



HAL
open science

High-dimensional Bayesian inference via the Unadjusted Langevin Algorithm

Alain Durmus, Éric Moulines

► **To cite this version:**

Alain Durmus, Éric Moulines. High-dimensional Bayesian inference via the Unadjusted Langevin Algorithm. 2016. hal-01304430v2

HAL Id: hal-01304430

<https://hal.science/hal-01304430v2>

Preprint submitted on 9 Dec 2016

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

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

High-dimensional Bayesian inference via the Unadjusted Langevin Algorithm

Alain Durmus¹ Éric Moulines²

December 9, 2016

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 normalisation factor $x \mapsto e^{-U(x)} / \int_{\mathbb{R}^d} e^{-U(y)} dy$. 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 the obtained bounds is studied to demonstrate the applicability of this method. 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 either Lipschitz continuous or measurable and bounded. An illustration to a Bayesian inference for binary regression is presented.

1 Introduction

There has been recently an increasing interest in Bayesian inference applied to high-dimensional models often motivated by machine learning applications. Rather than obtaining a point estimate typical in optimization setting, Bayesian methods attempt to

¹LTCI, Telecom ParisTech 46 rue Barrault, 75634 Paris Cedex 13, France. alain.durmus@telecom-paristech.fr

²Centre de Mathématiques Appliquées, UMR 7641, Ecole Polytechnique, France. eric.moulines@polytechnique.edu

sample the full posterior distribution over the parameters and possibly latent variables. This arguably provides a way to assert uncertainty in the model and prevents from overfitting [29], [38].

The problem can be formulated as follows. We aim at sampling a posterior distribution π on \mathbb{R}^d , $d \geq 1$, with density $x \mapsto e^{-U(x)} / \int_{\mathbb{R}^d} e^{-U(y)} dy$ w.r.t. the Lebesgue measure, where U is continuously differentiable. The Langevin stochastic differential equation associated with π is defined by:

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

where $(B_t)_{t \geq 0}$ is a d -dimensional Brownian motion 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 consider in this paper 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 [14] and [31] for molecular dynamics applications. Then it has been popularized in machine learning by [17], [18] and computational statistics by [29] and [33]. Following [33], in the sequel this method will be refer 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 of the cases differs from the distribution π . It has been proposed in [34] and [33] to use a Metropolis-Hastings step at each iteration to enforce reversibility w.r.t. π . This new algorithm is referred to as the Metropolis adjusted Langevin algorithm (MALA).

The ULA algorithm has already been studied in depth for constant sequence of step sizes in [36], [33] and [27]. In particular, [36, Theorem 4] gives an asymptotic expansion for the weak error between π and π_γ . For sequence of step sizes such that $\lim_{k \rightarrow +\infty} \gamma_k = 0$ and $\sum_{k=1}^{\infty} \gamma_k = \infty$, weak convergence of the weighted empirical distribution of the scheme have been shown in [23], [24] and [25]. Contrary to the previous cited works, we focus in this paper on non-asymptotic results. More precisely, we obtain explicit bounds between the distribution of the n^{th} iterate of the Markov chain $(X_k)_{k \geq 0}$, for $n \in \mathbb{N}$, defined by the Euler discretization and the target distribution π in Wasserstein and total variation distance for nonincreasing step sizes. When the sequence of step sizes is constant $\gamma_k = \gamma$ for all $k \geq 0$, fixed horizon and fixed precision strategies are considered. In addition, quantitative estimates between π and π_γ are obtained. When $\lim_{k \rightarrow +\infty} \gamma_k = 0$ decreases to zero 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 provides explicit convergence bounds. A particular attention is paid

to the dependency of the proposed bounds on the dimension. These results complete and improve the results obtained by [10] and [11].

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 provide non-asymptotic bounds of convergence of the weighted empirical measure applied to Lipschitz functions. In Section 4, we give non asymptotic bounds in total variation distance between the Euler discretization and π . This study is completed in Section 5 by non-asymptotic bounds of convergence of the weighted empirical measure applied to bounded and measurable functions. Our claims are supported in a Bayesian inference for a binary regression model in Section 6. The proofs are given in Section 7. Finally in Section 8, some results of independent interest, used in some proofs, on functional autoregressive models are gathered. Some technical proofs and derivations are postponed and carried out in a supplementary paper [12].

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 , $\zeta(A \times \mathbb{R}^d) = \mu(A)$ and $\zeta(\mathbb{R}^d \times A) = \nu(A)$. We denote by $\Pi(\mu, \nu)$ the set of transference plans of μ and ν . Furthermore, we say that a couple of \mathbb{R}^d -random variables (X, Y) is a coupling of μ and ν if there exists $\zeta \in \Pi(\mu, \nu)$ such that (X, Y) are distributed according to ζ . For two probability measures μ and ν , we define the Wasserstein distance of order $p \geq 1$ as

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

By [37, Theorem 4.1], for all μ, ν probability measures on \mathbb{R}^d , there exists a transference plan $\zeta^* \in \Pi(\mu, \nu)$ such that for any coupling (X, Y) distributed according to ζ^* , $W_p(\mu, \nu) = \mathbb{E}[\|X - Y\|^p]^{1/p}$. 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 $\mu \in \mathcal{P}_p(\mathbb{R}^d)$, $\int_{\mathbb{R}^d} \|x\|^p d\mu(x) < +\infty$. By [37, Theorem 6.16], $\mathcal{P}_p(\mathbb{R}^d)$ equipped with the Wasserstein distance W_p of order p is a complete separable metric space.

Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a Lipschitz function, namely there exists $C \geq 0$ such that for all $x, y \in \mathbb{R}^d$, $|f(x) - f(y)| \leq C \|x - y\|$. Then we denote

$$\|f\|_{\text{Lip}} = \inf\{|f(x) - f(y)| \|x - y\|^{-1} \mid x, y \in \mathbb{R}^d, x \neq y\}.$$

The Monge-Kantorovich theorem (see [37, Theorem 5.9]) implies that for all μ, ν probabilities measure on \mathbb{R}^d ,

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

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

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

where $D = \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d | x = y\}$. For all $x \in \mathbb{R}^d$ and $M > 0$, we denote by $B(x, M)$, the ball centered at x of radius M . For a subset $A \subset \mathbb{R}^d$, denote by A^c the complementary of 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 $\|M\|_F^2 = \text{Tr}(M^T M)$. Let $n, m \in \mathbb{N}^*$ and $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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 $i \in \{1, \dots, m\}$, the i -th component of $\vec{\Delta} F(x)$ is equals to $\sum_{j=1}^d (\partial^2 F_i / \partial x_j^2)(x)$. In the sequel, we take the convention that $\sum_p^n = 0$ 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 **H1**, for all $x \in \mathbb{R}^d$ by [22, 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 **H2**, [30, Theorem 2.1.8] shows that U has a unique minimizer $x^* \in \mathbb{R}^d$. If in addition **H1** holds, then [30, Theorem 2.1.12, Theorem 2.1.9] show that for all $x, y \in \mathbb{R}^d$:

$$\langle \nabla U(y) - \nabla U(x), y - x \rangle \geq \frac{\kappa}{2} \|y - x\|^2 + \frac{1}{m + L} \|\nabla U(y) - \nabla U(x)\|^2, \quad (3)$$

where

$$\kappa = \frac{2mL}{m + L}. \quad (4)$$

We first give a uniform bound in time on the second moment of the diffusion which implies a bound on the second moment of π .

Theorem 1. *Assume **H1** and **H2**.*

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

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

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

Proof. The proof is postponed to Section 7.1. □

In the next result we provide the geometric rate of convergence to stationarity of the semi-group in Wasserstein distance. It is worthwhile to note that these bounds do not depend on the dimension d .

Theorem 2. *Assume **H1** and **H2**.*

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

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

Proof. Most of the statement is well known; see [2] and the references therein. Nevertheless for completeness, we provide the proof in Section 7.2. □

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

$$\Gamma_{n,\ell} \stackrel{\text{def}}{=} \sum_{k=n}^{\ell} \gamma_k, \quad \Gamma_n = \Gamma_{1,n}. \quad (5)$$

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

$$R_\gamma(x, A) = \int_A (4\pi\gamma)^{-d/2} \exp\left(- (4\gamma)^{-1} \|y - x + \gamma \nabla U(x)\|^2\right) dy. \quad (6)$$

The process $(X_n)_{n \geq 0}$ given in (2) is an inhomogeneous Markov chain with respect to the family of Markov kernels $(R_{\gamma_n})_{n \geq 1}$. For $\ell, n \in \mathbb{N}^*$, $\ell \geq n$, define

$$Q_\gamma^{n,\ell} = R_{\gamma_n} \cdots R_{\gamma_\ell}, \quad Q_\gamma^n = Q_\gamma^{1,n} \quad (7)$$

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

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

Theorem 3. *Assume **H1** and **H2**.*

(i) Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Let x^* be the unique minimizer of U . Then for all $x \in \mathbb{R}^d$ and $n, \ell \in \mathbb{N}^*$, $\int_{\mathbb{R}^d} \|y - x^*\|^2 Q_\gamma^{n, \ell}(x, dy) \leq \varrho_{n, \ell}(x)$, where $\varrho_{n, \ell}(x)$ is given by

$$\varrho_{n, \ell}(x) = \prod_{k=n}^{\ell} (1 - \kappa \gamma_k) \|x - x^*\|^2 + 2d \sum_{k=n}^{\ell} \gamma_k \left\{ \prod_{i=k+1}^{\ell} (1 - \kappa \gamma_i) \right\}, \quad (8)$$

and κ is defined in (4).

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

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

Proof. The proof is postponed to Section 7.3. \square

We now proceed to establish the contraction property of the sequence $(Q_\gamma^n)_{n \geq 1}$ in W_2 . 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 is again independent of the dimension.

Theorem 4. Assume **H1** and **H2**. Then,

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

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

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

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

Proof. The proof is straightforward using the synchronous coupling but is postponed to Section 7.4 for completeness. \square

We now proceed to establish explicit bounds for $W_2(\delta_x Q_\gamma^n, \pi)$, with $x \in \mathbb{R}^d$. 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 $(Y_t, \bar{Y}_t)_{t \geq 0}$ defined for all $n \geq 0$ and $t \in [\Gamma_n, \Gamma_{n+1})$ by

$$\begin{cases} Y_t = Y_{\Gamma_n} - \int_{\Gamma_n}^t \nabla U(Y_s) ds + \sqrt{2}(B_t - B_{\Gamma_n}) \\ \bar{Y}_t = \bar{Y}_{\Gamma_n} - \nabla U(\bar{Y}_{\Gamma_n})(t - \Gamma_n) + \sqrt{2}(B_t - B_{\Gamma_n}), \end{cases} \quad (9)$$

with Y_0 is distributed according to ν_0 , $\bar{Y}_0 = x$ and $(\Gamma_n)_{n \geq 1}$ is given in (5). 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 $(\delta_x Q_\gamma^n)_{n \geq 0}$ and the stationary measure π of the Langevin diffusion (1).

Theorem 5. Assume **H1** and **H2**. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$. Then for all $x \in \mathbb{R}^d$ and $n \geq 1$,

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

where

$$u_n^{(1)}(\gamma) = 2 \prod_{k=1}^n (1 - \kappa \gamma_k / 2) \quad (10)$$

κ is defined in (4) and

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

Proof. The proof is postponed to Section 7.5. \square

Corollary 6. Assume **H1** and **H2**. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$. Assume that $\lim_{k \rightarrow \infty} \gamma_k = 0$ and $\lim_{n \rightarrow +\infty} \Gamma_n = +\infty$. Then for all $x \in \mathbb{R}^d$, $\lim_{n \rightarrow \infty} W_2(\delta_x Q_\gamma^n, \pi) = 0$.

Proof. The proof is postponed to Section 7.6. \square

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 **H1** and **H2**. 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

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

Proof. The proof is postponed to Section 7.7. \square

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 $x, y \in \mathbb{R}^d$, $\|\nabla^2 U(x) - \nabla^2 U(y)\| \leq \tilde{L} \|x - y\|$.

Note that under **H1** and **H3**, we have that for all $x, y \in \mathbb{R}^d$,

$$\|\nabla^2 U(x)y\| \leq L \|y\|, \quad \left\| \vec{\Delta}(\nabla U)(x) \right\|^2 \leq d\tilde{L}^2. \quad (12)$$

Theorem 8. Assume **H1**, **H2** and **H3**. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$. Then for all $x \in \mathbb{R}^d$ and $n \geq 1$,

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

where $u_n^{(1)}$ is given by (10), κ in (4) and

$$u_n^{(3)}(\gamma) = \sum_{i=1}^n \left[d\gamma_i^3 \left\{ 2L^2 + \kappa^{-1} \left(\frac{\tilde{L}^2}{3} + \gamma_i L^4 + \frac{4L^4}{3m} \right) + \gamma_i L^4 \left(\frac{\gamma_i}{6} + m^{-1} \right) \right\} \times \prod_{k=i+1}^n \left(1 - \frac{\kappa\gamma_k}{2} \right) \right]. \quad (13)$$

Proof. The proof is postponed to Section 7.8. \square

In the case of constant step sizes $\gamma_k = \gamma$ 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 **H1**, **H2** and **H3**. 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*

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

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

Note that the bounds provided by Theorem 5 and Theorem 8 scale linearly with the dimension d . Using Theorem 4-(ii) and Corollary 6 or Corollary 9, given $\varepsilon > 0$, we determine the smallest number of iterations n_ε and an associated step-size γ_ε to approach the stationary distribution in the Wasserstein distance with a precision ε . The dependencies on the dimension d and the precision ε of n_ε based on Theorem 5 and Theorem 8 are reported in Table 1. Details and further discussions are included in the supplementary paper [12].

