Improved deviations inequalities for log-concave measures

Kevin Tanguy
University of Angers, France

Note of August 30, 2018

Abstract

This short note deals with elementary semigroup arguments together with Harris’s negative association inequality in order to obtain improved deviation inequalities (left tail) for log-concave measure under some monotony and convexity assumption.

1 Introduction

As an introduction we recall some facts about Gaussian concentration of measure (cf. [10]) 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 produces, for \( f : \mathbb{R}^n \to \mathbb{R} \) a Lipschitz function with Lipschitz constant \( \| f \|_{Lip} \),

\[
\gamma_n \left( |f - \int_{\mathbb{R}^n} f d\gamma_n| \geq t \right) \leq 2e^{-\frac{t^2}{2\| f \|_{Lip}^2}}, \quad t \geq 0, \tag{1.1}
\]

with \( \gamma_n \) the standard Gaussian measure on \( \mathbb{R}^n \). Another instance of concentration of measure is the Poincaré’s inequality (which yields exponential concentration) satisfied by \( \gamma_n \). Namely, for \( f \in L^2(\gamma_n) \) smooth enough, it says

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

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 case. As instance, consider the 1-Lipschitz function on \( \mathbb{R}^n \) \( f(x) = \max_{i=1,...,n} x_i \). At the level of the variance, (1.2) gives

\[
\text{Var}(M_n) \leq 1,
\]

with \( M_n = \max_{i=1,...,n} X_i \) where \( (X_1, \ldots, X_n) \) stands for a standard Gaussian random vector in \( \mathbb{R}^n \), whereas it has been proven that \( \text{Var}(M_n) \leq C/\log n \) with \( C > 0 \) a numerical constant. At an exponential level, (1.1) is not satisfying
either. Indeed, it is well known in Extreme theory (cf. [9]) that $M_n$ can be renormalized by some numerical constant, $a_n = \sqrt{2\log n}$ and $b_n = a_n - \frac{\log 4 + \log \log n}{2a_n}$, $n \geq 1$, such that

$$a_n (M_n - b_n) \to \Lambda_0$$

in law, as $n \to \infty$, where $\Lambda_0$ corresponds to the Gumbel distribution:

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

Then, it is clear that the asymptotics of $\Lambda_0$ 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) leads to sub-optimal results for the convex function $x \mapsto \max_{i=1,\ldots,n} x_i$. When such phenomenon happens it is referred as Superconcentration phenomenon (cf. [7]). This kind of phenomenon happens for different functionals of Gaussian random variables (and also, as we will see, for other laws of probability) and as been studied in [5, 15, 16, 17, 14]. . . .

Recently, convexity assumption has 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, this two authors improved the left tail of convex function thanks to Ehrard’s inequality in [13]. 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

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

(1.3)

where $c > 0$ is a universal constant.

**Remark.** Of course, the improvements stays in the fact that $\text{Var}_{\gamma_n}(f) \leq \text{Lip}(f)^2$ as we have just seen on the basic example of the maximum of $n$ independent standard Gaussian random variables. Ehrard’s inequality has also been used by Valettas in [18] where he proved that (1.1) is tight if the convex function $f$ is not superconcentrated.

The purpose of this note is to show how semigroup’s arguments can provide elementary proofs of results similar as Theorem 1.1. Although our proof seems to be simpler, we had to assume $f$ to be monotone. Let us precise that the monotony of a function $f : \mathbb{R}^n \to \mathbb{R}$ is understood coordinate wise (while the others are fixed) throughout all the present article. This assumption can not permit us to consider $L^p$ norms of standard Gaussian vectors as Paouris and Valettas did in [13]. On the bright side, semigroup’s arguments can be used for a larger class of measure. Indeed, log-concave measure $d\mu(x) = e^{-V(x)}dx$ on $\mathbb{R}^n$ under a lower bound condition ($V''(x) \geq -\kappa$, $\kappa \in \mathbb{R}$) on the potential can be considered (cf. section four), whereas the use of Ehrard’s inequality is limited to Gaussian measures.

Let us briefly describe our main result, let $d\mu_i(x) = e^{-V_i(x)}dx$, $i = 1, \ldots, n$ on $\mathbb{R}$ be hypercontractive measures (cf. section two for the terminology) with
constant $\rho_i > 0$. Then, set $d\mu = d\mu_1 \otimes \ldots \otimes d\mu_n$ on $\mathbb{R}^n$. We obtained the following

**Theorem 1.2.** Under the preceding framework, assume that $V''_i \geq -\kappa$, $i = 1, \ldots, n$ for some $\kappa \in \mathbb{R}$. Then, for any smooth non-decreasing and convex function $f : \mathbb{R}^n \to \mathbb{R}$ and $\theta \geq 0$ we have

1. $\text{Var}_\mu(e^{-\theta f/2}) \leq C\theta^2 \|\nabla f\|_\phi E_\mu[e^{-\theta f}]$,
2. $\text{Ent}_\mu(e^{-\theta f/2}) \leq C\theta^2 \|\nabla f\|_\phi E_\mu[e^{-\theta f}]$.

where $\|\nabla f\|_\phi$ is a shorthand for $\sum_{i=1}^n \|\partial_i f\|_\phi^2$ and $\|\cdot\|_\phi$ stands for the Orlicz norm induced by the Young function $\phi$ (see below).

