Submitted to Bernoulli

High-dimensional Bayesian inference via the Unadjusted Langevin Algorithm

ALAIN DURMUS [*] and ÉRIC MOULINES [**]

CMLA, ENS Cachan, CNRS, Université Paris-Saclay,
94235 Cachan, France
E-mail: [*]alain.durmus@cmla.ens-cachan.fr

Centre de Mathématiques Appliquées, UMR 7641,
Ecole Polytechnique,
route de Saclay, 91128 Palaiseau cedex, France.
E-mail: [**]eric.moulines@polytechnique.edu

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.

AMS 2000 subject classifications: primary 65C05, 60F05, 62L10; secondary 65C40, 60J05,93E35.
Keywords: total variation distance, Langevin diffusion, Markov Chain Monte Carlo, Metropolis Adjusted Langevin Algorithm, Rate of convergence.

1. Introduction

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

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:

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

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

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

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 [?] and [?] for molecular dynamics applications. Then it has been popularized in machine learning by [?], [?] and computational statistics by [?] and [?]. Following [?], 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 [?] and [?] 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 [?], [?] and [?]. In particular, [?, 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 [?], [?] and [?].

Contrary to these reported works, we focus in this paper on non-asymptotic results. These questions have been addressed previously in [?] and [?]. [?] 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 [?] were later sharpened in [?], using different technical arguments. In particular, [?] 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. [?] also extends the results of [?] 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 [?] and [?], 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 [**?**] and [**?**] 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 [**?**].
- 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 [**?**]. 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 [**?**] 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 [**?**]. 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 a supplementary paper [**?**].

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 [**?**, 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 [**?**, 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 [**?**, 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 [**?**, Theorem 2.5, Theorem 2.9 Chapter 5] there exists a unique strong solution $(Y_t)_{t \geq 0}$ to (1) with $Y_0 = x$. Denote by $(P_t)_{t \geq 0}$ the semi-group associated with (1). It is well-known that π is its (unique) invariant probability. To get geometric convergence of $(P_t)_{t \geq 0}$ 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, [**?**, 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. [**?**].

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 [**?**, **??**]. □

Note that the convergence rate in Proposition 1-(iv) does not depend on the dimension. Let $(\gamma_k)_{k \geq 1}$ 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 $(X_k)_{k \geq 0}$ given in (2) is an inhomogeneous Markov chain with respect to the family of Markov kernels $(R_{\gamma_k})_{k \geq 1}$. 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

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

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

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

Proof. The proof is postponed to [?, ??]. □

We now proceed to establish that Q_γ^n is a strict contraction in W_2 for any $n \ge 1$. This result implies the geometric convergence of the sequence $(\delta_x R_\gamma^n)_{n \ge 1}$ 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 $(\gamma_k)_{k \ge 1}$ be a non-increasing sequence with $\gamma_1 \le 2/(m+L)$. For all $x, y \in \mathbb{R}^d$ and $\ell \ge n \ge 1$,

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

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

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

Proof. The proof is postponed to [?, ??] □

Corollary 4. *Assume **H** 1 and **H** 2. Let $(\gamma_k)_{k \ge 1}$ be a non-increasing sequence with $\gamma_1 \le 2/(m+L)$. Then for all Lipschitz functions $f : \mathbb{R}^d \to \mathbb{R}$ and $\ell \ge n \ge 1$, $Q_\gamma^{n,\ell} f$ is a Lipschitz function with $\|Q_\gamma^{n,\ell} f\|_{\mathrm{Lip}} \le \prod_{k=n}^{\ell} (1 - \kappa\gamma_k)^{1/2} \|f\|_{\mathrm{Lip}}$.*

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

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

□

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 $(\gamma_k)_{k \ge 1}$ be a non-increasing sequence with $\gamma_1 \le 1/(m+L)$. Then for all $x \in \mathbb{R}^d$ and $n \ge 1$,*

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

where

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

κ is defined in (6) and

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

Proof. The proof is postponed to [?, ??]. □

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 [?, ??]. □

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 [?, ??] applied with $\ell = 1$. □

