Electronic Journal of Probability

Electron. J. Probab. **18** (2013), no. 65, 1–26.
ISSN: 1083-6489 DOI: 10.1214/EJP.v18-2425

Measure concentration through non-Lipschitz observables and functional inequalities

Arnaud Guillin* Aldéric Joulin†

Abstract

Non-Gaussian concentration estimates are obtained for invariant probability measures of reversible Markov processes. We show that the functional inequalities approach combined with a suitable Lyapunov condition allows us to circumvent the classical Lipschitz assumption of the observables. Our method is general and offers an unified treatment of diffusions and pure-jump Markov processes on unbounded spaces.

Keywords: Concentration; invariant measure; reversible Markov process; Lyapunov condition; functional inequality; carré du champ; diffusion process; jump process.
AMS MSC 2010: 46E35, 60E15, 60J27, 60J60, 60K35.
Submitted to EJP on November 8, 2012, final version accepted on June 7, 2013.

1 Introduction

In the last few decades, the concentration of measure phenomenon has attracted a lot of attention. Given a metric probability space $(\mathcal{X}, d, \mu)$ and a sufficiently large class of functions defined on this space (we call them observables), the concentration of measure occurs when, observed through these functions, the space seems to be actually smaller than it is. In other words, there exists a non-decreasing continuous function $\alpha : [0, \infty) \to [0, \infty)$, null at the origin and tending to infinity at infinity, such that for a given class $\mathcal{C}$ of observables $f : \mathcal{X} \to \mathbb{R}$,

$$\mu\left(\left\{x \in \mathcal{X} : f(x) - \int_{\mathcal{X}} f\, d\mu > r\right\}\right) \quad \leq \quad \exp\left(-\alpha(r)\right), \quad r \geq 0.$$

The concentration is said to be Gaussian when α is quadratic-like. In connection with isoperimetry theory, the class $\mathcal{C}$ is usually taken to be the space of Lipschitz functions on $(\mathcal{X}, d, \mu)$, say $\mathrm{Lip}(\mathcal{X})$. A good review on the subject is the monograph of Ledoux [35] where the interested reader will find a clear introduction to the topic. One may mention

*Université Blaise Pascal, Clermont-Ferrand II and Institut Universitaire de France (IUF), France.
E-mail: guillin@math.univ-bpclermont.fr
†Institut de Mathématiques de Toulouse, Université de Toulouse, France.
E-mail: ajoulin@insa-toulouse.fr

also the recent progress in the area through mass transportation techniques. See the survey of Gozlan and Léonard [23].

In this paper, we emphasize a dynamical point of view on concentration of measure. Given the invariant measure μ of an ergodic and reversible continuous-time Markov process $(X_t)_{t\geq 0}$ with *carré du champ* operator Γ (see below for the definition), we provide concentration properties of μ through observables which depend on the dynamics. As it is sometimes the case in previous studies, our starting point is to assume that the pair (μ, Γ) satisfies a convenient functional inequality such as Poincaré or the entropic inequality. We refer to the notes of Ledoux [34] for precise credit and historical references for this large body of work. Such functional inequalities, which are verified by a wide variety of examples, are closely related to the long-time behaviour of the process. In particular, this approach offers an unified treatment of Markov processes on continuous and discrete space settings even if, in essence, these two situations are rather different from each other. In both cases the carré du champ refers to a natural distance related to the dynamics and, within this notion of distance, the Lipschitz observables under which Gaussian concentration estimates are obtained are the ones with a bounded carré du champ, that is to say the space $\mathrm{Lip}_\Gamma(\mathcal{X})$ of functions f such that $\Gamma(f,f)$ is bounded. However it is quite common in applications to need a control of the concentration through non-Lipschitz observables. Then a natural question arises: which type of measure concentration can we obtain beyond the space $\mathrm{Lip}_\Gamma(\mathcal{X})$? In particular in discrete space settings, such a study basically makes sense on an unbounded state space $\mathcal{X}$. Using the notion of Ricci curvature for Markov chains (the so-called Wasserstein curvature in continuous-time), a first result of this kind was given by Ollivier [37], in which he obtains concentration bounds involving a mixed Gaussian-exponential regime, i.e. $\alpha(r)$ is quadratic/linear for small/large deviation level r. In our language, he requires that the carré du champ $\Gamma(f,f)$ belongs to the space $\mathrm{Lip}(\mathcal{X})$. Despite this interesting and new result, which is sufficiently robust to be extended to additive functionals, see e.g. [30] and [31], it seems to the authors that there is no satisfactory treatment yet to this question and we hope to give (the beginning of) an answer to this problem with the present article.

Our idea is to use a Lyapunov condition on the observables. Namely we will consider the class $\mathcal{L}_V(a,b)$ of observables f such that

$$\Gamma(f,f) \leq -a\,\frac{\mathcal{L}V}{V} + b,$$

where a,b are two positive constants and V is a convenient test function. Our Lyapunov condition is somewhat different from the classical ones which have been successfully used for proving various types of functional inequalities, cf. [2, 13, 14, 25] and for concentration estimates of additive functionals, see for instance [12, 22, 17], since it applies directly on the observables. When a vanishes the class $\mathcal{L}_V(a,b)$ reduces to the space $\mathrm{Lip}_\Gamma(\mathcal{X})$ and the classical concentration results apply, cf. [34]. In particular, the behaviour of the carré du champ $\Gamma(f,f)$ depends now on the growth of the term $-\mathcal{L}V/V$, which has no reason to be bounded. Certainly, there is a price to pay for such an improvement: the concentration for large deviation level r is no longer Gaussian but only of exponential type under this class of observables.

The paper is organized as follows. In Section 2, we recall some basic material on reversible Markov processes and functional inequalities. Two types of processes are considered in our study: diffusions and pure-jump Markov processes. Next we state in Section 3 our main results of the paper, Theorems 3.6 and 3.12, in which some mixed Gaussian-exponential concentration properties of μ are obtained through observables satisfying the Lyapunov condition above and under the assumption of a convenient functional inequality satisfied by the dynamics (μ, Γ). As a result, such new concentration inequalities extend the classical estimates obtained when the observables belong to the space $\mathrm{Lip}_\Gamma(\mathcal{X})$, corresponding to the case where a vanishes. Finally, Section 4 is devoted to numerous examples in continuous and discrete settings.

2 Preliminaries

2.1 Functional inequalities

Let $(\mathcal{X}, d, \mu)$ be a metric probability space endowed with the corresponding Borel σ-field $\mathscr{B}$. Denote $\mathcal{A}_0$ a suitable algebra of real-valued functions defined on $\mathcal{X}$ and let $\mathcal{A}$ be an algebra extending $\mathcal{A}_0$, containing the constants, being stable under the action of smooth multivariate functions and such that for any $f \in \mathcal{A}$, we have $fg \in \mathcal{A}_0$ for any $g \in \mathcal{A}_0$. In the sequel we denote $L^p(\mu) := L^p(\mathcal{X}, \mathscr{B}, \mu)$ for $p \in [1, \infty]$. One of the main protagonists of the present paper is the *carré du champ* Γ, which is a bilinear symmetric operator defined on $\mathcal{A} \times \mathcal{A}$ by

$$\Gamma(f, g) \quad := \quad \frac{1}{2} \left(\mathcal{L}(fg) - f\mathcal{L}g - g\mathcal{L}f \right),$$

where $\mathcal{L}$ is an operator defined on $\mathcal{A}$ which is assumed to be symmetric on $\mathcal{A}_0$ in $L^2(\mu)$. As mentioned in the Introduction, there is a natural pseudo-distance associated to the operator Γ which can be defined as

$$d_\Gamma(x, y) := \sup \left\{ |f(x) - f(y)| : f \in \mathcal{A},\ \|\Gamma(f, f)\|_{L^\infty(\mu)} \leq 1 \right\}, \quad x, y \in \mathcal{X}.$$

Although this distance can be infinite, it is well-defined in the situations of interest and carries a lot of information. In the sequel, we denote $\mathrm{Lip}_\Gamma(\mathcal{X})$ the space of Lipschitz functions with respect to d_Γ. Certainly, there is no reason *a priori* that the space $\mathrm{Lip}_\Gamma(\mathcal{X})$ coincides with the usual Lipschitz space $\mathrm{Lip}(\mathcal{X})$, i.e. the space of Lipschitz functions on $\mathcal{X}$ with finite Lipschitz seminorm with respect to the given distance d,

$$\|f\|_{\mathrm{Lip}} \quad := \quad \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)} < \infty.$$

Consider the pre-Dirichlet form defined on $\mathcal{A}_0 \times \mathcal{A}_0$ by

$$\mathcal{E}_\mu(f, g) \quad := \quad \int_{\mathcal{X}} \Gamma(f, g)\, d\mu = -\int_{\mathcal{X}} f\, \mathcal{L}g\, d\mu = -\int_{\mathcal{X}} g\, \mathcal{L}f\, d\mu.$$

We assume in the remainder of the paper that this form is closable, that is, it can be extended to a true Dirichlet form (still denoted $\mathcal{E}_\mu$) on a domain $\mathcal{D}(\mathcal{E}_\mu)$ in which $\mathcal{A}_0$ is dense for the associated norm

$$\|f\|_{\mathcal{E}_\mu} := \sqrt{\|f\|^2_{L^2(\mu)} + \mathcal{E}_\mu(f, f)}.$$

In other words, the space $\mathcal{A}_0$ is a core of the domain of the Dirichlet form. In particular, the Donsker-Varadhan information of any probability measure ν on $\mathcal{X}$ with respect to the invariant measure μ is defined as

$$I(\nu|\mu) := \begin{cases} \mathcal{E}_\mu\left(\sqrt{f}, \sqrt{f}\right) & \text{if} \quad d\nu = f d\mu, \quad \sqrt{f} \in \mathcal{D}(\mathcal{E}_\mu); \\ \infty & \text{otherwise.} \end{cases}$$

Denote $(\mathcal{L}, \mathcal{D}_2(\mathcal{L}))$ the self-adjoint extension of the operator $(\mathcal{L}, \mathcal{A}_0)$ corresponding to the generator of a strongly continuous symmetric Markov semigroup $(P_t)_{t \geq 0}$ on $L^2(\mu)$.

In the probabilistic language, we have an $\mathcal{X}$-valued *càdlàg* ergodic Markov process $\{(X_t)_{t\geq 0}, (\mathbb{P}_x)_{x\in\mathcal{X}}\}$ defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$, with reversible invariant measure (or stationary distribution) μ. A key point in the forthcoming analysis is that the functional $\nu \mapsto I(\nu|\mu)$ defined above is nothing but the rate function governing the Large Deviation Principle in large time of the empirical measure of $(X_t)_{t\geq 0}$. However in the non-reversible case, it is given by a contraction form of the Donsker-Varadhan entropy which is different from the Donsker-Varadhan information, so that our study will not extend to the non-symmetric case, unfortunately.

Now let us introduce the functional inequalities we will focus on in the paper. Given an integrable function $f \in L^1(\mu)$, we denote $\mu(f) := \int_{\mathcal{X}} f d\mu$. Let I be an open interval of $\mathbb{R}$ and for a convex function $\phi : I \to \mathbb{R}$ we define the ϕ-entropy of a function $f : \mathcal{X} \to I$ with $\phi(f) \in L^1(\mu)$ as

$$\mathrm{Ent}_\mu^\phi(f) := \mu\left(\phi(f)\right) - \phi\left(\mu(f)\right).$$

The dynamics (μ, Γ) satisfies a ϕ-entropy inequality with constant $C_\phi > 0$ if for any I-valued function $f \in \mathcal{D}(\mathcal{E}_\mu)$ such that $\phi'(f) \in \mathcal{D}(\mathcal{E}_\mu)$,

$$C_\phi \,\mathrm{Ent}_\mu^\phi(f) \leq \frac{1}{2}\,\mathcal{E}_\mu(f, \phi'(f)).$$

See for instance the work of Chafaï [16] for a careful investigation of the properties of ϕ-entropies. The latter inequality is satisfied if and only if the following entropy dissipation of the semigroup holds: for any I-valued function f such that $\phi(f) \in L^1(\mu)$,

$$\mathrm{Ent}_\mu^\phi(P_t f) \leq e^{-2C_\phi t}\,\mathrm{Ent}_\mu^\phi(f), \quad t \geq 0.$$

In this paper we will consider three cases:

(i) the Poincaré inequality: $\phi(u) = u^2$ with $I = \mathbb{R}$ and the ϕ-entropy inequality rewrites as

$$\lambda\,\mathrm{Var}_\mu(f) \leq \mathcal{E}_\mu\left(f, f\right), \tag{2.1}$$

where the variance of f under μ is given by

$$\mathrm{Var}_\mu(f) := \mu(f^2) - \mu(f)^2.$$

The optimal constant λ_1 (say) is nothing but the spectral gap in $L^2(\mu)$ of the operator $-\mathcal{L}$. Estimating λ_1 allows us to obtain the optimal rate of convergence of the semigroup in $L^2(\mu)$.

