Electronic Journal of Probability

Electron. J. Probab. **18** (2013), no. 67, 1–36.
ISSN: 1083-6489 DOI: 10.1214/EJP.v18-2586

Transport-Entropy inequalities and deviation estimates for stochastic approximations schemes

Max Fathi* Noufel Frikha†

Abstract

We obtain new transport-entropy inequalities and, as a by-product, new deviation estimates for the laws of two kinds of discrete stochastic approximation schemes. The first one refers to the law of an Euler like discretization scheme of a diffusion process at a fixed deterministic date and the second one concerns the law of a stochastic approximation algorithm at a given time-step. Our results notably improve and complete those obtained in [10]. The key point is to properly quantify the contribution of the diffusion term to the concentration regime. We also derive a general non-asymptotic deviation bound for the difference between a function of the trajectory of a continuous Euler scheme associated to a diffusion process and its mean. Finally, we obtain non-asymptotic bound for stochastic approximation with averaging of trajectories, in particular we prove that averaging a stochastic approximation algorithm with a slow decreasing step sequence gives rise to optimal concentration rate.

Keywords: deviation bounds; transportation-entropy inequalities; Euler scheme; stochastic approximation algorithms; stochastic approximation with averaging.
AMS MSC 2010: 60H35 ; 65C30 ; 65C05.
Submitted to EJP on January 31, 2013, final version accepted on June 19, 2013.

1 Introduction

In this work, we derive transport-entropy inequalities and, as a consequence, non-asymptotic deviation estimates for the laws at a given time step of two kinds of discrete-time and d-dimensional stochastic evolution scheme of the form

$$X_{n+1} = X_n + \gamma_{n+1} H(n, X_n, U_{n+1}),\ n \geq 0, X_0 = x \in \mathbb{R}^d, \tag{1.1}$$

where $(\gamma_n)_{n\geq 1}$ is a deterministic positive sequence of time steps, the $(U_i)_{i\in\mathbb{N}^*}$ are i.i.d. $\mathbb{R}^q$-valued random variables defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with law μ and the function $H : \mathbb{N}\times\mathbb{R}^d\times\mathbb{R}^q \to \mathbb{R}^d$ is a measurable function satisfying for all $x \in \mathbb{R}^d$, for all $n \in \mathbb{N}$, $H(n, x, .) \in \mathcal{L}^1(\mu)$, and $\mu(du)$-a.s., $H(n, ., u)$ is continuous. Here and below, we will also assume that μ satisfies a *Gaussian concentration property*, that is there exists

*LPMA, Université Pierre et Marie Curie, Paris, France.
E-mail: max.fathi@etu.upmc.fr
†LPMA, Université Paris Diderot, Paris, France.
E-mail: frikha@math.univ-paris-diderot.fr http://www.proba.jussieu.fr/pageperso/frikha/

$\beta > 0$ such that for every real-valued 1-Lipschitz function f defined on $\mathbb{R}^q$ and for all $\lambda \geq 0$:

$$\mathbb{E}[\exp(\lambda f(U_1))] \leq \exp(\lambda \mathbb{E}[f(U_1)] + \frac{\beta\lambda^2}{4}). \qquad (GC(\beta))$$

It is well known that $(GC(\beta))$ implies the following deviation bound

$$\mathbb{P}[f(U_1) - \mathbb{E}[f(U_1)] \geq r] \leq \exp(-\frac{r^2}{\beta}) \ \ \forall r \geq 0,$$

Examples of random variables satisfying this property include Gaussians, as well as bounded random variables. A characterization of $(GC(\beta))$ due to Djellout, Guillin and Wu [8] is given by Gaussian tail of U_1, that is there exists $\varepsilon > 0$ such that $\mathbb{E}[\exp(\varepsilon|U_1|^2)] < +\infty$, see also Bolley and Villani [6] for another proof with a simple link between the involved constants. The two claims are actually equivalent.

We are interested in furthering the discussion, initiated in [10], about giving non asymptotic deviation bounds for two specific problems related to evolution schemes of the form (1.1). The first one is the deviation between a function of an Euler like discretization scheme of a diffusion process at a fixed deterministic date and its mean. The second one refers to the deviation between a stochastic approximation algorithm at a given time-step and its target. Under some mild assumptions, in particular the assumption that the function $u \mapsto H(n, x, u)$ is lipschitz uniformly in space and time, it is proved in [10] that both recursive schemes share the Gaussian concentration property of the innovation.

In the present work, we point out the contribution of the diffusion term to the concentration rate which to our knowledge is new. This covers many situations and gives rise to different regimes ranging from exponential to Gaussian. We also derive a general non-asymptotic deviation bound for the difference between a function of the trajectory of a *continuous Euler scheme* associated to a diffusion process and its mean. It turns out that, under mild assumptions, the concentration regime is log-normal. Finally, we study non-asymptotic deviation bound for stochastic approximation with averaging of trajectories according to the *averaging principle of Ruppert & Polyak*, see e.g. [21] and [18].

1.1 Euler like Scheme of a Diffusion Process

We consider a Brownian diffusion process $(X_t)_{t\geq 0}$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, satisfying the usual conditions, and solution to the following stochastic differential equation (SDE)

$$X_t = x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \qquad (SDE_{b,\sigma})$$

where $(W_t)_{t\geq 0}$ is a q-dimensional $(\mathcal{F}_t)_{t\geq 0}$ Brownian motion and the coefficients b, σ are assumed to be uniformly Lipschitz continuous in space and measurable in time.

A basic problem in Numerical Probability is to compute quantities like $\mathbb{E}_x[f(X_T)]$ for a given Lipschitz continuous function f and a fixed deterministic time horizon T using Monte Carlo simulation. For instance, it appears in mathematical finance and represents the price of a European option with maturity T when the dynamics of the underlying asset is given by $(SDE_{b,\sigma})$. To this end, we first introduce some discretization schemes of $(SDE_{b,\sigma})$ that can be easily simulated. For a fixed time step $\Delta = T/N,\ N \in \mathbb{N}^*$, we set $t_i := i\Delta$, for all $i \in \mathbb{N}$ and define an Euler like scheme by

$$X_0^\Delta = x,\ \forall i \in [\![0, N-1]\!], X_{t_{i+1}}^\Delta = X_{t_i}^\Delta + b(t_i, X_{t_i}^\Delta)\Delta + \sigma(t_i, X_{t_i}^\Delta)\Delta^{1/2}U_{i+1}, \qquad (1.2)$$

where $(U_i)_{i\in\mathbb{N}^*}$ is a sequence of $\mathbb{R}^q$-valued i.i.d. random variables with law μ satisfying: $\mathbb{E}[U_1] = 0_q,\ \mathbb{E}[U_1U_1^*] = I_q$, where U_1^* denotes the transpose of the column vector U_1 and $0_q, I_q$ respectively denote the zero vector of $\mathbb{R}^q$ and the identity matrix of $\mathbb{R}^q \otimes \mathbb{R}^q$. We also assume that μ satisfies $(GC(\beta))$ for some $\beta > 0$. The main advantage of such a situation is that it includes the case of the standard Euler scheme where $U_1 \overset{\mathrm{d}}{=} \mathcal{N}(0, I_q)$ and the case of the Bernoulli law where $U_1 \overset{\mathrm{d}}{=} (B_1, \cdots, B_q),\ (B_k)_{k\in[\![1,q]\!]}$ are i.i.d random variables with law $\mu = \frac{1}{2}(\delta_{-1} + \delta_1)$, both satisfying $(GC(\beta))$ with $\beta = 2$.

The weak error $\mathcal{E}_D(f, \Delta, T, b, \sigma) = \mathbb{E}_x[f(X_T)] - \mathbb{E}_x[f(X_T^\Delta)]$ corresponds to the discretization error when replacing the diffusion X by its Euler scheme X^Δ for the computation of $\mathbb{E}_x[f(X_T)]$. Since the seminal work of [22], it is known that, under smoothness assumption on the coefficients $b,\ \sigma$, the *standard Euler scheme* produces a weak error of order Δ. In a hypoelliptic setting for the coefficients b and σ and for a bounded measurable function f, Bally and Talay [2] obtained the expected order using Malliavin calculus. Let us also mention the recent work [1] where the authors study the *weak trajectorial error* using coupling techniques. More precisely, they prove that the Wasserstein distance between the law of a uniformly elliptic and one-dimensional diffusion process and the law of its *continuous Euler scheme* $X^{c,\Delta}$ with time step $\Delta := T/N$ is smaller than $\mathcal{O}(N^{-2/3+\epsilon}),\ \forall\epsilon > 0$.

The expansion of $\mathcal{E}_D$ also allows to improve the convergence rate to 0 of the discretization error using Richardson-Romberg extrapolation techniques, see e.g. [22].

In order to have a global control of the numerical procedure for the computation of $\mathbb{E}_x[f(X_T)]$, it remains to approximate the expectation $\mathbb{E}_x[f(X_T^\Delta)]$ using a Monte Carlo estimator $M^{-1} \times \sum_{k=1}^M f((X_T^\Delta)^j)$ where the $((X_T^\Delta)^j)_{j\in[\![1,M]\!]}$ are M independent copies of the scheme (1.2) starting at the initial value x at time 0. This gives rise to an *empirical error* defined by $\mathcal{E}_{Emp}(M, f, \Delta, T, b, \sigma) = \mathbb{E}_x[f(X_T^\Delta)] - M^{-1} \times \sum_{j=1}^M f((X_T^\Delta)^j)$. Consequently, the global error associated to the computation of $\mathbb{E}_x[f(X_T)]$ writes as

$$\begin{aligned}\mathcal{E}_{Glob}(M, \Delta) &= \mathbb{E}_x[f(X_T)] - \mathbb{E}_x[f(X_T^\Delta)] + \mathbb{E}_x[f(X_T^\Delta)] - \frac{1}{M} \times \sum_{j=1}^M f((X_T^\Delta)^j)\\ &:= \mathcal{E}_D(f, \Delta, T, b, \sigma) + \mathcal{E}_{Emp}(M, f, \Delta, T, b, \sigma).\end{aligned}$$

It is well-known that if $f(X_T^\Delta)$ belongs to $L^2(\mathbb{P})$ the central limit theorem provides an *asymptotic* rate of convergence of order $M^{1/2}$. Moreover, if $f(X_T^\Delta) \in L^3(\mathbb{P})$, a non-asymptotic result is given by the Berry-Essen theorem. However, in practical implementation, one is interested in obtaining deviation bounds in probability for a fixed M and a given threshold $r > 0$, that is explicitly controlling $\mathbb{P}\left(|\mathcal{E}_{Emp}(M, f, \Delta, T, b, \sigma)| \geq r\right)$.

In this context, Malrieu and Talay [17] obtained Gaussian deviation bounds in an ergodic framework and for a constant diffusion coefficient. Using optimal transportation techniques, Blower and Bolley [4] obtained Gaussian concentration inequalities and transportation inequalities for the joint law of the first n positions of a stochastic processes with state space some Polish space. Concerning the *standard* Euler scheme, Menozzi and Lemaire [16] obtained two-sided Gaussian bounds up to a systematic bias under the assumptions that the diffusion coefficient is uniformly elliptic, $\sigma\sigma^*$ is Hölder-continuous, bounded and that b is bounded. Frikha and Menozzi [10], getting rid of the non-degeneracy assumption on σ, recently obtained Gaussian deviation bound under the mild smoothness condition that $b,\ \sigma$ are uniformly Lipschitz-continuous in space (uniformly in time) and that σ is bounded. It should be noted that it is the boundedness of σ that gives rise to the Gaussian concentration regime for the deviation of the *empirical error*.

In the current work, we get rid of the boundedness of σ and we only need the Gaussian concentration property of the innovation. We suppose that the coefficients satisfy the following smoothness and domination assumptions

(HS) The coefficients b, σ are uniformly Lipschitz continuous in space uniformly in time.

(HD$_\alpha$) There exists a $\mathcal{C}^2(\mathbb{R}^d, \mathbb{R}_+^*)$ function V satisfying $\exists C_V > 0, |\nabla V|^2 \leq C_V V,\ \eta := \frac{1}{2}\sup_{x\in\mathbb{R}^d} \left\|\nabla^2 V(x)\right\| < +\infty$ and $\exists \alpha \in (0,1]$, such that for all $x \in \mathbb{R}^d$,

$$\exists C_b > 0, \sup_{t\in[0,T]} |b(t,x)|^2 \leq C_b V(x), \ \ , \exists C_\sigma > 0, \sup_{t\in[0,T]} Tr(a(t,x)) \leq C_\sigma V^{1-\alpha}(x).$$

where $a = \sigma\sigma^*$.

The idea behind assumption **(HD$_\alpha$)** is to parameterize the growth of the diffusion coefficient in order to quantify its contribution to the concentration regime. Indeed, under **(HS)** and **(HD$_\alpha$)**, with $\alpha \in [1/2, 1]$, and if the innovations satisfy $(GC(\beta))$, for some positive β, we derive non-asymptotic deviation bounds for the empirical error $\mathcal{E}_{Emp}(M, f, \Delta, T, b, \sigma)$ ranging from exponential (if $\alpha = 1/2$) to Gaussian (if $\alpha = 1$) regimes. Therefore, we greatly improve the results obtained in [10].

Our approach here is different from [10]. Indeed, in [10], the key tool consists in writing the deviation using the same kind of decompositions that are exploited in [22] for the analysis of the discretization error. In the current work, we will use the fact that the Euler-like scheme (1.2) defines an inhomogenous Markov chain having Feller transitions P_k, $k = 0, \cdots, N-1$, defined for non negative or bounded Borel function $f : \mathbb{R}^d \to \mathbb{R}$ by

$$P_k(f)(x) = \mathbb{E}\left[f(X^\Delta_{t_{k+1}}) \Big| X^\Delta_{t_k} = x \right] = \mathbb{E}\left[f\left(x + b(t_k, x)\Delta + \sigma(t_k, x)\Delta^{1/2} U \right)\right].$$

For every $k,\ p \in \{0, \cdots, N-1\}$, $k \leq p$, we also define the iterative kernels $P_{k,p}$ by

$$P_{k,p}(f)(x) = P_k \circ \cdots \circ P_{p-1}(f)(x) = \mathbb{E}\left[f(X^\Delta_{t_p}) \Big| X^\Delta_{t_k} = x \right].$$

Now using that the law μ of the innovation satisfies $(GC(\beta))$ for some positive β, for every 1-Lipschitz function f and for all $\lambda \geq 0$, we obtain

$$\begin{aligned} P_{N-1}(\exp(\lambda f))(x) &= \mathbb{E}\left[\exp\left(\lambda f \left(x + b(t_{N-1}, x)\Delta + \sigma(t_{N-1}, x)\Delta^{1/2} U \right)\right)\right] \\ &\leq \exp\left(\lambda P_{N-1}(f)(x) + \beta \frac{\lambda^2}{4} \Delta |\sigma(t_{N-1}, x)|^2 \right) \end{aligned}$$

If σ is bounded, the Gaussian concentration property will readily follow provided the iterated kernel functions $P_{k,p}(f)$ are uniformly Lipschitz. Under the mild smoothness assumption **(HS)**, this can be easily derived, see Proposition 3.5. Otherwise, using **(HD$_\alpha$)**, we obtain

$$P_{N-1}(\exp(\lambda f))(x) \leq \exp\left(\lambda P_{N-1}(f)(x) + \frac{C_\sigma \beta \Delta}{4} \lambda^2 V^{1-\alpha}(x) \right). \tag{1.3}$$

The last inequality is the first step of our analysis. To investigate the empirical error, the key idea is to exploit recursively from (1.3) that the increments of the scheme (1.2) satisfy $(GC(\beta))$ and to adequately quantify the contribution of the diffusion term $V^{1-\alpha}(x)$ to the concentration rate. Under **(HS)** and **(HD$_\alpha$)**, the latter is addressed using flow techniques and integrability results on the law of the scheme (1.2), see Propositions 3.1 and 3.6.

1.2 Stochastic Approximation Algorithm

Beyond concentration bounds of the empirical error for Euler-like schemes, we want to look at non asymptotic bounds for stochastic approximation algorithms. Introduced by H. Robbins and S. Monro [19], these recursive algorithms aim at finding a zero of a continuous function $h : \mathbb{R}^d \to \mathbb{R}^d$ which is unknown to the experimenter but can only be estimated through experiments. Successfully and widely investigated since this seminal work, such procedures are now commonly used in various contexts such as convex optimization since minimizing a function amounts to finding a zero of its gradient.

To be more specific, the aim of such an algorithm is to find a solution θ^* to the equation $h(\theta) := \mathbb{E}[H(\theta, U)] = 0$, where $H : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$ is a Borel function and U is a given $\mathbb{R}^q$-valued random variable with law μ. The function h is generally not computable, at least at a reasonable cost. Actually, it is assumed that the computation of h is costly compared to the computation of H for any couple $(\theta, u) \in \mathbb{R}^d \times \mathbb{R}^q$ and to the simulation of the random variable U.

A stochastic approximation algorithm corresponds to the following simulation-based recursive scheme

$$\theta^\gamma_{n+1} = \theta^\gamma_n - \gamma_{n+1} H(\theta^\gamma_n, U_{n+1}),\ n \geq 0,\ \theta_0 \in \mathbb{R}^d, \tag{1.4}$$

where $(U_n)_{n\geq 1}$ is an i.i.d. $\mathbb{R}^q$-valued sequence of random variables with law μ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\gamma = (\gamma_n)_{n\geq 1}$ is a sequence of non-negative deterministic steps satisfying the usual assumption

$$\sum_{n\geq 1} \gamma_n = +\infty, \text{ and } \sum_{n\geq 1} \gamma_n^2 < +\infty. \tag{1.5}$$

When the function h is the gradient of a potential, the recursive procedure (1.4) is a stochastic gradient algorithm. Indeed, replacing $H(\theta^\gamma_n, U_{n+1})$ by $h(\theta^\gamma_n)$ in (1.4) leads to the usual deterministic descent gradient method. When $h(\theta) = M(\theta) - \ell$, $\theta \in \mathbb{R}$, where M is a monotone function, say increasing, we can write $M(\theta) = \mathbb{E}[N(\theta, U)]$ where $N : \mathbb{R} \times \mathbb{R}^q \to \mathbb{R}$ is a Borel function and ℓ is a given constant such that the equation $M(\theta) = \ell$ has a solution. Setting $H = N - \ell$, the recursive procedure (1.4) then corresponds to the seminal Robbins-Monro algorithm and aims at computing the level of the function M.

In the present paper, we make no attempt to provide a general discussion concerning convergence results of stochastic approximation algorithms. We refer readers to [9], [14] for some general results on the $a.s.$ convergence of such procedures under the existence of a so-called *Lyapunov function*, $i.e.$ a continuously differentiable function $L : \mathbb{R}^d \to \mathbb{R}_+$ such that ∇L is Lipschitz, $|\nabla L|^2 \leq C(1 + L)$ for some positive constant C and

$$\langle \nabla L, h\rangle \geq 0.$$

See also [15] for a convergence theorem under the existence of a *pathwise Lyapunov function*. For the sake of simplicity, in the sequel it is assumed that θ^* is the unique solution of the equation $h(\theta) = 0$ and that the sequence $(\theta^\gamma_n)_{n\geq 0}$ defined by (1.4) converges $a.s.$ towards θ^*.

We assume that the law μ of the innovation satisfies $(GC(\beta))$ for some $\beta > 0$ and that the step sequence $(\gamma_n)_{n\geq 1}$ satisfies (1.5). We also suppose that the following assumptions on the function H are in force:

(HL) For all $u \in \mathbb{R}^q$, the function $H(., u)$ is Lipschitz-continuous with a Lipschitz modulus having linear growth in the variable u, that is:

$$\exists C_H > 0,\ \forall u \in \mathbb{R}^q,\ \sup_{(\theta,\theta')\in(\mathbb{R}^d)^2} \frac{|H(\theta, u) - H(\theta', u)|}{|\theta - \theta'|} \leq C_H(1 + |u|).$$

(HLS)$_\alpha$ (*Lyapunov Stability-Domination*) There exists a $\mathcal{C}^2(\mathbb{R}^d, \mathbb{R}_+^*)$ function L satisfying $\exists C_L > 0, |\nabla L|^2 \leq C_L L,\ \eta := \frac{1}{2}\sup_{x\in\mathbb{R}^d} \left\|\nabla^2 L(x)\right\| < +\infty$ such that

$$\forall \theta \in \mathbb{R}^d,\ \ \langle \nabla L(\theta), h(\theta)\rangle \geq 0, \quad \text{and} \quad \exists C_h > 0,\ \forall \theta\in\mathbb{R}^d,\ |h(\theta)|^2 \leq C_h L(\theta).$$

and $\exists \alpha \in (0,1]$,

$$\exists C_\alpha > 0,\ \forall\theta\in\mathbb{R}^d,\ \sup_{(u,u')\in(\mathbb{R}^q)^2} \frac{|H(\theta,u)-H(\theta,u')|}{|u-u'|} \leq C_\alpha L^{\frac{1-\alpha}{2}}(\theta)$$

(HUA) (*Uniform Attractivity*) The map $h : \theta \in \mathbb{R}^d \mapsto \mathbb{E}[H(\theta,U)]$ is continuously differentiable in θ and there exists $\underline{\lambda} > 0$ s.t. $\forall\theta\in\mathbb{R}^d,\ \forall\xi\in\mathbb{R}^d,\ \underline{\lambda}|\xi|^2 \leq \langle Dh(\theta)\xi, \xi\rangle$.

Compared to [10], our assumptions are weaker. Indeed, it is assumed in [10] that the map $(\theta,u)\in\mathbb{R}^d\times\mathbb{R}^q \mapsto H(\theta,u)$ is uniformly Lipschitz continuous. In our current framework, this latter assumption is replaced by **(HL)** and **(HLS)**$_\alpha$.

The last assumption **(HUA)**, which already appeared in [10], is introduced to derive a sharp estimate of the concentration rate in terms of the step sequence. Let us note that such assumption appears in the study of the weak convergence rate order for the sequence $(\theta_n)_{n\geq 1}$ as described in [9] or [14]. Indeed, it is commonly assumed that the matrix $Dh(\theta^*)$ is *uniformly attractive* that is $\mathcal{R}e(\lambda_{min}) > 0$ where λ_{min} is the eigenvalue with the smallest real part. In our current framework, this local condition on the Jacobian matrix of h at the equilibrium is replaced by the uniform assumption **(HUA)**. This allows to derive sharp estimates for the concentration rate of the sequence $(\theta_n)_{n\geq1}$ around its target θ^* and to provide a sensitivity analysis for the bias $\delta_n := \mathbb{E}[|\theta_n - \theta^*|]$ with respect to the starting point θ_0.

Let us note that under **(HUA)** and the *linear growth assumption*

$$\forall\theta\in\mathbb{R}^d,\ \ \mathbb{E}\left[|H(\theta,U)|^2\right] \leq C(1+|\theta-\theta^*|^2),$$

which is satisfied if **(HL)** and **(HLS)**$_\alpha$, with $\alpha\in[0,1]$, hold and if μ satisfies $(GC(\beta))$ for some $\beta > 0$, the function $L : \theta \mapsto \frac{1}{2}\left|\theta-\theta^*\right|^2$ is a Lyapunov function for the recursive procedure defined by (1.4) so that one easily deduces that $\theta_n^\gamma \to \theta^*$, $a.s.$ as $n\to+\infty$.

