Journal of Functional Analysis **163**, 1–28 (1999)
Article ID jfan.1998.3326, available online at http://www.idealibrary.com on IDEAL®

Exponential Integrability and Transportation Cost Related to Logarithmic Sobolev Inequalities

S. G. Bobkov*

Department of Mathematics, Syktyvkar University, 167001 Syktyvkar, Russia and Department of Mathematics, Bielefeld University, 33501 Bielefeld, Germany

and

F. Götze

Department of Mathematics, Bielefeld University, 33501 Bielefeld, Germany

We study some problems on exponential integrability, concentration of measure, and transportation cost related to logarithmic Sobolev inequalities. On the real line, we then give a characterization of those probability measures which satisfy these inequalities © 1999 Academic Press

Key Words: logarithmic Sobolev inequalities; exponential integrability; concentration of measure; transportation inequalities.

1. INTRODUCTION

Logarithmic Sobolev inequalities are an essential tool in the study of various problems (cf., e.g., [G2], [D-SC], [L3]). The main purpose of this paper is to refine some known connections between logarithmic Sobolev inequalities, exponential integrability of "smooth" functions and the concentration of measure. Provided a logarithmic Sobolev inequality holds, we estimate exponential moments of functions in terms of the "modulus of their gradient" and deduce a transportation inequality with a corresponding concentration inequality. Furthermore, we shall describe all probability measures satisfying logarithmic Sobolev inequalities on the real line.

Let us describe the general scheme of logarithmic Sobolev inequalities. Let (Ω, μ) denote a probability space, and assume there is an operator Γ defined on a set $\mathscr{A}$ of bounded measurable functions on Ω with the following properties:

* Research supported in part by Alexander von Humboldt Foundation.

0022-1236/99 $30.00

(1) for any $f \in \mathscr{A}$, $\Gamma(f)$ is a non-negative measurable function on Ω;
(2) for all $f \in \mathscr{A}$ and for all $a \in \mathbf{R}$, $b \geqslant 0$, $a + bf \in \mathscr{A}$ and $\Gamma(a+bf) = b\Gamma(f)$.

Introduce the entropy functional:

$$\operatorname{Ent}(g) \equiv \int g \log g \, d\mu - \int g \, d\mu \log \int g \, d\mu$$

$$= \int g \log \frac{g}{\int g \, d\mu} \, d\mu, \qquad g \geqslant 0.$$

We will say that (Ω, μ, Γ) satisfies a logarithmic Sobolev inequality with constant $c \geqslant 0$ (for short, LSI_c) if, for all $f \in \mathscr{A}$,

$$\operatorname{Ent}(e^f) \leqslant \frac{c}{2} \int \Gamma(f)^2 \, e^f \, d\mu. \tag{1.1}$$

If Γ is a derivation, i.e., if for all $f \in \mathscr{A}$ with values in (a, b) and $u \in C^\infty[a, b]$, we have $u(f) \in \mathscr{A}$ and $\Gamma(u(f)) = |u'(f)| \, \Gamma(f)$, then (1.1) is equivalent to the more familiar version

$$\operatorname{Ent}(g^2) \leqslant 2c \int \Gamma(g)^2 \, d\mu, \tag{1.2}$$

whenever $g \in \mathscr{A}$ satisfies $\inf g > 0$. To be more precise, (1.1) represents an exponential (or, modified) form of (1.2) which is usually called logarithmic Sobolev inequality (the normalization $\frac{1}{2}$ in (1.1) is chosen such that $c = 1$ in the Gaussian case). In the most interesting cases of derivations Γ, this operator and the inequality (1.2) itself (as well as (1.1)) are easily extended to a larger class of functions including unbounded "smooth" functions. For example, when Ω is a metric space with metric d, one may consider the following natural generalization of the modulus of the usual gradient:

$$\Gamma(f)(x) = |\nabla f(x)| = \limsup_{d(x, y) \to 0+} \frac{|f(x) - f(y)|}{d(x, y)}.$$

In this case, we may define Γ on the class of all Lipschitz functions f (i.e., with $\|f\|_{\mathrm{Lip}} < \infty$), and the inequalities (1.1) and (1.2) hold for this class if they hold for all bounded Lipschitz functions. Thus, for the Euclidean space $\Omega = \mathbf{R}^n$ with the usual Euclidean metric, we arrive at the usual definition of logarithmic Sobolev inequalities.

One of the questions of interest is to determine whether or not, a given probability measure μ satisfies LSI_c with some finite c. In the case of the real line $\Omega = \mathbf{R}$, we provide the following characterization. Let $F(x) = \mu((-\infty, x])$, $x \in \mathbf{R}$, denote the distribution function of μ, and let p be the density of the absolutely continuous part of μ with respect to Lebesgue measure. Let m denote a median of μ. Set

$$D_0 = \sup_{x<m} \left(F(x) \log \frac{1}{F(x)} \right) \int_x^m \frac{1}{p(t)}\, dt,$$

$$D_1 = \sup_{x>m} \left((1-F(x)) \log \frac{1}{1-F(x)} \right) \int_m^x \frac{1}{p(t)}\, dt,$$

defining D_0 and D_1 to be zero in case $\mu((-\infty, m)) = 0$ or $\mu((m, +\infty)) = 0$, respectively.

THEOREM 1.1. *Let μ be an arbitrary probability measure on $\mathbf{R}$. For some constant c, the log-Sobolev inequality* $\mathrm{Ent}(g^2) \leqslant 2c \int |g'|^2\, d\mu$ *holds in the class of all smooth functions g on $\mathbf{R}$ if and only if $D_0 + D_1 < +\infty$. In this case, the optimal value of c satisfies*

$$K_0(D_0 + D_1) \leqslant c \leqslant K_1(D_0 + D_1),$$

where K_0 and K_1 are certain absolute positive constants.

The proof of Theorem 1.1 uses a result of M. Artola, G. Talenti, and G. Tomaselli on Hardy-type inequalities with weights on the real line (cf. [Mu]) and is given in Sections 4 and 5 (cf. Theorem 5.3). In Section 4, we transform logarithmic Sobolev inequalities (1.2), up to an absolute multiplicative constant, into Sobolev-type inequalities $\|f - \int f\, d\mu\|_N^2 \leqslant c \int \Gamma(f)^2\, d\mu$ in a suitable Orlicz space $L_N(\Omega, \mu)$ with Orlicz function $N(x) = x^2 \log(1 + x^2)$. On the real line this allows to reduce LSI_c to a Hardy-type inequality (cf. Propositions 4.1 and 4.2).

In Section 2 we study in a general setting exponential moments of a function $f \in \mathscr{A}$ in terms of the distribution of $\Gamma(f)$ under μ provided LSI_c holds. As a main result (cf. Theorem 2.1), we prove, in particular:

THEOREM 1.2. *For any function $f \in \mathscr{A}$ with $\int f\, d\mu = 0$,*

$$\int e^f\, d\mu \leqslant \int e^{c\Gamma(f)^2}\, d\mu. \tag{1.3}$$

This estimate implies inequalities for exponential moments of "Lipschitz" functions similar to those proved by S. Aida, T. Masuda and I. Shigekawa [A-M-S] and M. Ledoux [L1], [L2]. In the Gaussian case, (1.3) improves an exponential inequality due to G. Pisier [P1]. In the case of the discrete cube, it affirmatively answers his question (p. 182) on the validity of the discrete analogue of the Gaussian variant of (1.3). These examples and applications will be discussed in more detail in Corollaries 2.2–2.4.

In Section 3, we study for a metric space (Ω, d) equipped with a probability measure μ inequalities of the following type:

$$W_\alpha(\mu, \nu) \leqslant \sqrt{2c \int \log \frac{d\nu}{d\mu}\, d\nu}. \tag{1.4}$$

Here $W_\alpha(\mu, \nu)$ denotes the Kantorovich–Rubinstein distance between μ and ν (we shall consider the cases $\alpha = 1$ and $\alpha = 2$, only) which is defined as the infimum of

$$\left(\iint d(x, y)^\alpha \, d\pi(x, y)\right)^{1/\alpha}$$

over all probability measures π on the product space $\Omega \times \Omega$ with marginal distributions μ and ν (ν is an arbitrary probability measure on Ω which is absolutely continuous with respect to μ with density $d\nu/d\mu$). Thus, (1.4) relates the minimal transportation cost needed to transport ν into μ to the so-called informational divergence $D(\nu \parallel \mu) = \int \log(d\nu/d\mu)\, d\nu$. Such transportation inequalities have been introduced for $\alpha = 1$ by K. Marton [Ma1]. On the basis of (1.4), she established concentration inequalities for discrete product measures (and, furthermore, for distributions of certain Markov processes, [Ma2]), of the form

$$1 - \mu(A^h) \leqslant \exp\left(-\frac{1}{2c}\left(h - \sqrt{2c \log \frac{1}{\mu(A)}}\right)^2\right), \tag{1.5}$$

where A^h denotes h-neighborhood of a set $A \subset \Omega$. This approach has been studied as well by M. Talagrand in [T4] proving, in particular, that the canonical Gaussian measure $\mu = \gamma_n$ on the Euclidean space satisfies an inequality (1.4) for the W_2-metric:

$$W_2(\gamma_n, \nu) \leqslant \sqrt{2 \int \log \frac{d\nu}{d\gamma_n}\, d\nu}. \tag{1.6}$$

As a consequence, M. Talagrand derives from (1.6) using Marton's line of arguments, a sharp Gaussian concentration inequality (1.5) with $c = 1$ (which is similar to an isoperimetric inequality).