In particular, the following holds

$$\mu \left( f - \int_{\mathbb{R}^n} f \, d\mu \leq -t \right) \leq e^{-c_{\rho,\lambda} \max\left(t \sqrt{\|\nabla f\|_\phi}, t^2/2\right)} \|\nabla f\|_\phi^2, \quad t \geq 0 \quad (1.4)$$

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

**Remark.**

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

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

2. When, the Gaussian measure measure is considered (with the quadratic potential $V_i(x) = x_i^2/2$, $i = 1, \ldots, n$) the quantity $\|\nabla f\|_\phi$ can be replaced by the variance $\text{Var}_{\nu_\phi}(f)$ which is smaller.

3. Of course, similar results holds for the right tail if one consider non-increasing concave function instead.

The article is organized as follow, in the next section we recall some facts about the interplay between functional inequalities, concentration of measure and semigroup interpolation’s arguments. We also recall Harris’s negative association inequality that will be one of the major argument in the proof. In section three, for pedagogical reasons, we focus on the Gaussian measure and exposed the scheme of proof of our method and provide an application of our main result for order statistics. In section four, we show how to easily prove Theorem 1.2 along the same lines as the Gaussian case. Finally, in section five, we discuss some potential extension in a more general setting.

2 Tools

For more details on semigroups, functional inequalities and concentration of measures, the reader is referred to [1, 2, 10, 6].
In the sequel, we will mainly deal with continuous probability measures of the form \( d\mu(x) = e^{-V(x)}dx \) on Borel sets of \( \mathbb{R}^n \) where \( V \) is some (smooth) potential. Such measures will be seen as invariant and reversible measures of the associated diffusion operators \( L = \Delta - \nabla V \cdot \nabla \).

Let \( V: \mathbb{R}^n \to \mathbb{R} \) be such that \( \int_{\mathbb{R}^n} e^{-V(x)}dx = 1 \), under mild smoothness and growth conditions, the second order operator \( L = \Delta - \nabla V \cdot \nabla \) admits \( d\mu(x) = e^{-V(x)}dx \) as symmetric and invariant probability measure. The operator \( L \) generates the Markov semigroup of operators \( (P_t)_{t \geq 0} \) and defines by integration by parts the Dirichlet form

\[
\mathcal{E}(f, g) = \int_{\mathbb{R}^n} f(-Lg)d\mu = \int_{\mathbb{R}^n} \nabla f \cdot \nabla gd\mu
\]

for smooth functions \( f, g \) on \( \mathbb{R}^n \).

Given such a couple \( (L, \mu) \), 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

\[
\lambda \text{Var}_\mu(f) \leq \mathcal{E}(f, f).
\] (2.1)

Similarly, it satisfies a logarithmic Sobolev inequality if there exists a constant \( \rho > 0 \) such that for all functions \( f \) of the Dirichlet domain,

\[
\rho \text{Ent}_\mu(f^2) \leq 2\mathcal{E}(f, f).
\] (2.2)

One speaks of the spectral gap constant (of \( (L, \mu) \)) as the best \( \lambda > 0 \) for which (2.1) holds, and of the logarithmic Sobolev constant (of \( (L, \mu) \)) as the best \( \rho > 0 \) for which (2.2) holds. We still use \( \lambda \) and \( \rho \) to design these constants. It is classical that \( \rho \leq \lambda \).

Both spectral gap and logarithmic Sobolev inequalities translate equivalently on the associated semigroup \( (P_t)_{t \geq 0} \). Namely, the spectral inequalities (2.1) is equivalent to saying that

\[
\|P_tf\|_2 \leq e^{-\lambda t}\|f\|_2 \quad \text{for every} \quad t \geq 0
\] (2.3)

for every mean zero function \( f \) in \( L^2(\mu) \). Similarly, the logarithmic Sobolev inequalities is equivalently stated as a exponential decay of the entropy along the semigroup. Namely,

\[
\text{Ent}_\mu(P_tf) \leq e^{-\rho t}\text{Ent}_\mu(f) \quad \text{for every} \quad t \geq 0
\] (2.4)

and every positive function \( f \) in \( L^1(\mu) \).

