The Tamed Unadjusted Langevin Algorithm

Nicolas Brosse, Alain Durmus, Éric Moulines, Sotirios Sabanis

The Tamed Unadjusted Langevin Algorithm

Nicolas Brosse [1], Alain Durmus [2], Éric Moulines [1] and Sotirios Sabanis [3]

October 23, 2017

Abstract

In this article, we consider the problem of sampling from a probability measure π having a density on $\mathbb{R}^d$ known up to a normalizing constant, $x \mapsto \mathrm{e}^{-U(x)} / \int_{\mathbb{R}^d} \mathrm{e}^{-U(y)} \mathrm{d}y$. The Euler discretization of the Langevin stochastic differential equation (SDE) is known to be unstable in a precise sense, when the potential U is superlinear, *i.e.* $\liminf_{\|x\| \to +\infty} \|\nabla U(x)\| / \|x\| = +\infty$. Based on previous works on the taming of superlinear drift coefficients for SDEs, we introduce the Tamed Unadjusted Langevin Algorithm (TULA) and obtain non-asymptotic bounds in V-total variation norm and Wasserstein distance of order 2 between the iterates of TULA and π, as well as weak error bounds. Numerical experiments are presented which support our findings.

1 Introduction

The Unadjusted Langevin Algorithm (ULA) first introduced in the physics literature by [Par81] and popularised in the computational statistics community by [Gre83] and [GM94] is a technique to sample complex and high-dimensional probability distributions. This issue has far-reaching consequences in Bayesian statistics and machine learning [And+03], [Cot+13], aggregation of estimators [DT12] and molecular dynamics [LS16]. More precisely, let π be a probability distribution on $\mathbb{R}^d$ which has density (also denoted by π) with respect to the Lebesgue measure given for all $x \in \mathbb{R}^d$ by,

$$\pi(x) = \mathrm{e}^{-U(x)} / \int_{\mathbb{R}^d} \mathrm{e}^{-U(y)} \mathrm{d}y \;, \quad \text{with} \quad \int_{\mathbb{R}^d} \mathrm{e}^{-U(y)} \mathrm{d}y < +\infty \;.$$

Assuming that $U : \mathbb{R}^d \to \mathbb{R}$ is continuously differentiable, the overdamped Langevin stochastic differential equation (SDE) associated with π is given by

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

[1] Centre de Mathématiques Appliquées, UMR 7641, Ecole Polytechnique, France.
Emails: nicolas.brosse@polytechnique.edu, eric.moulines@polytechnique.edu
[2] Ecole Normale Supérieure CMLA 61, Av. du Président Wilson 94235 Cachan Cedex, France
Email: alain.durmus@cmla.ens-cachan.fr
[3] University of Edinburgh, Scotland, UK. Email: s.sabanis@ed.ac.uk

where $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion. The discrete time Markov chain associated with the ULA algorithm is obtained by the Euler-Maruyama discretization scheme of the Langevin SDE defined for $k \in \mathbb{N}$ by,

$$X_{k+1} = X_k - \gamma \nabla U(X_k) + \sqrt{2\gamma} Z_{k+1} \ , \ X_0 = x_0 \ , \tag{2}$$

where $x_0 \in \mathbb{R}^d$, $\gamma > 0$ and $(Z_k)_{k\in\mathbb{N}}$ are i.i.d. standard d-dimensional Gaussian variables. Under adequate assumptions on a globally Lipschitz ∇U, non-asymptotic bounds in total variation and Wasserstein distances between the distribution of $(X_k)_{k\in\mathbb{N}}$ and π can be found in [Dal17], [DM17], [DM16]. However, the ULA algorithm is unstable if ∇U is superlinear *i.e.* $\liminf_{\|x\|\to+\infty} \|\nabla U(x)\| / \|x\| = +\infty$, see [RT96, Theorem 3.2], [MSH02] and [HJK11]. This is illustrated with a particular example in [MSH02, Lemma 6.3] where, the SDE (1) is considered in one dimension with $U(x) = x^4/4$ along with the associated Euler discretization (2) and it is shown that for all $\gamma > 0$, if $\mathbb{E}\left[X_0^2\right] \geq 2/\gamma$, one obtains $\lim_{n\to+\infty} \mathbb{E}\left[X_n^2\right] = +\infty$. Moreover, the sample path $(X_n)_{n\in\mathbb{N}}$ diverges to infinity with positive probability.

Until recently, either implicit numerical schemes, e.g. see [MSH02] and [HMS02], or adaptive stepsize schemes, e.g. see [LMS07], were used to address this problem. However, in the last few years, a new generation of explicit numerical schemes, which are more efficient computationally, has been introduced by "taming" appropriately the superlinearly growing drift, see [HJK12] and [Sab13] for more details. This methodology has been extended to Milstein-type schemes, see [KS17] and [WG13], which achieve optimal rate of convergence 1 and include the class of tamed Euler approximations for SDEs with constant diffusion coefficients. A more refined methodology in the latter direction appears in [KS16] which inspires some of the weak assumptions used in this article.

Nonetheless, at the exception of [MSH02], these works focus on the discretization of SDEs with superlinear coefficients in finite time. We aim at extending these techniques in the purpose of sampling from π, the invariant measure of (1). To deal with the superlinear drift ∇U, we introduce a family of drift coefficients $(G_\gamma)_{\gamma>0}$ with $G_\gamma : \mathbb{R}^d \to \mathbb{R}^d$ indexed by the step size γ which are close approximations of ∇U in a sense made precise below. Consider then the following Markov chain $(X_k)_{k\in\mathbb{N}}$ defined for all $k \in \mathbb{N}$ by

$$X_{k+1} = X_k - \gamma G_\gamma(X_k) + \sqrt{2\gamma} Z_{k+1} \ , \ X_0 = x_0 \ . \tag{3}$$

Under appropriate assumptions on G_γ and for a Lyapunov function $V : \mathbb{R}^d \to [1, +\infty)$, we show that $(X_k)_{k\in\mathbb{N}}$ is V-geometrically ergodic of invariant measure π_γ close to π. We suggest two different explicit formulations for the family $(G_\gamma)_{\gamma>0}$ based on previous studies on the tamed Euler scheme [HJK12], [Sab13], [HJ15]. Define for all $\gamma > 0$, $\mathsf{G}_\gamma, \mathsf{G}_{\gamma,c} : \mathbb{R}^d \to \mathbb{R}^d$ for all $x \in \mathbb{R}^d$ by

$$\mathsf{G}_\gamma(x) = \frac{\nabla U(x)}{1 + \gamma \|\nabla U(x)\|} \quad \text{and} \quad \mathsf{G}_{\gamma,c}(x) = \left(\frac{\partial_i U(x)}{1 + \gamma |\partial_i U(x)|}\right)_{i\in\{1,\ldots,d\}} , \tag{4}$$

where $\partial_i U$ is the i^{th}-coordinate of ∇U. The Euler scheme (3) with $G_\gamma = \mathsf{G}_\gamma$, respectively $G_\gamma = \mathsf{G}_{\gamma,c}$, is referred to as the Tamed Unadjusted Langevin Algorithm (TULA), respectively the Tamed Unadjusted Langevin Algorithm coordinate-wise (TULAc).

Another line of works has focused on the Metropolis Adjusted Langevin Algorithm (MALA) that consists in adding a Metropolis Hastings step to the ULA algorithm. [BH13] provides a detailed analysis of MALA in the case where the drift coefficient is superlinear. Note also that a normalization of the gradient was suggested in [RT96, Section 1.4.3] calling it MALTA (Metropolis Adjusted Langevin Truncated Algorithm) and analyzed in [Atc06] and [BV10].

The article is organized as follows. In Section 2, the Markov chain $(X_k)_{k\in\mathbb{N}}$ defined by (3) is shown to be V-geometrically ergodic w.r.t. an invariant measure π_γ. Non-asymptotic bounds between the distribution of $(X_k)_{k\in\mathbb{N}}$ and π in total variation and Wasserstein distances are provided, as well as weak error bounds. In Section 3, the methodology is illustrated through numerical examples. Finally, proofs of the main results appear in Section 4.

Notations

Let $\mathcal{B}(\mathbb{R}^d)$ denote the Borel σ-field of $\mathbb{R}^d$. Moreover, let $\mathrm{L}^1(\mu)$ be the set of μ-integrable functions for μ a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Further, $\mu(f) = \int_{\mathbb{R}^d} f(x)\mathrm{d}\mu(x)$ for an $f \in \mathrm{L}^1(\mu)$. Given a Markov kernel R on $\mathbb{R}^d$, for all $x \in \mathbb{R}^d$ and f integrable under $R(x,\cdot)$, denote by $Rf(x) = \int_{\mathbb{R}^d} f(y)R(x,\mathrm{d}y)$. Let $V : \mathbb{R}^d \to [1,\infty)$ be a measurable function. The V-total variation distance between μ and ν is defined as $\|\mu-\nu\|_V = \sup_{|f|\leq V} |\mu(f) - \nu(f)|$. If $V = 1$, then $\|\cdot\|_V$ is the total variation denoted by $\|\cdot\|_{\mathrm{TV}}$. Let μ and ν be two probability measures on a state space Ω with a given σ-algebra. If $\mu \ll \nu$, we denote by $\mathrm{d}\mu/\mathrm{d}\nu$ the Radon-Nikodym derivative of μ w.r.t. ν. In that case, the Kullback-Leibler divergence of μ w.r.t. to ν is defined as

$$\mathrm{KL}(\mu|\nu) = \int_\Omega \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right) \mathrm{d}\nu \, .$$

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

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

By [Vil09, Theorem 4.1], for all μ,ν probability measure on $\mathbb{R}^d$, there exists a transference plan $\zeta^\star \in \Pi(\mu,\nu)$ such that for any coupling (X,Y) distributed according to $\zeta^\star$, $W_p(\mu,\nu) = \mathbb{E}[\|X-Y\|^p]^{1/p}$.

For $u,v \in \mathbb{R}^d$, define the scalar product $\langle u,v\rangle = \sum_{i=1}^d u_i v_i$ and the Euclidian norm $\|u\| = \langle u,u\rangle^{1/2}$. Denote by $\mathbb{S}(\mathbb{R}^d) = \left\{u \in \mathbb{R}^d : \|u\| = 1\right\}$. For $k \in \mathbb{N}$, $m,m' \in \mathbb{N}^*$ and

Ω, Ω' two open sets of $\mathbb{R}^m, \mathbb{R}^{m'}$ respectively, denote by $\mathrm{C}^k(\Omega, \Omega')$, the set of k-times continuously differentiable functions. For $f \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R})$, denote by ∇f the gradient of f, Δf the Laplacian of f and $\nabla^2 f$ the Hessian of f. Define then for $x \in \mathbb{R}^d$, $\left\|\nabla^2 f(x)\right\| = \sup_{u \in \mathbb{R}^d, \|u\|=1} \left\|\nabla^2 f(x) u\right\|$. For $k \in \mathbb{N}$ and $f \in \mathrm{C}^k(\mathbb{R}^d, \mathbb{R})$, denote by $\mathrm{D}^i f$ the i-th derivative of f for $i \in \{0, \ldots, k\}$, *i.e.* $\mathrm{D}^i f$ is a symmetric i-linear map defined for all $x \in \mathbb{R}^d$ and $j_1, \ldots, j_i \in \{1, \ldots, d\}$ by $\mathrm{D}^i f(x)[e_{j_1}, \ldots, e_{j_l}] = \partial_{j_1 \ldots j_i} f(x)$ where $e_1, \ldots, e_d$ is the canonical basis of $\mathbb{R}^d$. For $x \in \mathbb{R}^d$ and $i \in \{1, \ldots, k\}$, define $\left\|\mathrm{D}^0 f(x)\right\| = |f(x)|$, $\left\|\mathrm{D}^i f(x)\right\| = \sup_{u_1, \ldots, u_i \in \mathbb{S}(\mathbb{R}^d)} \mathrm{D}^i f(x)[u_1, \ldots, u_i]$. Note that $\left\|\mathrm{D}^1 f(x)\right\| = \|\nabla f(x)\|$ and $\left\|\mathrm{D}^2 f(x)\right\| = \left\|\nabla^2 f(x)\right\|$. For $m, m' \in \mathbb{N}^*$, define

$$
\mathrm{C}_{\mathrm{poly}}(\mathbb{R}^m, \mathbb{R}^{m'}) = \Big\{ \mathsf{P} \in \mathrm{C}(\mathbb{R}^m, \mathbb{R}^{m'}) | \exists C_q, q \geq 0, \forall x \in \mathbb{R}^m, \\
\|\mathsf{P}(x)\| \leq C_q(1 + \|x\|^q) \Big\} .
$$

For $a \in \mathbb{R}_+$, $\lfloor a \rfloor$ is the integer part of a and $\lceil a \rceil = \lfloor a \rfloor + 1$. For $u \in \mathbb{R}^d$, $\operatorname{diag}(u)$ is a diagonal matrix in $\mathbb{R}^{d \times d}$. For all $x \in \mathbb{R}^d$ and $M > 0$, we denote by $\mathrm{B}(x, M)$ (respectively $\overline{\mathrm{B}}(x, M)$), the open (respectively close) ball centered at x of radius M. In the sequel, we take the convention that for $n, p \in \mathbb{N}$, $n < p$ then $\sum_p^n = 0$ and $\prod_p^n = 1$.

2 Ergodicity and convergence analysis

We assume below that U is continuously differentiable. Let introduce the following assumptions on U.

H1. *There exist $\ell, L \in \mathbb{R}_+$ such that for all $x, y \in \mathbb{R}^d$,*

$$
\|\nabla U(x) - \nabla U(y)\| \leq L \left\{ 1 + \|x\|^\ell + \|y\|^\ell \right\} \|x - y\| \ .
$$

H2. *1.* $\liminf_{\|x\| \to +\infty} \|\nabla U(x)\| = +\infty$.

2. $\liminf_{\|x\| \to +\infty} \left\langle \frac{x}{\|x\|}, \frac{\nabla U(x)}{\|\nabla U(x)\|} \right\rangle > 0$.

H2 is a standard assumption to show the ergodicity of the Markov chain targeting π, see e.g. [JH00, equation (34)] in the framework of the Random Walk Metropolis Hastings algorithm. Note that under **H2**, $\liminf_{\|x\| \to +\infty} U(x) = +\infty$, U has a minimum $x^\star$ and $\nabla U(x^\star) = 0$. Without loss of generality, it is assumed that $x^\star = 0$. It implies under **H1** that for all $x \in \mathbb{R}^d$,

$$
\|\nabla U(x)\| \leq 2L \left\{ 1 + \|x\|^{\ell+1} \right\} \ . \tag{5}
$$

Besides, under **H2**-2, there exists $C \in \mathbb{R}$ such that for all $x \in \mathbb{R}^d$, $\langle -\nabla U(x), x \rangle \leq C$. By [MT93, Theorem 2.1], [IW89, Chapter IV, Theorems 2.3, 3.1] and [RT96, Theorem 2.1], (1) has a unique strong solution. The distribution of $(Y_t)_{t \geq 0}$ defines a strongly Markovian semigroup $(P_t)_{t \geq 0}$ given for all $t \geq 0$, $x \in \mathbb{R}^d$ and $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$ by $P_t(x, \mathsf{A}) = \mathbb{E}_x \left[\mathbb{1}_{\{Y_t \in \mathsf{A}\}} \right]$.

Consider the infinitesimal generator $\mathscr{A}$ associated with (1) defined for all $h \in \mathrm{C}^2(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$ by

$$\mathscr{A} h(x) = -\langle \nabla U(x), \nabla h(x) \rangle + \Delta h(x) \, , \tag{6}$$

and for any $a \in \mathbb{R}_+^*$, define the Lyapunov function $V_a : \mathbb{R}^d \to [1, +\infty)$ for all $x \in \mathbb{R}^d$ by

$$V_a(x) = \exp\left(a(1 + \|x\|^2)^{1/2} \right) \, . \tag{7}$$

Foster-Lyapunov conditions enable to control the moments of the diffusion process $(Y_t)_{t\geq 0}$, see e.g. [MT93, Section 6] or [RT96, Theorem 2.2]. This methodology was applied for example in [MSH02, Lemma 5.2].

Proposition 1. *Assume* **H** 1 *and* **H** 2*. For all $a \in \mathbb{R}_+^*$, there exists $b_a \in \mathbb{R}_+$ (given explicitly in the proof) such that for all $x \in \mathbb{R}^d$,*

$$\mathscr{A} V_a(x) \leq -a V_a(x) + a b_a \, , \tag{8}$$

and,

$$\sup_{t \geq 0} P_t V_a(x) \leq V_a(x) + b_a \, .$$

Moreover, there exist $C_a \in \mathbb{R}_+$ and $\rho_a \in (0,1)$ such that for all $t \in \mathbb{R}_+$ and probability measures μ_0, ν_0 on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ satisfying $\mu_0(V_a) + \nu_0(V_a) < +\infty$,

$$\|\mu_0 P_t - \nu_0 P_t\|_{V_a} \leq C_a \rho_a^t \|\mu_0 - \nu_0\|_{V_a} \, , \; \|\mu_0 P_t - \pi\|_{V_a} \leq C_a \rho_a^t \mu_0(V_a) \, . \tag{9}$$

Proof. The proof is postponed to Section 4.2. □

The Markov chain $(X_k)_{k\in\mathbb{N}}$ defined in (3) is a discrete-time approximation of the diffusion $(Y_t)_{t\geq 0}$. To control the total variation and Wasserstein distances of the marginal distributions of $(X_k)_{k\in\mathbb{N}}$ and $(Y_t)_{t\geq 0}$, it is necessary to assume that for $\gamma > 0$ small enough, G_γ and ∇U are close. This is formalized by **A** 1. The counterpart of **H** 2 for the Markov chain $(X_k)_{k\in\mathbb{N}}$ is **A** 2; under this assumption, we obtain the stability and ergodicity of $(X_k)_{k\in\mathbb{N}}$. Lemma 2 below shows that both assumptions are satisfied for G_γ and $\mathsf{G}_{\gamma,c}$ defined in (4) assuming **H** 1 and **H** 2.

