Bernstein type's concentration inequalities for symmetric Markov processes

Fuqing Gao, Arnaud Guillin, Liming Wu

BERNSTEIN TYPE'S CONCENTRATION INEQUALITIES FOR SYMMETRIC MARKOV PROCESSES

FUQING GAO, ARNAUD GUILLIN, AND LIMING WU

ABSTRACT. Using the method of transportation-information inequality introduced in [28], we establish Bernstein type's concentration inequalities for empirical means $\frac{1}{t}\int_0^t g(X_s)ds$ where g is a unbounded observable of the symmetric Markov process (X_t). Three approaches are proposed : functional inequalities approach ; Lyapunov function method ; and an approach through the Lipschitzian norm of the solution to the Poisson equation. Several applications and examples are studied.

Keywords : Bernstein's concentration inequality, transportation-information inequality, functional inequality.

MSC 2000 : 60E15; 62J25, 35A23.

1. INTRODUCTION

1.1. **Bernstein's concentration inequality for sequences of i.i.d.r.v.** Let us begin with the classical Bernstein's concentration inequality in the i.i.d. case. Consider a sequence of real valued independent and identically distributed (i.i.d.) random variables (r.v.) $(\xi_k)_{k\geq 1}$, copies of some r.v. ξ, all defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{E}\xi = 0$ and $\mathbb{E}\xi^2 = \sigma^2 > 0$.

Theorem 1.1. *If there is some constant $M \geq 0$ such that*

$$\Lambda(\lambda) := \log \mathbb{E}e^{\lambda\xi} \leq \frac{\lambda^2\sigma^2}{2(1-\lambda M)}, \ \lambda \in (0, 1/M). \tag{1.1}$$

Then for any $r > 0$ and $n \geq 1$,

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^n \xi_k > r\right) \leq \exp\left(-n\frac{2r^2}{\sigma^2\left(\sqrt{1+\frac{2Mr}{\sigma^2}}+1\right)^2}\right), \ r > 0 \tag{1.2}$$

or equivalently for any $x > 0$ and $n \geq 1$,

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^n \xi_k > \sigma\sqrt{2x} + Mx\right) \leq e^{-nx}. \tag{1.3}$$

In particular

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^n \xi_k > r\right) \leq \exp\left(-\frac{nr^2}{2(\sigma^2+Mr)}\right), \ r > 0. \tag{1.4}$$

Date: May 2009.

The last inequality (1.4) is the original version of Bernstein's inequality. The proof of (1.2) is very easy : just apply Chebychev's inequality to obtain : $\forall r, \lambda > 0$,

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k > r\right) \le e^{-n\lambda r}\mathbb{E}\exp\left(\lambda\sum_{k=1}^{n}\xi_k\right) \le e^{-n[\lambda r - \Lambda(\lambda)]}$$

and then optimize over $\lambda \in (0, 1/M)$. We refer to E. Rio [44] or P. Massart [38] for known sufficient conditions for the verification of (1.1). For instance (1.1) is verified with $M = \|\xi^+\|_\infty/3$ if ξ is upper bounded, or for some not very explicit constant $M > 0$ if $\Lambda(\lambda) < +\infty$ for **some** $\lambda > 0$. Bernstein's concentration inequality is one of the most powerful concentration inequalities in probability, which is sharp both in the central limit theorem scale and the moderate deviation scale. This type of inequalities have had many applications, and are now particularly used in (non asymptotic) model selection problem, see Massart [38] or Baraud [7].

There are already many works on the generalization of Bernstein's inequality in the dependent case: Markov process or weakly dependent one. The strategy however remains the same : control the Laplace transform of partial sums. In the markovian context, Lezaud [34] used Kato's perturbation theory to get result in presence of a spectral gap, whereas Cattiaux-Guillin [15] (building on Wu [51]) used functional inequalities for the Laplace control or for the control of the mixing coefficients. More recently, Adamczak [1], Bertail-Clémençon [8], Merlevède-Peligrad-Rio [39] used a block strategy and then results in the independent case. Note however that, except the symmetric Markov processes case studied by Lezeaud [34], the known results do not reach the tight form (1.2) or (1.4). Our major objective is to give practical conditions ensuring this sharp form (1.2) in the context of integral functional of symmetric Markov processes.

There are two modern approaches to concentration inequalities. The first one, initiated by Ledoux, relies on functional inequalities, such as Poincaré or logarithmic Sobolev inequality (see for example [2] or [33]) and has attracted a lot of attention in the past decade: Wu [51] or Cattiaux-Guillin [15] used them in the continuous time context to get precise control of the Laplace transform of the partial sums, see also Massart [38] for the entropy method for various type of dependance in the discrete time case; another approach was to get a functional inequality for the whole law of the process and Herbst's like argument, note however that at this level of generality, the precise form of Bernstein's inequality has not been achieved yet.

The second approach is centered on the use of transportation inequalities (see precise definition in section 2 below): bounding Wasserstein's distance by some type of information (Kullback or Fisher). If originally investigated by Marton [36, 37] or Talagrand [46] for concentration, its systematic study is more recent, starting from the pioneer work of Bobkov-Gotze [10], followed by an abundant litterature, see [42, 9, 18, 12, 14, 26] with Kullback information, and [28, 29, 30] for Fisher information. If the use of Kullback information at the process level may lead to deviation inequality for integral functional of Markov processes (see [18] for example), the precise form of Bernstein's inequality is not reachable. We will therefore use here transportation inequalities with respect to the Fisher information, which are more natural for Markov processes : the Fisher information is exactly the large deviations rate in the Donsker-Varadhan theorem for symmetric Markov processes (see [20, 21, 22, 52, 32]).

But before going further into the details, let us present the framework on symmetric Markov processes.

1.2. Symmetric Markov processes.

Let $\mathcal{X}$ be a Polish space with Borel field $\mathcal{B}$. Let $(X_t)_{t\geq 0}$ be a $\mathcal{X}$-valued *càdlàg* Markov process with transition probability semigroup (P_t) which is symmetric and strongly continuous on $L^2(\mu) := L^2(E, \mathcal{B}, \mu)$, defined on $(\Omega, \mathcal{F}, (\mathbb{P}_x)_{x\in\mathcal{X}})$ $(\mathbb{P}_x(X_0 = x) = 1, \forall x \in \mathcal{X})$, where μ is a probability measure on $(\mathcal{X}, \mathcal{B})$, written as $\mu \in \mathcal{M}_1(\mathcal{X})$. For a given initial distribution $\beta \in \mathcal{M}_1(\mathcal{X})$, write $\mathbb{P}_\beta := \int_\mathcal{X} \beta(dx)\mathbb{P}_x(\cdot)$. Let $\mathcal{L}$ be the generator of (P_t), whose domain in $L^p(\mu) = L^p(\mathcal{X}, \mathcal{B}, \mu)$ $(p \in [1, +\infty])$ is denoted by $\mathbb{D}_p(\mathcal{L})$. It is self-adjoint, definitely non-positive on $L^2(\mu)$. Let

$$-\mathcal{L} = \int_0^{+\infty} \lambda dE_\lambda$$

be the spectral decomposition of $-\mathcal{L}$ on $L^2(\mu)$. The Dirichlet form $\mathcal{E}(f, g)$ is defined by

$$\mathbb{D}(\mathcal{E}) = \mathbb{D}_2(\sqrt{-\mathcal{L}}) = \left\{ h \in L^2(\mu); \int_0^{+\infty} \lambda d\langle E_\lambda h, h\rangle_\mu < +\infty \right\}$$

$$\mathcal{E}(f, g) = \langle \sqrt{-\mathcal{L}}f, \sqrt{-\mathcal{L}}g\rangle_\mu = \int_0^{+\infty} \lambda d\langle E_\lambda f, g\rangle_\mu, \ f, g \in \mathbb{D}(\mathcal{E})$$

where $\langle f, g\rangle_\mu = \int_\mathcal{X} fg d\mu$ is the standard inner product on $L^2(\mu)$.

We will study here deviation inequalities for

$$\frac{1}{t}\int_0^t g(X_s)ds$$

for some μ-centered function g (observable). It is quite natural to expect conditions relying on an interplay between the type of ergodicity of our Markov process and the type of boundedness or integrability of the function g.

That is why a long standing assumption in this paper will be the following Poincaré inequality : for some finite nonnegative best constant c_P,

$$\mathrm{Var}_\mu(f) \leq c_P\mathcal{E}(f, f), \ \forall f \in \mathbb{D}(\mathcal{E}). \tag{1.5}$$

Here and hereafter $\mu(f) := \int_\mathcal{X} f d\mu$ and $\mathrm{Var}_\mu(f) = \mu(f^2) - \mu(f)^2$ is the variance of f under μ. Poincaré's inequality is equivalent to the exponential decay of P_t to the equilibrium invariant measure μ in $L^2(\mu)$:

$$\mathrm{Var}_\mu(P_t f) \leq e^{-2t/c_P}\mathrm{Var}_\mu(f), \ \forall f \in L^2(\mu).$$

It is also equivalent to say that the spectral gap

$$\lambda_1 := \sup\{\lambda \geq 0; \ E_\lambda - E_0 = 0\} = \frac{1}{c_P} > 0.$$

Let us first show why this Poincaré inequality condition is natural in our context. Indeed, the first class of test function g that can be considered is the class of bounded ones. Using Kato's theory about perturbation of operators combined with ingenious and difficult combinatory calculus, Lezaud [34] proved the following Bernstein type's concentration inequality.

Theorem 1.2. ([34]) *Let g be a bounded and measurable function (say $g \in b\mathcal{B}$) such that $\mu(g) = 0$. Then for $\beta \ll \mu$,*

$$\mathbb{P}_\beta \left(\frac{1}{t} \int_0^t g(X_s) ds > r \right) \leq \|\frac{d\beta}{d\mu}\|_2 \exp \left(-\frac{2tr^2}{\sigma^2 \left(\sqrt{1 + \frac{2Mr}{\sigma^2}} + 1 \right)^2} \right) \tag{1.6}$$

$$\leq \|\frac{d\beta}{d\mu}\|_2 \exp \left(-\frac{tr^2}{2(\sigma^2 + Mr)} \right), \ \forall \ t, r > 0$$

where $M = M(g) = c_P \|g\|_\infty$ and σ^2 is the asymptotic variance (in the CLT) of the observable $g \in L^2(\mu)$, given by

$$\sigma^2 = \sigma^2(g) := \lim_{t \to +\infty} \frac{1}{t} \mathrm{Var}_{\mathbb{P}_\mu} \left(\int_0^t g(X_s) ds \right) = 2 \int_0^{+\infty} \langle P_t g, g \rangle_\mu dt. \tag{1.7}$$

For generalization of this result see Cattiaux-Guillin [15], Guillin-Léonard-Wu-Yao [28] etc. Notice a remarkable point : (1.6) is sharp both for the central limit theorem (CLT) scale $r \propto 1/\sqrt{t}$ (since $\frac{1}{\sqrt{t}} \int_0^t g(X_s) ds$ converges in law to the centered Gaussian distribution with variance $\sigma^2(g)$, see [31]), and for the moderate deviation scale (i.e. $1/\sqrt{t} \ll r \ll 1$) by the moderate deviation principle due to [50].

Notice that if $\sigma^2(g) \leq C\|g\|_\infty^2$ for some constant $C > 0$ and for all $g \in b\mathcal{B}$ with $\mu(g) = 0$, then the Bernstein's concentration inequality (1.6) implies the Poincaré inequality (1.5), by [28, Theorem 3.1]. In other words the Poincaré inequality is a minimal assumption for Bernstein's concentration inequality for **all** bounded observables g.

Remark 1.3. Let us point out that for bounded g, the assumption that $\sigma^2(g) \leq C\|g\|_\infty^2$ is a weak one, as by definition (1.7)

$$\sigma^2(g) \leq 4\|g\|_\infty \int_0^t \mathrm{Var}_\mu(P_t g)^{1/2} dt.$$

Assume now that a weak Poincaré inequality holds (see [5] for example), or a Lyapunov condition, i.e. $\mathcal{L}V \leq -\phi(V) + b1_C$ for some sub linear ϕ (see [23] for details), ensuring that $\mathrm{Var}_\mu(P_t g) \leq \psi(t)\|g\|_\infty^2$ with $\int_0^s \psi(s)^{1/2} ds < \infty$, then the Poincaré inequality holds under Bernstein's type inequality. We refer to the last section for some examples of this Lyapunov condition.

