorem 7.2.4. Let $f(x) = a\|x\|$ with $a = p(2e\alpha\beta\sigma_{\|\cdot\|,X}(p))^{-1}$ as in Lemma 7.4.2. Testing the property convex IC(β) with f and applying Lemma 7.4.2 yields

$$\mathbb{E}\, e^{a\|X\|}\, \mathbb{E}\, e^{-a\|X\|} \leq e^p.$$

By Jensen's inequality, both $\mathbb{E}\, e^{a(\|X\| - \mathbb{E}\,\|X\|)}$ and $\mathbb{E}\, e^{a(-\|X\| + \mathbb{E}\,\|X\|)}$ are bounded above by e^p. Thus Markov's inequality implies the tail bound

$$\mathbb{P}\left(a\big|\|X\| - \mathbb{E}\,\|X\|\big| > t\right) \leq 2e^{-t}e^p \leq 2e^{-t/2}, \quad t \geq 2p.$$

Consequently,

$$\begin{aligned} a^p\, \mathbb{E}\big|\|X\| - \mathbb{E}\,\|X\|\big|^p &= \int_0^\infty pt^{p-1}\mathbb{P}\left(a\big|\|X\| - \mathbb{E}\,\|X\|\big| > t\right) dt \\ &\leq (2p)^p + 2\int_0^\infty pt^{p-1}e^{-t/2} dt = (2p)^p + 2\cdot 2^p p\Gamma(p) \\ &\leq 2(2p)^p. \end{aligned}$$

Plugging in the value of a gives the result (we can take $C = 4\sqrt{2}e < 16$). $\Box$

7.5 An example

Let X be a symmetric random variable defined by $\mathbb{P}(|X| > t) = T(t)$, where

$$T(t) := 1_{[0,2)}(t) + \sum_{k=1}^\infty e^{-2^k} 1_{[2^k, 2^{k+1})}(t), \quad t \geq 0, \tag{7.5.1}$$

or, in other words, let $|X|$ have the distribution

$$(1 - e^{-2})\delta_2 + \sum_{k=2}^\infty \left(e^{-2^{k-1}} - e^{-2^k}\right)\delta_{2^k}.$$

Let us first show that the moments of X grow 3-regularly, but X does not satisfy IC(β) for any $\beta < \infty$ (we also prove a slightly stronger statement later).

Let Y be a symmetric exponential random variable. Then Y has log-concave tails, so the moments of Y grow 1-regularly (see Remark 7.2.3).

Moreover, if X and Y are constructed in the standard way by the inverses of their cumulative distribution functions on the probability space $(0,1)$, then

$$|Y| \le |X| \le 2|Y| + 2.$$

Therefore, for $p \ge q \ge 2$,

$$\|X\|_p \le 2\|Y\|_p + 2 \le 2\,\frac{p}{q}\|Y\|_q + 2 \le 3\,\frac{p}{q}\|X\|_q$$

(we used the fact that $|X| \ge 2$ in the last inequality). Thus the moments of X grow 3-regularly.

On the other hand, for every $h > 0$ there exists $t > 0$ such that

$$\mathbb{P}(|X| \ge t + h) = \mathbb{P}(|X| \ge t).$$

Therefore by Theorem 1 of [30] there does not exist a constant C such that the pair $(X, \varphi(\cdot/C))$, where $\varphi(x) = \frac{1}{2}x^2 1_{\{|x|\le 1\}} + (|x| - 1/2)1_{\{|x|>1\}}$, satisfies the convex infimum convolution inequality. But, by symmetry and the 3-regularity of moments of X,

$$\begin{aligned}\Lambda_X(s) &\le \ln\Big(1 + \sum_{k\ge1} \frac{s^{2k}\,\mathbb{E}\,X^{2k}}{(2k)!}\Big) \le \ln\Big(1 + \sum_{k\ge1} \frac{s^{2k}(3k)^{2k}\big(\mathbb{E}\,X^2\big)^k}{(2k)!}\Big)\\ &\le \ln\Big(1 + \sum_{k\ge1} s^{2k}(3e/2)^{2k}\big(\mathbb{E}\,X^2\big)^k\Big) = \ln\Big(1 + \sum_{k\ge1}\big(9e^2s^2\,\mathbb{E}\,X^2/4\big)^k\Big).\end{aligned}$$