For further purposes, notice that (2.3) and (2.4) imply the following inequalities, for any \( T \geq 0 \),

\[
\text{Var}_\mu(f) \leq \frac{1}{1 - e^{-\lambda T}} \left[ \|f\|_2^2 - \|P_Tf\|_2^2 \right]
\] (2.5)
and
\[
\text{Ent}_\mu(f) \leq \frac{1}{1 - e^{-2\rho t}} \left[ \|f \log f\|_1 - \|P_t f\|_1 \log \|P_t f\|_1 \right] \tag{2.6}
\]

As another feature of the logarithmic Sobolev inequalities is the (equivalent) hypercontractive property if the semigroup. Precisely, the logarithmic Sobolev inequality (2.2) is equivalent to saying that, whenever \( p \geq 1 + e^{2\rho t} \), for all functions \( f \) in \( L^p(\mu) \),
\[
\|P_t f\|_2 \leq \|f\|_p \tag{2.7}
\]

For simplicity, we say below that a probability measure \( \mu \), in this context, is hypercontractive with constant \( \rho \).

### 2.1 Functional inequalities and concentration of measure

Functional inequalities, such as Poincaré’s inequality or logarithmic Sobolev inequalities, can be used to estimate from above the Laplace transform of a measure. Then, such bounds easily yield deviation inequalities thanks to Chernoff’s argument. In each case, the upper bounds on the Laplace transform are obtained by differential inequality which are the result of the substitution of \( f \) by \( e^{\theta f / 2} \), for some parameter \( \theta \in \mathbb{R} \), in (2.1) or (2.2).

At the level of the logarithmic Sobolev inequality it is known as Herbst’s argument and can be stated as follow :

**Lemma 2.1.** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a function smooth enough and \( \theta \geq 0 \). Assume that, the following holds
\[
\text{Ent}_\mu(e^{-\theta f}) \leq \frac{\nu \theta^2}{2} \mathbb{E}_\mu[e^{-\theta f}] \tag{2.8}
\]
for some \( \nu > 0 \), then the following deviation inequality holds
\[
\mu \left( f - \int_{\mathbb{R}^n} f d\mu \leq -t \right) \leq e^{-\frac{t^2}{2\nu}}, \quad t \geq 0
\]

Similarly, at the level of Poincaré’s inequality, Aida and Stroock proved the following

**Lemma 2.2.** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a function smooth enough and \( \theta \geq 0 \). Assume that, the following holds
\[
\text{Var}_\mu(e^{-\theta f}) \leq \frac{\nu \theta^2}{4} \mathbb{E}_\mu[e^{-\theta f}] \tag{2.9}
\]
for some \( \nu > 0 \), then the following deviation inequality holds
\[
\mu \left( f - \int_{\mathbb{R}^n} f d\mu \leq -t \right) \leq 3e^{-t\sqrt{\nu}}, \quad t \geq 0
\]

The key step to prove our main results will be to establish (2.8) or (2.9) by semigroups arguments under some monotony and convexity assumptions.
2.2 Semigroup interpolation

As it will be needed in the sequel, we state below some representation of the variance or the entropy of a function along the semigroup \((P_t)_{t \geq 0}\).

\[
\text{Var}_\mu(f) = 2 \int_0^\infty \int_{\mathbb{R}^n} |\nabla P_t f|^2 d\mu dt \quad (2.10)
\]

and

\[
\text{Ent}_\mu(f^2) = \int_0^\infty \int_{\mathbb{R}^n} \frac{|\nabla P_t f|^2}{P_t f^2} d\mu dt \quad (2.11)
\]

As it is exposed in [8], when \(\mu\) satisfies a Poincaré inequality (respectively a logarithmic Sobolev inequalities) there is no need to deal with large value of \(t\) in (2.10) (respectively (2.11)). Indeed, the combination of the preceding representation by semigroup together with the exponential decay of the variance (respectively the entropy) along the semigroup (2.5) (respectively (2.6)) we have, for any \(T \geq 0\),

\[
\text{Var}_\mu(f) \leq \frac{2}{1 - e^{-2\lambda T}} \int_0^T \int_{\mathbb{R}^n} |\nabla P_t f|^2 d\mu dt \quad (2.12)
\]

and

\[
\text{Ent}_\mu(f^2) \leq \frac{1}{1 - e^{-2\rho T}} \int_0^T \int_{\mathbb{R}^n} \frac{|\nabla P_t f|^2}{P_t f^2} d\mu dt \quad (2.13)
\]

2.3 Negative association inequalities and semigroup

As our main results will rest on negative association inequalities, we state below Harris’s Lemma and see how it can be used on semigroups. Recall that monotony or convexity properties of a function \(f : \mathbb{R}^n \rightarrow \mathbb{R}\) are understood coordinate-wise.