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

H3. *The potential U is three times continuously differentiable and there exists $\tilde{L}$ such that for all $x, y \in \mathbb{R}^d$, $\left\| \nabla^2 U(x) - \nabla^2 U(y) \right\| \leq \tilde{L} \left\| x - y \right\|$.*

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

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

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 $n \geq 1$,*

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

where $u_n^{(1)}$ is given by (8), κ in (6) 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 [**?**, **??**]. □

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 $W_2(\delta_{x^\star} R^n_{\gamma_\varepsilon}, \pi) \leq \varepsilon$ for all $n \geq n_\varepsilon$. The precise expression of n_ε directly computed using Theorem 5 and Theorem 8 are also given in [**?**, **??**-**??**]. 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 [**?**] 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 [**?**] (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).

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

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

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 $W_2^2(\delta_{x^\star} R^n_{\gamma_n}, \pi) = \mathcal{O}(\log(n)/n)$ and $W_2^2(\delta_{x^\star} R^n_{\gamma_n}, \pi) \leq \mathcal{O}(\log(n)/n)^2$ if **H**3 holds by Theorem 8. The precise statement of these results are given by [**?**, **??**-**??**] in [**?**, **??**-**??**].

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 $u_n^{(3)} = d^2\mathcal{O}(n^{-2\alpha})$ (see [**?**, **??**-**??**] 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 [**?**, **??** in **??**]. Using this result in [**?**, **??**], we get that under **H**1 and **H**2, that $W_2^2(\delta_{x^\star} Q^n_\gamma, \pi) = \mathcal{O}(n^{-1})$ for $\gamma_1 > 2\kappa^{-1}$. If in addition **H**3 holds, we have $W_2^2(\delta_{x^\star} Q^n_\gamma, \pi) = \mathcal{O}(n^{-1})$ 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 [?, 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 [?, ??]
Theorem 8	$d^2\,\mathcal{O}(n^{-2\alpha})$	$d^2\,\mathcal{O}(n^{-2})$ for $\gamma_1 > 4\kappa^{-1}$ see [?, ??]

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 [?, ?? -??]. 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 [?, Section 3] or [?, Example 3.7] for its construction, and [?, ?, ?] 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 $\mathbf{\Phi}$ 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 **H**1 and **H**2.*

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

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

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

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

Then by **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 [**?**, Formula 2.0.2, page 542]. For completeness, this formula is given in [**?**, **??**].

(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 [**?**, Section 3.3] for Gaussian random walks. This construction has been used in [**?**] 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 (**??**) 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

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

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 $s > 0$, $1/2 - \boldsymbol{\Phi}(-s) \leq (2\pi)^{-1/2} s$ 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* $\boldsymbol{H}$*1 and* $\boldsymbol{H}$*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* $\ell, n \in \mathbb{N}^*$, $\ell > n$,

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

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

Proof. The proof is postponed to [**?**, **??**]. □

Consider the case of decreasing step sizes of the form $\gamma_k = \gamma_1/k^\alpha$ for $k \geq 1$ 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 $i = 2,3$,

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

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 $k \geq 1$ and $\alpha = 1$, setting $n = \ell - \lfloor \ell/2 \rfloor$, $\ell \in \mathbb{N}^*$, $\ell > 2$, we have for $i = 2,3$,

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

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 [**?**, Proposition 3], for potentials satisfying **H**1 but not necessarily convex since [**?**, Proposition 3] only requires the additional assumption that $(P_t)_{t\geq 0}$ is geometrically ergodic in total variation.

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

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

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.

	$\alpha \in (0,1)$	$\alpha = 1$
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

When $\gamma_k = \gamma \in (0, 1/(m+L))$ for all $k \geq 1$, under **H**1 and **H**2, for $\ell > \lceil \gamma^{-1} \rceil$ choosing $n = \ell - \lceil \gamma^{-1} \rceil$ implies that (see the supplementary document [**?**, **??**])

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

where

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

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 [**?**].

