Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 135 (2021) 103–138

stochastic processes and their applications

www.elsevier.com/locate/spa

Concentration inequalities for additive functionals: A martingale approach

Bob Pepin

Received 28 October 2018; received in revised form 28 December 2020; accepted 11 January 2021
Available online 4 February 2021

Abstract

This work shows how exponential concentration inequalities for additive functionals of stochastic processes over a finite time interval can be derived from concentration inequalities for martingales. The approach is entirely probabilistic and naturally includes time-inhomogeneous and non-stationary processes as well as initial laws concentrated on a single point. The class of processes studied includes martingales, Markov processes and general square integrable càdlàg processes. The general approach is complemented by a simple and direct method for martingales, diffusions and discrete-time Markov processes. The method is illustrated by deriving concentration inequalities for the Polyak–Ruppert algorithm, SDEs with time-dependent drift coefficients "contractive at infinity" with both Lipschitz and squared Lipschitz observables, some classical martingales and non-elliptic SDEs.

Keywords: Time average; Additive functional; Inhomogeneous functional; Concentration inequality; Time-inhomogeneous Markov process; Martingale

1. Introduction

In this work we consider concentration inequalities for additive functionals of the form

$$\int_0^T X_t \, dt$$

where X is a real-valued stochastic process. The methods we develop apply to a broad class of processes, and we will give theorems and examples that go beyond the classical setting of stationary Markov processes. We will treat in depth the cases where X is a martingale or $X_t = f(t, Y_t)$ for a Markov process Y and f in an appropriate class of functions. The concentration

E-mail address: bob@pepin.io.

https://doi.org/10.1016/j.spa.2021.01.004
0304-4149/

inequalities will be derived for the additive functionals centered around their expectation, which allows us to naturally treat non-stationary processes such as time-inhomogeneous diffusions.

We proceed to give an overview of the main results. In Section 2 we derive some representative results using short, self-contained proofs based on direct calculations. First, we show in Proposition 2.1 that for any continuous local martingale X such that $X_0 = 0$ we have the following concentration inequality for any $T, R > 0$:

$$\mathbb{P}\left(\int_0^T X_u\,du \geq R\ ;\ \int_0^T (T-u)^2\,d[X]_u \leq \sigma^2\right) \leq \exp\left(-\frac{R^2}{2\sigma^2}\right).$$

To the author's knowledge, the systematic treatment of concentration inequalities for additive functionals of martingales has not appeared in the literature before.

We will then move on to solutions to SDEs of the form

$$dX_t = b(t, X_t)dt + \sigma\,dB_t,$$

with X_0 deterministic. In particular, denoting $P_{s,t}$ the Markov transition operator associated to X, we will show in Corollary 2.7 that if there exist constants $c, \kappa > 0$ such that

$$|\sigma^\top \nabla P_{t,u} f(x)| \leq c e^{-\kappa(u-t)}, \quad x \in \mathbb{R}^n, 0 \leq t \leq u$$

then we have the following Gaussian concentration inequality for all $R, T > 0$:

$$\mathbb{P}\left(\frac{1}{T}\int_0^T f(u, X_u)\,du - \mathbb{E}\left[\frac{1}{T}\int_0^T f(u, X_u)\,du\right] \geq R\right) \leq \exp\left(-\frac{\kappa^2 R^2 T}{2c^2}\right).$$

The main novelty in the corollary is the treatment of time-inhomogeneous SDEs and the method of proof. The Proposition from which the corollary is derived also provides a novel, refined statement in terms of a bound of $|\sigma^\top \nabla P_{t,u} f|$ along trajectories of X.

Section 2 concludes with the case of discrete-time processes in Proposition 2.12, again giving careful consideration to the time-inhomogeneous case and controlling the relevant quantities along trajectories of the process. Concretely, we show that for a discrete-time stochastic process X_t and function f such that

$$|X_t - X_{t-1}| \leq C_t, \quad t \geq 1,$$
$$|P_{s,t} f(x) - P_{s,t} f(y)| \leq \sigma_s (1-\kappa_s)^{t-s}\,|x-y|, \quad x, y \in \mathbb{R}^n, 0 \leq s \leq t.$$

we have

$$\mathbb{P}\left(\sum_{u=1}^t f(u, X_u) - \mathbb{E}\left[\sum_{u=1}^t f(u, X_u)\right] \geq R\ ;\ \sum_{t=1}^T \frac{\sigma_t^2 C_t^2}{\kappa_t^2} \leq a^2\right) \leq \exp\left(-\frac{R^2}{8a^2}\right).$$

The careful treatment of the time-inhomogeneous case and control on the level of trajectories enables in particular for the first time the derivation of concentration inequalities for the Polyak–Ruppert algorithm of the correct order in Section 4.1 (concentration inequalities for the linear case were published concurrently with this work in [30]).

Section 3 is dedicated to a wide-ranging generalization of the results from the previous section. In Section 3.1 we introduce a family of auxiliary martingales $Z_t^u = E^{\mathcal{F}_t} X_u$ and show that for general square integrable processes, the concentration properties of $\int_0^T X_u\,du$ are intimately linked to the predictable quadratic covariation $\langle Z^u, Z^v\rangle$ and jumps ΔZ^u of the auxiliary martingales. In Section 3.2 we recover and extend the martingale results from Section 2 to the discontinuous setting using the general method. In Section 3.3 we apply the general method to general Markov processes and recover expressions for $\langle Z^u, Z^v\rangle$ in terms of the squared field operator Γ, generalizing the results on Markov processes from Section 2 to general Markov processes on Polish spaces. All of the results from the preceding subsections are novel in their generality. We conclude Section 3 with Section 3.4 where we show how to incorporate arbitrary distributions of X_0 and recall a number of martingale inequalities. The subsection also includes in Corollary 3.16 a novel approach to obtain Bernstein-type inequalities in some "self-bounding" cases.

The final Section 4 illustrates how to apply the results from the preceding section on a number of concrete cases. Section 4.1 contains the novel results on Polyak–Ruppert mentioned above.

Section 4.2 provides a concrete example of an SDE case with explicit conditions on the drift and diffusion coefficients as well as the observable function. Using known results on gradient bounds for $P_{s,t}$ when the drift coefficient is "contractive at infinity" characterized by constants ρ, κ, we show that for any initial law ν satisfying a $T_1(C)$ transport inequality and Lipschitz function f, we have

$$\mathbb{P}\Big(\frac{1}{T}\int_0^T f(X_t)\,dt - \frac{1}{T}\int_0^T \mu_t(f)\,dt \geq R + \rho\|f\|_{\mathrm{Lip}}\frac{1-e^{-\kappa T}}{\kappa T} W_1(\mu_0,\nu) \Big) \leq \exp\left(-\frac{\kappa^2 R^2 T}{2\rho^2\|f\|_{\mathrm{Lip}}^2\left(1 + C\,\frac{1-e^{-\kappa T}}{T}\right)}\right)$$

for a unique evolution system of measures μ_t (if the process has a stationary measure μ then $\mu_t = \mu$ for all t).

Section 4.3 provides some concrete examples using classical martingales as integrands: Brownian motion, the compensated Poisson process and compensated squared Brownian motion $B_t^2 - t$.

Sections 4.4 and 4.4 treat the cases where the integrand X is either the squared Ornstein–Uhlenbeck process or more generally the square of a Lipschitz function. These cases go beyond the scope of most previously published approaches to concentration inequalities for additive functionals. The final Section 4.6 presents a simple case of a highly non-elliptic SDE, which yields easily to the probabilistic methods presented here but is outside of the scope of previous approaches based for example on Poisson equations.

About the literature. In the Markovian setting, our approach is most closely related to the work of Joulin [25], and we recover and extend the results from that work (Propositions 3.7 and 4.10). The cases of martingales and general square integrable processes do not seem to have been systematically studied in the literature.

Most previous results on concentration inequalities for functionals of the form S_t have been obtained for time-homogeneous Markov processes using functional inequalities. The works [7,19,21] require the existence of a stationary measure and an initial distribution that has an integrable density with respect to the stationary measure. The same holds true for the combinatorial and perturbation arguments in the classic paper [28]. In [40] the authors establish concentration inequalities around the expectation using stochastic calculus and Girsanov's theorem under strong contractivity conditions. Some concentration inequalities for inhomogeneous functionals have previously been established in [20]. A different approach using renewal processes has been used in the work [29] to establish concentration inequalities for functionals with bounded integrands.

For Markov processes, the mixing conditions in this work are most naturally formulated in terms of bounds on either the Lipschitz seminorm, gradient or squared field (carré du champs) operator of the semigroup. Bounds on the Lipschitz seminorm are closely related to contractivity in the L^1 transportation distance and can for instance be found in [1,15] for elliptic diffusions, in [39] for the Riemannian case, in [16] for Langevin dynamics or in [23] for stochastic delay equations. See also [26,31] for the discrete-time case and a large number of examples in both discrete and continuous time. Gradient estimates for semigroups can be obtained using Bismut-type formulas, see for example [12,17,36], the works [9,10] for the non-autonomous case as well as the textbook [37]. Finally, in terms of the squared field operator, our mixing conditions are a relaxation of the Bakry–Emery curvature-dimension condition [2] since we allow for a prefactor strictly greater than 1.

Notation. For a right-continuous process $(X_t)_{t\geq 0}$ with left limits we write $X_{t-} = \lim_{\varepsilon\to 0^+} X_{t-\varepsilon}$ and $\Delta X_t = X_t - X_{t-}$. For a σ-field $\mathcal{F}$ and a random variable X, we denote $\mathbb{E}^{\mathcal{F}} X$ the conditional expectation of X with respect to $\mathcal{F}$.

2. Direct approach

2.1. Continuous martingales

In this and the following subsections we will establish concentration inequalities by focusing on the cases of continuous local martingales, continuous solutions to SDEs and discrete-time stochastic processes, all with their initial law concentrated on a single point. In the first two cases, we provide alternative and more direct proofs of results that also follow from the general approach in Section 3. Compared to the general approach, the direct proofs also provide an explicit martingale representation for the additive functionals under consideration. The discrete-time case introduces the ideas of Section 3 in a technically simpler setting. The focus on a restricted class of processes keeps the proofs short and self-contained and formal computations easily justifiable. The concentration inequalities presented here will be considerably generalized to broader classes of processes, integrands and initial laws in Section 3.

We begin with the case of a continuous local martingale. Even though it is the simplest scenario in our framework, it has not been studied systematically in the literature, which tends to concentrate on concentration inequalities for additive functionals of stationary Markov processes.

Proposition 2.1. *For a continuous local martingale X such that $X_0 = 0$ we have the following concentration inequality for any $T, R > 0$:*

$$\mathbb{P}\left(\int_0^T X_u\,du \geq R\ ;\ \int_0^T (T-u)^2\,d[X]_u \leq \sigma^2\right) \leq \exp\left(-\frac{R^2}{2\sigma^2}\right).$$

Furthermore for $0 \leq t \leq T$

$$\int_0^t X_u\,du = \int_0^t (T-u)\,dX_u - (T-t)X_t. \tag{2.1}$$

Proof. Fix $T > 0$ and define a local martingale M_t^T by

$$M_t^T = \int_0^t (T-u)\,dX_u.$$

Using integration by parts we obtain (2.1)

$$M_t^T = (T-t)X_t + \int_0^t X_u\,du$$

so that in particular M^T coincides with $\int_0^{\cdot} X_u\,du$ at time T:

$$M_T^T = \int_0^T X_u\,du. \tag{2.2}$$

It is well-known that for any continuous local martingale M and $t \geq 0$

$$\mathbb{P}\left(M_t \geq R; [M]_t \leq \sigma^2\right) \leq \exp\left(-\frac{R^2}{2\sigma^2}\right). \tag{2.3}$$

Indeed, for $\lambda > 0$ let $\mathcal{E}^\lambda(M)_t = \exp\left(\lambda M_t - \frac{\lambda^2}{2}[M]_t\right)$. Since $d\mathcal{E}^\lambda(M)_t = \mathcal{E}^\lambda(M)_t\,dM_t$ the process $\mathcal{E}^\lambda(M)_t$ is a positive local martingale and therefore a supermartingale so that $\mathbb{E}\,\mathcal{E}^\lambda(M)_T \leq \mathbb{E}\,\mathcal{E}^\lambda(M)_0 = 1$. By Chebyshev's inequality, for any $R, \sigma^2 > 0$,

$$\begin{aligned}\mathbb{P}\left(M_t \geq R; [M]_t \leq \sigma^2\right) &\leq \exp\left(-\lambda R + \tfrac{\lambda^2}{2}\sigma^2\right)\,\mathbb{E}\left[\exp\left(\lambda M_t - \tfrac{\lambda^2}{2}\sigma^2\right); [M]_t \leq \sigma^2\right] \\ &\leq \exp\left(-\lambda R + \tfrac{\lambda^2}{2}\sigma^2\right)\,\mathbb{E}\,\mathcal{E}^\lambda(M)_t \leq \exp\left(-\lambda R + \tfrac{\lambda^2}{2}\sigma^2\right)\end{aligned}$$

where $\lambda > 0$ is arbitrary. The inequality (2.3) now follows by optimizing over λ.

From the definition of M^T and elementary properties of the quadratic variation, we get

$$[M^T]_t = \int_0^t (T-u)^2 d[X]_u. \tag{2.4}$$

By inserting (2.2) and (2.4) into (2.3) we finally obtain

$$\mathbb{P}\left(\int_0^T X_u\,du \geq R\,;\int_0^T (T-u)^2\,d[X]_u \leq \sigma^2\right) \leq \exp\left(-\frac{R^2}{2\sigma^2}\right). \quad \square$$

Corollary 2.2. *If $d[X]_u \leq \sigma^2(T-u)^\alpha\,du$ for some $\sigma^2 > 0, \alpha \in (-3, +\infty)$ and all $0 \leq u \leq T$ then for all $R > 0$*

$$\mathbb{P}\left(\frac{1}{T^{2+\alpha/2}}\int_0^T X_u\,du \geq R\right) \leq \exp\left(-\frac{(3+\alpha)R^2T}{2\sigma^2}\right).$$

Proof. We have

$$\int_0^T (T-u)^2\,d[X]_u \leq \sigma^2\int_0^T (T-u)^{2+\alpha}\,du = \frac{\sigma^2T^{3+\alpha}}{3+\alpha}$$

so that

$$\begin{aligned}&\mathbb{P}\left(\frac{1}{T^{2+\alpha/2}}\int_0^T X_u\,du \geq R\right) \\ &= \mathbb{P}\left(\int_0^T X_u\,du \geq RT^{2+\alpha/2}\,;\int_0^T (T-u)^2\,d[X]_u \leq \frac{\sigma^2T^{3+\alpha}}{3+\alpha}\right) \\ &\leq \exp\left(-\frac{(3+\alpha)R^2T}{2\sigma^2}\right). \quad \square\end{aligned}$$

Corollary 2.3. *If $d[X]_u \leq e^{-(T-u)}\,du$ for all $0 \leq u \leq T$ then for all $R > 0$*

$$\mathbb{P}\left(\frac{1}{\sqrt{T}}\int_0^T X_u\,du \geq R\right) \leq \exp\left(-\frac{R^2T}{4}\right).$$

Proof. We have

$$\int_0^T (T-u)^2\, d[X]_u \le \int_0^T (T-u)^2 e^{-(T-u)}\, du = 2-(T^2+2T+2)e^{-T} \le 2$$

and the result follows. $\square$

Remark 2.4. The continuity assumption primarily serves to keep the exposition self-contained by providing a concise proof of the concentration inequality (2.3) for local martingales. The method extends to discontinuous local martingales by making use of the more advanced martingale inequalities described further below in Section 3.4.