1.3. **Main question and organization.** The main question we will focus on in this paper will be: *what is the interplay between the ergodic properties of the symmetric Markov process and the test function g?* Or more precisely, how to bound the constant M (appearing in (1.6)) by means of other quantities than $\|g\|_\infty$ and c_P?

In fact we shall answer this question by a very simple approach : instead of a direct control of the Laplace transform of partial sums, we use the method of transportation-information inequality introduced by Guillin-Léonard-Wu-Yao [28].

This paper is organized as follows. In the next section we describe the strategy and the main idea of this work, giving by the way another proof of Theorem 1.2 with a better estimate of M. The goal of the three following sections is to generalize Bernstein's inequality to unbounded case. We present three approaches : (1) functional inequalities such as log-Sobolev inequality or Φ-Sobolev inequality ; (2) the Lipschitzian norm $\|(-\mathcal{L})^{-1}g\|_{Lip}$;

and (3) Meyn-Tweedie's Lyapunov function method. Finally the last section is dedicated to the case where Poincaré inequality does not hold anymore, and the class of bounded test functions is now too large. Once again, the approach via Lyapunov function will be particularly efficient.

Note that, from *Section 2 through 5*, we assume implicitly that the previous Poincaré inequality is satisfied.

Before going to the job let us fix some more notations. For $p \in [1, +\infty]$, $\| \cdot \|_p$ is the standard norm of $L^p(\mu) := L^p(\mathcal{X}, \mathcal{B}, \mu)$, and $L^p_0(\mu) := \{g \in L^p(\mu); \ \mu(g) = 0\}$. The quantity σ^2 denotes always the asymptotic variance $\sigma^2(g)$ in the CLT, given by (1.7). The empirical measure $\frac{1}{t} \int_0^t \delta_{X_s} ds$ (δ_x being the Dirac measure at point x) is denoted by L_t, so that $\frac{1}{t} \int_0^t g(X_s) ds = L_t(g)$.

2. A transportation-information look at Bernstein's inequality

2.1. **The strategy and the main idea.** As in [28], our starting point is

Theorem 2.1. (Wu [51]) *Let $g \in L^1_0(\mu)$. Then*

$$\mathbb{P}_\beta \left(\frac{1}{t} \int_0^t g(X_s) ds > r \right) \le \|\frac{d\beta}{d\mu}\|_2 e^{-tI(r-)}, \ \forall t, r > 0 \tag{2.1}$$

where

$$I(r) := \inf\{I(\nu|\mu); \ \nu(|g|) < +\infty, \nu(g) = r\}, \ I(r-) := \lim_{\varepsilon \to 0+} I(r - \varepsilon), \ r \in \mathbb{R}$$

and

$$I(\nu|\mu) := \begin{cases} \mathcal{E}\left(\sqrt{f}, \sqrt{f}\right), & \text{if } \nu = f\mu, \sqrt{f} \in \mathbb{D}(\mathcal{E}), \\ +\infty, & \text{otherwise} \end{cases} \tag{2.2}$$

is the Fisher-Donsker-Varadhan's information of ν with respect to (w.r.t.) μ.

By the large deviations in Donsker-Varadhan [20, 21] (in the regular case) and Wu [52] (in full generality), $\nu \to I(\nu|\mu)$ is the rate function in the large deviations of the empirical measures $L_t := \frac{1}{t} \int_0^t \delta_{X_s} ds$, and the Cramer type's inequality (2.1) is sharp for large time t. The main problem now is to estimate the rate function $I(r)$ in the large deviations of $\frac{1}{t} \int_0^t g(X_s) ds$: that is exactly a role that the transportation-information inequality plays.

Theorem 2.2. ([28, Theorem 2.4]) *Let $g \in L^1_0(\mu)$ and $\alpha : \mathbb{R} \to [0, +\infty]$ be a nondecreasing left-continuous convex function with $\alpha(0) = 0$. The following properties are equivalent :*

(a) $\alpha(\nu(g)) \le I(\nu|\mu), \ \forall \nu \in \mathcal{M}_1(\mathcal{X})$ such that $\nu(|g|) < +\infty$.

(b) $\nu(g) \le \alpha^{-1}(I(\nu|\mu)), \ \forall \nu \in \mathcal{M}_1(\mathcal{X})$ such that $\nu(|g|) < +\infty$, where $\alpha^{-1}(x) := \inf\{r \in \mathbb{R}; \ \alpha(r) > x\}$ is the right inverse of α.

(c) It holds that

$$\mathbb{P}_\beta \left(\frac{1}{t} \int_0^t g(X_s) ds > r \right) \le \|\frac{d\beta}{d\mu}\|_2 e^{-t\alpha(r)}, \ \forall t, r > 0. \tag{2.3}$$

(d) It holds that

$$\mathbb{P}_\beta \left(\frac{1}{t} \int_0^t g(X_s) ds > \alpha^{-1}(x) \right) \le \|\frac{d\beta}{d\mu}\|_2 e^{-tx}, \ \forall t, x > 0. \tag{2.4}$$

(e) For any $\lambda > 0$,

$$\Lambda(\lambda g) := \sup\left\{\int_{\mathcal{X}} \lambda g h^2 d\mu - \mathcal{E}(h,h) | h \in \mathbb{D}(\mathcal{E}), \mu(h^2) = 1\right\} \leq \alpha^*(\lambda) \qquad (2.5)$$

where $\alpha^*(\lambda) := \sup_{r \geq 0}\{\lambda r - \alpha(r)\}$ is the (semi)-Legendre transformation of α.

It is not completely contained in [28, Theorem 2.4] (the condition (A2) therein is not satisfied), but the proof there works. Indeed $(a) \Leftrightarrow (b)$ and $(c) \Leftrightarrow (d)$ are obvious. We give the proof of the crucial implication $(a) \implies (c)$ for its simplicity. In fact by the transportation-information inequality in (a), we have for $r > 0$,

$$I(r) = \inf\{I(\nu|\mu); \ \nu(|g|) < +\infty, \ \nu(g) = r\} \geq \alpha(r)$$

and then $I(r-) \geq \alpha(r)$ by the left-continuity of α. Hence the concentration inequality (2.3) follows immediately from (2.1).

Remark 2.3. By Rayleigh's principle, $\Lambda(\lambda g)$ is the supremum of the spectrum of the Schrödinger operator $\mathcal{L} + \lambda g$ (in the sum-form sense).

Bernstein's inequality (1.6) is just (2.3) with

$$\alpha(r) = 1_{r \geq 0} \frac{2r^2}{\sigma^2 \left(\sqrt{1 + \frac{2Mr}{\sigma^2}} + 1\right)^2}.$$

Since $\alpha^{-1}(x) = \sqrt{2\sigma^2 x} + Mx$ for $x \geq 0$, by Theorem 2.2, Bernstein's inequality (1.6) is equivalent to

$$\nu(g) \leq \sqrt{2\sigma^2 I} + MI, \ I := I(\nu|\mu), \ \forall \nu \in \mathcal{M}_1(\mathcal{X}) \text{ so that } \nu(|g|) < +\infty. \qquad (2.6)$$

That is the strategy of this work.

Now let us present a very simple proof of Lezaud's result, which illustrates also the main idea for our approaches to establish (2.6). Assume $g \in L_0^2(\mu)$ so that $g^+ \in L^\infty(\mu)$.

Let $\nu = f\mu$ and $h = \sqrt{f} \in \mathbb{D}(\mathcal{E})$ (trivial otherwise for $I = +\infty$) such that $\nu(|g|) < +\infty$. Our main idea resides in the following simple but key decomposition :

$$\nu(g) = \int_{\mathcal{X}} g h^2 d\mu = \int_{\mathcal{X}} g\left[(h - \mu(h))^2 + 2\mu(h)h\right] d\mu \quad (\text{since } \mu(g) = 0)$$
$$= 2\mu(h)\langle g, h\rangle_\mu + \int_{\mathcal{X}} g(h - \mu(h))^2 d\mu =: A + B. \qquad (2.7)$$

Bounding A.
For the first term $A = 2\mu(h)\langle g, h\rangle_\mu$, note that $\mu(h) \leq \sqrt{\mu(h^2)} = 1$. Let $(-\mathcal{L})^{-1}g = \int_0^{+\infty} P_t g \, dt$ be the Poisson operator (the integral is absolutely convergent in $L^2(\mu)$ for all $g \in L_0^2(\mu)$ by the Poincaré inequality). Hence

$$\sigma^2 = \sigma^2(g) = 2\int_0^\infty \langle P_t g, g\rangle dt = 2\langle(-\mathcal{L})^{-1}g, g\rangle_\mu.$$

By Cauchy-Schwarz, we have

$$|\langle g, h\rangle_\mu| \leq \sqrt{\langle(-\mathcal{L})^{-1}g, g\rangle \mathcal{E}(h,h)} = \sqrt{\frac{\sigma^2}{2}I} \qquad (2.8)$$

Hence $|A| \leq \sqrt{2\sigma^2 I}$, in other words, the term A is always bounded by the first term at the right hand side of the inequality (2.6).

Remark 2.4. *Even without the hypothesis of the Poincaré inequality, (2.8) is still true for $g \in L_0^2(\mu)$ by Kipnis-Varadhan [31] once if $\sigma^2(g) = 2 \int_0^\infty \langle g, P_t g \rangle dt < +\infty$. The latter condition is the famous sufficient condition of Kipnis-Varadhan for the CLT of $\int_0^t g(X_s)ds$.*

Bounding B.
Now for (2.6) it remains to prove that the second term B satisfies

$$B = \int_{\mathcal{X}} g[h - \mu(h)]^2 d\mu \leq M\mathcal{E}(h, h) = MI. \tag{2.9}$$

It is indeed very easy in terms of $\|g\|_\infty$: letting $g^+ = \max\{g, 0\}$, we have by Poincaré,

$$B = \int_{\mathcal{X}} g[h - \mu(h)]^2 d\mu \leq \int_{\mathcal{X}} g^+[h - \mu(h)]^2 d\mu \leq \|g^+\|_\infty \mathrm{Var}_\mu(h) \leq c_P \|g^+\|_\infty I.$$

In other words we have proven (2.6) with $M = c_P\|g^+\|_\infty$, which is a little better than Lezaud's estimate $M = c_P\|g\|_\infty$. We summarize the discussion above as

Proposition 2.5. *Let $g \in b\mathcal{B}$ with $\mu(g) = 0$. Then (2.6) holds with $M = c_P\|g^+\|_\infty$, or equivalently Bernstein's inequality (1.6) holds with such M.*

Our remained task consists in proving (2.9) with some constant $M = M(g)$ for various classes of functions g under different ergodicity conditions for the process. Remark that the best constant $M(g)$ for (2.9) (or (2.6)) is positively homogeneous, i.e. $M(cg) = cM(g)$ for all $c \geq 0$.

2.2. **Approach by transportation-information inequality T_cI.** Let us introduce our first approach by means of the transportation-information inequality T_cI in [28].

Consider a cost function $c : \mathcal{X}^2 \to [0, +\infty]$ which is always lower semi-continuous (l.s.c.) and $c(x, x) = 0$ for all $x \in \mathcal{X}$, here $c(x, y)$ represents the cost of transporting a unit mass from x to y. Now given two probability measures $\nu, \mu \in \mathcal{M}_1(\mathcal{X})$, we define the *transportation cost from ν to μ* by

$$T_c(\nu, \mu) := \inf_{\pi \in \mathcal{C}(\nu, \mu)} \iint_{\mathcal{X}^2} c(x, y)\pi(dx, dy) \tag{2.10}$$

where $\mathcal{C}(\nu, \mu)$ is the family of all couplings of (ν, μ), i.e. all probability measures π on $\mathcal{X}^2$ such that $\pi(A \times \mathcal{X}) = \nu(A)$, $\pi(\mathcal{X} \times B) = \mu(B)$ for all $A, B \in \mathcal{B}$.