Parameter	d	ε
Theorem 5 and Theorem 4-(ii)	$\mathcal{O}(d \log(d))$	$\mathcal{O}(\varepsilon^{-2} \log(\varepsilon))$
Theorem 8 and Theorem 4-(ii)	$\mathcal{O}(d^{1/2} \log(d))$	$\mathcal{O}(\varepsilon^{-1} \log(\varepsilon))$

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

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 $\alpha \in (0, 1]$. The rate of the bounds provided by Theorem 5 and Theorem 8 are given in Table 2, see [12] for details.

	$\alpha \in (0, 1)$	$\alpha = 1$
Theorem 5	$d \mathcal{O}(n^{-\alpha})$	$d \mathcal{O}(n^{-1})$ for $\gamma_1 > 2\kappa^{-1}$ see [12, Section 3]
Theorem 8	$d \mathcal{O}(n^{-2\alpha})$	$d \mathcal{O}(n^{-2})$ for $\gamma_1 > 2\kappa^{-1}$ see [12, Section 2]

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

3 Mean square error and concentration for Lipschitz functions

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

$$\hat{\pi}_n^N(f) = \sum_{k=N+1}^{N+n} \omega_{k,n}^N f(X_k), \quad \omega_{k,n}^N = \gamma_{k+1} \Gamma_{N+2, N+n+1}^{-1}. \quad (14)$$

where $N \geq 0$ is the length of the burn-in period, $n \geq 1$ is the number of samples, and for $n, p \in \mathbb{N}$, $\Gamma_{n,p}$ is given by (5). In all this section, \mathbb{P}_x and \mathbb{E}_x denote the probability and the expectation respectively, induced on $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\mathbb{N}})$ by the Markov chain $(X_n)_{n \geq 0}$ started at $x \in \mathbb{R}^d$. We first compute an explicit bound for the Mean Squared Error (MSE) of this estimator defined by:

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

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

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

Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$ and recall that x^* is the unique minimizer of U . Using Theorem 5 and Theorem 8, we end up with the following bound:

$$\left\{ \mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f) \right\}^2 \leq \|f\|_{\text{Lip}}^2 \sum_{k=N+1}^{N+n} \omega_{k,n}^N \left\{ (\|x - x^*\|^2 + d/m) u_k^{(1)}(\gamma) + w_k(\gamma) \right\}, \quad (15)$$

where $u_n^{(1)}(\gamma)$ is given in (10) and $w_n(\gamma)$ is equal to $u_n^{(2)}(\gamma)$ defined by (11) if **H1-H2** hold, and to $u_n^{(3)}(\gamma)$, defined by (13), if **H1-H2** and **H3** hold.

Consider now the variance term. To control this term, we adapt the proof of [21, Theorem 2] for homogeneous Markov chain to our inhomogeneous setting, and we have:

Theorem 10. *Assume **H1** and **H2**. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Then for all $N \geq 0$, $n \geq 1$ and Lipschitz functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$, we get $\text{Var}_x \{ \hat{\pi}_n^N(f) \} \leq 8\kappa^{-2} \|f\|_{\text{Lip}}^2 \Gamma_{N+2, N+n+1}^{-1} v_{N,n}(\gamma)$, where*

$$v_{N,n}(\gamma) \stackrel{\text{def}}{=} \left\{ 1 + \Gamma_{N+2, N+n+1}^{-1} (\kappa^{-1} + 2/(m+L)) \right\}. \quad (16)$$

Proof. The proof is postponed to Section 7.9. \square

It is worth to observe that the bound for the variance is independent from the dimension.

We may now discuss the bounds on the MSE (obtained by combining the bounds for the squared bias (15) and the variance Theorem 10) for step sizes given for $k \geq 1$ by $\gamma_k = \gamma_1 k^{-\alpha}$ where $\alpha \in [0, 1]$ and $\gamma_1 < 1/(m + L)$. Details of these calculations are included in the supplementary paper [12, Section 4.1-4.2]. The order of the bounds (up to numerical constants) of the MSE are summarized in Table 3 as a function of γ_1 , n and N . Note that in the infinite horizon setting, it is optimal to take $\alpha = 1/2$ under **H 1** and **H2**, and $\alpha = 1/3$ under **H1**, **H2** and **H3**.

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

Table 3: Bound for the MSE for $\gamma_k = \gamma_1 k^{-\alpha}$ for fixed γ_1 and N under **H1** and **H2**

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

Table 4: Bound for the MSE for $\gamma_k = \gamma_1 k^{-\alpha}$ for fixed γ_1 and N under **H1**, **H2** and **H3**

If the total number of iterations $n + N$ is held fixed (fixed horizon setting), we may optimize the value of the step size γ_1 but also of the burn-in period N to minimize the upper bound on the MSE. The order (in n) for different values of $\alpha \in [0, 1]$ are summarized in Table 5 (we display the order in n but not the constants, which are quite involved).

Let us discuss first the bounds based on Theorem 5. This time for any $\alpha \in [0, 1/2)$, we can always achieve a MSE of order $n^{-1/2}$ by choosing appropriately γ_1 and N (for $\alpha = 1/2$ we have only $\log(n) n^{-1/2}$). For $\alpha \in (1/2, 1]$, the best strategy is to take $N = 0$ and the largest possible value for $\gamma_1 = 1/(m + L)$, which leads to a MSE of order $n^{\alpha-1}$ for $\alpha \in (0, 1/2)$ and $\log(n)$ for $\alpha = 1$. We now discuss the bounds provided by Theorem 8. It appears that, for any $\alpha \in [0, 1/3)$, we can always achieved the order $n^{-2/3}$ by choosing appropriately γ_1 and N (for $\alpha = 1/3$ we have only $\log^{1/3}(n) n^{-2/3}$). The worst case is for $\alpha \in (1/3, 1]$, where in fact the best strategy is to take $N = 0$ and the largest possible value for $\gamma_1 = 1/(m + L)$.

	H1, H2 and H3	H1, H2 and H3
$\alpha = 0$	$n^{-1/2}$	$n^{-2/3}$
$\alpha \in (0, 1/2)$	$n^{-1/2}$	$n^{-2/3}$
$\alpha = 1/2$	$\log(n)n^{-1/2}$	$\log^{1/3}(n)n^{-2/3}$
$\alpha \in (1/2, 1)$	$n^{\alpha-1}$	$n^{\alpha-1}$
$\alpha = 1$	$\log(n)$	$\log(n)$

Table 5: Optimal bound for the MSE by choosing γ_1

We can also follow the proof of [21, Theorem 5] to establish an exponential deviation inequality for $\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)]$ given by (14).

Theorem 11. *Assume **H1** and **H2**. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m + L)$. Then for all $N \geq 0$, $n \geq 1$, $r > 0$ and Lipschitz functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$:*

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

where $v_{N,n}(\gamma)$ is defined by (16).

Proof. The proof is postponed to the supplementary document Section 7.10. \square

If we apply this result to the sequence $(\gamma_k)_{k \geq 1}$ defined for all $k \geq 1$ by $\gamma_k = \gamma_1 k^{-\alpha}$, for $\alpha \in [0, 1]$, we end up with a concentration of order $\exp(-Cr^2 \gamma_1 n^{1-\alpha})$ for $\alpha \in [0, 1)$, for some constant $C \geq 0$ independent of γ_1 and n .

4 Quantitative bounds in total variation distance

We deal in this section with quantitative bounds in total variation distance. For Bayesian inference application, this kinds of bounds are of utmost interest 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 $(P_t)_{t \geq 0}$. Under **H2**, define χ_m for all $t \geq 0$ by

$$\chi_m(t) = \sqrt{(4/m)(e^{2mt} - 1)}.$$

Theorem 12. *Assume **H1** and **H2**.*

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

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

where Φ is the cumulative distribution function of the standard normal distribution.

(ii) *For any $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ and $t > 0$, $\|\mu P_t - \nu P_t\|_{\text{TV}} \leq W_1(\mu, \nu)/\chi_m(t)$.*

(iii) For any $x \in \mathbb{R}^d$ and $t \geq 0$, $\|\pi - \delta_x P_t\|_{\text{TV}} \leq \{(d/m)^{1/2} + \|x - x^*\|\}/\chi_m(t)$.

Proof. The proof is based on the reflexion coupling introduced in [26] and is given in Section 7.11. \square

Since for all $s > 0$, $s \leq e^s - 1$, note that Theorem 12-(ii) implies that for all $t > 0$ and $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$,

$$\|\mu P_t - \nu P_t\|_{\text{TV}} \leq (4\pi t)^{-1/2} W_1(\mu, \nu). \quad (17)$$

For all $n, \ell \geq 1$, $n < \ell$ denote by

$$\Lambda_{n,\ell} = \kappa^{-1} \left\{ \prod_{j=n}^{\ell} (1 - \kappa\gamma_j)^{-1} - 1 \right\}, \quad \Lambda_\ell = \Lambda_{1,\ell}. \quad (18)$$

Theorem 13. Assume **H1** and **H2**.

(i) Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing 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

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

(ii) Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence satisfying $\gamma_1 \leq 2/(m+L)$. For all $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ and $\ell, n \in \mathbb{N}^*$, $n < \ell$, $\|\mu Q_\gamma^{n,\ell} - \nu Q_\gamma^{n,\ell}\|_{\text{TV}} \leq (4\pi\Lambda_{n,\ell})^{-1/2} W_1(\mu, \nu)$.

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

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

Proof. The proof is postponed to Section 7.12. \square

We can combine this result and Theorem 5 or Theorem 8 to get explicit bounds in total variation between the Euler-Maruyama discretization and the target distribution π . For this we will always use the following decomposition, for all nonincreasing sequence $(\gamma_k)_{k \geq 0}$, initial point $x \in \mathbb{R}^d$ and $n \geq 0$:

$$\|\pi - \delta_x Q_\gamma^n\|_{\text{TV}} \leq \|\pi - \delta_x P_{\Gamma_n}\|_{\text{TV}} + \|\delta_x P_{\Gamma_n} - \delta_x Q_\gamma^n\|_{\text{TV}}. \quad (19)$$

The first term is dealt with Theorem 12-(iii). It remains to bound the second term in (19). Since we will use Theorem 5 and Theorem 8, we have two different results depending on the assumptions on U again. Define for all $x \in \mathbb{R}^d$ and $n, \ell \in \mathbb{N}$,

$$\begin{aligned} \vartheta_{n,\ell}^{(1)}(x) = L^2 \sum_{i=1}^n \gamma_i^2 \prod_{k=i+1}^n (1 - \kappa\gamma_k/2) & \left[\{\kappa^{-1} + \gamma_i\} (2d + dL^2\gamma_i^2/6) \right. \\ & \left. + L^2\gamma_i\delta_{i,n,\ell}(x) \{\kappa^{-1} + \gamma_i\} \right] \end{aligned} \quad (20)$$

$$\begin{aligned} \vartheta_{n,\ell}^{(2)}(x) &= \sum_{i=1}^n \gamma_i^3 \prod_{k=i+1}^n (1 - \kappa\gamma_k/2) [L^4\delta_{i,n,\ell}(x)(4\kappa^{-1}/3 + \gamma_{n+1}) \\ &\quad + d \left\{ 2L^2 + 4\kappa^{-1}(\tilde{L}^2/12 + \gamma_{n+1}L^4/4) + \gamma_{n+1}^2L^4/6 \right\}] , \end{aligned} \quad (21)$$

where $\varrho_{n,p}(x)$ is given by (8) and

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

Theorem 14. Assume **H1** and **H2**. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$. Then for all $x \in \mathbb{R}^d$ and $\ell, n \in \mathbb{N}^*$, $\ell > n$,

$$\begin{aligned} \|\delta_x P_{\Gamma_\ell} - \delta_x Q_\gamma^\ell\|_{\text{TV}} &\leq (\vartheta_n(x)/(4\pi\Gamma_{n+1,\ell}))^{1/2} \\ &\quad + 2^{-3/2}L \left(\sum_{k=n+1}^{\ell} \{(\gamma_k^3 L^2/3)\varrho_{1,k-1}(x) + d\gamma_k^2\} \right)^{1/2} . \end{aligned} \quad (22)$$

where $\varrho_{1,n}(x)$ is defined by (8), $\vartheta_n(x)$ is equal to $\vartheta_{n,0}^{(2)}(x)$ given by (21), if **H3** holds, and to $\vartheta_{n,0}^{(1)}(x)$ given by (20) otherwise.

Proof. The proof is postponed to Section 7.13. \square

Consider the case of decreasing step sizes defined: $\gamma_k = \gamma_1 k^{-\alpha}$ for $k \geq 1$ and $\alpha \in (0, 1]$. Under **H1** and **H2**, choosing $n = \ell - \lfloor \ell^\alpha \rfloor$ in the bound given by Theorem 14 and using Table 2 implies that $\|\delta_{x^*} Q_\gamma^\ell - \pi\|_{\text{TV}} = d^{1/2} \mathcal{O}(\ell^{-\alpha/2})$. Note that for $\alpha = 1$, this rate is true only for $\gamma_1 > 2\kappa^{-1}$. If in addition **H3** holds, choosing $n = \ell - \lfloor \ell^{\alpha/2} \rfloor$ in the bound given by (22) and using Table 2, (19) and Theorem 12-(iii) implies that $\|\delta_{x^*} Q_\gamma^\ell - \pi\|_{\text{TV}} = d^{1/2} \mathcal{O}(\ell^{-(3/4)\alpha})$. These conclusions and the dependency on the dimension are summarized in Table 6.