(ii) the entropic inequality: $\phi(u) = u\,\log u$ with $I = (0,\infty)$ and the ϕ-entropy inequality is given by

$$\rho\,\mathrm{Ent}_\mu(f) \leq \mathcal{E}_\mu\left(f, \log f\right), \tag{2.2}$$

where the entropy under μ of the smooth positive function f is defined by

$$\mathrm{Ent}_\mu(f) := \mu(f\,\log f) - \mu(f)\,\log\mu(f).$$

We have skipped in the inequality the constant $1/2$ for convenience in future computations. Once again, the best constant ρ_0 in (2.2) gives the optimal exponential decay of the entropy along the semigroup.

(iii) the Beckner-type inequality: $\phi(u) = u^p$ with $p \in (1,2]$ and $I = (0,\infty)$. We have in this case

$$\alpha_p\,\left(\mu(f^p) - \mu(f)^p\right) \leq \frac{p}{2}\,\mathcal{E}_\mu\left(f, f^{p-1}\right). \tag{2.3}$$

Estimating α_p gives the optimal rate of convergence of the semigroup in $L^p(\mu)$.

The entropic and Beckner-type inequalities (in the case $1 < p < 2$) are stronger than the Poincaré inequality (apply these inequalities to the function $1 + \varepsilon f$ and let $\varepsilon \to 0$). Moreover it reduces to the Poincaré inequality (2.1) if $p = 2$, whereas dividing both sides by $p - 1$ and taking the limit as $p \to 1$ we obtain the entropic inequality (2.2).

In this paper we will mainly consider two general classes of reversible Markov processes: diffusions and pure jump Markov processes, to which we turn now.

2.2 Diffusion processes

A diffusion process on the Euclidean space $\mathcal{X} = \mathbb{R}^d$ corresponds to a path continuous Markov process on $\mathbb{R}^d$ whose generator $\mathcal{L}$ is a second order differential operator defined initially on $\mathcal{A} := \mathcal{C}^\infty(\mathbb{R}^d)$, the space of infinitely differentiable real-valued functions on $\mathbb{R}^d$:

$$\mathcal{L}f(x) = \sum_{i,j=1}^d a_{i,j}(x) \, \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_{i=1}^d b_i(x) \, \frac{\partial f}{\partial x_i}(x), \quad x \in \mathbb{R}^d.$$

Here $\mathcal{L}$ is assumed to be symmetric on $\mathcal{A}_0 := \mathcal{C}_0^\infty(\mathbb{R}^d)$ with respect to some probability measure μ, where $\mathcal{C}_0^\infty(\mathbb{R}^d)$ is the subspace of $\mathcal{C}^\infty(\mathbb{R}^d)$ consisting of compactly supported functions. Moreover $a := \sigma\sigma^*$ is a measurable and locally bounded function from $\mathbb{R}^d$ to the space of $d \times d$ symmetric positive definite matrices with smooth entries, σ^* being the transpose of the matrix σ, and the measurable drift $b : \mathbb{R}^d \to \mathbb{R}^d$ is also assumed to be smooth. In this case the carré du champ is given by

$$\begin{aligned} \Gamma(f,g) &= \sum_{i,j=1}^d a_{i,j} \, \frac{\partial f}{\partial x_i} \, \frac{\partial f}{\partial x_j} \\ &= < \sigma^* \nabla f, \sigma^* \nabla g >, \end{aligned}$$

where $< \cdot, \cdot >$ stands for the Euclidean scalar product in $\mathbb{R}^d$ and ∇ is the usual gradient operator. In particular when σ is the identity matrix, the spaces $\mathrm{Lip}(\mathbb{R}^d)$ and $\mathrm{Lip}_\Gamma(\mathbb{R}^d)$ might be identified.

In contrast to the jump case introduced below, Γ is a differentiation, i.e. for any functions $(f_k)_{1 \le k \le n}, f$ in $\mathcal{C}^\infty(\mathbb{R}^d)$ and any smooth enough function $\phi : \mathbb{R}^n \to \mathbb{R}$,

$$\Gamma(\phi(f_1, \ldots, f_n), f) = \sum_{i=1}^n \frac{\partial \phi}{\partial x_i}(f_1, \ldots, f_n) \, \Gamma(f_i, f). \tag{2.4}$$

Due to this chain rule derivation formula, the entropic inequality (2.2) rewrites in the diffusion case as the famous log-Sobolev inequality

$$\rho \, \mathrm{Ent}_\mu(f^2) \le 4 \, \mathcal{E}_\mu(f,f), \tag{2.5}$$

which is the original inequality (up to the extra factor 4) derived by Gross [24] to study hypercontractivity of the underlying semigroup. When we will consider diffusion processes in the sequel, we will use the terminology “log-Sobolev inequality" instead of “entropic inequality".

On the other hand, letting $p = 2/q$ for $q \in [1,2)$ and $f = g^q$, the Beckner-type inequality (2.3) rewrites as the so-called standard Beckner inequality:

$$\alpha_{2/q} \, \left(\mu(g^2) - \mu(g^q)^{2/q} \right) \le (2-q) \, \mathcal{E}_\mu(g,g). \tag{2.6}$$

Such an inequality was introduced by Beckner [3] for the Gaussian measure. In particular, the limiting case $q \to 2$ recovers the classical log-Sobolev inequality. Note however that the inequality (2.6) is weaker than the log-Sobolev inequality, cf. [33].

2.3 Markov jump processes

Dealing with a pure-jump Markov process, the generator $\mathcal{L}$ is defined on the space $\mathcal{A}$ of real-valued bounded functions on the discrete space $\mathcal{X}$ by

$$\mathcal{L}f(x) = \int_{\mathcal{X}} (f(y) - f(x))\, Q_x(dy), \quad x \in \mathcal{X},$$

where the transition kernel $x \mapsto Q_x$ is a measurable mapping from $\mathcal{X}$ to the set of Radon measures on $\mathcal{X}$ endowed with the corresponding Borel σ-field, and which satisfies the following stability assumption:

$$\int_{\mathcal{X}} Q_x(dy) < \infty, \quad x \in \mathcal{X}, \tag{2.7}$$

so that the process is piecewise constant. If the transition kernel has finite support then one can take for $\mathcal{A}$ the space of all real-valued functions on $\mathcal{X}$. Here, reversibility means that the following detailed balance condition is satisfied:

$$Q_x(dy)\, \mu(dx) = Q_y(dx)\, \mu(dy). \tag{2.8}$$

The carré du champ operator Γ admits an explicit expression given for any $f, g \in \mathcal{A}$ by

$$\Gamma(f,g)(x) = \frac{1}{2} \int_{\mathcal{X}} (f(y) - f(x))\, (g(y) - g(x))\; Q_x(dy),$$

and we have

$$\Gamma(f,f)(x) = \frac{1}{2} \int_{\mathcal{X}} (f(y) - f(x))^2\; Q_x(dy).$$

In particular, the spaces $\mathrm{Lip}(\mathcal{X})$ and $\mathrm{Lip}_\Gamma(\mathcal{X})$ have no reason to coincide since the kernel of the generator may be unbounded, i.e.

$$\sup_{x \in \mathcal{X}} \int_{\mathcal{X}} Q_x(dy) = \infty. \tag{2.9}$$

Finally the pre-Dirichlet form is defined initially on the space $\mathcal{A}_0 \subset \mathcal{A}$ of functions with finite support and after extension the Dirichlet form is given for any $f, g \in \mathcal{D}(\mathcal{E}_\mu)$ by

$$\begin{aligned} \mathcal{E}_\mu(f,g) &= \frac{1}{2} \int_{\mathcal{X}} \int_{\mathcal{X}} (f(y) - f(x))\, (g(y) - g(x))\; Q_x(dy)\mu(dx) \\ &= \iint_{f(x) > f(y)} (f(y) - f(x))\, (g(y) - g(x))\; Q_x(dy)\mu(dx), \end{aligned}$$

where in the last line the reversibility is used. In our jump framework, the entropic inequality (2.2) corresponds to one of the so-called modified log-Sobolev inequalities introduced by Bobkov and Ledoux [7]. However, due to the lack of chain rule for discrete gradients, this inequality is different from the discrete version of the log-Sobolev inequality (2.5), and the same remark holds between the Beckner-type inequality (2.3) and the standard Beckner inequality (2.6). We refer to [21, 7, 10] for historical and tutorial references on these discrete functional inequalities, together with a hierarchy of the various modified log-Sobolev inequalities.

3 Main results

In this paper, we emphasize a dynamical point of view on the concentration of measure phenomenon. As announced in the Introduction, we obtain concentration properties of the invariant measure μ through observables which are not required to belong to the spaces $\mathrm{Lip}(\mathcal{X})$ nor $\mathrm{Lip}_\Gamma(\mathcal{X})$, but which satisfy a Lyapunov condition. In order to state this condition properly, let us introduce first the extended domain of the generator. Denote the probability measure $\mathbb{P}_\nu(\cdot) := \int_{\mathcal{X}} \mathbb{P}_x(\cdot)\,\nu(dx)$ where ν is an arbitrary initial probability distribution. Recall that if $\mathcal{X}$ is discrete then any function from $\mathcal{X}$ to $\mathbb{R}$ is continuous. A continuous function f is said to belong to the extended domain $\mathcal{D}_e(\mathcal{L})$ of the generator $\mathcal{L}$ if there exists some measurable function $g : \mathcal{X} \to \mathbb{R}$ such that for any $t \geq 0$, $\int_0^t |g(X_s)|\,ds < \infty$, $\mathbb{P}_\mu$-a.s. and the process

$$M_t^f = f(X_t) - f(X_0) - \int_0^t g(X_s)\,ds, \quad t \geq 0,$$

is a local $\mathbb{P}_\mu$-martingale. In this case we write $f \in \mathcal{D}_e(\mathcal{L})$ and $\mathcal{L}f = g$.

The first result on which our analysis is based is closely related to the theory of large deviations, see for instance [25], Lemma 5.6 for a proof in the general case of reversible Markov processes.

Lemma 3.1. *For any continuous function $1 \leq V \in \mathcal{D}_e(\mathcal{L})$ such that $-\mathcal{L}V/V$ is bounded from below μ-a.s., we have for any probability measure ν on $\mathcal{X}$,*

$$\int_{\mathcal{X}} -\frac{\mathcal{L}V}{V}\,d\nu \;\leq\; I(\nu|\mu). \tag{3.1}$$

Now we are able to state the Lyapunov condition we will focus on along this paper.

Definition 3.2. *Let a, b be two positive constants and let $V \in \mathcal{D}_e(\mathcal{L})$ be a test function with values in $[1, \infty)$. A function $f \in \mathcal{D}_2(\mathcal{L})$ belongs to the class $\mathcal{L}_V(a, b)$ if the following inequality is satisfied μ-a.s.:*

$$\Gamma(f, f) \leq -a\,\frac{\mathcal{L}V}{V} + b. \tag{3.2}$$

Remark 3.3. *In the examples of Section 4, the test function V will always be chosen sufficiently smooth and close to be non-integrable with respect to μ. Indeed it allows us in general to consider the largest possible ratio $-\mathcal{L}V/V$ and thus the largest possible class $\mathcal{L}_V(a, b)$ of observables for which our concentration results will be available.*

Remark 3.4. *The Poincaré inequality can be seen as a minimal assumption in our study of concentration by means of the Lyapunov condition (3.2). Indeed, if there exists a function $f \in \mathcal{L}_V(a, b)$ such that $\Gamma(f, f)$ is lower bounded by a positive constant at infinity, and this the case in the main examples of interest (except in the Cauchy-like case appearing briefly in Section 4), then the Poincaré inequality is satisfied, cf. [15]. Moreover, integrating with respect to μ both sides of the inequality (3.2) and using the Poincaré inequality yield $\mathrm{Var}_\mu(f) \leq b/\lambda_1$. In other words the constant b/λ_1 plays the role of the variance of the observable f in our work.*

Before stating our first main result, let us provide a key lemma. In the remainder of this paper, we only give the proofs in the jump case since the diffusion framework requires no additional difficulties and is even simpler, according to the chain rule derivation formula (2.4) satisfied by the carré du champ.

