

Concentration Properties of Restricted Measures with Applications to Non-Lipschitz Functions

S. G. Bobkov*, P. Nayar[†] and P. Tetali[‡]

April 24, 2016

2010 Mathematics Subject Classification. Primary 60Gxx

Keywords and phrases. Subgaussian constant, spread constant, restricted measures, concentration of measure.

Abstract

For any metric probability space (M, d, μ) , with a subgaussian constant $\sigma^2(\mu)$ and any Borel measurable set $A \subset M$, we show that $\sigma^2(\mu_A) \leq c \log(e/\mu(A)) \sigma^2(\mu)$, where μ_A is a normalized restriction of μ to the set A and c is a universal constant. As a consequence, we deduce concentration inequalities for non-Lipschitz functions.

1 Introduction

It is known that many high-dimensional probability distributions μ on the Euclidean space \mathbb{R}^n (and other metric spaces, including graphs) possess strong concentration properties. In a functional language, this may informally be stated as the assertion that any sufficiently smooth function f on \mathbb{R}^n , e.g., having a bounded Lipschitz semi-norm, is almost a constant on almost all space. There are several ways to quantify such a property. One natural approach proposed by N. Alon, R. Boppana and J. Spencer [A-B-S] associates, with a given metric probability space (M, d, μ) , its *spread constant*,

$$s^2(\mu) = \sup \text{Var}_\mu(f) = \sup \int (f - m)^2 d\mu,$$

where $m = \int f d\mu$, and the sup is taken over all functions f on M with $\|f\|_{\text{Lip}} \leq 1$. More information is contained in the so-called subgaussian constant $\sigma^2 = \sigma^2(\mu)$ which is defined as the infimum over all σ^2 such that

$$\int e^{tf} d\mu \leq e^{\sigma^2 t^2/2}, \quad \text{for all } t \in \mathbb{R}, \quad (1.1)$$

in the class \mathcal{L}_0 of all f on M with $m = 0$ and $\|f\|_{\text{Lip}} \leq 1$ (cf. [B-G-H]). Describing the diameter of \mathcal{L}_0 in the Orlicz space $L^{\psi_2}(\mu)$ for the Young function $\psi_2(t) = e^{t^2} - 1$ (within

*School of Mathematics, University of Minnesota, Minneapolis, MN 55455; research supported in part by the Humboldt Foundation, NSF and BSF grants

[†]Institute of Mathematics & Applications, Minneapolis, MN 55455; E-mail address: nayar@mimuw.edu.pl (corresponding author); research supported in part by NCN grant DEC-2012/05/B/ST1/00412

[‡]School of Mathematics and School of Computer Science, Georgia Institute of Technology, Atlanta, GA 30332; research supported in part by NSF DMS-1407657.

universal factors), the quantity $\sigma^2(\mu)$ appears as a parameter in a subgaussian concentration inequality for the class of all Borel subsets of M . As an equivalent approach, it may also be introduced via the transport-entropy inequality connecting the classical Kantorovich distance and the relative entropy from an arbitrary probability measure on M to the measure μ (cf. [B-G]).

While in general $s^2 \leq \sigma^2$, the latter characteristic allows one to control subgaussian tails under the probability measure μ uniformly in the entire class of Lipschitz functions on M . More generally, when $\|f\|_{\text{Lip}} \leq L$, (1.1) yields

$$\mu\{|f - m| \geq t\} \leq 2e^{-t^2/(2\sigma^2 L^2)}, \quad t > 0. \quad (1.2)$$

Classical and well-known examples include the standard Gaussian measure on $M = \mathbb{R}^n$ in which case $s^2 = \sigma^2 = 1$, and the normalized Lebesgue measure on the unit sphere $M = S^{n-1}$ with $s^2 = \sigma^2 = \frac{1}{n-1}$. The last example was a starting point in the study of the concentration of measure phenomena, a fruitful direction initiated in the early 1970s by V. D. Milman.

Other examples come often after verification that μ satisfies certain Sobolev-type inequalities such as Poincaré-type inequalities

$$\lambda_1 \text{Var}_\mu(u) \leq \int |\nabla u|^2 d\mu,$$

and logarithmic Sobolev inequalities

$$\rho \text{Ent}_\mu(u^2) = \rho \left[\int u^2 \log u^2 d\mu - \int u^2 d\mu \log \int u^2 d\mu \right] \leq 2 \int |\nabla u|^2 d\mu,$$

where u may be any locally Lipschitz function on M , and the constants $\lambda_1 > 0$ and $\rho > 0$ do not depend on u . Here the modulus of the gradient may be understood in the generalized sense as the function

$$|\nabla u(x)| = \limsup_{y \rightarrow x} \frac{|u(x) - u(y)|}{d(x, y)}, \quad x \in M$$

(this is the so-called ‘‘continuous setting’’), while in the discrete spaces, e.g., graphs, we deal with other naturally defined gradients. In both cases, one has respectively the well-known upper bounds

$$s^2(\mu) \leq \frac{1}{\lambda_1}, \quad \sigma^2(\mu) \leq \frac{1}{\rho}. \quad (1.3)$$

For example, $\lambda_1 = \rho = n - 1$ on the unit sphere (best possible values, [M-W]), which can be used to make a corresponding statement about the spread and Gaussian constants.

One of the purposes of this note is to give new examples by involving the family of normalized restricted measures

$$\mu_A(B) = \frac{\mu(A \cap B)}{\mu(A)}, \quad B \subset M \text{ (Borel)},$$

where a Borel measurable set $A \subset M$ is fixed and has a positive measure. As an example, returning to the standard Gaussian measure μ on \mathbb{R}^n , it is known that $\sigma^2(\mu_A) \leq 1$ for any convex body $A \subset \mathbb{R}^n$. This remarkable property, discovered by D. Bakry and M. Ledoux [B-L] in a sharper form of a Gaussian-type isoperimetric inequality, has nowadays several proofs and generalizations, cf. [B1, B2]. Of course, in general, the set A may have a rather disordered structure, for instance, to be disconnected. And then there is no hope for validity of a Poincaré-type inequality for the measure μ_A . Nevertheless, it turns out that the concentration property of μ_A is inherited from μ , unless the measure of A is too small. In particular, we have the following observation about abstract metric probability spaces.

Theorem 1.1. *For any measurable set $A \subset M$ with $\mu(A) > 0$, the subgaussian constant $\sigma^2(\mu_A)$ of the normalized restricted measure satisfies*

$$\sigma^2(\mu_A) \leq c \log \left(\frac{e}{\mu(A)} \right) \sigma^2(\mu), \quad (1.4)$$

where c is an absolute constant.

Although this assertion is technically simple, we will describe two approaches: one is direct and refers to estimates on the ψ_2 -norms over the restricted measures, and the other one uses a general comparison result due to F. Barthe and E. Milman on the concentration functions ([B-M]).

One may further generalize Theorem 1.1 by defining the subgaussian constant $\sigma_{\mathcal{F}}^2(\mu)$ within a given fixed subclass \mathcal{F} of functions on M , by using the same bound (1.1) on the Laplace transform. This is motivated by a possible different level of concentration for different classes; indeed, in case of $M = \mathbb{R}^n$, the concentration property may considerably be strengthened for the class \mathcal{F} of all convex Lipschitz functions. In particular, one result of M. Talagrand [T1, T2] provides a dimension-free bound $\sigma_{\mathcal{F}}^2(\mu) \leq C$ for an arbitrary product probability measure μ on the n -dimensional cube $[-1, 1]^n$. Hence, a more general version of Theorem 1.1 yields the bound

$$\sigma_{\mathcal{F}}^2(\mu_A) \leq c \log \left(\frac{e}{\mu(A)} \right)$$

with some absolute constant c , which holds for any Borel subset A of $[-1, 1]^n$ (cf. Section 6 below).

According to the very definition, the quantities $\sigma^2(\mu)$ and $\sigma^2(\mu_A)$ might seem responsible for deviations of only Lipschitz functions f on M and A , respectively. However, the inequality (1.4) may also be used to control deviations of non-Lipschitz f – on large parts of the space and under certain regularity hypotheses. Assume, for example, $\int |\nabla f| d\mu \leq 1$ (which is kind of a normalization condition) and consider

$$A = \{x \in M : |\nabla f(x)| \leq L\}. \quad (1.5)$$

If $L \geq 2$, this set has the measure $\mu(A) \geq 1 - \frac{1}{L} \geq \frac{1}{2}$, and hence, $\sigma^2(\mu_A) \leq c\sigma^2(\mu)$ with some absolute constant c . If we assume that f has a Lipschitz semi-norm $\leq L$ on A , then, according to (1.2),

$$\mu_A\{x \in A : |f - m| \geq t\} \leq 2e^{-t^2/(c\sigma^2(\mu)L^2)}, \quad t > 0, \quad (1.6)$$

where m is the mean of f with respect to μ_A . It is in this sense one may say that f is almost a constant on the set A .

This also yields a corresponding deviation bound on the whole space,

$$\mu\{x \in M : |f - m| \geq t\} \leq 2e^{-t^2/c\sigma^2(\mu)L^2} + \frac{1}{L}.$$

Stronger integrability conditions posed on $|\nabla f|$ can considerably sharpen the conclusion. By a similar argument, Theorem 1.1 yields, for example, the following exponential bound, known in the presence of a logarithmic Sobolev inequality for the space (M, d, μ) , and with σ^2 replaced by $1/\rho$ (cf. [B-G]).

Corollary 1.2. *Let f be a locally Lipschitz function on M with Lipschitz semi-norms $\leq L$ on the sets (1.5). If $\int e^{|\nabla f|^2} d\mu \leq 2$, then f is μ -integrable, and moreover,*

$$\mu\{x \in M : |f - m| \geq t\} \leq 2e^{-t/c\sigma(\mu)}, \quad t > 0,$$

where m is the μ -mean of f and c is an absolute constant.