A 1. *For all $\gamma > 0$, G_γ is continuous. There exist $\alpha \geq 0$, $C_\alpha < +\infty$ such that for all $\gamma > 0$ and $x \in \mathbb{R}^d$,*

$$\|G_\gamma(x) - \nabla U(x)\| \leq \gamma C_\alpha \left(1 + \|x\|^\alpha\right) \, .$$

A 2. *For all $\gamma > 0$, $\liminf_{\|x\|\to+\infty} \left\langle \frac{x}{\|x\|}, G_\gamma(x) \right\rangle - \frac{\gamma}{2\|x\|} \|G_\gamma(x)\|^2 > 0$.*

Lemma 2. *Assume* **H** 1 *and* **H** 2*. Let $\gamma > 0$ and G_γ be defined by G_γ or $\mathsf{G}_{\gamma,c}$. Then* **A** 1 *and* **A** 2 *are satisfied.*

Proof. The proof is postponed to Section 4.1. □

The Markov kernel R_γ associated with (3) is given for all $\gamma > 0$, $x \in \mathbb{R}^d$ and $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$ by

$$R_\gamma(x, \mathsf{A}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \mathbb{1}_{\mathsf{A}} \left(x - \gamma G_\gamma(x) + \sqrt{2\gamma} z \right) \mathrm{e}^{-\|z\|^2/2} \mathrm{d}z \ . \tag{10}$$

We then obtain the counterpart of Proposition 1 for the Markov chain $(X_k)_{k\in\mathbb{N}}$.

Proposition 3. *Assume **H** 1, **A** 1 and **A** 2. For all $\gamma \in \mathbb{R}_+^*$, there exist $M, æ, b \in \mathbb{R}_+^*$ (given explicitly in the proof) satisfying for all $x \in \mathbb{R}^d$*

$$R_\gamma V_æ(x) \leq \mathrm{e}^{-æ^2\gamma} V_æ(x) + \gamma b \mathbb{1}_{\overline{\mathrm{B}}(0,M)}(x) \ , \tag{11}$$

and we have for all $n \in \mathbb{N}$,

$$R_\gamma^n V_æ(x) \leq \mathrm{e}^{-æ^2 n\gamma} V_æ(x) + (b/æ^2) \mathrm{e}^{æ^2\gamma} \ . \tag{12}$$

In addition, for all $\gamma > 0$, R_γ has a unique invariant measure π_γ and for all $\varsigma \in (0, 1]$, R_γ is $V_æ^\varsigma$-geometrically ergodic w.r.t. π_γ.

Proof. The proof is postponed to Section 4.3. □

In the following result, we compare the discrete and continuous time processes $(X_k)_{k\in\mathbb{N}}$ and $(Y_t)_{t\geq 0}$ using Girsanov's theorem and Pinsker's inequality, see e.g. [Dal17] and [DM17, Theorem 10] for similar arguments.

Theorem 4. *Assume **H** 1, **H** 2, **A** 1 and **A** 2. Let $\gamma_0 > 0$. There exist $C > 0$ and $\lambda \in (0, 1)$ such that for all $\gamma \in (0, \gamma_0]$, $x \in \mathbb{R}^d$ and $n \in \mathbb{N}$,*

$$\left\| \delta_x R_\gamma^n - \pi \right\|_{V_æ^{1/2}} \leq C \left(\lambda^{n\gamma} V_æ(x) + \sqrt{\gamma} \right) \ , \tag{13}$$

where æ is defined in Proposition 3 and for all $\gamma \in (0, \gamma_0]$,

$$\| \pi_\gamma - \pi \|_{V_æ^{1/2}} \leq C\sqrt{\gamma} \ . \tag{14}$$

Proof. The proof is postponed to Section 4.4. □

For the Wasserstein distance of order 2, we introduce the following additional assumption on U.

H 3. *U is strongly convex, i.e. there exists $m > 0$ such that for all $x, y \in \mathbb{R}^d$,*

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

In this context, the assumption is classical, see e.g. [Ebe15], [DM16], [Dal17]. By coupling $(Y_t)_{t\geq 0}$ and the linear interpolation of $(X_k)_{k\in\mathbb{N}}$ with the same Brownian motion, the following result is obtained.

Theorem 5. *Assume* **A** 1, **A** 2, **H** 1, **H** 2 *and* **H** 3. *Let* $\gamma_0 > 0$. *There exist* $C > 0$ *and* $\lambda \in (0,1)$ *such that for all* $x \in \mathbb{R}^d$, $\gamma \in (0, \gamma_0]$ *and* $n \in \mathbb{N}$,

$$W_2^2(\delta_x R_\gamma^n, \pi) \leq C \left(\lambda^{n\gamma} V_{æ}(x) + \gamma\right) , \tag{15}$$

where æ *is defined in Proposition* 3 *and for all* $\gamma \in (0, \gamma_0]$,

$$W_2^2(\pi_\gamma, \pi) \leq C\gamma . \tag{16}$$

Proof. The proof is postponed to Section 4.5. □

If $U \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R})$ and under the following assumption on $\nabla^2 U$, the bound can be improved.

H 4. *U is twice continuously differentiable and there exist* $\nu, L_H \in \mathbb{R}_+$ *and* $\beta \in [0,1]$ *such that for all* $x, y \in \mathbb{R}^d$,

$$\left\|\nabla^2 U(x) - \nabla^2 U(y)\right\| \leq L_H \left\{1 + \|x\|^\nu + \|y\|^\nu\right\} \|x - y\|^\beta .$$

It is shown in Section 4.5 that **H**4 implies **H**1.

Theorem 6. *Assume* **A** 1, **A** 2, **H** 2, **H** 3 *and* **H** 4. *Let* $\gamma_0 > 0$. *There exist* $C > 0$ *and* $\lambda \in (0,1)$ *such that for all* $x \in \mathbb{R}^d$, $\gamma \in (0, \gamma_0]$ *and* $n \in \mathbb{N}$,

$$W_2^2(\delta_x R_\gamma^n, \pi) \leq C \left(\lambda^{n\gamma} V_{æ}(x) + \gamma^{1+\beta}\right) , \tag{17}$$

where æ *is defined in Proposition* 3 *and for all* $\gamma \in (0, \gamma_0]$,

$$W_2^2(\pi_\gamma, \pi) \leq C\gamma^{1+\beta} . \tag{18}$$

Proof. The proof is postponed to Section 4.5. □

The exponent of γ in (17) is improved from 1 to $1+\beta$. In particular, if $\nabla^2 U$ is locally Lipschitz, $\beta = 1$, and [DM16, Theorem 8] is recovered. For f a π-integrable function, the Poisson equation associated with the generator $\mathscr{A}$ defined in (6) is given for all $x \in \mathbb{R}^d$ by

$$\mathscr{A}\phi(x) = -\left(f(x) - \pi(f)\right) ,$$

where ϕ, if it exists, is called the solution of the Poisson equation. This equation has proved to be a useful tool to analyze additive functionals of diffusion processes, see e.g. [CCG12] and references therein. The existence and regularity of the solution of the Poisson equation is investigated in [GM96], [PV01], [Kop15], [Gor+16].

Let $(X_k)_{k\in\mathbb{N}}$ be the Markov chain defined in (3). To analyze the empirical average $(1/n)\sum_{k=0}^{n-1}\{f(X_k) - \pi(f)\}$ for $n \in \mathbb{N}^*$, we follow a method introduced in [MST10] and based on the Poisson equation. We first state a result on the solution of the Poisson equation under appropriate assumptions on U and f.

H 5. $U \in \mathrm{C}^4(\mathbb{R}^d, \mathbb{R})$ *and* $\left\|\mathrm{D}^i U\right\| \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R}_+)$ *for* $i \in \{1, \ldots, 4\}$.

Proposition 7. *Assume* **H**2 *and* **H**5*. Let* $f \in \mathrm{C}^3(\mathbb{R}^d, \mathbb{R})$ *be such that* $\left\| \mathrm{D}^i f \right\| \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R}_+)$ *for* $i \in \{0, \ldots, 3\}$*. Then, there exists* $\phi \in \mathrm{C}^4(\mathbb{R}^d, \mathbb{R})$*, such that for all* $x \in \mathbb{R}^d$ *and* $i \in \{0, \ldots, 4\}$*,*

$$\mathscr{A}\phi(x) = -\left(f(x) - \pi(f)\right) \quad \textit{and} \quad \left\| \mathrm{D}^i \phi \right\| \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R}_+) \; .$$

Proof. The proof is postponed to Section 4.6. □

Theorem 8. *Assume* **H**2*,* **H**5*,* **A**1 *and* **A**2*. Let* $f \in \mathrm{C}^3(\mathbb{R}^d, \mathbb{R})$ *be such that* $\left\| \mathrm{D}^i f \right\| \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R}_+)$ *for* $i \in \{0, \ldots, 3\}$*. Let* $\gamma_0 > 0$ *and* $(X_k)_{k \in \mathbb{N}}$ *be the Markov chain defined by* (3) *and starting at* $X_0 = 0$*. There exists* $C > 0$ *such that for all* $\gamma \in (0, \gamma_0]$ *and* $n \in \mathbb{N}^*$*,*

$$\left| \mathbb{E}\left[\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) - \pi(f) \right] \right| \leq C \left(\gamma + \frac{1}{n\gamma} \right) \; , \tag{19}$$

and

$$\mathbb{E}\left[\left(\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) - \pi(f) \right)^2 \right] \leq C \left(\gamma^2 + \frac{1}{n\gamma} \right) \; . \tag{20}$$

Proof. The proof is postponed to Section 4.7. □

Note that the standard rates of convergence are recovered, see [MST10, Theorems 5.1, 5.2].

3 Numerical examples

We illustrate our theoretical results using three numerical examples.

Multivariate Gaussian variable in high dimension We first consider a multivariate Gaussian variable in dimension $d \in \{100, 1000\}$ of mean 0 and covariance matrix $\Sigma = \mathrm{diag}\left((i)_{i \in \{1, \ldots, d\}} \right)$. The potential $U : \mathbb{R}^d \to \mathbb{R}$ defined for all $x \in \mathbb{R}^d$ by $U(x) = (1/2) x^{\mathrm{T}} \Sigma^{-1} x$ is d^{-1}-strongly convex and 1-gradient Lipschitz. The assumptions **H**1, **H**2, **H**3, **H**4 with $\beta = 1$ and **H**5 are thus satisfied. Note that in this case, ULA is stable and the analysis of [Dal17], [DM17], [DM16] valid.

Double well The potential is defined for all $x \in \mathbb{R}^d$ by $U(x) = (1/4)\left\| x \right\|^4 - (1/2)\left\| x \right\|^2$. We have $\nabla U(x) = \left(\left\| x \right\|^2 - 1 \right) x$ and $\nabla^2 U(x) = \left(\left\| x \right\|^2 - 1 \right) \mathrm{Id} + 2 x x^{\mathrm{T}}$. We get $\left\| \nabla^2 U(x) \right\| = 3\left\| x \right\|^2 - 1$, $\langle x, \nabla U(x) \rangle = \left\| x \right\| \left\| \nabla U(x) \right\|$ for $\left\| x \right\| \geq 1$ and

$$\left\| \nabla^2 U(x) - \nabla^2 U(y) \right\| \leq 3 \left(\left\| x \right\| + \left\| y \right\| \right) \left\| x - y \right\| \; ,$$

so that **H**1, **H**2, **H**4 with $\beta = 1$ and **H**5 are satisfied.

Ginzburg-Landau model This model of phase transitions in physics [LFR17, Section 6.2] is defined on a three-dimensional $d = p^3$ lattice for $p \in \mathbb{N}^*$ and the potential is given for $x = (x_{ijk})_{i,j,k \in \{1,\dots,p\}} \in \mathbb{R}^d$ by

$$U(x) = \sum_{i,j,k=1}^{p} \left\{ \frac{1-\tau}{2} x_{ijk}^2 + \frac{\tau\alpha}{2} \left\| \tilde{\nabla} x_{ijk} \right\|^2 + \frac{\tau\lambda}{4} x_{ijk}^4 \right\} ,$$

where $\alpha, \lambda, \tau > 0$ and $\tilde{\nabla} x_{ijk} = (x_{i_+jk} - x_{ijk}, x_{ij_+k} - x_{ijk}, x_{ijk_+} - x_{ijk})$ with $i_\pm = i \pm 1 \mod p$ and similarly for $j_\pm, k_\pm$. In the simulations, p is equal to 10. We have

$$\nabla U(x) = \Big\{ \tau\alpha \left(6x_{ijk} - x_{i_+jk} - x_{ij_+k} - x_{ijk_+} - x_{i_-jk} - x_{ij_-k} - x_{ijk_-} \right) \\ + (1-\tau) x_{ijk} + \tau\lambda x_{ijk}^3 \Big\}_{i,j,k \in \{1,\dots,p\}} ,$$

and

$$\nabla^2 U(x) = \operatorname{diag}\left(\left(1 - \tau + 6\tau\alpha + 3\tau\lambda x_{ijk}^2 \right)_{i,j,k \in \{1,\dots,p\}} \right) + M ,$$

where $M \in \mathbb{R}^{d \times d}$ is a constant matrix. **H**1, **H**4 with $\beta = 1$ and **H**5 are thus satisfied. Using that $x \mapsto \sum_{i,j,k=1}^{p} \left\| \tilde{\nabla} x_{ijk} \right\|^2$ is convex, we have for all $x \in \mathbb{R}^d$,

$$\langle x, \nabla U(x) \rangle \geq \sum_{i,j,k=1}^{p} \{ (1-\tau) x_{ijk}^2 + \tau\lambda x_{ijk}^4 \} .$$

By Cauchy-Schwarz inequality, $\left\{ \sum_{i,j,k=1}^{p} x_{ijk}^2 \right\}^2 \leq d \sum_{i,j,k=1}^{p} x_{ijk}^4$, and for all $x \in \mathbb{R}^d$, $\|x\|^2 \geq (2 |1-\tau| d)/(\tau\lambda)$, we get $\langle x, \nabla U(x) \rangle \geq \{(\tau\lambda)/2\} \sum_{i,j,k=1}^{p} x_{ijk}^4$. Besides, we have

$$\|\nabla U(x)\| \leq (|1-\tau| + 12\tau\alpha) \|x\| + \tau\lambda \left\| (x_{ijk}^3)_{i,j,k \in \{1,\dots,p\}} \right\| .$$

Let $a, b, c \in \{1, \dots, p\}$ be such that $|x_{abc}| = \max |x_{ijk}|$. We get

$$\|x\| \left\| (x_{ijk}^3)_{i,j,k \in \{1,\dots,p\}} \right\| \leq d x_{abc}^4 \leq d \sum_{i,j,k=1}^{p} x_{ijk}^4 .$$

Finally, for $\|x\|^2 \geq \max\{1, (2 |1-\tau| d)/(\tau\lambda)\}$, we obtain

$$\|x\| \|\nabla U(x)\| \leq (1 + (24\alpha d)/\lambda + 2d) \langle x, \nabla U(x) \rangle ,$$

and **H**2 is satisfied.

We benchmark TULA and TULAc against ULA given by (2), MALA and a Random Walk Metropolis-Hastings with a Gaussian proposal (RWM). TMALA (Tamed Metropolis Adjusted Langevin Algorithm) and TMALAc (Tamed Metropolis Adjusted Langevin Algorithm coordinate-wise), the Metropolized versions of TULA and TULAc, are also included in the numerical tests. Their theoretical analysis is similar to the one of MALTA [Atc06, Proposition 2.1].

The double well and Ginzburg-Landau model are rotationnally invariant and the results are provided only for their first coordinate. The Markov chains are started at $X_0 = 0, (10, 0^{\otimes(d-1)}), (100, 0^{\otimes(d-1)}), (1000, 0^{\otimes(d-1)})$ and for the multivariate Gaussian at a random vector of norm $0, 10, 100, 1000$. For the Gaussian and double well examples, for each initial condition, algorithm, step size $\gamma \in \{10^{-3}, 10^{-2}, 10^{-1}, 1\}$, we run 100 independent Markov chains started at X_0 of 10^6 samples (respectively 10^5) in dimension $d = 100$ (respectively $d = 1000$). For the Ginzburg-Landau model, we run 100 independent Markov chains started at X_0 of 10^5 samples. Then, we compute the boxplots of the errors on the 1st and 2nd moment for the first and the last coordinate, *i.e.* $\mathbb{E}[X_i] - \pi(X_i)$ and $\mathbb{E}\left[X_i^2\right] - \pi(X_i^2)$ for $i \in \{1, d\}$. For ULA, if the norm of X_k for $k \in \mathbb{N}$ exceeds 10^5, the chain is stopped and for this step size γ the trajectory of ULA is not taken into account. For MALA, RWM, TMALA and TMALAc, if the acceptance ratio is below 0.05, we similarly do not take into account the corresponding trajectories.

For the three examples and for $i \in \{1, \ldots, d\}$, $\mathbb{E}[X_i] = 0$. By symmetry, for the double well, we have for $i \in \{1, \ldots, d\}$ and $r \in \mathbb{R}_+$,

$$\mathbb{E}\left[X_i^2\right] = d^{-1} \int_{\mathbb{R}_+} r^2 \nu(r) \mathrm{d}r \Big/ \int_{\mathbb{R}_+} \nu(r) \mathrm{d}r \;, \quad \nu(r) = r^{d-1} \exp\left\{(r^2/2) - (r^4/4)\right\} \;.$$

A Random Walk Metropolis run of 10^7 samples gives $\mathbb{E}\left[X_i^2\right] = 0.104 \pm 0.001$ for $d = 100$ and $\mathbb{E}\left[X_i^2\right] = 0.032 \pm 0.001$ for $d = 1000$.

