BOUNDS ON THE DEFICIT IN THE LOGARITHMIC SOBOLEV INEQUALITY

S. G. BOBKOV, N. GOZLAN, C. ROBERTO AND P.-M. SAMSON

ABSTRACT. The deficit in the logarithmic Sobolev inequality for the Gaussian measure is considered and estimated by means of transport and information-theoretic distances.

1. INTRODUCTION

Let γ denote the standard Gaussian measure on the Euclidean space $\mathbf{R}^n$, thus with density

$$\frac{d\gamma(x)}{dx} = \frac{1}{(2\pi)^{n/2}}\, e^{-|x|^2/2}$$

with respect to the Lebesgue measure. (Here and in the sequel $|x|$ stands for the Euclidean norm of a vector $x \in \mathbf{R}^n$.) One of the basic results in the Gaussian Analysis is the celebrated logarithmic Sobolev inequality

$$\int f \log f \, d\gamma - \int f \, d\gamma \, \log \int f \, d\gamma \leq \frac{1}{2} \int \frac{|\nabla f|^2}{f} \, d\gamma, \tag{1.1}$$

holding true for all positive smooth functions f on $\mathbf{R}^n$ with gradient ∇f. In this explicit form it was obtained in the work of L. Gross [G], initiating fruitful investigations around logarithmic Sobolev inequalities and their applications in different fields. See e.g. a survey by M. Ledoux [L1] and the book [L2] for a comprehensive account of such activities up to the end of 90's. One should mention that in an equivalent form – as a relation between the entropy power and the Fisher information, (1.1) goes back to the work by A. J. Stam [S].

The inequality (1.1) is homogeneous in f, so, the restriction $\int f \, d\gamma = 1$ does not loses generality. It is sharp in the sense that the equality is attained, namely for all $f(x) = e^{l(x)}$ with arbitrary affine functions l on $\mathbf{R}^n$ (in which case the measures $\mu = f\gamma$ are still Gaussian). It is nevertheless of a certain interest to realize how large the difference between both sides of (1.1) is. This problem has many interesting aspects. For example, as was shown by E. Carlen in [C], which was perhaps a first address of the sharpness problem, for $f = |u|^2$ with a smooth complex-valued u such that $\int |u|^2 \, d\gamma = 1$, (1.1) may be strengthened to

$$\int |u|^2 \log |u|^2 \, d\gamma + \int |Wu|^2 \log |Wu|^2 \, d\gamma \leq 2 \int |\nabla u|^2 \, d\gamma,$$

where W denotes the Wiener transform of u. That is, a certain non-trivial functional may be added to the left-hand side of (1.1).

Key words and phrases. Logarithmic Sobolev inequality, Entropy, Fisher Information, Transport Distance, Gaussian measures.

The first author was partially supported by NSF grant DMS-1106530. The other authors were partially supported by the "Agence Nationale de la Recherche" through the grants ANR 2011 BS01 007 01 and ANR 10 LABX-58.

One may naturally wonder how to bound from below the deficit in (1.1), that is, the quantity

$$\delta(f) = \frac{1}{2} \int \frac{|\nabla f|^2}{f}\, d\gamma - \Big[\int f \log f \, d\gamma - \int f \, d\gamma \, \log \int f \, d\gamma\Big],$$

in terms of more explicit, like distribution-dependent characteristics of f showing its closeness to the extremal functions e^l (when $\delta(f)$ is small). Recently, results of this type have been obtained by A. Cianchi, N. Fusco, F. Maggi and A. Pratelli [C-F-M-P] in their study of the closely related isoperimetric inequality for the Gaussian measure. The work by E. Mossel and J. Neeman [M-N] deals with dimension-free bounds for the deficit in one functional form of the Gaussian isoperimetric inequality appearing in [B]. See also the subsequent paper by R. Eldan [E] where almost tight two-sided robustness bounds have been derived.

As for (1.1), one may also to involve distance-like quantities between the measures $\mu = f\gamma$ and γ. This approach looks even more natural, when the logarithmic Sobolev inequality is treated as the relation between classical information-theoretic distances as

$$D(X|Z) \leq \frac{1}{2}\, I(X|Z). \tag{1.2}$$

To clarify this inequality, let us recall standard notations and definitions. If random vectors X and Z in $\mathbf{R}^n$ have distributions μ and ν with densities p and q, and μ is absolutely continuous with respect to ν, the relative entropy of μ with respect to ν is defined by

$$D(X|Z) = D(\mu|\nu) = \int p(x) \, \log \frac{p(x)}{q(x)}\, dx.$$

Moreover, if p and q are smooth, one defines the relative Fisher information

$$I(X|Z) = I(\mu|\nu) = \int \Big|\frac{\nabla p(x)}{p(x)} - \frac{\nabla q(x)}{q(x)}\Big|^2 p(x)\, dx.$$

Both quantities are non-negative, and although non-symmetric in (μ, ν), they may be viewed as strong distances of μ to ν. This is already demonstrated by the well-known Pinsker-type inequality, connecting D with the total variation norm:

$$D(\mu|\nu) \geq \frac{1}{2}\, \|\mu - \nu\|_{\rm TV}^2.$$

In the sequel, we mainly consider the particular case where Z is standard normal, so that $\nu = \gamma$ in the above formulas. And in this case, as easy to see, for $d\mu = f\, d\gamma$ with $\int f\, d\gamma = 1$, the logarithmic Sobolev inequality (1.1) turns exactly into (1.2).

The aim of this note is to develop several lower bounds on the deficit in this inequality, $\frac{1}{2}\, I(X|Z) - D(X|Z)$, by involving also transport metrics such as the quadratic Kantorovich distance

$$W_2(X,Z) = W_2(\mu,\gamma) = \inf_\pi \Big(\iint |x-z|^2 \, d\pi(x,z)\Big)^{1/2}$$

(where the infimum runs over all probability measures on $\mathbf{R}^n \times \mathbf{R}^n$ with marginals μ and γ). More generally, one may consider the optimal transport cost

$$\mathcal{T}(X,Z) = \mathcal{T}(\mu,\gamma) = \inf_\pi \iint c(x-z)\, d\pi(x,z)$$

for various "cost" functions $c(x-z)$.

The metric W_2 is of weak type in the sense that it metrizes the weak topology in the space of probability measures on $\mathbf{R}^n$ (under proper moment constraints). It may be connected with the relative entropy by virtue of M. Talagrand's transport-entropy inequality

$$W_2(X,Z)^2 \leq 2\, D(X|Z), \tag{1.3}$$

cf. [T]. In view of (1.2), this also gives an apriori weaker transport-Fisher information inequality

$$W_2(X,Z) \le \sqrt{I(X|Z)}. \tag{1.4}$$

In formulations below, we use the non-negative convex function

$$\Delta(t) = t - \log(1+t), \qquad t > -1,$$

and denote by Z a random vector in $\mathbf{R}^n$ with the standard normal law.

Theorem 1.1. *For any random vector X in $\mathbf{R}^n$ with a smooth density, such that $I(X|Z)$ is finite,*

$$I(X|Z) - 2D(X|Z) \ \ge\ n\Delta\Big(\frac{I(X)}{n} - 1\Big). \tag{1.5}$$

Moreover,

$$I(X|Z) - 2D(X|Z) \ \ge\ \big(\sqrt{I(X|Z)} - W_2(X,Z)\big)^2 + n\Delta\left(\frac{W_2(X,Z)}{\sqrt{I(X|Z)}}\,\Big(\frac{I(X)}{n} - 1\Big)\right). \tag{1.6}$$

As is common,

$$I(X) = \int \frac{|\nabla p(x)|^2}{p(x)}\,dx$$

stands for the usual (non-relative) Fisher information. Thus, (1.5)-(1.6) represent certain sharpenings of the logarithmic Sobolev inequality. An interesting feature of the bound (1.6) is that, by removing the last term in it, we arrive at the Gaussian case in the so-called HWI inequality due to F. Otto and C. Villani [O-V],

$$D(X|Z) \le W_2(X,Z)\sqrt{I(X|Z)} - \frac{1}{2}\,W_2^2(X,Z). \tag{1.7}$$

As for (1.5), its main point is that, when $\mathbf{E}\,|X|^2 \le n$, then necessarily $I(X) \ge n$, and moreover, one can use the lower bound

$$\frac{1}{n}\,I(X) - 1 = \frac{1}{n}\,I(X|Z) - \frac{1}{n}\,\mathbf{E}\,|X|^2 + 1 \ge \frac{1}{n}\,I(X|Z).$$

Since $\Delta(t)$ is increasing for $t \ge 0$, (1.5) is then simplified to

$$I(X|Z) - 2D(X|Z) \ge n\Delta\Big(\frac{1}{n}\,I(X|Z)\Big). \tag{1.8}$$

In fact, this estimate is rather elementary in that it surprisingly follows from the logarithmic Sobolev inequality itself by virtue of rescaling (as will be explained later on). Here, let us only stress that the right-hand side of (1.8) can further be bounded from below. For example, by (1.2)-(1.3), we have

$$I(X|Z) - 2D(X|Z) \ \ge\ n\Delta\Big(\frac{2}{n}\,D(X,Z)\Big) \ \ge\ n\Delta\Big(\frac{1}{n}\,W_2^2(X,Z)\Big).$$

But, $\frac{1}{n}\,W_2^2(X,Z) \le \frac{1}{n}\,\mathbf{E}\,|X-Z|^2 \le 4$, and using $\Delta(t) \sim \frac{t^2}{2}$ for small t, the above yields a simpler bound.