Equivalently (up to an absolute factor), we have a Sobolev-type inequality

$$\|f - m\|_{\psi_1} \leq c\sigma(\mu) \|\nabla f\|_{\psi_2},$$

connecting the ψ_1 -norm of $f - m$ with the ψ_2 -norm of the modulus of the gradient of f . We prove a more general version of this corollary in Section 6 (cf. Theorem 6.1). As will be explained in the same section, similar assertions may also be made about convex f and product measures μ on $M = [-1, 1]^n$, thus extending Talagrand's theorem to the class of non-Lipschitz functions.

In view of the right bound in (1.3) and (1.4), the spread and subgaussian constants for restricted measures can be controlled in terms of the logarithmic Sobolev constant ρ , via

$$s^2(\mu_A) \leq \sigma^2(\mu_A) \leq c \log\left(\frac{e}{\mu(A)}\right) \frac{1}{\rho}.$$

However, it may happen that $\rho = 0$ and $\sigma^2(\mu) = \infty$, while $\lambda_1 > 0$ (e.g., for the product exponential distribution on \mathbb{R}^n). Then one may wonder whether one can estimate the spread constant of a restricted measure in terms of the spectral gap. In such a case there is a bound similar to (1.4).

Theorem 1.3. *Assume the metric probability space (M, d, μ) satisfies a Poincaré-type inequality with $\lambda_1 > 0$. For any $A \subset M$ with $\mu(A) > 0$, with some absolute constant c*

$$s^2(\mu_A) \leq c \log^2\left(\frac{e}{\mu(A)}\right) \frac{1}{\lambda_1}. \quad (1.7)$$

It should be mentioned that the logarithmic terms in (1.4) and (1.7) may not be removed and are actually asymptotically optimal as functions of $\mu(A)$, as $\mu(A)$ is getting small, see Section 7.

Our contribution below is organized into sections as follows:

2. Bounds on ψ_α -norms for restricted measures.
3. Proof of Theorem 1.1. Transport-entropy formulation.
4. Proof of Theorem 1.3. Spectral gap.
5. Examples.
6. Deviations for non-Lipschitz functions.
7. Optimality.
8. Appendix.

2 Bounds on ψ_α -norms for restricted measures

A measurable function f on the probability space (M, μ) is said to have a finite ψ_α -norm, $\alpha \geq 1$, if for some $r > 0$,

$$\int e^{(|f|/r)^\alpha} d\mu \leq 2.$$

The infimum over all such r represents the ψ_α -norm $\|f\|_{\psi_\alpha}$ or $\|f\|_{L^{\psi_\alpha}(\mu)}$, which is just the Orlicz norm associated with the Young function $\psi_\alpha(t) = e^{|t|^\alpha} - 1$.

We are mostly interested in the particular cases $\alpha = 1$ and $\alpha = 2$. In this section we recall well-known relations between the ψ_1 and ψ_2 -norms and the usual L^p -norms $\|f\|_p = \|f\|_{L^p(\mu)} = (\int |f|^p d\mu)^{1/p}$. For the readers' convenience, we include the proof in the appendix.

Lemma 2.1. *We have*

$$\sup_{p \geq 1} \frac{\|f\|_p}{\sqrt{p}} \leq \|f\|_{L^{\psi_2}(\mu)} \leq 4 \sup_{p \geq 1} \frac{\|f\|_p}{\sqrt{p}}, \quad (2.1)$$

$$\sup_{p \geq 1} \frac{\|f\|_p}{p} \leq \|f\|_{L^{\psi_1}(\mu)} \leq 6 \sup_{p \geq 1} \frac{\|f\|_p}{p}. \quad (2.2)$$

Given a measurable subset A of M with $\mu(A) > 0$, we consider the normalized restricted measure μ_A on M , i.e.,

$$\mu_A(B) = \frac{\mu(A \cap B)}{\mu(A)}, \quad B \subset M.$$

Our basic tool leading to Theorem 1.1 will be the following assertion.

Proposition 2.2. *For any measurable function f on M ,*

$$\|f\|_{L^{\psi_2}(\mu_A)} \leq 4e \log^{1/2} \left(\frac{e}{\mu(A)} \right) \|f\|_{L^{\psi_2}(\mu)}. \quad (2.3)$$

Proof. Assume that $\|f\|_{L^{\psi_2}(\mu)} = 1$ and fix $p \geq 1$. By the left inequality in (2.1), for any $q \geq 1$,

$$q^{q/2} \geq \int |f|^q d\mu \geq \mu(A) \int |f|^q d\mu_A,$$

so

$$\frac{\|f\|_{L^q(\mu_A)}}{\sqrt{q}} \leq \left(\frac{1}{\mu(A)} \right)^{1/q}.$$

But by the right inequality in (2.1),

$$\|f\|_{\psi_2} \leq 4 \sup_{q \geq 1} \frac{\|f\|_q}{\sqrt{q}} \leq 4\sqrt{p} \sup_{q \geq p} \frac{\|f\|_q}{\sqrt{q}}.$$

Applying it on the space (M, μ_A) , we then get

$$\begin{aligned} \|f\|_{L^{\psi_2}(\mu_A)} &\leq 4\sqrt{p} \sup_{q \geq p} \frac{\|f\|_{L^q(\mu_A)}}{\sqrt{q}} \\ &\leq 4\sqrt{p} \sup_{q \geq p} \left(\frac{1}{\mu(A)} \right)^{1/q} = 4\sqrt{p} \left(\frac{1}{\mu(A)} \right)^{1/p}. \end{aligned}$$

The obtained inequality,

$$\|f\|_{L^{\psi_2}(\mu_A)} \leq 4\sqrt{p} \left(\frac{1}{\mu(A)} \right)^{1/p},$$

holds true for any $p \geq 1$ and therefore may be optimized over p . Choosing $p = \log \frac{e}{\mu(A)}$, we arrive at (2.3). \square

A possible weak point in the bound (2.3) is that the means of f are not involved. For example, in applications, if f were defined only on A and had μ_A -mean zero, we might need to find an extension of f to the whole space M keeping the mean zero with respect to μ . In fact, this should not create any difficulty, since one may work with the symmetrization of f .

More precisely, we may apply Proposition 2.2 on the product space $(M \times M, \mu \otimes \mu)$ to the product sets $A \times A$ and functions of the form $f(x) - f(y)$. Then we get

$$\|f(x) - f(y)\|_{L^{\psi_2}(\mu_A \otimes \mu_A)} \leq 4e \log^{1/2} \left(\frac{e}{\mu(A)^2} \right) \|f(x) - f(y)\|_{L^{\psi_2}(\mu \otimes \mu)}.$$

Since $\log \left(\frac{e}{\mu(A)^2} \right) \leq 2 \log \left(\frac{e}{\mu(A)} \right)$, we arrive at:

Corollary 2.3. *For any measurable function f on M ,*

$$\|f(x) - f(y)\|_{L^{\psi_2}(\mu_A \otimes \mu_A)} \leq 4e\sqrt{2} \log^{1/2} \left(\frac{e}{\mu(A)} \right) \|f(x) - f(y)\|_{L^{\psi_2}(\mu \otimes \mu)}.$$

Let us now derive an analog of Proposition 2.2 for the ψ_1 -norm, using similar arguments. Assume that $\|f\|_{L^{\psi_1}(\mu)} = 1$ and fix $p \geq 1$. By the left inequality in (2.2), for any $q \geq 1$,

$$q^q \geq \int |f|^q d\mu \geq \mu(A) \int |f|^q d\mu_A,$$

so

$$\frac{\|f\|_{L^q(\mu_A)}}{q} \leq \left(\frac{1}{\mu(A)}\right)^{1/q}.$$

But, by the inequality (2.2),

$$\|f\|_{L^{\psi_1}} \leq 6 \sup_{q \geq 1} \frac{\|f\|_q}{q} \leq 6p \sup_{q \geq p} \frac{\|f\|_q}{q}.$$

Applying it on the space (M, μ_A) , we get

$$\begin{aligned} \|f\|_{L^{\psi_1}(\mu_A)} &\leq 6p \sup_{q \geq p} \frac{\|f\|_{L^q(\mu_A)}}{q} \\ &\leq 6p \sup_{q \geq p} \left(\frac{1}{\mu(A)}\right)^{1/q} = 6p \left(\frac{1}{\mu(A)}\right)^{1/p}. \end{aligned}$$

The obtained inequality,

$$\|f\|_{L^{\psi_1}(\mu_A)} \leq 6p \left(\frac{1}{\mu(A)}\right)^{1/p},$$

holds true for any $p \geq 1$ and therefore may be optimized over p . Choosing $p = \log \frac{e}{\mu(A)}$, we arrive at:

Proposition 2.4. *For any measurable function f on M , we have*

$$\|f\|_{L^{\psi_1}(\mu_A)} \leq 6e \log \left(\frac{e}{\mu(A)}\right) \|f\|_{L^{\psi_1}(\mu)}.$$

Similarly to Corollary 2.3, one may write down this relation on the product probability space $(M \times M, \mu \otimes \mu)$ with functions of the form $\tilde{f}(x, y) = f(x) - f(y)$ and the product sets $\tilde{A} = A \times A$. Then we get

$$\|f(x) - f(y)\|_{L^{\psi_1}(\mu_A \otimes \mu_A)} \leq 12e \log \left(\frac{e}{\mu(A)}\right) \|f(x) - f(y)\|_{L^{\psi_1}(\mu \otimes \mu)}. \quad (2.4)$$

3 Proof of Theorem 1.1. Transport-entropy formulation

The finiteness of the subgaussian constant for a given metric probability space (M, d, μ) means that ψ_2 -norms of Lipschitz functions on M with mean zero are uniformly bounded. Equivalently, for any (for all) $x_0 \in M$, we have that, for some $\lambda > 0$,

$$\int e^{d(x, x_0)^2 / \lambda^2} d\mu(x) < \infty.$$

Definition (1.1) of $\sigma^2(\mu)$ inspires to consider another norm-like quantity

$$\sigma_f^2 = \sup_{t \neq 0} \left[\frac{1}{t^2/2} \log \int e^{tf} d\mu \right].$$

Here is a well-known relation (with explicit numerical constants) which holds in the setting of an abstract probability space (M, μ) . Once again, we include a proof in the appendix for completeness.