A natural question arising in connection with (1.4) is to find an appropriate functional form for (1.4) and to relate it to other classes of inequalities. When $\alpha = 1$, we shall prove:

THEOREM 1.3. *Assume* $\int d(x, x_0)\, d\mu(x) < +\infty$, *for some* $x_0 \in \Omega$ (*where* d *denotes the metric on* Ω). *The transportation inequality* (1.4) *holds for all probability measures* ν *on* Ω *which are absolutely continuous with respect to* μ *if and only if, for all* $t \in \mathbf{R}$ *and all functions* f *on* Ω *with* $\int f\, d\mu = 0$ *and* $\|f\|_{\mathrm{Lip}} \leqslant 1$,

$$\int e^{tf}\, d\mu \leqslant e^{ct^2/2}.$$

In particular, (1.4) *together with the concentration inequality* (1.5) *always holds assuming the* $\mathrm{LSI_c}$*-property.*

Such a functional description shows as well that, in turn, (1.5) implies (1.4) although with a worse constant. Thus, (1.5) and (1.4) are essentially equivalent. However, we do not know, whether or not (1.4) with $\alpha > 1$ is implied by a logarithmic Sobolev inequality. A main argument we use in the case $\alpha = 1$ is the well known Kantorovich–Rubinstein's theorem representing the metric W_1 in terms of Lipshitz functions on Ω, and it seems there is no such "Lipshitz" representation for W_α. The question of a functional representation for metrics of Kantorovich–Rubinstein-type (raised by R. M. Dudley) was open for some time until V. L. Levin and S. T. Rachev proved in particular that ([Lev], [Ra])

$$W_2^2(\nu, \mu) = \sup \int g\, d\nu - \int d\mu,$$

where the supremum is taken over all pairs of bounded continuous functions (g, f) such that $g(y) - f(x) \leqslant d(x, y)^2$, for all $x, y \in \Omega$. This functional description allows one to give an equivalent representation for the transportation inequality (1.4) with $\alpha = 2$ (and with $c = 1$) by a relation between the distribution of an arbitrary function f and the distribution of the function $(Sf)(x) = \inf \{ f(y) + \frac{1}{2} d(x, y)^2 : y \in \Omega \}$. Namely, (1.4) is equivalent to the inequality

$$\int e^{Sf}\, d\mu \leqslant e^{\int f\, d\mu}. \tag{1.7}$$

Inequalities of this type were introduced by B. Maurey [Mau] as a functional approach to some of M. Talagrand's isoperimetric inequalities ([T1], [T2]). They are now referred to as inf-convolution inequalities. Using this approach, we shall give a simple alternative proof of Talagrand's transportation inequality (1.6) via (1.7) with $\mu = \gamma_n$. In this case, the inequality (1.7) may be viewed as a generalization of a Tsirel'son exponential inequality for the supremum of a bounded Gaussian process.

2. EXPONENTIAL INTEGRABILITY

First we emphasize some particular cases and applications of:

THEOREM 2.1. *Assume* (Ω, μ, Γ) *satisfies* LSI_c. *Then, for any* $f \in \mathscr{A}$ *with mean* $\int f\, d\mu = 0$, *and for all* $\alpha > c/2$,

$$\int e^f\, d\mu \leqslant \left(\int e^{\alpha \Gamma(f)^2}\, d\mu\right)^{c/(2\alpha - c)}. \tag{2.1}$$

In addition, for all $f \in \mathscr{A}$ *and* $\lambda \in (0, 1]$,

$$\int e^f\, d\mu \leqslant \left(\int e^{\lambda f + (1-\lambda)\, c\Gamma(f)^2/2}\, d\mu\right)^{1/\lambda}. \tag{2.2}$$

For $\lambda = 1$, there is an equality in (2.2), and comparing the derivatives of the both sides at $\lambda = 1$, we obtain exactly (1.1). Hence (2.2) is another version of the LSI_c-property. In this sense, (2.2) is stronger than (2.1).

Assuming that $\Gamma(f) \leqslant 1$ μ-a.e. and applying (2.1) to functions tf with $\alpha \to +\infty$, we get, for all $t \geqslant 0$,

$$\int e^{tf}\, d\mu \leqslant e^{ct^2/2}. \tag{2.3}$$

When Γ is a derivation, one can also apply (2.1) to functions $t(f^2 - \int f^2)$ with $0 \leqslant t < 1/(4c)$. Then $\Gamma(f^2 - \int f^2) = 2\,|f|\, \Gamma(f) \leqslant 2\,|f|$, so that by (2.1) with $\alpha = 1/(4t)$,

$$\int e^{tf^2}\, d\mu \leqslant \exp\left(\frac{t(1-2ct)}{1-4ct} \int f^2\, d\mu\right).$$

This is not as sharp as (2.3) and can be improved by virtue of (2.2). In the same way, applying (2.2) to $sf^2/(2c)$ (instead of f) with $0 < s < 1$ and $\lambda = (p-s)/(1-s)$, where $p \in (s, 1]$ is arbitrary, we get

$$\int e^{sf^2/(2c)}\, d\mu \leqslant \left(\int e^{psf^2/(2c)}\, d\mu\right)^{(1-s)/(p-s)}.$$

This inequality becomes an equality for $p = 1$, so comparing the derivatives of the logarithm of the both sides at $p = 1$ one gets

$$0 \geqslant \frac{1}{1-s}\left[\int \frac{sf^2}{2c}\, e^{sf^2/(2c)}(1-s) - \int e^{sf^2/(2c)} \log \int e^{sf^2/(2c)}\right].$$

This means that the function $u(s)=\log \int e^{sf^2/(2c)}\,d\mu$ satisfies the differential inequality $u'(s)/u(s) \leqslant 1/(s(1-s))$, which expresses the fact that the function

$$v(s)=\exp\left\{\frac{1-s}{s}\,u(s)\right\}=\left(\int e^{sf^2/(2c)}\right)^{(1-s)/s}$$

does not increase in $0<s<1$. Comparing $v(s)$ with $v(0^+)$ and introducing $t=s/(2c)$, we arrive at

$$\int e^{tf^2}\,d\mu \leqslant \exp\left(\frac{t}{1-2ct}\int f^2\,d\mu\right), \qquad 0<t<\frac{1}{2c}. \tag{2.4}$$

The inequality (2.3) has been shown to follow from (1.1) by M. Ledoux [L1], [L2], and (2.4) was deduced from (1.1) by. S. Aida, T. Masuda and I. Shigekawa [A-M-S], see also [A-S] (according to [D-S], the original idea goes back to I. Herbst). The above deduction of (2.4) from (2.2) essentially repeates an argument of [A-M-S]. As for the proof of (2.1), we develop an argument of M. Ledoux which was based on a another differential inequality. Below we will also show that (2.3) is implied by (2.4) (this requires however that Γ is a derivation; cf. Remark 2.5).

For $\alpha=c$, (2.1) becomes

$$\int e^f\,d\mu \leqslant \int e^{c\Gamma(f)^2}\,d\mu. \tag{2.5}$$

This inequality seems to be new except for the case of Gaussian measures where it had a different constant in the exponent. More precisely, let $(\Omega,\mu)=(\mathbf{R}^n,\gamma_n)$ be the Euclidean space equipped with the canonical Gaussian measure with density $(2\pi)^{-n/2}e^{-|x|^2/2}$, and let $\Gamma(f)=|\nabla f|$ denote the length of the usual gradient of a smooth function f. Given a convex function Φ on $\mathbf{R}$, G. Pisier [P1] proved that, for smooth functions f with $\int f\,d\gamma_n=0$.

$$\int_{\mathbf{R}^n}\Phi(f)\,d\gamma_n \leqslant \int_{\mathbf{R}^n}\int_{\mathbf{R}^n}\Phi\left(\frac{\pi}{2}\langle\nabla f(x),y\rangle\right)d\gamma_n(x)\,d\gamma_n(y),$$

where the constant $\pi/2$ appears to be optimal for the choice $\Phi(x)=|x|$. For $\Phi(x)=e^x$, this inequality yields

$$\int_{\mathbf{R}^n} e^f\,d\gamma_n \leqslant \int_{\mathbf{R}^n}\exp\left(\frac{\pi^2}{8}|\nabla f|^2\right)d\gamma_n. \tag{2.6}$$

On the other hand, by Gross' logarithmic inequality [G1], (1.2) holds for γ_n with $c=1$. Therefore, according to (2.5), the constant in the exponent can be somewhat improved:

COROLLARY 2.2. *For any integrable smooth function f on $\mathbf{R}^n$ with $\int f\,d\gamma_n=0$,*

$$\int_{\mathbf{R}^n} e^f\,d\gamma_n \leqslant \int_{\mathbf{R}^n} e^{|\nabla f|^2}\,d\gamma_n.$$

From Theorem 2.1, we also get $\int_{\mathbf{R}^n} e^{|f|}\,d\gamma_n < +\infty$, whenever $\int_{\mathbf{R}^n} e^{\alpha|\nabla f|^2}\,d\gamma_n < +\infty$, for some $\alpha > \frac{1}{2}$. Clearly, the condition $\alpha \geqslant \frac{1}{2}$ is necessary to prove this claim. Thus, it is natural to ask whether or not, even for one dimension, the above implication holds for $\alpha = \frac{1}{2}$. This turns out to be the case (an observation due to M. Talagrand [T5]).