Remark 2.5. Compared to Proposition 3.3, the result presented here is a priori more general because it applies to local martingales, whereas Proposition 3.3 is stated for martingales. On the other hand, the process M^T constructed in the proof of this section is also only a local martingale, whereas the general method yields a process M^T which is a martingale by construction. Under the hypotheses of Proposition 3.3 that X_t is a true martingale with $\mathbb{E}X_t^2 < \infty$ for all $t \ge 0$, the present result also yields a true martingale, since then $\mathbb{E}[M^T]_t \le T^2\,\mathbb{E}[X]_t < \infty$ for all $t \ge 0$.

2.2. Stochastic differential equations

In this section, we consider the case of a solution X to the following elliptic SDE on $\mathbb{R}^n$:

$$dX_t = b(t, X_t)dt + \sigma dB_t, \quad X_0 = x_0 \tag{2.5}$$

for $x_0 \in \mathbb{R}^n$ fixed, $b\colon [0,\infty)\times\mathbb{R}^n \to \mathbb{R}^n$ locally bounded, once differentiable in its first argument and twice differentiable in the second with bounded first derivative, σ a real $n\times n$ matrix such that $\sigma\sigma^\top$ is positive definite and B a standard n-dimensional Brownian motion.

For a bounded function f on $[0,\infty)\times\mathbb{R}^n$, twice continuously differentiable, and $0 \le t \le u$ we define the two-parameter semigroup associated to X and f,

$$P_{t,u} f(x) = \mathbb{E}[f(u, X_u)|X_t = x].$$

If the coefficients of (2.5) are time-homogeneous, we have $P_{t,u} f = P_{0,u-t} f = P_{u-t} f$, where the latter is just the usual (time-homogeneous) Markov semigroup.

The result in Corollary 2.7 is known in the time-homogeneous case, and can for example be deduced using the arguments in [25]. The refined formulation in Proposition 2.6 and the inclusion of the time-inhomogeneous setting are new.

Proposition 2.6. *For all $T, R, c > 0$ and $f\colon [0,\infty)\times\mathbb{R}^n \to \mathbb{R}$, bounded and twice continuously differentiable, we have*

$$\mathbb{P}\left(\int_0^T f(u,X_u)\,du - \mathbb{E}\int_0^T f(u,X_u)\,du \ge R\,;\, \int_0^T |\sigma^\top \nabla R_t^T(X_t)|^2\,dt \le c^2\right) \le \exp\left(-\frac{R^2}{2c^2}\right)$$

with

$$R_t^T(x) = \int_t^T P_{t,u} f(x)\,du, \quad x\in\mathbb{R}^n, 0\le t\le T.$$

Furthermore we have the decomposition

$$\int_0^t f(u, X_u)\,du = \int_0^t \nabla R_s^T(X_s)\cdot\sigma\,dBs - R_t^T(X_t), \quad 0 \le t \le T.$$

Proof. For $T > 0$ fixed define a martingale M^T by

$$M_t^T = \mathbb{E}^{\mathcal{F}_t}\int_0^T f(u, X_u)\,du, \quad 0 \le t \le T$$

where $\mathcal{F}_t = \sigma(\{X_s\}_{s\le t})$ is the natural filtration of X. Using Fubini's theorem with the bounded integrand f, the fact that $f(u, X_u)$ is $\mathcal{F}_t$-measurable for all $u \le t$ and the Markov property we get

$$\begin{aligned} M_t^T &= \int_0^T \mathbb{E}^{\mathcal{F}_t} f(u, X_u)\,du = \int_0^t f(u, X_u)\,du + \int_t^T \mathbb{E}^{\mathcal{F}_t} f(u, X_u)\,du \\ &= \int_0^t f(u, X_u)\,du + R_t^T(X_t) \end{aligned} \tag{2.6}$$

with

$$R_t^T(x) = \int_t^T P_{t,u} f(x)du.$$

By our regularity assumptions on f and the coefficients of (2.5), we have the Kolmogorov backward equation $(\partial_t + L)P_{t,u} f = 0$ where L denotes the infinitesimal generator of X. It follows from standard properties of the integral that

$$(\partial_t + L)R_t^T(x) = -P_{t,t} f(x) + \int_t^T (\partial_t + L)P_{t,u} f(x)du = -f(t, x), \quad t > 0, x \in \mathbb{R}^n. \tag{2.7}$$

Using the Itô formula, we get

$$\begin{aligned} dR_t^T(X_t) &= (\partial_t + L)R_t^T(X_t)\,dt + \nabla R_t^T(X_t)\cdot\sigma\,dB_t \\ &= -f(t, X_t)\,dt + \nabla R_t^T(X_t)\cdot\sigma\,dB_t. \end{aligned}$$

By (2.6) we also have

$$dR_t^T(X_t) = -f(t, X_t)\,dt + dM_t^T \tag{2.8}$$

so that we can identify

$$dM_t^T = \nabla R_t^T(X_t)\cdot\sigma\,dB_t. \tag{2.9}$$

From elementary properties of the stochastic integral and Brownian motion

$$d[M^T]_t = |\sigma^\top \nabla R_t^T(X_t)|^2\,dt.$$

In the proof of Proposition 2.1 we saw that for any continuous local martingale M with $M_0 = 0$ we have

$$\mathbb{P}(M_t \ge x, [M]_t \le y) \le \exp\left(-\frac{x^2}{2y}\right).$$

Now the result follows by applying this inequality to $(M_T^T - M_0^T)$ and noting that $M_T^T = \int_0^T f(u, X_u)\,du$ and $M_0^T = \mathbb{E}^{\mathcal{F}_0}\int_0^T f(u, X_u)\,du = \mathbb{E}\int_0^T f(u, X_u)\,du$, by our assumption that X_0 is concentrated on a single point $x_0 \in \mathbb{R}^n$. □

Corollary 2.7. *If there exist constants $c, \kappa > 0$ such that*

$$|\sigma^\top \nabla P_{t,u} f(x)| \leq c e^{-\kappa(u-t)} \quad x \in \mathbb{R}^n, 0 \leq t \leq u$$

then we have the following Gaussian concentration inequality for all $R > 0, T > 0$:

$$\mathbb{P}\left(\frac{1}{T}\int_0^T f(u, X_u)\, du - \mathbb{E}\left[\frac{1}{T}\int_0^T f(u, X_u)\, du\right] \geq R\right) \leq \exp\left(-\frac{\kappa^2 R^2 T}{2c^2}\right).$$

Proof. From our assumption we can estimate for all $x \in \mathbb{R}^n, 0 \leq t \leq T$

$$|\sigma^\top \nabla R_t^T(x)| \leq \int_t^T |\sigma^\top \nabla P_{t,u} f(x)|\, du \leq c \int_t^T e^{-\kappa(t-u)} du \leq \frac{c}{\kappa}$$

where the regularity assumptions on the coefficients of (2.5) ensure well-posedness of the Kolmogorov backward equation and by extension sufficient smoothness to differentiate under the integral for the first inequality.

Now the result follows directly from Proposition 2.6 since

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{T}\int_0^T f(u, X_u)\, du - \mathbb{E}\left[\frac{1}{T}\int_0^T f(u, X_u)\, du\right] \geq R\right) \\
&= \mathbb{P}\left(\int_0^T f(u, X_u)\, du - \mathbb{E}\left[\int_0^T f(u, X_u)\, du\right] \geq RT \;;\; \int_0^T |\sigma^\top \nabla R_t^T(X_t)|^2\, dt \right. \\
&\left. \leq \frac{c^2}{\kappa^2} T\right). \quad \square
\end{aligned}$$

Remark 2.8. The approach presented here, based on stochastic calculus, requires the existence of an Itô-type formula and the well-posedness of the Kolmogorov backward equation. The approach essentially operates on the level of the individual trajectories of X, and in return yields an explicit martingale representation for $\int_0^T f(u, X_u)\, du$. In contrast, the results on Markov processes from Proposition 3.7 rely on knowing the quadratic variation of a certain auxiliary martingale, for which it is sufficient to have a characterization of X in terms of a martingale problem.

Remark 2.9. Regarding the law of X_0, in contrast to the analytic approaches present in the literature, the case where X_0 is concentrated on a single point is the most natural for our probabilistic approach. In particular, this case is outside of the scope of many existing results in the literature such as [19,22], which require the law of X_0 to be absolutely continuous with respect to a stationary measure of the process. In Section 3.4 we will see how to deal with more general initial laws in the context of the approach presented here.

Remark 2.10. We avoid the question of stationarity altogether by deriving concentration inequalities of $\frac{1}{T}\int_0^T f(u, X_u)\, du$ centered around its expectation, as opposed to the usual formulation which shows concentration around $\int f d\mu$ for a stationary measure μ. In Section 4.2 we will see an example of how to derive concentration inequalities around $\int f d\mu$.

Remark 2.11. The essential ingredient in the proof is equation (2.7), which states that $(t, x) \mapsto \int_t^T P_{t,u} f(x)\, du$ is a solution to the following PDE in g on $[0, T] \times \mathbb{R}^n$:

$$Lg + \partial_t g = -f.$$

This observation is not new and the solution g notably features prominently in the classic book on martingale problems by Stroock and Varadhan [35]. However, the application to additive functionals seems to be new, if not entirely surprising. Indeed, a popular approach to additive functionals of Markov processes [6,21] involves the solution g to the Poisson equation on $\mathbb{R}^n$,

$$Lg = -f.$$

Under certain strong ergodicity conditions, requiring in particular the stationarity of X, a solution to the Poisson equation can be shown to exist and is then given by the well-known resolvent formula as $x \mapsto \int_0^\infty P_{0,u} f(x)\, du$. In contrast, the definition of $R_t^T(x) = \int_t^T P_{t,u} f(x)\, du$ makes sense in a very general setting.

2.3. Discrete time Markov process

In this section, we consider the case of a discrete-time Markov chain in order to build some probabilistic intuition for our assumptions and to highlight some issues that appear in the presence of jumps. For examples of processes satisfying the conditions below, see Section 4.1 or the articles [26,31].

Consider a discrete-time Markov Process $(X_t)_{t\in\mathbb{N}}$ taking values in $\mathbb{R}$ with $X_0 = x_0 \in \mathbb{R}$. Fix a measurable function $f : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}|f(t, X_t)| < \infty$ for all $t \in \mathbb{N}$ and define the associated two-parameter semigroup $P_{t,u} f$ by

$$P_{t,u} f(x) = \mathbb{E}[f(u, X_u)|X_t = x], \quad 0 \le t \le u; u, t \in \mathbb{N}.$$

Let

$$S_t = \sum_{u=1}^{t} f(u, X_u), \quad t \ge 1.$$

Proposition 2.12. *Assume that for each $t \in \mathbb{N}$ there exist positive constants σ_t, κ_t with $\kappa_t < 1$ and a bounded $\mathcal{F}_{t-1}$-measurable random variable C_t such that for $s, t \in \mathbb{N}$*

$$\begin{aligned} |X_t - X_{t-1}| &\le C_t, \quad t \ge 1, \\ |P_{s,t} f(x) - P_{s,t} f(y)| &\le \sigma_s (1-\kappa_s)^{t-s} |x - y|, \quad x, y \in \mathbb{R}, 0 \le s \le t. \end{aligned}$$

Then for all $T, R, a > 0$ we have

$$\mathbb{P}\left(S_T - \mathbb{E}S_T \ge R\ ;\ \sum_{t=1}^{T} \frac{\sigma_t^2 C_t^2}{\kappa_t^2} \le a^2\right) \le \exp\left(-\frac{R^2}{8a^2}\right).$$

Proof. Fix $T > 0$ and define a martingale M^T by

$$M_t^T = \mathbb{E}^{\mathcal{F}_t} S_T = \sum_{u=1}^{T} \mathbb{E}^{\mathcal{F}_t} f(u, X_u)$$

so that

$$M_t^T - M_{t-1}^T = \sum_{u=1}^{T} \left(\mathbb{E}^{\mathcal{F}_t} f(u, X_u) - \mathbb{E}^{\mathcal{F}_{t-1}} f(u, X_u)\right).$$

Note that by our assumptions on f, we have for all $t \in \mathbb{N}$

$$\mathbb{E}|M_t^T| \leq T \max_{u \leq T} \mathbb{E}|f(u, X_u)| < \infty$$

so that M^T is indeed a martingale.

We have by the Markov property, adaptedness of X_t and our assumptions

$$\begin{aligned}
&\mathbb{E}^{\mathcal{F}_t} f(u, X_u) - \mathbb{E}^{\mathcal{F}_{t-1}} f(u, X_u) \\
&\quad = P_{t,u} f(X_t) - \mathbb{E}^{\mathcal{F}_{t-1}} P_{t,u} f(X_t) \\
&\quad = P_{t,u} f(X_t) - P_{t,u} f(X_{t-1}) - \mathbb{E}^{\mathcal{F}_{t-1}} \left[P_{t,u} f(X_t) - P_{t,u} f(X_{t-1}) \right] \\
&\quad \leq \sigma_t (1 - \kappa_t)^{u-t} \left(|X_t - X_{t-1}| + \mathbb{E}^{\mathcal{F}_{t-1}} |X_t - X_{t-1}| \right) \\
&\quad \leq 2\sigma_t (1 - \kappa_t)^{u-t} C_t
\end{aligned}$$

This shows that the increments of the martingale M_t^T are uniformly bounded by a predictable process independent of T:

$$M_t^T - M_{t-1}^T \leq 2\, C_t\, \sigma_t \sum_{u=t}^{T} (1 - \kappa_t)^{u-t} \leq \frac{2\, C_t\, \sigma_t}{\kappa_t}.$$

An extension of the classical Azuma–Hoeffding inequality (see for example Theorem 3.4 in [3]) states that for any square-integrable martingale M with $M_0 = 0$ and such that $M_t - M_{t-1} \leq D_t$ for bounded, $\mathcal{F}_{t-1}$-measurable random variables D_t we have the inequality

$$\mathbb{P}(M_T \geq x\ ;\ \sum_{t=1}^{T} D_t^2 \leq y) \leq \exp\left(-\frac{x^2}{2y} \right).$$

Since we assumed X_0 to be deterministic, we have $M_T^T - M_0^T = S_T - \mathbb{E}^{\mathcal{F}_0} S_T = S_T - \mathbb{E} S_T$. By applying the preceding martingale inequality to M^T we get finally

$$\mathbb{P}\left(S_T - \mathbb{E} S_T \geq R\ ;\ \sum_{t=1}^{T} \frac{\sigma_t^2\, C_t^2}{\kappa_t^2} \leq a^2 \right) \leq \exp\left(-\frac{R^2}{8a^2} \right). \quad \square$$

3. Martingale and concentration inequalities

3.1. Processes bounded in $L^2(\Omega)$

Consider a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \geq 0})$ satisfying the usual conditions from the general theory of semimartingales, meaning that $\mathcal{F}$ is $\mathbb{P}$-complete, $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets in $\mathcal{F}$ and $\mathcal{F}_t$ is right-continuous. In this section $(X_t)_{t \geq 0}$ will denote a real-valued stochastic process adapted to $\mathcal{F}_t$, bounded in $L^2(\Omega)$ in the sense that $\sup_t \mathbb{E} X_t^2 < \infty$.

Define an adapted continuous finite-variation process $(S_t)_{t \geq 0}$ by

$$S_t = \int_0^t X_u \, du.$$

Fix $T > 0$ and define a martingale $(M_t^T)_{t \geq 0}$ by

$$M_t^T = \mathbb{E}^{\mathcal{F}_t} S_T.$$

By the boundedness and adaptedness assumptions on X, M^T is a square integrable martingale (by Doob's maximal inequality) which we can and will choose to be right-continuous with left limits so that the predictable quadratic variation $\langle M^T \rangle$ and the jumps $(\Delta M_t^T)_{t\geq 0}$ are well-defined.