The global error between the stochastic approximation procedure θ_n^γ at a given time step n and its target θ^* can be decomposed as *an empirical error* and *a bias* as follows

$$\begin{aligned} |\theta_n^\gamma - \theta^*| &= |\theta_n^\gamma-\theta^*| - \mathbb{E}_{\theta_0}[|\theta_n^\gamma-\theta^*|] + \mathbb{E}_{\theta_0}[|\theta_n^\gamma-\theta^*|] \\ &:= \mathcal{E}_{Emp}(\gamma,n,H,\underline{\lambda},\alpha) + \delta_n \end{aligned} \tag{1.6}$$

where we introduced the notations $\mathcal{E}_{Emp}(\gamma,n,H,\underline{\lambda},\alpha) = |\theta_n^\gamma-\theta^*| - \mathbb{E}_{\theta_0}[|\theta_n^\gamma-\theta^*|]$ and $\delta_n := \mathbb{E}_{\theta_0}[|\theta_n^\gamma-\theta^*|]$.

The *empirical error* $\mathcal{E}_{Emp}(\gamma,n,H,\underline{\lambda},\alpha)$ is the difference between the absolute value of the error at time n and its mean whereas *the bias* δ_n corresponds to the mean of the absolute value of the difference between the sequence $(\theta_n^\gamma)_{n\geq0}$ at time n and its target θ^*. Unlike the Euler like scheme, a bias systematically appears since we want to derive a deviation bound for the difference between θ_n^γ and its target θ^*. This term strongly depends on the choice of the step sequence $(\gamma_n)_{n\geq1}$ and the initial point θ_0, see Proposition 4.7 for a sensitivity analysis.

As for Euler like schemes, our strategy is different from [10]. Indeed, we exploit again the fact that the stochastic approximation scheme (1.4) defines an inhomogenous Markov chain having Feller transitions P_k, $k=0,\cdots,N-1$, defined for non negative or bounded Borel function $f : \mathbb{R}^d \to \mathbb{R}$ by

$$P_k(f)(\theta) = \mathbb{E}\left[f(\theta_{k+1}^\gamma)\middle|\,\theta_k^\gamma = \theta\right] = \mathbb{E}\left[f\left(\theta - \gamma_{k+1}H(\theta,U)\right)\right].$$

For every $k,\ p \in \{0, \cdots, N-1\}$, $k \leq p$, we also define the iterative kernels $P_{k,p}$ by

$$P_{k,p}(f)(\theta) = P_k \circ \cdots \circ P_{p-1}(f)(\theta) = \mathbb{E}\left[f(\theta_p^\gamma) \middle| \theta_k^\gamma = \theta \right].$$

For a 1-Lipschitz function f and for all $\lambda \geq 0$, using **(HLS)**$_\alpha$ and that the law μ of the innovation satisfies ($GC(\beta)$) for some positive β, we obtain

$$\begin{aligned} P_{N-1}(\exp(\lambda f))(\theta) &= \mathbb{E}\left[\exp\left(\lambda f\left(\theta - \gamma_N H(\theta, U)\right)\right)\right] \\ &\leq \exp\left(\lambda P_{N-1}(f)(\theta) + \beta \frac{\lambda^2}{4} C_\alpha^2 \gamma_N^2 L^{1-\alpha}(\theta)\right) \end{aligned} \tag{1.7}$$

Let us note the similarity between (1.3) and (1.7). If **(HLS)**$_\alpha$ holds with $\alpha = 1$ then the last term appearing in the right hand side of the last inequality is uniformly bounded in θ. This latter assumption corresponds to the framework developed in [10] and leads to a Gaussian concentration bound.

Otherwise, the problem is more challenging. Under the mild domination assumption **(HLS)**$_\alpha$, the key idea consists again in exploiting recursively from (1.7) that the increments of the stochastic approximation algorithm (1.4) satisfy ($GC(\beta)$) and in properly quantifying the contribution of the diffusion term $L^{1-\alpha}(\theta)$ to the concentration rate.

As already noticed in [10], the concentration rate and the bias strongly depends on the choice of the step sequence. In particular, if $\gamma_n = \frac{c}{n}$, with $c > 0$ then the optimal concentration rate and bias is achieved if $c > \frac{1}{2\underline{\lambda}}$, see Theorem 2.2. in [10]. Otherwise, they are sub-optimal. This kind of behavior is well-known concerning the weak convergence rate for stochastic approximation algorithm. Indeed, if $c > \frac{1}{2\mathcal{R}e(\lambda_{min})}$ we know that a Central Limit Theorem holds for the sequence $(\theta_n)_{n\geq 1}$ (see e.g. [9]). Let us note that the condition $c > \frac{1}{2\underline{\lambda}}$ as well as $c > \frac{1}{2\mathcal{R}e(\lambda_{min})}$ is difficult to handle and may lead to a blind choice in practical implementation.

To circumvent such a difficulty, it is fairly well-known that the key idea is to carefully smooth the trajectories of a converging stochastic approximation algorithm by averaging according to the *Ruppert & Polyak averaging principle*, see e.g. [21] and [18]. It consists in devising the original stochastic approximation algorithm (1.4) with a slow decreasing step $\gamma = (\gamma_n)_{n\geq 1}$, namely

$$\gamma_n = \left(\frac{c}{b+n}\right)^\nu, \ \ \nu \in \left(\frac{1}{2}, 1\right), c, b > 0,$$

and to simultaneously compute the empirical mean $(\bar{\theta}_n^\gamma)_{n\geq 1}$ of the sequence $(\theta_n^\gamma)_{n\geq 0}$ by setting

$$\bar{\theta}_n^\gamma = \frac{\theta_0 + \theta_1^\gamma + \cdots + \theta_{n-1}^\gamma}{n} = \bar{\theta}_{n-1}^\gamma - \frac{1}{n}\left(\bar{\theta}_{n-1}^\gamma - \theta_{n-1}^\gamma\right). \tag{1.8}$$

We will not enter into the technicalities of the subject but under mild assumptions (see e.g. [9], p.169) one shows that

$$\sqrt{n}(\bar{\theta}_n^\gamma - \theta^*) \overset{\mathcal{L}}{\to} \mathcal{N}(0, \Sigma^*),\ n \to +\infty,$$

where Σ^* is the optimal covariance matrix. For instance, for $d = 1$, one has $\Sigma^* = \frac{Var(H(\theta^*, U))}{(h'(\theta^*))^2}$. Hence, the optimal weak rate of convergence $\sqrt{n}$ is achieved for free without any condition on the constants c or b. However, this result is only asymptotic and so far, to our best knowledge, non-asymptotic estimates for the deviation between the empirical mean sequence $(\bar{\theta}_n^\gamma)_{n\geq 0}$ at given time step and its target θ^*, that is non-asymptotic averaging principle were not investigated.

The sequence $(z_n^\gamma)_{n\geq 0}$ defined by $z_n^\gamma := (\bar{\theta}_{n+1}^\gamma, \theta_n^\gamma)$ is $\mathcal{F}$-adapted, i.e. for all $n \geq 0$, z_n^γ is $\mathcal{F}_n$-measurable, where $\mathcal{F}_n := \sigma(\theta_0, U_k, k \leq n)$. Moreover, it defines an inhomogenous Markov chain having Feller transitions K_k, $k = 0, \cdots, N-1$, defined for non negative or bounded Borel function $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ by

$$\begin{aligned} K_k(f)(z) &= \mathbb{E}[\, f(z_{k+1}^\gamma) \big|\, z_k^\gamma = z] = \mathbb{E}[\, f(\bar{\theta}_{k+2}^\gamma, \theta_{k+1}^\gamma) \big|\, (\bar{\theta}_{k+1}^\gamma, \theta_k^\gamma) = (z_1, z_2)], \\ &= \mathbb{E}\left[f\left(\frac{k+1}{k+2} z_1 + \frac{1}{k+2}(z_2 - \gamma_{k+1} H(z_2, U)), z_2 - \gamma_{k+1} H(z_2, U) \right) \right]. \end{aligned}$$

For every $k, p \in \{0, \cdots, N-1\}$, $k \leq p$, we also define the iterative kernels $K_{k,p}$ by

$$K_{k,p}(f)(z) = K_k \circ \cdots K_{p-1}(f)(z) = \mathbb{E}[\, f(z_p^\gamma) \big|\, z_k^\gamma = z].$$

Hence, for any 1-Lipschitz function and for all $\lambda \geq 0$, using again **(HLS)**$_\alpha$ and that the law μ of the innovation satisfies $(GC(\beta))$ for some positive β, one has for all $k \in \{0, \cdots, N-1\}$

$$\begin{aligned} K_k(\exp(\lambda f))(z) &= \mathbb{E}\left[\exp\left(\lambda f\left(z_{k+1}^\gamma \right) \right) \big|\, z_k^\gamma = z \right] \\ &\leq \exp\left(\lambda K_k(f)(z) + \beta \frac{\lambda^2}{4} \left(C_\alpha \gamma_{k+1} \left(\frac{1}{k+2} + 1\right) L^{\frac{1-\alpha}{2}}(z_2) \right)^2 \right) \\ &\leq \exp\left(\lambda K_k(f)(z) + \beta \lambda^2 C_\alpha^2 \gamma_{k+1}^2 L^{1-\alpha}(z_2) \right) \end{aligned} \tag{1.9}$$

where we used that the functions $u \mapsto f\left(\frac{k+1}{k+2} z_1 + \frac{1}{k+2}(z_2 - \gamma_{k+1} H(z_2, u)), z_2 - \gamma_{k+1} H(z_2, u) \right)$ are Lipschitz-continuous with Lipschitz modulus equals to $C_\alpha \gamma_{k+1}(\frac{1}{k+2} + 1) L^{\frac{1-\alpha}{2}}(z_2)$ for all $(z_1, z_2) \in \mathbb{R}^d \times \mathbb{R}^d$.

Here again, (1.7) and (1.9) are quite similar and if $\alpha = 1$ the concentration regime turns out to be Gaussian. Otherwise, an analysis along the lines of the methodology developed so far provides the concentration regime of the stochastic approximation algorithm with averaging of trajectories.

1.3 Transport-Entropy inequalities

As a by-product of our analysis, we derive transport-entropy inequalities for the law of both stochastic approximation schemes. We recall here basic definitions and properties. For a complete overview and recent developments in the theory of transport inequalities, the reader may refer to the recent survey [12]. We will denote by $\mathcal{P}(\mathbb{R}^d)$ the set of probability measures on $\mathbb{R}^d$.

For $p \geq 1$, we consider the set $\mathcal{P}_p(\mathbb{R}^d)$ of probability measures with finite moment of order p. The Wasserstein metric $W_p(\mu, \nu)$ of order p between two probability measures $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ is defined by

$$W_p^p(\mu, \nu) = \inf\left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \pi(dx, dy) : \ \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d), \ \pi_0 = \mu, \ \pi_1 = \nu \right\}$$

where π_0 and π_1 are two probability measures standing for the first and second marginals of $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. For $\mu \in \mathcal{P}(\mathbb{R}^d)$, we define the relative entropy w.r.t $\nu \in \mathcal{P}(\mathbb{R}^d)$ as

$$H(\mu, \nu) = \int_{\mathbb{R}^d} \log\left(\frac{d\mu}{d\nu} \right) d\mu$$

if $\mu \ll \nu$ and $H(\mu, \nu) = +\infty$ otherwise. We are now in position to define the notion of transport-entropy inequality. Here as below, $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ is a convex, increasing function with $\Phi(0) = 0$.

Definition 1.1. *A probability measure μ on $\mathbb{R}^d$ satisfies a transport-entropy inequality with function Φ if for all $\nu \in \mathcal{P}(\mathbb{R}^d)$, one has*

$$\Phi(W_1(\nu,\mu)) \leq H(\nu,\mu)$$

For the sake of simplicity, we will write that μ satisfies T_Φ.

The following proposition comes from Corollary 3.4. of [12].

Proposition 1.2. *The following propositions are equivalent:*

- *The probability measure μ satisfies T_Φ.*
- *For all 1-Lipschitz function f, one has*

$$\forall \lambda \geq 0, \quad \int \exp(\lambda f) d\mu \leq \exp\left(\lambda \int f d\mu + \Phi^*(\lambda)\right),$$

where Φ^* is the monotone conjugate of Φ defined on $\mathbb{R}_+$ as $\Phi^*(\lambda) = \sup_{\rho \geq 0} \{\lambda\rho - \Phi(\rho)\}$.

Such transport-entropy inequalities are very attractive especially from a numerical point of view since they are related to the concentration of measure phenomenon which allows to establish non-asymptotic deviation estimates. The three next results put an emphasis on this point. Suppose that $(X_n)_{n\geq 1}$ is a sequence of i.i.d. $\mathbb{R}^d$-valued random variables with common law μ.

Corollary 1.3. *If μ satisfies T_Φ then for all 1-Lipschitz function f and for all $r \geq 0$, for all $M \geq 1$, one has*

$$\mathbb{P}\left(|\frac{1}{M}\sum_{k=1}^{M} f(X_k) - \mathbb{E}[f(X_1)]| \geq r\right) \leq 2\exp(-M\Phi(r))$$

Deriving non-asymptotic deviation bounds for $W_1(\mu_M, \mu)$ is of interest for many applications in the fields of numerical probability and statistic. In its present form, next result is due to Gozlan and Leonard [11], Theorem 12.

Proposition 1.4. *If μ satisfies T_Φ then the empirical measure μ_M defined as $\mu_M = \frac{1}{M}\sum_{k=1}^{M} \delta_{X_k}$ satisfies the following concentration bound*

$$\mathbb{P}\left(W_1(\mu_M,\mu) \geq \mathbb{E}[W_1(\mu_M,\mu)] + r\right) \leq \exp\left(-M\Phi(r)\right).$$

where for $x \in \mathbb{R}^d$, δ_x stands for the Dirac mass at point x.

The quantity $\mathbb{E}[W_1(\mu_M,\mu)]$ will go to zero as M goes to infinity, by convergence of empirical measures, but we still need quantitative bounds. The next result is an adaptation of Theorem 10.2.1 in [20] on similar bounds but for the distance W_2. For sake of completeness, we provide a proof in Appendix A.

Proposition 1.5. *Assume that μ has a finite moment of order $d+3$. Then, one has*

$$\mathbb{E}[W_1(\mu_M,\mu)] \leq C(d,\mu) M^{-1/(d+2)}$$

where

$$C(d,\mu) := 4\sqrt{d} + 2\sqrt{\int_{\mathbb{R}^d} (1+|x|^{d+1})^{-1} dx}\sqrt{2^{-2d} + 2^{3-d}\int |y|^{d+3}\mu(dy) + 2^{3-d} d(d+3)!}.$$

This bound is not optimal in general, but has the advantage of having very explicit constants. In the case of a distribution with compact support, it has been shown in [3], Section 7, that $\mathbb{E}[W_1(\mu_M,\mu)]$ is of order $O(M^{-1/d})$, and that this is the optimal exponent in d when $d \geq 3$.
In view of Kantorovich-Rubinstein duality formula, namely

$$W_1(\mu,\nu) = \sup\left\{\int f d\mu - \int f d\nu : [f]_1 \leq 1\right\}$$

where $[f]_1$ denotes the Lipschitz-modulus of f, the latter result provides the following concentration bounds $\forall r \geq 0,\ \forall M \geq 1$

$$\mathbb{P}\left(\sup_{f:[f]_1\leq 1}\left(\frac{1}{M}\sum_{k=1}^{M} f(X_k) - \mathbb{E}[f(X_1)]\right) \geq C(d,\mu)M^{-1/(d+2)} + r\right) \leq \exp\left(-M\Phi(r)\right).$$

Similar results were first obtained for different concentration regimes by Bolley, Guillin, Villani [7] relying on a non-asymptotic version of Sanov's Theorem. Some of these results have also been derived by Boissard [5] using concentration inequalities, and were also extended to ergodic Markov chains up to some contractivity assumptions in the Wasserstein metric on the transition kernel.

Some applications are proposed in [7]. Such results can indeed provide non-asymptotic deviation bounds for the estimation of the density of the invariant measure of a Markov chain. Let us note that the (possibly large) constant $C(d,\mu)$ appears as a trade-off to obtain uniform deviations over all Lipschitz functions.

As a consequence of the transport-entropy inequalities obtained for the laws at a given time step of Euler like schemes and stochastic approximation algorithm, we will derive non-asymptotic deviation bounds in the Wasserstein metric.

2 Main Results

2.1 Euler like schemes and diffusions

Theorem 2.1 (Transport-Entropy inequalities for Euler like schemes). *Denote by X_T^Δ the value at time T of the scheme* (1.2) *associated to the diffusion* ($SDE_{b,\sigma}$) *starting from x at time 0. Denote the Lipschitz modulus of b and σ appearing in the diffusion process* ($SDE_{b,\sigma}$) *by $[b]_1$ and $[\sigma]_1$, respectively and by μ_T^Δ the law of X_T^Δ. Assume that the innovations $(U_i)_{i\geq 1}$ in* (1.2) *satisfy* ($GC(\beta)$) *for some $\beta > 0$ and that the coefficients b,σ satisfy* **(HS)** *and* **(HD$_\alpha$)** *for $\alpha \in [\frac{1}{2}, 1]$.*

Then, μ_T^Δ satisfies $T_{\Phi_\alpha^*}$ with $\Phi_\alpha^*(\lambda) = \sup_{\rho\geq 0}\{\lambda\rho - \Phi_\alpha(\rho)\}$ and one has:

- *If $\alpha \in (\frac{1}{2}, 1]$, for all $\rho \geq 0$*

$$\Phi_\alpha(\rho) = \Psi_\alpha(T,\Delta,b,\sigma,x)(\rho^2 \vee \rho^{\frac{2\alpha}{2\alpha-1}}),$$

- *If $\alpha = \frac{1}{2}$, for all $\rho \in [0, \varphi(T,b,\sigma,\Delta)^{-1/2}\lambda_{3.2})$*

$$\Phi_{1/2}(\rho) = K_{3.2}\frac{(\rho\varphi(T,b,\sigma,\Delta)^{1/2}/\lambda_{3.2})^2}{1-(\rho\varphi(T,b,\sigma,\Delta)^{1/2}/\lambda_{3.2})}.$$

Moreover, we have $\Psi_\alpha(T,\Delta,b,\sigma,x) = K_{3.1}(\varphi(T,b,\sigma,\Delta)^2 \vee \varphi(T,b,\sigma,\Delta)^{\frac{\alpha}{2\alpha-1}})$, $\varphi(T,b,\sigma,\Delta) = C_\sigma\beta\frac{(1+C(\Delta)\Delta)}{4C(\Delta)}e^{3C(\Delta)T}$, $C(\Delta) := 2[b]_1 + [\sigma]_1^2 + \Delta[b]_1^2$, the constants $K_{3.1}$, $\lambda_{3.2}$ and $K_{3.2}$ being defined in Corollaries 3.2 and 3.4 respectively.

Note that in the above theorem, we do not need any non-degeneracy condition on the diffusion coefficient.

In the case $\alpha \in (\frac{1}{2}, 1]$, one easily gets the following explicit formula:

- If $\lambda \in [0, 2\Psi]$, then $\Phi_\alpha^*(\lambda) = \frac{1}{4\Psi}\lambda^2$;
- If $\lambda \in [\frac{2\alpha}{2\alpha-1}\Psi, +\infty)$, then $\Phi_\alpha^*(\lambda) = \frac{1}{2\alpha}\left(\frac{2\alpha-1}{2\alpha\Psi}\right)^{2\alpha-1}\lambda^{2\alpha}$;
- If $\lambda \in (2\Psi, \frac{2\alpha}{2\alpha-1}\Psi)$,then $\Phi_\alpha^*(\lambda) = \lambda - \Psi$.

Let us note that the linear behavior of Φ_α^* on a small interval is due to the fact that Φ_α is not $\mathcal{C}^1$. One may want to replace $\rho^2 \vee \rho^{\frac{2\alpha}{2\alpha-1}}$ by $\rho^2 + \rho^{\frac{2\alpha}{2\alpha-1}}$ (up to a factor 2) in the expression of Φ_α. However, in this case, an explicit expression for Φ_α^* does not exist (except for the case $\alpha = 1$) and only its asymptotic behavior can be derived so that one is led to compute it numerically in practical situations.

In the case $\alpha = 1/2$, tedious but simple computations show that

$$\Phi_{1/2}^*(\lambda) = \left(\left(1 + \frac{\lambda_{3.2}}{K_{3.2}\varphi(T, b, \sigma, \Delta)^{1/2}}\lambda\right)^{\frac{1}{2}} - 1\right)^2.$$

This behavior corresponds to a concentration profile that is Gaussian at short distance, and exponential at large distance.

Remark 2.2. *The order of magnitude of our bounds is actually optimal in α under our general assumptions. For example, if we consider the diffusion process $dX_t = (1 + X_t^2)^{(1-\alpha)/2}dB_t$, then the process $Y_t = V(X_t)$, with $V(x) := \int_0^x (1 + s^2)^{(\alpha-1)/2}ds$, satisfies the SDE $dY_t = dB_t + b(Y_t)dt$, where b is a bounded drift. This process therefore has the same concentration properties as a Brownian motion, which are known to be Gaussian. From this, we deduce*

$$\mathbb{P}_x(X_t \geq r) = \mathbb{P}_x(Y_t \geq V(r)) \leq \exp(-cV(r)^2).$$

This is indeed the order of magnitude of the concentration bounds given by Theorem 2.1.

Corollary 2.3. *(Non-asymptotic deviation bounds) Under the same assumptions as Theorem 2.1, one has:*

- *for all real-valued 1-Lipschitz function f defined on $\mathbb{R}^d$, for all $\alpha \in [1/2, 1]$ for all $M \geq 1$ and all $r \geq 0$,*

$$\mathbb{P}_x\left(|\frac{1}{M}\sum_{k=1}^{M} f((X_T^\Delta)^k) - \mathbb{E}_x[f(X_T^\Delta)]| \geq r\right) \leq 2\exp(-M\Phi_\alpha^*(r)),$$

- *for all $\alpha \in [1/2, 1]$, for all $M \geq 1$ and all $r \geq 0$,*

$$\mathbb{P}_x\left(\sup_{f:[f]_1 \leq 1}\left(\frac{1}{M}\sum_{k=1}^{M} f((X_T^\Delta)^k) - \mathbb{E}_x[f(X_T^\Delta)]\right) \geq \frac{C(d, \mu_T^\Delta)}{M^{1/(d+2)}} + r\right) \leq \exp\left(-M\Phi_\alpha^*(r)\right),$$

where the $((X_T^\Delta)^k)_{1\leq k\leq M}$ are M independent copies of the scheme (1.2).

The constant $C(d, \mu_T^\Delta)$ depends on the moment of order $d+3$ of μ_T^Δ. Hence, an explicit control in terms of x, b, σ, Δ can be easily obtained under our general assumptions. We leave the computational details to the reader.