Because of lack of space, we only display some boxplots in Figures 1 to 4. The Python code and all the figures are available at `https://github.com/nbrosse/TULA`. We remark that TULA, TULAc and to a lesser extent, TMALA and TMALAc, have a stable behaviour even with large step sizes and starting far from the origin. This is particularly visible in Figures 2 and 4 where ULA diverges (*i.e.* $\liminf_{k \to +\infty} \mathbb{E}[\|X_k\|] = +\infty$) and MALA does not move even for small step sizes $\gamma = 10^{-3}$. Note however the existence of a bias for ULA, TULA and TULAc in Figure 3. Finally, comparison of the results shows that TULAc is preferable to TULA.

Note that other choices are possible for G_γ, depending on the model under study. For example, in the case of the double well, we could "tame" only the superlinear part of ∇U, *i.e.* consider for all $\gamma > 0$ and $x \in \mathbb{R}^d$,

$$G_\gamma(x) = \frac{\|x\|^2 x}{1 + \gamma \|x\|^2} - x \;. \tag{21}$$

A1 is satisfied and we have

$$\begin{aligned}
\left\langle \frac{x}{\|x\|}, G_\gamma(x) \right\rangle - \frac{\gamma}{2\|x\|} \|G_\gamma(x)\|^2 &= \frac{\|x\|^3}{1 + \gamma \|x\|^2} \left\{1 + \gamma - \frac{\gamma}{2} \frac{\|x\|^2}{1 + \gamma \|x\|^2}\right\} \\
&\qquad - \|x\| \left\{1 + (\gamma/2)\right\} \;, \\
\liminf_{\|x\| \to +\infty} \left\langle \frac{x}{\|x\|^2}, G_\gamma(x) \right\rangle - \frac{\gamma}{2\|x\|^2} \|G_\gamma(x)\|^2 &= \frac{\gamma^{-1} - \gamma}{2} \;.
\end{aligned}$$

A2 is satisfied if and only if $\gamma \in (0,1)$. It is striking to see that this theoretical threshold is clearly visible on the simulations, see Figure 5 where the algorithm (3) with G_γ defined by (21) is denoted by sTULA (for "smart" TULA).

4 Proofs

4.1 Proof of Lemma 2

Let $\gamma > 0$. We have for all $x \in \mathbb{R}^d$, $\|\mathsf{G}_\gamma(x) - \nabla U(x)\| \le \gamma \|\nabla U(x)\|^2$ and

$$\|\mathsf{G}_{\gamma,c}(x) - \nabla U(x)\| \le \gamma \left\{ \sum_{i=1}^{d} (\partial_i U(x))^4 \right\}^{1/2} \le \gamma \|\nabla U(x)\|^2 \ .$$

By (5), **A**1 is satisfied. Define for all $x \in \mathbb{R}^d$, $x \neq 0$,

$$\begin{aligned} A_\gamma(x) &= \left\langle \frac{x}{\|x\|}, \mathsf{G}_\gamma(x) \right\rangle - \frac{\gamma}{2\|x\|} \|\mathsf{G}_\gamma(x)\|^2 \ , \\ B_\gamma(x) &= \left\langle \frac{x}{\|x\|}, \mathsf{G}_{\gamma,c}(x) \right\rangle - \frac{\gamma}{2\|x\|} \|\mathsf{G}_{\gamma,c}(x)\|^2 \ . \end{aligned}$$

By **H**2-2, there exist $M_1, \kappa > 0$ such that for all $x \in \mathbb{R}^d$, $\|x\| \ge M_1$, $\langle x, \nabla U(x) \rangle \ge \kappa \|x\| \|\nabla U(x)\|$. We get then for all $x \in \mathbb{R}^d$, $\|x\| \ge M_1$,

$$\begin{aligned} A_\gamma(x) &= \frac{1}{1+\gamma\|\nabla U(x)\|} \frac{1}{2\|x\|} \left\{ 2\langle x, \nabla U(x)\rangle - \|\nabla U(x)\| \frac{\gamma\|\nabla U(x)\|}{1+\gamma\|\nabla U(x)\|} \right\} \\ &\ge \frac{\|\nabla U(x)\|}{1+\gamma\|\nabla U(x)\|} \frac{1}{2\|x\|} (2\kappa\|x\| - 1) \ . \end{aligned}$$

By **H**2-1, there exist $M_2, C > 0$ such that for all $x \in \mathbb{R}^d$, $\|x\| \ge M_2$, $\|\nabla U(x)\| \ge C$. Using that $s \mapsto s(1+\gamma s)^{-1}$ is non-decreasing for $s \ge 0$, we get for all $x \in \mathbb{R}^d$, $\|x\| \ge \max(\kappa^{-1}, M_1, M_2)$, $A_\gamma(x) \ge (\kappa C)/\{2(1+\gamma C)\}$.

For B_γ, we have for all $x \in \mathbb{R}^d$, $\gamma \|\mathsf{G}_{\gamma,c}(x)\| \le \sqrt{d}$ and for all $x \in \mathbb{R}^d$, $\|x\| \ge M_1$,

$$\left\langle x, \left(\frac{\partial_i U(x)}{1+\gamma|\partial_i U(x)|} \right)_{i\in\{1,\ldots,d\}} \right\rangle \ge \frac{\kappa\|x\|\|\nabla U(x)\|}{1+\gamma\max_{i\in\{1,\ldots,d\}}|\partial_i U(x)|} \ .$$

Using that $s \mapsto s(1+\gamma s)^{-1}$ is non-decreasing for $s \ge 0$, we have

$$\begin{aligned} \left\| \left(\frac{\partial_i U(x)}{1+\gamma|\partial_i U(x)|} \right)_{i\in\{1,\ldots,d\}} \right\| &\le \frac{\sqrt{d}\max_{i=1,\ldots,d}|\partial_i U(x)|}{1+\gamma\max_{i\in\{1,\ldots,d\}}|\partial_i U(x)|} \\ &\le \frac{\sqrt{d}\|\nabla U(x)\|}{1+\gamma\max_{i\in\{1,\ldots,d\}}|\partial_i U(x)|} \ , \end{aligned}$$

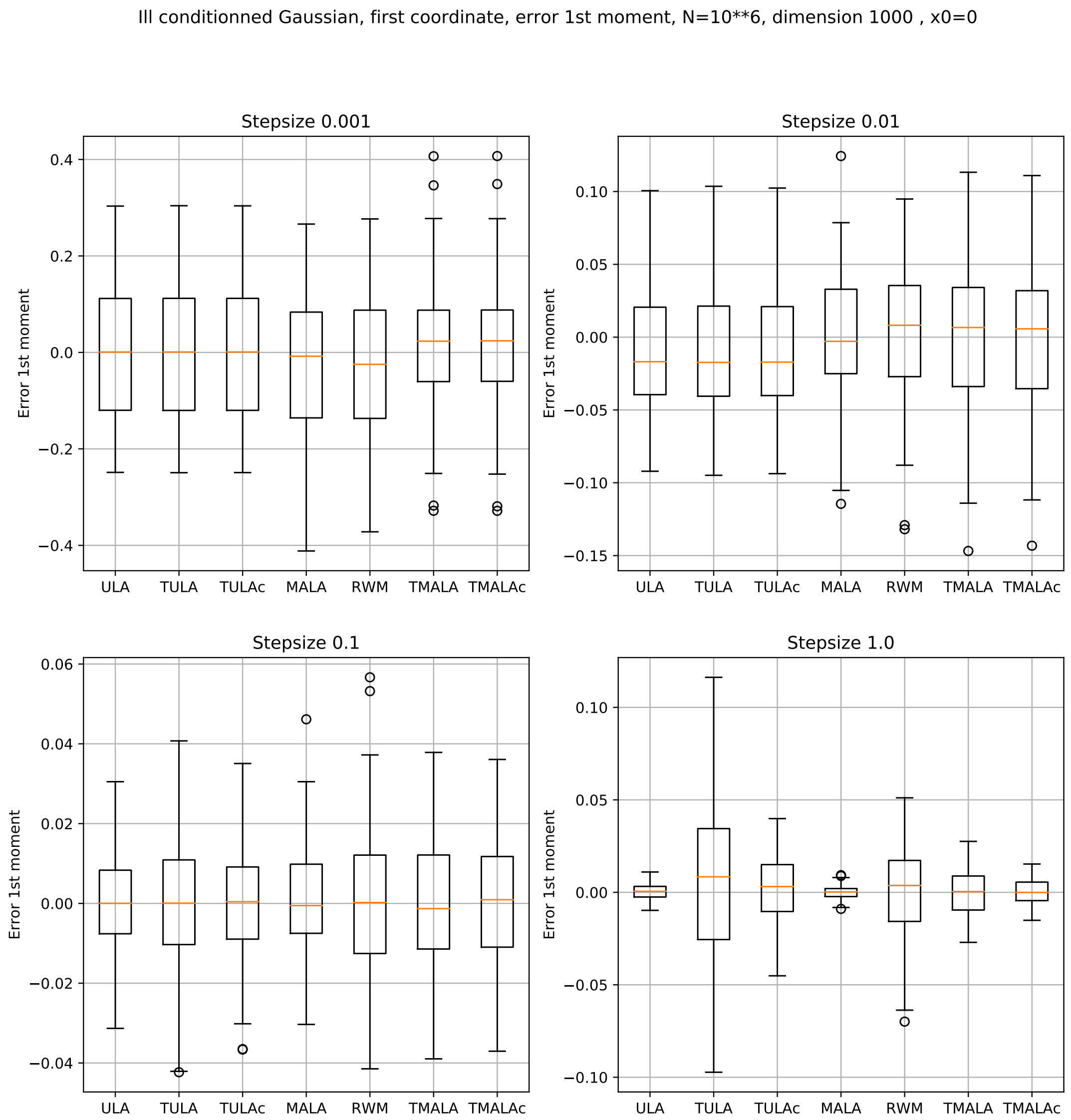

Figure 1: Boxplots of the first order error for the multivariate Gaussian (first coordinate) in dimension 1000 starting at 0 for different step sizes.

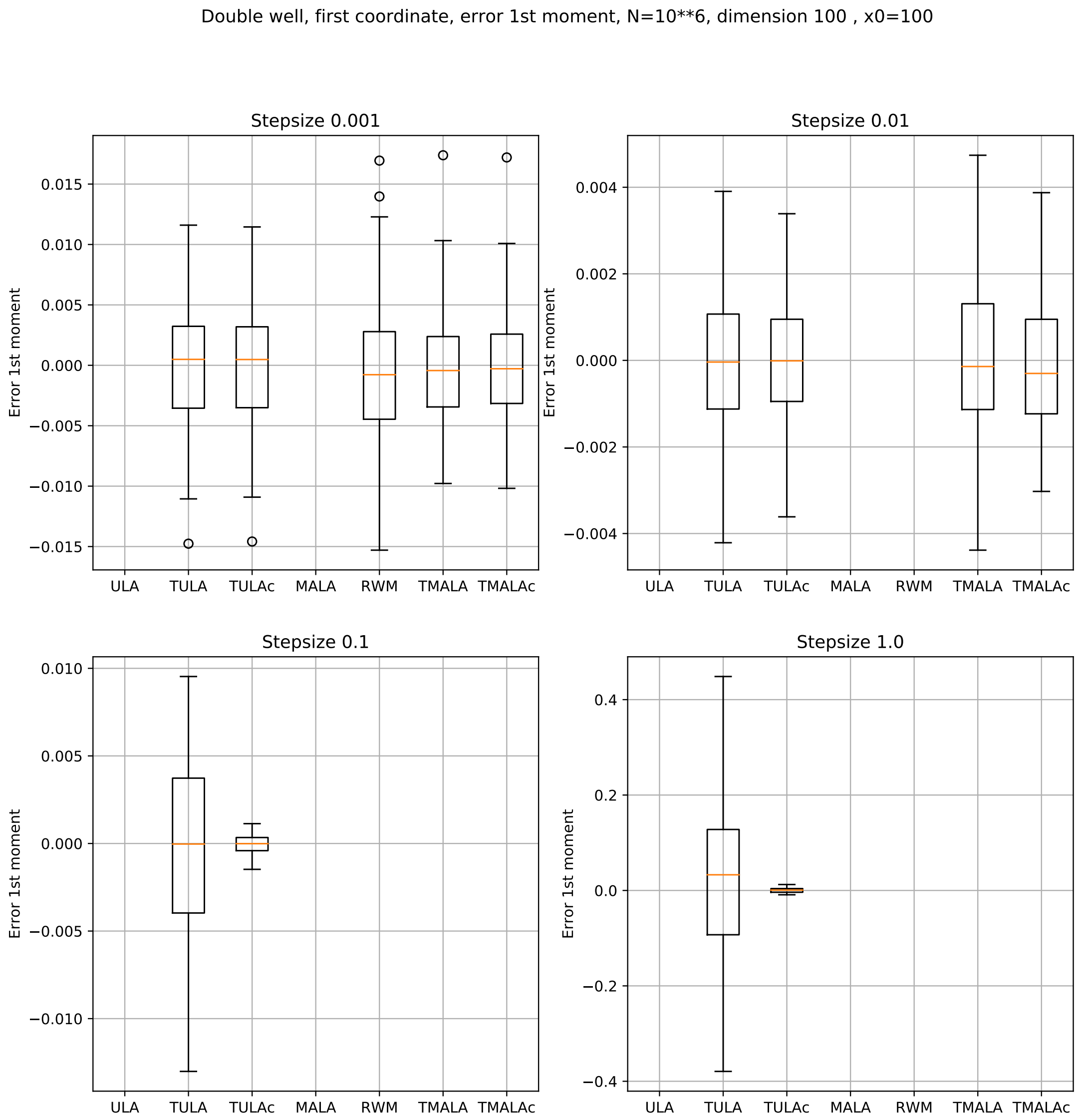

Figure 2: Boxplots of the first order error for the double well in dimension 100 starting at $(100, 0^{\otimes 99})$ for different step sizes.

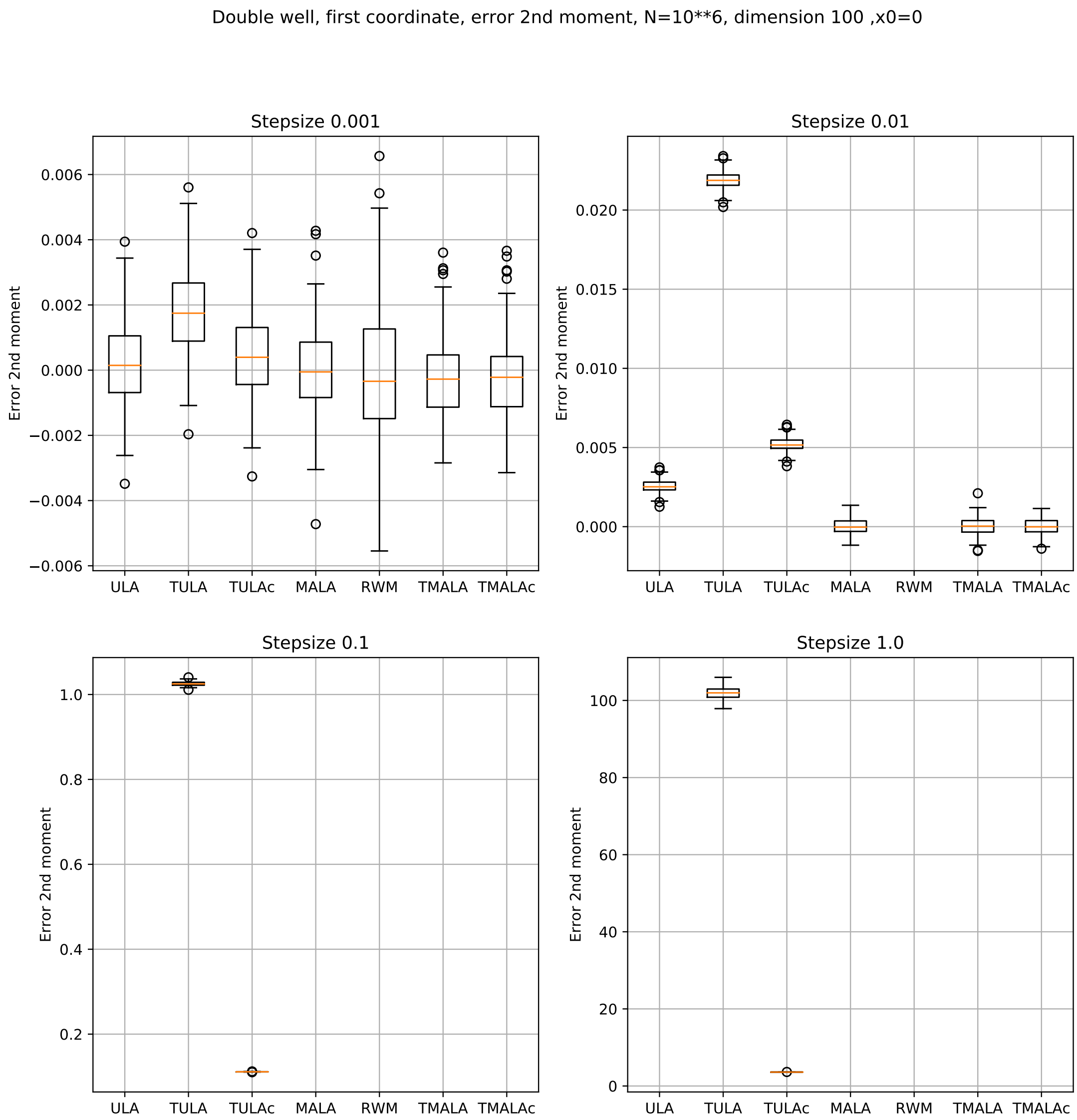

Figure 3: Boxplots of the second order error for the double well in dimension 100 starting at 0 for different step sizes.