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*

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

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, [**?**, 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 [**?**, **??**]. □

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 [**?**, Theorem 10], with respect to the step size γ, but Theorem 14 requires that **H**3 holds contrary to Corollary 13 and [**?**, 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 [**?**, 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 [**?**] 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 [**?**, 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, **H**2	**H**1, **H**2 and **H**3
γ_ε	$d^{-1}\mathcal{O}(\varepsilon^2)$	$d^{-1}\log^{-1}(d)\mathcal{O}(\varepsilon\left\lvert\log^{-1}(\varepsilon)\right\rvert)$
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 $\|\delta_{x^\star} R_{\gamma_\varepsilon}^{n_\varepsilon} - \pi\|_{\mathrm{TV}} \le \varepsilon$ for $\varepsilon > 0$.

4. Mean square error and concentration for bounded measurable functions

Let $(X_k)_{k\ge 0}$ be the Euler discretization of the Langevin diffusion (2) associated with the sequence of non-increasing step sizes $(\gamma_k)_{k\ge 1}$. 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 $f : \mathbb{R}^d \to \mathbb{R}$, $N, n \in \mathbb{N}$, $n \ge 1$ 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 $(X_n)_{n\ge 0}$ 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:

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

Similarly, if f is bounded,

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

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 $(\gamma_k)_{k\geq 1}$. 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 [?, Theorem 2] for homogeneous Markov chain to our inhomogeneous setting.

Theorem 15. *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 Lipschitz function. Then for all $N \geq 0$ and $n \geq 1$, we get $\operatorname{Var}_x\{\hat{\pi}_n^N(f)\} \leq 8\kappa^{-2} \|f\|_{\mathrm{Lip}}^2 \Gamma_{N+2,N+n+1}^{-1} v_{N,n}(\gamma)$, 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 [?, ??]. □

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 $k \geq 1$ by $\gamma_k = \gamma_1/k^\alpha$ where $\alpha \in [0,1]$ and $\gamma_1 < 1/(m+L)$. Details of these calculations are postponed to [?, ????]. 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 [?] 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
$\alpha = 0$	$\gamma_1 + (\gamma_1 n)^{-1}\{1 + \exp(-\kappa\gamma_1 N/2)\}$
$\alpha \in (0,1/2)$	$\gamma_1 n^{-\alpha} + (\gamma_1 n^{1-\alpha})^{-1}\left\{1 + \exp(-\kappa\gamma_1 N^{1-\alpha}/(2(1-\alpha)))\right\}$
$\alpha = 1/2$	$\gamma_1 \log(n) n^{-1/2} + (\gamma_1 n^{1/2})^{-1}\left\{1 + \exp(-\kappa\gamma_1 N^{1/2}/4)\right\}$
$\alpha \in (1/2,1)$	$n^{\alpha-1}\left[\gamma_1 + \gamma_1^{-1}\left\{1 + \exp(-\kappa\gamma_1 N^{1-\alpha}/(2(1-\alpha)))\right\}\right]$
$\alpha = 1$	$\mathcal{O}(\log(n)^{-1})$ 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*

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

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

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

Proof. The proof is postponed to **??**.

□

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 [**?**, **??**], 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 $f : \mathbb{R}^d \to \mathbb{R}$:*