Corollary 1.2. *For any random vector X in $\mathbf{R}^n$ with a smooth density and such that $\mathbf{E}\,|X|^2 \le n$, we have*

$$I(X|Z) - 2D(X|Z) \ \ge\ \frac{c}{n}\,W_2^4(X,Z), \tag{1.9}$$

up to an absolute constant $c > 0$.

Remark. Dimensional refinements of the HWI inequality (1.7) similar to (1.6) were recently considered by several authors. For instance, F-Y. Wang obtained in [W] some HWI type inequalities involving the dimension and the quadratic Kantorovich distance under the assumption that the reference measure enjoys some curvature dimension condition $\mathrm{CD}(-K, N)$ with $K \geq 0$ and $N \geq 0$ (see [B-E] for the definition). The standard Gaussian measure does not enter directly the framework of [W], but we believe that it might be possible to use similar semigroup arguments to derive (1.6). In the same spirit, D. Bakry, F. Bolley and I. Gentil [B-B-G] used semigroup techniques to prove a dimensional reinforcement of Talagrand's transport-entropy inequality.

Returning to (1.9), we note that, after a certain recentering of X, one may give some refinement over this bound, especially when $D(X|Z)$ is small. Given a random vector X in $\mathbf{R}^n$ with finite absolute moment, define the recentered random vector $\bar{X} = (\bar{X}_1, \dots, \bar{X}_n)$ by putting $\bar{X}_1 = X_1 - \mathbf{E}X_1$ and

$$\bar{X}_k = X_k - \mathbf{E}\,(X_k | X_1, \dots, X_{k-1}), \qquad k \geq 2,$$

where we use standard notations for the conditional expectations.

Theorem 1.3. *For any random vector X in $\mathbf{R}^n$ with a smooth density, such that $I(X|Z)$ is finite, the deficit in* (1.2) *satisfies*

$$\frac{1}{2}\, I(X|Z) - D(X|Z) \geq c\, \frac{\mathcal{T}^2(\bar{X}, Z)}{D(\bar{X}|Z)}. \tag{1.10}$$

Here the optimal transport cost $\mathcal{T}$ corresponds to the cost function $\Delta(|x - z|)$, c is a positive absolute constant and one uses the convention $0/0 = 0$ in the right hand side.

In particular, in dimension one, if a random vector X has mean zero, we get that

$$\frac{1}{2}\, I(X|Z) - D(X|Z) \geq c\, \frac{\mathcal{T}^2(X, Z)}{D(X|Z)}. \tag{1.11}$$

The bound (1.10) allows one to recognize the cases of equality in (1.2) – this is only possible when the random vector X is a translation of the standard random vector Z (an observation of E. Carlen [C] who used a different proof). The argument is sketched in Appendix D.

It is worthwhile noting that the transport distance $\mathcal{T}$ may be connected with the classical Kantorovich transport distance W_1 based on the cost function $c(x, z) = |x - z|$. More precisely, due to the convexity of Δ, there are simple bounds

$$W_1(X, Z) \,\geq\, \mathcal{T}(X, Z) \,\geq\, \Delta(W_1(X, Z)) \sim \min\{W_1(X, Z), W_1^2(X, Z)\}.$$

Hence, if $D(\bar{X}|Z) \leq 1$, then according to (1.3), $W_1^2(X, Z) \leq W_2^2(X, Z) \leq 2$, and (1.10) is simplified to

$$\frac{1}{2}\, I(X|Z) - D(X|Z) \geq c'\, \frac{W_1^4(\bar{X}, Z)}{D(\bar{X}|Z)}, \tag{1.12}$$

for some other absolute constant c'.

In connection with such bounds, let us mention a recent preprint by E. Indrei and D. Marcon [I-M], which we learned about while the current work was in progress. It is proved therein (Theorem 1.1 and Corollary 1.2) that, if a random vector X on $\mathbf{R}^n$ has a smooth density $p = e^{-V}$ satisfying $\varepsilon \le V'' \le M$, for some $0 < \varepsilon < M$, where V'' denotes the matrix of second partial derivatives of V, then

$$\frac{1}{2}\, I(X|Z) - D(X|Z) \,\ge\, c\, W_2^2(X - \mathbf{E}\, X, Z) \tag{1.13}$$

with some constants $c = c(\varepsilon, M)$. In certain cases it is somewhat stronger than (1.11). We will show that a slight adaptation of our proof of (1.11) leads to a bound similar to (1.13).

Theorem 1.4. *Let X be a random vector in $\mathbf{R}^n$ with a smooth density $p = e^{-V}$ with respect to Lebesgue measure such that $\frac{\partial^2 V}{\partial x_i^2} \ge \varepsilon$, for all $i = 1, \dots, n$ with some $\varepsilon > 0$. Then, the deficit in* (1.2) *satisfies*

$$\frac{1}{2}\, I(X|Z) - D(X|Z) \ge c\, \min(1, \varepsilon)\, W_2^2(\bar{X}, Z), \tag{1.14}$$

for some absolute constant c.

Note that Theorem 1.4 holds under less restrictive assumptions on p than the result from [I-M]. In particular, in dimension 1, we see that the constant c in (1.13) can be taken independent on M. In higher dimensions however, it is not clear how to compare $W_2(\bar{X}, Z)$ and $W_2(X - \mathbf{E}\, X, Z)$ in general. One favorable case is, for instance, when the distribution of X is unconditional (i.e., when its density p satisfies $p(x) = p(\varepsilon_1 x_1, \dots, \varepsilon_n x_n)$, for all $x \in \mathbf{R}^n$ and all $\varepsilon_i = \pm 1$). In this case, $\mathbf{E}\, X = 0$ and $\bar{X} = X$, and thus (1.14) reduces to (1.13) with a constant c independent on M.

Let us mention that in Theorem 1.3 of [I-M], the assumption $V'' \le M$ can be relaxed into an integrability condition of the form $\int \|V''\|^r\, dx \le M$, for some $r > 1$, but only at the expense of a constant c depending on the dimension n and of an exponent greater than 2 in the right hand side of (1.13).

Finally, let us conclude this introduction by showing optimality of the bounds (1.11), (1.12), (1.14) for mean zero Gaussian random vectors with variance close to 1. An easy calculation shows that, if Z is a standard Gaussian random vector in $\mathbf{R}^n$, then for any $\sigma > 0$,

$$D(\sigma Z|Z) = \frac{n}{2}\, \big((\sigma^2 - 1) - 2\log\sigma\big)\,, \qquad I(\sigma Z|Z) = n\sigma^2 \Big(\frac{1}{\sigma^2} - 1\Big)^2,$$

so that

$$\frac{1}{2}\, I(X|Z) - D(X|Z) \,=\, \frac{n}{2}\, \Big(\frac{1}{\sigma^2} - 1 + 2\log\sigma\Big) \,\sim\, n(\sigma - 1)^2, \qquad \text{as } \ \sigma \to 1.$$

On the other hand,

$$W_2^2(\sigma Z, Z) = n(\sigma - 1)^2, \qquad W_1(\sigma Z, Z) = |\sigma - 1|\, \mathbf{E}\, |Z| \,\simeq\, |\sigma - 1|\sqrt{n},$$

and thus the three quantities $W_2^2(\sigma Z, Z)$, $\mathcal{T}^2(\sigma Z, Z)/D(\sigma Z|Z)$ and $W_1^4(\sigma Z, Z)/D(\sigma Z|Z)$ are all of the same order $n(\sigma - 1)^2$, when σ goes to 1.

The paper is organized in the following way. In Section 2 we recall Stam's formulation of the logarithmic Sobolev inequality in the form of an "isoperimetric inequality for entropies" and discuss the involved improved variants of (1.1). Theorem 1.1 is proved in Section 3. In Section 4 we consider sharpened transport-entropy inequalities in dimension one, which are used to derive bounds on the deficit like those in (1.11)-(1.13). For general dimensions Theorems 1.3 and 1.4 are proved in Section 5. For the reader's convenience, and so as to get a more self-contained exposition, we move to Appendices several known results and arguments.

2. Self-improvement of the logarithmic Sobolev inequality

To start with, let us return to the history and remind the reader Stam's information-theoretic formulation of the logarithmic Sobolev inequality. As a base for the derivation, one may take (1.2) and rewrite it in terms of the Fisher information $I(X)$ and the (Shannon) entropy

$$h(X) = -\int p(x)\,\log p(x)\,dx,$$

where X is a random vector in $\mathbf{R}^n$ with density p. Here the integral is well-defined, as long as X has finite second moment. Introduce also the entropy power

$$N(X) = \exp\{2h(X)/n\},$$

which is a homogeneous functional of order 2. The basic connections between the relative and non-relative information quantities are given by

$$D(X|Z) = h(Z) - h(X), \qquad I(X|Z) = I(X) - I(Z),$$

where Z has a normal distribution, and provided that $\mathbf{E}\,|X|^2 = \mathbf{E}\,|Z|^2$.