Lemma 3.5. *Let f belong to the class $\mathcal{L}_V(a, b)$. Given $\lambda \in (0, 2/\sqrt{a})$, let μ_λ be the probability measure with density $f_\lambda := e^{\lambda f}/Z_\lambda$ with respect to μ, where Z_λ is the appropriate normalization constant, which is assumed to be finite. We assume moreover that $\sqrt{f_\lambda} \in \mathcal{D}(\mathcal{E}_\mu)$. Then we have the inequality*

$$I(\mu_\lambda|\mu) \leq \frac{\lambda^2 b}{4 - \lambda^2 a}, \quad 0 < \lambda < \frac{2}{\sqrt{a}}.$$

Proof. Since $f \in \mathcal{L}_V(a,b)$, we have for any $\lambda \in (0, 2/\sqrt{a})$:

$$\begin{aligned} I(\mu_\lambda|\mu) &= \frac{1}{Z_\lambda} \iint_{f(x)>f(y)} \left(e^{\lambda f(x)/2} - e^{\lambda f(y)/2}\right)^2 Q_x(dy)\mu(dx) \\ &= \iint_{f(x)>f(y)} \left(1 - e^{-\lambda(f(x)-f(y))/2}\right)^2 f_\lambda(x)\, Q_x(dy)\mu(dx) \\ &\leq \frac{\lambda^2}{4} \int_{\mathcal{X}} \Gamma(f,f)\, d\mu_\lambda \\ &\leq \frac{\lambda^2}{4} \int_{\mathcal{X}} \left(-a\,\frac{\mathcal{L}V}{V} + b\right) d\mu_\lambda \\ &\leq \frac{\lambda^2}{4} \left(aI(\mu_\lambda|\mu) + b\right), \end{aligned}$$

where in the last line we used Lemma 3.1. Finally rearranging the terms allows us to obtain the desired inequality. □

We turn now to our first main and new result which exhibits a non-Gaussian concentration estimate through observables belonging to the class $\mathcal{L}_V(a,b)$. Due to the approach we will use, the numerical constants in the estimates below have no reason to be sharp.

Theorem 3.6. *Assume that the pair* (μ, Γ) *satisfies the entropic inequality (2.2). Let* $f \in \mathcal{L}_V(a,b)$ *and let*

$$r_{\max} := \frac{8b}{3\rho_0\sqrt{a}}$$

be the size of the Gaussian window. Then the invariant measure μ *has the following concentration property: for any deviation level* $0 \leq r \leq r_{\max}$*, the deviation is of Gaussian-type:*

$$\mu\left(\{x \in \mathcal{X} : f(x) - \mu(f) > r\}\right) \leq \exp\left(-\frac{3\rho_0 r^2}{16b}\right), \tag{3.3}$$

and for any $r \geq r_{\max}$*, the decay is exponential:*

$$\mu\left(\{x \in \mathcal{X} : f(x) - \mu(f) > r\}\right) \leq \exp\left(-\frac{r}{2\sqrt{a}}\right). \tag{3.4}$$

Remark 3.7. *In the sequel, a concentration property such as (3.3)-(3.4) will be called Gaussian-exponential concentration.*

Proof. Denote $L_\lambda := \lambda^{-1}\log Z_\lambda$, where $Z_\lambda := \int_{\mathcal{X}} e^{\lambda f}\, d\mu$, with $\lambda \in (0, 1/\sqrt{a})$, and let μ_λ be the absolutely continuous probability measure with density $f_\lambda := e^{\lambda f}/Z_\lambda$ with respect to μ. Using a standard approximation procedure one may assume that the observable $f \in \mathcal{L}_V(a,b)$ is bounded so that $Z_\lambda < \infty$ and $\sqrt{f_\lambda} \in \mathcal{D}(\mathcal{E}_\mu)$. The following proof is a modification of the famous Herbst method popularized by Ledoux. Using the entropic inequality (2.2),

$$\begin{aligned}
\frac{d}{d\lambda} L_\lambda &= \frac{1}{\lambda^2 Z_\lambda} \, \mathrm{Ent}_\mu(e^{\lambda f}) \\
&\leq \frac{1}{\rho_0 \lambda^2 Z_\lambda} \, \mathcal{E}_\mu\left(\lambda f, e^{\lambda f}\right) \\
&= \frac{1}{\rho_0 \lambda} \iint_{f(x)>f(y)} (f(x)-f(y)) \left(1 - e^{-\lambda(f(x)-f(y))}\right) f_\lambda(x) \, Q_x(dy)\mu(dx) \\
&\leq \frac{1}{\rho_0} \int_{\mathcal{X}} \int_{\mathcal{X}} \Gamma(f,f) \, d\mu_\lambda \\
&\leq \frac{1}{\rho_0} \int_{\mathcal{X}} \left(-a \frac{\mathcal{L}V}{V} + b\right) d\mu_\lambda \\
&\leq \frac{1}{\rho_0} \left(a I(\mu_\lambda|\mu) + b\right),
\end{aligned}$$

where we used that $f \in \mathcal{L}_V(a,b)$ and then Lemma 3.1 in the two last lines. Thus Lemma 3.5 entails the inequality

$$\frac{d}{d\lambda} L_\lambda \leq \frac{4b}{3\rho_0}, \quad 0 < \lambda < \frac{1}{\sqrt{a}},$$

and therefore the following log-Laplace estimate is available for any $0 < \lambda < 1/\sqrt{a}$:

$$\log \int_{\mathcal{X}} e^{\lambda f} \, d\mu \leq \lambda \mu(f) + \frac{4b\lambda^2}{3\rho_0}. \tag{3.5}$$

Finally using Chebyshev's inequality and optimizing in $\lambda \in (0, 1/\sqrt{a})$ yields the tail estimates (3.3) and (3.4). The proof of Theorem 3.6 is thus complete. □

Remark 3.8. *Two deviation regimes appear, Gaussian and exponential, with continuous transition from one to the other. In contrast to the classical Herbst method where the observables belong to* $\mathrm{Lip}_\Gamma(\mathcal{X})$*, i.e.* $a = 0$ *in the Lyapunov condition (3.2), our assumption allows us to go beyond this Lipschitz property. Although the Gaussian concentration is preserved for small deviation levels, this feature is lost for large* r *and is replaced by an exponential tail which reveals to be sharp, as it will be developed in the examples of Section 4.*

Remark 3.9. *By the Central Limit Theorem, the order of magnitude in the Gaussian regime is correct in terms of all the parameters of interest. Since the entropic inequality entails a Poincaré inequality, we have* $\rho_0 \leq \lambda_1$ *and thus for any observable* $f \in \mathcal{L}_V(a,b)$*, we get* $\mathrm{Var}_\mu(f) \leq b/\rho_0$*. Therefore, if* $\mathcal{X}$ *is a product space and the process* $(X_t)_{t\geq 0}$ *has independent and identically distributed coordinates, then under the observable* $f(x) := \sum_{k=1}^d \phi(x_k)$*,* $x = (x_1, \ldots, x_d) \in \mathcal{X}$*, we obtain the following concentration*

$$\mu\left(\left\{x \in \mathcal{X} : f(x) - \mu(f) > r\sqrt{d}\right\}\right) \leq \exp\left(-\frac{3\rho_0 r^2}{16\tilde{b}}\right), \quad 0 \leq r \leq \frac{8\tilde{b}\sqrt{d}}{3\rho_0\sqrt{a}},$$

which is sharp for large d*. Here the important point is that the positive parameter* $\tilde{b}$*, given by the Lyapunov condition on the univariate function* ϕ*, is independent of* d*.*

Remark 3.10. *The most naive approach to obtain concentration of measure under a given non-Lipschitz observable* f *is controlling* f *by a monotone function of a Lipschitz function and then making a change of variable of the deviation level in the concentration estimates. For instance if the concentration under Lipschitz observables is Gaussian* *then it is of exponential type under quadratic like observables. However this naive approach is not always feasible since this method requires monotonicity, and when it is, one obtains the correct order of magnitude for large deviation levels but without the sharp dependence with respect to the important parameters, in particular if the expectation of f is not explicitly computable. Therefore one deduces that in most of the cases of interest such an approach is not convenient.*

Remark 3.11. *The method is sufficiently robust to get, for large deviation level r, other regimes than exponential under particular observables. For example, assume that we consider $f \in \mathcal{L}_V(a,b)$ such that $\Gamma(f,f) \ll -a\mathcal{L}V/V + b$ but that there exists two functions $\phi, \psi : (0,\infty) \to (0,\infty)$ such that for all $\varepsilon > 0$,*

$$\Gamma(f,f) \leq \phi(\varepsilon)\left(-a\,\frac{\mathcal{L}V}{V} + b\right) + \psi(\varepsilon).$$

Then plugging this estimate in the previous proofs of Lemma 3.5 and Theorem 3.6, one has for all $\varepsilon > 0$,

$$I(\mu_\lambda|\mu) \leq \frac{\lambda^2}{4}\left(a\phi(\varepsilon) I(\mu_\lambda|\mu) + b\phi(\varepsilon) + \psi(\varepsilon)\right).$$

Optimizing in $\varepsilon > 0$ enables to get for some function $\Phi : [0,\infty) \to [0,\infty]$ and all $\lambda > 0$,

$$I(\mu_\lambda|\mu) \leq \Phi(\lambda),$$

leading then to super-exponential regime for large r. We will illustrate this on an example in Section 4.

Inspired by the method of Otto and Villani [38], who studied the links between log-Sobolev and transportation inequalities on continuous spaces (see also Sammer [40] in the finite state space case), let us recover Theorem 3.6 by using a semigroup proof. Once again we focus our attention on the jump case. Let h be a smooth density with respect to μ. Given $t > 0$, denote ν_t the probability measure with density $P_t h$ with respect to μ. We assume that the Donsker-Varadhan information $I(\nu_t|\mu)$ is well-defined, i.e. $\sqrt{P_t h} \in \mathcal{D}(\mathcal{E}_\mu)$. Using Cauchy-Schwarz's inequality and then reversibility,

$$\begin{aligned}
\mathcal{E}_\mu(P_t h, f) &= \frac{1}{2}\int_{\mathcal{X}}\int_{\mathcal{X}} (P_t h(x) - P_t h(y))\,(f(x) - f(y))\,Q_x(dy)\mu(dx) \\
&\leq \sqrt{I(\nu_t|\mu)}\sqrt{\frac{1}{2}\int_{\mathcal{X}}\int_{\mathcal{X}} \left(\sqrt{P_t h(x)} + \sqrt{P_t h(y)}\right)^2 (f(x) - f(y))^2\, Q_x(dy)\mu(dx)} \\
&\leq 2\sqrt{I(\nu_t|\mu)}\sqrt{\int_{\mathcal{X}} \Gamma(f,f)\, d\nu_t} \\
&\leq 2\sqrt{I(\nu_t|\mu)}\sqrt{\int_{\mathcal{X}} \left(-a\,\frac{\mathcal{L}V}{V} + b\right) d\nu_t} \\
&\leq 2\sqrt{I(\nu_t|\mu)}\sqrt{aI(\nu_t|\mu) + b},
\end{aligned}$$

where in the two last lines we used that $f \in \mathcal{L}_V(a,b)$ and then Lemma 3.1. Using now the elementary inequality $2(a-b)^2 \leq (a^2 - b^2)\log(a/b)$ available for any $a, b > 0$ and then the entropic inequality (2.2), we get

$$\begin{aligned}
\mathcal{E}_\mu(P_t h, f) &\leq \mathcal{E}_\mu(P_t h, \log P_t h)\left(\frac{\sqrt{a}}{2} + \sqrt{\frac{b}{\mathcal{E}_\mu(P_t h, \log P_t h)}}\right) \\
&\leq \mathcal{E}_\mu(P_t h, \log P_t h)\left(\frac{\sqrt{a}}{2} + \sqrt{\frac{b}{\rho_0 \mathrm{Ent}_\mu(P_t h)}}\right).
\end{aligned}$$

