A Stein deficit for
the logarithmic Sobolev inequality

Michel Ledoux[*]   Ivan Nourdin[†]   Giovanni Peccati[‡]

September 3, 2016

Abstract

We provide some lower bounds on the deficit in the Gaussian logarithmic Sobolev inequality in terms of the so-called Stein characterization of the Gaussian distribution. The techniques are based on the representation of the relative Fisher information along the Ornstein-Uhlenbeck semigroup by the Minimum Mean-Square Error from information theory.

Keywords: Deficit, logarithmic Sobolev inequality, Ornstein-Uhlenbeck semigroup, Minimum Mean-Square Error, Stein kernel.

2000 Mathematics Subject Classification: 60E15, 26D10, 60B10

1   Introduction and main results

The classical logarithmic Sobolev inequality for the standard Gaussian measure

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

on the Borel sets of $\mathbb{R}^n$ expresses that for any smooth probability density f with respect to γ,

$$\mathrm{H}(f) = \int_{\mathbb{R}^n} f \log f \, d\gamma \leq \frac{1}{2} \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} \, d\gamma = \frac{1}{2} \mathrm{I}(f) \tag{1}$$

where $H(f)$ is the relative entropy of the measure $f d\gamma$ with respect to γ and $I(f)$ is its Fisher information.

It is a classical result, due to E. Carlen [10, 11], that the exponential densities

$$e_b(x) = e^{b \cdot x - |b|^2/2}, \quad x \in \mathbb{R}^n, \ b \in \mathbb{R}^n, \tag{2}$$

are saturating the inequality (1) and are the only ones. Note that the probability density e_b with respect to γ has mean b and covariance matrix the identity Id.

Modulo smoothness assumptions on the underlying density f, a proof of this result may be given by interpolation along the Ornstein-Uhlenbeck semigroup (cf. [27]). Let $(P_t)_{t \geq 0}$ be the Ornstein-Uhlenbeck semigroup with integral representation

$$P_t g(x) = \int_{\mathbb{R}^n} g\big(e^{-t}x + \sqrt{1 - e^{-2t}}\, y\big) d\gamma(y), \quad t \geq 0, \ x \in \mathbb{R}^n, \tag{3}$$

for any suitable $g : \mathbb{R}^n \to \mathbb{R}$. By expansion along this semigroup, the Bakry-Émery calculus (see [2, 3, 4, 27] and below) yields that

$$H(f) = \frac{1}{2} I(f) - \int_0^\infty \int_{\mathbb{R}^n} P_t f \left|\mathrm{Hess}(\log P_t f)\right|^2 d\gamma\, dt. \tag{4}$$

Here and throughout this work, $|\cdot|$ denotes the Euclidean norm on vectors and matrices (Hilbert-Schmidt norm). Hence, if there is equality in (1), for almost every $t \geq 0$ and $x \in \mathbb{R}^n$, $\mathrm{Hess}(\log P_t f)(x) = 0$ so that $\log f(x)$ is affine.

Following recent investigations for classical Sobolev and isoperimetric inequalities, both for the Lebesgue and Gaussian measures, the question has been raised to quantify the deficit in the logarithmic Sobolev inequality via a suitable distance to the saturating exponential densities. In the broader context of stability results for functional inequalities, when looking at a functional inequality with known optimal constants and optimizers, a natural question is indeed whether functions that are close to achieving the optimum are close to some optimizer. The task is to bound from below the deficit by some functional that measures how far we are from some optimizer (typically, a distance). Such quantitative estimates both strengthen the underlying functional inequality and provide physical measurements of the quality of the inequality with respect to the saturating functions. Functional stability inequalities are then in turn critical in quantitative convergence to equilibrium. Examples are the recent quantitative stability estimates for Sobolev [13, 21], Brunn-Minkowski [19, 18] and isoperimetric inequalities [23, 20, 30, 15, 6]. See also the survey [17].

Towards this goal, introduce therefore, for a (smooth) probability density f with respect to γ, the deficit

$$\delta(f) = \frac{1}{2} I(f) - H(f) \geq 0 \tag{5}$$

in the logarithmic Sobolev inequality (1) for the density f. We speak equivalently of the deficit of the probability $d\mu = f d\gamma$. Relevant lower bounds on the deficit may then be interpreted as a stability estimate on the functional inequality with respect to the extremizers.

2

Stability in the logarithmic Sobolev has therefore motivated recently a number of investigations. However, the various conclusions so far do not appear fully satisfactory, in particular with respect to dimension free bounds which should reasonably be expected (as the logarithmic Sobolev inequality itself does not depend on the dimension of the underlying state space). Note that the corresponding study of the deficit in the Gaussian isoperimetric inequality has produced (optimal) dimension free bounds [30, 6, 15] with respect to a natural distance to the extremal sets (half-spaces). While the logarithmic Sobolev inequality may be derived from the Gaussian isoperimetric inequality (cf. [4]), the derivation does not seem to preserve any information on the deficit.

To briefly survey some of the recent conclusions on the deficit in the logarithmic Sobolev inequality, note first the lower bound, under the condition $\int_{\mathbb{R}^n} |x|^2 d\mu \leq n$,

$$\delta(f) \geq \frac{1}{100\,n}\,W_2(\mu, \gamma)^4 \tag{6}$$

emphasized in [7] after an inequality of [5]. Here $W_2(\mu, \gamma)$ is the Monge-Kantorovich distance between μ and γ given by

$$W_2(\mu, \gamma) = \inf\left(\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |x - y|^2 d\pi(x, y)\right)^{1/2},$$

the infimum being taken over all couplings π on $\mathbb{R}^n \times \mathbb{R}^n$ with respective marginals μ and γ. The proof of (6) put forward in [7] relies on the dimensional self-improved form of the logarithmic Sobolev inequality [5] (see also [4]) expressing that for any smooth density f with respect to γ,

$$H(f) \leq \frac{1}{2}\int_{\mathbb{R}^n} \Delta f\, d\gamma + \frac{n}{2}\log\left(1 + \frac{I(f)}{n} - \frac{1}{n}\int_{\mathbb{R}^n} \Delta f\, d\gamma\right). \tag{7}$$

(It is part of the result that the expression inside the logarithm is positive.) Hence, after a simple rewriting,

$$\delta(f) \geq \frac{n}{2}\,\theta\left(\frac{I(f) - \int_{\mathbb{R}^n} \Delta f d\gamma}{n}\right)$$

where $\theta(r) = r - \log(1 + r)$, $r > -1$. Now, by a double integration by parts with respect to the Gaussian density,

$$\int_{\mathbb{R}^n} |x|^2 d\mu = \int_{\mathbb{R}^n} |x|^2 f\, d\gamma = n + \int_{\mathbb{R}^n} \Delta f\, d\gamma.$$

Hence, whenever $\int_{\mathbb{R}^n} |x|^2 d\mu \leq n$, then $\int_{\mathbb{R}^n} \Delta f d\gamma \leq 0$ and since θ is increasing,

$$\delta(f) \geq \frac{n}{2}\,\theta\left(\frac{I(f)}{n}\right).$$

Next we may use again the logarithmic Sobolev inequality $I(f) \geq 2\,H(f)$ together with the Talagrand [35] quadratic transportation cost inequality (cf. e.g. [4, 33, 36])

$$2\,H(f) \geq W_2(\mu, \gamma)^2. \tag{8}$$

Under the condition $\int_{\mathbb{R}^n} |x|^2 d\mu \leq n$, $W_2(\mu, \gamma)^2 \leq 4n$, and since $\theta(r) \geq \frac{r^2}{50}$ (for example) on the interval $[0, 4]$, the lower bound (6) follows. In Section 6, we will provide an independent proof of (6) based on the information theoretic tools developed in this work.

In another direction, the dimension free lower bound

$$\delta(f) \geq c(\lambda) \, W_2(\mu, \gamma)^2 \tag{9}$$

has been established in the recent [16] but under the further assumption that μ centered satisfies a Poincaré inequality with constant $\lambda > 0$.

One drawback of (6) is of course, besides the dimensional condition $\int_{\mathbb{R}^n} |x|^2 d\mu \leq n$, that the lower bound depends on n and vanishes as $n \to \infty$. It is mentioned in [16] that one cannot expect a dimension free lower bound only in terms of the Monge-Kantorovich metric W_2. In addition, for the extremal e_b of (2), $\int_{\mathbb{R}^n} |x|^2 e_b d\gamma = n + |b|^2$, so that the condition $\int_{\mathbb{R}^n} |x|^2 d\mu \leq n$ rules out all extremizers but the centered one (that is γ itself). It is therefore of interest to look for a measure to the extremizers which may produce stability estimates as much as possible independent of the dimension and moreover suitably identifying the extremal densities.

The papers [7, 16, 14] contain further stability results involving related transport distances between modifications of μ and γ, however still dimensional. The note [22] presents a lower bound on the deficit based on a distance (modulo translation) in dimension n starting with a distance in dimension one first introduced in [8].

In another direction, investigations have also concerned upper bounds on the deficit. The perspective is somewhat different here and gave recently rise to lower bounds on the (Euclidean) entropy gap for log-concave densities relevant to the study of affine-isoperimetric inequalities (cf. [1, 9]).

