



**HAL**  
open science

# Non-asymptotic convergence analysis for the Unadjusted Langevin Algorithm

Alain Durmus, Eric Moulines

► **To cite this version:**

Alain Durmus, Eric Moulines. Non-asymptotic convergence analysis for the Unadjusted Langevin Algorithm. 2015. hal-01176132v1

**HAL Id: hal-01176132**

**<https://hal.science/hal-01176132v1>**

Preprint submitted on 17 Jul 2015 (v1), last revised 19 Dec 2016 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Non-asymptotic convergence analysis for the Unadjusted Langevin Algorithm

Alain Durmus <sup>\*</sup> and Éric Moulines <sup>\*\*</sup>

*LTCI, Telecom ParisTech & CNRS, 46 rue Barrault, 75634 Paris Cedex 13, France,*  
e-mail: <sup>\*</sup>[alain.durmus@telecom-paristech.fr](mailto:alain.durmus@telecom-paristech.fr); <sup>\*\*</sup>[eric.moulines@telecom-paristech.fr](mailto:eric.moulines@telecom-paristech.fr)

**Abstract:** : Sampling distribution over high-dimensional state-space is a problem which has recently attracted a lot of research efforts; applications include Bayesian non-parametrics, Bayesian inverse problems and aggregation of estimators. All these problems boil down to sample a target distribution  $\pi$  having a density w.r.t. the Lebesgue measure on  $\mathbb{R}^d$ , known up to a normalisation factor (see [9] for details)  $x \mapsto e^{-U(x)} / \int_{\mathbb{R}^d} e^{-U(y)} dy$  where  $U$  is continuously differentiable and smooth. In this paper, we study a sampling technique based on the Euler discretization of the Langevin stochastic differential equation. Contrary to the Metropolis Adjusted Langevin Algorithm (MALA), we do not apply a Metropolis-Hastings correction. We obtain for both constant and decreasing step sizes in the Euler discretization, non-asymptotic bounds for the convergence to stationarity in both total variation and Wasserstein distances. A particular attention is paid on the dependence on the dimension of the state space, to demonstrate the applicability of this method in the high dimensional setting, at least when  $U$  is convex. These bounds are based on recently obtained estimates of the convergence of the Langevin diffusion to stationarity using Poincaré and log-Sobolev inequalities. These bounds improve and extend the results of [10]. We also investigate the convergence of an appropriately weighted empirical measure and we report sharp bounds for the mean square error and exponential deviation inequality for Lipschitz functions. A limited Monte Carlo experiment is carried out to support our findings.

**AMS 2000 subject classifications:** primary 65C05, 60F05, 62L10; secondary 65C40, 60J05, 93E35.

**Keywords and phrases:** Wasserstein distance, Langevin diffusion, Markov Chain Monte Carlo, Metropolis Adjusted Langevin Algorithm, Rate of convergence.

## 1. Introduction

We study the sampling over  $\mathbb{R}^d$  of a target distribution  $\pi$  with density  $x \mapsto e^{-U(x)} / \int_{\mathbb{R}^d} e^{-U(y)} dy$  w.r.t. the Lebesgue measure, where  $U$  is a smooth function. We consider a sampling method based on the Euler discretization of the Langevin stochastic differential equation

$$dY_t = -\nabla U(Y_t)dt + \sqrt{2}dB_t, \quad (1)$$

where  $(B_t)_{t \geq 0}$  is a  $d$ -dimensional Brownian motion. It is well-known that the Markov semi-group associated with the Langevin diffusion  $(Y_t)_{t \geq 0}$  is reversible w.r.t.  $\pi$ . Under suitable conditions, the convergence to  $\pi$  takes place at geometrical rate. Precise quantitative estimates of the rate of convergence with a generally mild dependence on the dimension of the state space have been recently obtained using Poincaré and log-Sobolev inequalities; see [3, 7] [4]. The Euler-Maruyama discretization scheme associated to the Langevin diffusion is the discrete time-Markov chain given by

$$X_{k+1} = X_k - \gamma_{k+1} \nabla U(X_k) + \sqrt{2\gamma_{k+1}} Z_{k+1} \quad (2)$$

where  $(Z_k)_{k \geq 1}$  is an i.i.d. sequence of standard Gaussian random variables and  $(\gamma_k)_{k \geq 1}$  is a sequence of stepsizes, which can either be held constant or be chosen to decrease to 0. The idea of using the

Markov chain  $(X_n)_{n \geq 0}$  to sample (approximately) from the target  $\pi$  has been first introduced in the physics literature by [27] and popularized in the computational statistics community by [15] and [16]. It has been studied in depth by [29], which proposed to use a Metropolis-Hastings step at each iteration to enforce reversibility w.r.t.  $\pi$  leading to the Metropolis Adjusted Langevin Algorithm (MALA). They coin the term *unadjusted* Langevin algorithm (ULA) when the Metropolis-Hastings step is avoided.

The purpose of this paper is to study the convergence of the ULA algorithm. The emphasis is put on non-asymptotic computable bounds; we pay a particular attention to the way these bounds scale with the dimension  $d$  and constants characterizing the smoothness and tail behavior of the potential  $U$ . Our study covers both the constant and decreasing step sizes and we analyse both the "finite horizon" (where the total number of simulations is specified before running the algorithm) and "any-time" settings (where the algorithm can be stopped after any iteration).

When the stepsize  $\gamma_k = \gamma$  is constant, under appropriate conditions related to [29], the Markov chain  $(X_n)_{n \geq 0}$  is  $V$ -uniformly geometrically ergodic with stationary distribution  $\pi_\gamma$  (the function  $V$  being taken to be  $\pi^{-\delta}$  for some appropriate value of  $\delta > 0$ ). With few exceptions, the stationary distribution  $\pi_\gamma$  is no longer equal to the target  $\pi$ . It is of course sensible to expect that if the stepsize  $\gamma$  is small enough, then the stationary distribution of this chain is in some sense close to  $\pi$ . We provide non-asymptotic bounds of the  $V$ -total variation distance between the law of the iterates and the target distribution  $\pi$ , with explicit dependence on the stepsize  $\gamma$ , the dimension  $d$  and the number of simulations. Our results complete and extend the recent works by [11] and [10]. We first consider the case of densities which are super-exponential in the tails, for which we prove convergence in total variation. We also address the case where  $\pi$  is globally log-concave for which we find ergodicity constants growing only linearly in the dimension  $d$ , for both the total variation and the Wasserstein distances.

When  $(\gamma_k)_{k \geq 1}$  decreases to zero, then  $(X_n)_{n \geq 0}$  is a non-homogeneous Markov chain. If in addition  $\sum_{k=1}^{\infty} \gamma_k = \infty$ , we show that the marginal distribution of this non-homogeneous chain converges, under some mild additional conditions, to the target distribution  $\pi$ . We also provide explicit expression for the convergence rate, emphasizing the role of the dimension and the rate of decrease of the sequence  $(\gamma_k)_{k \geq 1}$ . Compared to the related works by [21], [22], [23] and [24], we establish not only the weak convergence of the weighted empirical measure of the path to the target distribution, but establish a much stronger convergence in total variation and Wasserstein distances.

Assuming that the target distribution is globally log-concave, we also provide explicit mean-square error and exponential deviation bounds for a class of estimators of  $\pi(f)$  where  $f$  is a Lipschitz function satisfying some appropriate moment conditions. We show that these bounds depend only linearly in the dimension  $d$  and in the sample size.

The paper is organized as follows. In Section 2, we summarize the existing results on the rate of convergence of the Langevin diffusion to its stationary distribution. In Section 3, we then study the convergence in total variation and Wasserstein distances of the Euler discretization for constant and decreasing stepsizes. In Section 4 we provide non-asymptotic bounds of convergence of the weighted empirical measure. Some numerical illustrations are given Section 5 to support our claims. The proofs are given in Section 6. Some technical derivations are carried out in a supplementary paper [12].

### Notations and conventions

Denote by  $\mathcal{B}(\mathbb{R}^d)$  the Borel  $\sigma$ -field of  $\mathbb{R}^d$ . Denote by  $\mathbb{F}(\mathbb{R}^d)$  the set of all Borel measurable functions on  $\mathbb{R}^d$  and for  $f \in \mathbb{F}(\mathbb{R}^d)$ ,  $\|f\|_\infty = \sup_{x \in \mathbb{R}^d} |f(x)|$ . For  $\mu$  a probability measure on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  and  $f \in \mathbb{F}(\mathbb{R}^d)$  a  $\mu$ -integrable function, denote by  $\mu(f)$  the integral of  $f$  w.r.t.  $\mu$ . Let  $V : \mathbb{R}^d \rightarrow [1, \infty)$  be a measurable function. For  $f \in \mathbb{F}(\mathbb{R}^d)$ , the  $V$ -norm of  $f$  is given by  $\|f\|_V = \sup_{x \in \mathbb{R}^d} |f(x)|/V(x)$ . For two probability measures  $\mu$  and  $\nu$  on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ , the  $V$ -total variation distance of  $\mu$  and  $\nu$  is defined as

$$\|\mu - \nu\|_V = \sup_{f \in \mathbb{F}(\mathbb{R}^d), \|f\|_V \leq 1} \left| \int_{\mathbb{R}^d} f(x) d\mu(x) - \int_{\mathbb{R}^d} f(x) d\nu(x) \right|$$

If  $V \equiv 1$ , then  $\|\cdot\|_V$  is the total variation denoted by  $\|\cdot\|_{\text{TV}}$ .

For  $p \geq 1$ , denote by  $L^p(\pi)$  the set of measurable functions such that  $\pi(|f|^p) < \infty$ . For  $f \in L^2(\pi)$ , the variance of  $f$  under  $\pi$  is denoted by  $\text{Var}_\pi\{f\}$ . For all function  $f$  such that  $f \log(f) \in L^1(\pi)$ , the entropy with respect to  $\pi$  for  $f$  is defined by

$$\text{Ent}_\pi(f) = \int_{\mathbb{R}^d} f(x) \log(f(x)) d\pi(x). \quad (3)$$

Let  $\mu$  and  $\nu$  be two probability measures on  $\mathbb{R}^d$ . If  $\mu \ll \nu$ , we denote by  $d\mu/d\nu$  the Radon-Nikodym derivative of  $\mu$  w.r.t.  $\nu$ . We say that  $\zeta$  is a transference plan of  $\mu$  and  $\nu$  if it is a probability measure on  $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d))$  such that for all measurable set  $A$  of  $\mathbb{R}^d$ ,  $\zeta(A \times \mathbb{R}^d) = \mu(A)$  and  $\zeta(\mathbb{R}^d \times A) = \nu(A)$ . We denote by  $\Pi(\mu, \nu)$  the set of transference plans of  $\mu$  and  $\nu$ . Furthermore, we say that a couple of  $\mathbb{R}^d$ -random variables  $(X, Y)$  is a coupling of  $\mu$  and  $\nu$  if there exists  $\zeta \in \Pi(\mu, \nu)$  such that  $(X, Y)$  are distributed according to  $\zeta$ . For two probability measures  $\mu$  and  $\nu$ , we define the Wasserstein distance of order  $p \geq 1$  as

$$W_p(\mu, \nu) \stackrel{\text{def}}{=} \left( \inf_{\zeta \in \Pi(\mu, \nu)} \int_{\mathbb{R} \times \mathbb{R}} \|x - y\|^p d\zeta(x, y) \right)^{1/p}.$$

By [31, Theorem 4.1], for all  $\mu, \nu$  probability measure on  $\mathbb{R}^d$ , there exists a transference plan  $\zeta^* \in \Pi(\mu, \nu)$  such that for any coupling  $(X, Y)$  distributed according to  $\zeta^*$ ,  $W_p(\mu, \nu) = \mathbb{E}[\|X - Y\|^p]^{1/p}$ . This kind of transference plan (respectively coupling) will be called an optimal transference plan (respectively optimal coupling) associated with  $W_p$ . We denote by  $\mathcal{P}_p(\mathbb{R}^d)$  the set of probability measures with finite  $p$ -moment: for all  $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ ,  $\int_{\mathbb{R}^d} \|x\|^p \mu(dx) < +\infty$ . By [31, Theorem 6.16],  $\mathcal{P}_p(\mathbb{R}^d)$  equipped with the Wasserstein distance  $W_p$  of order  $p$  is a complete separable metric space.

Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a Lipschitz function, namely there exists  $C \geq 0$  such that for all  $x, y \in \mathbb{R}^d$ ,  $|f(x) - f(y)| \leq C \|x - y\|$ . Then we denote  $\|f\|_{\text{Lip}} = \inf\{|f(x) - f(y)| \|x - y\|^{-1} \mid x, y \in \mathbb{R}^d, x \neq y\}$ . The Monge-Kantorovich theorem (see [31, Theorem 5.9]) implies that for all  $\mu, \nu$  probabilities measure on  $\mathbb{R}^d$ ,

$$W_1(\mu, \nu) = \sup \left\{ \int_{\mathbb{R}^d} f(x) \mu(dx) - \int_{\mathbb{R}^d} f(x) \nu(dx) \mid f : \mathbb{R}^d \rightarrow \mathbb{R}; \|f\|_{\text{Lip}} \leq 1 \right\}. \quad (4)$$

For all  $x \in \mathbb{R}^d$  and  $M > 0$ , we denote by  $B(x, M)$ , the ball centered at  $x$  of radius  $M$ . In the sequel, we take the convention that for  $n, p \in \mathbb{N}$ ,  $n < p$  then  $\sum_p^n = 0$  and  $\prod_p^n = 1$ .

## 2. Ergodicity of the Langevin diffusion

Consider the following assumption on the potential  $U$ :

**H1.** *The function  $U$  is lower-bounded, twice continuously differentiable on  $\mathbb{R}^d$  and is gradient Lipschitz, i.e. there exists  $L \geq 0$  such that for all  $x, y \in \mathbb{R}^d$ ,*

$$\|\nabla U(x) - \nabla U(y)\| \leq L \|x - y\| .$$

Under **H1**, if  $\mu_0$  is a probability measure satisfying  $\int \|x\|^2 \mu_0(dx) < \infty$  then by [20, Theorem 2.5, Theorem 2.9 Chapter 5] there exists a unique strong solution  $(Y_t)_{t \geq 0}$  to (1) with initial distribution  $\mu_0$ . Define for all  $t \geq 0$ ,  $x \in \mathbb{R}^d$  and  $A \in \mathcal{B}(\mathbb{R}^d)$ ,  $(Y_s)_{s \geq 0}$ ,  $P_t(x, A) = \mathbb{P}_x [Y_t \in A]$ . The semi-group  $(P_t)_{t \geq 0}$  is reversible w.r.t.  $\pi$ , and hence admits  $\pi$  as its (unique) invariant measure. By [29, Theorem 2.1], under **H1** the diffusion is irreducible, aperiodic, strong Feller and all compact sets are small which implies, for any initial distribution  $\mu_0$ , that

$$\lim_{t \rightarrow +\infty} \|\mu_0 P_t - \pi\|_{\text{TV}} = 0 . \quad (5)$$

The generator  $\mathcal{A}$  and the carré du champ operator (see [4, Definition 1.4.2]) associated with the Langevin diffusion (1) are given for all  $(f, g) \in C^2(\mathbb{R}^d)$  by

$$\begin{aligned} \mathcal{A}f &= -\langle \nabla U, \nabla f \rangle + \Delta f , \\ \mathcal{G}(f, g) &= 2^{-1}(\mathcal{A}fg - f\mathcal{A}g - g\mathcal{A}f) = \langle \nabla f, \nabla g \rangle , \end{aligned} \quad (6)$$

where  $\Delta$  denote the Laplacian. We will use the short-hand notation  $\mathcal{G}(f) = \mathcal{G}(f, f)$ .

We are interested in when the convergence in (5) is exponential. A function  $V \in C^2(\mathbb{R}^d)$  is a *Lyapunov function* for  $(P_t)_{t \geq 0}$ , if  $V \geq 1$  and if there exists  $\theta > 0$ ,  $b \geq 0$  and  $K > 0$  such that,

$$\mathcal{A}V \leq -\theta V + b1_{B(0, K)} . \quad (7)$$

It is shown in [29, Theorem 2.2] that, for any Lyapunov function  $V$ , the Langevin diffusion is  $V$ -uniformly geometrically ergodic. [29, Theorem 2.3] shows that if there exists  $a \in (0, 1)$ ,  $c > 0$ , and  $K \geq 0$  such that  $a \|\nabla U(x)\|^2 - \Delta U(x) \geq c$  for all  $\|x\| \geq K$ , then (7) is satisfied with constants that can be computed explicitly. Another condition is that  $\pi$  is (super)-exponential in the tails.

**H2** ( $\alpha$ ). *The potential  $U$  is differentiable and there exist  $\rho > 0$ ,  $\alpha \geq 1$  and  $M_\rho \geq 0$  such that for all  $y \in \mathbb{R}^d$ ,  $\|y\| \geq M_\rho$ ,  $\langle \nabla U(y), y \rangle \geq \rho \|y\|^\alpha$*

**Proposition 1.** *Assume **H1** and **H2**( $\alpha$ ). Then,*

(i) *Under **H2**(1), the drift condition (7) is satisfied with the Lyapunov function  $W(x) = \exp((\rho/4)(\|x\|^2 + 1)^{1/2})$ ,  $\theta_1 = \rho^2/8$ ,  $K = K_{\rho,1} = \max(M_\rho, 4d/\rho)$  and*

$$b_{\rho,1} = (\rho/4)(\rho/4 + d + \sup_{\{\|y\| \leq K_{\rho,1}\}} \{\|\nabla U(y)\|\}) \max(1, (K_{\rho,1}^2 + 1)^{-1/2} \exp(\rho(K_{\rho,1}^2 + 1)^{1/2}/4)) .$$

(ii) *Under **H2**( $\alpha$ ) with  $\alpha > 1$ , the drift condition (7) is satisfied with the Lyapunov function  $V_s(x) = \exp(sU(x))$ , for  $s \in (0, 1)$ ,  $\theta_{\alpha,s} = sdL$ ,  $K = K_{\rho,\alpha,s} = \max(\{2dL/(\rho(1-s))\}^{1/(2(\alpha-1))}, M_\rho)$  and  $b_{\rho,\alpha,s} = sdL \sup_{\{\|y\| \leq K_{\rho,\alpha,s}\}} \{V_U(y)\}$ .*

*Proof.* The proof is postponed to Section 6.1. □

**Corollary 2** ([25, Theorem 6.1]). *Assume **H1** and **H2**( $\alpha$ ) for some  $\alpha \geq 1$ . Set  $V(x) = W(x)$  if  $\alpha = 1$  and  $V(x) = V_s(x)$ , for  $s \in (0, 1)$  if  $\alpha > 1$ . Then, there exists constant  $C_\alpha < \infty$  and  $v_\alpha > 0$  such that for all probability measures  $\mu_0$  and  $\nu_0$  on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ , satisfying  $\mu_0(V) + \nu_0(V) < +\infty$ , and  $t \in \mathbb{R}_+$ ,*

$$\|\mu_0 P_t - \nu_0 P_t\|_V \leq C_\alpha e^{-v_\alpha t} \|\mu_0 - \nu_0\|_V, \quad \|\mu_0 P_t - \pi\|_V \leq C_\alpha e^{-v_\alpha t} \mu_0(V).$$

The constants  $C_\alpha$  and  $v_\alpha$  depends on the drift and minorization conditions. Computable bounds under **H2**( $\alpha$ ) have been developed in the literature but these estimates are in general very conservative. Much better rates of convergence can be obtained using either Poincaré (or spectral gap) and log-Sobolev inequalities; see [3], [7] and [4, Chapter 4] and the references therein. We follow here [7]. The starting point is the following extension of the Pinsker inequality.