Lemma 3.1. *If f has mean zero and finite ψ_2 -norm, then*

$$\frac{1}{\sqrt{6}} \|f\|_{\psi_2}^2 \leq \sigma_f^2 \leq 4 \|f\|_{\psi_2}^2.$$

One can now relate the subgaussian constant of the restricted measure to the subgaussian constant of the original measure. Let now (M, d, μ) be a metric probability space. First, Lemma 3.1 immediately yields an equivalent description in terms of ψ_2 -norms, namely

$$\frac{1}{\sqrt{6}} \sup_f \|f\|_{\psi_2}^2 \leq \sigma^2(\mu) \leq 4 \sup_f \|f\|_{\psi_2}^2, \quad (3.1)$$

where the supremum is running over all $f : M \rightarrow \mathbb{R}$ with μ -mean zero and $\|f\|_{\text{Lip}} \leq 1$. Here, one can get rid of the mean zero assumption by considering functions of the form $f(x) - f(y)$ on the product space $(M \times M, \mu \otimes \mu, d_1)$, where d_1 is the l_1 -type metric given by $d_1((x_1, y_1), (x_2, y_2)) = d(x_1, x_2) + d(y_1, y_2)$. If f has mean zero, then, by Jensen's inequality,

$$\iint e^{(f(x)-f(y))^2/r^2} d\mu(x) d\mu(y) \geq \int e^{f(x)^2/r^2} d\mu(x),$$

which implies that

$$\|f(x) - f(y)\|_{L^{\psi_2}(\mu \otimes \mu)} \geq \|f\|_{L^{\psi_2}(\mu)}.$$

On the other hand, by the triangle inequality,

$$\|f(x) - f(y)\|_{L^{\psi_2}(\mu \otimes \mu)} \leq 2 \|f\|_{L^{\psi_2}(\mu)}.$$

Hence, we arrive at another, more flexible relation, where the mean zero assumption may be removed.

Lemma 3.2. *We have*

$$\frac{1}{4\sqrt{6}} \sup_f \|f(x) - f(y)\|_{L^{\psi_2}(\mu \otimes \mu)}^2 \leq \sigma^2(\mu) \leq 4 \sup_f \|f(x) - f(y)\|_{L^{\psi_2}(\mu \otimes \mu)}^2,$$

where the supremum is running over all functions f on M with $\|f\|_{\text{Lip}} \leq 1$.

Proof of Theorem 1.1. We are prepared to make last steps for the proof of the inequality (1.4). We use the well-known Kirszbraun's theorem: Any function $f : A \rightarrow \mathbb{R}$ with Lipschitz semi-norm $\|f\|_{\text{Lip}} \leq 1$ on A admits a Lipschitz extension to the whole space ([K], [MS]). Namely, one may put

$$\tilde{f}(x) = \inf_{a \in A} [f(a) + d(a, x)], \quad x \in M.$$

Applying first Corollary 2.3 and then the left inequality of Lemma 3.2 to \tilde{f} , we get

$$\begin{aligned} \|f(x) - f(y)\|_{L^{\psi_2}(\mu_A \otimes \mu_A)}^2 &= \|\tilde{f}(x) - \tilde{f}(y)\|_{L^{\psi_2}(\mu_A \otimes \mu_A)}^2 \\ &\leq (4e\sqrt{2})^2 \log\left(\frac{e}{\mu(A)}\right) \|\tilde{f}(x) - \tilde{f}(y)\|_{L^{\psi_2}(\mu \otimes \mu)}^2 \\ &\leq (4e\sqrt{2})^2 \log\left(\frac{e}{\mu(A)}\right) \cdot (4\sqrt{6})^2 \sigma^2(\mu). \end{aligned}$$

Another application of Lemma 3.2 – in the space (A, d, μ_A) (now the right inequality) yields

$$\sigma^2(\mu_A) \leq 4 \cdot (4e\sqrt{2})^2 \log\left(\frac{e}{\mu(A)}\right) \cdot (4\sqrt{6})^2 \sigma^2(\mu).$$

This is exactly (1.4) with constant $c = 4 \cdot (4e\sqrt{2})^2 (4\sqrt{6})^2 = 3 \cdot 2^{12} e^2 = 90796.72\dots$ \square

Remark 3.3. Let us also record the following natural generalization of Theorem 1.1, which is obtained along the same arguments. Given a collection \mathcal{F} of (integrable) functions on the probability space (M, μ) , define $\sigma_{\mathcal{F}}^2(\mu)$ as the infimum over all σ^2 such that

$$\int e^{t(f-m)} d\mu \leq e^{\sigma^2 t^2/2}, \quad \text{for all } t \in \mathbb{R},$$

for any $f \in \mathcal{F}$, where $m = \int f d\mu$. Then with the same constant c as in Theorem 1.1, for any measurable $A \subset M$, $\mu(A) > 0$, we have

$$\sigma_{\mathcal{F}_A}^2(\mu_A) \leq c \log \left(\frac{e}{\mu(A)} \right) \sigma_{\mathcal{F}}^2(\mu),$$

where \mathcal{F}_A denotes the collection of restrictions of functions f from \mathcal{F} to the set A .

Let us now mention an interesting connection of the subgaussian constant with the Kantorovich distance

$$W_1(\mu, \nu) = \inf \iint d(x, y) \pi(x, y)$$

and relative entropy

$$D(\nu||\mu) = \int \log \frac{d\nu}{d\mu} d\nu,$$

(also called Kullback-Leibler distance or informational divergence). Here, ν is a probability measure on M , which is absolutely continuous with respect to μ (for short, $\nu \ll \mu$), and the infimum in the definition of W_1 is running over all probability measures π on the product space $M \times M$ with marginal distributions μ and ν , i.e., such that

$$\pi(B \times M) = \mu(B), \quad \pi(M \times B) = \nu(B) \quad (\text{Borel } B \subset M).$$

As was shown in [B-G], if (M, d) is a Polish space (complete separable), the subgaussian constant $\sigma^2 = \sigma^2(\mu)$ may be described as an optimal value in the transport-entropy inequality

$$W_1(\mu, \nu) \leq \sqrt{2\sigma^2 D(\nu||\mu)}. \quad (3.2)$$

Hence, we obtain from the inequality (1.4) a similar relation for measures ν supported on given subsets of M .

Corollary 3.4. *Given a Borel probability measure μ on a Polish space (M, d) and a closed set A in M such that $\mu(A) > 0$, for any Borel probability measure ν supported on A ,*

$$W_1^2(\mu_A, \nu) \leq c\sigma^2(\mu) \log \left(\frac{e}{\mu(A)} \right) D(\nu||\mu_A),$$

where c is an absolute constant.

This assertion is actually equivalent to Theorem 1.1. Note that, for ν supported on A , there is an identity $D(\nu||\mu_A) = \log \mu(A) + D(\nu||\mu)$. In particular, $D(\nu||\mu_A) \leq D(\nu||\mu)$, so the relative entropies decrease when turning to restricted measures.

For another (almost equivalent) description of the subgaussian constant, introduce the concentration function

$$\mathcal{K}_\mu(r) = \sup [1 - \mu(A^r)] \quad (r > 0),$$

where $A^r = \{x \in M : d(x, a) < r \text{ for some } a \in A\}$ denotes an open r -neighbourhood of A for the metric d , and the sup is running over all Borel sets $A \subset M$ of measure $\mu(A) \geq \frac{1}{2}$. As is well-known, the transport-entropy inequality (3.2) gives rise to a concentration inequality

on (M, d, μ) of a subgaussian type (K. Marton's argument), but this can also be seen by a direct application of (1.1). Indeed, for any function f on M with $\|f\|_{\text{Lip}} \leq 1$, it implies

$$\iint e^{t(f(x)-f(y))} d\mu(x) d\mu(y) \leq e^{\sigma^2 t^2}, \quad t \in \mathbb{R},$$

and, by Chebyshev's inequality, we have a deviation bound

$$(\mu \otimes \mu)\{(x, y) \in M \times M : f(x) - f(y) \geq r\} \leq e^{-r^2/4\sigma^2}, \quad r \geq 0.$$

In particular, one may apply it to the distance functions $f(x) = d(A, x) = \inf_{a \in A} d(a, x)$. Assuming that $\mu(A) \geq \frac{1}{2}$, the measure on the left-hand side is greater than or equal to $\frac{1}{2}(1 - \mu(A^r))$, so that we obtain a concentration inequality

$$1 - \mu(A^r) \leq 2e^{-r^2/4\sigma^2}.$$

Therefore,

$$\mathcal{K}_\mu(r) \leq \min\left\{\frac{1}{2}, 2e^{-r^2/4\sigma^2}\right\} \leq e^{-r^2/8\sigma^2}.$$