Consider another important example. Let μ be an arbitrary product probability measure on the cube $\Omega=[-1,1]^n$. Recently, M. Ledoux [L3] established for such a measure, in the spirit of some of Talagrand's concentration inequalies ([T1], [T3]), a logarithmic Sobolev inequality (1.1) for the class $\mathscr{A}$ of all convex smooth functions on $[-1,1]^n$. He also determined the optimal constant as $c=4$. Together with (2.5) this yields:

COROLLARY 2.3. *For any smooth convex function f on $[-1,1]^n$ with $\int f\,d\gamma_n=0$,*

$$\int_{[-1,1]^n} e^f\,d\mu \leqslant \int_{[-1,1]^n} e^{4\,|\nabla f|^2}\,d\mu. \tag{2.7}$$

Note that, in such a situation, the inequality (1.1) may not follow from (1.2) since the latter does not hold with a universal constant even in dimension one for the class of all convex functions.

Let us now specialize to the canonical Bernoulli (i.e., normalizing counting) measure μ_n on the discrete cube $\{-1,1\}^n$. In this case, (1.1) holds true for convex functions with optimal constant $c=2$ (cf. [B]). This can easily be shown using the Gross' discrete logarithmic inequality [G1]. Thus the constant 4 in the exponent in (2.7) can be replaced by 2. Now we are able to state a discrete version of Pisier's inequality (2.6). Take for $\Gamma(f)$ the length of a discrete gradient of f:

$$[\mathrm{D}(f)(x)]^2 = \sum_{k=1}^{n} \left|\frac{f(s_i(x))-f(x)}{2}\right|^2.$$

Here f is an arbitrary function f on $\{-1,1\}^n$, and $s_i(x)$ denotes the neighbor of the point $x \in \{-1,1\}^n$ on the i-th coordinate: $s_i(x)_j = x_j$, for

$j \neq i$, and $s_i(x)_i = -x_i$. As in the Gaussian case, $(\Omega, \mu, \mathrm{D})$ satisfies (1.2) with $c = 1$ [G1]. It is easily verified, with the same constant (although D is not a derivation), that the triple $(\Omega, \mu, \mathrm{D})$ satisfies (1.1). Thus, we obtain a discrete version of Corollary 2.2:

COROLLARY 2.4. *For any function f on $\{-1, 1\}^n$ with $\int f \, d\mu_n = 0$, we have*

$$\int_{\{-1,1\}^n} e^f \, d\mu_n \leqslant \int_{\{-1,1\}^n} e^{\mathrm{D}(f)^2} \, d\mu_n.$$

Proof of Theorem 2.1. It is well known that the entropy functional can be represented as

$$\mathrm{Ent}(g) = \sup \int gh \, d\mu$$

where the supremum is taken over all functions h with $\int e^h \, d\mu \leqslant 1$. In particular,

$$\int e^h \, d\mu = 1 \qquad \text{implies} \qquad \int gh \, d\mu \leqslant \mathrm{Ent}(g).$$

Put $\beta = \log \int e^{\alpha \Gamma(f)^2} \, d\mu$ so that $\int e^h \, d\mu = 1$ for $h = \alpha \Gamma(f)^2 - \beta$. Hence, for any on-negative measurable function g,

$$\int (\alpha \Gamma(f)^2 - \beta) \; g \, d\mu \leqslant \mathrm{Ent}(g).$$

Take $g = e^f$ so that

$$\alpha \int \Gamma(f)^2 \, e^f \, d\mu - \beta \int e^f \, d\mu \leqslant \mathrm{Ent}(e^f). \tag{2.8}$$

Now, applying (1.1) in order to estimate the first term on left hand side of (2.8), we get

$$\mathrm{Ent}(e^f) \leqslant \frac{\beta c}{2\alpha - c} \int e^f \, d\mu. \tag{2.9}$$

We now apply (2.9) to functions of the form tf, $t \geqslant 0$, $f \in \mathscr{A}$. Hence, put

$$\beta(t) = \log \int e^{\alpha t \Gamma(f)^2} \, d\mu$$

and define the function u by

$$\int e^{tf}\,d\mu = e^{tu(t)}.$$

Since $\mathrm{Ent}(e^{tf}) = t^2 u'(t)\, e^{tu(t)}$, (2.9) will transform into

$$u'(t) \leqslant \frac{c}{2\alpha - c}\,\frac{\beta(t^2)}{t^2}, \qquad t > 0.$$

Note that $\beta(t)$ is a convex function. In addition, $\beta(0) = 0$ and $\beta'(0) \geqslant 0$, hence $\beta(t)$ is non-negative and non-decreasing in $t \geqslant 0$. Consequently, the function $\beta(t)/t$ does not decrease in $t > 0$. Thus, we can conclude that, for $0 \leqslant t \leqslant 1$, $u'(t) \leqslant c\beta(1)/(2\alpha - c) = c\beta/(2\alpha - c)$. Recalling that $\int f\,d\mu = 0$, we also have $u(0) = 0$. As a result,

$$u(1) \leqslant \frac{c}{2\alpha - c}\,\beta, \qquad \text{and} \qquad \int e^{f}\,d\mu = e^{u(1)} \leqslant \exp\left(\frac{c}{2\alpha - c}\,\beta\right).$$

This proves (2.1). To prove (2.2), take $\beta = \log \int e^{\lambda f + (1-\lambda)\, c\Gamma(f)^2/2}\,d\mu$. As in (2.8) we have

$$\int \left(\lambda f + \tfrac{1}{2}(1-\lambda)\, c\Gamma(f)^2 - \beta\right) g\,d\mu \leqslant \mathrm{Ent}(g).$$

Applying (1.1) to this inequality with $g = e^f$ yields the result.

Remark 2.5. Given a function f on the probability space (Ω, μ) and a constant $c > 0$, assume that, for all $a \in \mathbf{R}$ and $0 < t < 1/2c$

$$\int e^{t(f+a)^2}\,d\mu \leqslant \exp\left(\frac{t}{1-2ct}\int (f+a)^2\,d\mu\right).$$

Then, for all $t \in \mathbf{R}$.

$$\int e^{tf}\,d\mu \leqslant e^{t\int f\,d\mu + ct^2/2}.$$

Indeed, set $f_a = f + a$. As in (2.8), for any probability density g (with respect to the measure μ), we have

$$\int \left(tf_a^2 - \frac{t}{1-2ct}\int f_a^2\,d\mu\right) g\,d\mu \leqslant \mathrm{Ent}(g).$$

Optimizing over $t \in [0, 1/2c)$, we arrive at

$$\|f_a\|_{L^2(g\,d\mu)} - \|f_a\|_{L^2(d\mu)} \leqslant \sqrt{2c\,\mathrm{Ent}(g)}.$$

Letting $a \to -\infty$, and also $a \to +\infty$, we get

$$\left|\int fg\,d\mu - \int f\,d\mu\right| \leqslant \sqrt{2c \int g \log g\,d\mu}.$$

This implies (3.8) which is equivalent to (3.6) (cf. Proof of Theorem 3.1 below).

3. TRANSPORTATION INEQUALITIES

Let Ω be a separable metric space with metric d. Given $\alpha \geqslant 1$ and two Borel probability measures μ and ν on Ω, we defined the quantity

$$W_\alpha(\mu, \nu) = \inf\left(\iint d(x, y)^\alpha\,d\pi(x, y)\right)^{1/\alpha},$$

where the infimum is taken over all probability measures π on the product space $\Omega \times \Omega$ with marginal distributions μ and ν. This quantity is commonly refered to as L^α-Wasserstein distance between μ and ν. When $\alpha = 1$, $W_1(\mu, \nu)$ represents the classical Kantorovich–Rubinstein distance. The measure ν will be assumed to be absolutely continuous with respect to μ, with density $d\nu/d\mu$. In this case, the so-called informational divergence of ν with respect to μ,

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

is well defined. As shown by K. Marton [Ma1], [Ma2], suitable upper estimates for $W_1(\mu, \nu)$ in terms of $D(\nu \,\|\, \mu)$ like

$$W_1(\mu, \nu) \leqslant \sqrt{2cD(\nu \,\|\, \mu)} \tag{3.1}$$

turn out to be well adapted to derive from them sharp concentration inequalities for μ (assuming that ν in (3.1) is arbitrary). To start with, let us recall Marton's argument. Given Borel sets A and B of positive μ-measure, define the conditional restrictions of μ by $\mu_A(C) = \mu(C \cap A)/\mu(A)$, $\mu_B(C) = \mu(C \cap B)/\mu(B)$, for any Borel set C of Ω. By the triangle inequality and by (3.1),

$$W_1(\mu_A, \mu_B) \leqslant W_1(\mu, \mu_A) + W_1(\mu, \mu_B) \leqslant \sqrt{2cD(\mu_A \parallel \mu)} + \sqrt{2cD(\mu_B \parallel \mu)}$$
$$= \sqrt{2c \log \frac{1}{\mu(A)}} + \sqrt{2c \log \frac{1}{\mu(B)}}.$$

