arXiv:1605.01559v4 [math.ST] 15 Jul 2018

High-dimensional Bayesian inference via the Unadjusted Langevin Algorithm

Alain Durmus[1] and Éric Moulines[2]

[1]CMLA - École normale supérieure Paris-Saclay, CNRS, Université Paris-Saclay, 94235 Cachan, France.
[2]Centre de Mathématiques Appliquées, UMR 7641, Ecole Polytechnique, France.

July 17, 2018

Keywords: total variation distance, Langevin diffusion, 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 high-dimensional probability distribution π having a density w.r.t. the Lebesgue measure on $\mathbb{R}^d$, known up to a normalization constant $x \mapsto \pi(x) = \mathrm{e}^{-U(x)} / \int_{\mathbb{R}^d} \mathrm{e}^{-U(y)} \mathrm{d}y$. Such problem naturally occurs for example in Bayesian inference and machine learning. Under the assumption that U is continuously differentiable, ∇U is globally Lipschitz and U is strongly convex, we obtain non-asymptotic bounds for the convergence to stationarity in Wasserstein distance of order 2 and total variation distance 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 these bounds is explicit. The convergence of an appropriately weighted empirical measure is also investigated and bounds for the mean square error and exponential deviation inequality are reported for functions which are measurable and bounded. An illustration to Bayesian inference for binary regression is presented to support our claims.

1 Introduction

Interest for Bayesian inference methods for high-dimensional models has recently received renewed attention often motivated by machine learning applications. Rather

[1]Email: alain.durmus@cmla.ens-cachan.fr
[2]eric.moulines@polytechnique.edu

than obtaining a point estimate, Bayesian methods attempt to sample the full posterior distribution over the parameters and possibly latent variables which provides a way to assert uncertainty in the model and prevents from overfitting [33], [42].

The problem can be formulated as follows. We aim at sampling a posterior distribution π 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. The Langevin stochastic differential equation associated with π is defined by:

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

where $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, satisfying the usual conditions. Under mild technical conditions, the Langevin diffusion admits π as its unique invariant distribution.

We study the sampling method based on the Euler-Maruyama discretization of (1). This scheme defines the (possibly) non-homogeneous, discrete-time Markov chain $(X_k)_{k\geq 0}$ given by

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

where $(Z_k)_{k\geq 1}$ is an i.i.d. sequence of d-dimensional standard Gaussian random variables and $(\gamma_k)_{k\geq 1}$ is a sequence of step sizes, which can either be held constant or be chosen to decrease to 0. This algorithm has been first proposed by [16] and [35] for molecular dynamics applications. Then it has been popularized in machine learning by [20], [21] and computational statistics by [33] and [37]. Following [37], in the sequel this method will be referred to as the *unadjusted* Langevin algorithm (ULA). When the step sizes are held constant, under appropriate conditions on U, the homogeneous Markov chain $(X_k)_{k\geq 0}$ has a unique stationary distribution π_γ, which in most cases differs from the distribution π. It has been proposed in [38] and [37] to use a Metropolis-Hastings step at each iteration to enforce reversibility w.r.t. π. This algorithm is referred to as the Metropolis adjusted Langevin algorithm (MALA).

The ULA algorithm has already been studied in depth for constant step sizes in [40], [37] and [31]. In particular, [40, Theorem 4] gives an asymptotic expansion for the weak error between π and π_γ. When $\lim_{k\to+\infty} \gamma_k = 0$ and $\sum_{k=1}^{\infty} \gamma_k = \infty$, weak convergence of the weighted empirical distribution of the ULA algorithm has been established in [27], [28] and [29].

Contrary to these reported works, we focus in this paper on non-asymptotic results. These questions have been addressed previously in [10] and [12]. [10] establishes explicit bounds on the total variation distance between the distribution of the n-th iterate of the Markov chain defined in (2) and the target distribution π for fixed step size and a strongly convex potential U. It is shown that if the initial distribution is an appropriately chosen Gaussian or if a warm-start is used, the number of iterations required to get a sample ϵ-close to π in total variation is of order $\mathcal{O}(d^3\varepsilon^{-2})$ and $\mathcal{O}(d\varepsilon^{-2})$ respectively. The results of [10] were later sharpened in [12], using different technical arguments. In particular, [12] shows that starting from a minimizer of U, the number of iterations to get a sample ε-close from π in total variation is of order $\mathcal{O}(d\varepsilon^{-2})$ and that therefore a warm start is not necessary. [12] also extends the results of [10] to non-convex potentials and non-increasing sequences of step sizes. It also establish some bounds between π and π_γ in V-norm which scale as $\gamma^{1/2}$ as $\gamma \to 0$.

In this work, we focus on the case where U is strongly convex. Compared to [10] and [12], our contributions are as follows.

- We give explicit bounds between the distribution of the n-th iterate of the Markov chain defined in (2) and the target distribution π in Wasserstein and total variation distance for fixed and non-increasing step sizes. The obtained bounds improve those reported in [10] and [12] for the total variation distance.

- For fixed step sizes ($\gamma_k = \gamma$ for all $k \geq 0$), we analyse both fixed horizon (the total computational budget is fixed and the step size is chosen to minimize the upper bound on the Wasserstein or total variation distance) and fixed precision (for a fixed target precision, the number of iterations and the step size are optimized simultaneously to meet this constraint). For a fixed precision $\varepsilon > 0$, we show that the number of iterations $n \geq 0$, for ULA to get a sample ε-close to π in Wasserstein distance / total variation of order $\mathcal{O}(d\varepsilon^{-2})$ or $\mathcal{O}(d\varepsilon^{-1})$ (up to logarithmic terms), depending on the smoothness of U. We show that our result is optimal (up to logarithmic factors again) for d-dimensional Gaussian distribution. We show in the finite horizon setting that if the total number of iterations is n, we may choose the step size $\gamma = \gamma_n > 0$ such that the Wasserstein distance between the distribution of the n-th iterate and π is bounded by $\mathcal{O}(n^{-1/2})$ and $\mathcal{O}(n^{-1})$ depending on the smoothness of U.

- When $\lim_{k\to+\infty} \gamma_k = 0$ and $\sum_{k=1}^{\infty} \gamma_k = \infty$, we show that the marginal distribution of the non-homogeneous Markov chain $(X_k)_{k\geq 0}$ converges to the target distribution π and provide explicit convergence bounds in the case $\gamma_k = \gamma_1 k^{-\alpha}$, $\alpha \in (0,1]$. The optimal rate of convergence derived from our bounds for the Wasserstein/total variation distance is obtained for $\alpha = 1$ with $\gamma_1 > 0$ large enough. The convergence rates we report, improve those given in [12].

- Quantitative estimates between π and π_γ are obtained in Wasserstein and total variation distance. The bound on the total variation distance between π and π_γ we derive improves the one reported in [12]. In particular, when U is smooth enough, $\|\pi - \pi_\gamma\|_{\mathrm{TV}}$ scales as γ as $\gamma \to 0$.

- Convergence of weighted empirical measure is studied through bounds on the mean square error and exponential deviation of an estimator of $\int_{\mathbb{R}^d} f(x)\mathrm{d}\pi(x)$, for functions $f : \mathbb{R}^d \to \mathbb{R}$ which are either Lipschitz or bounded and measurable. When f is Lipschitz, U is smooth enough and in the any-time setting, the optimal rate of convergence for the MSE, using non-increasing sequences $\gamma_k = \gamma_1/k^\alpha$, is obtained for $\alpha = 1/3$ (which coincides with the rate used in [27] to derive a central limit theorem). If the step size is held constant, we get that the number of iterations for the mean square error to be smaller than $\varepsilon > 0$ is of order $\mathcal{O}(d\varepsilon^{-4})$ or $\mathcal{O}(d\varepsilon^{-3})$, depending on the smoothness of U. The case where f is bounded and measurable is an important result in Bayesian statistics to estimate credibility regions. For that purpose, we study the convergence of the Euler-Maruyama discretization towards its stationary distribution in total variation using a discrete time version of reflection coupling introduced in [5]. For fixed step size, the conclusion on the sufficient number of iterations for the mean square error to be smaller than $\varepsilon > 0$ is the same (up to logarithmic terms) as for Lipschitz functions.

In this paper, a special attention is paid to the dependency of the obtained bounds on the dimension of the state space, since we are particularly interested in the applications of this method to sampling in high-dimension.

The paper is organized as follows. In Section 2, we study the convergence in the Wasserstein distance of order 2 of the Euler discretization for constant and decreasing step sizes. In Section 3, we give non asymptotic bounds in total variation distance between the Euler discretization and π. This study is completed in Section 4 by non-asymptotic bounds of convergence of the weighted empirical measure applied to functions which are either Lipschitz or bounded and measurable. Our claims are supported in a Bayesian inference for a binary regression model in Section 5. Finally in Section 6, some results of independent interest, used in the proofs, on functional autoregressive models are gathered. Most proofs and derivations are postponed and carried out in Appendices and a supplementary paper [11].

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

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

By [41, Theorem 4.1], for all μ, ν probability measures 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 $\mu \in \mathcal{P}_p(\mathbb{R}^d)$, $\int_{\mathbb{R}^d} \|x\|^p \, \mathrm{d}\mu(x) < +\infty$. By [41, 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 [41, Theorem 5.9]) implies that for all μ, ν probability measures on $\mathbb{R}^d$,

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

Denote by $\mathbb{F}_b(\mathbb{R}^d)$ the set of all bounded Borel measurable functions on $\mathbb{R}^d$. For $f \in \mathbb{F}_b(\mathbb{R}^d)$ set $\mathrm{osc}(f) = \sup_{x,y\in\mathbb{R}^d} |f(x) - f(y)|$. For two probability measures μ and ν on $\mathbb{R}^d$, the total variation distance distance between μ and ν is defined by $\|\mu - \nu\|_{\mathrm{TV}} = \sup_{\mathtt{A}\in\mathcal{B}(\mathbb{R}^d)} |\mu(\mathtt{A}) - \nu(\mathtt{A})|$. By the Monge-Kantorovich theorem the total variation distance between μ and ν can be written on the form:

$$\|\mu - \nu\|_{\mathrm{TV}} = \inf_{\zeta\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathbb{1}_{\mathsf{D}^c}(x, y)\mathrm{d}\zeta(x, y) ,$$

where $\mathsf{D} = \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \, | x = y\}$. 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. For a subset $\mathtt{A} \subset \mathbb{R}^d$, denote by $\mathtt{A}^c$ the complementary of $\mathtt{A}$. Let $n \in \mathbb{N}^*$ and M be a $n \times n$-matrix, then denote by M^{T} the transpose of M and $\|M\|$ the operator norm associated with M defined by $\|M\| = \sup_{\|x\|=1} \|Mx\|$. Define the Frobenius norm associated with M by $\|M\|_{\mathrm{F}}^2 = \mathrm{Tr}(M^T M)$. 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 vector 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 $\vec{\Delta} F(x)$ equals to $\sum_{j=1}^d (\partial^2 F_i/\partial x_j^2)(x)$. In the sequel, we take the convention that $\sum_p^n = 0$ and $\prod_p^n = 1$ for $n, p \in \mathbb{N}$, $n < p$.

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 gradient Lipschitz: 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, for all $x \in \mathbb{R}^d$ by [25, Theorem 2.5, Theorem 2.9 Chapter 5] there exists a unique strong solution $(Y_t)_{t\geq0}$ to (1) with $Y_0 = x$. Denote by $(P_t)_{t\geq0}$ the semi-group associated with (1). It is well-known that π is its (unique) invariant probability. To get geometric convergence of $(P_t)_{t\geq0}$ to π in Wasserstein distance of order 2, we make the following additional assumption on the potential U.

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, [34, Theorem 2.1.8] shows that U has a unique minimizer $x^\star \in \mathbb{R}^d$. We briefly summarize some background material on the stability and the convergence in W_2 of the overdamped Langevin diffusion under **H**1 and **H**2. Most of the statements in Proposition 1 are known and are recalled here for ease of references; see e.g. [6].

Proposition 1. *Assume **H**1 and **H**2.*

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

$$\int_{\mathbb{R}^d} \|y - x^\star\|^2 P_t(x, \mathrm{d}y) \leq \|x - x^\star\|^2 \mathrm{e}^{-2mt} + (d/m)(1 - \mathrm{e}^{-2mt}) \ .$$

(ii) The stationary distribution π satisfies $\int_{\mathbb{R}^d} \|x - x^\star\|^2 \pi(\mathrm{d}x) \leq d/m$.

(iii) For any $x, y \in \mathbb{R}^d$ and $t > 0$, $W_2(\delta_x P_t, \delta_y P_t) \leq \mathrm{e}^{-mt} \|x - y\|$.

(iv) For any $x \in \mathbb{R}^d$ and $t > 0$, $W_2(\delta_x P_t, \pi) \leq \mathrm{e}^{-mt} \left\{\|x - x^\star\| + (d/m)^{1/2}\right\}$.

Proof. The proof is given in the supplementary document Appendix A.1. □

Note that the convergence rate in Proposition 1-(iv) does not depend on the dimension. Let $(\gamma_k)_{k\geq1}$ be a sequence of positive and non-increasing step sizes and for $n, \ell \in \mathbb{N}$, denote by

$$\Gamma_{n,\ell} = \sum_{k=n}^{\ell} \gamma_k \ , \qquad \Gamma_n = \Gamma_{1,n} \ . \tag{3}$$

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

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

The process $(X_k)_{k\geq0}$ given in (2) is an inhomogeneous Markov chain with respect to the family of Markov kernels $(R_{\gamma_k})_{k\geq1}$. For $\ell, n \in \mathbb{N}^*$, $\ell \geq n$, define

$$Q_\gamma^{n,\ell} = R_{\gamma_n} \cdots R_{\gamma_\ell} \ , \qquad Q_\gamma^n = Q_\gamma^{1,n} \tag{5}$$

with the convention that for $n, \ell \in \mathbb{N}$, $\ell < n$, $Q_\gamma^{n,\ell}$ is the identity operator.

We first derive a Foster-Lyapunov drift condition for $Q_\gamma^{n,\ell}$, $\ell, n \in \mathbb{N}^*$, $\ell \geq n$. Set

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

where m and L are defined in **H**1

Proposition 2. *Assume **H**1 and **H**2.*

(i) *Let $(\gamma_k)_{k\geq 1}$ be a non-increasing sequence with $\gamma_1 \leq 2/(m+L)$. Let $x^\star$ be the unique minimizer of U. Then for all $x \in \mathbb{R}^d$ and $n, \ell \in \mathbb{N}^*$,*

$$\int_{\mathbb{R}^d} \|y - x^\star\|^2 \, Q_\gamma^{n,\ell}(x, \mathrm{d}y) \leq \varrho_{n,\ell}(x) ,$$

where $\varrho_{n,\ell}(x)$ is given by

$$\varrho_{n,\ell}(x) = \prod_{k=n}^{\ell}(1-\kappa\gamma_k)\, \|x - x^\star\|^2 + 2d\kappa^{-1}\left\{1 - \kappa^{-1}\prod_{i=n}^{\ell}(1-\kappa\gamma_i)\right\} , \quad (7)$$

(ii) *For any $\gamma \in (0, 2/(m+L)]$, R_γ has a unique stationary distribution π_γ and*

$$\int_{\mathbb{R}^d} \|x - x^\star\|^2 \, \pi_\gamma(\mathrm{d}x) \leq 2d\kappa^{-1} .$$

Proof. The proof is postponed to Appendix A.2. □

We now proceed to establish that Q_γ^n is a strict contraction in W_2 for any $n \geq 1$. This result implies the geometric convergence of the sequence $(\delta_x R_\gamma^n)_{n\geq 1}$ to π_γ in W_2 for all $x \in \mathbb{R}^d$. Note that the convergence rate again does not depend on the dimension.

Proposition 3. *Assume **H**1 and **H**2. Then,*

(i) *Let $(\gamma_k)_{k\geq 1}$ be a non-increasing sequence with $\gamma_1 \leq 2/(m+L)$. For all $x, y \in \mathbb{R}^d$ and $\ell \geq n \geq 1$,*

$$W_2(\delta_x Q_\gamma^{n,\ell}, \delta_y Q_\gamma^{n,\ell}) \leq \left\{\prod_{k=n}^{\ell}(1-\kappa\gamma_k)\right\}^{1/2} \|x - y\| .$$

(ii) *For any $\gamma \in (0, 2/(m+L))$, for all $x \in \mathbb{R}^d$ and $n \geq 1$,*

$$W_2(\delta_x R_\gamma^n, \pi_\gamma) \leq (1-\kappa\gamma)^{n/2}\left\{\|x - x^\star\|^2 + 2\kappa^{-1}d\right\}^{1/2} .$$

Proof. The proof is postponed to Appendix A.3. □

Corollary 4. *Assume **H**1 and **H**2. Let $(\gamma_k)_{k\geq 1}$ be a non-increasing 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 $\|Q_\gamma^{n,\ell} f\|_{\mathrm{Lip}} \leq \prod_{k=n}^{\ell}(1-\kappa\gamma_k)^{1/2}\|f\|_{\mathrm{Lip}}$.*

Proof. The proof follows from Proposition 3-(i) using

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

□

We now proceed to establish explicit bounds for $W_2(\delta_x Q_\gamma^n, \pi)$, with $x \in \mathbb{R}^d$.

Theorem 5. *Assume **H** 1 and **H** 2. Let $(\gamma_k)_{k\geq 1}$ be a non-increasing sequence with $\gamma_1 \leq 1/(m+L)$. Then for all $x \in \mathbb{R}^d$ and $n \geq 1$,*

$$W_2^2(\delta_x Q_\gamma^n, \pi) \leq u_n^{(1)}(\gamma) \left\{ \|x - x^\star\|^2 + d/m \right\} + u_n^{(2)}(\gamma) ,$$

where

$$u_n^{(1)}(\gamma) = 2 \prod_{k=1}^{n} (1 - \kappa\gamma_k/2) \tag{8}$$

κ is defined in (A) and

$$u_n^{(2)}(\gamma) \quad = \quad L^2 d \sum_{i=1}^{n} \left[\gamma_i^2 \left\{ \kappa^{-1} + \gamma_i \right\} \left\{ 2 + \frac{L^2 \gamma_i}{m} + \frac{L^2 \gamma_i^2}{6} \right\} \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2) \right] \quad . \tag{9}$$

Proof. The proof is postponed to Appendix A.4. □

Corollary 6. *Assume **H** 1 and **H** 2. Let $(\gamma_k)_{k\geq 1}$ be a non-increasing 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 $x \in \mathbb{R}^d$, $\lim_{n\to\infty} W_2(\delta_x Q_\gamma^n, \pi) = 0$.*

Proof. The proof is postponed to Appendix A.5. □

In the case of constant step sizes $\gamma_k = \gamma$ for all $k \geq 1$, we can deduce from Theorem 5, a 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 $\gamma \leq 1/(m+L)$. 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. Since by Proposition 3, 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 5 and [12, Lemma 23] applied with $\ell = 1$. □

We can improve the bound provided by Theorem 5 under additional regularity assumptions on the potential U.

H 3. *The potential U is three times continuously differentiable and there exists $\tilde{L}$ such that for all $x, y \in \mathbb{R}^d$, $\left\|\nabla^2 U(x) - \nabla^2 U(y)\right\| \leq \tilde{L} \|x - y\|$.*

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

$$\left\|\nabla^2 U(x) y\right\| \leq L \|y\| \ , \ \left\|\vec{\Delta}(\nabla U)(x)\right\|^2 \leq d^2 \tilde{L}^2 . \tag{10}$$

Theorem 8. *Assume* **H**1*,* **H**2 *and* **H**3*. Let* $(\gamma_k)_{k\geq 1}$ *be a non-increasing sequence with* $\gamma_1 \leq 1/(m+L)$*. Then for all* $x \in \mathbb{R}^d$ *and* $n \geq 1$*,*

$$W_2^2(\delta_x Q_\gamma^n, \pi) \leq u_n^{(1)}(\gamma)\left\{\|x - x^\star\|^2 + d/m\right\} + u_n^{(3)}(\gamma) ,$$

where $u_n^{(1)}$ *is given by* (8)*,* κ *in* (A) *and*

$$u_n^{(3)}(\gamma) = \sum_{i=1}^n \left[d\gamma_i^3 \left\{ 2L^2 + \gamma_i L^4 \left(\frac{\gamma_i}{6} + m^{-1}\right) + \kappa^{-1}\left(\frac{4d\tilde{L}^2}{3} + \gamma_i L^4 + \frac{4L^4}{3m}\right)\right\} \times \prod_{k=i+1}^n \left(1 - \frac{\kappa\gamma_k}{2}\right)\right] . \quad (11)$$

Proof. The proof is postponed to Appendix A.6. □

If $\gamma_k = \gamma$ for all $k \geq 1$, we can deduce from Theorem 8, a sharper bound between π and the stationary distribution π_γ of R_γ.

Corollary 9. *Assume* **H**1*,* **H**2 *and* **H**3*. Let* $(\gamma_k)_{k\geq 1}$ *be a constant sequence* $\gamma_k = \gamma$ *for all* $k \geq 1$ *with* $\gamma \leq 1/(m+L)$*. Then*

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

Proof. The proof follows the same line as the proof of Corollary 7 and is omitted. □

Using Proposition 3-(ii) and Corollary 6 or Corollary 9, given $\varepsilon > 0$, we determine the number of iterations n_ε and an associated step size γ_ε to ensure that $W_2(\delta_{x^\star} R_{\gamma_\varepsilon}^n, \pi) \leq \varepsilon$ for all $n \geq n_\varepsilon$. The precise expression of n_ε directly computed using Theorem 5 and Theorem 8 are also given in [11, Section 1.1-Section 2.1]. Dependencies in dimension d and precision ε of n_ε are reported in Table 1. Under **H**1 and **H**2, the complexity matches the results reported in [12] for the total variation distance. Under **H**3, the dependency in the precision ε can be improved. If $\tilde{L} = 0$ (for example for non-degenerate d-dimensional Gaussian distributions), then the dependency in d given by Theorem 8 is of order $\mathcal{O}(d^{1/2}\log(d))$.

