Sampling from a strongly log-concave distribution with the Unadjusted Langevin Algorithm

Alain Durmus, Éric Moulines

Sampling from a strongly log-concave distribution with the Unadjusted Langevin Algorithm

Alain Durmus[1] Éric Moulines [2]

April 19, 2016

Keywords: Markov Chain Monte Carlo, Metropolis Adjusted Langevin Algorithm, Rate of convergence

AMS subject classification (2010): primary 65C05, 60F05, 62L10; secondary 65C40, 60J05,93E35

Abstract

We consider in this paper the problem of sampling a probability distribution π having a density w.r.t. the Lebesgue measure on $\mathbb{R}^d$, known up to a normalisation factor $x \mapsto \mathrm{e}^{-U(x)}/\int_{\mathbb{R}^d} \mathrm{e}^{-U(y)}\mathrm{d}y$. Under the assumption that U is continuously differentiable, ∇U is globally Lipshitz and U is strongly convex, we obtain non-asymptotic bounds for the convergence to stationarity in Wasserstein distances of the sampling method based on the Euler discretization of the Langevin stochastic differential equation for both constant and decreasing step sizes. The dependence on the dimension of the state space of the obtained bounds is studied to demonstrate the applicability of this method in the high dimensional setting. The convergence of an appropriately weighted empirical measure is also investigated and bounds for the mean square error and exponential deviation inequality for Lipschitz functions are reported. Some numerical results are presented to illustrate our findings.

1 Introduction

Let π be a probability distribution on $\mathbb{R}^d$, $d \geq 1$, with density $x \mapsto \mathrm{e}^{-U(x)}/\int_{\mathbb{R}^d} \mathrm{e}^{-U(y)}\mathrm{d}y$ w.r.t. the Lebesgue measure, where U is continuously differentiable, gradient Lipshitz and strongly convex. Consider the Langevin stochastic differential equation associated with π:

$$\mathrm{d}Y_t = -\nabla U(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \;, \tag{1}$$

where $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion. Under the stated assumptions on U, π satisfies a log-Sobolev inequalities (see [2, 6, 3]) and the Markov semi-group associated with the Langevin diffusion $(Y_t)_{t\geq 0}$ converges exponentially fast to π with a rate independent of the dimension of the state space. We study in this paper the sampling method based on the Euler-Maruyama discretization scheme associated to the Langevin diffusion, which defines a (possibly) non-homogeneous, discrete-time Markov chain given by

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

[1]LTCI, Telecom ParisTech 46 rue Barrault, 75634 Paris Cedex 13, France. alain.durmus@telecom-paristech.fr
[2]Centre de Mathématiques Appliquées, UMR 7641, Ecole Polytechnique, France. eric.moulines@polytechnique.edu

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.

This method was originally proposed in the physics literature by [19] and introduced in the computational statistics community by [11] and [12]. It has been studied in depth by [21], which proposed to use a Metropolis-Hastings step at each iteration to enforce reversibility w.r.t. π leading to the Metropolis Adjusted Langevin Algorithm (MALA). They coin the term *unadjusted* Langevin algorithm (ULA) to stress the fact that the Metropolis-Hastings step is avoided.

We obtain in this paper non-asymptotic and computable bounds between the marginal laws of the Markov chain $(X_n)_{n\geq 0}$ defined by the Euler discretization and the target distribution π in Wasserstein distance in nonincreasing step sizes. When the sequence of step sizes is constant $\gamma_k = \gamma$ for all $k \geq 0$, the Markov chain $(X_n)_{n\geq 0}$ has a unique stationary distribution π_γ (see [21]), which in most of the cases differs from the distribution π. Quantitative estimates between π and π_γ is obtained. When $(\gamma_k)_{k\geq 1}$ decreases to zero and $\sum_{k=1}^{\infty} \gamma_k = \infty$ then we show that the marginal distribution of the non-homogeneous Markov chain $(X_n)_{n\geq 0}$ converges to the target distribution π with explicit expression for the convergence rate.

The paper is organized as follows. In Section 2, we study the convergence in Wasserstein distances of the Euler discretization for constant and decreasing stepsizes. In Section 2.1 we provide non-asymptotic bounds of convergence of the weighted empirical measure. Some numerical illustrations are given Section 3 to support our claims. The proofs are given in Section 4. Some technical derivations are carried out in a supplementary paper [8].

Notations and conventions

Denote by $\mathcal{B}(\mathbb{R}^d)$ the Borel σ-field of $\mathbb{R}^d$, $\mathbb{F}(\mathbb{R}^d)$ 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 μ a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $f \in \mathbb{F}(\mathbb{R}^d)$ a μ-integrable function, denote by $\mu(f)$ the integral of f w.r.t. μ. We say that ζ is a transference plan of μ and ν 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 $\mathbb{R}^d$, $\zeta(\mathrm{A} \times \mathbb{R}^d) = \mu(\mathrm{A})$ and $\zeta(\mathbb{R}^d \times \mathrm{A}) = \nu(\mathrm{A})$. We denote by $\Pi(\mu,\nu)$ the set of transference plans of μ and ν. Furthermore, we say that a couple of $\mathbb{R}^d$-random variables (X, Y) is a coupling of μ and ν if there exists $\zeta \in \Pi(\mu,\nu)$ such that (X,Y) are distributed according to ζ. For two probability measures μ and ν, we define the Wasserstein distance of order $p \geq 1$ as

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

By [23, Theorem 4.1], for all μ,ν probability measure on $\mathbb{R}^d$, there exists a transference plan $\zeta^\star \in \Pi(\mu,\nu)$ such that for any coupling (X,Y) distributed according to $\zeta^\star$, $W_p(\mu,\nu) = \mathbb{E}[\|X-Y\|^p]^{1/p}$. 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 $\mu \in \mathcal{P}_p(\mathbb{R}^d)$, $\int_{\mathbb{R}^d} \|x\|^p \, \mu(\mathrm{d}x) < +\infty$. By [23, 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 $f : \mathbb{R}^d \to \mathbb{R}$ 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 $\|f\|_{\mathrm{Lip}} = \inf\{|f(x)-f(y)| \, \|x-y\|^{-1} \mid x,y \in \mathbb{R}^d, x \neq y\}$. The Monge-Kantorovich theorem (see [23, Theorem 5.9]) implies that for all μ,ν probabilities measure on $\mathbb{R}^d$,

$$W_1(\mu,\nu) = \sup\left\{ \int_{\mathbb{R}^d} f(x)\mu(\mathrm{d}x) - \int_{\mathbb{R}^d} f(x)\nu(\mathrm{d}x) \mid f : \mathbb{R}^d \to \mathbb{R} \; ; \; \|f\|_{\mathrm{Lip}} \leq 1 \right\} . \tag{3}$$

For all $x \in \mathbb{R}^d$ and $M > 0$, we denote by $\mathrm{B}(x, M)$, the ball centered at x of radius M. Let $n, m \in \mathbb{N}^*$ and M be a $n \times n$-matrix, then denote by M^{T} the transpose of M and $\|M\|$ the Frobenius associated with M defined by $\|M\| = \mathrm{Tr}(M^{\mathrm{T}} M)$. Let $n, m \in \mathbb{N}^*$ and $F : \mathbb{R}^n \to \mathbb{R}^m$ be a twice continuously differentiable function. Denote by ∇F and $\nabla^2 F$ the Jacobian and the Hessian of F respectively. Denote also by $\vec{\Delta} F$ the vectorial Laplacian of F defined by: for all $x \in \mathbb{R}^d$, $\vec{\Delta} F(x)$ is the vector of $\mathbb{R}^m$ such that for all $i \in \{1, \cdots, m\}$, the i-th component of $\vec{\Delta} F(x)$ is equals to $\sum_{j=1}^d (\partial^2 F_i / \partial x_j^2)(x)$. 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 Non-asymptotic bounds in Wasserstein distance of order 2 for ULA

Consider the following assumption on the potential U:

H1. *The function U is 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 **H**1, if μ_0 is a probability measure satisfying $\int \|x\|^2 \mu_0(\mathrm{d}x) < \infty$ then by [17, Theorem 2.5, Theorem 2.9 Chapter 5] there exists a unique strong solution $(Y_t)_{t \geq 0}$ to (1) with initial distribution μ_0. Denote by $(P_t)_{t \geq 0}$ the semi-group associated with (1), which is reversible w.r.t. π, and hence admits π as its (unique) invariant measure.

H2. *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 **H**2, [18, Theorem 2.1.8] shows that U has a unique minimizer $x^\star \in \mathbb{R}^d$. If in addition **H**1 holds, then [18, Theorem 2.1.12, Theorem 2.1.9] shows that for all $x, y \in \mathbb{R}^d$:

$$\langle \nabla U(y) - \nabla U(x), y - x \rangle \geq (\kappa/2) \|y - x\|^2 + \frac{1}{m + L} \|\nabla U(y) - \nabla U(x)\|^2 \ , \tag{4}$$

$$\langle \nabla U(y) - \nabla U(x), y - x \rangle \geq m \|y - x\|^2 \ , \tag{5}$$

where

$$\kappa = \frac{2mL}{m + L} \ . \tag{6}$$

Note that **H**1 and (5) imply that $L \geq m$. We first obtain the geometric rate of convergence to stationarity of the semi-group in Wasserstein distance. It is worthwhile to note that these bounds do not depend on the dimension d.