	H1, H2	H1, H2 and H3
$\gamma_k = \gamma_1 k^{-\alpha}, \alpha \in (0, 1]$	$d^{1/2} \mathcal{O}(\ell^{-\alpha/2})$	$d^{1/2} \mathcal{O}(\ell^{-3\alpha/4})$

Table 6: Order of convergence of $\|\delta_{x^*} Q_\gamma^\ell - \pi\|_{\text{TV}}$ for $\gamma_k = \gamma_1 k^{-\alpha}$ based on Theorem 14

When $\gamma_k = \gamma \in (0, 1/(m+L))$ for all $k \geq 1$, under **H1** and **H2**, for $\ell > \lceil \gamma^{-1} \rceil$ choosing $n = \ell - \lceil \gamma^{-1} \rceil$ implies that (see the supplementary document [12, Section 1.1])

$$\|\delta_x R_\gamma^\ell - \delta_x P_{\ell\gamma}\|_{\text{TV}} \leq (4\pi)^{-1/2} [\gamma D_1(\gamma, d) + \gamma^3 D_2(\gamma) D_3(\gamma, d, x)]^{1/2} + D_4(\gamma, d, x) , \quad (23)$$

where

$$D_1(\gamma, d) = 2L^2 \kappa^{-1} (\kappa^{-1} + \gamma) (2d + L^2 \gamma^2/6) , D_2(\gamma) = L^4 (\kappa^{-1} + \gamma) \quad (24)$$

$$\begin{aligned}
D_3(\gamma, d, x) &= \left\{ (\ell - \lceil \gamma^{-1} \rceil) e^{-m\gamma(\ell - \lceil \gamma^{-1} \rceil - 1)} \|x - x^*\|^2 + 2d(\kappa\gamma m)^{-1} \right\} \\
D_4(\gamma, d, x) &= 2^{-3/2} L [d(1 + \gamma) \\
&\quad + (L^2\gamma^3/3) \left\{ (1 + \gamma^{-1})(1 - \kappa\gamma)^{\ell - \lceil \gamma^{-1} \rceil} \|x - x^*\|^2 + 2(1 + \gamma)\kappa^{-1}d \right\}]^{1/2}.
\end{aligned}$$

Using this bound and Theorem 12-(iii), the minimal number of iterations $\ell_\varepsilon > 0$ for a target precision $\varepsilon > 0$ to get $\|\delta_{x^*} R_{\gamma_\varepsilon}^{\ell_\varepsilon} - \pi\|_{\text{TV}} \leq \varepsilon$ 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 8). In addition, letting ℓ go to infinity in (23) we get the following result.

Corollary 15. *Assume H1 and H2. Let $\gamma \in (0, 1/(m + L)]$. Then it holds*

$$\begin{aligned}
\|\pi_\gamma - \pi\|_{\text{TV}} &\leq 2^{-3/2} L [d(1 + \gamma) + 2(L^2\gamma^3/3)(1 + \gamma)\kappa^{-1}d]^{1/2} \\
&\quad + (4\pi)^{-1/2} [\gamma D_1(\gamma, d) + 2d\gamma^2 D_2(\gamma)(\kappa m)^{-1}]^{1/2},
\end{aligned}$$

where $D_1(\gamma)$ and $D_2(\gamma)$ are given in (24).

Note that the bound provided by Corollary 15 is of order $d^{1/2} \mathcal{O}(\gamma^{1/2})$.

If in addition H3 holds, for constant step sizes, we can improve the bounds given by Corollary 15.

Theorem 16. *Assume H1, H2 and H3. Let $\gamma \in (0, 1/(m + L)]$. Then it holds*

$$\begin{aligned}
\|\pi_\gamma - \pi\|_{\text{TV}} &\leq (4\pi)^{-1/2} \left\{ \gamma^2 E_1(\gamma, d) + 2d\gamma^2 E_2(\gamma)/(\kappa m) \right\}^{1/2} \\
&\quad + (4\pi)^{-1/2} \lceil \log(\gamma^{-1}) / \log(2) \rceil \left\{ \gamma^2 E_1(\gamma, d) + \gamma^2 E_2(\gamma)(2\kappa^{-1}d + d/m) \right\}^{1/2} \\
&\quad + 2^{-3/2} L \left\{ 2d\gamma^3 L^2 / (3\kappa) + d\gamma^2 \right\}^{1/2},
\end{aligned}$$

where $E_1(\gamma, d)$ and $E_2(\gamma)$ are defined by

$$\begin{aligned}
E_1(\gamma, d) &= 2d\kappa^{-1} \left\{ 2L^2 + 4\kappa^{-1}(\tilde{L}^2/12 + \gamma L^4/4) + \gamma^2 L^4/6 \right\} \\
E_2(\gamma) &= L^4(4\kappa^{-1}/3 + \gamma).
\end{aligned}$$

Proof. The proof is postponed to Section 7.14. □

Note that the bound provided by Theorem 16 is of order $d^{1/2} \mathcal{O}(\gamma |\log(\gamma)|)$ and this result is sharp up to a logarithmic factor. Indeed, if π is the d -dimensional standard Gaussian distribution, then π_γ is also a d -dimensional standard Gaussian distribution with zero-mean and covariance matrix $\sigma_\gamma^2 \mathbf{I}_d$, with $\sigma_\gamma^2 = (1 - \gamma/2)^{-1}$. Therefore using the Pinsker inequality, we get $\|\pi - \pi_\gamma\|_{\text{TV}} \leq 2^{-3/2} \gamma d^{1/2}$.

We can also for a precision target $\varepsilon > 0$ choose $\gamma_\varepsilon > 0$ and the number of iterations $n_\varepsilon > 0$ to get $\|\delta_x R_{\gamma_\varepsilon}^{n_\varepsilon} - \pi\|_{\text{TV}} \leq \varepsilon$. By Theorem 12-(iii), Theorem 13-(iii) and Theorem 16, the minimal number of iterations ℓ_ε is of order $d^{1/2} \log^2(d) \mathcal{O}(\varepsilon^{-1} \log^2(\varepsilon))$ for a well chosen step size γ_ε . The discussions on the bounds for constant sequences of step sizes are summarized in Table 7 and Table 8.

	H1, H2	H1, H2 and H3
$\ \pi - \pi_\gamma\ _{\text{TV}}$	$d^{1/2}\mathcal{O}(\gamma^{1/2})$	$d^{1/2}\mathcal{O}(\gamma \log(\gamma))$

Table 7: 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	$d^{-1}\mathcal{O}(\varepsilon^2)$	$d^{-1/2}\log^{-1}(d)\mathcal{O}(\varepsilon \log^{-1}(\varepsilon))$
n_ε	$d\log(d)\mathcal{O}(\varepsilon^{-2} \log(\varepsilon))$	$d^{1/2}\log^2(d)\mathcal{O}(\varepsilon^{-1}\log^2(\varepsilon))$

Table 8: Order of the step size $\gamma_\varepsilon > 0$ and the number of iterations $n_\varepsilon \in \mathbb{N}^*$ to get $\|\delta_{x^\star}R_{\gamma_\varepsilon}^{n_\varepsilon} - \pi\|_{\text{TV}} \leq \varepsilon$ for $\varepsilon > 0$.

5 Mean square error and concentration for bounded measurable functions

The result of the previous section allows us to study the approximation of $\int_{\mathbb{R}^d} f(y)\pi(dy)$ by the weighted average estimator $\hat{\pi}_n^N(f)$ defined by (14) for a measurable and bounded function $f : \mathbb{R}^d \rightarrow \mathbb{R}$. We use the same notation as in Section 3. We first obtain an elementary bound for the bias term in the MSE. By the Jensen inequality and because f is bounded, we have:

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

Using the results of Section 4, we can deduce different bounds for the bias, depending on the assumptions on U and the sequence of step sizes $(\gamma_k)_{k \geq 1}$.

The following result gives a bound on the variance term.

Theorem 17. *Assume **H1** and **H2**. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Then for all $N \geq 0$, $n \geq 1$, $x \in \mathbb{R}^d$ and $f \in \mathbb{F}_b(\mathbb{R}^d)$, we get $\text{Var}_x\{\hat{\pi}_n^N(f)\} \leq 2\text{osc}(f)^2\{\Gamma_{N+2, N+n+1}^{-1} + u_{N,n}^{(4)}(\gamma)\}$, where*

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

for $n_1, n_2 \in \mathbb{N}$, Λ_{n_1, n_2} is given by (18).

Proof. The proof is postponed to Section 7.15. □

We now establish an exponential deviation inequality for $\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)]$ given by (14) for a bounded measurable function f .

Theorem 18. Assume **H 1** and **H 2**. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Let $(X_n)_{n \geq 0}$ be given by (2) and started at $x \in \mathbb{R}^d$. Then for all $N \geq 0$, $n \geq 1$, $r > 0$, and functions $f \in \mathbb{F}_b(\mathbb{R}^d)$:

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

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 in the supplementary document to Section 7.16. \square

6 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 success probability $\{\varrho(\beta^T X_i)\}_{i=1}^p$, where ϱ is the logistic function defined for $z \in \mathbb{R}$ by $\varrho(z) = e^z / (1 + e^z)$ and $\{X_i\}_{i=1}^p$ and β are d dimensional vectors of known covariates and unknown regression coefficients, respectively. The prior distribution for the parameter β is a zero-mean Gaussian distribution with covariance matrix Σ_β . The density of the posterior distribution of β is up to a proportionality constant given by

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

Bayesian inference for the logistic regression model has long been recognized as a numerically involved problem, due to the analytically inconvenient form of the likelihood function. Several algorithms have been proposed, trying to mimick the data-augmentation (DA) approach of [1] for probit regression; see [20], [15] and [16]. Recently, a very promising DA algorithm has been proposed in [32], using the Polya-Gamma distribution in the DA part. This algorithm has been shown to be uniformly ergodic for the total variation by [8, Proposition 1], which provides an explicit expression for the ergodicity constant. This constant is exponentially small in the dimension of the parameter space and the number of samples. 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 $\pi_\beta(\cdot | \{(X_i, Y_i)\}_{i=1}^p)$. The gradient of its log-density may be expressed as

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

Therefore $-\log \pi_{\beta}(\cdot | \{X_i, Y_i\}_{i=1}^p)$ is strongly convex **H2** with $m = \lambda_{\max}^{-1}(\Sigma_{\beta})$ and satisfies **H1** with $L = (1/4) \sum_{i=1}^p X_i^T X_i + \lambda_{\min}^{-1}(\Sigma_{\beta})$, where $\lambda_{\min}(\Sigma_{\beta})$ and $\lambda_{\max}(\Sigma_{\beta})$ denote the minimal and maximal eigenvalues of Σ_{β} , respectively. We first compare the histograms produced by ULA and the Pólya-Gamma Gibbs sampling from [32]. 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 $\Sigma_{\beta}^{-1} = (dp)^{-1}(\sum_{i=1}^p X_i^T X_i) I_d$. We produce 10^8 samples from the Pólya-Gamma sampler using the R package `BayesLogit` [39]. 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 1. The same procedure is also applied with the decreasing step size sequence $(\gamma_k)_{k \geq 1}$ defined by $\gamma_k = \gamma_1 k^{-1/2}$, with $\gamma_1 = 10(\kappa \log(n)^{1/2})^{-1}$ and for the burn in period $N = \log(n)$, see also Figure 1. In addition, we also compare MALA and ULA on five real data sets, which are

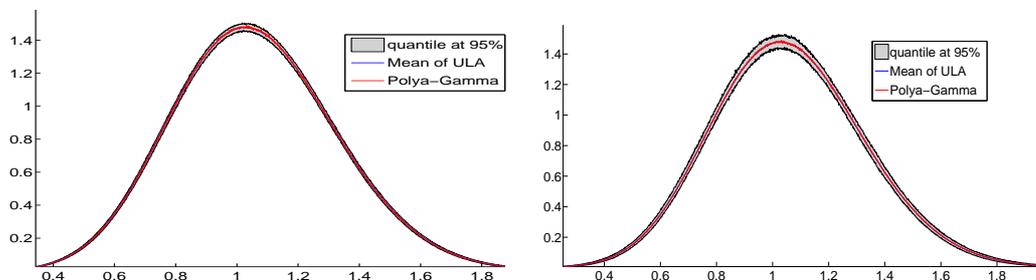


Figure 1: Empirical distribution comparison between the Pólya-Gamma Gibbs Sampler and ULA. Left panel: constant step size $\gamma_k = \gamma_1$ for all $k \geq 1$; right panel: decreasing step size $\gamma_k = \gamma_1 k^{-1/2}$ for all $k \geq 1$

summarized in Table 9. Note that for the Australian credit data set, the ordinal covariates have been stratified by dummy variables. Furthermore, we normalized the data sets and consider the Zellner prior setting $\Sigma^{-1} = (\pi^2 d/3) \Sigma_X^{-1}$ where $\Sigma_X = p^{-1} \sum_{i=1}^p X_i X_i^T$; see [35], [19] and the references therein. Also, we apply a pre-conditioned version of MALA and ULA, targeting the probability density $\tilde{\pi}_{\beta}(\cdot) \propto \pi_{\beta}(\Sigma_X^{1/2} \cdot)$. Then, we obtain samples from π_{β} 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 $i \in \{1, \dots, d\}$ 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

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

This quantity has already been considered in [6] and [9] to compare approximate samplers. To estimate the d marginal posterior distributions, we run $2 \cdot 10^7$ iterations of the Pólya-Gamma Gibbs sampler. Then 100 runs of MALA and ULA (10^6 iterations

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 9: Dimension of the data sets

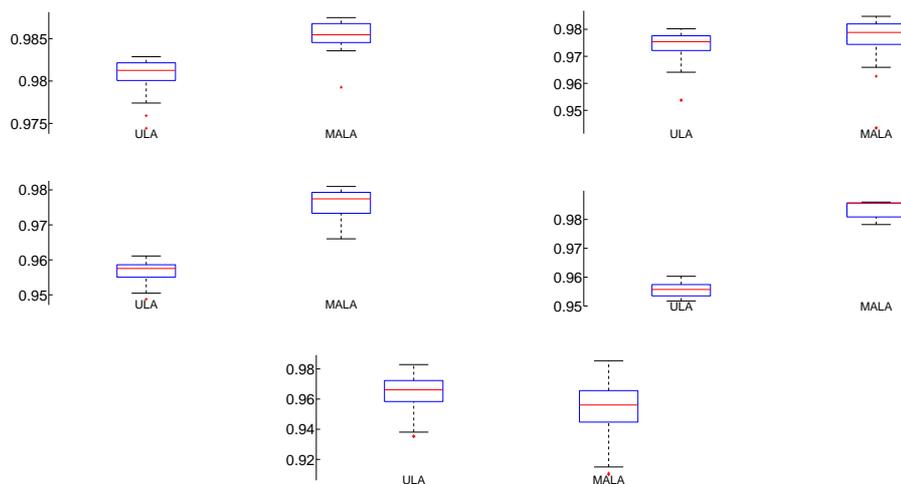


Figure 2: Marginal accuracy across all the dimensions.