Let $d(x, y)$ be a l.s.c. metric on $\mathcal{X}$, which does not necessarily generate the topology of $\mathcal{X}$. For any $p \geq 1$, the quantity

$$W_{p,d}(\nu, \mu) := (T_{d^p}(\nu, \mu))^{1/p} = \left(\inf_{\pi \in \mathcal{C}(\nu, \mu)} \iint_{\mathcal{X}^2} d^p(x, y)\pi(dx, dy) \right)^{1/p} \tag{2.11}$$

is the so called L^p-*Wasserstein distance* between ν and μ. $W_{p,d}$ is a metric on $\mathcal{M}_1^{d,p}(\mathcal{X}) := \{\nu \in \mathcal{M}_1(\mathcal{X}); \left(\int_{\mathcal{X}} d^p(x_0, x)\nu(dx) \right)^{1/p} < +\infty\}$ ($x_0 \in \mathcal{X}$ is some fixed point). We refer to the recent books of Villani [48, 47] for more on this subject.

An important particular case is $d(x, y) = 1_{x \neq y}$, the trivial metric on $\mathcal{X}$. In that case

$$W_{1,d}(\nu, \mu) = \frac{1}{2}\|\nu - \mu\|_{TV} = \sup_{A \in \mathcal{B}} |\nu(A) - \mu(A)| \tag{2.12}$$

where $\|m\|_{TV} = \sup_{f \in b\mathcal{B}, |f| \leq 1} |m(f)|$ is the total variation of a signed bounded measure m on $\mathcal{X}$. More generally given a positive continuous weight function ϕ, consider the distance $d_\phi(x, y) = 1_{x \neq y}[\phi(x) + \phi(y)]$, then (cf. [26])

$$W_{1,d_\phi}(\nu, \mu) = \|\phi(\nu - \mu)\|_{TV}.$$

Theorem 2.6. *Assume the following transportation-information inequality*

$$\alpha(T_c(\nu, \mu)) \leq I(\nu|\mu), \ \forall \nu \in \mathcal{M}_1(\mathcal{X}) \tag{2.13}$$

where α is nonnegative, nondecreasing convex and left continuous with $\alpha(0) = 0$ such that its right inverse α^{-1} is concave and $\alpha^{-1}(0) = 0$. Then for every measurable $g \in L_0^2(\mu)$ such that its sup-convolution

$$g^*(y) = \sup_{x \in \mathcal{X}} (g(x) - c(x, y)), \ y \in \mathcal{X} \tag{2.14}$$

is in $L^1(\mu)$, (2.6) and Bernstein's inequality (1.6) hold with

$$M(g) = \mu(g^*)c_P + c_P \alpha^{-1}\left(\frac{1}{c_P}\right). \tag{2.15}$$

In particular if the $W_1 I$-transportation-information inequality below holds

$$W_{1,d}^2(\nu, \mu) \leq 2c_G I(\nu|\mu), \ \forall \nu \in \mathcal{M}_1(\mathcal{X}) \tag{2.16}$$

then (2.6) holds for every d-Lipschitzian function g (with $\mu(g) = 0$) with

$$M(g) = \|g\|_{Lip(d)}\sqrt{2c_P c_G}.$$

Proof. At first $g^*(y) \geq g(y)$, $y \in \mathcal{X}$, so $\mu(g^*) \geq \mu(g) = 0$. For (2.6) we may assume that $\nu = h^2\mu$ with $0 \leq h \in \mathbb{D}(\mathcal{E})$ and $Var_\mu(h) \neq 0$ (trivial otherwise for $\nu = \mu$). Letting $\tilde{h} = h - \mu(h)$ and $\tilde{\nu} := \tilde{h}^2\mu/Var_\mu(h)$, we have by the very definition of T_c,

$$\int_{\mathcal{X}} g(x)\tilde{\nu}(dx) \leq \int_{\mathcal{X}} g^*(y)\mu(dy) + T_c(\tilde{\nu}, \mu)$$

$$\leq \mu(g^*) + \alpha^{-1}(I(\tilde{\nu}|\mu)) \leq \mu(g^*) + \alpha^{-1}\left(\frac{\mathcal{E}(h, h)}{Var_\mu(h)}\right)$$

where we have used $\mathcal{E}(|\tilde{h}|, |\tilde{h}|) \leq \mathcal{E}(\tilde{h}, \tilde{h}) = \mathcal{E}(h, h)$. It follows by the concavity of α^{-1},

$$B = \int_{\mathcal{X}} g\tilde{h}^2 d\mu \leq \mu(g^*)Var_\mu(h) + Var_\mu(h)\alpha^{-1}\left(\frac{\mathcal{E}(h, h)}{Var_\mu(h)}\right) \leq \mu(g^*)c_P I + c_P I \alpha^{-1}(1/c_P)$$

the desired (2.9).

For the last particular case we may assume that $\|g\|_{Lip(d)} = 1$. In that case $g^* = g$, and then one can apply (2.15). $\square$

Remark 2.7. By the preceding result, one can apply the criteria for $T_c I$ or $W_1 I$-transportation information inequalities in [28] to obtain Bernstein's inequality.

3. FUNCTIONAL INEQUALITIES APPROACH

3.1. **Log-Sobolev inequality.** Recall that for $0 \leq f \in L^1(\mu)$, the entropy of f w.r.t. μ is defined by

$$\text{Ent}_\mu(f) = \mu(f \log f) - \mu(f) \log \mu(f). \tag{3.1}$$

The log-Sobolev inequality ([3, 33]) says

$$\text{Ent}_\mu(h^2) \leq 2c_{LS}\mathcal{E}(h, h), \ \forall h \in \mathbb{D}(\mathcal{E}), \tag{3.2}$$

where c_{LS} is the best constant, called *log-Sobolev constant*. It is well known that $c_P \leq c_{LS}$.

Theorem 3.1. *Assume the log-Sobolev inequality (3.2). Let $g \in L_0^2(\mu)$ satisfy $\Lambda(\lambda) := \log \int_{\mathcal{X}} e^{\lambda g} d\mu < +\infty$ for some $\lambda > 0$.*

Then the transportation-information inequality (2.6) holds with

$$M = \inf_{\lambda > 0} \frac{1}{\lambda} \left[c_P \Lambda(\lambda) + 2c_{LS} \right] \leq c_P(\Lambda^*)^{-1}\left(\frac{2c_{LS}}{c_P}\right) \tag{3.3}$$

where $\Lambda^* : \mathbb{R}^+ \to [0, +\infty]$ is the Legendre transform of Λ and $(\Lambda^*)^{-1}$ is the right inverse. In particular Bernstein's inequality (1.6) holds with this constant M.

Proof. We may assume that $\nu = h^2\mu$ with $0 \leq h \in \mathbb{D}(\mathcal{E})$. We have to bound the term $B = \int_{\mathcal{X}} g[h - \mu(h)]^2 d\mu$ in the decomposition (2.7). Writing $\tilde{h} = h - \mu(h), I = I(\nu|\mu) = \mathcal{E}(h, h)$, we have for any constant $\lambda > 0$ such that $\Lambda(\lambda) < +\infty$, $\int e^{\lambda g - a} d\mu = 1$ where $a = \Lambda(\lambda) \geq 0$, and then

$$\begin{aligned}
B &= \frac{1}{\lambda} \left(\int_{\mathcal{X}} (\lambda g - a)\tilde{h}^2 d\mu + a \int \tilde{h}^2 d\mu \right) \\
&\leq \frac{1}{\lambda} \left(\text{Ent}_\mu(\tilde{h}^2) + ac_P I \right) \\
&\leq \frac{1}{\lambda} \left[2c_{LS} + \Lambda(\lambda)c_P \right] \cdot I
\end{aligned}$$

where the second inequality relies on $\text{Ent}_\mu(f) = \sup_{g:\mu(e^g)\leq 1} \int_{\mathcal{X}} fg d\mu$ (Donsker-Varadhan's variational formula) and the Poincaré inequality, and the third one on the log-Sobolev inequality. Optimizing over $\lambda > 0$ yields (2.6) with M given in (3.3). $\square$

It is a surprise : the explicit estimate of $M = M(g)$ above is not available even in the i.i.d. case under the exponential integrability condition.

Let us give a more explicit estimate of M in the diffusion case. We assume that

(H_Γ) $(\mathcal{E}, \mathbb{D}(\mathcal{E}))$ is given by the carré-du-champs $\Gamma : \mathbb{D}(\mathcal{E}) \times \mathbb{D}(\mathcal{E}) \to L^1(\mu)$ (symmetric, bilinear definite nonnegative form):

$$\mathcal{E}(h, h) = \int_{\mathcal{X}} \Gamma(h, h) \, d\mu, \ \forall h \in \mathbb{D}(\mathcal{E}). \tag{3.4}$$

Diffusion framework. We shall assume that Γ is a differentiation (or equivalently the sample paths of (X_t) are continuous, $\mathbb{P}_\mu - a.s.$, cf. Bakry [3]), that is: for all $(h_k)_{1 \leq k \leq n} \subset$

$\mathbb{D}(\mathcal{E}), g \in \mathbb{D}(\mathcal{E})$ and $F \in C_b^1(\mathbb{R}^n)$,

$$\Gamma(F(h_1, \cdots, h_n), g) = \sum_{i=1}^n \partial_i F(h_1, \cdots, h_n)\Gamma(h_i, g).$$

Write $\Gamma(f) := \Gamma(f, f)$ simply.

Corollary 3.2. *Assume (H_Γ) and that Γ is a differentiation. If the log-Sobolev inequality holds, then for any $g \in \mathbb{D}(\mathcal{E})$ so that $\Gamma(g)$ is bounded and $\mu(g) = 0$, the transportation-information inequality (2.6) holds with*

$$M = 2c_{LS}\sqrt{c_P \|\Gamma(g)\|_\infty}. \tag{3.5}$$

Proof. By Ledoux [33] or Bobkov-Götze [10], in the actual diffusion case the log-Sobolev inequality implies that

$$\Lambda(\lambda) = \log \int_{\mathcal{X}} e^{\lambda g} d\mu \le \frac{1}{2} c_{LS} \lambda^2 \|\Gamma(g)\|_\infty, \ \forall \lambda > 0.$$

Plugging it into (3.3), we get $M \le 2c_{LS}\sqrt{c_P\|\Gamma(g)\|_\infty}$. $\qquad \square$

Example 3.3. (Ornstein-Uhlenbeck processes) Let $\mu = \mathcal{N}(0, \theta)$, the Gaussian measure with zero mean and variance $\theta > 0$ on $\mathcal{X} = \mathbb{R}$, and $\mathcal{L}f = f'' - \theta^{-1}x \cdot f'$. It is well known that $c_P = c_{LS} = \theta$.

For every Lipschitzian function g with $\mu(g) = 0$, $\sqrt{\|\Gamma(g)\|_\infty} = \|\nabla g\|_\infty = \|g\|_{Lip}$ (the Lipschitzian coefficient w.r.t. the Euclidean metric). By Corollary 3.2, Bernstein's inequality (1.6) holds with $M = 2c_{LS}\sqrt{c_P}\|g\|_{Lip} = 2\theta^{3/2}\|g\|_{Lip}$. It is worth mentioning that for the special observable $g(x) = x$, (2.6) and then Bernstein's inequality (1.6) hold with $M = 0$ (i.e. the corresponding Gaussian concentration inequality holds); and for general g with $\mu(g) = 0$,

$$\nu(g) \le \|g\|_{Lip}\sqrt{2\theta I}$$

holds by [28, Proposition 2.9].

But by Theorem 3.1, for every μ-centered function g such that $\int e^{\delta g} d\mu < +\infty$ (for instance if $g \le C(1 + |x|^2)$), Bernstein's inequality (1.6) holds with $M = M(g)$ given in (3.3). Though natural, that was not known before up to our knowledge. It is easy to see that Bernstein inequality is false for observable $g(x)$ such that $\lim_{x \to \infty} \frac{g(x)}{x^2} = +\infty$.