**Proposition 2.1 (Harris’s negative association inequality).** Let \(f : \mathbb{R}^n \rightarrow \mathbb{R}\) non-increasing and \(g : \mathbb{R}^n \rightarrow \mathbb{R}\) non-decreasing, then

\[
\mathbb{E}[f(X)g(X)] \leq \mathbb{E}[f(X)]\mathbb{E}[g(X)] \quad \text{for} \quad X = (X_1, \ldots, X_n) \quad (2.14)
\]

with \(X_i\) independent random variables.

**Proof.** For the sake of completeness, we briefly present the argument in dimension one. The general case is handled by induction and conditional expectation. Let \(Y\) be a independent copy of the random vector \(X\). By assumption on \(f\) and \(g\), observe that the following holds

\[
(f(X) - f(Y)) \times (g(X) - g(Y)) \leq 0.
\]

It is enough to integrate the preceding inequality with respect to \(\mu\) to conclude.
Remark. In the sequel, this proposition will be freely used for the invariant measure $\mu$ and also at the level of the semigroup: let $t \geq 0$ and $x \in \mathbb{R}^n$ be fixed and consider $f$ and $g$ two functions satisfying the hypothesis of Harris’s negative association inequality, then

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

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

**Lemma 2.3.** 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 [11], under our log-concave setting, we have the following representation of $\nabla P_t f(x)$ for any $x \in \mathbb{R}^n$ and $t \geq 0$.

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

(2.15)

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

**Remark.** 1. In the Gaussian setting (when $V(x) = \frac{|x|^2}{2}$) this property is obvious thanks to the Mehler’s formula which gives an explicit representation formula of the Ornstein-Uhlenbeck semigroup:

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

(2.16)

2. Representation as (2.15) are part of the so-called intertwinnings relation between a semigroup with some differential operator (cf.[4, 3] and references therein).

3. The fact that a semigroup preserves some monotone properties of a function has been investigated in [12].

### 3 Gaussian measure

In this section, we will focus on the Gaussian measure together with the Ornstein-Uhlenbeck semigroup. We will see that negative association inequalities together with the representation formula yields improved deviation inequalities for increasing and convex functions $f : \mathbb{R}^n \to \mathbb{R}$. One crucial feature that will be needed in the sequel is the commutation property between $(P_t)_{t \geq 0}$ and the partial derivatives operator $\partial_i$, $i = 1, \ldots, n$. As it can be easily seen on (2.16), for $f : \mathbb{R}^n \to \mathbb{R}$ smooth enough we have

$$\partial_i P_t(f) = e^{-t} P_t(\partial_i f), \quad i = 1, \ldots, n, \quad t \geq 0$$

(3.1)
3.1 Proof of the main results

This subsection gives a proof of our main result in a Gaussian setting. In particular, it exposes the scheme of proof that will be used for log-concave measure. As we have mentioned it before, the arguments involving Harris’s negative association inequality 2.1 will be freely used in the sequel at the level of the invariant measure or the semigroup. Furthermore, although we will not state it, it is straightforward to extend the results below to Gaussian measure $\mu$ on $\mathbb{R}^n$ with covariance structure $M$ such that $M_{ij} \geq 0$, $\forall i, j \in \{1, \ldots, n\}$.

**Theorem 3.1.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a non-decreasing and convex function. Then, for any $\theta \geq 0$, we have

$$\text{Var}_{\gamma_n}(e^{-\frac{\theta}{2}f}) \leq \frac{\theta^2}{4} \text{Var}_{\gamma_n}(f) \mathbb{E}_{\gamma_n}[e^{-\theta f}]$$

In particular, the following deviation inequalities holds

$$\gamma_n \left(f - \int_{\mathbb{R}^n} f d\gamma_n \leq -t\right) \leq 3e^{-t\sqrt{\text{Var}_{\gamma_n}(f)}}, \ t \geq 0$$

**Proof.** Start with the representation formula for the variance along the semigroup

$$\text{Var}_{\gamma_n}(f) = 2 \int_0^\infty \int_{\mathbb{R}^n} |\nabla P_t f|^2 d\gamma_n dt = 2 \int_0^\infty \sum_{i=1}^n \int_{\mathbb{R}^n} [\partial_i P_t f]^2 d\gamma_n dt.$$

Thanks to the exact commutation (3.1), it is equivalent as

$$\text{Var}_{\gamma_n}(f) = 2 \int_0^\infty e^{-2t} \sum_{i=1}^n \int_{\mathbb{R}^n} P_t^2(\partial_i f) d\gamma_n dt.$$

Now, substitute $f$ by $e^{-\frac{\theta}{2}f}$ with $\theta \geq 0$. We obtain

$$\text{Var}_{\gamma_n}(e^{-\frac{\theta}{2}f}) = \frac{\theta^2}{2} \int_0^\infty e^{-2t} \sum_{i=1}^n \int_{\mathbb{R}^n} P_t^2(\partial_i f e^{-\frac{\theta}{2}f}) d\gamma_n dt$$

$$\leq \frac{\theta^2}{2} \int_0^\infty e^{-2t} \sum_{i=1}^n \int_{\mathbb{R}^n} P_t^2(\partial_i f) P_t^2(e^{-\frac{\theta}{2}f}) d\gamma_n dt$$

where we used Harris’s negative association inequalities 2.1 to get the upper bound. Indeed, $f$ is separately convex and non-decreasing thus $\partial_i f$ is a positive and non decreasing function for any $i = 1, \ldots, n$.