More generally, assuming that Z is standard normal, and $\mathbf{E}\,|X|^2 < \infty$, the first above equality should be replaced with

$$D(X|Z) = -h(X) + \mathbf{E}\left(\frac{n}{2}\,\log(2\pi) + \frac{|X|^2}{2}\right),$$

while, as was mentioned before, under mild regularity assumptions on p,

$$I(X|Z) = I(X) + \mathbf{E}\,|X|^2 - 2n.$$

Inserting these expressions into the inequality (1.2), the second moment is cancelled, and it becomes

$$I(X) + 2h(X) \geq 2n + n\log(2\pi).$$

However, this inequality is not homogeneous in X. So, one may apply it to λX in place of X with arbitrary $\lambda > 0$ and then optimize. The function

$$v(\lambda) = I(\lambda X) + 2h(\lambda X) = \frac{I(X)}{\lambda^2} + n\log\lambda^2 + 2h(X)$$

is minimized for $\lambda^2 = I(X)/n$, and at this point the inequality becomes:

Theorem 2.1 ([S]). *If a random vector X in $\mathbf{R}^n$ has a smooth density and finite second moment, then*

$$I(X)\,\frac{N(X)}{2\pi e} \geq n. \tag{2.1}$$

This relation was obtained by Stam with a different proof and is sometimes referred to as the isoperimetric inequality for entropies, cf. e.g. [D-C-T]. Stam's original argument is based on the general entropy power inequality

$$N(X+Y) \geq N(X) + N(Y), \tag{2.2}$$

which holds for all independent random vectors X and Y in $\mathbf{R}^n$ with finite second moments (so that the involved entropies do exist, cf. also [Bl], [Li]). Then, (2.1) can be obtained by taking $Y = \sqrt{t}\, Z$ with Z having a standard normal law and combining (2.2) with the de Bruijn identity

$$\frac{d}{dt}\, h(X + \sqrt{t}\, Z) = \frac{1}{2}\, I(X + \sqrt{t}\, Z) \qquad (t > 0). \tag{2.3}$$

Note that in the derivation (1.2) $\Rightarrow$ (2.1) the argument may easily be reversed, so these inequalities are in fact equivalent (as noticed by E. Carlen [C]). On the other hand, the isoperimetric inequality for entropies can be viewed as a certain sharpening of (1.1)-(1.2). Indeed, let us rewrite (2.1) explicitly as

$$\int p(x)\, \log p(x)\, dx \, \leq \, \frac{n}{2}\, \log\Big(\frac{1}{2\pi e\, n}\, \int \frac{|\nabla p(x)|^2}{p(x)}\, dx\Big). \tag{2.4}$$

It is also called an optimal Euclidean logarithmic Sobolev inequality; cf. [B-L] for a detail discussion including deep connections with dimensional lower estimates on heat kernel measures. In terms of the density $f(x) = p(x)/\varphi(x)$ of X with respect γ we have

$$\int p(x)\, \log p(x)\, dx = \frac{n}{2}\, \log \frac{1}{2\pi} - \frac{1}{2} \int |x|^2 f(x)\, d\gamma(x) + \int f\, \log f\, d\gamma,$$

while

$$\int \frac{|\nabla p(x)|^2}{p(x)}\, dx = \int \frac{|\nabla f(x)|^2}{f(x)}\, d\gamma(x) - \int |x|^2 f(x)\, d\gamma(x) + 2n.$$

Inserting these two equalities in (2.4), we arrive at the following reformulation of Theorem 2.1.

Corollary 2.2. *For any positive smooth function f on $\mathbf{R}^n$ such that $\int f\, d\gamma = 1$, putting $b = \frac{1}{n} \int |x|^2 f(x)\, d\gamma(x)$, we have*

$$\int f\, \log f\, d\gamma \, \leq \, \frac{n}{2}\, \log\Big(\frac{1}{n} \int \frac{|\nabla f|^2}{f}\, d\gamma + (2 - b)\Big) + \frac{n}{2}\, (b - 1). \tag{2.5}$$

In particular, if $b \leq 1$,

$$\int f\, \log f\, d\gamma \, \leq \, \frac{n}{2}\, \log\Big(\frac{1}{n} \int \frac{|\nabla f|^2}{f}\, d\gamma + 1\Big). \tag{2.6}$$

An application of $\log t \leq t - 1$ on the right-hand side of (2.5) returns us to the original logarithmic Sobolev inequality (1.1). It is in this sense, the inequality (2.5) is stronger, although it was derived on the basis of (1.1). In particular, the point of self-improvement is that the log-value of

$$I = \int \frac{|\nabla f|^2}{f}\, d\gamma$$

may be much smaller than the integral itself. This can be used, for example, in bounding the deficit $\delta(f)$ in (1.1). Indeed, when $b \leq 1$, (2.6) yields

$$2\delta(f) \geq I - n\, \log\Big(\frac{1}{n}\, I + 1\Big).$$

That is, using again the function $\Delta(t) = t - \log(t + 1)$, we have

$$2\delta(f) \geq n\, \Delta\Big(\frac{1}{n} \int \frac{|\nabla f|^2}{f}\, d\gamma\Big).$$

But this is exactly the information-theoretic bound (1.8), mentioned in Section 1 as a direct consequence of (1.5).

3. HWI inequality and its sharpening

We now turn to the remarkable HWI inequality of F. Otto and C. Villani and state it in full generality. Assume the probability measure ν on $\mathbf{R}^n$ has density

$$\frac{d\nu(x)}{dx} = e^{-V(x)}$$

with a twice continuously differentiable $V : \mathbf{R}^n \to \mathbf{R}$. We denote by $V''(x)$ the matrix of second partial derivatives of V at the point x, and use comparison of symmetric matrices in the usual matrix sense. Let $\mathrm{I_n}$ denote the identity $n \times n$ matrix.

Theorem 3.1 ([O-V]). *Assume that $V''(x) \geq \kappa\, \mathrm{I_n}$ for all $x \in \mathbf{R}^n$ with some $\kappa \in \mathbf{R}$. Then, for any probability measure μ on $\mathbf{R}^n$ with finite second moment,*

$$D(\mu|\nu) \leq W_2(\mu|\nu)\sqrt{I(\mu|\nu)} - \frac{\kappa}{2}\, W_2^2(\mu,\nu). \tag{3.1}$$

This inequality connects together all three important distances: the relative entropy (which sometimes is denoted by H), the relative Fisher information I, and the quadratic transport distance W_2. It may equivalently be written as

$$D(\mu|\nu) \leq \frac{1}{2\varepsilon}\, I(\mu|\nu) + \frac{\varepsilon - \kappa}{2}\, W_2^2(\mu,\nu) \tag{3.2}$$

with an arbitrary $\varepsilon > 0$. So, taking $\varepsilon = \kappa$ one gets

$$D(\mu|\nu) \leq \frac{1}{2\kappa}\, I(\mu|\nu). \tag{3.3}$$

If $\nu = \gamma$, we arrive in (3.3) at the logarithmic Sobolev inequality (1.1) for the Gaussian measure, and thus the HWI inequality represents its certain refinement. In particular, (3.1) may potentially be used in the study of the deficit in (1.1), as is pointed in Theorem 1.1.

In the proof of the latter, we will use two results. The following lemma, reversing the transport-entropy inequality, may be found in the survey by Raginsky and Sason [R-S], Lemma 15. It is due to Y. Wu who used it to prove a weak version of the Gaussian HWI inequality (without the curvature term $-\frac{1}{2}W_2^2(X,Z)$ appearing in (1.7)). The proof of Lemma 3.2 is reproduced in Appendix A.

For a random vector X in $\mathbf{R}^n$ with finite second moment, put

$$X_t = X + \sqrt{t}\, Z \qquad (t \geq 0),$$

where Z is a standard normal random vector in $\mathbf{R}^n$, independent of X.

Lemma 3.2. *Given independent random vectors X and Y in $\mathbf{R}^n$ with finite second moments, for all $t > 0$,*

$$D(X_t|Y_t) \leq \frac{1}{2t}\, W_2^2(X,Y).$$

We will also need a convexity property of the Fisher information in the form of the Fisher information inequality. As a full analog of the entropy power inequality (2.2), it was apparently first mentioned by Stam [S].

Lemma 3.3. *Given independent random vectors X and Y in $\mathbf{R}^n$ with smooth densities,*

$$\frac{1}{I(X+Y)} \geq \frac{1}{I(X)} + \frac{1}{I(Y)}. \tag{3.4}$$

Proof of Theorem 1.1. Let Z be standard normal. We recall that, if Y is a normal random vector with mean zero and covariance matrix $\sigma^2 \, \mathrm{I}_n$, then

$$D(X|Y) = h(Y) - h(X) + \frac{1}{2\sigma^2} \left(\mathbf{E}\,|X|^2 - \mathbf{E}\,|Y|^2\right).$$

In particular,

$$D(X|Z) = h(Z) - h(X) + \frac{1}{2} \left(\mathbf{E}\,|X|^2 - \mathbf{E}\,|Z|^2\right),$$

where $\mathbf{E}\,|Z|^2 = n$. Using de-Bruijn's identity (2.3), $\frac{d}{dt}\, h(X_t) = \frac{1}{2}\, I(X_t)$, we therefore obtain that, for all $t > 0$,

$$\begin{aligned} D(X_t|Z_t) &= h(Z_t) - h(X_t) + \frac{1}{2(1+t)} \left(\mathbf{E}\,|X_t|^2 - \mathbf{E}\,|Z_t|^2\right) \\ &= h(Z_t) - h(X_t) + \frac{1}{2(1+t)} \left(\mathbf{E}\,|X|^2 - \mathbf{E}\,|Z|^2\right) \\ &= (h(Z) - h(X)) + \frac{1}{2}\int_0^t (I(Z_\tau) - I(X_\tau))\, d\tau + \frac{1}{2(1+t)} \left(\mathbf{E}\,|X|^2 - \mathbf{E}\,|Z|^2\right) \\ &= D(X|Z) + \frac{1}{2}\int_0^t (I(Z_\tau) - I(X_\tau))\, d\tau - \frac{t}{2(1+t)} \left(\mathbf{E}\,|X|^2 - \mathbf{E}\,|Z|^2\right). \end{aligned}$$

Equivalently,

$$D(X|Z) \,=\, D(X_t|Z_t) + \frac{1}{2}\int_0^t (I(X_\tau) - I(Z_\tau))\, d\tau + \frac{t}{2(1+t)} \left(\mathbf{E}\,|X|^2 - \mathbf{E}\,|Z|^2\right). \tag{3.5}$$