To argue in the opposite direction, suppose that the concentration function admits a bound of the form $\mathcal{K}_\mu(r) \leq e^{-r^2/b^2}$ for all $r > 0$ with some constant $b > 0$. Given a function f on M with $\|f\|_{\text{Lip}} \leq 1$, let m be a median of f under μ . Then the set $A = \{f \leq m\}$ has measure $\mu(A) \geq \frac{1}{2}$, and by the Lipschitz property, $A^r \subset \{f < m + r\}$ for all $r > 0$. Hence, by the concentration hypothesis,

$$\mu\{f - m \geq r\} \leq \mathcal{K}_\mu(r) \leq e^{-r^2/b^2}.$$

A similar deviation bound also holds for the function $-f$ with its median $-m$, so that

$$\mu\{|f - m| \geq r\} \leq 2e^{-r^2/b^2}, \quad r > 0.$$

This is sufficient to properly estimate ψ_2 -norm of $f - m$ on (M, μ) . Namely, for any $\lambda < 1/b^2$,

$$\begin{aligned} \int e^{\lambda|f-m|^2} d\mu &= 1 + 2\lambda \int_0^\infty r e^{\lambda r^2} \mu\{|f-m| \geq r\} dr \\ &\leq 1 + 2\lambda \int_0^\infty r e^{\lambda r^2} e^{-r^2/b^2} dr = 1 + \frac{2\lambda}{\frac{1}{b^2} - \lambda} = 2, \end{aligned}$$

where in the last equality the value $\lambda = \frac{1}{3b^2}$ is chosen. Thus, $\int e^{|f-m|^2/(3b^2)} d\mu \leq 2$, which means that $\|f - m\|_{\psi_2} \leq \sqrt{3}b$. The latter gives $\|f(x) - f(y)\|_{L^{\psi_2}(\mu \otimes \mu)} \leq 2\sqrt{3}b$. Taking the supremum over f , it remains to apply Lemma 3.2, and then we get $\sigma^2(\mu) \leq 48b^2$.

Let us summarize.

Proposition 3.5. *Let $b = b(\mu)$ be an optimal value such that the concentration function of the space (M, d, μ) satisfies a subgaussian bound $\mathcal{K}_\mu(r) \leq e^{-r^2/b^2}$ ($r > 0$). Then*

$$\frac{1}{8} b^2(\mu) \leq \sigma^2(\mu) \leq 48 b^2(\mu).$$

Once this description of the subgaussian constant is recognized, one may give another proof of Theorem 1.1, by relating the concentration function \mathcal{K}_{μ_A} to \mathcal{K}_μ . In this connection let us state below, as a lemma, a general observation due to F. Barthe and E. Milman (cf. [B-M], Lemma 2.1, p. 585).

Lemma 3.6. *Let a Borel probability measure ν on M be absolutely continuous with respect to μ and have density p . Suppose that, for some right-continuous, non-increasing function $R : (0, 1/4] \rightarrow (0, \infty)$, such that $\beta(\varepsilon) = \varepsilon/R(\varepsilon)$ is increasing, we have*

$$\nu\{x \in M : p(x) > R(\varepsilon)\} \leq \varepsilon \quad \left(0 < \varepsilon \leq \frac{1}{4}\right).$$

Then

$$K_\nu(r) \leq 2\beta^{-1}(K_\mu(r/2)), \quad \text{for all } r \geq 2K_\mu^{-1}(\beta(1/4)).$$

Here β^{-1} denotes the inverse function, and $K_\mu^{-1}(\varepsilon) = \inf\{r > 0 : K_\mu(r) < \varepsilon\}$.

The 2nd proof of Theorem 1.1. The normalized restricted measure $\nu = \mu_A$ has density $p = \frac{1}{\mu(A)} 1_A$ (thus taking only one non-zero value), and an optimal choice of R is the constant function $R(\varepsilon) = \frac{1}{\mu(A)}$. Hence, Lemma 3.6 yields the relation

$$K_{\mu_A}(r) \leq \frac{2}{\mu(A)} K_\mu(r/2), \quad \text{for } r \geq 2K_\mu^{-1}(\mu(A)/4).$$

In particular, if $K_\mu(r) \leq e^{-r^2/b^2}$, then

$$K_{\mu_A}(r) \leq \frac{2}{\mu(A)} e^{-r^2/(4b^2)}, \quad \text{for } r \geq 2b\sqrt{\log(4/\mu(A))}.$$

Necessarily $K_{\mu_A}(r) \leq \frac{1}{2}$, so the last relation may be extended to the whole positive half-axis. Moreover, at the expense of a factor in the exponent, one can remove the factor $\frac{2}{\mu(A)}$; more precisely, we get $K_{\mu_A}(r) \leq e^{-r^2/\tilde{b}^2}$ with $\tilde{b}^2 = \frac{4b^2}{\log 2} \log \frac{4}{\mu(A)}$, that is,

$$b^2(\mu_A) \leq \frac{4}{\log 2} \log \frac{4}{\mu(A)} b^2(\mu).$$

It remains to apply the two-sided bound of Proposition 3.5. □

4 Proof of Theorem 1.3. Spectral gap

Theorem 1.1 insures, in particular, that, for any function f on the metric probability space (M, d, μ) with Lipschitz semi-norm $\|f\|_{\text{Lip}} \leq 1$,

$$\text{Var}_{\mu_A}(f) \leq c \log \left(\frac{e}{\mu(A)} \right) \sigma^2(\mu),$$

up to some absolute constant c . In fact, in order to reach a similar concentration property of the restricted measures, it is enough to start with a Poincaré-type inequality on M ,

$$\lambda_1 \text{Var}_\mu(f) \leq \int |\nabla f|^2 d\mu.$$

Under this hypothesis, a well-known theorem due to Gromov-Milman and Borovkov-Utev asserts that mean zero Lipschitz functions f have bounded ψ_1 -norm. One may use a variant of this theorem proposed by Aida and Strook [A-S], who showed that

$$\int e^{\sqrt{\lambda_1} f} d\mu \leq K_0 = 1.720102\dots \quad (\|f\|_{\text{Lip}} \leq 1).$$

Hence

$$\int e^{\sqrt{\lambda_1}|f|} d\mu \leq 2K_0 \quad \text{and} \quad \int e^{\frac{1}{2}\sqrt{\lambda_1}|f|} d\mu \leq \sqrt{2K_0} < 2,$$

thus implying that $\|f\|_{\psi_1} \leq \frac{2}{\sqrt{\lambda_1}}$. In addition,

$$\iint e^{\sqrt{\lambda_1}(f(x)-f(y))} d\mu(x)d\mu(y) \leq K_0^2 \quad \text{and} \quad \iint e^{\sqrt{\lambda_1}|f(x)-f(y)|} d\mu(x)d\mu(y) \leq 2K_0^2 < 6.$$

From this,

$$\int e^{\frac{1}{3}\sqrt{\lambda_1}|f(x)-f(y)|} d\mu(x)d\mu(y) < 6^{1/3} < 2,$$

which means that $\|f(x) - f(y)\|_{\psi_1} \leq \frac{3}{\sqrt{\lambda_1}}$ with respect to the product measure $\mu \otimes \mu$ on the product space $M \times M$. This inequality is translation invariant, so the mean zero assumption may be removed. Thus, we arrive at:

Lemma 4.1. *Under the Poincaré-type inequality with spectral gap $\lambda_1 > 0$, for any mean zero function f on (M, d, μ) with $\|f\|_{\text{Lip}} \leq 1$,*

$$\|f\|_{\psi_1} \leq \frac{2}{\sqrt{\lambda_1}}.$$

Moreover, for any f with $\|f\|_{\text{Lip}} \leq 1$,

$$\|f(x) - f(y)\|_{L^{\psi_1}(\mu \otimes \mu)} \leq \frac{3}{\sqrt{\lambda_1}}. \quad (4.1)$$

This is a version of the concentration of measure phenomenon (with exponential integrability) in the presence of a Poincaré-type inequality. Our goal is therefore to extend this property to the normalized restricted measures μ_A . This can be achieved by virtue of the inequality (2.4) which when combined with (4.1) yields an upper bound

$$\|f(x) - f(y)\|_{L^{\psi_1}(\mu_A \otimes \mu_A)} \leq 36e \log\left(\frac{e}{\mu(A)}\right) \frac{1}{\sqrt{\lambda_1}}.$$

Moreover, if f has μ_A -mean zero, the left norm dominates $\|f\|_{L^{\psi_1}(\mu_A)}$ (by Jensen's inequality). We can summarize, taking into account once again Kirszbraun's theorem, as we did in the proof of Theorem 1.1.

Proposition 4.2. *Assume the metric probability space (M, d, μ) satisfies a Poincaré-type inequality with constant $\lambda_1 > 0$. Given a measurable set $A \subset M$ with $\mu(A) > 0$, for any function $f : A \rightarrow \mathbb{R}$ with μ_A -mean zero and such that $\|f\|_{\text{Lip}} \leq 1$ on A ,*

$$\|f\|_{L^{\psi_1}(\mu_A)} \leq 36e \log\left(\frac{e}{\mu(A)}\right) \frac{1}{\sqrt{\lambda_1}}.$$

Theorem 1.3 is now easily obtained with the constant $c = 2(36e)^2$, by noting that L^2 -norms are dominated by L^{ψ_1} -norms. More precisely, since $e^{|t|} - 1 \geq \frac{1}{2}t^2$, one has $\|f\|_{\psi_1}^2 \geq \frac{1}{2}\|f\|_2^2$.

Remark 4.3. Related stability results are known for various classes of probability distributions on the Euclidean spaces $M = \mathbb{R}^n$ (and even in a more general situation, where μ_A is replaced by an absolutely continuous measure with respect to μ). See, in particular, the works by E. Milman [Mi1, Mi2] on convex bodies and log-concave measures.

5 Examples

Theorems 1.1 and 1.3 involve a lot of interesting examples. Here are some natural ones.