Remark 2.4 (Extension to smooth functions of a finite number of time step)**.** *The previous transport-inequalities and non-asymptotic bounds could be extended to smooth functions of a finite number of time step such as the maximum of a scalar Euler like scheme. In that case, it suffices to introduce the additional state variable $(M_{t_i}^\Delta)_{i\geq 1} := (\max_{k\in[\![0,i]\!]} X_{t_k}^\Delta)_{i\geq 1}$. Now, the couple $(X_{t_i}^\Delta, M_{t_i}^\Delta)_{1\leq i\leq N}$ is Markovian and similar arguments could be easily extended to the couple for Lipschitz functions of both variables.*

Remark 2.5 (Transport-Entropy inequalities for the law of a diffusion process)**.** *The previous transport-inequalities and non-asymptotic bounds could be extended to the law at time T of the diffusion process solution to $(SDE_{b,\sigma})$ by passing to the limit $\Delta \to 0$. Indeed, it is well-known that under* **(HS)***, one has $X_T^{\Delta} \overset{a.s.}{\longrightarrow} X_T$, as $\Delta \to 0$ and by Lebesgue theorem, one deduces from the first result of Corollary 2.3 that the empirical error (empirical mean) of X_T itself satisfies a non-asymptotic deviation bound with a similar deviation function (just pass to the limit $\Delta \to 0$ in all constants). Then, using Corollary 5.1 in [12] (equivalence between deviation of the empirical mean and transport-entropy inequalities), one easily derives that the law of X_T satisfies a similar transport-entropy inequalities when $\alpha \in (1/2, 1]$.*

We want to point out that it is the growth of σ that gives the concentration regime ranging from Gaussian concentration bound if $\alpha = 1$ to exponential when $\alpha = \frac{1}{2}$. However, in many popular models in finance, the diffusion coefficient is linear, for instance practitioners often have to deal with Black-Scholes like dynamics of the form

$$X_t = x_0 + \int_0^t b(X_s)X_s ds + \int_0^t \sigma(X_s)X_s dW_s$$

for smooth, bounded coefficients b, σ. This corresponds to assumption **(HD$_\alpha$)** where $\alpha = 0$ and $V(x) = 1 + |x|^2$, $x \in \mathbb{R}^d$. For the estimation of $\mathbb{E}_x[f(X_T^{\Delta})]$ for a Lipschitz function $f : \mathbb{R}^d \to \mathbb{R}$, or even in more general situations, the estimation of $\mathbb{E}_x[f(X^{\Delta})]$ for a Lipschitz function $f : \mathcal{C} \to \mathbb{R}$, where $\mathcal{C} := \mathcal{C}([0, T], \mathbb{R}^d)$ stands for the space of $\mathbb{R}^d$-valued continuous functions on $[0, T]$, equipped with the uniform norm $||f||_\infty := \sup_{0\le t\le T} |f(t)|$, the expected concentration is the log-normal one. To deal with the latter case, we consider the continuous Euler scheme $X^{c,\Delta}$ associated to $(SDE_{b,\sigma})$ and writing

$$\forall t \in [0, T],\ X_t^{c,\Delta} = x + \int_0^t b(\phi(s), X_{\phi(s)}^{c,\Delta})ds + \int_0^t \sigma(\phi(s), X_{\phi(s)}^{c,\Delta})dW_s, \quad x \in \mathbb{R}^d. \tag{2.1}$$

where we set $\phi(t) := t_i$ for $t_i \le t < t_{i+1}$, $i \in \mathbb{N}$. The next result provides a general non-asymptotic deviation bound for the empirical error under very mild assumptions.

Theorem 2.6 (General non-asymptotic deviation bounds)**.** *Denote by $X^{c,\Delta} := (X_t^{c,\Delta})_{0\le t\le T}$ the path of the scheme (2.1) with step Δ starting from point x at time 0. Assume that $\forall t \in [0, T]$, the coefficients $b(t, .)$ and $\sigma(t, .)$ are continuous functions in x and that they satisfy the linear growth assumption:*

$$\forall x \in \mathbb{R}^d, \quad \sup_{t\in[0,T]} |b(t, x)| \le C_b(1 + |x|), \quad \sup_{t\in[0,T]} Tr(a(t, x)) \le C_\sigma(1 + |x|^2).$$

Then, for all 1-Lipschitz function $f : \mathcal{C} \to \mathbb{R}$, for all $M \in \mathbb{N}^*$, for all $r \ge 0$, one has

$$\mathbb{P}_x\left(|\frac{1}{M}\sum_{k=1}^M f((X^{c,\Delta})^k) - \mathbb{E}_x[f(X^{c,\Delta})]| \ge r \right) \le \begin{cases} 2e^{-\frac{r^2 M}{(2(1+|x|))^2 \exp(2\kappa(b,\sigma,T))}}, \text{ if } \frac{r\sqrt{M}}{2(1+|x|)} \le e^{\kappa(b,\sigma,T)} \\ 2e^{-\frac{1}{4\kappa(b,\sigma,T)}\log\left(\frac{r^2 M}{(2(1+|x|))^2}\right)^2}, \text{ otherwise} \end{cases}$$

where $\kappa(b, \sigma, T) := 28(1 + (C_\sigma \vee C_b)T)$ and $((X^{c,\Delta})^k)_{1\le k\le M}$ are M independent copies of the scheme (2.1)*. The result remains valid when one considers the path of the diffusion X solution to $(SDE_{b,\sigma})$ instead of the continuous Euler scheme.*

Remark 2.7. *We want to point out that though the constants appearing in the above non-asymptotic deviation bound are all-purpose and rough estimates, the decay in r is optimal. Indeed, if we select $b(t, x) = 0$, $\sigma(t, x) = \sigma x$, $\sigma > 0$, so that $X_t = x_0 \exp(\sigma W_t - \sigma^2 t/2)$, $M = 1$ and $f = \Pi_T$, where Π_T denotes the projection at time T, sharp bounds can be easily derived and it is plain to see that in this simple example the concentration regime for large values of r is the log-normal one and gaussian for small values of r.*

2.2 Stochastic approximation algorithms

Theorem 2.8 (Transport-Entropy inequalities for stochastic approximation algorithms)**.** *Let $N \in \mathbb{N}^*$. Assume that the function H of the recursive procedure $(\theta_n^\gamma)_{0\le n\le N}$ (with starting point $\theta_0 \in \mathbb{R}^d$) defined by* (1.4) *satisfies* **(HL)***,* **(HUA)** *and* **(HLS)**$_\alpha$ *for $\alpha \in [\frac{1}{2}, 1]$, and that the step sequence $\gamma = (\gamma_n)_{n\ge 0}$ satisfies* (1.5)*. Suppose that the law of the innovation satisfies $(GC(\beta))$, $\beta > 0$. Denote by μ_N^γ the law of θ_N.*

Then, μ_N^γ satisfies $T_{\Phi_\alpha^*}$ with $\Phi_{\alpha,N}^*(\lambda) = \sup_{\rho\ge 0}\{\lambda\rho - \Phi_{\alpha,N}(\rho)\}$ and one has:

- *If $\alpha \in (\frac{1}{2}, 1]$, for all $\rho \ge 0$*

$$\Phi_{\alpha,N}(\rho) = \varphi_\alpha(\gamma, H, \theta_0)(C_N^\gamma \rho^2 \vee C_N^{\gamma,\alpha} \rho^{\frac{2\alpha}{2\alpha-1}}).$$

- *If $\alpha = \frac{1}{2}$, for all $\rho \in [0, \lambda_{4.1}/\tilde{s}_N)$,*

$$\Phi_{1/2,N}(\rho) = 2\varphi_{1/2}(\gamma, H, \theta_0) C_N^\gamma \frac{(\rho/\lambda_{4.1})^2}{1 - (\rho\tilde{s}_N/\lambda_{4.1})}.$$

Moreover the three concentration rate sequences are defined for $N \in \mathbb{N}^*$ by

$$C_N^\gamma := \sum_{k=0}^{N-1} \gamma_{k+1}^2 \frac{\Pi_{1,N}}{\Pi_{1,k}},$$

$$C_N^{\gamma,\alpha} := \sum_{k=0}^{N-1} \gamma_{k+1}^{\frac{2\alpha}{2\alpha-1}} (\frac{\Pi_{1,N}}{\Pi_{1,k}})^{\frac{2\alpha}{2\alpha-1}} ((k+1)\log^2(k+4))^{\frac{1-\alpha}{2\alpha-1}}$$

$$\tilde{s}_N := \max_{0\le k\le N-1} (k+1)^{1/2} \log(k+4)\gamma_{k+1} \left(\frac{\Pi_{1,N}}{\Pi_{1,k}}\right)^{\frac{1}{2}} \exp(\sum_{p=0}^{N-1} \frac{1}{(p+1)\log^2(p+4)})$$

with $\Pi_{1,N} := \prod_{k=0}^{N-1}(1 - 2\underline{\lambda}\gamma_{k+1} + C_{H,\mu}\gamma_{k+1}^2)$, the constants $C_{H,\mu}$ and $\varphi_\alpha(\gamma, H, \theta_0)$ being explicitly given in Propositions 4.4 and 4.5 respectively.

As in the case of Euler like schemes, for $\alpha \in (\frac{1}{2}, 1]$, we have:

- if $\lambda \in [0, 2\varphi(C_N^\gamma/(C_N^{\gamma,\alpha})^{2\alpha-1})^{\frac{1}{2(1-\alpha)}}]$, then $\Phi_{\alpha,N}^*(\lambda) = \lambda^2/(4\varphi C_N^\gamma)$;
- If $\lambda \in [\frac{2\alpha}{2\alpha-1}\varphi(C_N^\gamma/(C_N^{\gamma,\alpha})^{2\alpha-1})^{\frac{1}{2(1-\alpha)}}, +\infty)$, then $\Phi_{\alpha,N}^*(\lambda) = \frac{1}{2\alpha}\left(\frac{2\alpha-1}{2\alpha\varphi}\right)^{2\alpha-1} \frac{\lambda^{2\alpha}}{(C_N^{\gamma,\alpha})^{2\alpha-1}}$;
- If $\lambda \in (2\varphi(C_N^\gamma/(C_N^{\gamma,\alpha})^{2\alpha-1})^{\frac{1}{2(1-\alpha)}}, \frac{2\alpha}{2\alpha-1}\varphi(C_N^\gamma/(C_N^{\gamma,\alpha})^{2\alpha-1})^{\frac{1}{2(1-\alpha)}})$, then $\Phi_{\alpha,N}^*(\lambda) = (\frac{C_N^\gamma}{C_N^{\gamma,\alpha}})^{\frac{2\alpha-1}{2(1-\alpha)}}\lambda - \varphi\frac{(C_N^\gamma)^{\frac{\alpha}{1-\alpha}}}{(C_N^{\gamma,\alpha})^{\frac{2\alpha-1}{1-\alpha}}}$.

For $\alpha = \frac{1}{2}$, we obtain the following explicit bound for the Legendre transform of $\Phi_{1/2,N}$

$$\forall\lambda \ge 0, \quad \Phi_{1/2,N}^*(\lambda) = \frac{2\varphi C_N^\gamma}{\tilde{s}_N^2}\left(\left(1 + \frac{\tilde{s}_N\lambda_{4.1}\lambda}{2\varphi C_N^\gamma}\right)^{\frac{1}{2}} - 1\right)^2$$

Hence, for $N \ge 1$ being fixed, the following simple asymptotic behaviors can be easily derived:

- When λ is small, $\Phi_{1/2,N}^*(\lambda) \sim \lambda_{4.1}^2\lambda^2/(2\varphi C_N^\gamma)$;
- When λ goes to infinity, $\Phi_{1/2}^*(\lambda) \sim \lambda_{4.1}\lambda/\tilde{s}_N$.

Corollary 2.9. *(Non-asymptotic deviation bounds) Under the same assumptions as Theorem 2.8, one has*

$$\mathbb{P}_{\theta_0}\left(|\theta_N^\gamma - \theta^*| \geq r + \delta_N\right) \leq \exp\left(-\Phi^*_{\alpha,N}(r)\right)$$

and $\delta_N := \mathbb{E}_{\theta_0}\left[|\theta_N^\gamma - \theta^*|\right]$*. Moreover, the bias* δ_N *at step* N *satisfies*

$$\delta_N \leq e^{-\underline{\lambda}\Gamma_{1,N}+C_{\alpha,\mu}\Gamma_{2,N}}\,|\theta_0-\theta^*|+(2C_{\alpha,\mu})^{\frac{1}{2}}\left(\sum_{k=0}^{N-1}\gamma_{k+1}^2 e^{-2\underline{\lambda}(\Gamma_{1,N}-\Gamma_{1,k+1})+2C_{\alpha,\mu}(\Gamma_{2,N}-\Gamma_{2,k+1})}\right)^{\frac{1}{2}},$$

where $\Gamma_{1,N} := \sum_{k=1}^N \gamma_k$*,* $\Gamma_{2,N} := \sum_{k=1}^N \gamma_k^2$*,* $C_{\alpha,\mu} := \underline{\lambda}^2/2 + 2C_\alpha K\mathbb{E}[|U|^2]$ *with* $K > 0$*.*

Now, we investigate the impact of the step sequence $(\gamma_n)_{n\geq 1}$ on the concentration rate sequences C_N^γ, $C_N^{\gamma,\alpha}$, $\tilde{s}_N$ and the bias δ_N. Let us note that a similar analysis has been performed in [10]. We obtain the following results:

- If we choose $\gamma_n = \frac{c}{n}$, with $c > 0$. Then $\delta_N \to 0$, $N \to +\infty$, $\Gamma_{1,N} = c\log(N) + c_1' + r_N$, $c_1' > 0$ and $r_N \to 0$, so that $\Pi_{1,N} = \mathcal{O}(N^{-2c\underline{\lambda}})$.
  - If $c < \frac{1}{2\underline{\lambda}}$, the series $\sum_{k=1}^N \gamma_k^2/\Pi_{1,k}$, $\sum_{k=0}^{N-1}\gamma_{k+1}^{\frac{2\alpha}{2\alpha-1}}(1/\Pi_{1,k}^{\frac{2\alpha}{2\alpha-1}})((k+1)\log^2(k+4))^{\frac{1-\alpha}{2\alpha-1}}$ converge so that we obtain $C_N^\gamma = \mathcal{O}(N^{-2c\underline{\lambda}})$, $C_N^{\gamma,\alpha} = \mathcal{O}(N^{-\frac{2\alpha}{2\alpha-1}c\underline{\lambda}})$, $\tilde{s}_N = \mathcal{O}(N^{-c\underline{\lambda}})$.
  - If $c > \frac{1}{2\underline{\lambda}}$, a comparison between the series and the integral yields $C_N^\gamma = \mathcal{O}(N^{-1})$, $C_N^{\gamma,\alpha} = \mathcal{O}((\log(N))^{2\frac{1-\alpha}{2\alpha-1}}N^{-\frac{\alpha}{2\alpha-1}})$, $\tilde{s}_N = \mathcal{O}(\log(N)N^{-\frac{1}{2}})$.

  Let us notice that we find the same critical level for the constant c as in the Central Limit Theorem for stochastic algorithms. Indeed, if $c > \frac{1}{2\mathcal{R}e(\lambda_{min})}$ where λ_{min} denotes the eigenvalue of $Dh(\theta^*)$ with the smallest real part then we know that a Central Limit Theorem holds for $(\theta_n^\gamma)_{n\geq 1}$ (see e.g. [9], p.169). Such behavior was already observed in [10].

  The associated bound for the bias is the following:

$$\delta_N \leq K\left(\frac{|\theta_0-\theta^*|}{N^{\underline{\lambda}c}} + \frac{(2C_{\alpha,\mu})^{\frac{1}{2}}}{N^{\underline{\lambda}c\wedge\frac{1}{2}}}\right).$$

- If we choose $\gamma_n = \frac{c}{n^\rho}$, $c > 0$, $\frac{1}{2} < \rho < 1$, then $\delta_N \to 0$, $\Gamma_{1,N} \sim \frac{c}{1-\rho}N^{1-\rho}$ as $N \to +\infty$ and elementary computations show that there exists $C > 0$ s.t. for all $N \geq 1$, $\Pi_{1,N} \leq C\exp(-2\underline{\lambda}\frac{c}{1-\rho}N^{1-\rho})$. Hence, for all $\epsilon \in (0, 1-\rho)$ we have:

$$\begin{aligned} C_N^\gamma = \Pi_{1,N}\sum_{k=1}^N \gamma_k^2\Pi_{1,k}^{-1} &\leq c^2\left\{\Pi_{1,N}\Pi_{1,N-N^{\rho+\epsilon}}^{-1}\sum_{k=1}^{N-N^{\rho+\epsilon}}\frac{1}{k^{2\rho}} + \sum_{k=N-N^{\rho+\epsilon}+1}^{N}\frac{1}{k^{2\rho}}\right\} \\ &\leq c^2\left\{Ce^{-2\underline{\lambda}\frac{c}{1-\rho}(N^{1-\rho}-(N-N^{\rho+\epsilon})^{1-\rho})} + \frac{N^{\rho+\epsilon}}{(N-N^{\rho+\epsilon}+1)^{2\rho}}\right\} \\ &\leq c^2\left\{Ce^{-2\underline{\lambda}cN^\epsilon} + \frac{1}{N^{\rho-\epsilon}}\right\}. \end{aligned}$$

  Up to a modification of ϵ, this yields $C_N^\gamma = \Pi_{1,N}\sum_{k=1}^N\gamma_k^2\Pi_{1,k}^{-1} = o(N^{-\rho+\epsilon})$, $\epsilon \in (0, 1-\rho)$. Similar computations show that $C_N^{\gamma,\alpha} = o(N^{-\frac{(\rho-(1-\alpha))}{2\alpha-1}-\epsilon})$ and we clearly get $\tilde{s}_N = \mathcal{O}\left(\log(N)N^{-(\rho-\frac{1}{2})}\right)$.

Concerning the bias, from Corollary 2.9, we directly obtain the following bound:

$$\delta_N \leq K\left(\exp\left(-\frac{\underline{\lambda} c}{1-\rho}N^{1-\rho}\right)|\theta_0-\theta^*| + \frac{(2C_{\alpha,\mu})^{\frac{1}{2}}}{N^{\frac{\rho}{2}-\epsilon}}\right),\ \forall \epsilon > 0.$$

The impact of the initial difference $|\theta_0-\theta^*|$ is exponentially smaller compared to the case $\gamma_n = \frac{c}{n}$. This is natural since the step sequence is decreasing slower to 0.

Theorem 2.10 (Transport-Entropy inequalities for stochastic approximation with averaging of trajectories)**.** *Let $N \in \mathbb{N}^*$. Assume that the function H of the recursive procedure $(\theta_n^\gamma)_{0\leq n\leq N}$ (with starting point $\theta_0 \in \mathbb{R}^d$) defined by* (1.4) *satisfies* **(HL)**, **(HUA)** *and* **(HLS)**$_\alpha$ *for $\alpha \in [\frac{1}{2},1]$, and that the step sequence $\gamma = (\gamma_n)_{n\geq 1}$ satisfies* (1.5)*. Suppose that the law of the innovation satisfies $(GC(\beta))$, $\beta > 0$. Denote by $\bar{\mu}_N^\gamma$ the law of $(\bar{\theta}_n^\gamma)_{0\leq n\leq N}$ defined by* (1.8)*. Then, $\bar{\mu}_N^\gamma$ satisfies $T_{\bar{\Phi}^*_{\alpha,N}}$ with $\bar{\Phi}^*_{\alpha,N}(\lambda) = \sup_{\rho\geq 0}\left\{\lambda\rho - \bar{\Phi}_{\alpha,N}(\rho)\right\}$ and one has:*

- *If $\alpha \in (\frac{1}{2},1]$, for all $\rho \geq 0$*

$$\bar{\Phi}_{\alpha,N}(\rho) = \varphi_\alpha(\gamma,H,\theta_0)(\bar{C}_N^\gamma\rho^2 \vee \bar{C}_N^{\gamma,\alpha}\rho^{\frac{2\alpha}{2\alpha-1}})$$

- *If $\alpha = \frac{1}{2}$, for all $\rho \in [0,\lambda_{4.1}/\hat{s}_N)$,*

$$\bar{\Phi}_{1/2,N}(\rho) = 2\varphi_{1/2}(\gamma,H,\theta_0)\bar{C}_N^\gamma\frac{(\rho/\lambda_{4.1})^2}{1-(\rho\hat{s}_N/\lambda_{4.1})}$$

Moreover the three concentration rate sequences are defined for $N \in \mathbb{N}^*$ by

$$\bar{C}_N^\gamma := \sum_{k=1}^{N-1}\bar{\gamma}_{k,N}^2, \quad \bar{C}_N^{\gamma,\alpha} := \sum_{k=1}^{N-1}\bar{\gamma}_{k,N}^{\frac{2\alpha}{2\alpha-1}}((k+1)\log^2(k+4))^{\frac{1-\alpha}{2\alpha-1}},$$
$$\hat{s}_N := \max_{1\leq k\leq N-1}(k+1)^{\frac{1}{2}}\log(k+4)\bar{\gamma}_{k,N}e^{\sum_{p=0}^{N-1}\frac{1}{(p+1)\log^2(p+4)}}$$

with $\bar{\gamma}_{k,N} := \frac{\gamma_k}{N}(1+\sum_{j=k+1}^{N-1}(\frac{\Pi_{1,j}}{\Pi_{1,k}})^{\frac{1}{2}})$, and $\Pi_{1,N} := \prod_{p=0}^{N-1}(1-2\underline{\lambda}\gamma_{p+1}+C_{H,\mu}\gamma_{p+1}^2)$, the constants $\varphi_\alpha(\gamma,H,\theta_0)$ and $\lambda_{4.1}$ being defined in Section 4.2 and Proposition 4.5 respectively.

As regards the explicit computation of the Legendre transform of $\bar{\Phi}_{\alpha,N}$, similarly to the previous theorem, we have:

- for $\alpha \in (\frac{1}{2},1]$:
  - if $\lambda \in [0, 2\varphi(\bar{C}_N^\gamma/(\bar{C}_N^{\gamma,\alpha})^{2\alpha-1})^{\frac{1}{2(1-\alpha)}}]$, then $\bar{\Phi}^*_{\alpha,N}(\lambda) = (\lambda^2/4\varphi\bar{C}_N^\gamma)$;
  - If $\lambda \in [\frac{2\alpha}{2\alpha-1}\varphi(\bar{C}_N^\gamma/(\bar{C}_N^{\gamma,\alpha})^{2\alpha-1})^{\frac{1}{2(1-\alpha)}}, +\infty)$, then $\bar{\Phi}^*_{\alpha,N}(\lambda) = \frac{1}{2\alpha}\left(\frac{2\alpha-1}{2\alpha\varphi}\right)^{2\alpha-1}\frac{\lambda^{2\alpha}}{\bar{C}_N^{\gamma,\alpha})^{2\alpha-1}}$;
  - If $\lambda \in (2\varphi(\bar{C}_N^\gamma/(\bar{C}_N^{\gamma,\alpha})^{2\alpha-1})^{\frac{1}{2(1-\alpha)}}, \frac{2\alpha}{2\alpha-1}\varphi(\bar{C}_N^\gamma/(\bar{C}_N^{\gamma,\alpha})^{2\alpha-1})^{\frac{1}{2(1-\alpha)}})$, then $\bar{\Phi}^*_{\alpha,N}(\lambda) = (\frac{\bar{C}_N^\gamma}{\bar{C}_N^{\gamma,\alpha}})^{\frac{2\alpha-1}{2(1-\alpha)}}\lambda - \varphi\frac{(\bar{C}_N^\gamma)^{\frac{\alpha}{1-\alpha}}}{(\bar{C}_N^{\gamma,\alpha})^{\frac{2\alpha-1}{1-\alpha}}}$.
- for $\alpha = \frac{1}{2}$,

$$\forall \lambda \geq 0,\quad \bar{\Phi}^*_{1/2,N}(\lambda) = \frac{2\varphi\bar{C}_N^\gamma}{\hat{s}_N^2}\left(\left(1+\frac{\hat{s}_N\lambda_{4.1}\lambda}{2\varphi\bar{C}_N^\gamma}\right)^{\frac{1}{2}}-1\right)^2$$

Hence, for $N \geq 1$ being fixed, the following simple asymptotic behaviors can be easily derived:

- When λ is small, $\bar{\Phi}^*_{1/2,N}(\lambda) \sim \lambda_{4.1}^2 \lambda^2/(2\varphi \bar{C}_N^{\gamma})$;
- When λ goes to infinity, $\bar{\Phi}^*_{1/2}(\lambda) \sim \lambda_{4.1}\lambda/\hat{s}_N$.

Corollary 2.11. *(Non-asymptotic deviation bounds) Under the same assumptions as Theorem 2.10, for all $N \geq 1$ for all $r \geq 0$, one has*

$$\mathbb{P}_{\theta_0}\left(\left|\bar{\theta}_N^{\gamma} - \theta^*\right| \geq r + \bar{\delta}_N\right) \leq \exp\left(-\Phi^*_{\alpha,N}(r)\right)$$

and $\bar{\delta}_N := \mathbb{E}_{\theta_0}\left[\left|\bar{\theta}_N^{\gamma} - \theta^*\right|\right]$.

