Improved one-sided deviation inequalities under regularity assumptions for product measures

Kevin Tanguy
University of Angers, France

Note of June 27, 2019

Abstract

This note is concerned with lower tail estimates for product measures. Some improved deviation inequalities are obtained for functions satisfying some regularity and monotonicity assumptions. The arguments are based on semigroup interpolation together with Harris's negative association inequality and hypercontractive estimates.

1 Introduction

As an introduction we recall some facts about Gaussian concentration of measure (cf. [14]) and Superconcentration theory (cf. [7]).

It is well known that concentration of measure is an effective tool in various mathematical areas (cf. [6]). In a Gaussian setting, classical concentration results typically state that, for a Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ with Lipschitz constant $\|f\|_{\mathrm{Lip}}$,

$$\gamma_n\big(|f - \mathbb{E}_{\gamma_n}[f]| \geq t\big) \leq 2e^{-\frac{t^2}{2\|f\|^2_{Lip}}}, \quad t \geq 0, \tag{1.1}$$

with γ_n the standard Gaussian measure on $\mathbb{R}^n$. Another example of concentration of measure is the Poincaré inequality satisfied by γ_n. Namely, for $f \in L^2(\gamma_n)$ smooth enough :

$$\mathrm{Var}_{\gamma_n}(f) \leq \int_{\mathbb{R}^n} |\nabla f|^2 d\gamma_n, \tag{1.2}$$

where $|\cdot|$ stands for the Euclidean norm on $\mathbb{R}^n$. As effective as (1.1) and (1.2) are, their generality can lead to sub-optimal bounds in some particular cases. For instance, consider the 1-Lipschitz function on $\mathbb{R}^n$ $f(x) = \max_{i=1,\dots,n} x_i$. At the level of the variance, (1.2) gives

$$\mathrm{Var}(M_n) \leq 1,$$

with $M_n = \max_{i=1,\dots,n} X_i$ where $(X_1, \dots, X_n)$ stands for a standard Gaussian random vector in $\mathbb{R}^n$, whereas it has been proven that $\mathrm{Var}(M_n) \leq C/\log n$ with $C > 0$ a numerical constant. At an exponential level (1.1) is not satisfying

either. Indeed, it is well known in Extreme Value theory (cf. [12], pages $14-15$) that M_n can be renormalized by some numerical constants $a_n = \sqrt{2\log n}$ and $b_n = a_n - \frac{\log 4\pi + \log\log n}{2a_n}$, $n \geq 1$, such that

$$a_n(M_n - b_n) \to \Lambda_0 \quad \text{as} \quad n \to \infty$$

in distribution where Λ_0 corresponds to the Gumbel distribution with cumulative distribution function :

$$\mathbb{P}(\Lambda_0 \leq x) = \exp(-e^{-x}), \quad x \in \mathbb{R}.$$

Then, it is clear that the asymptotics of Λ_0 are not Gaussian but rather exponential on the right tail and double exponential on the left tail. It is now obvious that (1.1) and (1.2) lead to sub-optimal results for the function $f(x) = \max_{i=1,\dots,n} x_i$. This is referred to as Superconcentration phenomenon (cf. [7]). This kind of phenomenon occurs for different functionals of Gaussian random variables and has been studied in [5, 20, 21, 19, 24]....

Recently, additional convexity assumptions have been fruitfully used by Paouris and Valettas in order to improve the concentration inequality (1.1). In the context of small ball probabilities and random Dvoretzky's Theorem, these two authors improved the lower tail of any convex function, thanks to Ehrard's inequality, in [17]. More precisely, they obtained

Theorem 1.1. *[Paouris, Valettas] Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, then the following holds*

$$\gamma_n\left(f - \int_{\mathbb{R}^n} f d\gamma_n \leq -t\right) \leq e^{-c\frac{t^2}{\mathrm{Var}_{\gamma_n}(f)}}, \quad t > 1 \tag{1.3}$$

where $c > 0$ is a universal constant.

Remark. Of course, the improvements stays in the fact that $\mathrm{Var}_{\gamma_n}(f) \leq \|f\|^2_{\mathrm{Lip}}$ as we have just seen on the basic example of the maximum of n independent standard Gaussian random variables. Ehrhard's inequality has also been used by Valettas in [24] where he proved that (1.1) is tight if the convex function f is not superconcentrated.

Besides, the work from [17] has been used by Valettas to produce some variations of Theorem 1.1. Indeed, as consequence of his inequality with Paouris, combined with transportation-type arguments, he obtained (cf. [24], section 2.1.3) concentration inequalities for nondecreasing, convex functions in a log-concave measures setting.

A similar result as Theorem 1.1 or 1.3 has also been obtained in [9]. Instead of convexity, the author of [9] assumes that f belongs to the set

$$\mathcal{H}_+ = \{f \in C^2(\mathbb{R}^n, \mathbb{R}) \quad ; \quad \text{monotone with} \quad \partial^2_{ij} f \geq 0, \ \forall i, j = 1, \dots, n\}$$

and obtained the following lower deviation estimate

Theorem 1.2 (Nguyen Tien)**.** *Let $f \in \mathcal{H}_+$ be, then the following holds*

$$\gamma_n\left(f - \int_{\mathbb{R}^n} f d\gamma_n \leq -t\right) \leq e^{-\frac{t^2}{\mathrm{Var}_{\gamma_n}(f)}}, \quad t \geq 0. \tag{1.4}$$