In a recent work [9] (based on a previous version of this paper), an improvement of the proof of Theorem 5 has been proposed for constant step size. Whereas the constants are sharper, dependency in dimension d and precision $\varepsilon > 0$ is the same (first line of Table 1).

Under **H**1 and **H**2, by Theorem 5, in the finite horizon setting, then for any $n \geq 1$, we may choose a step size $\gamma = \gamma_n > 0$ such that $W_2^2(\delta_{x^\star} R_{\gamma_n}^n, \pi) = \mathcal{O}(\log(n)/n)$ and $W_2^2(\delta_{x^\star} R_{\gamma_n}^n, \pi) \leq \mathcal{O}(\log(n)/n)^2$ if **H**3 holds by Theorem 8. The precise statement of these results are given by [11, Corollary S2-Corollary S5] in [11, Section 1.3-Section 2.3].

For simplicity, 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 $\alpha \in (0,1)$. Then for $n \geq 1$, $u_n^{(1)} = \mathcal{O}(\mathrm{e}^{-\kappa\Gamma_n/2})$, $u_n^{(2)} = d\mathcal{O}(n^{-\alpha})$ and

Parameter	d, ε
Theorem 5 and Proposition 3-(ii)	$\mathcal{O}(d\log(d)\varepsilon^{-2}\left\|\log(\varepsilon)\right\|)$
Theorem 8 and Proposition 3-(ii)	$\mathcal{O}(d\log(d)\varepsilon^{-1}\left\|\log(\varepsilon)\right\|)$

Table 1: Dependencies of the number of iterations n_ε to get $W_2(\delta_{x^\star} R_{\gamma_\varepsilon}^{n_\varepsilon}, \pi) \leq \varepsilon$

$u_n^{(3)} = d^2\mathcal{O}(n^{-2\alpha})$ (see [11, Section 1.2-Section 2.2] for details). For $\gamma_k = \gamma_1/k$, we need to extend Theorem 5 and Theorem 8 to non-increasing sequence such that there exists $n_1 \geq 1$ such that $\gamma_{n_1} < 1/(m+L)$. It is done in [11, Theorem S11 in Section 3]. Using this result in [11, Section 3.1], we get that under **H**1 and **H**2, that $W_2^2(\delta_{x^\star} Q_\gamma^n, \pi) = \mathcal{O}(n^{-1})$ for $\gamma_1 > 2\kappa^{-1}$. If in addition **H**3 holds, we have $W_2^2(\delta_{x^\star} Q_\gamma^n, \pi) = \mathcal{O}(n^{-1})$ for $\gamma_1 > 4\kappa^{-1}$. However, note that the constants are exponential in γ_1. The conclusions of this discussion are summarized in Table 2.

Note that these rates are explicit compared to those reported in [12, Proposition 3]. In addition, two regimes can be observed as in stochastic approximation in the case $\alpha = 1$.

	$\alpha \in (0,1)$	$\alpha = 1$
Theorem 5	$d\,\mathcal{O}(n^{-\alpha})$	$d\,\mathcal{O}(n^{-1})$ for $\gamma_1 > 2\kappa^{-1}$ see [11, Section 3.1]
Theorem 8	$d^2\,\mathcal{O}(n^{-2\alpha})$	$d^2\,\mathcal{O}(n^{-2})$ for $\gamma_1 > 4\kappa^{-1}$ see [11, Section 3.1]

Table 2: Order of convergence of $W_2^2(\delta_{x^\star} Q_\gamma^n, \pi)$ for $\gamma_k = \gamma_1/k^\alpha$

Details and further discussions are included in [11, Section 1 -Section 2]. In particular, the dependencies of the obtained bounds with respect to the constants m and L which appear in **H**1, **H**2 are evidenced.

3 Quantitative bounds in total variation distance

We develop in this section quantitative bounds in total variation distance. For Bayesian inference application, total variation bounds are useful for computing highest posterior density (HPD) credible regions and intervals. For computing such bounds we will use the results of Section 2 combined with the regularizing property of the semigroup $(P_t)_{t\geq 0}$.

The first key result consists in upper-bounding the total variation distance $\|\mu P_t - \nu P_t\|_{\mathrm{TV}}$ for $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$. To that purpose, we use the coupling by reflection; see [30, Section 3] or [6, Example 3.7] for its construction, and [14, 15, 4] for applications. It is defined as the unique strong solution $(\mathsf{X}_t, \mathsf{Y}_t)_{t\geq 0}$ of the SDE:

$$\begin{cases} \mathrm{d}\mathsf{X}_t &= -\nabla U(\mathsf{X}_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t^d \\ \mathrm{d}\mathsf{Y}_t &= -\nabla U(\mathsf{Y}_t)\mathrm{d}t + \sqrt{2}(\mathrm{Id} - 2e_t e_t^T)\mathrm{d}B_t^d \end{cases}, \quad \text{where } e_t = \mathsf{e}(\mathsf{X}_t - \mathsf{Y}_t) \tag{12}$$

with $\mathsf{X}_0 = x$, $\mathsf{Y}_0 = y$, $\mathsf{e}(z) = z/\|z\|$ for $z \neq 0$ and $\mathsf{e}(0) = 0$ otherwise. Define the coupling time $T_c = \inf\{s \geq 0 \mid \mathsf{X}_s = \mathsf{Y}_s\}$. By construction $\mathsf{X}_t = \mathsf{Y}_t$ for $t \geq T_c$. Using

Levy's characterization, $\tilde{B}_t^d = \int_0^t (\mathrm{Id} - 2e_s e_s^T)\mathrm{d}B_s^d$ is a d-dimensional Brownian motion, therefore $(\mathsf{X}_t)_{t\geq 0}$ and $(\mathsf{Y}_t)_{t\geq 0}$ are weak solutions to (1) started at x and y respectively. Then by Lindvall's inequality, for all $t > 0$ we have $\|P_t(x,\cdot) - P_t(y,\cdot)\|_{\mathrm{TV}} \leq \mathbb{P}(\mathsf{X}_t \neq \mathsf{Y}_t)$.

Denote by $\boldsymbol{\Phi}$ the cumulative distribution function of the standard normal distribution. For $a > 0$, define χ_a for all $t \geq 0$ by

$$\chi_a(t) = \sqrt{(4/a)(\mathrm{e}^{2at} - 1)} \ . \tag{13}$$

Theorem 10. *Assume* $\boldsymbol{H}1$ *and* $\boldsymbol{H}2$*.*

(i) For any $x, y \in \mathbb{R}^d$ and $t > 0$, it holds

$$\|P_t(x,\cdot) - P_t(y,\cdot)\|_{\mathrm{TV}} \leq 1 - 2\boldsymbol{\Phi}\{-\|x-y\|/\chi_m(t)\} \ ,$$

where χ_m is defined in (13) and m is the strong convexity constant.

(ii) For any $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ and $t > 0$,

$$\|\mu P_t - \nu P_t\|_{\mathrm{TV}} \leq 2^{1/2}\, W_1(\mu,\nu)\Big/(\pi^{1/2}\chi_m(t)) \ .$$

(iii) For any $x \in \mathbb{R}^d$ and $t \geq 0$,

$$\|\pi - \delta_x P_t\|_{\mathrm{TV}} \leq 2^{1/2}\left\{(d/m)^{1/2} + \|x - x^\star\|\right\}\Big/(\pi^{1/2}\chi_m(t)) \ .$$

Proof. (i) Denote for $t > 0$, $\mathsf{B}_t^1 = \int_0^t \mathbb{1}_{\{s<T_c\}} e_s^T \mathrm{d}B_s^d$. We compute a bound for the coupling time. On $\{t < T_c\}$, by (12), we get

$$\mathrm{d}\{\mathsf{X}_t - \mathsf{Y}_t\} = -\{\nabla U(\mathsf{X}_t) - \nabla U(\mathsf{Y}_t)\}\,\mathrm{d}t + 2\sqrt{2}e_t \mathrm{d}\mathsf{B}_t^1 \ .$$

Itô's formula on $\{t < T_c\}$ yields

$$\begin{aligned}\mathrm{e}^{mt}\|\mathsf{X}_t - \mathsf{Y}_t\| = \|x-y\| + m\int_0^t \mathrm{e}^{ms}\|\mathsf{X}_s - \mathsf{Y}_s\|\,\mathrm{d}s \\ - \int_0^t \mathrm{e}^{ms}\langle \nabla U(\mathsf{X}_s) - \nabla U(\mathsf{Y}_s), e_s\rangle\,\mathrm{d}s + 2\sqrt{2}\int_0^t \mathrm{e}^{ms}\mathrm{d}\mathsf{B}_s^1 \ .\end{aligned}$$

Then by $\mathbf{H}2$, we obtain on $\{t < T_c\}$, $\|\mathsf{X}_t - \mathsf{Y}_t\| \leq \mathsf{U}_t$, where $(\mathsf{U}_t)_{t\in(0,T_c)}$ is the one-dimensional Ornstein-Uhlenbeck process defined by

$$\mathsf{U}_t = \mathrm{e}^{-mt}\|x-y\| + 2\sqrt{2}\int_0^t \mathrm{e}^{m(s-t)}\mathrm{d}\mathsf{B}_s^1 \ .$$

Therefore, for all $x, y \in \mathbb{R}^d$ and $t \geq 0$, we get

$$\mathbb{P}(T_c > t) \leq \mathbb{P}\left(\min_{0\leq s\leq t} \mathsf{U}_s > 0\right) \ .$$

Finally the proof follows from [2, Formula 2.0.2, page 542]. For completeness, this formula is given in Appendix D.2.

(ii) Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ and $\xi \in \Pi(\mu,\nu)$ be an optimal transference plan for (μ,ν) w.r.t. W_1. Since for all $s > 0$, $1/2 - \mathbf{\Phi}(-s) \leq (2\pi)^{-1/2} s$, (i) implies that for all $x, y \in \mathbb{R}^d$ and $t > 0$,

$$\|\mu P_t - \nu P_t\|_{\mathrm{TV}} \leq 2 \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{\|x-y\|}{(2\pi)^{1/2} \chi_m(t)} \mathrm{d}\xi(x,y) \ ,$$

which is the desired result.

(iii) The proof is a straightforward consequence of (ii) and Proposition 1-(iv).

□

Since for all $s > 0$, $s \leq \mathrm{e}^s - 1$, note that Theorem 10-(ii) implies that for all $t > 0$ and $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$,

$$\|\mu P_t - \nu P_t\|_{\mathrm{TV}} \leq (4\pi t)^{-1/2} W_1(\mu,\nu) \ . \tag{14}$$

Therefore for all bounded measurable function f, $P_t f$ is a Lipschitz function for all $t > 0$ with Lipshitz constant

$$\|P_t f\|_{\mathrm{Lip}} \leq (4\pi t)^{-1/2} \mathrm{osc}(f) \ . \tag{15}$$

We will now study the contraction of $Q_\gamma^{n,\ell}$ in total variation for non-increasing sequences $(\gamma_k)_{k \geq 1}$. Strikingly, we are able to derive results which closely parallel Theorem 10. The proof is nevertheless completely different because the reflection coupling is no longer applicable in discrete time. We use a coupling construction inspired by the method of [5, Section 3.3] for Gaussian random walks. This construction has been used in [13] to establish convergence of homogeneous Markov chain in Wasserstein distances using different method of proof. So as not to interrupt the argument, this construction is postponed to Section 6.

For all $n, \ell \geq 1$, $n < \ell$ and $(\gamma_k)_{k \geq 1}$ a non-increasing sequence denote by

$$\Lambda_{n,\ell}(\gamma) = \kappa^{-1} \left\{ \prod_{j=n}^{\ell} (1 - \kappa\gamma_j)^{-1} - 1 \right\} \ , \qquad \Lambda_\ell(\gamma) = \Lambda_{1,\ell}(\gamma) \ . \tag{16}$$

Theorem 11. *Assume* ***H****1 and* ***H****2.*

(i) Let $(\gamma_k)_{k \geq 1}$ be a non-increasing sequence satisfying $\gamma_1 \leq 2/(m+L)$. Then for all $x, y \in \mathbb{R}^d$ and $n, \ell \in \mathbb{N}^*$, $n < \ell$, we have

$$\|\delta_x Q_\gamma^{n,\ell} - \delta_y Q_\gamma^{n,\ell}\|_{\mathrm{TV}} \leq 1 - 2\mathbf{\Phi}\{-\|x-y\| / \{8\, \Lambda_{n,\ell}(\gamma)\}^{1/2}\} \ .$$

(ii) Let $(\gamma_k)_{k \geq 1}$ be a non-increasing sequence satisfying $\gamma_1 \leq 2/(m+L)$. Then, for all $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ and $\ell, n \in \mathbb{N}^*$, $n < \ell$, we have

$$\|\mu Q_\gamma^{n,\ell} - \nu Q_\gamma^{n,\ell}\|_{\mathrm{TV}} \leq \{4\pi \Lambda_{n,\ell}(\gamma)\}^{-1/2} W_1(\mu,\nu) \ .$$

(iii) Let $\gamma \in (0, 2/(m+L)]$. Then for any $x \in \mathbb{R}^d$ and $n \geq 1$,

$$\|\pi_\gamma - \delta_x R_\gamma^n\|_{\mathrm{TV}} \leq \{4\pi\kappa(1-(1-\kappa\gamma)^{n/2})\}^{-1/2} (1-\kappa\gamma)^{n/2} \left\{ \|x - x^\star\| + (2\kappa^{-1} d)^{1/2} \right\} \ .$$

Proof. (i) By (42) for all x, y and $k \geq 1$, we have

$$\|x - \gamma_k \nabla U(x) - y + \gamma_k \nabla U(y)\| \leq (1 - \kappa\gamma_k)^{1/2} \|x - y\| \ .$$

Let $n, \ell \geq 1$, $n < \ell$, then applying Theorem 19 in Section 6, we get

$$\|\delta_x Q_\gamma^{n,\ell} - \delta_y Q_\gamma^{n,\ell}\|_{\mathrm{TV}} \leq 1 - 2\mathbf{\Phi}\left(- \|x - y\| / \{8\, \Lambda_{n,\ell}(\gamma)\}^{1/2}\right) \ ,$$

(ii) Let $f \in \mathbb{F}_b(\mathbb{R}^d)$ and $\ell > n \geq 1$. For all $x, y \in \mathbb{R}^d$ by definition of the total variation distance and (i), we have

$$\begin{aligned}
\left| Q_\gamma^{n,\ell} f(x) - Q_\gamma^{n,\ell} f(y) \right| &\leq \operatorname{osc}(f) \|\delta_x Q_\gamma^{n,\ell} - \delta_y Q_\gamma^{n,\ell}\|_{\mathrm{TV}} \\
&\leq \operatorname{osc}(f) \left\{ 1 - 2\mathbf{\Phi}\left(- \|x - y\| / \{8\, \Lambda_{n,\ell}(\gamma)\}^{1/2}\right) \right\} \ ,
\end{aligned}$$

Using that for all $s > 0$, $1/2 - \mathbf{\Phi}(-s) \leq (2\pi)^{-1/2} s$ concludes the proof.

(iii) The proof follows from (iii), the bound for all $s > 0$, $1/2 - \mathbf{\Phi}(-s) \leq (2\pi)^{-1/2} s$ and Proposition 2-(ii).

□

We can combine Theorem 5 or Theorem 8 with Theorem 10 and Theorem 11 to obtain explicit bounds in total variation between the Euler-Maruyama discretization and the target distribution π. To that purpose, we use the following decomposition, for all non-increasing sequence $(\gamma_k)_{k \geq 1}$, initial point $x \in \mathbb{R}^d$ and $\ell \geq 0$:

$$\|\pi - \delta_x Q_\gamma^\ell\|_{\mathrm{TV}} \leq \|\pi - \delta_x P_{\Gamma_\ell}\|_{\mathrm{TV}} + \|\delta_x P_{\Gamma_\ell} - \delta_x Q_\gamma^\ell\|_{\mathrm{TV}} \ . \tag{17}$$

The first term is dealt with Theorem 10-(iii). It remains to bound the second term in (17). Since we will use Theorem 5 and Theorem 8, we have two different results depending on the assumptions on U. Define for all $x \in \mathbb{R}^d$ and $n, p \in \mathbb{N}$,

$$\begin{aligned}
\vartheta_{n,p}^{(1)}(x) = L^2 \sum_{i=1}^{n} \gamma_i^2 \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2) &\left[\left\{ \kappa^{-1} + \gamma_i \right\} (2d + dL^2\gamma_i^2/6) \right. \\
&\left. + L^2 \gamma_i \delta_{i,n,p}(x) \left\{ \kappa^{-1} + \gamma_i \right\} \right]
\end{aligned} \tag{18}$$

$$\begin{aligned}
\vartheta_{n,p}^{(2)}(x) = \sum_{i=1}^{n} \gamma_i^3 \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2) &\left[L^4 \delta_{i,n,p}(x) (4\kappa^{-1}/3 + \gamma_{n+1}) \right. \\
&\left. + d\left\{ 2L^2 + 4\kappa^{-1}(d\tilde{L}^2/3 + \gamma_{n+1} L^4/4) + \gamma_{n+1}^2 L^4/6 \right\} \right] \ ,
\end{aligned} \tag{19}$$

where

$$\delta_{i,n,p}(x) = \mathrm{e}^{-2m\Gamma_{i-1}} \varrho_{n,p}(x) + (1 - \mathrm{e}^{-2m\Gamma_{i-1}})(d/m) \ ,$$

and $\varrho_{n,p}(x)$ is given by (7).

Theorem 12. *Assume* **H** *1 and* **H** *2. Let $(\gamma_k)_{k\ge1}$ be a non-increasing sequence with $\gamma_1 \le 1/(m+L)$. Then for all $x \in \mathbb{R}^d$ and $\ell, n \in \mathbb{N}^*$, $\ell > n$,*

$$\|\delta_x P_{\Gamma_\ell} - \delta_x Q_\gamma^\ell\|_{\mathrm{TV}} \le (\vartheta_n(x)/(4\pi\Gamma_{n+1,\ell}))^{1/2} \\ + 2^{-3/2} L \left(\sum_{k=n+1}^{\ell} \left\{ (\gamma_k^3 L^2/3) \varrho_{1,k-1}(x) + d\gamma_k^2 \right\} \right)^{1/2} , \quad (20)$$

where $\varrho_{1,n}(x)$ is defined by (7), $\vartheta_n(x)$ is equal to $\vartheta_{n,0}^{(2)}(x)$ given by (19), if **H** *3 holds, and to $\vartheta_{n,0}^{(1)}(x)$ given by (18) otherwise.*

Proof. The proof is postponed to Appendix B.1. □

Consider the case of decreasing step sizes of the form $\gamma_k = \gamma_1/k^\alpha$ for $k \ge 1$ and $\alpha \in (0,1)$. Under **H**1 and **H**2, setting $n = \ell - \lfloor \ell^\alpha \rfloor$, $\ell \in \mathbb{N}^*$, we have for $i = 2, 3$,

$$\lim_{n\to+\infty} \Gamma_{n,\ell} = 1 , \quad \sum_{k=n+1}^{\ell} \gamma_k^i \le \gamma_{n+1}^i (\ell - n) \le \gamma_1^i \lfloor \ell^\alpha \rfloor / (\ell - \lfloor \ell^\alpha \rfloor)^{i\alpha} . \quad (21)$$

In addition, by Table 2, $\vartheta_n(x) = d\mathcal{O}(\ell^{-\alpha})$. Therefore combining this result and (21) in the bound of Theorem 12, we get that $\|\delta_{x^\star} Q_\gamma^\ell - \pi\|_{\mathrm{TV}} = d^{1/2}\mathcal{O}(\ell^{-\alpha/2})$. In the case $\gamma_k = \gamma_1/k^\alpha$ for $k \ge 1$ and $\alpha = 1$, setting $n = \ell - \lfloor \ell/2 \rfloor$, $\ell \in \mathbb{N}^*$, $\ell > 2$, we have for $i = 2, 3$,

$$\lim_{n\to+\infty} \Gamma_{n,\ell} = 1/2 , \quad \sum_{k=n+1}^{\ell} \gamma_k^i \le \gamma_{n+1}^i (\ell - n) \le \gamma_1^i / (\ell/2 - 1) . \quad (22)$$

In addition, by Table 2, $\vartheta_n(x) = d\mathcal{O}(\ell^{-1})$, for $\gamma_1 > 2\kappa^{-1}$. Therefore combining this result and (22) in the bound of Theorem 12, we get that $\|\delta_{x^\star} Q_\gamma^\ell - \pi\|_{\mathrm{TV}} = d^{1/2}\mathcal{O}(\ell^{-1/2})$.

Note that these rates for $\gamma_k = \gamma_1/k^\alpha$, $k \in \mathbb{N}^*$ and $\alpha \in (0,1]$ improve those obtained in [12, Proposition 3], for potentials satisfying **H** 1 but not necessarily convex since [12, Proposition 3] only requires the additional assumption that $(P_t)_{t\ge0}$ is geometrically ergodic in total variation.