Hence, using the inequality $\ln(1+x) \le x$ and summing the geometric series, we arrive at

$$\Lambda_X(s) \le \frac{9e^2\,\mathbb{E}\,X^2}{4 - 9e^2s^2\,\mathbb{E}\,X^2}\cdot s^2,$$

provided that $|s| \le \varepsilon$ for some small enough ε. Thus for some $A, \varepsilon > 0$ we have $\Lambda_X(s) \le As^2$ for $|s| \le \varepsilon$. Having chosen ε and possibly increasing A, we can also guarantee that $2A\varepsilon^2 \ge 1$. Hence

$$\begin{aligned}\Lambda_X^*(t) &\ge \sup_{|s|\le\varepsilon}\{st - As^2\} = \tfrac{1}{4A}t^2 1_{\{|t|\le 2A\varepsilon\}} + (\varepsilon|t| - A\varepsilon^2)1_{\{|t|>2A\varepsilon\}}\\ &= 2A\varepsilon^2\varphi\big(t/(2A\varepsilon)\big) \ge \varphi\big(t/(2A\varepsilon)\big).\end{aligned}$$

We conclude that X cannot satisfy IC(β) for any β.

Remark 7.5.1. Let us also sketch an alternative approach. Take $a, c > 0$, $b \in \mathbb{R}$, and denote $\varphi(x) = \min\{x^2, |x|\}$, $f(x) = f_{a,b}(x) = a(x-b)_+$ for $x \in \mathbb{R}$. One can check that

$$\big(f\Box\varphi(c\cdot)\big)(x) = \begin{cases} 0 & \text{if } x \le b,\\ c^2(x-b)^2 & \text{if } b < x \le b + 1/c,\\ c(x-b) & \text{if } x > b + 1/c,\end{cases}$$

if $a > 2c$. It is rather elementary but cumbersome to show that for any $c > 0$ there exist $a > 0$ and $b \in \mathbb{R}$ such that (7.1.1) is violated by the test function f. We omit the details.

In fact, the above example shows that even a slightly stronger statement is true: for vectors with independent coordinates with α-regular growth of moments the comparison of weak and strong moments of norms does not need to hold with the constant 1 at the first strong moment. More precisely, let $X_1, X_2, \ldots$ be independent random variables with distribution given by (7.5.1). We claim that there does *not* exist any $K < \infty$ such that

$$\big(\mathbb{E}\max_{i\le n}|X_i|^p\big)^{1/p} \le \mathbb{E}\max_{i\le n}|X_i| + K\sup_{\|t\|_1\le 1}\Big(\mathbb{E}\Big|\sum_{i=1}^n t_iX_i\Big|^p\Big)^{1/p} \tag{7.5.2}$$

holds for every $p \ge 2$ and every positive integer n (note that we chose the ℓ^∞-norm as our norm). We shall estimate the three expressions appearing in (7.5.2).

We have

$$\sup_{\|t\|_1\le 1}\Big(\mathbb{E}\Big|\sum_{i=1}^n t_iX_i\Big|^p\Big)^{1/p} \le \sup_{\|t\|_1\le 1}\sum_{i=1}^n |t_i|\|X_i\|_p = \|X_1\|_p \tag{7.5.3}$$

(this inequality is in fact an equality). Since the moments of X_1 grow 3-regularly, the last term in (7.5.2) is bounded by $\widetilde{K}p$ for some $\widetilde{K} < \infty$.

To estimate the remaining two terms we need the following standard fact.

Lemma 7.5.2. *For independent events $A_1, \ldots, A_n$,*

$$(1-e^{-1})\Big(1 \wedge \sum_{i=1}^n \mathbb{P}(A_i)\Big) \le \mathbb{P}\Big(\bigcup_{i=1}^n A_i\Big) \le 1 \wedge \sum_{i=1}^n \mathbb{P}(A_i).$$

In particular, for i.i.d. non-negative random variables $Y_1, \ldots, Y_n$,

$$(1-e^{-1})\int_0^\infty \Big[1 \wedge n\mathbb{P}(Y_1 > t)\Big]dt \le \mathbb{E}\max_{i\le n} Y_i \le \int_0^\infty \Big[1 \wedge n\mathbb{P}(Y_1 > t)\Big]dt.$$