The purpose of this note is the following : semigroup's arguments together with Harris negative association Lemma and hypercontractive estimates will be used to obtain a deviation inequality for the lower tail of functions belonging to $\mathcal{F}_+$ where

$$\mathcal{F}_+ = \{f \in C^2(\mathbb{R}^n, \mathbb{R}) \quad ; \quad \partial_i f \geq 0, \quad \partial^2_{ij} f \geq 0, \ \forall i, j = 1, \ldots, n\}.$$

The obtained deviation inequalities can be seen as an extension, for the lower tail, of Theorem 1.2. The cost of this extension is the larger quantity $\|\nabla f\|^2_\phi$ (instead of $\mathrm{Var}_\mu(f)$) in the exponential. As it will be explained in remark 1, if the measures $(\mu_i)_{i=1,\ldots,n}$ are symmetric, one can substitute $\mathcal{F}_+$ by the larger set $\mathcal{H}_+$.

At this stage, let us also notice that there exists some functions in $\mathcal{F}_+$ which are not convex. In dimension 2, f is not convex if $\mathrm{Det}\big(\mathrm{Hess}\, f(x,y)\big) < 0$ for some $(x,y) \in \mathbb{R}^2$ (i.e. $A = (x,y)$ is a saddle point and $\mathrm{Hess} f$ is not positive semi-definite).

For instance, consider the real function $f(x) = \max(x, 0)$ and set $h(x,y) = f(x)^2 + 4f(x)f(y) + f(y)^2$ for any $(x,y) \in \mathbb{R}^2$. Then h is an element of $\mathcal{F}_+$ which is not convex. As another example, one can consider the function

$$f(x) = \begin{cases} x+1 & \text{when} \quad x \geq 0, \\ e^x & \text{otherwise.} \end{cases}$$

Then, set $g(x,y) = f(x)^2 + 22f(x)f(y) + f(y)^2$. It is a simple matter to check that $g \in \mathcal{F}_+$. Besides, g is not convex on $\mathbb{R}^2$ since $\mathrm{Det}\big(\mathrm{Hess}\, g(x,y)\big) < 0$ on $\mathbb{R}^2_+$. In conclusion, $\mathcal{H}_+$ and $\mathcal{F}_+$ are not restricted classes of the set of convex functions.

Now, let us describe in more details our setting and state our main result.

Let $n \geq 1$ be fixed and consider $\mu = \mu_1 \otimes \ldots \otimes \mu_n$ where, for any $i = 1, \ldots, n$, $d\mu_i = e^{-V_i(x)}dx$ are probability measures on $\mathcal{B}(\mathbb{R})$, the Borel σ-algebra of $\mathbb{R}$, and $V_i : \mathbb{R} \to \mathbb{R}$ are smooth potentials. In the sequel, we will assume that there exists $\kappa_i \in \mathbb{R}$ such that

$$V_i''(x) \geq -\kappa_i \quad \forall x \in \mathbb{R} \quad \text{and} \quad i = 1, \ldots, n$$

and will denote by $\kappa = \max_{i=1,\ldots,n} \kappa_i$.

Now, let us recall some facts about functional inequalities and their links with related semigroups. General references on semigroups, functional inequalities and concentration of measures are [1, 14, 6].

In our setting, $d\mu(x) = e^{-V(x)}dx$ is a probability measure on $\mathcal{B}(\mathbb{R}^n)$, the Borel σ-algebra of $\mathbb{R}^n$, with

$$V(x) = \sum_{i=1}^{n} V_i(x_i) \quad \text{and} \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

It is classical that such measures can be seen as an invariant and reversible measure of the associated diffusion operator $L = \Delta - \nabla V \cdot \nabla$. The operator L generates the Markov semigroup of operators $(P_t)_{t\geq 0}$ and defines by integration by parts the Dirichlet form

$$\mathcal{E}(f,g) = \int_{\mathbb{R}^n} f(-Lg)d\mu = \int_{\mathbb{R}^n} \nabla f \cdot \nabla g d\mu \tag{1.5}$$

for some smooth functions f, g on $\mathbb{R}^n$. The set of functions for which the preceding expression make sense is called the Dirichlet domain of L. We denote by $\mathcal{D}(L)$ such set.

Given such a couple (L, μ), it is said to satisfy a spectral gap, or Poincaré, inequality if there is a constant $\lambda > 0$ such that for all functions f of the Dirichlet domain

$$\lambda \mathrm{Var}_\mu(f) \leq \mathcal{E}(f,f). \tag{1.6}$$

with $\mathrm{Var}_\mu(f) = \int_{\mathbb{R}^n} f^2 d\mu - (\int_{\mathbb{R}^n} f d\mu)^2$. Similarly, it satisfies a logarithmic Sobolev inequality if there exists a constant $\rho > 0$ such that for all functions f of the Dirichlet domain,

$$\rho \mathrm{Ent}_\mu(f^2) \leq 2\mathcal{E}(f,f). \tag{1.7}$$

with $\mathrm{Ent}_\mu(f) = \int_{\mathbb{R}^n} f \log f d\mu - (\int_{\mathbb{R}^n} f d\mu)(\log \int_{\mathbb{R}^n} f d\mu)$ and $f > 0$.

One speaks of the spectral gap constant (of (L, μ)) as the largest $\lambda > 0$ for which (1.6) holds, and of the logarithmic Sobolev constant (of (L, μ)) as the best $\rho > 0$ for which (1.7) holds. We still use λ and ρ to designate these constants. It is classical (cf. [14]) that $\rho \leq \lambda$.