Integrating time between 0 and infinity entails the covariance inequality:

$$\begin{aligned}
\mathrm{Cov}_\mu(f,h) &:= \mu(fh) - \mu(f)\,\mu(h) \\
&= \int_0^\infty \mathcal{E}_\mu(P_t h, f)\,dt \\
&\leq \int_0^\infty \mathcal{E}_\mu(P_t h, \log P_t h) \left(\frac{\sqrt{a}}{2} + \sqrt{\frac{b}{\rho_0 \mathrm{Ent}_\mu(P_t h)}} \right) dt \\
&= \frac{\sqrt{a}}{2}\,\mathrm{Ent}_\mu(h) + 2\sqrt{\frac{b\,\mathrm{Ent}_\mu(h)}{\rho_0}},
\end{aligned}$$

which in turn yields the inequality

$$\alpha\left(\mathrm{Cov}_\mu(f,h)\right) \leq \mathrm{Ent}_\mu(h),$$

where α is the function

$$\alpha(r) = \frac{\rho_0 r^2}{4b + 2\rho_0 \sqrt{a}\, r}, \quad r > 0.$$

Finally using Theorem 3.2 in [23], we obtain the following concentration estimate through the observable $f \in \mathcal{L}_V(a,b)$:

$$\mu\left(\{x \in \mathcal{X} : f(x) - \mu(f) > r\}\right) \leq \exp\left(-\alpha(r)\right), \quad r \geq 0.$$

One deduces that, up to numerical constants, this result is similar to that emphasized in Theorem 3.6 since for small deviation level, $\alpha(r) = O(\rho_0 r^2/b)$ whereas $\alpha(r) = O(r/\sqrt{a})$ for large r.

As we have seen above, the entropic inequality (2.2) entails on the one hand a concentration property for the invariant measure μ through observables $f \in \mathcal{L}_V(a,b)$. On the other hand and as announced in Remark 3.4, the Poincaré inequality can be seen as a minimal assumption in our study. Hence one may wonder if the Beckner-type inequality (2.3), which interpolates between both, provides qualitative concentration estimates through observables in $\mathcal{L}_V(a,b)$. Our second main result, Theorem 3.12, goes in this way.

Theorem 3.12. *Assume that there exists $p \in (1,2]$ such that the pair (μ, Γ) satisfies the Beckner-type inequality (2.3). Moreover, assume that the observable $f \in \mathcal{L}_V(a,b)$ with the restriction $\alpha_p \leq 2b(p-1)/(3pa)$ and let*

$$r_{\max} := \sqrt{\frac{32bp}{27(p-1)\alpha_p}}.$$

Then the following tail estimates hold: for any deviation level $0 \leq r \leq r_{\max}$,

$$\mu\left(\{x \in \mathcal{X} : f(x) - \mu(f) > r\}\right) \leq \exp\left(-\frac{9\alpha_p r^2}{32b}\right), \tag{3.6}$$

whereas for any $r \geq r_{\max}$,

$$\mu\left(\{x \in \mathcal{X} : f(x) - \mu(f) > r\}\right) \leq \exp\left(-r\sqrt{\frac{3p\alpha_p}{32b(p-1)}}\right). \tag{3.7}$$

Proof. The proof is adapted from the method of Aida and Stroock introduced in [1]. Assume without loss of generality that f is centered and bounded and for any $\lambda \in (0, \lambda_0)$, where

$$\lambda_0 := \sqrt{\frac{3p\alpha_p}{2b(p-1)}} \leq \frac{1}{\sqrt{a}},$$

denote once again $Z_\lambda := \int_{\mathcal{X}} e^{\lambda f}\, d\mu$ which is finite and let μ_λ be the probability measure with density $f_\lambda := e^{\lambda f}/Z_\lambda$ with respect to μ, which satisfies $\sqrt{f_\lambda} \in \mathcal{D}(\mathcal{E}_\mu)$. We have by the Beckner-type inequality (2.3) applied to the function $e^{\lambda f/p}$,

$$\begin{aligned} Z_\lambda - Z_{\lambda/p}^p &\leq \frac{p}{2\alpha_p}\, \mathcal{E}_\mu(e^{\lambda f/p}, e^{\lambda(1-1/p)f}) \\ &= \frac{p}{2\alpha_p} \iint_{f(x)>f(y)} e^{\lambda f(x)} \left(1 - e^{-\lambda(f(x)-f(y))/p}\right)\left(1 - e^{-\lambda(1-1/p)(f(x)-f(y))}\right) Q_x(dy)\mu(dx) \\ &\leq \frac{\lambda^2 (p-1) Z_\lambda}{2p\alpha_p} \int_{\mathcal{X}} \Gamma(f,f)\, d\mu_\lambda \\ &\leq \frac{\lambda^2 (p-1) Z_\lambda}{2p\alpha_p} \int_{\mathcal{X}} \left(-a\,\frac{\mathcal{L}V}{V} + b\right) d\mu_\lambda \\ &\leq \frac{\lambda^2 (p-1) Z_\lambda}{2p\alpha_p}\, (aI(\mu_\lambda|\mu) + b) \\ &\leq \left(\frac{\lambda}{\lambda_0}\right)^2 Z_\lambda, \end{aligned}$$

where we used that $f \in \mathcal{L}_V(a,b)$ and Lemmas 3.1-3.5 in the last three lines. Hence rearranging the terms above and iterating the procedure yields for every $n \geq 1$,

$$Z_\lambda \leq \prod_{k=0}^{n-1} \left(\frac{\lambda_0^2}{\lambda_0^2 - \lambda^2/p^{2k}}\right)^{p^k} (Z_{\lambda/p^n})^{p^n}.$$

Since f is centered, the quantity $Z_{\lambda/p^n}^{p^n}$ goes to 1 as $n \to \infty$ and from the latter inequality we obtain after taking logarithm,

$$\begin{aligned} \log Z_\lambda &\leq -\sum_{k=0}^{\infty} p^k \log\left(1 - \frac{(\lambda/\lambda_0)^2}{p^{2k}}\right) \\ &= \sum_{k=0}^{\infty} \frac{p^{2k+1}}{p^{2k+1}-1} \times \frac{(\lambda/\lambda_0)^{2(k+1)}}{k+1} \\ &\leq -\frac{p}{p-1} \log\left(1 - \left(\frac{\lambda}{\lambda_0}\right)^2\right). \end{aligned}$$

In the last inequality we used the trivial inequality $p^{2k+1} \leq (\frac{p}{p-1})\,(p^{2k+1}-1)$ available for any integer k because $p \in (1,2]$. We thus obtain for any $0 < \lambda \leq \lambda_0/2$,

$$\begin{aligned} Z_\lambda &\leq \left(1 + \frac{\lambda^2}{\lambda_0^2 - \lambda^2}\right)^{p/(p-1)} \\ &\leq \exp\left(\frac{p\lambda^2}{(p-1)(\lambda_0^2 - \lambda^2)}\right) \\ &\leq \exp\left(\frac{4p\lambda^2}{3(p-1)\lambda_0^2}\right). \end{aligned}$$

Finally using the exponential Chebyshev inequality and optimizing in λ entails the desired result. □

Remark 3.13. *Taking* $r = r_{\max}$ *in the inequality* (3.6) *above entails the upper bound* $\exp(-p/3(p-1))$ *which is independent of all the parameters of interest. In other words there is essentially no Gaussian window since we obtain* $r_{\max} = O(\sqrt{b})$ *instead of the* *correct order of magnitude $O(b)$. But this issue was expected: our proof is adapted from the method of Aida and Stroock, which is known in the classical case of Lipschitz observables (i.e. the case $a = 0$ in our Lyapunov condition) to capture the optimal concentration behaviour only for large deviation levels. Although we worked quite a bit to obtain the expected Gaussian decay for small deviation levels (through a modified entropic inequality in the spirit of Section 3 of Bobkov and Ledoux [6]), it seems that Theorem 3.12 still leaves room for improvement, in particular the assumption relying α_p to the parameters a, b which is technical but cannot be avoided for the moment (however it will be satisfied as soon as b is taken sufficiently large, or a small enough).*

Remark 3.14. *As already mentioned, the Beckner-type inequality is stronger than the Poincaré inequality, i.e. $\alpha_p \leq \lambda_1$. However Theorem 3.12 does not entail a better concentration estimate than that obtained under the Poincaré inequality, except maybe when focusing on the constants depending on p (this is clearly not our interest here). The reason is due to the approach emphasized above which is exactly the same for any $p \in (1, 2]$, in contrast to Theorem 3.6 where the Herbst method is used.*

4 Examples

4.1 Diffusion processes

Let us apply now Theorems 3.6 and 3.12 to diffusion processes. In this part, the function U is a smooth potential such that e^{-U} is Lebesgue integrable, and denote μ the Boltzmann probability measure with density e^{-U}/Z with respect to the Lebesgue measure, where Z is the appropriate normalization factor.

The first example of interest is the so-called Kolmogorov process with generator given for any $f \in \mathcal{C}^\infty(\mathbb{R}^d)$ by

$$\mathcal{L}f = \Delta f - < \nabla U, \nabla f > .$$

One easily checks that μ is reversible for this process and the carré du champ is $\Gamma(f, f) = \|\nabla f\|^2$ where $\|\cdot\|$ stands for the Euclidean norm in $\mathbb{R}^d$. Hence by Rademacher's theorem, the spaces $\mathrm{Lip}(\mathbb{R}^d)$ and $\mathrm{Lip}_\Gamma(\mathbb{R}^d)$ coincide. Moreover the domain $\mathcal{D}(\mathcal{E}_\mu)$ of the Dirichlet form is $H^1(\mu)$.

4.1.1 Ornstein-Uhlenbeck process and the standard Gaussian distribution

Let us consider first the Ornstein-Uhlenbeck process which is a special case of the Kolmogorov process with potential $U(x) = \|x\|^2/2$. It has the standard Gaussian distribution as invariant measure. By the famous Gross theorem [24], the pair (μ, Γ) satisfies the log-Sobolev inequality, i.e. the entropic inequality with (optimal) constant $\rho_0 = 2$. Hence Theorem 3.6 will apply for observables in $\mathcal{L}_V(a, b)$ for some good test function V. For instance if $f(x) = \|x\|^2$, then choose the test function $V = e^{cU}$ with $c \in (0, 1)$, i.e. V is at the boundary of non-integrability. Then we have

$$-\frac{\mathcal{L}V(x)}{V(x)} \quad = \quad -cd + c(1-c)\,\|x\|^2.$$

Thus with the choice $c = 1/2$ we get $f \in \mathcal{L}_V(a, b)$ with $a = 16$ and $b = 8d$ and by Theorem 3.6, we obtain for any $0 \leq r \leq 8d/3$,

$$\mu\left(\{x \in \mathbb{R}^d : \|x\|^2 > d + r\}\right) \quad \leq \quad \exp\left(-\frac{3r^2}{64d}\right),$$

which is sharp up to a numerical constant since in this case $\mathrm{Var}_\mu(f) = 2d$. In the exponential regime we get for any $r \geq 8d/3$,

$$\mu\left(\{x \in \mathbb{R}^d : \|x\|^2 > d+r\}\right) \quad \leq \quad \exp\left(-\frac{r}{8}\right).$$

Actually, such a behaviour is expected since under μ, the variable $f(x)$ is distributed as a chi-squared random variable with d degrees of freedom, for which the decay is known to be only exponential for large deviation levels. Moreover we mention that we are in the (rare) cases where the naive approach discussed in Remark 3.10 is convenient. Denoting the expectation $m := \int_{\mathbb{R}^d} \|x\| \, \mu(dx)$, we have

$$\begin{aligned} \mu\left(\{x \in \mathbb{R}^d : \|x\|^2 > d+r\}\right) &= \mu\left(\{x \in \mathbb{R}^d : \|x\| - m > \sqrt{d+r} - m\}\right) \\ &\leq \exp\left(-\frac{(\sqrt{d+r}-\sqrt{d})^2}{2}\right), \end{aligned}$$

an estimate which can be transformed into a similar Gaussian-exponential concentration as above. Here we used Cauchy-Schwartz' inequality to get $m \leq \sqrt{d}$ and then the classical Gaussian deviation estimate satisfied by Lipschitz functions. The reason for which this naive method is sharp resides in the following facts: on the one hand the observable $x \mapsto \|x\|^2$ can be written as an increasing function of a (non-negative) Lipschitz function and on the other hand replacing the mean m by $\sqrt{d}$ is optimal in large dimension since standard computations yield

$$m = \sqrt{2}\, \frac{\Gamma((d+1)/2)}{\Gamma(d/2)},$$

which behaves like $\sqrt{d}$ for large d (together with Cauchy-Schwartz, an alternative proof is to apply the Poincaré inequality to the function $x \to \|x\|$ to obtain $m \geq \sqrt{d-1}$). Despite its simplicity, it reveals that the naive approach requires restrictive arguments which are hardly satisfied in the cases of interest.