Now, we analyze the impact of the step sequence on the concentration rate sequences $\bar{C}_N^{\gamma}$, $\bar{C}_N^{\gamma,\alpha}$, $\hat{s}_N$ and the bias $\bar{\delta}_N$. We first simplify the expression of the concentration rate. Let us note that since the step sequence $(\gamma_n)_{n\geq 1}$ satisfies (1.5), there exists a positive constant $K > 0$ such that $(\Pi_{1,j}\Pi_{1,k}^{-1})^{\frac{1}{2}} \leq K \exp(-\underline{\lambda}(\Gamma_{1,j} - \Gamma_{1,k+1}))$, $k < j$. Moreover, since the function $x \mapsto \exp(-\underline{\lambda}x)$ is decreasing on $[\Gamma_{1,p}, \Gamma_{1,p+1}]$, one clearly gets for all $i, j \in \{0, \cdots, N-1\}$, $i < j$

$$M_j - M_i := \sum_{p=i}^{j-1} e^{-\underline{\lambda}\Gamma_{1,p+1}}\gamma_{p+1} = \sum_{p=i}^{j-1}\int_{\Gamma_{1,p}}^{\Gamma_{1,p+1}} e^{-\underline{\lambda}\Gamma_{1,p+1}}dx \leq \frac{1}{\underline{\lambda}}(e^{-\underline{\lambda}\Gamma_{1,i}} - e^{-\underline{\lambda}\Gamma_{1,j})}$$

so that, using the latter bound and an Abel transform, we obtain

$$\begin{aligned}\sum_{j=k+1}^{N-1}\exp(-\underline{\lambda}\Gamma_{1,j+1}) = \sum_{j=k+1}^{N-1}(M_{j+1} - M_j)\gamma_{j+1}^{-1} &\leq -\frac{1}{\underline{\lambda}}\left(\sum_{j=k+1}^{N-1}(e^{-\underline{\lambda}\Gamma_{1,j+1}} - e^{-\underline{\lambda}\Gamma_{1,j})}\gamma_{j+1}^{-1}\right) \\ &\leq -\frac{1}{\underline{\lambda}}\left(e^{-\underline{\lambda}\Gamma_{1,N}}\gamma_{N+1}^{-1} - e^{-\underline{\lambda}\Gamma_{1,k+1}}\gamma_{k+2}^{-1} - \sum_{p=k+1}^{N-1} e^{-\underline{\lambda}\Gamma_{1,p+1}}(\gamma_{p+2}^{-1} - \gamma_{p+1}^{-1})\right)\end{aligned}$$

which finally leads to the following bound

$$\bar{\gamma}_{k,N} \leq \frac{K}{\underline{\lambda}}\left(\frac{\gamma_k\gamma_{k+2}^{-1}}{N} + \frac{\gamma_k}{N}\sum_{p=k+1}^{N-1} e^{-\underline{\lambda}(\Gamma_{1,p} - \Gamma_{1,k+1})}(\gamma_{p+2}^{-1} - \gamma_{p+1}^{-1})\right). \tag{2.2}$$

Now, we are in position to study the impact of the step sequence $(\gamma_n)_{n\geq 1}$ on the concentration rate sequences:

- If we select $\gamma_n = \frac{c}{n}$ with $c > 0$, then, using that $\Gamma_{1,N} = c\log(N) + c_1' + r_N$, $c_1' > 0$ with $r_N \to 0$, one easily derives from (2.2) that there exists $C > 0$ such that

$$\bar{\gamma}_{k,N} \leq C\left(\frac{1}{N} + \frac{1}{k^{1-c\underline{\lambda}}}\frac{1}{N}\sum_{p=k}^{N-1}\frac{1}{p^{\underline{\lambda}c}}\right),$$

  and a comparison between the series and the integral yields the following bounds:

  - If $\underline{\lambda}c < \frac{1}{2}$, one has: $\bar{C}_N^{\gamma} = \mathcal{O}(N^{-2c\underline{\lambda}})$, $\bar{C}_N^{\gamma,\alpha} = \mathcal{O}(N^{-\frac{2\alpha}{2\alpha-1}c\underline{\lambda}})$ and $\hat{s}_N = \mathcal{O}(N^{-c\underline{\lambda}})$.
  - If $\underline{\lambda}c > \frac{1}{2}$, one has: $\bar{C}_N^{\gamma} = \mathcal{O}(N^{-1})$, $\bar{C}_N^{\gamma,\alpha} = \mathcal{O}((\log(N))^{2\frac{1-\alpha}{2\alpha-1}}N^{-\frac{\alpha}{2\alpha-1}})$ and $\hat{s}_N = \mathcal{O}(N^{-\frac{1}{2}})$.

  Hence, we clearly see that for the case $\gamma_n = \frac{c}{n}$, averaging the trajectories of a stochastic approximation algorithm is not the key to circumvent the lake of robustness concerning the choice of the constant c.

The bound for the bias is obtained by averaging the bound previously obtained for δ_N. We easily get:

$$\bar{\delta}_N \leq \frac{1}{N}\sum_{k=0}^{N-1}\mathbb{E}_{\theta_0}[|\theta_k^\gamma - \theta^*|] \leq K\left(\frac{|\theta_0-\theta^*|}{N^{\underline{\lambda} c}} + \frac{(2C_{\alpha,\mu})^{\frac{1}{2}}}{N^{\underline{\lambda} c \wedge \frac{1}{2}}}\right)$$

- If we choose $\gamma_n = \frac{c}{n^\rho}$, $c > 0$, $\frac{1}{2} < \rho < 1$ then we have for $k \leq p$

$$\begin{aligned}\Gamma_{1,p} - \Gamma_{1,k} = \sum_{j=k+1}^{p} j^{-\rho} = \sum_{j=k+1}^{p}\int_j^{j+1}\frac{1}{j^\rho}dx \geq \int_{k+1}^{p+1}\frac{1}{x^\rho}dx \\ \geq \frac{1}{1-\rho}\left((p+1)^{1-\rho} - (k+1)^{1-\rho}\right)\end{aligned}$$

so that for some positive constant C which may vary from line to line

$$\begin{aligned}\sum_{p=k+1}^{N-1} e^{-\underline{\lambda}(\Gamma_{1,p}-\Gamma_{1,k+1})}(\gamma_{p+2}^{-1} - \gamma_{p+1}^{-1}) &\leq C e^{\frac{\lambda}{1-\rho}(k+1)^{1-\rho}}\left(\sum_{p=k+1}^{N-1} e^{-\frac{\lambda}{1-\rho}(p+1)^{1-\rho}}\frac{1}{(p+1)^{1-\rho}}\right) \\ &\leq C e^{\frac{\lambda}{1-\rho}(k+1)^{1-\rho}}\int_{k+1}^{N} e^{-\frac{\lambda}{1-\rho}x^{1-\rho}} x^{-(1-\rho)}dx \\ &\leq C e^{\frac{\lambda}{1-\rho}(k+1)^{1-\rho}}\int_{(k+1)^{1-\rho}}^{N^{1-\rho}} e^{-\frac{\lambda}{1-\rho}x} x^{\frac{2\rho-1}{1-\rho}}dx\end{aligned}$$

where we use a change of variable in the latter integral. For k large enough, the function $x \mapsto e^{-\frac{\lambda}{1-\rho}x}x^{\frac{2\rho}{1-\rho}}$ is decreasing on $[k, +\infty)$ which implies

$$\begin{aligned}e^{\frac{\lambda}{1-\rho}(k+1)^{1-\rho}}\int_{(k+1)^{1-\rho}}^{(N-1)^{1-\rho}} e^{-\frac{\lambda}{1-\rho}x}x^{\frac{2\rho}{1-\rho}}\frac{1}{x^{\frac{1}{1-\rho}}}dx &\leq C(k+1)^{2\rho}\left[-\frac{1-\rho}{\rho}x^{-\frac{\rho}{1-\rho}}\right]_{(k+1)^{1-\rho}}^{+\infty} \\ &\leq C(k+1)^\rho.\end{aligned}$$

Hence, we finally have $\bar{\gamma}_{k,N} = \mathcal{O}(N^{-1})$ and $\bar{C}_N^\gamma = \mathcal{O}(N^{-1})$, $\bar{C}_N^{\gamma,\alpha} = \mathcal{O}((\log(N))^{2\frac{1-\alpha}{2\alpha-1}}N^{-\frac{\alpha}{2\alpha-1}})$ and $\hat{s}_N = \mathcal{O}(\log(N)N^{-\frac{1}{2}})$. Hence, averaging has allowed the concentration rate to go from the slow concentration rates $o(N^{-\rho+\epsilon})$, $o(N^{-\frac{\rho-(1-\alpha)}{2\alpha-1}-\epsilon})$ for all $\epsilon > 0$ and $\mathcal{O}\left(\log(N)N^{-(\rho-\frac{1}{2})}\right)$ to the optimal rates $\mathcal{O}(N^{-1})$, $\mathcal{O}((\log(N))^{2\frac{1-\alpha}{2\alpha-1}}N^{-\frac{\alpha}{2\alpha-1}})$ and $\hat{s}_N = \mathcal{O}(\log(N)N^{-\frac{1}{2}})$ for free, i.e. without any condition on the step sequence parameter c.

Concerning the bias, by averaging the bias sequence $(\delta_k)_{1\leq k\leq N-1}$ we directly obtain the following bound

$$\bar{\delta}_N \leq K\left(\frac{|\theta_0-\theta^*|}{N} + \frac{(2C_{\alpha,\mu})^{\frac{1}{2}}}{N^{\frac{\rho}{2}-\epsilon}}\right),\ \forall \epsilon > 0$$

Hence, we see that there is no sub-exponential decreasing of the impact of the initial condition but a decay at rate $\mathcal{O}(N^{-1})$. Consequently, this leads us to say that a stochastic approximation algorithm must be averaged after few iterations in practical implementations and not directly from the first step.

3 Euler Scheme: Proof of the Main Results

In this section we will assume that **(HS)** and **(HD$_\alpha$)** are in force.

3.1 Proof of Theorem 2.1

The proof of Theorem 2.1 is divided into several propositions. Next proposition stresses the key role played by the Gaussian concentration property of the innovations law, that is a weaker concentration regime will lead to a lower integrability rate with respect to α.

Proposition 3.1. *Denote by $X^\Delta := (X^\Delta_{t_k})_{0\le k\le N}$ the scheme* (1.2) *with time step $\Delta = T/N$, $N \in \mathbb{N}^*$ associated to the diffusion $(SDE_{b,\sigma})$ starting from x at time 0. Assume that the innovations $(U_i)_{i\ge 1}$ of* (1.2) *satisfy $(GC(\beta))$ for some $\beta > 0$. Then, there exists $\varepsilon_\beta > 0$ which only depends on the law μ such that for all $\lambda < \min(1, \varepsilon_\beta(2\eta\alpha C_\sigma T\exp(CT))^{-1})$, one has*

$$\sup_{0\le n\le N} \log\left(\mathbb{E}_x\left[\exp(\lambda V^\alpha(X^\Delta_{t_n}))\right]\right) \le \lambda\exp(CT)V^\alpha(x) + \frac{1}{2}\log\left(\mathbb{E}\left[\exp\left(\lambda 2\eta\alpha C_\sigma T\exp(CT)|U_1|^2\right)\right]\right).$$

with $C := C(b,\sigma,V,\alpha,\Delta) = \alpha(C_V C_b)^{\frac{1}{2}} + \beta C_\sigma\alpha^2(1+2\eta\Delta)^2(C_V + C_b) + \alpha\eta C_b\Delta$.

Proof. Using the concavity of $x \mapsto x^\alpha$, $\alpha \in (0,1]$, we have for all $k \ge 0$

$$V^\alpha(X^\Delta_{t_{k+1}}) - V^\alpha(X^\Delta_{t_k}) \le \alpha V^{\alpha-1}(X^\Delta_{t_k})(V(X^\Delta_{t_{k+1}}) - V(X^\Delta_{t_k})).$$

A Taylor expansion of order 2 of the function V, recalling that $2\eta = \sup_{x\in\mathbb{R}^d}\left\|\nabla^2 V(x)\right\| < +\infty$, yields

$$V(X^\Delta_{t_{k+1}}) - V(X^\Delta_{t_k})) \le \nabla V(X^\Delta_{t_k}).(X^\Delta_{t_{k+1}} - X^\Delta_{t_k}) + \eta|X^\Delta_{t_{k+1}} - X^\Delta_{t_k}|^2,$$

which together with the previous inequality leads to

$$\begin{aligned} V^\alpha(X^\Delta_{t_{k+1}}) - V^\alpha(X^\Delta_{t_k}) \le{}& \alpha\Delta\frac{\nabla V(X^\Delta_{t_k}).b(t_k,X^\Delta_{t_k})}{V^{1-\alpha}(X^\Delta_{t_k})} + \alpha\Delta^{\frac{1}{2}}\frac{\nabla V(X^\Delta_{t_k}).\sigma(t_k,X^\Delta_{t_k})U_{k+1}}{V^{1-\alpha}(X^\Delta_{t_k})} \\ &+ \alpha\eta\Delta^2\frac{|b(t_k,X^\Delta_{t_k})|^2}{V^{1-\alpha}(X^\Delta_{t_k})} + 2\alpha\eta\Delta^{\frac{3}{2}}\frac{b(t_k,X^\Delta_{t_k}).\sigma(t_k,X^\Delta_{t_k})U_{k+1}}{V^{1-\alpha}(X^\Delta_{t_k})} \\ &+ \alpha\eta\Delta\frac{|\sigma(t_k,X^\Delta_{t_k})U_{k+1}|^2}{V^{1-\alpha}(X^\Delta_{t_k})}. \end{aligned}$$

From **(HD$_\alpha$)**, for all $(x,u) \in \mathbb{R}^d \times \mathbb{R}^q$, we clearly have $\sup_{t\in[0,T]}|\nabla V(x).b(t,x)| \le (C_V C_b)^{\frac{1}{2}}V(x)$ and $\sup_{t\in[0,T]}|\sigma(t,x)u|^2 \le C_\sigma V^{1-\alpha}(x)|u|^2$ which yields

$$\begin{aligned} V^\alpha(X^\Delta_{t_{k+1}}) \le{}& V^\alpha(X^\Delta_{t_k})(1 + \alpha(C_V C_b)^{\frac{1}{2}}\Delta + \alpha\eta C_b\Delta^2) \\ &+ \alpha\Delta^{\frac{1}{2}}(1+2\eta\Delta)\frac{(\nabla V(X^\Delta_{t_k}) + b(X^\Delta_{t_k})).\sigma(X^\Delta_{t_k})U_{k+1}}{V^{1-\alpha}(X^\Delta_{t_k})} + C_\sigma\alpha\eta\Delta|U_{k+1}|^2. \end{aligned}$$

Using **(HD$_\alpha$)**, $\forall x \in \mathbb{R}^d$ the functions $g(x,.) : u \mapsto \frac{(\nabla V(x)+b(x)).\sigma(x)u}{V^{1-\alpha}(x)}$ are Lipschitz, and more precisely satisfy

$$\forall x \in \mathbb{R}^d, \qquad \sup_{(u,u')\in(\mathbb{R}^q)^2}\frac{|g(x,u) - g(x,u')|}{|u-u'|} \le (C_V^{1/2} + C_b^{1/2})C_\sigma^{1/2}V^{\frac{\alpha}{2}}(x).$$

Hence, from the Cauchy Schwarz inequality and since the law of the innovations satisfy $(GC(\beta))$ for some $\beta > 0$, there exists $\epsilon_\beta > 0$ such that for $\lambda < \min(1, \varepsilon_\beta(2\eta\alpha C_\sigma\Delta)^{-1})$, one has

$$\begin{aligned}\mathbb{E}\left[\exp(\lambda V^{\alpha}(X^{\Delta}_{t_{k+1}}))\Big|\,\mathcal{F}_{t_k}\right] &\le \exp(\lambda V^{\alpha}(X^{\Delta}_{t_k})(1+\alpha(C_V C_b)^{\frac12}\Delta+\alpha\eta C_b\Delta^2))\\ &\quad\times \mathbb{E}\left[\exp(2\lambda\alpha\Delta^{\frac12}(1+2\eta\Delta)g(X^{\Delta}_{t_k},U_{k+1}))\Big|\,\mathcal{F}_{t_k}\right]^{\frac12}\\ &\quad\times \mathbb{E}\left[\exp(2\lambda\eta\alpha C_\sigma\Delta|U_{k+1}|^2)\big|\,\mathcal{F}_{t_k}\right]^{\frac12}\\ &\le \exp(\lambda V^{\alpha}(X^{\Delta}_{t_k})(1+\alpha(C_V C_b)^{\frac12}\Delta+\alpha\eta C_b\Delta^2))\\ &\quad\times \exp(\lambda^2\beta\alpha^2\Delta(1+2\eta\Delta)^2(C_V+C_b)C_\sigma V^{\alpha}(X^{\Delta}_{t_k}))\\ &\quad\times \mathbb{E}\left[\exp(2\lambda\eta\alpha C_\sigma\Delta|U_1|^2)\right]^{\frac12}\\ &\le \exp(\lambda C(\Delta)V^{\alpha}(X^{\Delta}_{t_k}))\mathbb{E}\left[\exp(2\lambda\eta\alpha C_\sigma\Delta|U_1|^2)\right]^{\frac12},\end{aligned}$$

where $C(\Delta):=1+\Delta\left(\alpha(C_VC_b)^{\frac12}+\beta C_\sigma\alpha^2(1+2\eta\Delta)^2(C_V+C_b)+\alpha\eta C_b\Delta\right)$. Now define $V_k=\frac{V^{\alpha}(X^{\Delta}_{t_k})}{C(\Delta)^k}$, for $k\in\{0,\cdots,N\}$. Taking expectation in both sides of the previous inequality clearly implies

$$\mathbb{E}\left[\exp(\lambda V_{k+1})\right]\le \mathbb{E}\left[\exp(\lambda V_k)\right]\mathbb{E}\left[\exp\left(\lambda\frac{2\eta\alpha C_\sigma\Delta}{C(\Delta)^{k+1}}|U_1|^2\right)\right]^{\frac12}$$

and by a straightforward induction, for $n\in\{0,\cdots,N\}$ we have

$$\mathbb{E}\left[\exp(\lambda V_n)\right]\le \exp(\lambda V_0)\prod_{k=0}^{n-1}\mathbb{E}\left[\exp\left(\lambda\frac{2\eta\alpha C_\sigma\Delta}{C(\Delta)^{k+1}}|U_1|^2\right)\right]^{\frac12},$$

which finally yields, for $\lambda<\min(1,\varepsilon_\beta(2\eta\alpha C_\sigma\Delta C(\Delta)^n)^{-1})$,

$$\mathbb{E}\left[\exp(\lambda V^{\alpha}(X^{\Delta}_{t_n}))\right]\le \exp(\lambda C(\Delta)^nV^{\alpha}(x))\prod_{k=0}^{n-1}\mathbb{E}\left[\exp\left(\lambda 2\eta\alpha C_\sigma\Delta C(\Delta)^{k+1}|U_1|^2\right)\right]^{\frac12}.$$

Observe now that $C(\Delta)^N\le\exp(CT)$ with $C:=C(b,\sigma V,\alpha,\Delta)=\alpha(C_VC_b)^{\frac12}+\beta C_\sigma\alpha^2(1+2\eta\Delta)^2(C_V+C_b)+\alpha\eta C_b\Delta$. Using Jensen's inequality, the latter bound clearly provides the following control of the quantity of interest for $\lambda<\min(1,\varepsilon_\beta(2\eta\alpha C_\sigma T\exp(CT))^{-1})$

$$\sup_{0\le n\le N}\log\left(\mathbb{E}\left[\exp(\lambda V^{\alpha}(X^{\Delta}_{t_n}))\right]\right)\le\lambda\exp(CT)V^{\alpha}(x)+\frac12\log\left(\mathbb{E}\left[\exp\left(\lambda 2\eta\alpha C_\sigma T\exp(CT)|U_1|^2\right)\right]\right).$$

□

Corollary 3.2. *Assume that the assumptions of Proposition 3.1 are satisfied. Then, there exists a constant $K_{3.1}$ such that for all $\alpha\in(\frac12,1]$, one has*

$$\forall\lambda\ge0,\quad\sup_{0\le n\le N}\log\left(\mathbb{E}_x\left[\exp(\lambda V^{1-\alpha}(X^{\Delta}_{t_n}))\right]\right)\le K_{3.1}(\lambda\vee\lambda^{\frac{\alpha}{2\alpha-1}}).$$

Remark 3.3. *The constant $K_{3.1}$ can be explicitly computed. Indeed, one has $K_{3.1}:=$*

$\max\left(\Psi_1(T,\Delta,x,b,\sigma),\Psi_2(T,\Delta,x,b,\sigma)\right)$ *with*

$$\Psi_1(T,\Delta,x,b,\sigma) := e^{\frac{2\alpha-1}{\alpha}\underline{\rho}^{-\frac{1-\alpha}{2\alpha-1}}}\exp\left(\underline{\rho}\frac{1-\alpha}{\alpha}e^{CT}V^\alpha(x)+\frac{1}{2}\log\mathbb{E}[e^{\frac{\epsilon_\beta(1-\alpha)}{2\alpha}|U|^2}]\right)$$
$$+\left(V^{1-\alpha}(x)+\left(\frac{C_\sigma\mathbb{E}[|U|^2]}{K}\right)^{\frac{1-\alpha}{\alpha}}\right)e^{(1-\alpha)KT},$$
$$\Psi_2(T,\Delta,x,b,\sigma) := \underline{\rho}^{-\frac{1-\alpha}{2\alpha-1}}\frac{2\alpha-1}{\alpha}+\underline{\rho}\frac{1-\alpha}{\alpha}e^{CT}V^\alpha(x)+\frac{1}{2}\log\mathbb{E}\left[\exp\left(\frac{\epsilon_\beta(1-\alpha)}{2\alpha}|U|^2\right)\right],$$
$$\underline{\rho} := \frac{1}{2}\min(1,\varepsilon_\beta(2\eta\alpha C_\sigma T\exp(CT))^{-1}),$$
$$C := C(b,\sigma V,\alpha,\Delta) = \alpha(C_VC_b)^{\frac{1}{2}}+\beta C_\sigma\alpha^2(1+2\eta\Delta)^2(C_V+C_b)+\alpha\eta C_b\Delta$$
$$K := K(V,b,\Delta) = (C_VC_b)^{\frac{1}{2}}+\eta C_b\Delta$$

Proof. For $\lambda\in[0,1]$, one has

$$\begin{aligned}\mathbb{E}_x[\exp(\lambda V^{1-\alpha}(X^\Delta_{t_n}))] &= 1+\lambda\mathbb{E}_x[V^{1-\alpha}(X^\Delta_{t_n})]+\sum_{k\geq 2}\frac{\lambda^k}{k!}\mathbb{E}_x[V^{(1-\alpha)k}(X^\Delta_{t_n})]\\ &\leq 1+\lambda\mathbb{E}_x[V^{1-\alpha}(X^\Delta_{t_n})]+\lambda\sum_{k\geq 0}\frac{1}{k!}\mathbb{E}_x[V^{(1-\alpha)k}(X^\Delta_{t_n})]\\ &\leq \exp\left(\lambda(\mathbb{E}_x[V^{1-\alpha}(X^\Delta_{t_n})]+\mathbb{E}_x[\exp(V^{1-\alpha}(X^\Delta_{t_n}))])\right),\end{aligned}$$

Tedious but simple computations, in the spirit of Proposition 3.1, show that

$$\mathbb{E}_x[V^{1-\alpha}(X^\Delta_{t_n})]\leq\mathbb{E}_x[V^\alpha(X^\Delta_{t_n})]^{\frac{1-\alpha}{\alpha}}\leq\left(V^{1-\alpha}(x)+\left(\frac{C_\sigma\mathbb{E}[|U|^2]}{K}\right)^{\frac{1-\alpha}{\alpha}}\right)e^{(1-\alpha)KT}.$$

with $K:=K(V,b,\Delta)=(C_VC_b)^{\frac{1}{2}}+\eta C_b\Delta$.
Thanks to the following Young inequality, for all $(\rho,x)\in\mathbb{R}_+\times\mathbb{R}^d$, $V^{1-\alpha}(x)\leq\frac{1-\alpha}{\alpha}\rho V^\alpha(x)+\frac{2\alpha-1}{\alpha}\rho^{-\frac{1-\alpha}{2\alpha-1}}$, which is valid if $\alpha\in(\frac{1}{2},1]$, one has for $\rho=\underline{\rho}:=\frac{1}{2}\min(1,\varepsilon_\beta(2\eta\alpha C_\sigma T\exp(CT))^{-1})$

$$\begin{aligned}\sup_{0\leq n\leq N}\mathbb{E}_x[\exp(V^{1-\alpha}(X^\Delta_{t_n}))] &\leq e^{\frac{2\alpha-1}{\alpha}\underline{\rho}^{-\frac{1-\alpha}{2\alpha-1}}}\sup_{0\leq n\leq N}\mathbb{E}_x\left[\exp\left(\frac{1-\alpha}{\alpha}\rho V^\alpha(X^\Delta_{t_n})\right)\right]\\ &\leq e^{\frac{2\alpha-1}{\alpha}\underline{\rho}^{-\frac{1-\alpha}{2\alpha-1}}}e^{\underline{\rho}\frac{1-\alpha}{\alpha}e^{CT}V^\alpha(x)+\frac{1}{2}\log\mathbb{E}[\exp(\frac{\epsilon_\beta(1-\alpha)}{2\alpha}|U|^2)]}\end{aligned}$$

where we used Proposition 3.1 for the last inequality.