Let $(P_t)_{t\geq 0}$ be a Markov semigroup with generator L acting on a suitable class of functions on $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\big)$. A particular feature of the logarithmic Sobolev inequalities is the (equivalent, cf. [10]) hypercontractive property of the semigroup $(P_t)_{t\geq 0}$. Precisely, the logarithmic Sobolev inequality (1.7) is equivalent to saying that, whenever $p \geq 1 + e^{2\rho t}$, for all functions f in $L^p(\mu)$,

$$\|P_t(f)\|_2 \leq \|f\|_p. \tag{1.8}$$

For simplicity, we say below that a probability measure μ, in this context, is hypercontractive with constant ρ.

Finally, let us also recall that an Orlicz norm $\|\cdot\|_\phi$ is defined as follow : given a Young function ϕ, set

$$\|f\|_\phi = \inf\left\{c > 0\,;\, \int_{\mathbb{R}^n} \phi\left(\frac{|f|}{c}\right) d\mu \geq 1\right\}$$

the associated Orlicz norm of a measurable function $f : \mathbb{R}^n \to \mathbb{R}$. In the sequel, let $\phi : \mathbb{R}+ \to \mathbb{R}_+$ be convex such that $\phi(x) = \frac{x^2}{\log(e+x)}$ for $x \geq 1$ and $\phi(0) = 0$. To ease the notation, we set $\|\nabla f\|_\phi^2$ as a shorthand for $\sum_{i=1}^n \|\partial_i f\|_\phi^2$ where ∂_i, for any $i \in \{1, \ldots, n\}$, stands for the i-th partial derivative operator.

In this context, the following Theorem is our main result.

Theorem 1.3. *Within the preceding framework, assume that $(\mu_i)_{i=1,\ldots,n}$ are hypercontractive with constant ρ. Then, for any smooth $f \in \mathcal{F}_+$ we have*

$$\mathrm{Ent}_\mu(e^{-f}) \leq C_{\rho,\kappa} \|\nabla f\|_\phi^2 \mathbb{E}_\mu[e^{-f}] \tag{1.9}$$

where $C_{\rho,\kappa} = \frac{2e^{[1+(\kappa/\rho)]_+}}{\rho(1-e^{-1})}$. In particular, the following holds

$$\mu\left(f - \int_{\mathbb{R}^n} f d\mu \leq -t\right) \leq e^{-c_{\rho,\lambda} \frac{t^2}{2\|\nabla f\|_\phi^2}}, \quad t \geq 0 \tag{1.10}$$

where $c_{\rho,\lambda} > 0$ is a universal constant.

Remark. 1. In practice, it is classical to bound (cf. [8])$\|\nabla f\|_\phi^2$ by the following quantity :

$$\|\nabla f\|_\phi^2 = \sum_{i=1}^n \|\partial_i f\|_\phi^2 \leq C \sum_{i=1}^n \frac{\|\partial_i f\|_2^2}{1 + \log \|\partial_i f\|_1 / \|\partial_i f\|_2}$$

with $C > 0$ a numerical constant.

2. When, the standard Gaussian measure is considered

$$i.e. \quad V_i(x) = \frac{x^2}{2}, \quad i = 1, \ldots, n \quad \text{and} \quad x \in \mathbb{R}$$

the quantity $\|\nabla f\|_\phi^2$ can be replaced by the variance $\mathrm{Var}_{\gamma_n}(f)$ which is smaller. This is essentially Tien's result 1.2.

3. When the measures $(\mu_i)_{i=1,\ldots,n}$ are symmetric (e.g. the Gaussian measure γ_n), one can consider the set $\mathcal{H}_+$ instead of $\mathcal{F}_+$. Indeed, suppose $f \in \mathcal{H}_+$ is not increasing then it is enough to perform a change of variable and consider $f(-x) \in \mathcal{F}_+$.

We want to highlight the fact that only $\kappa \in \mathbb{R}$ is required here, it appears as a mild property shared by numerous potentials such as, for example, double-wells potentials on the line of the form $V(x) = ax^4 - bx^2$, $a, b > 0$. The stronger strict convexity assumption $V'' \geq \rho > 0$ (satisfied by the standard Gaussian measure γ_n) actually implies that μ satisfies a logarithmic Sobolev inequality, and thus hypercontractivity, with constant ρ (cf. [1]).

To better understand where the improvement lies in Theorem 1.3 Let us recall some facts : for a smooth function $f : \mathbb{R}^n \to \mathbb{R}$ it is known (cf. the introduction of [24] and references therein) that

$$\mathrm{Var}_\mu(f) \leq \|\nabla f\|_\phi^2 \leq \mathbb{E}_\mu[|\nabla f|^2] \leq \|f\|_{\mathrm{Lip}}^2$$

and each terms can be different from one another. For instance (cf. [7, 22, 5]), in a Gaussian case, if $f(x) = \max_{i=1,\dots,n} x_i$

$$\mathrm{Var}_{\gamma_n}(f) \simeq \frac{1}{\log n} \simeq \|\nabla f\|_\phi^2 \quad \text{and} \quad \mathbb{E}_{\gamma_n}[|\nabla f|^2] = \|f\|_{\mathrm{Lip}}^2 = 1.$$

If $f(x) = \mathrm{Med}(x_1, \dots, x_n)$, we have

$$\mathrm{Var}_{\gamma_n}(f) \simeq \frac{1}{n}, \quad \|\nabla f\|_\phi^2 \simeq \frac{1}{\log n} \quad \text{and} \quad \mathbb{E}_{\gamma_n}[|\nabla f|^2] = \|f\|_{\mathrm{Lip}}^2 = 1.$$

Let us mention that (1.1) has already been improved for convex functions, with $\mathbb{E}_{\gamma_n}[|\nabla f|^2]$ instead of $\|f\|_{\mathrm{Lip}}^2$ in [23, 4] (cf. [18] section 5.2). Thus, in Theorem 1.3, we obtain something slightly better. However, this bound is a priori larger (except for the Gaussian case) than the one involving $\mathrm{Var}_\mu(f)$ which would be the desired one for every μ.