As illustrated in Section 2, our goal is to derive concentration inequalities for S_T from concentration inequalities for the martingale M^T by using the relation $S_T = M_T^T$. In this subsection we take the first step towards that goal by characterizing M^T (and thus S_T) based on properties of the underlying process X, using a family of auxiliary martingales. The next two subsections will give explicit expressions for the auxiliary martingales in a martingale and Markov process context. The final subsection will complete the approach by recalling known concentration inequalities for martingales in terms of their (predictable) quadratic variations and jumps.

Define the family of auxiliary martingales $(Z^u)_{u\geq 0}$ by

$$Z_t^u = \mathbb{E}^{\mathcal{F}_t} X_u,$$

which will be chosen right-continuous with left limits. Each Z^u is square integrable so that the predictable quadratic covariation $\langle Z^u, Z^v \rangle$ is well-defined.

Formally the next result is just a consequence of the (bi)linearity of the integral and (predictable) quadratic variation. The proof shows that the formal calculation is justified under our assumption that X is square integrable. The main interest of the result lies in the fact that we can often find explicit expressions for $\langle Z^u, Z^v \rangle$ and ΔZ^u, as we will see in the next two sections.

Theorem 3.1. *For any $T > 0$ we have*

$$M_t^T = \int_0^T Z_t^u du, \tag{3.1}$$

$$\Delta M_t^T = \int_0^T \Delta Z_t^u du, \tag{3.2}$$

$$[M^T]_t = \int_0^T \int_0^T [Z^u, Z^v]_t \, du \, dv, \tag{3.3}$$

$$\langle M^T \rangle_t = \int_0^T \int_0^T \langle Z^u, Z^v \rangle_t \, du \, dv. \tag{3.4}$$

Proof. Since the theorem only involves values of X_t and Z_t^u for $u, t \leq T$, we can assume without loss of generality that $Z_t^u = Z_{t\wedge T}^{u\wedge T}$ and $X_t = X_{t\wedge T}$. We proceed with some preliminary estimates. First, the family $(Z_\tau^u)_{u,\tau}$, $u \geq 0$, τ stopping time, is uniformly bounded in L^2 since, by optional sampling and Jensen's inequality,

$$\mathbb{E}(Z_\tau^u)^2 = \mathbb{E}(Z_{\tau\wedge T}^{u\wedge T})^2 = \mathbb{E}(\mathbb{E}^{\mathcal{F}_{\tau\wedge T}} \mathbb{E}^{\mathcal{F}_T} X_{u\wedge T})^2 \leq \mathbb{E}X_{u\wedge T}^2 \leq \sup_{0\leq t\leq T} \mathbb{E}X_t^2 < \infty.$$

It follows that $([Z^u, Z^v]_\tau)_{u,v,\tau}$ and $(\langle Z^u, Z^v \rangle_\tau)_{u,v,\tau}$ are uniformly bounded in L^1 since, by elementary properties of the quadratic variation, Cauchy–Schwartz and uniform integrability of the martingales Z^u,

$$\mathbb{E}\left|[Z^u, Z^v]_\tau\right| \leq \mathbb{E}([Z^u]_\tau [Z^v]_\tau)^{1/2} \leq (\mathbb{E}(Z_\tau^u)^2)^{1/2}(\mathbb{E}(Z_\tau^v)^2)^{1/2}.$$

Furthermore the local martingale $Z^u Z^v - [Z^u, Z^v]$ is actually a uniformly integrable martingale since

$$\mathbb{E}\left|Z_\infty^u Z_\infty^v - [Z^u, Z^v]_\infty\right| = \mathbb{E}\left|Z_T^u Z_T^v - [Z^u, Z^v]_T\right| < \infty.$$

The same holds for $Z^u Z^v - \langle Z^u, Z^v \rangle$ by an identical argument.

We proceed to show (3.1), i.e. for fixed $T > 0$ and all $t \geq 0$

$$\mathbb{E}^{\mathcal{F}_t} \int_0^T X_u \, du = \int_0^T Z_t^u \, du. \tag{3.5}$$

Since $(Z_t^u)_{u \geq 0}$ is uniformly bounded in L^1 we get by Fubini's theorem that for arbitrary $G \in \mathcal{F}_t$

$$\mathbb{E}\left[\int_0^T Z_t^u \, du; G\right] = \int_0^T \mathbb{E}[Z_t^u; G] du = \int_0^T \mathbb{E}[X_u; G] du = \mathbb{E}\left[\int_0^T X_u \, du; G\right]$$

which is equivalent to (3.5).

To show (3.2) note that since $(Z_t^u)_{t \geq 0}$ is uniformly integrable we have $Z_{t-1/n}^u \to Z_{t-}^u$ in L^1. Since $(Z_t^u)_{u,t \geq 0}$ is uniformly bounded in L^1 we can use Fubini's theorem and dominated convergence to show that

$$M_{t-1/n}^T = \int_0^T Z_{t-1/n}^u \, du \to \int_0^T Z_{t-}^u \, du \text{ in } L^1.$$

Since the almost sure limit of the left-hand side is M_{t-}^T by definition and the almost sure and L^1 limits coincide when they both exist, this proves (3.2).

To show (3.3), we are going to use the characterization of $[M^T]$ as the unique adapted and *càdlàg* process A with paths of finite variation on compacts such that $(M^T)^2 - A$ is a local martingale and $\Delta A = (\Delta M^T)^2$, $A_0 = (M_0^T)^2$.

Let

$$A_t = \int_0^T \int_0^T [Z^u, Z^v]_t \, du \, dv.$$

Clearly $A_0 = (M_0^T)^2$ since $[Z^u, Z^v]_0 = Z_0^u Z_0^v$. Since sums and limits preserve measurability, A is adapted since all the integrands $[Z^u, Z^v]$ are. By polarization

$$A_t = \frac{1}{4} \int_0^T \int_0^T [Z^u + Z^v]_t \, du \, dv - \frac{1}{4} \int_0^T \int_0^T [Z^u - Z^v]_t \, du \, dv$$

so that the paths of A_t can be written as a difference between two increasing functions and are thus of finite variation on compacts. Furthermore, since $0 \leq [Z^u \pm Z^v]_t \leq [Z^u \pm Z^v]_{u \vee v}$ for all $t \geq 0$ and $[Z^u \pm Z^v]_t$ is increasing in t, pathwise left limits and right continuity follow from the monotone convergence theorem applied to the preceding decomposition.

Since $[Z^u, Z^v]_t$ is bounded in L^1, uniformly in u, v, t, and uniformly integrable in t, it follows as in the proof of (3.2) that

$$A_{t-} = \int_0^T \int_0^T [Z^u, Z^v]_{t-} \, du \, dv$$

so that

$$\Delta A_t = \int_0^T \int_0^T \Delta [Z^u, Z^v]_t \, du \, dv = \int_0^T \Delta Z_t^u \, du \int_0^T \Delta Z_t^v \, dv = \left(\Delta M_t^T\right)^2.$$

To show the martingale part of the characterization, recall that an adapted càdlàg process Y is a uniformly integrable martingale if for all stopping times τ, $\mathbb{E}\,|Y|_\tau < \infty$ and $\mathbb{E} Y_\tau = 0$. Now, for any stopping time τ, by our preliminary estimates above we get the integrability so that we can use Fubini's theorem and optional stopping to obtain

$$\mathbb{E}\left[\left(M_\tau^T\right)^2 - A_\tau\right] = \int_0^T \int_0^T \mathbb{E}\left(Z_\tau^u Z_\tau^v - [Z^u, Z^v]_\tau\right) du\, dv = 0$$

which finishes the proof that $[M^T]_t = A_t$.

It remains to identify $\langle M^T \rangle$ as the compensator of $[M^T]$, i.e. the unique finite variation predictable process $\tilde{A}$ such that $A - \tilde{A}$ is a local martingale. Let

$$\tilde{A}_t = \int_0^T \int_0^T \langle Z^u, Z^v \rangle_t \, du\, dv.$$

Then $\tilde{A}$ is predictable since measurability is preserved by sums and limits, and thus by integrals. Using polarization, we see that $\tilde{A}$ is the difference between two increasing processes and therefore of finite variation. Finally, for all stopping times τ, we again get from our preliminary estimates that $\mathbb{E}|A_\tau - \tilde{A}_\tau| < \infty$ and by optional stopping

$$\mathbb{E}[A_\tau - \tilde{A}_\tau] = \int_0^T \int_0^T \mathbb{E}\left([Z^u, Z^v]_\tau - \langle Z^u, Z^v \rangle_\tau\right) du\, dv = 0$$

which shows that $A - \tilde{A}$ is a martingale and thereby concludes the proof of the theorem. □

Proposition 3.2. *If there exist real-valued processes σ_t, J_t and a constant $\kappa \geq 0$ such that for all $0 \leq t \leq u \leq T$*

$$d\langle Z^u \rangle_t \leq \sigma_t^2 e^{-2\kappa(u-t)} dt,$$

$$|\Delta Z_t^u| \leq |\Delta J_t| e^{-\kappa(u-t)}$$

then

$$\langle M^T \rangle_t \leq \int_0^t \frac{\sigma_s^2}{\kappa^2} \left(1 - e^{-\kappa(T-s)}\right)^2 ds,$$

$$|\Delta M_t^T| \leq \frac{|\Delta J_t|}{\kappa} \left(1 - e^{-\kappa(T-t)}\right)$$

where the case $\kappa = 0$ is to be understood in the sense of the limit as $\kappa \to 0$.

Proof. The second inequality is immediate from Theorem 3.1 and the observation that $\Delta Z_t^u = 0$ for $t > u$. We now proceed to prove the first one. For $t \leq u \wedge v$ we have

$$\langle Z^u \rangle_t d\langle Z^v \rangle_t \leq e^{-2\kappa u} \int_0^t \sigma_s^2 e^{2\kappa s} ds\, e^{-2\kappa v} \sigma_t^2 e^{2\kappa t} = \frac{1}{2} d\left(\int_0^t \sigma_s^2 e^{-\kappa(u-s)} e^{-\kappa(v-s)} ds\right)^2$$

which is symmetric in u and v. Using Cauchy–Schwarz for the predictable quadratic variation, the fact that Z_t^u is constant for $t \geq u$ and integration by parts together with the previous inequality we get

$$\langle Z^u, Z^v \rangle_t \leq (\langle Z^u \rangle \langle Z^v \rangle)_{t \wedge u \wedge v}^{1/2} = \left(\int_0^{t \wedge u \wedge v} \langle Z^u \rangle_s d\langle Z^v \rangle_s + \int_0^{t \wedge u \wedge v} \langle Z^v \rangle_s d\langle Z^u \rangle_s\right)^{1/2}$$

$$\leq \int_0^t \mathbb{1}_{\{s \leq u \wedge v\}} \sigma_s^2 e^{-\kappa(u-s)} e^{-\kappa(v-s)} ds.$$

Therefore by Fubini, for any $0 \leq t \leq T$

$$\begin{aligned}
\langle M^T \rangle_t &= \int_0^T \int_0^T \langle Z^u, Z^v \rangle_t du dv \\
&\leq \int_0^T \int_0^T \int_0^t \mathbb{1}_{\{s \leq u \wedge v\}} \sigma_s^2 e^{-\kappa(u-s)} e^{-\kappa(v-s)} ds\, du\, dv \\
&= \int_0^t \sigma_s^2 \left(\int_s^T e^{-\kappa(u-s)} du \right)^2 ds.
\end{aligned}$$

which is the result. □

3.2. Martingales

Proposition 3.3. *If X is a square integrable real-valued martingale then*

$$\begin{aligned}
dM_t^T &= (T-t)\, dX_t \\
\Delta M_t^T &= (T-t)\, \Delta X_t \\
d[M^T]_t &= (T-t)^2\, d[X]_t \\
d\langle M^T \rangle_t &= (T-t)^2\, d\langle X \rangle_t
\end{aligned}$$

Proof. Since

$$Z_t^u = \mathbb{E}^{\mathcal{F}_t} X_u = X_{t \wedge u}$$

we have by (3.1)

$$M_t^T = \int_0^T X_{t \wedge u}\, du = \int_0^t X_u\, du + (T-t) X_t.$$

Using integration by parts

$$dM_t^T = X_t\, dt + (-X_t\, dt + (T-t)\, dX_t) = (T-t) dX_t$$

and the remaining equalities now follow directly from stochastic calculus. □

Remark 3.4. From integration by parts (and similarly for $\langle M^T \rangle$)

$$[M^T]_T = 2 \int_0^T (T-t)[X]_t\, dt.$$

From the fact that the Doléans–Dade exponential is a positive local martingale and therefore a supermartingale we immediately get the following corollary.

Corollary 3.5. *If the martingale X is continuous and $X_0 = x \in \mathbb{R}$ then for all $\lambda \in \mathbb{C}$ and $T > 0$*

$$\mathbb{E} \exp \left(\lambda (S_T - \mathbb{E} S_T) - \lambda^2 \int_0^T (T-u)^2\, d\langle X \rangle_u \right) \leq 1.$$

Remark 3.6 (*Central Limit Theorem*)**.** From the Doléans–Dade exponential it is also possible to derive a central limit theorem for a suitably normalized family of random variables $G_T = \frac{1}{\sqrt{\mathbb{E}[M^T]_T}} \int_0^T X_t\, dt$. A thorough investigation is beyond the scope of this work, but an outline of the argument goes as follows. Suppose that X is continuous. Denote $\mathcal{E}$ the Doléans–Dade exponential and for two families of random variables X_T, Y_T write $X_T \overset{P}{\sim} Y_T$ if $X_T/Y_T \to 1$ in probability as $T \to \infty$. Now suppose that $[M^T]_T \overset{P}{\sim} \mathbb{E}[M^T]_T$, which can for example follow from an ergodic theorem since $[M^T]_T$ is usually an additive functional itself. Then for $\lambda \in \mathbb{R}$ we have for the characteristic function of G_T that

$$\begin{aligned} e^{i\lambda G_T} &= \exp\left(i\lambda \frac{M_T^T}{\sqrt{\mathbb{E}[M^T]_T}} + \frac{\lambda^2}{2\mathbb{E}[M^T]_T}[M^T]_T - \frac{\lambda^2}{2\mathbb{E}[M^T]_T}[M^T]_T \right) \\ &= \mathcal{E}\left(\frac{i\lambda M_T^T}{\sqrt{\mathbb{E}[M^T]_T}} \right) \exp\left(-\frac{\lambda^2[M^T]_T}{2\mathbb{E}[M^T]_T} \right) \overset{P}{\sim} \mathcal{E}\left(\frac{i\lambda M_T^T}{\sqrt{\mathbb{E}[M^T]_T}} \right) e^{-\frac{\lambda^2}{2}}. \end{aligned}$$

Now if the Doléans–Dade exponential (which can be negative since we have complex exponent) is a true martingale and the family $\mathcal{E}\left(\frac{i\lambda M_T^T}{\sqrt{\mathbb{E}[M^T]_T}} \right)$ is uniformly integrable then for all $\lambda \in \mathbb{R}$

$$\lim_{T\to\infty} \mathbb{E}e^{i\lambda G_T} = \lim_{T\to\infty} \mathbb{E}\,\mathcal{E}\left(\frac{i\lambda M_T^T}{\sqrt{\mathbb{E}[M^T]_T}} \right) e^{-\frac{\lambda^2}{2}} = e^{-\frac{\lambda^2}{2}}$$

so that G_T converges in distribution to a standard normal random variable.

3.3. Markov processes

Consider a continuous-time Markov process $(Y_t)_{t\geq 0}$ with natural filtration $(\mathcal{F}_t)_{t\geq 0}$, taking values in a Polish space E and with trajectories that are right-continuous with left limits. Denote $\mathcal{B}$ the set of Borel functions on $\mathbb{R}^+ \times E$.