Let us look at the particularly interesting observable $g(x) = g_0(x) := x^2 - \theta$ for which we can get sharp Bernstein inequality. Indeed since $-\mathcal{L}g_0 = -2\theta^{-1}g_0$,

$$\sigma^2(g_0) = 2\langle(-\mathcal{L})^{-1}g_0, g_0\rangle_\mu = \theta \text{Var}_\mu(g_0) = 2\theta^3.$$

On the other hand observe that for each real number $a < \frac{1}{2}$, $U(x) := \exp\left(\frac{ax^2}{2\theta}\right) \in L^2(\mu)$, and

$$\left[\mathcal{L} + \frac{a - a^2}{\theta^2}g_0\right]U = \frac{a^2}{\theta}U.$$

In other words U is a positive eigenfunction of the Schrödinger operator $\mathcal{L} + \frac{a-a^2}{\theta^2}g_0$ associated with eigenvalue a^2/θ, which implies that (by Perron-Frobenius theorem and Rayleigh's formula)

$$\Lambda\left(\frac{a - a^2}{\theta^2}g_0\right) = \frac{a^2}{\theta}, \ a < \frac{1}{2}.$$

Hence for all $\lambda < \lambda_0 := \frac{1}{4\theta^2}$, taking $a = a_- := \frac{1}{2}\left(1 - \sqrt{1 - 4\theta^2\lambda}\right) < 1/2$, we have

$$\Lambda(\lambda g_0) = \frac{1}{4\theta}\left(1 - \sqrt{1 - 4\theta^2\lambda}\right)^2$$

Since $\lambda \to \Lambda(\lambda g_0)$ from $\mathbb{R}$ to $(-\infty, +\infty]$ is convex and lower semi-continuous, and its left derivative at λ_0 is $+\infty$, we conclude that

$$\Lambda(\lambda) := \Lambda(\lambda g_0) = \frac{1}{4\theta}\left(1 - \sqrt{1 - 4\theta^2\lambda}\right)^2 \ \text{ if } \ \lambda \le \lambda_0 = \frac{1}{4\theta^2};\ +\infty, \ \text{ if } \lambda > \lambda_0. \qquad (3.6)$$

From the previous explicit expression we obtain (by the fact that the geometric mean is not greater than the arithmetic mean)

$$\Lambda(\lambda) = \frac{\sigma^2(g_0)\lambda^2}{2(\frac{1}{2}[1 + \sqrt{1 - 4\theta^2\lambda}]^2} \le \frac{\sigma^2(g_0)\lambda^2}{2(1 - 4\theta^2\lambda)}, \ \lambda \in (0, \lambda_0)$$

where it follows that $g_0(x) = x^2 - \theta$ satisfies the Bernstein inequality (1.6) with the sharp constant $M = 4\theta^2$.

Notice that (3.6) will give, by Theorem 2.2, the concentration inequality for the estimator $\frac{1}{t}\int_0^t X_s^2 ds$ of θ, which is not only sharp for the CLT and moderate deviation scales, but also for large deviations.

3.2. Φ-Sobolev inequality.

Let $\Phi : \mathbb{R}^+ \to [0, +\infty]$ be a Young function, i.e. a convex, increasing and left continuous function with $\Phi(0) = 0$ and $\lim_{x \to +\infty} \Phi(x) = +\infty$. Consider the Orlicz space $L^\Phi(\mu)$ of those measurable functions g on $\mathcal{X}$ so that its gauge norm

$$N_\Phi(g) := \inf\{c > 0; \int \Phi(|g|/c)d\mu \le 1\}$$

is finite, where the convention $\inf \emptyset := +\infty$ is used. The Orlicz norm of g is defined by

$$\|g\|_\Phi := \sup\{\int gu\, d\mu;\ N_\Psi(u) \le 1\}$$

where

$$\Psi(r) := \sup_{\lambda \ge 0}(\lambda r - \Phi(\lambda)),\ r \ge 0 \qquad (3.7)$$

is the convex conjugate of Φ. It is well known that ([43, Proposition 4, p.61])

$$N_\Phi(g) \le \|g\|_\Phi \le 2N_\Phi(g).$$

The Φ-Sobolev inequality says that

$$\|(h - \mu(h))^2\|_\Phi \le c_{P,\Phi}\mathcal{E}(h, h),\ \forall h \in \mathbb{D}(\mathcal{E}) \qquad (3.8)$$

called sometimes Orlicz-Poincaré inequality, where $c_{P,\Phi}$ is the best constant. There is a rich theory of long history for this subject, see [17, 33, 49].

Set $\tilde{\Phi}(x) := \Phi(x^2), x \ge 0$ and let $\tilde{\Psi}$ be the Legendre transform of $\tilde{\Phi}$.

Lemma 3.4. *Assume the Φ-Sobolev inequality (3.8). If $g \in L^{\tilde{\Psi}}(\mu)$ so that $\mu(g) = 0$, then $\int_0^t g(X_s)ds \in L^2(\mathbb{P}_\mu)$ and it holds that*

$$\sigma^2(g) = \lim_{t \to +\infty} \frac{1}{t}\mathrm{Var}_{\mathbb{P}_\mu}\left(\int_0^t g(X_s)ds\right) \le c_{P,\Phi}\|g\|_{\tilde{\Psi}}^2. \qquad (3.9)$$

Moreover

$$\langle g, h\rangle_\mu^2 \leq \frac{1}{2}\sigma^2(g)\mathcal{E}(h,h), \ \forall h \in \mathbb{D}(\mathcal{E}). \tag{3.10}$$

Proof. At first for $g \in L_0^2(\mu)$, notice that by the spectral decomposition and Cauchy-Schwarz,

$$\sqrt{\langle g, (-\mathcal{L})^{-1}g\rangle_\mu} = \sup_{h \in \mathbb{D}(\mathcal{E}), \mathcal{E}(h,h)\leq 1} \langle g, h\rangle_\mu$$

and

$$|\langle g, h\rangle_\mu| = |\langle g, h - \mu(h)\rangle_\mu| \leq \|g\|_{\tilde{\Psi}} N_{\tilde{\Phi}}(h - \mu(h)).$$

Furthermore by the Φ-Sobolev inequality (3.8),

$$N_{\tilde{\Phi}}(h - \mu(h)) = \sqrt{N_\Phi((h - \mu(h))^2)} \leq \sqrt{\|(h - \mu(h))^2\|_\Phi} \leq \sqrt{c_{P,\Phi}\mathcal{E}(h,h)}$$

therefore

$$\langle g, (-\mathcal{L})^{-1}g\rangle_\mu \leq c_{P,\Phi}\|g\|_{\tilde{\Psi}}^2, \ g \in L_0^2(\mu). \tag{3.11}$$

Now take a sequence (g_n) in $L_0^\infty(\mu)$ converging to g in $L^\Psi(\mu)$, we have for any $t > 0$,

$$\frac{1}{t}\mathrm{Var}_{\mathbb{P}_\mu}\left(\int_0^t (g_n - g_m)(X_s)ds\right) \leq \sigma^2(g_n - g_m) = 2\langle g_n - g_m, (-\mathcal{L})^{-1}(g_n - g_m)\rangle_\mu$$

$$\leq 2c_{P,\Phi}\|g_n - g_m\|_{\tilde{\Psi}}^2.$$

This implies not only "$\int_0^t g(X_s)ds \in L^2(\mathbb{P}_\mu)$" but also (3.9). The last claim (3.10) holds for g_n in place of g then remains true for g by letting $n \to \infty$. $\qquad\square$

Theorem 3.5. *Assume the Φ-Sobolev inequality (3.8) and let Ψ be the convex conjugate of Φ given above. If $g \in L^{\tilde{\Psi}}(\mu)$ and $g^+ \in L^\Psi(\mu)$ with $\mu(g) = 0$, then the transportation-information inequality (2.6) holds with $\sigma^2 = \sigma^2(g)$ given by (3.9) and*

$$M = N_\Psi(g^+) \cdot c_{P,\Phi}. \tag{3.12}$$

In particular Bernstein's inequality (1.6) holds with that constant M.

Proof. The proof is even easier than that of Theorem 3.1. For (2.6) we may assume that $\nu = h^2\mu$ with $0 \leq h \in \mathbb{D}(\mathcal{E})$. By Lemma 3.4, $\sigma^2 = \sigma^2(g)$ given by (3.9) is finite. The term A in (2.6) is bounded by $\sqrt{2\sigma^2 I}$ by (3.10). For the term $B = \int_\mathcal{X} g[h - \mu(h)]^2 d\mu$ we have

$$B \leq N_\Psi(g^+)\|[h - \mu(h)]^2\|_\Phi \leq c_{P,\Phi} N_\Psi(g^+) I$$

where the desired result follows. $\qquad\square$

Remark 3.6. When $\Phi(x) = |x|$, $\Psi(x) = +\infty \cdot 1_{x>1}$, $N_\Psi(h) = \|h\|_\infty$. Then this result generalizes Proposition 2.5.

Remark 3.7. For one-dimensional diffusions, an explicit necessary and sufficient condition for the Φ-Sobolev inequality (3.8) is available, see the book of M.F. Chen [17]. For Φ-Sobolev inequality in high dimension, see the book of F.Y. Wang [49] for numerous known results.

Example 3.8. As a well known fact (see Saloff-Coste [45]), for the Brownian Motion (B_t) on a compact connected Riemannian manifold M of dimension n with the invariant measure μ given by the normalized Riemannian measure $\dfrac{dx}{V(M)}$ (where $V(M)$ is the volume of M), the Dirichlet form $\int |\nabla f|^2 d\mu$ satisfies the Φ-Sobolev inequality (3.8) with

$$\Phi(t) = \begin{cases} +\infty I_{(1,\infty)}(|t|), & \text{if} \quad n = 1, \\[2mm] \exp(C|t|) - 1, & \text{if} \quad n = 2, \\[2mm] |t|^{\frac{2n}{n-2}}, & \text{if} \quad n \geq 3. \end{cases}$$

Hence Bernstein's inequality (1.6) holds for $g \in L_0^1(\mu)$ satisfying

$$g \in \begin{cases} L^1(\mu), & \text{if} \quad n = 1, \\[2mm] L^1 \log L^1, & \text{if} \quad n = 2, \\[2mm] L^{\frac{2n}{n+2}}(\mu), & \text{if} \quad n \geq 3. \end{cases}$$

Those still hold for diffusion generated by $\Delta - \nabla V \cdot \nabla$ with C^2-smooth function V on a connected compact manifold.

Example 3.9. Consider the measure $\mu_\beta(dx) = \dfrac{\exp(-|x|^\beta)}{Z_\beta}$ (where Z_β is the normalized constant), and $\beta > 1$. For the diffusion process corresponding to the Dirichlet form $\langle -\mathcal{L}f, f \rangle_\mu = \int |\nabla f|^2 d\mu$, it satisfies Φ-Sobolev inequality (3.8) with

$$\Phi_\alpha(x) = x \log^\alpha(1 + x), \quad \alpha = 2(1 - 1/\beta)$$

according to Barthe, Cattiaux and Roberto [6, section 7]. Hence Bernstein's inequality (1.6) holds for $g \in L_0^2(\mu)$ satisfying

$$\int \exp\left(\lambda (g^+)^{\beta/(2\beta-2)}\right) d\mu < +\infty, \text{ for some } \lambda > 0. \tag{3.13}$$

Those two examples show that for Bernstein's inequality to hold, the integrability condition on the observable g in the continuous time symmetric Markov processes case may be much weaker than the exponential integrability condition in the i.i.d. case.

4. Lyapunov function method

Sometimes functional inequalities are difficult to check. In that situation the easy-to-check Lyapunov function method will be very helpful.