Proof. The upper bound is just the union bound. The lower bound follows from de Morgan's laws combined with independence, which imply that

$$\mathbb{P}\Big(\bigcup_{i=1}^n A_i\Big) = 1 - \prod_{i=1}^n\big(1-\mathbb{P}(A_i)\big),$$

and from the inequalities $1 - x \le e^{-x}$ and $1 - e^{-y} \ge (1-e^{-1})y$ for $x \in \mathbb{R}$, $y \in [0, 1]$. □

Fix $m \geq 2$ and let $e^{2^{m-1}} \leq n < e^{2^m}$. Then

$$1 \wedge nT(t) = \begin{cases} 1 & \text{if } 0 < t < 2^m, \\ nT(t) & \text{if } t \geq 2^m. \end{cases}$$

By the above lemma,

$$\begin{aligned} \mathbb{E} \max_{i \leq n} |X_i| &\leq \int_0^{2^m} dt + n \int_{2^m}^{\infty} T(t)dt = 2^m + n \sum_{j=m}^{\infty} e^{-2^j}(2^{j+1} - 2^j) \\ &= 2^m + n \sum_{j=m}^{\infty} e^{-2^j} 2^j \leq 2^m + ne^{-2^m} 2^m \sum_{j=0}^{\infty} (2e^{-2^m})^j \\ &= 2^m + \frac{ne^{-2^m} 2^m}{1 - 2e^{-2^m}}. \end{aligned}$$

Set $\theta = \theta(m, n) = ne^{-2^m} \in [e^{-2^{m-1}}, 1)$. Then

$$\mathbb{E} \max_{i \leq n} |X_i| \leq 2^m \Big(1 + \frac{\theta}{1 - 2e^{-2^m}}\Big). \tag{7.5.4}$$

Similarly,

$$\begin{aligned} \mathbb{E} \max_{i \leq n} |X_i|^p &\geq (1 - e^{-1}) \int_0^{\infty} 1 \wedge T(t^{1/p})dt \\ &= (1 - e^{-1}) \Big[\int_0^{2^{mp}} dt + n \int_{2^{mp}}^{\infty} T(t^{1/p})dt \Big] \\ &= (1 - e^{-1}) \Big[2^{mp} + n \sum_{j=m}^{\infty} e^{-2^j} \big(2^{(j+1)p} - 2^{jp}\big) \Big]. \end{aligned}$$

Hence

$$\mathbb{E} \max_{i \leq n} |X_i|^p > (1 - e^{-1}) ne^{-2^m} \big(2^{(m+1)p} - 2^{mp}\big) = (1 - e^{-1}) \theta 2^{mp} (2^p - 1). \tag{7.5.5}$$

Putting (7.5.3), (7.5.4), and (7.5.5) together, we see that (7.5.2) would imply

$$(1 - e^{-1})^{1/p} \theta^{1/p} 2^m (2^p - 1)^{1/p} \leq 2^m \Big(1 + \frac{\theta}{1 - 2e^{-2^m}}\Big) + \widetilde{K} p$$

for every $p \geq 2$, $m \geq 2$, and $\theta \in [e^{-2^{m-1}}, 1)$ of the form ne^{-2^m}, $n \in \mathbb{N}$. Take $p = 1/\theta$ and $\theta \sim 1/m$ to get

$$(1 - e^{-1})^{\theta} \theta^{\theta} (2^{1/\theta} - 1)^{\theta} \leq 1 + \frac{\theta}{1 - 2e^{-2^m}} + \frac{\widetilde{K}}{2^m \theta}.$$

Since $\theta \to 0$ and $2^m \theta \to \infty$ as $m \to \infty$ this inequality yields $2 \leq 1$, which is a contradiction. Hence inequality (7.5.2) cannot hold for all $p \geq 2$ and all positive integer n.

Appendix A

Facts related to Hamilton–Jacobi equations used in Chapter 5

Below we 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 (5.4.3), which have been exploited in the proof of Lemma 5.4.1. 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 Chapter 3 of [29] can be modified to yield the properties we need. Alternatively, as in [38], one could rely on modification of the results from [36], where the theory of Hamilton–Jacobi equations is extended to the setting of metric spaces.

Proposition A.0.1. *Let C, L be positive constants and let α be defined by* (5.4.3)*. Assume that $f\colon \mathbb{R}^n \to \mathbb{R}$ is either bounded from below or L-Lipschitz and let $u\colon (0,\infty)\times\mathbb{R}^n \to \mathbb{R}$ 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 $\alpha^*\colon \mathbb{R}^n \to \mathbb{R}$ is the Legendre transform of α, given explicitly by the formula

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

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 [29, Chapter 3.3.2, Lemma 1], however in our setting one needs to work with infima rather then minima.