Fix a function $f \in \mathcal{B}$ such that $\sup_t \mathbb{E} f(t, Y_t)^2 < \infty$ so that we are in the setting of Section 3.1 with $X_t = f(t, Y_t)$, $S_T = \int_0^T f(u, X_u)du$ and $M_t^T = \mathbb{E}^{\mathcal{F}_t} S_T$. Define the two-parameter semigroup $(P_{t,u} f)_{0\leq t\leq u}$ on E by

$$P_{t,u} f(y) = \mathbb{E}[f(u, Y_u)|Y_t = y].$$

Suppose that there is a set $\mathcal{D}(\Gamma) \subset \mathcal{B} \times \mathcal{B}$ and a map $\Gamma : \mathcal{D}(\Gamma) \to \mathcal{B}$ such that for each $(f, g) \in \mathcal{D}(\Gamma)$ we have, setting $F_t = f(t, Y_t)$, $G_t = g(t, Y_t)$, that

$$\langle F, G\rangle_t = 2\int_0^t \Gamma(f, g)(s, Y_s)\,ds, \quad 0 \leq t \leq T. \tag{3.6}$$

Usually, Γ corresponds to (an extension of) the squared field operator $\Gamma(f, g) = \frac{1}{2}(Lfg - fLg - gLf)$, see Remark 3.8.

Proposition 3.7. *If $(P_{\cdot,u} f, P_{\cdot,v} f)_{0\leq u,v\leq T} \in \mathcal{D}(\Gamma)$ we have for $0 \leq t \leq T$*

$$d\langle M^T\rangle_t = 2\int_t^T \int_t^T \Gamma(P_{t,u} f, P_{t,v} f)(t, Y_t)\,du\,dv\,dt, \tag{3.7}$$

$$\Delta M_t^T = \int_t^T \left(P_{t,u} f(Y_t) - P_{t,u} f(Y_{t-}) \right) du. \tag{3.8}$$

Proof. For all $0 \leq t \leq u \leq T$ we have

$$Z_t^u = \mathbb{E}^{\mathcal{F}_t} f(u, Y_u) = P_{t,u} f(Y_t).$$

Since f can depend on u we can always consider $f(u, y) - P_{0,u} f(y)$ instead of f and therefore without loss of generality suppose that $Z_0^u = P_{0,u} f(Y_0) = 0$. By (3.6) for $0 \leq t \leq u \wedge v \leq T$

$$d\langle Z^u, Z^v\rangle_t = 2\Gamma(P_{t,u} f, P_{t,v} f)(t, Y_t)dt, \quad 0 \leq t \leq u \wedge v \leq T.$$

For $t \geq v \wedge u$ either Z_t^u or Z_t^v is constant so that

$$d\langle Z^u, Z^v\rangle_t = 0, \quad u \wedge v \leq t \leq T,$$

and thus

$$\langle Z^u, Z^v\rangle_t = \int_0^t \mathbb{1}_{\{s \leq u \wedge v\}} 2\Gamma(P_{s,u} f, P_{s,v} f)(s, Y_s)ds.$$

By (3.4) and Fubini's theorem

$$\begin{aligned}\langle M^T\rangle_t &= \int_0^T \int_0^T \int_0^t \mathbb{1}_{\{s \leq u \wedge v\}} 2\Gamma(P_{s,u} f, P_{s,v} f)(s, Y_s)\, ds\, du\, dv \\ &= \int_0^t \int_s^T \int_s^T 2\Gamma(P_{s,u} f, P_{s,v} f)(s, Y_s)\, du\, dv\, ds.\end{aligned}$$

This proves the first equality (3.7). Equality (3.8) follows directly from (3.2) together with the observation that Z_t^u is constant for $t \geq u$ and the fact that $P_{t,u} f$ is continuous in t. □

Remark 3.8. When Y is a Markov process with infinitesimal generator $(L, \mathcal{D}(L) \subset \mathcal{B})$ then Γ in (3.7) corresponds to the usual squared field operator whenever the latter is well-defined,

$$\Gamma(f, g) = \frac{1}{2}(Lfg - fLg - gLf), \quad f \in \mathcal{D}(L), g \in \mathcal{D}(L), fg \in \mathcal{D}(L).$$

Indeed, suppose that for $f \in \mathcal{D}(L)$

$$f(t, Y_t) - f(0, Y_0) - \int_0^t (\partial_s f + Lf(s, \cdot))(s, Y_s)\, ds \tag{3.9}$$

is a local martingale. As before we can assume $P_{0,u} f(Y_0) = 0$. Now if $P_{t,u} f$, $P_{t,v} f$ and their product $P_{t,u} f P_{t,v} f$ is in $\mathcal{D}(L)$ then

$$P_{t,u} f(Y_t) P_{t,v} f(Y_t) - \int_0^t (\partial_s(P_{s,u} f P_{s,v} f) + L(P_{s,u} f P_{s,v} f))(s, Y_s)\, ds \tag{3.10}$$

is a local martingale. Since we assumed $P_{t,u} f$ to be in $\mathcal{D}(L)$ it solves the Kolmogorov backward equation

$$\partial_t P_{t,u} f(y) = -LP_{t,u}(t, y), \quad y \in E, 0 \leq t \leq u$$

and the same holds true for $P_{t,v} f$. Thus

$$\partial_t(P_{t,u} f P_{t,v} f) = P_{t,u} f \partial_t P_{t,v} f + P_{t,v} f \partial_t P_{t,u} f = -P_{t,u} f LP_{t,v} f - P_{t,v} f LP_{t,u} f.$$

Substituting this into the integral in (3.10) shows that indeed

$$P_{t,u} f(Y_t) P_{t,v} f(Y_t) - 2\int_0^t \frac{1}{2}(L(P_{s,u} f P_{s,v} f) - P_{s,u} f LP_{s,v} f - P_{s,v} f LP_{s,u} f)(s, Y_s)\, ds$$

is a local martingale.

Corollary 3.9. *If Y has continuous trajectories and Y_0 is constant then we have the following inequality for all $\lambda \in \mathbb{C}$ and $T > 0$ fixed:*

$$\mathbb{E}\exp\left[\lambda\,(S_T f - \mathbb{E}S_T f) - \lambda^2 \int_0^T \int_t^T \int_t^T \Gamma(P_{t,u}f, P_{t,v}f)(t, Y_t)\,du\,dv\,dt\right] \le 1.$$

Proof. This follows directly from the Chernoff bound and the fact that the Doléans–Dade exponential is a supermartingale. □

Remark 3.10 (*Central Limit Theorem*)**.** The considerations in Remark 3.6 for deriving a central limit theorem for additive functionals of martingales apply to continuous Markov processes as well.

3.4. Martingale inequalities

Let

$$S_T = \int_0^T X_u\,du$$

for some square integrable càdlàg process X and $M_t^T = \mathbb{E}^{\mathcal{F}_t} S_T$ as in the previous sections. Our key observation is that $S_T - \mathbb{E}^{\mathcal{F}_0} S_T = M_T^T - M_0^T$. Concentration inequalities for S_T then follow from concentration inequalities for martingales. The goal of this section is to show how to pass from $\mathbb{E}^{\mathcal{F}_0} S_T$ to $\mathbb{E}S_T$ and to recall some concentration inequalities for martingales.

For a real-valued random variable Y, denote $\Psi_Y(\lambda)$ the logarithm of the moment-generating function of Y and $\Psi_Y^*(x)$ its associated Cramér transform:

$$\Psi_Y(\lambda) = \log \mathbb{E}e^{\lambda Y}, \quad \lambda \in \mathbb{R},$$
$$\Psi_Y^*(x) = \sup_{\lambda\in\mathbb{R}}(\lambda x - \Psi_Y(\lambda)), \quad x \in \mathbb{R}.$$

Denote $\Lambda(\lambda)$ the logarithm of the moment-generating function of the centered random variable $\mathbb{E}^{\mathcal{F}_0} S_T - \mathbb{E}S_T$ and I its domain:

$$\Lambda(\lambda) = \Psi_{\mathbb{E}^{\mathcal{F}_0} S_T - \mathbb{E}S_T} = \log \mathbb{E}\left[\exp \lambda \left(\mathbb{E}^{\mathcal{F}_0} S_T - \mathbb{E}S_T\right)\right],$$
$$I = \{\lambda \in \mathbb{R} : \Lambda(\lambda) < \infty\}.$$

In particular if $X_0 = x \in \mathbb{R}$ is a deterministic constant then $\mathbb{E}^{\mathcal{F}_0} S_T = \mathbb{E}S_T$ and $\Lambda(\lambda) = 0$.

Following [14], define

$$\varphi(x) = e^x - 1 - x,$$
$$\varphi_a(x) = \varphi(ax)/a^2, \quad a \ge 0,$$
$$H_t^a = \sum_{s\le t}(\Delta M_s^T)^2 \mathbb{1}_{\{|\Delta M_s^T| > a\}} + \langle M^T\rangle_t, \quad a, t \ge 0$$

where for $a = 0$ we set $\varphi_0(x) = x^2/2$ and we have $H_t^0 = \sum_{s\le t}\left(\Delta M_s^T\right)^2 + \langle M^T\rangle_t$.

The next lemma allows us to extend our framework from processes with initial measures concentrated on a single point to more general classes of initial measures when we can control Λ.

Lemma 3.11. *For $a \geq 0, \lambda \in I$*

$$\mathbb{E}\exp\Big(\lambda(S_T - \mathbb{E}S_T) - \varphi_a(|\lambda|)H_T^a - \Lambda(\lambda)\Big) \leq 1.$$

Proof. In [14] Corollary 3.1 it is shown that for any square integrable martingale M and for all $a \geq 0, \lambda \geq 0$ the process

$$\exp\left(\lambda M_t - \varphi_a(|\lambda|)H_t^a\right)$$

is a supermartingale. Applying this to M^T and $-M^T$ together with the supermartingale property yields that for $a \geq 0, \lambda \in \mathbb{R}$

$$\mathbb{E}^{\mathcal{F}_0}\exp\left(\lambda(M_t^T - M_0^T) - \varphi_a(|\lambda|)H_t^a\right) \leq 1.$$

By definition we have furthermore that for $\lambda \in I$

$$\mathbb{E}\exp(\lambda(M_0^T - \mathbb{E}M_0^T)) = \exp \Lambda(\lambda).$$

Therefore for $\lambda \in I$ and all $t \in [0, T]$

$$\begin{aligned} &\mathbb{E}\exp\left(\lambda(M_t^T - \mathbb{E}M_t^T) - \varphi_a(|\lambda|)H_t^a - \Lambda(\lambda)\right) \\ &\quad = \mathbb{E}\left\{\mathbb{E}^{\mathcal{F}_0}\left[\exp\Big(\lambda(M_t^T - M_0^T) - \varphi_a(|\lambda|)H_t^a\Big)\right]\exp\Big(\lambda(M_0^T - \mathbb{E}M_0^T) - \Lambda(\lambda)\Big)\right\} \\ &\quad \leq \mathbb{E}\exp\left(\lambda(M_0^T - \mathbb{E}M_0^T) - \Lambda(\lambda)\right) = 1. \end{aligned}$$

We conclude by taking the inequality at $t = T$ and noting that $M_T^T = S_T, \mathbb{E}M_T^T = \mathbb{E}S_T$. □

In the previous sections, we saw how to estimate the quantities ΔM^T and $\langle M^T\rangle$, and thus H^a, for different classes of processes X. We will now recall some martingale inequalities involving H^a, which then lead directly to inequalities for $S_T - \mathbb{E}S_T$.

From Markov's inequality applied to $e^{\lambda Y}$ we immediately get Chernoff's inequality

$$\mathbb{P}\{Y \geq x\} \leq \exp(-\Psi_Y^*(x)).$$

By combining this with Lemma 3.11 and bounds on Λ we can immediately deduce the following Hoeffding, Bennett and Bernstein-type inequalities. The approach is classical and we follow [4].

Corollary 3.12. *If $\Lambda(\lambda) \leq \frac{\lambda^2}{2}\rho^2$ for some $\rho \geq 0$ then*

$$\mathbb{P}\left(S_T - \mathbb{E}S_T \geq R\,;\, H_T^0 \leq \sigma^2\right) \leq \exp\left(-\frac{R^2}{2(\rho^2+\sigma^2)}\right).$$

Proof. On the set $\{H_T^0 \leq \sigma^2\}$, using $\varphi_0(\lambda) = \frac{\lambda^2}{2}$, $\Psi_{S_T-\mathbb{E}S_T}$ is upper bounded by the logarithmic MGF of a centered Gaussian random variable with variance $\rho^2+\sigma^2$: $\Psi_{S_T-\mathbb{E}S_T}(\lambda) \leq \frac{(\rho^2+\sigma^2)\lambda^2}{2}$. This implies that $\Psi^*_{S_T-\mathbb{E}S_T}$ is lower bounded by the corresponding Cramér transform, $\Psi^*_{S_T-\mathbb{E}S_T}(x) \geq \frac{x^2}{2(\rho^2+\sigma^2)}$, and the result follows immediately from Chernoff's inequality. □

Corollary 3.13. *If $\Lambda(\lambda) \leq \nu\varphi_a(\lambda)$ for some $a, \nu \geq 0$ then*

$$\mathbb{P}\left(S_T - \mathbb{E}S_T \geq R\,;\, H_T^a \leq \mu\right) \leq \exp\left(-\frac{\mu+\nu}{a^2}h\left(\frac{aR}{\mu+\nu}\right)\right)$$

with

$$h(x) = (1+x)\log(1+x) - x, \quad x \geq -1.$$

Proof. On the set $\{H_T^a \leq \mu\}$, $\Psi_{S_T - \mathbb{E}S_T}(\cdot/a)$ is upper bounded by the logarithmic MGF of a centered Poisson random variable with parameter $\frac{\mu+\nu}{a^2}$: $\Psi_{S_T-\mathbb{E}S_T}(\lambda/a) \leq \frac{(\mu+\nu)\varphi(\lambda)}{a^2}$. This implies that

$$\Psi^*_{S_T-\mathbb{E}S_T}(ax) = \sup_{\lambda\geq 0}(\lambda a x - \Psi_{S_T-\mathbb{E}S_T}(\lambda)) = \sup_{\lambda \geq 0}(\lambda x - \Psi_{S_T-\mathbb{E}S_T}(\lambda/a))$$

is lower bounded by the corresponding Cramér transform, $\Psi^*_{S_T-\mathbb{E}S_T}(ax) \geq \frac{\mu+\nu}{a^2} h\left(\frac{a^2 x}{\mu+\nu}\right)$, and the result follows from Chernoff's inequality after rescaling by a. □

Corollary 3.14. *If $\Lambda(\lambda) \leq \frac{\lambda^2 \nu}{2(1-b\lambda)}$ for some $b, \nu \geq 0$ and all $\lambda < 1/b$ then*

$$\mathbb{P}\left(S_T - \mathbb{E}S_T \geq R\,;\, H_T^0 \leq \mu\right) \leq \exp\left(-\frac{\mu+\nu}{b^2} h_1\left(\frac{bR}{\mu+\nu}\right)\right)$$

with

$$h_1(x) = 1 + x - \sqrt{1+2x}.$$

Proof. On the set $\{H_T^0 \leq \mu\}$ using $\varphi_0(\lambda) = \frac{\lambda^2}{2} \leq \frac{\lambda^2}{2(1-\lambda b)}$, $\Psi_{S_T-\mathbb{E}S_T}$ is upper bounded by the (rescaled) logarithmic MGF of a sub-Gamma random variable (using the terminology of [4]) with parameter $(\mu+\nu, b)$: $\Psi_{S_T-\mathbb{E}S_T}(\lambda) \leq \frac{(\mu+\nu)\lambda^2}{2(1-b\lambda)}$. This implies that $\Psi^*_{S_T-\mathbb{E}S_T}$ is lower bounded by the corresponding Cramér transform, $\Psi^*_{S_T-\mathbb{E}S_T}(x) \geq \frac{\mu+\nu}{b^2} h_1\left(\frac{bx}{\mu+\nu}\right)$, and the result follows as before from Chernoff's inequality. □

Going beyond the Chernoff inequality, we have for example the following result which follows directly from Lemma 3.11 and an inequality on self-normalized processes in [32] Theorem 2.1.