Assume **H**1, **H**2 and **H**3 and that $\gamma_k = \gamma_1/k^\alpha$ for $k \ge 1$ and $\alpha \in (0,1]$. setting $n = \ell - \lfloor \ell^{\alpha/2} \rfloor$, $\ell \in \mathbb{N}^*$, we have for $i = 2, 3$,

$$\lim_{n\to+\infty} \Gamma_{n,\ell} = 1 , \quad \sum_{k=n+1}^{\ell} \gamma_k^i \le \gamma_{n+1}^i (\ell - n) \le \gamma_1^i \lfloor \ell^{\alpha/2} \rfloor / (\ell - \lfloor \ell^{\alpha/2} \rfloor)^{i\alpha} . \quad (23)$$

In addition (see Table 2) $\vartheta_n(x) = d^2\mathcal{O}(\ell^{-2\alpha})$, with $\gamma_1 > 4\kappa^{-1}$ in the case $\alpha = 1$. Therefore combining this result and (23) in the bound of Theorem 12, we get that $\|\delta_{x^\star} Q_\gamma^\ell - \pi\|_{\mathrm{TV}} = d^{1/2}\mathcal{O}(\ell^{-3\alpha/4})$. These discussions are summarized in Table 3.

When $\gamma_k = \gamma \in (0, 1/(m+L))$ for all $k \ge 1$, under **H**1 and **H**2, for $\ell > \lceil \gamma^{-1} \rceil$ choosing $n = \ell - \lceil \gamma^{-1} \rceil$ implies that (see Appendix B.2)

$$\|\delta_x R_\gamma^\ell - \delta_x P_{\ell\gamma}\|_{\mathrm{TV}} \le (4\pi)^{-1/2} \left[\gamma \mathsf{D}_1(\gamma, d) + \gamma^3 \mathsf{D}_2(\gamma) \mathsf{D}_3(\gamma, d, x) \right]^{1/2} + \mathsf{D}_4(\gamma, d, x) , \quad (24)$$

	$\alpha \in (0,1)$	$\alpha = 1$
Theorem 5	$d^{1/2}\,\mathcal{O}(\ell^{-\alpha/2})$	$d^{1/2}\,\mathcal{O}(\ell^{-1/2})$ for $\gamma_1 > 2\kappa^{-1}$
Theorem 8	$d^{1/2}\,\mathcal{O}(\ell^{-3\alpha/4})$	$d^{1/2}\,\mathcal{O}(\ell^{-3/4})$ for $\gamma_1 > 4\kappa^{-1}$

Table 3: Order of convergence of $\|\delta_{x^\star} Q_\gamma^\ell - \pi\|_{\mathrm{TV}}$ for $\gamma_k = \gamma_1/k^\alpha$ based on Theorem 12

where

$$\begin{aligned}
&\mathsf{D}_1(\gamma, d) = 2L^2\kappa^{-1}\left(\kappa^{-1}+\gamma\right)\left(2d + L^2\gamma^2/6\right) , \mathsf{D}_2(\gamma) = L^4\left(\kappa^{-1}+\gamma\right) \\
&\mathsf{D}_3(\gamma, d, x) = \left\{(\ell - \lceil \gamma^{-1}\rceil)\mathrm{e}^{-m\gamma(\ell - \lceil \gamma^{-1}\rceil - 1)}\|x - x^\star\|^2 + 2d(\kappa\gamma m)^{-1}\right\} \\
&\mathsf{D}_4(\gamma, d, x) = 2^{-3/2}L\left[d\gamma(1+\gamma)\right. \\
&\left.+(L^2\gamma^3/3)\left\{(1+\gamma^{-1})(1-\kappa\gamma)^{\ell - \lceil \gamma^{-1}\rceil}\|x-x^\star\|^2 + 2(1+\gamma)\kappa^{-1}d\right\}\right]^{1/2} .
\end{aligned} \tag{25}$$

Using this bound and Theorem 10-(iii), the number of iterations $\ell_\varepsilon > 0$ to achieve $\|\delta_{x^\star} R_{\gamma_\varepsilon}^{\ell_\varepsilon} - \pi\|_{\mathrm{TV}} \leq \varepsilon$ is of order $d\log(d)\mathcal{O}(|\log(\varepsilon)|\,\varepsilon^{-2})$ (the proper choice of the step size γ_ε is given in Table 5). This result is the same than the one obtained in [12].

Letting ℓ go to infinity in (24) we get the following result.

Corollary 13. *Assume* **H***1 and* **H***2. Let* $\gamma \in (0, 1/(m+L)]$. *Then it holds*

$$\|\pi_\gamma - \pi\|_{\mathrm{TV}} \leq 2^{-3/2}L\left[d\gamma(1+\gamma) + 2(L^2\gamma^3/3)(1+\gamma)\kappa^{-1}d\right]^{1/2} + (4\pi)^{-1/2}\left[\gamma\mathsf{D}_1(\gamma,d) + 2d\gamma^2\mathsf{D}_2(\gamma)(\kappa m)^{-1}\right]^{1/2} ,$$

where $\mathsf{D}_1(\gamma)$ *and* $\mathsf{D}_2(\gamma)$ *are given in* (25).

Note that Corollary 13 shows that $\|\pi_\gamma - \pi\|_{V^{1/2}} \leq C_1\gamma^{1/2}$ for some constant $C_1 \geq 0$. Under **H**1 and the assumption and R_γ and $(P_t)_{t\geq 0}$ are V-uniformly geometrically ergodic, [12, Theorem 10] establishes that $\|\pi_\gamma - \pi\|_{V^{1/2}} \leq C_2\gamma^{1/2}$ for some explicit constant $C_2 \geq 0$. In the case where U satisfies **H**2, then we can take $V = \|\cdot\|^2$ and C_2 is very similar to C_1. In particular both C_1 and C_2 are of order $d^{1/2}$.

However, if **H**3 holds, for constant step sizes, we can improve with respect to the step size γ, the bounds given by Corollary 13.

Theorem 14. *Assume* **H***1,* **H***2 and* **H***3. Let* $\gamma \in (0, 1/(m+L)]$. *Then it holds*

$$\begin{aligned}
\|\pi_\gamma - \pi\|_{\mathrm{TV}} &\leq (4\pi)^{-1/2}\left\{\gamma^2\mathsf{E}_1(\gamma,d) + 2d\gamma^2\mathsf{E}_2(\gamma)/(\kappa m)\right\}^{1/2} \\
&+ (4\pi)^{-1/2}\left\lceil \log\left(\gamma^{-1}\right)/\log(2)\right\rceil\left\{\gamma^2\mathsf{E}_1(\gamma,d) + \gamma^2\mathsf{E}_2(\gamma)(2\kappa^{-1}d + d/m)\right\}^{1/2} \\
&+ 2^{-3/2}L\left\{2d\gamma^3L^2/(3\kappa) + d\gamma^2\right\}^{1/2} ,
\end{aligned}$$

where $\mathsf{E}_1(\gamma,d)$ *and* $\mathsf{E}_2(\gamma)$ *are defined by*

$$\begin{aligned}
\mathsf{E}_1(\gamma,d) &= 2d\kappa^{-1}\left\{2L^2 + 4\kappa^{-1}(d\tilde{L}^2/3 + \gamma L^4/4) + \gamma^2L^4/6\right\} \\
\mathsf{E}_2(\gamma) &= L^4(4\kappa^{-1}/3 + \gamma) .
\end{aligned}$$

Proof. The proof is postponed to Appendix B.3. □

Note that the bound provided by Theorem 14 is of order $d\mathcal{O}(\gamma\,|\log(\gamma)|)$, improving the dependency given by Corollary 13 and [12, Theorem 10], with respect to the step size γ, but Theorem 14 requires that **H**3 holds contrary to Corollary 13 and [12, Theorem 10]. Furthermore when $\tilde{L} = 0$, this bound given by Theorem 14 is of order $d^{1/2}\mathcal{O}(\gamma\,|\log(\gamma)|)$ and is sharp up to a logarithmic factor. Indeed, assume that π is the d-dimensional standard Gaussian distribution. In such case, the ULA sequence $(X_k)_{k\geq 0}$ is the autoregressive process given for all $k \geq 0$ by $X_{k+1} = (1-\gamma)X_k + \sqrt{2\gamma}Z_{k+1}$. For $\gamma \in (0,1)$, this sequence has a stationary distribution π_γ, which is a d-dimensional Gaussian distribution with zero-mean and covariance matrix $\sigma_\gamma^2\,\mathrm{I}_d$, with $\sigma_\gamma^2 = (1-\gamma/2)^{-1}$. Therefore, using [26, Lemma 4.9] (or the Pinsker inequality), we get the following upper bound: $\|\pi - \pi_\gamma\|_{\mathrm{TV}} \leq Cd^{1/2}|\sigma_\gamma^2 - 1| = Cd^{1/2}\gamma/2$, where C is a universal constant.

We can also for a precision target $\varepsilon > 0$ choose $\gamma_\varepsilon > 0$ and the number of iterations $n_\varepsilon > 0$ to get $\|\delta_x R_{\gamma_\varepsilon}^{n_\varepsilon} - \pi\|_{\mathrm{TV}} \leq \varepsilon$. By Theorem 10-(iii), Theorem 11-(iii) and Theorem 14, a sufficient number of iterations ℓ_ε is of order $d\log^2(d)\mathcal{O}(\varepsilon^{-1}\log^2(\varepsilon))$ for a well chosen step size γ_ε. This result improves the conclusion of [12] and Corollary 13 with respect to the precision parameter ε, which provides an upper bound of the number of iterations of order $d\log(d)\mathcal{O}(\varepsilon^{-2}\log^2(\varepsilon))$. We can also compare our reported upper bound with the one obtained for the d-dimensional standard Gaussian distribution. If the initial distribution is the Dirac mass at zero (the minimum of the potential $U(x) = \|x\|^2/2$) and $\gamma \in (0,1)$, the distribution of the ULA sequence after n iterations is zero-mean Gaussian with covariance $(1-(1-\gamma)^{2(n+1)})/(1-\gamma/2)\,\mathrm{I}_d$. If we use [26, Lemma 4.9] again, we get for $\gamma \in (0,1)$,

$$\|\delta_0 R_\gamma^n - \pi\|_{\mathrm{TV}} \leq Cd^{1/2}\gamma|1 - 2\gamma^{-1}(1-\gamma)^{2(n+1)}|\;,$$

where C is a universal constant. To get an ε precision we need to choose $\gamma_\varepsilon = d^{-1/2}\varepsilon/(2C)$ and then $n_\varepsilon = \lceil (1/2)\log(\gamma_\varepsilon/4)/\log(1-\gamma_\varepsilon)\rceil = d^{1/2}\log(d)\mathcal{O}(\varepsilon^{-1}|\log(\varepsilon)|)$. On the other hand since $\tilde{L} = 0$, based on the bound given by Theorem 14, a sufficient number of iterations to get $\|\delta_x R_{\gamma_\varepsilon}^{n_\varepsilon} - \pi\|_{\mathrm{TV}} \leq \varepsilon$ is of order $d^{1/2}\log^2(d)\mathcal{O}(\varepsilon^{-1}\log^2(\varepsilon))$. It follows that our upper bound for the step size and the optimal number of iterations is again sharp up to a logarithmic factor in the dimension and the precision. The discussions on the bounds for constant sequences of step sizes are summarized in Table 4 and Table 5.

	H1, **H**2	**H**1, **H**2 and **H**3
$\|\pi - \pi_\gamma\|_{\mathrm{TV}}$	$d^{1/2}\mathcal{O}(\gamma^{1/2})$	$d\mathcal{O}(\gamma\,|\log(\gamma)|)$

Table 4: Order of the bound between π and π_γ in total variation function of the step size $\gamma > 0$ and the dimension d.

	H1, H2	H1, H2 and H3
γ_ε	$d^{-1}\mathcal{O}(\varepsilon^2)$	$d^{-1}\log^{-1}(d)\mathcal{O}(\varepsilon \lvert\log^{-1}(\varepsilon)\rvert)$
n_ε	$d\log(d)\mathcal{O}(\varepsilon^{-2}\lvert\log(\varepsilon)\rvert)$	$d\log^2(d)\mathcal{O}(\varepsilon^{-1}\log^2(\varepsilon))$

Table 5: Order of the step size $\gamma_\varepsilon > 0$ and the number of iterations $n_\varepsilon \in \mathbb{N}^*$ to get $\|\delta_{x^\star} R_{\gamma_\varepsilon}^{n_\varepsilon} - \pi\|_{\mathrm{TV}} \leq \varepsilon$ for $\varepsilon > 0$.

4 Mean square error and concentration for bounded measurable functions

Let $(X_k)_{k\geq 0}$ be the Euler discretization of the Langevin diffusion (2) associated with the sequence of non-increasing step sizes $(\gamma_k)_{k\geq 1}$. The result of the previous section allows us to study the approximation of $\pi(f)$ by the weighted average estimator $\hat{\pi}_n^N(f)$ defined, for $f : \mathbb{R}^d \to \mathbb{R}$, $N, n \in \mathbb{N}$, $n \geq 1$ by

$$\hat{\pi}_n^N(f) = \sum_{k=N+1}^{N+n} \omega_{k,n}^N f(X_k) , \quad \omega_{k,n}^N = \gamma_{k+1}\Gamma_{N+2,N+n+1}^{-1} . \tag{26}$$

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 $x \in \mathbb{R}^d$. First we derive a bound on the mean-square error, defined as

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

for $f : \mathbb{R}^d \to \mathbb{R}$, which is either Lipschitz or measurable and bounded. This quantity can be decomposed as the sum of the squared bias and variance:

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

We first obtain a bound for the bias for f Lipschitz. For all $k \in \{N+1, \ldots, N+n\}$, denote by ξ_k the optimal transference plan between $\delta_x Q_\gamma^k$ and π for W_2, *i.e.* $W_2^2(\delta_x Q_\gamma^k, \pi) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \|x-y\|^2 \mathrm{d}\xi_k(x,y)$. Then by the Jensen inequality and because f is Lipschitz, we have:

$$\begin{aligned}
\left\{\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right\}^2 &= \left(\sum_{k=N+1}^{N+n} \omega_{k,n}^N \int_{\mathbb{R}^d\times\mathbb{R}^d} \{f(z) - f(y)\}\xi_k(\mathrm{d}z, \mathrm{d}y)\right)^2 \\
&\leq \|f\|_{\mathrm{Lip}}^2 \sum_{k=N+1}^{N+n} \omega_{k,n}^N \int_{\mathbb{R}^d\times\mathbb{R}^d} \|z-y\|^2 \xi_k(\mathrm{d}z, \mathrm{d}y) \\
&\leq \|f\|_{\mathrm{Lip}}^2 \sum_{k=N+1}^{N+n} \omega_{k,n}^N W_2^2(\delta_x Q_\gamma^k, \pi) .
\end{aligned} \tag{27}$$

Similarly, if f is bounded,

$$\left(\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right)^2 \le \operatorname{osc}(f)^2 \sum_{k=N+1}^{N+n} \omega_{k,n}^N \|\delta_x Q_\gamma^k - \pi\|_{\mathrm{TV}}^2 \;;$$

Using the results of Sections 2 and 3, we can deduce different bounds for the bias, depending on the assumptions on U and the sequence of step sizes $(\gamma_k)_{k\ge1}$. We now derive a bound for the variance. We get then two different results depending on the class to which the function f belongs. In the case of Lipschitz function, we adapt the proof of [24, Theorem 2] for homogeneous Markov chain to our inhomogeneous setting.

Theorem 15. *Assume* **H***1 and* **H***2. Let $(\gamma_k)_{k\ge1}$ be a non-increasing sequence with $\gamma_1 \le 2/(m+L)$ and $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz function. Then for all $N \ge 0$ and $n \ge 1$, we get $\operatorname{Var}_x\{\hat{\pi}_n^N(f)\} \le 8\kappa^{-2}\|f\|_{\mathrm{Lip}}^2 \Gamma_{N+2,N+n+1}^{-1} v_{N,n}(\gamma)$, where*

$$v_{N,n}(\gamma) = \left\{1 + \Gamma_{N+2,N+n+1}^{-1}(\kappa^{-1} + 2/(m+L))\right\} \;. \tag{28}$$

Proof. The proof is postponed to Appendix C.1.1. □

It is noteworthy to observe that the bound for the variance does not depend on the dimension. We may now discuss the bounds on the MSE (obtained by combining the bounds for the squared bias (27) from Theorems 5 and 8, and the variance Theorem 15) for step sizes given for $k \ge 1$ by $\gamma_k = \gamma_1/k^\alpha$ where $\alpha \in [0,1]$ and $\gamma_1 < 1/(m+L)$. Details of these calculations are postponed to [11, Sections 4.1 and 4.2]. The order of the bounds (up to numerical constants) of the MSE are summarized in Table 6 as a function of γ_1, n and N. Then, we can conclude that in the infinite horizon setting, it is optimal to take $\alpha = 1/2$ under **H**1 and **H**2, and $\alpha = 1/3$ under **H**1, **H**2 and **H**3. Note that [27] shows also that the optimal value for α is $1/3$ by studying the asymptotic behaviour of $\hat{\pi}_n^0(f)$ as $n \to +\infty$ for smooth functions $f : \mathbb{R}^d \to \mathbb{R}$.

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

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

In the case $\gamma_k = \gamma$ for all $k \in \mathbb{N}^*$ and the total number of iterations $n + N$ is held fixed (fixed horizon setting), we optimize the value of the step size γ but also of the burn-in period N to get an upper bound of order $n^{-1/2}$ under **H**1 and **H**2, and $n^{-2/3}$ under **H**1, **H**2 and **H**3.

In the case where f is measurable and bounded, we have the following result.

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

Table 7: 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

Theorem 16. *Assume* ***H*** *1 and* ***H*** *2. Let* $(\gamma_k)_{k\geq 1}$ *be a non-increasing sequence with* $\gamma_1 \leq 2/(m+L)$ *and* $f : \mathbb{R}^d \to \mathbb{R}$ *be a measurable and bounded function. Then for all* $N \geq 0$, $n \geq 1$, $x \in \mathbb{R}^d$, *we get*

$$\mathrm{Var}_x\{\hat{\pi}_n^N(f)\} \leq \mathrm{osc}(f)^2\{2\gamma_1 \Gamma_{N+2,N+n+1}^{-1} + u_{N,n}^{(4)}(\gamma)\}$$

$$u_{N,n}^{(4)}(\gamma) = \sum_{k=N}^{N+n-1} \gamma_{k+1} \left\{ \sum_{i=k+2}^{N+n} \frac{\omega_{i,n}^N}{(\pi \Lambda_{k+2,i}(\gamma))^{1/2}} \right\}^2 + \kappa^{-1} \left\{ \sum_{i=N+1}^{N+n} \frac{\omega_{i,n}^N}{(4\pi \Lambda_{N+1,i}(\gamma))^{1/2}} \right\}^2 , \quad (29)$$

for $n_1, n_2 \in \mathbb{N}$, $\Lambda_{n_1,n_2}(\gamma)$ *is given by* (16).

Proof. The proof is postponed to Appendix C.1.2. ∎

To illustrate the result Theorem 16, we first illustrate numerically the behaviour $(u_{N,n}^{(4)})_{n\geq 1}$ for $\kappa = 1$ $N = 0$, and four different non-increasing sequences of step sizes $(\gamma_k)_{k\geq 1}$, $\gamma_k = (1+k)^{-\alpha}$ for $\alpha = 1/4, 1/2, 3/4$ and $\gamma_k = 1/2$ for $k \geq 1$. These results are gathered in Figure 1, where it can be observed that $(\Gamma_n u_{0,n}^{(4)}(\gamma))_{n\geq 1}$ converges to a limit as $n \to +\infty$. In Appendix C.2, we show that there exist $C_1, C_2 > 0$ independent of $(\gamma_k)_{k\geq 1}$, such that $C_1 \Gamma_n^{-1} \leq u_{0,n}^{(4)}(\gamma) \leq C_2 \Gamma_n^{-1}$, for non-increasing sequence $(\gamma_k)_{k\geq 1}$ satisfying $\lim_{k\to+\infty} \gamma_k = 0$ and $\lim_{k\to+\infty} \Gamma_k = +\infty$. Therefore, the consequences of Theorem 16 are similar to those of Theorem 15 and are omitted.

We now establish an exponential deviation inequality for $\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)]$ given by (26) for a bounded measurable function f.

Theorem 17. *Assume* ***H*** *1 and* ***H*** *2. Let* $(\gamma_k)_{k\geq 1}$ *be a non-increasing sequence with* $\gamma_1 \leq 2/(m+L)$. *Then for all* $N \geq 0$, $n \geq 1$, $r > 0$ *and Lipschitz functions* $f : \mathbb{R}^d \to \mathbb{R}$:

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

where $\upsilon_{N,n}(\gamma)$ *is defined by* (28).