1) The standard Gaussian measure $\mu = \gamma$ on \mathbb{R}^n satisfies a logarithmic Sobolev inequality on $M = \mathbb{R}^n$ with a dimension-free constant $\rho = 1$. Hence, from Theorem 1.1 we get:

Corollary 5.1. *For any measurable set $A \subset \mathbb{R}^n$ with $\gamma(A) > 0$, the subgaussian constant $\sigma^2(\gamma_A)$ of the normalized restricted measure γ_A satisfies*

$$\sigma^2(\gamma_A) \leq c \log \left(\frac{e}{\gamma(A)} \right),$$

where c is an absolute constant.

As it was already mentioned, if A is convex, there is a sharper bound $\sigma^2(\gamma_A) \leq 1$. However, it may not hold without convexity assumption. Nevertheless, if $\gamma(A)$ is bounded away from zero, we obtain a more universal principle.

Clearly, Corollary 5.1 extends to all product measures $\mu = \nu^n$ on \mathbb{R}^n such that ν satisfies a logarithmic Sobolev inequality on the real line, and with constants c depending on ρ , only. A characterization of the property $\rho > 0$ in terms of the distribution function of the measure ν and the density of its absolutely continuous component may be found in [B-G].

2) Consider a uniform distribution ν on the shell

$$A_\varepsilon = \{x \in \mathbb{R}^n : 1 - \varepsilon \leq |x| \leq 1\}, \quad 0 \leq \varepsilon \leq 1 \quad (n \geq 2).$$

Corollary 5.2. *The subgaussian constant of ν satisfies $\sigma^2(\nu) \leq \frac{c}{n}$, up to some absolute constant c .*

In other words, mean zero Lipschitz functions f on A_ε are such that $\sqrt{n} f$ are subgaussian with universal constant factor. This property is well-known in the extreme cases – on the unit Euclidean ball $A = B_n$ ($\varepsilon = 1$) and on the unit sphere $A = S^{n-1}$ ($\varepsilon = 0$).

Let μ denote the normalized Lebesgue measure on B_n . In the case $\varepsilon \geq \frac{1}{n}$, the shell A_ε represents the part of B_n of measure

$$\mu(A_\varepsilon) = 1 - \left(1 - \frac{1}{n}\right)^n \geq 1 - \frac{1}{e}.$$

Since the logarithmic Sobolev constant of the unit ball is of order $\frac{1}{n}$, and therefore $\sigma^2(\mu) \leq \frac{c}{n}$, the assertion of Corollary 5.2 immediately follows from Theorem 1.1.

In case $\varepsilon \leq \frac{1}{n}$, the assertion follows from a similar concentration property of the uniform distribution σ_{n-1} on the unit sphere. Indeed, with every Lipschitz function f on A_ε one may associate its restriction to S^{n-1} , which is also Lipschitz (with respect to the Euclidean distance). We have $|f(r\theta) - f(\theta)| \leq |r - 1| \leq \varepsilon \leq \frac{1}{n}$, for any $r \in [1 - \varepsilon, 1]$ and $\theta \in S^{n-1}$. Hence,

$$\begin{aligned} |f(r'\theta') - f(r\theta)| &\leq |f(\theta') - f(\theta)| + |f(r'\theta) - f(\theta')| + |f(r\theta) - f(\theta)| \\ &\leq |f(\theta') - f(\theta)| + \frac{2}{n}, \end{aligned}$$

whenever $r, r' \in [1 - \varepsilon, 1]$ and $\theta, \theta' \in S^{n-1}$, which implies

$$|f(r'\theta') - f(r\theta)|^2 \leq 2|f(\theta') - f(\theta)|^2 + \frac{8}{n^2}.$$

But the map $(r, \theta) \rightarrow \theta$ pushes forward ν onto σ_{n-1} , so, we obtain that, for any $c > 0$,

$$\begin{aligned} \iint \exp\{cn |f(r'\theta') - f(r\theta)|^2\} d\nu(r', \theta') d\nu(r, \theta) &\leq \\ e^{8/n} \iint \exp\{2cn |f(\theta') - f(\theta)|^2\} d\sigma_{n-1}(\theta') d\sigma_{n-1}(\theta). \end{aligned}$$

Here, for a certain numerical constant $c > 0$, the right-hand side is bounded by a universal constant, thus proving the claim.

3) The two-sided product exponential measure μ on \mathbb{R}^n with density $2^{-n} e^{-(|x_1| + \dots + |x_n|)}$ satisfies a Poincaré-type inequality on $M = \mathbb{R}^n$ with a dimension-free constant $\lambda_1 = 1/4$. Hence, from Proposition 4.2 we get:

Corollary 5.3. *For any measurable set $A \subset \mathbb{R}^n$ with $\mu(A) > 0$, and for any function $f : A \rightarrow \mathbb{R}$ with μ_A -mean zero and $\|f\|_{\text{Lip}} \leq 1$, we have*

$$\|f\|_{L^{\psi_1}(\mu_A)} \leq c \log \left(\frac{e}{\mu(A)} \right),$$

where c is an absolute constant. In particular,

$$s^2(\mu_A) \leq c \log^2 \left(\frac{e}{\mu(A)} \right).$$

Clearly, Corollary 5.3 extends to all product measures $\mu = \nu^n$ on \mathbb{R}^n such that ν satisfies a Poincaré-type inequality on the real line, and with constants c depending on λ_1 , only. A characterization of the property $\lambda_1 > 0$ may also be given in terms of the distribution function of ν and the density of its absolutely continuous component (cf. [B-G]).

4a) Let us take the metric probability space $(\{0, 1\}^n, d_n, \mu)$, where d_n is the Hamming distance, that is, $d_n(x, y) = \#\{i : x_i \neq y_i\}$, equipped with the uniform measure μ . For this particular space, Marton established the transport-entropy inequality (3.2) with an optimal constant $\sigma^2 = \frac{n}{4}$, cf. [Mar]. Using the relation (3.2) as an equivalent definition of the subgaussian constant, we obtain from Theorem 1.1:

Corollary 5.4. *For any non-empty set $A \subset \{0, 1\}^n$, the subgaussian constant $\sigma^2(\mu_A)$ of the normalized restricted measure μ_A satisfies, up to an absolute constant c ,*

$$\sigma^2(\mu_A) \leq cn \log \left(\frac{e}{\mu(A)} \right). \quad (5.1)$$

4b) Let us now assume that A is *monotone*, i.e., A satisfies the condition

$$(x_1, \dots, x_n) \in A \implies (y_1, \dots, y_n) \in A, \text{ whenever } y_i \geq x_i, i = 1, \dots, n.$$

Recall that the discrete cube can be equipped with a natural graph structure: there is an edge between x and y whenever they are of Hamming distance $d_n(x, y) = 1$. For monotone sets A , the graph metric d_A on the subgraph of A is equal to the restriction of d_n to $A \times A$. Indeed, we have:

$$d_n(x, y) \leq d_A(x, y) \leq d_A(x, x \wedge y) + d_A(y, x \wedge y) = d_n(x, x \wedge y) + d_n(y, x \wedge y) = d_n(x, y),$$

where $x \wedge y = (x_1 \wedge y_1, \dots, x_n \wedge y_n)$. Thus,

$$s^2(\mu_A, d_A) \leq \sigma^2(\mu_A, d_A) \leq cn \log \left(\frac{e}{\mu(A)} \right).$$

This can be compared with what follows from a recent result of Ding and Mossel (see [D-M]). The authors proved that the conductance (Cheeger constant) of (A, μ_A) satisfies $\phi(A) \geq \frac{\mu(A)}{16n}$. However, these type of isoperimetric results may not imply sharp concentration bounds. Indeed, by using Cheeger inequality, the above inequality leads to $\lambda_1 \geq c\mu(A)^2/n^2$ and $s^2(\mu_A, d_A) \leq 1/\lambda_1 \leq cn^2/\mu(A)^2$, which is even worse than the trivial estimate $s^2(\mu_A, d_A) \leq \frac{1}{2} \text{diam}(A)^2 \leq n^2/2$.

5) Let (M, d, μ) be a (separable) metric probability space with finite subgaussian constant $\sigma^2(\mu)$. The previous example can be naturally generalized to the product space (M^n, μ^n) , when it is equipped with the ℓ^1 -type metric

$$d_n(x, y) = \sum_{i=1}^n d(x_i, y_i), \quad x = (x_1, \dots, x_n), \quad y = (y_1, \dots, y_n) \in M^n.$$

This can be done with the help of the following elementary observation.

Proposition 5.5. *The subgaussian constant of the space (M^n, d_n, μ^n) is related to the subgaussian constant of (M, d, μ) by the equality $\sigma^2(\mu^n) = n\sigma^2(\mu)$.*

Indeed, one may argue by induction on n . Let f be a function on M^n . The Lipschitz property $\|f\|_{\text{Lip}} \leq 1$ with respect to d_n is equivalent to the assertion that f is coordinatewise Lipschitz, that is, any function of the form $x_i \rightarrow f(x)$ has a Lipschitz semi-norm ≤ 1 on M for all fixed coordinates $x_j \in M$ ($j \neq i$). Hence, in this case, for all $t \in \mathbb{R}$,

$$\int_M e^{tf(x)} d\mu(x_n) \leq \exp \left\{ t \int_M f(x) d\mu(x_n) + \frac{\sigma^2 t^2}{2} \right\},$$

where $\sigma^2 = \sigma^2(\mu)$. Here the function $(x_1, \dots, x_{n-1}) \rightarrow \int_M f(x) d\mu(x_n)$ is also coordinatewise Lipschitz. Integrating the above inequality with respect to $d\mu^{n-1}(x_1, \dots, x_{n-1})$ and applying the induction hypothesis, we thus get

$$\int_{M^n} e^{tf(x)} d\mu^n(x) \leq \exp \left\{ t \int_{M^n} f(x) d\mu^n(x) + n \frac{\sigma^2 t^2}{2} \right\}.$$

But this means that $\sigma^2(\mu^n) \leq n\sigma^2(\mu)$.