Theorem 1. *Assume **H**1 and **H**2.*

(i) For all $p \geq 2$, probability measures μ and $\nu \in \mathcal{P}_p(\mathbb{R}^d)$ and $t \geq 0$,

$$W_p(\mu P_t, \nu P_t) \leq \mathrm{e}^{-mt} W_p(\mu, \nu)$$

(ii) The stationary distribution π satisfies

$$\int_{\mathbb{R}^d} \|x - x^\star\|^2 \pi(\mathrm{d}x) \leq d/m \ . \tag{7}$$

Proof. Most of the statement is well known; see [4] and the references therein. Nevertheless for completeness, we provide the proof in Section 4.1. □

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

$$\Gamma_{n,p} \stackrel{\text{def}}{=} \sum_{k=n}^{p} \gamma_k \;, \qquad \Gamma_n = \Gamma_{1,n} \;. \tag{8}$$

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

$$R_\gamma(x,\mathrm{A}) = \int_{\mathrm{A}} (4\pi\gamma)^{-d/2} \exp\left(-(4\gamma)^{-1} \left\|y - x + \gamma \nabla U(x)\right\|^2\right) \mathrm{d}y \;. \tag{9}$$

Under **H**1 R_γ is strongly Feller, irreducible, strongly aperiodic. 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

$$Q_\gamma^{n,p} = R_{\gamma_n} \cdots R_{\gamma_p} \;, \qquad Q_\gamma^n = Q_\gamma^{1,n} \tag{10}$$

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 [21, Section 3]; We generalize these results to multidimensional diffusions and decreasing stepsizes.

Theorem 2. *Assume **H**1 and **H**2. For any $\gamma \in (0, 2/(m+L))$, R_γ has a unique stationary distribution π_γ. Moreover, for all $p \geq 1$, $\pi_\gamma \in \mathcal{P}_p(\mathbb{R}^d)$ and for all probability measure $\mu \in \mathcal{P}_{2p}(\mathbb{R}^d)$, we have for all $n \geq 0$:*

$$W_{2p}(\mu R_\gamma^n, \pi_\gamma) \leq (1-\kappa\gamma)^{np} W_{2p}(\mu, \pi_\gamma) \;. \tag{11}$$

Proof. The proof is postponed to Section 4.2. □

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 π 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 between the diffusion and the linear interpolation of the Euler discretization. In the strongly convex case, an obvious candidate is the synchronous coupling $(Y_t, \overline{Y}_t)_{t\geq 0}$ for all $n \geq 0$ and $t \in [\Gamma_n, \Gamma_{n+1})$ by

$$\begin{cases} Y_t = Y_{\Gamma_n} - \int_{\Gamma_n}^t \nabla U(Y_s)\mathrm{d}s + \sqrt{2}(B_t - B_{\Gamma_n}) \\ \bar{Y}_t = \bar{Y}_{\Gamma_n} - \nabla U(\bar{Y}_{\Gamma_n})(t - \Gamma_n) + \sqrt{2}(B_t - B_{\Gamma_n}) \;, \end{cases} \tag{12}$$

where $(\Gamma_n)_{n\geq 1}$ is given in (8). 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 μ_0 and ν_0 are the marginals of ζ_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 π of the Langevin diffusion (1).

Theorem 3. *Assume **H**1 and **H**2. 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 $n \geq 1$,*

$$W_2^2(\mu_0 Q_\gamma^n, \pi) \leq u_n^{(1)}(\gamma) W_2^2(\mu_0, \pi) + u_n^{(2)}(\gamma) \;, \tag{13}$$

where

$$u_n^{(1)}(\gamma) \stackrel{\text{def}}{=} \prod_{k=1}^{n} (1 - \kappa\gamma_k/2) \tag{14}$$

and

$$u_n^{(2)}(\gamma) \stackrel{\text{def}}{=} L^2 \sum_{i=1}^{n} \gamma_i^2 \left\{\kappa^{-1} + \gamma_i\right\} (2d + dL^2\gamma_i/m + dL^2\gamma_i^2/6) \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2) \;, \tag{15}$$

where κ *is defined in* (6) *.*

Proof. The proof is postponed to Section 4.3. □

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

Corollary 4. *Assume* ***H****1 and* ***H****2. Let* $(\gamma_k)_{k\geq 1}$ *be a nonincreasing sequence with* $\gamma_1 \leq 1/(m+L)$*. Assume that* $\lim_{k\to\infty} \gamma_k = 0$ *and* $\lim_{n\to+\infty} \Gamma_n = +\infty$*. Then for all* $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$*,*

$$\lim_{n\to\infty} W_2(\mu_0 Q_\gamma^n, \pi) = 0 \;.$$

Proof. The proof is postponed to Section 4.4. □

In the case of constant stepsizes $\gamma_k = \gamma$ for all $k \geq 1$, we can deduce from Theorem 3, a bound between π and the stationary distribution π_γ of R_γ.

Corollary 5. *Assume* ***H****1 and* ***H****2. Let* $(\gamma_k)_{k\geq 1}$ *be a constant sequence* $\gamma_k = \gamma$ *for all* $k \geq 1$ *with* $\gamma_1 \leq 1/(m+L)$*. Then*

$$W_2^2(\pi, \pi_\gamma) \leq 2\kappa^{-1} L^2 \gamma \left\{\kappa^{-1} + \gamma\right\} (2d + dL^2\gamma/m + dL^2\gamma^2/6) \;.$$

Proof. The proof is postponed to Section 4.5. □

We can improve these bound under additional regularity assumptions on the potential U.

H3. *The potential* U *is three times continuously differentiable and there exists* $\tilde{L}$ *such that for all* $x, y \in \mathbb{R}^d$*:*

$$\left\|\nabla^2 U(x) - \nabla^2 U(y)\right\| \leq \tilde{L} \left\|x - y\right\| \;. \tag{16}$$

Note that under **H**1 and **H**3, we have that for all $x, y \in \mathbb{R}^d$,

$$\left\|\nabla^2 U(x) y\right\| \leq L \left\|y\right\| \;, \quad \left\|\vec{\Delta}(\nabla U)(x)\right\|^2 \leq d\tilde{L}^2 \;. \tag{17}$$

Theorem 6. *Assume* ***H****1,* ***H****2 and* ***H****3. 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* $n \geq 1$*,*

$$W_2^2(\mu_0 Q_\gamma^n, \pi) \leq u_n^{(1)}(\gamma) W_2^2(\mu_0, \pi) + u_n^{(3)}(\gamma) \;, \tag{18}$$

where $u_n^{(1)}$ *is given by* (14) *and*

$$u_n^{(3)}(\gamma) \stackrel{\text{def}}{=} \sum_{i=1}^{n} d\gamma_i^3 \left\{2L^2 + \kappa^{-1}(\tilde{L}^2/3 + \gamma_i L^4 + 4L^4/(3m)) + \gamma_i L^4(\gamma_i/6 + m^{-1})\right\} \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2) \;, \tag{19}$$

where κ *is defined in* (6) *.*

Proof. The proof is postponed to Section 4.6. □

In the case of constant stepsizes $\gamma_k = \gamma$ for all $k \geq 1$, we can deduce from Theorem 6, a sharper bound between π and the stationary distribution π_γ of R_γ.

Corollary 7. *Assume* **H**1 *and* **H**2*. Let* $(\gamma_k)_{k\geq 1}$ *be a constant sequence* $\gamma_k = \gamma$ *for all* $k \geq 1$ *with* $\gamma_1 \leq 1/(m+L)$*. Then*

$$W_2^2(\pi, \pi_\gamma) \leq 2\kappa^{-1} d\gamma^2 \left\{ 2L^2 + \kappa^{-1}(\tilde{L}^2/3 + \gamma L^4 + 4L^4/(3m)) + \gamma L^4(\gamma/6 + m^{-1}) \right\} .$$

Proof. The proof follows the same line as the proof of Corollary 5 and is omitted. □

Let $x^\star$ be the unique minimizer of U. Since for all $y \in \mathbb{R}^d$ $\|x - y\|^2 \leq 2(\|x - x^\star\|^2 + \|x^\star - y\|^2)$, using (7), we get:

$$W_2^2(\delta_x, \pi) \leq 2(\|x - x^\star\|^2 + d/m) . \tag{20}$$

If $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$, we have $W_2^2(\mu_0, \pi) \leq \int \mu_0(\mathrm{d}x) W_2^2(\delta_x, \pi)$. Hence, the right hand side of (13) and (18) scales linearly with the dimension d. When $\gamma_k = \gamma$ for all $k \geq 1$, (14), (15) (19) imply

$$\begin{cases} u_n^{(1)}(\gamma) = (1 - \kappa\gamma/2)^n , \quad u_n^{(2)}(\gamma) \leq 2\kappa^{-1}\gamma \left\{\kappa^{-1} + \gamma\right\} (2d + dL^2\gamma/m + dL^2\gamma^2/6) , \\ u_n^{(3)}(\gamma) \leq 2\kappa^{-1} d\gamma^2 \left\{ 2L^2 + \kappa^{-1}(\tilde{L}^2/3 + \gamma L^4 + 4L^4/(3m)) + \gamma L^4(\gamma/6 + m^{-1}) \right\} . \end{cases} \tag{21}$$

Using this bound, given $\epsilon > 0$, we may determine the smallest number of iterations and an associated step-size γ, starting from x, to approach the stationary distribution in the Wasserstein distance $W_2(\delta_x Q_n^\gamma, \pi)$ with a precision ϵ. Details and further discussions are included in the supplementary paper [8].