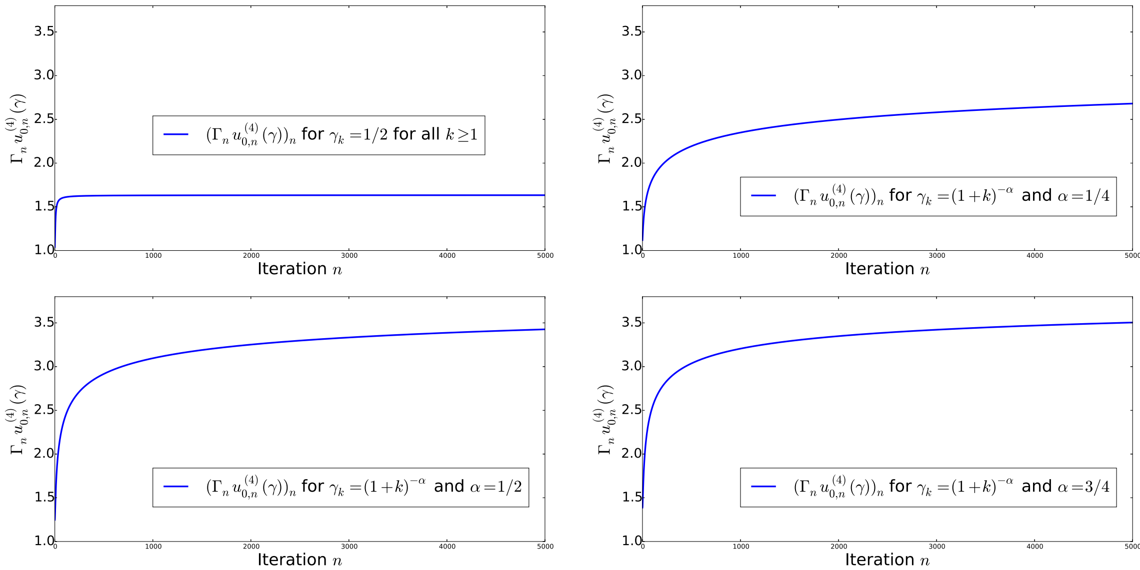

Figure 1: Plots of $(u_{0,n}^{(4)})_{n\geq 1}\Gamma_n$ for four sequences of step sizes $(\gamma_k)_{k\geq 1}$, $\gamma_k = (1+k)^{-\alpha}$ for $\alpha = 0, 1/4, 1/2, 3/4$

Proof. The proof is postponed to Appendix C.3. ☐

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 $\exp(-Cr^2\gamma_1 n^{1-\alpha})$ for $\alpha \in [0,1)$, for some constant $C \geq 0$ independent of γ_1 and n.

Theorem 18. *Assume **H** 1 and **H** 2. Let $(\gamma_k)_{k\geq 1}$ be a non-increasing 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 functions $f \in \mathbb{F}_b(\mathbb{R}^d)$:*

$$\mathbb{P}_x\left[\hat{\pi}_n^N(f) \geq \mathbb{E}_x[\hat{\pi}_n^N(f)] + r\right] \leq \mathrm{e}^{-\left\{r - \mathrm{osc}(f)(\Gamma_{N+2,N+n+1})^{-1}\right\}^2 / \{2\mathrm{osc}(f)^2 u_{N,n}^{(5)}(\gamma)\}} ,$$

where

$$u_{N,n}^{(5)}(\gamma) = \sum_{k=N}^{N+n-1} \gamma_{k+1}\left\{\sum_{i=k+2}^{N+n} \frac{\omega_{i,n}^N}{(\pi\Lambda_{k+2,i})^{1/2}}\right\}^2 + \kappa^{-1}\left\{\sum_{i=N+1}^{N+n} \frac{\omega_{i,n}^N}{(\pi\Lambda_{N+1,i})^{1/2}}\right\}^2 .$$

Proof. The proof is postponed to Appendix C.4. ☐

Note that $u_{N,n}^{(5)}(\gamma)$ is up to numerical constants similar to $u_{N,n}^{(4)}(\gamma)$ given in (29). Therefore, using the same calculations as in Appendix C.2, there exist $C_1, C_2 > 0$ such that $C_1\Gamma_n^{-1} \leq u_{0,n}^{(5)}(\gamma) \leq C_2\Gamma_n^{-1}$, for $\gamma_k = \gamma_1/k^{-\alpha}$, $\alpha \in [0,1]$. Then, if we apply Theorem 18 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 $\exp(-Cr^2\gamma_1 n^{1-\alpha})$ for $\alpha \in [0,1)$, for some constant $C \geq 0$ independent of γ_1 and n.

5 Numerical experiments

Consider a binary regression set-up in which the binary observations (responses) $\{Y_i\}_{i=1}^p$ are conditionally independent Bernoulli random variables with parameters $\{\varrho(\boldsymbol{\beta}^T X_i)\}_{i=1}^p$, where ϱ is the logistic function defined for $z \in \mathbb{R}$ by $\varrho(z) = \mathrm{e}^z/(1+\mathrm{e}^z)$ and $\{X_i\}_{i=1}^p$ and $\boldsymbol{\beta}$ are d dimensional vectors of known covariates and unknown regression coefficients, respectively. The prior distribution for the parameter $\boldsymbol{\beta}$ is a zero-mean Gaussian distribution with covariance matrix $\Sigma_{\boldsymbol{\beta}}$. The density of the posterior distribution of $\boldsymbol{\beta}$ is up to a proportionality constant given by

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

Bayesian inference for the logistic regression model has long been recognized as a numerically involved problem. Several algorithms have been proposed, trying to mimick the data-augmentation (DA) approach of [1] for probit regression; see [23], [18] and [19]. Recently, a very promising DA algorithm has been proposed in [36], 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. 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 $\pi_{\boldsymbol{\beta}}(\cdot|\{(X_i, Y_i)\}_{i=1}^p)$. The gradient of its log-density may be expressed as

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

Therefore $-\log \pi_{\boldsymbol{\beta}}(\cdot|\{X_i, Y_i\}_{i=1}^p)$ is strongly convex **H**2 with $m = \lambda_{\max}^{-1}(\Sigma_{\boldsymbol{\beta}})$ and satisfies **H**1 with $L = (1/4)\sum_{i=1}^p X_i^{\mathrm{T}} X_i + \lambda_{\min}^{-1}(\Sigma_{\boldsymbol{\beta}})$, where $\lambda_{\min}(\Sigma_{\boldsymbol{\beta}})$ and $\lambda_{\max}(\Sigma_{\boldsymbol{\beta}})$ denote the minimal and maximal eigenvalues of $\Sigma_{\boldsymbol{\beta}}$, respectively. We first compare the histograms produced by ULA and the Pòlya-Gamma Gibbs sampling from [36]. For that purpose, we take $d = 5$, $p = 100$, generate synthetic data $(Y_i)_{1\le i\le p}$ and $(X_i)_{1\le i\le p}$, and set $\Sigma_{\boldsymbol{\beta}}^{-1} = (dp)^{-1}(\sum_{i=1}^p X_i^{\mathrm{T}} X_i)\,\mathrm{I}_d$. We produce 10^8 samples from the Pólya-Gamma sampler using the R package BayesLogit [43]. Next, we make 10^3 runs of the Euler approximation scheme with $n = 10^6$ effective iterations, with a constant sequence $(\gamma_k)_{k\ge1}$, $\gamma_k = 10(\kappa n^{1/2})^{-1}$ for all $k \ge 0$ and a burn-in period $N = n^{1/2}$. The histogram of the Pólya-Gamma Gibbs sampler for first component, the corresponding mean of the obtained histograms for ULA and the 0.95 quantiles are displayed in Figure 2. The same procedure is also applied with the decreasing step size sequence $(\gamma_k)_{k\ge1}$ 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 2. In addition, we also compare MALA and ULA on five real data sets, which are summarized in Table 8. Note that for the Australian credit data set, the ordinal covariates have been stratified by dummy variables. Furthermore, we normalized the data sets

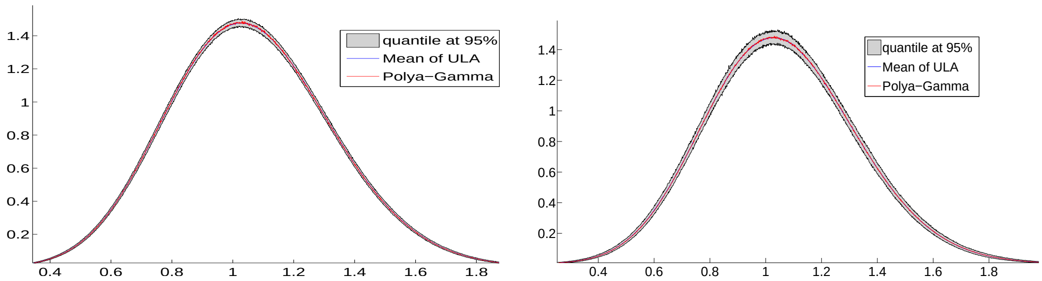

Figure 2: Empirical distribution comparison between the Polya-Gamma Gibbs Sampler and ULA. Left panel: constant step size $\gamma_k = \gamma_1$ for all $k \geq 1$; right panel: decreasing step size $\gamma_k = \gamma_1 k^{-1/2}$ for all $k \geq 1$

Data set \ Dimensions	Observations p	Covariates d
German credit [1]	1000	25
Heart disease [2]	270	14
Australian credit[3]	690	35
Pima indian diabetes[4]	768	9
Musk[5]	476	167

Table 8: Dimension of the data sets

and consider the Zellner prior setting $\Sigma^{-1} = (\pi^2 d/3)\Sigma_X^{-1}$ where $\Sigma_X = p^{-1}\sum_{i=1}^{p} X_i X_i^T$; see [39], [22] 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 $\Sigma_X^{1/2}$. We compare MALA and ULA for each data sets by estimating for each component $i \in \{1, \ldots, d\}$ the marginal accuracy between their d marginal empirical distributions and the d marginal posterior distributions, where the marginal accuracy between two probability measure μ, ν on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is defined by

$$\mathrm{MA}(\mu, \nu) = 1 - (1/2)\|\mu - \nu\|_{\mathrm{TV}} .$$

This quantity has already been considered in [17] and [8] to compare approximate samplers. To estimate the d marginal posterior distributions, we run $2 \cdot 10^7$ iterations of the Polya-Gamma Gibbs sampler. Then 100 runs of MALA and ULA (10^6 iterations per run) have been performed. For MALA, the step size is chosen so that the acceptance probability at stationarity is approximately equal to 0.5 for all the data sets. For ULA, we choose the same constant step size than MALA. We display the boxplots of the mean of the estimated marginal accuracy across all the dimensions in Figure 3. These results all imply that ULA is an alternative to the Polya-Gibbs sampler and the MALA algorithm.

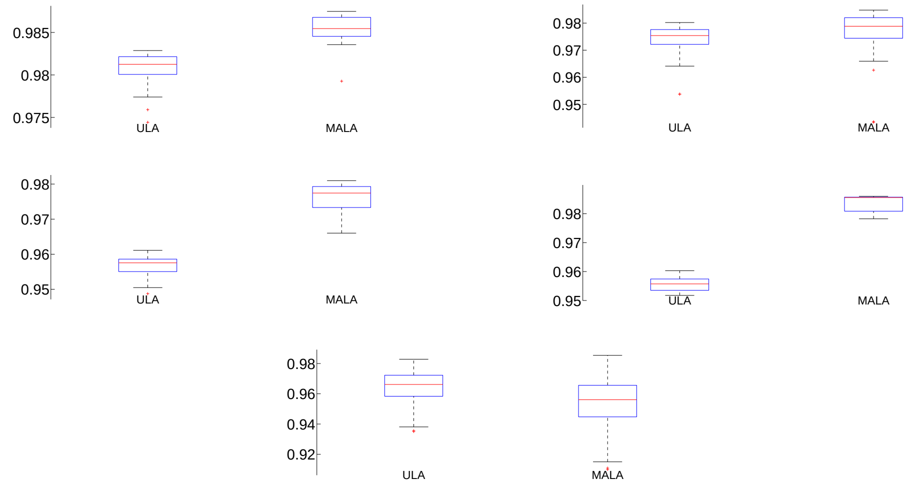

Figure 3: Marginal accuracy across all the dimensions.
Upper left: German credit data set. Upper right: Australian credit data set. Lower left: Heart disease data set. Lower right: Pima Indian diabetes data set. At the bottom: Musk data set

6 Contraction in total variation for functional autoregressive models

In this section, we consider functional autoregressive models defined for $k \geq 0$ by

$$\mathrm{X}_{k+1} = h_{k+1}(\mathrm{X}_k) + \sigma_{k+1}\mathrm{Z}_{k+1} \,, \tag{30}$$

where $(\mathrm{Z}_k)_{k\geq 1}$ is a sequence of i.i.d. d dimensional standard Gaussian random variables, $(\sigma_k)_{k\geq 1}$ is a sequence of positive real numbers and $(h_k)_{k\geq 1}$ is a sequence of measurable functions from $\mathbb{R}^d$ to $\mathbb{R}^d$ which satisfies the following assumption:

AR1. *For all $k \geq 1$, h_k is ϖ_k-Lipschitz.*

The sequence $\{\mathrm{X}_k,\ k \in \mathbb{N}\}$ is an inhomogeneous Markov chain with Markov kernels $(\mathsf{P}_k)_{k\geq 1}$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ given for all $x \in \mathbb{R}^d$ and $\mathtt{A} \in \mathbb{R}^d$ by

$$\mathsf{P}_k(x, \mathtt{A}) = \frac{1}{(2\pi\sigma_k^2)^{d/2}} \int_{\mathtt{A}} \exp\left(-\left\|y - h_k(x)\right\|^2 / (2\sigma_k^2)\right) \mathrm{d}y \,. \tag{31}$$

We denote for all $n \geq 1$ by Q^n the marginal distribution of X_n given by

$$\mathsf{Q}^n = \mathsf{P}_1 \cdots \mathsf{P}_n \,. \tag{32}$$

[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
[5] https://archive.ics.uci.edu/ml/datasets/Musk+(Version+1)

In this section we compute an upper bound of $\|\delta_x \mathsf{Q}^n - \delta_y \mathsf{Q}^n\|_{\mathrm{TV}}$ which does not depend on the dimension d. Define for $x, y \in \mathbb{R}^d$

$$\mathrm{E}_k(x,y) = h_k(y) - h_k(x) \;, \mathrm{e}_k(x,y) = \begin{cases} \mathrm{E}_k(x,y)/\left\|\mathrm{E}_k(x,y)\right\| & \text{if } \mathrm{E}_k(x,y) \neq 0 \\ 0 & \text{otherwise} \;, \end{cases} \tag{33}$$

For all $x, y, z \in \mathbb{R}^d$, $x \neq y$, define

$$\mathrm{F}_k(x,y,z) = h_k(y) + \left(\mathrm{Id} - 2\mathrm{e}_k(x,y)\mathrm{e}_k(x,y)^{\mathrm{T}}\right) z \tag{34}$$

$$\alpha_k(x,y,z) = \frac{\boldsymbol{\varphi}_{\sigma_k^2}(\|\mathrm{E}_k(x,y)\| - \langle \mathrm{e}_k(x,y), z\rangle)}{\boldsymbol{\varphi}_{\sigma_k^2}(\langle \mathrm{e}_k(x,y), z\rangle)} \;, \tag{35}$$

where $\boldsymbol{\varphi}_{\sigma_k^2}$ is the probability density of a zero-mean gaussian variable with variance σ_k^2. Let Z_1 be a standard d-dimensional Gaussian random variable. Set $\mathrm{X}_1 = h_k(x) + \sigma_k \mathrm{Z}_1$ and

$$\mathrm{Y}_1 = \begin{cases} h_k(y) + \sigma_k \mathrm{Z}_1 & \text{if } \mathrm{E}_k(x,y) = 0 \\ B_1\, \mathrm{X}_1 + (1 - B_1)\, \mathrm{F}_k(x,y,\mathrm{Z}_1) & \text{if } \mathrm{E}_k(x,y) \neq 0 \;, \end{cases}$$

where B_1 is a Bernoulli random variable independent of Z_1 with success probability

$$p_k(x,y,z) = 1 \wedge \; \alpha_k(x,y,z) \;.$$

The construction above defines for all $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$ the Markov kernel K_k on $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d))$ given for all $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$ and $\mathtt{A} \in \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d)$ by

$$\begin{aligned} \mathsf{K}_k((x,y),\mathtt{A}) = \frac{\mathbb{1}_{\mathsf{D}}(h_k(x), h_k(y))}{(2\pi\sigma_k^2)^{d/2}} \int_{\mathbb{R}^d} \mathbb{1}_{\mathtt{A}}(\tilde{x},\tilde{x}) \mathrm{e}^{-\|\tau_k(\tilde{x},x)\|^2/(2\sigma_k^2)} \mathrm{d}\tilde{x} \\ + \frac{\mathbb{1}_{\mathsf{D}^{\mathrm{c}}}(h_k(x), h_k(y))}{(2\pi\sigma_k^2)^{d/2}} \left[\int_{\mathbb{R}^d} \mathbb{1}_{\mathtt{A}}(\tilde{x},\tilde{x}) p_k(x,y,\tau_k(\tilde{x},x)) \mathrm{e}^{-\|\tau_k(\tilde{x},x)\|^2/(2\sigma_k^2)} \mathrm{d}\tilde{x} \right. \\ \left. + \int_{\mathbb{R}^d} \mathbb{1}_{\mathtt{A}}(\tilde{x}, \mathrm{F}_k(x,y,\tau_k(\tilde{x},x))) \left\{1 - p_k(x,y,\tau_k(\tilde{x},x))\right\} \mathrm{e}^{-\|\tau_k(\tilde{x},x)\|^2/(2\sigma_k^2)} \mathrm{d}\tilde{x} \right] \;, \end{aligned} \tag{36}$$

where for all $\tilde{x} \in \mathbb{R}^d$, $\tau_k(\tilde{x},x) = \tilde{x} - h_k(x)$ and $\mathsf{D} = \left\{ (\tilde{x},\tilde{y}) \in \mathbb{R}^d \times \mathbb{R}^d \,\middle|\, \tilde{x} = \tilde{y} \right\}$. It is shown in [5, Section 3.3] that for all $x, y \in \mathbb{R}^d$ and $k \geq 1$, $\mathsf{K}_k((x,y),\cdot)$ is a transference plan of $\mathsf{P}_k(x,\cdot)$ and $\mathsf{P}_k(y,\cdot)$. For completeness, the proof is given in Appendix D.1. Furthermore, we have for all $x, y \in \mathbb{R}^d$ and $k \geq 1$

$$\mathsf{K}_k((x,y),\mathsf{D}) = 2\boldsymbol{\Phi}\left(-\frac{\|\mathrm{E}_k(x,y)\|}{2\sigma_k} \right) \;. \tag{37}$$

For all initial distribution μ_0 on $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d))$, $\widetilde{\mathbb{P}}_{\mu_0}$ and $\widetilde{\mathbb{E}}_{\mu_0}$ denote the probability and the expectation respectively, associated with the sequence of Markov kernels $(\mathsf{K}_k)_{k \geq 1}$ defined in (36) and μ_0 on the canonical space $((\mathbb{R}^d \times \mathbb{R}^d)^{\mathbb{N}}, (\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d))^{\otimes \mathbb{N}})$, $\{(\mathrm{X}_i, \mathrm{Y}_i),\ i \in \mathbb{N}\}$ denotes the canonical process and $\{\widetilde{\mathcal{F}}_i,\ i \in \mathbb{N}\}$ the corresponding filtration. Then if $(\mathrm{X}_0, \mathrm{Y}_0) = (x,y) \in \mathbb{R}^d \times \mathbb{R}^d$, for all $k \geq 1$ $(\mathrm{X}_k, \mathrm{Y}_k)$ is a coupling of $\delta_x \mathsf{Q}^k$ and $\delta_y \mathsf{Q}^k$. Using Lindvall's inequality, bounding $\|\delta_x \mathsf{Q}^n - \delta_y \mathsf{Q}^n\|_{\mathrm{TV}}$ amounts to evaluate $\widetilde{\mathbb{P}}_{(x,y)}(\mathrm{X}_n \neq \mathrm{Y}_n)$.

Theorem 19. *Assume* $\boldsymbol{AR}1$*. Then for all* $x, y \in \mathbb{R}^d$ *and* $n \geq 1$*,*

$$\|\delta_x \mathrm{Q}^n - \delta_y \mathrm{Q}^n\|_{\mathrm{TV}} \leq \mathbb{1}_{\mathrm{D}^c}((x,y)) \left\{ 1 - 2\boldsymbol{\Phi}\left(-\frac{\|x-y\|}{2\Xi_n^{1/2}} \right) \right\} ,$$

where $(\Xi_i)_{i\geq 1}$ *is defined for all* $k \geq 1$ *by* $\Xi_k = \sum_{i=1}^k \{\sigma_i^2 / \prod_{j=1}^i \varpi_j^2\}$*.*

We preface the proof by a technical Lemma.