**Lemma 3** ([7, Lemma 1.1]). *Let  $\psi$  be a  $C^2$  convex function defined on  $\mathbb{R}_+$ . Assume that  $\psi$  is uniformly convex on all bounded intervals,  $\psi(1) = 0$  and  $\lim_{u \rightarrow \infty} \psi(u)/u = +\infty$ . Then, for all probability measures  $(\mu, \nu)$  on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  such that  $\mu \ll \nu$ ,*

$$\|\mu - \nu\|_{\text{TV}} \leq c_\psi I_\psi^{1/2}(\mu|\nu), \quad \text{where } I_\psi(\mu, \nu) = \int \psi \left( \frac{d\mu}{d\nu} \right) d\nu,$$

and  $c_\psi$  is a universal constant.

If  $\psi(u) = (u - 1)^2$ , then  $I_\psi(\mu, \nu)$  is the chi-square distance,  $c_\psi = 1$  and Lemma 3 shows that  $\|\mu - \nu\|_{\text{TV}} \leq \text{Var}_\nu^{1/2} \{d\mu/d\nu\}$ . If  $\psi(u) = u \ln(u)$ , then  $I_\psi(\mu, \nu)$  is the Kullback-Leibler divergence,  $c_\psi = \sqrt{2}$  and Lemma 3 implies  $\|\mu - \nu\|_{\text{TV}} \leq (2 \text{KL}(\mu|\nu))^{1/2}$ .

**Proposition 4** ([7, Proposition 1.8]). *Let  $\psi$  be a function satisfying the conditions of Lemma 3. Assume that there exists a constant  $C_\psi$  such that for any density function  $h$  in the extended domain of the generator satisfying  $\int \psi(h) d\pi < \infty$ ,*

$$\int \psi(h) d\pi \leq C_\psi \int \psi''(h) \mathcal{G}(h) d\pi. \quad (8)$$

Then, for all  $t \geq 0$ , and any initial distribution  $\mu_0$  such that  $\mu_0 \ll \pi$ ,

$$\|\mu_0 P_t - \pi\|_{\text{TV}} \leq c_\psi e^{-t/C_\psi} I_\psi^{1/2}(\mu_0|\pi).$$

If we choose  $\psi(u) = (u - 1)^2$ , (8) boils down to the *Poincaré inequality*. If there exists a constant  $C_P$  such that, for any density  $h \in \text{D}(\mathcal{A})$   $\text{Var}_\pi \{h\} \leq C_P \int \mathcal{G}(h) d\pi$ , then Proposition 4 implies that for all  $t \geq 0$ ,

$$\|\mu_0 P_t - \pi\|_{\text{TV}} \leq \exp(-t/C_P) (\text{Var}_\pi \{d\mu_0/d\pi\})^{1/2}. \quad (9)$$

The best constant  $C_P$  for which this inequality holds is referred to as the *Poincaré constant*. To determine upper bounds of the Poincaré constants, some additional conditions are required.

**Proposition 5** ([2, Theorem 1.4]). *Assume **H1** and that there exists a Lyapunov function satisfying (7). Then  $\pi$  satisfies a Poincaré inequality with constant*

$$C_{\text{lyap}} = \theta^{-1} \left\{ 1 + (4bK^2/\pi^2) e^{\text{osc}_K(U)} \right\} \quad (10)$$

where  $\text{osc}_K(U) = \sup_{\text{B}(0,K)}(U) - \inf_{\text{B}(0,K)}(U)$  is the oscillation of  $U$  on  $\text{B}(0, K)$ .

Combining (10) with Proposition 1 and Proposition 4 yields explicit convergence bounds for target distribution with exponential tails. Typically, the term  $e^{\text{osc}_R(U)}$  scales exponentially with the dimension. Better Poincaré constant may be obtained by assuming stronger conditions on the distribution  $\pi$ .

**Proposition 6** ([5, Theorem 2.1]). *Assume **H1** and  $U$  is convex. Then,  $\pi$  satisfies a Poincaré inequality with constant  $C_{\text{cvx}}$  given by*

$$C_{\text{cvx}} = 432 \int_{\mathbb{R}^d} \left\{ x - \int_{\mathbb{R}^d} y d\pi(y) \right\}^2 d\pi(x) .$$

Applying Proposition 4 with  $\psi(u) = u \log(u)$ , (8) leads to the log-Sobolev inequality. If there exists some constant  $C_{\text{LS}}$  such that, for any density  $h \in D(\mathcal{A})$ ,  $\text{Ent}_\pi(h) \leq C_{\text{LS}} \int h^{-1} \mathcal{G}(h) d\pi$  then for all  $t \geq 0$ ,

$$\|\mu_0 P_t - \pi\|_{\text{TV}} \leq \exp(-t/C_{\text{LS}}) (2\text{Ent}_\pi(d\mu_0/d\pi))^{1/2} . \quad (11)$$

To successfully use the Log-Sobolev inequality, it is required to strengthen the conditions on the potential  $U$ .

**H3.**  $U$  is strongly convex, i.e. there exists  $m > 0$  such that for all  $x, y \in \mathbb{R}^d$ ,

$$U(y) \geq U(x) + \langle \nabla U(x), y - x \rangle + (m/2) \|x - y\|^2 .$$

Under **H3**, by [26, Theorem 2.1.8], there exists a unique  $x^* \in \mathbb{R}^d$  such that  $x^* \stackrel{\text{def}}{=} \arg \min_{\mathbb{R}^d} U$ . By [26, Theorem 2.1.12, Theorem 2.1.9], condition **H3** implies that for all  $x, y \in \mathbb{R}^d$ :

$$\langle \nabla U(y) - \nabla U(x), y - x \rangle \geq (\kappa/2) \|y - x\|^2 + \frac{1}{m+L} \|\nabla U(y) - \nabla U(x)\|^2 , \quad (12)$$

$$\langle \nabla U(y) - \nabla U(x), y - x \rangle \geq m \|y - x\|^2 , \quad (13)$$

where

$$\kappa = \frac{2mL}{m+L} . \quad (14)$$

Note that **H1** and (13) imply that  $L \geq m$ .

**Theorem 7.** *Assume **H1** and **H3**. Then for all probability measure  $\mu_0 \ll \pi$  such that  $d\mu_0/d\pi \log(d\mu_0/d\pi) \in L^1(\pi)$ , we have*

$$\|\mu_0 P_t - \pi\|_{\text{TV}} \leq e^{-mt} \left( 2\text{Ent}_\pi \left( \frac{d\mu_0}{d\pi} \right) \right)^{1/2} . \quad (15)$$

It is worthwhile to note that, in such case, the convergence rate does not depend on the dimension  $d$  but only on the convexity constants  $m$ . In such case, is also possible to devise convergence in Wasserstein distance. This type of convergence will prove to be very important to obtain precise estimates in the deviation inequalities. Here again, the rate of convergence does not depend on the dimension.

**Theorem 8.** *Assume **H1** and **H3**.*

(i) *The stationary distribution  $\pi$  satisfies*

$$\int_{\mathbb{R}^d} \|x - x^*\|^2 \pi(dx) \leq d/m . \quad (16)$$

(ii) For all probability measures  $\mu_0$  and  $\nu_0 \in \mathcal{P}_2(\mathbb{R}^d)$  and  $t \geq 0$ ,

$$W_2(\mu_0 P_t, \nu_0 P_t) \leq e^{-mt} W_2(\mu_0, \nu_0)$$

*Proof.* The proof is postponed to Section 6.2.  $\square$

### 3. Euler approximations of the Langevin equation

Let  $(\gamma_k)_{k \geq 1}$  be a sequence of positive and non-increasing step sizes and for  $n, p \in \mathbb{N}$ , denote by

$$\Gamma_{n,p} \stackrel{\text{def}}{=} \sum_{k=n}^p \gamma_k, \quad \Gamma_n = \Gamma_{1,n}. \quad (17)$$

For  $\gamma > 0$ , consider the Markov kernel  $R_\gamma$  given for all  $A \in \mathcal{B}(\mathbb{R}^d)$  and  $x \in \mathbb{R}^d$  by

$$R_\gamma(x, A) = \int_A (4\pi\gamma)^{-d/2} \exp\left(- (4\gamma)^{-1} \|y - x + \gamma \nabla U(x)\|^2\right) dy. \quad (18)$$

The sequence  $(X_n)_{n \geq 0}$  given in (2) is a Markov chain with respect to the sequence of Markov kernels  $(R_{\gamma_n})_{n \geq 1}$ . For  $p, n \geq 1, p \geq n$ , define

$$Q_\gamma^{n,p} = R_{\gamma_n} \cdots R_{\gamma_p}, \quad Q_\gamma^n = Q_\gamma^{1,n} \quad (19)$$

with the convention that for  $n, p \geq 0, n < p, Q_\gamma^{p,n}$  is the identity operator. The stability of the Euler discretization of a one-dimensional Langevin diffusion with constant step size has been studied in [29, Section 3]; We generalize these results to multidimensional diffusions and decreasing stepsizes. We first establish a geometric drift condition for  $R_\gamma$  with explicit constants.

**Proposition 9.** *Assume **H1** and **H2**( $\alpha$ ) with  $\alpha > 1$ . Then for all  $\bar{\gamma} \in (0, L^{-1})$ , we have for all  $\gamma \in (0, \bar{\gamma}]$  and  $x \in \mathbb{R}^d$ ,*

$$R_\gamma V_{1/2}(x) \leq e^{-s_\gamma \gamma} V_{1/2}(x) + \gamma c_{\bar{\gamma}, \rho, \alpha} \mathbf{1}_{B(0, K_{\bar{\gamma}, \rho, \alpha})},$$

where  $V_{1/2}(x) = \exp(U(x)/2)$ ,  $s_\gamma = dL(2(1 - L\bar{\gamma}))^{-1}$ ,  $K_{\bar{\gamma}, \rho, \alpha} = \max(M_\rho, (8s_\gamma/\rho)^{1/(2(\alpha-1))})$  and

$$c_{\bar{\gamma}, \rho, \alpha} = 2s_\gamma e^{s_\gamma \bar{\gamma}} \sup_{\{\|y\| \leq K_{\bar{\gamma}, \rho, \alpha}\}} V_{1/2}(y).$$

*Proof.* The proof is postponed to Section 6.3.  $\square$

**Corollary 10.** *Assume **H1** and **H2**( $\alpha$ ) with  $\alpha > 1$ . Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 < L^{-1}$ . Then for all  $n \geq 1$*

$$Q_\gamma^n V_{1/2}(x) \leq e^{-s_{\gamma_1} \Gamma_{1,n}} V_{1/2}(x) + c_{\gamma_1, \rho, \alpha} s_{\gamma_1}^{-1} e^{s_{\gamma_1} \gamma_1}.$$

*Proof.* The proof is postponed to Section 6.3.  $\square$

The total variation distance of the target probability measure  $\pi$  with the sequence  $(\mu_0 Q_\gamma^n)_{n \geq 0}$ , where  $\mu_0$  is the initial distribution, is decomposed as follows: for all  $n \geq 0$ ,  $p \geq 1$  and  $n < p$ ,

$$\|\mu_0 Q_\gamma^p - \pi\|_{\text{TV}} \leq \|\mu_0 Q_\gamma^n Q_\gamma^{n+1,p} - \mu_0 Q_\gamma^n P_{\Gamma_{n+1,p}}\|_{\text{TV}} + \|\mu_0 Q_\gamma^n P_{\Gamma_{n+1,p}} - \pi\|_{\text{TV}}. \quad (20)$$

To control the first term on the right hand side, we use a method borrowed from [11]. For  $0 \leq s \leq t$ , let  $C([s, t], \mathbb{R}^d)$  be the space of continuous function on  $[s, t]$  taking values in  $\mathbb{R}^d$ . For all  $x \in \mathbb{R}^d$ , denote by  $\mu_{n,p}^x$  and  $\bar{\mu}_{n,p}^x$  the laws on  $C([\Gamma_n, \Gamma_p], \mathbb{R}^d)$  of the Langevin diffusion  $(Y_t)_{\Gamma_n \leq t \leq \Gamma_p}$  and of the continuously-interpolated Euler discretisation  $(\bar{Y}_t)_{\Gamma_n \leq t \leq \Gamma_p}$  both started at  $x$  at time  $\Gamma_n$ . In the sequel, we use the synchronous coupling  $(Y_t, \bar{Y}_t)_{\Gamma_n \leq t \leq \Gamma_p}$  of these two processes. For any  $\zeta_0 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ , consider the diffusion  $(Y_t, \bar{Y}_t)_{t \geq 0}$  with initial distribution equals to  $\zeta_0$ , and defined for  $t \geq 0$  by

$$\begin{cases} dY_t = -\nabla U(Y_t)dt + \sqrt{2}dB_t \\ d\bar{Y}_t = -\overline{\nabla U}(\bar{Y}_t)dt + \sqrt{2}dB_t \end{cases} \quad (21)$$

where  $(\Gamma_n)_{n \geq 1}$  is given in (17) and for any continuous function  $y : \mathbb{R}_+ \rightarrow \mathbb{R}^d$ ,  $t \mapsto y_t$

$$\overline{\nabla U}(y) = \sum_{k=0}^{\infty} \nabla U(y_{\Gamma_n}) 1_{[\Gamma_n, \Gamma_{n+1})}(t) \quad (22)$$

The Girsanov theorem [20, Theorem 5.1, Corollary 5.16, Chapter 3] shows that  $\mu_{n,p}^x$  and  $\bar{\mu}_{n,p}^x$  are mutually absolutely continuous and in addition,

$$\frac{d\mu_{n,p}^x}{d\bar{\mu}_{n,p}^x}(\bar{Y}) = \exp \left( \frac{1}{2} \int_{\Gamma_n}^{\Gamma_p} \langle \nabla U(\bar{Y}_s) - \overline{\nabla U}(\bar{Y}_s), d\bar{Y}_s \rangle - \frac{1}{4} \int_{\Gamma_n}^{\Gamma_p} \left\{ \|\nabla U(\bar{Y}_s)\|^2 - \|\overline{\nabla U}(\bar{Y}_s)\|^2 \right\} ds \right) \quad (23)$$

where  $\overline{\nabla U}$  is defined in (22). The Pinsker inequality implies that for all  $x \in \mathbb{R}^d$

$$\|\delta_x Q_\gamma^{n+1,p} - \delta_x P_{\Gamma_{n+1,p}}\|_{\text{TV}} \leq 2^{-1/2} \left( \int_{\Gamma_n}^{\Gamma_p} \mathbb{E}_x \left[ \|\nabla U(\bar{Y}_s) - \overline{\nabla U}(\bar{Y}_s)\|^2 \right] ds \right)^{1/2}.$$

Under **H1**, we get

$$\|\delta_x Q_\gamma^{n+1,p} - \delta_x P_{\Gamma_{n+1,p}}\|_{\text{TV}} \leq 2^{-1/2} L \left( \sum_{k=n}^{p-1} \left\{ (\gamma_{k+1}^3/3) \mathbb{E}_x \left[ \|\nabla U(X_k)\|^2 \right] + d\gamma_{k+1}^2 \right\} \right)^{1/2}. \quad (24)$$

Therefore, the first term of (20) goes to 0 as  $p$  goes to  $+\infty$  provided that the remainder  $\sum_{k=n}^p \gamma_k^2$  goes to 0 as  $n, p$  goes to  $+\infty$ . If a bound for  $\sup_{k \geq 0} \mathbb{E}_x [\|\nabla U(X_k)\|^2]$  is known then the convergence given by (24) is quantitative. Such a bound is implied by drift conditions for the kernels  $(R_{\gamma_k})_{k \geq 1}$ ; see Corollary 10. On the other hand, under weak additional conditions, the second term on the right hand side of (20),  $\|\mu_0 Q_\gamma^n P_{\Gamma_{n+1,p}} - \pi\|_{\text{TV}}$ , goes to 0 if  $\Gamma_{n+1,p}$  goes to  $+\infty$  as  $n, p \rightarrow \infty$ .

**Theorem 11.** Assume **H1** and **H2**( $\alpha$ ) for some  $\alpha > 1$ . Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 < L^{-1}$ . Then, for all  $n \geq 0$ ,  $p \geq 1$ ,  $n < p$ , and  $x \in \mathbb{R}^d$

$$\|\delta_x Q_\gamma^p - \pi\|_{\text{TV}} \leq C_\alpha e^{-v_\alpha \Gamma_{n+1,p}} \{e^{-s_\gamma \Gamma_{1,n}} V_{1/2}(x) + c_{\gamma,\rho,\alpha} s_\gamma^{-1} e^{\gamma_1 s_\gamma}\} + B_{n,p}(x; \gamma),$$

where  $C_\alpha$  and  $v_\alpha$  are given by Corollary 2,  $s_\gamma$  and  $c_{\gamma,\rho,\alpha}$  by Proposition 9 and

$$B_{n,p}(x; \gamma) = 2^{-1/2} L \left( \sum_{k=n}^{p-1} \left\{ (\gamma_{k+1}^3/3) \|\nabla U\|_{V_{1/2}}^2 (V_{1/2}(x) + c_{\gamma,\rho,\alpha} s_\gamma^{-1} e^{\gamma_1 s_\gamma}) + d\gamma_{k+1}^2 \right\} \right)^{1/2}. \quad (25)$$

If in addition  $\lim_{k \rightarrow \infty} \gamma_k = 0$  and  $\lim_{n \rightarrow +\infty} \Gamma_n = +\infty$ , then for all  $x \in \mathbb{R}^d$ , we get  $\lim_{n \rightarrow \infty} \|\delta_x Q_\gamma^n - \pi\|_{\text{TV}} = 0$ .

*Proof.* The proof is postponed to Section 6.4.  $\square$

When  $\pi$  satisfies a Poincaré inequality an explicit bound for  $\|\delta_x Q_\gamma^p - \pi\|_{\text{TV}}$  for all  $p \geq 1$  and  $x \in \mathbb{R}^d$  can be obtained.

**Theorem 12.** Assume **H1** and  $\pi$  satisfies a Poincaré inequality with constant  $C_P$ . Let  $(\gamma_k)_{k \geq 1}$  be a non increasing sequence with  $\gamma_1 < L^{-1}$ .

(i) Under **H2**( $\alpha$ ), then for all  $n \geq 0$ ,  $p \geq 1$ ,  $n < p$  and  $x \in \mathbb{R}^d$ ,

$$\|\delta_x Q_\gamma^p - \pi\|_{\text{TV}} \leq A_{n,p}(x; \gamma) e^{-\Gamma_{n+1,p}/C_P} + B_{n,p}(x; \gamma)$$

where  $B_{n,p}(x; \gamma)$  is given by (25),

$$A_{n,p}(x; \gamma) = 2^{d/4} \left( e^{U(0)} \left( \frac{8\sqrt{d}}{3\rho} \right)^d + \|\exp\{-U\}\|_\infty \frac{\pi^{d/2} (2 \max(1, M_\rho))^d}{\Gamma(d/2 + 1)} \right)^{1/2} \\ \times \left( 4\pi \left\{ \prod_{k=1}^n (1 - L\gamma_k) \right\}^2 \sum_{i=1}^n \gamma_i (1 - L\gamma_i)^{-1} \right)^{-d/2} V_{1/2}(x),$$

and  $\Gamma : \mathbb{R}_+^* \rightarrow \mathbb{R}_+$  is the Gamma function defined for  $t > 0$  by  $\Gamma(t) = \int_0^{+\infty} s^{t-1} e^{-s} ds$ .

(ii) If  $\nabla U$  is bounded, then then for all  $n \geq 0$ ,  $p \geq 1$ ,  $n < p$  and  $x \in \mathbb{R}^d$ ,

$$\|\delta_x Q_\gamma^p - \pi\|_{\text{TV}} \leq A_{n,p}(x; \gamma) e^{-\Gamma_{n+1,p}/C_P} + 2^{-1/2} L \left( \sum_{k=n}^{p-1} \{(\gamma_{k+1}^3/3) \|\nabla U\|_\infty^2 + d\gamma_{k+1}^2\} \right)^{1/2},$$

*Proof.* The proof is postponed to Section 6.5.  $\square$

Corollary 2 shows that  $(P_t)_{t \geq 0}$  is a contraction operator on the space of probability measure  $\mu$ ,  $\mu(V_{1/4}) < +\infty$ , endowed with  $\|\cdot\|_{V_{1/4}}$ . It is therefore possible to control  $\|\delta_x Q_\gamma^p - \pi\|_{V_{1/4}}$ . To simplify the notations, we limit our discussion to constant stepsizes.