Based on Theorem 3 and Theorem 6, we can obtain explicit bounds for $W_2^2(\delta_x Q_\gamma^n, \pi)$ 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 and Table 2, see [8, Section 1-2] for details. Two regimes can be observed as in stochastic approximation in the case of Theorem 3.

	$\alpha \in (0,1)$	$\alpha = 1$
Order of convergence	$\mathcal{O}(n^{-\alpha})$	$\mathcal{O}(n^{-1})$ for $\gamma_1 > 2\kappa^{-1}$ see [8, Section 3]

Table 1: Order of convergence of $W_2(\delta_x Q_\gamma^n, \pi)$ for $\gamma_k = \gamma_1 k^{-\alpha}$ under **H**1 and **H**2

	$\alpha \in (0,1)$	$\alpha = 1$
Order of convergence	$\mathcal{O}(n^{-2\alpha})$	$\mathcal{O}(n^{-2})$

Table 2: Order of convergence of $W_2(\delta_x Q_\gamma^n, \pi)$ for $\gamma_k = \gamma_1 k^{-\alpha}$ under **H**1, **H**2 and **H**3

We now consider the 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 γ_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 3, see [8, Section 1-2] for details.

	Optimal choice of γ_1	Bound on $W_2^2(\delta_x Q_\gamma^n, \pi)$
$\alpha \in [0,1)$	$\mathcal{O}(n^{\alpha-1}\log(n))$	$\mathcal{O}(dn^{-1}\log(n))$

Table 3: Order of the optimal choice of γ_1 for the fixed horizon setting and implied bound on $W_2^2(\delta_x Q_\gamma^n, \pi)$ based on Theorem 3

Moreover, these bounds for a fixed number of iterations implies using the doubling trick (see [14]) an anytime algorithm which guarantees for all $n \geq 1$ and $x \in \mathbb{R}^d$ that $W_2(\delta_x Q_\gamma^n, \pi)$ is $\mathcal{O}((\log(n)n^{-1})^{1/2})$ or $\mathcal{O}((\log(n)n^{-1})$.

	Optimal choice of γ_1	Bound on $W_2^2(\delta_x Q_\gamma^n, \pi)$
$\alpha \in [0,1)$	$\mathcal{O}(n^{\alpha-1}\log(n))$	$\mathcal{O}(dn^{-2}\log^2(n))$

Table 4: Order of the optimal choice of γ_1 for the fixed horizon setting and implied bound on $W_2^2(\delta_x Q_\gamma^n, \pi)$ based on Theorem 6

2.1 Mean square error and concentration

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

$$\hat{\pi}_n^N(f) = \sum_{k=N+1}^{N+n} \omega_{k,n}^N f(X_k) \;, \quad \omega_{k,n}^N = \gamma_{k+1}\Gamma_{N+2,N+n+1}^{-1} \;. \tag{22}$$

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 (8). 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:

$$\mathrm{MSE}_f(N,n) = \mathbb{E}_x\left[\left|\hat{\pi}_n^N(f) - \pi(f)\right|^2\right] = \left\{\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right\}^2 + \mathrm{Var}_x\left\{\hat{\pi}_n^N(f)\right\} \;. \tag{23}$$

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

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

Using Theorem 3, we end up with the following bound.

Proposition 8. *Assume* ***H****1 and* ***H****2. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$. Let $x^\star$ 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 $f : \mathbb{R}^d \to \mathbb{R}$:*

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

where $u_n^{(1)}(\gamma)$ is given in (14) and $w_n(\gamma)$ is equal to $u_n^{(2)}(\gamma)$ defined by (15) and to $u_n^{(3)}(\gamma)$, defined by (19), if ***H****3 holds.*

Consider now the variance term. To control this term, we adapt the proof of [16, Theorem 2] for homogeneous Markov chain to our inhomogeneous setting, and we have:

Theorem 9. *Assume* ***H****1 and* ***H****2. 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 $f : \mathbb{R}^d \to \mathbb{R}$, we get*

$$\mathrm{Var}_x\left\{\hat{\pi}_n^N(f)\right\} \leq 8\kappa^{-2}\|f\|_{\mathrm{Lip}}^2 \Gamma_{N+2,N+n+1}^{-1} v_{N,n}(\gamma) \;,$$

where

$$v_{N,n}(\gamma) \overset{def}{=} \left\{1 + \Gamma_{N+2,N+n+1}^{-1}(\kappa^{-1} + 2/(m+L))\right\} . \quad (24)$$

Proof. The proof is postponed to Section 4.7. □

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 8 and the variance Theorem 9) 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 [8, Section 5]. The order of the bounds (up to numerical constants) of the MSE are summarized in Table 5 as a function of γ_1, n and N. If the total number of iterations $n+N$ is held fixed

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

Table 5: Bound for the MSE for $\gamma_k = \gamma_1 k^{-\alpha}$ for fixed γ_1 and N under **H**1 and **H**2

	Bound for the MSE
$\alpha = 0$	$\gamma_1^2 + (\gamma_1 n)^{-1}\exp(-\kappa\gamma_1 N/2)$
$\alpha \in (0,1/3)$	$\gamma_1^2 n^{-2\alpha} + (\gamma_1 n^{1-\alpha})^{-1}\exp(-\kappa\gamma_1 N^{1-\alpha}/(2(1-\alpha)))$
$\alpha = 1/3$	$\gamma_1^2\log(n)n^{-2/3} + (\gamma_1 n^{2/3})^{-1}\exp(-\kappa\gamma_1 N^{1/2}/4)$
$\alpha \in (1/3,1)$	$n^{\alpha-1}\left\{\gamma_1^2 + \gamma_1^{-1}\exp(-\kappa\gamma_1 N^{1-\alpha}/(2(1-\alpha)))\right\}$
$\alpha = 1$	$\log(n)^{-1}\left\{\gamma_1^2 + \gamma_1^{-1}N^{-\gamma_1\kappa/2}\right\}$

Table 6: Bound for the MSE for $\gamma_k = \gamma_1 k^{-\alpha}$ for fixed γ_1 and N under **H**1, **H**2 and **H**3

(fixed horizon setting), as in Section 2, we may optimize the value of the step size γ_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 8 and Table 7 (we display the order in n but not the constants, which are quite involved and not overly informative).

We observe two differents bounds based on Theorem 3 and Theorem 6. Let us discuss first, the bounds obtained by the last one. It appears that, for any $\alpha \in [0,1/3)$, we can always achieved the order $n^{-2/3}$ by choosing appropriately γ_1 and N (for $\alpha = 1/3$ we have only $\log^{1/3}(n)n^{-2/3}$). The worst case is for $\alpha \in (1/3,1]$, 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 [8, Section 5.2], 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 [14]) an anytime algorithm which guarantees for all $n \geq 1$, a MSE of order $\mathcal{O}(n^{-2/3})$.

Now let us discuss the bounds based on Theorem 3. This time for any $\alpha \in [0,1/2)$, we can always achieved the order $n^{-1/2}$ by choosing appropriately γ_1 and N (for $\alpha = 1/2$ we have only $\log(n)n^{-1/2}$). For $\alpha \in (1/2,1]$, 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 [8], that constant step sizes ($\alpha = 0$) are again optimal.

We can also follow the proof of [16, Theorem 5] to establish an exponential deviation inequality for $\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)]$ given by (22)

	Optimal choice of γ_1	Optimal choice of N	Bound for the MSE
$\alpha = 0$	$n^{-1/3}$	$n^{1/3}$	$n^{-2/3}$
$\alpha \in (0, 1/2)$	$n^{\alpha-1/3}$	$n^{(1/3-\alpha)/(1-\alpha)}$	$n^{-2/3}$
$\alpha = 1/2$	$(\log(n))^{-1/3}$	$\log^{1/2}(n)$	$\log^{1/3}(n)n^{-2/3}$
$\alpha \in (1/2, 1)$	$1/(m + L)$	0	$n^{1-\alpha}$
$\alpha = 1$	$1/(m + L)$	0	$\log(n)$

Table 7: Bound for the MSE for $\gamma_k = \gamma_1 k^{-\alpha}$ for fixed n under **H**1, **H**2 and **H**3

	Optimal choice of γ_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 8: Bound for the MSE for $\gamma_k = \gamma_1 k^{-\alpha}$ for fixed n under **H**1 and **H**2

Theorem 10. *Assume* ***H****1 and* ***H****2. 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* $f : \mathbb{R}^d \to \mathbb{R}$*:*

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

Proof. Using the Markov inequality and Proposition 25, for all $\lambda > 0$, we have:

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

Then the result follows from taking $\lambda = (r\kappa^2\Gamma_{N+2,N+n+1})/(8\left\|f\right\|_{\mathrm{Lip}}^2 \upsilon_{N,n}(\gamma))$. □

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 $\alpha = 1$.