Then, as $P_t$ preserves the monotony and the sign of any function (for any $t \geq 0$). In particular, by assumption on $f$, the function $x \mapsto P_t^2 \partial_i f(x)$ is non-decreasing whereas $x \mapsto P_t^2(e^{-\theta f/2})(x)$ is non-increasing. Therefore, by Harris’s negative association inequality we obtain
Var_{\gamma_n}(e^{-\frac{\theta}{2}f}) \leq \frac{\theta^2}{2} \int_0^\infty e^{-2t} \sum_{i=1}^n \left( \int_{\mathbb{R}^n} P_t^2(\partial_i f) d\gamma_n \right) \left( \int_{\mathbb{R}^n} P_t^2(e^{-\frac{\theta}{2}f}) d\gamma_n \right) dt
\leq \frac{\theta^2}{2} \mathbb{E}_{\gamma_n}[e^{-\theta f}] \times \int_0^\infty e^{-2t} \sum_{i=1}^n \int_{\mathbb{R}^n} P_t^2(\partial_i f) d\gamma_n dt.

Observe that the last inequality is obtained thanks to Jensen’s inequality and the fact that \(\gamma_n\) is the invariant measure of \((P_t)_{t \geq 0}\). To conclude, use again the exact commutation property (3.1) and the representation formula of the variance (2.10) to get

\[\var{\gamma_n}(e^{-\frac{\theta}{2}f}) \leq \frac{\theta^2}{4} \var{\gamma_n}(f) \mathbb{E}_{\gamma_n}[e^{-\theta f}]\]

The deviation inequality is obtained thanks to Lemma 2.2.

The exact same arguments can also be developed at the level of the Entropy.

**Theorem 3.2.** Let \(f : \mathbb{R}^n \to \mathbb{R}\) be a non-decreasing and convex function. Then, for any \(\theta \geq 0\) we have

\[\operatorname{Ent}_{\gamma_n}(e^{-\frac{\theta}{2}f}) \leq \frac{\theta^2}{4} \var{\gamma_n}(f) \mathbb{E}_{\gamma_n}[e^{-\theta f}]\]

In particular, the following deviation inequalities holds

\[\gamma_n\left(f - \int_{\mathbb{R}^n} f d\gamma_n \leq -t\right) \leq 3e^{-\frac{\lambda^2}{4}\var{\gamma_n}[e^{-\lambda f}]}, \quad t \geq 0\]

**Proof.** Start with the representation formula (2.11)

\[\operatorname{Ent}_{\gamma_n}(f) = \int_0^\infty \int_{\mathbb{R}^n} \frac{|\nabla P_t f|^2}{P_t f} d\gamma_n dt\]

and apply it to \(e^{-\theta f}, \theta \geq 0\). With the exact same argument, we obtain

\[\operatorname{Ent}_{\gamma_n}(e^{-\theta f}) = \theta^2 \int_0^\infty e^{-2t} \sum_{i=1}^n \left( \int_{\mathbb{R}^n} P_t^2(\partial_i f e^{-\theta f}) \right) P_t^2(e^{-\theta f}) d\gamma_n dt \leq \theta^2 \int_0^\infty e^{-2t} \sum_{i=1}^n \int_{\mathbb{R}^n} P_t^2(\partial_i f) \frac{P_t^2(e^{-\theta f})}{P_t(e^{-\theta f})} d\gamma_n dt \leq \theta^2 \mathbb{E}_{\gamma_n}[e^{-\theta f}] \times \int_0^\infty e^{-2t} \sum_{i=1}^n \int_{\mathbb{R}^n} P_t^2(\partial_i f) d\gamma_n dt\]

Finally, we have proven

\[\operatorname{Ent}_{\gamma_n}(e^{-\lambda f}) \leq \frac{\lambda^2}{2} \var{\gamma_n}(f) \mathbb{E}_{\gamma_n}[e^{-\lambda f}]\]

We end the proof with Lemma 2.1.
The preceding method can also be directly done at the level of the Laplace transform.