**Theorem 13.** Assume **H1** and **H2**( $\alpha$ ) for some  $\alpha > 1$  and that  $\gamma_k = \gamma$  for all  $k \geq 1$ , where  $\gamma < L^{-1}$ . Then, for all  $n \geq 1$  and  $x \in \mathbb{R}^d$

$$\begin{aligned} \|\delta_x Q_\gamma^p - \pi\|_{V_{1/4}} &\leq C_\alpha e^{-v_\alpha \gamma^n} V_{1/4}(x) \\ &+ 2^{1/2} L \max(1, C_\alpha) (1 - e^{-v_\alpha})^{-1} \left( 2(V_{1/2}(x) + c_{\gamma, \rho, \alpha} s_\gamma^{-1} e^{\gamma s_\gamma}) + b_{\rho, \alpha, 1/2} / \theta_{\alpha, 1/2} \right)^{1/2} \\ &\quad \times \left( (\gamma + \gamma^2)d + 3^{-1}(\gamma^2 + \gamma^3) \|\nabla U\|_{V_{1/2}^{1/2}}^2 (V_{1/2}(x) + c_{\gamma, \rho, \alpha} s_\gamma^{-1} e^{\gamma s_\gamma}) \right)^{1/2}, \end{aligned}$$

where  $C_\alpha, v_\alpha$  are defined in Corollary 2,  $s_\gamma, c_{\gamma, \rho, \alpha}$  in Proposition 9, and  $b_{\rho, \alpha, 1/2}, \theta_{\alpha, 1/2}$  in Proposition 1-(ii).

*Proof.* The proof is postponed to Section 6.6.  $\square$

This result gives a bound of order  $\mathcal{O}(\gamma^{1/2})$  as  $n$  goes to infinity. If we assume that the total number of iterations is fixed (finite horizon setting), a straightforward optimization in  $\gamma$  leads to a bound of order  $\mathcal{O}((\log(n)n^{-1})^{1/2})$ . Using the "doubling trick" (see [18]) this can be converted to an anytime algorithm with a bound on  $\|\delta_x Q_\gamma^n - \pi\|_{\text{TV}}$  of order  $\mathcal{O}((\log(n)n^{-1})^{1/2})$  for all  $x \in \mathbb{R}^d$  and  $n \geq 1$ .

Under **H3**, much better estimates can be obtained. We show that discretization (for fixed as well as vanishing step sizes) is stable as soon as the stepsize is small enough. We then obtain explicit computable bounds for these moments.

**Theorem 14.** Assume **H1** and **H3**. Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 \leq 2/(m+L)$ . Let  $x^*$  be the unique minimizer of  $U$ . Then for all  $n, p \geq 1$ ,  $n \leq p$

$$\int_{\mathbb{R}^d} \|x - x^*\|^2 \mu_0 Q_\gamma^{n,p}(dx) \leq \prod_{k=n}^p (1 - \kappa \gamma_k) \int_{\mathbb{R}^d} \|x - x^*\|^2 \mu_0(dx) + 2d\kappa^{-1},$$

where  $\kappa$  is defined in (14)

*Proof.* The proof is postponed to Section 6.7.  $\square$

As seen in Theorem 7, **H3** implies exponential convergence of  $(\mu_0 P_t)_{t \geq 0}$  to  $\pi$  for a suitable initial probability measure  $\mu_0$ , with explicit constant independent of the dimension. We now deal with case when  $\pi$  satisfies **H3**, and deduce some new bounds for  $\|\delta_x Q_\gamma^n - \pi\|_{\text{TV}}$ . In particular, the obtained bounds are now linear in the dimension. We use again the decomposition given by (20), but use the log-Sobolev inequality instead.

**Theorem 15.** Assume **H1** and **H3**. Let  $(\gamma_k)_{k \geq 1}$  be a non increasing sequence with  $\gamma_1 \leq 2/(m+L)$ . Let  $x^*$  be the unique minimizer of  $U$ . Then for all  $n \geq 0$ ,  $p \geq 1$ ,  $n < p$  and  $x \in \mathbb{R}^d$ ,

$$\|\delta_x Q_\gamma^p - \pi\|_{\text{TV}} \leq C_{n,p}(x, \gamma) e^{-\Gamma_{n+1,p}/m} + D_{n,p}(x, \gamma)$$

where

$$C_{n,p}(x, \gamma) = (L/2) \left\{ \prod_{k=1}^n (1 - \kappa\gamma_k) \right\} \|x - x^*\|^2 + dL\kappa^{-1}(1 - \kappa\gamma_n) \\ - (d/2)(1 + \log(2\gamma_n/m) - 2L\gamma_n) \\ D_{n,p}(x, \gamma) = 2^{-1/2}L \left( \sum_{k=n}^{p-1} (\gamma_{k+1}^3 L^2/3) \left\{ \prod_{i=1}^k (1 - \kappa\gamma_i) \|x - x^*\|^2 + 2d\kappa^{-1} \right\} + d\gamma_{k+1}^2 \right)^{1/2}.$$

*Proof.* The proof is postponed to Section 6.8.  $\square$

We now proceed to establish explicit bounds for  $W_2(\mu_0 Q_\gamma^n, \pi)$ , with  $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ . Since  $\pi$  is invariant for  $P_t$  for all  $t \geq 0$ , it suffices to get some bounds on  $W_2(\mu_0 Q_\gamma^n, \nu_0 P_{\Gamma_n})$ , with  $\nu_0 \in \mathcal{P}_2(\mathbb{R}^d)$  and take  $\nu_0 = \pi$ . To do so, we construct a coupling. In the strongly convex case, an obvious candidate is the synchronous coupling (21).

Therefore since for all  $n \geq 0$ ,  $W_2^2(\mu_0 P_{\Gamma_n}, \nu_0 Q_\gamma^n) \leq \mathbb{E}[\|Y_{\Gamma_n} - \bar{Y}_{\Gamma_n}\|^2]$ , where  $\mu_0$  and  $\nu_0$  are the marginals of  $\zeta_0$ , we compute an explicit bound of the Wasserstein distance between the sequence of distributions  $(\mu_0 Q_\gamma^n)_{n \geq 0}$  and the stationary measure  $\pi$  of the Langevin diffusion (1).

**Theorem 16.** *Assume H1 and H3. Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 \leq 1/(m+L)$ . Then for all  $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$  and  $n \geq 1$ ,*

$$W_2^2(\mu_0 Q_\gamma^n, \pi) \leq u_n^{(1)}(\gamma) W_2^2(\mu_0, \pi) + u_n^{(2)}(\gamma), \quad (26)$$

where

$$u_n^{(1)}(\gamma) \stackrel{\text{def}}{=} \prod_{k=1}^n (1 - \kappa\gamma_k/2) \quad (27)$$

and

$$u_n^{(2)}(\gamma) \stackrel{\text{def}}{=} L^2 \sum_{i=1}^n \gamma_i^2 \{ \kappa^{-1} + \gamma_i \} (2d + dL^2\gamma_i/m + dL^2\gamma_i^2/6) \prod_{k=i+1}^n (1 - \kappa\gamma_k/2), \quad (28)$$

where  $\kappa$  is defined in (14).

*Proof.* The proof is postponed to Section 6.9.  $\square$

Let  $x^*$  be the unique minimizer of  $U$ . Since for all  $y \in \mathbb{R}^d$   $\|x - y\|^2 \leq 2(\|x - x^*\|^2 + \|x^* - y\|^2)$ , using (16), we get:

$$W_2^2(\delta_x, \pi) \leq 2(\|x - x^*\|^2 + d/m). \quad (29)$$

If  $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ , we have  $W_2^2(\mu_0, \pi) \leq \int \mu_0(dx) W_2^2(\delta_x, \pi)$ . Hence, the right hand side of (26) scales linearly with the dimension  $d$ . When  $\gamma_k = \gamma$  for all  $k \geq 1$ , (27) and (28) imply

$$u_n^{(1)}(\gamma) = (1 - \kappa\gamma/2)^n, \quad u_n^{(2)}(\gamma) \leq 2\kappa^{-1}\gamma \{ \kappa^{-1} + \gamma \} (2d + dL^2\gamma/m + dL^2\gamma^2/6). \quad (30)$$

Using this bound, given  $\epsilon > 0$ , we may determine the smallest number of iterations and an associated step-size  $\gamma$ , starting from  $x$ , to approach the stationary distribution in the Wasserstein distance

$W_2(\delta_x Q_n^\gamma, \pi)$  with a precision  $\epsilon$ . Details and further discussions are included in the supplementary paper [12].

We now consider decreasing stepsizes. Under this additional assumption, we may establish the convergence of the sequence  $(\mu_0 Q_n^\gamma)_{n \geq 0}$  to  $\pi$ .

**Corollary 17.** *Assume **H1** and **H3**. Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 \leq 1/(m+L)$ . Assume that  $\lim_{k \rightarrow \infty} \gamma_k = 0$  and  $\lim_{n \rightarrow +\infty} \Gamma_n = +\infty$ . Then for all  $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $\lim_{n \rightarrow \infty} W_2(\mu_0 Q_n^\gamma, \pi) = 0$ .*

*Proof.* The proof is postponed to Section 6.9.  $\square$

Based on Theorem 16, we can obtain explicit bounds for  $W_2^2(\delta_x Q_n^\gamma, \pi)$  for all  $x \in \mathbb{R}^d$ . For simplicity, we consider sequences  $(\gamma_k)_{k \geq 1}$  defined for all  $k \geq 1$  by  $\gamma_k = \gamma_1 k^{-\alpha}$ , for  $\gamma_1 < 1/(m+L)$  and  $\alpha \in (0, 1]$ . The order of these bounds is given in Table 1, see [12] for details. Two regimes can be observed as in stochastic approximation. When  $\alpha \in (0, 1)$ , the rate of convergence is of order  $n^{-\alpha}$ ; for  $\alpha = 1$ , the rate of convergence is  $n^{-1}$ , for  $\gamma_1 > 2\kappa^{-1}$ , see [12, Section 2]. We now consider the

	$\alpha \in (0, 1)$	$\alpha = 1$
Order of convergence	$2^{2+\alpha} \kappa^{-2} \gamma_1 L^2 d n^{-\alpha}$	$\mathcal{O}(n^{-1})$ for $\gamma_1 > 2\kappa^{-1}$

TABLE 1

Order of convergence of  $W_2(\delta_x Q_n^\gamma, \pi)$  for  $\gamma_k = \gamma_1 k^{-\alpha}$

fixed horizon setting. Assuming here that the step sizes  $(\gamma_k)_{k \geq 1}$  are defined for  $k \geq 1$  by  $\gamma_k = \gamma_1 k^{-\alpha}$  for  $\alpha \in [0, 1)$ , we determine the value of  $\gamma_1$  minimizing the upper bound  $u_n^{(1)}(\gamma) W_2^2(\mu_0, \pi) + u_n^{(2)}(\gamma)$ . The results are summarized in Table 2, see [12] for details. In essence, this result shows that choosing

	Optimal choice of $\gamma_1$	Bound on $W_2^2(\delta_x Q_n^\gamma, \pi)$
$\alpha \in [0, 1)$	$2(1-\alpha)\kappa^{-1}(2/n)^{1-\alpha} \log(\kappa n / (2(1-\alpha)))$	$16(1-\alpha)L^2 \kappa^{-3} d n^{-1} \log(\kappa n / (2(1-\alpha)))$

TABLE 2

Order of the optimal choice of  $\gamma_1$  for the fixed horizon setting and implied bound on  $W_2^2(\delta_x Q_n^\gamma, \pi)$

$\gamma_1$  of order  $\log(n)n^{\alpha-1}$  leads a bound for  $W_2^2(\delta_x Q_n^\gamma, \pi)$  of order  $\log(n)/n$ . Moreover, these bounds for a fixed number of iterations implies using the doubling trick (see [18]) an anytime algorithm which guarantees for all  $n \geq 1$  and  $x \in \mathbb{R}^d$  that  $W_2(\delta_x Q_n^\gamma, \pi)$  is  $\mathcal{O}((\log(n)n^{-1})^{1/2})$ .

#### 4. Mean square error and concentration for strongly log-concave distribution

Let be  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  and  $(X_n)_{n \geq 0}$  the Euler discretization of the Langevin diffusion. In this section we study the approximation of  $\int_{\mathbb{R}^d} f(y) \pi(dy)$  by the weighted average estimator

$$\hat{\pi}_n^N(f) = \sum_{k=N+1}^{N+n} \omega_{k,n}^N f(X_k), \quad \omega_{k,n}^N = \gamma_{k+1} \Gamma_{N+2, N+n+1}^{-1}. \quad (31)$$

where  $N \geq 0$  is the length of the burn-in period,  $n \geq 1$  is the number of samples, and for  $n, p \in \mathbb{N}$ ,  $\Gamma_{n,p}$  is given by (17). We restrict the discussion to Lipschitz functions  $f$ . In all this section,  $\mathbb{P}_x$

and  $\mathbb{E}_x$  denote the probability and the expectation respectively, induced on  $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\mathbb{N}})$  by the Markov chain  $(X_n)_{n \geq 0}$  started at  $x \in \mathbb{R}^d$ . We first compute an explicit bounds for the Mean Squared Error (MSE) of this estimator defined by:

$$\text{MSE}_f(N, n) = \mathbb{E}_x \left[ |\hat{\pi}_n^N(f) - \pi(f)|^2 \right] = \left\{ \mathbb{E}_x [\hat{\pi}_n^N(f) - \pi(f)] \right\}^2 + \text{Var}_x \{ \hat{\pi}_n^N(f) \}. \quad (32)$$

We first obtain an elementary bound for the bias. For all  $k \in \{N+1, \dots, N+n\}$ , let  $\xi_k$  be the optimal transference plan between  $\delta_x Q_\gamma^k$  and  $\pi$  for  $W_2$ . Then by the Jensen inequality and because  $f$  is Lipschitz, we have:

$$\begin{aligned} \left( \mathbb{E}_x [\hat{\pi}_n^N(f)] - \pi(f) \right)^2 &= \left( \sum_{k=N+1}^{N+n} \omega_{k,n}^N \int_{\mathbb{R}^d \times \mathbb{R}^d} \{f(z) - f(y)\} \xi_k(dz, dy) \right)^2 \\ &\leq \|f\|_{\text{Lip}}^2 \sum_{k=N+1}^{N+n} \omega_{k,n}^N \int_{\mathbb{R}^d \times \mathbb{R}^d} \|z - y\|^2 \xi_k(dz, dy). \end{aligned}$$

Using Theorem 16, we end up with the following bound.

**Proposition 18.** *Assume **H1** and **H3**. Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 \leq 1/(m+L)$ . Let  $x^*$  be the unique minimizer of  $U$ . Let  $(X_n)_{n \geq 0}$  be given by (2) and started at  $x \in \mathbb{R}^d$ . Then for all  $n, N \geq 0$  and Lipschitz function  $f: \mathbb{R}^d \rightarrow \mathbb{R}$ :*

$$\left\{ \mathbb{E}_x [\hat{\pi}_n^N(f)] - \pi(f) \right\}^2 \leq \|f\|_{\text{Lip}}^2 \sum_{k=N+1}^{N+n} \omega_{k,n}^N \left\{ 2(\|x - x^*\|^2 + d/m) u_k^{(1)}(\gamma) + u_k^{(2)}(\gamma) \right\},$$

where  $u_n^{(1)}(\gamma)$  and  $u_n^{(2)}(\gamma)$  are given in (27) and (28), respectively.

Consider now the variance term. Our main tool is the Gaussian Poincaré inequality [6, Theorem 3.20] (see also [4, Theorem 4.1.1]) which states that if  $Z = (Z_1, \dots, Z_d)$  is a Gaussian vector with identity covariance matrix, then  $\text{Var} \{g(Z)\} \leq \|g\|_{\text{Lip}}^2$ . The Gaussian Poincaré inequality may be applied to  $R_\gamma$  defined by (18) noticing that for all  $y \in \mathbb{R}^d$ ,  $R_\gamma(y, \cdot)$  is a Gaussian distribution with mean  $y - \gamma \nabla U(y)$  and covariance matrix  $2\gamma \text{Id}$ .

**Lemma 19.** *Assume **H1**. Let  $g: \mathbb{R}^d \rightarrow \mathbb{R}$  be a Lipschitz function. Then for all  $\gamma > 0$ ,  $y \in \mathbb{R}^d$ ,*

$$0 \leq R_\gamma \{g(\cdot) - R_\gamma g(y)\}^2(y) = \int R_\gamma(y, dz) \{g(z) - R_\gamma g(y)\}^2 \leq 2\gamma \|g\|_{\text{Lip}}^2.$$

To go further, we decompose  $\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)]$  as the sum of martingale increments,

$$\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)] = \sum_{k=N}^{N+n-1} \left\{ \mathbb{E}_x^{\mathcal{G}_{k+1}} [\hat{\pi}_n^N(f)] - \mathbb{E}_x^{\mathcal{G}_k} [\hat{\pi}_n^N(f)] \right\} + \mathbb{E}_x^{\mathcal{G}_N} [\hat{\pi}_n^N(f)] - \mathbb{E}_x[\hat{\pi}_n^N(f)], \quad (33)$$

where  $(\mathcal{G}_n)_{n \geq 0}$  here is the natural filtration associated with Euler approximation  $(X_n)_{n \geq 0}$ . This implies that the variance may be expressed as the following sum

$$\begin{aligned} \text{Var}_x \{ \hat{\pi}_n^N(f) \} &= \sum_{k=N}^{N+n-1} \mathbb{E}_x \left[ \left( \mathbb{E}_x^{\mathcal{G}_{k+1}} [\hat{\pi}_n^N(f)] - \mathbb{E}_x^{\mathcal{G}_k} [\hat{\pi}_n^N(f)] \right)^2 \right] \\ &\quad + \mathbb{E}_x \left[ \left( \mathbb{E}_x^{\mathcal{G}_N} [\hat{\pi}_n^N(f)] - \mathbb{E}_x[\hat{\pi}_n^N(f)] \right)^2 \right]. \quad (34) \end{aligned}$$

Because  $\hat{\pi}_n^N(f)$  is an additive functional, the martingale increment  $\mathbb{E}_x^{\mathcal{G}^{k+1}} [\hat{\pi}_n^N(f)] - \mathbb{E}_x^{\mathcal{G}^k} [\hat{\pi}_n^N(f)]$  has a simple expression. For  $k = N + n - 1, \dots, N + 1$ , define backward in time the function

$$\Phi_{n,k}^N : x_k \mapsto \omega_{k,n}^N f(x_k) + R_{\gamma_{k+1}} \Phi_{n,k+1}^N(x_k), \quad (35)$$

where  $\Phi_{n,N+n}^N : x_{N+n} \mapsto \Phi_{n,N+n}^N(x_{N+n}) = \omega_{N+n,n}^N f(x_{N+n})$ . Denote finally

$$\Psi_n^N : x_N \mapsto R_{\gamma_{N+1}} \Phi_{n,N+1}^N(x_N). \quad (36)$$

Note that for  $k \in \{N, \dots, N + n - 1\}$ , by the Markov property,

$$\Phi_{n,k+1}^N(X_{k+1}) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(X_k) = \mathbb{E}_x^{\mathcal{G}^{k+1}} [\hat{\pi}_n^N(f)] - \mathbb{E}_x^{\mathcal{G}^k} [\hat{\pi}_n^N(f)], \quad (37)$$

and  $\Psi_n^N(X_N) = \mathbb{E}_x^{\mathcal{G}^N} [\hat{\pi}_n^N(f)]$ . With these notations, (34) may be equivalently expressed as

$$\begin{aligned} \text{Var}_x \{ \hat{\pi}_n^N(f) \} &= \sum_{k=N}^{N+n-1} \mathbb{E}_x \left[ \mathbb{E}_x^{\mathcal{G}^k} \left[ \{ \Phi_{n,k+1}^N(X_{k+1}) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(X_k) \}^2 \right] \right] + \text{Var}_x \{ \Psi_n^N(X_N) \} \\ &= \sum_{k=N}^{N+n-1} \mathbb{E}_x \left[ R_{\gamma_{k+1}} \left\{ \Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(X_k) \right\}^2 (X_k) \right] + \text{Var}_x \{ \Psi_n^N(X_N) \}. \end{aligned} \quad (38)$$

Now for  $k = N + n, \dots, N + 1$ , we will use the Gaussian Poincaré inequality (Lemma 19) to the sequence of function  $\Phi_{n,k}^N$  to prove that  $x \mapsto R_{\gamma_{k+1}} \{ \Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(x) \}^2(x)$  is uniformly bounded. It is required to bound the Lipschitz constant of  $\Phi_{n,k}^N$ . For  $k \in \{N, \dots, N + n - 1\}$  and for all  $y, z \in \mathbb{R}^d$ , we have

$$| \Phi_{n,k+1}^N(y) - \Phi_{n,k+1}^N(z) | = \left| \omega_{k+1,n}^N \{ f(y) - f(z) \} + \sum_{i=k+2}^{N+n} \omega_{i,n}^N \{ Q_\gamma^{k+2,i} f(y) - Q_\gamma^{k+2,i} f(z) \} \right|. \quad (39)$$

To bound  $\| Q_\gamma^{n,p} f \|_{\text{Lip}}$  for all  $n, p \in \mathbb{N}^*$ , we will bound for all initial distribution  $\mu_0, \nu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $W_2(\mu_0 Q_\gamma^{n,p}, \nu_0 Q_\gamma^{n,p})$ . This can be done using the discrete time-version of the synchronous coupling.