Lemma 20. *For all* $\varsigma, \mathrm{a} > 0$ *and* $t \in \mathbb{R}_+$*, the following identity holds*

$$\int_{\mathbb{R}} \boldsymbol{\varphi}_{\varsigma^2}(y) \left\{ 1 - 1 \wedge \frac{\boldsymbol{\varphi}_{\varsigma^2}(t-y)}{\boldsymbol{\varphi}_{\varsigma^2}(y)} \right\} \left\{ 1 - 2\boldsymbol{\Phi}\left(-\frac{|2y-t|}{2\mathrm{a}} \right) \right\} \mathrm{d}y \\ = 1 - 2\boldsymbol{\Phi}\left(-\frac{t}{2(\varsigma^2 + \mathrm{a}^2)^{1/2}} \right) .$$

Proof. Let $\varsigma, \mathrm{a} > 0$ and $t \in \mathbb{R}_+$. Let us denote by I the integral on the left hand side in the expression above. Then,

$$\begin{aligned} I &= \int_{-\infty}^{t/2} \{\boldsymbol{\varphi}_{\varsigma^2}(y) - \boldsymbol{\varphi}_{\varsigma^2}(t-y)\} \left\{ 1 - 2\boldsymbol{\Phi}\left(\frac{2y-t}{2\mathrm{a}} \right) \right\} \mathrm{d}y \\ &= \int_{-\infty}^{t/2} \boldsymbol{\varphi}_{\varsigma^2}(y) \left\{ 1 - 2\boldsymbol{\Phi}\left(\frac{2y-t}{2\mathrm{a}} \right) \right\} \mathrm{d}y \\ &\qquad - \int_{-\infty}^{-t/2} \boldsymbol{\varphi}_{\varsigma^2}(y) \left\{ 1 - 2\boldsymbol{\Phi}\left(\frac{t+2y}{2\mathrm{a}} \right) \right\} \mathrm{d}y , \end{aligned} \tag{38}$$

Now to simplify the proof, we give a probabilistic interpretation of this two integrals. Let X and Y be two real Gaussian random variables with zero mean and variance a^2 and ς^2 respectively. Since for all $u \in \mathbb{R}_+$, $1 - 2\boldsymbol{\Phi}(-u/(2\mathrm{a})) = \mathbb{P}[|X| \leq u/2]$, we have by (38)

$$I = \mathbb{P}\left(Y \leq t/2, X + Y \leq t/2, Y - X \leq t/2 \right) \\ - \mathbb{P}\left(Y \geq t/2, X + Y \geq t/2, Y - X \geq t/2 \right) .$$

Using that Y and $-Y$ have the same law in the second term, we get $I = I_1 + I_2$ where

$$\begin{aligned} I_1 &= \mathbb{P}\left(Y \leq t/2, X + Y \leq t/2, Y - X \leq t/2, X \geq 0 \right) \\ &\qquad - \mathbb{P}\left(Y \leq -t/2, X - Y \geq t/2, Y + X \leq -t/2, X \geq 0 \right) \\ &= \mathbb{P}\left(|X + Y| \leq t/2, X \geq 0 \right) , \end{aligned} \tag{39}$$

and

$$I_2 = \mathbb{P}\left(Y \leq t/2, X + Y \leq t/2, Y - X \leq t/2, X \leq 0 \right) \\ - \mathbb{P}\left(Y \leq -t/2, X - Y \geq t/2, Y + X \leq -t/2, X \leq 0 \right) .$$

Using again that Y and $-Y$ have the same law in the two terms we have

$$
\begin{aligned}
I_2 = \mathbb{P}\left(Y \geq -t/2, X - Y \leq t/2, Y + X \geq -t/2, X \leq 0\right) \\
- \mathbb{P}\left(Y \geq t/2, X + Y \geq t/2, X - Y \leq -t/2, X \leq 0\right) \\
= \mathbb{P}\left(|X + Y| \leq t/2, X \leq 0\right) \ . \qquad (40)
\end{aligned}
$$

Combining (39), (40), we get $I = \mathbb{P}(|X+Y| \leq t/2)$. The proof follows from the fact that $X+Y$ is a real Gaussian random variable with mean zero and variance $\mathrm{a}^2 + \varsigma^2$, since X and Y are independent.

$\square$

Proof of Theorem 19. Since for all $k \geq 1$, $(\mathrm{X}_k, \mathrm{Y}_k)$ is a coupling of $\delta_x \mathsf{Q}^k$ and $\delta_y \mathsf{Q}^k$, $\|\delta_x \mathsf{Q}^k - \delta_y \mathsf{Q}^k\|_{\mathrm{TV}} \leq \widetilde{\mathbb{P}}_{(x,y)}(\mathrm{X}_k \neq \mathrm{Y}_k)$.

Define for all $k_1, k_2 \in \mathbb{N}^*$, $k_1 \leq k_2$, $\Xi_{k_1,k_2} = \sum_{i=k_1}^{k_2} \{\sigma_i^2 / \prod_{j=k_1}^{i} \varpi_j^2\}$. Let $n \geq 1$. We show by backward induction that for all $k \in \{0, \cdots, n-1\}$,

$$
\widetilde{\mathbb{P}}_{(x,y)}(\mathrm{X}_n \neq \mathrm{Y}_n) \leq \widetilde{\mathbb{E}}_{(x,y)}\left[\mathbb{1}_{\mathsf{D}^c}(\mathrm{X}_k, \mathrm{Y}_k)\left[1 - 2\boldsymbol{\Phi}\left\{-\frac{\|\mathrm{X}_k - \mathrm{Y}_k\|}{2\left(\Xi_{k+1,n}\right)^{1/2}}\right\}\right]\right] \ , \qquad (41)
$$

Note that the inequality for $k = 0$ will conclude the proof.

Since $\mathrm{X}_n \neq \mathrm{Y}_n$ implies that $\mathrm{X}_{n-1} \neq \mathrm{Y}_{n-1}$, the Markov property and (37) imply

$$
\begin{aligned}
\widetilde{\mathbb{P}}_{(x,y)}(\mathrm{X}_n \neq \mathrm{Y}_n) &= \widetilde{\mathbb{E}}_{(x,y)}\left[\mathbb{1}_{\mathsf{D}^c}(\mathrm{X}_{n-1}, \mathrm{Y}_{n-1}) \widetilde{\mathbb{E}}_{(\mathrm{X}_{n-1}, \mathrm{Y}_{n-1})}\left[\mathbb{1}_{\mathsf{D}^c}(\mathrm{X}_1, \mathrm{Y}_1)\right]\right] \\
&\leq \widetilde{\mathbb{E}}_{(x,y)}\left[\mathbb{1}_{\mathsf{D}^c}(\mathrm{X}_{n-1}, \mathrm{Y}_{n-1})\left[1 - 2\boldsymbol{\Phi}\left\{-\frac{\|\mathrm{E}_{n-1}(\mathrm{X}_{n-1}, \mathrm{Y}_{n-1})\|}{2\sigma_n}\right\}\right]\right]
\end{aligned}
$$

Using **AR**1 and (33), $\|\mathrm{E}_n(\mathrm{X}_{n-1}, \mathrm{Y}_{n-1})\| \leq \varpi_n \|\mathrm{X}_{n-1} - \mathrm{Y}_{n-1}\|$, showing (41) holds for $k = n-1$.

Assume that (41) holds for $k \in \{1, \ldots, n-1\}$. On $\{\mathrm{X}_k \neq \mathrm{Y}_k\}$, we have

$$
\|\mathrm{X}_k - \mathrm{Y}_k\| = \left|-\|\mathrm{E}_k(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1})\| + 2\sigma_k \mathrm{e}_k(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1})^{\mathrm{T}} \mathrm{Z}_k\right| \ ,
$$

which implies

$$
\begin{aligned}
&\mathbb{1}_{\mathsf{D}^c}(\mathrm{X}_k, \mathrm{Y}_k)\left[1 - 2\boldsymbol{\Phi}\left\{-\frac{\|\mathrm{X}_k - \mathrm{Y}_k\|}{2\Xi_{k+1,n}^{1/2}}\right\}\right] \\
&= \mathbb{1}_{\mathsf{D}^c}(\mathrm{X}_k, \mathrm{Y}_k)\left[1 - 2\boldsymbol{\Phi}\left\{-\frac{\left|2\sigma_k \mathrm{e}_k(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1})^{\mathrm{T}} \mathrm{Z}_k - \|\mathrm{E}_k(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1})\|\right|}{2\Xi_{k+1,n}^{1/2}}\right\}\right] \ .
\end{aligned}
$$

Since Z_k is independent of $\widetilde{\mathcal{F}}_{k-1}$, $\sigma_k \mathrm{e}_k(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1})^{\mathrm{T}} \mathrm{Z}_k$ is a real Gaussian random variable with zero mean and variance σ_k^2, therefore by Lemma 20, we get

$$\widetilde{\mathbb{E}}^{\widetilde{\mathcal{F}}_{k-1}}_{(x,y)} \left[\mathbb{1}_{\mathrm{D}^{\mathrm{c}}}(\mathrm{X}_k, \mathrm{Y}_k) \left[1 - 2\boldsymbol{\Phi} \left\{ -\frac{\|\mathrm{X}_k - \mathrm{Y}_k\|}{2\Xi_{k+1,n}^{1/2}} \right\} \right] \right] \\ \leq \mathbb{1}_{\mathrm{D}^{\mathrm{c}}}(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1}) \left[1 - 2\boldsymbol{\Phi} \left\{ -\frac{\|\mathrm{E}_k(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1})\|}{2\left(\sigma_k^2 + \Xi_{k+1,n}\right)^{1/2}} \right\} \right] .$$

Using by **AR**1 that $\|\mathrm{E}_k(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1})\| \leq \varpi_k \|\mathrm{X}_{k-1} - \mathrm{Y}_{k-1}\|$ concludes the induction. □

Acknowledgements

The authors would like to thank Arnak Dalalyan for helpful discussions. The work of A.D. and E.M. is supported by the Agence Nationale de la Recherche, under grant ANR-14-CE23-0012 (COSMOS), Initiative Data Science from Ecole Polytechnique and Chaire BayeScale "P. Laffitte".

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] A. N. Borodin and P. Salminen. *Handbook of Brownian motion—facts and formulae*. Probability and its Applications. Birkhäuser Verlag, Basel, second edition, 2002.

[3] 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.

[4] S. Bubeck, R Eldan, and J. Lehec. Finite-time analysis of projected langevin monte carlo. In *Proceedings of the 28th International Conference on Neural Information Processing Systems*, NIPS'15, pages 1243–1251, Cambridge, MA, USA, 2015. MIT Press.

[5] R. Bubley, M. Dyer, and M. Jerrum. An elementary analysis of a procedure for sampling points in a convex body. *Random Structures Algorithms*, 12(3):213–235, 1998.

[6] M. F. Chen and S. F. Li. Coupling methods for multidimensional diffusion processes. *Ann. Probab.*, 17(1):151–177, 1989.

[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] N. Chopin and Ridgway J. Leave Pima Indians alone: binary regression as a benchmark for Bayesian computation. *Statist. Sci.*, 32(1):64–87, 2017.

[9] A. S. Dalalyan. Further and stronger analogy between sampling and optimization: Langevin monte carlo and gradient descent. In *Proceedings of the 30th Annual Conference on Learning Theory.*

[10] A. S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(3):651–676, 2017.

[11] A. Durmus and É. Moulines. Supplement to “high-dimensional bayesian inference via the unadjusted langevin algorithm”, 2015. https://hal.inria.fr/hal-01176084/.

[12] A. Durmus and É. Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.*, 27(3):1551–1587, 2017.

[13] A. Eberle. Quantitative contraction rates for Markov chains on continuous state spaces. In preparation.

[14] A. Eberle. Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields*, pages 1–36, 2015.

[15] A. Eberle, A. Guillin, and R. Zimmer. Quantitative Harris type theorems for diffusions and McKean-Vlasov processes. To appear in Trans. Am. Math. Soc., 2018.

[16] D. L Ermak. A computer simulation of charged particles in solution. i. technique and equilibrium properties. *The Journal of Chemical Physics*, 62(10):4189–4196, 1975.

[17] C. Faes, J. T. Ormerod, and M. P. Wand. Variational Bayesian inference for parametric and nonparametric regression with missing data. *Journal of the American Statistical Association*, 106(495):959–971, 2011.

[18] 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.

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

[20] U. Grenander. Tutorial in pattern theory. Division of Applied Mathematics, Brown University, Providence, 1983.

[21] 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.

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

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

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

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

[26] B. Klartag. A central limit theorem for convex sets. *Invent. Math.*, 168(1):91–131, 2007.

[27] D. Lamberton and G. Pagès. Recursive computation of the invariant distribution of a diffusion. *Bernoulli*, 8(3):367–405, 2002.

[28] 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.

[29] V. Lemaire. *Estimation de la mesure invariante d'un processus de diffusion*. PhD thesis, Université Paris-Est, 2005.

[30] T. Lindvall and L. C. G. Rogers. Coupling of multidimensional diffusions by reflection. *Ann. Probab.*, 14(3):860–872, 1986.

[31] J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.*, 101(2):185–232, 2002.

[32] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, New York, NY, USA, 2nd edition, 2009.

[33] R. M. Neal. Bayesian learning via stochastic dynamics. In *Advances in Neural Information Processing Systems 5, [NIPS Conference]*, pages 475–482, San Francisco, CA, USA, 1993. Morgan Kaufmann Publishers Inc.

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

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

[36] 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.

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

[38] P. J. Rossky, J. D. Doll, and H. L. Friedman. Brownian dynamics as smart Monte Carlo simulation. *The Journal of Chemical Physics*, 69(10):4628–4633, 1978.

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

[40] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4):483–509 (1991), 1990.

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

[42] M. Welling and Y. W. Teh. Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 681–688, 2011.

[43] 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.

A Proofs of Section 2

In this section are gathered the postponed proofs of Section 2. If **H**1 holds, then [34, Theorem 2.1.12, Theorem 2.1.9] show that for all $x, y \in \mathbb{R}^d$:

$$\langle \nabla U(y) - \nabla U(x), y - x \rangle \geq \frac{\kappa}{2} \|y - x\|^2 + \frac{1}{m+L} \|\nabla U(y) - \nabla U(x)\|^2 \ , \tag{42}$$

where

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

A.1 Proof of Proposition 1

(i) The generator $\mathscr{A}$ associated with $(P_t)_{t\geq 0}$ is given, for all $f \in C^2(\mathbb{R}^d)$ and $y \in \mathbb{R}^d$, by:

$$\mathscr{A} f(y) = -\langle \nabla U(y), \nabla f(y) \rangle + \Delta f(y) \ . \tag{43}$$

Denote for all $y \in \mathbb{R}^d$ by $V(y) = \|y - x^\star\|^2$. Let $x \in \mathbb{R}^d$ and $(Y_t)_{t\geq 0}$ be a solution of (1) started at x. Under **H**1 $\sup_{t\in[0,T]} \mathbb{E}[\|Y_t\|^2] < +\infty$ 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. Denote for all $t \geq 0$ and $x \in \mathbb{R}^d$ by $v(t,x) = P_t V(x)$. Then we have, $\partial v(t,x)/\partial t = P_t \mathscr{A} V(x)$.

Since $\nabla U(x^\star) = 0$ and by **H**2, $\langle \nabla U(x) - \nabla U(x^\star), x - x^\star \rangle \geq m \|x - x^\star\|^2$, we have

$$\mathscr{A} V(x) = 2\left(-\langle \nabla U(x) - \nabla U(x^\star), x - x^\star\rangle + d\right) \leq 2\left(-mV(x) + d\right) . \tag{44}$$

Therefore, we get

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

and the proof follows from the Grönwall inequality.

(ii) Set $V(x) = \|x - x^\star\|^2$. By Proposition 1-(i), using that $\pi P_t = \pi$ for all $t > 0$ and that the function $z \mapsto z \wedge c$ is concave for all $c > 0$, we get using the Jensen inequality

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

Using Lebesgue's dominated convergence theorem and taking the limit as $t \to +\infty$, we get $\pi(V \wedge c) \leq d/m$. Using the monotone convergence theorem and taking the limit as $c \to +\infty$ concludes the proof.

(iii) Let $x, y \in \mathbb{R}^d$. Consider the following SDE in $\mathbb{R}^d \times \mathbb{R}^d$:

$$\begin{cases} \mathrm{d}Y_t &= -\nabla U(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t ,\\ \mathrm{d}\tilde{Y}_t &= -\nabla U(\tilde{Y}_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t ,\end{cases} \tag{45}$$

where $(Y_0, \tilde{Y}_0) = (x, y)$. Since ∇U is Lipschitz, then by [25, 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 (45),

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

which implies using **H**2 and Grönwall's inequality that

$$\left\| Y_t - \tilde{Y}_t \right\|^2 \leq \left\| Y_0 - \tilde{Y}_0 \right\|^2 - 2m \int_0^t \left\| Y_s - \tilde{Y}_s \right\|^2 \mathrm{d}s \leq \left\| Y_0 - \tilde{Y}_0 \right\|^2 \mathrm{e}^{-2mt} .$$

Since for all $t \geq 0$, the law of $(Y_t, \tilde{Y}_t)$ is a coupling between $\delta_x P_t$ and $\delta_y P_t$, by definition of W_2, $W_2(\delta_x P_t, \delta_y P_t) \leq \mathbb{E}[\|Y_t - \tilde{Y}_t\|^2]^{1/2}$, which concludes the proof.

(iv) The proof is a direct consequence of (ii) and (iii)

A.2 Proof of Proposition 2

(i) Note that the proof is trivial if $\ell < n$. Therefore we only need to consider the case $\ell \geq n$. For any $\gamma \in (0, 2/(m+L))$, we have for all $x \in \mathbb{R}^d$:

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

Using that $\nabla U(x^\star) = 0$, we get using the previous identity and (42):

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

where we have used for the last inequality that $\gamma \leq 2/(m+L)$. Then by definition (5) of $Q_\gamma^{n,\ell}$ for $\ell, n \geq 1$, $\ell \geq n$, the proof follows from a straightforward induction.

(ii) By (i), we have for all $x \in \mathbb{R}^d$ and $n \geq 1$,

$$\begin{aligned}
\int_{\mathbb{R}^d} \|y - x^\star\|^2 R_\gamma^n(x, \mathrm{d}y) &\leq (1 - \kappa\gamma)^n \|x - x^\star\|^2 + 2d \sum_{k=1}^{n} \gamma (1 - \kappa\gamma)^{n-k} \\
&= (1 - \kappa\gamma)^n \|x - x^\star\|^2 + 2\kappa^{-1} d (1 - (1 - \kappa\gamma)^n) \,. \qquad (46)
\end{aligned}$$

Since any compact set of $\mathbb{R}^d$ is accessible and small for R_γ, then [32, Theorem 15.0.1] implies that R_γ has a unique stationary distribution π_γ. Using (46), the proof is along the same lines as Proposition 1-(ii).

A.3 Proof of Proposition 3

(i) Let $(Z_k)_{k \geq 1}$ be a sequence of i.i.d. d-dimensional Gaussian random variables. For $n \in \mathbb{N}$, define the process $(X_k^{n,1}, X_k^{n,2})_{k \geq 0}$ as follows: $(X_0^{n,1}, X_0^{n,2}) = (x, y)$ and for $k \geq 0$,

$$X_{k+1}^{n,j} = X_k^{n,j} - \gamma_{k+n} \nabla U(X_k^{n,j}) + \sqrt{2\gamma_{k+n}} Z_{k+1} \qquad j = 1, 2 \,. \qquad (47)$$

Note that $X_\ell^{n,1}$ and $X_\ell^{n,2}$ are distributed according to $\delta_x Q_\gamma^{n,\ell}$ and $\delta_y Q_\gamma^{n,\ell}$ respectively. Therefore by definition of the Wasserstein distance of order 2, we get for any $\ell \geq n \geq 1$. $W_2^2(\delta_x Q_\gamma^{n,\ell}, \delta_y Q_\gamma^{n,\ell}) \leq \mathbb{E}[\|X_\ell^{n,1} - X_\ell^{n,2}\|^2]$ and (42) implies for $k \geq n - 1$,

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

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

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

(ii) Let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. For all $n \geq 0$, $\mu R_\gamma^n \in \mathcal{P}_2(\mathbb{R}^d)$. Then, by Proposition 3-(i) for $\gamma \leq 2/(m+L)$, R_γ is a strict contraction in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ and there is a unique fixed point π_γ which is the unique invariant distribution.

A.4 Proof of Theorem 5

We preface the proof by a technical Lemma.

Lemma 21. *Let $(Y_t)_{t\geq 0}$ be the solution of (1) started at $x \in \mathbb{R}^d$. For all $t \geq 0$ and $x \in \mathbb{R}^d$,*

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

Proof. Let $\mathscr{A}$ be the generator associated with $(P_t)_{t\geq 0}$ defined by (44). Denote for all $x, y \in \mathbb{R}^d$, $\tilde{V}_x(y) = \|y - x\|^2$. Note that 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. 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,

$$\frac{\partial \tilde{v}(t,x)}{\partial t} = P_t \mathscr{A} \tilde{V}_x(x) . \tag{48}$$

By **H**2, we have for all $y \in \mathbb{R}^d$, $\langle \nabla U(y) - \nabla U(x), y - x\rangle \geq m\|x - y\|^2$, which implies

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

Using (48), this inequality and that $\tilde{V}_x$ is positive, we get

$$\frac{\partial \tilde{v}(t,x)}{\partial t} = P_t \mathscr{A}\tilde{V}_x(x) \leq 2\left(d - \int_{\mathbb{R}^d} \langle \nabla U(x), y - x\rangle P_t(x, \mathrm{d}y)\right) . \tag{49}$$

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

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

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

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

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 this upper bound in (49), and since $2\|x - x^\star\| t^{3/2} d^{1/2} \leq t\|x - x^\star\|^2 + t^2 d$, we get

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

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

To show Theorem 5 and Theorem 8, since π is invariant for P_t for all $t \geq 0$, it suffices to get some bounds on $W_2(\delta_x Q_\gamma^n, \nu_0 P_{\Gamma_n})$, 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. An obvious candidate is the synchronous coupling $(Y_t, \overline{Y}_t)_{t \geq 0}$ defined for all $n \geq 0$ and $t \in [\Gamma_n, \Gamma_{n+1})$ by

$$\begin{cases} Y_t = Y_{\Gamma_n} - \int_{\Gamma_n}^t \nabla U(Y_s) \mathrm{d}s + \sqrt{2}(B_t - B_{\Gamma_n}) \\ \bar{Y}_t = \bar{Y}_{\Gamma_n} - \nabla U(\bar{Y}_{\Gamma_n})(t - \Gamma_n) + \sqrt{2}(B_t - B_{\Gamma_n}) \ , \end{cases} \tag{50}$$

with Y_0 is distributed according to ν_0, $\bar{Y}_0 = x$ and $(\Gamma_n)_{n \geq 1}$ is given in (3). Therefore since for all $n \geq 0$, $W_2^2(\delta_x Q_\gamma^n, \nu_0 P_{\Gamma_n}) \leq \mathbb{E}[\|Y_{\Gamma_n} - \bar{Y}_{\Gamma_n}\|^2]$, taking $\nu_0 = \pi$, we derive an explicit bound on the Wasserstein distance between the sequence of distributions $(\delta_x Q_\gamma^k)_{k \geq 0}$ and the stationary measure π of the Langevin diffusion (1).

Let $(\mathcal{F}'_t)_{t \geq 0}$ be the filtration associated with $(B_t)_{t \geq 0}$ and $(Y_0, \overline{Y}_0)$.