For an opposite bound, it is sufficient to test (1.1) for (M^n, d_n, μ^n) in the class of all coordinatewise Lipschitz functions of the form $f(x) = u(x_1) + \dots + u(x_n)$ with μ -mean zero functions u on M such that $\|u\|_{\text{Lip}} \leq 1$.

Corollary 5.6. *For any Borel set $A \subset M^n$ such that $\mu^n(A) > 0$, the subgaussian constant of the normalized restricted measure μ_A^n with respect to the ℓ^1 -type metric d_n satisfies*

$$\sigma^2(\mu_A^n) \leq cn\sigma^2(\mu) \log \left(\frac{e}{\mu^n(A)} \right),$$

where c is an absolute constant.

For example, if μ is a probability measure on $M = \mathbb{R}$ such that $\int_{-\infty}^{\infty} e^{x^2/\lambda^2} d\mu(x) \leq 2$ ($\lambda > 0$), then for the restricted product measures we have

$$\sigma^2(\mu_A^n) \leq cn\lambda^2 \log \left(\frac{e}{\mu^n(A)} \right) \tag{5.2}$$

with respect to the ℓ^1 -norm $\|x\|_1 = |x_1| + \dots + |x_n|$ on \mathbb{R}^n .

Indeed, by the integral hypothesis on μ , for any f on \mathbb{R} with $\|f\|_{\text{Lip}} \leq 1$,

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{(f(x)-f(y))^2/2\lambda^2} d\mu(x)d\mu(y) &\leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{(x-y)^2/2\lambda^2} d\mu(x)d\mu(y) \\ &\leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{(x^2+y^2)/\lambda^2} d\mu(x)d\mu(y) \leq 4. \end{aligned}$$

Hence, if f has μ -mean zero, by Jensen's inequality,

$$\int_{-\infty}^{\infty} e^{f(x)^2/4\lambda^2} d\mu(x) \leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{(f(x)-f(y))^2/4\lambda^2} d\mu(x)d\mu(y) \leq 2,$$

meaning that $\|f\|_{L^{\psi_2}(\mu)} \leq 2\lambda$. By Lemma 3.1, cf. (3.1), it follows that $\sigma^2(\mu) \leq 16\lambda^2$, so, (5.2) holds true by an application of Corollary 5.6.

6 Deviations for non-Lipschitz functions

Let us now turn to the interesting question on the relationship between the distribution of a locally Lipschitz function and the distribution of its modulus of the gradient. We still keep the setting of a metric probability space (M, d, μ) and assume it has a finite subgaussian constant $\sigma^2 = \sigma^2(\mu)$ ($\sigma \geq 0$).

Let us say that a continuous function f on M is locally Lipschitz, if $|\nabla f(x)|$ is finite for all $x \in M$. Recall that we consider the sets

$$A = \{x \in M : |\nabla f(x)| \leq L\}, \quad L > 0. \quad (6.1)$$

First we state a more general version of Corollary 1.2.

Theorem 6.1. *Assume that a locally Lipschitz function f on M has Lipschitz semi-norms $\leq L$ on sets of the form (6.1). If $\mu\{|\nabla f| \geq L_0\} \leq \frac{1}{2}$, then for all $t > 0$,*

$$(\mu \otimes \mu)\{|f(x) - f(y)| \geq t\} \leq 2 \inf_{L \geq L_0} \left[e^{-t^2/c\sigma^2 L^2} + \mu\{|\nabla f| > L\} \right], \quad (6.2)$$

where c is an absolute constant.

Proof. Although the argument is already mentioned in Section 1, let us replace (1.6) with a slightly different bound. First note that the Lipschitz semi-norm of f with respect to the metric d in M is the same as its Lipschitz semi-norm with respect to the metric on the set A induced from M (which is true for any non-empty subset of M). Hence, we are in a position to apply Theorem 1.1, and then Definition (1.1) for the normalized restriction μ_A yields a subgaussian bound:

$$\iint e^{t(f(x)-f(y))} d\mu_A(x)d\mu_A(y) \leq e^{c\sigma^2 L^2 t^2/2}, \quad \text{for all } t \in \mathbb{R},$$

where A is defined in (6.1) with $L \geq L_0$, and c is a universal constant. From this, for any $t > 0$,

$$(\mu_A \otimes \mu_A)\{(x, y) \in A \times A : |f(x) - f(y)| \geq t\} \leq 2e^{-t^2/(2c\sigma^2 L^2)},$$

and therefore

$$(\mu \otimes \mu)\{(x, y) \in A \times A : |f(x) - f(y)| \geq t\} \leq 2e^{-t^2/(2c\sigma^2 L^2)}.$$

The product measure of the complement of $A \times A$ does not exceed $2\mu\{|\nabla f(x)| > L\}$, and we obtain (6.2). \square

If $\int e^{|\nabla f|^2} d\mu \leq 2$, we have, by Chebyshev's inequality, $\mu\{|\nabla f| \geq L\} \leq 2e^{-L^2}$, so one may take $L_0 = \sqrt{\log 4}$. Theorem 6.1 then gives that, for any $L^2 \geq \log 4$,

$$(\mu \otimes \mu)\{|f(x) - f(y)| \geq t\} \leq 2e^{-t^2/c\sigma^2 L^2} + 4e^{-L^2}.$$

For $t \geq 2\sigma$, one may choose here $L^2 = \frac{t}{\sigma}$, leading to

$$(\mu \otimes \mu)\{|f(x) - f(y)| \geq t\} \leq 6e^{-t/c\sigma},$$

for some absolute constant $c > 1$. In case $0 \leq t \leq 2\sigma$, this inequality is fulfilled automatically, so it holds for all $t \geq 0$. As a result, with some absolute constant C ,

$$\|f(x) - f(y)\|_{\psi_1} \leq C\sigma,$$

which is an equivalent way to state the inequality of Corollary 1.2.

As we have already mentioned, with the same arguments inequalities like (6.2) can be derived on the basis of subgaussian constants defined for different classes of functions. For example, one may consider the subgaussian constant $\sigma_{\mathcal{F}}^2(\mu)$ for the class \mathcal{F} of all convex Lipschitz functions f on the Euclidean space $M = \mathbb{R}^n$ (which we equip with the Euclidean distance). Note that $|\nabla f(x)|$ is everywhere finite in the n -space, when f is convex. Keeping in mind Remark 4.3, what we need is the following analog of Kirszbraun's theorem:

Lemma 6.2. *Let f be a convex function on \mathbb{R}^n . For any $L > 0$, there exists a convex function g on \mathbb{R}^n such that $f = g$ on the set $A = \{x : |\nabla f(x)| \leq L\}$ and $|\nabla g| \leq L$ on \mathbb{R}^n .*

Accepting for a moment this lemma without proof, we get:

Theorem 6.3. *Assume that a convex function f on \mathbb{R}^n satisfies $\mu\{|\nabla f| \geq L_0\} \leq \frac{1}{2}$. Then for all $t > 0$,*

$$(\mu \otimes \mu)\{|f(x) - f(y)| \geq t\} \leq 2 \inf_{L \geq L_0} \left[e^{-t^2/c\sigma^2 L^2} + \mu\{|\nabla f| > L\} \right],$$

where $\sigma^2 = \sigma_{\mathcal{F}}^2(\mu)$ and c is an absolute constant.

For illustration, let $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ be an arbitrary product probability measure on the cube $[-1, 1]^n$. If f is convex and Lipschitz on \mathbb{R}^n , thus with $|\nabla f| \leq 1$, then

$$(\mu \otimes \mu)\{|f(x) - f(y)| \geq t\} \leq 2e^{-t^2/c}. \quad (6.3)$$

This is one of the forms of Talagrand's concentration phenomenon for the family of convex sets/functions (cf. [T1, T2], [M], [L]). That is, the subgaussian constants $\sigma_{\mathcal{F}}^2(\mu)$ are bounded for the class \mathcal{F} of convex Lipschitz f and product measures μ on the cube. Hence, using Theorem 6.3, Talagrand's deviation inequality (6.3) admits a natural extension to the class of non-Lipschitz convex functions:

Corollary 6.4. *Let μ be a product probability measure on the cube, and let f be a convex function on \mathbb{R}^n . If $\mu\{|\nabla f| \geq L_0\} \leq \frac{1}{2}$, then for all $t > 0$,*

$$(\mu \otimes \mu)\{|f(x) - f(y)| \geq t\} \leq 2 \inf_{L \geq L_0} \left[e^{-t^2/cL^2} + \mu\{|\nabla f| > L\} \right],$$

where c is an absolute constant.

In particular, we have a statement similar to Corollary 1.2 – for this family of functions, namely

$$\|f - m\|_{L^{\psi_1}(\mu)} \leq c \|\nabla f\|_{L^{\psi_2}(\mu)},$$

where m is the μ -mean of f .