On the other hand, all the measures π on $\Omega \times \Omega$ with marginals μ_A and μ_B must be supported in $A \times B$, i.e., $\pi(A \times B) = 1$. Therefore, by the very definition of W_1, we have $W_1(\mu_A, \mu_B) \geqslant d(A, B) = \inf \{d(x, y) : x \in A, y \in B\}$. Hence,

$$d(A, B) \leqslant \sqrt{2c \log \frac{1}{\mu(A)}} + \sqrt{2c \log \frac{1}{\mu(B)}}. \tag{3.2}$$

This is already an isoperimetric-type inequality. Given $h > 0$, let A^h denote the open h-neighborhood of A for d, $A^h = \{x \in \Omega : d(a, x) < h$ for some $a \in A\}$. Taking for B the complement of A^h, one obtains from (3.2) the following equivalent property: For every A and $h \geqslant \sqrt{2c \log(1/\mu(A))}$,

$$1 - \mu(A^h) \leqslant \exp\left(-\frac{1}{2c}\left(h - \sqrt{2c \log \frac{1}{\mu(A)}}\right)^2\right). \tag{3.3}$$

As an example, one may take the product space $\Omega = \Omega_1 \times \cdots \times \Omega_n$ with the Hamming metric $d(x, y) = \mathrm{card}\{i \leqslant n : x_i \neq y_i\}$ and with a product measure $\mu = \mu_1 \otimes \cdots \otimes \mu_n$. In this case, (3.1)–(3.3) hold with optimal constant $c = n/4$ [Ma2]. For the Hamming distance, concentration inequalities related to (3.3) appeared first in the work of R. Ahlswede, P. Gács, and J. Körner [A-G-K] under the name "blowing up" property. They can be proved in this special case by different methods (cf. [MS], [T3]), but, among these inequalities, (3.3) is apparently the sharpest one. The particular two-point case $\Omega = \{0, 1\}^n$ with uniform measure μ is the exception: The optimal bound on the left-hand side of (3.3) is known and given by Harper's theorem [H]. Another important example is the Gaussian measure $\mu = \gamma_n$ on $\Omega = \mathbf{R}^n$ with the usual Euclidean metric. In this case, the inequality (3.1), with $c = 1$, has recently been established by M. Talagrand [T4]. His approach is another transportation inequality

$$W_2(\mu, v) \leqslant \sqrt{2c\, D(v \parallel \mu)}, \tag{3.4}$$

which of course is stronger than (3.1) and has the advantage that, for the Euclidean-type distance, it can be extended by induction from one dimension to n dimensions without loss in the constant c. However, it is not clear whether (3.4) implies a sharper concentration inequality than (3.3). Another question of interest is whether or not (3.1) is indeed stronger than its consequence (3.3). It turns out that (3.1) has a simple functional description which shows that, up to an absolute constant, (3.1) itself is a consequence of (3.3), which in particular holds assuming the LSI_c-property.

THEOREM 3.1. *Let μ be a probability measure on (Ω, d) such that $\int d(x, x_0)\, d\mu(x) < +\infty$, for some $x_0 \in \Omega$. The inequality*

$$W_1(\mu, \nu) \leqslant \sqrt{2c\, D(\nu \,\|\, \mu)} \tag{3.5}$$

holds for any absolutely continuous probability measure ν (with respect to μ), if and only if, for every function f on Ω with $\|f\|_{\mathrm{Lip}} \leqslant 1$ and $\int f\, d\mu = 0$, and, for all $t \in \mathbf{R}$.

$$\int e^{tf}\, d\mu \leqslant e^{ct^2/2}. \tag{3.6}$$

The assumption on μ ensures that all Lipschitz functions on Ω are μ-integrable which is necessary to get exponential, integrability in the form (3.6). Finally, in order to connect (3.5) with logarithmic Sobolev inequalities, we recall that, for every function f on Ω with $\|f\|_{\mathrm{Lip}} < +\infty$, "the modulus of gradient" of f is given by

$$\Gamma f(x) = |\nabla f(x)| = \limsup_{d(x, y) \to 0+} \frac{|f(x) - f(y)|}{d(x, y)}$$

(with $|\nabla f(x)| = 0$ for isolated points x in Ω). Thus, we have by Theorem 3.1:

COROLLARY 3.2. *Given $c > 0$, assume that LSI_c is satisfied for the triple (Ω, μ, Γ). Then, (3.5) holds. In particular, for every Borel set $A \subset \Omega$ and $h \geqslant \sqrt{2c \log 1/\mu(A)}$,*

$$1 - \mu(A^h) \leqslant \exp\left(-\frac{1}{2c}\left(h - \sqrt{2c \log \frac{1}{\mu(A)}}\right)^2\right).$$

Indeed, (3.6) is exactly the exponential inequality (2.3) which is implied by LSI_c as soon as the assumption $\Gamma(f) \leqslant 1$ is fulfilled. This relation holds since $|\nabla f(x)| \leqslant \|f\|_{\mathrm{Lip}}$, for all $x \in \Omega$. This establishes (3.5) and its consequence (3.3).

Proof of Theorem 3.1. Without loss of generality we restrict ourselves to the case $t \geqslant 0$ in (3.6). As in the proof of Theorem 2.1, (3.6) holds if and only if, for any non-negative Borel measurable function g on Ω,

$$\int (tf - ct^2/2)\, g\, d\mu \leqslant \mathrm{Ent}(g).$$

This inequality is homogeneous in g, so g may be chosen as a probability density (with respect to μ). Recalling that f has mean 0, we conclude that (3.6) holds if and only if, for any probability density g,

$$\int (fg - f)\, d\mu \leqslant \frac{ct}{2} + \frac{1}{t} \int g \log g\, d\mu. \tag{3.7}$$

Minimizing the right hand side of (3.7) in $t > 0$, we may rewrite (3.7) equivalently as

$$\int (fg - f)\, d\mu \leqslant \sqrt{2c \int g \log g\, d\mu}. \tag{3.8}$$

Let g be the density of the measure ν. Then (3.8) takes the form

$$\int f\, d\nu - \int f\, d\mu \leqslant \sqrt{2c \int \log \frac{d\nu}{d\mu}\, d\nu} = \sqrt{2c\, D(\nu \,\|\, \mu)}. \tag{3.9}$$

It remains to note that, by the Kantorovich–Rubinstein theorem (cf., e.g., [D, p. 330]),

$$\sup \left| \int f\, d\nu - \int f\, d\mu \right| = W_1(\mu, \nu), \tag{3.10}$$

where the sup is taken over all Lipschitz functions f on Ω with $\|f\|_{\mathrm{Lip}} \leqslant 1$. To be more precise, let M_1 denote the space of all Borel probability measures ν on Ω such that $\int d(x, x_0)\, d\nu(x) < +\infty$, for some $x_0 \in \Omega$. Then (M_1, W_1) is a metric space. Furthermore, all Lipschitz functions are ν-integrable, for $\nu \in M_1$, and (3.10) holds whenever $\mu, \nu \in M_1$ (the Kantorovich–Rubinstein theorem). To prove the equivalence of (3.5) and (3.6), assume first that the property (3.6) or, equivalently, the inequality (3.9) with arbitrary ν is fulfilled. Assume $D(\nu \,\|\, \mu) < +\infty$. Then, by (3.9) and by the assumption on μ, we obtain $\nu \in M_1$, and thus, one may apply (3.10). Hence, we arrive at (3.5). Conversely, assume that (3.5) is satisfied. Via (3.10), we get (3.9), for any $\nu \in M_1$. That is, (3.7) is satisfied, provided $\int d(x, x_0)\, g(x)\, d\mu(x) < +\infty$, for some $x_0 \in \Omega$. In particular, by the assumption on μ, (3.7) holds for all bounded probability densities g. Therefore, it holds for all probability densities g. But (3.7) is equivalent to (3.6). Thus, Theorem 3.1 is proved.

Remark 3.3. The subgaussian property (3.5) (up to an absolute constant in the exponent) can be expressed in another way. Given a function f with $\int f\, d\mu = 0$, the optimal value of $c = c(f)$ in (3.5) satisfies

$$K_0 \|f\|_N^2 \leqslant c \leqslant K_1 \|f\|_N^2,$$

where $\|f\|_N$ denotes the Orlicz norm for $N(x)=e^{x^2}-1$, and where K_0 and K_1 denote universal constants (one may take $K_0=1/6$, $K_1=2$). Therefore, roughly speaking, $\sqrt{c}$ in (3.5) is a diameter of the family of all functions f on Ω with mean 0 and $\|f\|_{\text{Lip}}\leqslant 1$ in the Orlicz space $L_N(\Omega,\mu)$ (cf. section 4 for definitions). Obviously, this diameter is at most Kc, for some universal K, provided that the concentration inequality (3.3) is satisfied. Thus, using Theorem 3.1, (3.3) implies the transportation inequality (3.1) with a constant Kc.