Now, let us describe the organization of the article. Section 2 is concerned with semigroup facts and negative association. In section 4 we prove Theorem 1.3. Section 4 will describe some extensions. Finally, in section 5, we say a few words about Theorem 1.2.

In the sequel, we will always assume that the functions are sufficiently integrable with respect to μ in order that studied inequalities make sense and the commutation between integrals and derivatives are legit. Also, by convention, $C > 0$ is a numerical constant that may change at each occurence.

2 Tools

2.1 Semigroup properties

In this section, we present some tools needed to prove Theorem 1.3. In the context described in the introduction, let us collect some important properties of the semigroup $(P_t)_{t\geq 0}$. Again, for more details, the reader is referred to [1] (or [13], pages $306 - 328$, for a shorter exposition).

Proposition 2.1. *Within the preceding framework, the following holds*

- *For any smooth function $f : \mathbb{R}^n \to \mathbb{R}$, the semigroup $(P_t)_{t\geq 0}$ solves the heat equation associated to L.*

$$i.e. \quad \partial_t P_t(f) = LP_t(f) = P_t(Lf) \quad \text{for any} \quad t \geq 0. \tag{2.1}$$

- *$(P_t)_{t\geq 0}$ is ergodic : for any smooth function $f : \mathbb{R}^n \to \mathbb{R}$.*

$$\lim_{t\to+\infty} P_t(f) = \mathbb{E}_\mu[f]. \tag{2.2}$$

- *For any $i = 1, \dots, n$ and any smooth function $f : \mathbb{R}^n \to \mathbb{R}$, the uniform lower bound $V_i'' \geq -\kappa_i$, is equivalent to the following commutation property*
$$|\partial_i P_t(f)| \leq e^{\kappa t} P_t(|\partial_i f|) \quad \textit{for any} \quad t \geq 0 \tag{2.3}$$
where $\kappa = \max_{i=1,\dots,n} \kappa_i$.

Remark. 1. When $\mu = \gamma_n$ the commutation property (2.3) is exact (cf. [13, 1]) Namely, for any $i = 1, \dots, n$ and any smooth function $f : \mathbb{R}^n \to \mathbb{R}$

$$\partial_i P_t(f) = e^{-t} P_t(\partial_i f) \quad \text{for any} \quad t \geq 0. \tag{2.4}$$

This fact can also be checked on the representation formula (2.10) given in the sequel.

2. The uniform lower bound $V_i'' \geq -\kappa_i$, for any $i = 1, \dots, n$, can be seen as a strong gradient bound and is part of Bakry-Emery's theory (cf. [8], Corollary 3 and equation (22) therein in particular).

2.2 Semigroup representation of the Entropy

As it will be needed in the sequel, we state below some representation (cf. [1] section 5.5 or section 2.1 in [13]) of the entropy of a function along the semigroup $(P_t)_{t\geq 0}$.

$$\mathrm{Ent}_\mu(f^2) = \int_0^{+\infty} \int_{\mathbb{R}^n} \frac{|\nabla P_t(f^2)|}{P_t(f^2)} d\mu dt. \tag{2.5}$$

As it is exposed in [8], when μ satisfies a logarithmic Sobolev inequality there is no need to deal with large value of t in (2.5). Indeed, a logarithmic Sobolev inequality is equivalently stated as a exponential decay of the entropy along the semigroup. Namely,

$$\mathrm{Ent}_\mu\big(P_t(f)\big) \leq e^{-2t\rho} \mathrm{Ent}_\mu(f) \quad \text{for every} \quad t \geq 0 \tag{2.6}$$

and every positive function f in $L^1(\mu)$. Therefore, the combination of the preceding representation (2.5) by semigroup together with the exponential decay of the the entropy (cf. [1] page 244) along the semigroup we have, for any $T > 0$,

$$\mathrm{Ent}_\mu(f^2) \leq \frac{1}{1 - e^{-2\rho T}} \int_0^T \int_{\mathbb{R}^n} \frac{|\nabla P_t(f^2)|}{P_t(f^2)} d\mu dt. \tag{2.7}$$

In the sequel, we choose e.g. $T = \frac{1}{2\rho}$.

2.3 Semigroup and Harris inequality

As mentioned earlier, in order to investigate the lower tail, one has to use negative association inequality. Therefore we state below Harris's Lemma (cf.[6] page 43) and see how it can be combined with semigroups. Recall that monotonicity or convexity properties of a function $f : \mathbb{R}^n \to \mathbb{R}$ are understood coordinate-wise.

Proposition 2.2 (Harris's negative association inequality)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ two monotone functions with different monotonicity, then*

$$\mathbb{E}\big[f(X)g(X)\big] \leq \mathbb{E}\big[f(X)\big]\mathbb{E}\big[g(X)\big] \quad \textit{for} \quad X = (X_1, \dots, X_n) \tag{2.8}$$

with X_i independent random variables.