4.1. **General result.** A measurable function G is said to be in the μ-extended domain $\mathbb{D}_{e,\mu}(\mathcal{L})$ of the generator of the Markov process $((X_t), \mathbb{P}_\mu)$ if there is some measurable function g such that $\int_0^t |g|(X_s)\,ds < +\infty, \mathbb{P}_\mu$-a.s. and one $\mathbb{P}_\mu$-version of

$$M_t(G) := G(X_t) - G(X_0) + \int_0^t g(X_s)ds$$

is a local $\mathbb{P}_\mu$-martingale. It is obvious that g is uniquely determined up to μ-equivalence. In such case one writes $G \in \mathbb{D}_{e,\mu}(\mathcal{L})$ and $-\mathcal{L}G = g$. When the above properties hold for $\mathbb{P}_x$ instead of $\mathbb{P}_\mu$ for **every** $x \in \mathcal{X}$, we say that G belongs to the extended domain $\mathbb{D}_e(\mathcal{L})$. In the latter case $-\mathcal{L}G = g$ is determined uniquely up to $\int_0^\infty e^{-t} P_t(x, \cdot) dt$-equivalence for every $x \in \mathcal{X}$.

The Lyapunov condition can be stated now :

(H_L) There exist a measurable function $U : \mathcal{X} \to [1, +\infty)$ in $\mathbb{D}_{e,\mu}(\mathcal{L})$, a positive function ϕ and a constant $b > 0$ such that

$$-\frac{\mathcal{L}U}{U} \geq \phi - b, \ \mu\text{-a.s.}$$

When the process is irreducible and the constant b is replaced by $b1_C$ for some "small set" C, then it is well-known that the existence of a positive bounded ϕ such that $\inf_{\mathcal{X} \setminus C} \phi > 0$ in (H_L) is equivalent to Poincaré inequality (see [4, 5], for instance).

Lyapunov conditions are widely used to study the speed of convergence of Markov chains [41] or Markov processes [24, 23], large or moderate deviations and essential spectral radii [54, 27, 55] or sharp large deviations [32]. More recently, they have been used to study functional inequalities such as weak Poincaré inequality [5] or super-Poincaré inequality [16]. See Wang [49] on weak and super Poincaré inequalities.

For a given function f, let $K_\phi(f) \in [0, +\infty]$ be the minimal constant $C \in [0, +\infty]$ such that $|f| \leq C\phi$.

Theorem 4.1. *Assume the Lyapunov function condition* (H_L). *For* $g \in L_0^2(\mu)$, *if* $K_\phi(g^+) < +\infty$, *then the transportation-information inequality (2.6) holds with*

$$M = K_\phi(g^+)(bc_P + 1). \tag{4.1}$$

In particular Bernstein's inequality (1.6) holds with that constant M.

Proof. We are inspired by the elegant proof of Barthe-Bakry-Cattiaux-Guillin [4] for the Poincaré inequality. As before let $\nu = h^2\mu$ with $0 \leq h \in \mathbb{D}(\mathcal{E})$. For the term $B = \int_{\mathcal{X}} g[h - \mu(h)]^2 d\mu$ in (2.6) we have by (H_L),

$$B \leq K_\phi(g^+) \int_{\mathcal{X}} \phi[h - \mu(h)]^2 d\mu \leq K_\phi(g^+) \int_{\mathcal{X}} \left(b - \frac{\mathcal{L}U}{U}\right) [h - \mu(h)]^2 d\mu.$$

By a result in large deviations [28, Lemma 5.6], we have

$$\int_{\mathcal{X}} -\frac{\mathcal{L}U}{U} [h - \mu(h)]^2 d\mu \leq \mathcal{E}(h, h) = I.$$

Hence applying the Poincaré inequality, we get

$$B \leq K_\phi(g^+)(bc_P + 1) I$$

the desired result. $\qquad\square$

4.2. **Particular case : diffusions on** $\mathbb{R}^d$. Let $\mathcal{X} = \mathbb{R}^d$, $x \cdot y$ and $|x| = \sqrt{x \cdot x}$ be the Euclidean inner product and norm, respectively. Consider $\mathcal{L} = \Delta - \nabla V \cdot \nabla$ on $\mathbb{R}^d$, where V is lower bounded C^2-smooth such that $Z = \int_{\mathbb{R}^d} e^{-V} dx$ is finite. The corresponding semigroup P_t is symmetric on $L^2(\mu)$ for $\mu = \frac{1}{Z} e^{-V} dx$. From Theorem 4.1 we derive easily

Corollary 4.2. *In the framework above, let $\gamma > 0$ be some fixed constant. If one of the following conditions*

$$\exists a < 1, R, c > 0, \text{such that if } |x| > R, \qquad (1-a)|\nabla V|^2 - \Delta V \geq c\left(1 + |x|^\gamma\right) \qquad (4.2)$$

or

$$\exists R, c > 0, \text{such that } \forall |x| > R, \qquad |x|^{\gamma/2}\frac{x}{|x|} \cdot \nabla V(x) \geq c\left(1 + |x|^\gamma\right) \qquad (4.3)$$

is satisfied, then the Lyapunov function condition (H_L) is satisfied with $\phi(x) := c(1+|x|^\gamma)$, and then for any μ-centered function g such that $g(x) \leq C(1+|x|^\gamma)$, Bernstein's inequality (1.6) holds for some constant $M = M(g)$ given by (4.1).

Proof. Under (4.2), one takes $U = e^{aV}$; and under (4.3) one choose $U = e^{a|x|^{1+(\gamma/2)}}$ with small enough $a > 0$ (so that c may be arbitrary). One sees that condition (H_Γ) is satisfied in both cases. $\qquad\square$

Example 4.3. Let $V(x) = |x|^\beta$ ($\beta > 0$ is fixed) for $|x| > 1$ in the framework above.
  Case 1. $\beta \in (0, 1)$. In this case the Poincaré inequality does not hold (cf. [33]). And Bernstein's inequality (1.6) does not hold for **all** $g \in b\mathcal{B}$ (with $\mu(g) = 0$) as explained in the Introduction. Section 6 is devoted to such examples.
  Case 2. $\beta = 1$. For this exponential type's measure μ, the Poincaré inequality holds and one can apply Lezaud's result for bounded g. We do not believe that the Bernstein's inequality holds for unbounded g.
  Case 3. $\beta > 1$. Condition (4.3) is satisfied with $\gamma = 2(\beta - 1)$. Hence Bernstein's inequality (1.6) holds for μ-centered g such that $g \leq C(1 + |x|^{2(\beta-1)})$, in concordance with condition (3.13) in Example 3.9.

4.3. Particular case : birth-death processes. Let $\mathcal{X} = \mathbb{N}$ and

$$\mathcal{L}f(k) = b_k(f(k+1) - f(k)) + a_k(f(k-1) - f(k)), \ k \in \mathbb{N}$$

where $b_k > 0, k \geq 0$ are the birth rates, $a_k > 0, k \geq 1$ are the death rates respectively, and $f(-1) := f(0)$.
  We assume that the process is positive recurrent, i.e.,

$$\sum_{n\geq 0} \pi_n \sum_{i\geq n}(\pi_i b_i)^{-1} = \infty \quad \text{and} \quad C := \sum_{n=0}^{+\infty}\pi_n < +\infty,$$

where π_n is given by

$$\pi_0 = 1, \quad \pi_n = \frac{b_0 b_1 \cdots b_{n-1}}{a_1 a_2 \cdots a_n}, \quad n \geq 1$$

is an invariant measure of the process. Define the normalized probability μ of π by $\mu_n = \frac{\pi_n}{C}$ for any $n \geq 0$, which is actually the unique reversible invariant probability of the process.

Corollary 4.4. *Given a positive weight function ϕ_0 on $\mathbb{N}$ such that $\phi_0 \geq \delta > 0$. If there are some constant $\kappa > 1$ and some $N \geq 1$ so that*

$$a_n - \kappa b_n \geq \phi_0(n), \ n \geq N, \qquad (4.4)$$

then (H_Γ) holds with $\phi(n) := (1 - \kappa^{-1})\phi_0(n)$ (and some finite constant b). In particular the results in Theorem 4.1 holds true.

Proof. Let $U(n) = \kappa^n$, we have

$$-\frac{\mathcal{L}U}{U}(n) = \frac{\kappa-1}{\kappa}\left(a_n - \kappa b_n\right)$$

where it follows that $c_P < +\infty$ ([4, 5]) and so the desired result holds by Theorem 4.1. $\square$

Example 4.5. ($M/M/\infty$-**queue system**) Let $b_k = \lambda > 0$ $(k \geq 0)$ and $a_k = k$ $(k \geq 1)$. Then μ is the Poisson distribution with parameter λ. It is an ideal model for a queue system with a number of *serveurs* much larger than the number of clients. It is well known that $c_P = 1$ but the log-Sobolev inequality does not hold ([53]).

For $\phi_0(n) = n + \delta$ where $\delta > 0$ is fixed, taking $U(n) = \kappa^n$ $(\kappa > 1)$ as above and applying Theorem 4.1, we get by an optimization over $\kappa > 1$ that for all g so that $g \leq K(n+\delta)$ $(K > 0)$, $B \leq MI$ where

$$M = K[(\sqrt{\lambda}+1)^2 + \delta]. \tag{4.5}$$

Hence (2.6) and Bernstein's inequality (1.6) hold with such M. Notice that the growth of M for large λ is linear in λ.

An important observable is $g_0(n) = n - \lambda$ (then $L_t(g_0)$ is the difference between the mean number of clients in the queue system during time interval $[0, t]$ and the asymptotic mean λ). Since $(-\mathcal{L})^{-1}g_0 = g_0$, we have $\sigma^2(g_0) = 2\langle(-\mathcal{L})^{-1}g_0, g_0\rangle_\mu = 2\mathrm{Var}_\mu(g_0) = 2\lambda$. We want to get a better estimate of $M = M(g_0)$.

For $U(n) = \kappa^n$ $(\kappa > 0)$, we have

$$\left[\mathcal{L} + \frac{\kappa-1}{\kappa}g_0\right]U = \frac{(\kappa-1)^2}{\kappa}\lambda U.$$

In other words $0 < U \in L^2(\mu)$ is an eigenfunction of the Schrödinger operator $\mathcal{L} + \frac{\kappa-1}{\kappa}g_0$ with eigenvalue $\frac{(\kappa-1)^2}{\kappa}\lambda$. By Perron-Frobenius theorem and Raylaigh's principle,

$$\Lambda\left(\frac{\kappa-1}{\kappa}g_0\right) = \frac{(\kappa-1)^2}{\kappa}\lambda.$$

Thus if $s < 1$,

$$\Lambda(sg_0) = \frac{\lambda s^2}{1-s} = \frac{\sigma^2(g_0)s^2}{2(1-s)} \tag{4.6}$$

and then $\Lambda(sg_0) = +\infty$ for all $s \geq 1$ (by the convexity of $s \to \Lambda(sg_0)$).

By Theorem 2.2, for $g = g_0$, not only the Bernstein inequality (1.6) holds with the optimal constant $M(g_0) = 1$, and this inequality is itself sharp : indeed (4.6) implies by Proposition 2.1 and the large deviation lower bound in Wu [52, Theorem B.1],

$$\lim_{t\to\infty}\frac{1}{t}\log\mathbb{P}_\mu\left(\frac{1}{t}\int_0^t X_s ds > \lambda + r\right) = -\frac{r^2}{\lambda\left(\sqrt{1+\frac{r}{\lambda}}+1\right)^2}, \; r > 0.$$

The calculus above shows that the mean number of clients $\frac{1}{t}\int_0^t X_s ds$ does not possess any Poisson type's concentration inequality, contrary to the intuition that one might have for this standard process related with the Poisson measure.

5. A Lipschitzian approach

In this section we assume always the existence of the carré-du-champs operator Γ, i.e. (H_Γ) in §3. We suppose furthermore that $\Gamma = \Gamma_0 + \Gamma_1$ where $\Gamma_k : \mathbb{D}(\mathcal{E}^2) \to L^1(\mu)$, $k = 0, 1$ are both bilinear nonnegative definite forms, Γ_0 is a differentiation, Γ_1 is given by

$$\Gamma_1(f, g)(x) = \frac{1}{2} \int_{\mathcal{X}} (f(y) - f(x))(g(y) - g(x)) J(x, dy), \ f, g \in \mathbb{D}(\mathcal{E}).$$

Here Γ_0 corresponds to the continuous diffusion part of (X_t), and $J(x, dy)$ is a nonnegative jumps kernel (maybe σ-infinite) on $\mathcal{X}$ such that $J(x, \{x\}) = 0$ and $\mu(dx)J(x, dy)$ is symmetric on $\mathcal{X}^2$, describing the jumps rate of the process.