**Proposition 20.** *Assume H1 and H3. Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 \leq 2/(m+L)$ . Then*

(i) *for all  $\mu_0, \nu_0 \in \mathcal{P}_2(\mathbb{R}^d)$  and  $p \geq n \geq 1$ ,*

$$W_2^2(\mu_0 Q_\gamma^{n,p}, \nu_0 Q_\gamma^{n,p}) \leq \prod_{k=n}^p (1 - \kappa \gamma_k) W_2^2(\mu_0, \nu_0);$$

(ii) *for all Lipschitz functions  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  and  $p \geq n \geq 1$ ,  $Q_\gamma^{n,p} f$  is a Lipschitz function with*

$$\| Q_\gamma^{n,p} f \|_{\text{Lip}} \leq \prod_{k=n}^p (1 - \kappa \gamma_k)^{1/2} \| f \|_{\text{Lip}}.$$

*Proof.* The proof is postponed to Section 6.10. □

**Corollary 21.** Assume **H1** and **H3**. Let  $(\gamma_k)_{k \geq 1}$  be nonincreasing sequence with  $\gamma_1 \leq 2/(m+L)$ . Let  $N \geq 0$  and  $n \geq 1$ . Then for all  $y \in \mathbb{R}^d$ , Lipschitz function  $f$  and  $k \in \{N, \dots, N+n-1\}$ ,

$$R_{\gamma_{k+1}} \left\{ \Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(y) \right\}^2(y) \leq 8\gamma_{k+1} \|f\|_{\text{Lip}}^2 (\kappa \Gamma_{N+2, N+n+1})^{-2},$$

where  $\Phi_{n,k+1}^N$  is given by (35).

*Proof.* The proof is postponed to Section 6.10.  $\square$

Also to control the last term in right hand side of (38), we need to control the variance of  $\Psi_n^N(X_N)$  under  $\delta_x Q_\gamma^N$ . But similarly to the sequence of functions  $\Phi_{n,k}^N$ ,  $\Psi_n^N$  is Lipschitz by Proposition 20 since for all  $y, z \in \mathbb{R}^d$ , we have

$$|\Psi_n^N(y) - \Psi_n^N(z)| = \left| \sum_{i=N+1}^{N+n} \omega_{i,n}^N \{Q_\gamma^{N+1,i} f(y) - Q_\gamma^{N+1,i} f(z)\} \right|. \quad (40)$$

Therefore it suffices to find some bound for the variance of  $g$  under  $\delta_y Q_\gamma^{n,p}$ , for  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  a Lipschitz function,  $y \in \mathbb{R}^d$  and  $\gamma > 0$ , which is done in the following Lemma.

**Lemma 22.** Assume **H1** and **H3**. Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 \leq 2/(m+L)$ . Let  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  be a Lipschitz function. Then for all  $n, p \geq 1$ ,  $n \leq p$  and  $y \in \mathbb{R}^d$

$$0 \leq \int_{\mathbb{R}^d} Q_\gamma^{n,p}(y, dz) \{g^2(z) - Q_\gamma^{n,p} g(y)\}^2 \leq 2\kappa^{-1} \|g\|_{\text{Lip}}^2,$$

where  $Q_\gamma^{n,p}$  is given by (19).

*Proof.* The proof is postponed to Section 6.10.  $\square$

**Corollary 23.** Assume **H1** and **H3**. Let  $(\gamma_k)_{k \geq 1}$  be nonincreasing sequence with  $\gamma_1 \leq 2/(m+L)$ . Then for all Lipschitz function  $f$ ,

$$\text{Var}_x \{ \Psi_n^N(X_N) \} \leq 8\kappa^{-3} \|f\|_{\text{Lip}}^2 \Gamma_{N+2, N+n+1}^{-2},$$

where  $\Psi_n^N$  is given by (36).

*Proof.* The proof is postponed to Section 6.10  $\square$

Finally, we have:

**Theorem 24.** Assume **H1** and **H3**. Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 \leq 2/(m+L)$ . Then for all  $N \geq 0$ ,  $n \geq 1$  and Lipschitz functions  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ , we get  $\text{Var}_x \{ \hat{\pi}_n^N(f) \} \leq 8\kappa^{-2} \|f\|_{\text{Lip}}^2 \Gamma_{N+2, N+n+1}^{-1} u_{N,n}^{(3)}(\gamma)$  where

$$u_{N,n}^{(3)}(\gamma) \stackrel{\text{def}}{=} \left\{ 1 + \Gamma_{N+2, N+n+1}^{-1} (\kappa^{-1} + 2/(m+L)) \right\}. \quad (41)$$

*Proof.* Plugging the bounds given by Corollary 21 and Corollary 23 in (38), we have

$$\begin{aligned} \text{Var}_x \{ \hat{\pi}_n^N(f) \} &\leq 8\kappa^{-2} \|f\|_{\text{Lip}}^2 \left\{ \Gamma_{N+2, N+n+1}^{-2} \Gamma_{N+1, N+n} + \kappa^{-1} \Gamma_{N+2, N+n+1}^{-2} \right\} \\ &\leq 8\kappa^{-2} \|f\|_{\text{Lip}}^2 \left\{ \Gamma_{N+2, N+n+1}^{-1} + \Gamma_{N+2, N+n+1}^{-2} (\gamma_{N+1} + \kappa^{-1}) \right\}. \end{aligned}$$

Using that  $\gamma_{N+1} \leq 2/(m+L)$  concludes the proof.  $\square$

It is worth to observe that this bound is independent from the dimension. We may now discuss the bounds on the MSE (obtained by combining the bounds for the squared bias Proposition 18 and the variance Theorem 24) for step sizes given for  $k \geq 1$  by  $\gamma_k = \gamma_1 k^{-\alpha}$  where  $\alpha \in [0, 1]$  and  $\gamma_1 < 1/(m + L)$ . Details of these calculations are included in the supplementary paper [12]. The order of the bounds (up to numerical constants) of the MSE are summarized in Table 3 as a function of  $\gamma_1$ ,  $n$  and  $N$ . If the total number of iterations  $n + N$  is held fixed (fixed horizon setting),

	Bound for the MSE
$\alpha = 0$	$\gamma_1 + (\gamma_1 n)^{-1} \exp(-\kappa \gamma_1 N/2)$
$\alpha \in (0, 1/2)$	$\gamma_1 n^{-\alpha} + (\gamma_1 n^{1-\alpha})^{-1} \exp(-\kappa \gamma_1 N^{1-\alpha}/(2(1-\alpha)))$
$\alpha = 1/2$	$\gamma_1 \log(n) n^{-1/2} + (\gamma_1 n^{1/2})^{-1} \exp(-\kappa \gamma_1 N^{1/2}/4)$
$\alpha \in (1/2, 1)$	$n^{\alpha-1} \{ \gamma_1 + \gamma_1^{-1} \exp(-\kappa \gamma_1 N^{1-\alpha}/(2(1-\alpha))) \}$
$\alpha = 1$	$\log(n)^{-1} \{ \gamma_1 + \gamma_1^{-1} N^{-\gamma_1 \kappa/2} \}$

TABLE 3

Bound for the MSE for  $\gamma_k = \gamma_1 k^{-\alpha}$  for fixed  $\gamma_1$  and  $N$

as in Section 3, we may optimize the value of the step size  $\gamma_1$  but also of the burn-in period  $N$  to minimize the upper bound of the MSE. The order (in  $n$ ) for different values of  $\alpha \in [0, 1]$  are summarized in Table 4 (we display the order in  $n$  but not the constants, which are quite involved and not overly informative). It appears that, for any  $\alpha \in [0, 1/2)$ , we can always achieved the order  $n^{-1/2}$  by choosing appropriately  $\gamma_1$  and  $N$  (for  $\alpha = 1/2$  we have only  $\log(n) n^{-1/2}$ ). The worst case is for  $\alpha \in (1/2, 1]$ , where in fact the best strategy is to take  $N = 0$  and the largest possible value for  $\gamma_1 = 1/(m + L)$ . Finally, we note that from the explicit expression of the bound in [12], that constant step sizes ( $\alpha = 0$ ) are optimal. Finally, we mention that the bounds for  $\alpha \in [0, 1/2)$  for a fixed number of iterations implies using the doubling trick (see [18]) an anytime algorithm which guarantees for all  $n \geq 1$ , a MSE of order  $\mathcal{O}(n^{-1/2})$ .

	Optimal choice of $\gamma_1$	Optimal choice of $N$	Bound for the MSE
$\alpha = 0$	$n^{-1/2}$	$n^{1/2}$	$n^{-1/2}$
$\alpha \in (0, 1/2)$	$n^{\alpha-1/2}$	$n^{(1/2-\alpha)/(1-\alpha)}$	$n^{-1/2}$
$\alpha = 1/2$	$(\log(n))^{-1/2}$	$\log(n)$	$\log(n) n^{-1/2}$
$\alpha \in (1/2, 1)$	$1/(m + L)$	0	$n^{1-\alpha}$
$\alpha = 1$	$1/(m + L)$	0	$\log(n)$

TABLE 4

Bound for the MSE for  $\gamma_k = \gamma_1 k^{-\alpha}$  for fixed  $n$

In this part, we establish an exponential deviation inequality for  $\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)]$  given by (31), with  $N \geq 0$ ,  $n \geq 1$ ,  $x \in \mathbb{R}^d$  and  $f$  a Lipschitz function. To that purpose, we derive an upper bound of the Laplace transform of  $\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)]$ . Using the decomposition by martingale

increments (33)

$$\begin{aligned} & \mathbb{E}_x \left[ e^{\lambda \{ \hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)] \}} \right] \\ &= \mathbb{E}_x \left[ \exp \left( \lambda \{ \mathbb{E}_x^{\mathcal{G}_N} [\hat{\pi}_n^N(f)] - \mathbb{E}_x[\hat{\pi}_n^N(f)] \} + \sum_{k=N}^{N+n-1} \lambda \{ \mathbb{E}_x^{\mathcal{G}_{k+1}} [\hat{\pi}_n^N(f)] - \mathbb{E}_x^{\mathcal{G}_k} [\hat{\pi}_n^N(f)] \} \right) \right]. \end{aligned}$$

Now using (37) with the sequence of functions  $(\Phi_{n,k}^N)$  and  $\Psi_n^N$  given by (35) and (36), respectively, we have by the Markov property

$$\begin{aligned} & \mathbb{E}_x \left[ e^{\lambda \{ \hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)] \}} \right] \\ &= \mathbb{E}_x \left[ e^{\lambda \{ \Psi_n^N(X_n) - \mathbb{E}_x[\Psi_n^N(X_n)] \}} \prod_{k=N}^{N+n-1} R_{\gamma_{k+1}} \left\{ \exp \left( \lambda \{ \Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(X_k) \} \right) \right\} (X_k) \right] \quad (42) \end{aligned}$$

where  $R_\gamma$  is given by (18) for  $\gamma > 0$ . We use the same strategy to get concentration inequalities than to bound the variance term in the previous section, replacing the Gaussian Poincaré inequality by the log-Sobolev inequality to get uniform bound for  $R_{\gamma_{k+1}} \left\{ \exp \left( \lambda \{ \Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(X_k) \} \right) \right\} (X_k)$  w.r.t.  $X_k$ , for all  $k \in \{N+1, \dots, N+n\}$ . Indeed for all  $x \in \mathbb{R}^d$ , recall that  $R_\gamma(x, \cdot)$  is a Gaussian distribution with mean  $x - \gamma \nabla U(x)$  and covariance matrix  $2\gamma I_d$ . The log-Sobolev inequality provides a bound for the the Laplace transform of Lipschitz function  $g(Z) - R_\gamma g(x)$  where  $Z$  is distributed under  $R_\gamma(x, \cdot)$ .

**Lemma 25** ([6, Theorem 5.5]). *Assume H1. Then for all Lipschitz function  $g$ ,  $\gamma > 0$ ,  $x \in \mathbb{R}^d$  and  $\lambda > 0$ ,*

$$\int R_\gamma(x, dy) \left\{ \exp \left( \lambda \{ g(y) - R_\gamma g(x) \} \right) \right\} \leq \exp \left( \gamma \lambda^2 \|g\|_{\text{Lip}}^2 \right).$$

where  $R_\gamma$  is given by (18).

We deduced from this lemma, (39) and Proposition 20-(ii), an equivalent of Corollary 21 for the Laplace transform of  $\Phi_{n,k+1}^N$  under  $\delta_y R_{\gamma_{k+1}}$  for  $k \in \{N+1, \dots, N+n\}$  and all  $y \in \mathbb{R}^d$ .

**Corollary 26.** *Assume H1 and H3. Let  $(\gamma_k)_{k \geq 1}$  be nonincreasing sequence with  $\gamma_1 \leq 2/(m+L)$ . Let  $N \geq 0$  and  $n \geq 1$ . Then for all  $k \in \{N+1, \dots, N+n\}$ ,  $y \in \mathbb{R}^d$  and  $\lambda > 0$ ,*

$$R_{\gamma_{k+1}} \left\{ \exp \left( \lambda \{ \Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(y) \} \right) \right\} (y) \leq \exp \left( 4\gamma_{k+1} \lambda^2 \|f\|_{\text{Lip}}^2 (\kappa \Gamma_{N+2, N+n+1})^{-2} \right),$$

where  $\Phi_{n,k}^N$  is given by (35).

It remains to control the Laplace transform of  $\Psi_n^N$  under  $\delta_x Q_\gamma^N$ , where  $\delta_x Q_\gamma^N$  is defined by (19). For this, using again that by (40) and Proposition 20-(ii),  $\Psi_n^N$  is a Lipschitz function, we iterate Lemma 25 to get bounds on the Laplace transform of Lipschitz function  $g$  under  $Q_\gamma^{n,p}(y, \cdot)$  for all  $y \in \mathbb{R}^d$  and  $n, p \geq 1$ , since for all  $n, p \geq 1$ ,  $Q_\gamma^{n,p} g$  is a Lipschitz function by Proposition 20-(ii).

**Lemma 27.** *Assume H1 and H3. Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 \leq 2/(m+L)$ . Let  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  be a Lipschitz function, then for all  $n, p \geq 1$ ,  $n \leq p$ ,  $y \in \mathbb{R}^d$  and  $\lambda > 0$ :*

$$Q_\gamma^{n,p} \left\{ \exp \left( \lambda \{ g(\cdot) - Q_\gamma^{n,p} g(y) \} \right) \right\} (y) \leq \exp \left( \kappa^{-1} \lambda^2 \|g\|_{\text{Lip}}^2 \right), \quad (43)$$

where  $Q_{n,p}^\gamma$  is given by (19).

*Proof.* The proof is postponed to Section 6.10.  $\square$

Combining this result with (40) and Proposition 20-(ii), we get an analogue of Corollary 23 for the Laplace transform of  $\Psi_n^N$ :

**Corollary 28.** *Assume H1 and H3. Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 \leq 2/(m+L)$ . Let  $N \geq 0$  and  $n \geq 1$ . Then for all  $\lambda > 0$ ,*

$$\mathbb{E}_x \left[ e^{\lambda \{\Psi_n^N(X_n) - \mathbb{E}_x[\Psi_n^N(X_n)]\}} \right] \leq \exp \left( 4\kappa^{-3} \lambda^2 \|f\|_{\text{Lip}}^2 \Gamma_{N+2, N+n+1}^{-2} \right),$$

where  $\Psi_n^N$  is given by (36).

The Laplace transform of  $\hat{\pi}_n^N(f)$  can be explicitly bounded using Corollary 26 and Corollary 28 in (42).

**Proposition 29.** *Assume H1 and H3. Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 \leq 2/(m+L)$ . Then for all  $N \geq 0$ ,  $n \geq 1$ , Lipschitz functions  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  and  $\lambda > 0$ :*

$$\mathbb{E}_x \left[ e^{\lambda \{\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)]\}} \right] \leq \exp \left( 4\kappa^{-2} \lambda^2 \|f\|_{\text{Lip}}^2 \Gamma_{N+2, N+n+1}^{-1} u_{N,n}^{(3)}(\gamma) \right),$$

where  $u_{N,n}^{(3)}(\gamma)$  is given by and (41).

**Theorem 30.** *Assume H1 and H3. Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 \leq 2/(m+L)$ . Let  $(X_n)_{n \geq 0}$  be given by (2) and started at  $x \in \mathbb{R}^d$ . Then for all  $N \geq 0$ ,  $n \geq 1$ ,  $r > 0$  and Lipschitz functions  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ :*

$$\mathbb{P}_x \left[ \hat{\pi}_n^N(f) \geq \mathbb{E}_x[\hat{\pi}_n^N(f)] + r \right] \leq \exp \left( -\frac{r^2 \kappa^2 \Gamma_{N+2, N+n+1}}{16 \|f\|_{\text{Lip}}^2 u_{N,n}^{(3)}(\gamma)} \right).$$

*Proof.* Using the Markov inequality and Proposition 29, for all  $\lambda > 0$ , we have:

$$\mathbb{P}_x \left[ \hat{\pi}_n^N(f) \geq \mathbb{E}_x[\hat{\pi}_n^N(f)] + r \right] \leq \exp \left( -\lambda r + 4\kappa^{-2} \lambda^2 \|f\|_{\text{Lip}}^2 \Gamma_{N+2, N+n+1}^{-1} u_{N,n}^{(3)}(\gamma) \right).$$

Then the result follows from taking  $\lambda = (r\kappa^2 \Gamma_{N+2, N+n+1}) / (8 \|f\|_{\text{Lip}}^2 u_{N,n}^{(3)}(\gamma))$ .  $\square$

If we apply this result to the sequence  $(\gamma_k)_{k \geq 1}$  defined for all  $k \geq 1$  by  $\gamma_k = \gamma_1 k^{-\alpha}$ , for  $\alpha \in [0, 1]$ , we end up with a concentration of order  $\exp(-n^{1-\alpha})$  for  $\alpha \in [0, 1)$  and  $n^{-1}$  for  $\alpha = 1$ .