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

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 2. 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+\(German+Credit+Data\)](http://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data))

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

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

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

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

7 Proofs

7.1 Proof of Theorem 1

(i) The generator \mathcal{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:

$$\mathcal{A}f(y) = -\langle \nabla U(y), \nabla f(y) \rangle + \Delta f(y). \quad (25)$$

Denote for all $y \in \mathbb{R}^d$ by $V(y) = \|y - x^*\|^2$. Let $x \in \mathbb{R}^d$ and $(Y_t)_{t \geq 0}$ be a solution of (1) started at x . Under **H1** $\sup_{t \in [0, T]} \mathbb{E}[\|Y_t\|^2] < +\infty$ for all $T \geq 0$. Therefore, the process

$$\left(V(Y_t) - V(x) - \int_0^t \mathcal{A}V(Y_s) ds \right)_{t \geq 0}$$

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, $\partial v(t, x) / \partial t = P_t \mathcal{A}V(x)$.

Since $\nabla U(x^*) = 0$ and by **H2**, $\langle \nabla U(x) - \nabla U(x^*), x - x^* \rangle \geq m \|x - x^*\|^2$, we have

$$\mathcal{A}V(x) = 2(-\langle \nabla U(x) - \nabla U(x^*), x - x^* \rangle + d) \leq 2(-mV(x) + d). \quad (26)$$

Therefore, we get

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

and the proof follows from the Grönwall inequality.

(ii) Set $V(x) = \|x - x^*\|^2$. By Jensen's inequality and Lemma 19-(i), for all $c > 0$ and $t > 0$, we get

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

Taking the limit as $t \rightarrow +\infty$, we get $\pi(V \wedge c) \leq d/m$. Using the monotone convergence theorem and taking the limit as $c \rightarrow +\infty$ concludes the proof.

7.2 Proof of Theorem 2

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

$$\begin{cases} dY_t &= -\nabla U(Y_t) dt + \sqrt{2} dB_t, \\ d\tilde{Y}_t &= -\nabla U(\tilde{Y}_t) dt + \sqrt{2} dB_t, \end{cases} \quad (27)$$

where $(Y_0, \tilde{Y}_0) = (x, y)$. Since ∇U is Lipschitz, then by [22, 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 (27),

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

which implies using **H2** and Grönwall's inequality that

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

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.

7.3 Proof of Theorem 3

(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 $x \in \mathbb{R}^d$:

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

Using that $\nabla U(x^*) = 0$, and (3), we get from the previous inequality:

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

where we have used for the last inequality that $\gamma_1 \leq 2/(m+L)$ and $(\gamma_k)_{k \geq 1}$ is nonincreasing. Then by definition (7) 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 $n \geq 1$,

$$\begin{aligned} \int_{\mathbb{R}^d} \|y - x^*\|^2 R_\gamma^n(x, dy) & \leq (1 - \kappa\gamma)^n \|x - x^*\|^2 + 2d \sum_{k=1}^n \gamma (1 - \kappa\gamma)^{n-k} \\ & \leq (1 - \kappa\gamma)^n \|x - x^*\|^2 + 2\kappa^{-1}d(1 - (1 - \kappa\gamma)^n). \end{aligned} \quad (28)$$

Since any compact set of \mathbb{R}^d is small with respect to the Lebesgues for R_γ , then [28, Theorem 15.0.1] implies that R_γ has a unique stationary distribution π_γ . Using (28), the proof is along the same line as the one of Theorem 1-(ii).

7.4 Proof of Theorem 4

(i) $(Z_k)_{k \geq 1}$ be a sequence of i.i.d. d -dimensional Gaussian random variables. We consider the processes $(X_k^{n,1}, X_k^{n,2})_{k \geq 0}$ with $(X_0^{n,1}, X_0^{n,2}) = (x, y)$ and defined for $k \geq 0$ by

$$X_{k+1}^{n,j} = X_k^{n,j} - \gamma_{k+n} \nabla U(X_k^{n,j}) + \sqrt{2\gamma_{k+n}} Z_{k+1} \quad j = 1, 2. \quad (29)$$

Using (29), 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 (3) implies for $k \geq n - 1$,

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

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

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

(ii) Let $\mu \in \mathcal{P}_{2p}(\mathbb{R}^d)$ and $p \geq 1$. It is straightforward that for all $n \geq 0$, $\mu R_\gamma^n \in \mathcal{P}_{2p}(\mathbb{R}^d)$. Then, by Theorem 4 for $\gamma \leq 2/(m+L)$, R_γ is a strict contraction in $(\mathcal{P}_{2p}(\mathbb{R}^d), W_{2p})$ and there is a unique fixed point π_γ which is the unique invariant distribution. (ii) follows from Theorem 4.

7.5 Proof of Theorem 5

We preface the proof by two technical Lemmata.

Lemma 19. *Let $(Y_t)_{t \geq 0}$ be the solution of (1) started at $x \in \mathbb{R}^d$. For all $t \geq 0$ and $x \in \mathbb{R}^d$,*

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

Proof. Let \mathcal{A} be the generator associated with $(P_t)_{t \geq 0}$ defined by (26). Denote for all $x, y \in \mathbb{R}^d$, $\tilde{V}_x(y) = \|y - x\|^2$. Note that the process $(\tilde{V}_x(Y_t) - \tilde{V}_x(x) - \int_0^t \mathcal{A} \tilde{V}_x(Y_s) ds)_{t \geq 0}$, is a $(\mathcal{F}_t)_{t \geq 0}$ -martingale under \mathbb{P}_x . Denote for all $t \geq 0$ and $x \in \mathbb{R}^d$ by $\tilde{v}(t, x) = P_t \tilde{V}_x(x)$. Then we get,

$$\frac{\partial \tilde{v}(t, x)}{\partial t} = P_t \mathcal{A} \tilde{V}_x(x). \quad (30)$$

By **H2**, 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

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

Using (30), this inequality and that \tilde{V}_x is positive, we get

$$\frac{\partial \tilde{v}(t, x)}{\partial t} = P_t \mathcal{A} \tilde{V}_x(x) \leq 2 \left(d - \int_{\mathbb{R}^d} \langle \nabla U(x), y - x \rangle P_t(x, dy) \right). \quad (31)$$

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

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

Furthermore, by **H1** and Theorem 1-(i), we have

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

where we used for the last line that by the Taylor theorem with remainder term, for all $s \geq 0$, $(1 - e^{-2ms})/(2m) \leq s$ and $(2ms + e^{-2ms} - 1)/(2m) \leq ms^2$, and the inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$. Plugging this upper bound in (31), and since $2 \|x - x^*\| t^{3/2} d^{1/2} \leq t \|x - x^*\|^2 + t^2 d$, we get

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

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

Let $(\mathcal{F}_t)_{t \geq 0}$ be the filtration associated with $(B_t)_{t \geq 0}$ and (Y_0, \bar{Y}_0) .

Lemma 20. *Assume **H1** and **H2**. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$. Let $\zeta_0 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$, $(Y_t, \bar{Y}_t)_{t \geq 0}$ such that (Y_0, \bar{Y}_0) is distributed according to ζ_0 and given by (9). Then almost surely for all $n \geq 0$ and $\epsilon > 0$,*

$$\begin{aligned} \|Y_{\Gamma_{n+1}} - \bar{Y}_{\Gamma_{n+1}}\|^2 &\leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon)\} \|Y_{\Gamma_n} - \bar{Y}_{\Gamma_n}\|^2 \\ &\quad + (2\gamma_{n+1} + (2\epsilon)^{-1}) \int_{\Gamma_n}^{\Gamma_{n+1}} \|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2 ds, \end{aligned} \quad (32)$$

$$\begin{aligned} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\|Y_{\Gamma_{n+1}} - \bar{Y}_{\Gamma_{n+1}}\|^2 \right] &\leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon)\} \|Y_{\Gamma_n} - \bar{Y}_{\Gamma_n}\|^2 \\ &\quad + L^2 \gamma_{n+1}^2 (1/(4\epsilon) + \gamma_{n+1}) \left(2d + L^2 \gamma_{n+1} \|Y_{\Gamma_n} - x^*\|^2 + dL^2 \gamma_{n+1}^2 / 6 \right). \end{aligned} \quad (33)$$

Proof. Let $n \geq 0$ and $\epsilon > 0$, and set $\Theta_n = Y_{\Gamma_n} - \bar{Y}_{\Gamma_n}$. We first show (32). By definition we have:

$$\begin{aligned} \|\Theta_{n+1}\|^2 &= \|\Theta_n\|^2 + \left\| \int_{\Gamma_n}^{\Gamma_{n+1}} \{\nabla U(Y_s) - \nabla U(\bar{Y}_{\Gamma_n})\} ds \right\|^2 \\ &\quad - 2\gamma_{n+1} \langle \Theta_n, \nabla U(Y_{\Gamma_n}) - \nabla U(\bar{Y}_{\Gamma_n}) \rangle - 2 \int_{\Gamma_n}^{\Gamma_{n+1}} \langle \Theta_n, \{\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\} \rangle ds. \end{aligned} \quad (34)$$

Young's inequality and Jensen's inequality imply

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

Using (3), $\gamma_1 \leq 1/(m+L)$ and $(\gamma_k)_{k \geq 1}$ is nonincreasing, (34) becomes

$$\begin{aligned} \|\Theta_{n+1}\|^2 &\leq \{1 - \gamma_{n+1}\kappa\} \|\Theta_n\|^2 + 2\gamma_{n+1} \int_{\Gamma_n}^{\Gamma_{n+1}} \|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2 ds \\ &\quad - 2 \int_{\Gamma_n}^{\Gamma_{n+1}} \langle \Theta_n, \{\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\} \rangle ds. \end{aligned} \quad (35)$$

Using the inequality $|\langle a, b \rangle| \leq \epsilon \|a\|^2 + (4\epsilon)^{-1} \|b\|^2$ concludes the proof of (32).

We now prove (33). Note that (32) implies that

$$\begin{aligned} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\|\Theta_{n+1}\|^2 \right] &\leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon)\} \|\Theta_n\|^2 \\ &\quad + (2\gamma_{n+1} + (2\epsilon)^{-1}) \int_{\Gamma_n}^{\Gamma_{n+1}} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2 \right] ds. \end{aligned} \quad (36)$$

By **H1**, the Markov property of $(Y_t)_{t \geq 0}$ and Lemma 19, we have

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

The proof is then concluded plugging this bound in (36). \square

Proof of Theorem 5. Let $x \in \mathbb{R}^d$, $n \geq 1$ and $\zeta_0 = \pi \otimes \delta_x$. Let $(Y_t, \bar{Y}_t)_{t \geq 0}$ with (Y_0, \bar{Y}_0) distributed according to ζ_0 and defined by (9). By definition of W_2 and since for all $t \geq 0$, π is invariant for P_t , $W_2^2(\mu_0 Q^n, \pi) \leq \mathbb{E}_{\zeta_0} \left[\|Y_{\Gamma_n} - \bar{Y}_{\Gamma_n}\|^2 \right]$. Lemma 20 with $\epsilon = \kappa/4$, a straightforward induction and Theorem 1-(ii) imply for all $n \geq 0$

$$\mathbb{E}_{\zeta_0} \left[\|Y_{\Gamma_n} - \bar{Y}_{\Gamma_n}\|^2 \right] \leq u_n^{(1)}(\gamma) \int_{\mathbb{R}^d} \|y - x\|^2 \pi(dy) + A_n(\gamma), \quad (37)$$

where $(u_n^{(1)}(\gamma))_{n \geq 1}$ is given by (10), and

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

with

$$\tilde{\delta}_i = e^{-2m\Gamma_{i-1}} \mathbb{E}_{\zeta_0} \left[\|Y_0 - x^*\|^2 \right] + (1 - e^{-2m\Gamma_{i-1}})(d/m). \quad (38)$$

Since Y_0 is distributed according to π , Theorem 1-(ii) shows that for all $i \in \{1, \dots, n\}$, $\delta_i \leq d/m$ and therefore $A_n(\gamma) \leq$. In addition since for all $y \in \mathbb{R}^d$ $\|x - y\|^2 \leq 2(\|x - x^*\|^2 + \|x^* - y\|^2)$, using Theorem 1-(ii), we get $\int_{\mathbb{R}^d} \|y - x\|^2 \pi(dy) \leq \|x - x^*\|^2 + d/m$, which completes the proof. \square

7.6 Proof of Corollary 6

We preface the proof by a technical lemma.