**Theorem 3.3.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function. Assume that $f$ is non-decreasing and separately convex. Then, for any $\theta \geq 0$ and every $t \geq 0$, the following holds

$$
\mathbb{E}_{\gamma_n}[e^{-\theta f}] \leq e^{\frac{\theta^2}{2} \text{Var}_{\gamma_n}(f) - \theta \mathbb{E}_{\gamma_n}(f)}
$$

of Theorem 3.3. Assume that $f$ is also centered to ease the notation. Then, for any $\theta \geq 0$, set

$$
H(t) = \mathbb{E}_{\gamma_n}[e^{-\theta P_t(f)}] \quad \text{with} \quad t \geq 0.
$$

By ergodicity of $(P_t)_{t \geq 0}$, $H(t) = 1 - \int_{t}^{+\infty} H'(s)ds$. Then, by integration by parts, we obtain for any $s \geq 0$,

$$
H'(s) = \theta \int_{\mathbb{R}^n} \nabla P_s f \cdot \nabla e^{-\theta P_s f} d\gamma_n.
$$

which can be rewritten as

$$
H'(s) = -\theta^2 \int_{\mathbb{R}^n} |\nabla P_s f|^2 e^{-\theta P_s f} d\gamma_n.
$$

So, for any $t \geq 0$, $H(t) = 1 + \theta^2 \int_{t}^{+\infty} \int_{\mathbb{R}^n} |\nabla P_s f|^2 e^{-\theta P_s f} d\gamma_n ds$. Then, observe by the commutation property between $\nabla$ and $(P_s)_{s \geq 0}$ that

$$
\int_{\mathbb{R}^n} |\nabla P_s f|^2 e^{-\theta P_s f} d\gamma_n ds = \sum_{i=1}^{n} e^{-2s} \int_{\mathbb{R}^n} P_s^2(\partial_i f)e^{-\theta P_s f} d\gamma_n.
$$

Again, by Harris’s negative association inequality 2.1,

$$
\int_{\mathbb{R}^n} P_s^2(\partial_i f)e^{-\theta P_s f} d\gamma_n \leq \left( \int_{\mathbb{R}^n} P_s^2(\partial_i f) d\gamma_n \right) \left( \int_{\mathbb{R}^n} e^{-\theta P_s f} d\gamma_n \right), \quad \forall s \geq 0.
$$

To sum up, we have proven that, for any $t \geq 0$,

$$
H(t) \leq 1 + \theta^2 \int_{t}^{+\infty} e^{-2s} G(s) H(s) ds
$$

where $G(s) = \sum_{i=1}^{n} \int_{\mathbb{R}^n} P_s^2(\partial_i f) d\gamma_n$. Apply now Gronwall’s Lemma to get

$$
H(t) \leq e^{\theta^2 \int_{t}^{+\infty} e^{-2s} G(s) ds} \quad \text{for any} \quad t \geq 0 \quad \text{and} \quad \theta \geq 0.
$$

In particular,

$$
H(0) \leq e^{\frac{\theta^2}{2} \text{Var}_{\gamma_n}(f)}, \quad \theta \geq 0
$$

which is the desired conclusion. 

\[\square\]
3.2 Application to Order statistics

It is possible to apply our main results to order statistics. Consider a vector \( X = (X_1, \ldots, X_n) \) in \( \mathbb{R}^n \) with i.i.d. component \( X_i, i = 1, \ldots, n \), with law \( \mu \) and denote by \( X_{(1)} \leq \ldots \leq X_{(n)} \) the associated order statistics. Notice that such functional are non-decreasing and convex as it can be seen on the following representation:

\[
X_{(k)} = \sum_{i=1}^{n} X_i 1\{X_i = i\} \quad \text{for any} \quad k \in \{1, \ldots, n\}
\]

Observe that, up to regularization argument, the functions \( x \mapsto \partial_i x_{(k)} = 1\{x_{(k)} = i\}, i = 1, \ldots, n \) are also non-decreasing for any \( k \in \{1, \ldots, n\} \). Among these order statistics, we will focus on the Median and the Maximum. A direct application of Theorems 3.1 and 3.2 yields

\[
\gamma_n(M_n - \mathbb{E}[M_n] \leq -t) \leq e^{-\max(t\sqrt{\text{Var}(M_n)}, \frac{t^2}{2\gamma_n^2(\text{Var}(M_n))})} \quad t \geq 0 \tag{3.2}
\]

where \( M_n = \max_{i=1,\ldots,n} X_i \). Similarly, it also provides the following

\[
\gamma_n(\text{Med} - \mathbb{E}[\text{Med}] \leq -t) \leq e^{-\max(t\sqrt{\text{Var}(\text{Med})}, \frac{t^2}{2\gamma_n^2(\text{Var}(\text{Med}))})} \quad t \geq 0 \tag{3.3}
\]

where Med stands for the median of the sample \( X_1, \ldots, X_n \).

On one hand, as mentioned in the introduction, (3.2) improved upon known results but is not sharp since it is expected to have a decay in double exponential \( t \mapsto e^{-e^t} \). On the other hand, according to [5], the asymptotics of the left tail for the Median is sharp.