In order to estimate from above the last integral, we apply Lemma 3.3 to the couple $(X, \sqrt{\tau}\, Z)$, which gives

$$I(X_\tau) \leq \frac{1}{\frac{1}{I(X)} + \frac{1}{I(\sqrt{\tau}\, Z)}} = \frac{nI(X)}{n + \tau I(X)}.$$

Inserting also $I(Z_\tau) = \frac{n}{1+\tau}$, we get

$$\begin{aligned} \int_0^t (I(X_\tau) - I(Z_\tau))\, d\tau &\leq \int_0^t \left(\frac{nI(X)}{n + \tau I(X)} - \frac{n}{1+\tau}\right) d\tau \\ &= \frac{n}{2} \log \frac{n + tI(X)}{n(1+t)}. \end{aligned}$$

Thus, from (3.5),

$$D(X|Z) \,\leq\, D(X_t|Z_t) + \frac{n}{2}\, \log \frac{n + tI(X)}{n(1+t)} + \frac{t}{2(1+t)} \left(\mathbf{E}\,|X|^2 - n\right).$$

Furthermore, an application of Lemma 3.2 together with the identity

$$\mathbf{E}\,|X|^2 - n = I(X|Z) - I(X) + n$$

yield

$$D(X|Z) \,\leq\, \frac{1}{2t}\, W_2^2(X, Z) + \frac{n}{2}\, \log \frac{n + tI(X)}{n(1+t)} + \frac{t}{2(1+t)} \left(I(X|Z) - I(X) + n\right). \tag{3.6}$$

As t goes to infinity in (3.6), we get in the limit

$$D(X|Z) \le \frac{1}{2}\, I(X|Z) - \frac{n}{2}\,\Delta\Big(\frac{I(X)}{n} - 1\Big),$$

which is exactly the required inequality (1.5) of Theorem 1.1.

As for (1.6), let as restate (3.5) as the property that the deficit $I(X|Z) - 2D(X|Z)$ is bounded from below by

$$I(X|Z) - \frac{1}{t}\, W_2^2(X,Z) - n\,\log\frac{n + tI(X)}{n(1+t)} - \frac{t}{1+t}\,(I(X|Z) - I(X) + n)\,. \tag{3.7}$$

Assuming that X is not normal, we end the proof by choosing

$$t = \frac{W_2(X,Z)}{\sqrt{I(X|Z)} - W_2(X,Z)}.$$

Indeed, putting for short $W = W_2(X,Z)$, $I = I(X|Z)$, $I_0 = I(X)$, the right-hand side of (3.7) with this value of t turns into

$$\begin{aligned} I - W(\sqrt{I} - W) - n\,\log\frac{1 + \frac{W}{\sqrt{I}-W}\,\frac{I_0}{n}}{\frac{\sqrt{I}}{\sqrt{I}-W}} - \frac{W}{\sqrt{I}}\,(I - I_0 + n) \\ = (\sqrt{I} - W)^2 - n\,\log\Big(1 + \frac{W}{\sqrt{I}}\,\Big(\frac{I_0}{n} - 1\Big)\Big) + n\frac{W}{\sqrt{I}}\,\Big(\frac{I_0}{n} - 1\Big) \\ = (\sqrt{I} - W)^2 + n\Delta\Big(\frac{W}{\sqrt{I}}\,\Big(\frac{I_0}{n} - 1\Big)\Big). \end{aligned}$$

□

4. Sharpened transport-entropy inequalities on the line

Nowadays, Talagrand's transport-entropy inequality (1.2),

$$\frac{1}{2}\, W_2^2(\mu,\gamma) \le D(\mu|\gamma), \tag{4.1}$$

has many proofs (cf. e.g. [B-G]). In the one dimensional case it admits the following refinement, which is due to F. Barthe and A. Kolesnikov.

Theorem 4.1 ([B-K]). *For any probability measure μ on the real line with finite second moment, having the mean or median at the origin,*

$$\frac{1}{2}\, W_2^2(\mu,\gamma) + \frac{1}{4}\, \mathcal{T}'(\mu,\gamma) \le D(\mu|\gamma), \tag{4.2}$$

where the optimal transport cost $\mathcal{T}'$ is based on the cost function $c'(x-z) = \Delta\big(\frac{|x-z|}{\sqrt{2\pi}}\big)$.

It is also shown in [B-K] that the constant $\frac{1}{4}$ may be replaced with 1 under the median assumption. Anyhow, the deficit in (4.1) can be bounded in terms of the transport distance $\mathcal{T}$ which represents a slight weakening of W_2 (since the function $\Delta(t) = t - \log(t+1)$ is almost quadratic near zero).

In Appendix C we remind a basic argument, which actually works well for a larger family of measures ν in place of γ like in Theorem 3.1 (however with $\kappa > 0$). In order to work with the usual cost function $c(x-z) = \Delta(|x-z|)$, the inequality (4.2) will be modified to

$$\frac{1}{2}\,W_2^2(\mu,\gamma) + \frac{1}{8\pi}\,\mathcal{T}(\mu,\gamma) \leq D(\mu|\gamma) \tag{4.3}$$

under the assumption that μ has mean zero.

As a natural complement to Theorem 4.1, it is also shown there that, under an additional log-concavity assumption on μ, the transport distance $\mathcal{T}$ in the inequalities (4.2)-(4.3) may be replaced with W_2. That is, the constant $\frac{1}{2}$ in (4.1) may be increased.

Theorem 4.2. *Suppose that the probability measure μ on the real line has a twice continuously differentiable density $\frac{d\mu(x)}{dx} = e^{-v(x)}$ such that, for a given $\varepsilon > 0$,*

$$v''(x) \geq \varepsilon, \qquad x \in \mathbf{R}. \tag{4.4}$$

If μ has mean at the origin, then with some absolute constant $c > 0$ we have

$$\Big(\frac{1}{2} + c\min\{1,\sqrt{\varepsilon}\,\}\Big)\,W_2^2(\mu,\gamma) \leq D(\mu|\gamma). \tag{4.5}$$

Here, one may take $c = 1 - \log 2$.

Let us now explain how these refinements can be used in the problem of bounding the deficit in the one dimensional logarithmic Sobolev inequality. Returning to (4.3), we are going to combine this bound with the HWI inequality (3.1). Putting

$$W = W_2(\mu,\gamma), \quad D = D(\mu|\gamma), \quad I = I(\mu|\gamma),$$

we rewrite (3.1) as

$$I - 2D \geq (\sqrt{I} - W)^2.$$

On the other hand, applying the logarithmic Sobolev inequality $I \geq 2D$, (4.3) yields $I \geq W^2 + \frac{1}{4\pi}\,\mathcal{T}$, where $\mathcal{T} = \mathcal{T}(\mu,\gamma)$. Hence,

$$I - 2D \geq \bigg(\sqrt{W^2 + \frac{1}{4\pi}\,\mathcal{T}} - W\bigg)^2 = W^2\,\bigg(\sqrt{1 + \frac{\mathcal{T}}{4\pi\,W^2}} - 1\bigg)^2.$$

Here, by the very definition of the transport distance, one has $\mathcal{T} \leq W^2$, so $\varepsilon = \frac{\mathcal{T}}{4\pi\,W^2} \leq \frac{1}{4\pi}$. This implies that $\sqrt{1+\varepsilon} - 1 \geq c\varepsilon$ with $c = 4\pi\,\big(\sqrt{1+\frac{1}{4\pi}} - 1\big)$. Thus, up to a positive numerical constant,

$$D + c\,\frac{\mathcal{T}^2}{W^2} \leq \frac{1}{2}\,I. \tag{4.6}$$

In order to get a more flexible formulation, denote by μ_t the shift of the measure μ,

$$\mu_t(A) = \mu(A - t), \qquad A \subset \mathbf{R} \ \ \text{(Borel)},$$

which is the distribution of the random variable $X + t$ (with fixed $t \in \mathbf{R}$), when X has the distribution μ. As easy to verify,

$$\begin{aligned} D(\mu_t|\gamma) &= D(\mu|\gamma) + \frac{t^2}{2} + t\,\mathbf{E}X, \\ \frac{1}{2}\,I(\mu_t|\gamma) &= \frac{1}{2}\,I(\mu|\gamma) + \frac{t^2}{2} + t\,\mathbf{E}X. \end{aligned}$$

Hence, the deficit

$$\delta(\mu) = \frac{1}{2}\, I(\mu|\gamma) - D(\mu|\gamma)$$

in the logarithmic Sobolev inequality (1.2) is translation invariant: $\delta(\mu_t) = \delta(\mu)$. Applying (4.6) to μ_t with $t = -\int x\, d\mu(x)$, so that μ_t would have mean zero, therefore yields:

Corollary 4.3. *For any non-Gaussian probability measure μ on the real line with finite second moment, up to an absolute constant $c > 0$,*

$$D(\mu|\gamma) + c\, \frac{\mathcal{T}^2(\mu_t, \gamma)}{W_2^2(\mu_t, \gamma)} \le \frac{1}{2}\, I(\mu|\gamma), \tag{4.7}$$

where the optimal transport cost $\mathcal{T}$ is based on the cost function $\Delta(|x - z|)$, and where t is the mean of μ. In particular,

$$D(\mu|\gamma) + \frac{c}{2}\, \frac{\mathcal{T}^2(\mu_t, \gamma)}{D(\mu_t|\gamma)} \le \frac{1}{2}\, I(\mu|\gamma). \tag{4.8}$$

Here the second inequality follows from the first one by using $W_2^2 \le 2D$. It will be used in the next section to perform tensorisation for a multidimensional extension. Note that (4.8) may be derived directly from (4.3) with similar arguments. Indeed, one can write

$$\begin{aligned} I - 2D \ &\ge\ (\sqrt{I} - W)^2 \ \ge\ (\sqrt{2D} - W)^2 \\ &=\ \frac{(2D - W^2)^2}{(\sqrt{2D} + W)^2} \ \ge\ \frac{(2D - W^2)^2}{(2\sqrt{2D})^2} \ \ge\ \frac{\mathcal{T}^2}{128\,\pi^2 D^2}, \end{aligned}$$

thus proving (4.8) with constant $c = 1/(128\,\pi^2)$.