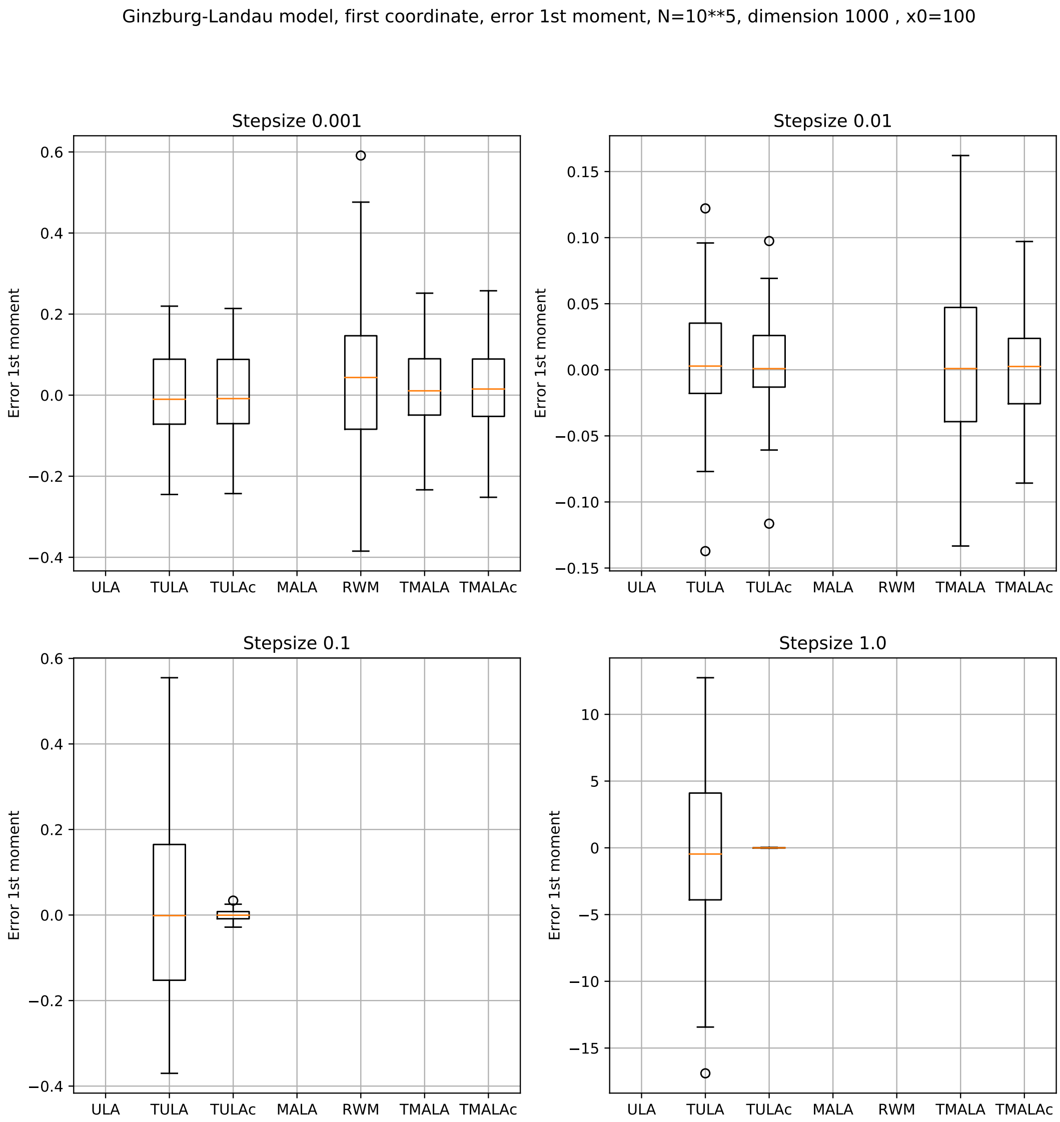

Figure 4: Boxplots of the first order error for the Ginzburg-Landau model in dimension 1000 starting at $(100, 0^{\otimes 999})$ for different step sizes.

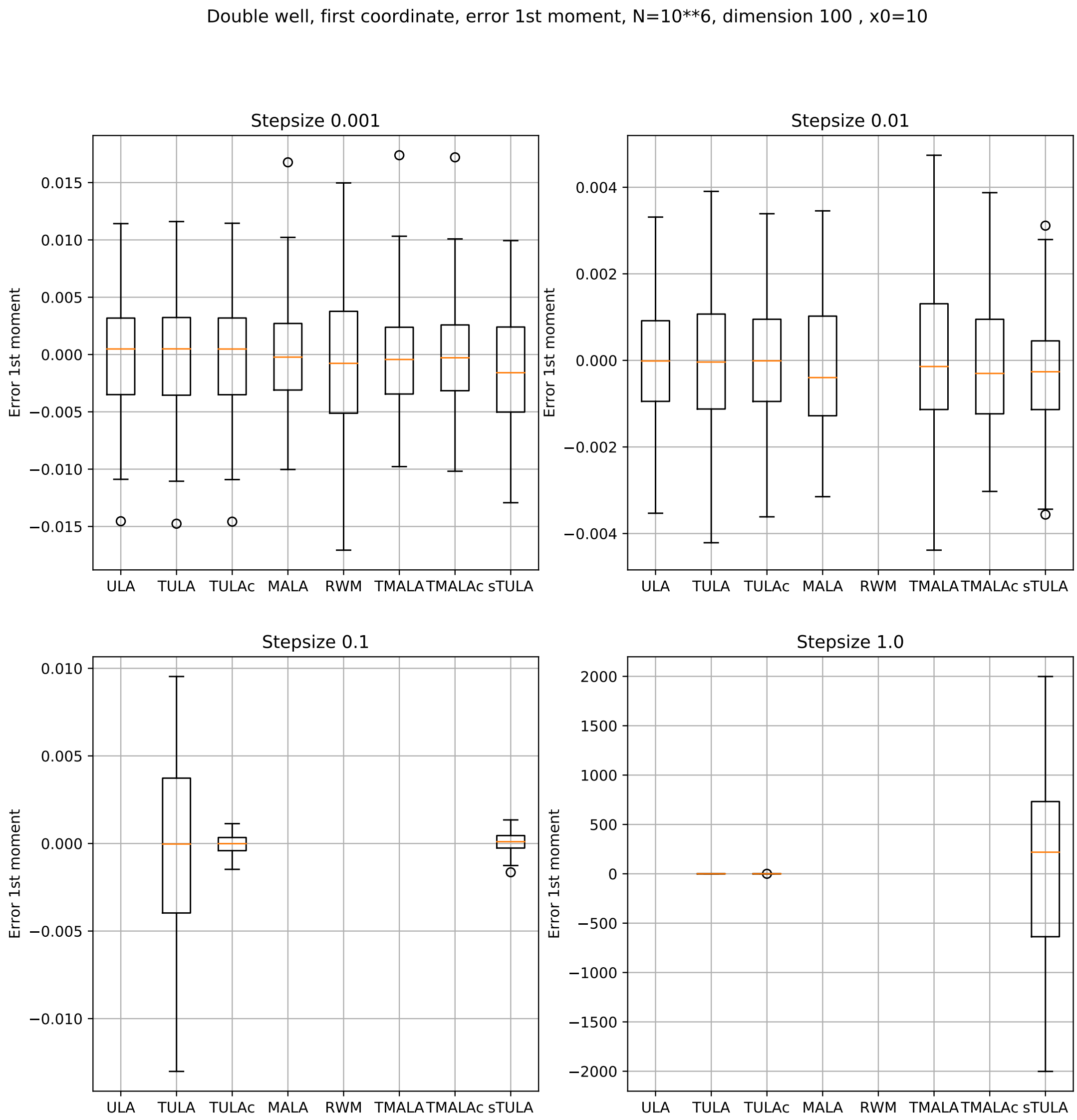

Figure 5: Boxplots of the first order error for the double well in dimension 100 starting at $(10, 0^{\otimes 99})$ for different step sizes, including the results of the "smart" TULA.

and combining these inequalities, we get for all $x \in \mathbb{R}^d$, $\|x\| \geq \max(\kappa^{-1} d, M_1)$,

$$B_\gamma(x) \geq \frac{\|\nabla U(x)\|}{1 + \gamma \max_{i \in \{1,\ldots,d\}} |\partial_i U(x)|} \frac{1}{2\|x\|} \{2\kappa \|x\| - d\} \geq \frac{\|\nabla U(x)\|}{1 + \gamma \|\nabla U(x)\|} \frac{\kappa}{2} ,$$

and for all $x \in \mathbb{R}^d$, $\|x\| \geq \max(\kappa^{-1} d, M_1, M_2)$, we get $B_\gamma(x) \geq (\kappa C)/\{2(1+\gamma C)\}$.

4.2 Proof of Proposition 1

We have for all $x \in \mathbb{R}^d$,

$$\frac{\mathscr{A} V_a(x)}{a V_a(x)} = -\left\langle \nabla U(x), \frac{x}{(1+\|x\|^2)^{1/2}} \right\rangle + \frac{a\|x\|^2}{1+\|x\|^2} + \frac{d}{(1+\|x\|^2)^{1/2}} - \frac{\|x\|^2}{(1+\|x\|^2)^{3/2}} . \quad (22)$$

By **H**2-2 and using $s \mapsto s/(1+s^2)^{1/2}$ is non-decreasing for $s \geq 0$, there exist $M_1, \kappa \in \mathbb{R}_+^*$ such that for all $x \in \mathbb{R}^d$, $\|x\| \geq M_1$, $\left\langle \nabla U(x), x(1+\|x\|^2)^{-1/2} \right\rangle \geq \kappa \|\nabla U(x)\|$. By **H**2-1, there exists $M_2 \geq M_1$ such that for all $x \in \mathbb{R}^d$, $\|x\| \geq M_2$, $\|\nabla U(x)\| \geq \kappa^{-1}\{1 + a + d(1 + M_1^2)^{-1/2}\}$. We then have for all $x \in \mathbb{R}^d$, $\|x\| \geq M_2$, $\mathscr{A} V_a(x) \leq -a V_a(x)$. Define

$$b_a = \exp(a(1+M_2^2)^{1/2})\{2L(1+M_2^{\ell+1}) + a + d\} .$$

Combining (5) and (22) gives (8). Applying [MT93, Theorem 1.1] with $V(x,t) = V_a(x)\mathrm{e}^{at}$, $g_-(t) = 0$ and $g_+(t) = a b_a \mathrm{e}^{at}$ for $x \in \mathbb{R}^d$ and $t \geq 0$, we get $P_t V_a(x) \leq \mathrm{e}^{-at} V_a(x) + b_a(1 - \mathrm{e}^{-at})$. Eq. (9) is a consequence of [RT96, Theorem 2.2] and [MT93, Theorem 6.1].

4.3 Proof of Proposition 3

Let $\gamma, a \in \mathbb{R}_+^*$. Note that the function $x \mapsto (1+\|x\|^2)^{1/2}$ is Lipschitz continuous with Lipschitz constant equal to 1. By the log-Sobolev inequality [BGL14, Proposition 5.5.1], and the Cauchy-Schwarz inequality, we have for all $x \in \mathbb{R}^d$ and $a > 0$

$$\begin{aligned} R_\gamma V_a(x) &\leq \mathrm{e}^{a^2\gamma} \exp\left\{ a \int_{\mathbb{R}^d} (1+\|y\|^2)^{1/2} R_\gamma(x, \mathrm{d}y) \right\} \\ &\leq \mathrm{e}^{a^2\gamma} \exp\left\{ a \left(1 + \|x - \gamma G_\gamma(x)\|^2 + 2\gamma d\right)^{1/2} \right\} . \end{aligned} \quad (23)$$

We now bound the term inside the exponential in the right hand side. For all $x \in \mathbb{R}^d$,

$$\|x - \gamma G_\gamma(x)\|^2 = \|x\|^2 - 2\gamma \left(\langle G_\gamma(x), x \rangle - (\gamma/2) \|G_\gamma(x)\|^2 \right) . \quad (24)$$

By **A** 2, there exist $M_1, \kappa \in \mathbb{R}_+^*$ such that for all $x \in \mathbb{R}^d$, $\|x\| \geq M_1$, $\langle x, G_\gamma(x) \rangle - (\gamma/2)\|G_\gamma(x)\|^2 \geq \kappa \|x\|$. Denote by $M = \max(M_1, 2d\kappa^{-1})$. For all $x \in \mathbb{R}^d$, $\|x\| \geq M$, we have

$$\|x - \gamma G_\gamma(x)\|^2 + 2\gamma d \leq \|x\|^2 - \gamma\kappa \|x\| .$$

Using for all $t \in [0,1]$, $(1-t)^{1/2} \leq 1 - t/2$ and $s \mapsto s/(1+s^2)^{1/2}$ is non-decreasing for $s \geq 0$, we have for all $x \in \mathbb{R}^d$, $\|x\| \geq M$,

$$\left(1 + \|x - \gamma G_\gamma(x)\|^2 + 2\gamma d\right)^{1/2} \leq \left(1 + \|x\|^2\right)^{1/2} \left(1 - \frac{\gamma\kappa \|x\|}{1 + \|x\|^2}\right)^{1/2}$$
$$\leq \left(1 + \|x\|^2\right)^{1/2} - \frac{\gamma\kappa M}{2(1+M^2)^{1/2}} \; .$$

Plugging this result in (23) shows that for all $x \in \mathbb{R}^d$, $\|x\| \geq M$,

$$R_\gamma V_{æ}(x) \leq \mathrm{e}^{-æ^2\gamma} V_{æ}(x) \quad \text{for} \quad æ = \frac{\kappa M}{4(1+M^2)^{1/2}} \; . \tag{25}$$

Define

$$c = æ^2 + æ\Big[M\left\{2L\left\{1 + \|M\|^{\ell+1}\right\} + \gamma C_\alpha \left(1 + \|M\|^\alpha\right)\right\}$$
$$+ \frac{\gamma}{2}\left\{2L\left\{1 + \|M\|^{\ell+1}\right\} + \gamma C_\alpha \left(1 + \|M\|^\alpha\right)\right\}^2 + d\Big] \; .$$

Under **H**1, **A**1 and by (5), we have for all $x \in \mathbb{R}^d$

$$\|G_\gamma(x)\| \leq 2L\left\{1 + \|x\|^{\ell+1}\right\} + \gamma C_\alpha \left(1 + \|x\|^\alpha\right) \; , \tag{26}$$

and $\max_{\|x\| \leq M} \|G_\gamma(x)\| \leq 2L\left\{1 + \|M\|^{\ell+1}\right\} + \gamma C_\alpha \left(1 + \|M\|^\alpha\right)$. Combining it with (23), (24), $s \mapsto s/(1+s^2)^{1/2}$ is non-decreasing for $s \geq 0$ and $(1+t_1+t_2)^{1/2} \leq (1+t_1)^{1/2} + t_2/2$ for $t_1 = \|x\|^2$, $t_2 = \gamma^2 \|G_\gamma(x)\|^2 + 2\gamma \|x\| \|G_\gamma(x)\| + 2\gamma d$, we have for all $x \in \mathbb{R}^d$, $\|x\| \leq M$,

$$R_\gamma V_{æ}(x) \leq \mathrm{e}^{\gamma c} V_{æ}(x) \; . \tag{27}$$

Then, using that for all $t \geq 0$, $1 - \mathrm{e}^{-t} \leq t$, we get for all $x \in \mathbb{R}^d$, $\|x\| \leq M$,

$$R_\gamma V_{æ}(x) - \mathrm{e}^{-æ^2\gamma} V_{æ}(x) \leq \mathrm{e}^{\gamma c}(1 - \mathrm{e}^{-\gamma(æ^2+c)}) V_{æ}(x) \leq \gamma \mathrm{e}^{\gamma c_E}(æ^2 + c) V_{æ}(x) \; , \tag{28}$$

which combined with (25) gives (11) with $b = \mathrm{e}^{\gamma c}(æ^2 + c)$. A straightforward induction gives for all $n \in \mathbb{N}$ and $x \in \mathbb{R}^d$,

$$R_\gamma^n V_{æ}(x) \leq \mathrm{e}^{-næ^2\gamma} V_{æ}(x) + \{(b\gamma)(1 - \mathrm{e}^{-næ^2\gamma})\}/(1 - \mathrm{e}^{-æ^2\gamma}) \; .$$

Using $1 - \mathrm{e}^{-æ^2\gamma} = \int_0^\gamma æ^2 \mathrm{e}^{-æ^2 t} \mathrm{d}t \geq \gamma æ^2 \mathrm{e}^{-æ^2\gamma}$, we get (12). Finally, using Jensen's inequality and $(s+t)^\varsigma \leq s^\varsigma + t^\varsigma$ for $\varsigma \in (0,1)$, $s, t \geq 0$ in (11), by [RT96, Section 3.1], for all $\gamma > 0$, R_γ has a unique invariant measure π_γ and R_γ is $V_{æ}^\varsigma$-geometrically ergodic w.r.t. π_γ.

4.4 Proof of Theorem 4

The proof is adapted from [DT12, Proposition 2] and [DM17, Theorem 10]. We first state a lemma.