3 Numerical experiments

Consider a binary regression set-up in which the binary observations (responses) $(Y_1, \ldots, Y_p)$ are conditionally independent Bernoulli random variables with success probability $\varrho(\boldsymbol{\beta}^T X_i)$, where ϱ is the logistic function defined for $z \in \mathbb{R}$ by $\varrho(z) = \mathrm{e}^z/(1 + \mathrm{e}^z)$ 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 Σ. The posterior density distribution of $\boldsymbol{\beta}$ is up to a proportionality constant given by

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

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 [15], [9] and [10]. Recently, a very promising DA algorithm has been proposed in [20], using the Polya-Gamma distribution in the DA part. This algorithm has been shown to be uniformly ergodic for the total variation by [7, 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

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

Therefore $-\log \pi_{\boldsymbol{\beta}}(\cdot|(X_i, Y_i)_{1\leq i\leq p})$ is strongly convex **H**2 with $m = \lambda_{\max}^{-1}(\Sigma)$ and satisfies **H**1 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 Σ, respectively. To assess the proposed algorithm, we first compare the histograms given by ULA and the Pòlya-Gamma Gibbs sampling from [20]. 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 $\Sigma_{\boldsymbol{\beta}} = (\sum_{i=1}^{p}\|X_i\|^2)(dp)^{-1}\,\mathrm{I}_d$. We produce 10^7 samples from the Pòlya-Gamma sampler using the R package BayesLogit [24]. 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

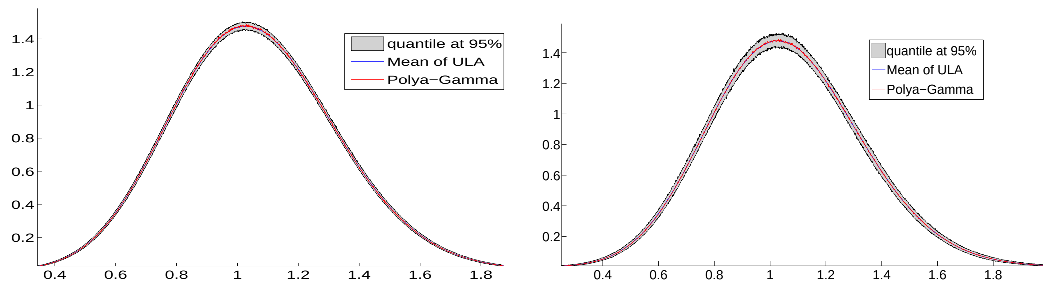

Figure 1: Empirical distribution comparison between the Polya-Gamma Gibbs Sampler and ULA. Left panel: constant step size $\gamma_k = \gamma_1$ for all $k \geq 1$; right panel: decreasing step size $\gamma_k = \gamma_1 k^{-1/2}$ for all $k \geq 1$

four real data sets, which are summarized in Table 9. 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 [22], [13] and the references therein. Also, we apply a pre-conditioned 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

Data set \ Dimensions	Observations p	Covariates d
German credit [1]	1000	25
Heart disease [2]	270	14
Australian credit[3]	690	35
Prima indian diabetes[4]	768	9

Table 9: Dimension of the data sets

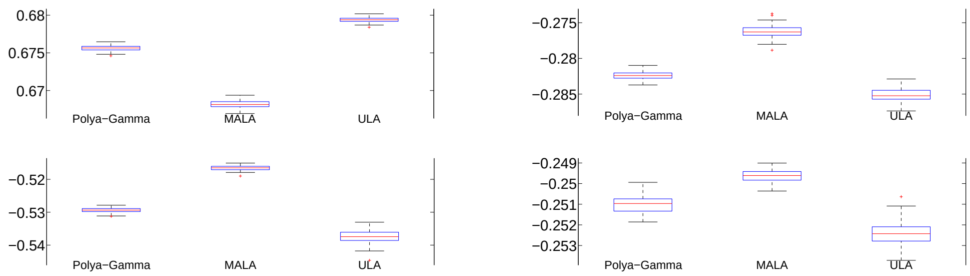

Figure 2: Upper left: German credit data set. Upper right: Australian credit data set. Lower left: Heart disease data set. Lower right: Prima Indian diabetes data set

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 of one component of $\boldsymbol{\beta}$ 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 Polya-Gamma Gibbs sampler and MALA algorithm.

4 Proofs

Let $(\mathcal{F}_t)_{t \geq 0}$ be the filtration associated with $(B_t)_{t\geq 0}$ with $\mathcal{F}_0$, the σ-field generated by $(Y_0, \overline{Y}_0)$.

4.1 Proof of Theorem 1

We preface the proof by a technical Lemma. Denote by $x^\star$ the unique minimizer of U. The generator $\mathscr{A}$ associated with $(P_t)_{t\geq 0}$ is given, for all $f \in C^2(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$, by:

$$\mathscr{A} f(x) = - \langle \nabla U(x), \nabla f(x) \rangle + \Delta f(x) \ . \tag{25}$$

[1] http://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data)
[2] http://archive.ics.uci.edu/ml/datasets/Statlog+(Heart)
[3] http://archive.ics.uci.edu/ml/datasets/Statlog+(Australian+Credit+Approval)
[4] http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes

Lemma 11. *Assume* $\boldsymbol{H}1$ *and* $\boldsymbol{H}2$.

(i) For all $t \geq 0$ *and* $x \in \mathbb{R}^d$,

$$\mathbb{E}_x\left[\|Y_t - x^\star\|^2\right] \leq \|x - x^\star\|^2 \mathrm{e}^{-2mt} + \frac{d}{m}(1 - \mathrm{e}^{-2mt}) .$$

(ii) For all $t \geq 0$ *and* $x \in \mathbb{R}^d$,

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

Proof. (i) Denote for all $x \in \mathbb{R}^d$ by $V(x) = \|x - x^\star\|^2$. Under **H**1 $\sup_{t\in[0,T]} \mathbb{E}_x[\|Y_t\|^2] < +\infty$ for all $T \geq 0$. Therefore, the process $\left(V(Y_t) - V(x) - \int_0^t \mathscr{A}V(Y_s)\mathrm{d}s\right)_{t\geq 0}$ is a $(\mathcal{F}_t)_{t\geq 0}$-martingale under $\mathbb{P}_x$. Since $\nabla U(x^\star) = 0$ and using (5), we have

$$\mathscr{A}V(x) = 2\left(-\langle \nabla U(x) - \nabla U(x^\star), x - x^\star\rangle + d\right) \leq 2\left(-mV(x) + d\right) . \tag{26}$$

Denote for all $t \geq 0$ and $x \in \mathbb{R}^d$ by $v(t, x) = P_tV(x)$. Then we have, $\partial v(t, x)/\partial t = P_t\mathscr{A}V(x)$. Using (26), we get

$$\frac{\partial v(t, x)}{\partial t} = P_t\mathscr{A}V(x) \leq -2mP_tV(x) + 2d = -2mv(t, x) + 2d ,$$

and the proof follows from the Grönwall inequality.

(ii) Denote for all $x, y \in \mathbb{R}^d$, $\tilde{V}_x(y) = \|y - x\|^2$. therefore the process $(\tilde{V}_x(Y_t) - \tilde{V}_x(x) - \int_0^t \mathscr{A}\tilde{V}_x(Y_s)\mathrm{d}s)_{t\geq 0}$, 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 we get,

$$\frac{\partial \tilde{v}(t, x)}{\partial t} = P_t\mathscr{A}\tilde{V}_x(x) . \tag{27}$$

By (5), we have for all $y \in \mathbb{R}^d$,

$$\mathscr{A}\tilde{V}_x(y) = 2\left(-\langle \nabla U(y), y - x\rangle + d\right) \leq 2\left(-m\tilde{V}_x(y) + d - \langle \nabla U(x), y - x\rangle\right) . \tag{28}$$

Using (27), this inequality and that $\tilde{V}_x$ is positive, we get

$$\frac{\partial \tilde{v}(t, x)}{\partial t} = P_t\mathscr{A}\tilde{V}_x(x) \leq 2\left(d - \int_{\mathbb{R}^d} \langle \nabla U(x), y - x\rangle P_t(x, \mathrm{d}y)\right) . \tag{29}$$