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

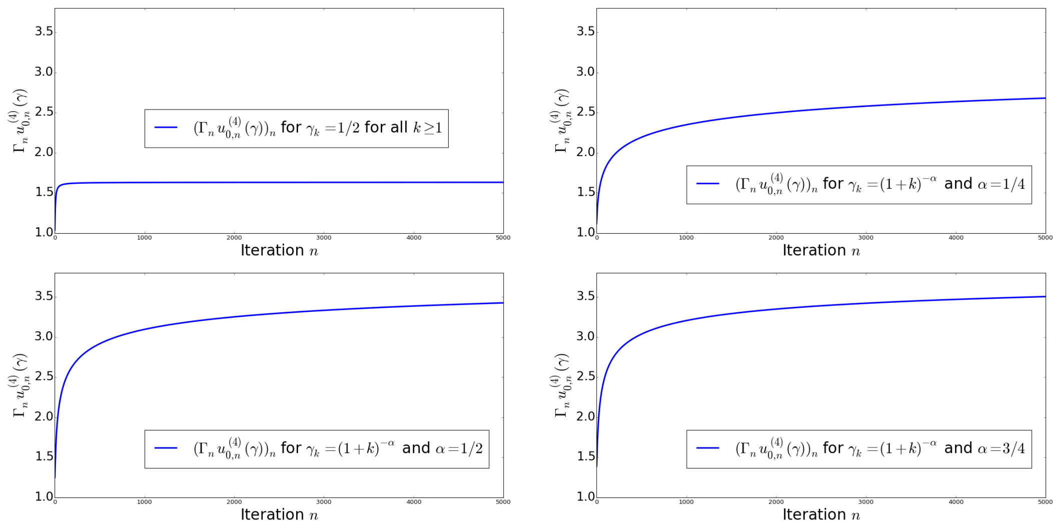

Figure 1. Plots of $(u^{(4)}_{0,n})_{n\geq 1}\Gamma_n$ for four sequences of step sizes $(\gamma_k)_{k\geq 1}$, $\gamma_k = (1+k)^{-\alpha}$ for $\alpha = 0, 1/4, 1/2, 3/4$

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

Proof. The proof is postponed to [?, ??]. □

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 $f \in \mathbb{F}_b(\mathbb{R}^d)$:*

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

where

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

Proof. The proof is postponed to [?, ??]. □

Note that $u^{(5)}_{N,n}(\gamma)$ is up to numerical constants similar to $u^{(4)}_{N,n}(\gamma)$ given in (29). Therefore, using the same calculations as in [?, ??], 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 [**?**] for probit regression; see [**?**], [**?**] and [**?**]. Recently, a very promising DA algorithm has been proposed in [**?**], using the Polya-Gamma distribution in the DA part. This algorithm has been shown to be uniformly ergodic for the total variation by [**?**, 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 [**?**]. For that purpose, we take $d = 5$, $p = 100$, generate synthetic data $(Y_i)_{1 \leq i \leq p}$ and $(X_i)_{1 \leq i \leq p}$, and set $\Sigma_{\boldsymbol{\beta}}^{-1} = (dp)^{-1} (\sum_{i=1}^p X_i^{\mathrm{T}} X_i) \, \mathrm{I}_d$. We produce 10^8 samples from the Pólya-Gamma sampler using the R package BayesLogit [**?**]. Next, we make 10^3 runs of the Euler approximation scheme with $n = 10^6$ effective iterations, with a constant sequence $(\gamma_k)_{k \geq 1}$, $\gamma_k = 10(\kappa n^{1/2})^{-1}$ for all $k \geq 0$ and a burn-in period $N = n^{1/2}$. The 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 $(\gamma_k)_{k\geq 1}$ defined by $\gamma_k = \gamma_1 k^{-1/2}$, with $\gamma_1 = 10(\kappa \log(n)^{1/2})^{-1}$ and for the burn in period $N = \log(n)$, see also Figure 2. In

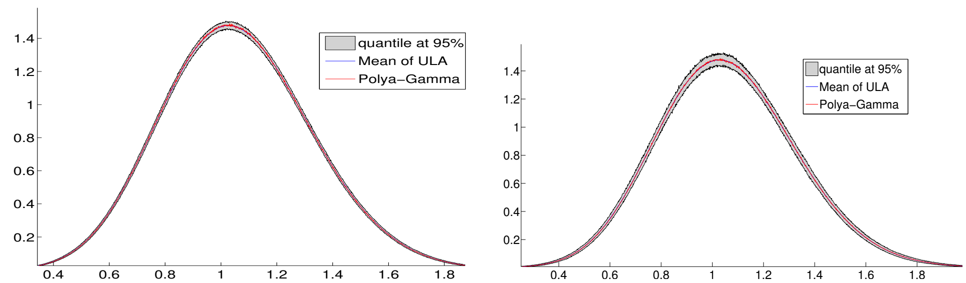

Figure 2. Empirical distribution comparison between the Polya-Gamma Gibbs Sampler and ULA. Left panel: constant step size $\gamma_k = \gamma_1$ for all $k \geq 1$; right panel: decreasing step size $\gamma_k = \gamma_1 k^{-1/2}$ for all $k \geq 1$

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 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 [?], [?] 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 [?] and [?] 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.