The aim of this work is to propose a lower bound on the deficit $\delta(f)$ in the logarithmic Sobolev inequality in terms of the Stein characterization of the standard normal distribution γ. Before addressing the result, let us first emphasize that, in order to make sense of $\delta(f)$, it is legitimate to assume that $H(f) < \infty$. Since the deficit should be small, it will also hold that $I(f) < \infty$. In particular, this condition entails the fact that the density f has some smoothness, and regularity will be implicitly assumed for the various expressions to be well-defined. In addition, the finiteness of $H(f)$ ensures by the entropic inequality (see e.g. [4, Section 5.1.1]) that

$$\int_{\mathbb{R}^n} |x|^2 d\mu = \int_{\mathbb{R}^n} |x|^2 f \, d\gamma \leq 4 \, H(f) + 4 \log \int_{\mathbb{R}^n} e^{|x|^2/4} d\gamma < \infty.$$

Throughout this study of the deficit $\delta(f)$ of the density f, it will therefore be assumed that $\int_{\mathbb{R}^n} |x|^2 d\mu < \infty$ (although this condition is not everywhere strictly necessary). In particular, we may consider the covariance matrix $\Gamma = (\Gamma_{ij})_{1 \leq i,j \leq n}$ of μ given for all $i, j = 1, \ldots, n$ by

$$\Gamma_{ij} = \int_{\mathbb{R}^n} x_i x_j \, d\mu - \int_{\mathbb{R}^n} x_i \, d\mu \int_{\mathbb{R}^n} x_j \, d\mu.$$

The investigation will therefore involve the Stein characterization of the normal distribution, and more generally ideas related to Stein's method (cf. [12, 31, 34]). Recall indeed the basic

4

integration by parts formula

$$\int_{\mathbb{R}^n} x\varphi \, d\gamma \;=\; \int_{\mathbb{R}^n} \nabla\varphi \, d\gamma$$

for any smooth $\varphi : \mathbb{R}^n \to \mathbb{R}$. This equation is characteristic of the Gaussian distribution γ in the sense that if μ is a probability measure on $\mathbb{R}^n$ (with finite mean) such that for any smooth $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$\int_{\mathbb{R}^n} x\varphi \, d\mu \;=\; \int_{\mathbb{R}^n} \nabla\varphi \, d\mu \tag{10}$$

(as vectors in $\mathbb{R}^n$), then it is necessarily equal to γ. Indeed, apply for example (10) to $\varphi(x) = e^{i\lambda \cdot x}$, $\lambda \in \mathbb{R}^n$, to get that the Fourier transform $F(\lambda) = \int_{\mathbb{R}^n} e^{i\lambda \cdot x} d\mu(x)$, $\lambda \in \mathbb{R}^n$, of μ satisfies $\nabla F = -\lambda F$, hence $F(\lambda) = e^{-|\lambda|^2/2}$ and $\mu = \gamma$.

According to this description, let, for a given probability measure μ on $\mathbb{R}^n$,

$$\mathcal{D}(\mu, \gamma) \;=\; \sup_{\varphi \in \mathcal{R}} \left| \int_{\mathbb{R}^n} \big[x\varphi - \nabla\varphi \big] d\mu \right| \tag{11}$$

where $\mathcal{R}$ is the class of the resolvents (for the Ornstein-Uhlenbeck semigroup $(P_t)_{t \geq 0}$)

$$\varphi \;=\; R\psi \;=\; 4 \int_0^\infty e^{-4t} P_t \psi \, dt$$

with $\psi : \mathbb{R}^n \to \mathbb{R}$ (smooth, for example C^1) such that $\int_{\mathbb{R}^n} \psi^2 P_t f d\gamma \leq 1$ for every $t \geq 0$. The value 4 has no particular meaning. It is shown in Section 2 that $\mathcal{R}$ is a determining class for γ, that is, if μ and γ agree on functions of the class $\mathcal{R}$, then $\mu = \gamma$.

The main result of this work is a stability estimate in the logarithmic Sobolev inequality by means of the Stein functional (11). If f is a probability density with respect to γ with mean b, define the shifted probability density

$$f_b(x) \;=\; f(x + b) \, e^{-(b \cdot x + |b|^2/2)}, \quad x \in \mathbb{R}^n, \tag{12}$$

which has mean zero with respect to γ. (In other words, if X is a random vector with distribution $f d\gamma$ and mean b, $X - b$ has distribution $f_b d\gamma$ and mean zero.) Then, whenever $d\mu_b = f_b d\gamma$ is close to γ, that is f_b is close to the constant 1 function, $f(x + b)$ is close to $e^{b \cdot x + |b|^2/2}$, hence after translation f is close to the extremal e_b of (2).

Theorem 1. *Let f be a probability density on $\mathbb{R}^n$ and let $d\mu = f d\gamma$. Assume that μ has barycenter b and covariance matrix $\Gamma \leq \mathrm{Id}$ (in the sense of symmetric matrices). Then,*

$$\delta(f) \;\geq\; \frac{1}{4} \, \mathcal{D}(\mu_b, \gamma)^4.$$

The functional $\mathcal{D}(\mu, \gamma)$ might not be fully transparent, and for the comparison with more classical distances, let us consider besides

$$\mathcal{D}'(\mu, \gamma) \;=\; \sup_{\varphi \in \mathcal{B}} \left| \int_{\mathbb{R}^n} \big[x\varphi - \nabla\varphi \big] d\mu \right| \tag{13}$$

where the supremum runs over the class $\mathcal{B}$ of smooth functions φ on $\mathbb{R}^n$ with

$$\|\varphi\|_\infty \leq 1, \quad \|\nabla\varphi\|_\infty \leq 1, \quad \big\|\mathrm{Hess}(\varphi)\big\|_\infty \leq 1$$

where $\|\varphi\|_\infty$, $\|\nabla\varphi\|_\infty$, $\|\mathrm{Hess}(\varphi)\|_\infty$ denote respectively the uniform norm of f and of the Euclidean norm of ∇f and $\mathrm{Hess}(f)$. The functionals $\mathcal{D}(\mu, \gamma)$ and $\mathcal{D}'(\mu, \gamma)$ are compared in Section 2 where it is shown that

$$\mathcal{D}'(\mu, \gamma) \leq 2\big(1 + \mathrm{I}(f)\big)^{1/2} \mathcal{D}(\mu, \gamma).$$

The main Theorem 1 therefore yields the following alternate lower bound on the deficit. Note that $\mathrm{I}(f_b) = \mathrm{I}(f) - |b|^2 \leq \mathrm{I}(f)$.

Theorem 2. *Let f be a probability density on $\mathbb{R}^n$ and let $d\mu = f d\gamma$. Assume that μ has barycenter b and covariance matrix $\Gamma \leq \mathrm{Id}$ (in the sense of symmetric matrices). Then,*

$$\delta(f) \geq \frac{1}{64(1 + \mathrm{I}(f_b))^2} \mathcal{D}'(\mu_b, \gamma)^4.$$

In the preceding statements, the covariance hypothesis $\Gamma \leq \mathrm{Id}$ is of course not very natural, although the aforementioned investigations implicitly encountered the same difficulty, and for example (6) assumes that $\int_{\mathbb{R}^n} |x|^2 d\mu \leq n$. However, with respect to (6), the lower bound in Theorem 1 does not involve specifically the dimension n in front of the metric. Moreover, the result identifies the extremal with mean b while, as we have seen, (6) is restricted to mean zero densities.

To somewhat appreciate the input of Theorem 1 with respect to previous bounds, let us consider a simple product example. Consider namely the distribution $d\mu = f d\gamma$ of a centered Gaussian vector with diagonal covariance matrix $(\sigma_i^2)_{1 \leq i \leq n}$ such that $\sigma_i \leq 1$ (hence the conditions $\Gamma \leq \mathrm{Id}$ and $\int_{\mathbb{R}^n} |x|^2 d\mu \leq n$ are satisfied). By the product structure and a direct computation, it is easily seen that

$$\delta(f) = \sum_{i=1}^n \left[\frac{1 - \sigma_i^2}{2\sigma_i^2} + \log\sigma_i\right].$$

For the matter of further comparison,

$$\frac{1}{4}\sum_{i=1}^n \frac{(1 - \sigma_i^2)^2}{\sigma_i^4} \geq \delta(f) = \frac{1}{2}\sum_{i=1}^n \left[\frac{1 - \sigma_i^2}{\sigma_i^2} + \log\sigma_i^2\right] \geq \frac{1}{4}\sum_{i=1}^n (1 - \sigma_i^2)^2. \tag{14}$$

By the obvious coupling,

$$W_2(\mu, \gamma) = \left(\sum_{i=1}^n (1 - \sigma_i)^2\right)^{1/2}.$$

On the other hand, by integration by parts, for any smooth $\varphi : \mathbb{R}^n \to \mathbb{R}$ and any $i = 1, \ldots, n$,

$$\int_{\mathbb{R}^n} [x_i\varphi - \partial_i\varphi]d\mu = (\sigma_i^2 - 1)\int_{\mathbb{R}^n} \partial_i\varphi \, d\mu.$$

6

Take $\psi(x) = \alpha \cdot x$, $x \in \mathbb{R}^n$, where $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ with $\sum_{i=1}^n \alpha_i^2 \leq 1$. For any $t \geq 0$,

$$
\begin{aligned}
\int_{\mathbb{R}^n} \psi^2 P_t f \, d\gamma &= \int_{\mathbb{R}^n} P_t(\psi^2) d\mu \\
&= \sum_{i,j=1}^n \alpha_i \alpha_j \int_{\mathbb{R}^n} P_t(x_i x_j) d\mu \\
&= \sum_{i,j=1}^n \alpha_i \alpha_j \int_{\mathbb{R}^n} \left[\mathrm{e}^{-2t} x_i x_j + (1 - \mathrm{e}^{-2t}) \delta_{ij} \right] d\mu \\
&\leq |\alpha|^2 \leq 1.
\end{aligned}
$$

Hence $\psi \in \mathcal{R}$. Now, if $\varphi = R\psi$, for any $i = 1, \ldots, n$,

$$
\partial_i \varphi = 4 \int_0^\infty \mathrm{e}^{-4t} \partial_i P_t \psi \, dt = 4\alpha_i \int_0^\infty \mathrm{e}^{-5t} dt = \frac{4}{5} \alpha_i.
$$

As a consequence,

$$
\mathcal{D}(\mu, \gamma) \geq \frac{4}{5} \left(\sum_{i=1}^n (1 - \sigma_i^2)^2 \alpha_i^2 \right)^{1/2},
$$

and taking the supremum over all α with $\sum_{i=1}^n \alpha_i^2 \leq 1$,

$$
\mathcal{D}(\mu, \gamma) \geq \frac{4}{5} \max_{1 \leq i \leq n} (1 - \sigma_i^2).
$$

Then, (6) produces the lower-bound

$$
\delta(f) \geq \frac{1}{100 \, n} \left(\sum_{i=1}^n (1 - \sigma_i^2) \right)^2
$$

while Theorem 1 yields

$$
\delta(f) \geq \frac{1}{10} \max_{1 \leq i \leq n} (1 - \sigma_i^2)^4.
$$

According to (14), it therefore appears that if $\sigma_i = \sigma < 1$, $i = 1, \ldots, n$, the lower-bound (6) is of the accurate order. Nevertheless, as soon as a large proportion of σ_i's are much closer (equal) to 1, then (6) deteriorates as $n \to \infty$ while Theorem 1 keeps the correct order as the true $\delta(f)$. Note that in this example, the lower-bound (9) from [16] is optimal.