In the sequel, this proposition will also be used at the level of the semigroup. That is to say for the underlying heat kernel measure $p_t(x, dy)$ which is defined (cf. [1] page 12) as

$$P_t(f)(x) = \int_{\mathbb{R}^n} f(y)p_t(x, dy) \quad \text{with} \quad t \geq 0 \quad \text{and} \quad x \in \mathbb{R}^n.$$

This is the content of the following Lemma.

Lemma 2.1. *Let $t \geq 0$ and $x \in \mathbb{R}^n$ be fixed and consider f and g two monotone functions with different monotonicity, then*

$$P_t(fg)(x) \leq P_t(f)(x)P_t(g)(x).$$

The following Lemma explains, in our context, that the semigroup $(P_t)_{t\geq 0}$ preserves monotonicity properties of a function.

Lemma 2.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be monotone, then $x \mapsto P_t(f)(x)$, $t \geq 0$ shares the same monotonicity properties as the function f.*

Proof. As it is exposed in [15], in our setting, we have the following representation of $\nabla P_t f(x)$ for any $x \in \mathbb{R}^n$ and $t \geq 0$.

$$\nabla P_t f(x) = \mathbb{E}\big[\nabla f(X_t)e^{-\int_0^t V''(X_s)ds}\big|X_0 = x\big]. \tag{2.9}$$

Thus, $x \mapsto P_t f(x)$ shares the same monotonicity properties as f. □

Remark.

1. In the Gaussian setting, for quadratic potentials, this property is obvious thanks to Mehler's formula which gives an explicit representation of the Ornstein-Uhlenbeck semigroup :

$$P_t f(x) = \int_{\mathbb{R}^n} f(xe^{-t} + \sqrt{1 - e^{-2t}}y)d\gamma_n(y), \quad t \geq 0,\, x \in \mathbb{R}^n. \tag{2.10}$$

2. Representation as (2.9) is part of the so-called intertwinnings relation between a semigroup with some differential operator (cf.[3, 2] and references therein).

3. The fact that a semigroup preserves the monotonicity of a function has also been investigate in [16].

3 Study of the lower tail - Proof of Theorem 1.3

Recall that the measures $(\mu_i)_{i=1,\dots,n}$ are assumed to be hypercontractive with constant ρ. In this section we prove Theorem 1.3 thanks to Lemma 2.1 and 2.2.

Proof. Let $f \in \mathcal{F}_+$ be. Then, start with the representation formula (2.7)

$$\mathrm{Ent}_\mu(f^2) \leq \frac{1}{1-e^{-2\rho T}} \int_0^T \int_{\mathbb{R}^n} \frac{|\nabla P_t(f^2)|}{P_t(f^2)} d\mu dt$$

and apply it to $e^{-f/2}$. We obtain, thanks to the commutation properties (2.3),

$$\mathrm{Ent}_\mu(e^{-f}) \quad \leq \quad \frac{1}{1-e^{-2\rho T}} \int_0^T e^{2\kappa t} \sum_{i=1}^n \int_{\mathbb{R}^n} \frac{P_t^2(\partial_i f e^{-f})}{P_t(e^{-f})} d\mu dt.$$

Notice that, for any $i \in \{1,\dots,n\}$, $\partial_i f$ and e^{-f} are monotone with different monotonicity. Therefore, by Lemma 2.2, this is also the case for $P_t(\partial_i f)$ and $P_t(e^{-f})$. Then, by applying Lemma 2.1 twice, we get

$$\begin{aligned}
\mathrm{Ent}_\mu(e^{-\theta f}) \quad &\leq \quad \frac{1}{1-e^{-2\rho T}} \int_0^T e^{2\kappa t} \sum_{i=1}^n \int_{\mathbb{R}^n} P_t^2(\partial_i f) \frac{P_t^2(e^{-f})}{P_t(e^{-f})} d\mu dt \\
&\leq \quad \frac{1}{1-e^{-2\rho T}} \mathbb{E}_\mu[e^{-f}] \times \int_0^T e^{2\kappa t} \sum_{i=1}^n \int_{\mathbb{R}^n} P_t^2(\partial_i f) d\mu dt
\end{aligned}$$

where in the last upper bound we used that μ is the invariant measure of $(P_t)_{t\geq 0}$. Namely, $\mathbb{E}_\mu[P_t(h)] = \mathbb{E}_\mu[h]$ for any smooth functions $h : \mathbb{R}^n \to \mathbb{R}$.

Finally, in the preceding inequality, the last factor can be upper bounded by hypercontractive arguments. To this task, we follow the proof of Talagrand's inequalities exposed in [8] (pages $8-9$) in order to obtain

$$\int_0^T e^{2\kappa t} \sum_{i=1}^n \int_{\mathbb{R}^n} P_t^2(\partial_i f) d\mu \leq \frac{2e^{[1+(\kappa/\rho)]_+}}{\rho} \sum_{i=1}^n \|\partial_i f\|_\phi^2 \tag{3.1}$$

To sum up, we have proven

$$\mathrm{Ent}_\mu(e^{-f}) \leq C_{\rho,\kappa} \|\nabla f\|_\phi^2 \mathbb{E}_\mu[e^{-f}].$$

with $C_{\rho,\kappa} = \frac{2e^{[1+(\kappa/\rho)]_+}}{\rho(1-e^{-2\rho T})}$. The deviation inequality is classically obtained by applying the preceding inequality to $e^{-\theta f}$ with $\theta \geq 0$. □