Now, for all $\lambda>1$, using the Young inequality $\lambda V^{1-\alpha}(X^\Delta_{t_n})\leq(\frac{2\alpha-1}{\alpha})\rho^{-\frac{1-\alpha}{2\alpha-1}}\lambda^{\frac{\alpha}{2\alpha-1}}+(\frac{1-\alpha}{\alpha})\rho V^\alpha(X^\Delta_{t_n})$, valid for all $\rho>0$ (to be chosen later on) and for all $\alpha\in(\frac{1}{2},1]$, one derives

$$\begin{aligned}\mathbb{E}_x[\exp(\lambda V^{1-\alpha}(X^\Delta_{t_n}))] &\leq \exp\left((\frac{2\alpha-1}{\alpha})\rho^{-\frac{1-\alpha}{2\alpha-1}}\lambda^{\frac{\alpha}{2\alpha-1}}\right)\mathbb{E}_x\left[\exp\left(\left(\frac{1-\alpha}{\alpha}\right)\rho V^\alpha(X^\Delta_{t_n})\right)\right]\\ &\leq \exp\left(K\lambda^{\frac{\alpha}{2\alpha-1}}\right)\end{aligned}$$

with $K(\rho):=\frac{2\alpha-1}{\alpha}\rho^{-\frac{1-\alpha}{2\alpha-1}}+\log(\mathbb{E}_x\left[e^{\left(\frac{1-\alpha}{\alpha}\right)\rho V^\alpha(X^\Delta_{t_n})}\right])$ and $\frac{1-\alpha}{\alpha}\rho<\min(1,\varepsilon_\beta(2\eta\alpha C_\sigma Te^{CT})^{-1})$. We select $\rho=\underline{\rho}$ in the last inequality to complete the proof and use Proposition 3.1 to bound the quantity $K(\underline{\rho})$. □

Corollary 3.4. *Under the same assumptions as Proposition 3.1, one has*

$$\forall\lambda\in[0,\lambda_{3.2}),\quad\sup_{0\leq n\leq N}\log\left(\mathbb{E}_x\left[\exp(\lambda^2V^{1/2}(X^\Delta_{t_n}))\right]\right)\leq K_{3.2}\frac{(\lambda/\lambda_{3.2})^2}{1-(\lambda/\lambda_{3.2})}$$

with $K_{3.2} := \lambda_{3.2}^2 e^{CT}(2V^{\frac{1}{2}}(x)+2\eta\alpha C_\sigma \mathbb{E}[|U_1|^2]T)$ *and* $\lambda_{3.2}$ *satisfies* $\mathbb{E}[e^{\lambda_{3.2}^2 2\eta\alpha C_\sigma T\exp(CT)|U_1|^2}] \leq 2$.

Proof. By definition of $\lambda_{3.2}$, we have $\forall k \geq 1, \lambda_{3.2}^{2k}(2\eta\alpha C_\sigma T\exp(CT))^k \mathbb{E}[|U_1|^{2k}] \leq 2k!$. Consequently, setting temporarily $C_1 := \exp(CT)V^{1/2}(x), C_2 := 2\eta\alpha C_\sigma T\exp(CT)$ for sake of simplicity, simple computations show that for $\lambda < \lambda_{3.2}$

$$\begin{aligned}\log\mathbb{E}\left[\exp\left(\lambda^2 C_2|U_1|^2\right)\right] - \lambda^2 C_2\mathbb{E}[|U_1|^2] &= \log\left(1+\sum_{k\geq 1}\frac{\lambda^{2k}C_2^k\mathbb{E}[|U_1|^{2k}]}{k!}\right) - \lambda^2 C_2\mathbb{E}[|U_1|^2]\\ &\leq \sum_{k\geq 2}\frac{\lambda^{2k}C_2^k\mathbb{E}[|U_1|^{2k}]}{k!}\\ &\leq 2\sum_{k\geq 2}\left(\frac{\lambda}{\lambda_{3.2}}\right)^{2k} \leq 2\frac{(\lambda/\lambda_{3.2})^2}{1-(\lambda/\lambda_{3.2})}\end{aligned}$$

hence, using Proposition 3.1 for $\alpha = \frac{1}{2}$, we clearly get

$$\begin{aligned}\sup_{0\leq n\leq N}\log\left(\mathbb{E}_x\left[\exp(\lambda^2 V^{1/2}(X_{t_n}^\Delta))\right]\right) &\leq \lambda_{3.2}^2\left(C_1+\frac{C_2\mathbb{E}[|U_1|^2]}{2}\right)(\lambda/\lambda_{3.2})^2 + \frac{(\lambda/\lambda_{3.2})^2}{1-(\lambda/\lambda_{3.2})}\\ &\leq 2\lambda_{3.2}^2\left(C_1+\frac{C_2\mathbb{E}[|U_1|^2]}{2}\right)\frac{(\lambda/\lambda_{3.2})^2}{1-(\lambda/\lambda_{3.2})}.\end{aligned}$$

This completes the proof. □

Proposition 3.5. *(Control of the Lipschitz modulus of iterative kernels) Denote the Lipschitz modulus of* b *and* σ *appearing in the diffusion process* ($SDE_{b,\sigma}$) *by* $[b]_1$ *and* $[\sigma]_1$*, respectively. Denote by* P_k *and* $P_{k,p} = P_k \circ \cdots \circ P_{p-1}$, $k,\ p \in \{0, \cdots, N-1\}$, $k \leq p$ *the (Feller) transition kernel and the iterative kernels of the Markov chain defined by the scheme* (1.2)*, respectively. Then, for all real-valued Lipschitz function* f *and for all* $k,\ p \in \{0,\cdots,N-1\}$, $k \leq p$ *the functions* $P_k(f)$ *are Lipschitz-continuous and one has*

$$[P_{k,p}(f)]_1 := \sup_{(x,x')\in(\mathbb{R}^d)^2}\frac{|P_{k,p}(f)(x)-P_{k,p}(f)(x')|}{|x-x'|} \leq [f]_1(1+C(b,\sigma,\Delta)\Delta)^{\frac{p-k}{2}}$$

where $[f]_1$ *stands for the Lipschitz modulus of the function* f *and* $C(b,\sigma,\Delta) = 2[b]_1 + [\sigma]_1^2 + \Delta[b]_1^2$.

Proof. Using the Cauchy Schwarz inequality and **(HS)**, for all $(x,y) \in (\mathbb{R}^d)^2$ and for all $k \in \{0,\cdots,N-1\}$, one has

$$\begin{aligned}|P_k(f)(x)-P_k(f)(y)| &\leq \mathbb{E}\left[\left|f(x+b(t_k,x)\Delta+\sigma(t_k,x)U_1) - f(y+b(t_k,y)\Delta+\Delta^{\frac{1}{2}}\sigma(t_k,y)U_1)\right|\right]\\ &\leq [f]_1\mathbb{E}\left[\left|x-y+(b(t_k,x)-b(t_k,y))\Delta+\Delta^{\frac{1}{2}}(\sigma(t_k,x)-\sigma(t_k,y))U_1\right|^2\right]^{\frac{1}{2}}\\ &\leq [f]_1(1+C(b,\sigma,\Delta)\Delta)^{\frac{1}{2}}|x-y|.\end{aligned}$$

A straightforward induction argument completes the proof.

□

Proposition 3.6. *(Control of the Laplace transform) Denote by* X_T^Δ *the value at time* T *of the scheme* (1.2) *associated to the diffusion* ($SDE_{b,\sigma}$)*. Assume that the innovations*

$(U_n)_{n\geq 1}$ in (1.2) satisfy ($GC(\beta)$) for some $\beta > 0$. Let f be a real-valued 1-Lipschitz-continuous function defined on $\mathbb{R}^d$. For all $\lambda \geq 0$ and for all $\alpha \in (\frac{1}{2}, 1]$, one has

$$\mathbb{E}_x\left[\exp(\lambda f(X_T^\Delta))\right] \leq \exp(\lambda \mathbb{E}_x\left[f(X_T^\Delta)\right])\exp\left(K_{3.1}(\varphi(T,b,\sigma,\Delta)\vee\varphi(T,b,\sigma,\Delta)^{\frac{\alpha}{2\alpha-1}})(\lambda^2\vee\lambda^{\frac{2\alpha}{2\alpha-1}})\right),$$

with $\varphi(T,b,\sigma,\Delta) := C_\sigma\beta\frac{(1+C(\Delta)\Delta)}{4C(\Delta)}e^{3C(\Delta)T}$ and $C(\Delta) := 2[b]_1 + [\sigma]_1^2 + \Delta[b]_1^2$.
If $\alpha = \frac{1}{2}$, for all $\lambda \in [0, \varphi(T,b,\sigma,\Delta)^{-1/2}\lambda_{3.2})$, one has

$$\mathbb{E}_x\left[\exp(\lambda f(X_T^\Delta))\right] \leq \exp(\lambda \mathbb{E}_x\left[f(X_T^\Delta)\right])\exp\left(K_{3.2}\frac{(\lambda\varphi(T,b,\sigma,\Delta)^{1/2}/\lambda_{3.2})^2}{1-(\lambda\varphi(T,b,\sigma,\Delta)^{1/2}/\lambda_{3.2})}\right).$$

Proof. As mentionned earlier on in the introduction, we begin our proof using that the law μ of the innovation satisfies ($GC(\beta)$) and **($\mathbf{HD}_\alpha$)**. Hence, for $\lambda \geq 0$ and $k \in \{0, \cdots, N-1\}$, one has

$$\begin{aligned} P_k(\exp(\lambda f))(x) &= \mathbb{E}\left[\exp\left(\lambda f\left(x + b(t_k,x)\Delta + \sigma(t_k,x)\Delta^{1/2}U_{k+1}\right)\right)\right] \\ &\leq \exp\left(\lambda P_k(f)(x) + \beta\frac{\lambda^2}{4}[f]_1^2\Delta|\sigma(t_k,x)|^2\right) \\ &\leq \exp\left(\lambda P_k(f)(x) + C_\sigma\beta\frac{\lambda^2}{4}[f]_1^2\Delta V^{1-\alpha}(x)\right). \end{aligned} \tag{3.1}$$

Taking expectation from both sides of the last inequality and using the Hölder inequality with conjugate exponents (p,q) (to be specified later on) leads to

$$\mathbb{E}_x\left[\exp(\lambda f(X_{t_{k+1}}^\Delta))\right] \leq \mathbb{E}_x\left[\exp(\lambda p P_k(f)(X_{t_k}^\Delta))\right]^{\frac{1}{p}}\mathbb{E}_x\left[\exp\left(\frac{qC_\sigma\beta}{4}\Delta\lambda^2[f]_1^2V^{1-\alpha}(X_{t_k}^\Delta)\right)\right]^{\frac{1}{q}}. \tag{3.2}$$

Now, we apply the last inequality for $f := P_{k+1,N}(f)$ and obtain

$$\begin{aligned} \mathbb{E}_x\left[\exp(\lambda P_{k+1,N}(f)(X_{t_{k+1}}^\Delta))\right] &\leq \mathbb{E}_x\left[\exp(\lambda p P_{k,N}(f)(X_{t_k}^\Delta))\right]^{\frac{1}{p}} \\ &\quad\times\mathbb{E}_x\left[\exp\left(\frac{qC_\sigma\beta}{4}\Delta\lambda^2[P_{k+1,N}(f)]_1^2V^{1-\alpha}(X_{t_k}^\Delta)\right)\right]^{\frac{1}{q}}. \end{aligned}$$

Consequently, an elementary induction yields

$$\begin{aligned} \mathbb{E}_x\left[\exp(\lambda f(X_T^\Delta))\right] &= \mathbb{E}_x\left[\exp(\lambda P_{N,N}(f)(X_{t_N}^\Delta))\right] \\ &\leq \mathbb{E}_x\left[\exp(\lambda p^N P_{0,N}(f)(x))\right]^{\frac{1}{p^N}} \\ &\quad\times\prod_{k=0}^{N-1}\left(\mathbb{E}_x\left[\exp\left(\frac{C_\sigma\beta}{4}\lambda^2qp^{2k}\Delta[P_{N-k,N}(f)]_1^2V^{1-\alpha}(X_{t_{N-k-1}}^\Delta)\right)\right]^{\frac{1}{q}}\right)^{\frac{1}{p^k}} \\ &\leq \exp(\lambda\mathbb{E}_x\left[f(X_T^\Delta)\right]) \\ &\quad\times\exp\left(\sum_{k=0}^{N-1}\frac{1}{p^k}\frac{1}{q}\sup_{0\leq n\leq N}\log\left(\mathbb{E}_x\left[e^{\frac{C_\sigma\beta}{4}\lambda^2\Delta qp^{2N}(1+C(\Delta)\Delta)^NV^{1-\alpha}(X_{t_n}^\Delta)}\right]\right)\right) \end{aligned}$$

where we used Proposition 3.5 for the last inequality. Observe now that since (p,q) are conjugate exponents, we have $\frac{1}{q}\sum_{k=0}^{N-1}\frac{1}{p^k} = \frac{1}{q}(1-\frac{1}{p^N})\frac{1}{1-\frac{1}{p}} \leq \frac{1}{q}\frac{p}{p-1} = 1$, so that

$$\mathbb{E}_x\left[\exp(\lambda f(X_T^\Delta))\right] \leq \exp(\lambda\mathbb{E}_x\left[f(X_T^\Delta)\right])e^{\sup_{0\leq n\leq N}\log\left(\mathbb{E}_x\left[e^{\frac{C_\sigma\beta}{4}\lambda^2\Delta qp^{2N}(1+C(\Delta)\Delta)^NV^{1-\alpha}(X_{t_n}^\Delta)}\right]\right)}.$$

Setting $p := 1 + C(\Delta)\Delta$, $q = \frac{p}{p-1} = \frac{1+C(\Delta)\Delta}{C(\Delta)\Delta}$ and using the straightforward inequality $(1 + C(\Delta)\Delta)^{3N} \le \exp(3C(\Delta)T)$, we derive

$$\mathbb{E}_x\left[\exp(\lambda f(X_T^\Delta))\right] \le \exp(\lambda \mathbb{E}_x\left[f(X_T^\Delta)\right]) e^{\sup_{0\le n\le N}\log\left(\mathbb{E}_x\left[e^{\frac{C_\sigma\beta(1+C(\Delta))}{4C(\Delta)}e^{3C(\Delta)T}\lambda^2 V^{1-\alpha}(X_{t_n}^\Delta)}\right]\right)}.$$

We set $\varphi(T, b, \sigma, \Delta) := C_\sigma \beta \frac{(1+C(\Delta)\Delta)}{4C(\Delta)} e^{3C(\Delta)T}$. For $\alpha \in (\frac{1}{2}, 1]$, Corollary 3.2 clearly implies

$$\mathbb{E}_x\left[\exp(\lambda f(X_T^\Delta))\right] \le \exp(\lambda \mathbb{E}_x\left[f(X_T^\Delta)\right]) \exp\left(K_{3.1}(\varphi(T, b, \sigma, \Delta) \vee \varphi(T, b, \sigma, \Delta)^{\frac{\alpha}{2\alpha-1}})(\lambda^2 \vee \lambda^{\frac{2\alpha}{2\alpha-1}})\right)$$

and for $\alpha = \frac{1}{2}$, according to Proposition 3.4, for $\lambda < \varphi(T, b, \sigma, \Delta)^{-1/2}\lambda_{3.2}$, one has

$$\mathbb{E}_x\left[\exp(\lambda f(X_T^\Delta))\right] \le \exp(\lambda \mathbb{E}_x\left[f(X_T^\Delta)\right]) \exp\left(K_{3.2}\frac{(\lambda\varphi(T, b, \sigma, \Delta)^{1/2}/\lambda_{3.2})^2}{1 - (\lambda\varphi(T, b, \sigma, \Delta)^{1/2}/\lambda_{3.2})}\right).$$

□

3.2 Proof of Theorem 2.6

We will prove the result for the process X solution of $(SDE_{b,\sigma})$. The proof for the continuous Euler scheme is similar. The following lemma is standard. As will become clear in the discussion below, the only interest it holds is to know the explicit behavior with respect to p. This behavior is optimal as it can be readily checked by taking $X_t = x_0 \exp(\sigma W_t - \sigma^2 t/2)$, $t \in [0, T]$, see also remark 2.7.

Lemma 3.7. *Under the assumptions of Theorem 2.6, for all $p \ge 1$, one has*

$$\mathbb{E}_x[\sup_{0\le t\le T} |X_t|^{2p}] \le (1 + |x|)^{2p} \exp(26p^2(1 + (C_b \vee C_\sigma)T)).$$

Proof. Let $g : x \mapsto \sqrt{1 + |x|^2}$ satisfying for all $x \in \mathbb{R}^d$, $\nabla g(x) = g^{-1}(x)x$, $\nabla^2 g(x) = g^{-1}(x)I_d - g^{-3}(x)xx^*$ and $V : x \mapsto g^{2p}(x)$. We apply Itô's formula to the process $V(X_t)$ with $\nabla V(x) = 2pg(x)^{2p-1}\nabla g(x)$, $\nabla^2 V(x) = 2pg(x)^{2p-1}\nabla^2 g(x) + 2p(2p-1)g(x)^{2p-2}\nabla g(x)\nabla g(x)^*$ noticing that for all $t \in [0, T]$

$$\begin{aligned}
\nabla V(x).b(t, x) + \frac{1}{2}Tr(\sigma^*\nabla^2 V\sigma)(t, x) &\le 2pC_b g(x)^{2p-1}(1 + |x|) + \frac{1}{2}C_\sigma(1 + |x|^2)||\nabla^2 V(x)|| \\
&\le 4pC_b g(x)^{2p} \\
&+ \frac{1}{2}C_\sigma(1 + |x|^2)(4pg(x)^{2p-2} + 2p(2p-1)g(x)^{2p-2}) \\
&\le 4p(C_b \vee C_\sigma)g(x)^{2p} + 2p(C_b \vee C_\sigma)g(x)^{2p} \\
&+ p(2p-1)(C_b \vee C_\sigma)g(x)^{2p} \\
&\le 8p^2(C_b \vee C_\sigma)V(x)
\end{aligned}$$

we clearly obtain,

$$V(X_t^{\tau_m}) \le V(x) + 8p^2(C_b \vee C_\sigma)\int_0^t V(X_s^{\tau_m})ds + \int_0^{t\wedge\tau_m}(\nabla V^*\sigma)(X_s^{\tau_m})dW_s, \tag{3.3}$$

where we classically introduced the stopping time $\tau_m := \inf\{t \ge 0 : |X_t - x| \ge m\}$ for $m \in \mathbb{N}^*$ and $X^{\tau_m} := (X_{t\wedge\tau_m})_{t\ge0}$. The stochastic integral $M_t^m := \int_0^{t\wedge\tau_m}(\nabla V^*\sigma)(X_s^{\tau_m})dW_s$ defines a continuous martingale so that taking expectation in the previous inequality clearly yields

$$\mathbb{E}_x[V(X_t^{\tau_m})] \le V(x) + 8p^2(C_b \vee C_\sigma)\int_0^t \mathbb{E}_x[V(X_s^{\tau_m})]ds.$$

Now, using Gronwall's lemma we derive

$$\forall m \in \mathbb{N}^*, \quad \sup_{t\in[0,T]} \mathbb{E}_x[V(X_t^{\tau_m})] \leq (1+|x|)^{2p} \exp(8p^2(C_b \vee C_\sigma)T)$$

As $\tau_m \to +\infty$ $a.s.$, as $m \to +\infty$ (since $\sup_{s\in[0,t]} |X_s| < +\infty$) using Fatou's lemma, we finally obtain for all $p \geq 1$

$$\sup_{0\leq t\leq T} \mathbb{E}_x[V(X_t)] = \sup_{0\leq t\leq T} \mathbb{E}_x[g(X_t)^{2p}] \leq (1+|x|)^{2p} \exp(8p^2(C_b \vee C_\sigma)T). \tag{3.4}$$

We then observe that Itô's formula also implies

$$\mathbb{E}_x[\sup_{0\leq s\leq t} V(X_t^{\tau_m})] \leq V(x) + 8p^2(C_b \vee C_\sigma) \int_0^t \mathbb{E}_x[\sup_{0\leq u\leq s} V(X_u^{\tau_m})]ds + \mathbb{E}_x[(M_t^m)^*] \tag{3.5}$$

where $(M_t^m)^* := \sup_{0\leq s\leq t} M_s^m$. Combining Jensen's and Doob's inequalities, one clearly gets

$$\begin{aligned}\mathbb{E}_x[(M_t^m)^*]^2 &\leq \mathbb{E}_x[((M_t^m)^*)^2] \leq 4\mathbb{E}_x[(M_t^m)^2] \leq 16p^2 C_\sigma \int_0^t \mathbb{E}_x[g(X_s^{\tau_m})^{4p}]ds \\ &\leq 16p^2 C_\sigma T(1+|x|)^{4p} \exp(32p^2(C_b \vee C_\sigma)T)\end{aligned}$$

where we used $\forall x \in \mathbb{R}^d,\ (\nabla V^* \sigma)^2(x) \leq 4p^2 C_\sigma g(x)^{4p-2}(1+|x|^2) = 4p^2 C_\sigma g(x)^{4p}$ and (3.4) for the last inequality. Consequently, plugging the latter estimate into (3.5), one has for all $t \in [0,T]$

$$\begin{aligned}\mathbb{E}_x[\sup_{0\leq s\leq t} V(X_t^{\tau_m})] &\leq V(x) + 4p(C_\sigma T)^{\frac{1}{2}}(1+|x|)^{2p} \exp(16p^2(C_b \vee C_\sigma)T) \\ &\quad + 8p^2(C_b \vee C_\sigma) \int_0^t \mathbb{E}_x[\sup_{0\leq u\leq s} V(X_u^{\tau_m})]ds \\ &\leq (1+|x|)^{2p}(1+4p(C_\sigma T)^{\frac{1}{2}} \exp(16p^2(C_b \vee C_\sigma)T)) \\ &\quad + 8p^2(C_b \vee C_\sigma) \int_0^t \mathbb{E}_x[\sup_{0\leq u\leq s} V(X_u^{\tau_m})]ds\end{aligned}$$

so that using Gronwall's lemma yields and passing to the limit $m \to +\infty$, for all $p \geq 1$

$$\mathbb{E}_x[\sup_{0\leq t\leq T} |X_t|^{2p}] \leq \mathbb{E}_x[\sup_{0\leq s\leq T} V(X_t)] \leq 2(1+|x|)^{2p} \exp(26p^2(1+(C_b \vee C_\sigma)T)).$$