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

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

Furthermore, by **H**1 and Lemma 11-(i), we have

$$\begin{aligned}\left|\int_{\mathbb{R}^d} \langle \nabla U(x), y-x\rangle \, P_t(x,\mathrm{d}y)\right| &\leq \sqrt{t} L^2 \left\|x-x^\star\right\| \left(\int_0^t \mathbb{E}_x\left[\left\|Y_s - x^\star\right\|^2\right] \mathrm{d}s\right)^{1/2} \\ &\leq \sqrt{t} L^2 \left\|x-x^\star\right\| \left(\frac{1-\mathrm{e}^{-2mt}}{2m}\left\|x-x^\star\right\|^2 + \frac{2tm+\mathrm{e}^{-2mt}-1}{2m}(d/m)\right)^{1/2} \\ &\leq L^2 \left\|x-x^\star\right\| \left(t\left\|x-x^\star\right\| + t^{3/2} d^{1/2}\right) , \end{aligned} \tag{30}$$

where we used for the last line that by the Taylor theorem with remainder term, for all $s \geq 0$, $(1-\mathrm{e}^{-2ms})/(2m) \leq s$ and $(2ms+\mathrm{e}^{-2ms}-1)/(2m) \leq ms^2$, and the inequality $\sqrt{a+b} \leq \sqrt{a}+\sqrt{b}$. Plugging (30) in (29), and since $2\left\|x-x^\star\right\| t^{3/2} d^{1/2} \leq t\left\|x-x^\star\right\|^2 + t^2 d$, we get

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

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

$\square$

Proof of Theorem 1. (i) Consider the following SDE in $\mathbb{R}^d \times \mathbb{R}^d$:

$$\begin{cases} \mathrm{d}Y_t &= -\nabla U(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \, , \\ \mathrm{d}\tilde{Y}_t &= -\nabla U(\tilde{Y}_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \, , \end{cases} \tag{31}$$

where $(Y_0, \tilde{Y}_0)$ is some coupling between μ and ν. Since μ and ν are in $\mathcal{P}_p(\mathbb{R}^d)$ and ∇U is Lipschitz, then by [17, 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 (31),

$$\left\|Y_t - \tilde{Y}_t\right\|^p = \left\|Y_0 - \tilde{Y}_0\right\|^p - p\int_0^t \left\|Y_s - \tilde{Y}_s\right\|^{p-2} \left\langle \nabla U(Y_s) - \nabla U(\tilde{Y}_s), Y_s - \tilde{Y}_s \right\rangle \mathrm{d}s \, ,$$

which implies using (5) and Grönwall's inequality that

$$\left\|Y_t - \tilde{Y}_t\right\|^p \leq \left\|Y_0 - \tilde{Y}_0\right\|^p - mp\int_0^t \left\|Y_s - \tilde{Y}_s\right\|^p \mathrm{d}s \leq \left\|Y_0 - \tilde{Y}_0\right\|^p \mathrm{e}^{-mpt} \, .$$

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

(ii) Set $V(x) = \left\|x-x^\star\right\|^2$. By Jensen's inequality and Lemma 11-(i), for all $c > 0$ and $t > 0$, we get

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

Taking the limit as $t \to +\infty$, we get $\pi(V\wedge c) \leq d/m$. Using the monotone convergence theorem, taking the limit as $c \to +\infty$, we finally obtain (7).

$\square$

4.2 Proof of Theorem 2

We preface the proof by a Lemma.

Lemma 12. *Assume **H**1 and **H**2. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$ and $p \in \mathbb{N}$, $p \geq 1$. Then for all $\mu_0, \nu_0 \in \mathcal{P}_{2p}(\mathbb{R}^d)$ and $\ell \geq n \geq 1$,*

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

Proof. Let ζ_0 be an optimal transference plan of μ_0 and ν_0 and $(Z_k)_{k\geq 1}$ be a sequence of i.i.d. d-dimensional Gaussian random variables. We consider the processes $(X_k^1, X_k^2)_{k\geq 0}$ with initial distributions equal to ζ_0 and defined for $k \geq 0$ by

$$X_{k+1}^j = X_k^j - \gamma_{k+n} \nabla U(X_k^j) + \sqrt{2\gamma_{k+n}} Z_{k+1} \qquad j = 1,2 \; . \tag{32}$$

Using (32), we get for any $p \geq n \geq 0$. $W_{2p}^{2p}(\mu_0 Q_\gamma^{n,\ell}, \nu_0 Q_\gamma^{n,\ell}) \leq \mathbb{E}\left[\left\| X_\ell^1 - X_\ell^2 \right\|^{2p} \right]$ and (4) implies for $k \geq n-1$,

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

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

$$\left\| X_\ell^1 - X_\ell^2 \right\|^2 \leq \prod_{k=n}^{\ell} (1-\kappa\gamma_k) \left\| X_0^1 - X_0^2 \right\|^2 \; .$$

□

Proof of Theorem 2. Let $\mu \in \mathcal{P}_{2p}(\mathbb{R}^d)$ and $p \geq 1$. It is straightforward that for all $n \geq 0$, $\mu R_\gamma^n \in \mathcal{P}_{2p}(\mathbb{R}^d)$. Then, by Lemma 12 is a strict contraction in $(\mathcal{P}_{2p}(\mathbb{R}^d), W_{2p})$ and there is a unique fixed point π_γ which is the unique invariant distribution. Equation (11) follows from Lemma 12. □

4.3 Proof of Theorem 3

We preface the proof by a technical Lemma.

Lemma 13. *Assume **H**1 and **H**2. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$. Let $\zeta_0 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$, $(Y_t, \overline{Y}_t)_{t\geq 0}$ such that $(Y_0, \overline{Y}_0)$ is distributed according to ζ_0 and given by (12). Then almost surely for all $n \geq 0$ and $\epsilon > 0$,*

$$\begin{aligned} \left\| Y_{\Gamma_{n+1}} - \overline{Y}_{\Gamma_{n+1}} \right\|^2 &\leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon)\} \left\| Y_{\Gamma_n} - \overline{Y}_{\Gamma_n} \right\|^2 \\ &\quad + (2\gamma_{n+1} + (2\epsilon)^{-1}) \int_{\Gamma_n}^{\Gamma_{n+1}} \left\| \nabla U(Y_s) - \nabla U(Y_{\Gamma_n}) \right\|^2 \mathrm{d}s \; , \end{aligned} \tag{33}$$

$$\begin{aligned} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\left\| Y_{\Gamma_{n+1}} - \overline{Y}_{\Gamma_{n+1}} \right\|^2 \right] &\leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon)\} \left\| Y_{\Gamma_n} - \overline{Y}_{\Gamma_n} \right\|^2 \\ &\quad + L^2 \gamma_{n+1}^2 (1/(4\epsilon) + \gamma_{n+1}) \left(2d + L^2 \gamma_{n+1} \left\| Y_{\Gamma_n} - x^\star \right\|^2 + dL^2 \gamma_{n+1}^2/6 \right) \; . \end{aligned} \tag{34}$$

Proof. Let $n \geq 0$ and $\epsilon > 0$, and set $\Theta_n = Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}$. We first show (33).

By definition we have:

$$\|\Theta_{n+1}\|^2 = \|\Theta_n\|^2 + \left\| \int_{\Gamma_n}^{\Gamma_{n+1}} \left\{ \nabla U(Y_s) - \nabla U(\overline{Y}_{\Gamma_n}) \right\} \mathrm{d}s \right\|^2 - 2\gamma_{n+1} \left\langle \Theta_n, \nabla U(Y_{\Gamma_n}) - \nabla U(\overline{Y}_{\Gamma_n}) \right\rangle - 2 \int_{\Gamma_n}^{\Gamma_{n+1}} \left\langle \Theta_n, \left\{ \nabla U(Y_s) - \nabla U(Y_{\Gamma_n}) \right\} \right\rangle \mathrm{d}s \; . \quad (35)$$

Young's inequality and Jensen's inequality imply

$$\left\| \int_{\Gamma_n}^{\Gamma_{n+1}} \left\{ \nabla U(Y_s) - \nabla U(\overline{Y}_{\Gamma_n}) \right\} \mathrm{d}s \right\|^2 \leq 2\gamma_{n+1}^2 \left\| \nabla U(Y_{\Gamma_n}) - \nabla U(\overline{Y}_{\Gamma_n}) \right\|^2 + 2\gamma_{n+1} \int_{\Gamma_n}^{\Gamma_{n+1}} \left\| \nabla U(Y_s) - \nabla U(Y_{\Gamma_n}) \right\|^2 \mathrm{d}s \; .$$

Using (4), $\gamma_1 \leq 1/(m+L)$ and $(\gamma_k)_{k\geq 1}$ is nonincreasing, (35) becomes

$$\|\Theta_{n+1}\|^2 \leq \left\{ 1 - \gamma_{n+1}\kappa \right\} \|\Theta_n\|^2 + 2\gamma_{n+1} \int_{\Gamma_n}^{\Gamma_{n+1}} \left\| \nabla U(Y_s) - \nabla U(Y_{\Gamma_n}) \right\|^2 \mathrm{d}s - 2 \int_{\Gamma_n}^{\Gamma_{n+1}} \left\langle \Theta_n, \left\{ \nabla U(Y_s) - \nabla U(Y_{\Gamma_n}) \right\} \right\rangle \mathrm{d}s \; . \quad (36)$$

Using the inequality $|\langle a, b\rangle| \leq \epsilon \|a\|^2 + (4\epsilon)^{-1}\|b\|^2$ concludes the proof of (33).