Corollary 3.15. *If $\Lambda(\lambda) \leq \frac{\lambda^2}{2}\rho^2$ for some $\rho \geq 0$ and all $\lambda \in \mathbb{R}$ then*

$$\mathbb{P}\left(\frac{|S_T - \mathbb{E}S_T|}{\sqrt{\frac{3}{2}(H_T^0 + \mathbb{E}H_T^0 + 2\rho^2)}} \geq R\right) \leq \min\{2^{1/3}, (2/3)^{2/3} R^{-2/3}\} \exp\left(-\frac{R^2}{2}\right).$$

Proof. By Theorem 2.1 in [32], for a pair of random variables (A, B) with $B > 0$ satisfying

$$\mathbb{E}\left[\exp\left(\lambda A - \frac{\lambda^2}{2} B^2\right)\right] \leq 1, \quad \lambda \in \mathbb{R}$$

and $\mathbb{E}B^2 = \mathbb{E}A^2 < \infty$ we have

$$\mathbb{P}\left(\frac{|A|}{\sqrt{\frac{3}{2}(B^2 + \mathbb{E}[A^2])}} \geq R\right) \leq \min\{2^{1/3}, (2/3)^{2/3} R^{-2/3}\} e^{-R^2/2}. \tag{3.11}$$

The corollary is now a direct consequence of Lemma 3.11 (with $a = 0$). □

Corollary 3.16. *If $\Lambda(\lambda) \leq \frac{\lambda^2}{2}\rho^2$ for some $\rho \geq 0$ and all $\lambda \in \mathbb{R}$ then*

$$\mathbb{P}\left(|S_T - \mathbb{E}S_T| \geq R\,;\, H_T^0 + \mathbb{E}H_T^0 + 2\rho^2 \leq C|S_T - \mathbb{E}S_T| + D\right)$$
$$\leq 2^{1/3}\left(1 \wedge \frac{CR + D}{3R^2}\right)^{1/3} \exp\left(-\frac{R^2}{3(CR+D)}\right), \quad C, D \geq 0.$$

Proof. On the set $\{|S_T - \mathbb{E}S_T| \geq R\}$ we have by monotonicity of $x \mapsto \frac{x}{1+x}$ that

$$\frac{|S_T - \mathbb{E}S_T|}{\sqrt{\frac{3}{2}(C|S_T - \mathbb{E}S_T| + D)}} \geq \frac{R}{\sqrt{\frac{3}{2}(CR + D)}}$$

and on $\{H_T^0 + \mathbb{E}H_T^0 + 2\rho^2 \leq C|S_T - \mathbb{E}S_T| + D\}$ we have

$$\frac{|S_T - \mathbb{E}S_T|}{\sqrt{\frac{3}{2}(H_T^0 + \mathbb{E}H_T^0 + 2\rho^2)}} \geq \frac{|S_T - \mathbb{E}S_T|}{\sqrt{\frac{3}{2}(C|S_T - \mathbb{E}S_T| + D)}}.$$

Together with the previous corollary we get the result

$$\mathbb{P}\left(|S_T - \mathbb{E}S_T| \geq R\,;\, H_T^0 + \mathbb{E}H_T^0 + 2\rho^2 \leq C|S_T - \mathbb{E}S_T| + D\right)$$
$$\leq \mathbb{P}\left(\frac{|S_T - \mathbb{E}S_T|}{\sqrt{\frac{3}{2}(H_T^0 + \mathbb{E}H_T^0 + 2\rho^2)}} \geq \frac{R}{\sqrt{\frac{3}{2}(CR + D)}}\right)$$
$$\leq 2^{1/3}\left(1 \wedge \frac{CR + D}{3R^2}\right)^{1/3} \exp\left(-\frac{R^2}{3(CR+D)}\right). \quad \square$$

Corollary 3.17. *If $\Lambda(\lambda) \leq \frac{\lambda^2}{2}\rho^2$ for some $\rho \geq 0$ and all $\lambda \in \mathbb{R}$ then*

$$\mathbb{P}\left(|S_T - \mathbb{E}S_T| \geq R\,;\, H_T^0 \leq C|S_T| + D\right)$$
$$\leq 2^{1/3}\left(1 \wedge \frac{CR + D'}{3R^2}\right)^{1/3} \exp\left(-\frac{R^2}{3(CR+D')}\right), \quad C, D \geq 0$$

with $D' = D + C|\mathbb{E}S_T| + \mathbb{E}H_T^0 + 2\rho^2$

Proof. On the set $\{H_T^0 \leq C|S_T| + D\}$ we have

$$H_T^0 \leq C(|S_T| - |\mathbb{E}S_T|) + C|\mathbb{E}S_T| + D \leq C|S_T - \mathbb{E}S_T| + C|\mathbb{E}S_T| + D$$

so that

$$H_T^0 + \mathbb{E}H_T^0 + 2\rho^2 \leq C|S_T - \mathbb{E}S_T| + D + C|\mathbb{E}S_T| + \mathbb{E}H_T^0 + 2\rho^2$$

and the result follows directly from the previous Corollary. $\square$

4. Applications

4.1. Polyak–Ruppert averages

In this section, we use the notation $\Delta X_t = X_t - X_{t-1}$ for a discrete-time process X. The symbols t, s, u, T will always denote time variables taking values in $\mathbb{Z}^+$.

Consider the real-valued process X defined by the recursion

$$X_t = X_{t-1} - \alpha_t\, g(X_{t-1}, W_t), \quad X_0 = x \tag{4.1}$$

with $x \in \mathbb{R}$, $(\alpha_t)_{t\in\mathbb{N}}$ a sequence in $\mathbb{R}$, W_t a sequence of independent identically distributed random variables with common law μ such that μ has compact support, and $g : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ is a function such that

$$0 < m(w) \leq \partial_x g(x, w) \leq M(w) < \infty, \quad x, w \in \mathbb{R} \tag{4.2}$$

for some functions $m : \mathbb{R} \to \mathbb{R}$ and $M : \mathbb{R} \to \mathbb{R}$.

The recursion (4.1) is an instance of the Robbins–Monroe algorithm for finding a root of the function $\bar{g}(x) = \int g(x, w)\mu(dw)$. In our case, the assumption (4.2) implies that $\bar{g}$ is the derivative of some strongly convex function and that $\bar{g}$ is Lipschitz continuous with Lipschitz constant $\bar{M} = \int M(w)\mu(dw)$. We also denote $\bar{m} = \int m(w)\mu(dw)$.

Under certain assumptions on g and the sequence of step sizes α_t, it can be shown that X_t converges almost surely to a limit x^* such that $\bar{g}(x^*) = 0$ [24,34]. It was later shown that the convergence rate of the algorithm could be improved by considering the Polyak–Ruppert averages $\frac{1}{T}\sum_{t=0}^{T-1} X_t$ [33].

Using the approach developed in Section 2.3 we now show how to obtain concentration inequalities for the Polyak–Ruppert averages around their expected value in the sense that if $\alpha_t = \lambda t^{-p}$ for λ sufficiently small and $p < 1/2$ then (see Corollary 4.4 for the precise statement)

$$\begin{aligned}&\mathbb{P}\left(\frac{1}{T}\sum_{t=0}^{T-1} X_t - \mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1} X_t\right] \geq R\,;\ \sup_{t\leq T}\sup_{w} |g(X_t, w)| \leq G\right)\\ &\leq \exp\left(-\frac{(1-2p)\,\bar{m}^2\,R^2\,T}{32\,G^2}\right).\end{aligned}$$

In particular, the order in T, the dependence of the numerator on $\partial_x g$ and of the denominator on g match the central limit theorem in [18]. By [24] we also have under some additional conditions that for $t \geq 1$, $\mathbb{E}\,|X_t - x^*| \leq C t^{-p/2}$ for some constant C, so that

$$\begin{aligned}\mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1} X_t\right] - x^* &\leq \frac{X_0 - x^*}{T} + \frac{C}{T}\sum_{t=1}^{T-1} t^{-p/2} \leq \frac{X_0 - x^*}{T} + \frac{C}{T}\frac{(T-1)^{1-p/2}}{1-p/2}\\ &\leq \frac{X_0 - x^*}{T} + \frac{2C}{T^{p/2}}\end{aligned}$$

and $\mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1} X_t\right] \to x^*$ as $T \to \infty$.

Lemma 4.1. *Let $g^*(x) = \sup_{w\in\operatorname{supp}(\mu)} g(x, w)$ and set $C_t = |\alpha_t\, g^*(X_{t-1})|$. The process C_t defines a sequence of $\mathcal{F}_{t-1}$-measurable bounded random variables such that $|\Delta X_t| \leq C_t$.*

Proof. The $\mathcal{F}_{t-1}$-measurability is clear, as is the inequality $|\Delta X_t| = |\alpha_t\, g(X_{t-1}, W_t)| \leq |\alpha_t\, g^*(X_{t-1})|$. Since g is continuous, in order to show that $g^*(X_{t-1})$ is bounded it is sufficient to show that X_{t-1} is bounded. This follows from a simple induction. Indeed, for an arbitrary $s > 0$, suppose that $|X_s| \leq R_s < \infty$ and let $R_{s+1} = R_s + \alpha_{s+1}\sup_{|x|\leq R_s} g^*(x)$. Then $|X_{s+1}| \leq R_s + \alpha_{s+1}|g(X_s, W_{s+1})| \leq R_s + \alpha_{s+1}\sup_{|x|\leq R_s} g^*(x) = R_{s+1} < \infty$. Since $R_0 = |x| < \infty$ the conclusion follows by induction. $\square$

Lemma 4.2. *For any Lipschitz function f with Lipschitz constant $\|f\|_{\mathrm{Lip}}$ and $x, y \in \mathbb{R}$, $0 \le s \le t$ we have*

$$\left|P_{s,t} f(x) - P_{s,t} f(y)\right| \le \|f\|_{\mathrm{Lip}}\ |x-y| \prod_{u=s+1}^{t} \sqrt{1 - 2\alpha_u \bar{m} + \alpha_u^2 \bar{M}}$$

Proof. Let X_t^x, X_t^y be the values at time t of the recursion (4.1) started from $X_s = x$ respectively $X_s = y$. Then by definition $P_{s,t} f(x) = \mathbb{E} f(X_t^x)$ so that

$$\left|P_{s,t} f(x) - P_{s,t} f(y)\right| = \left|\mathbb{E}\left(f(X_t^x) - f(X_t^y)\right)\right| \le \|f\|_{\mathrm{Lip}} \sqrt{\mathbb{E}\left(X_t^x - X_t^y\right)^2}.$$

Now from summation by parts and the bounds on $\partial_x g$ that we assumed we get

$$\begin{aligned} \Delta(X^x - X^y)_t^2 &= 2(X_{t-1}^x - X_{t-1}^y)\Delta(X^x - X^y)_t + \left[\Delta(X^x - X^y)_t\right]^2 \\ &\le -2\,\alpha_t\, m(W_t)(X_{t-1}^x - X_{t-1}^y)^2 + \alpha_t^2\, M(W_t)^2 (X_{t-1}^x - X_{t-1}^y)^2 \\ &= -(2\,\alpha_t\, m(W_t) - \alpha_t^2\, M(W_t)^2)(X_{t-1}^x - X_{t-1}^y)^2 \end{aligned}$$

so that by developing the recursion we obtain

$$(X^x - X^y)_t^2 \le (x-y)^2 \prod_{u=s+1}^{t} \left(1 - (2\,\alpha_u\, m(W_u) - \alpha_u^2\, M(W_u)^2)\right)$$

and the result follows by taking expectation, using the fact that the W_t are i.i.d. Note that the square root is well-defined since $1 - (2\,\alpha_u\, \bar{m} - \alpha_u^2\, \bar{M}^2) \ge 1 - 2\,\alpha_u\, \bar{M} + \alpha_u^2 \bar{M}^2 = (1 - \alpha_u \bar{M})^2 \ge 0$. $\square$

Corollary 4.3. *If $\alpha_1 + \alpha_T \le \frac{2\bar{m}}{\bar{M}^2}$ and $\alpha_{t+1} \le \alpha_t$ for all $t \ge 1$ then for any 1-Lipschitz function f*

$$\left|P_{s,t} f(x) - P_{s,t} f(y)\right| \le |x-y|\ \left(1 - \alpha_T\, \bar{m} + \tfrac{1}{2}\alpha_T^2 \bar{M}^2\right)^{t-s}$$

Proof. Let $\beta_t = \sqrt{1 - 2\alpha_t \bar{m} + \alpha_t^2 \bar{M}^2}$. Since $\beta_u^2 - \beta_T^2 = (\alpha_u - \alpha_T)((\alpha_u + \alpha_T)\bar{M}^2 - 2\bar{m})$, the assumptions on α imply that $\beta_u \le \beta_T$ for all $1 \le u \le T$ so that

$$\prod_{u=s+1}^{t} \beta_u \le \beta_T^{t-s}.$$

Since $\sqrt{1-x} \le 1 - x/2$ we also have

$$\beta_T \le 1 - \alpha_T\, \bar{m} + \tfrac{1}{2}\alpha_T^2 \bar{M}^2.$$

Together with the preceding Lemma we finally obtain

$$\begin{aligned} \left|P_{s,t} f(x) - P_{s,t} f(y)\right| &\le \|f\|_{\mathrm{Lip}}\ |x-y| \prod_{u=s+1}^{t} \beta_u \\ &\le |x-y|\ \left(1 - \alpha_T\, \bar{m} + \tfrac{1}{2}\alpha_T^2 \bar{M}^2\right)^{t-s}. \quad \square \end{aligned}$$

Corollary 4.4. *For any $T \in \mathbb{N}$ fixed, if $\alpha_t = \lambda t^{-p}$ for $p < 1/2$ and λ such that*

$$\lambda \le \frac{2\bar{m}}{\bar{M}^2} \frac{T^p}{1 + T^p}$$

we have

$$\mathbb{P}\left(\frac{1}{T}\sum_{t=0}^{T-1} X_t - \mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1} X_t\right] \geq R \; ; \; \sup_{t\leq T}\sup_{w}|g(X_t,w)| \leq G\right)$$
$$\leq \exp\left(-\frac{(1-2p)\,\bar{m}^2\,R^2\,T}{32\,G^2}\right).$$

Proof. We are in the situation of Proposition 2.12 with $\sigma_t^2 = 1$, $C_t = \lambda t^{-p} g^*(X_t)$ from Lemma 4.1 and $\kappa_t = \kappa_T = \lambda T^{-p}(\bar{m} - \frac{1}{2}\lambda T^{-p}\bar{M}^2)$ from Corollary 4.3. We also have by Lemma 4.1 that for all $t \in \mathbb{N}$

$$\mathbb{E}|X_t| \leq t \max_{u\leq t} \mathbb{E}|\Delta X_u| \leq t \max_{u\leq t} \mathbb{E}|C_u| < \infty.$$