Let us come back to the example announced in Remark 3.11. As observed above, our concentration result is convenient as soon as f is close to realizing the equality in the Lyapunov condition (3.2). However what happens for a function g such that $\|\nabla g\| \ll \|\nabla f\|$ at infinity ? For instance if $f(x) = \|x\|^2$ as above then how does the invariant measure concentrate through the observable $g(x) = \|x\|^{3/2}$? Let us investigate this point in detail now. Although the method we use below applies in the Gaussian case, we mention that it works in a more general framework. Assume that $f \in \mathcal{L}_V(a,b)$ and that there exists an observable g satisfying for any $\varepsilon > 0$,

$$\|\nabla g\|^2 \leq \varepsilon \|\nabla f\|^2 + \frac{1}{\varepsilon}. \tag{4.1}$$

Denote $a_\varepsilon := a\varepsilon$ and $b_\varepsilon := b\varepsilon + 1/\varepsilon$. Using the argument given in the proof of Lemma 3.5, we have for any $\lambda > 0$,

$$\begin{aligned} I(\mu_\lambda|\mu) &\leq \inf_{\varepsilon>0} \frac{\lambda^2}{4}\left(a_\varepsilon I(\mu_\lambda|\mu) + b_\varepsilon\right) \\ &= \frac{\lambda^2}{2}\sqrt{aI(\mu_\lambda|\mu) + b}, \end{aligned}$$

where in the definition of $I(\mu_\lambda|\mu)$ we replaced f by g. Hence we obtain

$$I(\mu_\lambda|\mu) \quad \leq \quad \frac{\lambda^2}{2}\left(\frac{a\lambda^2}{2} + \sqrt{b}\right).$$

Now the same argument as in the proof of Theorem 3.6 together with the latter inequality entail

$$\begin{aligned} \frac{d}{d\lambda} L_\lambda &\leq \inf_{\varepsilon>0} \frac{1}{\rho_0} \left(a_\varepsilon I(\mu_\lambda|\mu) + b_\varepsilon\right) \\ &= \frac{2}{\rho_0} \sqrt{a I(\mu_\lambda|\mu) + b} \\ &\leq \frac{2}{\rho_0} \left(\frac{a\lambda^2}{2} + \sqrt{b}\right), \end{aligned}$$

since $\sqrt{x(x+y)+y^2} \leq x+y$ for any $x, y \geq 0$. Hence we get for any $\lambda > 0$,

$$\begin{aligned} \log \int_{\mathcal{X}} e^{\lambda g}\, d\mu &\leq \lambda \mu(g) + \frac{2}{\rho_0} \left(\frac{a\lambda^4}{6} + \lambda^2 \sqrt{b}\right) \\ &\leq \lambda \mu(g) + \frac{4}{\rho_0} \max\left(\frac{a\lambda^4}{6}, \lambda^2\sqrt{b}\right). \end{aligned}$$

Finally using the exponential Chebyshev inequality and optimizing in λ entails for small deviations the Gaussian-type estimate

$$\mu\left(\{x \in \mathbb{R}^d : g(x) - \mu(g) > r\}\right) \leq \exp\left(-\frac{\rho_0\, r^2}{16\sqrt{b}}\right), \quad 0 \leq r \leq \frac{8\sqrt{6}\, b^{3/4}}{\rho_0 \sqrt{a}},$$

whereas for large deviations,

$$\mu\left(\{x \in \mathbb{R}^d : g(x) - \mu(g) > r\}\right) \leq \exp\left(-\left(\frac{6\rho_0}{a}\right)^{1/3} \frac{r^{4/3}}{4}\right), \quad r \geq \frac{8\sqrt{6}\, b^{3/4}}{\rho_0\sqrt{a}}.$$

Hence we have improved the decay in the concentration estimate for large deviation level r since it is no longer of order $\exp(-cr)$ but of order $\exp(-cr^{4/3})$. In terms of small deviations, the parameter $\sqrt{b}$ appears in the exponential at the denominator instead of b. But this is expected since by Poincaré inequality, the inequality (4.1) and the assumption $f \in \mathcal{L}_V(a,b)$, we obtain the following estimate on the variance of g:

$$\begin{aligned} \mathrm{Var}_\mu(g) &\leq \frac{1}{\rho_0} \left(\inf_{\varepsilon>0} \varepsilon \int_{\mathbb{R}^d} \|\nabla f\|^2\, d\mu + \frac{1}{\varepsilon}\right) \\ &\leq \frac{2}{\rho_0} \sqrt{\int_{\mathbb{R}^d} \|\nabla f\|^2\, d\mu} \\ &\leq \frac{2\sqrt{b}}{\rho_0}. \end{aligned}$$

Coming back to our example, if $f(x) = \|x\|^2$ and $g(x)$ is proportional to $\|x\|^{3/2}$, then as observed before we have $f \in \mathcal{L}_V(16, 8d)$ with $V(x) = e^{\|x\|^2/4}$ and moreover (4.1) is satisfied. Then we obtain the following inequalities: for small deviations $0 \leq r \leq \sqrt{6}\,(8d)^{3/4}$,

$$\mu\left(\{x \in \mathbb{R}^d : \|x\|^{3/2} - \int_{\mathbb{R}^d} \|x\|^{3/2}\, \mu(dx) > r\}\right) \leq \exp\left(-\frac{r^2}{8\sqrt{8d}}\right),$$

and for large deviations $r \geq \sqrt{6}\,(8d)^{3/4}$,

$$\mu\left(\{x \in \mathbb{R}^d : \|x\|^{3/2} - \int_{\mathbb{R}^d} \|x\|^{3/2}\, \mu(dx) > r\}\right) \leq \exp\left(-3^{1/3} \left(\frac{r}{4}\right)^{4/3}\right).$$

Now let us consider another type of observables. Take f as a quadratic form on $\mathbb{R}^d$, i.e. there exists a positive definite symmetric matrix $A = (a_{i,j})_{i,j=1,\ldots,d}$ of size d such $f(x) =< Ax, x >$, $x \in \mathbb{R}^d$. Then in the Gaussian regime we expect the variance of f in the denominator,

$$\mathrm{Var}_\mu(f) \,=\, 2 \sum_{i,j=1}^{d} a_{i,j}^2 \,=\, 2 \sum_{i=1}^{d} \lambda_i^2,$$

where (λ_i) are the real eigenvalues of the symmetric matrix A. In other words $\mathrm{Var}_\mu(f)$ is nothing but 2 times the square of the Hilbert-Schmidt norm $\|A\|_{\mathrm{HS}}$ of the matrix A. Using the same test function V as before would entail that $f \in \mathcal{L}_V(a,b)$ with $a = 16\,\|A\|_{\mathrm{op}}^2$ and $b = 8d\,\|A\|_{\mathrm{op}}^2$. Here $\|A\|_{\mathrm{op}}$ denotes the (Euclidean) operator norm of the matrix A, i.e. its spectral radius, $\max_i \lambda_i$. With this choice of parameters the inequality $\mathrm{Var}_\mu(f) \le b/\rho_0$ is too weak to provide a reasonable variance estimate since b behaves badly in terms of dimension. To obtain a qualitative concentration estimate, the idea is to change the test function and choose $V(x) = e^{c\|Ax\|^2}$ with $c > 0$ to be fixed later to ensure the integrability of V with respect to the Gaussian measure μ. Then we have $f \in \mathcal{L}_V(a,b)$ for some constants a and b if and only if for any $y \in \mathbb{R}^d$, the following inequality holds:

$$4ac^2 \sum_{i=1}^{d} \lambda_i^4\, y_i^2 \,\le\, 2\,(ac-2) \sum_{i=1}^{d} \lambda_i^2\, y_i^2 + b - 2ac \sum_{i=1}^{d} \lambda_i^2.$$

After a bit of analysis, choosing $c := 1/4\|A\|_{\mathrm{op}}^2$, the same a as before and

$$b \,:=\, 2ac \sum_{i=1}^{d} \lambda_i^2 \,=\, 8 \sum_{i=1}^{d} \lambda_i^2,$$

we get $f \in \mathcal{L}_V(a,b)$ and applying then Theorem 3.6, we obtain the following concentration estimate: for any $0 \le r \le 8\|A\|_{\mathrm{HS}}^2/3\|A\|_{\mathrm{op}}$,

$$\mu\left(\{x \in \mathbb{R}^d :< Ax, x > - \int_{\mathbb{R}^d} < Ax, x > \,\mu(dx) > r\}\right) \quad \le \quad \exp\left(-\frac{3r^2}{64\|A\|_{\mathrm{HS}}^2}\right),$$

whereas for any $r \ge 8\|A\|_{\mathrm{HS}}^2/3\|A\|_{\mathrm{op}}$,

$$\mu\left(\{x \in \mathbb{R}^d :< Ax, x > - \int_{\mathbb{R}^d} < Ax, x > \,\mu(dx) > r\}\right) \quad \le \quad \exp\left(-\frac{r}{8\|A\|_{\mathrm{op}}}\right).$$

This example emphasizes the importance of the choice of the function V in the condition $\mathcal{L}_V(a,b)$. See also for instance [26, 32] for some nice studies on the concentration properties of Gaussian-like quadratic forms and Gaussian chaoses.

4.1.2 Kolmogorov process and the Boltzmann invariant measure

Recall that the Boltzmann probability measure μ has density proportional to e^{-U} with respect to the Lebesgue measure and is reversible for the Kolmogorov process with generator

$$\mathcal{L}f \,=\, \Delta f - < \nabla U, \nabla f > .$$

To begin, assume that the measure μ is spherically log-concave, i.e. there exists a C^2 function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ convex non-decreasing with $\phi(0) = 0$ and such that $U(x) = \phi(\|x\|)$ for any $x \in \mathbb{R}^d$. By a famous result of Bobkov [5], the dynamics (μ, Γ) satisfy (at least) a Poincaré inequality. Let us consider the potential U as an observable and also the test function $V = e^{cU}$, which belongs to $L^1(\mu)$ for any $c \in (0,1)$ since ϕ is convex. Assume that there exists a truncation level $R > 0$ and $M = M(R) \in (0, 1-c)$ such that

$$\frac{\Delta U(x)}{\|\nabla U(x)\|^2} = \frac{(d-1)\,\phi'(\|x\|) + \|x\|\,\phi''(\|x\|)}{\|x\|\,\phi'(\|x\|)^2} \leq M, \quad \|x\| \geq R.$$

Since we have

$$\begin{aligned} -\frac{\mathcal{L}V(x)}{V(x)} &= -c\,\Delta U(x) + c(1-c)\,\|\nabla U(x)\|^2 \\ &= -c(d-1)\,\frac{\phi'(\|x\|)}{\|x\|} - c\,\phi''(\|x\|) + c(1-c)\,\phi'(\|x\|)^2, \end{aligned}$$

one deduces that U belongs to the class $\mathcal{L}_V(a,b)$ with $a = 1/c(1-c-M)$, the parameter $b = b(R)$ being chosen conveniently on B_R, the centered ball of radius R in $\mathbb{R}^d$. Actually one can take b as the maximum between $3a\lambda_1$ (a quantity appearing in the restriction of Theorem 3.12) and

$$\left\| \left(a\,\frac{\mathcal{L}V}{V} + \|\nabla U\|^2 \right) 1_{B_R} \right\|_{L^\infty(\mu)}.$$

As a result, one can apply Theorem 3.12 to obtain the following concentration estimate: for any $0 \leq r \leq \sqrt{64b/27\lambda_1}$,

$$\mu\left(\{x \in \mathbb{R}^d : U(x) - \mu(U) > r\}\right) \leq \exp\left(-\frac{9\lambda_1 r^2}{32b}\right),$$

whereas for any $r \geq \sqrt{64b/27\lambda_1}$,

$$\mu\left(\{x \in \mathbb{R}^d : U(x) - \mu(U) > r\}\right) \leq \exp\left(-r\sqrt{\frac{3\lambda_1}{16b}}\right).$$

As mentioned in Remark 3.13, only the exponential decay is interesting since we do not obtain the correct order of magnitude in terms of the variance of U in the Gaussian window. See also the recent and elegant article of Bobkov and Madiman [9] where a somewhat similar tail estimate is established via a different approach, but with the sharp Gaussian regime.