Lemma 21. *Let $(\gamma_k)_{k \geq 1}$ be a sequence of nonincreasing real numbers, $\varpi > 0$ and $\gamma_1 < \varpi^{-1}$. Then for all $n \geq 0$, $j \geq 1$ and $\ell \in \{1, \dots, n+1\}$,*

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

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

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

\square

Proof of Corollary 6. By Theorem 5, it suffices to show that $u_n^{(1)}$ and $u_n^{(2)}$, defined by (10) and (11) respectively, goes to 0 as $n \rightarrow +\infty$. Using the bound $1 + t \leq e^t$ for $t \in \mathbb{R}$, and $\lim_{n \rightarrow +\infty} \Gamma_n = +\infty$, we have $\lim_{n \rightarrow +\infty} u_n^{(1)} = 0$. Since $(\gamma_k)_{k \geq 0}$ is nonincreasing, note that to show that $\lim_{n \rightarrow +\infty} u_n^{(2)} = 0$, it suffices to prove $\lim_{n \rightarrow +\infty} \sum_{i=1}^n \prod_{k=i+1}^n (1 -$

$\kappa\gamma_k/2)\gamma_i^2 = 0$. But since $(\gamma_k)_{k \geq 1}$ is nonincreasing, there exists $c \geq 0$ such that $c\Gamma_n \leq n-1$ and by Lemma 21 applied with $\ell = \lfloor c\Gamma_n \rfloor$ the integer part of $c\Gamma_n$:

$$\sum_{i=1}^n \prod_{k=i+1}^n (1 - \kappa\gamma_k/2) \gamma_i^2 \leq 2\kappa^{-1} \gamma_{\lfloor c\Gamma_n \rfloor} + \exp(-2^{-1}\kappa(\Gamma_n - \Gamma_{\lfloor c\Gamma_n \rfloor})) \sum_{i=1}^{\lfloor c\Gamma_n \rfloor - 1} \gamma_i. \quad (39)$$

Since $\lim_{k \rightarrow +\infty} \gamma_k = 0$, by the Cesàro theorem, $\lim_{n \rightarrow +\infty} n^{-1}\Gamma_n = 0$. Therefore since $\lim_{n \rightarrow +\infty} \Gamma_n = +\infty$, $\lim_{n \rightarrow +\infty} (\Gamma_n)^{-1} \Gamma_{\lfloor c\Gamma_n \rfloor} = 0$, and the conclusion follows from combining in (39), this limit, $\lim_{k \rightarrow +\infty} \gamma_k = 0$, $\lim_{n \rightarrow +\infty} \Gamma_n = +\infty$ and $\sum_{i=1}^{\lfloor c\Gamma_n \rfloor - 1} \gamma_i \leq c\gamma_1 \Gamma_n$. \square

7.7 Proof of Corollary 7

Since by Theorem 4, for all $x \in \mathbb{R}^d$, $(\delta_x R_\gamma^n)_{n \geq 0}$ converges to π_γ as $n \rightarrow \infty$ in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$, the proof then follows from Theorem 5 and Lemma 21 applied with $\ell = 1$.

7.8 Proofs of Theorem 8

Lemma 22. *Assume H1, H2 and H3. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$. and $\zeta_0 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$. Let $(Y_t, \bar{Y}_t)_{t \geq 0}$ be defined by (9) such that (Y_0, \bar{Y}_0) is distributed according to ζ_0 . Then for all $n \geq 0$ and $\epsilon > 0$, almost surely*

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

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

$$\begin{aligned} \nabla U(Y_s) - \nabla U(Y_{\Gamma_n}) &= \int_{\Gamma_n}^s \left\{ \nabla^2 U(Y_u) \nabla U(Y_u) + \vec{\Delta}(\nabla U)(Y_u) \right\} du \\ &\quad + \sqrt{2} \int_{\Gamma_n}^s \nabla^2 U(Y_u) dB_u. \end{aligned} \quad (40)$$

Since Θ_n is \mathcal{F}_{Γ_n} -measurable and $(\int_0^s \nabla^2 U(Y_u) dB_u)_{s \in [0, \Gamma_{n+1}]}$ is a $(\mathcal{F}_s)_{s \in [0, \Gamma_{n+1}]}$ -martingale under H1, by (40) we have:

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

Combining this equality and $|\langle a, b \rangle| \leq \epsilon \|a\|^2 + (4\epsilon)^{-1} \|b\|^2$ in (35) we have

$$\begin{aligned} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\|\Theta_{n+1}\|^2 \right] &\leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon)\} \|\Theta_n\|^2 + (2\epsilon)^{-1} A \\ &\quad 2\gamma_{n+1} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\int_{\Gamma_n}^{\Gamma_{n+1}} \|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2 ds \right], \end{aligned} \quad (41)$$

where

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

We now separately bound the two last terms of the right hand side. By **H1**, the Markov property of $(Y_t)_{t \geq 0}$ and Lemma 19, we have

$$\begin{aligned} & \int_{\Gamma_n}^{\Gamma_{n+1}} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2 \right] ds \\ & \leq L^2 \left(d\gamma_{n+1}^2 + dL^2\gamma_{n+1}^4/12 + (1/2)L^2\gamma_{n+1}^3 \|Y_{\Gamma_n} - x^*\|^2 \right) . \end{aligned} \quad (42)$$

We now bound A . We get using Jensen's inequality, Fubini's theorem, $\nabla U(x^*) = 0$ and (12)

$$\begin{aligned} A & \leq 2 \int_{\Gamma_n}^{\Gamma_{n+1}} (s - \Gamma_n) \int_{\Gamma_n}^s \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\|\nabla^2 U(Y_u) \nabla U(Y_u)\|^2 \right] du ds \\ & \quad + 2^{-1} \int_{\Gamma_n}^{\Gamma_{n+1}} (s - \Gamma_n) \int_{\Gamma_n}^s \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\|\vec{\Delta}(\nabla U)(Y_u)\|^2 \right] du ds \\ & \leq 2 \int_{\Gamma_n}^{\Gamma_{n+1}} (s - \Gamma_n) L^4 \int_{\Gamma_n}^s \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\|Y_u - x^*\|^2 \right] du ds + \gamma_{n+1}^3 d\tilde{L}^2/6 . \end{aligned} \quad (43)$$

By Lemma 19-(i), the Markov property and for all $t \geq 0$, $1 - e^{-t} \leq t$, we have for all $s \in [\Gamma_n, \Gamma_{n+1}]$,

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

Using this inequality in (43) and for all $t \geq 0$, $1 - e^{-t} \leq t$, we get

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

Combining this bound and (42) in (41) concludes the proof. \square

Proof of Theorem 8. The proof of the Theorem is the same as the one of Theorem 5, using Lemma 22 in place of Lemma 20, and is omitted. \square

7.9 Proof of Theorem 10

Our main tool is the Gaussian Poincaré inequality [4, Theorem 3.20] which can be applied to $R_\gamma(y, \cdot)$ defined by (6), noticing that $R_\gamma(y, \cdot)$ is a Gaussian distribution with mean $y - \gamma \nabla U(y)$ and covariance matrix $2\gamma I_d$: for all Lipschitz function $g : \mathbb{R}^d \rightarrow \mathbb{R}$

$$R_\gamma \{g(\cdot) - R_\gamma g(y)\}^2(y) \leq 2\gamma \|g\|_{\text{Lip}}^2 . \quad (44)$$

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

$$\begin{aligned} \text{Var}_x \{ \hat{\pi}_n^N(f) \} &= \sum_{k=N}^{N+n-1} \mathbb{E}_x \left[\left(\mathbb{E}_x^{\mathcal{G}_{k+1}} [\hat{\pi}_n^N(f)] - \mathbb{E}_x^{\mathcal{G}_k} [\hat{\pi}_n^N(f)] \right)^2 \right] \\ &\quad + \mathbb{E}_x \left[\left(\mathbb{E}_x^{\mathcal{G}_N} [\hat{\pi}_n^N(f)] - \mathbb{E}_x [\hat{\pi}_n^N(f)] \right)^2 \right]. \end{aligned} \quad (45)$$

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

$$\Phi_{n,k}^N : x_k \mapsto \omega_{k,n}^N f(x_k) + R_{\gamma_{k+1}} \Phi_{n,k+1}^N(x_k), \quad (46)$$

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

$$\Psi_n^N : x_N \mapsto R_{\gamma_{N+1}} \Phi_{n,N+1}^N(x_N). \quad (47)$$

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

$$\Phi_{n,k+1}^N(X_{k+1}) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(X_k) = \mathbb{E}_x^{\mathcal{G}_{k+1}} [\hat{\pi}_n^N(f)] - \mathbb{E}_x^{\mathcal{G}_k} [\hat{\pi}_n^N(f)], \quad (48)$$

and $\Psi_n^N(X_N) = \mathbb{E}_x^{\mathcal{G}_N} [\hat{\pi}_n^N(f)]$. With these notations, (45) may be equivalently expressed as

$$\begin{aligned} \text{Var}_x \{ \hat{\pi}_n^N(f) \} &= \sum_{k=N}^{N+n-1} \mathbb{E}_x \left[R_{\gamma_{k+1}} \{ \Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(X_k) \}^2 (X_k) \right] \\ &\quad + \text{Var}_x \{ \Psi_n^N(X_N) \}. \end{aligned} \quad (49)$$

Now for $k = N + n - 1, \dots, N$, we will use the Gaussian Poincaré inequality (44) to the sequence of function $\Phi_{n,k+1}^N$ to prove that $x \mapsto R_{\gamma_{k+1}} \{ \Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(x) \}^2(x)$ is uniformly bounded. It is required to bound the Lipschitz constant of $\Phi_{n,k}^N$.

Lemma 23. *Assume **H1** and **H2**. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m + L)$. Then for all Lipschitz functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$ and $\ell \geq n \geq 1$, $Q_\gamma^{n,\ell} f$ is a Lipschitz function with $\|Q_\gamma^{n,\ell} f\|_{\text{Lip}} \leq \prod_{k=n}^{\ell} (1 - \kappa \gamma_k)^{1/2} \|f\|_{\text{Lip}}$.*

Proof. Recall that for all μ, ν probability measures on \mathbb{R}^d and $p \leq q$, $W_p(\mu, \nu) \leq W_q(\mu, \nu)$. Hence the Kantorovich–Rubinstein theorem implies, for all $y, z \in \mathbb{R}^d$,

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

The proof then follows from Theorem 4 with $p = 1$. \square

Lemma 24. Assume **H1** and **H2**. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Let $N \geq 0$ and $n \geq 1$. Then for all $y \in \mathbb{R}^d$, Lipschitz function f and $k \in \{N, \dots, N+n-1\}$,

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

where $\Phi_{n,k+1}^N$ is given by (46).

Proof. By (46), $\|\Phi_{n,k}^N\|_{\text{Lip}} \leq \sum_{i=k+1}^{N+n} \omega_{i,n}^N \|Q_\gamma^{k+2,i} f\|_{\text{Lip}}$. Using Lemma 23, 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 (14), we have

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

Finally, the proof follows from (44). \square

Also to control the last term in right hand side of (49), we need to control the variance of $\Psi_n^N(X_N)$ under $\delta_x Q_\gamma^N$. But similarly to the sequence of functions $\Phi_{n,k}^N$, Ψ_n^N is Lipschitz by Lemma 23 by definition, see (47). Therefore it suffices to find some bound for the variance of g under $\delta_y Q_\gamma^{n,p}$, for $g : \mathbb{R}^d \rightarrow \mathbb{R}$ a Lipschitz function, $y \in \mathbb{R}^d$ and $\gamma > 0$, which is done in the following Lemma.

Lemma 25. Assume **H1** and **H2**. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Let $g : \mathbb{R}^d \rightarrow \mathbb{R}$ be a Lipschitz function. Then for all $n, p \geq 1$, $n \leq p$ and $y \in \mathbb{R}^d$

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

where $Q_\gamma^{n,p}$ is given by (7).

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

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

(44) implies $\text{Var}_y^{\mathcal{G}_n} \{g(X_p)\} \leq 2 \sum_{k=n+1}^p \gamma_k \|Q_\gamma^{k+1,p}g\|_{\text{Lip}}^2$. The proof follows from Lemma 23 and Lemma 21, using the bound $(1-t)^{1/2} \leq 1-t/2$ for $t \in [0, 1]$. \square

Corollary 26. Assume **H1** and **H2**. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Then for all Lipschitz function f and $x \in \mathbb{R}^d$, $\text{Var}_x \{\Psi_n^N(X_N)\} \leq 8\kappa^{-3} \|f\|_{\text{Lip}}^2 \Gamma_{N+2, N+n+1}^{-2}$, where Ψ_n^N is given by (47).