Coming back to the condition $\Gamma \leq \mathrm{Id}$, it would already be of interest to understand how the deficit δ could control the proximity of the covariance matrix to the identity. In Section 3, we provide a variation on Theorem 1 that somehow takes into account this deficiency. In particular, it is shown there that

$$
2\delta(f) + \|\Gamma - \mathrm{Id}\|^2 \geq \frac{1}{4} \mathcal{D}(\mu_b, \gamma)^4 \tag{15}
$$

where, for an $n \times n$ matrix A,

$$
\|A\| = \sup_{|\alpha|=1} A\alpha \cdot \alpha.
$$

(despite the notation, observe that $\|\cdot\|$ is not a norm). A more precise version (Theorem 8) allows for a deficit for arbitrary sizes of $\|\Gamma - \mathrm{Id}\|$.

In another direction, the next result provides a kind of compactness argument to bound from below the deficit by an unknown constant depending on f.

Theorem 3. *Let f be a probability density on $\mathbb{R}^n$ and let $d\mu = f d\gamma$. Assume that μ has barycenter b and that $\mathrm{I}(f) < \infty$. Then*

$$\delta(f) \geq c(f)\,\mathcal{D}(\mu_b,\gamma)^4$$

where $c(f) > 0$ is a constant depending on f only via the uniform integrability of the family of measures $(\alpha \cdot x)^2 d\mu$, where α runs over the unit sphere of $\mathbb{R}^n$.

The distance, or rather measure of proximity in the sense of Stein, $\mathcal{D}'(\mu,\gamma)$ may be recast in terms of the Stein kernel associated with a given distribution, and compared to its discrepancy as emphasized in [29]. For a centered probability measure μ, let τ_μ be a Stein kernel (matrix) of μ in the sense that for any smooth $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$\int_{\mathbb{R}^n} x\varphi\, d\mu = \int_{\mathbb{R}^n} \tau_\mu \nabla\varphi\, d\mu$$

(as vectors in $\mathbb{R}^n$). Then

$$\mathcal{D}'(\mu,\gamma) = \sup_{\varphi \in \mathcal{B}} \left| \int_{\mathbb{R}^n} \left[(\tau_\mu - \mathrm{Id})\nabla\varphi \right] d\mu \right|.$$

Recalling the Stein discrepancy between μ and γ,

$$\mathrm{S}(\mu \,|\, \gamma) = \left(\int_{\mathbb{R}^n} |\tau_\mu - \mathrm{Id}|^2 d\mu \right)^{1/2}$$

(where we recall that $|\cdot|$ stands for the Hilbert-Schmidt norm when applied to matrices), it holds that

$$\mathcal{D}'(\mu,\gamma) \leq \mathrm{S}(\mu \,|\, \gamma). \tag{16}$$

For the matter of comparison, note from [29] that $\mathrm{W}_2(\mu,\gamma) \leq \mathrm{S}(\mu,\gamma)$.

As an additional link between the deficit and the Stein characterization, the recent [29] points out an improved form of the logarithmic Sobolev inequality involving the Stein discrepancy $\mathrm{S} = \mathrm{S}(\mu \,|\, \gamma)$ as

$$\mathrm{H}(f) \leq \frac{\mathrm{S}^2}{2} \log\left(1 + \frac{\mathrm{I}(f)}{\mathrm{S}^2} \right).$$

In terms of the deficit $\delta = \delta(f) = \frac{1}{2}\mathrm{I}(f) - \mathrm{H}(f)$,

$$\mathrm{H}(f) \leq \frac{\mathrm{S}^2}{2} \log\left(1 + \frac{2\,\mathrm{H}(f) + 2\delta}{\mathrm{S}^2} \right)$$

so that if $r = \frac{2\,\mathrm{H}(f)}{\mathrm{S}^2}$, then $r \leq \log\left(1 + r + \frac{2\delta}{\mathrm{S}^2}\right)$, that is

$$\frac{2\delta}{\mathrm{S}^2} \geq e^r - 1 - r \geq \frac{r^2}{2} = \frac{2\,\mathrm{H}(f)^2}{\mathrm{S}^4}.$$

Therefore

$$\delta(f) \geq \frac{\mathrm{H}(f)^2}{\mathrm{S}^2}.$$

Together with the transportation cost inequality (8), we may therefore state the following corollary, close in spirit to (6).

Proposition 4. *Let $d\mu = f d\gamma$ centered on $\mathbb{R}^n$ with Stein kernel τ_μ and associated discrepancy $\mathrm{S}(\mu\,|\,\gamma)$. Then*

$$\delta(f) \geq \frac{\mathrm{W}_2(\mu,\gamma)^4}{4\,\mathrm{S}(\mu\,|\,\gamma)^2}. \tag{17}$$

Remark 5. In [29, Theorem 3.2], it is proved that

$$\mathrm{W}_2(\mu,\gamma) \leq \mathrm{S}(\mu\,|\,\gamma)\arccos\left(e^{-\frac{\mathrm{H}(f)}{\mathrm{S}^2(\mu\,|\,\gamma)}}\right).$$

Such a relation allows one to infer that

$$\mathrm{H}(f) \geq \mathrm{S}^2(\mu\,|\,\gamma)\log\left(\frac{1}{\cos(\mathrm{W}_2(\mu,\gamma)\,\mathrm{S}(\mu\,|\,\gamma)^{-1})}\right),$$

so that the estimate (17) can be slightly improved as

$$\delta(f) \geq \mathrm{S}(\mu\,|\,\gamma)^2\log\left(\frac{1}{\cos(\mathrm{W}_2(\mu,\gamma)\,\mathrm{S}(\mu\,|\,\gamma)^{-1})}\right)^2. \tag{18}$$

Notice that, in view of $\mathrm{W}_2(\mu,\gamma) \leq \mathrm{S}(\mu\,|\,\gamma)$, one has that

$$\cos(1) \leq \cos\left(\mathrm{W}_2(\mu,\gamma)\,\mathrm{S}(\mu\,|\,\gamma)^{-1}\right) \leq 1.$$

The paper is organized as follows. In the next Section 2, we describe properties and comparisons of the functionals $\mathcal{D}(\mu,\gamma)$ and $\mathcal{D}'(\mu,\gamma)$. Section 3 provides the crucial information theoretic tools to analyze the deficit in terms of $\mathcal{D}(\mu,\gamma)$, and on which the proof of Theorem 1 relies. Theorems 1 and 3 are then established in Sections 4 and 5 respectively. The final section is devoted to an alternate proof of the lower bound (6) based on the tools of Section 3.

2   Properties of $\mathcal{D}(\mu,\gamma)$ and $\mathcal{D}'(\mu,\gamma)$

The purpose of this section to compare $\mathcal{D}(\mu,\gamma)$ and $\mathcal{D}'(\mu,\gamma)$. Before, we collect some general informations on $\mathcal{D}(\mu,\gamma)$ to get a better feeling about it.

Recall that we assume throughout the investigation that the density f is smooth and that $\int_{\mathbb{R}^n} |x|^2 d\mu = \int_{\mathbb{R}^n} |x|^2 f d\gamma < \infty$. Recall also that the Ornstein-Uhlenbeck semigroup $(P_t)_{t\geq 0}$ described by the integral representation (3) is invariant and symmetric with respect to γ. Its infinitesimal generator $\mathrm{L} = \Delta - x \cdot \nabla$ satisfies the integration by parts formula, for smooth functions $g, h : \mathbb{R}^n \to \mathbb{R}$,

$$\int_{\mathbb{R}^n} g \, \mathrm{L}h \, d\gamma = -\int_{\mathbb{R}^n} \nabla g \cdot \nabla h \, d\gamma.$$

It should be noted first that the integrals $\int_{\mathbb{R}^n} [x\varphi - \nabla\varphi] d\mu$ in the definition of $\mathcal{D}(\mu, \gamma)$ are well-defined. If $\varphi = R\psi$ with $\int_{\mathbb{R}^n} \psi^2 P_t f d\gamma = \int_{\mathbb{R}^n} P_t(\psi^2) d\mu \leq 1$ for every $t \geq 0$,

$$\int_{\mathbb{R}^n} |x\varphi| d\mu \leq 4 \int_0^\infty e^{-4t} \int_{\mathbb{R}^n} |x P_t \psi| d\mu \, dt$$

$$\leq 4 \left(\int_{\mathbb{R}^n} |x|^2 d\mu \right)^{1/2} \int_0^\infty e^{-4t} \left(\int_{\mathbb{R}^n} (P_t\psi)^2 d\mu \right)^{1/2} dt$$

$$\leq \left(\int_{\mathbb{R}^n} |x|^2 d\mu \right)^{1/2} < \infty.$$

Next, after integration by parts in the integral representation (3) of P_t, for every $x \in \mathbb{R}^n$,

$$|\nabla P_t \psi|^2(x) = \sum_{i=1}^n (\partial_i P_t \psi)^2(x)$$

$$= e^{-2t} \sum_{i=1}^n \left(\int_{\mathbb{R}^n} \partial_i \psi \big(e^{-t}x + \sqrt{1 - e^{-2t}}\, y \big) d\gamma(y) \right)^2$$

$$= \frac{e^{-2t}}{1 - e^{-2t}} \sum_{i=1}^n \left(\int_{\mathbb{R}^n} y_i \, \psi \big(e^{-t}x + \sqrt{1 - e^{-2t}}\, y \big) d\gamma(y) \right)^2$$

$$\leq \frac{e^{-2t}}{1 - e^{-2t}} P_t(\psi^2)(x)$$

so that

$$\int_{\mathbb{R}^n} |\nabla\varphi| d\mu \leq 4 \int_0^\infty e^{-4t} \int_{\mathbb{R}^n} |\nabla P_t \psi| d\mu \, dt \leq 4 \int_0^\infty \frac{e^{-5t}}{\sqrt{1 - e^{-2t}}} dt \leq 4.$$