The statement in Corollary 3.2 that LSI_c implies (3.5) can be sharpened:

THEOREM 3.4. *Given $c>0$, assume that* LSI_c *is satisfied by the triple* (Ω,μ,Γ). *Then, for any probability measure ν which is absolutely continuous with respect to μ,*

$$\sup_{\|f\|_{\text{Lip}}\leqslant 1} \|f\|_{L^2(d\nu)}-\|f\|_{L^2(d\mu)}\leqslant\sqrt{2c\,D(\nu\,\|\,\mu)}. \tag{3.11}$$

Clearly, (3.11) is stronger than (3.5): Insert in (3.11) Lipshitz functions $f+a$ with $|a|\to\infty$. The proof of this statement uses (2.4) and the same arguments as those given in Remark 2.5. Hence we omit it.

What is the interpretation of the left hand side of (3.11), say $Q(\nu,\mu)$? Indeed, in connection with Theorem 3.4, this is a natural question. Obviously, $W_1(\mu,\nu)\leqslant Q(\nu,\mu)\leqslant W_2(\mu,\nu)$, but we do not know when the equality $Q(\nu,\mu)=W_2(\mu,\nu)$ holds for every absolutely continuous ν. In the case of Gaussian measures $\mu=\gamma_n$ on $\Omega=\mathbf{R}^n$ (with the Euclidean metric d), this would imply Talagrand's inequality (3.4),

$$W_2(\gamma_n,\nu)\leqslant\sqrt{2c\,D(\nu\,\|\,\gamma_n)},\qquad \text{for}\quad c=1, \tag{3.12}$$

based on the Gross logarithmic inequality, only. Nevertheless, one can give a simple alternative proof of (3.12) using the Levin–Rachev functional description of the Wasserstein metric and the Maurey (-type) inf-convolution inequality for the Gaussian measure.

To be more precise, assume that (Ω,d) is a Polish space, and let μ and ν denote probability measures on Ω. Then,

$$W_2^2(\nu,\mu)=\sup\int g\,d\nu-\int f\,d\mu, \tag{3.13}$$

where the supremum is taken over all pairs of bounded continuous functions (g,f) such that $g(y)-f(x)\leqslant d(x,y)^2$, for all $x,y\in\Omega$. This functional form for W_2 is due to V. L. Levin [Lev] and S. T. Rachev [Ra] (they established (3.13) in an even more general setting). The requirements on (g, f) suggest to consider instead of g the so-called inf-convolution. Put in general

$$(Sf)(x) = (S_{d^2/2} f)(x) = \inf \{ f(y) + \tfrac{1}{2} d(x, y)^2 : y \in \Omega \}$$

(without discussing inessential questions of measurability here). Thus, (3.13) is equivalent to

$$\sup_f \left[\int Sf \, d\nu - \int f \, d\mu \right] = \tfrac{1}{2} W_2^2(\nu, \mu), \tag{3.14}$$

which holds for all (measurable) f. Therefore the inequality (3.12) is equivalent to

$$\int e^{Sf} \, d\mu \leqslant e^{\int f \, d\mu}, \qquad \text{for all } f, \qquad \text{where} \quad \mu = \gamma_n. \tag{3.15}$$

Indeed, (3.15) amounts to say that, for any probability density g of a measure ν with respect to $\mu = \gamma_n$,

$$\int Sf \, d\nu - \int f \, d\gamma_n = \int \left(Sf - \int f \, d\gamma_n \right) g \leqslant \operatorname{Ent} g = D(\nu \parallel \gamma_n),$$

which is exactly (3.12) due to (3.14). So, (3.15) is indeed a functional form for Talagrand's transportation inequality (3.12). It might be worthwhile to note that, for convex functions f on $\mathbf{R}^n$, (3.15) can be rewritten in an infinite dimensional setting as Tsirel'son's inequality [Ts]

$$\mathbf{E} \exp \{ \sup_t (x_t - \sigma_t^2/2) \} \leqslant \exp \{ \mathbf{E} \sup_t x_t \},$$

where x_t denotes an arbitrary bounded Gaussian process with variances σ_t^2.

In order to prove (3.15), we shall use an argument of B. Maurey [Mau]. We use the Brunn–Minkowski inequality in the functional form of A. Prékopa and L. Leindler ([Pr], [Lei], cf. also [P2, p. 3] for a simple proof due to K. Ball): Given (fixed) $\lambda \in (0, 1)$, for any measurable functions u, v and w on $\mathbf{R}^n$,

$$\left(\int e^{-u} \, dx \right)^{\lambda} \left(\int e^{-v} \, dx \right)^{1-\lambda} \leqslant \int e^{-w} \, dx \tag{3.16}$$

holds provided that u, v and w satisfy $w(\lambda x+(1-\lambda)\, y)\leqslant \lambda u(x)+(1-\lambda)\, v(y)$, for all $x, y\in \mathbf{R}^n$. For a given function f on $\mathbf{R}^n$, this condition is satisfied for all $\lambda\in(0,1)$ by the functions

$$w(x)=\tfrac{1}{2}\,|x|^2,\quad u(x)=\tfrac{1}{2}\,|x|^2+(1-\lambda)\, f(x),\quad v(x)=\tfrac{1}{2}\,|x|^2-\lambda(Sf)(x),$$

and for these functions (3.16) may be rewritten as

$$\left(\int e^{\lambda Sf}\,d\gamma_n\right)^{1-\lambda}\left(\int e^{-(1-\lambda)\,f}\,d\gamma_n\right)^{\lambda}\leqslant 1. \tag{3.17}$$

When $\lambda=\frac{1}{2}$, this is Maurey's inf-convolution inequality $\int e^{Sf/2}\,d\gamma_n \int e^{-f/2}\,d\gamma_n \leqslant 1$. There is equality in (3.17) when $\lambda=1$, and comparing derivatives at this point, we arrive at (3.15).

Remark 3.5. M. Talagrand established another transportation inequality of the form (3.12) for the product measure μ_n on $\mathbf{R}^n$ of the two-sided exponential measure with density $d\mu_n(x)/dx=e^{-(|x_1|+\cdots+|x_n|)}2^{-n}$. His inequality ([T4, Theorem 1.2]) is

$$W_2(\mu_n, \nu)\leqslant\sqrt{c\, D(\nu\,\|\,\mu_n)},\qquad c=\frac{\alpha}{1-\alpha},\qquad \alpha\in(0,1), \tag{3.18}$$

where the distance $W_2(\mu_n,\nu)$ corresponds to the (noncanonical) metric d_α in $\mathbf{R}^n$ given by

$$\tfrac{1}{2}\, d_\alpha^2(x,y)=\sum_{i=1}^{n} U(\alpha\,|x_i-y_i|),\qquad x, y\in\mathbf{R}^n,\qquad \text{where}\quad U(t)=t-1+e^{-t}.$$

The function $U(|t|)$ behaves like $\frac{1}{2}t^2$ for small t and like $|t|$ for large t. In fact, a stronger inequality, namely, the inequality (3.18) with respect to the cost function $\sum_{i=1}^n U(\alpha(x_i-y_i))$ was proved in [T4], and in this stronger inequality the condition $\alpha<1$ is essential, and there are cases of equality for each α.

As in the Gaussian case, the transportation inequality (3.18) can be connected with a known inf-convolution inequality. Using arguments as above, the inequality (3.15) with $\mu=\mu_n$ and with operator $S=S_{d_\alpha^2/c}$,

$$\int e^{S_{d^2_\alpha/c}f}\,d\mu_n\leqslant e^{\int f\,d\mu_n}, \tag{3.19}$$

represents a functional form for (3.18). In this case, an inf-convolution inequality related to (3.19) was obtained by B. Maurey in the same paper [Mau] but as a functional form for a different isoperimetric-type inequality for μ_n due to M. Talagrand ([T2]). Namely, B. Maurey proved that, for all functions f on $\mathbf{R}^n$,

$$\int e^{S_{\rho^2/2} f}\, d\mu_n \int e^{-f}\, d\mu_n \leqslant 1 \tag{3.20}$$

with respect to the metric ρ on $\mathbf{R}^n$ given by $\frac{1}{2}\rho^2(x, y) = \sum_{i=1}^n W(|x_i - y_i|)$, $x, y \in \mathbf{R}^n$, where $W(t) = \frac{1}{36} t^2$ for $|t| \leqslant 4$, and $W(t) = \frac{2}{9}(|t| - 2)$ otherwise. Since (3.20) implies

$$\int e^{S_{\rho^2/2}}\, d\mu_n \leqslant e^{\int f\, d\mu_n}, \tag{3.21}$$

and since the functions U and W can be estimated up to constants by each other, we roughly arrive at (3.18). Indeed, (3.21) implies (3.19) and thus (3.18) if and only if $S_{\rho^2/2} \geqslant S_{d_\alpha^2/c}$, that is, if and only if $W(t) \geqslant 2U(\alpha t)/c$, for all $t \geqslant 0$. By a simple computation, the optimal constant is determined to be

$$c = \sup_{t > 0} \frac{2U(\alpha t)}{W(t)} = 36\alpha^2$$

(the sup is attained at $t = 0$). Thus, (3.18) holds with this c for all $\alpha > 0$.