Proof. By (47) and Lemma 23, Ψ_n^N is Lipschitz function with

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

Using Lemma 23, 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 (14), we have

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

The proof follows from Lemma 25. \square

Proof of Theorem 10. Plugging the bounds given by Lemma 24 and Corollary 26 in (49), we have

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

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

7.10 Proof of Theorem 11

Let $N \geq 0$, $n \geq 1$, $x \in \mathbb{R}^d$ and f be a Lipschitz function. To prove Theorem 11, 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

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

Now using (48) with the sequence of functions $(\Phi_{n,k}^N)$ and Ψ_n^N given by (46) and (47), respectively, we have by the Markov property

$$\begin{aligned} &\mathbb{E}_x \left[e^{\lambda \{ \hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)] \}} \right] \\ &= \mathbb{E}_x \left[e^{\lambda \{ \Psi_n^N(X_n) - \mathbb{E}_x[\Psi_n^N(X_n)] \}} \prod_{k=N}^{N+n-1} R_{\gamma_{k+1}} \left[e^{\lambda \{ \Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(X_k) \}} \right] (X_k) \right], \quad (50) \end{aligned}$$

where R_γ is given by (6) 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

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

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 $2\gamma \text{Id}$. The log-Sobolev inequality [4, Theorem 5.5] shows that for all Lipschitz function $g : \mathbb{R}^d \rightarrow \mathbb{R}$, $x \in \mathbb{R}^d$, $\gamma > 0$ and $\lambda > 0$,

$$\int R_\gamma(x, dy) \{ \exp(\lambda \{g(y) - R_\gamma g(x)\}) \} \leq \exp\left(\gamma \lambda^2 \|g\|_{\text{Lip}}^2\right). \quad (51)$$

We deduced from this result, (48) and Lemma 23, an equivalent of Lemma 24 for the Laplace transform of $\Phi_{n,k+1}^N$ under $\delta_y R_{\gamma_{k+1}}$ for $k \in \{N+1, \dots, N+n\}$ and all $y \in \mathbb{R}^d$.

Corollary 27. *Assume H1 and H2. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Let $N \geq 0$ and $n \geq 1$. Then for all $k \in \{N, \dots, N+n-1\}$, $y \in \mathbb{R}^d$ and $\lambda > 0$,*

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

where $\Phi_{n,k}^N$ is given by (46).

It remains to control the Laplace transform of Ψ_n^N under $\delta_x Q_\gamma^N$, where $\delta_x Q_\gamma^N$ is defined by (7). For this, using again that by (47) and Lemma 23, Ψ_n^N is a Lipschitz function, we iterate (51) to get bounds on the Laplace transform of Lipschitz function g under $Q_\gamma^{n,\ell}(y, \cdot)$ for all $y \in \mathbb{R}^d$ and $n, \ell \geq 1$, since for all $n, \ell \geq 1$, $Q_\gamma^{n,\ell} g$ is a Lipschitz function by Lemma 23.

Lemma 28. *Assume H1 and H2. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Let $g : \mathbb{R}^d \rightarrow \mathbb{R}$ be a Lipschitz function, then for all $n, p \geq 1$, $n \leq p$, $y \in \mathbb{R}^d$ and $\lambda > 0$:*

$$Q_\gamma^{n,p} \{ \exp(\lambda \{g(\cdot) - Q_\gamma^{n,p} g(y)\}) \} (y) \leq \exp\left(\kappa^{-1} \lambda^2 \|g\|_{\text{Lip}}^2\right), \quad (52)$$

where $Q_{n,p}^\gamma$ is given by (7).

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

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

By the Gaussian log-Sobolev inequality (51), we get

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

The proof follows from Lemma 23 and Lemma 21, using the bound $(1-t)^{1/2} \leq 1-t/2$ for $t \in [0, 1]$. \square

Combining this result and $\|\Psi_n^N\|_{\text{Lip}} \leq 2\kappa^{-1} \|f\|_{\text{Lip}} \Gamma_{N+2, N+n+1}^{-1}$ by Lemma 23, we get an analogue of Corollary 26 for the Laplace transform of Ψ_n^N :

Corollary 29. *Assume H1 and H2. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Let $N \geq 0$ and $n \geq 1$. Then for all $\lambda > 0$ and $x \in \mathbb{R}^d$,*

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

where Ψ_n^N is given by (47).

The Laplace transform of $\hat{\pi}_n^N(f)$ can be explicitly bounded using Corollary 27 and Corollary 29 in (50).

Proposition 30. *Assume H1 and H2. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Then for all $N \geq 0$, $n \geq 1$, Lipschitz functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$, $\lambda > 0$ and $x \in \mathbb{R}^d$:*

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

where $u_{N,n}^{(3)}(\gamma)$ is given by (16).

Proof of Theorem 11. Using the Markov inequality and Proposition 30, for all $\lambda > 0$, we have:

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

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

7.11 Proof of Theorem 12

(i) To derive explicit bound for $\|P_t(x, \cdot) - P_t(y, \cdot)\|_{\text{TV}}$, we use the coupling by reflection; see [26] and [7, Example 3.7]. This coupling is defined as the unique solution $(X_t, Y_t)_{t \geq 0}$ of the SDE:

$$\begin{cases} dX_t &= -\nabla U(X_t)dt + \sqrt{2}dB_t^d \\ dY_t &= -\nabla U(Y_t)dt + \sqrt{2}(\text{Id} - 2e_t e_t^T)dB_t^d, \end{cases} \quad \text{where } e_t = \mathbf{e}(X_t - Y_t)$$

with $X_0 = x$, $Y_0 = y$, $\mathbf{e}(z) = z/\|z\|$ for $z \neq 0$ and $\mathbf{e}(0) = 0$ otherwise. Define the coupling time $T_c = \inf\{s \geq 0 \mid X_s \neq Y_s\}$. By construction $X_t = Y_t$ for $t \geq T_c$. By Levy's characterization, $\tilde{B}_t^d = \int_0^t (\text{Id} - 2e_s e_s^T) dB_s^d$ is a d -dimensional Brownian motion, therefore $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$ are weak solutions to (1) started at x and y respectively. Then by Lindvall's inequality, for all $t > 0$ we have $\|P_t(x, \cdot) - P_t(y, \cdot)\|_{\text{TV}} \leq \mathbb{P}(X_t \neq Y_t)$. We now bound for $t > 0$, $\mathbb{P}(X_t \neq Y_t)$. For $t < T_c$, denoting by $B_t^1 = \int_0^t \mathbb{1}_{\{s < T_c\}} e_s^T dB_s^d$, we get

$$d\{X_t - Y_t\} = -\{\nabla U(X_t) - \nabla U(Y_t)\} dt + 2\sqrt{2}e_t dB_t^1.$$

By Itô's formula and **H2**, we have for $t < T_c$,

$$\begin{aligned} \|X_t - Y_t\| &= \|x - y\| - \int_0^t \langle \nabla U(X_s) - \nabla U(Y_s), e_s \rangle ds + 2\sqrt{2}B_t^1 \\ &\leq \|x - y\| - m \int_0^t \|X_s - Y_s\| ds + 2\sqrt{2}B_t^1. \end{aligned}$$

By Grönwall's inequality, we have

$$\|X_t - Y_t\| \leq e^{-mt} \|x - y\| + 2\sqrt{2}B_t^1 - m2\sqrt{2} \int_0^t B_s^1 e^{-m(t-s)} ds.$$

Therefore by integration by part, $\|X_t - Y_t\| \leq U_t$ where $(U_t)_{t \in (0, T_c)}$ is the one-dimensional Ornstein-Uhlenbeck process defined by

$$U_t = e^{-mt} \|x - y\| + 2\sqrt{2} \int_0^t e^{m(s-t)} dB_s^1. = e^{-mt} \|x - y\| + \int_0^{8t} e^{m(s-t)} d\tilde{B}_s^1$$

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

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

Finally the proof follows from [3, Formula 2.0.2, page 542].

(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 $s > 0$, $1/2 - \Phi(-s) \leq (2\pi)^{-1/2} s$, (i) implies that for all $x, y \in \mathbb{R}^d$ and $t > 0$,

$$\|\mu P_t - \nu P_t\|_{\text{TV}} \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{\|x - y\|}{\sqrt{(2\pi/m)(e^{2mt} - 1)}} d\xi(x, y),$$

which is the desired result.

(iii) The proof is a straightforward consequence of (ii) and Theorem 2-(ii).

7.12 Proof of Theorem 13

(i) By (3) for all x, y and for all $k \geq 1$, 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 32 in Section 8, we get

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

where $\tilde{\Lambda}_{n,\ell} = \sum_{i=n}^{\ell} \gamma_i \{\prod_{j=n}^i (1 - \kappa \gamma_j)^{-1}\}$. The proof then follows since by a straightforward calculation $\tilde{\Lambda}_{n,\ell} = \Lambda_{n,\ell}$

(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

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

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

(iii) The proof follows from (iii), the bound for all $s > 0$, $1/2 - \Phi(-s) \leq (2\pi)^{-1/2} s$ and Theorem 3-(ii).

7.13 Proof of Theorem 14

Let $n, \ell \geq 1$, $\ell > n$. Applying Lemma 20 or Lemma 22, we get that for all $x \in \mathbb{R}^d$

$$W_1(\delta_x Q_\gamma^n, \delta_x P_{\Gamma_n}) \leq \{\vartheta_n(x)\}^{1/2} , \vartheta_n(x) = \begin{cases} \vartheta_{n,0}^{(1)}(x) & \text{(H1, H2)} , \\ \vartheta_{n,0}^{(2)}(x) & \text{(H1, H2, H3)} , \end{cases} \quad (53)$$

where $\vartheta_{n,0}^{(1)}(y)$ is given by (20) and $\vartheta_{n,0}^{(2)}(y)$ is given by (21). By the triangle inequality, we get

$$\begin{aligned} \left\| \delta_x P_{\Gamma_\ell} - \delta_x Q_\gamma^\ell \right\|_{\text{TV}} &\leq \left\| \{\delta_x P_{\Gamma_n} - \delta_x Q_\gamma^{1,n}\} P_{\Gamma_{n+1,\ell}} \right\|_{\text{TV}} \\ &\quad + \left\| \delta_x Q_\gamma^{1,n} \left\{ Q_\gamma^{n+1,\ell} - P_{\Gamma_{n+1,\ell}} \right\} \right\|_{\text{TV}} . \end{aligned} \quad (54)$$

Using (17) and (53), we have

$$\left\| \{\delta_x P_{\Gamma_n} - \delta_x Q_\gamma^{1,n}\} P_{\Gamma_{n+1,\ell}} \right\|_{\text{TV}} \leq (\vartheta_n(x) / (4\pi \Gamma_{n+1,\ell}))^{1/2} . \quad (55)$$

For the second term, by [10, Equation 11] (note that we have a different convention for the total variation distance) and the Pinsker inequality, we have

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

By **H1** and Theorem 3, we get

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

Combining the last inequality and (55) in (54) concludes the proof.

7.14 Proof of Theorem 16

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

$$n(\gamma) = \lceil \log(\gamma^{-1}) / \log(2) \rceil. \quad (56)$$

Lemma 31. *Assume **H1**, **H2** and **H3**. Let $\gamma \in (0, 1/(m+L))$. Then for all $x \in \mathbb{R}^d$ and $\ell \in \mathbb{N}^*$, $\ell > 2^{n(\gamma)}$,*

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

where $\varrho_{1,\ell-1}(x)$ is defined by (8) and for all $n_1, n_2 \in \mathbb{N}$, $\vartheta_{n_1,n_2}^{(2)}$ is given by (21).