## 5. Numerical experiments

Consider a binary regression set-up in which the binary observations (responses)  $(Y_1, \dots, Y_p)$  are conditionally independent Bernoulli random variables with success probability  $\varrho(\boldsymbol{\beta}^T X_i)$ , where  $\varrho$  is the logistic function defined for  $z \in \mathbb{R}$  by  $\varrho(z) = e^z / (1 + e^z)$  and  $X_i$  and  $\boldsymbol{\beta}$  are  $d$  dimensional vectors of known covariates and unknown regression coefficient, respectively. The prior distribution for the

parameter  $\boldsymbol{\beta}$  is a zero-mean Gaussian distribution with covariance matrix  $\Sigma$ . The posterior density distribution of  $\boldsymbol{\beta}$  is up to a proportionality constant given by

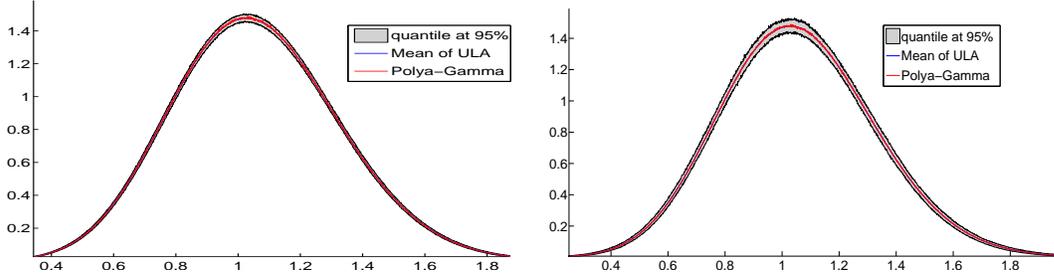
$$\pi_{\boldsymbol{\beta}}(\boldsymbol{\beta} | ((X_i, Y_i))_{1 \leq i \leq p}) \propto \exp \left( \sum_{i=1}^p Y_i \boldsymbol{\beta}^T X_i - \log(1 + e^{\boldsymbol{\beta}^T X_i}) - (1/2) \boldsymbol{\beta}^T \Sigma^{-1} \boldsymbol{\beta} \right).$$

Bayesian inference for the logistic regression model has long been recognized as a numerically involved problem, due to the analytically inconvenient form of the model's likelihood function. Several algorithms have been proposed, trying to mimick the data-augmentation (DA) approach of [1] for probit regression; see [19], [13] and [14]. Recently, a very promising DA algorithm has been proposed in [28], using the Polya-Gamma distribution in the DA part. This algorithm has been shown to be uniformly ergodic for the total variation by [8, Proposition 1], which provides an explicit expression for the ergodicity constant. This constant is exponentially small in the dimension of the parameter space and the number of samples (it is likely however that this constant is very conservative). Moreover, the complexity of the augmentation step is cubic in the dimension, which prevents from using this algorithm when the dimension of the regressor is large.

We apply ULA to sample from the posterior distribution  $\pi_{\boldsymbol{\beta}}(\cdot | ((X_i, Y_i))_{1 \leq i \leq p})$ . The gradient of its log-density may be expressed as

$$\nabla \log \{ \pi_{\boldsymbol{\beta}}(\boldsymbol{\beta} | ((X_i, Y_i))_{1 \leq i \leq p}) \} = \sum_{i=1}^p Y_i X_i - \frac{X_i}{1 + e^{-\boldsymbol{\beta}^T X_i}} - \Sigma^{-1} \boldsymbol{\beta},$$

Therefore  $-\log \pi_{\boldsymbol{\beta}}(\cdot | ((X_i, Y_i))_{1 \leq i \leq p})$  is strongly convex **H3** with  $m = \lambda_{\max}^{-1}(\Sigma)$  and satisfies **H1** with  $L = (1/4) \max_{1 \leq i \leq p} \{ \|X_i\| + \lambda_{\min}^{-1}(\Sigma) \}$ , where  $\lambda_{\min}(\Sigma)$  and  $\lambda_{\max}(\Sigma)$  are the minimal and maximal eigenvalues of  $\Sigma$ , respectively. To assess the proposed algorithm, we first compare the histograms given by ULA and the Pólya-Gamma Gibbs sampling from [28]. For this, we take  $d = 5$ ,  $p = 100$ , generate synthetic data  $(Y_i)_{1 \leq i \leq p}$  and  $(X_i)_{1 \leq i \leq p}$ , and set  $\Sigma_{\boldsymbol{\beta}} = (\sum_{i=1}^p \|X_i\|^2) (dp)^{-1} \mathbf{I}_d$ . We produce  $10^7$  samples from the Pólya-Gamma sampler using the R package `BayesLogit` [32]. Next, we make  $10^3$  runs of the Euler approximation scheme with  $n = 10^6$  effective iterations, with a constant sequence  $(\gamma_k)_{k \geq 1}$ ,  $\gamma_k = 10(\kappa n^{1/2})^{-1}$  for all  $k \geq 0$  and a burn-in period  $N = n^{1/2}$ . The plot of the histogram of the Pólya-Gamma Gibbs sampler for one component, the corresponding mean of the obtained histograms for ULA and the quantiles at 95% can be found in Figure 1. The same procedure is also applied with the decreasing step size sequence  $(\gamma_k)_{k \geq 1}$  defined by  $\gamma_k = \gamma_1 k^{-1/2}$ , with  $\gamma_1 = 10(\kappa \log(n)^{1/2})^{-1}$ , and for the burn in period  $N = \log(n)$ , see also Figure 1. In addition, we also compare the Pólya-Gamma Gibbs sampler, MALA and ULA on four real data sets, which are summarized in Table 5. Note that for the Australian credit data set, the ordinal covariates have been stratified by dummy variables. Furthermore, we normalized the data sets and consider the Zellner prior setting  $\Sigma^{-1} = (\pi^2 p / 3) \Sigma_X^{-1}$  where  $\Sigma_X = p^{-1} \sum_{i=1}^p X_i X_i^T$ ; see [30], [17] and the references therein. Also, we apply a pre-conditionned version of MALA and ULA, targeting the probability density  $\tilde{\pi}_{\boldsymbol{\beta}}(\cdot) \propto \pi_{\boldsymbol{\beta}}(\Sigma_X^{1/2} \cdot)$ . Then, we obtain samples from  $\pi_{\boldsymbol{\beta}}$  by post-multiplying the obtained draws by  $\Sigma_X^{1/2}$ . For each data sets, 100 runs of the Polya-Gamma Gibbs sampler ( $10^5$  iterations per run), and 100 runs of MALA and ULA ( $10^6$  iterations per run) have been performed. Despite the fact that longer runs are carried out, the computational time of ULA is still two orders of magnitude lower than the Pólya-Gamma simulator. For MALA, the step-size is chosen so that the acceptance probability in stationarity is approximately equal to 0.5. For ULA, we choose constant step-sizes  $\gamma = 5 \times 10^{-3}$  for all the data sets. We display the boxplots of the estimators for the mean



**Figure 1: Empirical distribution comparison between the Poly-Gamma Gibbs Sampler and ULA. Left panel: constant step size  $\gamma_k = \gamma_1$  for all  $k \geq 1$ ; right panel: decreasing step size  $\gamma_k = \gamma_1 k^{-1/2}$  for all  $k \geq 1$**

Data set	Dimensions	Observations $p$	Covariates $d$
German credit <sup>1</sup>		1000	25
Heart disease <sup>2</sup>		270	14
Australian credit <sup>3</sup>		690	35
Prima indian diabetes <sup>4</sup>		768	9

TABLE 5  
Dimension of the data sets

of one component of  $\beta$  in Figure 2. Note that there are some discrepancies between the posterior mean estimators obtained using either the DA, MALA and ULA. These differences are of order  $10^{-3}$  and are likely to be due to accumulations of numerical errors. These differences are negligible compared to the posterior variance of these estimators, which is of order  $10^{-1}$ . These results all imply that ULA is a much simpler and faster alternative to the Poly-Gamma Gibbs sampler and MALA algorithm.

## 6. Proofs

### 6.1. Proof of Proposition 1

(i) For all  $x \in \mathbb{R}^d$ , we have

$$\begin{aligned} \mathcal{A}W(x) &= (\rho/4)(\|x\|^2 + 1)^{-1/2}W(x)(-\langle \nabla U(x), x \rangle + (\rho/4)(\|x\|^2 + 1)^{-1/2} \|x\|^2 \\ &\quad - (\|x\|^2 + 1)^{-1} \|x\|^2 + d) \end{aligned}$$

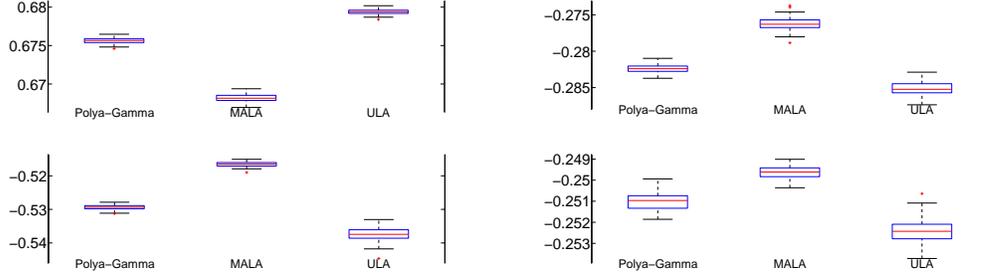
Then, for all  $x$ ,  $\|x\| \geq K_{\rho,1} = \max(M_\rho, 4d/\rho, 1)$ , since  $1 + \|x\|^2 \geq \|x\|^2$  and  $\langle \nabla U(x), x \rangle \geq \rho \|x\|$  under **H2**(1),  $\mathcal{A}W(x) \leq -(\rho^2/8)W(x)$ , and  $\sup_{\{\|x\| \leq K_{\rho,1}\}} \mathcal{A}W(x) \leq b_{\rho,1}$ .

<sup>1</sup>[http://archive.ics.uci.edu/ml/datasets/Statlog+\(German+Credit+Data\)](http://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data))

<sup>2</sup>[http://archive.ics.uci.edu/ml/datasets/Statlog+\(Heart\)](http://archive.ics.uci.edu/ml/datasets/Statlog+(Heart))

<sup>3</sup>[http://archive.ics.uci.edu/ml/datasets/Statlog+\(Australian+Credit+Approval\)](http://archive.ics.uci.edu/ml/datasets/Statlog+(Australian+Credit+Approval))

<sup>4</sup><http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>



**Figure 2: Upper left: German credit data set. Upper right: Australian credit data set. Lower left: Heart disease data set. Lower right: Prima indian diabetes data set**

(ii) For all  $x \in \mathbb{R}^d$ , we have

$$\mathcal{A}V_s(x) = s(1-s)(-\|\nabla U(x)\|^2 + (1-s)^{-1}\Delta U(x))V_s(x)$$

If  $\alpha > 1$ , then by the Cauchy-Schwarz inequality, under **H1**, for all  $x \in \mathbb{R}^d$ ,  $\Delta U(x) \leq dL$  and  $\|\nabla U(x)\| \geq \rho \|x\|^{\alpha-1}$  for  $\|x\| \geq M_\rho$ . Therefore,

$$\mathcal{A}V_s(x) \leq s(1-s)(-\rho \|x\|^{2(\alpha-1)} + (1-s)^{-1}dL)V_s(x) \leq -sdLV_s(x), \quad \text{for } \|x\| \geq K_{\rho,\alpha,s},$$

and  $\sup_{\{\|x\| \leq K_{\rho,\alpha,s}\}} \mathcal{A}V_s(x) \leq b_{\rho,\alpha}$ .

## 6.2. Proof of Theorem 8

(i) Set  $V_\star(x) = \|x - x^\star\|^2$ . By Jensen's inequality, for all  $c > 0$  and  $t > 0$ , we get

$$\begin{aligned} \pi(V_\star \wedge c) &= \pi P_t(V_\star \wedge c) \leq \pi(P_t V_\star \wedge c) = \int \pi(dx) \left\{ \|x - x^\star\|^2 e^{-2mt} + \frac{d}{m}(1 - e^{-2mt}) \wedge c \right\} \\ &\leq \pi(V_\star \wedge c)e^{-2mt} + (1 - e^{-2mt})d/m. \end{aligned}$$

Taking the limit as  $t \rightarrow +\infty$ , we get  $\pi(V_\star \wedge c) \leq d/m$ . Using the monotone convergence theorem, taking the limit as  $c \rightarrow +\infty$ , we finally obtain (16).

(ii) Consider the following SDE in  $\mathbb{R}^d \times \mathbb{R}^d$ :

$$\begin{cases} dY_t &= -\nabla U(Y_t)dt + \sqrt{2}dB_t, \\ d\tilde{Y}_t &= -\nabla U(\tilde{Y}_t)dt + \sqrt{2}dB_t, \end{cases} \quad (44)$$

where  $(Y_0, \tilde{Y}_0)$  is some coupling between  $\mu$  and  $\nu$ . Since  $\mu$  and  $\nu$  are in  $\mathcal{P}_2(\mathbb{R}^d)$  and  $\nabla U$  is Lipschitz, then by [20, Theorem 2.5, Theorem 2.9, Chapter 5], this SDE has a unique strong solution  $(Y_t, \tilde{Y}_t)_{t \geq 0}$  associated with  $(B_t)_{t \geq 0}$ . Moreover since  $(Y_t, \tilde{Y}_t)_{t \geq 0}$  is a solution of (44),

$$\|Y_t - \tilde{Y}_t\|^2 = \|Y_0 - \tilde{Y}_0\|^2 - 2 \int_0^t \langle \nabla U(Y_s) - \nabla U(\tilde{Y}_s), Y_s - \tilde{Y}_s \rangle ds,$$

which implies using (13) and Grönwall's inequality that

$$\|Y_t - \tilde{Y}_t\|^2 \leq \|Y_0 - \tilde{Y}_0\|^2 - 2m \int_0^t \|Y_s - \tilde{Y}_s\|^2 ds \leq \|Y_0 - \tilde{Y}_0\|^2 e^{-2mt}.$$

For all  $t \geq 0$ , the law of  $(Y_t, \tilde{Y}_t)$  is a coupling between  $\mu P_t$  and  $\nu P_t$ . Therefore by definition of  $W_2$ ,  $W_2(\mu P_t, \nu P_t) \leq \mathbb{E}[\|Y_t - \tilde{Y}_t\|^2]^{1/2}$  showing (ii).

### 6.3. Proof of Proposition 9 and Corollary 10

*Proof of Proposition 9.* Let  $\bar{\gamma} \in (0, L^{-1})$ . First note by **H2**( $\alpha$ ) and the Cauchy-Schwarz inequality, we have for all  $x \in \mathbb{R}^d$ ,  $\|x\| \geq M_\rho$ ,

$$\|\nabla U(x)\| \geq \rho \|x\|^{\alpha-1}. \quad (45)$$

Since under **H1**, for all  $x, y \in \mathbb{R}^d$ ,  $U(y) \leq U(x) + \langle \nabla U(x), y - x \rangle + (L/2)\|y - x\|^2$ , we have for all  $\gamma \in (0, \bar{\gamma})$  and  $x \in \mathbb{R}^d$ ,

$$\begin{aligned} R_\gamma V_{1/2}(x)/V_{1/2}(x) &= (4\pi\gamma)^{-d/2} \int_{\mathbb{R}^d} \exp\left(\frac{(U(y) - U(x))/2 - (4\gamma)^{-1}\|y - x + \gamma\nabla U(x)\|^2}{2}\right) dy \\ &\leq (4\pi\gamma)^{-d/2} \int_{\mathbb{R}^d} \exp\left(-4^{-1}\gamma\|\nabla U(x)\|^2 - (4\gamma)^{-1}(1 - \gamma L)\|y - x\|^2\right) dy \\ &\leq (1 - \gamma L)^{-d/2} \exp(-4^{-1}\gamma\|\nabla U(x)\|^2), \end{aligned} \quad (46)$$

where we used for the last line that  $\gamma < L^{-1}$ . Since  $\log(1 - L\gamma) = -L \int_0^\gamma (1 - Lt)^{-1} dt$ , for all  $\gamma \in (0, \bar{\gamma}]$ ,  $\log(1 - L\gamma) \geq -L\gamma(1 - L\bar{\gamma})^{-1}$ . Using this inequality, (46) becomes

$$R_\gamma V_{1/2}(x)/V_{1/2}(x) \leq \exp\left(\gamma\left\{s_{\bar{\gamma}} - 4^{-1}\|\nabla U(x)\|^2\right\}\right). \quad (47)$$

By (45), for all  $x \in \mathbb{R}^d$ ,  $\|x\| \geq K_{\bar{\gamma}, \alpha, \rho} = \max(M_\rho, (8s_{\bar{\gamma}}/\rho)^{1/(2(\alpha-1))})$ , we have

$$R_\gamma V_{1/2}(x) \leq e^{-s_{\bar{\gamma}}\gamma} V_{1/2}(x). \quad (48)$$

Also by (47) and since for all  $t \geq 0$ ,  $e^t - 1 \leq te^t$ , we get for all  $x \in \mathbb{R}^d$

$$R_\gamma V_{1/2}(x) - e^{-s_{\bar{\gamma}}\gamma} V_{1/2}(x) \leq e^{-s_{\bar{\gamma}}\gamma} (e^{2\gamma s_{\bar{\gamma}}} - 1) V_{1/2}(x) \leq 2s_{\bar{\gamma}}\gamma e^{s_{\bar{\gamma}}\bar{\gamma}} V_{1/2}(x).$$

The proof is completed combining the last inequality and (48).  $\square$

*Proof of Corollary 10.* By a straightforward induction of Proposition 9, we have for all  $x \in \mathbb{R}^d$  and  $n \geq 1$ ,

$$Q_\gamma^n V_{1/2}(x) \leq e^{-s_{\gamma_1} \Gamma_{1,n}} V_{1/2}(x) + c_{\gamma_1, \rho, \alpha} \sum_{i=1}^n \gamma_i e^{-s_{\gamma_1} \Gamma_{i+1,n}}. \quad (49)$$

In addition, since for all  $t \geq 0$ ,  $e^{-t} = 1 - \int_0^t e^{-u} du \leq 1 - te^{-t}$ , we have

$$\begin{aligned} \sum_{i=1}^n \gamma_i e^{-s_{\gamma_1} \Gamma_{i+1,n}} &\leq \sum_{i=1}^n \gamma_i \prod_{k=i+1}^n (1 - e^{-s_{\gamma_1} \gamma_1 s_{\gamma_1} \gamma_k}) \\ &\leq s_{\gamma_1}^{-1} e^{s_{\gamma_1} \gamma_1} \sum_{i=1}^n \left\{ \prod_{k=i+1}^n (1 - e^{-s_{\gamma_1} \gamma_1 s_{\gamma_1} \gamma_k}) - \prod_{k=i}^n (1 - e^{-s_{\gamma_1} \gamma_1 s_{\gamma_1} \gamma_k}) \right\} \leq s_{\gamma_1}^{-1} e^{s_{\gamma_1} \gamma_1}. \end{aligned}$$

Plugging this inequality in (49) gives the desired result.  $\square$

#### 6.4. Proof of Theorem 11

*Proof of Theorem 11.* Let  $n \geq 0, p \geq 1$  with  $n < p$ . We separately bound the two terms of the right hand side of (20). By Corollary 10, we get that for all  $k \geq 1$ ,

$$\mathbb{E}_x[\|\nabla U(X_k)\|^2] \leq \|\nabla U\|_{V_{1/2}^{1/2}}^2 Q_\gamma^k V_{1/2}(x) \leq \|\nabla U\|_{V_{1/2}^{1/2}}^2 (V_{1/2}(x) + c_{\gamma,\rho,\alpha} s_\gamma^{-1} e^{\gamma_1 s_\gamma}).$$