□

For all real-valued and 1-Lipschitz function f defined on $\mathcal{C}$ and for all $p \geq 1$, one has

$$\begin{aligned}\mathbb{E}_x[|f(X) - \mathbb{E}_x[f(X)]|^{2p}] &= \mathbb{E}_x[|f(X) - f(0) + f(0) - \mathbb{E}_x[f(X)]|^{2p}] \leq 2^{2p}\mathbb{E}_x[||X||_\infty^{2p}] \\ &\leq 2^{2p+1}(1+|x|)^{2p} \exp(26p^2(1+(C_b \vee C_\sigma)T))\end{aligned} \tag{3.6}$$

where we used Lemma 3.7 for the last inequality. Now, combining the Chebyshev and Rosenthal inequalities for independent zero-mean random variables (see e.g. [13]), for all $p \geq 1$, there exists $C_{2p} > 0$ such that

$$\begin{aligned}\mathbb{P}_x\left(\frac{1}{M}|\sum_{k=1}^M f(X^k) - \mathbb{E}_x[f(X)]| \geq r\right) &\leq \frac{\mathbb{E}_x[(\sum_{k=1}^M f(X^k) - \mathbb{E}_x[f(X)])^{2p}]}{r^{2p}M^{2p}} \\ &\leq C_{2p}\frac{\mathbb{E}_x[|f(X) - \mathbb{E}_x[f(X)]|^{2p}]}{r^{2p}M^p} \\ &\leq 2\frac{(2(1+|x|))^{2p}\exp(28p^2(1+(C_b \vee C_\sigma)T))}{r^{2p}M^p} \\ &:= 2\exp(-\varphi(p))\end{aligned}$$

with $\varphi(p) := -\kappa(b,\sigma,T)p^2 + p\log(\frac{r^2M}{(2(1+|x|))^2})$ and where we used for all $p \geq 1$, $C_{2p} \leq (2p)^{2p} \leq \exp(2p^2)$, see e.g. p.235-236 in [13], and (3.6) for the last inequality. Optimizing the latter inequality with respect to p with $p \geq 1$, i.e. selecting $p = \frac{1}{2\kappa(b,\sigma,T)}\log(\frac{r^2M}{(2(1+|x|))^2})$, we obtain

$$\mathbb{P}_x\left(\frac{1}{M}|\sum_{k=1}^{M} f(X^k) - \mathbb{E}_x[f(X)]| \geq r\right) \leq 2\exp\left(-\frac{1}{4\kappa(b,\sigma,T)}\log\left(\frac{r^2M}{(2(1+|x|))^2}\right)^2\right)$$

for $r^2M \geq (2(1+|x|))^2\exp(2\kappa(b,\sigma,T))$. Otherwise, using the Jensen and Rosenthal inequalities, one has for all $p \in [0,1]$

$$\begin{aligned}\mathbb{E}_x[(\sum_{k=1}^{M} f(X^k) - \mathbb{E}_x[f(X)])^{2p}] &\leq \mathbb{E}_x[(\sum_{k=1}^{M} f(X^k) - \mathbb{E}_x[f(X)])^2]^p \leq \left(MC_2\mathbb{E}_x[|f(X) - \mathbb{E}_x[f(X)]|^2]\right)^p \\ &\leq M^p\left(4(2(1+|x|))^2\exp(\kappa(b,\sigma,T))\right)^p\end{aligned}$$

where we used (3.6) for the last inequality. Now, noticing that we have $4e \leq \exp(\kappa(b,\sigma,T))$, Chebyshev's inequality yields

$$\mathbb{P}_x\left(\frac{1}{M}|\sum_{k=1}^{M} f(X^k) - \mathbb{E}_x[f(X)]| \geq r\right) \leq \frac{C^p}{r^{2p}M^p} \leq 2\frac{(Cp)^p}{r^{2p}M^p} \leq 2\exp(-\varphi(p))$$

with $\varphi(p) := -p\log(p) + p\log(\frac{r^2M}{C})$, $C := (2(1+|x|))^2\exp(2\kappa(b,\sigma,T) - 1)$ and where we used that for all $p \geq 0$, $C^p \leq 2(Cp)^p$ since the function $p \mapsto 2p^p$ is minimized for $p = \exp(-1)$ and $2\exp(-1/e)) > 1$. Consequently, optimizing over p such that $p \leq 1$, i.e. selecting $p = \frac{r^2M}{Ce}$, one has

$$\mathbb{P}_x\left(\frac{1}{M}|\sum_{k=1}^{M} f(X^k) - \mathbb{E}_x[f(X)]| \geq r\right) \leq 2\exp\left(-\frac{r^2M}{(2(1+|x|))^2\exp(2\kappa(b,\sigma,T))}\right)$$

for $r^2M \leq Ce = (2(1+|x|))^2\exp(2\kappa(b,\sigma,T))$. This completes the proof.

4 Stochastic Approximation Algorithm: Proof of the main Results

Throughout this section we will assume that **(HL)**, **(HLS)**$_\alpha$ and **(HUA)** are in force.

4.1 Proof of Theorem 2.8

The proof of Theorem 2.8 is divided into several propositions.

Proposition 4.1. *Denote by* $(\theta_n^\gamma)_{0\leq n\leq N}$ *the scheme* (1.4) *with step sequence* $\gamma = (\gamma_n)_{0\leq n\leq N}$ *satisfying* (1.5). *Assume that the innovations* $(U_i)_{i\geq 1}$ *of* (1.4) *satisfy* $(GC(\beta))$ *for some* $\beta > 0$. *Then, there exists* $\varepsilon_\beta > 0$ *which only depends on the law* μ *such that for all* $\lambda < \min(1, \varepsilon_\beta(8\eta\alpha C_\alpha^2\Pi_{2,N})^{-1})$, *one has*

$$\begin{aligned}\sup_{0\leq n\leq N}\log\left(\mathbb{E}_{\theta_0}\left[e^{\lambda L^\alpha(\theta_n^\gamma)}\right]\right) &\leq (L^\alpha(\theta_0) + \underline{C}\sum_{k=0}^{N-1}\gamma_{k+1}^2)\Pi_{2,N}\lambda \\ &\quad + \left(\frac{1}{2}\sum_{k=0}^{N-1}\gamma_{k+1}^2\right)\log\left(\mathbb{E}\left[e^{8\eta\alpha C_\alpha^2\Pi_{2,N}\lambda|U|^2}\right]\right).\end{aligned}$$

with $\Pi_{2,N} = \Pi_{2,N}(\alpha) := \prod_{k=0}^{N-1}(1 + (2\eta\alpha C_h + \frac{\beta}{2}\alpha^2C_\alpha^2)\gamma_{k+1}^2)$ *and* $\underline{C} = 4\eta\alpha C_\alpha^2\mathbb{E}[|U|^2]$.

Proof. The proof relies on similar arguments as those used in the proof of Proposition 3.1. Using the concavity of $x \mapsto x^\alpha$, $\alpha \in (0,1]$, a Taylor expansion of order 2 of the function L, and finally **(HLS)**$_\alpha$, for all $k \in \{0, \cdots, N-1\}$, we have

$$\begin{aligned} L^\alpha(\theta_{k+1}^\gamma) - L^\alpha(\theta_k^\gamma) &\leq \alpha L^{\alpha-1}(\theta_k^\gamma)\left(\nabla L(\theta_k^\gamma).(\theta_{k+1}^\gamma - \theta_k^\gamma) + \eta|\theta_{k+1}^\gamma - \theta_k^\gamma|^2\right), \\ &= -\gamma_{k+1}\alpha L^{\alpha-1}(\theta_k^\gamma)\left\langle \nabla L(\theta_k^\gamma), h(\theta_k^\gamma)\right\rangle \\ &- \gamma_{k+1}\alpha L^{\alpha-1}(\theta_k^\gamma)\left\langle \nabla L(\theta_k^\gamma), H(\theta_k^\gamma, U_{k+1}) - h(\theta_k^\gamma)\right\rangle \\ &+ \alpha\eta\gamma_{k+1}^2 L^{\alpha-1}(\theta_k^\gamma)|H(\theta_k^\gamma, U_{k+1})|^2, \\ &\leq -\gamma_{k+1}\alpha L^{\alpha-1}(\theta_k^\gamma)\left\langle \nabla L(\theta_k^\gamma), H(\theta_k^\gamma, U_{k+1}) - h(\theta_k^\gamma)\right\rangle \\ &+ 2\eta\alpha\gamma_{k+1}^2 L^{\alpha-1}(\theta_k^\gamma)|H(\theta_k^\gamma, U_{k+1}) - h(\theta_k^\gamma)|^2 + 2\eta\alpha\gamma_{k+1}^2 L^{\alpha-1}(\theta_k^\gamma)|h(\theta_k^\gamma)|^2. \end{aligned}$$

Let us note that **(HLS)**$_\alpha$ implies that $\forall(\theta, u) \in \mathbb{R}^d \times \mathbb{R}^q$, $|H(\theta,u) - h(\theta)|^2 = |H(\theta,u) - \mathbb{E}[H(\theta,U)]|^2 \leq 2C_\alpha^2 L^{1-\alpha}(\theta)(\mathbb{E}[|U|^2] + |u|^2)$ which leads to

$$\begin{aligned} L^\alpha(\theta_{k+1}^\gamma) - L^\alpha(\theta_k^\gamma) &\leq -\gamma_{k+1}\alpha L^{\alpha-1}(\theta_k^\gamma)\left\langle \nabla L(\theta_k^\gamma), H(\theta_k^\gamma, U_{k+1}) - h(\theta_k^\gamma)\right\rangle + 4\eta\alpha C_\alpha^2\gamma_{k+1}^2\mathbb{E}[|U|^2] \\ &+ 4\eta\alpha C_\alpha^2\gamma_{k+1}^2|U_{k+1}|^2 + 2\eta\alpha C_h\gamma_{k+1}^2 L^\alpha(\theta_k^\gamma). \end{aligned}$$

Using again **(HLS)**$_\alpha$, $\forall\theta \in \mathbb{R}^d$ the functions $g(\theta,.) : u \mapsto \frac{\langle \nabla L(\theta), H(\theta,u) - h(\theta)\rangle}{L^{1-\alpha}(\theta)}$ are Lipschitz and more precisely satisfy

$$\forall\theta \in \mathbb{R}^d, \quad \sup_{(u,u')\in(\mathbb{R}^q)^2} \frac{|g(\theta,u) - g(\theta,u')|}{|u-u'|} \leq C_\alpha L^{\frac{\alpha}{2}}(\theta).$$

Consequently, denoting $\underline{C} = 4\eta\alpha C_\alpha^2\mathbb{E}[|U|^2]$, from the Cauchy-Schwarz inequality and since the law of the innovation satisfies ($GC(\beta)$) for some $\beta > 0$, there exists $\epsilon_\beta > 0$ such that for $\lambda < \min(1, \varepsilon_\beta(8\eta\alpha C_\alpha^2\gamma_1^2)^{-1})$, one has

$$\begin{aligned} \mathbb{E}\left[\exp(\lambda L^\alpha(\theta_{k+1}^\gamma))\middle|\, \mathcal{F}_k\right] &\leq \exp(\lambda(1 + 2\eta\alpha C_h\gamma_{k+1}^2)L^\alpha(\theta_k^\gamma))\exp(\underline{C}\gamma_{k+1}^2\lambda) \\ &\times \mathbb{E}\left[\exp(-2\alpha\lambda\gamma_{k+1}g(\theta_k^\gamma, U_{k+1}))\middle|\, \mathcal{F}_k\right]^{\frac{1}{2}}\mathbb{E}\left[\exp(8\eta\alpha\lambda C_\alpha^2\gamma_{k+1}^2|U_{k+1}|^2)\middle|\, \mathcal{F}_k\right]^{\frac{1}{2}} \\ &\leq \exp(\lambda(1 + (2\eta\alpha C_h + \frac{\beta}{2}C_\alpha^2\alpha)\gamma_{k+1}^2)L^\alpha(\theta_k^\gamma))\exp(\underline{C}\gamma_{k+1}^2\lambda) \\ &\times \mathbb{E}\left[\exp(8\eta\alpha\lambda C_\alpha^2\gamma_{k+1}^2|U|^2)\right]^{\frac{1}{2}} \end{aligned}$$

In the aim of simplifying notations, we define $\Pi_{2,n} := \prod_{k=0}^{n-1}(1 + (2\eta\alpha C_h + \frac{\beta}{2}C_\alpha^2\alpha)\gamma_{k+1}^2)$ and temporarily set $L_k := L^\alpha(\theta_k^\gamma)/\Pi_{2,k}$, for $k \in \{0, \cdots, N\}$. Taking expectation in both sides of the previous inequality clearly implies

$$\mathbb{E}_{\theta_0}\left[\exp(\lambda L_{k+1})\right] \leq \mathbb{E}_{\theta_0}\left[\exp(\lambda L_k)\right]\exp\left(\underline{C}\frac{\gamma_{k+1}^2}{\Pi_{2,k+1}}\lambda\right)\mathbb{E}\left[\exp\left(8\eta\alpha C_\alpha^2\frac{\gamma_{k+1}^2}{\Pi_{2,k+1}}\lambda|U|^2\right)\right]^{\frac{1}{2}}$$

and by a straightforward induction, for $n \in \{0, \cdots, N\}$ we have

$$\mathbb{E}_{\theta_0}\left[\exp(\lambda L_n)\right] \leq \exp(\lambda L_0)\exp\left(\underline{C}\sum_{k=0}^{n-1}\frac{\gamma_{k+1}^2}{\Pi_{2,k+1}}\lambda\right)\prod_{k=0}^{n-1}\mathbb{E}\left[\exp\left(8\eta\alpha C_\alpha^2\frac{\gamma_{k+1}^2}{\Pi_{2,k+1}}\lambda|U|^2\right)\right]^{\frac{1}{2}},$$

which finally yields for $\lambda < \min(1, \varepsilon_\beta(8\eta\alpha C_\alpha^2\gamma_1^2)^{-1})$

$$\begin{aligned} \mathbb{E}_{\theta_0}\left[\exp(\lambda L^\alpha(\theta_n^\gamma))\right] &\leq \exp(\Pi_{2,n}L^\alpha(\theta_0)\lambda)\exp\left(\underline{C}\sum_{k=0}^{n-1}\frac{\Pi_{2,n}}{\Pi_{2,k+1}}\gamma_{k+1}^2\lambda\right) \\ &\times \prod_{k=0}^{n-1}\mathbb{E}\left[\exp\left(8\eta\alpha C_\alpha^2\frac{\Pi_{2,n}}{\Pi_{2,k+1}}\gamma_{k+1}^2\lambda|U|^2\right)\right]^{\frac{1}{2}}. \end{aligned}$$

Up to a modification of a constant, we can assume without loss of generality that $\sup_{0\le n\le N}\gamma_{n+1} = \gamma_1 \le 1$ so that using the Jensen's inequality, the latter bound clearly provides the following control of the quantity of interest for $\lambda < \min(1, \varepsilon_\beta(8\eta\alpha C_\alpha^2\Pi_{2,N})^{-1})$

$$\sup_{0\le n\le N}\log\left(\mathbb{E}_{\theta_0}\left[e^{\lambda L^\alpha(\theta_n^\gamma)}\right]\right) \le \left(L^\alpha(\theta_0) + \underline{C}\sum_{k=0}^{N-1}\gamma_{k+1}^2\right)\Pi_{2,N}\lambda + \left(\frac{1}{2}\sum_{k=0}^{N-1}\gamma_{k+1}^2\right)\log\left(\mathbb{E}\left[e^{8\eta\alpha C_\alpha^2\Pi_{2,N}\lambda|U|^2}\right]\right).$$

□

Corollary 4.2. *Assume that the assumptions of Proposition 4.1 are satisfied. Then, there exists a constant $K_{4.1}$ depending on γ,α,θ_0,H such that for all $\alpha\in(\frac{1}{2},1]$, one has*

$$\forall\lambda\ge 0,\quad \sup_{0\le n\le N}\log\left(\mathbb{E}_{\theta_0}\left[\exp(\lambda L^{1-\alpha}(\theta_n^\gamma))\right]\right) \le K_{4.1}(\lambda\vee\lambda^{\frac{\alpha}{2\alpha-1}}).$$

Remark 4.3. *The constant $K_{4.1}$ can be explicitly computed. Indeed, one has $K_{4.1} := \max(\Psi_1(\gamma,\alpha,\theta_0,H),\Psi_2(\gamma,\alpha,\theta_0,H))$ with*

$$\begin{aligned}\Psi_1(\gamma,\alpha,\theta_0,H) &= \left(L^{1-\alpha}(\theta_0) + (8\eta\alpha C_\alpha^2\mathbb{E}[|U|^2]\sum_{k=0}^{N-1}\gamma_{k+1}^2)^{\frac{1-\alpha}{\alpha}}\right)\prod_{k=0}^{N-1}(1+2\eta(1-\alpha)C_h\gamma_{k+1}^2)\\
&\quad + e^{\frac{2\alpha-1}{\alpha}\overline{\rho}^{-\frac{1-\alpha}{2\alpha-1}}}\\
&\quad \times e^{\left(L^\alpha(\theta_0)+2\alpha\underline{C}\sum_{k=0}^{N-1}\gamma_{k+1}^2\right)\Pi_{2,N}\overline{\rho}\frac{1-\alpha}{\alpha}+\left(\frac{1}{2}\sum_{k=0}^{N-1}\gamma_{k+1}^2\right)\log\left(\mathbb{E}\left[e^{\frac{\varepsilon_\beta(1-\alpha)}{2\alpha}|U|^2}\right]\right)}\\
\Psi_2(\gamma,\alpha,\theta_0,H) &= \frac{2\alpha-1}{\alpha}\overline{\rho}^{-\frac{1-\alpha}{2\alpha-1}} + \left(L^\alpha(\theta_0)+\underline{C}\sum_{k=0}^{N-1}\gamma_{k+1}^2\right)\Pi_{2,N}\overline{\rho}\frac{1-\alpha}{\alpha}\\
&\quad + \left(\frac{1}{2}\sum_{k=0}^{N-1}\gamma_{k+1}^2\right)\log\left(\mathbb{E}\left[e^{\frac{\varepsilon_\beta(1-\alpha)}{2\alpha}|U|^2}\right]\right)\\
\overline{\rho} &= \frac{1}{2}\min(1,\varepsilon_\beta(8\eta\alpha C_\alpha^2\Pi_{2,N})^{-1})\end{aligned}$$

Proof. We only give a sketch of proof since it is rather similar to the one of Corollary 3.2. For $\lambda\in[0,1]$, one has

$$\mathbb{E}_{\theta_0}[\exp\left(\lambda L^{1-\alpha}(\theta_n^\gamma)\right)] \le \exp\left(\lambda(\mathbb{E}_{\theta_0}[L^{1-\alpha}(\theta_n^\gamma)] + \mathbb{E}_{\theta_0}[\exp(L^{1-\alpha}(\theta_n^\gamma))]\right).$$

Tedious but simple computations in the spirit of Proposition 4.1 easily show that

$$\begin{aligned}\sup_{0\le n\le N}\mathbb{E}_{\theta_0}[L^{1-\alpha}(\theta_n^\gamma)] \le \sup_{0\le n\le N}\mathbb{E}_{\theta_0}[L^\alpha(\theta_n^\gamma)]^{\frac{1-\alpha}{\alpha}} &\le \left(L^{1-\alpha}(\theta_0) + (8\eta\alpha C_\alpha^2\mathbb{E}[|U|^2]\sum_{k=0}^{N-1}\gamma_{k+1}^2)^{\frac{1-\alpha}{\alpha}}\right)\\
&\quad\times\prod_{k=0}^{N-1}(1+2\eta(1-\alpha)C_h\gamma_{k+1}^2).\end{aligned}$$

Moreover, thanks to the Young inequality $L^{1-\alpha}(\theta) \le \frac{1-\alpha}{\alpha}\rho L^\alpha(\theta) + \frac{2\alpha-1}{\alpha}\rho^{-\frac{1-\alpha}{2\alpha-1}}$, for every $(\rho,\theta)\in\mathbb{R}_+^*\times\mathbb{R}^d$ and $\alpha\in(\frac{1}{2},1]$ and using Proposition 4.1, one obtains for $\rho=\overline{\rho} := \frac{1}{2}\min(1,\varepsilon_\beta(8\eta\alpha C_\alpha^2\Pi_{2,N})^{-1})$

$$\begin{aligned}\sup_{0\le n\le N}\mathbb{E}_{\theta_0}[\exp(L^{1-\alpha}(\theta_n^\gamma))] \le \exp\left(\frac{2\alpha-1}{\alpha}\overline{\rho}^{-\frac{1-\alpha}{2\alpha-1}}\right)\exp\Bigg(&\left(L^\alpha(\theta_0)+\underline{C}\sum_{k=0}^{N-1}\gamma_{k+1}^2\right)\Pi_{2,N}\overline{\rho}\frac{1-\alpha}{\alpha}\\
&+\left(\frac{1}{2}\sum_{k=0}^{N-1}\gamma_{k+1}^2\right)\log\left(\mathbb{E}\left[e^{\frac{\varepsilon_\beta(1-\alpha)}{2\alpha}|U|^2}\right]\right)\Bigg),\end{aligned}$$

so that for all $\lambda \in [0,1]$

$$\mathbb{E}_{\theta_0}[\exp\left(\lambda L^{1-\alpha}(\theta_n^\gamma)\right)] \leq \Psi_1(\gamma,\alpha,\theta_0,H)\lambda.$$

Now, for $\lambda > 1$, we use the Young inequality $\lambda L^{1-\alpha}(\theta_n^\gamma) \leq \frac{2\alpha-1}{\alpha}\rho^{-\frac{1-\alpha}{2\alpha-1}}\lambda^{\frac{\alpha}{2\alpha-1}} + \frac{1-\alpha}{\alpha}\rho L^{\alpha}(\theta_n^\gamma)$ to derive

$$\mathbb{E}_{\theta_0}[\exp(\lambda L^{1-\alpha}(\theta_n^\gamma))] \leq \exp\left(K\lambda^{\frac{\alpha}{2\alpha-1}}\right)$$

with $K(\rho) := \frac{2\alpha-1}{\alpha}\rho^{-\frac{1-\alpha}{2\alpha-1}} + \log\mathbb{E}_{\theta_0}\left[\exp\left(\left(\frac{1-\alpha}{\alpha}\right)\rho L^{\alpha}(\theta_n^\gamma)\right)\right]$ and $\frac{1-\alpha}{\alpha}\rho < \min(1, \varepsilon_\beta(8\eta\alpha C_\alpha^2\Pi_{2,N})^{-1})$. We select $\rho = \overline{\rho}$ in the last inequality and use Proposition 4.1 to bound the quantity $K(\overline{\rho})$. □

Proposition 4.4. *(Control of the Lipschitz modulus of iterative kernels) Denote by P_k and $P_{k,p} = P_k \circ \cdots \circ P_{p-1}$, $k,\ p \in \{0,\cdots,N-1\}$, $k \leq p$ the (Feller) transition kernel and the iterative kernels of the Markov chain defined by the scheme* (1.4)*. Then for all Lipschitz function f and for all $k,\ p \in \{0,\cdots,N-1\}$, $k \leq p$ the functions $P_{k,p}(f)$ are Lipschitz-continuous and one has*

$$[P_{k,p}(f)]_1 := \sup_{(\theta,\theta')\in(\mathbb{R}^d)^2} \frac{|P_{k,p}(f)(\theta) - P_{k,p}(f)(\theta')|}{|\theta-\theta'|} \leq [f]_1 \prod_{i=k}^{p-1}(1 - 2\underline{\lambda}\gamma_{i+1} + C_{H,\mu}\gamma_{i+1}^2)^{\frac{1}{2}}$$

where $[f]_1$ stands for the Lipschitz modulus of the function f and $C_{H,\mu} := 2C_H^2(1+\mathbb{E}[|U|^2])$.