Lemma 9. *Assume* **H**1, **H**2, **A** 1 *and* **A** 2*. Let* $\gamma_0 > 0$*,* $p \in \mathbb{N}^*$ *and* ν_0 *be a probability measure on* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$*. There exists* $C > 0$ *such that for all* $\gamma \in (0, \gamma_0]$

$$\mathrm{KL}(\nu_0 R_\gamma^p | \nu_0 P_{p\gamma}) \leq C\gamma^2 \int_{\mathbb{R}^d} \sum_{i=0}^{p-1} \left\{ \int_{\mathbb{R}^d} V_{\text{æ}}(z) R_\gamma^i(y, \mathrm{d}z) \right\} \nu_0(\mathrm{d}y) \ .$$

Proof. Let $y \in \mathbb{R}^d$ and $\gamma > 0$. Denote by μ_p^y and $\overline{\mu}_p^y$ the laws on $\mathcal{C}([0, p\gamma], \mathbb{R}^d)$ of the Langevin diffusion $(Y_t)_{t\in[0,p\gamma]}$ defined by (1) and of the linear interpolation $(\overline{Y}_t)_{t\in[0,p\gamma]}$ of $(X_k)_{k\in\mathbb{N}}$ defined by (3) both started at y. More precisely, denote by $(Y_t, \overline{Y}_t)_{t\geq 0}$ the unique strong solution of

$$\begin{cases} \mathrm{d}Y_t = -\nabla U(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \quad , \quad Y_0 = y \ , \\ \mathrm{d}\overline{Y}_t = -G_\gamma\left(\overline{Y}_{\lfloor t/\gamma \rfloor \gamma}\right)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \quad , \quad \overline{Y}_0 = y \ , \end{cases} \tag{29}$$

and by $(\mathcal{F}_t)_{t\geq 0}$ the filtration associated with $(B_t)_{t\geq 0}$. The following theorem states that μ_p^y and $\overline{\mu}_p^y$ are equivalent and gives the associated Rakon-Nikodym derivative.

Theorem 10 ([LS13, Theorem 7.19])**.** *Let* $(Y_t, \overline{Y}_t)_{t\in[0,p\gamma]}$ *be the unique strong solution of* (29)*. If*

$$\mathbb{P}\left(\int_0^{p\gamma} \|\nabla U(Y_t)\|^2 + \left\| G_\gamma\left(Y_{\lfloor t/\gamma \rfloor \gamma}\right) \right\|^2 \mathrm{d}t < +\infty \right) = 1 \ ,$$
$$\mathbb{P}\left(\int_0^{p\gamma} \left\|\nabla U(\overline{Y}_t)\right\|^2 + \left\| G_\gamma\left(\overline{Y}_{\lfloor t/\gamma \rfloor \gamma}\right) \right\|^2 \mathrm{d}t < +\infty \right) = 1 \ ,$$

then μ_p^y *and* $\overline{\mu}_p^y$ *are equivalent and* $\mathbb{P}$*-almost surely,*

$$\frac{\mathrm{d}\mu_p^y}{\mathrm{d}\overline{\mu}_p^y}((\overline{Y}_t)_{t\in[0,p\gamma]}) = \exp\left(\frac{1}{2}\int_0^{p\gamma} \left\langle -\nabla U(\overline{Y}_s) + G_\gamma\left(\overline{Y}_{\lfloor s/\gamma \rfloor \gamma}\right), \mathrm{d}\overline{Y}_s \right\rangle \right.$$
$$\left. - \frac{1}{4}\int_0^{p\gamma} \left\{ \left\|\nabla U(\overline{Y}_s)\right\|^2 - \left\| G_\gamma\left(\overline{Y}_{\lfloor s/\gamma \rfloor \gamma}\right) \right\|^2 \right\} \mathrm{d}s \right) \ .$$

By (5), (26) and Propositions 1 and 3, the assumptions of Theorem 10 are satisfied and we get

$$\begin{aligned} \mathrm{KL}(\overline{\mu}_p^y | \mu_p^y) &= \mathbb{E}\left[-\log\left\{ \frac{\mathrm{d}\mu_p^y}{\mathrm{d}\overline{\mu}_p^y}((\overline{Y}_t)_{t\in[0,p\gamma]}) \right\} \right] \\ &= (1/4)\int_0^{p\gamma} \mathbb{E}\left[\left\| \nabla U(\overline{Y}_s) - G_\gamma\left(\overline{Y}_{\lfloor s/\gamma \rfloor \gamma}\right) \right\|^2 \right] \mathrm{d}s \\ &= (1/4)\sum_{i=0}^{p-1} \int_{i\gamma}^{(i+1)\gamma} \mathbb{E}\left[\left\| \nabla U(\overline{Y}_s) - G_\gamma(\overline{Y}_{i\gamma}) \right\|^2 \right] \mathrm{d}s \ . \end{aligned}$$

For $i \in \{0, \ldots, p-1\}$ and $s \in [i\gamma, (i+1)\gamma)$, we have $\left\|\nabla U(\overline{Y}_s) - G_\gamma(\overline{Y}_{i\gamma})\right\|^2 \leq 2(A+B)$ where

$$A = \left\|\nabla U(\overline{Y}_s) - \nabla U(\overline{Y}_{i\gamma})\right\|^2 \ , \quad B = \left\|\nabla U(\overline{Y}_{i\gamma}) - G_\gamma(\overline{Y}_{i\gamma})\right\|^2 \ .$$

By **A**1, $B \leq \gamma^2 C_\alpha^2 \left(1 + \left\|\overline{Y}_{i\gamma}\right\|^\alpha\right)^2$ and by **H**1,

$$A \leq L^2 \left(1 + \left\|\overline{Y}_s\right\|^\ell + \left\|\overline{Y}_{i\gamma}\right\|^\ell\right)^2 \left\|\overline{Y}_s - \overline{Y}_{i\gamma}\right\|^2 \ .$$

On the other hand for $s \in [i\gamma, (i+1)\gamma)$,

$$\begin{aligned} \left\|\overline{Y}_s - \overline{Y}_{i\gamma}\right\|^2 &= (s - i\gamma)^2 \left\|G_\gamma(\overline{Y}_{i\gamma})\right\|^2 + 2\left\|B_s - B_{i\gamma}\right\|^2 \\ &\qquad - 2^{3/2}(s - i\gamma)\left\langle B_s - B_{i\gamma}, G_\gamma(\overline{Y}_{i\gamma})\right\rangle \ , \\ \left\|\overline{Y}_s\right\| &\leq \left\|\overline{Y}_{i\gamma}\right\| + \gamma\left\|G_\gamma(\overline{Y}_{i\gamma})\right\| + \sqrt{2}\left\|B_s - B_{i\gamma}\right\| \ . \end{aligned}$$

Define $\mathsf{P}_{\gamma,1}, \mathsf{P}_2 \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}_+, \mathbb{R}_+)$ for $t \in \mathbb{R}_+$ by

$$\begin{aligned} \mathsf{P}_{\gamma,1}(t) &= (2\pi)^{-d/2} L^2 \int_{\mathbb{R}^d} \left[2\left\|z\right\|^2 + \gamma\left\{2L(1 + t^{\ell+1}) + \gamma C_\alpha(1 + t^\alpha)\right\}^2\right] \\ &\times \left[1 + t^\ell + \left\{t + \gamma\left(2L(1 + t^{\ell+1}) + \gamma C_\alpha(1 + t^\alpha)\right) + \sqrt{2\gamma}\left\|z\right\|\right\}^\ell\right]^2 \mathrm{e}^{-\left\|z\right\|^2/2} \mathrm{d}z \ , \qquad (30) \\ \mathsf{P}_2(t) &= C_\alpha^2 \left(1 + t^\alpha\right)^2 \ . \qquad (31) \end{aligned}$$

By (26),

$$\int_{i\gamma}^{(i+1)\gamma} \mathbb{E}^{\mathcal{F}_{i\gamma}}\left[\left\|\nabla U(\overline{Y}_s) - G_\gamma(\overline{Y}_{i\gamma})\right\|^2\right] \mathrm{d}s \leq \gamma^2\left\{\mathsf{P}_{\gamma,1}(\left\|\overline{Y}_{i\gamma}\right\|) + 2\gamma \mathsf{P}_2(\left\|\overline{Y}_{i\gamma}\right\|)\right\} \ .$$

By [Kul97, Theorem 4.1, Chapter 2], we get

$$\mathrm{KL}(\delta_y R_\gamma^p | \delta_y P_{p\gamma}) \leq \mathrm{KL}(\overline{\mu}_p^y | \mu_p^y) \leq (\gamma^2/4) \sum_{i=0}^{p-1} \mathbb{E}_y\left[\mathsf{P}_{\gamma,1}(\left\|\overline{Y}_{i\gamma}\right\|) + 2\gamma \mathsf{P}_2(\left\|\overline{Y}_{i\gamma}\right\|)\right] \ .$$

By (30) and (31), there exists $C > 0$ such that for all $\gamma \in (0, \gamma_0]$ and $x \in \mathbb{R}^d$, $\mathsf{P}_{\gamma,1}(\|x\|) + 2\gamma \mathsf{P}_2(\|x\|) \leq 4CV_{æ}(x)$. Combining it with the chain rule for the Kullback-Leibler divergence concludes the proof. □

Proof of Theorem 4. Let $\gamma \in (0, \gamma_0]$. By Proposition 1, we have for all $n \in \mathbb{N}$ and $x \in \mathbb{R}^d$,

$$\left\|\delta_x R_\gamma^n - \pi\right\|_{V_{æ}^{1/2}} \leq C_{æ/2} \rho_{æ/2}^{n\gamma} V_{æ}^{1/2}(x) + \left\|\delta_x R_\gamma^n - \delta_x P_{n\gamma}\right\|_{V_{æ}^{1/2}} \ .$$

Denote by $k_\gamma = \lceil \gamma^{-1} \rceil$ and by q_γ, r_γ the quotient and the remainder of the Euclidian division of n by k_γ. We have $\|\delta_x R_\gamma^n - \delta_x P_{n\gamma}\|_{V_{æ}^{1/2}} \leq A + B$ where

$$\begin{aligned}
A &= \left\| \delta_x R_\gamma^{q_\gamma k_\gamma} P_{r_\gamma \gamma} - \delta_x R_\gamma^n \right\|_{V_{æ}^{1/2}} \\
B &= \sum_{i=1}^{q_\gamma} \left\| \delta_x R_\gamma^{(i-1)k_\gamma} P_{(n-(i-1)k_\gamma)\gamma} - \delta_x R_\gamma^{ik_\gamma} P_{(n-ik_\gamma)\gamma} \right\|_{V_{æ}^{1/2}} \\
&\leq \sum_{i=1}^{q_\gamma} C_{æ/2} \rho_{æ/2}^{(n-ik_\gamma)\gamma} \left\| \delta_x R_\gamma^{(i-1)k_\gamma} P_{k_\gamma \gamma} - \delta_x R_\gamma^{ik_\gamma} \right\|_{V_{æ}^{1/2}} .
\end{aligned}$$

For $i \in \{1, \ldots, q_\gamma\}$ we have by [DM17, Lemma 24],

$$\begin{aligned}
\left\| \delta_x R_\gamma^{(i-1)k_\gamma} P_{k_\gamma \gamma} - \delta_x R_\gamma^{ik_\gamma} \right\|_{V_{æ}^{1/2}}^2 \leq 2 \left\{ \delta_x R_\gamma^{(i-1)k_\gamma} P_{k_\gamma \gamma}(V_{æ}) + \delta_x R_\gamma^{ik_\gamma}(V_{æ}) \right\} \\
\times \mathrm{KL}(\delta_x R_\gamma^{ik_\gamma} | \delta_x R_\gamma^{(i-1)k_\gamma} P_{k_\gamma \gamma}) .
\end{aligned}$$

By Proposition 3 and lemma 9 and $k_\gamma \leq 1 + \gamma^{-1}$, we have for all $i \in \{1, \ldots, q_\gamma\}$

$$\begin{aligned}
\mathrm{KL}(\delta_x R_\gamma^{ik_\gamma} | \delta_x R_\gamma^{(i-1)k_\gamma} P_{k_\gamma \gamma}) &\leq C\gamma^2 \sum_{j=0}^{k_\gamma - 1} \int_{\mathbb{R}^d} V_{æ}(z) \delta_x R_\gamma^{(i-1)k_\gamma + j}(\mathrm{d}z) \\
&\leq C\gamma^2 (1 + \gamma^{-1}) \left\{ \mathrm{e}^{-æ^2 \gamma k_\gamma (i-1)} V_{æ}(x) + \frac{b}{æ^2} \mathrm{e}^{æ^2 \gamma} \right\} .
\end{aligned}$$

By Proposition 1, we have for $x \in \mathbb{R}^d$, $P_{k_\gamma \gamma} V_{æ}(x) \leq V_{æ}(x) + b_{æ}$ and by Proposition 3, we get for all $i \in \{1, \ldots, q_\gamma\}$

$$\delta_x R_\gamma^{(i-1)k_\gamma} P_{k_\gamma \gamma}(V_{æ}) + \delta_x R_\gamma^{ik_\gamma}(V_{æ}) \leq 2 \left\{ \mathrm{e}^{-æ^2 \gamma k_\gamma (i-1)} V_{æ}(x) + \frac{b}{æ^2} \mathrm{e}^{æ^2 \gamma} + b_{æ} \right\} . \tag{32}$$

We obtain

$$\begin{aligned}
B \leq 2C_{æ/2} C^{1/2} \gamma (1 + \gamma^{-1})^{1/2} \\
\times \sum_{i=0}^{q_\gamma - 1} \rho_{æ/2}^{(q_\gamma - 1 - i)\gamma k_\gamma} \left\{ \mathrm{e}^{-i\gamma k_\gamma æ^2/2} V_{æ}(x) + \left(b_{æ} + \frac{b}{æ^2} \mathrm{e}^{æ^2 \gamma} \right) \right\} .
\end{aligned}$$

Define $\kappa_m = \min(\rho_{æ/2}, \mathrm{e}^{-æ^2/2})$ and $\kappa_M = \max(\rho_{æ/2}, \mathrm{e}^{-æ^2/2})$. We have

$$\begin{aligned}
B \left\{ 2C_{æ/2} C^{1/2} \gamma (1 + \gamma^{-1})^{1/2} \right\}^{-1} \leq \left(b_{æ} + \frac{b}{æ^2} \mathrm{e}^{æ^2 \gamma} \right) \frac{1}{1 - \rho_{æ/2}^{k_\gamma \gamma}} \\
+ V_{æ}(x) \kappa_M^{(q_\gamma - 1)\gamma k_\gamma} \frac{1}{1 - (\kappa_m / \kappa_M)^{k_\gamma \gamma}} .
\end{aligned}$$

Bounding A along the same lines and using $k_\gamma \gamma \geq 1$, we get (13). By Proposition 3 and taking the limit $n \to +\infty$, we obtain (14). $\square$

4.5 Proofs of Theorems 5 and 6

We first state preliminary technical lemmas on the diffusion $(Y_t)_{t\geq 0}$. The proofs are postponed to the Appendix. Define for all $p \in \mathbb{N}^*$ and $k \in \{0, \cdots, p\}$,

$$a_{k,p} = m^{k-p} \prod_{i=k+1}^{p} \left\{ i(d+2(i-1))(i-k)^{-1} \right\} . \tag{33}$$

Lemma 11. *Assume* **H**3. *Let $p \in \mathbb{N}^*$, $x \in \mathbb{R}^d$ and $(Y_t)_{t\geq 0}$ be the solution of (1) started at x. For all $t \geq 0$,*

$$\mathbb{E}_x \left[\|Y_t\|^{2p} \right] \leq a_{0,p} \left(1 - \mathrm{e}^{-2pmt}\right) + \sum_{k=1}^{p} a_{k,p} \mathrm{e}^{-2kmt} \|x\|^{2k} ,$$

where for $k \in \{0, \cdots, p\}$, $a_{k,p}$ is given in (33).