4. LOGARITHMIC SOBOLEV INEQUALITIES AS POINCARÉ-TYPE INEQUALITIES IN ORLICZ SPACES

Since $\Gamma(f + a) = \Gamma(f)$, for all $a \in \mathbf{R}$, the inequality

$$\mathrm{Ent}(f^2) \leqslant c \int \Gamma(f)^2\, d\mu \tag{4.1}$$

may formally be strengthened as

$$\mathscr{L}(f) \equiv \sup_{a \in \mathbf{R}} \mathrm{Ent}((f + a)^2) \leqslant c \int \Gamma(f)^2\, d\mu. \tag{4.2}$$

Now (4.2) being equivalent to (4.1) is invariant under translations $f \to f + \mathrm{const}$, which leads to the question whether the functional $\mathscr{L}^{1/2}$ may be a seminorm. Indeed, we will connect it to an Orlicz space norm.

Let (Ω, μ) be a probability space. Given a Young function $N : \mathbf{R} \to [0, +\infty)$, i.e., an even, convex function with $N(0) = 0$, $N(x) > 0$ for

$x > 0$ the Orlicz space $L_N = L_N(\Omega, \mu)$ consists of all measurable functions f with

$$\|f\|_N = \sup\left\{\lambda > 0 : \int N(f/\lambda)\, d\mu\right\} < +\infty.$$

Any Young function N strictly increases on $[0, +\infty)$ so an inverse $N^{-1} : [0, +\infty) \to [0, +\infty)$ exists. When $N(x) = |x|^p (1 \leqslant p < +\infty)$, L_N is the usual Lebesgue space with norm $\|f\|_p$. In what follows, we consider the norms $\|f\|_N$ and $\|f\|_\Psi$ for the Young functions $N(x) = x^2 \log(1 + x^2)$ and $\Psi(x) = |x| \log(1 + |x|)$, respectively

Proposition 4.1. *For any function f in $L_N(\Omega, \mu)$,*

$$\tfrac{2}{3} \left\| f - \int f\, d\mu \right\|_N^2 \leqslant \mathscr{L}(f) \leqslant \tfrac{13}{4} \left\| f - \int f\, d\mu \right\|_N^2. \tag{4.3}$$

Thus, the Logarithmic Sobolev inequality (4.1) is equivalent, up to a numerical multiplicative constant, to an inequality

$$\left\| f - \int f\, d\mu \right\|_N \leqslant C\, \|\Gamma(f)\|_2 \tag{4.4}$$

which belongs to the class of Poincaré-type inequalities. When $\Omega = \mathbf{R}$, the estimates (4.3) can be further modified, and (4.4) may be reduced to a Hardy-type inequality for the Orlicz space norm $\|\cdot\|_\Psi$.

Let $m = m(\mu)$ denote a median of the probability measure μ on $\mathbf{R}$. For functions f on $\mathbf{R}$ introduce $f_0 = f \mathbf{1}_{(-\infty, m]}$ and $f_1 = f \mathbf{1}_{[m, +\infty)}$.

Proposition 4.2. *Assume that for any smooth function f on $\mathbf{R}$,*

$$\operatorname{Ent}(f^2) \leqslant c \int_{-\infty}^{\infty} f'(x)^2\, d\mu(x). \tag{4.5}$$

Then, for any smooth function f on $\mathbf{R}$ with $f(m) = 0$,

$$\|f_0^2\|_\Psi + \|f_1^2\|_\Psi \leqslant d \int_{-\infty}^{\infty} f'(x)^2\, d\mu(x) \tag{4.6}$$

with $d = 75c/2$. Conversely, (4.6) *implies* (4.5) *with $c = 117d/2$.*

The inequality (4.6) being an equivalent form for logarithmic Sobolev inequalities on the real line will be used in the next section to characterize probability measures μ satisfying (4.5). The proof of Propositions 4.1 and 4.2 is given in this section. Clearly, (4.5) as well (4.6) may be extended from the class of all smooth functions to the class of all absolutely continuous functions (with the same condition $f(m)=0$ in the case of (4.6)).

LEMMA 4.3. $\|f\|_1 \leqslant \|f\|_2 \leqslant (\sqrt{5}/2)\, \|f\|_N$, *for all* $f \in L_N(\Omega, \mu)$.

Proof. One may assume that $\|f\|_N = 1$. Since $\int \Psi(f^2)\, d\mu = \int N(f)\, d\mu = 1$, we have, by Jensen's inequality, $\Psi(\int f^2\, d\mu) \leqslant 1$. Hence, $\int f^2\, d\mu \leqslant \psi^{-1}(1) < 5/4$ since $\psi(5/4) = 1.014... > 1$ (where Ψ^{-1} denotes the inverse function).

Proof of Proposition 4.1. One may assume that $f \in L_N$, $\|f\|_N = 1$, and that $\int f\, d\mu = 0$. The second inequality in (4.3) is essentially a version of an inequality due to O. S. Rothaus who showed ([Ro], Lemma 10) that

$$\mathscr{L}(f) \leqslant \mathrm{Ent}(f^2) + 2 \int f^2\, d\mu \tag{4.7}$$

whenever $\int f\, d\mu = 0$. Now, introducing the function $U(x) = 2x - x \log x$, $x \geqslant 0$, and noting that

$$\int f^2 \log f^2\, d\mu \leqslant \int f^2 \log(1 + f^2)\, d\mu = \int N(f)\, d\mu = 1,$$

we obtain that

$$\mathrm{Ent}(f^2) + 2 \int f^2\, d\mu = \int f^2 \log f^2\, d\mu + U\left(\int f^2\, d\mu\right) \leqslant 1 + U\left(\int f^2\, d\mu\right). \tag{4.8}$$

The function $U(x)$ increases in $0 \leqslant x \leqslant e$. Hence, by Lemma 4.3,

$$U\left(\int f^2\, d\mu\right) < U(5/4) = 5/2 - \Psi(5/4) < 9/4.$$

Combining (4.7) and (4.8), we get $\mathscr{L}(f) \leqslant 13/4$. This proves the second inequality in (4.3).

In order to prove the first inequality in (4.3), assume that $f \in L_N$, $\int f\, d\mu = 0$, and by homogeneity that $\mathscr{L}(f) = 2$. Since in general

$$\int f^2\, d\mu - \left(\int f\, d\mu\right)^2 = \tfrac{1}{2} \lim_{|a| \to \infty} \mathrm{Ent}((f+a)^2) \leqslant \tfrac{1}{2} \mathscr{L}(f),$$

we get $\int f^2\,d\mu \leqslant 1$. Thus, $\int f^2\,d\mu \log \int f^2\,d\mu \leqslant 0$ and therefore

$$\begin{aligned}\int f^2 \log f^2\,d\mu &= \mathrm{Ent}(f^2) + \int f^2\,d\mu \log \int f^2\,d\mu \\ &\leqslant \mathrm{Ent}(f^2) \\ &\leqslant \mathcal{L}(f) = 2.\end{aligned}$$

But $\Psi(x) \leqslant 1 + x \log x$, for all $x \geqslant 0$. Therefore,

$$\int N(f)\,d\mu = \int \Psi(f^2)\,d\mu \leqslant 1 + \int f^2 \log f^2\,d\mu \leqslant 3.$$

Finally note that $N(x/\sqrt{3}) \leqslant \frac{1}{3} N(x)$, for all real x. Consequently, $\int N(f/\sqrt{3})\,d\mu \leqslant 1$, that is, $\|f/\sqrt{3}\|_N \leqslant 1$. Thus, $\|f\|_N^2 \leqslant 3 = \frac{3}{2}\mathcal{L}(f)$, and Proposition 4.1 follows.

To prove Proposition 4.2, we need an elementary lemma.

LEMMA 4.4. *For any function $f \in L_N(\mathbf{R}, \mu)$,*

$$\left\| f - \int f\,d\mu \right\|_N \leqslant 3\,\|f\|_N. \tag{4.9}$$

If $f = 0$ on $(-\infty, m(\mu))$, we also have

$$\|f\|_N \leqslant 5\,\left\| f - \int f\,d\mu \right\|_N. \tag{4.10}$$

Proof. By Lemma 4.3, $\|f - \int f\,d\mu\|_N \leqslant \|f\|_N + \|f\|_1 \leqslant 3\,\|f\|_N$. This proves (4.9). To prove (4.10), using the Cauchy–Schwarz inequality and once more by Lemma 4.3, we get

$$\left|\int f\,d\mu\right| = \left|\int f \mathbf{1}_{(-\infty, m)}\,d\mu\right| \leqslant \|f\|_2 \frac{1}{\sqrt{2}} \leqslant \sqrt{\frac{5}{8}}\,\|f\|_N.$$

Hence,

$$\|f\|_N \leqslant \left\| f - \int f\,d\mu \right\|_N + \left|\int f\,d\mu\right| \leqslant \left\| f - \int f\,d\mu \right\|_N + \sqrt{\frac{5}{8}}\,\|f\|_N.$$

Thus,

$$\|f\|_N \leqslant \frac{1}{1-\sqrt{5/8}} \left\| f - \int f\, d\mu \right\|_N \leqslant 5 \left\| f - \int f\, d\mu \right\|_N$$

and Lemma 4.4 is proved.