Lemma 22. *Assume* ***H*** *1 and* ***H*** *2. Let $(\gamma_k)_{k \geq 1}$ be a non-increasing 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 (50). Then almost surely for all $n \geq 0$ and $\epsilon > 0$,*

$$\begin{aligned} \left\|Y_{\Gamma_{n+1}} - \overline{Y}_{\Gamma_{n+1}}\right\|^2 &\leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon)\} \left\|Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}\right\|^2 \\ &\quad + (2\gamma_{n+1} + (2\epsilon)^{-1}) \int_{\Gamma_n}^{\Gamma_{n+1}} \|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2 \mathrm{d}s \ , \end{aligned} \tag{51}$$

$$\begin{aligned} \mathbb{E}^{\mathcal{F}'_{\Gamma_n}}\left[\left\|Y_{\Gamma_{n+1}} - \overline{Y}_{\Gamma_{n+1}}\right\|^2\right] &\leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon)\} \left\|Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}\right\|^2 \\ &\quad + L^2 \gamma_{n+1}^2 (1/(4\epsilon) + \gamma_{n+1}) \left(2d + L^2 \gamma_{n+1} \|Y_{\Gamma_n} - x^\star\|^2 + dL^2 \gamma_{n+1}^2/6\right) \ . \end{aligned} \tag{52}$$

Proof. We first show (51). Set $\Theta_n = Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}$. By definition we have:

$$\begin{aligned} \|\Theta_{n+1}\|^2 &= \|\Theta_n\|^2 + \left\| \int_{\Gamma_n}^{\Gamma_{n+1}} \{\nabla U(Y_s) - \nabla U(\overline{Y}_{\Gamma_n})\} \mathrm{d}s \right\|^2 \\ &- 2\gamma_{n+1} \langle \Theta_n, \nabla U(Y_{\Gamma_n}) - \nabla U(\overline{Y}_{\Gamma_n}) \rangle - 2 \int_{\Gamma_n}^{\Gamma_{n+1}} \langle \Theta_n, \{\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\} \rangle \mathrm{d}s \ . \end{aligned} \tag{53}$$

Young's inequality and Jensen's inequality imply

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

Using (42), $\gamma_1 \leq 1/(m+L)$ and $(\gamma_k)_{k \geq 1}$ is non-increasing, (53) becomes

$$\|\Theta_{n+1}\|^2 \leq \{1 - \gamma_{n+1}\kappa\} \|\Theta_n\|^2 + 2\gamma_{n+1} \int_{\Gamma_n}^{\Gamma_{n+1}} \|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2 \, \mathrm{d}s - 2 \int_{\Gamma_n}^{\Gamma_{n+1}} \langle \Theta_n, \{\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\} \rangle \, \mathrm{d}s \; . \tag{54}$$

Using the inequality $|\langle a, b \rangle| \leq \epsilon \|a\|^2 + (4\epsilon)^{-1} \|b\|^2$ concludes the proof of (51).

We now prove (52). Note that (51) implies that

$$\mathbb{E}^{\mathcal{F}'_{\Gamma_n}} \left[\|\Theta_{n+1}\|^2 \right] \leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon)\} \|\Theta_n\|^2 + (2\gamma_{n+1} + (2\epsilon)^{-1}) \int_{\Gamma_n}^{\Gamma_{n+1}} \mathbb{E}^{\mathcal{F}'_{\Gamma_n}} \left[\|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2 \right] \mathrm{d}s \; . \tag{55}$$

By **H**1, the Markov property of $(Y_t)_{t \geq 0}$ and Lemma 21, we have

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

The proof is then concluded plugging this bound in (55) . □

Proof of Theorem 5. Let $x \in \mathbb{R}^d$, $n \geq 1$ and $\zeta_0 = \pi \otimes \delta_x$. Let $(Y_t, \overline{Y}_t)_{t \geq 0}$ with $(Y_0, \overline{Y}_0)$ distributed according to ζ_0 and defined by (50). By definition of W_2 and since for all $t \geq 0$, π is invariant for P_t, $W_2^2(\mu_0 Q^n, \pi) \leq \mathbb{E}\left[\left\| Y_{\Gamma_n} - \overline{Y}_{\Gamma_n} \right\|^2 \right]$. Lemma 22 with $\epsilon = \kappa/4$, Proposition 1-(i) imply, using a straightforward induction, that for all $n \geq 0$

$$\mathbb{E}\left[\left\| Y_{\Gamma_n} - \overline{Y}_{\Gamma_n} \right\|^2 \right] \leq u_n^{(1)}(\gamma) \int_{\mathbb{R}^d} \|y - x\|^2 \, \pi(\mathrm{d}y) + A_n(\gamma) \; , \tag{56}$$

where $(u_n^{(1)}(\gamma))_{n \geq 1}$ is given by (8), and

$$A_n(\gamma) = L^2 \sum_{i=1}^n \gamma_i^2 \left\{ \kappa^{-1} + \gamma_i \right\} (2d + dL^2 \gamma_i^2/6) \prod_{k=i+1}^n (1 - \kappa\gamma_k/2) + L^4 \sum_{i=1}^n \tilde{\delta}_i \gamma_i^3 \left\{ \kappa^{-1} + \gamma_i \right\} \prod_{k=i+1}^n (1 - \kappa\gamma_k/2) \tag{57}$$

with

$$\tilde{\delta}_i = \mathrm{e}^{-2m\Gamma_{i-1}} \mathbb{E}\left[\|Y_0 - x^\star\|^2\right] + (1 - \mathrm{e}^{-2m\Gamma_{i-1}})(d/m) \le d/m \; .$$

Since Y_0 is distributed according to π, Proposition 1-(ii) shows that for all $i \in \{1, \cdots, n\}$,

$$\tilde{\delta}_i \le d/m \; . \tag{58}$$

In addition since for all $y \in \mathbb{R}^d$, $\|x - y\|^2 \le 2(\|x - x^\star\|^2 + \|x^\star - y\|^2)$, using Proposition 1-(ii), we get $\int_{\mathbb{R}^d} \|y - x\|^2 \pi(\mathrm{d}y) \le \|x - x^\star\|^2 + d/m$. Plugging this result, (58) and (57) in (56) completes the proof. □

A.5 Proof of Corollary 6

We preface the proof by a technical lemma.

Lemma 23. *Let $(\gamma_k)_{k\ge1}$ be a sequence of non-increasing real numbers, $\varpi > 0$ and $\gamma_1 < \varpi^{-1}$. Then for all $n \ge 0$, $j \ge 1$ and $\ell \in \{1, \dots, n+1\}$,*

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

Proof. Let $\ell \in \{1, \dots, n+1\}$. Since $(\gamma_k)_{k\ge1}$ is non-increasing and $\gamma_1 < \varpi^{-1}$,

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

□

Proof of Corollary 6. By Theorem 5, it suffices to show that $u_n^{(1)}(\gamma)$ and $u_n^{(2)}(\gamma)$, defined by (8) and (9) respectively, goes to 0 as $n \to +\infty$. Using the bound $1 + t \le \mathrm{e}^t$ for $t \in \mathbb{R}$, and $\lim_{n\to+\infty} \Gamma_n = +\infty$, we have $\lim_{n\to+\infty} u_n^{(1)}(\gamma) = 0$. Since $(\gamma_k)_{k\ge0}$ is non-increasing, note that to show that $\lim_{n\to+\infty} u_n^{(2)}(\gamma) = 0$, it suffices to prove $\lim_{n\to+\infty} \sum_{i=1}^n \prod_{k=i+1}^n (1 - \kappa\gamma_k/2)\gamma_i^2 = 0$. But since $(\gamma_k)_{k\ge1}$ is non-increasing, there exists $c \ge 0$ such that $c\Gamma_n \le n - 1$ and by Lemma 23 applied with $\ell = \lfloor c\Gamma_n \rfloor$ the integer part of $c\Gamma_n$:

$$\sum_{i=1}^{n} \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2)\, \gamma_i^2 \le 2\kappa^{-1}\gamma_{\lfloor c\Gamma_n \rfloor} + \exp\left(-2^{-1}\kappa(\Gamma_n - \Gamma_{\lfloor c\Gamma_n \rfloor})\right) \sum_{i=1}^{\lfloor c\Gamma_n \rfloor - 1} \gamma_i \; . \tag{59}$$

Since $\lim_{k\to+\infty}\gamma_k = 0$, by the Cesáro theorem, we have $\lim_{n\to+\infty} n^{-1}\Gamma_n = 0$. Then using that $\lim_{n\to+\infty}\Gamma_n = +\infty$, we get $\lim_{n\to+\infty}\Gamma_{\lfloor c\Gamma_n\rfloor}/\Gamma_n = 0$, and the conclusion follows from combining in (59), this limit, $\lim_{k\to+\infty}\gamma_k = 0$, $\lim_{n\to+\infty}\Gamma_n = +\infty$ and $\sum_{i=1}^{\lfloor c\Gamma_n\rfloor -1}\gamma_i \leq c\gamma_1\Gamma_n$. □

A.6 Proofs of Theorem 8

Lemma 24. *Assume* **H1**, **H2** *and* **H3**. *Let* $(\gamma_k)_{k\geq 1}$ *be a non-increasing sequence with* $\gamma_1 \leq 1/(m+L)$. *Let* $\zeta_0 \in \mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^d)$ *and* $(Y_t,\overline{Y}_t)_{t\geq 0}$ *be defined by* (50) *such that* $(Y_0,\overline{Y}_0)$ *is distributed according to* ζ_0. *Then for all* $n\geq 0$ *and* $\epsilon>0$, *almost surely*

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

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 $s\in[\Gamma_n,\Gamma_{n+1})$,

$$\begin{aligned}\nabla U(Y_s)-\nabla U(Y_{\Gamma_n}) = &\int_{\Gamma_n}^s\left\{\nabla^2U(Y_u)\nabla U(Y_u)+\vec{\Delta}(\nabla U)(Y_u)\right\}\mathrm{d}u\\ &+\sqrt{2}\int_{\Gamma_n}^s\nabla^2U(Y_u)\mathrm{d}B_u\,.\end{aligned}\tag{60}$$

Since Θ_n is $\mathcal{F}_{\Gamma_n}$-measurable and $(\int_0^s\nabla^2U(Y_u)\mathrm{d}B_u)_{s\in[0,\Gamma_{n+1}]}$ is a $(\mathcal{F}_s)_{s\in[0,\Gamma_{n+1}]}$-martingale under **H1**, by (60) we have:

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

Combining this equality and $|\langle a,b\rangle|\leq\epsilon\|a\|^2+(4\epsilon)^{-1}\|b\|^2$ in (54) we have

$$\begin{aligned}\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\|\Theta_{n+1}\|^2\right]\leq&\{1-\gamma_{n+1}(\kappa-2\epsilon)\}\|\Theta_n\|^2+(2\epsilon)^{-1}A\\ &+2\gamma_{n+1}\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\int_{\Gamma_n}^{\Gamma_{n+1}}\|\nabla U(Y_s)-\nabla U(Y_{\Gamma_n})\|^2\mathrm{d}s\right]\,,\end{aligned}\tag{61}$$

where

$$A=\int_{\Gamma_n}^{\Gamma_{n+1}}\left\|\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\int_{\Gamma_n}^s\nabla^2U(Y_u)\nabla U(Y_u)+\vec{\Delta}(\nabla U)(Y_u)\mathrm{d}u\right]\right\|^2\mathrm{d}s\,.$$

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 21, we have

$$\int_{\Gamma_n}^{\Gamma_{n+1}} \mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left\|\nabla U(Y_s)-\nabla U(Y_{\Gamma_n})\right\|^2\right]\mathrm{d}s \\ \leq L^2\left(d\gamma_{n+1}^2 + dL^2\gamma_{n+1}^4/12 + (1/2)L^2\gamma_{n+1}^3\left\|Y_{\Gamma_n}-x^\star\right\|^2\right) . \quad (62)$$

We now bound A. We get using Jensen's inequality, Fubini's theorem, $\nabla U(x^\star)=0$ and (10)

$$\begin{aligned} A &\leq 2\int_{\Gamma_n}^{\Gamma_{n+1}}(s-\Gamma_n)\int_{\Gamma_n}^{s}\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left\|\nabla^2 U(Y_u)\nabla U(Y_u)\right\|^2\right]\mathrm{d}u\,\mathrm{d}s \\ &\qquad + 2\int_{\Gamma_n}^{\Gamma_{n+1}}(s-\Gamma_n)\int_{\Gamma_n}^{s}\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left\|\vec{\Delta}(\nabla U)(Y_u)\right\|^2\right]\mathrm{d}u\,\mathrm{d}s \\ &\leq 2\int_{\Gamma_n}^{\Gamma_{n+1}}(s-\Gamma_n)L^4\int_{\Gamma_n}^{s}\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left\|Y_u-x^\star\right\|^2\right]\mathrm{d}u\,\mathrm{d}s + 2\gamma_{n+1}^3 d^2\tilde{L}^2/3 . \end{aligned} \quad (63)$$

By Lemma 21-(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 (63) and for all $t\geq 0$, $1-\mathrm{e}^{-t}\leq t$, we get

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

Combining this bound and (62) in (61) concludes the proof. □

Proof of Theorem 8. The proof of is along the same lines as Theorem 5, using Lemma 24 in place of Lemma 22. □

B Proofs of Section 3

In this section are gathered the postponed proofs of Section 3.

B.1 Proof of Theorem 12

Applying Lemma 22 or Lemma 24, we get that for all $x\in\mathbb{R}^d$

$$W_1(\delta_x Q_\gamma^n, \delta_x P_{\Gamma_n}) \leq \{\vartheta_n(x)\}^{1/2} \ , \vartheta_n(x) = \begin{cases} \vartheta_{n,0}^{(1)}(x) & (\mathbf{H}1,\ \mathbf{H}2) , \\ \vartheta_{n,0}^{(2)}(x) & (\mathbf{H}1,\ \mathbf{H}2,\ \mathbf{H}3) , \end{cases} \quad (64)$$

By the triangle inequality, we get

$$\left\| \delta_x P_{\Gamma_\ell} - \delta_x Q_\gamma^\ell \right\|_{\mathrm{TV}} \leq \left\| \left\{ \delta_x P_{\Gamma_n} - \delta_x Q_\gamma^n \right\} P_{\Gamma_{n+1,\ell}} \right\|_{\mathrm{TV}} + \left\| \delta_x Q_\gamma^n \left\{ Q_\gamma^{n+1,\ell} - P_{\Gamma_{n+1,\ell}} \right\} \right\|_{\mathrm{TV}} . \tag{65}$$

Using (14) and (64), we have

$$\left\| \left\{ \delta_x P_{\Gamma_n} - \delta_x Q_\gamma^{1,n} \right\} P_{\Gamma_{n+1,\ell}} \right\|_{\mathrm{TV}} \leq (\vartheta_n(x)/(4\pi \Gamma_{n+1,\ell}))^{1/2} . \tag{66}$$

For the second term, by [12, Equation 15] (note that we have a different convention for the total variation distance) and the Pinsker inequality, we have

$$\begin{aligned} &\left\| \delta_x Q_\gamma^{1,n} \left\{ Q_\gamma^{n+1,\ell} - P_{\Gamma_{n+1,\ell}} \right\} \right\|_{\mathrm{TV}}^2 \\ &\qquad \leq 2^{-3} L^2 \sum_{k=n+1}^{\ell} \left\{ (\gamma_k^3/3) \int_{\mathbb{R}^d} \left\| \nabla U(z) - \nabla U(x^\star) \right\|^2 Q_\gamma^{k-1}(x, \mathrm{d}z) + d\gamma_k^2 \right\} . \end{aligned}$$

By **H**1 and Proposition 2, we get

$$\left\| \delta_x Q_\gamma^{1,n} \left\{ Q_\gamma^{n+1,\ell} - P_{\Gamma_{n+1,\ell}} \right\} \right\|_{\mathrm{TV}}^2 \leq 2^{-3} L^2 \sum_{k=n+1}^{\ell} \left\{ (\gamma_k^3 L^2/3) \varrho_{1,k-1}(x) + d\gamma_k^2 \right\} .$$

Combining the last inequality and (66) in (65) concludes the proof.

B.2 Proof of (24)

Consider the constant sequence $\gamma_k = \gamma$ for all $k \in \mathbb{N}^*$ with $\gamma \in (0, 1/(m+L)]$. By (18), we have for all $n \in \mathbb{N}^*$ and $x \in \mathbb{R}^d$

$$\vartheta_{n,0}^{(1)}(x) \leq \gamma \mathsf{D}_1(\gamma, d) + \gamma^3 \mathsf{D}_2(\gamma) \sum_{i=1}^{n} (1 - \kappa\gamma/2)^{n-i} \delta_{i,n,0}(x) ,$$

where

$$\mathsf{D}_1(\gamma, d) = 2L^2 \kappa^{-1} \left(\kappa^{-1} + \gamma\right) \left(2d + L^2\gamma^2/6\right) , \quad \mathsf{D}_2(\gamma) = L^4 \left(\kappa^{-1} + \gamma\right) .$$

In addition, using that $\kappa \geq 2m$ and for all $t \geq 0$, $1 - t \leq \mathrm{e}^{-t}$,

$$\begin{aligned} \sum_{i=1}^{n} (1 - \kappa\gamma/2)^{n-i} \delta_{i,n,0}(x) &= \sum_{i=1}^{n} \left[(1 - \kappa\gamma/2)^{n-i} \left\{ \mathrm{e}^{-2m\gamma(i-1)} \left\| x - x^\star \right\|^2 \right.\right. \\ &\qquad \left.\left. + \left(1 - \mathrm{e}^{-2m\gamma(i-1)}\right) (d/m) \right\} \right] \\ &\leq n \mathrm{e}^{-m\gamma(n-1)} \left\| x - x^\star \right\|^2 + 2d(\kappa\gamma m)^{-1} . \end{aligned} \tag{67}$$

Therefore for all $n \geq 1$ and $x \in \mathbb{R}^d$ we get

$$\vartheta_{n,0}^{(1)}(x) \leq \gamma \mathsf{D}_1(\gamma) + \gamma^3 \mathsf{D}_2(\gamma) \left\{ n \mathrm{e}^{-m\gamma(n-1)} \left\| x - x^\star \right\|^2 + 2d(\kappa\gamma m)^{-1} \right\} . \tag{68}$$

Let now $\ell \in \mathbb{N}^*$, $\ell \geq \lceil \gamma^{-1} \rceil + 1$ and $n = \ell - \lceil \gamma^{-1} \rceil$. Then,

$$\begin{aligned}
&\sum_{k=n+1}^{\ell} \{(\gamma_k^3 L^2/3)\varrho_{1,k-1}(x) + d\gamma_k^2\} \\
&\leq (L^2\gamma^3/3)\left\{(1-\kappa\gamma)^n(\ell-n)\left\|x-x^\star\right\|^2 + 2\kappa^{-1}\gamma d(\ell-n)\right\} + d\gamma^2(\ell-n) \\
&\leq (L^2\gamma^3/3)\left\{(1+\gamma^{-1})(1-\kappa\gamma)^{\ell-\lceil\gamma^{-1}\rceil}\left\|x-x^\star\right\|^2 + 2(1+\gamma)\kappa^{-1}d\right\} + d\gamma(1+\gamma)\ .
\end{aligned}$$

Combining this inequality and (68) in the bound given by Theorem 12 shows (24).

B.3 Proof of Theorem 14

We preface the proof by a preliminary lemma. Define for all $\gamma > 0$, the function $\mathsf{n} : \mathbb{R}_+^* \to \mathbb{N}$ by

$$\mathsf{n}(\gamma) = \left\lceil \log\left(\gamma^{-1}\right) / \log(2) \right\rceil\ . \tag{69}$$

Lemma 25. *Assume **H**1, **H**2 and **H**3. Let $\gamma \in (0, 1/(m+L))$. Then for all $x \in \mathbb{R}^d$ and $\ell \in \mathbb{N}^*$, $\ell > 2^{\mathsf{n}(\gamma)}$,*

$$\begin{aligned}
\|\delta_x P_{\ell\gamma} - \delta_x R_\gamma^\ell\|_{\mathrm{TV}} &\leq (\vartheta^{(2)}_{\ell-2^{\mathsf{n}(\gamma)},0}(x)/(\pi 2^{\mathsf{n}(\gamma)+2}\gamma))^{1/2} \\
&\quad + 2^{-3/2}L\left\{(\gamma^3L^2/3)\varrho_{1,\ell-1}(x) + d\gamma^2\right\}^{1/2} + \sum_{k=1}^{\mathsf{n}(\gamma)}(\vartheta^{(2)}_{2^{k-1},\ell-2^k}(x)/(\pi 2^{k+1}\gamma))^{1/2}\ .
\end{aligned}$$

where $\varrho_{1,\ell-1}(x)$ is defined by (7) and for all $n_1, n_2 \in \mathbb{N}$, $\vartheta^{(2)}_{n_1,n_2}$ is given by (19).