Let us now turn to Theorem 4.2 with its additional hypothesis (4.4). Note that the property $v'' \ge 0$ describes the so-called log-concave probability distributions on the real line (with C^2-smooth densities), so (4.4) represents its certain quantitative strengthening. It is also equivalent to the property that X has a log-concave density with respect to the Gaussian measure with mean zero and variance ε.

Arguing as before, from (4.5) we have

$$I - 2D \ge W^2 \left(\sqrt{1 + c\, \min\{1, \sqrt{\varepsilon}\}} - 1\right)^2.$$

Hence, we obtain:

Corollary 4.4. *Let μ be a probability measure on the real line with mean zero, and satisfying (4.4) with some $\varepsilon > 0$. Then, up to an absolute constant $c > 0$,*

$$D(\mu|\gamma) + c\, \min\{1, \varepsilon\}\, W_2^2(\mu, \gamma) \le \frac{1}{2}\, I(\mu|\gamma), \tag{4.9}$$

5. Proof of Theorems 1.3 and 1.4

As the next step, it is natural to try to tensorize the inequality (4.7) so that to extend it to the multidimensional case.

If $x = (x_1, \dots, x_n) \in \mathbf{R}^n$, denote by $x_{1:i}$ the subvector $(x_1, \dots, x_i)$, $i = 1, \dots, n$. Given a probability measure μ on $\mathbf{R}^n$, denote by μ_1 its projection to the first coordinate, i.e., $\mu_1(A) = \mu(A \times \mathbf{R}^{n-1})$, for Borel sets $A \subset \mathbf{R}$. For $i = 2, \dots, n$, let $\mu_i(dx_i|x_{1:i})$ denote the conditional distribution of the i-th coordinate under μ knowing the first $i-1$ coordinates $x_1, \dots, x_{i-1}$. Under mild regularity assumptions on μ, all these conditional measures are well-defined, and we have a general formula for the "full expectation"

$$\int f(x)\,\mu(dx) = \int f(x_1, \dots, x_n)\,\mu_n(dx_n|x_{1:n-1}) \dots \mu_2(dx_2|x_1)\mu_1(dx_1). \tag{5.1}$$

For example, it suffices to require that μ has a smooth positive density, which is polynomially decaying at infinity. Then we will say that μ is regular. In many inequalities, the regularity assumption is only technical for purposes of the proof, and may easily be omitted in the resulting formulations.

The distance functionals D, I, and $\mathcal{T}$ satisfy the following tensorization relations with respect to product measures similarly to (5.1). To emphasize the dimension, we denote by γ_n the standard Gaussian measure on $\mathbf{R}^n$.

Lemma 5.1. *For any regular probability measure μ on $\mathbf{R}^n$ with finite second moment,*

$$\begin{aligned} D(\mu|\gamma_n) &= D(\mu_1|\gamma_1) + \sum_{i=2}^n \int D(\mu_i(\,\cdot\,|x_{1:i-1})\,|\gamma_1)\,\mu(dx), \\ I(\mu|\gamma_n) &\geq I(\mu_1|\gamma_1) + \sum_{i=2}^n \int I(\mu_i(\,\cdot\,|x_{1:i-1})\,|\gamma_1)\,\mu(dx), \\ \mathcal{T}(\mu,\gamma_n) &\leq \mathcal{T}(\mu_1,\gamma_1) + \sum_{i=2}^n \int \mathcal{T}(\mu_i(\,\cdot\,|x_{1:i-1}),\gamma_1)\,\mu(dx). \end{aligned}$$

This statement is rather simple, and so we omit the proof. It remains to hold also for other product references measures ν^n on $\mathbf{R}^n$ in place of γ_n (with necessary regularity assumptions for the case of Fisher information).

Applying the first two inequalities, we see that the deficit δ satisfies a similar property,

$$\delta(\mu) \,\geq\, \delta(\mu_1) + \sum_{i=2}^n \int \delta(\mu_i(\,\cdot\,|x_{1:i-1}))\,\mu(dx). \tag{5.2}$$

Proof of Theorem 1.3. Let us apply the one dimensional result (4.8) with constant $c = 1/(128\,\pi^2)$ in (5.2) to the measures μ_1 and $\mu_i(\,\cdot\,|x_{1:i-1})$. Put $t_1 = \int x_1\,\mu_1(dx_1)$, and

$$t_i(x) = t_i(x_1, \dots, x_{i-1}) = \int x_i\,\mu_i(dx_i|x_{1:i-1}), \qquad x = (x_1, \dots, x_n) \in \mathbf{R}^n,$$

and denote by $\tilde{\mu}_i(\,\cdot\,|x_{1:i-1})$ the corresponding shift of $\mu_i(\,\cdot\,|x_{1:i-1})$ as in Corollary 4.3. Then we have

$$256\,\pi^2\delta(\mu) \,\geq\, \frac{\mathcal{T}^2(\tilde{\mu}_1,\gamma)}{D(\tilde{\mu}_1|\gamma)} + \sum_{i=2}^n \int \frac{\mathcal{T}^2(\tilde{\mu}_i(\,\cdot\,|x_{1:i-1}),\gamma)}{D(\tilde{\mu}_i(\,\cdot\,|x_{1:i-1})|\gamma)}\,\mu(dx).$$

But the function $\psi(u,v) = u^2/v$ is convex in the upper half-plane $u \in \mathbf{R}$, $v \geq 0$. So, by Jensen's inequality,

$$\begin{aligned} 256\,\pi^2\delta(\mu) &\geq \frac{\mathcal{T}^2(\tilde{\mu}_1,\gamma)}{D(\tilde{\mu}_1|\gamma)} + \sum_{i=2}^n \frac{\big(\int \mathcal{T}(\tilde{\mu}_i(\,\cdot\,|x_{1:i-1}),\gamma)\,\mu(dx)\big)^2}{\int D(\tilde{\mu}_i(\,\cdot\,|x_{1:i-1})|\gamma)\,\mu(dx)} \\ &\geq \frac{\big(\mathcal{T}(\tilde{\mu}_1,\gamma) + \sum_{i=2}^n \int \mathcal{T}(\tilde{\mu}_i(\,\cdot\,|x_{1:i-1}),\gamma)\,\mu(dx)\big)^2}{D(\tilde{\mu}_1|\gamma) + \sum_{i=2}^n \int D(\tilde{\mu}_i(\,\cdot\,|x_{1:i-1})|\gamma)\,\mu(dx)}, \end{aligned}$$

where the last bound comes from the inequality

$$\sum_{i=1}^n \psi(u_i, v_i) \geq \psi\Big(\sum_{i=1}^n u_i, \sum_{i=1}^n v_i\Big),$$

which is due to the convexity of ψ and its 1-homogeneity.

Now consider the map $T : \mathbf{R}^n \to \mathbf{R}^n$ defined for all $x \in \mathbf{R}^n$ by

$$T(x) = [x_1 - t_1, x_2 - t_2(x_1), \dots, x_n - t_n(x_1, x_2, \dots, x_{n-1})].$$

By definition, T pushes forward μ onto $\bar{\mu}$. The map T is invertible and its inverse $U = (u_1, \dots, u_n)$ satisfies

$$\begin{aligned} u_1(x) &= x_1 + t_1, \\ u_2(x) &= x_2 + t_2(u_1(x)), \\ &\vdots \\ u_i(x) &= x_i + t_i(u_1(x), \dots, u_{i-1}(x)), \\ &\vdots \\ u_n(x) &= x_n + t_n(u_1(x), \dots, u_{n-1}(x)). \end{aligned}$$

It is not difficult to check that $\bar{\mu}_1 = \tilde{\mu}_1$ and for all $i \geq 2$, $\bar{\mu}_i(\,\cdot\,|x_{1:i-1}) = \tilde{\mu}_i(\,\cdot\,|u_1(x), \dots, u_{k-1}(x))$. Therefore, since U pushes forward $\bar{\mu}$ onto μ,

$$\begin{aligned} &\mathcal{T}(\tilde{\mu}_1, \gamma) + \sum_{i=2}^n \int \mathcal{T}(\tilde{\mu}_i(\,\cdot\,|x_{1:i-1}), \gamma)\, \mu(dx) \\ &\quad = \mathcal{T}(\bar{\mu}_1, \gamma) + \sum_{i=2}^n \int \mathcal{T}(\tilde{\mu}_i(\,\cdot\,|u_1(x), \dots, u_{i-1}(x)), \gamma)\, \bar{\mu}(dx) \\ &\quad = \mathcal{T}(\bar{\mu}_1, \gamma) + \sum_{i=2}^n \int \mathcal{T}(\bar{\mu}_i(\,\cdot\,|x_{1:i-1}), \gamma)\, \bar{\mu}(dx) \;\geq\; \mathcal{T}(\bar{\mu}, \gamma_n), \end{aligned}$$

where we made use of Lemma 5.1 on the last step. The same with equality sign holds true for the D-functional. As a result, in terms of the recentered measure $\bar{\mu}$, we arrive at the following

$$D(\mu|\gamma) + \frac{1}{256\,\pi^2}\, \frac{\mathcal{T}^2(\bar{\mu}, \gamma)}{D(\bar{\mu}|\gamma)} \leq \frac{1}{2}\, I(\mu|\gamma), \tag{5.3}$$