On the set $\{\sup_{t\leq T}|g^*(X_t)| \leq G\}$ we have $C_t^2 \leq \lambda^2 G^2 t^{-2p}$ and

$$\sum_{t=1}^{T}\frac{\sigma_t^2 C_t^2}{\kappa_t^2} \leq \frac{\lambda^2 G^2}{\kappa_T^2}\sum_{t=1}^{T} t^{-2p} \leq \frac{\lambda^2 G^2}{\kappa_T^2}\int_0^T t^{-2p}dt = \frac{\lambda^2 G^2 T^{1-2p}}{(1-2p)\kappa_T^2} \leq \frac{4G^2 T}{(1-2p)\bar{m}^2}$$

where the last inequality follows from the assumption that $\lambda \leq \frac{2\bar{m}}{\bar{M}^2}\frac{T^p}{1+T^p} \leq \frac{\bar{m}}{\bar{M}^2}T^p$ so that

$$\kappa_T = \lambda T^{-p}(\bar{m} - \tfrac{1}{2}\lambda T^{-p}\bar{M}^2) \geq \tfrac{1}{2}\lambda T^{-p}\bar{m}.$$

It remains to apply Proposition 2.12:

$$\mathbb{P}\left(\sum_{t=0}^{T-1} X_t - \mathbb{E}\sum_{t=0}^{T-1} X_t \geq RT \; ; \; \sup_{t\leq T}\left|g^*(X_t)\right| \leq G\right)$$
$$\leq \mathbb{P}\left(\sum_{t=0}^{T-1} X_t - \mathbb{E}\sum_{t=0}^{T-1} X_t \geq RT \; ; \; \sum_{t=1}^{T}\frac{\sigma_t^2 C_t^2}{\kappa_t^2} \leq \frac{4G^2 T}{(1-2p)\bar{m}^2}\right)$$
$$\leq \exp\left(-\frac{(1-2p)\,\bar{m}^2\,R^2\,T}{32\,G^2}\right). \quad \square$$

4.2. Lipschitz observables and SDEs contractive at infinity

We use the notations $\|f\|_{\mathrm{Lip}} = \sup_{x\neq y}\frac{|f(x)-f(y)|}{|x-y|}$ for the Lipschitz seminorm,

$$W_1(\nu_1,\nu_2) = \sup_{f:\|f\|_{\mathrm{Lip}}\leq 1}\left(\int f\,d\nu_1 - \int f\,d\nu_2\right)$$

for the L^1 transportation distance, $\mu P_{s,t} = \int P_{s,t}(x,\cdot)\mu(dx)$ and $\mu(f) = \int f\,d\mu$ for a function f, measures μ, ν_1, ν_2 and a transition kernel $P_{s,t}$.

Consider the SDE

$$dX_t = b(t,X_t)\,dt + dB_t, \quad X_0 \sim \nu$$

with $b\colon [0,\infty)\times\mathbb{R}^d \to \mathbb{R}^d$ a locally Lipschitz continuous function, B a d-dimensional Brownian motion and ν a probability measure on $\mathbb{R}^d$.

We make the following additional assumption on “contractivity at infinity”: there exist constants $D, K > 0$ such that for all $t \geq 0$ and $x, y \in \mathbb{R}^d$ with $|x-y| > D$ we have

$$(x-y)\cdot(b(t,x)-b(t,y)) \leq -K|x-y|^2.$$

In [15] it was shown (in the time-homogeneous setting, but the methods extend directly to the time-inhomogeneous case [10]) that the assumption on b implies the exponential contractivity of the transition kernels $P_{s,t}$ associated to X in L^1 transportation distance: there exist constants $\rho, \kappa > 0$ such that for any two probability measures ν_1 and ν_2 on $\mathbb{R}^d$ we have

$$W_1(\nu_1 P_{s,t}, \nu_2 P_{s,t}) \le \rho e^{-\kappa(t-s)} W_1(\nu_1, \nu_2) \tag{4.3}$$

or equivalently [27] for all Lipschitz functions f

$$\|\nabla P_{s,t} f\|_\infty \le \rho e^{-\kappa(t-s)} \|f\|_{\text{Lip}}.$$

The key estimate (4.3), and the results of this section, hold in fact for a large class of SDEs "contractive at infinity". The work [15] already includes the case of a diffusion coefficient which is not the identity as well as explicit values for the constants, see also [8]. The paper [38] treats the case with a non-constant diffusion matrix and generalizes the results to Riemannian manifolds. For the non-autonomous situation, in the general context of Riemannian manifolds with possibly time-dependent metric and with explicit constants, see [10]. Another approach, which provides exponential gradient estimates for SDEs with highly degenerate diffusion matrices, can be found in [11].

We assume furthermore that ν satisfies a T_1 inequality [40]: there exists a constant C such that for any 1-Lipschitz function f and $\lambda > 0$, we have

$$\int e^{\lambda(f - \int f d\nu)} d\nu \le e^{\frac{\lambda^2 C}{2}}. \tag{4.4}$$

Note that for $\nu = \delta_x$ the T_1 inequality holds with constant $C = 0$.

We are going to show that whenever (4.3) and (4.4) hold then for all Lipschitz functions f and $R > 0$

$$\mathbb{P}\left(\frac{1}{T}\int_0^T f(X_t)\,dt - \mathbb{E}\left[\frac{1}{T}\int_0^T f(X_t)\,dt\right] \ge R\right) \\ \le \exp\left(-\frac{\kappa^2 R^2 T}{2\rho^2 \|f\|_{\text{Lip}}^2 \left(1 + C\,\frac{1-e^{-\kappa T}}{T}\right)}\right).$$

In the time-homogeneous setting, X has a unique stationary measure μ and we have

$$\mathbb{P}\left(\frac{1}{T}\int_0^T f(X_t)\,dt - \int f\,d\mu \ge R + \rho\|f\|_{\text{Lip}} \frac{1-e^{-\kappa T}}{\kappa T} W_1(\mu, \nu)\right) \\ \le \exp\left(-\frac{\kappa^2 R^2 T}{2\rho^2 \|f\|_{\text{Lip}}^2 \left(1 + C\,\frac{1-e^{-\kappa T}}{T}\right)}\right).$$

In the time-inhomogeneous setting, we have the existence of a unique evolution system of measures μ_t [13] such that

$$\mathbb{P}\left(\frac{1}{T}\int_0^T f(X_t)\,dt - \frac{1}{T}\int_0^T \mu_t(f)\,dt \ge R + \rho\|f\|_{\text{Lip}} \frac{1-e^{-\kappa T}}{\kappa T} W_1(\mu_0, \nu)\right) \\ \le \exp\left(-\frac{\kappa^2 R^2 T}{2\rho^2 \|f\|_{\text{Lip}}^2 \left(1 + C\,\frac{1-e^{-\kappa T}}{T}\right)}\right).$$

Lemma 4.5. *For any two continuously differentiable functions f, g on $\mathbb{R}^n$*

$$\langle f(X), g(X)\rangle_t = \frac{1}{2}\int_0^t \nabla f(X_s)\cdot\nabla g(X_s)\,ds.$$

Proof. By polarization it is sufficient to show that

$$\langle f(X)\rangle_t = \frac{1}{2}\int_0^t |\nabla f(X_s)|^2\,ds.$$

For f twice continuously differentiable, the result follows directly from Itô's formula. To extend the result to general f by approximation, define the sequence of stopping times $\tau_k = \inf\{t \geq 0 : |X_t| \geq k\}$. For each k fixed, choose a sequence of smooth, compactly supported functions f_n such that $f_n \to f$ and $|\nabla f_n| \to |\nabla f|$ uniformly on $\{x : |x| \leq k\}$. Then $f_n(X_{t\wedge\tau_k})$ converges to $f(X_{t\wedge\tau_k})$ uniformly on compacts in probability (u.c.p.), meaning that $\sup_{t\in[0,T]}|f(X_{t\wedge\tau_k}) - f_n(X_{t\wedge\tau_k})| \to 0$ in probability for any $T \geq 0$. Indeed, $\sup_t|f(X_{t\wedge\tau_k}) - f_n(X_{t\wedge\tau_k})| \leq \sup_{|x|\leq k}|f(x) - f_n(x)| \to 0$ as $n \to \infty$. By continuity of the predictable quadratic variation

$$\langle f_n(X)\rangle_{t\wedge\tau_k} \to \langle f(X)\rangle_{t\wedge\tau_k} \text{ u.c.p.}$$

On the other hand, by the uniform convergence of $|\nabla f_n|$ to $|\nabla f|$ on $\{x : |x| \leq k\}$,

$$\langle f_n(X)\rangle_{t\wedge\tau_k} = \frac{1}{2}\int_0^{t\wedge\tau_k} |\nabla f_n(X_s)|^2\,ds \to \frac{1}{2}\int_0^{t\wedge\tau_k} |\nabla f(X_s)|^2\,ds \text{ a.s.}$$

so that

$$\langle f(X)\rangle_{t\wedge\tau_k} = \frac{1}{2}\int_0^{t\wedge\tau_k} |\nabla f(X_s)|^2\,ds.$$

Now the result follows by letting $k \to \infty$. $\square$

Lemma 4.6. *For a 1-Lipschitz function f, let $M_t^T = \mathbb{E}^{\mathcal{F}_t}\int_0^T f(X_s)\,ds$. Then*

$$\langle M^T\rangle_T \leq \frac{\rho^2 T}{\kappa^2}.$$

Proof. From (4.3) it follows that $P_t f$ is continuously differentiable and the Lipschitz condition on f ensures that $\mathbb{E}f(X_t)^2 < \infty$ for all t. By the preceding Lemma we can apply Proposition 3.7 with $\Gamma(f, g) = \frac{1}{2}\nabla f\cdot\nabla g$. Since by Cauchy–Schwarz $2\Gamma(f, g) \leq |\nabla f||\nabla g|$ we get

$$\begin{aligned} d\langle M^T\rangle_t &= \int_t^T\int_t^T 2\Gamma(P_{t,u}f, P_{t,v}f)(X_t)\,du\,dv \\ &\leq \left(\int_t^T \left|\nabla P_{t,u}f(X_t)\right|\,du\right)^2 dt \leq \rho^2\left(\int_0^{T-t} e^{-\kappa u}\,du\right)^2 dt \leq \frac{\rho^2}{\kappa^2}\,dt \end{aligned}$$

and $\langle M^T\rangle_T \leq \frac{\rho^2 T}{\kappa^2}$. $\square$

Lemma 4.7. *For all 1-Lipschitz functions f we have*

$$\Lambda(\lambda) \leq \frac{\lambda^2 C\rho^2}{2}\left(\frac{1 - e^{-\kappa T}}{\kappa}\right)^2$$

where

$$\Lambda(\lambda) = \log \mathbb{E}\, e^{\lambda\left(\mathbb{E}^{\mathcal{F}_0} S_T - \mathbb{E} S_T\right)}.$$

and

$$S_T = \int_0^T f(X_t)dt.$$

Proof. We have by Theorem 3.1 and the Markov property

$$\mathbb{E}^{\mathcal{F}_0} S_T - \mathbb{E} S_T = \int_0^T P_{0,t} f(X_0)\, dt - \mathbb{E} \int_0^T P_{0,t} f(X_0)\, dt = F(X_0) - \int F d\nu$$

with

$$F(x) = \int_0^T P_{0,t} f(x)\, dt.$$

From the triangle inequality for the Lipschitz seminorm and (4.3) we get

$$\|F\|_{\mathrm{Lip}} \le \int_0^T \|P_{0,t} f\|_{\mathrm{Lip}}\, dt \le \int_0^T \rho e^{-\kappa t}\, dt = \rho \frac{1 - e^{-\kappa T}}{\kappa}$$

and the result follows directly from (4.4). □

Proposition 4.8. *For all Lipschitz functions f and $R > 0$ we have*

$$\mathbb{P}\left(\frac{1}{T}\int_0^T f(X_t)\, dt - \mathbb{E}\left[\frac{1}{T}\int_0^T f(X_t)\, dt\right] \ge R\right) \le \exp\left(-\frac{\kappa^2 R^2 T}{2\rho^2 \|f\|_{\mathrm{Lip}}^2 \left(1 + C\, \frac{1-e^{-\kappa T}}{T}\right)}\right).$$

Proof. The result follows directly from the preceding Lemmas and Corollary 3.12. Indeed, noting that $f/\|f\|_{\mathrm{Lip}}$ is 1-Lipschitz and that $H_T^0 = \langle M^T \rangle_T$,

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{T}\int_0^T f(X_t)\, dt - \mathbb{E}\left[\frac{1}{T}\int_0^T f(X_t)\, dt\right] \ge R\right) \\
&\quad = \mathbb{P}\left(\int_0^T f(X_t)/\|f\|_{\mathrm{Lip}}\, dt - \mathbb{E}\int_0^T f(X_t)/\|f\|_{\mathrm{Lip}}\, dt \ge RT/\|f\|_{\mathrm{Lip}}\ ;\ H_T^0 \le \frac{\rho^2 T}{\kappa^2}\right) \\
&\quad \le \exp\left(-\frac{\kappa^2 R^2 T}{2\rho^2 \|f\|_{\mathrm{Lip}}^2 \left(1 + C\, \frac{1-e^{-\kappa T}}{T}\right)}\right). \quad \square
\end{aligned}$$

Lemma 4.9. *The process X admits an evolution system of measures μ_t such that for any 1-Lipschitz function f we have*

$$\mathbb{E}\left[\frac{1}{T}\int_0^T f(X_t)\, dt\right] - \frac{1}{T}\int_0^T \mu_t(f)\, dt \le \rho \frac{1 - e^{-\kappa T}}{\kappa T} W_1(\nu, \mu_0).$$

In the time-homogeneous case $\mu_t = \mu$ is the stationary measure associated to X.

Proof. Existence and uniqueness of an evolutionary system of measures is a special case of Theorem 3.5 in [10]. In the time-homogeneous case existence and uniqueness of a stationary measure is also shown in [15] and is a direct consequence of (4.3).