Remark: Let us mentioned that similar arguments has been already used (to reach deviation inequalities) for order statistics in [16]. In this article, the right tail is investigate and the semigroup approach is replaced by optimal transport arguments.

4 Hypercontractive, Log-concave measure

It is possible to easily extend the preceding scheme of proof to log-concave measures and demonstrate Theorem 1.2. To this task, let 
\[ d\mu_i(x) = e^{-V_i(x)} dx, i = 1,\ldots,n \] on \( \mathbb{R} \) be hypercontractive with constant \( \rho_i > 0 \). Then, set \( d\mu = d\mu_1 \otimes \ldots \otimes d\mu_n \) on \( \mathbb{R}^n \).

Let us recall the definition of the Orlicz norm \( \| \cdot \|_\phi \) and the statement of Theorem 1.2. Given a Young function \( \phi \), it is customary [1] to denote by

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

the associated Orlicz norm of a measurable function \( f : \mathbb{R}^n \to \mathbb{R} \).
Theorem. Under the preceding framework, assume that \( V''_i \geq -\kappa, \kappa \in \mathbb{R}, i = 1, \ldots, n \). Then, for any smooth non-decreasing and convex function \( f : \mathbb{R}^n \to \mathbb{R} \) and \( \theta \geq 0 \) we have

1. \( \text{Var}_\mu(e^{-\theta f/2}) \leq C\theta^2 \| \nabla f \|_\phi \mathbb{E}_\mu[e^{-\theta f}] \),
2. \( \text{Ent}_\mu(e^{-\theta f/2}) \leq C\theta^2 \| \nabla f \|_\phi \mathbb{E}_\mu[e^{-\theta f}] \),

where \( \| \nabla f \|_\phi \) is a shorthand for \( \sum_{i=1}^{n} \| \partial_i f \|_\phi^2 \) and \( \| \cdot \|_\phi \) stands for the Orlicz norm induced by the Young function \( \phi \) (see below).

In particular, the following holds

\[
\mu\left(f - \int_{\mathbb{R}^n} f \, d\mu \right) \leq e^{-c \max \left( t \sqrt{\| \nabla f \|_\phi^2} / 2 \| \nabla f \|_\phi \right), \quad t \geq 0
\]

where \( c > 0 \) is a universal constant.

Remark. Since 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 \( \gamma_n \)) actually implies that \( \mu \) satisfies a logarithmic Sobolev inequality, and thus hypercontractivity, with constant \( \rho \) (cf. [1]).

Proof. As it is exposed in [1], the representation formula of the variance (2.10) or the entropy (2.11) are still available. Also, as mentioned in section two, the semigroup \( (P_t)_{t \geq 0} \) still preserves (in this log-concave setting) the monotony of a function (cf. Lemma 2.15).

Besides, the assumption \( V'' \geq -\kappa \) is equivalent, thanks to Bakry and Emery’s Gamma 2 criterion, to

\[
\partial_t P_t(f) \leq e^{\kappa t}(P_t \partial f), \quad t \geq 0
\]

So, the exponential decay of the variance (2.5) (respectively of the entropy (2.6)) along the semigroup yields the following estimates (choosing \( T = \frac{1}{2\rho} \) for instance),

\[
\text{Var}_\mu(e^{-\theta f/2}) \leq \frac{\theta^2}{2} \mathbb{E}_\mu[e^{-\theta f}] \times \left( \sum_{i=1}^{n} \int_{0}^{T} e^{2\kappa t} \int_{\mathbb{R}^n} P_t^2(\partial_i f) \, d\mu \, dt \right)
\]

and

\[
\text{Ent}_\mu(e^{-\theta f/2}) \leq \frac{\theta^2}{4} \mathbb{E}_\mu[e^{-\theta f}] \times \left( \sum_{i=1}^{n} \int_{0}^{T} e^{2\kappa t} \int_{\mathbb{R}^n} P_t^2(\partial_i f) \, d\mu \, dt \right)
\]

However, it not possible to recover \( \text{Var}_\mu(f) \) from \( \int_{0}^{T} e^{2\kappa t} \int_{\mathbb{R}^n} |P_t(\nabla f)|^2 \, d\mu \, dt \) because the commutation property between \( \partial_i \) and \( (P_t)_{t \geq 0} \) is not exact.
Nevertheless, this quantity can easily be upper bounded by hypercontractive arguments. To this task, we can follow the proof of Talagrand’s inequalities exposed in [8] (page 8-9) in order to obtain

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

(4.1)

where $\|\cdot\|_\phi$ is the Orlicz norm induced by the Young function $\phi$. That is to say, as in [8], let $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be convex such that $\phi(x) = \frac{x^2}{\log(e+x)}$, $x \geq 1$ and $\phi(0) = 0$.