Proof of Lemma 6.2. An affine function $l_{a,v}(x) = a + \langle x, v \rangle$ ($v \in \mathbb{R}^n$, $a \in \mathbb{R}$) may be called a tangent function to f , if $f \geq l$ on \mathbb{R}^n and $f(x) = l_{a,v}(x)$ for at least one point x . It is well-known that

$$f(x) = \sup\{l_{a,v}(x) : l_{a,v} \in \mathcal{L}\},$$

where \mathcal{L} denotes the collection of all tangent functions $l_{a,v}$. Put,

$$g(x) = \sup\{l_{a,v}(x) : l_{a,v} \in \mathcal{L}, |v| \leq L\}.$$

By the construction, $g \leq f$ on \mathbb{R}^n and, moreover,

$$\begin{aligned} \|g\|_{\text{Lip}} &\leq \sup\{\|l_{a,v}\|_{\text{Lip}} : l_{a,v} \in \mathcal{L}, |v| \leq L\} \\ &= \sup\{|v| : l_{a,v} \in \mathcal{L}, |v| \leq L\} \leq L. \end{aligned}$$

It remains to show that $g = f$ on the set $A = \{|\nabla f| \leq L\}$. Let $x \in A$ and let $l_{a,v}$ be tangent to f and such that $l_{a,v}(x) = f(x)$. This implies that $f(y) - f(x) \geq \langle y - x, v \rangle$ for all $y \in \mathbb{R}^n$ and hence

$$|\nabla f(x)| = \limsup_{y \rightarrow x} \frac{|f(y) - f(x)|}{|y - x|} \geq \limsup_{y \rightarrow x} \frac{\langle y - x, v \rangle}{|y - x|} = v.$$

Thus, $|v| \leq L$, so that $g(x) \geq l_{a,v}(x) = f(x)$. \square

7 Optimality

Here we show that the logarithmic dependence in $\mu(A)$ in Theorems 1.1 and 1.3 is optimal, up to the universal constant c . We provide several examples.

Example 1. Let us return to Example 4, Section 5, of the discrete hypercube $M = \{0, 1\}^n$, which we equip with the Hamming distance d_n and the uniform measure μ . Let us test the inequality (5.1) of Corollary 5.4 on the set $A \subset \{-1, 1\}^n$ consisting of $n + 1$ points

$$(0, 0, 0, \dots, 0), (1, 0, 0, \dots, 0), (1, 1, 0, \dots, 0), \dots, (1, 1, 1, \dots, 1).$$

We have $\mu(A) = (n + 1)/2^n \geq 1/2^n$. The function $f : A \rightarrow \mathbb{R}$, defined by

$$f(x) = \#\{i : x_i = 1\} - \frac{n}{2},$$

has a Lipschitz semi-norm $\|f\|_{\text{Lip}} \leq 1$ with respect to d and the μ_A -mean zero. Moreover, $\int f^2 d\mu_A = \frac{n(n+2)}{12}$. Expanding the inequality $\int e^{tf} d\mu_A \leq e^{\sigma^2(\mu_A) t^2/2}$ at the origin yields $\int f^2 d\mu_A \leq \sigma^2(\mu_A)$. Hence, recalling that $\sigma^2(\mu) \leq \frac{n}{4}$, we get

$$\begin{aligned} \sigma^2(\mu_A) &\geq \int f^2 d\mu_A \geq \frac{n^2}{12} \\ &\geq \frac{n}{3} \sigma^2(\mu) \geq \frac{1}{3 \log 2} \sigma^2(\mu) \log \left(\frac{1}{\mu(A)} \right). \end{aligned}$$

This example shows the optimality of (5.1) in the regime $\mu(A) \rightarrow 0$.

Example 2. Let γ_n be the standard Gaussian measure on \mathbb{R}^n of dimension $n \geq 2$. We have $\sigma^2(\gamma_n) = 1$. Consider the normalized measure γ_{A_R} on the set

$$A_R = \{(x_1, x_2, \dots, x_n) \in \mathbb{R}^n : x_1^2 + x_2^2 \geq R^2\}, \quad R \geq 0.$$

Using the property that the function $\frac{1}{2}(x_1^2 + x_2^2)$ has a standard exponential distribution under the measure γ_n , we find that $\gamma_n(A_R) = e^{-R^2/2}$. Moreover,

$$\begin{aligned} s^2(\gamma_{A_R}) \geq \text{Var}_{\gamma_{A_R}}(x_1) &= \int x_1^2 d\gamma_{A_R}(x) = \frac{1}{2} \int (x_1^2 + x_2^2) d\gamma_{A_R}(x) \\ &= \frac{1}{e^{-R^2/2}} \int_{R^2/2}^{\infty} r e^{-r} dr = \frac{R^2}{2} + 1 = \log \left(\frac{e}{\gamma_n(A_R)} \right). \end{aligned}$$

Therefore,

$$\sigma^2(\gamma_{A_R}) \geq s^2(\gamma_{A_R}) \geq \log\left(\frac{e}{\gamma_n(A_R)}\right),$$

showing that the inequality (1.4) of Theorem 1.1 is optimal, up to the universal constant, for any value of $\gamma_n(A) \in [0, 1]$.

Example 3. A similar conclusion can be made about the uniform probability measure μ on the Euclidean ball $B(0, \sqrt{n})$ of radius \sqrt{n} , centred at the origin (asymptotically for growing dimension n). To see this, it is sufficient to consider the cylinders

$$A_\varepsilon = \{(x_1, y) \in \mathbb{R} \times \mathbb{R}^{n-1} : |x_1| \leq \sqrt{n - \varepsilon^2} \text{ and } |y| \leq \varepsilon\}, \quad 0 < \varepsilon \leq \sqrt{n},$$

and the function $f(x) = x_1$. We leave to the readers the corresponding computations.

Example 4. Let μ be the two-sided exponential measure on \mathbb{R} with density $\frac{1}{2}e^{-|x|}$. In this case $\sigma^2(\mu) = \infty$, but, as easy to see, $2 \leq s^2(\mu) \leq 4$ (recall that $\lambda_1(\mu) = \frac{1}{4}$). We are going to test optimality of the inequality (1.7) on the sets $A_R = \{x \in \mathbb{R} : |x| \geq R\}$ ($R \geq 0$). Clearly, $\mu(A_R) = e^{-R}$, and we find that

$$\begin{aligned} s^2(\mu_{A_R}) \geq \text{Var}_{\mu_{A_R}}(x) &= \int_{-\infty}^{\infty} x^2 d\mu_{A_R}(x) = \frac{1}{e^{-R}} \int_R^{\infty} r^2 e^{-r} dr \\ &= R^2 + 2R + 2 \geq (R + 1)^2 = \log^2\left(\frac{e}{\mu(A_R)}\right). \end{aligned}$$

Therefore,

$$s^2(\mu_{A_R}) \geq \log^2\left(\frac{e}{\mu(A_R)}\right),$$

showing that the inequality (1.7) is optimal, up to the universal constant, for any value of $\mu(A) \in (0, 1]$.

Acknowledgment. The authors gratefully acknowledge the support and hospitality of the Institute for Mathematics & Applications (IMA), and the University of Minnesota, Minneapolis, where much of this work was conducted. The second named author would like to acknowledge the hospitality of the Georgia Institute of Technology, Atlanta, during the period 02/8-13/2015.

References

- [A-S] Aida, S.; Strook, D. Moment estimates derived from Poincaré and logarithmic Sobolev inequalities. *Math. Research Letters* 1 (1994), 75–86.
- [A-B-S] Alon, N; Boppana, R; Spencer, J. An asymptotic isoperimetric inequality. *Geometric and Functional Anal.* 8 (1998), 411–436.
- [B-L] Bakry, D.; Ledoux, M. Lévy–Gromov’s isoperimetric inequality for an infinite dimensional diffusion generator. *Invent. Math.* 123 (1996), 259–281.
- [B-M] Barthe, F.; Milman, E. Transference principles for log-Sobolev and spectral-gap with applications to conservative spin systems. *Comm. Math. Phys.* 323 (2013), no. 2, 575–625.
- [B1] Bobkov, S. G. Localization proof of the isoperimetric Bakry-Ledoux inequality and some applications. *Teor. Veroyatnost. i Primenen.* 47 (2002), no. 2, 340–346. Translation in: *Theory Probab. Appl.* 47 (2003), no. 2, 308–314.
- [B2] Bobkov, S. G. Perturbations in the Gaussian isoperimetric inequality. *J. Math. Sciences (New York)*, 166 (2010), no. 3, 225–238. Translated from: *Problems in Math. Analysis* 45 (2010), 3–14.

- [B-G] Bobkov, S. G.; Götze, F. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* 163 (1999), no. 1, pp. 1–28.
- [B-G-H] Bobkov, S. G.; Houdré, C.; Tetali, P. The subgaussian constant and concentration inequalities. *Israel J. Math.* 156 (2006), 255–283.
- [D-M] Ding, J.; Mossel, E. Mixing under monotone censoring. *Electronic Comm. Probab.* 19 (2014), 1–6.
- [K] Kirszbraun, M. D. Über die zusammenziehende und Lipschitzsche Transformationen. *Fund. Math.* 22 (1934) 77–108.
- [L] Ledoux, M. The concentration of measure phenomenon. *Mathematical Surveys and Monographs* 89 (2001), Amer. Math. Soc., Providence, RI.
- [Mar] Marton, K. Bounding \bar{d} -distance by informational divergence: a method to prove measure concentration, *Ann. Probab.* 24 (1996), no. 2, 857–866.
- [M] Maurey, B. Some deviation inequalities. *Geom. Funct. Anal.* 1 (1991), 188–197.
- [MS] McShane, E. J. Extension of range of functions. *Bull. Amer. Math. Soc.* 40 (1934), no. 12, 837–842.
- [Mi1] Milman, E. On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.* 177 (2009), no. 1, 1–43.
- [Mi2] Milman, E. Properties of isoperimetric, functional and transport-entropy inequalities via concentration. *Probab. Theory Related Fields* 152 (2012), no. 3–4, 475–507.
- [M-W] Mueller, C. E.; Weissler, F. B. Hypercontractivity for the heat semigroup for ultraspherical polynomials and on the n -sphere. *J. Funct. Anal.* 48 (1992), 252–283.
- [T1] Talagrand, M. An isoperimetric theorem on the cube and the Khinchine–Kahane inequalities. *Proc. Amer. Math. Soc.* 104 (1988), 905–909.
- [T2] Talagrand, M. Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. I.H.E.S.* 81 (1995), 73–205.