This inequality and (24) imply that for all  $x \in \mathbb{R}^d$ ,

$$\begin{aligned} & \|\delta_x Q_\gamma^{n+1,p} - \delta_x P_{\Gamma_{n+1,p}}\|_{\text{TV}} \\ & \leq 2^{-1/2} L \left( \sum_{k=n}^{p-1} \left\{ (\gamma_{k+1}^3/3) \|\nabla U\|_{V_{1/2}^{1/2}}^2 (V_{1/2}(x) + c_{\gamma,\rho,\alpha} s_\gamma^{-1} e^{\gamma_1 s_\gamma}) + d\gamma_{k+1}^2 \right\} \right)^{1/2}. \end{aligned}$$

We now deal with the second term of (20). By Corollary 2,  $(P_t)_{t \geq 0}$  is  $V_{1/2}$ -geometrically ergodic, which implies with Corollary 10 that for all  $x \in \mathbb{R}^d$ ,

$$\|\delta_x Q_\gamma^n P_{\Gamma_{n+1,p}} - \pi\|_{\text{TV}} \leq C_\alpha e^{-v_\alpha \Gamma_{n+1,p}} \{e^{-s_\gamma} V_{1/2}(x) + c_{\gamma,\rho,\alpha} s_\gamma^{-1} e^{\gamma_1 s_\gamma}\}. \quad (50)$$

We now show the second part of the statement. Without loss of generality, since  $(\gamma_k)_{k \geq 1}$  converges to 0 and is nonincreasing, we can assume that for all  $k \geq 1$ ,  $\gamma_k + \log(\gamma_k)/v_\alpha < 0$ , where  $v_\alpha$  is given by Theorem 11. Since  $(\Gamma_k)_{k \geq 1}$  goes to infinity, there exists  $P \geq 1$  such that for all  $p \geq P$ ,  $\Gamma_p \geq -\log(\gamma_1)/v_\alpha$ . Then, we can define for all  $p \geq P$ ,

$$n(p) \stackrel{\text{def}}{=} \min \{k \in \{0, \dots, p-1\} \mid \Gamma_{k+1,p} + \log(\gamma_{k+1})/v_\alpha < 0\}. \quad (51)$$

Note that  $\Gamma_{n(p)+1,p} < -\log(\gamma_{n(p)+1,p})/v_\alpha$ . So,  $\sum_{k=n}^{p-1} \gamma_{k+1}^2 \leq -\gamma_{n(p)+1} \log(\gamma_{n(p)+1})/v_\alpha$ . We deduce from the last inequality and since  $(\gamma_k)_{k \geq 1}$  is nonincreasing, that there exists  $C \geq 0$ , such that for all  $p \geq P$ ,

$$B_{n(p),p}(x; \gamma) \leq C(-\log(\gamma_{n(p)+1})\gamma_{n(p)+1})^{1/2} \quad (52)$$

Also by (51) and definition of  $P$ ,  $n(p) > 0$  and  $\Gamma_{n(p),p} \geq -\log(\gamma_{n(p)})/v_\alpha$ . It follows that

$$e^{-v_\alpha \Gamma_{n(p),p}} \leq \gamma_{n(p)}. \quad (53)$$

Using again  $\Gamma_{n(p),p} < -\log(\gamma_{n(p)+1,p})/v_\alpha$ , we have  $\lim_{p \rightarrow \infty} n(p) = \infty$ . By this result,  $\lim_{k \rightarrow +\infty} \gamma_k = 0$  and plugging (52)-(53) in the bound of Theorem 11 completes the proof.  $\square$

#### 6.5. Proof of Theorem 12

Denote for  $\gamma > 0$ ,  $r_\gamma : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  the transition density of  $R_\gamma$  defined for  $x, y \in \mathbb{R}^d$  by

$$r_\gamma(x, y) = (4\pi\gamma)^{-1} \exp(-(4\gamma)^{-1} \|y - x + \gamma \nabla U(x)\|^2). \quad (54)$$

Then, for all  $n \geq 1$ , we denote by  $q_\gamma^n : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  the transition density associated with  $Q_\gamma^n$  defined by induction by: for all  $x, y \in \mathbb{R}^d$

$$q_\gamma^1(x, y) = r_{\gamma_1}(x, y), \quad q_\gamma^{n+1}(x, y) = \int_{\mathbb{R}^d} q_\gamma^n(x, z) r_{\gamma_{n+1}}(z, y) dz \text{ for } n > 1. \quad (55)$$

In the proof of the Theorem, we need to have a bound on  $\text{Var}_\pi \{d\delta_x Q_\gamma^n / d\pi\}$  for all  $n \geq 1$ . For this, we need some estimates for  $q_\gamma^n$  for all  $n$  and for the normalizing constant  $\int_{\mathbb{R}^d} e^{-U(y)} dy$ , which are given in the following two lemmas.

**Lemma 31.** *Assume **H1**. Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 < L$ . Then for all  $n \geq 1$  and  $x, y \in \mathbb{R}^d$ ,*

$$q_\gamma^n(x, y) \leq \frac{\exp\left(2^{-1}(U(x) - U(y)) - (2\sigma_{\gamma,n})^{-1} \|y - x\|^2\right)}{(2\pi\sigma_{\gamma,n} \prod_{i=1}^n (1 - L\gamma_i))^{d/2}},$$

where  $\sigma_{\gamma,n} = \sum_{i=1}^n 2\gamma_i(1 - L\gamma_i)^{-1}$ .

*Proof.* First under **H1**, we have for all  $x, y \in \mathbb{R}^d$ ,  $U(y) \leq U(x) - \langle \nabla U(x), y - x \rangle + (L/2) \|y - x\|^2$ , so for all  $\gamma \in (0, L^{-1})$

$$r_\gamma(x, y) \leq (4\pi\gamma)^{-d/2} \exp\left(2^{-1}(U(x) - U(y)) - (1 - L\gamma_1)(4\gamma_1)^{-1} \|y - x\|^2\right). \quad (56)$$

Then, the proof of the claimed inequality is by induction. By (56), the inequality holds for  $n = 1$ . Now assume that it holds for  $n \geq 1$ . By induction hypothesis and (56) applied for  $\gamma = \gamma_{n+1}$ , we have

$$\begin{aligned} q_\gamma^{n+1}(x, y) &\leq (4\pi\gamma_{n+1})^{-d/2} \left\{ 2\pi\sigma_{\gamma,n} \prod_{i=1}^n (1 - L\gamma_i) \right\}^{-d/2} \exp(2^{-1}(U(x) - U(y))) \\ &\quad \times \int_{\mathbb{R}^d} \exp\left(- (2\sigma_{\gamma,n})^{-1} \|z - x\|^2 - (1 - L\gamma_{n+1})(4\gamma_{n+1})^{-1} \|z - y\|^2\right) dz \\ &\leq (4\pi\gamma_{n+1})^{-d/2} \left\{ 2\pi\sigma_{\gamma,n} \prod_{i=1}^n (1 - L\gamma_i) \right\}^{-d/2} (4\pi(2\sigma_{\gamma,n}^{-1} + (1 - L\gamma_{n+1})(2\gamma_{n+1})^{-1})^{-1})^{d/2} \\ &\quad \times \exp\left(2^{-1}(U(x) - U(y)) - (2\sigma_{\gamma,n+1})^{-1} \|y - x\|^2\right). \end{aligned}$$

Rearranging terms in the last inequality concludes the proof.  $\square$

**Lemma 32.** *Assume **H1** and **H2**(1). Then  $\int_{\mathbb{R}^d} e^{-U(y)} dy \leq \vartheta_U$  where*

$$\vartheta_U \stackrel{\text{def}}{=} \left( e^{U(0)} \left( 8\sqrt{d}/(3\rho) \right)^d + \|\exp\{-U\}\|_\infty \frac{\pi^{d/2}(2M_\rho)^d}{\Gamma(d/2 + 1)} \right). \quad (57)$$

*Proof.* First, if **H2**(1) is satisfied, then for all  $y \in \mathbb{R}^d$ ,  $\|y\| \geq 2M_\rho$ , we have:

$$U(y) - U(0) = \int_0^1 \langle \nabla U(ty), y \rangle dt \geq \int_{1/2}^1 t \langle \nabla U(ty), y \rangle dt \geq (3/8)\rho \|y\|.$$

Therefore by the Cauchy-Schwarz inequality, we get for all  $y \in \mathbb{R}^d$ ,  $\|y\| \geq 2M_\rho$

$$U(y) - U(0) \geq (3/8)\rho \|y\| \geq (3/8)d^{-1/2} \|y\|_1,$$

where for  $z \in \mathbb{R}^d$ ,  $\|z\|_1 = \sum_{i=1}^d |z_i|$  and  $(z_i)_{1 \leq i \leq d}$  are the components of  $z$ . Therefore, using this result we have

$$\int_{\mathbb{R}^d} e^{-U(y)} dy \leq e^{U(0)} \int_{\mathbb{R}^d} e^{-(3/8)d^{-1/2}\rho \|y-x\|_1} dy + e^{\| -U \|_\infty} \int_{\mathbb{R}^d} 1_{\{\|y\| \leq 2M_\rho\}} dy.$$

Then the proof is concluded by a straightforward calculation and using that the volume of a ball of radius  $M$  in  $\mathbb{R}^d$  is given by  $\pi^{d/2}M^d/\Gamma(d/2 + 1)$ .  $\square$

**Corollary 33.** *Assume **H1** and **H2(1)**. Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 < L$ . Then for all  $n \geq 1$  and  $x \in \mathbb{R}^d$ ,*

$$\mathrm{Var}_\pi \left\{ \frac{d\delta_x Q_\gamma^n}{d\pi} \right\} \leq (\vartheta_U \exp(U(x))) \left( 2\pi \left\{ \prod_{k=1}^n (1 - L\gamma_k) \right\}^2 \sum_{i=1}^n \gamma_i (1 - L\gamma_i)^{-1} \right)^{-d/2},$$

where  $\vartheta_U$  is given by (57).

*Proof of Theorem 12.* For the two cases, we bound the two terms of the right hand side of (20). (i) The first term is dealt with the same reasoning than for the proof of Theorem 11. As for the second term, the claimed bound follows from (9) and Corollary 33. (ii) The second term of (20) is bounded as in the first point, but the bound for the first term is given by (24) and the assumption on  $\nabla U$ .  $\square$

### 6.6. Proof of Theorem 13

We preface the proof by the following preliminary lemma.

**Lemma 34.** *Assume **H1**. Let  $V : \mathbb{R}^d \rightarrow [1, +\infty)$  be a measurable function. Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence. Then for any probability measure  $\nu_0 \in \mathcal{P}_2(\mathbb{R}^d)$  on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  and  $n \geq 1$ ,*

$$\begin{aligned} \|\nu_0 Q_\gamma^n - \nu_0 P_{\Gamma_n}\|_V &\leq 2^{-1/2} L (\nu_0 Q_\gamma^n(V^2) + \nu_0 P_{\Gamma_n}(V^2))^{1/2} \\ &\quad \times \left( \sum_{k=0}^{n-1} \gamma_{k+1}^2 d + (\gamma_{k+1}^3/3) \nu_0 Q_\gamma^k(\|\nabla U\|^2) \right)^{1/2}. \end{aligned}$$

*Proof.* For all  $t \in [0, 1]$ ,  $t \log(t) - t + 1 = \int_t^1 (u - t)u^{-1} du \geq 2^{-1}(1 - t)^2$ , and on  $[1, +\infty)$ ,  $t \mapsto 2(1 + t)(t \log(t) - t + 1) - (1 - t)^2$  is nonincreasing. Therefore, for all  $t \geq 0$ ,

$$|1 - t| \leq (2(1 + t)(t \log(t) - t + 1))^{1/2}. \quad (58)$$

Now consider the synchronous coupling  $(Y_t, \bar{Y}_t)_{0 \leq t \leq \Gamma_n}$  defined by (21) with initial distribution  $\nu_0$ . For ease of notation denote by  $\mu$  and  $\bar{\mu}$  the law of  $(Y_t)_{0 \leq t \leq \Gamma_n}$  and  $(\bar{Y}_t)_{0 \leq t \leq \Gamma_n}$  on  $C([0, \Gamma_n], \mathbb{R}^d)$ , respectively. Equation (23) implies

$$\begin{aligned} \|\nu_0 Q_\gamma^n - \nu_0 P_{\Gamma_n}\|_V &= \sup_{f \in \mathbb{F}(\mathbb{R}^d), \|f\|_V \leq 1} \left| \mathbb{E} [f(\bar{Y}_{\Gamma_n}) - f(Y_{\Gamma_n})] \right| \\ &= \sup_{f \in \mathbb{F}(\mathbb{R}^d), \|f\|_V \leq 1} \left| \mathbb{E} \left[ f(Y_{\Gamma_n}) \left\{ \frac{d\bar{\mu}}{d\mu}(Y) - 1 \right\} \right] \right| \leq \mathbb{E} \left[ V(Y_{\Gamma_n}) \left| \frac{d\bar{\mu}}{d\mu}(Y) - 1 \right| \right]. \quad (59) \end{aligned}$$

Using (58) and the Cauchy-Schwarz inequality, we have

$$\|\nu_0 Q_\gamma^n - \nu_0 P_{\Gamma_n}\|_V \leq (2\mathbb{E} [V^2(Y_{\Gamma_n}) + V^2(\bar{Y}_{\Gamma_n})] \mathrm{KL}(\mu|\bar{\mu}))^{1/2}.$$

Combining this result and (24) in (59) completes the proof  $\square$

*Proof of Theorem 13.* First note that by the triangle inequality and Corollary 2, for all  $n \geq 1$

$$\|\pi - \delta_x Q_\gamma^n\|_{V_{1/4}} \leq C_\alpha e^{-v_\alpha n \gamma} V_{1/4}(x) + \|\delta_x P_{\Gamma_n} - \delta_x Q_\gamma^n\|_{V_{1/4}}. \quad (60)$$

We now bound the second term of the right hand side. Let  $k_\gamma = \lceil \gamma^{-1} \rceil$  and  $q_\gamma$  and  $r_\gamma$  be respectively the quotient and the remainder of the Euclidean division of  $n$  by  $k_\gamma$ . The triangle inequality implies  $\|\delta_x P_{\Gamma_n} - \delta_x Q_\gamma^n\|_{V_{1/4}} \leq A + B$  with

$$\begin{aligned} A &= \left\| \delta_x Q_\gamma^{(q_\gamma-1)k_\gamma} P_{\Gamma_{(q_\gamma-1)k_\gamma, n}} - \delta_x Q_\gamma^{(q_\gamma-1)k_\gamma} Q_\gamma^{(q_\gamma-1)k_\gamma+1, n} \right\|_{V_{1/4}} \\ B &= \sum_{i=1}^{q_\gamma-1} \left\| \delta_x Q_\gamma^{(i-1)k_\gamma} P_{\Gamma_{(i-1)k_\gamma+1, n}} - \delta_x Q_\gamma^{ik_\gamma} P_{\Gamma_{ik_\gamma+1, n}} \right\|_{V_{1/4}}. \end{aligned}$$

Proposition 1-(ii) implies by the proof of [25, Theorem 6.1] that  $\delta_y P_t(V_{1/2}) \leq V_{1/2}(y) + b_{\rho, \alpha, 1/2} / \theta_{\alpha, 1/2}$ . It follows from Corollary 2, this inequality, Lemma 34, Corollary 10 and  $k_\gamma \geq \gamma^{-1}$  that

$$\begin{aligned} B &\leq \sum_{i=1}^{q_\gamma-1} C_\alpha e^{-v_\alpha (q_\gamma-i)} \left\| \delta_x Q_\gamma^{(i-1)k_\gamma} P_{\Gamma_{(i-1)k_\gamma+1, ik_\gamma}} - \delta_x Q_\gamma^{ik_\gamma} \right\|_{V_{1/4}} \\ &\leq \sum_{i=1}^{q_\gamma-1} C_\alpha e^{-v_\alpha (q_\gamma-i)} \left\{ (2(V_{1/2}(x) + c_{\gamma, \rho, \alpha} s_\gamma^{-1} e^{\gamma s_\gamma}) + b_{\rho, \alpha, 1/2} / \theta_{\alpha, 1/2})^{1/2} \right. \\ &\quad \left. 2^{-1/2} L \left( (\gamma + \gamma^2)d + 3^{-1}(\gamma^2 + \gamma^3) \|\nabla U\|_{V_{1/2}}^2 (V_{1/2}(x) + c_{\gamma, \rho, \alpha} s_\gamma^{-1} e^{\gamma s_\gamma}) \right)^{-1/2} \right\} \quad (61) \end{aligned}$$

Finally,  $A$  can be bounded using the same technique.  $\square$

### 6.7. Proof of Theorem 14

For any  $\gamma \in (0, 2/(m+L))$ , we have for all  $x \in \mathbb{R}^d$ :

$$\int_{\mathbb{R}^d} \|y - x^*\|^2 R_\gamma(x, dy) = \|x - \gamma \nabla U(x) - x^*\|^2 + 2\gamma d.$$

Using that  $\nabla U(x^*) = 0$ , and (12), we get from the previous inequality:

$$\begin{aligned} \int_{\mathbb{R}^d} \|y - x^*\|^2 R_\gamma(x, dy) \\ \leq (1 - \kappa\gamma) \|x - x^*\|^2 + \gamma \left( \gamma - \frac{2}{m+L} \right) \|\nabla U(x) - \nabla U(x^*)\|^2 + 2\gamma d \end{aligned}$$

Since  $\gamma_1 \leq 2/(m+L)$  and  $(\gamma_k)_{k \geq 1}$  is nonincreasing, by a straightforward induction, we have by definition (19) of  $Q_\gamma^{n,p}$  for  $p, n \geq 1$ ,  $p \leq n$ ,

$$\int_{\mathbb{R}^d} \|x - x^*\|^2 \mu_0 Q_\gamma^{n,p}(dx) \leq \prod_{k=n}^p (1 - \kappa\gamma_k) \int_{\mathbb{R}^d} \|x - x^*\|^2 \mu_0(dx) + (2d) \sum_{i=n}^p \prod_{k=i+1}^p (1 - \kappa\gamma_k) \gamma_i, \quad (62)$$

Consider the second term in the right hand side of (62). Since  $\gamma_1 \leq 2/(m+L)$ ,  $m \leq L$  and  $(\gamma_k)_{k \geq 1}$  is nonincreasing,  $\max_{k \geq 1} \gamma_k \leq \kappa^{-1}$  and therefore:

$$\sum_{i=n}^p \prod_{k=i+1}^p (1 - \kappa\gamma_k)\gamma_i \leq \kappa^{-1} \sum_{i=n}^p \left\{ \prod_{k=i+1}^p (1 - \kappa\gamma_k) - \prod_{k=i}^p (1 - \kappa\gamma_k) \right\} \leq \kappa^{-1}. \quad (63)$$

### 6.8. Proof of Theorem 15

We preface the proof of the theorem by a preliminary lemma.