Since from the definition of an evolutionary system of measures $\mu_t = \mu_0 P_{0,t}$ we have by (4.3)

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{T}\int_0^T f(X_t)\,dt - \frac{1}{T}\int_0^T \mu_t(f)\,dt\right] \\
&= \frac{1}{T}\int_0^T \left(\nu P_{0,t} f - \mu_0 P_{0,t} f\right)\,dt \le \frac{1}{T}\int_0^T W_1(\nu P_{0,t}, \mu_0 P_{0,t})\,dt \\
&\le \frac{1}{T}\int_0^T \rho e^{-\kappa t} W_1(\nu, \mu_0)\,dt \le \rho\frac{1-e^{-\kappa T}}{\kappa T} W_1(\nu, \mu_0). \quad \square
\end{aligned}$$

Proposition 4.10. *For all Lipschitz functions f and $R > 0$ we have*

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{T}\int_0^T f(X_t)\,dt - \frac{1}{T}\int_0^T \mu_t(f)\,dt \ge R + \rho\|f\|_{\mathrm{Lip}}\frac{1-e^{-\kappa T}}{\kappa T} W_1(\mu_0, \nu)\right) \\
&\le \exp\left(-\frac{\kappa^2 R^2 T}{2\rho^2\|f\|_{\mathrm{Lip}}^2\left(1 + C\,\frac{1-e^{-\kappa T}}{T}\right)}\right).
\end{aligned}$$

Proof. By the preceding Lemma

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{T}\int_0^T f(X_t)\,dt - \frac{1}{T}\int_0^T \mu_t(f)\,dt \ge R + \rho\|f\|_{\mathrm{Lip}}\frac{1-e^{-CT}}{CT} W_1(\mu_0, \nu)\right) \\
&\le \mathbb{P}\left(\frac{1}{T}\int_0^T f(X_t)\,dt - \mathbb{E}\left[\frac{1}{T}\int_0^T f(X_t)\,dt\right] \ge R\right)
\end{aligned}$$

and the result follows immediately from the preceding Proposition. $\square$

4.3. Martingale integrands

Proposition 4.11. *For a Brownian motion B we have for all $R, T \ge 0$*

$$\mathbb{P}\left(\frac{1}{T^2}\int_0^T B_t\,dt \ge R\right) \le \exp(-3R^2 T).$$

Proof. Let $M_t^T = \mathbb{E}^{\mathcal{F}_t}\int_0^T B_u\,du$. Then by Proposition 3.3 we have

$$\langle M^T\rangle_T = \int_0^T (T-t)^2\,d\langle B\rangle_t = \int_0^T (T-t)^2\,dt = \frac{T^3}{3}$$

and by Corollary 3.12 with $H_T^0 = \langle M^T\rangle_T$

$$\mathbb{P}\left(\frac{1}{T^2}\int_0^T B_t\,dt \ge R\right) = \mathbb{P}\left(\int_0^T B_t\,dt \ge RT^2\,;\, H_T^0 \le \frac{T^3}{3}\right) \qquad \le \exp(-3R^2 T). \quad \square$$

Proposition 4.12. *For a Poisson process N we have for all $R, T > 0$*

$$\mathbb{P}\left(\frac{1}{T^2}\int_0^T N_t\,dt - \frac{1}{2} \geq R\right) \leq \exp\left(-\frac{T}{3}h(3R)\right) \leq \exp\left(-\frac{3R^2T}{2(1+R)}\right)$$

with

$$h(x) = (1+x)\log(1+x) - x.$$

Proof. Let $X_t = N_t - t$ and $M_t^T = \mathbb{E}^{\mathcal{F}_t}\int_0^T X_t\,dt$. We have by Proposition 3.3

$$\Delta M_t^T = (T-t)\Delta X_t \leq T, \quad 0 \leq t \leq T$$

and

$$\langle M^T\rangle_T = \int_0^T (T-t)^2\,d\langle X\rangle_t = \int_0^T (T-t)^2\,dt = \frac{T^3}{3}$$

so that

$$H_T^T = \langle M^T\rangle_T + \sum_0^T (\Delta M_t^T)^2 \mathbb{1}\{\Delta M_t^T > T\} = \langle M^T\rangle_T = \frac{T^3}{3}$$

and by Corollary 3.13

$$\begin{aligned}\mathbb{P}\left(\frac{1}{T^2}\int_0^T N_t\,dt - \frac{1}{2} \geq R\right) &= \mathbb{P}\left(\int_0^T X_t\,dt \geq RT^2\ ;\ H_T^T \leq \frac{T^3}{3}\right)\\ &\leq \exp\left(-\frac{T}{3}h(3R)\right).\end{aligned}$$

The second inequality in the result follows immediately from the elementary inequality $h(x) \geq \frac{x^2}{2(1+x/3)}$ for $x > 0$. □

Proposition 4.13. *For a Brownian motion B we have for all $R, T \geq 0$*

$$\mathbb{P}\left(\left|\frac{1}{T^2}\int_0^T B_t^2\,dt - \frac{1}{2}\right| \geq R\right) \leq 2^{1/3}\left(1 \wedge \frac{4R+3/4}{3R^2}\right)\exp\left(-\frac{R^2}{3(4R+3/4)}\right).$$

Proof. Consider the local martingale $X_t = B_t^2 - t$. We have using integration by parts that

$$X_t = 2\int_0^t B_s\,dB_s$$

so that

$$\langle X\rangle_t = 4\int_0^t B_s^2\,ds = 4\int_0^t (X_s + s)\,ds.$$

By (2.4) and Remark 3.4 we have

$$\langle M^T\rangle_T = \int_0^T (T-t)^2\,d\langle X\rangle_t = 4\int_0^T (T-t)^2\,(X_t + t)\,dt$$

and

$$\mathbb{E}\langle M^T\rangle_T = 4\int_0^T (T-t)^2\,t\,dt.$$

Finally

$$\langle M^T\rangle_T + \mathbb{E}\langle M^T\rangle_T = 4\int_0^T (T-t)^2\, X_t\, dt + 8\int_0^T (T-t)^2\, t\, dt$$
$$\leq 4T^2\int_0^T X_t\, dt + (3/4)T^4$$

and the result follows from Corollary 3.16 with $C = 4T^2$ and $D = (3/4)T^4$. $\square$

4.4. Squared Ornstein–Uhlenbeck process

Let X^x be the Ornstein–Uhlenbeck process on $\mathbb{R}^d$, solution to the SDE

$$dX_t^x = -\kappa X_t^x + dB_t, \quad X_0^x = x$$

with $\kappa > 0$ and B_t a d-dimensional Brownian motion.

In this section, we will derive concentration inequalities for additive functionals with the square of X^x as integrand. This case is challenging since for $\phi(x) = |x|^2$, $\nabla P_t\phi(x)$ cannot be bounded uniformly in x. We will make use of the special properties of the Ornstein–Uhlenbeck semigroup. In the next section we will extend the approach to a slightly more general situation, which however does not recover the bound developed in this section. The case of the squared Ornstein–Uhlenbeck process was previously studied in [28] (Example 4.2) and [19] (Example 3.1) using analytic methods, which require the initial law of X to be absolutely continuous with respect to the stationary measure of X.

Proposition 4.14. *We have for all $R, T > 0, x \in \mathbb{R}^d$*

$$\mathbb{P}\left(\left|\frac{1}{T}\int_0^T |X_t^x|^2\, dt - \mathbb{E}\left[\frac{1}{T}\int_0^T |X_t^x|^2\, dt\right]\right| \geq R\right)$$
$$\leq 2^{1/3}\left(1 \wedge \frac{R+D}{3\kappa^2 R^2 T}\right)^{1/3} \exp\left(-\frac{\kappa^2 R^2 T}{3(R+D)}\right)$$

with

$$D = \frac{|x|^2}{\kappa T} + \frac{d}{\kappa}$$

Proof. We have component-wise $(X_t^x)_i = x_i e^{-\kappa t} + \int_0^t e^{-\kappa(t-s)}\, dB_s^i$. Let $\phi(x) = |x|^2$ and for $t \geq 0$ set $P_t\phi(x) = \mathbb{E}\phi(X_t) = \mathbb{E}|X_t|^2$. Then for $|x| < 1/\varepsilon$ for some arbitrary $\varepsilon > 0$, we can differentiate under the expectation

$$\partial_i \mathbb{E}\phi(X_t^x) = \mathbb{E}\partial_i\phi(X_t^x) = 2x_i e^{-2\kappa t}, \quad |x| < 1/\varepsilon$$

so that

$$\nabla P_t\phi(x) = 2xe^{-2\kappa t}, \quad |x| \leq 1/\varepsilon. \tag{4.5}$$

By Lemma 4.5, for any two continuously differentiable f, g we have

$$\langle f(X^x), g(X^x)\rangle_t = \frac{1}{2}\int_0^t \nabla f(X_s^x)\nabla g(X_s^x)\, ds$$

so that by Proposition 3.7, on $A_\varepsilon := \{\sup_{0\le t\le T}|X_t^x| < 1/\varepsilon\}$,

$$\begin{aligned} d\langle M^T\rangle_t &= \int_t^T \int_t^T (\nabla P_{u-t}\phi \cdot \nabla P_{v-t}\phi)(X_t^x)\,du\,dv\,dt \\ &= 4|X_t^x|^2 \left(\int_0^{T-t} e^{-2\kappa u}du\right)^2 dt \\ &= \frac{(X_t^x)^2}{\kappa^2}\left(1-e^{-2\kappa(T-t)}\right)^2 dt. \end{aligned}$$

Integrating, we get

$$\langle M^T\rangle_T = \frac{1}{\kappa^2}\int_0^T |X_t^x|^2\left(1-e^{-2\kappa(T-t)}\right)^2 dt \le \frac{1}{\kappa^2}\int_0^T |X_t^x|^2\,dt$$

so that Corollary 3.17 applies with $C = \frac{1}{\kappa^2}$ and $D = 0$:

$$\begin{aligned} &\mathbb{P}\left(\left|\int_0^T |X_t^x|^2\,dt - \mathbb{E}\int_0^T |X_t^x|^2\,dt\right| \ge RT\ ;\ A_\varepsilon\right) \\ &\quad= \mathbb{P}\left(\left|\int_0^T |X_t^x|^2\,dt - \mathbb{E}\int_0^T |X_t|^2\,dt\right| \ge RT\ ;\ H_T^0 \le \frac{1}{\kappa^2}\int_0^T |X_t^x|^2\,dt\right) \\ &\quad\le 2^{1/3}\left(1\wedge \frac{R+D'}{3\kappa^2R^2T}\right)^{1/3}\exp\left(-\frac{\kappa^2R^2T}{3(R+D')}\right) \end{aligned}$$

with $D' = |\mathbb{E}S_T|/T + (\kappa^2/T)\mathbb{E}H_T^0 \le 2|\mathbb{E}S_T|/T$.

Since

$$\mathbb{E}|X_t^x|^2 = |x|^2 e^{-2\kappa t} + d\,\frac{1-e^{-2\kappa t}}{2\kappa}$$

we have that

$$D' \le \frac{2}{T}\int_0^T \mathbb{E}|X_t^x|^2\,dt \le \frac{|x|^2}{\kappa T} + \frac{d}{\kappa}.$$

We can lift the restriction to A_ε by noting that by Markov's inequality and Doob's maximal inequality

$$\begin{aligned} \mathbb{P}(A_\varepsilon^c) &= \mathbb{P}(\sup_t |X_t^x| > 1/\varepsilon) \le \mathbb{P}(\sup_{0\le t\le T} e^{\kappa t}|X_t^x| > 1/\varepsilon) = \mathbb{P}\left(\sup_{0\le t\le T}\left|\int_0^t e^{\kappa s}\,dB_s\right| > 1/\varepsilon\right) \\ &\le \varepsilon^2\mathbb{E}\sup_{0\le t\le T}\left|\int_0^t e^{\kappa s}\,dB_s\right|^2 \le 4\varepsilon^2\,\frac{e^{2\kappa T}-1}{\kappa} \end{aligned}$$

so that finally

$$\begin{aligned} &\mathbb{P}\left(\left|\int_0^T |X_t^x|^2\,dt - \mathbb{E}\int_0^T |X_t|^2\,dt\right| \ge RT\right) \\ &\quad\le \mathbb{P}\left(\left|\int_0^T |X_t^x|^2\,dt - \mathbb{E}\int_0^T |X_t|^2\,dt\right| \ge RT\ ;\ A_\varepsilon\right) + \mathbb{P}(A_\varepsilon^c) \\ &\quad\le 2^{1/3}\left(1\wedge \frac{R+D}{3\kappa^2R^2T}\right)^{1/3}\exp\left(-\frac{\kappa^2R^2T}{3(R+D)}\right) + O(\varepsilon^2) \end{aligned}$$

with

$$D = \frac{|x|^2}{\kappa T} + \frac{d}{\kappa}$$

and the result follows by letting $\varepsilon \to 0$. □

Remark 4.15. From the previous proposition, we also get that

$$\lim_{T\to\infty} T^{-1} \log \mathbb{P}\left(\frac{1}{T}\int_0^T |X_t^x|^2 \geq R\right) \leq -I(R) \tag{4.6}$$

with

$$I(R) = \frac{(d-2\kappa R)^2}{12(R+d/(2\kappa))}.$$

According to large deviation estimates from [5], the optimal bound in the right-hand side of (4.6) is obtained by replacing I with the good rate function

$$J(R) = \frac{(d-2\kappa R)^2}{8R}.$$

Tracing back the computations, we see that the discrepancy between the denominators of I and J, namely the factor 12 instead of 8 and the extra term $d/(2\kappa)$ for I, can be directly attributed to the factor $2/3$ and term $\mathbb{E}[A^2]$ in the self-normalized martingale inequality (3.11). This discrepancy is the same as the one observed between the sharp bound for normal random variables and the self-normalized bound in [32], see also Remark 2.2 in that paper. This suggests that the martingale M^T belongs to a subclass of martingales for which the bound in [32] can be sharpened.

4.5. Squared Lipschitz integrands

Let X be a Markov process with generator L, transition kernel $P_{s,t}$ and squared field operator Γ, i.e. $\Gamma(f) = \frac{1}{2}(Lf^2 - 2fLf)$.

We assume the following commutation property between the squared field operator and the transition kernel:

$$\Gamma(P_{s,t}f) \leq \sigma^2 \left(P_{s,t}\sqrt{\Gamma(f)}\right)^2 e^{-2\kappa(t-s)}$$

for some $\sigma, \kappa \geq 0$.

Using the inequalities on self-normalized martingales, we can derive a Bernstein-type inequality for time averages of positive functions g^2 such that $\Gamma(g^2) \leq 2g^2$.

Proposition 4.16. *For all twice continuously differentiable g^2 such that $\Gamma(g^2) \leq 2g^2$ and $Lg^2 \leq -2Cg^2 + D^2$ for some constants $C \in \mathbb{R}$, $D \geq 0$ we have for*

$$S_T = \int_0^T g^2(X_t)\,dt$$

the following Bernstein-type inequality:

$$\begin{aligned}&\mathbb{P}(|S_T - \mathbb{E}S_T| \geq RT)\\ &\leq 2^{1/3}\exp\left(-\frac{R^2T}{24\sigma^2\left(C_\kappa^2(T)R + 2C_\kappa^2(T)\mathbb{E}(S_T/T) + 2D_{\kappa,C}^2(T)\right)}\right)\end{aligned}$$

with

$$C_\kappa(T) = \int_0^T e^{-(\kappa+C)u}\, du$$
$$D_{\kappa,C}(T) = D \int_0^T e^{-\kappa u} \int_0^u e^{-C(u-v)}\, dv\, du.$$

Proof. Since g^2 is assumed twice continuously differentiable, we have $\partial_u P_{t,u} g^2 = P_{t,u} L g^2$ and we get from our assumption $Lg^2 \le -2Cg^2 + D^2$ by Gronwall's lemma that

$$P_{t,u} g^2 \le g^2 e^{-2C(u-t)} + D^2 \int_t^u e^{-2C(u-v)}\, dv.$$

Together with the assumption that $\Gamma(g^2) \le 2g^2$ we have

$$P_{t,u}(2\Gamma(g^2)) \le 4P_{t,u}(g^2) \le 4g^2 e^{-2C(u-t)} + D^2 \int_t^u e^{-2C(u-v)}\, dv$$

and by applying the assumption on commutation of Γ and P

$$\begin{aligned} 2\Gamma(P_{t,u} g^2) &\le \sigma^2 (P_{t,u}\sqrt{2\Gamma(g^2)})^2 e^{-2\kappa(u-t)} \le \sigma^2 P_{t,u}(2\Gamma(g^2)) e^{-2\kappa(u-t)} \\ &\le 4\sigma^2 g^2\, e^{-2(\kappa+C)(u-t)} + 4\sigma^2 D^2\, e^{-2\kappa(u-t)} \int_t^u e^{-2C(u-v)}\, dv. \end{aligned}$$

Now from Proposition 3.7 we have for all $T > 0$

$$\begin{aligned} d\langle M^T\rangle_t &= \int_t^T \int_t^T 2\Gamma(P_{t,u} g^2, P_{t,v} g^2)(X_t)\, dv\, du\, dt \\ &\le \left(\int_t^T \sqrt{2\Gamma(P_{t,u} g^2)(X_t)}\, du \right)^2 dt \\ &\le 4\sigma^2 \left(\sqrt{g^2(X_t)} \int_t^T e^{-(\kappa+C)(u-t)}\, du + D \int_t^T e^{-\kappa(u-t)} \int_t^u e^{-C(u-v)}\, dv\, du \right)^2 dt \\ &\le 8\sigma^2 \left(C_\kappa^2(T) g^2(X_t) + D_{\kappa,C}^2(T) \right) dt \end{aligned}$$

so that

$$\langle M^T\rangle_T \le 8\sigma^2 C_\kappa^2(T) \int_0^T g^2(X_t)\, dt + 8\sigma^2 D_{C,\kappa}^2(T)\, T$$

with $C_\kappa(T)$, $D_{C,\kappa}(T)$ as in the statement of the Proposition. The result now follows from Corollary 3.15. □

Remark 4.17. If X is a diffusion in the sense that $\Gamma(\Phi(f_1), f_2) = \Phi'(f_1)\Gamma(f_1, f_2)$ for all continuously differentiable Φ and f_1, f_2 in the domain of Γ, then for all continuously differentiable functions g such that $2\Gamma(g) \le 1$, we have $\Gamma(g^2) = (2g)^2\Gamma(g) \le 2g^2$. In particular, if $\Gamma(g) = \frac{1}{2}\|\nabla g\|^2$, then this holds when g is differentiable and 1-Lipschitz.