Proof. Let $\gamma \in (0, 1/(m+L))$ and $\ell \in \mathbb{N}^*$. For ease of notation, let $n = \mathsf{n}(\gamma)$, and assume that $\ell > 2^n$. Consider the following decomposition

$$\begin{aligned}
\left\|\delta_x P_{\ell\gamma} - \delta_x R_\gamma^\ell\right\|_{\mathrm{TV}} &\leq \left\|\left\{\delta_x P_{(\ell-2^n)\gamma} - \delta_x R_\gamma^{\ell-2^n}\right\} P_{2^n\gamma}\right\|_{\mathrm{TV}} \\
&\quad + \left\|\delta_x R_\gamma^{\ell-1}\left\{P_\gamma - R_\gamma\right\}\right\|_{\mathrm{TV}} + \sum_{k=1}^{n}\left\|\delta_x R_\gamma^{\ell-2^k}\left\{P_{2^{k-1}\gamma} - R_\gamma^{2^{k-1}}\right\} P_{2^{k-1}\gamma}\right\|_{\mathrm{TV}}\ .
\end{aligned} \tag{70}$$

We bound each term in the right hand side. First by (14) and Equation (64), we have

$$\left\|\left\{\delta_x P_{(\ell-2^n)\gamma} - \delta_x R_\gamma^{\ell-2^n}\right\} P_{2^n\gamma}\right\|_{\mathrm{TV}} \leq (\vartheta^{(2)}_{\ell-2^n,0}(x)/(\pi 2^{n+2}\gamma))^{1/2}\ , \tag{71}$$

where $\vartheta^{(2)}_{n,0}(x)$ is given by (19). Similarly but using in addition Proposition 2, we have for all $k \in \{1, \cdots, n\}$,

$$\left\|\delta_x R_\gamma^{\ell-2^k}\left\{P_{2^{k-1}\gamma} - R_\gamma^{2^{k-1}}\right\} P_{2^{k-1}\gamma}\right\|_{\mathrm{TV}} \leq (\vartheta^{(2)}_{2^{k-1},\ell-2^k}(x)/(\pi 2^{k+1}\gamma))^{1/2}\ , \tag{72}$$

where $\vartheta^{(2)}_{2^{k-1},\ell-2^k}(x)$ is given by (19). For the last term, by [10, Equation 11] and the Pinsker inequality, we have

$$\left\|\delta_x R_\gamma^{\ell-1}\left\{P_\gamma - R_\gamma\right\}\right\|_{\mathrm{TV}}^2 \leq 2^{-3}L^2\left\{(\gamma^3/3)\int_{\mathbb{R}^d}\|\nabla U(z)\|^2 R_\gamma^{\ell-1}(x,\mathrm{d}z) + d\gamma^2\right\} .$$

By **H**1 and Proposition 2, we get

$$\left\|\delta_x R_\gamma^{\ell-1}\left\{R_\gamma - P_\gamma\right\}\right\|_{\mathrm{TV}}^2 \leq 2^{-3}L^2\left\{(\gamma^3 L^2/3)\varrho_{1,\ell-1}(x) + d\gamma^2\right\} . \tag{73}$$

Combining (71), (72) and (73) in (70) concludes the proof. □

Proof of Theorem 14. First for all $n \geq 1$ and $x \in \mathbb{R}^d$, we have

$$\vartheta^{(2)}_{n,0}(x) \leq \gamma^2 \mathsf{E}_1(\gamma,d) + \gamma^3\mathsf{E}_2(\gamma)\sum_{i=1}^n\prod_{k=i+1}^n(1-\kappa\gamma_k/2)\delta_{i,n,0}(x) ,$$

where

$$\mathsf{E}_1(\gamma,d) = 2d\kappa^{-1}\left\{2L^2 + 4\kappa^{-1}(d\tilde{L}^2/3 + \gamma L^4/4) + \gamma^2 L^4/6\right\} , \mathsf{E}_2(\gamma) = L^4(4\kappa^{-1}/3+\gamma) .$$

By (67), we get for all $n \geq 1$ and $x \in \mathbb{R}^d$,

$$\vartheta^{(2)}_{n,0}(x) \leq \gamma^2\mathsf{E}_1(\gamma,d) + \gamma^3\mathsf{E}_2(\gamma)\left\{n\mathrm{e}^{-m\gamma(n-1)}\|x-x^\star\|^2 + 2d(\kappa\gamma m)^{-1}\right\} . \tag{74}$$

On the other hand, for all $x \in \mathbb{R}^d$, $\ell, n \in \mathbb{N}$, $n \geq 1$, $\ell > n$ we have using that $\kappa \geq 2m$ and for all $t \geq 0$, $1-t \leq \mathrm{e}^{-t}$,

$$\begin{aligned}
\vartheta^{(2)}_{n,\ell}(x) &\leq \gamma^3 n\mathsf{E}_1(\gamma) + \gamma^3 n\mathsf{E}_2(\gamma)\left\{\mathrm{e}^{-m\gamma(n-1)}\varrho_{n,\ell}(x) + d/m\right\} \\
&\leq \gamma^3 n\mathsf{E}_1(\gamma) + \gamma^3 n\mathsf{E}_2(\gamma)\left\{\mathrm{e}^{-m\gamma(n-1)}\left((1-\kappa\gamma)^{\ell-n}\|x-x^\star\|^2 + 2\kappa^{-1}d\right) + d/m\right\} \\
&\leq \gamma^3 n\mathsf{E}_1(\gamma) + \gamma^3 n\mathsf{E}_2(\gamma)\left\{\mathrm{e}^{-m\gamma(\ell-1)}\|x-x^\star\|^2 + 2\kappa^{-1}d + d/m\right\} .
\end{aligned} \tag{75}$$

Finally, for all $\ell \in \mathbb{N}^*$ and $x \in \mathbb{R}^d$, we have

$$(\gamma^3L^2/3)\varrho_{1,\ell-1}(x) + d\gamma^2 \leq (\gamma^3L^2/3)\left\{(1-\kappa\gamma)^{\ell-1}\|x-x^\star\|^2 + 2d\kappa^{-1}\right\} + d\gamma^2 . \tag{76}$$

Combining (74), (75) and (76) in the bound given by Lemma 25, and using that

$\gamma^{-1} \leq 2^{\mathsf{n}(\gamma)} \leq 2\gamma^{-1}$ we have for all $\ell \in \mathbb{N}^*$, $\ell > 2^{\mathsf{n}(\gamma)}$,

$$\|\delta_x P_{\ell\gamma} - \delta_x R_\gamma^\ell\|_{\mathrm{TV}} \leq 2^{-3/2} L \left[(\gamma^3 L^2/3)\left\{(1-\kappa\gamma)^{\ell-1}\|x-x^\star\|^2 + 2d\kappa^{-1}\right\} + d\gamma^2\right]^{1/2}$$
$$+ (4\pi)^{-1/2}\left[\gamma^2 \mathsf{E}_1(\gamma) + \gamma^3 \mathsf{E}_2(\gamma)\left\{(\ell-\gamma^{-1})\mathrm{e}^{-m\gamma(\ell-2\gamma^{-1}-1)}\|x-x^\star\|^2 + 2d(\kappa\gamma m)^{-1}\right\}\right]^{1/2}$$
$$+ \sum_{k=1}^{\mathsf{n}(\gamma)}\left[\frac{\gamma^3 2^{k-1}\mathsf{E}_1(\gamma) + \gamma^3 2^{k-1}\mathsf{E}_2(\gamma)\left\{\mathrm{e}^{-m\gamma(\ell-2^k-1)}\|x-x^\star\|^2 + 2\kappa^{-1}d + d/m\right\}}{\pi 2^{k+1}\gamma}\right]^{1/2}$$
$$\leq 2^{-3/2} L \left\{(\gamma^3 L^2/3)\left\{(1-\kappa\gamma)^{\ell-1}\|x-x^\star\|^2 + 2d\kappa^{-1}\right\} + d\gamma^2\right\}^{1/2}$$
$$+ (4\pi)^{-1/2}\left[\gamma^2 \mathsf{E}_1(\gamma) + \gamma^3 \mathsf{E}_2(\gamma)\left\{(\ell-\gamma^{-1})\mathrm{e}^{-m\gamma(\ell-2\gamma^{-1}-1)}\|x-x^\star\|^2 + 2d(\kappa\gamma m)^{-1}\right\}\right]^{1/2}$$
$$+ (4\pi)^{-1/2}\mathsf{n}(\gamma)\left[\gamma^2 \mathsf{E}_1(\gamma) + \gamma^2 \mathsf{E}_2(\gamma)\left\{\mathrm{e}^{-m\gamma(\ell-2\gamma^{-1}-1)}\|x-x^\star\|^2 + 2\kappa^{-1}d + d/m\right\}\right]^{1/2} .$$

Letting ℓ go to infinity, using Theorem 10-(iii) and Theorem 11-(iii), we get the desired conclusion. □

C Proof of Section 4

In this section are gathered the postponed proofs of Section 4.

C.1 Proof of Theorem 15 and Theorem 16

Our main tool in the proof of Theorem 15 and Theorem 16 is the Gaussian Poincaré inequality [3, Theorem 3.20] which can be applied to $R_\gamma(y,\cdot)$ defined by (4), noticing that $R_\gamma(y,\cdot)$ is a Gaussian distribution with mean $y - \gamma\nabla U(y)$ and covariance matrix $2\gamma\,\mathrm{I}_d$: for all Lipschitz function $g : \mathbb{R}^d \to \mathbb{R}$

$$R_\gamma\left\{g(\cdot) - R_\gamma g(y)\right\}^2(y) \leq 2\gamma\,\|g\|_{\mathrm{Lip}}^2 . \tag{77}$$

To go further, we decompose $\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)]$, for $f : \mathbb{R}^d \to \mathbb{R}$, Lipschitz or measurable and bounded, as the sum of martingale increments, w.r.t. $(\mathcal{G}_n)_{n\geq 0}$, the natural filtration associated with Euler approximation $(X_n)_{n\geq 0}$, and we get

$$\mathrm{Var}_x\left\{\hat{\pi}_n^N(f)\right\} = \sum_{k=N}^{N+n-1}\mathbb{E}_x\left[\left(\mathbb{E}_x^{\mathcal{G}_{k+1}}\left[\hat{\pi}_n^N(f)\right] - \mathbb{E}_x^{\mathcal{G}_k}\left[\hat{\pi}_n^N(f)\right]\right)^2\right] + \mathbb{E}_x\left[\left(\mathbb{E}_x^{\mathcal{G}_N}\left[\hat{\pi}_n^N(f)\right] - \mathbb{E}_x[\hat{\pi}_n^N(f)]\right)^2\right] . \tag{78}$$

Since $\hat{\pi}_n^N(f)$ 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

$$\Phi_{n,k}^N : x_k \mapsto \omega_{k,n}^N f(x_k) + R_{\gamma_{k+1}}\Phi_{n,k+1}^N(x_k) , \tag{79}$$

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

$$\Psi^N_n : x_N \mapsto R_{\gamma_{N+1}} \Phi^N_{n,N+1}(x_N) \,. \tag{80}$$

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

$$\Phi^N_{n,k+1}(X_{k+1}) - R_{\gamma_{k+1}} \Phi^N_{n,k+1}(X_k) = \mathbb{E}^{\mathcal{G}_{k+1}}_x \left[\hat{\pi}^N_n(f)\right] - \mathbb{E}^{\mathcal{G}_k}_x \left[\hat{\pi}^N_n(f)\right] \,, \tag{81}$$

and $\Psi^N_n(X_N) = \mathbb{E}^{\mathcal{G}_N}_x \left[\hat{\pi}^N_n(f)\right]$. With these notations, (78) may be equivalently expressed as

$$\begin{aligned}\operatorname{Var}_x \left\{\hat{\pi}^N_n(f)\right\} = \sum_{k=N}^{N+n-1} \mathbb{E}_x \left[R_{\gamma_{k+1}} \left\{\Phi^N_{n,k+1}(\cdot) - R_{\gamma_{k+1}} \Phi^N_{n,k+1}(X_k)\right\}^2 (X_k)\right] \\ + \operatorname{Var}_x \left\{\Psi^N_n(X_N)\right\} \,. \end{aligned} \tag{82}$$

Now for $k = N+n-1, \dots, N$, we will use the Gaussian Poincaré inequality (77) to the sequence of function $\Phi^N_{n,k+1}$ to prove that $x \mapsto R_{\gamma_{k+1}}\{\Phi^N_{n,k+1}(\cdot) - R_{\gamma_{k+1}} \Phi^N_{n,k+1}(x)\}^2(x)$ is uniformly bounded. It is required to bound the Lipschitz constant of $\Phi^N_{n,k}$.

C.1.1 Proof of Theorem 15

We preface the proof by two lemmas.

Lemma 26. *Assume **H** 1 and **H** 2. Let $(\gamma_k)_{k \ge 1}$ be a non-increasing sequence with $\gamma_1 \le 2/(m+L)$. Let $N \ge 0$ and $n \ge 1$. Then for all $y \in \mathbb{R}^d$, Lipschitz function f and $k \in \{N, \dots, N+n-1\}$,*

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

where $\Phi^N_{n,k+1}$ is given by (79).

Proof. By (79), $\|\Phi^N_{n,k}\|_{\mathrm{Lip}} \le \sum_{i=k+1}^{N+n} \omega^N_{i,n} \|Q^{k+2,i}_\gamma f\|_{\mathrm{Lip}}$. Using Corollary 4, the bound $(1-t)^{1/2} \le 1 - t/2$ for $t \in [0,1]$ and the definition of $\omega^N_{i,n}$ given by (26), we have

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

Finally, the proof follows from (77). □

Also to control the last term in right hand side of (82), 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 Corollary 4 by definition, see (80). 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 using the following result.

Lemma 27. *Assume* **H** *1 and* **H** *2. Let* $(\gamma_k)_{k\geq 1}$ *be a non-increasing 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* $y \in \mathbb{R}^d$

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

where $Q_\gamma^{n,p}$ *is given by* (5)*.*

Proof. 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

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

Equation (77) implies $\mathrm{Var}_y^{\mathcal{G}_n}\{g(X_p)\} \leq 2\sum_{k=n+1}^{p} \gamma_k \|Q_\gamma^{k+1,p} g\|_{\mathrm{Lip}}^2$. The proof follows from Corollary 4 and Lemma 23, using the bound $(1-t)^{1/2} \leq 1 - t/2$ for $t \in [0,1]$. □

Corollary 28. *Assume* **H** *1 and* **H** *2. Let* $(\gamma_k)_{k\geq 1}$ *be a non-increasing sequence with* $\gamma_1 \leq 2/(m+L)$*. Then for all Lipschitz function* f *and* $x \in \mathbb{R}^d$*,* $\mathrm{Var}_x\{\Psi_n^N(X_N)\} \leq 8\kappa^{-3} \|f\|_{\mathrm{Lip}}^2 \Gamma_{N+2,N+n+1}^{-2}$*, where* Ψ_n^N *is given by* (80)*.*

Proof. By (80) and Corollary 4, Ψ_n^N is Lipschitz function with $\|\Psi_n^N\|_{\mathrm{Lip}} \leq \sum_{i=N+1}^{N+n} \omega_{i,n}^N \|Q_\gamma^{N+1,i} f\|_{\mathrm{Lip}}$. Using Corollary 4, 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 (26), we have

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

The proof follows from Lemma 27. □

Plugging the bounds given by Lemma 26 and Corollary 28 in (82), 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 of Theorem 15.

C.1.2 Proof of Theorem 16

Let $k \in \{N, \ldots, N+n-1\}$. We cannot directly apply the Poincaré inequality (77) since the function $\Phi^N_{n,k}$, defined in (79), is not Lipschitz. However, Theorem 11-(ii) shows that for all $\ell, n \in \mathbb{N}^*$, $n < \ell$, $Q^{n,\ell}_\gamma f$ is a Lipschitz function with

$$\left\| Q^{n,\ell}_\gamma f \right\|_{\mathrm{Lip}} \leq \mathrm{osc}(f) / \{4\pi \Lambda_{n,\ell}(\gamma)\}^{1/2} . \tag{83}$$

Using (79), we may decompose $\Phi^N_{n,k} = \omega^N_{k+1,n} f + \tilde{\Phi}^N_{n,k}$, where $\tilde{\Phi}^N_{n,k} = \sum_{i=k+2}^{N+n} \omega^N_{i,n} Q^{k+2,i}_\gamma f$ which is Lipshitz with constant

$$\left\| \tilde{\Phi}^N_{n,k} \right\|_{\mathrm{Lip}} \leq \sum_{i=k+2}^{N+n} \omega^N_{i,n} \left\| Q^{k+2,i}_\gamma f \right\|_{\mathrm{Lip}} \leq \mathrm{osc}(f) \sum_{i=k+2}^{N+n} \omega^N_{i,n} / \left\{4\pi \Lambda_{k+2,i}(\gamma)\right\}^{1/2} . \tag{84}$$

Using the inequality $(a+b)^2 \leq 2a^2 + 2b^2$, (77), we finally get for any $y \in \mathbb{R}^d$

$$\begin{aligned} R_{\gamma_{k+1}} \left\{ \Phi^N_{n,k+1}(\cdot) - R_{\gamma_{k+1}} \Phi^N_{n,k+1}(y) \right\}^2 (y) \leq 2 (\omega^N_{k+1,n})^2 \mathrm{osc}(f)^2 \\ + \gamma_{k+1} \mathrm{osc}(f)^2 \left\{ \sum_{i=k+2}^{N+n} \omega^N_{i,n} / \{\pi \Lambda_{k+2,i}(\gamma)\}^{1/2} \right\}^2 . \end{aligned} \tag{85}$$

It remains to control $\mathrm{Var}_x \left\{ \Psi^N_n(X_N) \right\}$, where Ψ^N_n is defined in (80). Using (83), Ψ^N_n is a Lipschitz function with Lipschitz constant bounded by:

$$\left\| \Psi^N_n \right\|_{\mathrm{Lip}} \leq \sum_{i=N+1}^{N+n} \omega^N_{i,n} \left\| Q^{N+1,i}_\gamma f \right\|_{\mathrm{Lip}} \leq \mathrm{osc}(f) \sum_{i=N+1}^{N+n} \omega^N_{i,n} / \{4\pi \Lambda_{N+1,i}(\gamma)\}^{1/2} . \tag{86}$$

By Lemma 27, we have the following result which is the counterpart of Corollary 28: for all $y \in \mathbb{R}^d$,

$$\mathrm{Var}_y \left\{ \Psi^N_n(X_N) \right\} \leq 2\kappa^{-1} \left\| f \right\|^2_\infty \left\{ \sum_{i=N+1}^{N+n} \omega^N_{i,n} / (\pi \Lambda_{N+1,i})^{1/2} \right\}^2 . \tag{87}$$

Finally, the proof follows from combining (85) and (87) in (82).

C.2 Bounds on $u^{(4)}_{0,n}(\gamma)$

Let $(\gamma_k)_{k \geq 1}$ be a non-increasing sequence of step size such that $\lim_{k \to +\infty} \gamma_k = 0$ and $\lim_{k \to +\infty} \Gamma_k = +\infty$. In this section, we show that there exist $C_1, C_2 > 0$ independent of $(\gamma_k)_{k \geq 1}$ satisfying for any $n \in \mathbb{N}^*$

$$C_1 \Gamma_n^{-1} \leq u^{(4)}_{0,n}(\gamma) \leq C_2 \Gamma_n^{-1} , \tag{88}$$

for $u^{(4)}_{0,n}$ defined in (29). We consider the following decomposition of $u^{(4)}_{0,n}(\gamma)$

$$
\begin{aligned}
u^{(4)}_{N,n}(\gamma) &= w^1_n + w^2_n \;, \\
w^1_n &= \sum_{k=0}^{n-1} \gamma_{k+1} \left\{ \sum_{i=k+2}^{n} \frac{\omega^0_{i,n}}{(\pi \Lambda_{k+2,i}(\gamma))^{1/2}} \right\}^2 \;,\; w^2_n = \kappa^{-1} \left\{ \sum_{i=1}^{n} \frac{\omega^0_{i,n}}{(4\pi\Lambda_{1,i}(\gamma))^{1/2}} \right\}^2 \;.
\end{aligned}
$$

Since $\kappa\Lambda_{n,\ell} = \prod_{j=n}^{\ell}(1-\kappa\gamma_j)^{-1} - 1$ for $n, \ell \in \mathbb{N}^*$, using that for all $(a_i)_{i\in\{1,\ldots,k\}} \in [0,1)^k$, $k \in \mathbb{N}^*$, $\prod_{i=1}^{k}(1-a_i)^{-1} - 1 \geq \exp(\sum_{i=1}^{k} a_i) - 1 \geq \sum_{i=1}^{k} a_i$, we have

$$
\prod_{j=n}^{\ell}(1-\kappa\gamma_j) \leq 1/(\kappa\Lambda_{n,\ell}) \leq 1/(\kappa^2\Gamma_{n,\ell}) \;. \tag{89}
$$

From the left inequality, we conclude using the definition of $\omega^0_{i,n}$, $i \in \{1,\ldots,n\}$, in (26) and the bound $(1-t)^{1/2} \leq 1-t/2$ for $t \in [0,1]$, that there exists $C_1 > 0$ independent of $(\gamma_k)_{k\geq 1}$ such that for any $n \in \mathbb{N}^*$,

$$
C_1 \Gamma_{2,n+1} \leq w^1_n \;. \tag{90}
$$

Now from the right inequality in (89) and using $(a+b)^2 \leq 2(a^2+b^2)$, we have for any $n \in \mathbb{N}^*$,

$$
w^1_n = 2\sum_{k=0}^{n-1} \gamma_{k+1} \left\{ \sum_{i=k+2}^{p_k} \frac{\omega^0_{i,n}}{(\pi\kappa^2\Gamma_{k+2,i})^{1/2}} \right\}^2 + 2\sum_{k=0}^{n-1} \gamma_{k+1} \left\{ \sum_{i=p_k+1}^{n} \frac{\omega^0_{i,n}}{(\pi\Lambda_{k+2,i}(\gamma))^{1/2}} \right\}^2 \;, \tag{91}
$$

where $(p_k)_{k\in\mathbb{N}^*}$ is any sequence of integers. Also we have using that $(\gamma_j)_{j\geq 1}$ is non-increasing and an integral comparison test that there exists $C \geq 0$ independent of $(\gamma_k)_{k\geq 1}$ such that for any $k, p \in \mathbb{N}^*$, $k+2 \leq p$,

$$
\begin{aligned}
\sum_{i=k+2}^{p} \frac{\omega^0_{i,n}}{(\pi\kappa^2\Gamma_{k+2,i}(\gamma))^{1/2}} &\leq \Gamma^{-1}_{2,n+1} \sum_{i=k+2}^{p} \frac{\gamma_{i+1}}{(\pi\kappa^2\Gamma_{k+2,i})^{1/2}} \leq \Gamma^{-1}_{2,n+1} \sum_{i=k+2}^{p} \frac{\Gamma_{k+2,i} - \Gamma_{k+2,i-1}}{(\pi\kappa^2\Gamma_{k+2,i})^{1/2}} \\
&\leq C\Gamma^{-1}_{2,n+1}\Gamma^{1/2}_{k+1,p} \;.
\end{aligned}
$$