Ad (b). For fixed t, u is L-Lipschitz as the function of 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 [29, 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 [29, 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 $z = x + 2CLt(y-x)/|x-y|$,

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

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. $\square$

Bibliography

[1] R. Adamczak, *Logarithmic Sobolev inequalities and concentration of measure for convex functions and polynomial chaoses*, Bull. Pol. Acad. Sci. Math. **53** (2005), no. 2, 221–238. 29, 37, 40

[2] R. Adamczak, W. Bednorz, and P. Wolff, *Moment estimates implied by modified log-Sobolev inequalities*, ESAIM Probab. Stat. **21** (2017), 467–494. 79, 84

[3] R. Adamczak, M. Kotowski, B. Polaczyk, and M. Strzelecki, *A note on concentration for polynomials in the Ising model*, preprint (2018+), `arXiv:1809.03187`. 77

[4] R. Adamczak, R. Latała, A. E. Litvak, K. Oleszkiewicz, A. Pajor, and N. Tomczak-Jaegermann, *A short proof of Paouris' inequality*, Canad. Math. Bull. **57** (2014), no. 1, 3–8. 90, 92

[5] R. Adamczak and M. Strzelecki, *Modified log-Sobolev inequalities for convex functions on the real line. Sufficient conditions*, Studia Math. **230** (2015), no. 1, 59–93. 21, 30, 37, 38, 40, 42, 43, 82

[6] ———, *On the convex Poincaré inequality and weak transportation inequalities*, to appear in Bernoulli, preprint (2017+), `arXiv:1703.01765v2`. 7

[7] R. Adamczak and P. Wolff, *Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order*, Probab. Theory Related Fields **162** (2015), no. 3-4, 531–586. 79

[8] S. Aida and D. Stroock, *Moment estimates derived from Poincaré and logarithmic Sobolev inequalities*, Math. Res. Lett. **1** (1994), no. 1, 75–86. 76

[9] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer, *Sur les inégalités de Sobolev logarithmiques*,

Panoramas et Synthèses, vol. 10, Société Mathématique de France, Paris, 2000, With a preface by Dominique Bakry and Michel Ledoux. 37, 63, 71, 73

[10] S. Artstein-Avidan, A. Giannopoulos, and V. D. Milman, *Asymptotic geometric analysis. Part I*, Mathematical Surveys and Monographs, vol. 202, American Mathematical Society, Providence, RI, 2015. 2

[11] D. Bakry, I. Gentil, and M. Ledoux, *Analysis and geometry of Markov diffusion operators*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 348, Springer, Cham, 2014. 37

[12] F. Barthe, P. Cattiaux, and C. Roberto, *Interpolated inequalities between exponential and Gaussian, Orlicz hypercontractivity and isoperimetry*, Rev. Mat. Iberoam. **22** (2006), no. 3, 993–1067. 9, 15, 16, 17

[13] F. Barthe and A. V. Kolesnikov, *Mass transport and variants of the logarithmic Sobolev inequality*, J. Geom. Anal. **18** (2008), no. 4, 921–979. 18, 19

[14] F. Barthe and C. Roberto, *Sobolev inequalities for probability measures on the real line*, Studia Math. **159** (2003), no. 3, 481–497, Dedicated to Professor Aleksander Pełczyński on the occasion of his 70th birthday (Polish). 11, 12, 26

[15] ———, *Modified logarithmic Sobolev inequalities on* $\mathbb{R}$, Potential Anal. **29** (2008), no. 2, 167–193. 10, 11, 12, 40

[16] W. Beckner, *A generalized Poincaré inequality for Gaussian measures*, Proc. Amer. Math. Soc. **105** (1989), no. 2, 397–400. 9

[17] S. Bobkov and M. Ledoux, *Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution*, Probab. Theory Related Fields **107** (1997), no. 3, 383–400. 10, 23, 58, 62, 63, 82

[18] S.G. Bobkov, I. Gentil, and M. Ledoux, *Hypercontractivity of Hamilton-Jacobi equations*, J. Math. Pures Appl. (9) **80** (2001), no. 7, 669–696. 5, 57, 67