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

$$\begin{aligned} \|\delta_x P_{\ell\gamma} - \delta_x R_\gamma^\ell\|_{\text{TV}} & \leq \left\| \left\{ \delta_x P_{(\ell-2^n)\gamma} - \delta_x R_\gamma^{\ell-2^n} \right\} P_{2^n\gamma} \right\|_{\text{TV}} \\ & + \left\| \delta_x R_\gamma^{\ell-1} \{P_\gamma - R_\gamma\} \right\|_{\text{TV}} + \sum_{k=1}^n \left\| \delta_x R_\gamma^{\ell-2^k} \left\{ P_{2^{k-1}\gamma} - R_\gamma^{2^{k-1}} \right\} P_{2^{k-1}\gamma} \right\|_{\text{TV}}. \quad (57) \end{aligned}$$

We bound each term in the right hand side. First by (17) and Equation (53), we have

$$\left\| \left\{ \delta_x P_{(\ell-2^n)\gamma} - \delta_x R_\gamma^{\ell-2^n} \right\} P_{2^n\gamma} \right\|_{\text{TV}} \leq (\vartheta_{\ell-2^n,0}^{(2)}(x) / (\pi 2^{n+1}\gamma))^{1/2}, \quad (58)$$

where $\vartheta_{n,0}^{(2)}(x)$ is given by (21). Similarly but using in addition Theorem 3, we have for all $k \in \{1, \dots, n\}$,

$$\left\| \delta_x R_\gamma^{\ell-2^k} \left\{ P_{2^{k-1}\gamma} - R_\gamma^{2^{k-1}} \right\} P_{2^{k-1}\gamma} \right\|_{\text{TV}} \leq (\vartheta_{2^{k-1}, \ell-2^k}^{(2)}(x) / (\pi 2^{k+1} \gamma))^{1/2}, \quad (59)$$

where $\vartheta_{2^{k-1}, \ell-2^k}^{(2)}(x)$ is given by (21). For the last term, by [10, Equation 11] and the Pinsker inequality, we have

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

By H1 and Theorem 3, we get

$$\left\| \delta_x R_\gamma^{\ell-1} \{R_\gamma - P_\gamma\} \right\|_{\text{TV}}^2 \leq 2^{-3} L^2 \{(\gamma^3 L^2/3) \varrho_{1, \ell-1}(x) + d\gamma^2\}. \quad (60)$$

Combining (58), (59) and (60) in (57) concludes the proof. \square

Proof of Theorem 16. By Lemma 31, we get (see the supplementary document [12, Section 1.2] for the detailed calculation):

$$\begin{aligned} & \|\delta_x P_{\ell\gamma} - \delta_x R_\gamma^\ell\|_{\text{TV}} \\ & \leq 2^{-3/2} L \left\{ (\gamma^3 L^2/3) \left\{ (1 - \kappa\gamma)^{\ell-1} \|x - x^*\|^2 + 2\kappa^{-1}d \right\} + d\gamma^2 \right\}^{1/2} \\ & \quad + (4\pi)^{-1/2} [\gamma^2 \mathbf{E}_1(\gamma, d) + \gamma^3 \mathbf{E}_2(\gamma) \mathbf{E}_3(\gamma, d, x)]^{1/2} \\ & \quad + \mathfrak{n}(\gamma) (4\pi)^{-1/2} [\gamma^2 \mathbf{E}_1(\gamma, d) + \gamma^2 \mathbf{E}_2(\gamma) \mathbf{E}_4(\gamma, d, x)]^{1/2}, \end{aligned} \quad (61)$$

where

$$\begin{aligned} \mathbf{E}_3(\gamma, d, x) &= (\ell - \gamma^{-1}) e^{-m\gamma(\ell-2\gamma^{-1}-1)} \|x - x^*\|^2 + 2d/(\kappa\gamma m) \\ \mathbf{E}_4(\gamma, d, x) &= e^{-m\gamma(\ell-2\gamma^{-1}-1)} \|x - x^*\|^2 + 2\kappa^{-1}d + d/m. \end{aligned}$$

Letting ℓ go to infinity, using Theorem 12-(iii) and Theorem 13-(iii), we get the desired conclusion. \square

7.15 Proof of Theorem 17

Let $N \geq 0$, $n \geq 1$, $x \in \mathbb{R}^d$ and $f \in \mathbb{F}_b(\mathbb{R}^d)$. We will use again the decomposition of $\text{Var}_x \{\hat{\pi}_n^N(f)\}$ given in (49).

Let $k \in \{N, \dots, N+n-1\}$. We can not directly apply the Poincaré inequality since the function $\Phi_{n,k}^N$, defined in (46), is not Lipschitz anymore. However, Theorem 13-(ii) shows that for all $\ell, n \in \mathbb{N}^*$, $n < \ell$, $Q_\gamma^{n,\ell} f$ is a Lipschitz function with

$$\left\| Q_\gamma^{n,\ell} f \right\|_{\text{Lip}} \leq \text{osc}(f) / (4\pi \Lambda_{n,\ell})^{1/2}. \quad (62)$$

On the other hand, by definition (46), $\Phi_{n,k}^N = \omega_{k+1,n}^N f + \tilde{\Phi}_{n,k}^N$, where $\tilde{\Phi}_{n,k}^N = \sum_{i=k+2}^{N+n} \omega_{i,n}^N Q_\gamma^{k+2,i} f$. Therefore by (62) $\Phi_{n,k}^N$ has only one component which is not Lipschitz continuous and can be isolated. Using that f is bounded and the Young inequality, we get for any $y \in \mathbb{R}^d$

$$R_{\gamma_{k+1}} \left\{ \Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(y) \right\}^2(y) \leq 2(\omega_{k+1,n}^N)^2 \text{osc}(f)^2 + 2R_{\gamma_{k+1}} \left\{ \tilde{\Phi}_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \tilde{\Phi}_{n,k+1}^N(y) \right\}^2(y).$$

In addition by (62), we get

$$\left\| \tilde{\Phi}_{n,k}^N \right\|_{\text{Lip}} \leq \sum_{i=k+2}^{N+n} \omega_{i,n}^N \left\| Q_\gamma^{k+2,i} f \right\|_{\text{Lip}} \leq \text{osc}(f) \sum_{i=k+2}^{N+n} \omega_{i,n}^N / (\pi \Lambda_{k+2,i})^{1/2}. \quad (63)$$

Finally using (44), we end up with a counterpart Lemma 24: for all $y \in \mathbb{R}^d$,

$$R_{\gamma_{k+1}} \left\{ \Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(y) \right\}^2(y) \leq 2(\omega_{k+1,n}^N)^2 \text{osc}(f)^2 + 4\gamma_{k+1} \text{osc}(f)^2 \left\{ \sum_{i=k+2}^{N+n} \omega_{i,n}^N / (\pi \Lambda_{k+2,i})^{1/2} \right\}^2. \quad (64)$$

It remains to control the variance of Ψ_n^N under $\delta_x Q_\gamma^N$, where $\delta_x Q_\gamma^N$ is defined by (7) and Φ_n^N is given by (47). But using (62), Ψ_n^N is a Lipschitz function with Lipschitz constant bounded by:

$$\left\| \Psi_n^N \right\|_{\text{Lip}} \leq \sum_{i=N+1}^{N+n} \omega_{i,n}^N \left\| Q_\gamma^{N+1,i} f \right\|_{\text{Lip}} \leq 2 \|f\|_\infty \sum_{i=N+1}^{N+n} \omega_{i,n}^N / \Lambda_{N+1,i}^{1/2}. \quad (65)$$

Thus by Lemma 25, we have the following result which is the counterpart of Corollary 26: for all $y \in \mathbb{R}^d$,

$$\text{Var}_y \left\{ \Psi_n^N(X_N) \right\} \leq 2\kappa^{-1} \|f\|_\infty^2 \left\{ \sum_{i=N+1}^{N+n} \omega_{i,n}^N / (\pi \Lambda_{N+1,i})^{1/2} \right\}^2. \quad (66)$$

Finally, the proof follows from combining (64) and (66) in (49).

7.16 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 (50) again but combined with the decomposition of $\Phi_{n,k+1}^N$, for $k \in \{N, \dots, N+n-1\}$, into a Lipschitz component and a bounded measurable component as it is done in the proof of (64). Let $k \in \{N, \dots, N+n-1\}$. By definition (46),

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

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

By (63) and (51), we obtain for all $y \in \mathbb{R}^d$ and $\lambda > 0$,

$$\begin{aligned} R_{\gamma_{k+1}} \left\{ e^{\lambda \{\Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(y)\}} \right\} (y) \\ \leq \exp \left(\lambda \operatorname{osc}(f) \gamma_{k+2} (\Gamma_{N+2, N+n+1})^{-2} + (\lambda \operatorname{osc}(f))^2 \gamma_{k+1} \left(\sum_{i=k+2}^{N+n} \omega_{i,n}^N / (\pi \Lambda_{k+2,i})^{1/2} \right)^2 \right). \end{aligned} \quad (67)$$

It remains to control the Laplace transform of Ψ_n^N under $\delta_x Q_\gamma^N$. For this, note that by (65) Ψ_n^N is a Lipschitz function. Therefore using Lemma 28, we get an analogue of Corollary 29: for all $y \in \mathbb{R}^d$ and $\lambda > 0$,

$$\mathbb{E}_y \left[e^{\lambda \{\Psi_n^N(X_n) - \mathbb{E}_x[\Psi_n^N(X_n)]\}} \right] \leq \exp \left(\kappa^{-1} \lambda^2 \operatorname{osc}(f)^2 \left(\sum_{i=N+1}^{N+n} \omega_{i,n}^N / (\pi \Lambda_{N+1,i})^{1/2} \right)^2 \right), \quad (68)$$

Combining (67) and (68) in (50), the Laplace transform of $\hat{\pi}_n^N(f)$ can be explicitly bounded: for all $\lambda > 0$,

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

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

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

Then the proof follows from taking

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

8 Contraction results in total variation for some functional autoregressive models

In this section, we consider functional autoregressive models of the form: for all $k \geq 0$

$$X_{k+1} = h_{k+1}(X_k) + \sigma_{k+1} Z_{k+1}, \quad (69)$$

where $(Z_k)_{k \geq 1}$ 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 $(h_k)_{k \geq 1}$ are a sequence of measurable map from \mathbb{R}^d to \mathbb{R}^d which satisfies the following assumption:

AR1. For all $k \geq 1$, there exists $\varpi_k \in [0, 1]$ such that h_k is $1 - \varpi_k$ -Lipschitz, i.e. for all $x, y \in \mathbb{R}^d$, $\|h_k(x) - h_k(y)\| \leq (1 - \varpi_k) \|x - y\|$.

The sequence $\{X_k, k \in \mathbb{N}\}$ defines an inhomogeneous Markov chain associated with the sequence of Markov kernel $(P_k)_{k \geq 1}$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ given for all $x \in \mathbb{R}^d$ and $A \in \mathbb{R}^d$ by

$$P_k(x, A) = \frac{1}{(2\pi\sigma_k^2)^{d/2}} \int_A \exp\left(-\|y - h_k(x)\|^2 / (2\sigma_k^2)\right) dy. \quad (70)$$

We denote for all $n \geq 1$ by Q^n the marginal laws of the sequence $(X_k)_{k \geq 1}$ and given by

$$Q^n = P_1 \cdots P_n. \quad (71)$$

In this section we are interested in showing that for all $x, y \in \mathbb{R}^d$, the sequence $\{\|\delta_x Q^n - \delta_y Q^n\|_{TV}, n \in \mathbb{N}\}$ goes to 0 with an explicit rate depending on the assumption on the sequence $(h_k)_{k \geq 1}$, which does not depend on the dimension d . We also mention [13] which establishes convergence of homogeneous Markov chain in some Wasserstein distances using different method of proof but still relying on the same coupling. The proof is based on a coupling for Gaussian random walks proposed in [5, Section 3.3]. Let $x, y \in \mathbb{R}^d$. We consider for all $k \geq 1$ the following coupling (X_1, Y_1) between $P_k(x, \cdot)$ and $P_k(y, \cdot)$. Define the function E and e from $\mathbb{R}^d \times \mathbb{R}^d$ to \mathbb{R}^d by

$$E_k(x, y) = h_k(y) - h_k(x), e_k(x, y) = \begin{cases} E_k(x, y) / \|E_k(x, y)\| & \text{if } E_k(x, y) \neq 0 \\ 0 & \text{otherwise,} \end{cases} \quad (72)$$

Consider the following coupling. For all $\tilde{x}, \tilde{y}, \tilde{z} \in \mathbb{R}^d$, $\tilde{x} \neq \tilde{y}$, define

$$F_k(\tilde{x}, \tilde{y}, \tilde{z}) = h_k(\tilde{y}) + (\text{Id} - 2e_k(\tilde{x}, \tilde{y})e_k(\tilde{x}, \tilde{y})^T) \tilde{z} \quad (73)$$

$$\alpha_k(\tilde{x}, \tilde{y}, \tilde{z}) = \frac{\varphi_{\sigma_k^2}(\|E_k(\tilde{x}, \tilde{y})\| - \langle e_k(\tilde{x}, \tilde{y}), \tilde{z} \rangle)}{\varphi_{\sigma_k^2}(\langle e_k(\tilde{x}, \tilde{y}), \tilde{z} \rangle)}, \quad (74)$$

where $\varphi_{\sigma_k^2}$ is the probability density of a zero-mean gaussian variable with variance σ_k^2 . Let Z be a standard d -dimensional Gaussian random variable. Set $X_1 = h_k(x) + \sigma_k Z$ and

$$Y_1 = \begin{cases} h_k(y) + \sigma_k Z & \text{if } E_k(x, y) = 0 \\ B_1 X_1 + (1 - B_1) F_k(x, y, Z) & \text{if } E_k(x, y) \neq 0, \end{cases}$$

where B_1 is a Bernoulli random variable independent of Z 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 \mathbf{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 $\mathbf{A} \in \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d)$ by

$$\begin{aligned} \mathbf{K}_k((x, y), \mathbf{A}) &= \frac{\mathbb{1}_D(h_k(x), h_k(y))}{(2\pi\sigma_k^2)^{d/2}} \int_{\mathbb{R}^d} \mathbb{1}_A(\tilde{x}, \tilde{x}) e^{-\|\tau_k(\tilde{x}, x)\|^2/(2\sigma_k^2)} d\tilde{x} \\ &+ \frac{\mathbb{1}_{D^c}(h_k(x), h_k(y))}{(2\pi\sigma_k^2)^{d/2}} \left[\int_{\mathbb{R}^d} \mathbb{1}_A(\tilde{x}, \tilde{x}) p_k(x, y, \tau_k(\tilde{x}, x)) e^{-\|\tau_k(\tilde{x}, x)\|^2/(2\sigma_k^2)} d\tilde{x} \right. \\ &\left. + \int_{\mathbb{R}^d} \mathbb{1}_A(\tilde{x}, F_k(x, y, \tau_k(\tilde{x}, x))) \{1 - p_k(x, y, \tau_k(\tilde{x}, x))\} e^{-\|\tau_k(\tilde{x}, x)\|^2/(2\sigma_k^2)} d\tilde{x} \right], \end{aligned} \quad (75)$$

where for all $\tilde{x} \in \mathbb{R}^d$, $\tau_k(\tilde{x}, x) = \tilde{x} - h_k(x)$ and $D = \{(\tilde{x}, \tilde{y}) \in \mathbb{R}^d \times \mathbb{R}^d \mid \tilde{x} = \tilde{y}\}$. It is well known (see e.g. [5, Section 3.3]), that for all $x, y \in \mathbb{R}^d$ and $k \geq 1$, $\mathbf{K}_k((x, y), \cdot)$ is a transference plan of $\mathbf{P}_k(x, \cdot)$ and $\mathbf{P}_k(y, \cdot)$. Furthermore, we have for all $x, y \in \mathbb{R}^d$ and $k \geq 1$

$$\mathbf{K}_k((x, y), D) = 2\Phi\left(-\frac{\|\mathbf{E}_k(x, y)\|}{2\sigma_k}\right). \quad (76)$$