5.1. General result. Recall that $\Gamma(f) = \Gamma(f, f)$.

Theorem 5.1. *Assume that d is a lower semi-continuous metric on $\mathcal{X}$ (which does not necessarily generate the topology of $\mathcal{X}$), such that $\int_{\mathcal{X}} d(x, x_0)^2 d\mu(x) < +\infty$. Given $g \in L_0^2(\mu)$, let $G \in L_0^2(\mu) \bigcap \mathbb{D}_2(\mathcal{L})$ be the unique solution of the Poisson equation $-\mathcal{L}G = g$. If $\|\Gamma(G)\|_\infty < +\infty$, then the transportation-information inequality (2.6) holds with*

$$M = 2\sqrt{c_P\|\Gamma(G)\|_\infty}. \tag{5.1}$$

In particular Bernstein's inequality (1.6) holds with that constant M.

Proof. As before we may assume that $\nu = h^2\mu$ with $0 \le h \in \mathbb{D}(\mathcal{E}) \bigcap L^\infty(\mu)$. For the term $B = \int_{\mathcal{X}} g[h - \mu(h)]^2 d\mu$ in (2.7), setting $\tilde{h} = h - \mu(h)$ we write

$$B = \langle -\mathcal{L}G, \tilde{h}^2 \rangle_\mu = \int_{\mathcal{X}} \Gamma_0(G, \tilde{h}^2) d\mu + \int_{\mathcal{X}} \Gamma_1(G, \tilde{h}^2) d\mu.$$

For the Γ_0-term, we have

$$\int_{\mathcal{X}} \Gamma_0(G, \tilde{h}^2) d\mu \le \int_{\mathcal{X}} \sqrt{\Gamma_0(G)\Gamma_0(\tilde{h}^2)} d\mu = 2 \int_{\mathcal{X}} \sqrt{\Gamma_0(G)\tilde{h}^2\Gamma_0(h)} d\mu$$

The Γ_1-term above requires some more work. We proceed as follows.

$$\int_{\mathcal{X}} \Gamma_1(G, \tilde{h}^2) d\mu = \frac{1}{2} \iint_{\mathcal{X}^2} (G(y) - G(x))(\tilde{h}(y) + \tilde{h}(x))(\tilde{h}(y) - \tilde{h}(x)) \mu(dx)J(x, dy)$$

$$\le 2 \int_{\mathcal{X}} \mu(dx) \sqrt{\int_{\mathcal{X}} (\tilde{h}(y) - \tilde{h}(x))^2 \mu(dx)J(x, dy)}$$

$$\cdot \sqrt{\frac{1}{8} \int_{\mathcal{X}} (G(y) - G(x))^2 [\tilde{h}(y) + \tilde{h}(x)]^2 \mu(dx)J(x, dy)}.$$

Plugging those two estimates into the expression of B above, we get by Cauchy-Schwarz's inequality,

$$B \le 2 \sqrt{\int_{\mathcal{X}} \Gamma_0(G)\tilde{h}^2 d\mu + \frac{1}{8} \int_{\mathcal{X}} \int_{\mathcal{X}} (G(y) - G(x))^2 [\tilde{h}(y) + \tilde{h}(x)]^2 \mu(dx)J(x, dy)}$$

$$\cdot \sqrt{\int_{\mathcal{X}} (\Gamma_0(h) + \Gamma_1(h)) d\mu}.$$

The last factor is $\sqrt{I}$. Using the symmetry in (x, y) of $\mu(dx)J(x, dy)$ and $(a + b)^2 \leq 2(a^2 + b^2)$, the second term inside the first square root above can be bounded by

$$\frac{1}{4} \iint_{\mathcal{X}^2} (G(y) - G(x))^2 [\tilde{h}(y)^2 + \tilde{h}(x)^2] \mu(dx) J(x, dy)$$
$$= \frac{1}{2} \iint_{\mathcal{X}^2} (G(y) - G(x))^2 \tilde{h}(x)^2 \mu(dx) J(x, dy) = \int_{\mathcal{X}} \Gamma_1(G)(x) \tilde{h}(x)^2 \mu(dx).$$

Hence the sum inside the first square root above is not greater than $\int_{\mathcal{X}} \Gamma(G)(x) \tilde{h}(x)^2 \mu(dx)$. Thus we obtain

$$B = \int_{\mathcal{X}} g[h - \mu(h)]^2 d\mu \leq 2 \sqrt{\int_{\mathcal{X}} \Gamma(G)(x) \tilde{h}(x)^2 \mu(dx)} \cdot \sqrt{I}. \tag{5.2}$$

Now noting that $\int_{\mathcal{X}} \Gamma(G)(x) \tilde{h}(x)^2 \mu(dx) \leq \|\Gamma(G)\|_\infty \text{Var}_\mu(h) \leq c_P \|\Gamma(G)\|_\infty I$, we conclude that $B \leq 2\sqrt{c_P \|\Gamma(G)\|_\infty} I$, the desired result. $\qquad\square$

Some sharp estimates of $\|\Gamma(G)\|_\infty$ for diffusions are available : see Djellout and Wu [19] for one dimensional diffusions, and Wu [56] for elliptic diffusions on manifolds. Here we present examples of jumps processes.

5.2. **Birth-death processes continued.** The following two lemmas are taken from Liu and Ma [35].

Lemma 5.2. *Given a function g on $\mathbb{N}$ with $\mu(g) = 0$, consider the Poisson equation*

$$- \mathcal{L}G = g. \tag{5.3}$$

For any $k \geq 0$, the solution of the above equation (5.3) satisfies the following relation :

$$G(k + 1) - G(k) = - \frac{\sum_{j=0}^{k} \mu_j g(j)}{\mu_{k+1} a_{k+1}} = \frac{\sum_{j \geq k+1} \mu_j g(j)}{\mu_{k+1} a_{k+1}}. \tag{5.4}$$

Lemma 5.3. *Let $\rho : \mathbb{N} \to \mathbb{R}$ be an increasing function in $L^2(\mu)$. Provided that $\|g\|_{Lip(\rho)} := \sup_{k \in \mathbb{N}} \frac{|g(k+1) - g(k)|}{\rho(k+1) - \rho(k)} = 1$ with $\mu(g) = 0$, we have for any $k \geq 0$,*

$$\sum_{i \geq k} \mu_i g(i) \leq \sum_{i \geq k} \mu_i (\rho(i) - \mu(\rho)). \tag{5.5}$$

We can derive easily

Corollary 5.4. *Let $\rho : \mathbb{N} \to \mathbb{R}$ be an increasing function in $L^2(\mu)$. If*

$$K := \frac{1}{2} \sup_{n \geq 0} \left(1_{n \geq 1} \frac{1}{a_n \mu_n^2} \left[\sum_{i \geq n} \mu_i (\rho(i) - \mu(\rho)) \right]^2 + \frac{1}{b_n \mu_n^2} \left[\sum_{i \geq n+1} \mu_i (\rho(i) - \mu(\rho)) \right]^2 \right) \tag{5.6}$$

is finite, then for every g with $\mu(g) = 0$ and $\|g\|_{Lip(\rho)} < +\infty$, the transportation inequality (2.6) holds with $M = 2\sqrt{c_P K} \|g\|_{Lip(\rho)}$.

Proof. By Lemmas 5.2 and 5.3, the solution G of $-\mathcal{L}G = g$ satisfies $\|\Gamma(G)\|_\infty \leq K \|g\|_{Lip(\rho)}^2$ (using $a_{n+1} \mu_{n+1} = b_n \mu_n$). It remains to apply Theorem 5.1. $\qquad\square$

See [35] for convex concentration inequalities. Though we can give many examples to which Corollary 5.4 applies, we want to look at the $M/M/\infty$ queue system again.

Example 5.5. ($M/M/\infty$ **queue, continued**) The constant K in (5.6) above is infinite for $\rho(n) = n$, but finite for $\rho(n) = \sum_{k=0}^{n} 1/\sqrt{k+1}$ (a quite artificial choice). *What happens for $\rho(n) = \rho_0(n) := n$?* (In that case $\|g\|_{Lip(\rho_0)} =: \|g\|_{Lip}$ is the Lipschitzian coefficient w.r.t. the Euclidean metric.)

A crucial feature of this model is the commutation relation $DP_t = e^{-t}P_t D$ where $Df(n) := f(n+1) - f(n)$, a property shared by Ornstein-Uhlenbeck process for $D = \nabla$. From this fact one sees that

$$\|(-\mathcal{L})^{-1}g\|_{Lip} \leq \|g\|_{Lip}.$$

Then if $\|g\|_{Lip} \leq 1$, $G = (-\mathcal{L})^{-1}g$ satisfies

$$\Gamma(G)(n) = \frac{1}{2}\left(\lambda[G(n+1) - G(n)]^2 + n[G(n-1) - G(n)]^2\right) \leq \frac{1}{2}(\lambda + n).$$

Applying (5.2) in the proof of Theorem 5.1, we get by (4.5)

$$B \leq \sqrt{2\int_{\mathbb{N}}(\lambda + n)\tilde{h}^2\mu(dn)}\sqrt{I} \leq \sqrt{2[(\sqrt{\lambda}+1)^2 + \lambda]}\,I.$$

Thus we have proven

Corollary 5.6. *For the $M/M/\infty$ queue, if the Lipschitzian norm $\|g\|_{Lip}$ of g w.r.t. the Euclidean metric is finite (and $\mu(g) = 0$), then (2.6) and Bernstein's inequality (1.6) hold with*

$$M = \|g\|_{Lip}\sqrt{2\left[(\sqrt{\lambda}+1)^2 + \lambda\right]}.$$

6. The subgeometric case

6.1. **General result.** In this last section, we will suppose no more that a Poincaré inequality holds, and inspired by the Lyapunov function approach, we introduce a more classical version of Lyapunov condition

(H_{LC}) there exist a continuous function $U : \mathcal{X} \to [1, +\infty)$ in $\mathbb{D}_{e,\mu}(\mathcal{L})$, a measurable positive function ϕ, a set $C \in \mathcal{B}$ with $\mu(C) > 0$ and constant $b > 0$ such that

$$-\frac{\mathcal{L}U}{U} \geq \phi - b\mathbb{1}_C, \ \mu\text{-a.s.}$$

In our mind ϕ goes to 0 at infinity in this section.

We will also assume that a local Poincaré inequality holds for the set C in (H_{LC}): there exists some constant κ_C such that for all $g \in \mathbb{D}(\mathcal{E})$ such that $\mu(g\mathbb{1}_C) = 0$

$$\mu(g^2\mathbb{1}_C) \leq \kappa_C\mathcal{E}(g, g). \tag{6.1}$$

Note that for diffusions on $\mathbb{R}^d$, C is often a ball $B(0, R)$ and the local Poincaré inequality may then be easily deduced from the local Poincaré inequality for the Lebesgue measure on balls by a perturbation argument.

Theorem 6.1. *Assume the Lyapunov function condition (H_{LC}) and the local Poincaré inequality (6.1) for the set C. For $g \in L_0^2(\mu)$ such that $\sigma^2(g)$ is finite, if $K_\phi(g^+) < +\infty$, then the transportation-information inequality (2.6) holds with*

$$M = K_\phi(g^+)(b\kappa_C + 1). \tag{6.2}$$

In particular Bernstein's inequality (1.6) holds with that constant M.

Proof. In fact we have to slightly modify the key approach described in section 2: for a constant $c > 0$ to be chosen later,

$$\nu(g) = \int_{\mathcal{X}} gh^2 d\mu = \int_{\mathcal{X}} g([h-c]^2 + 2ch)d\mu$$
$$= 2c\langle g, h\rangle_\mu + \int_{\mathcal{X}} g[h-c]^2 d\mu =: A + B. \tag{6.3}$$

For the first term $A = 2c\langle g, h\rangle_\mu$, since $\sigma^2 = \sigma^2(g)$ is assumed to be finite, we have by Remark 2.4 that $|A| \le c\sqrt{2\sigma^2 I}$.