Proof of Proposition 4.2. First we derive (4.6) from (4.5). By Proposition 4.1, we have, for any smooth function f from $L_N(\mathbf{R}, \mu)$,

$$\tfrac{2}{3} \left\| f - \int f\, d\mu \right\|_N^2 \leqslant c \int_{-\infty}^{\infty} f'(x)^2\, d\mu(x). \tag{4.11}$$

As already noted, this inequality extends to all absolutely continuous functions f in $L_N(\mathbf{R}, \mu)$ in the sense that starting with an integrable function f', one defines f as indefinite integral of f'. This implies in particular that the measure μ on the right hand side of (4.11) may be replaced with its absolutely continuous component (with respect to Lebesgue measure), and therefore the integral in (4.11) can be taken over any set of full Lebesgue measure, e.g., over $\mathbf{R}\backslash\{m\}$. Hence, we can apply (4.11) to f_0 and f_1

$$\tfrac{2}{3} \left\| f_0 - \int f_0\, d\mu \right\|_N^2 \leqslant c \int_{-\infty}^{m} f'(x)^2\, d\mu(x),$$

$$\tfrac{2}{3} \left\| f_1 - \int f_1\, d\mu \right\|_N^2 \leqslant c \int_{m}^{+\infty} f'(x)^2\, d\mu(x).$$

Applying (4.10) and the general identity $\|g\|_N^2 = \|g^2\|_\Psi$, we get

$$\tfrac{2}{75} \|f_0^2\|_\Psi \leqslant c \int_{-\infty}^{m} f'(x)^2\, d\mu(x),$$

$$\tfrac{2}{75} \|f_1^2\|_\Psi \leqslant c \int_{m}^{+\infty} f'(x)^2\, d\mu(x).$$

Adding these inequalities, we obtain (4.6) with $d = 75c/2$.

To derive (4.5) from (4.6), we use (4.9). First assume $f(m) = 0$. Since $f = f_0 + f_1$, we have

$$\begin{aligned}
\left\| f - \int f\, d\mu \right\|_N^2 &\leqslant \left(\left\| f_0 - \int f_0\, d\mu \right\|_N + \left\| f_1 - \int f_1\, d\mu \right\|_N \right)^2 \\
&\leqslant 9(\|f_0\|_N + \|f_1\|_N)^2 \leqslant 18(\|f_0\|_N^2 + \|f_1\|_N^2) \\
&= 18(\|f_0^2\|_\Psi + \|f_1^2\|_\Psi) \leqslant 18d \int_{-\infty}^{\infty} f'(x)^2\, d\mu(x)
\end{aligned}$$

where we used (4.6) in the last inequality. Now, by Proposition 4.1,

$$\mathscr{L}(f) \leqslant \tfrac{13}{4} \left\| f - \int f \, d\mu \right\|_N^2 \leqslant \tfrac{117}{2} d \int_{-\infty}^{\infty} f'(x)^2 \, d\mu(x).$$

This inequality is invariant under translations $f \to f + \text{const}$, thus it holds without the condition $f(m) = 0$. This proves (4.5) for $c = 117d/2$.

5. PROBABILITY MEASURES SATISFYING LOGARITHMIC SOBOLEV INEQUALITIES ON THE REAL LINE

Using Proposition 4.2, we shall give in this section a direct characterization of probability measures μ on the real line $\mathbf{R}$ satisfying the logarithmic Sobolev inequality

$$\operatorname{Ent}(f^2) \leqslant c \int_{-\infty}^{\infty} f'(x)^2 \, d\mu(x) \tag{5.1}$$

with some (finite) constant c for all smooth functions f on $\mathbf{R}$. Our tool is the following theorem due to M. Artola, G. Talenti, and G. Tomaselli (cf. [Mu]) on the optimal constant $A = A(\nu, \lambda)$ in the Hardy-type inequality with weights

$$\int_0^{\infty} f(x)^2 \, d\nu(x) \leqslant A \int_0^{\infty} f'(x)^2 \, d\lambda(x).$$

Here f is supposed to be an arbitrary smooth function on $[0, +\infty)$ such that $f(0) = 0$, and μ and λ are (non-negative) Borel measures on $[0, +\infty)$. Denote by $p_\lambda = p_\lambda(x)$ the absolutely continuous component of λ with respect to Lebesgue measure, and define the constant $B = B(\nu, \lambda)$ as

$$B(\nu, \lambda) = \sup_{x > 0} \nu([x, +\infty)) \int_0^x \frac{dt}{p_\lambda(t)}.$$

THEOREM 5.1 [Mu]. $B \leqslant A \leqslant 4B$.

This theorem has the following natural generalization. Consider a Borel measure ν on $[0, +\infty)$ and a Banach space $(X, \|\cdot\|)$ of Borel measurable functions on $[0, +\infty)$ (with usual factorization with respect to measure ν) such that

(1) $f \leqslant |g|$ ν-a.e., $g \in X$ implies $f \in X$ and $\|f\| \leqslant \|g\|$, for all Borel measurable functions f;

(2) any pointwise non-decreasing sequence f_n of non-negative functions in X converging pointwise to a function $f \in X$ satisfies $\|f_n\| \to \|f\|$.

By property (1), X is an ideal Banach space, and property (2) is called order semicontinuity of the norm. For an ideal Banach space X, the last property is equivalent to a representation of the norm in X in the form

$$\|f\| = \sup_{g \in \mathcal{G}} \int_0^\infty |f(x)|\, g(x)\, d\nu(x), \tag{5.2}$$

for some family $\mathcal{G}$ of non-negative Borel measurable functions g on $[0, +\infty)$ (this statement holds in the setting of an abstract probability space (Ω, ν), cf., e.g., [KA, p. 190]).

For these Banach spaces X, one immediately obtains by Theorem 5.1:

Corollary 5.2. *Let $A = A(X, \lambda)$ be the optimal constant in the inequality*

$$\|f^2\| \leqslant A \int_0^\infty f'(x)^2\, d\lambda(x), \tag{5.3}$$

where $f \in X$ is an arbitrary smooth function such that $f(0) = 0$. Then, $B \leqslant A \leqslant 4B$, where

$$B = B(X, \lambda) = \sup_{x>0} \|\mathbf{1}_{[x, +\infty)}\| \int_0^x \frac{dt}{p_\lambda(t)}.$$

Indeed, the measures $\nu_g(dx) = g(x)\, \nu(dx)$ satisfy $B(\nu_g, \lambda) \leqslant A(\nu_g, \lambda) \leqslant 4B(\nu_g, \lambda)$. Using the definitions (5.2) and (5.3), we get

$$A(X, \lambda) = \sup_{g \in \mathcal{G}} A(\nu_g, \lambda), \quad B(X, \lambda) = \sup_{g \in \mathcal{G}} B(\nu_g, \lambda),$$

hence, $B(X, \lambda) \leqslant A(X, \lambda) \leqslant 4B(X, \lambda)$.

In particular, one may apply Corollary 5.2 to the Orlicz space $X = L_\Psi(\nu)$ which of course satisfies the properties (1) and (2) above. Recall that $\Psi(x) = |x| \log(1 + |x|)$. For indicator functions, we get by definition of the Orlicz norm

$$\|\mathbf{1}_{[x, +\infty)}\|_\Psi = \frac{1}{\Psi^{-1}(1/\nu([x, +\infty)))},$$

where Ψ^{-1} denotes the inverse function. Consequently, the optimal constant A in (5.3) for the norm $\|\cdot\| = \|\cdot\|_\psi$ can be estimated as follows:

$$\sup_{x>0} \frac{1}{\Psi^{-1}(1/\nu([x,+\infty)))} \int_0^x \frac{dt}{p_\lambda(t)}$$

$$\leqslant A \leqslant 4 \sup_{x>0} \frac{1}{\Psi^{-1}(1/\nu([x,+\infty)))} \int_0^x \frac{dt}{p_\lambda(t)}. \tag{5.4}$$

Using Proposition 4.2, we may now conclude the proof of the main result of this section.

Let μ be a Borel probability measure on $\mathbf{R}$ with distribution function $F(x) = \mu((-\infty, x])$, and density function $p = p(x)$, $x \in \mathbf{R}$, for its absolutely continuous part with respect to Lebesgue measure. Denote by m a median of μ. Define

$$D_0 = \sup_{x<m} \left(F(x) \log \frac{1}{F(x)} \right) \int_x^m \frac{1}{p(t)}\, dt,$$

$$D_1 = \sup_{x>m} \left((1-F(x)) \log \frac{1}{1-F(x)} \right) \int_m^x \frac{1}{p(t)}\, dt,$$

defining D_0 and D_1 to be zero in case $\mu((-\infty, m)) = 0$ or $\mu((m, +\infty)) = 0$, respectively.