Thus, we have established in (5.3) the desired inequality (1.10) with constant $c = \frac{1}{256\,\pi^2}$. $\square$

Remark 5.3. On order to relate the transport distance $\mathcal{T}$ to W_1, one may use the convexity of the function $\Delta(t) = t - \log(1+t)$ together with a simple lower bound $\Delta(t) \geq (1 - \log 2)\, \min\{t, t^2\}$ for $t \geq 0$. Applying Jensen's inequality, we therefore obtain that, for any random variable $\xi \geq 0$,

$$\mathbf{E}\Delta(\xi) \geq \Delta(\mathbf{E}\xi) \geq (1 - \log 2)\, \min\{\mathbf{E}\xi, (\mathbf{E}\xi)^2\}.$$

On the other hand, $\Delta(t) \leq t$, so $\mathbf{E}\Delta(\xi) \leq \mathbf{E}\xi$. Thus, by the very definition of the transport distances,

$$(1 - \log 2)\, \min\{W_1(\mu, \nu), W_1^2(\mu, \nu)\} \;\leq\; \mathcal{T}(\mu, \nu) \;\leq\; W_1(\mu, \nu),$$

for all probability measures μ and ν on $\mathbf{R}^n$.

Proof of Theorem 1.4. The proof is completely similar. The main point is that if μ has a smooth density $f = e^{-V}$ with respect to Lebesgue, with a V such that $(\partial/\partial x_i)^2\, V \geq \varepsilon$, for some $\varepsilon > 0$, then the first marginal μ_1 has a density of the form e^{-v_1} with $v_1'' \geq \varepsilon$ and for all $i \geq 2$ and all $x_1, \ldots, x_{i-1}$ the one dimensional conditional probability $\mu_i(\,\cdot\,|x_{1:i-1})$ has a density $e^{-v_i(x_i|x_{1:i-1})}$ with $(\partial/\partial x_i)^2\, v_i(x_i|x_{1:i-1}) \geq \varepsilon$. Indeed, by definition of conditional probabilities, it holds

$$v_i(x_i|x_{1:i-1}) = -\log\left(\int e^{-V(x_{1:i},y_{i+1:n})}\, dy_{i+1}\cdots dy_n\right) + w(x_{1:i-1}),$$

where $w(x_{1:i-1}) = \log\left(\int e^{-V(x_{1:i-1},y_{i:n})}\, dy_i dy_{i+1}\cdots dy_n\right)$ does not depend on x_i. A straightforward calculation shows that

$$\left(\frac{\partial}{\partial x_i}\right)^2 v_i(x_i|x_{1:i-1}) = \frac{\int \left(\frac{\partial}{\partial x_i}\right)^2 V(x_{1:i}, y_{i+1:n}) e^{-V(x_{1:i},y_{i+1:n})}\, dy_{i+1}\cdots dy_n}{\int e^{-V(x_{1:i},y_{i+1:n})}\, dy_{i+1}\ldots dy_n} \geq \varepsilon.$$

A similar calculation holds for μ_1. Therefore, μ_1 and the conditional probabilities $\mu_i(\,\cdot\,|x_{1:i-1})$ verify the assumption of Corollary 4.4. Thus, applying the tensorisation formula (5.2), we see that

$$\delta(\mu) \geq c\min\{1,\varepsilon\}\left(W_2^2(\tilde{\mu}_1,\gamma) + \sum_{i=2}^{n} W_2^2(\tilde{\mu}_i(\,\cdot\,|x_{1:i-1}),\gamma)\right),$$

where, as before, $\tilde{\mu}_i(\,\cdot\,|x_{1:i-1})$ is the shifting of $\mu_i(\,\cdot\,|x_{1:i-1})$ by its mean. Reasoning as in the proof of Theorem 1.3, one sees that the quantity into brackets is bounded from below by $W_2^2(\bar{\mu},\gamma_n)$, which completes the proof. □

6. Appendix A: The reversed transport-entropy inequality

Here we include a simple proof of the general inequality of Lemma 3.2,

$$D(X_t|Y_t) \leq \frac{1}{2t}\, W_2^2(X,Y), \qquad t > 0,$$

where X and Y are independent random vectors in $\mathbf{R}^n$ with finite second moments.

We denote by p_U the density of a random vector U and by $p_{U|V=v}$ the conditional density of U knowing the value of a random vector $V = v$. Note that the regularized random vectors $X_t = X + \sqrt{t}\, Z$ have smooth densities.

By the chain rule formula for the relative entropy, one has

$$D(X,Y,X_t|X,Y,Y_t) = D(X_t|Y_t) + \int D(p_{X,Y|X_t=v}|p_{X,Y|Y_t=v})\, p_{X_t}(v)\, dv,$$

and therefore

$$D(X,Y,X_t|X,Y,Y_t) \geq D(X_t|Y_t).$$

On the other hand, we also have

$$D(X,Y,X_t|X,Y,Y_t) = \iint D(p_{X_t|(X,Y)=(x,y)}|p_{Y_t|(X,Y)=(x,y)})\, p_{X,Y}(x,y)\, dxdy.$$

Now observe that $p_{X_t|(X,Y)=(x,y)}$ is the density of a normal law with mean x and covariance matrix tI_n, and similarly for $p_{Y_t|(X,Y)=(x,y)}$. But

$$D(x + \sqrt{t}Z\,|\,y + \sqrt{t}\,Z) = \frac{|x-y|^2}{2t},$$

so

$$D(X,Y,X_t|X,Y,Y_t) = \frac{1}{2t}\iint |x-y|^2 p_{X,Y}(x,y)\,dxdy = \frac{1}{2t}\,W_2^2(X,Y),$$

where the last equality follows by an optimal choice for the coupling density of X and Y.

7. Appendix B: Refinement of HWI with three measures

While there are several different proofs of Theorem 3.1, it also follows from an interesting relation due to D. Cordero-Erausquin, involving three measures. Let the probability measure ν on $\mathbf{R}^n$ have density $\frac{d\nu(x)}{dx} = e^{-V(x)}$ with a twice continuously differentiable $V : \mathbf{R}^n \to \mathbf{R}$.

Theorem 7.1 ([CE]). *Assume that $V''(x) \geq \kappa\,\mathrm{I_n}$ for all $x \in \mathbf{R}^n$ with some $\kappa \in \mathbf{R}$. Then, for all absolutely continuous probability measures μ_1 and μ_2 on $\mathbf{R}^n$,*

$$D(\mu_2|\nu) \leq D(\mu_1|\nu) + \int \langle \nabla h_1(x), T(x)-x\rangle\; d\nu(x) + \frac{\kappa}{2}\int |T(x)-x|^2\,d\nu(x), \tag{7.1}$$

where $T : \mathbf{R}^n \to \mathbf{R}^n$ is the Brenier map transporting μ_1 to μ_2, and $h = \frac{d\mu_1}{d\nu}$.

The Brenier map is of the form $T = \nabla f$, where f is a suitable unique convex function on $\mathbf{R}^n$ (cf. [MC] for a detail exposition).

Below we sketch the proof of Theorem 7.1 without discussing the regularity questions. Let p_i (resp. q_i) be the density of μ_i with respect to Lebesgue measure (resp. ν). First write the change of the variable formula (Monge-Ampère equation):

$$p_1(x) = p_2(Tx)\,|T'x|,$$

where $|T'(x)|$ stands for the Jacobian of T (i.e. an absolute value of the determinant of the matrix of the first partial derivatives of T). It follows that

$$q_1(x)e^{-V(x)} = q_2(Tx)e^{-V(Tx)}\,|T'x|,$$

so taking the log and integrating with respect to μ_1, it holds

$$\begin{aligned}\int \log q_1(x)\,d\mu_1(x) - \int V(x)\,d\mu_1(x) \quad &= \quad \int \log q_2(Tx)\,d\mu_1(x) - \int V(Tx)\,d\mu_1(x)\\ &\quad + \int \log|T'x|\,d\mu_1(x).\end{aligned}$$

By the very definition of the relative entropy, this is the same as

$$D(\mu_2|\nu) = D(\mu_1|\nu) + \int (V(Tx) - V(x))\,d\mu_1(x) - \int \log|T'x|\,d\mu_1(x).$$

By the assumption on V, for all $x, y \in \mathbf{R}^n$,

$$V(y) - V(x) + \langle \nabla V(x), y-x\rangle + \frac{\kappa}{2}\int |y-x|^2,$$

so

$$\begin{aligned}D(\mu_2|\nu) \quad \geq \quad & D(\mu_1|\nu) + \int \langle \nabla V(x), Tx - x\rangle\; d\mu_1(x)\\ & + \frac{\kappa}{2}\int |Tx-x|^2\,d\mu_1(x) - \int \log|T'x|\,d\mu_1(x).\end{aligned}$$

Next, integrating by parts, writing $T = (T_1, \dots, T_n)$, for the first integral we have

$$\begin{aligned}\int \langle \nabla V(x), Tx - x\rangle \, d\mu_1(x) \ &= \ \int \langle \nabla V(x), Tx - x\rangle \, q_1(x)\, e^{-V(x)}\, dx \\ &= \ \int \sum_{i=1}^n (\partial_i T(x) - 1)\, q_1(x)\, e^{-V(x)}\, dx + \int \sum_{i=1}^n (T_i x - x_i)\, \partial_i q_1(x)\, e^{-V(x)}\, dx \\ &= \ \int \mathrm{Tr}\,(T'x - \mathrm{I}_n)\, d\mu_1(x) + \int \langle \nabla q_1(x), Tx - x\rangle \, d\nu(x).\end{aligned}$$