Let consider now the second term

$$B = \int_{\mathcal{X}} g[h-c]^2 d\mu \le \int_{\mathcal{X}} g^+[h-c]^2 d\mu \le K_\phi(g^+)\int_{\mathcal{X}}\left(b1_C - \frac{\mathcal{L}U}{U}\right)[h-c]^2 d\mu.$$

By a result in large deviations [28, Lemma 5.6], we have

$$\int_{\mathcal{X}} -\frac{\mathcal{L}U}{U}[h-c]^2 d\mu \le \mathcal{E}(h, h) = I.$$

For the other term we apply the local Poincaré inequality, valid if we consider $c = \mu(h1_C)$ which leads to

$$B \le K_\phi(g^+)(b\kappa_C + 1) I.$$

Remark finally that $c = \mu(h1_C) \le 1$. $\square$

Now we present an easy sufficient condition for the finiteness of $\sigma^2(g)$ (and then for the CLT by Remark 2.4) by following Glynn and Meyn [25], which has its own interest.

Lemma 6.2. *Suppose that $R_1 = \int_0^\infty e^{-t}P_t dt$ is μ-irreducible (i.e. $\mu \ll R_1(x, \cdot)$ for every $x \in \mathcal{X}$) and Harris positive recurrent ([41]). Assume that there are*
  • *a (Lyapunov) continuous function $W : \mathcal{X} \to [1, +\infty)$ in the extended domain $\mathbb{D}_e(\mathcal{L})$ (see §4.1),*
  • *a measurable function $F : \mathcal{X} \to (0, +\infty)$,*
  • *a R_1-small set C with $\mu(C) > 0$, i.e. $R_1(x, A) \ge \delta\nu(A)$ for all $x \in C, A \in \mathcal{B}$ for some constant $\delta > 0$ and $\nu \in \mathcal{M}_1(\mathcal{X})$,*
  • *and a positive constant b*
such that W is bounded on C and

$$\mathcal{L}W \le -F + b1_C. \tag{6.4}$$

If $|g| \le cF$ for some constant $c > 0$ and $\mu(g) = 0$, then
  (1) *There exists some measurable function G such that $|G| \le cW$ for some constant $c > 0$, such that for any $t > 0$, $\int_0^t P_s|g|ds < +\infty$ and $P_tG - G = -\int_0^t P_sgds$ everywhere on $\mathcal{X}$ (in such case we say that G belongs to the extended domain in the strong sense $\mathbb{D}_s(\mathcal{L})$ of $\mathcal{L}$ and write $-\mathcal{L}G = g$).*

(2) *If furthermore $g \in L_0^p(\mu)$ and $W \in L^q(\mu)$ where $p \in [2, +\infty]$ and $1/p + 1/q = 1$, then $\sigma^2(g)$ is finite.*

Its proof is postponed to the Appendix.

6.2. **Particular case: diffusions on $\mathbb{R}^d$.** We study here the diffusion in $\mathbb{R}^d$ with generator $\mathcal{L} = \Delta - \nabla V \cdot \nabla$ and $\mu = e^{-V} dx/Z$, presented in Section 4. The first thing to remark is that any compact set is a small set, and thus balls are small sets. A local Poincaré inequality is then available. We then have

Corollary 6.3. *Suppose that there exists a positive and bounded function $\tilde{\phi}$ such that*

$$\exists a < 1, R, c > 0, \text{such that if} \ \ |x| > R, \qquad (1-a)|\nabla V|^2 - \Delta V \geq \tilde{\phi}(x). \qquad (6.5)$$

Then the weak Lyapunov condition (H_{LC}) is satisfied with $U = e^{aV}$ with $\phi = a\tilde{\phi}$ and $C = B(0, R)$; and if $\int e^{(a-1)V} dx < +\infty$ (i.e. $\mu(U) < +\infty$), then for any μ centered bounded function g such that $|g| \leq c_1 \tilde{\phi} U$ and $g(x) \leq c_2 \tilde{\phi}$ for some positive constants c_1, c_2, the asymptotic variance $\sigma^2(g)$ is finite by Lemma 6.2 and Bernstein's inequality holds.

Note that, in parallel to the second condition of Corollary 4.2, one may also consider Lyapunov function of the form $U(|x|)$, but the result is then not as explicit and we prefer to illustrate such an approach through examples.

Example 6.4. (sub-exponential measure) Let $V(x) = |x|^\beta$ (if $|x| > 1$) for $\beta \in (0, 1)$ such that no Poincaré inequality holds. However, one may apply the previous corollary with $U(x) = e^{a|x|^\beta}$ and $\tilde{\phi}(x) = (1 - a - \delta)\beta^2(1 + |x|)^{2(\beta-1)}$ ($a, \delta \in (0, 1)$, $a + \delta < 1$). Hence by Corollary 6.3, Bernstein's inequality holds for μ centered bounded function g such that for large $|x|$, $g(x) \leq c/(1 + |x|)^{2(1-\beta)}$.

Example 6.5. (Cauchy type measure) Let $V(x) = \frac{1}{2}(d + \beta) \log(1 + |x|^2)$ for $\beta > 0$. The condition (H_{LC}) holds with $U = e^{aV} = (1 + |x|^2)^{a(d+\beta)/2}$ and $\tilde{\phi}(x) = c/(1 + |x|^2)$ for some constant $c > 0$, where $a \in (0, 1)$ so that $(1 - a)(d + \beta) > d$ (for $\mu(U) < +\infty$). So Bernstein's inequality holds for μ centered bounded function g such that for large $|x|$, $g(x) \leq K/(1 + |x|^2)$ for some constant $K > 0$, by Corollary 6.3.

Remark 6.6. One may be surprised that the upper bound for the test function is the same for every Cauchy type measure. One may find the beginning of an answer in recent results of Bobkov-Ledoux [11] (see also Cattiaux-Gozlan-Guillin-Roberto [13]). Indeed, in their work they prove that this type of measures satisfy a weighted Poincaré type inequality where the weight is the same for every Cauchy-type measure.

6.3. **Particular case : birth-death processes.** We adopt here the notations of subsection 4.3, and assume once again that the process is positive recurrent. We suppose for simplicity that for large enough n, the death rate a_n is larger than the birth rate b_n.

Corollary 6.7. *If there are $m > 0$, $N \geq 1$ and a positive sequence $(c_n)_{n \in \mathbb{N}}$ such that*
   (1) for all $n \geq N$, $a_n - b_n \geq c_n > 0$;
   (2) $\sum_n n^m \mu_n < +\infty$,
then Bernstein's inequality is valid for every μ centered bounded function g such that for large n, $|g(n)| \leq cn^{m-1} c_n$ and $g(n) \leq Kc_n/n$ for some constants $c, K > 0$.

Proof. Let $U(n) = (1 + n)^m$, then for large n,

$$-\frac{\mathcal{L}U(n)}{U(n)} \geq m(a_n - b_n)\left(\frac{1}{n} + o(\frac{1}{n})\right).$$

Hence the Lyapunov condition (H_{LC}) holds for $\phi(n) = (m-\delta)c_n/(1+n)$ where $\delta \in (0, m)$. The local Poincaré inequality is always valid in this context and a precise estimation of the constant may be found in Chen [17]. Since $\mu(U)$ is finite, we can apply Lemma 6.2 to conclude that $\sigma^2(g) < +\infty$ for $|g| \leq c\phi U$. It remains to apply Theorem 6.1.

$\square$

Example 6.8. Let $b_n \equiv 1$ and $a_n = 1 + a/(n+1)$ where $a > 0$. Then $c_n := a_n - b_n = a/(n+1)$ and π_n behaves as $\frac{1}{n^a}$ for large n. Thus the process is positive recurrent if and only if $a > 1$. For $a > 1$, take $m \in (0, a-1)$, we see that the conditions in Corollary 6.7 are all satisfied. Hence Bernstein's inequality holds for μ-centered g such that $|g(n)| \leq K/n^2$ for large n. This is quite similar as in the Cauchy measure case.

7. APPENDIX

Proof of Lemma 6.2. Let us first prove part (2) by admitting part (1). Let G be the strong solution of $-\mathcal{L}G = g$ given in part (1). Since $W \in L^q(\mu)$, considering $G - \mu(G)$ if necessary we may and will assume that $\mu(G) = 0$. Now for any $\varepsilon > 0$, let $R_\varepsilon = \int_0^\infty e^{-\varepsilon t}P_t dt = (\varepsilon - \mathcal{L})^{-1}$ be the resolvent. By the resolvent equation, $G - R_\varepsilon g = \varepsilon R_\varepsilon G$ which tends to $\mu(G) = 0$ in $L^q(\mu)$ as $\varepsilon \to 0$ by the ergodic theorem, we have

$$\lim_{\varepsilon \to 0}\langle R_\varepsilon g, g\rangle_\mu = \int Gg d\mu < +\infty.$$

This relation yields that $\sigma^2(g)$ in (1.7) exists and $\sigma^2(g) = 2\int Gg d\mu$ (in the actual symmetric case).

We turn now to prove part (1). This is due to Glynn and Meyn [25, Theorem 3.2] when F is bounded from below by a positive constant. Let us modify slightly their proof for the general case.

Step 1 (Reduction to the discrete time case). At first since $e^{-bt}W(X_t)$ is a local super-martingale, then a super-martingale, so $P_t W \leq e^{bt}W$ for all $t \geq 0$. Moreover for any $\lambda > 0$, by Itô's formula,

$$M_t = e^{-\lambda t}W(X_t) - W(X_0) + \int_0^t e^{-\lambda s}(\lambda W - \mathcal{L}W)(X_s)ds$$

is a $\mathbb{P}_x$-local martingale for every $x \in \mathcal{X}$. Hence taking a sequence of stopping times (τ_n) increasing to $+\infty$ such that $\mathbb{E}^x M_{\tau_n} = 0$, we have for every $x \in \mathcal{X}$,

$$\mathbb{E}^x \int_0^{\tau_n} e^{-\lambda s}(\lambda W + F - b1_C)(X_s)ds \leq \mathbb{E}^x \int_0^{\tau_n} e^{-\lambda s}(\lambda W - \mathcal{L}W)(X_s)ds \leq W(x).$$

Letting n go to infinity, we obtain by monotone convergence

$$\lambda R_\lambda W + R_\lambda F \leq W + bR_\lambda 1_C.$$

Consider the Markov kernel $Q = R_1$. The relation above says that

$$QW \leq W - QF + bQ1_C. \tag{7.1}$$

Assume that one can prove that there is G such that $|G| \leq cW$ (for some constant $c > 0$) such that

$$(1 - Q)G = Qg. \tag{7.2}$$

Then $G = R_1(G + g) \in \mathbb{D}_s(\mathcal{L})$ and $R_1(-\mathcal{L})G = (1 - R_1)G = R_1 g$. Consequently $-\mathcal{L}G = (1 - \mathcal{L})R_1(-\mathcal{L})G = (I - \mathcal{L})R_1 g = g$, the desired claim in part (1).

Therefore it remains to solve (7.2) under the condition (7.1).

Step 2 (atom case). Let us suppose at first that the small set C in (7.1) is an atom of Q, i.e., $Q(x, \cdot) = Q(y, \cdot)$ for all $x, y \in C$. In this case one solution to (7.2) is given by

$$G(x) = \mathbb{E}^x \sum_{k=0}^{\sigma_C} Qg(Y_k) \tag{7.3}$$

where $(Y_n)_{n \geq 0}$ is the Markov chain with transition probability kernel Q defined on $(\Omega, (\mathcal{F}_n), \mathbb{Q}_x)$ equipped with the shift θ (so that $Y_n(\theta\omega) = Y_{n+1}(\omega)$), $\sigma_C = \inf\{n \geq 0; \ Y_n \in C\}$.