The family $\mathcal{R}$ is a determining class in the sense that whenever $\mathcal{D}(\mu, \gamma) = 0$, then $\mu = \gamma$. To check this claim, choose $\psi(x) = e^{i\lambda \cdot x}$, $\lambda \in \mathbb{R}^n$ (rather their real and imaginary parts), so that

$$\varphi(x) = R\psi(x) = 4 e^{-|\lambda|^2/2} \int_0^1 u^3 \, e^{iu\lambda \cdot x + |\lambda|^2 u^2/2} du, \quad x \in \mathbb{R}^n.$$

With F the Fourier transform of μ,

$$\int_{\mathbb{R}^n} [x\varphi - \nabla\varphi] d\mu = -4i \, e^{-|\lambda|^2/2} \int_0^1 u^3 \, e^{|\lambda|^2 u^2/2} \big[\nabla F(u\lambda) + u\lambda F(u\lambda) \big] du.$$

If the left-hand side of this identity is zero, after the change of u into $\frac{u}{|\lambda|}$,

$$\int_0^\rho u^3 \, e^{u^2/2} \big[\nabla F(u\theta) + u\lambda F(u\theta) \big] du = 0$$

where $\rho = |\lambda|$ and $\theta = \frac{\lambda}{|\lambda|}$, $\lambda \neq 0$. This relation holding true for any $\rho > 0$ and $\theta \in \mathbb{S}^{n-1}$, it follows that $\nabla F(w) + w F(w) = 0$ for any $w \in \mathbb{R}^n$, and thus F is the Fourier transform of the standard normal γ on $\mathbb{R}^n$.

The next proposition is the announced comparison between $\mathcal{D}(\mu, \gamma)$ and $\mathcal{D}'(\mu, \gamma)$.

Proposition 6. *Let $d\mu = f d\gamma$. Then*

$$\mathcal{D}'(\mu, \gamma) \leq 2 \big(1 + \mathrm{I}(f)\big)^{1/2} \mathcal{D}(\mu, \gamma).$$

Proof. Given $\varphi \in \mathcal{B}$, it is straightforward to check that $\varphi = R\psi$ where

$$\psi = -\frac{1}{4} (\mathrm{L} - 4\,\mathrm{Id})\varphi.$$

Hence the condition $\int_{\mathbb{R}^n} \psi^2 P_t f d\gamma \leq 1$ for every $t \geq 0$ in the definition of $\mathcal{D}(\mu, \gamma)$ turns into

$$\int_{\mathbb{R}^n} (\mathrm{L}\varphi - 4\varphi)^2 P_t f \, d\gamma \leq 16.$$

Developing the square, we examine successively the three terms

$$\int_{\mathbb{R}^n} \varphi^2 P_t f d\gamma, \quad \int_{\mathbb{R}^n} \mathrm{L}\varphi \, \varphi P_t f d\gamma, \quad \int_{\mathbb{R}^n} (\mathrm{L}\varphi)^2 P_t f \, d\gamma$$

under the boundedness assumptions on φ and its derivatives.

First $\int_{\mathbb{R}^n} \varphi^2 P_t f d\gamma \leq \|\varphi\|_\infty \leq 1$. To handle

$$\int_{\mathbb{R}^n} \mathrm{L}\varphi \, \varphi P_t f d\gamma,$$

write by integration by parts that

$$\int_{\mathbb{R}^n} \mathrm{L}\varphi \, \varphi P_t f d\gamma = -\int_{\mathbb{R}^n} \nabla\varphi \cdot \nabla(\varphi P_t f) d\gamma$$

$$= -\int_{\mathbb{R}^n} |\nabla\varphi|^2 P_t f \, d\gamma - \int_{\mathbb{R}^n} \varphi \nabla\varphi \cdot \nabla P_t f \, d\gamma.$$

Now, if $\|\varphi\|_\infty \leq 1$ and $\|\nabla\varphi\|_\infty \leq 1$, by the Cauchy-Schwarz inequality and the exponential decay recalled in (23) below,

$$\left| \int_{\mathbb{R}^n} \varphi \nabla\varphi \cdot \nabla P_t f \, d\gamma \right| \leq \int_{\mathbb{R}^n} |\nabla P_t f| d\gamma \leq \sqrt{\mathrm{I}(P_t f)} \leq \mathrm{e}^{-t} \sqrt{\mathrm{I}(f)}.$$

As a consequence, for every $t \geq 0$,

$$\left| \int_{\mathbb{R}^n} \mathrm{L}\varphi \, \varphi P_t f \, d\gamma \right| \leq 1 + \sqrt{\mathrm{I}(f)}.$$

Finally, again by integration by parts,

$$\int_{\mathbb{R}^n} (\mathrm{L}\varphi)^2 P_t f \, d\gamma = -\int_{\mathbb{R}^n} \nabla\varphi \cdot \nabla(\mathrm{L}\varphi \, P_t f) d\gamma$$
$$= -\int_{\mathbb{R}^n} \mathrm{L}\varphi \, \nabla\varphi \cdot \nabla P_t f \, d\gamma - \int_{\mathbb{R}^n} \nabla\varphi \cdot \nabla\mathrm{L}\varphi \, P_t f \, d\gamma$$

Under $\|\nabla\varphi\|_\infty \leq 1$,

$$-\int_{\mathbb{R}^n} \mathrm{L}\varphi \, \nabla\varphi \cdot \nabla P_t f \, d\gamma \leq \int_{\mathbb{R}^n} |\mathrm{L}\varphi||\nabla P_t f| d\gamma \leq \frac{1}{2}\int_{\mathbb{R}^n} (\mathrm{L}\varphi)^2 P_t f \, d\gamma + \frac{1}{2}\,\mathrm{I}(P_t f).$$

Therefore

$$\int_{\mathbb{R}^n} (\mathrm{L}\varphi)^2 P_t f d\gamma \leq -2\int_{\mathbb{R}^n} \nabla\varphi \cdot \nabla\mathrm{L}\varphi \, P_t f \, d\gamma + \mathrm{I}(P_t f).$$

Now

$$-2\,\nabla\varphi \cdot \nabla\mathrm{L}\varphi = 2\,\Gamma_2(\varphi) - \mathrm{L}\big(|\nabla\varphi|^2\big)$$

where $\Gamma_2(\varphi) = |\mathrm{Hess}(\varphi)|^2 + |\nabla\varphi|^2 \leq 2$. Once more by integration by parts,

$$\int_{\mathbb{R}^n} \mathrm{L}\big(|\nabla\varphi|^2\big) P_t f d\gamma = -\int_{\mathbb{R}^n} \nabla\big(|\nabla\varphi|^2\big) \cdot \nabla P_t f d\gamma$$

and

$$\big|\nabla\big(|\nabla\varphi|^2\big)\big| \leq 2|\nabla\varphi|\,\big|\mathrm{Hess}(\varphi)\big| \leq 2.$$

Altogether, if follows that

$$\int_{\mathbb{R}^n} (\mathrm{L}\varphi)^2 P_t f d\gamma \leq 4 + 2\sqrt{\mathrm{I}(f)} + \mathrm{I}(f).$$

As a consequence of the preceding three upper bounds, and using that $\sqrt{I(f)} \leq \frac{1}{2}(1+I(f))$, for any $t \geq 0$,

$$\int_{\mathbb{R}^n} (\mathrm{L}\varphi - 4\varphi)^2 P_t f d\gamma \leq 33 + 6\,\mathrm{I}(f) \leq 64\big(1 + I(f)\big).$$

As a result, $\varphi \in \mathcal{B}$ implies that $\varphi \in 2\sqrt{1 + I(f)}\,\mathcal{R}$, and the proof of the proposition is completed by homogeneity. $\qquad\square$

3   Information theoretic representation

With this section we start addressing the proof of the main result, and in particular develop the information theoretic tools towards a suitable expression for the deficit in the semigroup formulation (4).