For all initial distribution μ_0 on $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d))$, $\tilde{\mathbb{P}}_{\mu_0}$ and $\tilde{\mathbb{E}}_{\mu_0}$ denote the probability and the expectation respectively, associated with the sequence of Markov kernels $(\mathbf{K}_k)_{k \geq 1}$ defined in (75) and μ_0 on the canonical space $((\mathbb{R}^d \times \mathbb{R}^d)^{\mathbb{N}}, (\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d))^{\otimes \mathbb{N}})$, $\{(X_i, Y_i), i \in \mathbb{N}\}$ denotes the canonical process and $\{\tilde{\mathcal{F}}_i, i \in \mathbb{N}\}$ the corresponding filtration. Then if $(X_0, Y_0) = (x, y) \in \mathbb{R}^d \times \mathbb{R}^d$, for all $k \geq 1$ (X_k, Y_k) is coupling of $\delta_x \mathbf{Q}^k$ and $\delta_y \mathbf{Q}^k$. Using Lindvall inequality, bounding $\|\delta_x \mathbf{Q}^n - \delta_y \mathbf{Q}^n\|_{\text{TV}}$ for $n \geq 1$ amounts to evaluate $\tilde{\mathbb{P}}_{(x, y)}(X_n \neq Y_n)$. It is the content of the proof of the main result of this section.

Theorem 32. *Assume **AR1**. Then for all $x, y \in \mathbb{R}^d$ and $n \geq 1$,*

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

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

We preface the proof by a technical Lemma.

Lemma 33. *For all $\varsigma, a > 0$ and $t \in \mathbb{R}_+$, the following identity holds*

$$\begin{aligned} \int_{\mathbb{R}} \varphi_{\varsigma^2}(y) \left\{ 1 - \min\left(1, \frac{\varphi_{\varsigma^2}(t - y)}{\varphi_{\varsigma^2}(y)}\right) \right\} \left\{ 1 - 2\Phi\left(-\frac{|2y - t|}{2a}\right) \right\} dy \\ = 1 - 2\Phi\left(-\frac{t}{2(\varsigma^2 + a^2)^{1/2}}\right). \end{aligned}$$

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

$$\begin{aligned}
I &= \int_{-\infty}^{t/2} \{\varphi_{\varsigma^2}(y) - \varphi_{\varsigma^2}(t-y)\} \left\{1 - 2\Phi\left(\frac{2y-t}{2a}\right)\right\} dy \\
&= \int_{-\infty}^{t/2} \varphi_{\varsigma^2}(y) \left\{1 - 2\Phi\left(\frac{2y-t}{2a}\right)\right\} dy \\
&\quad - \int_{-\infty}^{-t/2} \varphi_{\varsigma^2}(y) \left\{1 - 2\Phi\left(\frac{t+2y}{2a}\right)\right\} dy,
\end{aligned} \tag{77}$$

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 $u \in \mathbb{R}_+$, $1 - 2\Phi(-u/(2a)) = \mathbb{P}[|X| \leq u/2]$, we have by (77)

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

Using that Y and $-Y$ have the same law in the second term, we get

$$\begin{aligned}
I &= \mathbb{P}(Y \leq t/2, X + Y \leq t/2, Y - X \leq t/2) \\
&\quad - \mathbb{P}(Y \leq -t/2, X - Y \geq t/2, Y + X \leq -t/2) \\
&= I_1 + I_2,
\end{aligned} \tag{78}$$

where

$$\begin{aligned}
I_1 &= \mathbb{P}(Y \leq t/2, X + Y \leq t/2, Y - X \leq t/2, X \geq 0) \\
&\quad - \mathbb{P}(Y \leq -t/2, X - Y \geq t/2, Y + X \leq -t/2, X \geq 0) \\
&= \mathbb{P}(|X + Y| \leq t/2, X \geq 0),
\end{aligned} \tag{79}$$

and

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

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

$$\begin{aligned}
I_2 &= \mathbb{P}(Y \geq -t/2, X - Y \leq t/2, Y + X \geq -t/2, X \leq 0) \\
&\quad - \mathbb{P}(Y \geq t/2, X + Y \geq t/2, X - Y \leq -t/2, X \leq 0) \\
&= \mathbb{P}(|X + Y| \leq t/2, X \leq 0).
\end{aligned} \tag{80}$$

Combining (79), (80) in (78), we have $I = \mathbb{P}(|X + Y| \leq t/2)$. The proof follows from the fact that $X + Y$ is a real Gaussian random variable with mean zero and variance $a^2 + \varsigma^2$, since X and Y are independent. \square

Proof of Theorem 32. Since for all $k \geq 1$, (X_k, Y_k) is a coupling of $\delta_x \mathbf{Q}^k$ and $\delta_y \mathbf{Q}^k$, $\|\delta_x \mathbf{Q}^k - \delta_y \mathbf{Q}^k\|_{\text{TV}} \leq \tilde{\mathbb{P}}_{(x,y)}(X_k \neq Y_k)$.

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

$$\begin{aligned} \tilde{\mathbb{P}}_{(x,y)}(X_n \neq Y_n) & \leq \tilde{\mathbb{E}}_{(x,y)} \left[\mathbb{1}_{\mathbb{D}^c}(X_k, Y_k) \left[1 - 2\Phi \left\{ -\frac{\|X_k - Y_k\|}{2(\Xi_{k+1,n})^{1/2}} \right\} \right] \right], \quad (81) \end{aligned}$$

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

Since $X_n \neq Y_n$ implies that $X_{n-1} \neq Y_{n-1}$, the Markov property and (76) imply

$$\begin{aligned} \tilde{\mathbb{P}}_{(x,y)}(X_n \neq Y_n) & = \tilde{\mathbb{E}}_{(x,y)} \left[\mathbb{1}_{\mathbb{D}^c}(X_{n-1}, Y_{n-1}) \tilde{\mathbb{E}}_{(X_{n-1}, Y_{n-1})} [\mathbb{1}_{\mathbb{D}^c}(X_1, Y_1)] \right] \\ & \leq \tilde{\mathbb{E}}_{(x,y)} \left[\mathbb{1}_{\mathbb{D}^c}(X_{n-1}, Y_{n-1}) \left[1 - 2\Phi \left\{ -\frac{\|\mathbb{E}_{n-1}(X_{n-1}, Y_{n-1})\|}{2\sigma_n} \right\} \right] \right] \end{aligned}$$

Using **AR1** and (72), $\|\mathbb{E}_n(X_{n-1}, Y_{n-1})\| \leq (1 - \varpi_n) \|X_{n-1} - Y_{n-1}\|$, showing (81) holds for $k = n-1$.

Assume that (81) holds for $k \in \{1, \dots, n-1\}$. On $\{X_k \neq Y_k\}$, we have

$$\|X_k - Y_k\| = \left| -\|\mathbb{E}_k(X_{k-1}, Y_{k-1})\| + 2\sigma_k e_k(X_{k-1}, Y_{k-1})^\top Z_k \right|,$$

which implies

$$\begin{aligned} & \mathbb{1}_{\mathbb{D}^c}(X_k, Y_k) \left[1 - 2\Phi \left\{ -\frac{\|X_k - Y_k\|}{2\Xi_{k+1,n}^{1/2}} \right\} \right] \\ & = \mathbb{1}_{\mathbb{D}^c}(X_k, Y_k) \left[1 - 2\Phi \left\{ -\frac{|2\sigma_k e_k(X_{k-1}, Y_{k-1})^\top Z_k - \|\mathbb{E}_k(X_{k-1}, Y_{k-1})\||}{2\Xi_{k+1,n}^{1/2}} \right\} \right]. \end{aligned}$$

Since Z_k is independent of $\tilde{\mathcal{F}}_{k-1}$, $\sigma_k e_k(X_{k-1}, Y_{k-1})^\top Z_k$ is a real Gaussian random variable with zero mean and variance σ_k^2 , therefore by Lemma 33, we get

$$\begin{aligned} & \tilde{\mathbb{E}}_{(x,y)}^{\tilde{\mathcal{F}}_{k-1}} \left[\mathbb{1}_{\mathbb{D}^c}(X_k, Y_k) \left[1 - 2\Phi \left\{ -\frac{\|X_k - Y_k\|}{2\Xi_{k+1,n}^{1/2}} \right\} \right] \right] \\ & \leq \mathbb{1}_{\mathbb{D}^c}(X_{k-1}, Y_{k-1}) \left[1 - 2\Phi \left\{ -\frac{\|\mathbb{E}_k(X_{k-1}, Y_{k-1})\|}{2(\sigma_k^2 + \Xi_{k+1,n})^{1/2}} \right\} \right]. \end{aligned}$$

Using by **AR1** that $\|\mathbb{E}_k(X_{k-1}, Y_{k-1})\| \leq (1 - \varpi_k) \|X_{k-1} - Y_{k-1}\|$ concludes the induction. \square

Acknowledgements

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

References

- [1] J. H. Albert and S. Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679, 1993.
- [2] F. Bolley, I. Gentil, and A. Guillin. Convergence to equilibrium in Wasserstein distance for Fokker-Planck equations. *J. Funct. Anal.*, 263(8):2430–2457, 2012.
- [3] A. N. Borodin and P. Salminen. *Handbook of Brownian motion—facts and formulae*. Probability and its Applications. Birkhäuser Verlag, Basel, second edition, 2002.
- [4] 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.
- [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] Faes C., Ormerod J. T., and Wand M. P. Variational bayesian inference for parametric and nonparametric regression with missing data. *Journal of the American Statistical Association*, 106(495):959–971, 2011.
- [7] M. F. Chen and S. F. Li. Coupling methods for multidimensional diffusion processes. *Ann. Probab.*, 17(1):151–177, 1989.
- [8] H. M. Choi and J. P. Hobert. The Polya-Gamma Gibbs sampler for Bayesian logistic regression is uniformly ergodic. *Electron. J. Statist.*, 7:2054–2064, 2013.
- [9] N. Chopin and Ridgway J. Leave pima indians alone: binary regression as a benchmark for bayesian computation. Technical Report 0864, arXiv, une 2015.
- [10] A. S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pages n/a–n/a, 2016.
- [11] A. Durmus and É. Moulines. Non-asymptotic convergence analysis for the unadjusted langevin algorithm. Accepted for publication in *Ann. Appl. Probab.*
- [12] A. Durmus and É. Moulines. Supplement to “high-dimensional bayesian inference via the unadjusted langevin algorithm”, 2015. <https://hal.inria.fr/hal-01176084/>.

- [13] A. Eberle. Quantitative contraction rates for Markov chains on continuous state spaces. In preparation.
- [14] 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.
- [15] 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.
- [16] R. B. Gramacy and N. G. Polson. Simulation-based regularized logistic regression. *Bayesian Anal.*, 7(3):567–590, 09 2012.
- [17] U. Grenander. Tutorial in pattern theory. Division of Applied Mathematics, Brown University, Providence, 1983.
- [18] 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.
- [19] T. E. Hanson, A J. Branscum, and W. O. Johnson. Informative g -priors for logistic regression. *Bayesian Anal.*, 9(3):597–611, 2014.
- [20] C. C. Holmes and L. Held. Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Anal.*, 1(1):145–168, 03 2006.
- [21] A. Joulin and Y. Ollivier. Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.*, 38(6):2418–2442, 2010.
- [22] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer New York, 1991.
- [23] D. Lamberton and G. Pagès. Recursive computation of the invariant distribution of a diffusion. *Bernoulli*, 8(3):367–405, 2002.
- [24] 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.
- [25] V. Lemaire. *Estimation de la mesure invariante d'un processus de diffusion*. PhD thesis, Université Paris-Est, 2005.
- [26] T. Lindvall and L. C. G. Rogers. Coupling of multidimensional diffusions by reflection. *Ann. Probab.*, 14(3):860–872, 1986.

- [27] 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.
- [28] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, New York, NY, USA, 2nd edition, 2009.
- [29] 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.
- [30] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Applied Optimization. Springer, 2004.
- [31] G. Parisi. Correlation functions and computer simulations. *Nuclear Physics B*, 180:378–384, 1981.
- [32] 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.
- [33] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.
- [34] 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.
- [35] D. Sabanés Bové and L. Held. Hyper- g priors for generalized linear models. *Bayesian Anal.*, 6(3):387–410, 2011.
- [36] 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.
- [37] C. Villani. *Optimal transport : old and new*. Grundlehren der mathematischen Wissenschaften. Springer, Berlin, 2009.
- [38] 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.
- [39] 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.