Actually, spherically log-concave probability measures include the case of a potential U such that $U(x) = \|x\|^\beta$ with $\beta \geq 1$. Since the case $\beta = 2$ has already been considered, three different situations arise:

(i) the case $\beta = 1$, for which only the Poincaré inequality is satisfied.

(ii) the case $\beta \in (1,2)$: the standard Beckner inequality holds, cf. [33].

(iii) the case $\beta > 2$: using Wang's criterion [41], the log-Sobolev inequality is then verified.

In these three cases, one may choose the following parameters:

$$c := \frac{1}{2}, \quad M := \frac{1}{4}, \quad a := 8 \quad \text{and} \quad R := 4^{1/\beta}\left(\frac{d+\beta-2}{\beta}\right)^{1/\beta},$$

provided the restriction $d + \beta - 2 > 0$ holds. Finally, if $\beta > 2$ then the parameter b can be easily computed according to the previous choice, in contrast to the case $\beta \in [1,2)$ for which U is not C^2 at 0. Therefore, an approximation procedure is required to obtain a convenient constant b in this non-smooth situation. To do so, we apply the proof above with an increasing sequence of test functions $V_n = e^{cU_n}$ converging pointwise to V, where U_n is C^2 on $\mathbb{R}^d$. Then we use an easy perturbation argument to get the standard Beckner inequality (or the Poincaré inequality in the case $\beta = 1$) for the Boltzmann probability measure defined with respect to the potential U_n and finally, we apply Fatou's lemma in the concentration estimate obtained for the observable U_n.

One may also extend the result to the case of a general potential U, when for example a logarithmic Sobolev inequality holds. Let us assume for example that the potential U is such that its Hessian matrix, denoted $\operatorname{Hess} U$, is lower bounded and that the following Lyapunov condition holds:

$$\mathcal{L}V(x) \leq \left(-c_1 \|x\|^2 + c_2\right) V(x), \quad x \in \mathbb{R}^d,$$

where $c_1, c_2 > 0$ and $V \geq 1$ is C^2. Then a logarithmic Sobolev inequality holds, cf. [13], and one can apply Theorem 3.6 for observables f such that the norm of $\nabla f(x)$ is controlled by $\|x\|$, since in this case $f \in \mathcal{L}_V(a,b)$ with $a = 1/c_1$ and $b = c_2/c_1$. For instance, the Lyapunov condition above will be verified if at least one of the two conditions below is satisfied: there exist $\alpha \in (0,1)$ and $\beta > 0$ such that for sufficiently large x,

$$(1-\alpha) \|\nabla U(x)\|^2 - \Delta U(x) \ \geq \ \beta \|x\|^2 \quad \text{or} \quad < x, \nabla U(x) > \geq \ \beta \|x\|^2.$$

4.1.3 Log-Sobolev inequality for modified dynamics

Our last example concerns the case where a log-Sobolev inequality does not hold for the classical dynamics but for a slightly modified dynamics, both with the same Boltzmann invariant probability measure μ. We will focus here on the simple case $U(x) := \|x\|^\alpha$ for $1 < \alpha < 2$, so that the standard Beckner inequality (thus the Poincaré inequality) holds for the classical dynamics (μ, Γ), but not a log-Sobolev inequality. However according to [14], the following weighted log-Sobolev inequality holds:

$$\operatorname{Ent}_\mu(f^2) \leq C \int_{\mathbb{R}^d} \left(1 + \|x\|^{2-\alpha}\right) \|\nabla f(x)\|^2 \, \mu(dx),$$

where $C > 0$ is some constant depending on dimension d. Now consider the process with the following generator:

$$\mathcal{L}^{\sigma^2} f := \sigma^2 \, \Delta f + < \nabla(\sigma^2) - \sigma^2 \nabla U, \nabla f >,$$

where σ is some smooth function from $\mathbb{R}^d$ to $\mathbb{R}$. Once again the measure μ is reversible for this process, but the notable difference relies on the weight σ^2 in the carré du champ, i.e. $\Gamma^{\sigma^2}(f,f) := \sigma^2 \, \|\nabla f\|^2$, so that a Lipschitz function f may have an unbounded carré du champ $\Gamma^{\sigma^2}(f,f)$, in contrast to the example of the Kolmogorov process studied above. In particular, the domain of the Dirichlet form is a weighted H^1 space, i.e.

$$\mathcal{D}(\mathcal{E}_\mu) := \left\{ f \in L^2(\mu) : \int_{\mathbb{R}^d} \sigma^2 \, \|\nabla f\|^2 \, d\mu < \infty \right\}.$$

Letting now $\sigma(x)^2 := 1 + \|x\|^{2-\alpha}$, one observes that the weighted log-Sobolev inequality above rewrites as the log-Sobolev inequality for the new dynamics (μ, Γ^{σ^2}). Choosing the test function $V(x) = e^{c\|x\|^\alpha}$, which belongs to $L^1(\mu)$ for any $c \in (0,1)$, we have for any x outside a neighborhood of 0,

$$\begin{aligned} -\frac{\mathcal{L}^{\sigma^2} V(x)}{V(x)} &= -c\alpha(d+\alpha-2)\,(1+\|x\|^{2-\alpha})\|x\|^{\alpha-2} + \alpha^2 c(1-c)(1+\|x\|^{2-\alpha})\|x\|^{2(\alpha-1)} \\ & \quad -c\alpha(2-\alpha), \end{aligned}$$

which behaves like $\alpha^2 c(1-c)\|x\|^\alpha$ at infinity. Hence using the same reasoning as in the case of the Kolmogorov process above, one deduces that observables f having a gradient $\|\nabla f(x)\|$ controlled by $\|x\|^{\alpha-1}$ satisfy Theorem 3.6. In particular, the observable $f(x) = \|x\|^\alpha$ belongs to this class, as expected: according to a result of Latala and Oleszkiewicz [33], the measure μ concentrates like $\exp(-cr^\alpha)$ for large deviation level r through Lipschitz observables in $\mathrm{Lip}(\mathbb{R}^d)$. We point out that our results might be made more precise by following the approach provided in Remark 3.11. To that aim, one has to consider the functional inequality $I_\mu(a)$ involved in Latala and Oleszkiewicz's work [33], which is more general than the standard Beckner inequality emphasized above.

In fact using the modified dynamics (μ, Γ^{σ^2}), one can even consider interesting cases for which even the Poincaré inequality does not hold for the original dynamics (μ, Γ). For instance consider the generalized Cauchy measure μ with density proportional to $(1+\|x\|^2)^{-\beta}$, $\beta > d/2$. Such a measure satisfies some weighted Poincaré inequality [8] as well as the following weighted log-Sobolev inequality [14]:

$$\mathrm{Ent}_\mu(f^2) \leq C \int_{\mathbb{R}^d} \left(1+\|x\|^2\right) \log\left(e+\|x\|^2\right) \|\nabla f(x)\|^2 \, \mu(dx), \tag{4.2}$$

where C is some positive constant depending on β and d. Letting

$$\sigma(x)^2 := \left(1+\|x\|^2\right) \log\left(e+\|x\|^2\right),$$

then the weighted inequality (4.2) rewrites as the log-Sobolev inequality for the dynamics (μ, Γ^{σ^2}). Now let $V(x) = 1 + \|x\|^k$ for some $0 < k < 2\beta - d$, so that V lies in $L^1(\mu)$. Then we have for any x outside a neighborhood of 0,

$$\begin{aligned} -\frac{\mathcal{L}^{\sigma^2} V(x)}{V(x)} &= -k(d+k-2)\,\frac{1+\|x\|^2}{\|x\|^2} \log\left(e+\|x\|^2\right) - \frac{2k\left(1+\|x\|^2\right)}{e+\|x\|^2} \\ &\quad +2k(\beta-1) \log\left(e+\|x\|^2\right), \end{aligned}$$

which is of order $k(2\beta - d - k)\log\left(e+\|x\|^2\right)$ for large $\|x\|$. Then we obtain by Theorem 3.6 a Gaussian-exponential concentration estimate through observables f having their gradient $\|\nabla f(x)\|$ dominated by $1/\|x\|$. Note that the function $f(x) = \log(\|x\|)$ belongs to this class of observables, leading to the well-known heavy tail phenomenon satisfied by Cauchy-type measures, cf. [8]. Finally, we mention that one can take profit of Remark 3.11 to get intermediate concentration regime for observables not saturating the Lyapunov condition.

4.2 Birth-death processes

Let us begin the study of jump processes by considering a simple but however non trivial example, namely birth-death processes. Here $(X_t)_{t\geq 0}$ is a Markov process on the state space $\mathbb{N} := \{0, 1, 2, \ldots\}$ endowed with the classical metric $d(x,y) = |x-y|$, $x, y \in \mathbb{N}$. The transition probabilities are given by

$$\mathbb{P}_x(X_t = y) = \begin{cases} \lambda_x t + o(t) & \text{if } y = x+1, \\ \nu_x t + o(t) & \text{if } y = x-1, \\ 1-(\lambda_x+\nu_x)t + o(t) & \text{if } y = x, \end{cases}$$

where $\lim_{t\to 0} t^{-1} o(t) = 0$. The transition rates λ and ν are respectively called the birth and death rates and satisfy to $\lambda > 0$ on $\mathbb{N}$ and $\nu > 0$ on $\mathbb{N}^* := \{1, 2, \ldots\}$ and $\nu_0 = 0$, so that the process is irreducible. Although we assume that the stability condition (2.7), which rewrites as

$$\lambda_x + \nu_x < \infty, \quad x \in \mathbb{N},$$

is satisfied, the generator might be unbounded in the sense of (2.9), i.e.

$$\sup_{x\in\mathbb{N}} \lambda_x + \nu_x = \infty.$$

The process is positive recurrent and non-explosive when the rates satisfy to

$$\sum_{x=1}^{\infty} \frac{\lambda_0\lambda_1\cdots\lambda_{x-1}}{\nu_1\nu_2\cdots\nu_x} < \infty \quad \text{and} \quad \sum_{x=1}^{\infty} \left(\frac{1}{\lambda_x} + \frac{\nu_x}{\lambda_x\lambda_{x-1}} + \cdots + \frac{\nu_x\cdots\nu_1}{\lambda_x\cdots\lambda_1\lambda_0} \right) = \infty,$$

respectively. In this case the detailed balance condition (2.8) rewrites as

$$\lambda_x\, \mu(\{x\}) = \nu_{x+1}\, \mu(\{x+1\}), \quad x \in \mathbb{N},$$

where μ is the unique stationary distribution of the process given by

$$\mu(\{x\}) = \mu(\{0\}) \prod_{y=1}^{x} \frac{\lambda_{y-1}}{\nu_y},\ x \in \mathbb{N}, \tag{4.3}$$

$\mu(\{0\})$ being the normalization constant. In the situations of interest, the death rate ν has to be bigger than λ to ensure such criteria.