Recall the Ornstein-Uhlenbeck semigroup $(P_t)_{t\geq 0}$ from (3). Note that, as vector valued functions, provided $f : \mathbb{R}^n \to \mathbb{R}$ is smooth, $\nabla P_t f = \mathrm{e}^{-t} P_t(\nabla f)$. It is immediate on the integral representation (3) to observe that, by integration by parts, for every $t \geq 0$, as vectors in $\mathbb{R}^n$,

$$P_t(xf) = \mathrm{e}^{-t} x P_t f + (1 - \mathrm{e}^{-2t}) P_t(\nabla f) = \mathrm{e}^{-t} x P_t f + 2\,\mathrm{sh}(t)\nabla P_t f. \tag{19}$$

From (19) is deduced an alternate description of the Fisher information

$$I(P_t f) = \int_{\mathbb{R}^n} \frac{|\nabla P_t f|^2}{P_t f} \, d\gamma$$

along the semigroup as

$$4 \operatorname{sh}^2(t) \, I(P_t f) = \int_{\mathbb{R}^n} \frac{|P_t(xf) - e^{-t} x P_t f|^2}{P_t f} \, d\gamma.$$

Given a probability density f with respect to γ, let X be a random vector with distribution $d\mu = f \, d\gamma$. Let furthermore N be independent with law γ, and set, for every $t \geq 0$,

$$X_t = e^{-t} X + \sqrt{1 - e^{-2t}} \, N.$$

Note that X_t has distribution $P_t f \, d\gamma$ since for any bounded measurable $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}(\varphi(X_t)) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \varphi(e^{-t} x + \sqrt{1 - e^{-2t}} \, y) f(x) d\gamma(x) d\gamma(y)$$

$$= \int_{\mathbb{R}^n} P_t \varphi \, f \, d\gamma = \int_{\mathbb{R}^n} \varphi \, P_t f \, d\gamma.$$

The next observation is that if $u = \frac{P_t(xf)}{P_t f} : \mathbb{R}^n \to \mathbb{R}^n$, then

$$\mathbb{E}(X \mid X_t) = u(X_t).$$

Indeed, for any bounded measurable $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}(\varphi(X_t) \, \mathbb{E}(X \mid X_t)) = \mathbb{E}(X \varphi(X_t)) = \int_{\mathbb{R}^n} x P_t \varphi f \, d\gamma = \int_{\mathbb{R}^n} \varphi P_t(xf) d\gamma$$

while

$$\mathbb{E}(\varphi(X_t) u(X_t)) = \int_{\mathbb{R}^n} P_t(\varphi u) f \, d\gamma = \int_{\mathbb{R}^n} \varphi \, u P_t f \, d\gamma$$

from which the announced claim follows.

As a consequence, for every $t \geq 0$

$$4 \operatorname{sh}^2(t) \, I(P_t f) = \int_{\mathbb{R}^n} P_t f \left| \frac{P_t(xf)}{P_t f} - e^{-t} x \right|^2 d\gamma = \mathbb{E}\left(|\mathbb{E}(X \mid X_t) - e^{-t} X_t|^2 \right). \qquad (20)$$

In this Ornstein-Uhlenbeck context, the representation (20) of the Fisher information is the analogue of the Minimum Mean-Square Error (MMSE) emphasized in [25, 26, 28, 32]. The proximity with the linear estimator $X_t^\ell = e^{-t} X_t$ will turn out essential in the further developments. In particular, (20) rewrites as

$$4 \operatorname{sh}^2(t) \, I(P_t f) = \mathbb{E}\left(|\mathbb{E}(X - X_t^\ell \mid X_t)|^2 \right), \quad t \geq 0.$$

At this stage, it might be of interest to point out that if $\mathbb{E}(X \mid X_t) = X_t^\ell$ (for some $t > 0$), then X must be standard normal. Indeed, under this assumption, for any smooth $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}\big(\varphi(X_t)X\big) \;=\; \mathbb{E}\big(\varphi(X_t)X_t^\ell\big),$$

that is

$$(1 - e^{-2t})\,\mathbb{E}\big(X\,\varphi(X_t)\big) \;=\; e^{-t}\,\sqrt{1 - e^{-2t}}\;\mathbb{E}\big(N\,\varphi(X_t)\big).$$

After integration by parts with respect to N,

$$\mathbb{E}\big(X\,\varphi(X_t)\big) \;=\; e^{-t}\,\mathbb{E}\big(\nabla\varphi(X_t)\big).$$

For $\varphi(x) = e^{i\lambda \cdot x}$, $\lambda \in \mathbb{R}^n$, this amounts again to the differential equation $\nabla F = -\lambda F$ for the Fourier transform $F(\lambda) = \mathbb{E}(e^{i\lambda \cdot X})$ of X.

We next investigate the analogue of (20) for the time derivative of the Fisher information $I(P_t f)$. Actually, this derivative is at the root of the representation formula (4) that we recall here

$$\mathrm{H}(f) \;=\; \frac{1}{2}\,\mathrm{I}(f) - \int_0^\infty \int_{\mathbb{R}^n} P_t f\,\big|\mathrm{Hess}(\log P_t f)\big|^2 d\gamma\, dt. \tag{21}$$

Indeed, de Bruijn's formula first expresses that

$$\frac{d}{dt}\,\mathrm{H}(P_t f) \;=\; -\int_{\mathbb{R}^n} P_t f |\nabla \log P_t f|^2 d\gamma \;=\; -\,\mathrm{I}(P_t f)$$

so that

$$\mathrm{H}(f) \;=\; \int_0^\infty \mathrm{I}(P_t f)dt.$$

At the second order, following the Γ-calculus as exposed e.g. in [2, 4],

$$\begin{aligned}
\frac{d}{dt}\,\mathrm{I}(P_t f) &= -2 \int_{\mathbb{R}^n} P_t f\,\Gamma_2(\log P_t f)d\gamma \\
&= -2 \int_{\mathbb{R}^n} P_t f\Big[\big|\mathrm{Hess}(\log P_t f)\big|^2 + \big|\nabla(\log P_t f)\big|^2\Big]d\gamma \\
&= -2 \int_{\mathbb{R}^n} P_t f\,\big|\mathrm{Hess}(\log P_t f)\big|^2 d\gamma - 2\,\mathrm{I}(P_t f).
\end{aligned} \tag{22}$$

Note, as is classical, that this differential equation implies the exponential decay of the Fisher information

$$\mathrm{I}(P_t f) \;\leq\; e^{-2t}\,\mathrm{I}(f), \quad t \geq 0. \tag{23}$$

By integration by parts, it follows from (22) that

$$\mathrm{H}(f) \;=\; \int_0^\infty e^{-2t}\big(e^{2t}\,\mathrm{I}(P_t f)\big)dt \;=\; \frac{1}{2}\,\mathrm{I}(f) + \frac{1}{2}\int_0^\infty e^{-2t}\,\frac{d}{dt}\big(e^{2t}\,\mathrm{I}(P_t f)\big)dt$$

and hence (21).

Accordingly, in the study of the deficit $\delta(f) = \frac{1}{2}\,\mathrm{I}(f) - \mathrm{H}(f)$, we are therefore interested into

$$\int_{\mathbb{R}^n} P_t f\,\big|\mathrm{Hess}(\log P_t f)\big|^2 d\gamma \;=\; \int_{\mathbb{R}^n} P_t f\,\bigg|\frac{\mathrm{Hess}(P_t f)}{P_t f} - \frac{\nabla P_t f \otimes \nabla P_t f}{(P_t f)^2}\bigg|^2 d\gamma.$$

14

We analyze this expression as the Fisher information in (20). Taking partial derivative ∂_j in (19) first yields that

$$\partial_j P_t(x_i f) = e^{-t} \delta_{ij} P_t f + e^{-t} x_i \partial_j P_t f + 2 \operatorname{sh}(t) \partial_{ij} P_t f$$

for all $i, j = 1, \ldots, n$. After a further use of (19),

$$2 \operatorname{sh}(t) \partial_j P_t(x_i f) = (1 - e^{-2t}) \delta_{ij} P_t f + e^{-t} x_i P_t(x_j f) - e^{-2t} x_i x_j P_t f + 4 \operatorname{sh}^2(t) \partial_{ij} P_t f.$$

Applying then (19) one more time but to $x_i f$ for every i, we finally get that

$$4 \operatorname{sh}^2(t) \partial_{ij} P_t f = P_t(x_i x_j f) - (1 - e^{-2t}) \delta_{ij} P_t f - e^{-t} \big[x_i P_t(x_j f) + x_j P_t(x_i f) \big] + e^{-2t} x_i x_j P_t f.$$

On the other hand, always from (19), for all $i, j = 1, \ldots, n$,

$$4 \operatorname{sh}^2(t) \partial_i P_t f \partial_j P_t f = P_t(x_i f) P_t(x_j f) - e^{-t} \big[x_i P_t(x_j f) + x_j P_t(x_i f) \big] P_t f + e^{-2t} x_i x_j (P_t f)^2.$$

In compact notation, it follows that

$$4 \operatorname{sh}^2(t) \left[\frac{\operatorname{Hess}(P_t f)}{P_t f} - \frac{\nabla P_t f \otimes \nabla P_t f}{(P_t f)^2} \right] = \frac{P_t(x \otimes x f)}{P_t f} - \frac{P_t(x f)}{P_t f} \otimes \frac{P_t(x f)}{P_t f} - (1 - e^{-2t}) \operatorname{Id}.$$

Recall that if $u = \frac{P_t(x f)}{P_t f} : \mathbb{R}^n \to \mathbb{R}^n$, then $\mathbb{E}\big(X \mid X_t\big) = u(X_t)$ (as vectors). Exactly in the same way, if $v = \frac{P_t(x \otimes x f)}{P_t f}$, then

$$\mathbb{E}\big(X \otimes X \mid X_t\big) = v(X_t)$$

as $n \times n$ matrices. Hence (recall $\mathbb{E}(|X|^2) < \infty$), setting

$$Z_t = \operatorname{Cov}\big(X \mid X_t\big) = \mathbb{E}\big(X \otimes X \mid X_t\big) - \mathbb{E}\big(X \mid X_t\big) \otimes \mathbb{E}\big(X \mid X_t\big), \tag{24}$$

it holds

$$16 \operatorname{sh}^4(t) \int_{\mathbb{R}^n} P_t f \, \big| \operatorname{Hess}(\log P_t f) \big|^2 d\gamma = \mathbb{E}\Big(\big| Z_t - (1 - e^{-2t}) \operatorname{Id} \big|^2 \Big).$$

From (21), we may therefore emphasize the following identity which will be the cornerstone for the analysis of the deficit.

Proposition 7. *Under the preceding notation,*

$$\delta(f) = \int_0^\infty \frac{1}{16 \operatorname{sh}^4(t)} \mathbb{E}\Big(\big| Z_t - (1 - e^{-2t}) \operatorname{Id} \big|^2 \Big) dt. \tag{25}$$

4 Proof of Theorem 1

This section addresses the proof of Theorem 1, therefore based on the representation (25) of Proposition 7.