**Lemma 35.** *Assume **H1** and **H3**.*

(i) *Let  $\gamma \in (0, 2/(m+L))$ , then for all  $x \in \mathbb{R}^d$ ,*

$$\text{Ent}_\pi (\text{d}\delta_x R_\gamma / \text{d}\pi) \leq (L/2)(1 - \kappa\gamma) \|x - x^*\|^2 - (d/2)(1 + \log(2\gamma/m) - 2L\gamma).$$

(ii) *Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 \leq 2/(m+L)$ . Then for all  $x \in \mathbb{R}^d$  and  $n \geq 1$ ,*

$$\begin{aligned} \text{Ent}_\pi (\text{d}\delta_x Q_\gamma^n / \text{d}\pi) &\leq (L/2) \left\{ \prod_{k=1}^n (1 - \kappa\gamma_k) \right\} \|x - x^*\|^2 \\ &\quad + dL\kappa^{-1}(1 - \kappa\gamma_n) - (d/2)(1 + \log(2\gamma_n/m) - 2L\gamma_n). \end{aligned}$$

*Proof.* (i) Let  $\gamma \in (0, 2/(m+L))$  and  $r_\gamma$  be the transition density of  $R_\gamma$  given by (54). Under **H1** and using  $\nabla U(x^*) = 0$ , for all  $y \in \mathbb{R}^d$ ,  $U(y) \leq U(x^*) + (L/2) \|y - x^*\|^2$ . Therefore, by (3), we have for all  $x \in \mathbb{R}^d$

$$\begin{aligned} \text{Ent}_\pi (\text{d}\delta_x R_\gamma / \text{d}\pi) &= \int_{\mathbb{R}^d} \log(r_\gamma(x, y)/\pi(y)) r_\gamma(x, y) dx \\ &= \int_{\mathbb{R}^d} \left\{ U(y) + \log \left( \int_{\mathbb{R}^d} e^{-U(z)} dz \right) - (4\gamma)^{-1} \|y - x + \gamma \nabla U(x)\|^2 - (d/2) \log(4\pi\gamma) \right\} r_\gamma(x, y) dy \\ &\leq \int_{\mathbb{R}^d} \left\{ U(x^*) + (L/2) \|y - x^*\|^2 + \log \left( \int_{\mathbb{R}^d} e^{-U(z)} dz \right) - (d/2)(1 + \log(4\pi\gamma)) \right\} r_\gamma(x, y) dy \\ &\leq (L/2) \|x - \gamma \nabla U(x) - x^*\|^2 + \log \left( \int_{\mathbb{R}^d} e^{-U(z) + U(x^*)} dz \right) - (d/2)(1 + \log(4\pi\gamma) - 2L\gamma). \quad (64) \end{aligned}$$

Using **H3** and (12), we have for all  $z \in \mathbb{R}^d$

$$U(z) \geq U(x^*) + (m/2) \|z - x^*\|^2 \quad \text{and} \quad \|z - \gamma \nabla U(z) - x^*\|^2 \leq (1 - \kappa\gamma) \|z - x^*\|^2.$$

Plugging these two inequalities in (64) gives the desired result.

(ii) For  $n = 1$  the result is just a consequence of (i). For  $n \geq 2$ , by (3), (55) and the Jensen inequality applied to the convex function  $t \mapsto t \log(t)$ , we have for all  $x \in \mathbb{R}^d$  and  $n \geq 1$ ,

$$\begin{aligned} \text{Ent}_\pi (\text{d}\delta_x Q_\gamma^n / \text{d}\pi) &= \int_{\mathbb{R}^d} \log \left\{ \pi^{-1}(y) \int_{\mathbb{R}^d} q_\gamma^{n-1}(x, z) r_{\gamma_n}(z, y) dz \right\} \int_{\mathbb{R}^d} q_\gamma^{n-1}(x, z) r_{\gamma_n}(z, y) dz dy \\ &\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \log \{ r_{\gamma_n}(z, y) \pi^{-1}(y) \} q_\gamma^{n-1}(x, z) r_{\gamma_n}(z, y) dz dy. \end{aligned}$$

Using Fubini's theorem, (i) and Theorem 14 in the last inequality concludes the proof.  $\square$

*Proof of Theorem 15.* Let  $n \geq 0, p \geq 1$  with  $n < p$ . We separately bound the two terms of the right hand side of (20). Using **H1**,  $\nabla U(x^*) = 0$  and Theorem 14, we get that for all  $k \geq 1$ ,

$$\mathbb{E}_x[\|\nabla U(X_k)\|^2] \leq L^2 \mathbb{E}_x[\|X_k - x^*\|^2] \leq L^2 \left( \prod_{i=1}^k (1 - \kappa \gamma_i) \|x - x^*\|^2 + 2d\kappa^{-1} \right).$$

Combining this inequality and (24) gives the bound for the first term of (20). Then, the bound for the second term is a direct consequence of Theorem 7 and Lemma 35-(ii).  $\square$

### 6.9. Proof of Theorem 16 and Corollary 17

We preface the proof by two technical Lemmas. Recall that  $x^*$  is the unique minimizer of  $U$ .

**Lemma 36.** *Assume **H1** and **H3**. Then, for all  $t \geq 0$  and  $x \in \mathbb{R}^d$ ,*

$$\mathbb{E}_x \left[ \|Y_t - x^*\|^2 \right] \leq \|x - x^*\|^2 e^{-2mt} + \frac{d}{m} (1 - e^{-2mt}), \quad (65)$$

$$\mathbb{E}_x \left[ \|Y_t - x\|^2 \right] \leq dt(2 + L^2 t^2/3) + (3/2)t^2 L^2 \|x - x^*\|^2. \quad (66)$$

*Proof.* Denote for all  $x \in \mathbb{R}^d$  by  $V_*(x) = \|x - x^*\|^2$ . The process  $\left( V_*(Y_t) - V_*(x) - \int_0^t \mathcal{A}V_*(Y_s) ds \right)_{t \geq 0}$  is a  $(\mathcal{F}_t)_{t \geq 0}$ -martingale under  $\mathbb{P}_x$ . Since  $\nabla U(x^*) = 0$  and using (13), we have

$$\mathcal{A}V_*(x) = 2(-\langle \nabla U(x) - \nabla U(x^*), x - x^* \rangle + d) \leq 2(-mV_*(x) + d). \quad (67)$$

Denote for all  $t \geq 0$  and  $x \in \mathbb{R}^d$  by  $v(t, x) = P_t V_*(x)$ . Then using Dynkin's formula,  $\partial v(t, x)/\partial t = P_t \mathcal{A}V_*(x)$ . Using (67), we get

$$\frac{\partial v(t, x)}{\partial t} = P_t \mathcal{A}V_*(x) \leq -2mP_t V_*(x) + 2d = -2mv(t, x) + 2d,$$

and by the the Grönwall inequality, the previous equation shows (65).

Denote for all  $x, y \in \mathbb{R}^d$ ,  $\tilde{V}_x(y) = \|y - x\|^2$ . Then under **H1**, for all  $T \geq 0$ ,  $\sup_{t \in [0, T]} \mathbb{E}_x[\|Y_t\|^2] < +\infty$ , therefore the process  $(\tilde{V}_x(Y_t) - \tilde{V}_x(x) - \int_0^t \mathcal{A}\tilde{V}_x(Y_s) ds)_{t \geq 0}$ , is a  $(\mathcal{F}_t)_{t \geq 0}$ -martingale under  $\mathbb{P}_x$ . Denote for all  $t \geq 0$  and  $x \in \mathbb{R}^d$  by  $\tilde{v}(t, x) = P_t \tilde{V}_x(x)$ . Then using Dynkin's formula,

$$\frac{\partial \tilde{v}(t, x)}{\partial t} = P_t \mathcal{A}\tilde{V}_x(x). \quad (68)$$

By (13), we have for all  $y \in \mathbb{R}^d$ ,

$$\mathcal{A}\tilde{V}_x(y) = 2(-\langle \nabla U(y), y - x \rangle + d) \leq 2(-m\tilde{V}_x(y) + d - \langle \nabla U(x), y - x \rangle). \quad (69)$$

Using (68), this inequality and that  $\tilde{V}_x$  is positive, we get

$$\frac{\partial \tilde{v}(t, x)}{\partial t} = P_t \mathcal{A}\tilde{V}_x(x) \leq 2 \left( d - \int_{\mathbb{R}^d} \langle \nabla U(x), y - x \rangle P_t(x, dy) \right). \quad (70)$$

By the Cauchy-Schwarz inequality,  $\nabla U(x^*) = 0$ , (1) and the Jensen inequality, we have,

$$\begin{aligned} |\mathbb{E}_x [\langle \nabla U(x), Y_t - x \rangle]| &\leq \|\nabla U(x)\| \|\mathbb{E}_x [Y_t - x]\| \\ &\leq \|\nabla U(x)\| \left\| \mathbb{E}_x \left[ \int_0^t \{\nabla U(Y_s) - \nabla U(x^*)\} ds \right] \right\| \\ &\leq \sqrt{t} \|\nabla U(x) - \nabla U(x^*)\| \left( \int_0^t \mathbb{E}_x [\|\nabla U(Y_s) - \nabla U(x^*)\|^2] ds \right)^{1/2}. \end{aligned}$$

Furthermore, by **H1** and (65), we have

$$\begin{aligned} \left| \int_{\mathbb{R}^d} \langle \nabla U(x), y - x \rangle P_t(x, dy) \right| &\leq \sqrt{t} L^2 \|x - x^*\| \left( \int_0^t \mathbb{E}_x [\|Y_s - x^*\|^2] ds \right)^{1/2} \\ &\leq \sqrt{t} L^2 \|x - x^*\| \left( \frac{1 - e^{-2mt}}{2m} \|x - x^*\|^2 + \frac{2tm + e^{-2mt} - 1}{2m} (d/m) \right)^{1/2} \\ &\leq L^2 \|x - x^*\| (t \|x - x^*\| + t^{3/2} d^{1/2}), \end{aligned} \quad (71)$$

where we used for the last line that by the Taylor theorem with remainder term, for all  $s \geq 0$ ,  $(1 - e^{-2ms})/(2m) \leq s$  and  $(2ms + e^{-2ms} - 1)/(2m) \leq ms^2$ , and the inequality  $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ . Plugging (71) in (70), and since  $2 \|x - x^*\| t^{3/2} d^{1/2} \leq t \|x - x^*\|^2 + t^2 d$ , we get

$$\frac{\partial \tilde{v}(t, x)}{\partial t} \leq 2d + 3L^2 t \|x - x^*\|^2 + L^2 t^2 d$$

Since  $\tilde{v}(0, x) = 0$ , the proof is completed by integrating this result.  $\square$

**Lemma 37.** *Assume **H1** and **H3**. Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 \leq 1/(m+L)$ . Let  $\zeta_0 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ ,  $(Y_t, \bar{Y}_t)_{t \geq 0}$  such that  $(Y_0, \bar{Y}_0)$  is distributed according to  $\zeta_0$  and given by (21). Let  $(\mathcal{F}_t)_{t \geq 0}$  be the filtration associated with  $(B_t)_{t \geq 0}$  with  $\mathcal{F}_0$ , the  $\sigma$ -field generated by  $(Y_0, \bar{Y}_0)$ . Then for all  $n \geq 0$  and  $\epsilon > 0$ ,*

$$\begin{aligned} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[ \left\| Y_{\Gamma_{n+1}} - \bar{Y}_{\Gamma_{n+1}} \right\|^2 \right] &\leq \{1 - \gamma_{n+1} (\kappa - 2\epsilon)\} \|Y_{\Gamma_n} - \bar{Y}_{\Gamma_n}\|^2 \\ &\quad + L^2 \gamma_{n+1}^2 (1/(4\epsilon) + \gamma_{n+1}) \left( 2d + L^2 \gamma_{n+1} \|Y_{\Gamma_n} - x^*\|^2 + dL^2 \gamma_{n+1}^2 / 6 \right). \end{aligned}$$

*Proof.* Let  $n \geq 0$  and  $\epsilon > 0$ , and set  $\Delta_n = Y_{\Gamma_n} - \bar{Y}_{\Gamma_n}$  by definition we have:

$$\begin{aligned} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[ \|\Delta_{n+1}\|^2 \right] &= \|\Delta_n\|^2 + \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[ \left\| \int_{\Gamma_n}^{\Gamma_{n+1}} \{\nabla U(Y_s) - \nabla U(\bar{Y}_{\Gamma_n})\} ds \right\|^2 \right] \\ &\quad - 2\gamma_{n+1} \langle \Delta_n, \nabla U(Y_{\Gamma_n}) - \nabla U(\bar{Y}_{\Gamma_n}) \rangle - 2 \int_{\Gamma_n}^{\Gamma_{n+1}} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} [\langle \Delta_n, \{\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\} \rangle ds]. \end{aligned}$$

Using the two inequalities  $|\langle a, b \rangle| \leq \epsilon \|a\|^2 + (4\epsilon)^{-1} \|b\|^2$  and (12), we get

$$\begin{aligned} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[ \|\Delta_{n+1}\|^2 \right] &\leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon)\} \|\Delta_n\|^2 - 2\gamma_{n+1}/(m+L) \|\nabla U(Y_{\Gamma_n}) - \nabla U(\bar{Y}_{\Gamma_n})\|^2 \\ &+ \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[ \left\| \int_{\Gamma_n}^{\Gamma_{n+1}} \{\nabla U(Y_s) - \nabla U(\bar{Y}_{\Gamma_n})\} ds \right\|^2 \right] + \frac{1}{2\epsilon} \int_{\Gamma_n}^{\Gamma_{n+1}} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[ \|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2 \right] ds. \end{aligned} \quad (72)$$

The Young's inequality and the Jensen's inequality imply

$$\begin{aligned} &\mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[ \left\| \int_{\Gamma_n}^{\Gamma_{n+1}} \{\nabla U(Y_s) - \nabla U(\bar{Y}_{\Gamma_n})\} ds \right\|^2 \right] \\ &\leq 2\gamma_{n+1}^2 \|\nabla U(Y_{\Gamma_n}) - \nabla U(\bar{Y}_{\Gamma_n})\|^2 + 2\gamma_{n+1} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[ \int_{\Gamma_n}^{\Gamma_{n+1}} \|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2 ds \right] \end{aligned}$$

Using this inequality, that  $\gamma_1 \leq 1/(m+L)$  and  $(\gamma_k)_{k \geq 1}$  is nonincreasing in (72), imply,

$$\begin{aligned} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[ \|\Delta_{n+1}\|^2 \right] &\leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon)\} \|\Delta_n\|^2 \\ &+ (2\gamma_{n+1} + (2\epsilon)^{-1}) \int_{\Gamma_n}^{\Gamma_{n+1}} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[ \|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2 \right] ds. \end{aligned} \quad (73)$$

By **H1**, the Markov property of  $(Y_t)_{t \geq 0}$  and (66), we have

$$\begin{aligned} &\int_{\Gamma_n}^{\Gamma_{n+1}} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[ \|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2 \right] ds \\ &\leq L^2 \left( d\gamma_{n+1}^2 + dL^2\gamma_{n+1}^4/12 + (1/2)L^2\gamma_{n+1}^3 \|Y_{\Gamma_n} - x^*\|^2 \right). \end{aligned}$$

Plugging this bound in (73) concludes the proof.  $\square$

**Proposition 38.** Assume **H1** and **H3**. Let  $(\gamma_k)_{k \geq 1}$  be a nonincreasing sequence with  $\gamma_1 \leq 1/(m+L)$ . Let  $\zeta_0 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ ,  $(Y_t, \bar{Y}_t)_{t \geq 0}$  such that  $(Y_0, \bar{Y}_0)$  is distributed according to  $\zeta_0$  and given by (21). Then for all  $n \geq 0$  and  $t \in [\Gamma_n, \Gamma_{n+1}]$ :

$$\mathbb{E} \left[ \|Y_t - \bar{Y}_t\|^2 \right] \leq u_n^{(1)}(\gamma) \mathbb{E} \left[ \|Y_0 - \bar{Y}_0\|^2 \right] + u_n^{(4)}(\gamma) + u_{t,n}^{(5)}(\gamma),$$

where  $(u_n^{(1)}(\gamma))_{n \geq 1}$  is given by (27), and

$$\begin{aligned} u_n^{(4)}(\gamma) &\stackrel{\text{def}}{=} L^2 \sum_{i=1}^n \gamma_i^2 \{\kappa^{-1} + \gamma_i\} (2d + dL^2\gamma_i^2/6) \prod_{k=i+1}^n (1 - \kappa\gamma_k/2) \\ &\quad + L^4 \sum_{i=1}^n \delta_i \gamma_i^3 \{\kappa^{-1} + \gamma_{i+1}\} \prod_{k=i+1}^n (1 - \kappa\gamma_k/2) \end{aligned}$$

with

$$\delta_i = e^{-2m\Gamma_{i-1}} \mathbb{E} \left[ \|Y_0 - x^*\|^2 \right] + (1 - e^{-2m\Gamma_{i-1}})(d/m),$$

and

$$u_{t,n}^{(5)}(\gamma) \stackrel{\text{def}}{=} \frac{m+L}{2} \left( \frac{(t-\Gamma_n)^3 L^2}{3} \left\{ \prod_{k=1}^n (1-\kappa\gamma_k) \mathbb{E} \left[ \|\bar{Y}_0 - x^*\|^2 \right] + 2d\kappa^{-1} \right\} + (t-\Gamma_n)^2 d \right).$$

*Proof.* Lemma 37 with  $\epsilon = \kappa/4$ , a straightforward induction and (65) imply for all  $n \geq 0$

$$\mathbb{E} \left[ \|Y_{\Gamma_n} - \bar{Y}_{\Gamma_n}\|^2 \right] \leq u_n^{(1)}(\gamma) \mathbb{E} \left[ \|Y_0 - \bar{Y}_0\|^2 \right] + u_n^{(4)}(\gamma). \quad (74)$$

Now let  $n \geq 0$  and  $t \in [\Gamma_n, \Gamma_{n+1}]$ . By (21),

$$\|Y_t - \bar{Y}_t\|^2 = \|Y_{\Gamma_n} - \bar{Y}_{\Gamma_n}\|^2 - 2 \int_{\Gamma_n}^t \langle \nabla U(Y_s) - \nabla U(\bar{Y}_{\Gamma_n}), Y_s - \bar{Y}_s \rangle ds. \quad (75)$$

Moreover for all  $s \in [\Gamma_n, \Gamma_{n+1}]$ , by (12) we get

$$\begin{aligned} \langle \nabla U(Y_s) - \nabla U(\bar{Y}_{\Gamma_n}), Y_s - \bar{Y}_s \rangle &= \langle \nabla U(Y_s) - \nabla U(\bar{Y}_{\Gamma_n}), Y_s - \bar{Y}_{\Gamma_n} + \bar{Y}_{\Gamma_n} - \bar{Y}_s \rangle \\ &\geq (m+L)^{-1} \|\nabla U(Y_s) - \nabla U(\bar{Y}_{\Gamma_n})\|^2 + \langle \nabla U(Y_s) - \nabla U(\bar{Y}_{\Gamma_n}), \bar{Y}_{\Gamma_n} - \bar{Y}_s \rangle. \end{aligned} \quad (76)$$

Since  $|\langle a, b \rangle| \leq (m+L)^{-1} \|a\|^2 + (m+L) \|b\|^2/4$ , we have

$$\begin{aligned} \langle \nabla U(Y_s) - \nabla U(\bar{Y}_{\Gamma_n}), \bar{Y}_{\Gamma_n} - \bar{Y}_s \rangle \\ \geq -(m+L)^{-1} \|\nabla U(Y_s) - \nabla U(\bar{Y}_{\Gamma_n})\|^2 - (m+L) \|\bar{Y}_s - \bar{Y}_{\Gamma_n}\|^2/4. \end{aligned}$$

Using this inequality in (76), we get

$$\langle \nabla U(Y_s) - \nabla U(\bar{Y}_{\Gamma_n}), Y_s - \bar{Y}_s \rangle \geq -(m+L) \|\bar{Y}_s - \bar{Y}_{\Gamma_n}\|^2/4,$$

and (75) becomes

$$\|Y_t - \bar{Y}_t\|^2 \leq \|Y_{\Gamma_n} - \bar{Y}_{\Gamma_n}\|^2 + ((m+L)/2) \int_{\Gamma_n}^t \|\bar{Y}_s - \bar{Y}_{\Gamma_n}\|^2 ds \quad (77)$$

Therefore, by the previous inequality, it remains to bound the expectation of  $\|\bar{Y}_s - \bar{Y}_{\Gamma_n}\|^2$ . By (21) and using  $\nabla U(x^*) = 0$ ,

$$\|\bar{Y}_s - \bar{Y}_{\Gamma_n}\|^2 = \left\| -(s-\Gamma_n)(\nabla U(\bar{Y}_{\Gamma_n}) - \nabla U(x^*)) + \sqrt{2}(B_s - B_{\Gamma_n}) \right\|^2.$$