Using this result in (91), we obtain that for any $n \in \mathbb{N}^*$,

$$
w^1_n \leq 2C\Gamma^{-2}_{2,n+1} \sum_{k=0}^{n-1} \gamma_{k+1}\Gamma_{k+1,p_k} + 2\sum_{k=0}^{n-1} \gamma_{k+1} \left\{ \sum_{i=p_k+1}^{n} \frac{\omega^0_{i,n}}{(\pi\Lambda_{k+2,i}(\gamma))^{1/2}} \right\}^2 \;. \tag{92}
$$

Now taking for any $n \in \mathbb{N}^*$, $k \in \{0,\ldots,n-1\}$,

$$
p_k = n \wedge \inf\{p \in \{k+1,\ldots,n-1\} \,:\, \Gamma_{k+1,p} \geq 1\} \;, \tag{93}
$$

with the convention $\inf \emptyset = +\infty$, we have for any $i \in \{p_k + 1, \ldots, n\}$, $p_k + 1 \leq n$, using for $t \geq 0$, $1 - t \leq \mathrm{e}^{-t}$,

$$\kappa \Lambda_{k+1,i}(\gamma) = \left\{ \prod_{j=k+1}^{i} (1 - \kappa\gamma_j) \right\}^{-1} \left\{ 1 - \prod_{j=k+1}^{i} (1 - \kappa\gamma_j) \right\} \geq \left\{ \prod_{j=k+1}^{i} (1 - \kappa\gamma_j) \right\}^{-1} (1 - \mathrm{e}^{-\kappa\Gamma_{k+1,p_k}})$$
$$\geq \left\{ \prod_{j=k+1}^{i} (1 - \kappa\gamma_j) \right\}^{-1} (1 - \mathrm{e}^{-\kappa}) \ .$$

Using this result, we get by (92) and $(1-t)^{1/2} \leq 1 - t/2$ for $t \in [0,1]$, that there exists $\tilde{C} \geq 0$ independent of $(\gamma_k)_{k \geq 1}$ such that for any $n \in \mathbb{N}^*$,

$$w_n^1 \leq 2C\Gamma_{2,n+1}^{-2} \sum_{k=0}^{n-1} \gamma_{k+1}\Gamma_{k+1,p_k} + 2(1 - \mathrm{e}^{-\kappa})^{-1} \sum_{k=0}^{n-1} \gamma_{k+1} \left\{ \sum_{i=p_k+1}^{n} \omega_{i,n}^0 \prod_{j=k+1}^{i} (1 - \kappa\gamma_j)^{1/2} \right\}^2$$
$$\leq 2C\Gamma_{2,n+1}^{-1} \sum_{k=0}^{n-1} \gamma_{k+1}\Gamma_{k+1,p_k} + 2(1 - \mathrm{e}^{-\kappa})^{-1} \sum_{k=0}^{n-1} \gamma_{k+1} \left\{ \sum_{i=p_k+1}^{n} \omega_{i,n}^0 \prod_{j=k+1}^{i} (1 - \kappa\gamma_j/2) \right\}^2$$
$$\leq 2C\Gamma_{2,n+1}^{-1} \sum_{k=0}^{n-1} \gamma_{k+1}\Gamma_{k+1,p_k} + 2(1 - \mathrm{e}^{-\kappa})^{-1}\tilde{C}\Gamma_{2,n+1}^{-1} \ ,$$

Since $\Gamma_{k+1,p_k} \leq 1 + \gamma_1$, for any $n \in \mathbb{N}^*$, $k \in \{0, \ldots, n-1\}$ and definition of p_k (93), we obtain that there exists $C \geq 0$ such that for any $n \in \mathbb{N}^*$,

$$w_n^1 \leq C\Gamma_{2,n+1}^{-1} \ . \tag{94}$$

Similarly, we have that there exists $C \geq 0$ independent of $(\gamma_k)_{k \geq 1}$ satisfying for any $n \in \mathbb{N}^*$, $w_n^2 \leq C\Gamma_{2,n+1}^{-1}$ Combining this result, (90) and (94) concludes the proof of (88).

C.3 Proof of Theorem 17

Let $N \geq 0$, $n \geq 1$, $x \in \mathbb{R}^d$ and f be a Lipschitz function. To prove Theorem 17, we derive an upper bound of the Laplace transform of $\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)]$. Consider the decomposition by martingale increments

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

Now using (81) with the sequence of functions $(\Phi_{n,k}^N)$ and Ψ_n^N given by (79) and (80), respectively, we have by the Markov property

$$\mathbb{E}_x\left[\mathrm{e}^{\lambda\{\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)]\}}\right]$$
$$= \mathbb{E}_x\left[\mathrm{e}^{\lambda\left\{\Psi_n^N(X_n) - \mathbb{E}_x\left[\Psi_n^N(X_n)\right]\right\}} \prod_{k=N}^{N+n-1} R_{\gamma_{k+1}}\left[\mathrm{e}^{\lambda\{\Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}}\Phi_{n,k+1}^N(X_k)\}}\right](X_k)\right] \ , \tag{95}$$

where R_γ is given by (4) 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 on

$$R_{\gamma_{k+1}}\{\exp(\lambda\{\Phi^N_{n,k+1}(\cdot) - R_{\gamma_{k+1}}\Phi^N_{n,k+1}(X_k)\})\}(X_k)$$

w.r.t. X_k, for all $k \in \{N+1, \dots, N+n\}$. Indeed for all $x \in \mathbb{R}^d$ and $\gamma > 0$, 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 [3, Theorem 5.5] shows that for all Lipschitz function $g : \mathbb{R}^d \to \mathbb{R}$, $x \in \mathbb{R}^d$, $\gamma > 0$ and $\lambda > 0$,

$$\int R_\gamma(x, \mathrm{d}y)\left\{\exp\left(\lambda\{g(y) - R_\gamma g(x)\}\right)\right\} \le \exp\left(\gamma\lambda^2 \|g\|^2_{\mathrm{Lip}}\right) . \tag{96}$$

We deduced from this result, (81) and Corollary 4, an equivalent of Lemma 26 for the Laplace transform of $\Phi^N_{n,k+1}$ under $\delta_y R_{\gamma_{k+1}}$ for $k \in \{N+1, \dots, N+n\}$ and all $y \in \mathbb{R}^d$.

Corollary 29. *Assume* ***H*** *1 and* ***H*** *2. Let $(\gamma_k)_{k\ge1}$ be a non-increasing sequence with $\gamma_1 \le 2/(m+L)$. Let $N \ge 0$ and $n \ge 1$. Then for all $k \in \{N, \dots, N+n-1\}$, $y \in \mathbb{R}^d$ and $\lambda > 0$,*

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

where $\Phi^N_{n,k}$ is given by (79).

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 (5). For this, using again that by (80) and Corollary 4, Ψ^N_n is a Lipschitz function, we iterate (96) 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 $n, \ell \ge 1$, since for all $n, \ell \ge 1$, $Q^{n,\ell}_\gamma g$ is a Lipschitz function by Corollary 4.

Lemma 30. *Assume* ***H*** *1 and* ***H*** *2. Let $(\gamma_k)_{k\ge1}$ be a non-increasing sequence with $\gamma_1 \le 2/(m+L)$. Let $g : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz function, then for all $n, p \ge 1$, $n \le p$, $y \in \mathbb{R}^d$ and $\lambda > 0$:*

$$Q^{n,p}_\gamma\left\{\exp\left(\lambda\{g(\cdot) - Q^{n,p}_\gamma g(y)\}\right)\right\}(y) \le \exp\left(\kappa^{-1}\lambda^2\|g\|^2_{\mathrm{Lip}}\right) , \tag{97}$$

where $Q^\gamma_{n,p}$ is given by (5).

Proof. Let $(X_n)_{n\ge0}$ 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 (96), we get

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

The proof follows from Corollary 4 and Lemma 23, using the bound $(1-t)^{1/2} \leq 1-t/2$ for $t \in [0,1]$.

□

Combining this result and $\|\Psi_n^N\|_{\mathrm{Lip}} \leq 2\kappa^{-1}\|f\|_{\mathrm{Lip}}\Gamma_{N+2,N+n+1}^{-1}$ by Corollary 4, we get an analogue of Corollary 28 for the Laplace transform of Ψ_n^N:

Corollary 31. *Assume **H**1 and **H**2. Let $(\gamma_k)_{k\geq 1}$ be a non-increasing sequence with $\gamma_1 \leq 2/(m+L)$. Let $N \geq 0$ and $n \geq 1$. Then for all $\lambda > 0$ and $x \in \mathbb{R}^d$,*

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

where Ψ_n^N is given by (80).

The Laplace transform of $\hat{\pi}_n^N(f)$ can be explicitly bounded using Corollary 29 and Corollary 31 in (95).

Proposition 32. *Assume **H**1 and **H**2. Let $(\gamma_k)_{k\geq 1}$ be a non-increasing sequence with $\gamma_1 \leq 2/(m+L)$. Then for all $N \geq 0$, $n \geq 1$, Lipschitz functions $f : \mathbb{R}^d \to \mathbb{R}$, $\lambda > 0$ and $x \in \mathbb{R}^d$:*

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

where $u_{N,n}^{(3)}(\gamma)$ is given by (28).

Proof of Theorem 17. Using the Markov inequality and Proposition 32, for all $\lambda > 0$, we have:

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

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

C.4 Proof of Theorem 18

Let $N \geq 0$, $n \geq 1$, $x \in \mathbb{R}^d$ and $f \in \mathbb{F}_b(\mathbb{R}^d)$. The main idea of the proof is to consider the decomposition (95) again but combined with the decomposition of $\Phi_{n,k+1}^N$, for $k \in \{N,\ldots,N+n-1\}$, into a Lipschitz component and a bounded measurable component as it is done in the proof of (85). Let $k \in \{N,\ldots,N+n-1\}$. By definition (79),

$\Phi^N_{n,k} = \omega^N_{k+1,n} f + \tilde{\Phi}^N_{n,k}$, where $\tilde{\Phi}^N_{n,k} = \sum_{i=k+2}^{N+n} \omega^N_{i,n} Q_\gamma^{k+2,i} f$. Using that f is bounded, we get for all $y \in \mathbb{R}^d$ and $\lambda > 0$,

$$R_{\gamma_{k+1}} \left\{ \mathrm{e}^{\lambda\{\Phi^N_{n,k+1}(\cdot) - R_{\gamma_{k+1}}\Phi^N_{n,k+1}(y)\}} \right\}(y) \\ \leq \mathrm{e}^{\lambda \operatorname{osc}(f)\, \gamma_{k+2}(\Gamma_{N+2,N+n+1})^{-2}} R_{\gamma_{k+1}} \left\{ \mathrm{e}^{\lambda\{\tilde{\Phi}^N_{n,k+1}(\cdot) - R_{\gamma_{k+1}}\tilde{\Phi}^N_{n,k+1}(y)\}} \right\}(y)$$

By (84) and (96), we obtain for all $y \in \mathbb{R}^d$ and $\lambda > 0$,

$$R_{\gamma_{k+1}} \left\{ \mathrm{e}^{\lambda\{\Phi^N_{n,k+1}(\cdot) - R_{\gamma_{k+1}}\Phi^N_{n,k+1}(y)\}} \right\}(y) \\ \leq \exp\left(\lambda \operatorname{osc}(f)\, \gamma_{k+2}(\Gamma_{N+2,N+n+1})^{-2} + (\lambda \operatorname{osc}(f))^2 \gamma_{k+1} \left(\sum_{i=k+2}^{N+n} \omega^N_{i,n} / (\pi \Lambda_{k+2,i})^{1/2} \right)^2 \right) . \tag{98}$$

It remains to control the Laplace transform of Ψ^N_n under $\delta_x Q^N_\gamma$. For this, note that by (86) Ψ^N_n is a Lipschitz function. Therefore using Lemma 30, we get an analogue of Corollary 31: for all $y \in \mathbb{R}^d$ and $\lambda > 0$,

$$\mathbb{E}_y \left[\mathrm{e}^{\lambda\{\Psi^N_n(X_n) - \mathbb{E}_x[\Psi^N_n(X_n)]\}} \right] \leq \exp\left(\kappa^{-1} \lambda^2 \operatorname{osc}(f)^2 \left(\sum_{i=N+1}^{N+n} \omega^N_{i,n} / (\pi \Lambda_{N+1,i})^{1/2} \right)^2 \right) , \tag{99}$$

Combining (98) and (99) in (95), the Laplace transform of $\hat{\pi}^N_n(f)$ can be explicitly bounded: for all $\lambda > 0$,

$$\mathbb{E}_x \left[\mathrm{e}^{\lambda\{\hat{\pi}^N_n(f) - \mathbb{E}_x[\hat{\pi}^N_n(f)]\}} \right] \leq \mathrm{e}^{\lambda \operatorname{osc}(f)(\Gamma_{N+2,N+n+1})^{-1} + (\lambda \operatorname{osc}(f))^2 u^{(5)}_{N,n}(\gamma)} .$$

Using this result and the Markov inequality, for all $\lambda > 0$, we have:

$$\mathbb{P}_x \left[\hat{\pi}^N_n(f) \geq \mathbb{E}_x[\hat{\pi}^N_n(f)] + r \right] \\ \leq \exp\left(-\lambda r + \lambda \operatorname{osc}(f)(\Gamma_{N+2,N+n+1})^{-1} + (\lambda \operatorname{osc}(f))^2 u^{(5)}_{N,n}(\gamma) \right) .$$

Then the proof follows from taking

$$\lambda = (r - \operatorname{osc}(f)(\Gamma_{N+2,N+n+1})^{-1}) / (2 \operatorname{osc}(f)^2 u^{(5)}_{N,n}(\gamma)) .$$

D Additional technical results

D.1 Coupling

Lemma 33. *Assume* ***AR*** *1. For all $x, y \in \mathbb{R}^d$ and $k \geq 1$, $\mathsf{K}_k((x,y),\cdot)$ is a transference plan of $\mathsf{P}_k(x,\cdot)$ and $\mathsf{P}_k(y,\cdot)$*

Proof. By construction, $\mathsf{K}_k((x,y),\cdot\times\mathbb{R}^d)=\mathsf{P}_k(x,\cdot)$ for all $x,y\in\mathbb{R}^d$ and $\mathsf{K}_k((x,y),\mathbb{R}^d\times\cdot)=\mathsf{P}_k(y,\cdot)$ for all (x,y) such that $h_k(x)=h_k(y)$. Therefore, it remains to show that $\mathsf{K}_k((x,y),\mathbb{R}^d\times\cdot)=\mathsf{P}_k(y,\cdot)$ for any $(x,y)\in\mathbb{R}^d\times\mathbb{R}^d$, $h_k(x)\neq h_k(y)$. First for all $\mathtt{A}\in\mathcal{B}(\mathbb{R}^d)$, we have

$$\mathsf{K}_k((x,y),\mathbb{R}^d\times\mathtt{A})=\frac{1}{(2\pi\sigma_k^2)^{d/2}}\int_{\mathbb{R}^d}\mathbb{1}_{\mathtt{A}}(\tilde{x})p_k(x,y,\tilde{x}-h_k(x))\mathrm{e}^{-\|\tilde{x}-h_k(x)\|^2/(2\sigma_k^2)}\mathrm{d}\tilde{x} \tag{100}$$
$$+\frac{1}{(2\pi\sigma_k^2)^{d/2}}\int_{\mathbb{R}^d}\mathbb{1}_{\mathtt{A}}(\mathrm{F}_k(x,y,\tilde{x}-h_k(x)))\left\{1-p_k(x,y,\tilde{x}-h_k(x))\right\}\mathrm{e}^{-\|\tilde{x}-h_k(x)\|^2/(2\sigma_k^2)}\mathrm{d}\tilde{x}\,.$$

Since $(\mathrm{Id}-2\mathrm{e}_k(x,y)\mathrm{e}_k(x,y)^{\mathrm{T}})$ is an orthogonal matrix, making the change of variable $\tilde{y}=\mathrm{F}_k(x,y,\tilde{x}-h_k(x))$ and using that

$$\langle\mathrm{e}_k(x,y),h_k(y)-\tilde{y}\rangle=\langle\mathrm{e}_k(x,y),\tilde{x}-h_k(x)\rangle$$

we get that

$$\int_{\mathbb{R}^d}\mathbb{1}_{\mathtt{A}}(\mathrm{F}_k(x,y,\tilde{x}-h_k(x)))\left\{1-p_k(x,y,\tilde{x}-h_k(x))\right\}\mathrm{e}^{-\|\tilde{x}-h_k(x)\|^2/(2\sigma_k^2)}\mathrm{d}\tilde{x}$$
$$=\int_{\mathbb{R}^d}\mathbb{1}_{\mathtt{A}}(\tilde{y})\left\{1-p_k(x,y,h_k(y)-\tilde{y})\right\}\mathrm{e}^{-\|\tilde{y}-h_k(y)\|^2/(2\sigma_k^2)}\mathrm{d}\tilde{y}\,. \tag{101}$$

By definition of α_k (35), we have for all $\tilde{x}\in\mathbb{R}^d$,

$$\alpha_k\left(x,y,\tilde{x}-h_k(x)\right)=\frac{\boldsymbol{\varphi}_{\sigma_k^2}\left(\langle\mathrm{e}_k(x,y),h_k(y)-\tilde{x}\rangle\right)}{\boldsymbol{\varphi}_{\sigma_k^2}\left(\|\mathrm{E}_k(x,y)\|-\langle\mathrm{e}_k(x,y),h_k(y)-\tilde{x}\rangle\right)}=\frac{1}{\alpha_k\left(x,y,h_k(y)-\tilde{x}\right)}\,. \tag{102}$$

In addition using that

$$\|\tilde{x}-h_k(x)\|^2=\|\tilde{x}-h_k(y)\|^2-2\left\langle h_k(y)-\tilde{x},\mathrm{E}_k(x,y)\right\rangle+\|\mathrm{E}_k(x,y)\|^2\;,$$

we obtain

$$p_k(x,y,\tilde{x}-h_k(x))\mathrm{e}^{-\|\tilde{x}-h_k(x)\|^2/(2\sigma_k^2)}=p_k(x,y,h_k(y)-\tilde{x})\mathrm{e}^{-\|\tilde{x}-h_k(y)\|^2/(2\sigma_k^2)}\,. \tag{103}$$

Plugging (101) and (103) into (100) implies that $\mathsf{K}_k((x,y),\mathbb{R}^d\times\mathtt{A})=\mathsf{P}_k(y,\mathtt{A})$. ☐

D.2 Distribution of hitting time of 0 for Ornstein-Ulhenbeck processes

Consider the one-dimensional Ornstein-Ulhenbeck process $(\tilde{\mathsf{U}}_t)_{t\geq 0}$ defined for $t\geq 0$ by

$$\tilde{\mathsf{U}}_t=a\mathrm{e}^{-\theta t}+\sigma\int_0^t\mathrm{e}^{\theta(s-t)}\mathrm{d}B_s^1=a\mathrm{e}^{-\theta t}+\frac{\sigma}{\sqrt{2\theta}}\mathrm{e}^{-\theta t}B^1_{\mathrm{e}^{2\theta t}-1}\;,$$

where $a\in\mathbb{R}$, $\theta,\sigma>0$ and $(B_t^1)_{t\geq 0}$ is a one-dimensional Brownian motion. Note that with our convention, $(\tilde{\mathsf{U}}_t)_{t\geq 0}$ is the solution of the SDE $\mathrm{d}\tilde{\mathsf{U}}_t=-\theta\,\tilde{\mathsf{U}}_t\mathrm{d}t+\sigma\mathrm{d}B_t^1$ with initial condition $\tilde{\mathsf{U}}_0=a$. Define the hitting time of $(\tilde{\mathsf{U}}_t)_{t\geq 0}$ of 0 by $\tilde{T}_0=\inf\{t\geq 0\,:\,\tilde{\mathsf{U}}_t=0\}$.

Proposition 34 ([2, Formula 2.0.2, page 542])**.** *For all $a \in \mathbb{R}$, $\theta, \sigma > 0$, and $t > 0$, it holds*

$$\mathbb{P}\left(\min_{0\leq s\leq t} \tilde{\mathsf{U}}_s > 0\right) = \mathbb{P}\left(\tilde{T}_0 > t\right) = 1 - 2\mathbf{\Phi}\left(-\frac{\sqrt{2\theta}\,|a|}{\sigma\sqrt{\mathrm{e}^{2\theta t} - 1}}\right),$$

where $\mathbf{\Phi}$ is the cumulative distribution function of the standard normal distribution.

This figure "var_tv_0_v2.png" is available in "png" format from:

http://arxiv.org/ps/1605.01559v4

This figure "var_tv_1_2_v2.png" is available in "png" format from:

http://arxiv.org/ps/1605.01559v4

This figure "var_tv_1_4_v2.png" is available in "png" format from:

http://arxiv.org/ps/1605.01559v4

This figure "var_tv_3_4_v2.png" is available in "png" format from:

http://arxiv.org/ps/1605.01559v4