Since $\delta(f) = \delta(f_b)$ where f_b is the shifted density from (12), it is enough to deal with the centered case $b = 0 = \mathbb{E}(X)$.

For every $t \geq 0$, we bound from below the expectation $\mathbb{E}\big(\big|Z_t - (1 - e^{-2t})\mathrm{Id}\big|^2\big)$. Recall the linear estimator $X_t^\ell = e^{-t}X_t$. Observe that

$$Z_t = \mathrm{Cov}(X \mid X_t) = \mathrm{Cov}(X - X_t^\ell \mid X_t)$$

so that

$$\mathbb{E}\Big(\big|Z_t - (1 - e^{-2t})\mathrm{Id}\big|^2\Big) = \mathbb{E}\Big(\big|\mathrm{Cov}(X - X_t^\ell \mid X_t) - (1 - e^{-2t})\mathrm{Id}\big|^2\Big). \tag{26}$$

By Jensen's inequality

$$\mathbb{E}\Big(\big|Z_t - (1 - e^{-2t})\mathrm{Id}\big|^2\Big) \geq \sum_{i,j=1}^n \Big[\mathbb{E}(U_i U_j) - (1 - e^{-2t})^2(\Gamma_{ij} - \delta_{ij})\Big]^2 \tag{27}$$

where U_i, $i = 1, \ldots n$, are the coordinates of the vector $U = \mathbb{E}(X - X_t^\ell \mid X_t)$.

Assume therefore that $\Gamma \leq \mathrm{Id}$. Hence

$$\mathbb{E}\Big(\big|Z_t - (1 - e^{-2t})\mathrm{Id}\big|^2\Big) \geq \sum_{i,j=1}^n \big[\mathbb{E}(U_i U_j)\big]^2.$$

In particular, for every unit vector $\alpha = (\alpha_1, \ldots, \alpha_n)$ in $\mathbb{R}^n$,

$$\sum_{i,j=1}^n \big[\mathbb{E}(U_i U_j)\big]^2 \geq \Big[\mathbb{E}\Big(\sum_{i,j=1}^n \alpha_i \alpha_j U_i U_j\Big)\Big]^2$$

so that

$$\mathbb{E}\Big(\big|Z_t - (1 - e^{-2t})\mathrm{Id}\big|^2\Big) \geq \Big[\mathbb{E}\Big(\big[\alpha \cdot \mathbb{E}(X - X_t^\ell \mid X_t)\big]^2\Big)\Big]^2.$$

Let now $\psi : \mathbb{R}^n \to \mathbb{R}$ smooth such that $\mathbb{E}(\psi(X_t)^2) \leq 1$. For α fixed, by the Cauchy-Schwarz inequality,

$$\mathbb{E}\Big(\big[\alpha \cdot \mathbb{E}(X - X_t^\ell \mid X_t)\big]^2\Big) \geq \Big[\mathbb{E}\Big(\psi(X_t)\, \alpha \cdot \big[\mathbb{E}(X - X_t^\ell \mid X_t)\big]\Big)\Big]^2.$$

Now,

$$\begin{aligned}
\mathbb{E}\Big(\psi(X_t)\, \alpha \cdot \big[\mathbb{E}(X - X_t^\ell \mid X_t)\big]\Big) \\
&= \mathbb{E}\big(\psi(X_t)\, \alpha \cdot [X - e^{-t} X_t]\big) \\
&= (1 - e^{-2t})\, \mathbb{E}\big(\alpha \cdot X\, \psi(X_t)\big) - e^{-t} \sqrt{1 - e^{-2t}}\, \mathbb{E}\big(\alpha \cdot N\, \psi(X_t)\big) \\
&= (1 - e^{-2t})\, \alpha \cdot \mathbb{E}\big(X\psi(X_t) - e^{-t}\, \nabla\psi(X_t)\big)
\end{aligned}$$

where integration by parts with respect to N is performed in the last step. Taking the supremum over all unit vectors α, it therefore follows that for any (smooth) $\psi : \mathbb{R}^n \to \mathbb{R}$ such that $\mathbb{E}(\psi(X_t)^2) \leq 1$,

$$\mathbb{E}\Big(\big|Z_t - (1 - e^{-2t})\mathrm{Id}\big|^2\Big) \geq (1 - e^{-2t})^4 \Big|\mathbb{E}\big(X\psi(X_t) - e^{-t}\, \nabla\psi(X_t)\big)\Big|^4.$$

16

Switching back to semigroup notation and recalling that X_t has distribution $P_t f d\gamma$, for any $t \geq 0$,

$$\mathbb{E}\left(\left|Z_t - (1 - e^{-2t})\mathrm{Id}\right|^2\right) \geq (1 - e^{-2t})^4 \left|\int_{\mathbb{R}^n} [x P_t \psi - \nabla P_t \psi] d\mu\right|^4$$

where we recall that $d\mu = f d\gamma$ and $\int_{\mathbb{R}^n} \psi^2 P_t f d\gamma = \mathbb{E}(\psi(X_t)^2) \leq 1$. From (25), we therefore obtain that

$$\delta(f) \geq \int_0^\infty e^{-4t} \left|\int_{\mathbb{R}^n} [x P_t \psi - \nabla P_t \psi] d\mu\right|^4 dt.$$

By Jensen's inequality in the t variable,

$$\delta(f) \geq \frac{1}{4} \left|\int_{\mathbb{R}^n} [x\varphi - \nabla \varphi] d\mu\right|^4$$

where $\varphi = R\psi$. Considering the definition of $\mathcal{D}(\mu, \gamma)$, the proof of Theorem 1 is complete.

To conclude this section, we present a variation on Theorems 1 and 2 which somehow takes into account the covariance condition.

Consider, for each $0 < \varepsilon \leq 1$, the modified class $\mathcal{R}_\varepsilon$ consisting of the functions

$$\varphi = R_\varepsilon \psi = 4 \int_s^\infty e^{-4t} P_t \psi \, dt$$

with $e^{-4s} = \varepsilon$ and $\int_{\mathbb{R}^n} \psi^2 P_t f d\gamma \leq 1$ for every $t \geq 0$. Note that $\mathcal{R}_1 = \mathcal{R}$. It is easily seen that $\mathcal{R}_\varepsilon$ is a determining class for any ε. Define accordingly $\mathcal{D}_\varepsilon(\mu, \gamma)$. The following statement covers in particular (15), and with the flexibility on $\varepsilon > 0$ actually allows for a lower bound on the deficit independently of the size of $\|\Gamma - \mathrm{Id}\| = \sup_{|\alpha|=1}(\Gamma - \mathrm{Id})\alpha \cdot \alpha$.

Theorem 8. *Let f be a probability density on $\mathbb{R}^n$ and let $d\mu = f d\gamma$. Assume that μ has barycenter b and covariance matrix Γ. Then, for every $0 < \varepsilon \leq 1$,*

$$2\delta(f) + \varepsilon \|\Gamma - \mathrm{Id}\|^2 \geq \frac{1}{4\,\varepsilon^3} \mathcal{D}_\varepsilon(\mu_b, \gamma)^4. \tag{28}$$

Whenever $\delta(f) > 0$, a sensible choice for ε could be

$$\varepsilon = \varepsilon(\delta) = \min\left(1, \frac{\delta(f)}{\|\Gamma - \mathrm{Id}\|^2}\right)$$

yielding

$$3\delta(f) \geq \frac{1}{4\,\varepsilon(\delta)^3} \mathcal{D}_{\varepsilon(\delta)}(\mu_b, \gamma)^4 \geq \frac{1}{4} \mathcal{D}_{\varepsilon(\delta)}(\mu_b, \gamma)^4. \tag{29}$$

Proof. Assume that $b = 0$ and start from (27). For any unit vector α in $\mathbb{R}^n$,

$$\mathbb{E}\left(\left|Z_t - (1 - e^{-2t})\mathrm{Id}\right|^2\right) \geq \left[\mathbb{E}\left([\alpha \cdot U]^2\right) - (1 - e^{-2t})^2 \left[(\Gamma - \mathrm{Id})\alpha \cdot \alpha\right]\right]^2$$

17

itself lower-bounded by

$$\frac{1}{2}\left[\mathbb{E}\big([\alpha \cdot U]^2\big)\right]^2 - 2(1 - e^{-2t})^4\big[(\Gamma - \mathrm{Id})\alpha \cdot \alpha\big]^2.$$

Arguing as above,

$$\mathbb{E}\Big(\big|Z_t - (1 - e^{-2t})\mathrm{Id}\big|^2\Big) \geq (1 - e^{-2t})^4\left[\frac{1}{2}\left|\int_{\mathbb{R}^n}\big[xP_t\psi - \nabla P_t\psi\big]d\mu\right|^4 - 2\|\Gamma - \mathrm{Id}\|^2\right].$$

Now, from (25) of Proposition 7, we may bound from below the deficit for every $s > 0$ by

$$\delta(f) \geq \int_s^\infty \frac{1}{16\,\mathrm{sh}^4(t)}\,\mathbb{E}\Big(\big|Z_t - (1 - e^{-2t})\mathrm{Id}\big|^2\Big)dt.$$

Hence, by the preceding,

$$\delta(f) \geq \frac{1}{2}\int_s^\infty e^{-4t}\left|\int_{\mathbb{R}^n}\big[xP_t\psi - \nabla P_t\psi\big]d\mu\right|^4 dt - 2\int_s^\infty e^{-4t}\,\|\Gamma - \mathrm{Id}\|^2 dt$$

$$\geq \frac{1}{2}\int_s^\infty e^{-4t}\left|\int_{\mathbb{R}^n}\big[xP_t\psi - \nabla P_t\psi\big]d\mu\right|^4 dt - \frac{1}{2}e^{-4s}\,\|\Gamma - \mathrm{Id}\|^2.$$