For any function $f : \mathbb{N} \to \mathbb{R}$, the generator $\mathcal{L}$ of the process is given by

$$\mathcal{L}f(x) \quad = \quad \lambda_x\ (f(x+1) - f(x)) + \nu_x\ (f(x-1) - f(x))\,, \quad x \in \mathbb{N},$$

and the carré du champ is

$$\Gamma(f,f)(x) \quad = \quad \frac{1}{2} \left\{ \lambda_x\ (f(x+1) - f(x))^2 + \nu_x\, (f(x-1) - f(x))^2 \right\}, \quad x \in \mathbb{N}.$$

In particular, the Dirichlet form is given by

$$\mathcal{E}_\mu(f,g) := \sum_{x\in\mathbb{N}} \lambda_x\ (f(x+1) - f(x))\ (g(x+1) - g(x))\ \mu(\{x\}),$$

where f, g belong to the space $\mathcal{D}(\mathcal{E}_\mu)$ of functions $u : \mathbb{N} \to \mathbb{R}$ such that $\mathcal{E}_\mu(u,u)$ is finite.

Let us recall two recent results on the concentration of invariant measures of birth-death processes under Lipschitz observables in $\mathrm{Lip}(\mathbb{N})$. On the one hand, under a convenient ergodic condition (a positive Wasserstein - or Ricci - curvature of the process, both definitions referring more or less to the same object), Joulin [29] gives some concentration estimates of Poisson-type (i.e. similar to the tail of the Poisson measure) for processes with bounded rates λ and ν. On the other hand, when we apply Ollivier's result [37] to birth-death processes with positive Ricci curvature, his Gaussian-exponential concentration property is available for processes having (at most) linear rates λ and ν. As we will see below, Theorems 3.6 and 3.12 apply for observables f such that the associated carré du champ $\Gamma(f,f)$ has a growth comparable to that of ν. In particular when focusing on Lipschitz observables in $\mathrm{Lip}(\mathbb{N})$, our concentration estimates hold without restriction on the growth of the rates (except that induced by the functional inequalities assumed in these two theorems), in contrast to the two aforementioned results of Joulin and Ollivier. Note however that our results are not directly comparable to theirs since the assumptions are not the same.

First let us provide some basic conditions which ensure an entropic or Poincaré inequality. The following condition is due to Caputo, Dai Pra and Posta [11] and has been recently recovered by Chafaï and Joulin [17] by using a semigroup approach: if λ is non-increasing and ν is non-decreasing and there exists $\alpha > 0$ such that

$$\inf_{x\in\mathbb{N}} \lambda_x - \lambda_{x+1} + \nu_{x+1} - \nu_x \geq \alpha, \tag{4.4}$$

then the entropic inequality (2.2) is satisfied with constant α, or equivalently $\rho_0 \geq \alpha$. Such an assumption exhibits very asymmetric rates. More precisely, it enforces the rates λ and ν to be bounded and super-linear, respectively, excluding some interesting cases which can be however considered for the Poincaré inequality. Indeed, Miclo [36] states that the spectral gap λ_1 is positive if and only if

$$\delta := \sup_{x\geq 1} \sum_{k=0}^{x-1} \frac{1}{\lambda_k \mu(\{k\})} \sum_{l\geq x} \mu(\{l\}) \quad < \quad \infty, \tag{4.5}$$

and in this case we have $1/\delta \geq \lambda_1 \geq 1/4\delta$, i.e. λ_1 is of order $1/\delta$. In particular, in contrast to the entropic inequality, one may find examples satisfying the Poincaré inequality with an unbounded birth rate λ. Now assume that the test function $V(x) := \kappa^x$ is in $L^1(\mu)$ for some constant $\kappa > 1$ depending on λ, ν. Then an observable f belongs to $\mathcal{L}_V(a,b)$ if and only if

$$\Gamma(f,f) \quad \leq \quad \frac{a(\kappa-1)}{\kappa} \; (\nu - \kappa\,\lambda) + b,$$

showing that on a large scale the behaviour of $\Gamma(f,f)$ is controlled by the growth of the death rate ν. Certainly, the parameters a, b cannot be specified explicitly without any other information on the observable f. According to this observation and to make the link with the results of Joulin and Ollivier mentioned previously, assume that the observable $f \in \mathrm{Lip}(\mathbb{N})$. Then two extreme situations may appear when the death rate ν is unbounded:

(i) a small birth rate λ, i.e. λ is bounded. In this case one may choose the following parameters to ensure that $f \in \mathcal{L}_V(a,b)$:

$$a := \frac{\kappa}{2(\kappa-1)} \quad \text{and} \quad b := \frac{(1+\kappa)\|\lambda\|_{L^\infty(\mu)}}{2}.$$

(ii) a birth rate λ of the order of ν. Let $x_0 \in \mathbb{N}^*$ and assume that $\lambda_x \leq c\nu_x$ for all $x \geq x_0$, where $c \in (0,1)$ is some parameter. Then in order to get $f \in \mathcal{L}_V(a,b)$, one can choose for $\kappa \in (1, 1/c)$,

$$a := \frac{\kappa(1+c)}{2(1-c\kappa)(\kappa-1)},$$

$$\text{and} \qquad b := \left\| \left(\frac{\lambda+\nu}{2} + a\,\frac{\mathcal{L}V}{V} \right) 1_{[0,x_0]} \right\|_{L^\infty(\mu)} = \frac{1+\kappa}{2(1-c\kappa)} \left\| (\lambda - c\nu)\, 1_{[0,x_0]} \right\|_{L^\infty(\mu)}.$$

In both cases (i) and (ii) there exist plenty of examples satisfying Poincaré inequality and thus Theorem 3.12, whereas only the case (i) may satisfy the entropic inequality and so Theorem 3.6. Of course b has also to be at least $3a\lambda_1$ if only the Poincaré inequality is satisfied (due to the restriction appearing in the statement of Theorem 3.12). For instance in the case (i), the desired Gaussian-exponential concentration result obtained from Theorem 3.6 is the following: if the rates λ and ν are respectively non-increasing and non-decreasing and satisfy the inequality (4.4), then for any observable $f \in \mathrm{Lip}(\mathbb{N})$ and any deviation level $r \geq 0$,

$$\mu\left(\{x \in \mathbb{N} : f(x) - \mu(f) > r\}\right) \leq \exp\left(- \min\left\{ \frac{3\alpha\, r^2}{8(1+\kappa)\lambda_0}, \sqrt{\frac{\kappa-1}{2\kappa}}\, r \right\} \right). \tag{4.6}$$

Here the parameter $\kappa > 1$ is free and an optimization could be done to improve the constants in the estimate.

To conclude with the birth-death example, let us focus our attention on a model which mimics the diffusion case, namely ultra log-concave distributions on $\mathbb{N}$, see for instance [28, 11]. We say that a probability measure μ on $\mathbb{N}$ is ultra log-concave (resp. log-concave) if it satisfies for any $x \in \mathbb{N}^*$,

$$x\,\mu(\{x\})^2 \geq (x+1)\,\mu(\{x+1\})\,\mu(\{x-1\}) \quad \text{(resp.} \quad \mu(\{x\})^2 \geq \mu(\{x+1\})\,\mu(\{x-1\})).$$

For instance the Poisson distribution is ultra log-concave whereas the geometric measure is only log-concave. Assume that the measure μ has density e^{-U}/Z with respect to the counting measure on $\mathbb{N}$, where U is some nice function and Z is the normalization constant. Denote ΔU the discrete Laplacian of the potential U, i.e.

$$\Delta U(x) := U(x+1) - 2U(x) + U(x-1), \quad x \in \mathbb{N}^*.$$

Then μ is ultra log-concave (resp. log-concave) if and only if $\Delta U(x) \geq \log(1 + 1/x)$ for any integer $x \in \mathbb{N}^*$ (resp. ΔU is non-negative).

From a dynamical point of view, measure μ is the stationary distribution of the birth-death process with rates

$$\lambda_x = 1 \quad \text{and} \quad \nu_x = e^{U(x)-U(x-1)}\,1_{\{x\neq 0\}}, \quad x \in \mathbb{N}.$$

Then under the ultra log-concavity assumption, we have for any integer $x \geq 2$,

$$\begin{aligned}
\lambda_x - \lambda_{x+1} + \nu_{x+1} - \nu_x &= \left(e^{\Delta U(x)} - 1\right) e^{\sum_{k=1}^{x-1} \Delta U(k) + U(1) - U(0)} \\
&\geq \prod_{k=1}^{x-1} \left(1 + \frac{1}{k}\right) \frac{e^{U(1)-U(0)}}{x} \\
&= e^{U(1)-U(0)} \\
&= \nu_1.
\end{aligned}$$

One deduces that the rate ν is non-decreasing and moreover (4.4) is satisfied with $\alpha = \nu_1$ (the cases $x \in \{0,1\}$ being straightforward). Hence the concentration estimate (4.6) applies. Finally, note that the log-concavity assumption only is not sufficient to ensure an entropic inequality since one obtains in this case the inequality

$$\inf_{x \in \mathbb{N}} \lambda_x - \lambda_{x+1} + \nu_{x+1} - \nu_x \geq 0.$$

However, as in the diffusion case, one may find examples of log-concave distributions on $\mathbb{N}$ satisfying the Poincaré inequality by using condition (4.5), which simply rewrites as

$$\sup_{x \in \mathbb{N}^*} \sum_{0 \leq k \leq x-1 < l} e^{U(k)-U(l)} < \infty,$$

so that Theorem 3.12 can be used.

4.3 Glauber dynamics for unbounded particles

We consider the situation where $\mathcal{X}$ is the unbounded configuration space $\mathbb{N}^\Lambda$, where Λ is a bounded subset of $\mathbb{Z}^d$. For each site $x \in \Lambda$, denote η_x the number of particles located at x. Given the activity $z > 0$, let π be the Poisson measure on $\mathbb{N}^\Lambda$ given by

$$\pi(\{\eta\}) = e^{-z|\Lambda|} \prod_{x \in \Lambda} \frac{z^{\eta_x}}{\eta_x!}, \quad \eta \in \mathbb{N}^\Lambda,$$

where $|\Lambda|$ denotes the cardinality of Λ. We equip $\mathbb{N}^\Lambda$ with the total variation distance, i.e. if η and $\bar\eta$ are two configurations in $\mathbb{N}^\Lambda$, then

$$d(\eta, \bar\eta) := \sum_{x \in \Lambda} |\eta_x - \bar\eta_x|.$$

Our definition is a straightforward generalization of the classical notion of total variation distance between probability measures, since it coincides with the usual definition when the configurations are normalized by their total masses. For any function $f : \mathbb{N}^\Lambda \to \mathbb{R}$, the discrete gradient operators are defined by

$$D_x^+ f(\eta) \quad := \quad f(\eta + \delta_x) - f(\eta), \quad D_x^- f(\eta) \quad := \quad f(\eta - \delta_x) - f(\eta), \quad \eta \in \mathbb{N}^\Lambda,$$

where δ_x is the Dirac mass at point $x \in \Lambda$ and by convention $D_x^- f(\emptyset) := 0$. Note that by [20], a given function f belongs to the space $\mathrm{Lip}(\mathbb{N}^\Lambda)$ if and only if

$$\sup_{(\eta,x) \in \mathbb{N}^\Lambda \times \Lambda} |D_x^+ f(\eta)| < \infty.$$

Now let $\phi : \mathbb{Z}^d \to [0, \infty)$ be an even function, null at the origin and satisfying $\sum_{x \in \mathbb{Z}^d} \phi(x) < \infty$. We define the Hamiltonian $H : \mathbb{N}^\Lambda \to \mathbb{R}_+$ as

$$H(\eta) \quad := \quad \frac{1}{2} \sum_{x,y \in \Lambda} \phi(x-y)\, \eta_x\, \eta_y.$$

Then the Gibbs measure μ at inverse temperature $\beta > 0$ is the probability measure on $\mathbb{N}^\Lambda$ given by

$$\mu(\{\eta\}) \quad = \quad \frac{1}{Z}\, e^{-\beta\, H(\eta)}\, \pi(\{\eta\}),$$

where Z is the normalization constant. As observed below, our study is based on the configuration space $\mathbb{N}^\Lambda$ since our model exhibits free boundary condition, that is to say Λ is, in some sense, disconnected from the lattice $\mathbb{Z}^d$. However the aforementioned model might be extended outside Λ by introducing an appropriate boundary condition.