Proof. Using the Cauchy-Schwarz inequality, **(HUA)** then **(HL)**, for all $(\theta,\theta') \in (\mathbb{R}^d)^2$, one has

$$\begin{aligned}
|P_k(f)(\theta) - P_k(f)(\theta')| &\leq \mathbb{E}\left[|f(\theta - \gamma_{k+1}H(\theta,U_{k+1})) - f(\theta' - \gamma_{k+1}H(\theta',U_{k+1}))|\right] \\
&\leq [f]_1\mathbb{E}\left[|\theta-\theta' - \gamma_{k+1}(H(\theta,U_{k+1}) - H(\theta',U_{k+1}))|^2\right]^{\frac{1}{2}} \\
&\leq [f]_1\left(|\theta-\theta'|^2 - 2\gamma_{k+1}\langle\theta-\theta', h(\theta)-h(\theta')\rangle\right. \\
&\quad \left.+\gamma_{k+1}^2\mathbb{E}\left[|H(\theta,U_{k+1}) - H(\theta',U_{k+1})|^2\right]\right)^{\frac{1}{2}} \\
&\leq [f]_1(1 - 2\underline{\lambda}\gamma_{k+1} + 2C_H^2(1+\mathbb{E}[|U|^2])\gamma_{k+1}^2)^{\frac{1}{2}}|\theta-\theta'|.
\end{aligned}$$

A straightforward induction argument completes the proof.

□

Proposition 4.5. *(Control of the Laplace transform) Denote by θ_N^γ the value at step N of the stochastic approximation algorithm* (1.4) *with step sequence $\gamma := (\gamma_n)_{n\geq1}$ satisfying* (1.5)*. Assume that the innovations $(U_n)_{n\geq1}$ in* (1.4) *satisfy* ($GC(\beta)$) *for some $\beta > 0$. Let f be a real-valued 1-Lipschitz-continuous function defined on $\mathbb{R}^d$. Then, for all $\lambda \geq 0$, for all $N \geq 1$, for all $\alpha \in (\frac{1}{2},1]$, one has*

$$\mathbb{E}_{\theta_0}[e^{\lambda f(\theta_N^\gamma)}] \leq e^{\mathbb{E}_{\theta_0}[\lambda f(\theta_N^\gamma)]}e^{\varphi_\alpha(\gamma,H,\theta_0)(C_N^\gamma\lambda^2 \vee C_N^{\gamma,\alpha}\lambda^{\frac{2\alpha}{2\alpha-1}})}$$

with the two concentration rates $C_N^\gamma := \sum_{k=0}^{N-1}\gamma_{k+1}^2\frac{\Pi_{1,N}}{\Pi_{1,k}}$, with $\Pi_{1,N} := \prod_{k=0}^{N-1}(1-2\underline{\lambda}\gamma_{k+1} + C_{H,\mu}\gamma_{k+1}^2)$ and $C_N^{\gamma,\alpha} := \sum_{k=0}^{N-1}\gamma_{k+1}^{\frac{2\alpha}{2\alpha-1}}(\frac{\Pi_{1,N}}{\Pi_{1,k}})^{\frac{2\alpha}{2\alpha-1}}((k+1)\log^2(k+4))^{\frac{1-\alpha}{2\alpha-1}}$ for all $N \geq 1$ and where $\varphi_\alpha(\gamma,H,\theta_0) := K_{4.1}2^{\frac{1-\alpha}{2\alpha-1}}\frac{\beta C_\alpha^2}{4} \vee (\frac{\beta C_\alpha^2}{4})^{\frac{\alpha}{2\alpha-1}}\exp\left(\frac{1}{2\alpha-1}\sum_{k=0}^{N-1}\frac{1}{(k+1)\log^2(k+4)}\right)$.

If $\alpha = \frac{1}{2}$, then there exists two positive constants $\lambda_{4.1}$ and $\varphi_{1/2}(\gamma,H,\theta_0)$ such that

$$\forall\lambda \in [0,\lambda_{4.1}/\tilde{s}_N),\ \mathbb{E}_{\theta_0}[e^{\lambda f(\theta_N^\gamma)}] \leq e^{\lambda\mathbb{E}_{\theta_0}[f(\theta_N^\gamma)]}e^{2\varphi_{1/2}(\gamma,H,\theta_0)C_N^\gamma\frac{(\lambda/\lambda_{4.1})^2}{1-(\lambda\tilde{s}_N/\lambda_{4.1})}}$$

with $\tilde{s}_N := \max_{0\leq k\leq N-1}(k+1)^{1/2}\log(k+4)\gamma_{k+1}\left(\frac{\Pi_{1,N}}{\Pi_{1,k}}\right)^{\frac{1}{2}}e^{\sum_{p=0}^{N-1}\frac{1}{(p+1)\log^2(p+4)}}$.

Proof. The proof relies on similar arguments as those used for the proof of Proposition 3.6. For $\lambda \geq 0$ and $k \in \{0, \cdots, N-1\}$, one has

$$P_k(\exp(\lambda f))(\theta) \leq \exp\left(\lambda P_k(f) + \frac{\lambda^2}{4}\beta\gamma_{k+1}^2[f]_1^2 C_\alpha^2 L^{1-\alpha}(\theta)\right)$$

Taking expectation on both sides of the last inequality with $\theta = \theta_k^\gamma$ and applying the Hölder inequality with conjugate exponents (p_k, q_k) (to be fixed later on), one obtains

$$\mathbb{E}_{\theta_0}[\exp(\lambda f(\theta_k^\gamma))] \leq \mathbb{E}_{\theta_0}\left[\exp\left(\lambda p_k P_k(f)(\theta_k^\gamma)\right)\right]^{\frac{1}{p_k}} \mathbb{E}_{\theta_0}\left[\exp\left(q_k \frac{\lambda^2}{4}\beta\gamma_{k+1}^2[f]_1^2 C_\alpha^2 L^{1-\alpha}(\theta_k^\gamma)\right)\right]^{\frac{1}{q_k}}$$

and applying the last inequality to $f := P_{k+1,N}(f)$ yields

$$\mathbb{E}_{\theta_0}[\exp(\lambda P_{k+1,N}(f)(\theta_k^\gamma))] \leq \mathbb{E}_{\theta_0}\left[\exp\left(\lambda p_k P_{k,N}(f)(\theta_k^\gamma)\right)\right]^{\frac{1}{p_k}}$$
$$\times \mathbb{E}_{\theta_0}\left[\exp\left(q_k \frac{\lambda^2}{4}\beta\gamma_{k+1}^2[P_{k+1,N}(f)]_1^2 C_\alpha^2 L^{1-\alpha}(\theta_k^\gamma)\right)\right]^{\frac{1}{q_k}}. \quad (4.1)$$

We use Corollary 4.2 to obtain for $\alpha \in (\frac{1}{2}, 1]$

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left[\exp\left(q_k \frac{\lambda^2}{4}\beta\gamma_{k+1}^2[P_{k+1,N}(f)]_1^2 C_\alpha^2 L^{1-\alpha}(\theta_k^\gamma)\right)\right]^{\frac{1}{q_k}} \\
&\leq \exp\left(K_{4.1}\frac{\beta C_\alpha^2}{4} \vee \left(\frac{\beta C_\alpha^2}{4}\right)^{\frac{\alpha}{2\alpha-1}} (\gamma_{k+1}^2[P_{k+1,N}(f)]_1^2\lambda^2 \vee \gamma_{k+1}^{\frac{2\alpha}{2\alpha-1}}[P_{k+1,N}(f)]_1^{\frac{2\alpha}{2\alpha-1}} q_k^{\frac{1-\alpha}{2\alpha-1}}\lambda^{\frac{2\alpha}{2\alpha-1}})\right) \\
&:= f_k(\lambda)
\end{aligned}$$

Now, an elementary induction argument leads to

$$\begin{aligned}
\mathbb{E}_{\theta_0}[\exp(\lambda f(\theta_N^\gamma))] &= \mathbb{E}_{\theta_0}[\exp(\lambda P_{N,N} f(\theta_N^\gamma))] \\
&\leq \mathbb{E}_{\theta_0}[\exp(\lambda \prod_{k=0}^{N-1} p_k P_{0,N}(f)(\theta_0))]^{\frac{1}{\Pi_{k=0}^{N-1} p_k}} \prod_{k=0}^{N-1} f_{N-1-k}\left(\lambda \prod_{i=1}^{k} p_{N-i}\right)^{\frac{1}{\Pi_{i=1}^{k} p_{N-i}}}
\end{aligned} \quad (4.2)$$

We select $p_k := 1 + \frac{1}{(k+1)\log^2(k+4)}$, $q_k = (1 + \frac{1}{(k+1)\log^2(k+4)})(k+1)\log^2(k+4) \leq 2(k+1)\log^2(k+4)$, $k = 0, \cdots, N-1$ so that $\prod_{k=0}^{N-1} p_k$ converges and more precisely we have $\prod_{k=0}^{N-1} p_k < \exp(\sum_{k=0}^{N-1}\frac{1}{(k+1)\log^2(k+4)}) < \infty$.
We set $\varphi_\alpha(\gamma, H, \theta_0) := K_{4.1} 2^{\frac{1-\alpha}{2\alpha-1}} \frac{\beta C_\alpha^2}{4} \vee (\frac{\beta C_\alpha^2}{4})^{\frac{\alpha}{2\alpha-1}} \exp\left(\frac{1}{2\alpha-1}\sum_{k=0}^{N-1}\frac{1}{(k+1)\log^2(k+4)}\right)$. Now, using Proposition 4.4 and Corollary 4.2, we easily derive from (4.2)

$$\forall\lambda \geq 0, \quad \mathbb{E}_{\theta_0}[\exp(\lambda f(\theta_N^\gamma))] \leq \exp\left(\mathbb{E}_{\theta_0}[\lambda f(\theta_N^\gamma)]\right)\exp\left(\varphi_\alpha(\gamma, H, \theta_0)(C_N^\gamma\lambda^2 \vee C_N^{\gamma,\alpha}\lambda^{\frac{2\alpha}{2\alpha-1}})\right)$$

with $C_N^{\gamma,\alpha} := \sum_{k=0}^{N-1}\gamma_{k+1}^{\frac{2\alpha}{2\alpha-1}}(\frac{\Pi_{1,N}}{\Pi_{1,k}})^{\frac{2\alpha}{2\alpha-1}}((k+1)\log^2(k+4))^{\frac{1-\alpha}{2\alpha-1}}$.

For $\alpha = \frac{1}{2}$, we start from (4.1). First, we use the control obtained in Proposition 4.1 to derive

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left[\exp\left(q_k\frac{\lambda^2}{4}\beta\gamma_{k+1}^2[P_{k+1,N}(f)]_1^2 C_{1/2}^2 L^{\frac{1}{2}}(\theta_k^\gamma)\right)\right]^{\frac{1}{q_k}} \\
&\leq \exp\left(\left(L^{\frac{1}{2}}(\theta_0) + \underline{C}\sum_{p=0}^{N-1}\gamma_{p+1}^2\right)\Pi_{2,N}(1/2)\frac{\beta C_{1/2}^2}{4}\gamma_{k+1}^2[P_{k+1,N}(f)]_1^2\lambda^2 + \frac{1}{q_k}\left(\frac{1}{2}\sum_{p=0}^{N-1}\gamma_{p+1}^2\right)\right. \\
&\left.\times\ \log\mathbb{E}\left[\exp\left(\beta\eta C_{1/2}^4\Pi_{2,N}(1/2)\,q_k\gamma_{k+1}^2[P_{k+1,N}(f)]_1^2\lambda^2|U|^2\right)\right]\right).
\end{aligned}$$

To simplify the latter bound, that is to obtain an explicit and computable formula for the second term appearing in the right hand side, we will need the following lemma:

Lemma 4.6. *For all* $\lambda \in [0, \lambda_{4.1}/s_N^{1/2})$*, one has*

$$\log \mathbb{E}\left[e^{\beta\eta C_{1/2}^4 \Pi_{2,N}(1/2) q_k \gamma_{k+1}^2 [P_{k+1,N}(f)]_1^2 \lambda^2 |U|^2}\right] \leq \beta\eta C_{1/2}^4 \Pi_{2,N}\left(1/2\right) \mathbb{E}[|U|^2] q_k \gamma_{k+1}^2 [P_{k+1,N}(f)]_1^2 \lambda^2 + 2 q_k \gamma_{k+1}^2 [P_{k+1,N}(f)]_1^2 \frac{(\lambda/\lambda_{4.1})^2}{1-(\lambda s_N^{1/2}/\lambda_{4.1})},$$

with $s_N := \max_{0\leq k\leq N-1} q_k \gamma_{k+1}^2 \frac{\Pi_{1,N}}{\Pi_{1,k}}$ *and* $\lambda_{4.1}$ *satisfies* $\mathbb{E}[\exp(\lambda_{4.1}^2 \beta\eta C_{1/2}^4 \Pi_{2,N}\left(1/2\right)|U|^2)] \leq 2$.

Proof. The proof is similar to the proof of Corollary 3.4. By definition of $\lambda_{4.1}$, we have $\lambda_{4.1}^{2p}(\beta\eta C_{1/2}^4 \Pi_{2,N}(1/2)/2)^p \mathbb{E}[|U|^{2p}] \leq 2p!$, $\forall p \geq 1$. Hence, setting $C_1 := \beta\eta C_{1/2}^4 \Pi_{2,N}\left(1/2\right)$ we easily deduce for $\lambda < \lambda_{4.1}/s_N^{1/2}$,

$$\begin{aligned}
\log \mathbb{E}\left[e^{\lambda^2 C_1 q_k \gamma_{k+1}^2 [P_{k+1,N}(f)]_1^2 |U|^2}\right] &- \lambda^2 C_1 q_k \gamma_{k+1}^2 [P_{k+1,N}(f)]_1^2 \mathbb{E}[|U|^2] \\
&\leq \sum_{p\geq 2} \frac{\lambda^{2p} C_1^p (q_k \gamma_{k+1}^2 [P_{k+1,N}(f)]_1^2)^p \mathbb{E}[|U|^{2p}]}{p!} \\
&\leq 2\sum_{p\geq 2}\left(\frac{\lambda^2 q_k \gamma_{k+1}^2 [P_{k+1,N}(f)]_1^2}{\lambda_{4.1}^2}\right)^p \\
&\leq 2 q_k \gamma_{k+1}^2 [P_{k+1,N}(f)]_1^2 \frac{(\lambda/\lambda_{4.1})^2}{1-(\lambda s_N^{1/2}/\lambda_{4.1})}
\end{aligned}$$

This completes the proof. □

Using the previous lemma, we obtain for all $\lambda \in [0, \lambda_{4.1}/s_N^{1/2})$,

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left[\exp\left(q_k \frac{\lambda^2}{4} \beta \gamma_{k+1}^2 [P_{k+1,N}(f)]_1^2 C_{1/2}^2 L^{\frac{1}{2}}(\theta_k^\gamma)\right)\right]^{\frac{1}{q_k}} \\
&\quad \leq \exp\left(\Psi(N,\gamma,\theta_0)\gamma_{k+1}^2 [P_{k+1,N}(f)]_1^2 \lambda^2 + \left(\sum_{p=0}^{N-1} \gamma_{p+1}^2\right) \gamma_{k+1}^2 [P_{k+1,N}(f)]_1^2 \frac{(\lambda/\lambda_{4.1})^2}{1-(\lambda s_N^{1/2}/\lambda_{4.1})}\right),
\end{aligned}$$

where we introduced the notation $\Psi(N,\gamma,\theta_0) := \left(L^{\frac{1}{2}}(\theta_0) + \underline{C}\sum_{p=0}^{N-1}\gamma_{p+1}^2\right)\Pi_{2,N}\left(1/2\right)\frac{\beta C_{1/2}^2}{4} + \beta\eta C_{1/2}^4 \Pi_{2,N}\left(1/2\right)\mathbb{E}[|U|^2]$.

Now, as for $\alpha \in (\frac{1}{2}, 1]$, an induction argument in the spirit of (4.2) yields for all $\lambda \in [0, \lambda_{4.1}/\tilde{s}_N)$

$$\begin{aligned}
\mathbb{E}_{\theta_0}[\exp(\lambda f(\theta_N^\gamma))] &\leq \exp\left(\lambda \mathbb{E}_{\theta_0}[f(\theta_N^\gamma)]\right) \exp\left(C_N^\gamma \Psi\left(N,\gamma,\theta_0\right) e^{\sum_{k=0}^{N-1}\frac{1}{(k+1)\log^2(k+4)}} \lambda^2 \right. \\
&\qquad \left. + e^{\sum_{k=0}^{N-1}\frac{1}{(k+1)\log^2(k+4)}} \left(\sum_{p=0}^{N-1}\gamma_{p+1}^2\right) C_N^\gamma \frac{(\lambda/\lambda_{4.1})^2}{1-(\lambda\tilde{s}_N/\lambda_{4.1})}\right), \\
&\leq \exp\left(\lambda \mathbb{E}_{\theta_0}[f(\theta_N^\gamma)]\right) \exp\left(2\varphi_{1/2}(\gamma,H,\theta_0) C_N^\gamma \left((\lambda/\lambda_{4.1})^2 \vee \frac{(\lambda/\lambda_{4.1})^2}{1-(\lambda\tilde{s}_N/\lambda_{4.1})}\right)\right) \\
&= \exp\left(\lambda \mathbb{E}_{\theta_0}[f(\theta_N^\gamma)]\right) \exp\left(2\varphi_{1/2}(\gamma,H,\theta_0) C_N^\gamma \frac{(\lambda/\lambda_{4.1})^2}{1-(\lambda\tilde{s}_N/\lambda_{4.1})}\right)
\end{aligned}$$

with $\varphi_{1/2}(\gamma,H,\theta_0) := \exp(\sum_{k=0}^{N-1}\frac{1}{(k+1)\log^2(k+4)})(\lambda_{4.1}^2\Psi(N,\gamma,\theta_0) + \sum_{p=0}^{N-1}\gamma_{p+1}^2)$, and $\tilde{s}_N := s_N^{1/2}\exp(\sum_{k=0}^{N-1}\frac{1}{(k+1)\log^2(k+4)})$, and where we used again $\prod_{k=0}^{N-1} p_k < \exp(\sum_{k=0}^{N-1}\frac{1}{(k+1)\log^2(k+4)})$. □

In contrast to Euler like schemes, a bias appears in the non-asymptotic deviation bound for the stochastic approximation algorithm. Consequently, it is crucial to have a control on it. At step n of the algorithm, it is given by $\delta_n := \mathbb{E}[|\theta_n^\gamma - \theta^*|]$. Under the current assumptions **(HL)**, **(HLS)**$_\alpha$, **(HUA)**, we have the following proposition.

Proposition 4.7 (Control of the bias)**.** *For all $n \geq 1$, we have*

$$\delta_n \leq e^{-\underline{\lambda}\Gamma_{1,n}+C_{\alpha,\mu}\Gamma_{2,n}}\,|\theta_0-\theta^*| + (2C_{\alpha,\mu})^{\frac{1}{2}}\left(\sum_{k=0}^{n-1}\gamma_{k+1}^2 e^{-2\underline{\lambda}(\Gamma_{1,n}-\Gamma_{1,k+1})+2C_{\alpha,\mu}(\Gamma_{2,n}-\Gamma_{2,k+1})}\right)^{\frac{1}{2}},$$

where $\Gamma_{1,n} := \sum_{k=1}^n \gamma_k$, $\Gamma_{2,n} := \sum_{k=1}^n \gamma_k^2$, $C_{\alpha,\mu} := \underline{\lambda}^2/2 + 2C_\alpha K\mathbb{E}[|U|^2]$ with $K>0$.

Proof. With the notations of Section 1.2, we define for all $n \geq 1$, $\Delta M_n := h(\theta_{n-1}^\gamma) - H(\theta_{n-1}^\gamma, U_n) = \mathbb{E}[H(\theta_{n-1}^\gamma,U_n)|\,\mathcal{F}_{n-1}] - H(\theta_n^\gamma, U_n)$. Recalling that $(U_n)_{n\geq1}$ is a sequence of i.i.d. random variables we have that $(\Delta M_n)_{n\geq 1}$ is a sequence of martingale increments w.r.t. the natural filtration $\mathcal{F} := (\mathcal{F}_n := \sigma(\theta_0, U_1, \cdots, U_n,); n \geq 1)$.

From the dynamic (1.4), we now write for all $n \geq 0$,

$$\begin{aligned} z_{n+1} &:= \theta_{n+1}^\gamma - \theta^* = \theta_n^\gamma - \theta^* - \gamma_{n+1}\{h(\theta_n^\gamma) - \Delta M_{n+1}\} \\ &= \theta_n^\gamma - \theta^* - \gamma_{n+1}\int_0^1 d\lambda Dh(\theta^*+\lambda(\theta_n^\gamma-\theta^*))(\theta_n^\gamma-\theta^*) + \gamma_{n+1}\Delta M_{n+1}, \end{aligned}$$

where we used that $h(\theta^*)=0$ for the last equality. Setting $J_n := \int_0^1 d\lambda Dh(\theta^*+\lambda(\theta_n^\gamma-\theta^*))$, we obtain $z_{n+1} = (I-\gamma_{n+1}J_n)z_n + \gamma_{n+1}\Delta M_{n+1}$ which yields

$$\begin{aligned} \mathbb{E}_{\theta_0}[|z_{n+1}|^2] &= \mathbb{E}_{\theta_0}[|I-\gamma_{n+1}J_n|^2|z_n|^2] + 2\gamma_{n+1}\mathbb{E}_{\theta_0}[(I-\gamma_{n+1}J_n)\Delta M_{n+1}] + \gamma_{n+1}^2\mathbb{E}_{\theta_0}[|\Delta M_{n+1}|^2] \\ &= \mathbb{E}_{\theta_0}[|I-\gamma_{n+1}J_n|^2|z_n|^2] + \gamma_{n+1}^2\mathbb{E}_{\theta_0}[|\Delta M_{n+1}|^2]. \end{aligned}$$

From assumption **(HLS)**$_\alpha$, we deduce that $\forall(\theta,u) \in \mathbb{R}^d \times \mathbb{R}^q$, $|h(\theta)-H(\theta,u)|^2 \leq 2C_\alpha^2 L^{1-\alpha}(\theta)(\mathbb{E}[|U|^2]+|u|^2)$ which combined with the independence of θ_n and U_{n+1} clearly implies

$$\mathbb{E}_{\theta_0}[|h(\theta_n^\gamma) - H(\theta_n^\gamma, U_{n+1})|^2] \leq 4C_\alpha^2\mathbb{E}[|U|^2]\mathbb{E}_{\theta_0}[L^{1-\alpha}(\theta_n^\gamma)].$$

Now, let us notice that L has sub-quadratic growth so that there exists a constant $K>0$ such that

$$\begin{aligned} \mathbb{E}_{\theta_0}\left[|\Delta M_{n+1}|^2\right] = \mathbb{E}_{\theta_0}\left[|h(\theta_n^\gamma) - H(\theta_n^\gamma,U_{n+1})|^2\right] &\leq 4C_\alpha^2\mathbb{E}[|U|^2]\mathbb{E}_{\theta_0}\left[L^{1-\alpha}(\theta_n^\gamma)\right] \\ &\leq 4KC_\alpha^2\mathbb{E}[|U|^2](1+\mathbb{E}_{\theta_0}[|z_n|^2]), \end{aligned}$$

which provides the following bound

$$\begin{aligned} \mathbb{E}_{\theta_0}[|z_{n+1}|^2] &\leq (1-\underline{\lambda}\gamma_{n+1})^2\mathbb{E}_{\theta_0}[|z_n|^2] + 4KC_\alpha^2\mathbb{E}[|U|^2]\gamma_{n+1}^2\mathbb{E}_{\theta_0}[|z_n|^2] \\ &\leq \left(1-2\underline{\lambda}\gamma_{n+1}+2C_{\alpha,\mu}\gamma_{n+1}^2\right)\mathbb{E}_{\theta_0}[|z_n|^2] + 2C_{\alpha,\mu}\gamma_{n+1}^2. \end{aligned}$$

Temporarily setting $\tilde{\Pi}_n = \prod_{p=0}^{n-1}(1-2\underline{\lambda}\gamma_{p+1}+2C_{\alpha,\mu}\gamma_{p+1}^2)$, a straightforward induction argument provides

$$\begin{aligned} \mathbb{E}_{\theta_0}[|z_n|^2] &\leq \tilde{\Pi}_n|\theta_0-\theta^*|^2 + 2C_{\alpha,\mu}\sum_{k=0}^{n-1}\gamma_{k+1}^2\tilde{\Pi}_n\tilde{\Pi}_{k+1}^{-1} \\ &\leq e^{-2\underline{\lambda}\Gamma_{1,n}+2C_{\alpha,\mu}\Gamma_{2,n}}|\theta_0-\theta^*|^2 + 2C_{\alpha,\mu}\sum_{k=0}^{n-1}\gamma_{k+1}^2 e^{-2\underline{\lambda}(\Gamma_{1,n}-\Gamma_{1,k+1})+2C_{\alpha,\mu}(\Gamma_{2,n}-\Gamma_{2,k+1})} \end{aligned}$$

where we used the elementary inequality, $1+x \leq \exp(x)$, $x \in \mathbb{R}$. This completes the proof.