[19] S.G. Bobkov and F. Götze, *Discrete isoperimetric and Poincaré-type inequalities*, Probab. Theory Related Fields **114** (1999), no. 2, 245–277. 50, 51, 58, 73, 74, 75

[20] ______, *Exponential integrability and transportation cost related to logarithmic Sobolev inequalities*, J. Funct. Anal. **163** (1999), no. 1, 1–28. 21, 26, 33

[21] S.G. Bobkov, P. Nayar, and P. Tetali, *Concentration properties of restricted measures with applications to non-Lipschitz functions*, Geometric aspects of functional analysis, Lecture Notes in Math., vol. 2169, Springer, Cham, 2017, pp. 25–53. 84, 85

[22] C. Borell, *The Brunn-Minkowski inequality in Gauss space*, Invent. Math. **30** (1975), no. 2, 207–216. 2

[23] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities*, Oxford University Press, Oxford, 2013, A nonasymptotic theory of independence, With a foreword by Michel Ledoux. 3

[24] P. Cattiaux and A. Guillin, *On quadratic transportation cost inequalities*, J. Math. Pures Appl. (9) **86** (2006), no. 4, 341–361. 5, 26

[25] P. Cattiaux, A Guillin, and L.-M. Wu, *Some remarks on weighted logarithmic Sobolev inequality*, Indiana Univ. Math. J. **60** (2011), no. 6, 1885–1904. 25

[26] D. Chafaï and F. Malrieu, *On fine properties of mixtures with respect to concentration of measure and Sobolev type inequalities*, Ann. Inst. Henri Poincaré Probab. Stat. **46** (2010), no. 1, 72–96. 73

[27] V. H. de la Peña, M. J. Klass, and T. L. Lai, *Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws*, Ann. Probab. **32** (2004), no. 3A, 1902–1933. 82

[28] P. Diaconis and L. Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, Ann. Appl. Probab. **6** (1996), no. 3, 695–750. 37

[29] L. C. Evans, *Partial differential equations*, second ed., Graduate Studies in Mathematics, vol. 19, American Mathematical Society, Providence, RI, 2010. 33, 47, 103, 104

[30] N. Feldheim, A. Marsiglietti, P. Nayar, and J. Wang, *A note on the convex infimum convolution inequality*, Bernoulli **24** (2018), no. 1, 257–270. 38, 41, 50, 58, 100

[31] I. Gentil, A. Guillin, and L. Miclo, *Modified logarithmic Sobolev inequalities and transportation inequalities*, Probab. Theory Related Fields **133** (2005), no. 3, 409–436. 8, 9, 25

[32] E. D. Gluskin and S. Kwapień, *Tail and moment estimates for sums of independent random variables with logarithmically concave tails*, Studia Math. **114** (1995), no. 3, 303–309. 79

[33] N. Gozlan, *A characterization of dimension free concentration in terms of transportation inequalities*, Ann. Probab. **37** (2009), no. 6, 2480–2498. 5

[34] ______, *Poincaré inequalities and dimension free concentration of measure*, Ann. Inst. Henri Poincaré Probab. Stat. **46** (2010), no. 3, 708–739. 14

[35] ______, *Transport-entropy inequalities on the line*, Electron. J. Probab. **17** (2012), no. 49, 1–18. 5, 21, 48, 55

[36] N. Gozlan, C. Roberto, and P.-M. Samson, *Hamilton Jacobi equations on metric spaces and transport entropy inequalities*, Rev. Mat. Iberoam. **30** (2014), no. 1, 133–163. 103

[37] ______, *From dimension free concentration to the Poincaré inequality*, Calc. Var. Partial Differential Equations **52** (2015), no. 3-4, 899–925. 5, 6, 15, 30, 57, 58, 60, 61, 72, 73, 76

[38] N. Gozlan, C. Roberto, P.-M. Samson, Y. Shu, and P. Tetali, *Characterization of a class of weak transport-entropy inequalities on the line*, Ann. Inst. Henri Poincaré Probab. Stat. **54** (2018), no. 3, 1667–1693. 30, 31, 32, 34, 35, 37, 38, 40, 48, 50, 53, 55, 58, 68, 84, 89, 93, 103

[39] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali, *Kantorovich duality for general transport costs and applications*, J. Funct. Anal. **273** (2017), no. 11, 3327–3405. 30, 31, 32, 33, 35, 41, 42, 43, 44, 54, 55, 89