Proof. The proof is postponed to Appendix A. □

Lemma 12. *Assume* **H**3 *and let $p \in \mathbb{N}^*$. We have $\int_{\mathbb{R}^d} \|y\|^{2p} \pi(\mathrm{d}y) \leq a_{0,p}$.*

Proof. Under **H**3, by [BGG12], (1) has a unique reversible measure π and $\lim_{t\to+\infty} W_{2p}(\delta_0 P_t, \pi) = 0$. [Vil09, Theorem 6.9] and Lemma 11 conclude the proof. □

Let $\gamma > 0$ and set $N = \lceil (\ell+1)/2 \rceil$ under **H**1. Define $\mathsf{P}_{\gamma,3} \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}_+, \mathbb{R}_+)$ for $s \in \mathbb{R}_+$ by

$$\mathsf{P}_{\gamma,3}(s) = 2d + 8L^2(1+s^{\ell+1}) \left\{ \frac{\gamma}{2} \left(2 + \sum_{k=1}^{N} a_{k,N} s^{2k} \right) + N m a_{0,N} \frac{\gamma^2}{3} \right\} . \tag{34}$$

Lemma 13. *Assume* **H**1 *and* **H**3. *Let $x \in \mathbb{R}^d$, $\gamma > 0$ and $(Y_t)_{t\geq 0}$ be the solution of (1) started at x. For all $t \in [0, \gamma]$, we have $\mathbb{E}_x \left[\|Y_t - x\|^2 \right] \leq t \mathsf{P}_{\gamma,3}(\|x\|)$, where $\mathsf{P}_{\gamma,3}$ is defined in (34).*

Proof. The proof is postponed to Appendix B. □

For $p \in \mathbb{N}$ and $\gamma > 0$, define $\mathsf{Q}_{\gamma,p} \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}_+, \mathbb{R}_+)$ for all $s \in \mathbb{R}_+$ by,

$$\begin{aligned} \mathsf{Q}_{\gamma,p}(s) = &\left\{ \prod_{i=1}^{p} 2i(d+3i-2) \right\} \left[2d \frac{\gamma^p}{(p+1)!} + 8L^2(1+s^{\ell+1}) \right. \\ &\left. \times \left\{ (2 + \sum_{k=1}^{N} a_{k,N} s^{2k}) \frac{\gamma^{p+1}}{(p+2)!} + 2N m a_{0,N} \frac{\gamma^{p+2}}{(p+3)!} \right\} \right] \\ &+ 2 \sum_{k=1}^{p} \left\{ \prod_{i=k+1}^{p} 2i(d+3i-2) \right\} \left\{ d + 4 + \frac{L^2(1+s^{\ell+1})^2}{m(k+1)} \right\} \\ &\times \left\{ \left(\sum_{i=1}^{k} a_{i,k} s^{2i} \right) \frac{\gamma^{p-k}}{(p+1-k)!} + 2km a_{0,k} \frac{\gamma^{p+1-k}}{(p+2-k)!} \right\} . \end{aligned} \tag{35}$$

Lemma 14. *Assume* **H**1 *and* **H**3*. Let $p \in \mathbb{N}$, $\gamma > 0$, $x \in \mathbb{R}^d$ and $(Y_t)_{t \geq 0}$ be the solution of* (1) *started at x. For all $t \in [0, \gamma]$, we have $\mathbb{E}_x\left[\|Y_t\|^{2p}\|Y_t - x\|^2\right] \leq t\mathsf{Q}_{\gamma,p}(\|x\|)$, where $\mathsf{Q}_{\gamma,p}$ is defined in* (35)*.*

Proof. The proof is postponed to Appendix C. □

Lemma 15. *Assume* **H**4*.*

a) For all $x \in \mathbb{R}^d$, $\|\nabla^2 U(x)\| \leq C_H\{1 + \|x\|^{\nu+\beta}\}$ where $C_H = \max(2L_H, \|\nabla^2 U(0)\|)$.

b) For all $x, y \in \mathbb{R}^d$,

$$\|\nabla U(x) - \nabla U(y) - \nabla^2 U(y)(x-y)\| \leq \frac{2L_H}{1+\beta}\{1 + \|x\|^\nu + \|y\|^\nu\}\|x-y\|^{1+\beta} .$$

Proof. a) By **H**4, we get for all $x \in \mathbb{R}^d$

$$\begin{aligned}\|\nabla^2 U(x)\| &\leq \|\nabla^2 U(x) - \nabla^2 U(0)\| + \|\nabla^2 U(0)\| \\ &\leq L_H\{1 + \|x\|^\nu\}\|x\|^\beta + \|\nabla^2 U(0)\| .\end{aligned}$$

The proof then follows from the upper bound for all $x \in \mathbb{R}^d$, $\|x\|^\beta \leq 1 + \|x\|^{\nu+\beta}$.

b) Let $x, y \in \mathbb{R}^d$. By **H**4,

$$\begin{aligned}&\|\nabla U(x) - \nabla U(y) - \nabla^2 U(y)(x-y)\| \\ &\quad\leq \int_0^1 \|\nabla^2 U(tx + (1-t)y) - \nabla^2 U(y)\| \mathrm{d}t\, \|x-y\| \\ &\quad\leq L_H \int_0^1 \{1 + \|y\|^\nu + \|tx + (1-t)y\|^\nu\}\|t(x-y)\|^\beta \mathrm{d}t\, \|x-y\| ,\end{aligned}$$

and the proof follows from $\|tx + (1-t)y\|^\nu \leq \|x\|^\nu + \|y\|^\nu$.

□

If U is twice continuously differentiable and there exist $\ell, L \geq 0$ such that for all $x \in \mathbb{R}^d$,

$$\|\nabla^2 U(x)\| \leq L\left\{1 + \|x\|^\ell\right\} , \tag{36}$$

then **H**1 is satisfied. Note that **H**4 implies **H**1 by Lemma 15-a) and (36) with $L = C_H$ and $\ell = \nu + \beta$.

For all $n \in \mathbb{N}$, we now bound the Wasserstein distance W_2 between π and the distribution of the n^{th} iterate of X_n defined by (3). The strategy consists given two initial conditions (x, y), in coupling X_n and $Y_{\gamma n}$ solution of (1) at time γn, using the same Brownian motion. Similarly to (29), for $\gamma > 0$, consider the unique strong solution $(Y_t, \overline{Y}_t)_{t \geq 0}$ of

$$\begin{cases} \mathrm{d}Y_t = -\nabla U(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \quad , \quad Y_0 = y , \\ \mathrm{d}\overline{Y}_t = -G_\gamma\left(\overline{Y}_{\lfloor t/\gamma \rfloor \gamma}\right)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \quad , \quad \overline{Y}_0 = x , \end{cases} \tag{37}$$

where $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion. Note that for $n \in \mathbb{N}$, $\overline{Y}_{n\gamma} = X_n$ and let $(\mathcal{F}_t)_{t\geq 0}$ be the filtration associated with $(B_t)_{t\geq 0}$.

Lemma 16. *Assume **A** 1, **A** 2, **H** 1 and **H** 3. Let $\gamma_0 > 0$. Define $(Y_t)_{t\geq 0}$, $(\overline{Y}_t)_{t\geq 0}$ by (37) and $X_n = \overline{Y}_{n\gamma}$ for $n \in \mathbb{N}$. Then there exists $C > 0$ such that for all $n \in \mathbb{N}$ and $\gamma \in (0, \gamma_0]$, almost surely,*

$$\mathbb{E}^{\mathcal{F}_{n\gamma}}\left[\left\|Y_{(n+1)\gamma} - X_{n+1}\right\|^2\right] \leq \mathrm{e}^{-m\gamma}\left\|Y_{n\gamma} - X_n\right\|^2 + C\gamma^2 V_{æ}(X_n) \; .$$

Proof. Without loss of generality, we show the result for $n = 0$. Define for $t \in [0, \gamma)$, $\Theta_t = Y_t - \overline{Y}_t$. By Itô's formula, we have for all $t \in [0, \gamma)$,

$$\left\|\Theta_t\right\|^2 = \left\|y - x\right\|^2 - 2\int_0^t \langle \Theta_s, \nabla U(Y_s) - G_\gamma(x)\rangle \, \mathrm{d}s \; .$$

By (5) and Lemma 11, the family of random variables $(\langle \Theta_s, \nabla U(Y_s) - G_\gamma(x)\rangle)_{s\in[0,\gamma)}$ is uniformly integrable. Pathwise continuity implies then for $s \in [0, \gamma)$ the continuity of $s \mapsto \mathbb{E}\left[\langle \Theta_s, \nabla U(Y_s) - G_\gamma(x)\rangle\right]$. Taking the expectation and deriving, we have for $t \in [0, \gamma)$,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[\left\|\Theta_t\right\|^2\right] &= -2\mathbb{E}\left[\langle \Theta_t, \nabla U(Y_t) - G_\gamma(x)\rangle\right] \\
&= -2\mathbb{E}\left[\left\langle \Theta_t, \nabla U(Y_t) - \nabla U(\overline{Y}_t)\right\rangle\right] - 2A_1 - 2A_2 \\
&\leq -2m\mathbb{E}\left[\left\|\Theta_t\right\|^2\right] - 2A_1 - 2A_2 \; ,
\end{aligned} \tag{38}$$

where

$$A_1 = \mathbb{E}\left[\left\langle \Theta_t, \nabla U(\overline{Y}_t) - \nabla U(x)\right\rangle\right] \; , \quad A_2 = \mathbb{E}\left[\langle \Theta_t, \nabla U(x) - G_\gamma(x)\rangle\right] \; . \tag{39}$$

Using that $|\langle a, b\rangle| \leq (m/4)\left\|a\right\|^2 + m^{-1}\left\|b\right\|^2$ for all $a, b \in \mathbb{R}^d$,

$$|A_1| \leq (m/4)\mathbb{E}\left[\left\|\Theta_t\right\|^2\right] + m^{-1}\mathbb{E}\left[\left\|\nabla U(\overline{Y}_t) - \nabla U(x)\right\|^2\right] \; .$$

Similarly to the proof of Lemma 9, we have $\mathbb{E}\left[\left\|\nabla U(\overline{Y}_t) - \nabla U(x)\right\|^2\right] \leq t\mathsf{P}_{\gamma,1}(\|x\|)$ where $\mathsf{P}_{\gamma,1}$ is defined in (30). For A_2, we have

$$|A_2| \leq (m/4)\mathbb{E}\left[\left\|\Theta_t\right\|^2\right] + m^{-1}\left\|\nabla U(x) - \nabla G_\gamma(x)\right\|^2 \tag{40}$$

and $\left\|\nabla U(x) - G_\gamma(x)\right\|^2 \leq \gamma^2\mathsf{P}_2(\|x\|)$ where P_2 is defined in (31). We get for $t \in [0, \gamma)$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[\left\|\Theta_t\right\|^2\right] \leq -m\mathbb{E}\left[\left\|\Theta_t\right\|^2\right] + 2m^{-1}\left\{t\mathsf{P}_{\gamma,1}(\|x\|) + \gamma^2\mathsf{P}_2(\|x\|)\right\} \; .$$

Using Grönwall's lemma and $1 - \mathrm{e}^{-s} \leq s$ for all $s \geq 0$, we obtain

$$\mathbb{E}\left[\left\|Y_\gamma - X_1\right\|^2\right] \leq \mathrm{e}^{-m\gamma}\left\|y - x\right\|^2 + m^{-1}\gamma^2\left\{\mathsf{P}_{\gamma,1}(\|x\|) + 2\gamma\mathsf{P}_2(\|x\|)\right\} \; .$$

Finally, by (30) and (31), there exists $C > 0$ such that for all $x \in \mathbb{R}^d$, $\mathsf{P}_{\gamma,1}(\|x\|) + 2\gamma\mathsf{P}_2(\|x\|) \leq CmV_{æ}(x)$. □

Lemma 17. *Assume* **A** *1,* **A** *2,* **H** *3 and* **H** *4. Let $\gamma_0 > 0$. Define $(Y_t)_{t\geq 0}$, $(\overline{Y}_t)_{t\geq 0}$ by (37) and $X_n = \overline{Y}_{n\gamma}$ for $n \in \mathbb{N}$. Then there exists $C > 0$ such that for all $n \in \mathbb{N}$ and $\gamma \in (0, \gamma_0]$, almost surely,*

$$\mathbb{E}^{\mathcal{F}_{n\gamma}}\left[\left\|Y_{(n+1)\gamma} - X_{n+1}\right\|^2\right] \leq \mathrm{e}^{-m\gamma}\left\|Y_{n\gamma} - X_n\right\|^2 + C\gamma^{2+\beta}V_{æ}(X_n) + C\gamma^3 V_{æ}(Y_{n\gamma}) .$$

Remark 18. *The calculations in the proof show that the dependence w.r.t. X_n and $Y_{n\gamma}$ is in fact polynomial but their exact expressions are very involved. For the sake of simplicity, we bound these polynomials by $V_{æ}$. The same remark applies equally to Lemma 16.*

Proof. Note first that **H**4 implies **H**1 by Lemma 15-a) and Equation (36) with $L = C_H$ and $\ell = \nu + \beta$. Without loss of generality, we show the result for $n = 0$. The proof is a refinement of Lemma 16 and we use the same notations. We have to improve the bound on A_1 defined in (39). We decompose $A_1 = A_{11} + A_{12}$ where

$$A_{11} = \mathbb{E}\left[\left\langle \Theta_t, \nabla U(\overline{Y}_t) - \nabla U(x) - \nabla^2 U(x)(\overline{Y}_t - x)\right\rangle\right] ,$$
$$A_{12} = \mathbb{E}\left[\left\langle \Theta_t, \nabla^2 U(x)(\overline{Y}_t - x)\right\rangle\right] .$$

Using $|\langle a, b\rangle| \leq (m/6)\|a\|^2 + \{3/(2m)\}\|b\|^2$ for all $a, b \in \mathbb{R}^d$,

$$|A_{11}| \leq \frac{m}{6}\mathbb{E}\left[\|\Theta_t\|^2\right] + \frac{3}{2m}\mathbb{E}\left[\left\|\nabla U(\overline{Y}_t) - \nabla U(x) - \nabla^2 U(x)(\overline{Y}_t - x)\right\|^2\right] . \quad (41)$$

By Lemma 15-b),

$$\left\|\nabla U(\overline{Y}_t) - \nabla U(x) - \nabla^2 U(x)(\overline{Y}_t - x)\right\|^2 \leq \frac{4L_H^2}{(1+\beta)^2}\left(1 + \|x\|^\nu + \left\|\overline{Y}_t\right\|^\nu\right)^2 \left\|\overline{Y}_t - x\right\|^{2(1+\beta)} .$$

Following the proof of Lemma 9, we have

$$\mathbb{E}\left[\left\|\nabla U(\overline{Y}_t) - \nabla U(x) - \nabla^2 U(x)(\overline{Y}_t - x)\right\|^2\right] \leq t^{1+\beta}\mathsf{P}_{\gamma,4}(\|x\|) . \quad (42)$$

where $\mathsf{P}_{\gamma,4} \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}_+, \mathbb{R}_+)$ is defined for all $s \in \mathbb{R}_+$ by,

$$\mathsf{P}_{\gamma,4}(s) = \frac{4L_H^2}{(1+\beta)^2}\int_{\mathbb{R}^d}\left[\sqrt{2}\|z\| + \sqrt{\gamma}\left\{2L(1+s^{\ell+1}) + \gamma C_\alpha(1+s^\alpha)\right\}\right]^{2(1+\beta)}$$
$$\times\left[1 + s^\nu + \left\{s + \gamma\left(2L(1+s^{\ell+1}) + \gamma C_\alpha(1+s^\alpha)\right) + \sqrt{2\gamma}\|z\|\right\}^\nu\right]^2 \frac{\mathrm{e}^{-\|z\|^2/2}}{(2\pi)^{d/2}}\mathrm{d}z . \quad (43)$$

We decompose A_{12} in $A_{12} = A_{121} + A_{122}$ where

$$A_{121} = \mathbb{E}\left[\left\langle \Theta_t, -t\nabla^2 U(x)G_\gamma(x)\right\rangle\right] , \; A_{122} = \sqrt{2}\mathbb{E}\left[\left\langle \Theta_t, \nabla^2 U(x)B_t\right\rangle\right] .$$

Define $\mathsf{P}_{\gamma,5} \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}_+, \mathbb{R}_+)$ for $s \in \mathbb{R}_+$ by,

$$\mathsf{P}_{\gamma,5}(s) = C_H^2 \left(1 + s^{\nu+\beta}\right)^2 \left\{2L(1 + s^{\ell+1}) + \gamma C_\alpha(1 + s^\alpha)\right\}^2 . \tag{44}$$

By Lemma 15-a) and (26),

$$|A_{121}| \le (m/6)\mathbb{E}\left[\|\Theta_t\|^2\right] + \{3/(2m)\} t^2 \mathsf{P}_{\gamma,5}(\|x\|) , \tag{45}$$

and by Cauchy-Schwarz inequality,

$$\begin{aligned} |A_{122}| &= \sqrt{2}\left|\mathbb{E}\left[\left\langle \int_0^t \{\nabla U(Y_s) - \nabla U(y)\}\, \mathrm{d}s, \nabla^2 U(x) B_t \right\rangle\right]\right| \\ &\le \sqrt{2}\mathbb{E}\left[\left\| \int_0^t \{\nabla U(Y_s) - \nabla U(y)\}\, \mathrm{d}s \right\|^2\right]^{1/2} \mathbb{E}\left[\left\|\nabla^2 U(x) B_t\right\|^2\right]^{1/2} . \end{aligned} \tag{46}$$

By Lemma 15-a), $\mathbb{E}\left[\left\|\nabla^2 U(x) B_t\right\|^2\right]^{1/2} \le \sqrt{dt} C_H \left(1 + \|x\|^{\nu+\beta}\right)$. By **H** 1, Cauchy-Schwarz inequality and using $\left(1 + \|y\|^\ell + \|Y_s\|^\ell\right)^2 \le 3\left(2 + \|y\|^{2\ell} + \|Y_s\|^{2\lceil \ell \rceil}\right)$ for $s \in [0, \gamma)$, we have

$$\begin{aligned} \mathbb{E}\left[\left\| \int_0^t \{\nabla U(Y_s) - \nabla U(y)\}\, \mathrm{d}s \right\|^2\right] &\le 3tL^2(2 + \|y\|^{2\ell}) \int_0^t \mathbb{E}\left[\|Y_s - y\|^2\right] \mathrm{d}s \\ &\quad + 3tL^2 \int_0^t \mathbb{E}\left[\|Y_s\|^{2\lceil \ell \rceil} \|Y_s - y\|^2\right] \mathrm{d}s . \end{aligned}$$

By Lemmas 13 and 14, we get

$$\mathbb{E}\left[\left\| \int_0^t \{\nabla U(Y_s) - \nabla U(y)\}\, \mathrm{d}s \right\|^2\right] \le \frac{3t^3 L^2}{2} \left\{\left(2 + \|y\|^{2\ell}\right) \mathsf{P}_{\gamma,3}(\|y\|) + \mathsf{Q}_{\gamma,\lceil \ell \rceil}(\|y\|)\right\} ,$$

where $\mathsf{P}_{\gamma,3}, \mathsf{Q}_{\gamma,\lceil \ell \rceil} \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}_+, \mathbb{R}_+)$ are defined in (34) and (35), and by (46)

$$|A_{122}| \le t^2 \sqrt{3d} C_H L \left(1 + \|x\|^{\nu+\beta}\right) \left\{\left(2 + \|y\|^{2\ell}\right) \mathsf{P}_{\gamma,3}(\|y\|) + \mathsf{Q}_{\gamma,\lceil \ell \rceil}(\|y\|)\right\}^{1/2} . \tag{47}$$