THEOREM 5.3. *For some positive absolute constants K_0 and K_1, the optimal value of c in the logarithmic Sobolev inequality* (5.1) *satisfies*

$$K_0(D_0 + D_1) \leqslant c \leqslant K_1(D_0 + D_1).$$

Actually, one may choose $K_0 = 1/150$ and $K_1 = 468$. To simplify the expression in (5.4), we first prove:

LEMMA 5.4. *Let $c_1 = \frac{1}{2}$, $c_2 = 2$. For all $t \geqslant 2$,*

$$c_1 \frac{t}{\log t} \leqslant \Psi^{-1}(t) \leqslant c_2 \frac{t}{\log t}. \tag{5.5}$$

Proof. The first inequality in (5.5), that is,

$$\Psi\left(\frac{c_1 t}{\log t}\right) = \frac{c_1 t}{\log t} \log\left(1 + \frac{c_1 t}{\log t}\right) \leqslant t$$

will follow, due to $c_1/\log t \leqslant 1$ for $t \geqslant 2$, from the inequality

$$\frac{c_1 t}{\log t} \log(1+t) \leqslant t,$$

which is equivalent to $\log(1+t) \leqslant 2\log t$, that is, to $1+t \leqslant t^2$. The last is evident. The second inequality in (5.5),

$$\frac{c_2 t}{\log t} \log\left(1+\frac{c_2 t}{\log t}\right) \geqslant t,$$

may be rewritten as

$$1+\frac{c_2 t}{\log t} \geqslant t^{1/c_2}$$

and will follow from $c_2 t/\log t \geqslant t^{1/c_2}$. For $c_2 = 2$, the last is equivalent to $u(t) = \log t/(2\sqrt{t}) \leqslant 1$. The function u attains its maximum on the interval $[2, +\infty)$ at the point $t = e^2$, and $u(e^2) = 1/e < 1$. Hence Lemma 5.4 follows.

Proof of Theorem 5.3. Without loss of generality, let $m = 0$. The inequality (4.6) may be divided into the two inequalities:

$$\|f_0^2\|_{\Psi} \leqslant d \int_{-\infty}^{0} f_0'(x)^2 \, d\mu(x), \tag{5.6}$$

$$\|f_1^2\|_{\Psi} \leqslant d \int_{0}^{\infty} f_1'(x)^2 \, d\mu(x), \tag{5.7}$$

where f_0 and f_1 are arbitrary smooth functions defined on $(-\infty, 0]$ and $[0, +\infty)$, respectively, with $f_0(0) = f_1(0) = 0$. According to (5.4) with $\nu = \lambda$ being the restriction of μ to $[0, +\infty)$, and by Lemma 5.5, we have, for the optimal constant d_1 in (5.7),

$$\frac{1}{2} \sup_{x>0} \mu([x, +\infty)) \log \frac{1}{\mu([x, +\infty))} \int_0^x \frac{dt}{p(t)}$$
$$\leqslant d_1 \leqslant 8 \sup_{x>0} \mu([x, +\infty)) \log \frac{1}{\mu([x, +\infty))} \int_0^x \frac{dt}{p(t)}. \tag{5.8}$$

Here we have used the fact that $m = 0$ is a median, hence, $t = 1/\mu([x, +\infty)) \geqslant 2$, for all $x > 0$. It is obvious that the supremum in (5.8) will not change if one replaces the expression $\mu([x, +\infty))$ by $\mu((x, +\infty))$. But $\mu((x, +\infty)) = 1 - F(x)$, and using definition of D_1, we may rewrite (5.8) as

$$\tfrac{1}{2} D_1 \leqslant d_1 \leqslant 8 D_1.$$

By a similar reasoning, we get, for the optimal constant d_0 in (5.6),

$$\tfrac{1}{2} D_0 \leqslant d_0 \leqslant 8 D_0.$$

By Proposition 4.2, (5.1) implies (5.6)–(5.7) with $d = 75c/2$. Since $d = \max(d_0, d_1)$ is the optimal constant satisfying both inequalities (5.6) and (5.7), we get

$$\tfrac{1}{2} \max(D_0, D_1) \leqslant 75c/2.$$

Therefore, $D_0 + D_1 \leqslant 150c$ which implies Theorem 5.3 with $K_0 = 1/150$. On the other hand, again by Proposition 4.2, $c \leqslant 117d/2 = 117 \max(d_0, d_1)/2$, so, $c \leqslant 468 \max(D_0, D_1) \leqslant 468(D_0 + D_1)$. Thus, one may choose $K_1 = 360$ which proves Theorem 5.3.

REFERENCES

[A-G-K] R. Ahlswede, P. Gács, and J. Körner, Bounds on conditional probabilities with applications in multi-user communication, *Z. Wahrsch. Verw. Gebiete* **34** (1976), 157–177.

[A-M-S] S. Aida, T. Masuda, and I. Shigekawa, Logarithmic Sobolev inequalities and exponential integrability, *J. Funct. Anal.* **126** (1994), 83–101.

[A-S] S. Aida and D. Stroock, Moment estimates derived from Poincaré and logarithmic Sobolev inequalities, *Math. Research Letters* **1** (1994), 75–86.

[B] S. G. Bobkov, On Gross' and Talagrand's inequalities on the discrete cube, *Vestnik Syktyvkar Univ.* **1**(1) (1995), 12–19. [in Russian]

[D-S] E. B. Davies and B. Simon, Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians, *J. Funct. Anal.* **59** (1984), 335–395.

[D-SC] P. Diaconis and L. Saloff-Coste, Logarithmic Sobolev inequalities for finite Marlov chains, *Ann. Appl. Probab.* **6**(3) (1996), 695–750.

[D] R. M. Dudley, "Real Analysis and Probability," Pacific Grove, CA, 1989.

[G1] L. Gross, Logarithmic Sobolev inequalities, *Amer. J. Math.* **97** (1975), 1061–1083.

[G2] L. Gross, Logarithmic Sobolev inequalities and contractivity properties of semi-groups, Varenna, 1992, *Lecture Notes in Math.* **1563** (1993), 54–88.

[H] L. H. Harper, Optimal numbering and isopermimetric problems on graphs, *J. Combin. Theory* **1** (1966), 385–393.

[K-A] L. V. K. Kantorovich and G. P. Akilov, "Functional Analysis," Pergamon, Oxford, 1982.

[L1] M. Ledoux, Isoperimetry and Gaussian analysis, *Ecole d'été de Probabilités de Saint-Flour*, 1994, *Lecture Notes in Math.* **1648** (1996), 165–294.

[L2] M. Ledoux, Remarks on logarithmic Sobolev constants, exponential integrability and bounds on the diameter, *J. Math. Kyoto Univ.* **35** (1995), 211–220.

[L3] M. Ledoux, On Talagrand's deviation inequalities for product measures, *ESAIM-Probab. Statist.* **1** (1996), 63–87.

[Lei] L. Leindler, On a certain converse of Hölder's inequality, *Acta Sci. Math. Szeged.* **33** (1972), 217–223.

[Lev] V. L. Levin, The problem of mass transfer in a topological space, and probability measures having given marginal measures on the product of two spaces, *Soviet Math. Dokl.* **29** (1984), 638–643.

[Ma1] K. Marton, A simple proof of the blowing-up lemma, *IEEE Trans. Inform. Theory* **IT-32** (1986), 445–446.

[Ma2] K. Marton, Bounding $\bar{d}$-distance by informational divergence: a method to prove measure concentration, *Ann. Probab.* **24** (1996), 857–866.

[Mau] B. Maurey, Some deviation inequalities, *Geom. Funct. Anal.* **1** (1991), 188–197.

[Mu] B. Muckenhoupt, Hardy's inequality with weights, *Studia Math.* **XLIV** (1972), 31–38.

[M-S] V. D. Milman and G. Schechtman, "Asymptotic Theory of Finite Dimensional Normed Spaces," Lecture Notes in Math., Vol. 1200, Springer-Verlag, Berlin, 1986.

[P1] G. Pisier, Probabilistic methods in the geometry of Banach spaces, *in* "Probability and Analysis, Varenna, Italy, 1985," Lecture Notes in Math., Vol. 1206, pp. 167–241, Springer-Verlag, Berlin, 1986.

[P2] G. Pisier, "The Volume of Convex Bodies and Banach Space Geometry," Tracts in Math., Cambridge Univ. Press, Cambridge, UK, 1989.

[Pr] A. Précopa, On logarithmic concave measures and functions, *Acta Sci. Math. Szeged.* **34** (1973), 335–343.

[Ra] S. T. Rachev, The Monge–Kantorovich mass transference problem and its stochastic applications, *Theory Probab. Appl.* **24** (1984), 647–671.

[Ro] O. S. Rothaus, Analytic inequalities, isoperimetric inequalities and logarithmic Sobolev inequalities, *J. Funct. Anal.* **64** (1985), 296–313.

[T1] M. Talagrand, An isoperimetric theorem on the cube and the Khinchine–Kahane inequalities, *Proc. Amer. Math. Soc.* **104** (1988), 905–909.

[T2] M. Talagrand, A new isoperimetric inequality and the concentration of measure phenomenon, *Lecture Notes in Math.* **1469** (1991), 94–124.

[T3] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, *Publ. Math. I.H.E.S.* **81** (1995), 73–205.

[T4] M. Talagrand, Transportation cost for Gaussian and other product measures, *Geom. Funct. Anal.* **6** (1996), 587–600.

[T5] M. Talagrand, Private communication, 1996.

[Ts] B. S. Tsirel'son, A geometric approach to a maximum likelyhood estimation for infinite-dimensional Gaussian location, II, *Theory Probab. Appl.* **30** (1985), 820–828.