[40] M. Gromov and V. D. Milman, *A topological application of the isoperimetric inequality*, Amer. J. Math. **105** (1983), no. 4, 843–854. 76

[41] L. Gross, *Logarithmic Sobolev inequalities*, Amer. J. Math. **97** (1975), no. 4, 1061–1083. 3, 36

[42] O. Guédon, P. Nayar, and T. Tkocz, *Concentration inequalities and geometry of convex bodies*, Analytical and probabilistic methods in the geometry of convex bodies, IMPAN Lect. Notes, vol. 2, Polish Acad. Sci. Inst. Math., Warsaw, 2014, pp. 9–86. 91

[43] R. Latała, *On some problems concerning log-concave random vectors*, to appear in IMA Volume "Discrete Structures: Analysis and Applications", Springer. 90

[44] R. Latała and K. Oleszkiewicz, *Between Sobolev and Poincaré*, Geometric aspects of functional analysis, Lecture Notes in Math., vol. 1745, Springer, Berlin, 2000, pp. 147–168. 8, 9

[45] R. Latała and M. Strzelecka, *Weak and strong moments of ℓ_r-norms of log-concave vectors*, Proc. Amer. Math. Soc. **144** (2016), no. 8, 3597–3608. 92

[46] ———, *Comparison of weak and strong moments for vectors with independent coordinates*, Mathematika **64** (2018), no. 1, 211–229. 90, 92, 93

[47] R. Latała and T. Tkocz, *A note on suprema of canonical processes based on random variables with regular moments*, Electron. J. Probab. **20** (2015), no. 36, 17. 92

[48] R. Latała and J. O. Wojtaszczyk, *On the infimum convolution inequality*, Studia Math. **189** (2008), no. 2, 147–187. 90, 91, 92, 93

[49] M. Ledoux, *On Talagrand's deviation inequalities for product measures*, ESAIM Probab. Statist. **1** (1995/97), 63–87 (electronic). 29, 37, 41

[50] ———, *Concentration of measure and logarithmic Sobolev inequalities*, Séminaire de Probabilités, XXXIII, Lecture Notes in Math., vol. 1709, Springer, Berlin, 1999, pp. 120–216. 82

[51] ———, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. 3, 21, 29, 63, 68

[52] M. Ledoux and M. Talagrand, *Probability in Banach spaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 23, Springer-Verlag, Berlin, 1991, Isoperimetry and processes. 2

[53] P. Lévy, *Problèmes concrets d'analyse fonctionnelle. Avec un complément sur les fonctionnelles analytiques par F. Pellegrino*, Gauthier-Villars, Paris, 1951, 2d ed. 1

[54] B. Maurey, *Some deviation inequalities*, Geom. Funct. Anal. **1** (1991), no. 2, 188–197. 5, 29, 33, 44, 54, 78, 83, 89, 91

[55] L. Miclo, *Quand est-ce que des bornes de Hardy permettent de calculer une constante de Poincaré exacte sur la droite?*, Ann. Fac. Sci. Toulouse Math. (6) **17** (2008), no. 1, 121–192. 21

[56] E. Milman, *On the role of convexity in isoperimetry, spectral gap and concentration*, Invent. Math. **177** (2009), no. 1, 1–43. 60, 82

[57] V. D. Milman, *A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies*, Funkcional. Anal. i Priložen. **5** (1971), no. 4, 28–37. 2

[58] V. D. Milman and G. Schechtman, *Asymptotic theory of finite-dimensional normed spaces*, Lecture Notes in Mathematics, vol. 1200, Springer-Verlag, Berlin, 1986, With an appendix by M. Gromov. 2

[59] K. Oleszkiewicz, *Precise moment and tail bounds for Rademacher sums in terms of weak parameters*, Israel J. Math. **203** (2014), no. 1, 429–443. 92

[60] F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), no. 2, 361–400. 5

[61] ———, *Comment on: "Hypercontractivity of Hamilton-Jacobi equations" [J. Math. Pures Appl. (9)* **80** *(2001), no. 7, 669–696; MR1846020 (2003b:47073)] by S. G. Bobkov, I. Gentil and M. Ledoux*, J. Math. Pures Appl. (9) **80** (2001), no. 7, 697–700. 5