4.2 Proof of Theorem 2.10

Proposition 4.8. *(Control of the Lipschitz modulus of iterative kernels) Denote by* K_k *and* $K_{k,p} = K_k \circ \cdots \circ K_{p-1}$, $k,\ p \in \{0, \cdots, N-1\}$, $k \le p$ *the (Feller) transition kernel and the iterative kernels of the Markov chain* $z = (\bar{\theta}, \theta)$ *defined by the scheme* (1.4), (1.8). *Let* $f : \mathbb{R}^d \to \mathbb{R}$ *be a* 1*-Lipschitz function. Then for all* $k,\ p \in \{0, \cdots, N-1\}$, $k \le p$ *the functions* $K_{k,p}(f) : z \mapsto \mathbb{E}[\, f(\bar{\theta}^{\gamma}_{p+1}) |\, z_k = z]$ *are Lipschitz-continuous. In particular, for all* $(z, z') \in (\mathbb{R}^d \times \mathbb{R}^d)^2$*, one has*

$$|K_{k,p}(f)(z) - K_{k,p}(f)(z')| \le \frac{k+1}{p+1}|z_1 - z_1'| + \frac{1}{p+1} \sum_{j=k+1}^{p} \left(\frac{\Pi_{1,j}}{\Pi_{1,k}}\right)^{\frac{1}{2}} |z_2 - z_2'|$$

where $\Pi_{1,p} = \prod_{k=0}^{p-1}(1 - 2\underline{\lambda}\gamma_{k+1} + C_{H,\mu}\gamma_{k+1}^2)$.

Proof. Let $(z, z') \in (\mathbb{R}^d \times \mathbb{R}^d)^2$. We denote by $z^{k,z}_{p,1} = \bar{\theta}^{k,z}_{p+1}$ and $z^{k,z}_{p,2} = \theta^{k,z}_{p}$ the values at step p of the two components of the stochastic approximation algorithm $(z_n)_{n \ge 0}$ starting at point z at step k. Using (1.8) and a straightforward induction, one easily derives

$$\bar{\theta}^{k,z}_{p+1} = \frac{k+1}{p+1} z_1 + \frac{1}{p+1} \sum_{j=k+1}^{p} \theta^{k,z}_j,$$

so that taking conditional expectation in the previous equality and using Proposition 4.4,we obtain

$$\begin{aligned} |K_{k,p}(f)(z) - K_{k,p}(f)(z')| &= |\mathbb{E}[f(\bar{\theta}^{k,z}_{p+1})] - \mathbb{E}[f(\bar{\theta}^{k,z'}_{p+1})]| \le \mathbb{E}[|\bar{\theta}^{k,z}_{p+1} - \bar{\theta}^{k,z'}_{p+1}|] \\ &\le \frac{k+1}{p+1}|z_1 - z_1'| + \frac{1}{p+1} \sum_{j=k+1}^{p} \mathbb{E}[|\theta^{k,z}_j - \theta^{k,z'}_j|] \\ &\le \frac{k+1}{p+1}|z_1 - z_1'| + \frac{1}{p+1} \sum_{j=k+1}^{p} \left(\frac{\Pi_{1,j}}{\Pi_{1,k}}\right)^{\frac{1}{2}} |z_2 - z_2'| \end{aligned}$$

□

Let $k \in \{0, \cdots, N-1\}$ and f be a real-valued 1-Lipschitz function defined on $\mathbb{R}^d$. Using that the law of the innovations of the scheme satisfies ($GC(\beta)$), for all $\lambda \ge 0$, one has

$$\begin{aligned} \mathbb{E}\left[e^{\lambda K_{k,N-1} f(z_k)} \middle|\, z_{k-1} = z \right] &= \mathbb{E}\left[e^{\lambda K_{k,N-1} f(\frac{k}{k+1}\bar{\theta}^{\gamma}_k + \frac{1}{k+1}\theta^{\gamma}_k, \theta^{\gamma}_k)} \middle|\, (\bar{\theta}^{\gamma}_k, \theta^{\gamma}_{k-1}) = (z_1, z_2) \right] \\ &\le e^{\lambda K_{k-1,N-1}(f)(z)} e^{\lambda^2 \frac{\beta}{4} [g]_1^2} \end{aligned}$$

where $g : u \mapsto K_{k,N-1}(f)\left(\frac{k}{k+1} z_1 + \frac{1}{k+1} z_2 - \frac{\gamma_k}{k+1} H(z_2, u), z_2 - \gamma_k H(z_2, u)\right)$. Combining Proposition 4.8 and **(HLS)**$_\alpha$, one easily obtains

$$[g]_1 \le C_\alpha L^{\frac{1-\alpha}{2}}(z_2)\gamma_k \left(\frac{1}{N} + \frac{1}{N} \sum_{j=k+1}^{N-1} \left(\frac{\Pi_{1,j}}{\Pi_{1,k}}\right)^{\frac{1}{2}} \right)$$

so we deduce that

$$\mathbb{E}\left[\exp(\lambda K_{k,N-1} f(z_k)) |\, z_{k-1}\right] \le \exp(\lambda K_{k-1,N-1}(f)(z_{k-1})) \exp\left(\lambda^2 \frac{\beta}{4} C_\alpha^2 L^{1-\alpha}(z_{k-1}) \tilde{\gamma}_{k,N}^2 \right)$$

where we introduced the notation $\tilde{\gamma}_{k,N} := \frac{\gamma_k}{N}\left(1+\sum_{j=k+1}^{N-1}(\Pi_{1,j}/\Pi_{1,k})^{\frac{1}{2}}\right)$. Hence, taking expectation in the previous inequality and using the Hölder inequality with conjugate exponents (p_k, q_k), one clearly gets

$$\mathbb{E}_{\theta_0}\left[e^{\lambda K_{k,N-1}(f)(z_k)}\right] \le \mathbb{E}_{\theta_0}\left[e^{\lambda p_k K_{k-1,N-1}(f)(z_{k-1})}\right]^{\frac{1}{p_k}} \mathbb{E}_{\theta_0}\left[e^{\lambda^2 \frac{\beta}{4} C_\alpha^2 q_k L^{1-\alpha}(\theta_{k-1}^\gamma)\tilde{\gamma}_{k,N}^2}\right]^{\frac{1}{q_k}}$$

Similarly to the proof of Proposition 4.5, we set $p_k = 1 + \frac{1}{(k+1)\log^2(k+4)}$, $q_k = (1 + \frac{1}{(k+1)\log^2(k+4)})(k+1)\log^2(k+4) \le 2(k+1)\log^2(k+4)$ and use Corollary 4.2 to obtain for $\alpha \in (\frac{1}{2}, 1]$

$$\mathbb{E}_{\theta_0}\left[e^{\lambda^2 \frac{\beta}{4} C_\alpha^2 q_k L^{1-\alpha}(\theta_{k-1}^\gamma)\tilde{\gamma}_{k,N}^2}\right]^{\frac{1}{q_k}} \le e^{K_{4.1} 2^{\frac{1-\alpha}{2\alpha-1}} \frac{\beta C_\alpha^2}{4} \vee \left(\frac{\beta C_\alpha^2}{4}\right)^{\frac{\alpha}{2\alpha-1}} (\tilde{\gamma}_{k,N}^2 \lambda^2 \vee \tilde{\gamma}_{k,N}^{\frac{2\alpha}{2\alpha-1}} q_k^{\frac{1-\alpha}{2\alpha-1}} \lambda^{\frac{2\alpha}{2\alpha-1}})}.$$

An elementary induction argument allows to conclude

$$\begin{aligned}\mathbb{E}_{\theta_0}\left[\exp(\lambda f(\bar{\theta}_N^\gamma))\right] &= \mathbb{E}_{\theta_0}\left[\exp(\lambda K_{N-1,N-1}(f)(z_{N-1}))\right] \\ &\le \exp(\lambda \mathbb{E}_{\theta_0}[f(\bar{\theta}_N^\gamma)]) \exp\left(\varphi_\alpha(\gamma, H, \theta_0)(\bar{C}_N^\gamma \lambda^2 \vee \bar{C}_N^{\gamma,\alpha} \lambda^{\frac{2\alpha}{2\alpha-1}})\right)\end{aligned}$$

with $\bar{C}_N^\gamma := \sum_{k=1}^{N-1} \tilde{\gamma}_{k,N}^2$, $\bar{C}_N^{\gamma,\alpha} := \sum_{k=1}^{N-1} \tilde{\gamma}_{k,N}^{\frac{2\alpha}{2\alpha-1}} ((k+1)\log^2(k+4))^{\frac{1-\alpha}{2\alpha-1}}$ and where we introduced $\varphi_\alpha(\gamma, H, \theta_0) := K_{4.1} 2^{\frac{1-\alpha}{2\alpha-1}} \frac{\beta C_\alpha^2}{4} \vee \left(\frac{\beta C_\alpha^2}{4}\right)^{\frac{\alpha}{2\alpha-1}} e^{\frac{1}{2\alpha-1}\sum_{k=0}^{N-1} \frac{1}{(k+1)\log^2(k+4)}}$.

For $\alpha = \frac{1}{2}$, similarly to the proof of Proposition 4.5 (we actually use again Lemma 4.6), we derive for all $\lambda \in [0, \lambda_{4.1}/\bar{s}_N^{1/2})$

$$\mathbb{E}_{\theta_0}\left[e^{q_k \frac{\lambda^2}{4} \beta \tilde{\gamma}_{k,N}^2 C_{1/2}^2 L^{1/2}(\theta_{k-1}^\gamma)}\right]^{\frac{1}{q_k}} \le e^{\Psi(N,\gamma,\theta_0)\tilde{\gamma}_{k,N}^2 \lambda^2 + (\sum_{p=0}^{N-1}\gamma_{p+1}^2)\tilde{\gamma}_{k,N}^2 \frac{(\lambda/\lambda_{4.1})^2}{1-(\lambda \bar{s}_N^{1/2}/\lambda_{4.1})}}$$

with $\Psi(N,\gamma,\theta_0) := \left(L^{\frac{1}{2}}(\theta_0) + \underline{C}\sum_{p=0}^{N-1}\gamma_{p+1}^2\right)\Pi_{2,N}(1/2)\frac{\beta C_{1/2}^2}{4} + \beta\eta C_{1/2}^4 \Pi_{2,N}(1/2)\mathbb{E}[|U|^2]$ and $\bar{s}_N := \max_{1\le k\le N-1}(k+1)\log^2(k+4)\tilde{\gamma}_{k,N}^2$. Then an elementary induction argument clearly yields

$$\forall \lambda \in [0, \lambda_{4.1}/\hat{s}_N), \quad \mathbb{E}_{\theta_0}[e^{\lambda f(\bar{\theta}_N^\gamma)}] \le e^{\lambda \mathbb{E}_{\theta_0}[f(\bar{\theta}_N^\gamma)]} e^{2\varphi_{1/2}(\gamma,H,\theta_0)\bar{C}_N^\gamma \frac{(\lambda/\lambda_{4.1})^2}{1-(\lambda \hat{s}_N/\lambda_{4.1})}}$$

with $\varphi_{1/2}(\gamma, H, \theta_0) := e^{\sum_{k=0}^{N-1}\frac{1}{(k+1)\log^2(k+4)}}(\lambda_{4.1}^2 \Psi(N,\gamma,\theta_0) + \sum_{p=0}^{N-1}\gamma_{p+1}^2)$, and also $\hat{s}_N := \bar{s}_N^{1/2} e^{\sum_{k=0}^{N-1}\frac{1}{(k+1)\log^2(k+4)}}$.

A Proof of Proposition 1.5

Let $e_\sigma := \frac{1}{2\sigma}\exp(-|x|/\sigma)$ be the density of the exponential distribution with variance $2\sigma^2$ on $\mathbb{R}$. If μ is a probability measure on $\mathbb{R}^d$, we define μ^σ as the convolution of μ with $e_\sigma^{\otimes d}$, that is

$$\mu^\sigma(dx) := \int \prod_{i=1}^d \frac{1}{2\sigma}\exp(-|x_i - y_i|/\sigma)\mu(dy).$$

Lemma A.1. *If μ is a probability measure on $\mathbb{R}^d$ with finite first moment, then $W_1(\mu, \mu^\sigma) \le \sqrt{2d}\sigma$.*

Proof. Let X and Y be independent random vectors with laws μ and $e_\sigma^{\otimes d}$ respectively. Then $(X, X+Y)$ is a coupling of μ and μ^σ, and

$$W_1(\mu, \mu^\sigma) \le \mathbb{E}[|Y|] \le \mathbb{E}[|Y|^2]^{1/2} \le \sqrt{2d}\sigma.$$

□

We therefore have the bound

$$W_1(\mu_M,\mu) \le W_1(\mu_M,\mu_M^\sigma) + W_1(\mu_M^\sigma,\mu^\sigma) + W_1(\mu^\sigma,\mu) \le W_1(\mu_M^\sigma,\mu^\sigma) + \sqrt{8d}\sigma, \qquad \text{(A.1)}$$

so what is left is to bound $\mathbb{E}[W_1(\mu_M^\sigma,\mu^\sigma)]$ and to optimize with respect to σ.

The density of μ_M^σ with respect to the Lebesgue measure is given by $g_{1,\sigma,M}(x) := \frac{1}{M}\sum_{k=1}^M e_\sigma^{\otimes d}(x-X_k)$, and the density of μ^σ is $g_{2,\sigma}(x) := \mathbb{E}_\mu(e_\sigma^{\otimes d}(x-X))$.

By the Kantorovitch-Rubinstein duality formula, we have

$$W_1(\mu_M^\sigma,\mu^\sigma) = \sup_{f:[f]_1\le 1} \int f(x)g_{1,\sigma,M}(x)dx - \int f(x)g_{2,\sigma}(x)dx \le \int |x||g_{1,\sigma,M}(x) - g_{2,\sigma}(x)|dx$$

To bound this quantity, we use the Cauchy-Schwarz inequality, namely for any non-negative measurable function f on $\mathbb{R}^d$, we have

$$\int f(x)dx \le C_d\sqrt{\int (1+|x|^{d+1})f(x)^2dx}, \quad C_d := \sqrt{\int_{\mathbb{R}^d} \frac{1}{1+|x|^{d+1}}dx}.$$

Using this inequality, we get the bound

$$\begin{aligned} W_1(\mu_M^\sigma,\mu^\sigma) &\le C_d\sqrt{\int (1+|x|^{d+1})|x|^2|g_{1,\sigma,M}(x) - g_{2,\sigma}(x)|^2dx} \qquad \text{(A.2)} \\ &\le C_d\sqrt{\int (1+2|x|^{d+3})|g_{1,\sigma,M}(x) - g_{2,\sigma}(x)|^2dx}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E}[W_1(\mu_M^\sigma,\mu^\sigma)] &\le C_d\mathbb{E}\left[\sqrt{\int (1+2|x|^{d+3})|\frac{1}{M}\sum_{k=1}^M e_\sigma^{\otimes d}(x-X_k) - \mathbb{E}_\mu(e_\sigma^{\otimes d}(x-X))|^2dx}\right] \\ &\le \frac{C_d}{\sqrt{M}}\sqrt{\int (1+2|x|^{d+3})\mathrm{Var}_\mu(e_\sigma^{\otimes d}(x-X))dx} \\ &\le \frac{C_d}{\sqrt{M}}\sqrt{\int (1+2|x|^{d+3})\mathbb{E}[e_\sigma^{\otimes d}(x-X)^2]dx}. \end{aligned}$$

Note that $e_\sigma^{\otimes d}(x)^2 = 2^{-2d}\sigma^{-d}e_{\sigma/2}^{\otimes d}(x)$, so that we get

$$\begin{aligned} \mathbb{E}[W_1(\mu_M^\sigma,\mu^\sigma)] &\le \frac{C_d}{2^d\sigma^{d/2}\sqrt{M}}\sqrt{\int (1+2|x|^{d+3})\int e_{\sigma/2}^{\otimes d}(x-y)\mu(dy)dx} \\ &\le \frac{C_d}{2^d\sigma^{d/2}\sqrt{M}}\sqrt{\int\int (1+2|u+y|^{d+3})e_{\sigma/2}^{\otimes d}(u)du\mu(dy)} \\ &\le \frac{C_d}{2^d\sigma^{d/2}\sqrt{M}}\sqrt{\int\int (1+2^{d+3}(|u|^{d+3}+|y|^{d+3}))e_{\sigma/2}^{\otimes d}(u)du\mu(dy)} \\ &\le \frac{C_d}{2^d\sigma^{d/2}\sqrt{M}}\sqrt{1+2^{d+3}\int |y|^{d+3}\mu(dy) + 2^{d+3}\int |u|^{d+3}e_{\sigma/2}^{\otimes d}(u)du} \\ &\le \frac{C_d}{2^d\sigma^{d/2}\sqrt{M}}\sqrt{1+2^{d+3}\int |y|^{d+3}\mu(dy) + \sigma^{d+3}\int |u|^{d+3}e_1^{\otimes d}(u)du} \\ &\le \frac{C_d}{2^d\sigma^{d/2}\sqrt{M}}\sqrt{1+2^{d+3}\int |y|^{d+3}\mu(dy) + 2^{d+3}\sigma^{d+3}d(d+3)!} \end{aligned}$$

In the end, assuming $\sigma \leq 1$, we obtain

$$\mathbb{E}[W_1(\mu,\mu^\sigma)] \leq \sqrt{8d}\sigma + \frac{C_d}{2^d\sigma^{d/2}\sqrt{M}}\sqrt{1+2^{d+3}\int |y|^{d+3}\mu(dy) + 2^{d+3}\sigma^{d+3}d(d+3)!}$$
$$\leq C(d,\mu)(\sigma + \frac{\sigma^{-d/2}}{\sqrt{M}})$$

Taking $\sigma = M^{-1/(d+2)}$, we get the upper bound we were aiming for.

References

[1] A. Alfonsi and A. Kohatsu-Higa and B. Jourdain: Pathwise optimal transport bounds between a one-dimensional diffusion and its Euler scheme. arXiv:1209.0576

[2] V. Bally and D. Talay: The law of the Euler scheme for stochastic differential equations, II. Convergence rate of the density. *Monte-Carlo methods and Appl.*, **2**, (1996), 93–128, MR-1401964

[3] F. Barthe and C. Bordenave: Combinatorial optimization over two random point sets. *To appear in Séminaire de Probabilités*, arXiv:1103.2734

[4] G. Blower and F. Bolley: Concentration inequalities on product spaces with applications to Markov processes. *Studia Mathematica*, **175-1** (2006), 47–72, MR-2261699.

[5] E. Boissard: Simple bounds for the convergence of empirical and occupation measures in 1-Wasserstein distance, *Electronic Journal of Probability*, **16** (2011), 2296–2333, MR-2861675.

[6] F. Bolley and C. Villani: Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities, *Annales de la Faculté des Sciences de Toulouse*, (6) **14(3)** (2005), 331–352. MR-2172583.

[7] F. Bolley, A. Guillin, and C. Villani: Quantitative concentration inequalities for empirical measures on non-compact spaces, *Prob. Th. Rel. Fields*, **137** (2007), 541–593. MR-2280433.

[8] H. Djellout, A. Guillin, and L. Wu: Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Prob*, **32:no. 3B** (2004), 2702–2732, MR-2078555.

[9] M. Duflo, Algorithmes stochastiques, *Mathématiques & Applications* (Berlin) [Mathematics & Applications], vol. 23, Springer-Verlag, Berlin, 1996. MR-1612815

[10] N. Frikha, S. Menozzi: Concentration bounds for stochastic approximations, *Electronic Communications in Probability*, **17** (2012), 1–15, MR-2988393.

[11] N. Gozlan and C. Léonard: A large deviation approach to some transportation cost inequalities. *Prob. Th. Rel. Fields*, **139(1-2)** (2007), 235–283. MR-2322697.

[12] N. Gozlan and C. Léonard: Transport inequalities. A survey, *Markov Process. Related Fields*, **16(4)** (2010), 635–736. MR-2895086.

[13] W. B. Johnson, G. Schechtman, and J. Zinn: Best constants in moment inequalities for linear combinations of independent and exchangeable random variables, *Ann. Probab.*, **13(1)** (1985), 234–253. MR-770640.

[14] Harold J. Kushner and G. George Yin, *Stochastic approximation and recursive algorithms and applications*, second ed., Applications of Mathematics (New York), vol. 35, Springer-Verlag, New York, 2003, Stochastic Modelling and Applied Probability. MR-1993642

[15] S. Laruelle and G. Pagès, Stochastic approximation with averaging innovation applied to finance, *Monte Carlo Methods Appl.* **18–1** (2012), 1–52. MR-2908605.

[16] V. Lemaire and S. Menozzi, On some non asyptotic bounds for the Euler scheme, *Electronic Journal of Probability* **15** (2010), 1645–1681. MR-2735377.

[17] F. Malrieu and D. Talay: Concentration inequalities for Euler schemes, *Monte Carlo and Quasi-Monte Carlo Methods 2004*, H. Niederreiter and D. Talay (Eds.) (2006), 355–371. MR-2208718.

[18] B. T. Polyak and A. B. Juditsky: Acceleration of stochastic approximation by averaging, *SIAM J. Control Optim.*, **30(4)**, (1992), 838–855. MR-1167814

[19] H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Statistics*,**22** (1951), 400–407. MR-0042668.

[20] S. T. Rachev and L. Rüschendorf: Mass transportation problems. Vol. II. Probability and its Applications (New York). Springer-Verlag, New York, 1998. Applications. MR-1619171.

[21] D. Ruppert: Stochastic approximations. In *Handbook of sequential analysis*, volume 118 of *Statist. Textbooks Monogr.*, pages 503–529. Dekker, New York, 1991. MR-1174318.

[22] D. Talay and L. Tubaro: Expansion of the global error for numerical schemes solving stochastic differential equations, *Stoch. Anal. and App.* **8-4** (1990), 94–120. MR-1091544.