Thus,

$$\begin{aligned}D(\mu_2|\nu) \ &\geq \ D(\mu_1|\nu) + \int \langle \nabla q_1(x), Tx - x\rangle \, d\nu(x) \\ &\quad + \frac{\kappa}{2} \int |Tx - x|^2 \, d\mu_1(x) + \int \big(\mathrm{Tr}\,(T'x - \mathrm{I}_n) - \log|T'x|\big)\, d\mu_1(x). \end{aligned} \tag{7.2}$$

But the eigenvalues $\lambda_1(x), \dots, \lambda_n(x)$ of $T'(x)$ are non-negative (as eigenvalues of the gradient of a convex function). Hence, using $\log t \leq t - 1$,

$$\log|T'x| = \sum_{i=1}^n \log \lambda_i(x) \leq \sum_{i=1}^n (\lambda_i(x) - 1) = \mathrm{Tr}\,(T'x - \mathrm{I}_n).$$

Therefore, the last term in (7.2) is non-negative, and (7.1) follows. □

8. Appendix C: Reinforced transport-entropy inequalities

Finally, let us explain how to derive Theorem 4.1 in the form (4.3). This may be done in a more general situation by using the previous transport argument resulted in the above inequality (7.2) in dimension one.

Let the probability measure ν on the real line have the density $\frac{d\nu(x)}{dx} = e^{-V(x)}$ with a twice continuously differentiable "potential" $V : \mathbf{R} \to \mathbf{R}$, such that $V''(x) \geq \kappa$, for all $x \in \mathbf{R}$ with some $\kappa > 0$. Choosing the measures $\mu_1 = \nu$ and $\mu_2 = \mu$, we have $D(\mu_1|\nu) = 0$ and $q_1(x) = 1$, so $\nabla q_1(x) = q_1'(x) = 0$. Hence, (7.2) reads

$$\begin{aligned}D(\mu|\nu) \ &\geq \ \frac{\kappa}{2}\, W_2^2(\mu,\nu) + \int \big(T'x - 1 - \log T'x)\big)\, d\nu(x) \\ &= \ \frac{\kappa}{2}\, W_2^2(\mu,\nu) + \int \Delta(T'x - 1\big)\, d\nu(x).\end{aligned} \tag{8.1}$$

Since $V'' \geq \kappa$, the measure ν may be obtained as the image of the standard Gaussian measure γ under an increasing map whose Lipschitz norm $\leq 1/\sqrt{\kappa}$. As a one dimensional statement, it is rather simple and can be verified directly. On the other hand, γ is known to satisfy the Cheeger-type analytic inequality

$$\lambda \int |f - m(f)|\, d\gamma \leq \int |f'|\, d\gamma$$

with optimal constant $\lambda = \sqrt{\frac{2}{\pi}}$. Here, $f : \mathbf{R} \to \mathbf{R}$ may be an arbitrary locally Lipschitz function with Radon-Nikodym derivative f', and $m(f)$ denotes a median of f under γ. Hence,

$$\lambda\sqrt{\kappa} \int |f - m(f)|\, d\nu \leq \int |f'|\, d\nu \tag{8.2}$$

with the median functional understood with respect to ν. According to Theorem 3.1 of [B-H], (8.2) can be generalized as

$$\int L(f - m(f))\, d\nu \leq \int L(c_L f'/(\lambda\sqrt{\kappa}))\, d\nu \tag{8.3}$$

with an arbitrary even convex function $L : \mathbf{R} \to [0, \infty)$, such that $L(0) = 0$, $L(t) > 0$ for $t > 0$, and

$$c_L = \sup_{t>0} \frac{tL'(t)}{L(t)} < \infty,$$

where $L'(t)$ may be understood as the right derivative at t.

We apply (8.3) with $L(t) = \Delta(|t|) = |t| - \log(1 + |t|)$ in which case $c_L = 2$, so that

$$\int \Delta(f - m(f))\, d\nu \leq \int \Delta(2f'/(\lambda\sqrt{\kappa}))\, d\nu. \tag{8.4}$$

It will be convenient to replace here the median with the mean $\nu(f) = \int f\, d\nu$. First observe that, by Jensen's inequality, (8.4) yields

$$\Delta(\nu(f) - m(f)) \leq \int \Delta(2f'/(\lambda\sqrt{\kappa}))\, d\nu. \tag{8.5}$$

Hence, using once more the convexity of Δ together with (8.4)-(8.5) for the function $2f$, we get

$$\begin{aligned} \int \Delta(f - \nu(f))\, d\nu \;&\leq\; \frac{1}{2} \int \Delta\big(2(f - m(f))\big)\, d\nu + \frac{1}{2}\, \Delta\big(2(\nu(f) - m(f))\big) \\ &\leq\; \int \Delta(4f'(\lambda\sqrt{\kappa}))\, d\nu. \end{aligned}$$

Equivalently,

$$\int \Delta(f')\, d\nu \geq \int \Delta\Big(\frac{\lambda\sqrt{\kappa}}{4}\, (f - \nu(f)\Big)\, d\nu.$$

To further simplify, one may use an elementary bound

$$\Delta(ct) \geq \min(c, c^2)\, \Delta(t),$$

holding for all $t \geq 0$ and $c \geq 0$. This follows easily from the representation

$$\Delta(ct) = \int_0^{ct} \Delta'(s)\, ds = \int_0^{ct} \frac{s}{1+s}\, ds = c^2 \int_0^t \frac{u}{1+cu}\, du.$$

Hence, we get

$$\int \Delta(f')\, d\nu \geq \min\Big(\frac{\lambda\sqrt{\kappa}}{4}, \Big(\frac{\lambda\sqrt{\kappa}}{4}\Big)^2\Big) \int \Delta(f - \gamma(f))\, d\nu.$$

It remains to apply the latter with $f(x) = Tx - x$ when estimating the last integral in (8.1). This thus gives

$$D(\mu|\nu) \geq \frac{1}{2}\, W_2^2(\mu, \nu) + \min\Big(\frac{\lambda\sqrt{\kappa}}{4}, \Big(\frac{\lambda\sqrt{\kappa}}{4}\Big)^2\Big) \int \Delta(Tx - x)\, d\nu(x),$$

where, since μ has mean zero, the last integral equals $\mathcal{T}(\mu, \nu)$. Let us summarize.

Theorem 8.1. *Let ν be a probability measure on the real line with density $\frac{d\nu(x)}{dx} = e^{-V(x)}$ such that $V''(x) \geq \kappa$, for all $x \in \mathbf{R}$ with some $\kappa > 0$. Then, for any probability measure μ on $\mathbf{R}$ with mean zero,*

$$D(\mu|\nu) \geq \frac{1}{2}\, W_2^2(\mu, \nu) + \min\big(\lambda\sqrt{\kappa}, (\lambda\sqrt{\kappa})^2\big)\, \mathcal{T}(\mu, \nu),$$

where $\lambda = \sqrt{\frac{1}{8\pi}}$.

In the Gaussian case $\nu = \gamma$, we have $\kappa = 1$, and the above inequality turns into (4.3).

Proof of Theorem 4.2. In this (Gaussian) case, let us return to the inequality (8.1), i.e.,

$$D(\mu|\gamma) \geq \frac{1}{2}\, W_2^2(\mu,\gamma) + \int \Delta\big(T'x - 1\big)\, d\gamma(x). \tag{8.6}$$

The basic assumption (4.4) ensures that T has a Lipschitz norm $\leq \frac{1}{\sqrt{\varepsilon}}$, so $T'x \leq \frac{1}{\sqrt{\varepsilon}}$. But, on bounded intervals the function Δ satisfies

$$\Delta(t) \geq ct^2, \qquad -1 < t \leq a \ \ (a \geq 0).$$

More precisely, $\Delta(t) \geq \frac{1}{2}\, t^2$ for $t \leq 0$, while for $0 \leq t \leq a$ the optimal value of the constant c corresponds to the endpoint $t = a$, which is due to the concavity of the function $\Delta(\sqrt{x})$. Thus, for all $0 \leq t \leq a$, we have $\Delta(t) \geq \frac{\Delta(a)}{a^2}\, t^2$. Using these bounds in (8.6), we obtain that

$$D(\mu|\gamma) \geq \frac{1}{2}\, W_2^2(\mu,\gamma) + c(\varepsilon) \int (T'x - 1)^2\, d\gamma(x), \tag{8.7}$$

where

$$c(\varepsilon) = \frac{1}{2}, \ \ \text{for } \varepsilon \geq 1, \qquad c(\varepsilon) = \frac{\Delta(\frac{1}{\sqrt{\varepsilon}} - 1)}{(\frac{1}{\sqrt{\varepsilon}} - 1)^2}, \ \ \text{for } 0 < \varepsilon < 1.$$

On the other hand, applying the Poincaré-type inequality for the Gaussian measure

$$\mathrm{Var}_\gamma(f) \leq \int f'^2\, d\gamma$$

with $f(x) = Tx - x$, together with the assumption that $\int x\, d\mu(x) = \int Tx\, d\gamma(x) = 0$, the last integral in (8.7) can be bounded from below by

$$\int (Tx - x)^2\, d\gamma(x) = W_2^2(\mu,\gamma).$$

It remains to use, for $0 < \varepsilon < 1$, the bound $\Delta(a) \geq (1 - \log 2)\, \min\{a, a^2\}$. The inequality (4.5) is proved. ☐

9. Appendix D: Equality cases in the logarithmic Sobolev inequality for the standard Gaussian measure

In this last section, we show how Theorem 1.3 can be used to recover the following result by E. Carlen.