Remark. Let us notice that the preceding scheme of proof can also be done at the level of the variance with the dynamical representation (used in [8])

$$\mathrm{Var}_\mu(f) = 2 \int_0^\infty \int_{\mathbb{R}^n} |\nabla P_t(f)|^2 d\mu dt.$$

Furthermore, when $\mu = \gamma_n$, one can choose $T = +\infty$. Then, thanks to the exact commutation property (2.4) between ∇ and $(P_t)_{t\geq 0}$ together with the preceding dynamical representation of the variance, we get

$$\mathrm{Ent}_{\gamma_n}(e^{-f}) \leq \mathbb{E}_{\gamma_n}[e^{-f}]\mathrm{Var}_{\gamma_n}(f).$$

4 Some extensions

Let us say a few words about some potential extensions. As it was emphasized in [8], one key features of the preceding methodology is the following. Given a Markov semigroup $(P_t)_{t\geq 0}$ with generator L and invariant measure μ. Assume that (L,μ) is hypercontractive and that the associated Dirichlet form $\mathcal{E}$ may be decomposed along directions Γ_i acting on functions on some state space E as

$$\mathcal{E}(f,f) = \sum_{i=1}^n \int_E \Gamma_i^2(f)d\mu$$

in a way that, for each $i = 1,\ldots,n$, Γ_i commutes to $(P_t)_{t\geq 0}$ in the sense that, for some constant $\kappa \in \mathbb{R}$, every $t \geq 0$ and f smooth enough,

$$\Gamma_i(P_t f) \leq e^{\kappa t} P_t\big(\Gamma_i(f)\big). \tag{4.1}$$

In the current article, this commutation property is obtained as a strong gradient bound from Bakry and Emery's Gamma 2 criterion and is stated in (2.3).

As a first example, one can investigate the standard exponential measure (or, more generally, gamma measure) $d\mu = e^{-\sum_{i=1}^n x_i} 1_{\{x_1\geq 0\}}\ldots 1_{\{x_n\geq 0\}} dx_1\ldots dx_n$ on $\mathbb{R}_+^n$ with the direction $\Gamma_i(f) = \sqrt{x_i}\partial_i$. According to [1, 20], the commutation properties (4.1) is satisfied with $\kappa = -1$. Now, observe that the operator Γ_i, $i = 1,\ldots,n$ preserves the key features of the function f. More precisely, assume $f \in \mathcal{F}_+$, then it is easy to check that $x_i \mapsto \Gamma_i(f) \in \mathcal{F}_+$. Besides the following identity, for any $\theta \in \mathbb{R}$, holds

$$\Gamma_i(e^{\theta f}) = \theta e^{\theta f}\Gamma_i(f).$$

Therefore, it is possible to apply Harris's negative association 2.2 in this situation.

Indeed, in this setting, it is then easy to extend slightly the result of the current article. Following the lines of the proof of our main result, we obtain

$$\begin{aligned}
\mathrm{Ent}_\mu(e^{-f}) &\leq \frac{1}{2-e^{-2\lambda T}}\int_0^T e^{2\kappa t}\sum_{i=1}^n \int_{\mathbb{R}^n} P_t^2\big(\Gamma_i(f)e^{-f}\big)d\mu dt \\
&\leq \mathbb{E}_\mu[e^{-f}] \times \left(\frac{1}{2-e^{-2\lambda T}}\int_0^T e^{2\kappa t}\sum_{i=1}^n \int_{\mathbb{R}^n} P_t^2(\Gamma_i(f))d\mu dt\right) \\
&\leq \mathbb{E}_\mu[e^{-f}] \times C_{\rho,\kappa}\|\Gamma(f)\|_\phi^2
\end{aligned}$$

where $\|\Gamma(f)\|_\phi^2$ is a shorthand for $\sum_{i=1}^n \|\Gamma_i(f)\|_\phi^2$. Notice also, according to [8], that hypercontractive estimates also yields the following upper bound

$$\|\Gamma(f)\|_\phi^2 \leq C \sum_{i=1}^n \frac{\|\Gamma_i f\|_2^2}{1+\log\left(\|\Gamma_i f\|_1/\|\Gamma_i f\|_2\right)}$$

with $C > 0$ a numerical constant. It is obvious that the same proof holds at the level of the variance. This bounds has to be compared with Proposition 2.20 (for $k = 1$) in [17].

As exposed in [8], non-product measures can also be investigated. For instance, if μ stands for the uniform probability measure on the sphere $\mathbb{S}^{n-1}$, one may consider the following fact

$$\mathcal{E}(f,f) = \int_{\mathbb{S}^{n-1}} f(-\Delta f)d\mu = \frac{1}{2}\sum_{i,j=1} \int_{\mathbb{S}^{n-1}} (D_{i,j}f)^2 d\mu$$

where the direction $D_{ij} = x_i\partial_j - x_j\partial_i$, $i,j = 1,\ldots,n$. The operators D_{ij} commute in an essential way to the spherical Laplacian $\Delta = \frac{1}{2}\sum_{i,j=1}^n D_{ij}^2$ so that (4.1) holds with $\kappa = 0$. However, the monotone properties needed in the proof (in order to apply Harris's negative association inequality) seems to be difficult to characterize.