Then taking the expectation, using the Markov property of  $(B_t)_{t \geq 0}$  and **H1**, we have

$$\mathbb{E} \left[ \|\bar{Y}_s - \bar{Y}_{\Gamma_n}\|^2 \right] \leq (s-\Gamma_n)^2 L^2 \mathbb{E} \left[ \|\bar{Y}_{\Gamma_n} - x^*\|^2 \right] + 2(s-\Gamma_n)d. \quad (78)$$

The proof follows from taking the expectation in (77), combining (74)-(78), and using Theorem 14.  $\square$

*Proof of Theorem 16.* Let  $\zeta_0$  be an optimal transference plan of  $\mu_0$  and  $\pi$ . Let  $(Y_t, \bar{Y}_t)_{t \geq 0}$  with  $(Y_0, \bar{Y}_0)$  distributed according to  $\zeta_0$  and defined by (21). By definition of  $W_2$  and since for all  $t \geq 0$ ,  $\pi$  is invariant for  $P_t$ ,  $W_2^2(\mu_0 Q^n, \pi) \leq \mathbb{E}[\|Y_{\Gamma_n} - X_{\Gamma_n}\|^2]$ . Then the proof follows from Proposition 38 since  $Y_0$  is distributed according to  $\pi$  and by (16), which shows that for all  $i \in \{1, \dots, n\}$ ,  $\delta_i \leq d/m$ .  $\square$

**Lemma 39.** *Let  $(\gamma_k)_{k \geq 1}$  be a sequence of nonincreasing real numbers,  $\varpi > 0$  and  $\gamma_1 < \varpi^{-1}$ . Then for all  $n \geq 0$ ,  $j \geq 1$  and  $\ell \in \{1, \dots, n+1\}$ ,*

$$\sum_{i=1}^{n+1} \prod_{k=i+1}^{n+1} (1 - \varpi \gamma_k) \gamma_i^j \leq \prod_{k=\ell}^{n+1} (1 - \varpi \gamma_k) \sum_{i=1}^{\ell-1} \gamma_i^j + \frac{\gamma_\ell^{j-1}}{\varpi}.$$

*Proof.* Let  $\ell \in \{1, \dots, n+1\}$ . Since  $(\gamma_k)_{k \geq 1}$  is non-increasing and for every  $t \in \mathbb{R}$ ,  $(1+t) \leq e^t$ ,

$$\begin{aligned} \sum_{i=1}^{n+1} \prod_{k=i+1}^{n+1} (1 - \varpi \gamma_k) \gamma_i^j &= \sum_{i=1}^{\ell-1} \prod_{k=i+1}^{n+1} (1 - \varpi \gamma_k) \gamma_i^j + \sum_{i=\ell}^{n+1} \prod_{k=i+1}^{n+1} (1 - \varpi \gamma_k) \gamma_i^j \\ &\leq \prod_{k=\ell}^{n+1} (1 - \varpi \gamma_k) \sum_{i=1}^{\ell-1} \gamma_i^j + \gamma_\ell^{j-1} \sum_{i=\ell}^{n+1} \prod_{k=i+1}^{n+1} (1 - \varpi \gamma_k) \gamma_i \\ &\leq \prod_{k=\ell}^{n+1} (1 - \varpi \gamma_k) \sum_{i=1}^{\ell-1} \gamma_i^j + \frac{\gamma_\ell^{j-1}}{\varpi}. \end{aligned}$$

$\square$

*Proof of Corollary 17.* By Theorem 16, it suffices to show that  $u_n^{(1)}$  and  $u_n^{(2)}$ , defined by (27) and (28) respectively, goes to 0 as  $n \rightarrow +\infty$ . Using the bound  $1+t \leq e^t$  for  $t \in \mathbb{R}$ , and  $\lim_{n \rightarrow +\infty} \Gamma_n = +\infty$ , we have  $\lim_{n \rightarrow +\infty} u_n^{(1)} = 0$ . Now to show that  $\lim_{n \rightarrow +\infty} u_n^{(2)} = 0$ , a sufficient condition since  $(\gamma_k)_{k \geq 0}$  is non-increasing, is that

$$\lim_{n \rightarrow +\infty} \sum_{i=1}^n \prod_{k=i+1}^n (1 - \kappa \gamma_k / 2) \gamma_i^2 = 0.$$

But since  $(\gamma_k)_{k \geq 1}$  is nonincreasing, there exists  $c \geq 0$  such that  $c\Gamma_n \leq n-1$  and by Lemma 39 applied with  $\ell = \lfloor c\Gamma_n \rfloor$  the integer part of  $c\Gamma_n$ :

$$\sum_{i=1}^n \prod_{k=i+1}^n (1 - \kappa \gamma_k / 2) \gamma_i^2 \leq 2\gamma_{\lfloor c\Gamma_n \rfloor} / \kappa + \exp(-\kappa \Gamma_n (1 - \Gamma_n^{-1} \Gamma_{\lfloor c\Gamma_n \rfloor}) / 2) \sum_{i=1}^{\lfloor c\Gamma_n \rfloor - 1} \gamma_i. \quad (79)$$

Since  $\lim_{k \rightarrow +\infty} \gamma_k = 0$ , by the Cesàro theorem,  $\lim_{n \rightarrow +\infty} (\Gamma_n)^{-1} \Gamma_{\lfloor c\Gamma_n \rfloor} = 0$ , and the conclusion follows from combining in (79), this limit,  $\lim_{k \rightarrow +\infty} \gamma_k = 0$ ,  $\lim_{n \rightarrow +\infty} \Gamma_n = +\infty$  and  $\sum_{i=1}^{\lfloor c\Gamma_n \rfloor - 1} \gamma_i \leq c\gamma_1 \Gamma_n$ .  $\square$

### 6.10. Proofs of Proposition 20, Corollary 21, Lemma 22, and Lemma 27

*Proof of Proposition 20.* (i) Let  $\zeta_0$  be an optimal transference plan of  $\mu_0$  and  $\nu_0$  and  $(Z_k)_{k \geq n-1}$  be a sequence of i.i.d.  $d$ -dimensional Gaussian random variables. We consider the processes  $(X_{n-1,k}^1, X_{n-1,k}^2)_{k \geq n-1}$  with initial distributions equal to  $\zeta_0$  and defined for  $k \geq n-1$  by

$$X_{n-1,k+1}^j = X_{n-1,k}^j - \gamma_{k+1} \nabla U(X_{n-1,k}^j) + \sqrt{2} \gamma_{k+1} Z_{k+1} \quad j = 1, 2. \quad (80)$$

Using (80), we get for any  $p \geq n \geq 0$ ,  $W_2^2(\mu_0 Q_\gamma^{n,p}, \nu_0 Q_\gamma^{n,p}) \leq \mathbb{E} \left[ \|X_{n-1,p}^1 - X_{n-1,p}^2\|^2 \right]$  and (12) implies for  $k \geq n-1$ ,

$$\begin{aligned} \|X_{n-1,k+1}^1 - X_{n-1,k+1}^2\|^2 &= \|X_{n-1,k}^1 - X_{n-1,k}^2\|^2 + \gamma_{k+1}^2 \|\nabla U(X_{n-1,k}^1) - \nabla U(X_{n-1,k}^2)\|^2 \\ &\quad - 2\gamma_{k+1} \langle X_{n-1,k}^1 - X_{n-1,k}^2, \nabla U(X_{n-1,k}^1) - \nabla U(X_{n-1,k}^2) \rangle \leq (1 - \kappa \gamma_{k+1}) \|X_{n-1,k}^1 - X_{n-1,k}^2\|^2. \end{aligned}$$

Therefore by a straightforward induction we get for all  $p \geq n$ ,

$$\|X_{n-1,p}^1 - X_{n-1,p}^2\|^2 \leq \prod_{k=n}^p (1 - \kappa \gamma_k) \|X_{n-1,n-1}^1 - X_{n-1,n-1}^2\|^2.$$

(ii) Recall that for all  $\mu, \nu$  probability measures on  $\mathbb{R}^d$  and  $p \leq q$ ,  $W_p(\mu, \nu) \leq W_q(\mu, \nu)$ . Hence, for all  $y, z \in \mathbb{R}^d$ , the Monge-Kantorovich theorem (4):

$$|Q_\gamma^{n,p} f(y) - Q_\gamma^{n,p} f(z)| \leq \|f\|_{\text{Lip}} W_1(\delta_y Q_\gamma^{n,p}, \delta_z Q_\gamma^{n,p}) \leq \|f\|_{\text{Lip}} W_2(\delta_y Q_\gamma^{n,p}, \delta_z Q_\gamma^{n,p}).$$

The proof then follows from (i). □

*Proof of Corollary 21.* By (39),  $\|\Phi_{n,k}^N\|_{\text{Lip}} \leq \sum_{i=k+1}^{N+n} \omega_{i,n}^N \|Q_\gamma^{k+2,i} f\|_{\text{Lip}}$ . Using Proposition 20-(ii), the bound  $(1-t)^{1/2} \leq 1-t/2$  for  $t \in [0, 1]$  and the definition of  $\omega_{i,n}^N$  given by (31), we have

$$\begin{aligned} \|\Phi_{n,k}^N\|_{\text{Lip}} &\leq \|f\|_{\text{Lip}} \sum_{i=k+1}^{N+n} \omega_{i,n}^N \prod_{j=k+2}^i (1 - \kappa \gamma_j / 2) \\ &\leq 2 \|f\|_{\text{Lip}} (\kappa \Gamma_{N+2, N+n+1})^{-1} \sum_{i=k+1}^{N+n} \left\{ \prod_{j=k+2}^i (1 - \kappa \gamma_j / 2) - \prod_{j=k+2}^{i+1} (1 - \kappa \gamma_j / 2) \right\} \\ &\leq 2 \|f\|_{\text{Lip}} (\kappa \Gamma_{N+2, N+n+1})^{-1}. \end{aligned}$$

Finally, the proof follows from Lemma 19. □

*Proof of Lemma 22.* By decomposing  $g(X_p) - \mathbb{E}_y^{\mathcal{G}_n} [g(X_p)] = \sum_{k=n+1}^p \{ \mathbb{E}_y^{\mathcal{G}_k} [g(X_p)] - \mathbb{E}_y^{\mathcal{G}_{k-1}} [g(X_p)] \}$ , and using  $\mathbb{E}_y^{\mathcal{G}_k} [g(X_p)] = Q_\gamma^{k+1,p} g(X_k)$ , we get

$$\begin{aligned} \text{Var}_y^{\mathcal{G}_n} \{g(X_p)\} &= \sum_{k=n+1}^p \mathbb{E}_y^{\mathcal{G}_n} \left[ \mathbb{E}_y^{\mathcal{G}_{k-1}} \left[ \left( \mathbb{E}_y^{\mathcal{G}_k} [g(X_p)] - \mathbb{E}_y^{\mathcal{G}_{k-1}} [g(X_p)] \right)^2 \right] \right] \\ &= \sum_{k=n+1}^p \mathbb{E}_y^{\mathcal{G}_n} \left[ R_{\gamma_k} \left\{ Q_\gamma^{k+1,p} g(\cdot) - R_{\gamma_k} Q_\gamma^{k+1,p} g(X_{k-1}) \right\}^2 (X_{k-1}) \right]. \end{aligned}$$

Lemma 19 implies  $\text{Var}_y^{\mathcal{G}^n} \{g(X_p)\} \leq 2 \sum_{k=n+1}^p \gamma_k \|Q_\gamma^{k+1,p} g\|_{\text{Lip}}^2$ . The proof follows from Proposition 20-(ii) and Lemma 39, using the bound  $(1-t)^{1/2} \leq 1-t/2$  for  $t \in [0, 1]$ .  $\square$

*Proof of Corollary 23.* By (40),  $\Psi_n^N$  is Lipschitz function with

$$\|\Psi_n^N\|_{\text{Lip}} = \|f\|_{\text{Lip}} \sum_{i=N+1}^{N+n} \omega_{i,n}^N \|Q_\gamma^{N+1,i} f\|_{\text{Lip}}.$$

Using Proposition 20-(ii), the bound  $(1-t)^{1/2} \leq 1-t/2$  for  $t \in [0, 1]$  and the definition of  $\omega_{i,n}^N$  given by (31), we have

$$\begin{aligned} \|\Psi_n^N\|_{\text{Lip}} &\leq \|f\|_{\text{Lip}} \sum_{i=N+1}^{N+n} \omega_{i,n}^N \prod_{j=N+2}^i (1 - \kappa\gamma_j/2) \\ &\leq 2 \|f\|_{\text{Lip}} (\kappa\Gamma_{N+2, N+n+1})^{-1} \sum_{i=N+1}^{N+n} \left\{ \prod_{j=N+2}^i (1 - \kappa\gamma_j/2) - \prod_{j=N+2}^{i+1} (1 - \kappa\gamma_j/2) \right\} \\ &\leq 2 \|f\|_{\text{Lip}} (\kappa\Gamma_{N+2, N+n+1})^{-1}. \end{aligned}$$

The proof follows from Lemma 22.  $\square$

*Proof of Lemma 27.* Let  $(X_n)_{n \geq 0}$  the Euler approximation given by (2) and started at  $y \in \mathbb{R}^d$ . By decomposing  $g(X_p) - \mathbb{E}_y^{\mathcal{G}^n} [g(X_p)] = \sum_{k=n+1}^p \{\mathbb{E}_y^{\mathcal{G}^k} [g(X_p)] - \mathbb{E}_y^{\mathcal{G}^{k-1}} [g(X_p)]\}$ , and using  $\mathbb{E}_y^{\mathcal{G}^k} [g(X_p)] = Q_\gamma^{k+1,p} g(X_k)$ , we get

$$\begin{aligned} &\mathbb{E}_y^{\mathcal{G}^n} [\exp(\lambda \{g(X_p) - \mathbb{E}_y^{\mathcal{G}^n} [g(X_p)]\})] \\ &= \mathbb{E}_y^{\mathcal{G}^n} \left[ \prod_{k=n+1}^p \mathbb{E}_y^{\mathcal{G}^{k-1}} [\exp(\lambda \{\mathbb{E}_y^{\mathcal{G}^k} [g(X_p)] - \mathbb{E}_y^{\mathcal{G}^{k-1}} [g(X_p)]\})] \right] \\ &= \mathbb{E}_y^{\mathcal{G}^n} \left[ \prod_{k=n+1}^p R_{\gamma_k} \exp(\lambda \{Q_\gamma^{k+1,p} g(\cdot) - R_{\gamma_k} Q_\gamma^{k+1,p} g(X_{k-1})\}) (X_{k-1}) \right]. \end{aligned}$$

By the Gaussian log-Sobolev inequality Lemma 25, we get

$$\mathbb{E}_y^{\mathcal{G}^n} [\exp(\lambda \{g(X_p) - \mathbb{E}_y^{\mathcal{G}^n} [g(X_p)]\})] \leq \exp \left( \lambda^2 \sum_{k=n+1}^p \gamma_k \|Q_\gamma^{k+1,p} g\|_{\text{Lip}}^2 \right).$$

The proof follows from Proposition 20-(ii) and Lemma 39, using the bound  $(1-t)^{1/2} \leq 1-t/2$  for  $t \in [0, 1]$ .  $\square$

## Acknowledgements

The authors are indebted to Arnaud Guillin for sharing his knowledge of Poincaré and log-Sobolev inequalities.

## References

- [1] J. H. Albert and S. Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679, 1993.
- [2] D. Bakry, F. Barthe, P. Cattiaux, and A. Guillin. A simple proof of the Poincaré inequality for a large class of probability measures. *Electronic Communications in Probability [electronic only]*, 13:60–66, 2008.
- [3] D. Bakry, P. Cattiaux, and A. Guillin. Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.*, 254(3):727–759, 2008.
- [4] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.
- [5] S. G. Bobkov. Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.*, 27(4):1903–1921, 10 1999.
- [6] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.
- [7] P. Cattiaux and A. Guillin. Trends to equilibrium in total variation distance. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(1):117–145, 2009.
- [8] H. M. Choi and J. P. Hobert. The Polya-Gamma Gibbs sampler for Bayesian logistic regression is uniformly ergodic. *Electron. J. Statist.*, 7:2054–2064, 2013.
- [9] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. *Statist. Sci.*, 28(3):424–446, 2013.
- [10] A. Dalalyan. Theoretical guarantees for approximate sampling from a smooth and log-concave density. submitted 1412.7392, arXiv, December 2014.
- [11] A. S. Dalalyan and A. B. Tsybakov. Sparse regression learning by aggregation and Langevin Monte-Carlo. *J. Comput. System Sci.*, 78(5):1423–1443, 2012.
- [12] A. Durmus and Moulines E. Supplement supplement to “non-asymptotic convergence analysis for the unadjusted langevin algorithm”, 2015.
- [13] S. Frühwirth-Schnatter and R. Frühwirth. Data augmentation and MCMC for binary and multinomial logit models statistical modelling and regression structures. In Thomas Kneib and Gerhard Tutz, editors, *Statistical Modelling and Regression Structures*, chapter 7, pages 111–132. Physica-Verlag HD, Heidelberg, 2010.
- [14] R. B. Gramacy and N. G. Polson. Simulation-based regularized logistic regression. *Bayesian Anal.*, 7(3):567–590, 09 2012.
- [15] Grenander. Tutorial in pattern theory. Division of Applied Mathematics, Brown University, Providence.
- [16] U. Grenander and M. I. Miller. Representations of knowledge in complex systems. *J. Roy. Statist. Soc. Ser. B*, 56(4):549–603, 1994. With discussion and a reply by the authors.
- [17] T. E. Hanson, A J. Branscum, and W. O. Johnson. Informative  $g$ -priors for logistic regression. *Bayesian Anal.*, 9(3):597–611, 2014.
- [18] E. Hazan and S Kale. Beyond the regret minimization barrier: optimal algorithms for stochastic strongly-convex optimization. *Journal of Machine Learning Research*, 15(1):2489–2512, 2014.
- [19] C. C. Holmes and L. Held. Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Anal.*, 1(1):145–168, 03 2006.
- [20] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer New York, 1991.

- [21] D. Lamberton and G. Pagès. Recursive computation of the invariant distribution of a diffusion. *Bernoulli*, 8(3):367–405, 2002.
- [22] D. Lamberton and G. Pagès. Recursive computation of the invariant distribution of a diffusion: the case of a weakly mean reverting drift. *Stoch. Dyn.*, 3(4):435–451, 2003.
- [23] Lemaire. *Estimation de la mesure invariante d'un processus de diffusion*. PhD thesis, Université Paris-Est, 2005.
- [24] V. Lemaire and S. Menozzi. On some non asymptotic bounds for the Euler scheme. *Electron. J. Probab.*, 15:no. 53, 1645–1681, 2010.
- [25] S. P. Meyn and R. L. Tweedie. Stability of Markovian Processes III: Foster-Lyapunov Criteria for Continuous-Time Processes. *Advances in Applied Probability*, 25(3), 1993.
- [26] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Applied Optimization. Springer, 2004.
- [27] G. Parisi. Correlation functions and computer simulations. *Nuclear Physics B*, 180:378–384, 1981.
- [28] N. G. Polson, J. G. Scott, and J. Windle. Bayesian inference for logistic models using Polya-Gamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349, 2013.
- [29] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.
- [30] D. Sabanés Bové and L. Held. Hyper- $g$  priors for generalized linear models. *Bayesian Anal.*, 6(3):387–410, 2011.
- [31] C. Villani. *Optimal transport : old and new*. Grundlehren der mathematischen Wissenschaften. Springer, Berlin, 2009.
- [32] J. Windle, N. G. Polson, and J. G. Scott. Bayeslogit: Bayesian logistic regression, 2013. <http://cran.r-project.org/web/packages/BayesLogit/index.html> R package version 0.2.