Theorem 9.1. *Let μ be a probability measure on $\mathbf{R}^n$ such that $D(\mu|\gamma) < \infty$. We have*

$$D(\mu|\gamma) = \frac{1}{2}\, I(\mu|\gamma),$$

if and only if μ is a translation of γ.

In what follows, we denote by $\mathcal{S}_n$ the set of permutations of $\{1,\dots,n\}$. If μ is a probability measure on $\mathbf{R}^n$, we denote by μ_σ its image under the permutation map

$$(x_1,\dots,x_n) \mapsto (x_{\sigma(1)},\dots,x_{\sigma(n)}).$$

If μ has density f with respect to the standard n-dimensional Gaussian measure γ, then the density of μ_σ with respect to γ is given by

$$f_\sigma(x_1,\dots,x_n)=f(x_{\sigma^{-1}(1)},\dots,x_{\sigma^{-1}(n)}).$$

Obviously,

$$I(\mu_\sigma|\gamma)=I(\mu|\gamma) \quad \text{and} \quad D(\mu_\sigma|\gamma)=D(\mu|\gamma).$$

Hence, we have the following automatic improvement of Theorem 1.3.

Theorem 9.2. *Let X be a random vector in $\mathbf{R}^n$ with law μ. Then,*

$$D(\mu|\gamma)+c\max_{\sigma\in\mathcal{S}_n}\frac{\mathcal{T}^2(\overline{\mu_\sigma},\gamma)}{D(\overline{\mu_\sigma}|\gamma)}\leq \frac{1}{2}I(\mu|\gamma),$$

where $\overline{\mu_\sigma}$ is the law of the random vector Y^σ defined by

$$Y_i^\sigma=X_{\sigma(i)}-\mathbf{E}(X_{\sigma(i)}|X_{\sigma(1)},\dots,X_{\sigma(i-1)}).$$

Proof of Theorem 9.1. To avoid complicated notations, we will restrict ourselves to the dimension $n=2$. We may assume that μ has a smooth density p with respect to the Lebesgue measure such that $D(\mu|\gamma)=\frac{1}{2}\,I(\mu|\gamma)<\infty$. Necessarily, μ has a finite second moment, and moreover, $\overline{\mu_\sigma}=\gamma$, for all $\sigma\in\mathcal{S}_2$, i.e., for $\sigma=\mathrm{id}=(12)$ and $\sigma=(21)$.

For a random vector X with law μ, put $m_1=\mathbf{E}X_1$, $m_2=\mathbf{E}X_2$, $a(X_1)=\mathbf{E}\,(X_2|X_1)$ and $b(X_2)=\mathbf{E}\,(X_1|X_2)$. The probability measure $\gamma=\overline{\mu_{\mathrm{id}}}$ represents the image of μ under the map $(x_1,x_2)\mapsto(x_1-m_1,x_2-a(x_1))$. It then easily follows that

$$p(x_1,x_2)=\frac{1}{2\pi}\exp\left(-\frac{1}{2}\,(x_1-m_1)^2-\frac{1}{2}\,(x_2-a(x_1))^2\right)$$

for almost all $(x_1,x_2)\in\mathbf{R}^2$. Since also $\gamma=\overline{\mu_{(2,1)}}$, the same reasoning yields

$$p(x_1,x_2)=\frac{1}{2\pi}\exp\left(-\frac{1}{2}\,(x_2-m_2)^2-\frac{1}{2}\,(x_1-b(x_2))^2\right),$$

for almost all $(x_1,x_2)\in\mathbf{R}^2$. Therefore, for almost all $(x_1,x_2)\in\mathbf{R}^2$, it holds

$$(x_1-m_1)^2+(x_2-a(x_1))^2=(x_2-m_2)^2+(x_1-b(x_2))^2.$$

Let us denote by A the set of all couples (x_1,x_2) for which there is equality, and for $x_1\in\mathbf{R}$, let $A_{x_1}=\{x_2\in\mathbf{R}:(x_1,x_2)\in A\}$ denote the corresponding section of A. By Fubini's theorem,

$$0=|\mathbf{R}^2\setminus A|=\int_{-\infty}^{\infty}|\mathbf{R}\setminus A_{x_1}|\,dx_1,$$

where $|\cdot|$ stands for the Lebesgue measure of a set in the corresponding dimension. Hence, for almost all x_1, the set $\mathbf{R}\setminus A_{x_1}$ is of Lebesgue measure 0. For any such x_1,

$$2x_2(m_2-a(x_1))+a(x_1)^2-m_2^2+(x_1-m_1)^2\geq 0,\qquad \forall\,x_2\in A_{x_1}.$$

Thus, $a(x_1)=m_2$ (otherwise letting $x_2\to\pm\infty$ would lead to a contradiction). This proves that $a=m_2$ almost everywhere, and therefore, the random vector $(X_1-\mathbf{E}X_1,X_2-\mathbf{E}X_2)$ is standard Gaussian. But this means that μ is a translation of γ. □

Acknowledgement. We would like to thank M. Ledoux for pointing to the paper by F-Y. Wang.

References

[B-B-G] Bakry, D., Bolley, F., Gentil, I. Dimension dependent hypercontractivity for Gaussian kernels. Probab. Theory Related Fields 154 (2012), no. 3-4, 845–874.

[B-E] Bakry, D., Émery, M. Diffusions hypercontractives. Seminaire de probabilites, XIX, 1983/84, 177–206, Lecture Notes in Math., 1123, Springer, Berlin, 1985.

[B-L] Bakry, D., Ledoux, M. A logarithmic Sobolev form of the Li-Yau parabolic inequality. Rev. Mat. Iberoam. 22 (2006), no. 2, 683-702.

[B-K] Barthe, F., Kolesnikov, A. V. Mass transport and variants of the logarithmic Sobolev inequality. J. Geom. Anal. 18 (2008), no. 4, 921-979.

[Bl] Blachman, N. M. The convolution inequality for entropy powers. IEEE Trans. Inform. Theory 11 (1965), 267–271.

[B] Bobkov, S. G. An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space. Ann. Probab. 25 (1997), no. 1, 206–214.

[B-G] Bobkov, S. G., Gotze, F. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. J. Funct. Anal. 163 (1999), no. 1, 1–28.

[B-H] Bobkov, S. G., Houdré, C. Isoperimetric constants for product probability measures. Ann. Probab. 25 (1997), no. 1, 184–205.

[C] Carlen, E. A. Superadditivity of Fisher's information and logarithmic Sobolev inequalities. J. Funct. Anal. 101 (1991), no. 1, 194–211.

[C-F-M-P] Cianchi, A., Fusco, N., Maggi, F., Pratelli, A. On the isoperimetric deficit in Gauss space. Amer. J. Math. 133(1):131-186, 2011.

[CE] Cordero-Erausquin, D. Some applications of mass transport to Gaussian-type inequalities. Arch. Ration. Mech. Anal. 161 (2002), no. 3, 257-269.

[D-C-T] Dembo, A., Cover, T. M., Thomas, J. A. Information-theoretic inequalities. IEEE Trans. Inform. Theory 37 (1991), no. 6, 1501–1518.

[E] Eldan, R. A two-sided estimate for the Gaussian noise stability deficit. Preprint (2013), arXiv:1307.2781 [math.PR].

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

[I-M] Indrei, E., Marcon, D. A quantitative log-Sobolev inequality for a two parameter family of functions. To appear in Int. Math. Res. Not. (2013).

[L1] Ledoux, M. Concentration of measure and logarithmic Sobolev inequalities. Seminaire de Probabilites XXXIII. Lecture Notes in Math. 1709 (1999), 120-216, Springer.

[L2] Ledoux, M. The concentration of measure phenomenon. Math. Surveys and monographs, vol. 89, AMS, 2001.

[Li] Lieb, E. H. Proof of an entropy conjecture of Wehrl. Comm. Math. Phys. 62 (1978), no. 1, 35–41.

[MC] McCann, R. J. Existence and uniqueness of monotone measure-preserving maps. Duke Math. J. 80 (1995), no. 2, 309-323.

[M-N] Mossel, E., Neeman, J. Robust dimension free isoperimetry in Gaussian space. Preprint (2012). To appear in Ann. Probab.

[O-V] Otto, F., Villani, C. Generalization of an inequality by Talagrand, and links with the logarithmic Sobolev inequality. J. Funct. Anal. 173 (2000), 361-400.

[R-S] Raginsky, M., Sason, I. Concentration of measure inequalities in Information Theory. Communications and Coding, ArXiv e-prints, December 2012.

[S] Stam, A. J. Some inequalities satisfied by the quantities of information of Fisher and Shannon. Information and Control 2 (1959), 101–112.

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

[W] Wang, F-Y. Generalized transportation-cost inequalities and applications. Potential Anal. 28 (2008), no. 4, 321–334.

School of Mathematics, University of Minnesota, USA
E-mail address: bobkov@math.umn.edu

Université Paris Est Marne la Vallée - Laboratoire d'Analyse et de Mathématiques Appliquées (UMR CNRS 8050), 5 bd Descartes, 77454 Marne la Vallée Cedex 2, France
E-mail address: nathael.gozlan@univ-mlv.fr, paul-marie.samson@univ-mlv.fr

Université Paris Ouest Nanterre la Défense, MODAL'X, EA 3454, 200 avenue de la République 92000 Nanterre, France
E-mail address: croberto@math.cnrs.fr