5 Some remarks about theorem 1.2

We briefly want to highlight the fact that the arguments used in [9] can be easily expressed in terms of semigroup arguments. As we focus on the Gaussian case, notice that $(P_t)_{t\geq 0}$ stands for the Ornstein-Uhlenbeck semigroup. This reformulation gives shorter proof as we will show in the sequel. Unfortunately, the strategy presented below relies on exact commutation and can not be extended to the measure μ.

Following [9], introduce the operator T_g defined as follows

$$T_g(y) = \int_0^\infty \mathbb{E}_{\gamma_n}[\nabla f(X)\cdot\nabla P_t(g)(y)]dt \quad \text{with} \quad y \in \mathbb{R}^n$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is fixed, $g : \mathbb{R}^n \to \mathbb{R}$ is centered under γ_n and $\mathcal{L}(X) = \gamma_n$.

Lemma 5.1. *With the preceding notations, for any $\theta \geq 0$, we have*

$$\mathbb{E}_{\gamma_n}[e^{\theta f} g] = \theta \mathbb{E}_{\gamma_n}[e^{\theta f} T_g].$$

Proof. Since g is centered under γ_n and by ergodicity (2.2) of $(P_t)_{t\geq 0}$, we have

$$\mathbb{E}_{\gamma_n}[e^{\theta f}(g - \mathbb{E}_{\gamma_n}[g])] = \mathbb{E}_{\gamma_n}[e^{\theta f}(P_0(g) - P_\infty(g))].$$

Thus, by the fundamental Theorem of calculus, we have

$$\begin{aligned}
\mathbb{E}_{\gamma_n}[e^{\theta f} g] &= \mathbb{E}_{\gamma_n}\left[e^{\theta f}\left(-\int_0^\infty \frac{d}{dt} P_t(g) dt\right)\right] \\
&= \int_0^\infty \mathbb{E}_{\gamma_n}\big[e^{\theta f}\big(-LP_t(g)\big)\big] dt \qquad \text{by} \quad (2.1) \\
&= \int_0^\infty \mathbb{E}_{\gamma_n}[\nabla e^{\theta f} \cdot \nabla P_t(g)] dt \qquad \text{by} \quad (1.5) \\
&= \theta \mathbb{E}_{\gamma_n}[e^{\theta f} T_g].
\end{aligned}$$

□

Remark. The use of the operator T_g was the main idea of the article [9], we state it in a slightly different way which avoids a lot of calculus. For further purposes, observe that $\mathbb{E}_{\gamma_n}[T_g] = \mathrm{Cov}_{\gamma_n}(f, g)$. In particular,

$$\mathbb{E}_{\gamma_n}[T_f] = \mathrm{Var}_{\gamma_n}(f). \tag{5.1}$$

As in [9], the proof of Theorem 1.2 relies on Lemma 5.1. To stay as close as possible to the original proof, consider $g \in \mathcal{F}_+$ and set $f = -g$.

For notational convenience, set $m = \int_{\mathbb{R}^n} f d\gamma_n$. Then, for any $\theta \geq 0$ define $\psi(\theta) = \mathbb{E}_{\gamma_n}[e^{\theta(f-m)}]$. From Lemma 5.1, we have

$$\psi'(\theta) = \mathbb{E}_{\gamma_n}[e^{\theta(f-m)}(f-m)] = \theta \mathbb{E}_{\gamma_n}[e^{\theta(f-m)} T_f].$$

Besides,

$$\theta \mathbb{E}_{\gamma_n}[e^{\theta(f-m)} T_f] = \theta \mathbb{E}_{\gamma_n}\left[e^{\theta(f-m)}\big(T_f - \mathrm{Var}_{\gamma_n}(f)\big)\right] + \theta \mathrm{Var}_{\gamma_n}(f) \mathbb{E}_{\gamma_n}[e^{\theta(f-m)}].$$

To conclude, it is enough to show that $\mathbb{E}_{\gamma_n}\left[e^{\theta(f-m)}\big(T_f - \mathrm{Var}_{\gamma_n}(f)\big)\right] \leq 0$. Indeed, if it is the case we have

$$\psi'(\theta) \leq \theta \mathrm{Var}_{\gamma_n}(f) \psi(\theta).$$

Once integrated, this differential inequality yields

$$\mathbb{E}_{\gamma_n}[e^{\theta(f-m)}] \leq e^{\mathrm{Var}_{\gamma_n}(f)\frac{\theta^2}{2}} \quad \text{for all} \quad \theta \geq 0. \tag{5.2}$$

Finally, the deviation inequality from Theorem 1.2 is obtained by classical arguments : one has to use Chernoff inequality and optimize in $\theta \geq 0$.