[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)

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

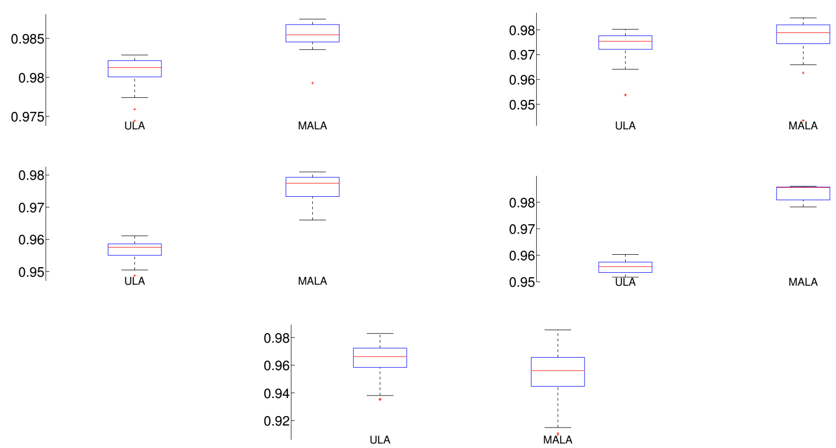

Figure 3. Marginal accuracy across all the dimensions.
Upper left: German credit data set. Upper right: Australian credit data set. Lower left: Heart disease data set. Lower right: Pima Indian diabetes data set. At the bottom: Musk data set

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

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

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

$$\mathrm{F}_k(x,y,z) = h_k(y) + 0\sigma_k \left(\mathrm{Id} - 2\mathrm{e}_k(x,y)\mathrm{e}_k(x,y)^{\mathrm{T}}\right) z \tag{34}$$

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

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 $0B_1$ is given Z_1, a Bernoulli random variable with success probability

$$p_k(x,y,\mathrm{Z}_1) = 1 \wedge \; \alpha_k(x,y,\mathrm{Z}_1) \; .$$

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

$$
\begin{aligned}
\mathsf{K}_k((x,y),\mathtt{A}) &= \frac{\mathbb{1}_{\mathsf{D}}(h_k(x),h_k(y))}{(2\pi\sigma_k^2)^{d/2}} \int_{\mathbb{R}^d} \mathbb{1}_{\mathtt{A}}(\tilde{x},\tilde{x}) \mathrm{e}^{-\|\tau_k(\tilde{x},x)\|^2/(2\sigma_k^2)} \mathrm{d}\tilde{x} \qquad (36)\\
&+ \frac{\mathbb{1}_{\mathsf{D}^{\mathrm{c}}}(h_k(x),h_k(y))}{(2\pi\sigma_k^2)^{d/2}} \left[\int_{\mathbb{R}^d} \mathbb{1}_{\mathtt{A}}(\tilde{x},\tilde{x}) p_k(x,y,0\sigma_k^{-1}\tau_k(\tilde{x},x)) \mathrm{e}^{-\|\tau_k(\tilde{x},x)\|^2/(2\sigma_k^2)} \mathrm{d}\tilde{x} \right.\\
&\left. + \int_{\mathbb{R}^d} \mathbb{1}_{\mathtt{A}}(\tilde{x}, \mathrm{F}_k(x,y,0\sigma_k^{-1}\tau_k(\tilde{x},x))) \left\{1 - p_k(x,y,0\sigma_k^{-1}\tau_k(\tilde{x},x))\right\} \mathrm{e}^{-\|\tau_k(\tilde{x},x)\|^2/(2\sigma_k^2)} \mathrm{d}\tilde{x} \right] ,
\end{aligned}
$$

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 [**?**, 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 [**?**, **??**]. Furthermore, we have for all $x,y \in \mathbb{R}^d$ and $k \geq 1$

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

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

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

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)

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

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

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

and

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

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

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

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.

□

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 $k \in \{0, \cdots, n-1\}$,

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

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

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

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".

Supplementary Material

Most proofs and derivations are postponed and carried out in a supplementary paper. (doi: COMPLETED BY THE TYPESETTER; .pdf).

References