We now prove (34). Note that (33) implies that

$$\mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\|\Theta_{n+1}\|^2 \right] \leq \left\{ 1 - \gamma_{n+1}(\kappa - 2\epsilon) \right\} \|\Theta_n\|^2 + (2\gamma_{n+1} + (2\epsilon)^{-1}) \int_{\Gamma_n}^{\Gamma_{n+1}} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\left\| \nabla U(Y_s) - \nabla U(Y_{\Gamma_n}) \right\|^2 \right] \mathrm{d}s \; . \quad (37)$$

By **H**1, the Markov property of $(Y_t)_{t\geq 0}$ and Lemma 11-(ii), we have

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

The proof is then concluded plugging this bound in (37) . □

Proof of Theorem 3. Let ζ_0 be an optimal transference plan of μ_0 and π. Let $(Y_t, \overline{Y}_t)_{t\geq 0}$ with $(Y_0, \overline{Y}_0)$ distributed according to ζ_0 and defined by (12). By definition of W_2 and since for all $t \geq 0$, π is invariant for P_t, $W_2^2(\mu_0 Q^n, \pi) \leq \mathbb{E}_{\zeta_0}\left[\left\| Y_{\Gamma_n} - \overline{Y}_{\Gamma_n} \right\|^2 \right]$. Lemma 13-(34) with $\epsilon = \kappa/4$, a straightforward induction and Lemma 11-(i) imply for all $n \geq 0$

$$\mathbb{E}_{\zeta_0}\left[\left\| Y_{\Gamma_n} - \overline{Y}_{\Gamma_n} \right\|^2 \right] \leq u_n^{(1)}(\gamma) \mathbb{E}_{\zeta_0}\left[\left\| Y_0 - \overline{Y}_0 \right\|^2 \right] + A_n(\gamma) \; , \quad (38)$$

where $(u_n^{(1)}(\gamma))_{n\geq 1}$ is given by (14), and

$$A_n(\gamma) \stackrel{\text{def}}{=} L^2 \sum_{i=1}^{n} \gamma_i^2 \left\{\kappa^{-1} + \gamma_i\right\} (2d + dL^2\gamma_i^2/6) \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2) + L^4 \sum_{i=1}^{n} \delta_i \gamma_i^3 \left\{\kappa^{-1} + \gamma_i\right\} \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2)$$

with

$$\delta_i = \mathrm{e}^{-2m\Gamma_{i-1}} \mathbb{E}_{\zeta_0}\left[\|Y_0 - x^\star\|^2\right] + (1 - \mathrm{e}^{-2m\Gamma_{i-1}})(d/m) . \tag{39}$$

Then the proof follows since Y_0 is distributed according to π and by (7), which shows that for all $i \in \{1, \cdots, n\}$, $\delta_i \leq d/m$. □

4.4 Proof of Corollary 4

Lemma 14. *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 $\ell \in \{1, \ldots, n+1\}$,*

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

Proof. Let $\ell \in \{1, \ldots, n+1\}$. Since $(\gamma_k)_{k\geq 1}$ is non-increasing,

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

□

Proof of Corollary 4. By Theorem 3, it suffices to show that $u_n^{(1)}$ and $u_n^{(2)}$, defined by (14) and (15) respectively, goes to 0 as $n \to +\infty$. Using the bound $1 + t \leq \mathrm{e}^t$ for $t \in \mathbb{R}$, and $\lim_{n\to+\infty} \Gamma_n = +\infty$, we have $\lim_{n\to+\infty} u_n^{(1)} = 0$. Now to show that $\lim_{n\to+\infty} u_n^{(2)} = 0$, a sufficient condition since $(\gamma_k)_{k\geq 0}$ is nonincreasing, is that $\lim_{n\to+\infty} \sum_{i=1}^{n} \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2)\, \gamma_i^2 = 0$. 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 14 applied with $\ell = \lfloor c\Gamma_n \rfloor$ the integer part of $c\Gamma_n$:

$$\sum_{i=1}^{n} \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2)\, \gamma_i^2 \leq 2\gamma_{\lfloor c\Gamma_n \rfloor}/\kappa + \exp\left(-\kappa\Gamma_n(1 - \Gamma_n^{-1}\Gamma_{\lfloor c\Gamma_n \rfloor})/2\right) \sum_{i=1}^{\lfloor c\Gamma_n \rfloor - 1} \gamma_i . \tag{40}$$

Since $\lim_{k\to+\infty} \gamma_k = 0$, by the Cesáro theorem, $\lim_{n\to+\infty} n^{-1}\Gamma_n = 0$. Therefore since $\lim_{n\to+\infty} \Gamma_n = +\infty$, $\lim_{n\to+\infty} (\Gamma_n)^{-1}\Gamma_{\lfloor c\Gamma_n \rfloor} = 0$, and the conclusion follows from combining in (40), this limit, $\lim_{k\to+\infty} \gamma_k = 0$, $\lim_{n\to+\infty} \Gamma_n = +\infty$ and $\sum_{i=1}^{\lfloor c\Gamma_n \rfloor - 1} \gamma_i \leq c\gamma_1\Gamma_n$. □

4.5 Proof of Corollary 5

Since by Theorem 2, for all $x \in \mathbb{R}^d$, $(\delta_x R_\gamma^n)_{n\geq 0}$ converges to π_γ as $n \to \infty$ in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$, the proof then follows from Theorem 3 and Lemma 14 applied with $\ell = 1$.

4.6 Proofs of Theorem 6

Lemma 15. *Assume **H**1, **H**2 and **H**3. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$. and $\zeta_0 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$. Let $(Y_t, \overline{Y}_t)_{t\geq 0}$ be defined by (12) such that $(Y_0, \overline{Y}_0)$ is distributed according to ζ_0. Then for all $n \geq 0$ and $\epsilon > 0$, almost surely*

$$\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left\|Y_{\Gamma_{n+1}} - \overline{Y}_{\Gamma_{n+1}}\right\|^2\right] \leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon)\}\left\|Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}\right\|^2 \tag{41}$$
$$+ \gamma_{n+1}^3\left\{d(2L^2 + \epsilon^{-1}(\tilde{L}^2/12 + \gamma_{n+1}L^4/4) + \gamma_{n+1}^2 L^4/6) + L^4(\epsilon^{-1}/3 + \gamma_{n+1})\left\|Y_{\Gamma_n} - x^\star\right\|^2\right\} .$$

Proof. Let $n \geq 0$ and $\epsilon > 0$, and set $\Theta_n = Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}$. Using Itô's formula, we have for all $s \in [\Gamma_n, \Gamma_{n+1})$,

$$\nabla U(Y_s) - \nabla U(Y_{\Gamma_n}) = \int_{\Gamma_n}^s \left\{\nabla^2 U(Y_u)\nabla U(Y_u) + \vec{\Delta}(\nabla U)(Y_u)\right\}\mathrm{d}u + \sqrt{2}\int_{\Gamma_n}^s \nabla^2 U(Y_u)\mathrm{d}B_u . \tag{42}$$

Since Θ_n is $\mathcal{F}_{\Gamma_n}$-measurable and $(\int_0^s \nabla^2 U(Y_u)\mathrm{d}B_u)_{s\in[0,\Gamma_{n+1}]}$ is a $(\mathcal{F}_s)_{s\in[0,\Gamma_{n+1}]}$-martingale under **H**1, by (42) we have:

$$\left|\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\langle \Theta_n, \nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\rangle\right]\right|$$
$$= \left|\left\langle \Theta_n, \mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\int_{\Gamma_n}^s \left\{\nabla^2 U(Y_u)\nabla U(Y_u) + \vec{\Delta}(\nabla U)(Y_u)\right\}\mathrm{d}u\right]\right\rangle\right|$$

Combining this equality and $|\langle a, b\rangle| \leq \epsilon\|a\|^2 + (4\epsilon)^{-1}\|b\|^2$ in (36) we have

$$\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\|\Theta_{n+1}\|^2\right] \leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon)\}\|\Theta_n\|^2 + 2\gamma_{n+1}\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\int_{\Gamma_n}^{\Gamma_{n+1}} \|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2\mathrm{d}s\right]$$
$$+ (2\epsilon)^{-1}\int_{\Gamma_n}^{\Gamma_{n+1}} \left\|\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\int_{\Gamma_n}^s \left\{\nabla^2 U(Y_u)\nabla U(Y_u) + (1/2)\vec{\Delta}(\nabla U)(Y_u)\right\}\mathrm{d}u\right]\right\|^2 \mathrm{d}s . \tag{43}$$

We now separately bound the two last terms of the right hand side. By **H**1, the Markov property of $(Y_t)_{t\geq 0}$ and Lemma 11-(ii), we have

$$\int_{\Gamma_n}^{\Gamma_{n+1}} \mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2\right]\mathrm{d}s$$
$$\leq L^2\left(d\gamma_{n+1}^2 + dL^2\gamma_{n+1}^4/12 + (1/2)L^2\gamma_{n+1}^3\|Y_{\Gamma_n} - x^\star\|^2\right) . \tag{44}$$