The preceding proposition applies in particular to the Ornstein–Uhlenbeck process on $\mathbb{R}^d$ with $g(x)^2 = |x|^2$. Indeed, for $\kappa > 0$, consider the d-dimensional Ornstein–Uhlenbeck process with generator

$$Lf(x) = -\kappa x \cdot \nabla f + \frac{1}{2}\Delta f.$$

By a direct calculation we have $Lg^2 = -2\kappa g^2 + d$. The corresponding squared field operator is $\Gamma(f) = \frac{1}{2}|\nabla f|^2$ so that $\Gamma(g^2) = 2g^2$. We also have

$$\Gamma(P_{s,t} f) \leq \left(P_{s,t}(\sqrt{\Gamma(f)})\right)^2 e^{-2\kappa(t-s)},$$

see for example [2]. From the expression for Lg^2 we get furthermore $\partial_t P_{0,t} g^2 = P_{0,t} L g^2 = -2\kappa P_{0,t} g^2 + d$ so that for $X_0 = 0$ we have $\mathbb{E} g^2(X_t) = P_{0,t} g^2(X_0) = d/(2\kappa)(1 - e^{-2\kappa t})$ and $\mathbb{E} S_T = \int_0^T \mathbb{E} g^2(X_t)\, dt \leq d/(2\kappa) T$. In the notation of the preceding Proposition, we have $C_\kappa \leq 1/(2\kappa)$, $D_{\kappa,C} \leq d/(\kappa^2)$ so that

$$\mathbb{P}\left(\left|\frac{1}{T}\int_0^T |X_t|^2\, dt - \mathbb{E}\left[\frac{1}{T}\int_0^T |X_t|^2\, dt\right]\right| \geq R\right)$$
$$\leq 2^{1/3} \exp\left(-\frac{\kappa^2 R^2 T}{6\left(R + d/\kappa + 8d^2/(\kappa^2)\right)}\right).$$

This bound is looser than the one obtained in Proposition 4.14 above, which relies on the special property of the Ornstein–Uhlenbeck process (4.5) to obtain the tighter bound.

4.6. SDEs with degenerate diffusion matrix

For $\alpha, \beta > 0$ let $(X^{x,y}, Y^{x,y})_{x,y\in\mathbb{R}}$ be the family of solutions to

$$dX_t^{x,y} = -\alpha X_t^{x,y} dt + dB_t, \quad X_0^{x,y} = x,$$
$$dY_t^{x,y} = -\beta Y_t^{x,y} dt + X_t^{x,y} dt, \quad Y_0^{x,y} = y.$$

In other words, $X^{x,y}$ is an Ornstein–Uhlenbeck process and $Y_t^{x,y}$ can be written $Y_t^{x,y} = y\, e^{-\beta t} + \int_0^t e^{-\beta(t-s)} X_s^x\, ds$. The associated semigroup is $P_t f(x, y) = \mathbb{E} f(X_t^{x,y}, Y_t^{x,y})$ and the squared field operator is

$$\Gamma(f)(x, y) = |\partial_x f(x, y)|^2.$$

In particular, since Y has differentiable trajectories, this example illustrates that our concentration inequalities do not rely on "roughness" of the trajectories.

Proposition 4.18. *For all 1-Lipschitz functions f and $R, T > 0$*

$$\mathbb{P}\left(\frac{1}{T}\int_0^T f(Y_t^{x,y})\, dt - \mathbb{E}\left[\frac{1}{T}\int_0^T f(Y_t^{x,y})\, dt\right] \geq R\right)$$
$$\leq \exp\left(\frac{-R^2\, T\, (\alpha \wedge \beta)^2\, |\alpha - \beta|^2}{4\left(1 - e^{-(\alpha\wedge\beta)T}\right)^2 \left(1 - e^{-|\alpha-\beta|T}\right)^2}\right).$$

Proof. We have by Itô's formula that

$$e^{\beta t} Y_t^{x,y} = y + \int_0^t e^{(\beta-\alpha)s} e^{\alpha s} X_s^{x,y} ds$$
$$= y + x\int_0^t e^{(\beta-\alpha)s} ds + \int_0^t e^{(\beta-\alpha)s}\left(\int_0^s e^{\alpha r} dB_r\right) ds$$

so that for all $h, x, y \in \mathbb{R}$, by canceling terms that do not depend on x, integrating and rearranging,

$$\frac{Y^{x+h,y} - Y^{x,y}}{h} = \frac{e^{-\alpha t} - e^{-\beta t}}{\beta - \alpha} = \frac{e^{-(\alpha\wedge\beta)t}(1 - e^{-|\alpha-\beta|t})}{|\alpha - \beta|}.$$

In particular, for a 1-Lipschitz function $f(y)$, using the Lipschitz property and the bound from above

$$\begin{aligned}\sqrt{\Gamma(P_t f)(x, y)} = |\partial_x P_t f(x, y)| &= \lim_{h\to 0} \frac{1}{|h|}|\mathbb{E} f(Y_t^{x+h,y}) - f(Y_t^{x,y})| \\ &\le \lim_{h\to 0} \mathbb{E}\frac{|Y_t^{x+h,y} - Y_t^{x,y}|}{|h|} \\ &= \frac{e^{-(\alpha\wedge\beta)t}(1 - e^{-|\alpha-\beta|t})}{|\alpha - \beta|}.\end{aligned}$$

By Proposition 3.7, time-homogeneity of P_t and the Cauchy–Schwartz inequality for Γ

$$\begin{aligned}\langle M^T\rangle_T &= \int_0^T \int_t^T \int_t^T \Gamma(P_{t,u} f, P_{t,v} f)(X_t^{x,y}, Y_t^{x,y})\, du\, dv\, dt \\ &\le 2\int_0^T \left(\int_0^{T-t} \sqrt{\Gamma(P_u f)(X_t^{x,y}, Y_t^{x,y})}\, du\right)^2 dt.\end{aligned}$$

We have from the bound on $\sqrt{\Gamma(P_t f)(x, y)}$ derived above

$$\begin{aligned}\int_0^{T-t} \sqrt{\Gamma(P_u f)(X_t^{x,y}, Y_t^{x,y})} du &\le \int_0^T e^{-(\alpha\wedge\beta)u} du\, \frac{(1 - e^{-|\alpha-\beta|T})}{|\alpha - \beta|} \\ &\le \frac{\left(1 - e^{-(\alpha\wedge\beta)T}\right)}{(\alpha\wedge\beta)} \frac{\left(1 - e^{-|\alpha-\beta|T}\right)}{|\alpha - \beta|}\end{aligned}$$

so that finally

$$\langle M^T\rangle_T \le 2T \frac{\left(1 - e^{-(\alpha\wedge\beta)T}\right)^2}{(\alpha\wedge\beta)^2} \frac{\left(1 - e^{-|\alpha-\beta|T}\right)^2}{|\alpha - \beta|^2}$$

and the result follows. □

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.spa.2021.01.004.

Funding

This work was supported by the National Research Fund, Luxembourg.

References

[1] Roland Assaraf, Benjamin Jourdain, Tony Lelièvre, Raphaël Roux, Computation of sensitivities for the invariant measure of a parameter dependent diffusion, Stoch. Partial Differ. Equ. Anal. Comput. (2017).

[2] Dominique Bakry, Ivan Gentil, Michel Ledoux, Analysis and Geometry of Markov Diffusion Operators, in: Grundlehren der Mathematischen Wissenschaften, vol. 348, Springer International Publishing, Cham, 2014.

[3] Bernard Bercu, Bernard Delyon, Emmanuel Rio, Concentration Inequalities for Sums and Martingales, in: SpringerBriefs in Mathematics, Springer International Publishing, Cham, 2015.

[4] Stéphane Boucheron, Gábor Lugosi, Pascal Massart, Concentration inequalities: a nonasymptotic theory of independence, first ed., Oxford University Press, Oxford, 2013, p. 481.

[5] Wlodzimierz Bryc, Amir Dembo, Large deviations for quadratic functionals of Gaussian processes, J. Theor. Probab. 10 (2) (1997) 307–332, Publisher: Springer.

[6] Patrick Cattiaux, Djalil Chafaı, Arnaud Guillin, Central limit theorems for additive functionals of ergodic Markov diffusions processes, ALEA Lat. Am. J. Probab. Math. Stat. 9 (2) (2012) 337–382.

[7] Patrick Cattiaux, Arnaud Guillin, Deviation bounds for additive functionals of markov processes, ESAIM: Probab. Statist. 12 (2008) 12–29.

[8] P. Cattiaux, A. Guillin, Semi Log-Concave Markov diffusions, in: Catherine Donati-Martin, Antoine Lejay, Alain Rouault (Eds.), Séminaire de Probabilités XLVI, Vol. 2123, Springer International Publishing, Cham, 2014, pp. 231–292.

[9] Li-Juan Cheng, Anton Thalmaier, Evolution systems of measures and semigroup properties on evolving manifolds, Electron. J. Probab. 23 (2018) 27.

[10] Li-Juan Cheng, Anton Thalmaier, Shao-Qin Zhang, Exponential contraction in wasserstein distance on static and evolving manifolds, 2020, In: arXiv:2001.06187 [math]. arXiv:2001.06187.

[11] D. Crisan, P. Dobson, M. Ottobre, Uniform in time estimates for the weak error of the Euler method for SDEs and a Pathwise Approach to Derivative Estimates for Diffusion Semigroups, 2019, arXiv:1905.03524 [math]. arXiv:1905.03524.

[12] Dan Crisan, Michela Ottobre, Pointwise gradient bounds for degenerate semigroups (of UFG type), Proc. R. Soc. A 472 (2195) (2016).

[13] Giuseppe Da Prato, Michael Röckner, A note on evolution systems of measures for time-dependent stochastic differential equations, in: Seminar on Stochastic Analysis, Random Fields and Applications V, Springer, 2007, pp. 115–122.

[14] K. Dzhaparidze, J.H. van Zanten, On Bernstein-type inequalities for martingales, Stoch. Process. Appl. 93 (1) (2001) 109–117.

[15] Andreas Eberle, Reflection couplings and contraction rates for diffusions, Probab. Theory Related Fields (2015).

[16] Andreas Eberle, Arnaud Guillin, Raphael Zimmer, Couplings and quantitative contraction rates for Langevin dynamics, Ann. Probab. 47 (4) (2019) 1982–2010, Publisher: Institute of Mathematical Statistics.

[17] K.D. Elworthy, Xue-Mei Li, Formulae for the derivatives of heat semigroups, J. Funct. Anal. 125 (1) (1994) 252–286.

[18] Gersende Fort, Central limit theorems for stochastic approximation with controlled Markov chain dynamics, ESAIM: Probab. Statist. 19 (2015) 60–80, Publisher: EDP Sciences.

[19] Fuqing Gao, Arnaud Guillin, Liming Wu, Bernstein-type concentration inequalities for symmetric Markov processes, Theory Probab. Appl. 58 (3) (2014) 358–382.

[20] Arnaud Guillin, Moderate deviations of inhomogeneous functionals of Markov processes and application to averaging, Stoch. Process. Appl. 92 (2) (2001) 287–313.

[21] A. Guillin, C. Leonard, F.-Y. Wang, L. Wu, Transportation-information inequalities for Markov processes (II) : relations with other functional inequalities, 2009, In: ArXiv e-prints.

[22] Arnaud Guillin, Christian Léonard, Liming Wu, Nian Yao, Transportation-information inequalities for Markov processes, Probab. Theory Related Fields 144 (3) (2009) 669–695.

[23] Martin Hairer, Jonathan C. Mattingly, Michael Scheutzow, Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations, Probab. Theory Related Fields 149 (1) (2011) 223–259.

[24] Arnulf Jentzen, Benno Kuckuck, Ariel Neufeld, Philippe von Wurstemberger, Strong error analysis for stochastic gradient descent optimization algorithms, 2018, arXiv:1801.09324 [math]. arXiv:1801.09324.

[25] Aldéric Joulin, A new Poisson-type deviation inequality for Markov jump processes with positive Wasserstein curvature, Bernoulli 15 (2) (2009) 532–549.

[26] Aldéric Joulin, Yann Ollivier, Curvature, concentration and error estimates for Markov chain Monte Carlo, Ann. Probab. 38 (6) (2010) 2418–2442.

[27] Kazumasa Kuwada, Duality on gradient estimates and Wasserstein controls, J. Funct. Anal. 258 (11) (2010) 3758–3774.

[28] Pascal Lezaud, Chernoff and Berry–Esséen inequalities for Markov processes, ESAIM: Probab. Statist. 5 (2001) 183–201.
[29] Eva Löcherbach, Dasha Loukianova, Deviation inequalities for centered additive functionals of recurrent harris processes having general state space, J. Theor. Probab. 25 (1) (2012) 231–261.
[30] Wenlong Mou, Chris Junchi Li, Martin J. Wainwright, Peter L. Bartlett, Michael I. Jordan, On Linear Stochastic Approximation: Fine-grained Polyak-Ruppert and Non-Asymptotic Concentration, 2020, _eprint: 2004.04719.
[31] Yann Ollivier, Ricci curvature of Markov chains on metric spaces, J. Funct. Anal. 256 (3) (2009) 810–864.
[32] Victor de la Peña, Guodong Pang, Exponential inequalities for self-normalized processes with applications, Electron. Commun. Probab. 14 (0) (2009) 372–381.
[33] B.T. Polyak, A.B. Juditsky, Acceleration of stochastic approximation by averaging, SIAM J. Control Optim. 30 (4) (1992) 838–855.
[34] Herbert Robbins, Sutton Monro, A stochastic approximation method, Ann. Math. Statist. 22 (3) (1951) 400–407, Publisher: The Institute of Mathematical Statistics.
[35] Daniel W. Stroock, S.R.S. Varadhan, Multidimensional diffusion processes, in: Grundlehren der mathematischen Wissenschaften, vol. 233, Springer, Berlin ; New York, 2006, p. 338.
[36] Anton Thalmaier, On the differentiation of heat semigroups and Poisson integrals, Stoch. Int. J. Probab. Stoch. Process. 61 (3) (1997) 297–321.
[37] Feng-Yu Wang, Analysis for Diffusion Processes on Riemannian Manifolds, World Scientific Pub. Co, 2014, p. 379.
[38] Feng-Yu Wang, Exponential contraction in wasserstein distances for diffusion semigroups with negative curvature, Potential Anal. (2020) 1–22, Publisher: Springer.
[39] Liming Wu, Gradient estimates of Poisson equations on Riemannian manifolds and applications, J. Funct. Anal. 257 (12) (2009) 4015–4033.
[40] Liming Wu, Arnaud Guillin, Hacene Djellout, Transportation cost-information inequalities and applications to random dynamical systems and diffusions, Ann. Probab. 32 (3) (2004) 2702–2732.