By the same Jensen's inequality argument, but now on the interval (s, ∞),

$$\delta(f) \geq \frac{1}{8}e^{12s}\left|\int_{\mathbb{R}^n}\big[x\varphi - \nabla\varphi\big]d\mu\right|^4 - \frac{1}{2}e^{-4s}\,\|\Gamma - \mathrm{Id}\|^2$$

where now $\varphi = R_\varepsilon\psi$ with $\varepsilon = e^{-4s}$. Theorem 8 then follows. $\qquad\square$

5 Proof of Theorem 3

As for the preceding theorems, we may and do assume that $b = 0$. From the latter (28) (with $\varepsilon = 1$), we get that

$$2\delta(f) \geq \frac{1}{4}\mathcal{D}(\mu, \gamma)^4 - \big\|\Gamma - \mathrm{Id}\big\|^2. \tag{30}$$

The challenge now is to control the covariance matrix by the deficit. Assume therefore that $\|\Gamma - \mathrm{Id}\| > 0$, otherwise apply Theorem 1. Recalling that $Z_t = \mathrm{Cov}(X \mid X_t)$, for any $t \geq 0$ and any unit vector α in $\mathbb{R}^n$, by Jensen's inequality,

$$\mathbb{E}\Big(\big|Z_t - (1 - e^{-2t})\mathrm{Id}\big|^2\Big) = \mathbb{E}\Big(\big|\mathrm{Cov}(X \mid X_t) - (1 - e^{-2t})\mathrm{Id}\big|^2\Big)$$

$$\geq \Big[(\Gamma - \mathrm{Id})\alpha \cdot \alpha + e^{-2t} - \mathbb{E}\Big(\big[\mathbb{E}(\alpha \cdot X \mid X_t)\big]^2\Big)\Big]^2.$$

By a rough estimate,

$$\mathbb{E}\Big(\big|Z_t - (1 - e^{-2t})\mathrm{Id}\big|^2\Big) \geq \frac{1}{2}\big[(\Gamma - \mathrm{Id})\alpha \cdot \alpha\big]^2 - 2\,e^{-4t} - 2\Big[\mathbb{E}\Big(\big[\mathbb{E}(\alpha \cdot X \mid X_t)\big]^2\Big)\Big]^2$$

18

and taking the supremum over α, for any $t \geq 0$,

$$\mathbb{E}\Big(\big|Z_t - (1 - e^{-2t})\mathrm{Id}\big|^2\Big) \geq \frac{1}{2}\|\Gamma - \mathrm{Id}\|^2 - 2\,e^{-4t} - 2\sup_{|\alpha|=1}\Big[\mathbb{E}\Big(\big[\mathbb{E}(\alpha \cdot X \mid X_t)\big]^2\Big)\Big]^2. \qquad (31)$$

The following lemma is the compactness argument from which the conclusion will follow.

Lemma 9. *Set*

$$\rho(t) \;=\; \sup_{|\alpha|=1} \mathbb{E}\Big(\big[\mathbb{E}(\alpha \cdot X \mid X_t)\big]^2\Big), \quad t \geq 0.$$

Then, whenever X is centered and $\mathrm{I}(f) < \infty$,

$$\lim_{t\to\infty} \rho(t) = 0.$$

Proof. Fix first a unit vector $\alpha \in \mathbb{R}^n$, and let $t \geq 0$. We develop the proof in the semigroup language. As discussed in Section 3, $\mathbb{E}(\alpha \cdot X \mid X_t) = u_\alpha(X_t)$ where $u_\alpha = \frac{P_t(\alpha \cdot x f)}{P_t f}$. Hence

$$\mathbb{E}\Big(\big[\mathbb{E}(\alpha \cdot X \mid X_t)\big]^2\Big) \;=\; \int_{\mathbb{R}^n} \frac{P_t(\alpha \cdot x f)^2}{P_t f}\, d\gamma.$$

Now,

$$\begin{aligned}
\int_{\mathbb{R}^n} \frac{P_t(\alpha \cdot x f)^2}{P_t f}\, d\gamma \;&\leq\; \int_{\{|P_t(\alpha \cdot xf)|\leq P_t f\}} \frac{P_t(\alpha \cdot x f)^2}{P_t f}\, d\gamma + \int_{\{|P_t(\alpha \cdot xf)|\geq P_t f\}} \frac{P_t(\alpha \cdot x f)^2}{P_t f}\, d\gamma \\
&\leq\; \int_{\mathbb{R}^n} \big|P_t(\alpha \cdot x f)\big|\,d\gamma + \int_{A_t} \frac{P_t(\alpha \cdot x f)^2}{P_t f}\, d\gamma
\end{aligned} \qquad (32)$$

where $A_t = \{|P_t(\alpha \cdot x f)| \geq P_t f\}$.

Observe that

$$\begin{aligned}
\gamma(A_t) \;&\leq\; \gamma\big(2|P_t(\alpha \cdot x f)| \geq 1\big) + \gamma\big(2|P_t f - 1| \geq 1\big) \\
&\leq\; 2\int_{\mathbb{R}^n} \big|P_t(\alpha \cdot x f)\big|\,d\gamma + 2\int_{\mathbb{R}^n} \big|P_t f - 1\big|\,d\gamma.
\end{aligned}$$

Since $P_t(\alpha \cdot x f)$ is centered with respect to γ, by the Gaussian L^1-Poincaré inequality,

$$\int_{\mathbb{R}^n} \big|P_t(\alpha \cdot x f)\big|\,d\gamma \;\leq\; 2\int_{\mathbb{R}^n} \big|\nabla\big(P_t(\alpha \cdot x f)\big)\big|\,d\gamma \;\leq\; 2\,e^{-t}\int_{\mathbb{R}^n} \big|\nabla(\alpha \cdot x f)\big|\,d\gamma.$$

Now, since α is a unit vector,

$$\int_{\mathbb{R}^n} \big|\nabla(\alpha \cdot x f)\big|\,d\gamma \;\leq\; 1 + \int_{\mathbb{R}^n} |\alpha \cdot x|\,|\nabla f|\,d\gamma \;\leq\; 1 + \Big(\int_{\mathbb{R}^n} |x|^2 f\, d\gamma\Big)^{1/2} \mathrm{I}(f)^{1/2}.$$

Again by the Gaussian L^1-Poincaré and Cauchy-Schwarz inequalities,

$$\int_{\mathbb{R}^n} \big|P_t f - 1\big|\,d\gamma \;\leq\; 2\int_{\mathbb{R}^n} |\nabla P_t f|\,d\gamma \;\leq\; 2\sqrt{\mathrm{I}(P_t f)},$$

19

so that, by (23),

$$\int_{\mathbb{R}^n} |P_t f - 1| d\gamma \;\leq\; 2\,e^{-t}\sqrt{I(f)}\,.$$

These estimates already ensure that, uniformly in $|\alpha| = 1$,

$$\lim_{t\to 0} \int_{\mathbb{R}^n} \big|P_t(\alpha\cdot xf)\big| d\gamma \;=\; 0$$

and $\lim_{t\to 0} \gamma(A_t) = 0$.

To handle the second term in (32), note that by the Cauchy-Schwarz inequality (for P_t),

$$
\begin{aligned}
\int_{A_t} \frac{P_t(\alpha\cdot xf)^2}{P_t f}\, d\gamma
&\leq \int_{A_t} P_t\big((\alpha\cdot x)^2 f\big) d\gamma \\
&= \int_{\mathbb{R}^n} P_t(1_{A_t})(\alpha\cdot x)^2 f\, d\gamma \\
&\leq \int_{\mathbb{R}^n} P_t(1_{A_t})|x|^2 f\, d\gamma.
\end{aligned}
$$

Since $\int_{\mathbb{R}^n} P_t(1_{A_t}) d\gamma = \gamma(A_t)$, the conclusion follows by dominated convergence. Lemma 9 is established. $\qquad\square$

On the basis of Lemma 9, we may now conclude the proof of Theorem 3. Going back to (31), for every $t \geq 0$

$$\mathbb{E}\Big(\big|Z_t - (1 - e^{-2t})\mathrm{Id}\big|^2\Big) \;\geq\; \frac{1}{2}\|\Gamma - \mathrm{Id}\|^2 - 2\,e^{-4t} - 2\,\rho(t).$$

By Lemma 9, choose t_0 large enough so that

$$2\,e^{-4t} + 2\,\rho(t) \;\leq\; \frac{1}{4}\|\Gamma - \mathrm{Id}\|^2$$

for every $t \geq t_0$. Then, for $t \geq t_0$,

$$\mathbb{E}\Big(\big|Z_t - (1 - e^{-2t})\mathrm{Id}\big|^2\Big) \;\geq\; \frac{1}{4}\|\Gamma - \mathrm{Id}\|^2$$

and

$$\delta(f) \;\geq\; \int_{t_0}^{\infty} \frac{1}{16\,\mathrm{sh}^4(t)}\, \mathbb{E}\Big(\big|Z_t - (1 - e^{-2t})\mathrm{Id}\big|^2\Big) \;\geq\; \frac{e^{-4t_0}}{16}\,\|\Gamma - \mathrm{Id}\|^2. \tag{33}$$

The conclusion of the proof of Theorem 3 is then a suitable combination of (33) and (30).