Also by (17), we get using $\nabla U(x^\star)=0$, Jensen's inequality and Fubini's theorem,

$$\begin{aligned}
A &\stackrel{\text{def}}{=} \int_{\Gamma_n}^{\Gamma_{n+1}} \left\| \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\int_{\Gamma_n}^{s} \nabla^2 U(Y_u)\nabla U(Y_u) + (1/2)\vec{\Delta}(\nabla U)(Y_u)\mathrm{d}u \right] \right\|^2 \mathrm{d}s \\
&\leq \int_{\Gamma_n}^{\Gamma_{n+1}} (s-\Gamma_n) \int_{\Gamma_n}^{s} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\left\| \left\{ \nabla^2 U(Y_u)\nabla U(Y_u) + (1/2)\vec{\Delta}(\nabla U)(Y_u) \right\} \right\|^2 \right] \mathrm{d}u\, \mathrm{d}s \\
&\leq 2\int_{\Gamma_n}^{\Gamma_{n+1}} (s-\Gamma_n) \int_{\Gamma_n}^{s} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\left\| \nabla^2 U(Y_u)\nabla U(Y_u) \right\|^2 + (1/4)\left\| \vec{\Delta}(\nabla U)(Y_u) \right\|^2 \right] \mathrm{d}u\, \mathrm{d}s \\
&\leq 2\int_{\Gamma_n}^{\Gamma_{n+1}} (s-\Gamma_n) L^4 \int_{\Gamma_n}^{s} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\left\| Y_u - x^\star \right\|^2 \right] \mathrm{d}u\, \mathrm{d}s + \gamma_{n+1}^3 d\tilde{L}^2/6 \;. \qquad (45)
\end{aligned}$$

By Lemma 11-(i), the Markov property and for all $t \geq 0$, $1-\mathrm{e}^{-t} \leq t$, we have for all $s \in [\Gamma_n, \Gamma_{n+1}]$,

$$\int_{\Gamma_n}^{s} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\left\| Y_u - x^\star \right\|^2 \right] \mathrm{d}u \leq (2m)^{-1}(1-\mathrm{e}^{-2m(s-\Gamma_n)}) \left\| Y_{\Gamma_n} - x^\star \right\|^2 + d(s-\Gamma_n)^2 \;.$$

Using this inequality in (45) and for all $t \geq 0$, $1-\mathrm{e}^{-t} \leq t$, we get

$$A \leq (2L^4\gamma_{n+1}^3/3) \left\| Y_{\Gamma_n} - x^\star \right\|^2 + L^4 d\gamma_{n+1}^4/2 + \gamma_{n+1}^3 d\tilde{L}^2/6 \;.$$

Combining this bound and (44) in (43) concludes the proof. □

Proof of Theorem 6. The proof of the Theorem is the same as the one of Theorem 3, using Lemma 15 in place of Lemma 13, and is omitted. □

4.7 Proof of Theorem 9

Our main tool is the Gaussian Poincaré inequality [5, Theorem 3.20] (see also [3, Theorem 4.1.1]) which states that if $Z=(Z_1,\ldots,Z_d)$ is a Gaussian vector with identity covariance matrix, then $\mathrm{Var}\left\{g(Z)\right\} \leq \left\|g\right\|_{\mathrm{Lip}}^2$. The Gaussian Poincaré inequality may be applied to R_γ defined by (9) 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 $2\gamma\,\mathrm{I}_d$.

Lemma 16. *Assume **H**1. Let $g:\mathbb{R}^d \to \mathbb{R}$ be a Lipschitz function. Then for all $\gamma>0$, $y \in \mathbb{R}^d$,*

$$0 \leq R_\gamma \left\{ g(\cdot) - R_\gamma g(y) \right\}^2 (y) = \int R_\gamma(y, \mathrm{d}z) \left\{ g(z) - R_\gamma g(y) \right\}^2 \leq 2\gamma \left\|g\right\|_{\mathrm{Lip}}^2 \;.$$

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

$$\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)] = \sum_{k=N}^{N+n-1} \left\{ \mathbb{E}_x^{\mathcal{G}_{k+1}}\left[\hat{\pi}_n^N(f)\right] - \mathbb{E}_x^{\mathcal{G}_k}\left[\hat{\pi}_n^N(f)\right] \right\} + \mathbb{E}_x^{\mathcal{G}_N}\left[\hat{\pi}_n^N(f)\right] - \mathbb{E}_x[\hat{\pi}_n^N(f)] \;, \quad (46)$$

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

$$\begin{aligned}
\mathrm{Var}_x\left\{\hat{\pi}_n^N(f)\right\} = \sum_{k=N}^{N+n-1} \mathbb{E}_x\left[\left(\mathbb{E}_x^{\mathcal{G}_{k+1}}\left[\hat{\pi}_n^N(f)\right] - \mathbb{E}_x^{\mathcal{G}_k}\left[\hat{\pi}_n^N(f)\right] \right)^2 \right] \\
+ \mathbb{E}_x\left[\left(\mathbb{E}_x^{\mathcal{G}_N}\left[\hat{\pi}_n^N(f)\right] - \mathbb{E}_x[\hat{\pi}_n^N(f)] \right)^2 \right] \;. \quad (47)
\end{aligned}$$

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

$$\Phi_{n,k}^N : x_k \mapsto \omega_{k,n}^N f(x_k) + R_{\gamma_{k+1}} \Phi_{n,k+1}^N(x_k) \; , \tag{48}$$

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

$$\Psi_n^N : x_N \mapsto R_{\gamma_{N+1}} \Phi_{n,N+1}^N(x_N) \; . \tag{49}$$

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

$$\Phi_{n,k+1}^N(X_{k+1}) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(X_k) = \mathbb{E}_x^{\mathcal{G}_{k+1}}\left[\hat{\pi}_n^N(f)\right] - \mathbb{E}_x^{\mathcal{G}_k}\left[\hat{\pi}_n^N(f)\right] \; , \tag{50}$$

and $\Psi_n^N(X_N) = \mathbb{E}_x^{\mathcal{G}_N}\left[\hat{\pi}_n^N(f)\right]$. With these notations, (47) may be equivalently expressed as

$$\operatorname{Var}_x\left\{\hat{\pi}_n^N(f)\right\} = \sum_{k=N}^{N+n-1} \mathbb{E}_x\left[R_{\gamma_{k+1}}\left\{\Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}}\Phi_{n,k+1}^N(X_k)\right\}^2(X_k)\right] + \operatorname{Var}_x\left\{\Psi_n^N(X_N)\right\} \; . \tag{51}$$

Now for $k = N+n, \ldots, N+1$, we will use the Gaussian Poincaré inequality (Lemma 16) 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 $k \in \{N, \ldots, N+n-1\}$ and for all $y, z \in \mathbb{R}^d$, we have

$$\left|\Phi_{n,k+1}^N(y) - \Phi_{n,k+1}^N(z)\right| = \left|\omega_{k+1,n}^N\left\{f(y) - f(z)\right\} + \sum_{i=k+2}^{N+n} \omega_{i,n}^N\left\{Q_\gamma^{k+2,i} f(y) - Q_\gamma^{k+2,i} f(z)\right\}\right| \; . \tag{52}$$

Lemma 17. *Assume* ***H****1 and* ***H****2. Let* $(\gamma_k)_{k\geq 1}$ *be a nonincreasing sequence with* $\gamma_1 \leq 2/(m+L)$*. Then for all Lipschitz functions* $f : \mathbb{R}^d \to \mathbb{R}$ *and* $\ell \geq n \geq 1$, $Q_\gamma^{n,\ell} f$ *is a Lipschitz function with*

$$\left\|Q_\gamma^{n,\ell} f\right\|_{\mathrm{Lip}} \leq \prod_{k=n}^{\ell} (1 - \kappa\gamma_k)^{1/2} \left\|f\right\|_{\mathrm{Lip}} \; .$$

Proof. Recall that for all μ, ν 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 (3):

$$\left|Q_\gamma^{n,\ell} f(y) - Q_\gamma^{n,\ell} f(z)\right| \leq \|f\|_{\mathrm{Lip}} W_1(\delta_y Q_\gamma^{n,\ell}, \delta_z Q_\gamma^{n,\ell}) \leq \|f\|_{\mathrm{Lip}} W_2(\delta_y Q_\gamma^{n,\ell}, \delta_z Q_\gamma^{n,\ell}) \; .$$

The proof then follows from Lemma 12 with $p = 1$. □

Lemma 18. *Assume* ***H****1 and* ***H****2. 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* $k \in \{N, \ldots, N+n-1\}$*,*