Combining (41), (42), (45) and (47), we get

$$\begin{aligned} |A_1| &\le (m/3)\mathbb{E}\left[\|\Theta_t\|^2\right] + \{3/(2m)\}\left\{t^{1+\beta} \mathsf{P}_{\gamma,4}(\|x\|) + t^2 \mathsf{P}_{\gamma,5}(\|x\|)\right\} \\ &\quad + t^2 \sqrt{3d} C_H L \left(1 + \|x\|^{\nu+\beta}\right) \left\{\left(2 + \|y\|^{2\ell}\right) \mathsf{P}_{\gamma,3}(\|y\|) + \mathsf{Q}_{\gamma,\lceil \ell \rceil}(\|y\|)\right\}^{1/2} , \end{aligned}$$

and by (40), $|A_2| \le (m/6)\mathbb{E}\left[\|\Theta_t\|^2\right] + \{3/(2m)\}\gamma^2 \mathsf{P}_2(\|x\|)$, where $\mathsf{P}_2 \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}_+, \mathbb{R}_+)$ is defined in (31). Combining these inequalities in (38), we get

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[\|\Theta_t\|^2\right] &\le -m\mathbb{E}\left[\|\Theta_t\|^2\right] + 3m^{-1}\left\{\gamma^2 \mathsf{P}_2(\|x\|) + t^{1+\beta} \mathsf{P}_{\gamma,4}(\|x\|) + t^2 \mathsf{P}_{\gamma,5}(\|x\|)\right\} \\ &\quad + 2t^2 \sqrt{3d} C_H L \left(1 + \|x\|^{\nu+\beta}\right) \left\{\left(2 + \|y\|^{2\ell}\right) \mathsf{P}_{\gamma,3}(\|y\|) + \mathsf{Q}_{\gamma,\lceil \ell \rceil}(\|y\|)\right\}^{1/2} . \end{aligned}$$

Using Grönwall's lemma and $1-\mathrm{e}^{-s} \leq s$ for all $s \geq 0$, we obtain

$$\mathbb{E}^{\mathcal{F}_{n\gamma}}\left[\left\|Y_{(n+1)\gamma}-X_{n+1}\right\|^2\right] \leq \mathrm{e}^{-m\gamma}\left\|Y_{n\gamma}-X_n\right\|^2 + 3m^{-1}\left\{\gamma^3\mathsf{P}_2(\|X_n\|)+\frac{\gamma^{2+\beta}}{2+\beta}\mathsf{P}_{\gamma,4}(\|X_n\|)+\frac{\gamma^3}{3}\mathsf{P}_{\gamma,5}(\|X_n\|)\right\}$$
$$+2\gamma^3\sqrt{d/3}C_H L\left(1+\|X_n\|^{\nu+\beta}\right)\left\{\left(2+\|Y_{n\gamma}\|^{2\ell}\right)\mathsf{P}_{\gamma,3}(\|Y_{n\gamma}\|)+\mathsf{Q}_{\gamma,\lceil\ell\rceil}(\|Y_{n\gamma}\|)\right\}^{1/2} .$$

Finally, by (31), (34), (43), (44) and (35), there exists $C>0$ such that for all $x \in \mathbb{R}^d$ and $\gamma \in (0,\gamma_0]$,

$$3m^{-1}\left\{\gamma^3\mathsf{P}_2(\|x\|)+\frac{\gamma^{2+\beta}}{2+\beta}\mathsf{P}_{\gamma,4}(\|x\|)+\frac{\gamma^3}{3}\mathsf{P}_{\gamma,5}(\|x\|)\right\} \leq C\gamma^{2+\beta}V_{æ}(x) ,$$
$$2\sqrt{d/3}C_H L\left(1+\|x\|^{\nu+\beta}\right) \leq C^{1/2}V_{æ}(x)^{1/2} ,$$
$$\left(2+\|x\|^{2\ell}\right)\mathsf{P}_{\gamma,3}(\|x\|)+\mathsf{Q}_{\gamma,\lceil\ell\rceil}(\|x\|) \leq CV_{æ}(x) .$$

□

Proof of Theorem 5. Let $\gamma \in (0,\gamma_0]$. Define $(Y_t)_{t\geq 0}$, $(\overline{Y}_t)_{t\geq 0}$ by (37) and $X_n = \overline{Y}_{n\gamma}$ for $n \in \mathbb{N}$. By Lemma 16 and Proposition 3, we have for all $n \in \mathbb{N}$,

$$\mathbb{E}\left[\|Y_{n\gamma}-X_n\|^2\right] \leq \mathrm{e}^{-nm\gamma}\|y-x\|^2 + C\gamma^2\sum_{k=0}^{n-1}\mathrm{e}^{-m\gamma(n-1-k)}\mathbb{E}\left[V_{æ}(X_k)\right]$$
$$\leq \mathrm{e}^{-nm\gamma}\|y-x\|^2+\frac{C\gamma^2}{1-\mathrm{e}^{-m\gamma}}\frac{b}{æ^2}\mathrm{e}^{æ^2\gamma}+C\gamma^2V_{æ}(x)\sum_{k=0}^{n-1}\mathrm{e}^{-m\gamma(n-1-k)}\mathrm{e}^{-æ^2\gamma k} . \tag{48}$$

Define $\kappa_m = \min(\mathrm{e}^{-m},\mathrm{e}^{-æ^2})$ and $\kappa_M = \max(\mathrm{e}^{-m},\mathrm{e}^{-æ^2})$. We have

$$\sum_{k=0}^{n-1}\mathrm{e}^{-m\gamma(n-1-k)}\mathrm{e}^{-æ^2\gamma k} \leq \kappa_M^{(n-1)\gamma}\frac{1}{1-(\kappa_m/\kappa_M)^\gamma} ,$$

and,

$$1-\mathrm{e}^{-m\gamma} \geq m\gamma\mathrm{e}^{-m\gamma} \quad , \quad 1-(\kappa_m/\kappa_M)^\gamma \geq \gamma\log(\kappa_M/\kappa_m)\mathrm{e}^{\gamma\log(\kappa_m/\kappa_M)} .$$

In eq. (48), integrating y with respect to π, for all $n \in \mathbb{N}$, $(Y_{n\gamma},X_n)$ is a coupling between π and $\delta_x R_\gamma^n$. By Lemma 12, we get (17). By Proposition 3 and [Vil09, Corollary 6.11], we have for all $x \in \mathbb{R}^d$, $\lim_{n\to+\infty} W_2(\delta_x R_\gamma^n,\pi) = W_2(\pi_\gamma,\pi)$ and we obtain (18). □

Proof of Theorem 6. Note that **H**4 implies **H**1 by Lemma 15-a) and Equation (36) with $L = C_H$ and $\ell = \nu+\beta$. Let $\gamma \in (0,\gamma_0]$. Define $(Y_t)_{t\geq 0}$, $(\overline{Y}_t)_{t\geq 0}$ by (37) and $X_n = \overline{Y}_{n\gamma}$ for $n \in \mathbb{N}$. By Lemma 17, we have for all $n \in \mathbb{N}$,

$$\mathbb{E}\left[\|Y_{n\gamma}-X_n\|^2\right] \leq \mathrm{e}^{-nm\gamma}\|y-x\|^2 + A_n + B_n ,$$

where

$$A_n = C\gamma^{2+\beta} \sum_{k=0}^{n-1} \mathrm{e}^{-m\gamma(n-1-k)} \mathbb{E}\left[V_{æ}(X_k)\right] \ ,$$

$$B_n = C\gamma^{3} \sum_{k=0}^{n-1} \mathrm{e}^{-m\gamma(n-1-k)} \mathbb{E}\left[V_{æ}(Y_{k\gamma})\right] \ .$$

Analysis similar to the proof of Theorem 5 using Proposition 1 instead of Proposition 3 for B_n shows then the result. □

4.6 Proof of Proposition 7

The proof is adapted from [PV01, Theorem 1] and follows the same steps. Define $\bar{f} = f - \pi(f)$. Note first that **H** 5 implies **H** 1. By **H** 2, (1) admits a unique strong solution $(Y_t^x)_{t\geq 0}$ for any initial condition $Y_0 = x \in \mathbb{R}^d$. By [Le 16, Theorem 8.5] valid under a local Lipschitz condition on ∇U, $(x,t) \mapsto Y_t^x$ is almost surely continuous for all $x \in \mathbb{R}^d$ and $t \geq 0$. For all compact set $\mathsf{K} \subset \mathbb{R}^d$, by Proposition 1, the families $(\bar{f}(Y_t^x))_{t\geq 0}$ for $x \in \mathbb{R}^d$ and $(\bar{f}(Y_t^x))_{x\in\mathsf{K}}$ for $t \geq 0$ are uniformly integrable. Therefore, $t \mapsto P_t\bar{f}(x)$ is continuous for every $x \in \mathbb{R}^d$ and $x \mapsto P_t\bar{f}(x)$ is continuous for all $t \geq 0$. By [PV01, Proof of Theorem 1, step (a)], there exist $C > 0$ and $p \in \mathbb{N}$ such that for all $x \in \mathbb{R}^d$,

$$\int_0^{+\infty} \left|P_t\bar{f}(x)\right| \mathrm{d}t \leq C\left(1 + \|x\|^p\right) \ .$$

We can then define $\phi : \mathbb{R}^d \to \mathbb{R}$ for all $x \in \mathbb{R}^d$ by

$$\phi(x) = \int_0^{+\infty} P_t\bar{f}(x)\mathrm{d}t \ , \quad \text{and} \quad \phi \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R}) \ .$$

We have $\lim_{N\to+\infty} \int_0^N P_t\bar{f}(x)\mathrm{d}t = \phi(x)$ locally uniformly in x and ϕ is hence continuous.

Let $x \in \mathbb{R}^d$ and consider the Dirichlet problem,

$$\mathscr{A}\hat{\phi}(y) = -\bar{f}(y) \quad \text{for} \quad y \in \mathrm{B}(x,1) \quad \text{and} \quad \hat{\phi}(y) = \phi(y) \quad \text{for} \quad y \in \partial\,\mathrm{B}(x,1) \ ,$$

where $\partial\,\mathrm{B}(x,1) = \overline{\mathrm{B}}(x,1) \setminus \mathrm{B}(x,1)$. By [GT15, Lemma 6.10, Theorem 6.17], there exists a solution $\hat{\phi} \in \mathrm{C}^4(\mathrm{B}(x,1),\mathbb{R}) \cap \mathrm{C}(\overline{\mathrm{B}}(x,1),\mathbb{R})$. Define the stopping time $\tau = \inf\{t \geq 0 : Y_t^x \notin \mathrm{B}(x,1)\}$ and for $n \in \mathbb{N}$, $n \geq 2$, $\tau_n = \inf\{t \geq 0 : Y_t^x \notin \mathrm{B}(x, 1-1/n)\}$. By [Fri12, Volume I, Chapter 6, equation (5.11)], $\mathbb{E}[\tau] < +\infty$. By Itô's formula, we have for all $n \geq 2$,

$$\hat{\phi}(x) = \mathbb{E}\left[\hat{\phi}(Y_{\tau_n}^x)\right] + \mathbb{E}\left[\int_0^{\tau_n} \bar{f}(Y_t^x)\mathrm{d}t\right] \ .$$

Using the continuity of $\hat{\phi}$ on $\overline{\mathrm{B}}(x,1)$ and $\lim_{n\to+\infty} \tau_n = \tau$ almost surely, we get

$$\hat{\phi}(x) = \mathbb{E}\left[\phi(Y_\tau^x)\right] + \mathbb{E}\left[\int_0^{\tau} \bar{f}(Y_t^x)\mathrm{d}t\right] \ .$$

By [KS91, Chapter 5, Theorem 4.20], $(Y_t^x)_{t\geq 0}$ is a strong Markov process. By [PV01, Proof of Theorem 1, step (d)], we obtain $\phi(x) = \hat{\phi}(x)$. Using [GT15, Problem 6.1 (a)], we get $\left\|\mathrm{D}^i\,\phi(x)\right\| \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R}_+)$ for $i \in \{0, \ldots, 4\}$.

4.7 Proof of Theorem 8

The proof is adapted from [MST10, Section 5.1] Let $\gamma \in (0, \gamma_0]$. In this Section, C is a positive constant which can change from line to line but does not depend on γ. For $k \in \mathbb{N}$, denote by

$$\delta_{k+1} = X_{k+1} - X_k = -\gamma G_\gamma(X_k) + \sqrt{2\gamma} Z_{k+1} \; .$$

By Proposition 7, there exists $\phi \in \mathrm{C}^4(\mathbb{R}^d, \mathbb{R})$, such that for all $x \in \mathbb{R}^d$ and $i \in \{0, \ldots, 4\}$,

$$\mathscr{A} \phi(x) = -\left(f(x) - \pi(f)\right) \quad \text{and} \quad \left\|\mathrm{D}^i \phi\right\| \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R}_+) \; .$$

By Taylor's formula, we have for $k \in \mathbb{N}$,

$$\begin{aligned} \phi(X_{k+1}) = \phi(X_k) &+ \mathrm{D}\, \phi(X_k)[\delta_{k+1}] + (1/2)\, \mathrm{D}^2 \phi(X_k)[\delta_{k+1}, \delta_{k+1}] \\ &+ (1/6)\, \mathrm{D}^3 \phi(X_k)[\delta_{k+1}, \delta_{k+1}, \delta_{k+1}] + r_k \; , \\ r_k = (1/6) &\int_0^1 (1-s)^3\, \mathrm{D}^4 \phi(X_k + s\delta_{k+1})[\delta_{k+1}, \delta_{k+1}, \delta_{k+1}, \delta_{k+1}] \mathrm{d}s \; . \end{aligned}$$

Using the expression of δ_{k+1} and (6), we get

$$\begin{aligned} \phi(X_{k+1}) &= \phi(X_k) + \gamma \mathscr{A} \phi(X_k) + \sqrt{2\gamma}\, \mathrm{D}\, \phi(X_k)[Z_{k+1}] \\ &+ \gamma \left\{\mathrm{D}^2 \phi(X_k)[Z_{k+1}, Z_{k+1}] - \Delta\phi(X_k)\right\} + \gamma\, \mathrm{D}\, \phi(X_k)[\nabla U(X_k) - G_\gamma(X_k)] \\ &+ (\gamma^2/2)\, \mathrm{D}^2 \phi(X_k)[G_\gamma(X_k), G_\gamma(X_k)] - \sqrt{2}\gamma^{3/2}\, \mathrm{D}^2 \phi(X_k)[G_\gamma(X_k), Z_{k+1}] \\ &+ (1/6)\, \mathrm{D}^3 \phi(X_k)[\delta_{k+1}, \delta_{k+1}, \delta_{k+1}] + r_k \; . \end{aligned}$$

Summing from $k = 0$ to $n-1$ for $n \in \mathbb{N}^\star$, dividing by $n\gamma$, we get

$$\frac{1}{n} \sum_{k=0}^{n-1} (f(X_k) - \pi(f)) = \frac{\phi(X_0) - \phi(X_n)}{n\gamma} + \frac{1}{n\gamma} \left(\sum_{i=0}^{3} M_{i,n} + \sum_{i=0}^{3} S_{i,n} \right) \; ,$$

where

$$\begin{aligned} M_{0,n} &= ((\sqrt{2}\gamma^{3/2})/6) \sum_{k=0}^{n-1} \{2\, \mathrm{D}^3 \phi(X_k)[Z_{k+1}, Z_{k+1}, Z_{k+1}] \\ &\qquad\qquad + 3\gamma\, \mathrm{D}^3 \phi(X_k)[G_\gamma(X_k), G_\gamma(X_k), Z_{k+1}]\} \; , \\ M_{1,n} &= \gamma \sum_{k=0}^{n-1} (\mathrm{D}^2 \phi(X_k)[Z_{k+1}, Z_{k+1}] - \Delta\phi(X_k)) \; , \\ M_{2,n} &= \sqrt{2\gamma} \sum_{k=0}^{n-1} \mathrm{D}\, \phi(X_k)[Z_{k+1}] \; , \\ M_{3,n} &= -\sqrt{2}\gamma^{3/2} \sum_{k=0}^{n-1} \mathrm{D}^2 \phi(X_k)[G_\gamma(X_k), Z_{k+1}] \; , \end{aligned}$$

and

$$S_{0,n} = -(\gamma^2/6)\sum_{k=0}^{n-1}\{6\,\mathrm{D}^3\,\phi(X_k)[G_\gamma(X_k), Z_{k+1}, Z_{k+1}] \\ + \gamma\,\mathrm{D}^3\,\phi(X_k)[G_\gamma(X_k), G_\gamma(X_k), G_\gamma(X_k)]\}\ ,$$

$$S_{1,n} = \gamma\sum_{k=0}^{n-1}\mathrm{D}\,\phi(X_k)[\nabla U(X_k) - G_\gamma(X_k)]\ ,$$

$$S_{2,n} = (\gamma^2/2)\sum_{k=0}^{n-1}\mathrm{D}^2\,\phi(X_k)[G_\gamma(X_k), G_\gamma(X_k)]\ ,$$

$$S_{3,n} = \sum_{k=0}^{n-1} r_k\ .$$

By **A** 1, we calculate for $n \in \mathbb{N}^*$, $|S_{1,n}| \le \gamma^2 C_\alpha \sum_{k=0}^{n-1}\|\mathrm{D}\,\phi(X_k)\|\,(1 + \|X_k\|^\alpha)$. By **H** 5, (26) and Proposition 7, there exist $p, q \ge 1$ and $C_q > 0$ such that the summands of $(M_{i,n})_{n\in\mathbb{N}}$ and $(S_{i,n})_{n\in\mathbb{N}}$ for $i \in \{0,\ldots,3\}$ are dominated by $C_q\,(1 + \|X_k\|^q)\,(1+\|Z_{k+1}\|^p)$ for $k \in \{0,\ldots,n-1\}$. Therefore, by Proposition 1, for $i \in \{0,\ldots,3\}$, $(M_{i,n})_{n\in\mathbb{N}}$ are martingales and for $n \in \mathbb{N}^*$, $\mathbb{E}\left[S_{i,n}^2\right] \le Cn^2\gamma^4$,

$$\mathbb{E}\left[M_{0,n}^2\right] \le Cn\gamma^3\ ,\ \mathbb{E}\left[M_{1,n}^2\right] \le Cn\gamma^2\ ,\ \mathbb{E}\left[M_{2,n}^2\right] \le Cn\gamma\ ,\ \mathbb{E}\left[M_{3,n}^2\right] \le Cn\gamma^3\ ,$$

which yield the result.