Now, let us show that $\mathbb{E}_{\gamma_n}\left[e^{\theta(f-m)}\big(T_f - \mathrm{Var}_{\gamma_n}(f)\big)\right] \leq 0$. To this task, use Lemma 5.1 with $g = T_f - \mathrm{Var}_{\gamma_n}(f)$ (which, according to (5.1), is centered under γ_n) to get

$$\mathbb{E}_{\gamma_n}\left[e^{\theta(f-m)}\big(T_f - \mathrm{Var}_{\gamma_n}(f)\big)\right] = \theta \mathbb{E}_{\gamma_n}[e^{\theta(f-m)} T_{T_f}].$$

Now, let us investigate T_{T_f}, thanks to the exact commutation property (2.4), we have, for any $y \in \mathbb{R}^n$ (omitted here),

$$\begin{aligned} T_{T_f} &= \int_0^\infty e^{-t}\mathbb{E}_{\gamma_n}[\nabla f \cdot P_t(\nabla T_f)]dt \\ &= \int_0^\infty e^{-t}\sum_{i=1}^n \mathbb{E}_{\gamma_n}[\partial_i f P_t(\partial_i T_f)]dt. \end{aligned}$$

Besides, for any $i = 1, \dots, n$,

$$\begin{aligned} \partial_i T_f &= \partial_i \int_0^\infty \sum_{j=1}^n \mathbb{E}_{\gamma_n}[\partial_j f \partial_j P_t(f)]dt \\ &= \int_0^\infty \sum_{j=1}^n \mathbb{E}_{\gamma_n}[\partial_{ij}^2 f \partial_j P_t(f) + \partial_j f \partial_{ij}^2 P_t(f)]dt \\ &= \int_0^\infty \sum_{j=1}^n e^{-t}\mathbb{E}_{\gamma_n}[\partial_{ij}^2 f P_t(\partial_j f) + e^{-t}\partial_j f P_t(\partial_{ij}^2 f)]dt \\ &\leq 0 \end{aligned}$$

by hypothesis on f. Thus, $\mathbb{E}_{\gamma_n}\left[e^{\theta(f-m)}\big(T_f - \mathrm{Var}_{\gamma_n}(f)\big)\right] \leq 0$ and the proof is complete.

Remark. Theorem 1.2 implicitly uses a covariance identity (through the operator T_f). Similar identities have been used in [11] for infinitely divisible random vectors having finite exponential moments. In particular, sharp deviation inequalities were obtained. We wonder if Theorem 1.3 can be extend to this level of generality.

Acknowledgment : we are greatly indebted to P. Valettas for several comments and precious remarks. We warmly thank the anonymous referee for helpful comments in improving the exposition. We also thank P. Monmarché who indicated to us functions h and g which are non-convex elements of $\mathcal{F}_+$.

References

[1] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators.* Grundlehren der Mathematischen Wissenschaften, 348, 2014.

[2] M. Bonnefont and A. Joulin. Intertwining relations for one-dimensional diffusions and application to functional inequalities. *Potential Anal.*, 41(4):1005–1031, 2014.

[3] M. Bonnefont, A. Joulin, and Y. Ma. A note on spectral gap and weighted Poincaré inequalities for some one-dimensional diffusions. *ESAIM Probab. Stat.*, 20:18–29, 2016.

[4] S. Bobkov. Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Prob*, 27(4) : 1903–1921, 1999.

[5] S. Boucheron and M. Thomas. Concentration inequalities for order statistics. *Electronic Communications in Probability*, 2012.

[6] T. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities : a nonasymptotic theory of independance.* Oxford University Press, 2013.

[7] S. Chatterjee. *Superconcentration and related topics.* Springer, 2014.

[8] D. Cordero-Erausquin and M. Ledoux. Hypercontractive Measures, Talagrand's inequality, and Influences. *Geometric aspects of functional analysis, 169-189, Lectures Notes in Math 2050*, 2012.

[9] Nguyen Tien Dung. An improved bound for the gaussian concentration inequality. *Preprint : arXiv:1904.03674v1*, 2019.

[10] L. Gross. Logarithmic sobolev inequalities. *Amer. J. Math.*, 1975.

[11] C. Houdré. Remarks on deviation inequalities for functions of infinitely divisible random vecteors. *The Annals of Probability*, 30(3):1223–1237, 2002.

[12] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes.* Springer Series in Statistics., 1983.

[13] M. Ledoux. The geometry of Markov diffusions operators. *Ann. Fac. Sci. Toulouse Math. (6) 9, no.2, 305-366*, 2000.

[14] M. Ledoux. *The concentration of measure phenomenon.* Mathematical Surveys and Monographs, 89, 2001.

[15] F. Malrieu and D. Talay. Concentration inequalities for Euler schemes. *Monte Carlo and quasi-Monte Carlo methods 2004*, 2006.

[16] G. Pagès. Functional co-monotony of processes with applications to peacocks and barrier options. *Séminaire de Probabilités XLV*, 2078, 2013.

[17] G. Paouris and P. Valettas. A gaussian small deviation inequality for convex functions. *Annals of Probability*, 46(3):1141–1454, 2018.

[18] G. Paouris and P. Valettas. Variance estimates and almost euclidean structure. *Advances in Geometry*, To appear.

[19] G. Paouris, P. Valettas, and J. Zinn. Random version of Dvoretzky's Theorem in l_p^n. *Stochastic Processes and their Applications*, 127(10):3187–3227, 2017.

[20] K. Tanguy. Some superconcentration inequalities for extrema of stationary gaussian processes. *Statistics and Probability Letters*, 106:239–246, November 2015.

[21] K. Tanguy. *Quelques inégalités de superconcentration : théorie et applications (in french).* PhD thesis, Institute of Mathematics of Toulouse, 2017.

[22] K. Tanguy. Non asymptotic variance bounds and deviation inequalities by optimal transport. *Electronic Journal of Probability*, 24(12):1–18, 2019.

[23] P. M. Samson. Concentration inequalities for convex functions on product spaces. *Stochastic inequalities and applications*, 56 : 33–52, 2003.

[24] P. Valettas. On the tightness of Gaussian concentration for convex functions. *Journal d'Analyse Mathématique*, To appear.