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

where $\Phi_{n,k+1}^N$ *is given by* (48)*.*

Proof. By (52), $\left\|\Phi^N_{n,k}\right\|_{\mathrm{Lip}} \leq \sum_{i=k+1}^{N+n} \omega^N_{i,n} \left\|Q^{k+2,i}_\gamma f\right\|_{\mathrm{Lip}}$. Using Lemma 17, the bound $(1-t)^{1/2} \leq 1-t/2$ for $t \in [0,1]$ and the definition of $\omega^N_{i,n}$ given by (22), we have

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

Finally, the proof follows from Lemma 16. $\square$

Also to control the last term in right hand side of (51), we need to control the variance of $\Psi^N_n(X_N)$ under $\delta_x Q^N_\gamma$. But similarly to the sequence of functions $\Phi^N_{n,k}$, Ψ^N_n is Lipschitz by Lemma 17 since for all $y, z \in \mathbb{R}^d$, we have

$$
\left|\Psi^N_n(y) - \Psi^N_n(z)\right| = \left| \sum_{i=N+1}^{N+n} \omega^N_{i,n} \left\{ Q^{N+1,i}_\gamma f(y) - Q^{N+1,i}_\gamma f(z) \right\} \right| . \tag{53}
$$

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

Lemma 19. *Assume **H**1 and **H**2. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Let $g : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz function. Then for all $n, p \geq 1$, $n \leq p$ and $y \in \mathbb{R}^d$*

$$
0 \leq \int_{\mathbb{R}^d} Q^{n,p}_\gamma(y, \mathrm{d}z) \left\{ g^2(z) - Q^{n,p}_\gamma g(y) \right\}^2 \leq 2\kappa^{-1} \|g\|^2_{\mathrm{Lip}} ,
$$

where $Q^{n,p}_\gamma$ is given by (10).

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

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

Lemma 16 implies $\mathrm{Var}^{\mathcal{G}_n}_y\{g(X_p)\} \leq 2\sum_{k=n+1}^{p} \gamma_k \left\|Q^{k+1,p}_\gamma g\right\|^2_{\mathrm{Lip}}$. The proof follows from Lemma 17 and Lemma 14, using the bound $(1-t)^{1/2} \leq 1-t/2$ for $t \in [0,1]$. $\square$

Corollary 20. *Assume **H**1 and **H**2. Let $(\gamma_k)_{k\geq 1}$ be nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Then for all Lipschitz function f,*

$$
\mathrm{Var}_x \left\{ \Psi^N_n(X_N) \right\} \leq 8\kappa^{-3} \|f\|^2_{\mathrm{Lip}} \Gamma^{-2}_{N+2,N+n+1} ,
$$

where Ψ^N_n is given by (49).

Proof. By (53), Ψ_n^N is Lipschitz function with

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

Using Lemma 17, 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 (22), we have

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

The proof follows from Lemma 19. □

Proof. Plugging the bounds given by Lemma 18 and Corollary 20 in (51), we have

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

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

4.8 Proof of Theorem 10

Let $N \geq 0$, $n \geq 1$, $x \in \mathbb{R}^d$ and f be a Lipschitz function. To prove Theorem 10, 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 (46)

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

Now using (50) with the sequence of functions $(\Phi_{n,k}^N)$ and Ψ_n^N given by (48) and (49), respectively, we have by the Markov property

$$\begin{aligned}
&\mathbb{E}_x\left[\mathrm{e}^{\lambda\{\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)]\}}\right] \\
&= \mathbb{E}_x\left[\mathrm{e}^{\lambda\{\Psi_n^N(X_n) - \mathbb{E}_x\left[\Psi_n^N(X_n)\right]\}} \prod_{k=N}^{N+n-1} R_{\gamma_{k+1}}\left\{\exp\left(\lambda\{\Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}}\Phi_{n,k+1}^N(X_k)\}\right)\right\}(X_k)\right]
\end{aligned} \tag{54}$$

where R_γ is given by (9) 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 $R_{\gamma_{k+1}}\{\exp(\lambda$

$\{\Phi^N_{n,k+1}(\cdot) - R_{\gamma_{k+1}}\Phi^N_{n,k+1}(X_k)\})\}(X_k)$ w.r.t. X_k, for all $k \in \{N+1, \ldots, N+n\}$. 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 $2\gamma \, \mathrm{I}_d$. The log-Sobolev inequality provides a bound for the Laplace transform of Lipschitz function $g(Z) - R_\gamma g(x)$ where Z is distributed under $R_\gamma(x, \cdot)$.

Lemma 21 ([5, Theorem 5.5])**.** *Assume* ***H1****. Then for all Lipschitz function g, $\gamma > 0$, $x \in \mathbb{R}^d$ and $\lambda > 0$,*

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

where R_γ is given by (9).

We deduced from this lemma, (52) and Lemma 17, an equivalent of Lemma 18 for the Laplace transform of $\Phi^N_{n,k+1}$ under $\delta_y R_{\gamma_{k+1}}$ for $k \in \{N+1, \ldots, N+n\}$ and all $y \in \mathbb{R}^d$.

Corollary 22. *Assume* ***H1*** *and* ***H2****. 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 $k \in \{N+1, \ldots, N+n\}$, $y \in \mathbb{R}^d$ and $\lambda > 0$,*

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

where $\Phi^N_{n,k}$ is given by (48).

It remains to control the Laplace transform of Ψ^N_n under $\delta_x Q^N_\gamma$, where $\delta_x Q^N_\gamma$ is defined by (10). For this, using again that by (53) and Lemma 17, Ψ^N_n is a Lipschitz function, we iterate Lemma 21 to get bounds on the Laplace transform of Lipschitz function g under $Q^{n,\ell}_\gamma(y, \cdot)$ for all $y \in \mathbb{R}^d$ and $n, \ell \geq 1$, since for all $n, \ell \geq 1$, $Q^{n,\ell}_\gamma g$ is a Lipschitz function by Lemma 17.

Lemma 23. *Assume* ***H1*** *and* ***H2****. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Let $g : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz function, then for all $n, p \geq 1$, $n \leq p$, $y \in \mathbb{R}^d$ and $\lambda > 0$:*

$$Q^{n,p}_\gamma\left\{\exp\left(\lambda\{g(\cdot) - Q^{n,p}_\gamma g(y)\}\right)\right\}(y) \leq \exp\left(\kappa^{-1}\lambda^2 \|g\|^2_{\mathrm{Lip}}\right) , \tag{55}$$

where $Q^\gamma_{n,p}$ is given by (10).

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

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

By the Gaussian log-Sobolev inequality Lemma 21, we get

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

The proof follows from Lemma 17 and Lemma 14, using the bound $(1-t)^{1/2} \leq 1 - t/2$ for $t \in [0,1]$.

□

Combining this result with (53) and Lemma 17, we get an analogue of Corollary 20 for the Laplace transform of Ψ_n^N:

Corollary 24. *Assume* ***H****1 and* ***H****2. 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 $\lambda > 0$,*

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

where Ψ_n^N is given by (49).

The Laplace transform of $\hat{\pi}_n^N(f)$ can be explicitly bounded using Corollary 22 and Corollary 24 in (54).

Proposition 25. *Assume* ***H****1 and* ***H****2. 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 $f : \mathbb{R}^d \to \mathbb{R}$ and $\lambda > 0$:*

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

where $u_{N,n}^{(3)}(\gamma)$ is given by and (24).

Acknowledgment

The work of A.D. and E.M. is supported by the Agence Nationale de la Recherche, under grant ANR-14-CE23-0012 (COSMOS).

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, P. Cattiaux, and A. Guillin. Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.*, 254(3):727–759, 2008.

[3] 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.

[4] François Bolley, Ivan Gentil, and Arnaud Guillin. Convergence to equilibrium in Wasserstein distance for Fokker-Planck equations. *J. Funct. Anal.*, 263(8):2430–2457, 2012.

[5] 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.

[6] P. Cattiaux and A. Guillin. Trends to equilibrium in total variation distance. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(1):117–145, 2009.

[7] 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.

[8] A. Durmus and Moulines E. Supplement supplement to "non-asymptotic convergence analysis for the unadjusted langevin algorithm", 2015.

[9] 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.

[10] R. B. Gramacy and N. G. Polson. Simulation-based regularized logistic regression. *Bayesian Anal.*, 7(3):567–590, 09 2012.

[11] U. Grenander. Tutorial in pattern theory. Division of Applied Mathematics, Brown University, Providence.

[12] 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.

[13] T. E. Hanson, A J. Branscum, and W. O. Johnson. Informative g-priors for logistic regression. *Bayesian Anal.*, 9(3):597–611, 2014.

[14] 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.

[15] C. C. Holmes and L. Held. Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Anal.*, 1(1):145–168, 03 2006.

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

[17] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer New York, 1991.

[18] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Applied Optimization. Springer, 2004.

[19] G. Parisi. Correlation functions and computer simulations. *Nuclear Physics B*, 180:378–384, 1981.

[20] 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.

[21] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

[22] D. Sabanés Bové and L. Held. Hyper-g priors for generalized linear models. *Bayesian Anal.*, 6(3):387–410, 2011.

[23] C. Villani. *Optimal transport : old and new*. Grundlehren der mathematischen Wissenschaften. Springer, Berlin, 2009.

[24] 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.