Acknowledgements

This work was supported by the École Polytechnique Data Science Initiative and the Alan Turing Institute under the EPSRC grant EP/N510129/1.

A Proof of Lemma 11

By **H**3, (1) has a unique strong solution $(Y_t)_{t\ge0}$ for any initial data $Y_0 = x \in \mathbb{R}^d$. Define for $p \in \mathbb{N}^*$, $V_p : \mathbb{R}^d \to \mathbb{R}_+$ by $V_p(y) = \|y\|^{2p}$ for $y \in \mathbb{R}^d$. We have using **H**3,

$$\begin{aligned}\mathscr{A} V_p(x) &= -2p\,\|x\|^{2(p-1)}\,\langle\nabla U(x), x\rangle + 2p(d + 2(p-1))\,\|x\|^{2(p-1)} \\ &\le -2pm\,\|x\|^{2p} + 2p\,\|x\|^{2(p-1)}\,(d + 2(p-1))\ .\end{aligned}$$

Applying [MT93, Theorem 1.1] with $V(x,t) = V_p(x)\mathrm{e}^{2pmt}$, $g_-(t) = 0$ and $g_+(x,t) = 2p(d+2(p-1))V_{p-1}(x)\mathrm{e}^{2pmt}$ for $x \in \mathbb{R}^d$ and $t \ge 0$, we get denoting by $v_p(t,x) = P_tV_p(x)$,

$$v_p(t,x) \le \mathrm{e}^{-2pmt}V_p(x) + 2p(d + 2(p-1))\int_0^t \mathrm{e}^{-2pm(t-s)}v_{p-1}(s,x)\mathrm{d}s\ .$$

A straightforward induction concludes the proof.

B Proof of Lemma 13

Define $\tilde{V}_x : \mathbb{R}^d \to \mathbb{R}_+$ for all $y \in \mathbb{R}^d$ by $\tilde{V}_x(y) = \|y-x\|^2$. By Lemma 11, the process $(\tilde{V}_x(Y_t) - \tilde{V}_x(x) - \int_0^t \mathscr{A}\tilde{V}_x(Y_s)\mathrm{d}s)_{t \geq 0}$, is a $(\mathcal{F}_t)_{t\geq 0}$-martingale. Denote for all $t \geq 0$ and $y \in \mathbb{R}^d$ by $\tilde{v}(t,x) = P_t\tilde{V}_x(x)$. Then we get,

$$\frac{\partial \tilde{v}(t,x)}{\partial t} = P_t \mathscr{A}\tilde{V}_x(x) \; . \tag{49}$$

By **H**3, we have for all $y \in \mathbb{R}^d$,

$$\mathscr{A}\tilde{V}_x(y) = 2\left(-\langle \nabla U(y), y-x\rangle + d\right) \leq 2\left(-m\tilde{V}_x(y) + d - \langle \nabla U(x), y-x\rangle\right) \; . \tag{50}$$

Using (49), this inequality and that $\tilde{V}_x$ is nonnegative, we get

$$\frac{\partial \tilde{v}(t,x)}{\partial t} = P_t \mathscr{A}\tilde{V}_x(x) \leq 2\left(d - \int_{\mathbb{R}^d} \langle \nabla U(x), y-x\rangle \, P_t(x, \mathrm{d}y)\right) \; . \tag{51}$$

Using (5) and (1), we have

$$\begin{aligned} \left|\mathbb{E}_x\left[\langle \nabla U(x), Y_t - x\rangle\right]\right| &\leq \|\nabla U(x)\| \, \|\mathbb{E}_x\left[Y_t - x\right]\| \\ &\leq \|\nabla U(x)\| \left\|\mathbb{E}_x\left[\int_0^t \{\nabla U(Y_s)\}\,\mathrm{d}s\right]\right\| \\ &\leq 2L\left\{1 + \|x\|^{\ell+1}\right\} \int_0^t \mathbb{E}_x\left[\|\nabla U(Y_s)\|\right]\mathrm{d}s \; . \end{aligned} \tag{52}$$

Using (5) again,

$$\begin{aligned} \int_0^t \mathbb{E}_x\left[\|\nabla U(Y_s)\|\right]\mathrm{d}s &\leq 2L \int_0^t \mathbb{E}\left[1 + \|Y_s\|^{\ell+1}\right]\mathrm{d}s \\ &\leq 2L\left\{2t + \int_0^t \mathbb{E}\left[\|Y_s\|^{2N}\right]\mathrm{d}s\right\} \; . \end{aligned} \tag{53}$$

Furthermore using that for all $s \geq 0$, $1 - \mathrm{e}^{-s} \leq s$, $s + \mathrm{e}^{-s} - 1 \leq s^2/2$, and Lemma 11 we get

$$\begin{aligned} \int_0^t \mathbb{E}_x\left[\|Y_s\|^{2N}\right]\mathrm{d}s &\leq a_{0,N}\frac{2Ntm + \mathrm{e}^{-2Nmt} - 1}{2Nm} + \sum_{k=1}^N a_{k,N}\|x\|^{2k}\frac{1 - \mathrm{e}^{-2mkt}}{2km} \\ &\leq t^2 N m a_{0,N} + t\sum_{k=1}^N a_{k,N}\|x\|^{2k} \; . \end{aligned}$$

Plugging this inequality in (53) and (52), we get

$$\left|\mathbb{E}_x\left[\langle \nabla U(x), Y_t - x\rangle\right]\right| \leq 4L^2(1 + \|x\|^{\ell+1})\left\{2t + Nma_{0,N}t^2 + t\sum_{k=1}^N a_{k,N}\|x\|^{2k}\right\} \; . \tag{54}$$

Using this bound in (51) and integrating the inequality gives

$$\tilde{v}(t,x) \leq 2dt + 8L^2(1 + \|x\|^{\ell+1})\left\{t^2 + Nma_{0,N}(t^3/3) + (t^2/2)\sum_{k=1}^N a_{k,N}\|x\|^{2k}\right\} \; . \tag{55}$$

C Proof of Lemma 14

We show the result by induction on p. The case $p = 0$ follows from (55). Suppose $p \geq 1$. Define for $y \in \mathbb{R}^d$, $W_{x,p} : \mathbb{R}^d \to \mathbb{R}_+$ by $W_{x,p}(y) = \|y\|^{2p} \|y - x\|^2$. We have

$$\begin{aligned}\mathscr{A} W_{x,p}(y) = -2\|y\|^{2p} \langle \nabla U(y), y - x\rangle - (2p)\|y\|^{2(p-1)} \|y - x\|^2 \langle \nabla U(y), y\rangle \\ + 2\|y\|^{2(p-1)} \left\{ d\|y\|^2 + 4p\langle y, y - x\rangle + p(d + 2p - 2)\|y - x\|^2 \right\} .\end{aligned}$$

By **H**3, (5) and using $|\langle a, b\rangle| \leq \eta \|a\|^2 + (4\eta)^{-1} \|b\|^2$ for all $\eta > 0$, we have

$$\begin{aligned}\mathscr{A} W_{x,p}(y) &\leq \frac{\|y\|^{2p} \|\nabla U(x)\|^2}{2m(p+1)} + 2\|y\|^{2(p-1)} \left\{ (d+4)\|y\|^2 + p(d + 3p - 2)\|y - x\|^2 \right\} \\ &\leq \|y\|^{2p} \left\{ 2(d+4) + \frac{2L^2(1 + \|x\|^{\ell+1})^2}{m(p+1)} \right\} + 2p(d + 3p - 2)\|y - x\|^2 \|y\|^{2(p-1)} . \qquad (56)\end{aligned}$$

By Lemma 11, the process $(W_{x,p}(Y_t) - W_{x,p}(x) - \int_0^t \mathscr{A} W_{x,p}(Y_s)\mathrm{d}s)_{t \geq 0}$ is a $(\mathcal{F}_t)_{t\geq 0}$-martingale. For $x \in \mathbb{R}^d$ and $t \geq 0$, denote by $w_{x,p}(x, t) = P_t W_{x,p}(x)$ and $v_p(x, t) = \mathbb{E}_x \left[\|Y_t\|^{2p}\right]$. Taking the expectation of (56) w.r.t. $\delta_x P_t$ and integrating w.r.t. t, we get

$$\begin{aligned}w_{x,p}(t, x) \leq 2\left\{ d + 4 + \frac{L^2(1 + \|x\|^{\ell+1})^2}{m(p+1)} \right\} \int_0^t v_p(s, x)\mathrm{d}s \\ + 2p(d + 3p - 2)\int_0^t w_{x,p-1}(s, x)\mathrm{d}s .\end{aligned}$$

By Lemma 11, $v_p(t, x) \leq 2pma_{0,p}t + \sum_{k=1}^p a_{k,p} \|x\|^{2k}$. A straightforward induction concludes the proof.

References

[And+03] C. Andrieu et al. “An introduction to MCMC for machine learning”. In: *Machine learning* 50.1-2 (2003), pp. 5–43.

[Atc06] Yves F. Atchadé. “An Adaptive Version for the Metropolis Adjusted Langevin Algorithm with a Truncated Drift”. In: *Methodology and Computing in Applied Probability* 8.2 (June 2006), pp. 235–254.

[BGG12] F. Bolley, I. Gentil, and A. Guillin. “Convergence to equilibrium in Wasserstein distance for Fokker-Planck equations”. In: *J. Funct. Anal.* 263.8 (2012), pp. 2430–2457.

[BGL14] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*. Vol. 348. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, Cham, 2014, pp. xx+552.

[BH13] N. Bou-Rabee and M. Hairer. "Nonasymptotic mixing of the MALA algorithm". In: *IMA Journal of Numerical Analysis* 33.1 (2013), pp. 80–110. eprint: /oup/backfile/content_public/journal/imajna/33/1/10.1093/imanum/drs003/2/drs003.pdf.

[BV10] Nawaf Bou-Rabee and Eric Vanden-Eijnden. "Pathwise accuracy and ergodicity of metropolized integrators for SDEs". In: *Communications on Pure and Applied Mathematics* 63.5 (2010), pp. 655–696.

[CCG12] Patrick Cattiaux, Djalil Chafaı, and Arnaud Guillin. "Central limit theorems for additive functionals of ergodic Markov diffusions processes". In: *ALEA* 9.2 (2012), pp. 337–382.

[Cot+13] S. L. Cotter et al. "MCMC methods for functions: modifying old algorithms to make them faster". In: *Statist. Sci.* 28.3 (2013), pp. 424–446.

[Dal17] Arnak S. Dalalyan. "Theoretical guarantees for approximate sampling from smooth and log-concave densities". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79.3 (2017), pp. 651–676.

[DM16] A. Durmus and E. Moulines. "High-dimensional Bayesian inference via the Unadjusted Langevin Algorithm". In: *ArXiv e-prints* (May 2016). arXiv: 1605.01559 [math.ST].

[DM17] Alain Durmus and Éric Moulines. "Nonasymptotic convergence analysis for the unadjusted Langevin algorithm". In: *Ann. Appl. Probab.* 27.3 (June 2017), pp. 1551–1587.

[DT12] A. S. Dalalyan and A. B. Tsybakov. "Sparse regression learning by aggregation and Langevin Monte-Carlo". In: *J. Comput. System Sci.* 78.5 (2012), pp. 1423–1443.

[Ebe15] A. Eberle. "Reflection couplings and contraction rates for diffusions". English. In: *Probab. Theory Related Fields* (2015), pp. 1–36.

[Fri12] Avner Friedman. *Stochastic differential equations and applications.* Courier Corporation, 2012.

[GM94] U. Grenander and M. I. Miller. "Representations of knowledge in complex systems". In: *J. Roy. Statist. Soc. Ser. B* 56.4 (1994). With discussion and a reply by the authors, pp. 549–603.

[GM96] Peter W. Glynn and Sean P. Meyn. "A Liapounov bound for solutions of the Poisson equation". In: *Ann. Probab.* 24.2 (Apr. 1996), pp. 916–931.

[Gor+16] J. Gorham et al. "Measuring Sample Quality with Diffusions". In: *ArXiv e-prints* (Nov. 2016). arXiv: 1611.06972 [stat.ML].

[Gre83] U. Grenander. "Tutorial in pattern theory". Division of Applied Mathematics, Brown University, Providence. 1983.

[GT15] David Gilbarg and Neil S Trudinger. *Elliptic partial differential equations of second order.* springer, 2015.

[HJ15] Martin Hutzenthaler and Arnulf Jentzen. *Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients*. Vol. 236. 1112. American Mathematical Society, 2015.

[HJK11] Martin Hutzenthaler, Arnulf Jentzen, and Peter E. Kloeden. "Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 467.2130 (2011), pp. 1563–1576. eprint: http://rspa.royalsocietypublishing.org/content/467/2130/1563.full.pdf.

[HJK12] Martin Hutzenthaler, Arnulf Jentzen, and Peter E. Kloeden. "Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients". In: *Ann. Appl. Probab.* 22.4 (Aug. 2012), pp. 1611–1641.

[HMS02] Desmond J. Higham, Xuerong Mao, and Andrew M. Stuart. "Strong Convergence of Euler-Type Methods for Nonlinear Stochastic Differential Equations". In: *SIAM Journal on Numerical Analysis* 40.3 (2002), pp. 1041–1063. eprint: https://doi.org/10.1137/S0036142901389530.

[IW89] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North-Holland Mathematical Library. Elsevier Science, 1989.

[JH00] Søren Fiig Jarner and Ernst Hansen. "Geometric ergodicity of Metropolis algorithms". In: *Stochastic Processes and their Applications* 85.2 (2000), pp. 341–361.

[Kop15] Marie Kopec. "Weak backward error analysis for overdamped Langevin processes". In: *IMA Journal of Numerical Analysis* 35.2 (2015), pp. 583–614. eprint: /oup/backfile/content_public/journal/imajna/35/2/10.1093/imanum/dru016/2/dru016.pdf.

[KS16] C. Kumar and S. Sabanis. "On Milstein approximations with varying coefficients: the case of super-linear diffusion coefficients". In: *ArXiv e-prints* (Jan. 2016). arXiv: 1601.02695 [math.PR].

[KS17] Chaman Kumar and Sotirios Sabanis. "On tamed milstein schemes of SDEs driven by Lévy noise". In: *Discrete and Continuous Dynamical Systems - Series B* 22.2 (2017), pp. 421–463.

[KS91] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer New York, 1991.

[Kul97] S. Kullback. *Information theory and statistics*. Reprint of the second (1968) edition. Dover Publications, Inc., Mineola, NY, 1997, pp. xvi+399.

[Le 16] Jean-François Le Gall. *Brownian motion, martingales, and stochastic calculus*. Vol. 274. Springer, 2016.

[LFR17] S. Livingstone, M. F. Faulkner, and G. O. Roberts. “Kinetic energy choice in Hamiltonian/hybrid Monte Carlo”. In: *ArXiv e-prints* (June 2017). arXiv: 1706.02649 [stat.CO].

[LMS07] H. Lamba, J. C. Mattingly, and A. M. Stuart. “An adaptive Euler–Maruyama scheme for SDEs: convergence and stability”. In: *IMA Journal of Numerical Analysis* 27.3 (2007), pp. 479–506. eprint: /oup/backfile/content_public/journal/imajna/27/3/10.1093/imanum/drl032/2/drl032.pdf.

[LS13] Robert Liptser and Albert N Shiryaev. *Statistics of random Processes: I. general Theory*. Vol. 5. Springer Science & Business Media, 2013.

[LS16] Tony Lelièvre and Gabriel Stoltz. “Partial differential equations and stochastic methods in molecular dynamics”. In: *Acta Numerica* 25 (2016), pp. 681–880.

[MSH02] J. C. Mattingly, A. M. Stuart, and D. J. Higham. “Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise”. In: *Stochastic Process. Appl.* 101.2 (2002), pp. 185–232.

[MST10] Jonathan C. Mattingly, Andrew M. Stuart, and M. V. Tretyakov. “Convergence of Numerical Time-Averaging and Stationary Measures via Poisson Equations”. In: *SIAM Journal on Numerical Analysis* 48.2 (2010), pp. 552–577. eprint: http://dx.doi.org/10.1137/090770527.

[MT93] S. P. Meyn and R. L. Tweedie. “Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes”. In: *Adv. in Appl. Probab.* 25.3 (1993), pp. 518–548.

[Par81] G. Parisi. “Correlation functions and computer simulations”. In: *Nuclear Physics B* 180 (1981), pp. 378–384.

[PV01] E. Pardoux and Yu. Veretennikov. “On the Poisson Equation and Diffusion Approximation. I”. In: *Ann. Probab.* 29.3 (July 2001), pp. 1061–1085.

[RT96] G. O. Roberts and R. L. Tweedie. “Exponential convergence of Langevin distributions and their discrete approximations”. In: *Bernoulli* 2.4 (1996), pp. 341–363.

[Sab13] Sotirios Sabanis. “A note on tamed Euler approximations”. In: *Electron. Commun. Probab.* 18 (2013), 10 pp.

[Vil09] C. Villani. *Optimal transport : old and new*. Grundlehren der mathematischen Wissenschaften. Berlin: Springer, 2009.

[WG13] Xiaojie Wang and Siqing Gan. “The tamed Milstein method for commutative stochastic differential equations with non-globally Lipschitz continuous coefficients”. In: *Journal of Difference Equations and Applications* 19.3 (2013), pp. 466–490. eprint: http://dx.doi.org/10.1080/10236198.2012.656617.