Now, let us introduce the Glauber dynamics associated to the Gibbs measure above, which can be seen as a spatial birth-death process, cf. Preston [39]. If η is the configuration of the system at time t, then a particle appears or disappears at site $x \in \Lambda$ with rates $z\, e^{-\beta D_x^+ H(\eta)} dt$ and dt, respectively. In particular, the case $H = 0$ corresponds to the non-interacting case. The generator $\mathcal{L}$ is thus of birth-death type and defined for any function $f : \mathbb{N}^\Lambda \to \mathbb{R}$ by

$$\mathcal{L} f(\eta) \quad := \quad \sum_{x \in \Lambda} \left(c^-(\eta, x)\, D_x^- f(\eta) + c^+(\eta, x)\, D_x^+ f(\eta) \right), \quad \eta \in \mathbb{N}^\Lambda,$$

where the rates of the dynamics c^+ and c^- are given by

$$\begin{cases} c^+(\eta, x) & = & z\, e^{-\beta D_x^+ H(\eta)} \;=\; z\, e^{-\beta \sum_{y \in \Lambda} \phi(x-y)\, \eta_y}; \\ c^-(\eta, x) & = & \eta_x. \end{cases}$$

In particular, the stability condition (2.7) is clearly satisfied since Λ is finite and moreover, according to the detailed balance condition (2.8) which in our context rewrites as

$$c^\pm(\eta, x)\, \mu(\{\eta\}) \quad = \quad c^\mp(\eta \pm \delta_x, x)\, \mu(\{\eta \pm \delta_x\}), \quad \eta_x > 0, \quad (\eta, x) \in \mathbb{N}^\Lambda \times \Lambda,$$

the Gibbs measure μ is reversible for these dynamics. Finally, the carré du champ of an observable f is given by

$$\Gamma(f,f)(\eta) \;=\; \frac{1}{2} \sum_{x \in \Lambda} \left\{ c^-(\eta, x)\, |D_x^- f(\eta)|^2 + c^+(\eta, x)\, |D_x^+ f(\eta)|^2 \right\}, \quad \eta \in \mathbb{N}^\Lambda.$$

Recently, the problem of finding the speed of convergence to equilibrium of this model has been addressed in several articles, cf. for instance [4] or Wu's paper [42] for a spectral method (i.e. related to Poincaré inequality) in the continuum $\mathbb{R}^d$, and also [18] for an approach through the entropic inequality. In all these papers, the objective is to find constants which are independent of Λ and of the boundary condition. In a recent work [19], Dai Pra and Posta established the entropic inequality with constant $\rho_0 \geq 1 - z\,\varepsilon(\beta)$, under the following Dobrushin-type uniqueness condition:

$$\varepsilon(\beta) := \sum_{x\in\mathbb{Z}^d} \left(1 - e^{-\beta\phi(x)}\right) \quad < \quad \frac{1}{z}. \tag{4.7}$$

Note that assumption (4.7) will be verified as soon as β is small enough, i.e. the temperature of the system is sufficiently high. If we choose for some $\kappa > 1$ the test function $V(\eta) := \kappa^{\sum_{x\in\Lambda}\eta_x}$ which is in $L^1(\mu)$, then an observable f belongs to the class $\mathcal{L}_V(a,b)$ if and only if

$$\Gamma(f,f)(\eta) \quad \leq \quad \frac{a(\kappa-1)}{\kappa} \sum_{x\in\Lambda} \left(\eta_x - \kappa\, z\, e^{-\beta D_x^+ H(\eta)}\right) + b, \quad \eta \in \mathbb{N}^\Lambda,$$

as in the context of birth-death processes above. Thus if the Dobrushin-type uniqueness condition (4.7) is satisfied, then the Gaussian-exponential concentration estimate of Theorem 3.6 applies under these observables. Finally, we have $D_x^+ H(\eta) \geq 0$ because ϕ is non-negative and if we consider an observable f in the space $\mathrm{Lip}(\mathbb{N}^\Lambda)$, then the parameters a, b are chosen independently of f by taking

$$a := \frac{\kappa}{2(\kappa-1)} \quad \text{and} \quad b := \frac{(1+\kappa)}{2}\, z|\Lambda|,$$

and we obtain the following Gaussian-exponential concentration estimate: for any $r \geq 0$,

$$\mu\left(\left\{\eta \in \mathbb{N}^\Lambda : f(\eta) - \mu(f) > r\right\}\right) \leq \exp\left(-\min\left\{\frac{3(1 - z\varepsilon(\beta))\, r^2}{8(1+\kappa)z|\Lambda|}, \sqrt{\frac{\kappa-1}{2\kappa}}\, r\right\}\right).$$

Once again an optimization in terms of the free parameter $\kappa > 1$ can be done to refine the estimate. Note however that we do not recover the Poisson-type behaviour expected when comparing to the non-interacting case, for which μ is nothing but the Poisson distribution π.

References

[1] S. Aida and D. Stroock. Moment estimates derived from Poincaré and logarithmic Sobolev inequalities. *Math. Res. Lett.* 1 (1994), 75-86. MR-1258492

[2] D. Bakry, F. Barthe, P. Cattiaux and A. Guillin. A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case. *Elec. Comm. Probab.* 13 (2008), 60-66. MR-2386063

[3] W. Beckner. A generalized Poincaré inequality for Gaussian measures. *Proc. Amer. Math. Soc.* 105 (1989), 397-400. MR-0954373

[4] L. Bertini, N. Cancrini and F. Cesi. The spectral gap for a Glauber-type dynamics in a continuous gas. *Ann. Inst. Henri Poincaré Probab. Stat.* 38 (2002), 91-108. MR-1899231

[5] S.G. Bobkov. Spectral gap and concentration for some spherically symmetric probability measures. *Geometric aspects of functional analysis*, 37-43, Lecture Notes in Math., 1807, Springer, Berlin, 2003. MR-2083386

[6] S.G. Bobkov and M. Ledoux. Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields* 107 (1997), 383-400. MR-1440138

[7] S.G. Bobkov and M. Ledoux. On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures. *J. Funct. Anal.* 156 (1998), 347-365. MR-1636948

[8] S.G. Bobkov and M. Ledoux. Weighted Poincaré-type inequalities for Cauchy and other convex measures. *Ann. Probab.* 37 (2009), 403-427. MR-2510011

[9] S.G. Bobkov and M. Madiman. Concentration of the information in data with log-concave distributions. *Ann. Probab.* 39 (2011), 1528-1543. MR-2857249

[10] S.G. Bobkov and P. Tetali. Modified logarithmic Sobolev inequalities in discrete settings. *J. Theor. Probab.* 19 (2006), 289-336. MR-2283379

[11] P. Caputo, P. Dai Pra and G. Posta. Convex entropy decay via the Bochner-Bakry-Émery approach. *Ann. Inst. Henri Poincaré Probab. Stat.* 45 (2009), 734-753. MR-2548501

[12] P. Cattiaux and A. Guillin. Deviation bounds for additive functionals of Markov processes. *ESAIM Probab. Stat.* 12 (2008), 12-29. MR-2367991

[13] P. Cattiaux, A. Guillin and L. Wu. A note on Talagrand transportation inequality and logarithmic Sobolev inequality. *Probab. Theory Related Fields* 148 (2010), 285-304. MR-2653230

[14] P. Cattiaux, A. Guillin and L. Wu. Some remarks on weighted logarithmic Sobolev inequality. *Indiana University Math. J.* 60 (2011), 1885-1904. MR-3008255

[15] P. Cattiaux, A. Guillin and P.A. Zitt. Poincaré inequalities and hitting times. *Ann. Inst. Henri Poincaré Probab. Stat.* 49 (2013), 95-118.

[16] D. Chafaï. Entropies, convexity, and functional inequalities. *J. Math. Kyoto Univ.* 44 (2004), 325-363. MR-2081075

[17] D. Chafaï and A. Joulin. Intertwining and commutation relations for birth-death processes. To appear in *Bernoulli*.

[18] P. Dai Pra, A.M. Paganoni and G. Posta. Entropy inequalities for unbounded spin systems. *Ann. Probab.* 30 (2002), 1969-1976. MR-1944012

[19] P. Dai Pra and G. Posta. Entropy decay for interacting systems via the Bochner-Bakry-Émery approach. *Electron. J. Probab.* 18 (2013), 1-21.

[20] L. Decreusefond, A. Joulin and N. Savy. Upper bounds on Rubinstein distances on configuration spaces and applications. *Comm. Stoch. Anal.* 4 (2010), 377-399. MR-2677197

[21] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* 6 (1996), 695-750. MR-1425868

[22] F. Gao, A. Guillin and L. Wu. Bernstein type's concentration inequalities for symmetric Markov processes. To appear in *SIAM Theory Probab. Appl.*

[23] N. Gozlan and C. Léonard. Transport inequalities. A survey. *Markov Process. Related Fields* 16 (2010), 635-736. MR-2895086

[24] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.* 97 (1975), 1061-1083. MR-0420249

[25] A. Guillin, C. Léonard, L. Wu and N. Yao. Transportation inequalities for Markov processes. *Probab. Theory Related Fields* 144 (2009), 669-695. MR-2496446

[26] D.L. Hanson and F.T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Statist.* 42 (1971), 1079-1083. MR-0279864

[27] C. Houdré. Remarks on deviation inequalities for functions of infinitely divisible random vectors. *Ann. Probab.* 30 (2002), 1223-1237. MR-1920106

[28] O. Johnson. Log-concavity and the maximum entropy property of the Poisson distribution. *Stochastic Process. Appl.* 117 (2007), 791-802. MR-2327839

[29] A. Joulin. Poisson-type deviation inequalities for curved continuous-time Markov chains. *Bernoulli* 13 (2007), 782-798. MR-2348750

[30] A. Joulin. A new Poisson-type deviation inequality for Markov jump processes with positive Wasserstein curvature. *Bernoulli* 15 (2009), 532-549. MR-2543873

[31] A. Joulin and Y. Ollivier. Curvature, concentration, and error estimates for Markov chain Monte Carlo. *Ann. Probab.* 38 (2010), 2418-2442. MR-2683634

[32] R. Latala. Estimates of moments and tails of Gaussian chaoses. *Ann. Probab.* 34 (2006), 2315-2331. MR-2294983

[33] R. Latala and K. Oleszkiewicz. Between Sobolev and Poincaré. *Geometric aspects of functional analysis*, 147-168, Lecture Notes in Math., 1745, Springer, Berlin, 2000. MR-1796718

[34] M. Ledoux. Concentration of measure and logarithmic Sobolev inequalities. *Séminaire de Probabilités* XXXIII, 120-216, Lecture Notes in Math., 1709, Springer, Berlin, 1999. MR-1767995

[35] M. Ledoux. *The concentration of measure phenomenon*. Mathematical Surveys and Monographs, 89. American Mathematical Society, Providence, RI, 2001. MR-1849347

[36] L. Miclo. An example of application of discrete Hardy's inequalities. *Markov Process. Related Fields* 5 (1999), 319-330. MR-1710983

[37] Y. Ollivier. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.* 256 (2009), 810-864. MR-2484937

[38] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* 173 (2000), 361-400. MR-1760620

[39] C. Preston. Spatial birth-and-death processes. Proceedings of the 40th session of the international statistical institute (Warsaw 1975). *Bull. Inst. Internat. Statist.* 46 (1975), 371-391, 405-408 (1975). MR-0474532

[40] M.D. Sammer. Aspects of mass transportation in discrete concentration inequalities. Ph.D. Thesis, Georgia Institute of Technology, 2005. Available at `http://hdl.handle.net/1853/7006`.

[41] F.Y. Wang. Logarithmic Sobolev inequalities on noncompact Riemannian manifolds. *Probab. Theory Related Fields* 109 (1997), 417-424. MR-1481127

[42] L. Wu. Estimate of spectral gap for continuous gas. *Ann. Inst. Henri Poincaré Probab. Stat.* 40 (2004), 387-409. MR-2070332

Acknowledgments. The authors are grateful to P. Dai Pra and G. Posta for useful discussions. They thank also the French ANR Projects Evol and GeMeCoD for financial support.