It may be observed that, in dimension one, the conclusion of Lemma 9 actually amounts, by (20), to the following result which we present as a statement of possible independent interest. From the classical exponential decay (23), $e^{2t}\,I(P_t f) \leq I(f)$ for every $t \geq 0$. Under the centering $\int_{\mathbb{R}} xf d\gamma = 0$, we actually have

Corollary 10. *Let f be a smooth probability density on the real line $\mathbb{R}$ with respect to γ such that $\int_{\mathbb{R}} xf d\gamma = 0$, $\int_{\mathbb{R}} |x|^2 f d\gamma < \infty$ and $I(f) < \infty$. Then*

$$\lim_{t\to\infty} e^{2t}\,I(P_t f) \;=\; 0.$$

6   An alternate proof of the estimate (6)

To conclude this work, we present in this section an alternate proof of the lower bound (6)

$$\delta(f) \geq \frac{c}{n} W_2(\mu, \gamma)^4 \tag{34}$$

under the condition $\int_{\mathbb{R}^n} |x|^2 d\mu \leq n$ (with the constant $c = \frac{1}{4}$).

Recall Z_t, $t \geq 0$, from (24). By the Cauchy-Schwarz inequality, for every t,

$$n\,\mathbb{E}\Big(\big|Z_t - (1 - e^{-2t})\mathrm{Id}\big|^2\Big) \geq \Big[\mathbb{E}\big(\mathrm{Tr}\big(Z_t - (1 - e^{-2t})\mathrm{Id}\big)\big)\Big]^2. \tag{35}$$

Now, after some straightforward calculations,

$$\mathbb{E}\big(\mathrm{Tr}\big(Z_t - (1 - e^{-2t})\mathrm{Id}\big)\big) = \mathbb{E}\big(|X|^2\big) - \sum_{i=1}^{n} \mathbb{E}\big(\mathbb{E}(X^i \mid X_t)^2\big) - n(1 - e^{-2t})$$

where X^i, $i = 1, \ldots, n$, are the coordinates of the vector X. On the other hand,

$$\mathbb{E}\Big(\big|\mathbb{E}(X \mid X_t) - e^{-t}X_t\big|^2\Big) = \sum_{i=1}^{n} \mathbb{E}\big(\mathbb{E}(X^i \mid X_t)^2\big) - e^{-2t}(2 - e^{-2t})\,\mathbb{E}\big(|X|^2\big) + n\,e^{-2t}(1 - e^{-2t}).$$

Therefore, whenever $\mathbb{E}(|X|^2) = \int_{\mathbb{R}^n} |x|^2 d\mu \leq n$,

$$-\,\mathbb{E}\big(\mathrm{Tr}\big(Z_t - (1 - e^{-2t})\mathrm{Id}\big)\big) \geq \mathbb{E}\Big(\big|\mathbb{E}(X \mid X_t) - e^{-t}X_t\big|^2\Big).$$

Hence, by (20) and (35),

$$n\,\mathbb{E}\Big(\big|Z_t - (1 - e^{-2t})\mathrm{Id}\big|^2\Big) \geq 16\,\mathrm{sh}^4(t)\,\mathrm{I}(P_t f)^2,$$

and combined with (25),

$$\delta(f) \geq \frac{1}{n}\int_0^\infty \mathrm{I}(P_t f)^2 dt. \tag{36}$$

Now write $d\mu_t = P_t f\, d\gamma$, $t \geq 0$, and define

$$w(t) = W_2(\mu, \mu_t).$$

In particular, $w(\infty) = W_2(\mu, \gamma)$ (while $w(0) = 0$). Recall from [33] that $w'(t) \leq \sqrt{\mathrm{I}(P_t f)}$, $t \geq 0$. From the transportation cost inequality (8) applied to μ_t, $w(t)^2 \leq 2\mathrm{H}(P_t f)$, and with the logarithmic Sobolev inequality $w(t)^2 \leq 2\mathrm{H}(P_t f) \leq \mathrm{I}(P_t f)$. Hence, combining these inequalities, for every $t \geq 0$,

$$w(t)^3 w'(t) \leq w(t)^3 \sqrt{\mathrm{I}(P_t f)} \leq \mathrm{I}(P_t f)^2.$$

Together with (36),

$$\delta(f) \geq \frac{1}{n}\int_0^\infty w(t)^3\, w'(t)dt = \frac{1}{4n}\, w(\infty)^4 = \frac{1}{4n}\, W_2(\mu, \gamma)^4$$

and therefore the desired conclusion (34) with $c = \frac{1}{4}$.

References

[1] S. Arstein-Avidan, B. Klartag, C. Schütt, E. Werner. Functional affine-isoperimetry and an inverse logarithmic Sobolev inequality. *J. Funct. Anal.* 262 (2012), 4181–4204.

[2] D. Bakry. L'hypercontractivité et son utilisation en théorie des semigroupes, *École d'Été de Probabilités de Saint-Flour, Lecture Notes in Math.* 1581 (1994), 1–114. Springer.

[3] D. Bakry, M. Émery. Diffusions hypercontractives. *Séminaire de Probabilités XIX, Lecture Notes in Math.* 1123 (1985), 177–206. Springer.

[4] D. Bakry, I. Gentil, M. Ledoux. *Analysis and geometry of Markov diffusion operators.* Grundlehren der mathematischen Wissenschaften 348. Springer (2014).

[5] D. Bakry, M. Ledoux. A logarithmic Sobolev form of the Li-Yau parabolic inequality. *Rev. Mat. Iberoam.* 22 (2006), 683–702.

[6] M. Barchiesi, A. Brancolini, V. Julin. Sharp dimension free quantitative estimates for the Gaussian isoperimetric inequality (2014). *Ann. Probab.*, to appear.

[7] S. Bobkov, N. Gozlan, C. Roberto, P.-M. Samson. Bounds on the deficit in the logarithmic Sobolev inequality. *J. Funct. Anal.* 267 (2014), 4110–4138.

[8] D. Bucur, I. Fragalà. Lower bounds for the Prékopa-Leindler deficit by some distances modulo translations. *J. Convex Anal.* 21 (2014), 289–305.

[9] U. Caglar, M. Fradelizi, O. Guédon, J. Lehec, C. Schütt, E. Werner. Functional versions of L^p-affine surface area and entropy inequalities. *Int. Math. Res. Not. IMRN* 2016 (2016), 1223–1250.

[10] E. Carlen. Some integral identities and inequalities for entire functions and their application to the coherent state transform. *J. Funct. Anal.* 97 (1991), 231–249.

[11] E. Carlen. Superadditivity of Fisher's information and logarithmic Sobolev inequalities. *J. Funct. Anal.* 101 (1991), 194–211.

[12] L. H. Y. Chen, L. Goldstein, Q.-M. Shao. *Normal approximation by Stein's method.* Probability and its Applications. Springer (2011).

[13] A. Cianchi, N. Fusco, F. Maggi, A. Pratelli. The sharp Sobolev inequality in quantitative form. *J. Eur. Math. Soc.* 11 (2009), 1105–1139.

[14] D. Cordero-Erausquin. Transport inequalities for log-concave measures, quantitative forms and applications (2016)

[15] R. Eldan. A two-sided estimate for the Gaussian noise stability deficit. *Invent. Math.* 201 (2015), 561–624.

[16] M. Fathi, E. Indrei, M. Ledoux. Quantitative logarithmic Sobolev inequalities and stability estimates (2014). *Discrete and Continuous Dynamical Systems*, to appear.

[17] A. Figalli. Stability in geometric and functional inequalities. *European Congress of Mathematics*, 585–599. Eur. Math. Soc. (2013).

[18] A. Figalli, D. Jerison. Quantitative stability for the Brunn-Minkowski inequality. *J. Eur. Math. Soc.*, to appear (2014).

[19] A. Figalli, F. Maggi, A. Pratelli. A refined Brunn-Minkowski inequality for convex sets. *Ann. Inst. H. Poincaré Anal. Non-Linéaire* 26 (2009), 2511–2519.

[20] A. Figalli, F. Maggi, A. Pratelli. A mass transportation approach to quantitative isoperimetric inequalities. *Invent. Math.* 182 (2010), 167–211.

[21] A. Figalli, F. Maggi, A. Pratelli. Sharp stability theorems for the anisotropic Sobolev and log-Sobolev inequalities on functions of bounded variation. *Adv. Math.* 242 (2013), 80–101.

[22] F. Feo, E. Indrei, M. R. Posteraro, C. Roberto. Some remarks on the stability of the Log-Sobolev inequality for the Gaussian measure (2015).

[23] N. Fusco, F. Maggi, A. Pratelli. The sharp quantitative isoperimetric inequality. *Ann. of Math.* 168 (2008), 941–980.

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

[25] D. Guo, S. Shamai, S. Verdú. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. Inform. Theory* 51 (2005), 1261–1282.

[26] D. Guo, Y. Wu, S. Shamai, S. Verdú. Estimation in Gaussian noise: properties of the minimum mean-square error. *IEEE Trans. Inform. Theory* 57 (2011), 2371–2385.

[27] M. Ledoux. On an integral criterion for hypercontractivity of diffusion semigroups and extremal functions. *J. Funct. Anal.* 105 (1992) , 444–465.

[28] M. Ledoux. Heat flow derivatives and Minimun Mean-Square Error in Gaussian noise. *IEEE Trans. Inform. Theory* 62 (2016), 3401–3409.

[29] M. Ledoux, I. Nourdin, G. Peccati. Stein's method, logarithmic Sobolev and transport inequalities. *Geom. and Funct. Anal.* 25 (2015), 256–306.

[30] E. Mossel, J. Neeman. Robust dimension free isoperimetry in Gaussian space. *Ann. Probab.* 43 (2015), 971–991.

[31] I. Nourdin, G. Peccati. *Normal approximations with Malliavin calculus: from Stein's method to universality.* Cambridge University Press (2012).

[32] I. Nourdin, G. Peccati, Y. Swan. Integration by parts and representation of information functionals. Proceedings of the 2014 IEEE International Symposium on Information Theory (ISIT), Honolulu (2014), 2217–2221.

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

[34] C. Stein. *Approximate computation of expectations.* Institute of Mathematical Statistics Lecture Notes – Monograph Series, **7**. Institute of Mathematical Statistics.

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

[36] C. Villani. *Optimal transport. Old and new.* Grundlehren der mathematischen Wissenschaften 338. Springer (2009).