[62] G. Paouris, *Concentration of mass on convex bodies*, Geom. Funct. Anal. **16** (2006), no. 5, 1021–1049. 90, 92

[63] G. Paouris and P. Valettas, *A Gaussian small deviation inequality for convex functions*, Ann. Probab. **46** (2018), no. 3, 1441–1454. 82

[64] G. Pisier, *Probabilistic methods in the geometry of Banach spaces*, Probability and analysis (Varenna, 1985), Lecture Notes in Math., vol. 1206, Springer, Berlin, 1986, pp. 167–241. 82

[65] R. T. Rockafellar, *Convex analysis*, Princeton Mathematical Series, No. 28, Princeton University Press, Princeton, N.J., 1970. 21

[66] O. S. Rothaus, *Analytic inequalities, isoperimetric inequalities and logarithmic Sobolev inequalities*, J. Funct. Anal. **64** (1985), no. 2, 296–313. 18

[67] P.-M. Samson, *Concentration of measure inequalities for Markov chains and Φ-mixing processes*, Ann. Probab. **28** (2000), no. 1, 416–461. 29, 72, 79, 82

[68] ———, *Concentration inequalities for convex functions on product spaces*, Stochastic inequalities and applications, Progr. Probab., vol. 56, Birkhäuser, Basel, 2003, pp. 33–52. 29, 90

[69] ———, *Infimum-convolution description of concentration properties of product probability measures, with applications*, Ann. Inst. H. Poincaré Probab. Statist. **43** (2007), no. 3, 321–338. 90

[70] E. Schmidt, *Der Brunn-Minkowskische Satze und sein Spiegeltheorem sowie die isoperimetrische Eigenschaft der Kugel in der euklidischen und hyperbolischen Geometrie*, Math. Ann. **120** (1948), 307–422. 1

[71] ———, *Die Brunn-Minkowskische Ungleichung und ihr Spiegelbild sowie die isoperimetrische Eigenschaft der Kugel in der euklidischen und nichteuklidischen Geometrie. I*, Math. Nachr. **1** (1948), 81–157. 1

[72] ———, *Die Brunn-Minkowskische Ungleichung und ihr Spiegelbild sowie die isoperimetrische Eigenschaft der Kugel in der euklidischen und nichteuklidischen Geometrie. II*, Math. Nachr. **2** (1949), 171–244. 1

[73] J. Shao, *Modified logarithmic Sobolev inequalities and transportation cost inequalities in $\mathbb{R}^n$*, Potential Anal. **31** (2009), no. 2, 183–202. 28

[74] Y. Shu and M. Strzelecki, *A characterization of a class of convex log-Sobolev inequalities on the real line*, Ann. Inst. Henri Poincaré Probab. Stat. **54** (2018), no. 4, 2075–2091. 7

[75] M. Strzelecka, M. Strzelecki, and T. Tkocz, *On the convex infimum convolution inequality with optimal cost function*, ALEA Lat. Am. J. Probab. Math. Stat. **14** (2017), no. 2, 903–915. 7

[76] V. N. Sudakov and B. S. Cirel′ son, *Extremal properties of half-spaces for spherically invariant measures*, Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **41** (1974), 14–24, 165, Problems in the theory of probability distributions, II. 2

[77] M. Talagrand, *A new isoperimetric inequality and the concentration of measure phenomenon*, Geometric aspects of functional analysis (1989–90), Lecture Notes in Math., vol. 1469, Springer, Berlin, 1991, pp. 94–124. 5, 6, 29

[78] ———, *Concentration of measure and isoperimetric inequalities in product spaces*, Inst. Hautes Études Sci. Publ. Math. (1995), no. 81, 73–205. 6, 29

[79] ———, *Transportation cost for Gaussian and other product measures*, Geom. Funct. Anal. **6** (1996), no. 3, 587–600. 3

[80] C. Villani, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009, Old and new. 4

[81] F.-Y. Wang, *Functional inequalities for empty essential spectrum*, J. Funct. Anal. **170** (2000), no. 1, 219–245. 16

[82] ———, *A generalization of Poincaré and log-Sobolev inequalities*, Potential Anal. **22** (2005), no. 1, 1–15. 9, 15, 16

[83] ———, *From super Poincaré to weighted log-Sobolev and entropy-cost inequalities*, J. Math. Pures Appl. (9) **90** (2008), no. 3, 270–285. 28