Appendix

Proof of Lemma 2.1. Using the homogeneity, in order to derive the right-hand side inequality in (2.1), we may assume that $\sup_{p \geq 1} \frac{\|f\|_p}{\sqrt{p}} \leq 1$. Then $\int |f|^p d\mu \leq p^{p/2}$ for all $p \geq 1$, and by Chebyshev’s inequality,

$$1 - F(t) \equiv \mu\{|f| \geq t\} \leq \left(\frac{\sqrt{p}}{t}\right)^p, \quad \text{for all } t > 0.$$

If $t \geq 2$, choose here $p = \frac{1}{4} t^2$, in which case $1 - F(t) \leq 2^{-\frac{1}{4} t^2}$. Integrating by parts, we have, for any $0 < \varepsilon < \frac{\log 2}{4}$,

$$\begin{aligned} \int e^{\varepsilon f^2} d\mu &= - \int_0^\infty e^{\varepsilon t^2} d(1 - F(t)) \\ &= 1 + 2\varepsilon \int_0^2 t e^{\varepsilon t^2} (1 - F(t)) dt + 2\varepsilon \int_2^\infty t e^{\varepsilon t^2} (1 - F(t)) dt \\ &\leq 1 + 2\varepsilon \int_0^2 t e^{\varepsilon t^2} dt + 2\varepsilon \int_2^\infty t e^{\varepsilon t^2} e^{-\frac{\log 2}{4} t^2} dt \\ &= e^{4\varepsilon} + \frac{\varepsilon}{\frac{\log 2}{4} - \varepsilon} e^{-(\log 2 - 4\varepsilon)} = e^{4\varepsilon} \left(1 + \frac{\varepsilon}{2(\frac{\log 2}{4} - \varepsilon)}\right). \end{aligned}$$

If $\varepsilon \leq \frac{\log 2}{8}$, the latter expression does not exceed $\frac{3}{2} e^{4\varepsilon}$ which does not exceed 2 for $\varepsilon \leq \frac{\log(4/3)}{4}$. Both inequalities are fulfilled for $\varepsilon = \frac{\log 2}{10}$, and with this value $\int e^{\varepsilon f^2} d\mu \leq 2$. Hence

$$\|f\|_{L^{\psi_2}(\mu)} \leq \frac{1}{\sqrt{\varepsilon}} = \sqrt{\frac{10}{\log 2}} < 4,$$

which yields the right inequality in (2.1). Conversely, if $\|f\|_{L^{\psi_2}(\mu)} = 1$, then $\int e^{\frac{3}{4}f^2} d\mu \leq 2$. Since $u(t) = t^p e^{-\frac{3}{4}t^2}$ is maximized in $t > 0$ at $t_0 = \sqrt{\frac{2p}{3}}$, we get

$$\|f\|_p^p = \int u(|f|) e^{f^2} d\mu \leq u(t_0) \cdot 2 = 2 \left(\sqrt{\frac{2p}{3e}} \right)^p.$$

Hence, $\frac{\|f\|_p}{\sqrt{p}} \leq 2^{1/p} \sqrt{\frac{2}{3e}} < 1$, which yields the left inequality.

Now, let us turn to (2.2) and assume that $\sup_{p \geq 1} \frac{\|f\|_p}{p} = 1$. Then $\int |f|^p d\mu \leq p^p$ for all $p \geq 1$, and by Chebyshev's inequality, for all $t > 0$,

$$1 - F(t) \equiv \mu\{|f| \geq t\} \leq \left(\frac{p}{t}\right)^p.$$

If $t \geq 2$, we may choose here $p = \frac{1}{2}t$ in which case $1 - F(t) \leq 2^{-\frac{1}{2}t}$, while for $1 \leq t < 2$ we choose $p = 1$, so that $1 - F(t) \leq \frac{1}{t}$. Arguing as before, we have, for any $0 < \varepsilon < \frac{\log 2}{2}$,

$$\begin{aligned} \int e^{\varepsilon|f|} d\mu &= 1 + \varepsilon \int_0^1 e^{\varepsilon t} (1 - F(t)) dt + \varepsilon \int_1^2 e^{\varepsilon t} (1 - F(t)) dt + \varepsilon \int_2^\infty e^{\varepsilon t} (1 - F(t)) dt \\ &\leq 1 + \varepsilon \int_0^1 e^{\varepsilon t} dt + \varepsilon \int_1^2 \frac{e^{\varepsilon t}}{t} dt + \varepsilon \int_2^\infty e^{\varepsilon t} e^{-\frac{\log 2}{2}t} dt. \end{aligned}$$

The pre-last integral can be bounded by $\int_1^2 \frac{e^{2\varepsilon}}{t} dt = e^{2\varepsilon} \log 2$, so

$$\int e^{\varepsilon|f|} d\mu \leq e^\varepsilon + \varepsilon e^{2\varepsilon} \log 2 + \frac{\varepsilon}{\frac{\log 2}{2} - \varepsilon} e^{-2(\frac{\log 2}{2} - \varepsilon)}.$$

For $\varepsilon = \frac{1}{6}$, the latter expression is equal to 1.98903902..., and thus $\int e^{\varepsilon|f|} d\mu < 2$. Hence

$$\|f\|_{L^{\psi_1}(\mu)} \leq \frac{1}{\varepsilon} = 6.$$

Conversely, if $\|f\|_{L^{\psi_1}(\mu)} = 1$, then $\int e^{\frac{3}{4}|f|} d\mu = 2$. Since $u(t) = t^p e^{-\frac{3}{4}t}$ is maximized at $t_0 = p$, we get

$$\|f\|_p^p = \int u(|f|) e^{\frac{3}{4}|f|} d\mu \leq u(t_0) \cdot 2 = 2 \left(\frac{4p}{3e} \right)^p.$$

Hence, $\frac{\|f\|_p}{p} \leq 2^{1/p} \frac{4}{3e} < 1$, which yields the left inequality. \square

Proof of Lemma 3.1. First assume that $\|f\|_{\psi_2} = 1$, in particular $\int e^{\frac{7}{8}f^2} d\mu \leq 2$. The function

$$u(t) = \log \int e^{tf} d\mu$$

is smooth, convex, with $u(0) = 0$ and

$$u'(t) = \frac{\int f e^{tf} d\mu}{\int e^{tf} d\mu}.$$

In particular, $u'(0) = 0$. Note that, by Jensen's inequality, $\int e^{tf} d\mu \geq 1$, so $u(t) \geq 0$. Further differentiation gives

$$u''(t) = \frac{\int f^2 e^{tf} d\mu - (\int f e^{tf} d\mu)^2}{(\int e^{tf} d\mu)^2} \leq \int f^2 e^{tf} d\mu.$$

Using $tf \leq \frac{t^2+f^2}{2}$ and the elementary inequality $x e^{-\frac{3}{8}x} \leq \frac{8}{3}e^{-1}$, we get, for $|t| \leq 1$,

$$\begin{aligned} \int f^2 e^{tf} d\mu &\leq \int f^2 e^{\frac{t^2+f^2}{2}} d\mu \\ &= e^{t^2/2} \int f^2 e^{-\frac{3}{8}f^2} e^{\frac{7}{8}f^2} d\mu \leq e^{1/2} \frac{8}{3} e^{-1} \int e^{\frac{7}{8}f^2} d\mu \leq 4. \end{aligned}$$

Thus, $u''(t) \leq 4$, and by Taylor's formula, $u(t) \leq 2t^2$.

On the hand, for $|t| \geq 1$, by Cauchy's inequality,

$$\begin{aligned} \int e^{tf} d\mu &\leq \int e^{\frac{t^2+f^2}{2}} d\mu = e^{t^2/2} \int e^{f^2/2} d\mu \\ &\leq e^{t^2/2} \left(\int e^{\frac{7}{8}f^2} d\mu \right)^{4/7} = 2^{4/7} e^{t^2/2} \leq e^{(\frac{1}{2} + \frac{4}{7} \log 2)t^2} \leq e^{t^2}. \end{aligned}$$

Hence, in this case $u(t) \leq t^2$. Thus,

$$\sigma_f^2 = \sup_{t \neq 0} \frac{u(t)}{t^2/2} \leq 4,$$

proving the right inequality of Lemma 3.1.

For the left inequality, let $\sigma_f^2 = 1$. Then $\int e^{tf} d\mu \leq e^{t^2/2}$ for all $t \in \mathbb{R}$, which implies

$$1 - F(t) \equiv \mu\{|f| \geq t\} \leq 2e^{-t^2/2}, \quad t \geq 0.$$

From this, integrating by parts, we have, for any $0 < \varepsilon < \frac{1}{2}$,

$$\begin{aligned} \int e^{\varepsilon f^2} d\mu &= \int_0^\infty e^{\varepsilon t^2} dF(t) = - \int_0^\infty e^{\varepsilon t^2} d(1 - F(t)) \\ &= 1 + 2\varepsilon \int_0^\infty t e^{\varepsilon t^2} (1 - F(t)) dt \\ &\leq 1 + 4\varepsilon \int_0^\infty t e^{\varepsilon t^2} e^{-t^2/2} dt = 1 + \frac{2\varepsilon}{\frac{1}{2} - \varepsilon}. \end{aligned}$$

The last expression is equal to 2 for $\varepsilon = \frac{1}{6}$, which means that $\|f\|_{\psi_2} \leq \sqrt{6}$. \square