To justify this fact which is one key in [25], notice

1) G given by (7.3) is well defined. In fact $|Qg| \leq cQF$. Using the condition (7.1) and the fact that

$$W(Y_n) - W(Y_0) + \sum_{k=0}^{n-1}(W - QW)(Y_k)$$

is a $\mathbb{Q}_x$-martingale, we obtain the following at first for $\sigma_C \wedge n$ and then for σ_C (by letting $n \to \infty$)

$$\mathbb{E}^x \sum_{0 \leq k \leq \sigma_C - 1} QF(Y_k) \leq b\mathbb{E}^x \sum_{0 \leq k \leq \sigma_C - 1} Q1_C(Y_k) + W(x)$$

$$= b\mathbb{E}^x \sum_{1 \leq k \leq \sigma_C} 1_C(Y_k) + W(x) \leq b + W$$

where the second equality for $\sigma_C \wedge n$ (instead of σ_C) follows by Doob's stopping time theorem. Consequently

$$\mathbb{E}^x \sum_{0 \leq k \leq \sigma_C} QF(Y_k) \leq \sup_{x \in C} QF(x) + \mathbb{E}^x \sum_{k=0}^{\sigma_C - 1} QF(Y_k) \leq \sup_{x \in C} QF(x) + b + W(x).$$

By (7.1), $QF \leq W + b$ is bounded on C. Therefore G is well defined and $|G| \leq c(b' + W)$.

2) Let $\tau_C := \inf\{n \geq 1; Y_n \in C\}$. We have $\sigma_C \circ \theta = \tau_C - 1$ on $[\sigma_C = 0]$ and $\sigma_C \circ \theta = \sigma_C - 1$ on $[\sigma_C \geq 1]$. Hence for $x \in C$

$$QG(x) = \mathbb{E}^x \sum_{k=0}^{\sigma_C \circ \theta} Qg(Y_{k+1}) = \mathbb{E}^x \sum_{k=1}^{\tau_C} Qg(Y_k)$$

which is constant on $x \in C$ and equals to $\mu(g)/\mu(C) = 0$, then $G(x) - QG(x) = G(x) = Qg(x)$ for $x \in C$. Now for $x \notin C$,

$$QG(x) = \mathbb{E}^x \sum_{k=0}^{\sigma_C \circ \theta} Qg(Y_{k+1}) = \mathbb{E}^x \sum_{k=0}^{\sigma_C - 1} Qg(Y_{k+1}) = G(x) - Qg(x).$$

So $G - QG = Qg$ everywhere on $\mathcal{X}$.

Step 3 (non-atom case). In the non-atom case one can consider the splitting chain in [25, Proof of Theorem 2.3] to reduce the problem to the atom case. $\qquad\square$

References

[1] R. Adamczak, A tail probability for suprema of unbounded empirical processes with applications to Markov chains. *Elect. J. Proba. 13 (2008), 1000-1034.*

[2] C. Ané &al. ,*Sur les ingalités de Sobolev logarithmiques.* Panoramas et Synthèses 10. Société Mathématique de France, Paris, 2000.

[3] D. Bakry. L'hypercontractivité et son utilisation en théorie des semigroupes. In *Ecole d'Eté de Probabilités de Saint-Flour (1992)*, number 1581 in Lecture Notes in Mathematics. Springer-Verlag, 1994.

[4] F. Barthe, D. Bakry, P. Cattiaux, and A. Guillin. A simple proof of the Poincar inequality for a large class of probability measures including the log-concave case. *Electron. Commun. Probab. 13 (2008), 60–66.*

[5] D. Bakry, P. Cattiaux, and A. Guillin. Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal. 254 (2008), no. 3, 727–759.*

[6] F. Barthe, P. Cattiaux and C. Roberto, Interpolated inequalities between exponential and Gaussian, Orlicz hypercontractivity and isoperimetry. *Rev. Mat. Iberoam. 22 (2006), no. 3, 993–1067.*

[7] Y. Baraud, A Bernstein-type inequality for suprema of random processes with an application to statistics. See ArXiv. 2009.

[8] P. Bertail and Clémençon, Sharp bounds for the tails of functionals of Markov chains. *Prépublication Equippe MODAL'X, Université Paris Ouest. No. 2008-21.*

[9] S.G. Bobkov, I. Gentil, and M. Ledoux. Hypercontractivity of Hamilton-Jacobi equations. *Journal de Mathématiques Pures et Appliquées*, 80(7):669–696, 2001.

[10] S.G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *Journal of Functional Analysis.*, 163:1–28, 1999.

[11] S.G. Bobkov and M. Ledoux, Weighted Poincar-type inequalities for Cauchy and other convex measures. *Ann. Prob..*, 37:403–427 (2009).

[12] F. Bolley and C. Villani. Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. *Annales de la Faculté des Sciences de Toulouse.*, 14:331–352, 2005.

[13] P. Cattiaux, N. Gozlan, A. Guillin, C. Roberto. Functional inequalities for heavy tails distributions and application to isoperimetry. Preprint, (2008).

[14] P. Cattiaux and A. Guillin. On quadratic transportation cost inequalities. *J. Math. Pures Appl. (9) 86 (2006), no. 4, 341–361.*

[15] P. Cattiaux and A. Guillin. Deviation bounds for additive functionals of Markov processes. *ESAIM Probab. Stat. 12 (2008), 12–29 (electronic).*

[16] P. Cattiaux, A. Guillin, L. Wu, and F.Y. Wang. Lyapunov conditions for super Poincar inequalities. *J. Funct. Anal. 256 (2009), no. 6, 1821–1841.*

[17] M.F. Chen. *Eigenvalues, inequalities, and ergodic theory.* Probability and its Applications. Springer-Verlag, 2005.

[18] H. Djellout, A. Guillin, and L. Wu. Transportation cost-information inequalities for random dynamical systems and diffusions. *Annals of Probability*, 32(3B):2702–2732, 2004.

[19] H. Djellout, L. Wu. Lipschitzian Lipschitzian norm estimate of one-dimensional Poisson equations and applications. To appear in *Ann. Inst. H. Poinc.*, 2007.

[20] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluations of certain Markov process expectations for large time, I. *Comm. Pure Appl. Math.*, 28:1–47, 1975.

[21] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluations of certain Markov process expectations for large time, III. *Comm. Pure Appl. Math.*, 29:389–461, 1976.

[22] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluations of certain Markov process expectations for large time, IV. *Comm. Pure Appl. Math.*, 36:183–212, 1983.

[23] R. Douc, G. Fort, and A. Guillin. Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stochastic Process. Appl. 119 (2009), no. 3, 897–923.*

[24] D. Down, S.P. Meyn, R.L. Tweedie,Exponential and uniform ergodicity of Markov processes. *Ann. Probab*, 23(4):1671–1691, 1995.

[25] P.W. Glynn and S.P. Meyn, A Lyapunov bound for solutions of the Poisson equation, *Ann. Probab. 1996, Vol. 24, No. 2, 916-931.*

[26] N. Gozlan and C. Léonard. A large deviation approach to some transportation cost inequalities. *Probab. Theory Related Fields 139 (2007), no. 1-2, 235–283.*

[27] A. Guillin Moderate deviations of inhomogeneous functionals of Markov processes and application to Averaging.*Stoch. Proc. Appl.,* 92(2):287–313, 2001.

[28] A. Guillin, Ch. Léonard, L. Wu and N. Yao, Transport-information inequalities for Markov processes (I), *Proba. Th. Rel. Fields (2009), 144: 669-695.*

[29] A. Guillin, Ch. Léonard, L. Wu and F.Y. Wang, Transportation-information inequalities for Markov processes (II) : relations with other functional inequalities. *Preprint 08. ArXiv.*

[30] A. Guillin, A. Joulin, Ch. Léonard and L. Wu, Transport-information inequalities for Markov processes (III): jumps processes case. Preprint 2008.

[31] C. Kipnis and S.R.S. Varadhan, Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions, *Comm. Math. Phys. 104, p1-19, 1986.*

[32] I. Kontoyiannis, S.P. Meyn, Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. Electron. J. Probab. 10 (2005), no. 3, 61–123

[33] M. Ledoux. *The Concentration of Measure Phenomenon.* Mathematical Surveys and Monographs 89. American Mathematical Society, Providence RI, 2001.

[34] P. Lezaud. Chernoff and Berry-Esséen inequalities for Markov processes. *ESAIM, Probab. Statist.,* pages 183–201, 2001. (electronic).

[35] W. Liu and Y. Ma. Spectral gap and convex concentration inequalities for birth-death processes. *Ann. Inst. Henri Poincar Probab. Stat. 45 (2009), no. 1, 58–69.*

[36] K. Marton. Bounding $\bar{d}$-distance by informational divergence: a way to prove measure concentration. *Annals of Probability*, 24:857–866, 1996.

[37] K. Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.,* 6:556–571, 1997.

[38] P. Massart. *Concentration inequalities and model selection.* Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003. Lecture Notes in Mathematics, 1896. Springer, Berlin, 2007.

[39] F. Merlevède, M. Peligrad and E. Rio, A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Preprint 09.*

[40] F. Merlevède, M. Peligrad and E. Rio, Bernstein inequality and moderate deviations under strong mixing conditions. *Preprint 09.*

[41] S.P. Meyn and R.L. Tweedie. *Markov chains and stochastic stability.* Communications and Control Engineering Series. Springer-Verlag, 1993.

[42] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *Journal of Functional Analysis*, 173:361–400, 2000.

[43] M.M. Rao and Z.D. Ren, *Theory of Orlicz Space*, Marcel Dekker, Inc. New York, 1991.

[44] E. Rio, *Théorie asymptotique des processus aléatoires faiblement dépendentes.* Mathématiques et Applications **31**. Springer. Berlin.

[45] L. Saloffe-Coste, *Aspects of Sobolev-Type Inequlities*, London Mathmatical Society Lecture Note Series, 289.

[46] M. Talagrand. Transportation cost for gaussian and other product measures. *Geometric and Functional Analysis*, 6:587–600, 1996.

[47] C. Villani. *Optimal transport, old and new.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 338. Springer-Verlag, Berlin, 2009.

[48] C. Villani. *Topics in Optimal Transportation.* Graduate Studies in Mathematics 58. American Mathematical Society, Providence RI, 2003.

[49] F.Y. Wang. *Functional inequalities, Markov Semigroup and Spectral Theory.* Chinese Sciences Press, Beijing, New York, 2005.

[50] L.M. Wu, Moderate deviations of dependent random variables related to CLT, *Ann. Probab. Vol.23, 420-445 (1995).*

[51] L. Wu. A deviation inequality for non-reversible Markov processes. *Ann. Inst. Henri Poincaré (série Probab. Stat.),* 36:435–445, 2000.

[52] L. Wu. Uniformly integrable operators and large deviations for Markov processes. *J. Funct. Anal*, 172:301–376, 2000.

[53] L. Wu, A new modified logarithmic Sobolev inequality for Poisson point processes and several applications. *Probab. Th. Rel. Fields* **118**,*p.427-438* (2000)

[54] L. Wu. Large and moderate deviations for stochastic damping Hamiltonian systems. *Stoch. Proc. Appl.*, 91:205–238, 2001.

[55] L. Wu. Essential spectral radius for Markov semigroups (I) : discrete time case. *Probab. Th. Rel. Fields*, 128:255–321, 2004.

[56] L. Wu, Gradient estimates of Poisson equations on Riemannian manifolds and applications, *J. Funct. Anal. 257 (2009), 4015-4033.*

FUQING GAO. SCHOOL OF MATHEMATICS AND STATISTICS, WUHAN UNIVERSITY, 430072 HUBEI, CHINA
E-mail address: `fqgao@whu.edu.cn`

ARNAUD GUILLIN. LABORATOIRE DE MATHÉMATIQUES APPLIQUÉES, CNRS-UMR 6620, UNIVERSITÉ BLAISE PASCAL, 63177 AUBIÈRE, FRANCE.
E-mail address: `Arnaud.Guillin@math.univ-bpclermont.fr`

LIMING WU. LABORATOIRE DE MATHÉMATIQUES APPLIQUÉES, CNRS-UMR 6620, UNIVERSITÉ BLAISE PASCAL, 63177 AUBIÈRE, FRANCE. AND INSTITUTE OF APPLIED MATHEMATICS, CHINESE ACADEMY OF SCIENCES, 100190 BEIJING, CHINA.
E-mail address: `Li-Ming.Wu@math.univ-bpclermont.fr`