Application for order statistics can also be performed (as in section two). However, the hypercontractive bound on the variance (involving Orlicz norm) is not optimal. For instance, if $V(x) = |x|^\alpha/\alpha$ with $\alpha > 0$ it is known that $\text{Var}(M_n) \sim 1/(\log n)^{2(\alpha-1)/\alpha}$ whereas (2.7) only gives a bound of order $1/\log n$ which is strictly bigger since $\alpha > 2$. Let us mentionned that sharper bounds has been obtained for the right tail in [16].

It might be interesting to see if the work of [4, 3] (involving intertwinnings and weighted gradient (that is to say $|\nabla \sigma f|^2(x) = \sum_{i=1}^n (\partial_i f)^2(x_i)$ with $x = (x_1, \ldots, x_n)$) can be used to produce sharper bounds on the left tail through semigroup arguments and weighted Poincaré’s inequalities.

5 Potential extension

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 $\mu$. Assume that $(L, \mu)$ is hypercontractive and that the associated Dirichlet form $E$ may be decomposed along directions $\Gamma_i$ acting on functions on some state space $E$ as

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

in a way that, for each $i = 1, \ldots, n$, $\Gamma_i$ 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,

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

(5.1)

In the current article, $\Gamma_i = \partial_i$ and the commutation property is obtained as a strong gradient bound from Bakry and Emery’s Gamma 2 criterion.

As a first example, one can investigate the standard exponential measure (or gamma measure) $d\mu = e^{-\sum_{i=1}^n x_i 1_{\{x_i \geq 0\}} \cdots 1_{\{x_n \geq 0\}}} dx_1 \cdots dx_n$ on $\mathbb{R}_+^n$ with the direction $\Gamma_i(f) = \sqrt{x_i} \partial_i$. According to [1, 15], the commutation properties (5.1) is satisfied with $\kappa = -1$. Now, observe that the operator $\Gamma_i$, $i = 1, \ldots, n$ preserves the key features of the function $f$. More precisely, assume $f$ to be
non-decreasing and separately convex, then it is easy to check that \( x_i \mapsto \Gamma_i(f) \) is a positive non-decreasing function. 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 2.1 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 (at the level of the variance).

\[
\begin{align*}
\text{Var}_\mu(e^{-\theta f}) & \leq \frac{\theta^2}{4} \mathbb{E}_\mu[e^{-\theta f}] \times \left( \frac{2}{2 - e^{-2\kappa T}} \int_0^T e^{2\kappa t} \sum_{i=1}^n \int_{\mathbb{R}^n} P_t^i(\Gamma_i(f)e^{-\theta f/2})d\mu dt \right) \\
& \leq \frac{\theta^2}{4} \mathbb{E}_\mu[e^{-\theta f}] \times C_{\rho, \kappa} \|\Gamma(f)\|_\phi
\end{align*}
\]
where \( \|\Gamma(f)\|_\phi \) is a shorthand for \( \sum_{i=1}^n \|\Gamma_i(f)\|_2^2 \). Notice also, according to [8], that hypercontractive estimates also yields the following upper bound
\[
\|\Gamma(f)\|_\phi \leq C \sum_{i=1}^n \frac{\|\Gamma_i f\|_2^2}{1 + \log \left( \frac{\|\Gamma_i f\|_1}{\|\Gamma_i f\|_2} \right)}
\]
It is obvious that the same proof holds at the level of the entropy.

As exposed in [8], non-product measures can also be investigate. For instance, if \( \mu \) stands for the uniform probability measure on the sphere \( S^{n-1} \), one may considered the following fact
\[
\mathcal{E}(f, f) = \int_{S^{n-1}} f(-\Delta f) d\mu = \frac{1}{2} \sum_{i,j=1}^n \int_{S^{n-1}} (D_{ij} f)^2 d\mu
\]
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 (5.1) holds with \( \kappa = 0 \). However, the monotone properties needed in the proof (in order to apply Harris’s negative association inequality) seems more complicated to easily characterized. Indeed, one has to required the functions \( x \mapsto D_{ij} f(x) \) to be positive and non-decreasing. Exhibiting a class of function satisfying such conditions seems to be difficult.

References

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

[2] D. Bakry, M. Ledoux, and L. Saloff-Coste. *Markov semigroups at Saint-Flour*. Probability at Saint-Flour. Springer, Heidelberg, 2012.

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

[4] 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.

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

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

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

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

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

[13] G. Paouris and P. Valettas. A small deviation inequality for convex functions. *Arxiv (preprint) 1611.01723v1*, 2016.

[14] G. Paouris, P. Valettas, and J. Zinn. Random version of Dvoretzky’s Theorem in $l^p_n$. 2015.

[15] K. Tanguy. Some superconcentration inequalities for extrema of stationary gaussian processes. *Statistics and Probability Letters*, 2015.

[16] K. Tanguy. Non asymptotic variance bounds and deviation inequalities by optimal transport. *Preprint, http://arxiv.org/abs/1708.08620*, 2017.

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

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