

Non-asymptotic convergence analysis for the Unadjusted Langevin Algorithm

Alain Durmus, Éric Moulines

▶ **To cite this version:**

Alain Durmus, Éric Moulines. Non-asymptotic convergence analysis for the Unadjusted Langevin Algorithm. 2016. hal-01176132v3

HAL Id: hal-01176132
https://hal.science/hal-01176132v3

Preprint submitted on 19 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.

Non-asymptotic convergence analysis for the Unadjusted Langevin Algorithm

Alain Durmus [*] **and Éric Moulines** [**]

LTCI, Telecom ParisTech & CNRS,
46 rue Barrault, 75634 Paris Cedex 13, France.
e-mail: [*]alain.durmus@telecom-paristech.fr

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

Abstract: : In this paper, we study a method to sample from a target distribution π over $\mathbb{R}^d$ having a positive density with respect to the Lebesgue measure, known up to a normalisation factor. This method is based on the Euler discretization of the overdamped Langevin stochastic differential equation associated with π. For both constant and decreasing step sizes in the Euler discretization, we obtain non-asymptotic bounds for the convergence to the target distribution π in total variation distance. A particular attention is paid to the dependency on the dimension d, to demonstrate the applicability of this method in the high dimensional setting. These bounds improve and extend the results of [12].

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

1. Introduction

Sampling distributions over high-dimensional state-spaces is a problem which has recently attracted a lot of research efforts in computational statistics and machine learning (see [11] and [1] for details); applications include Bayesian non-parametrics, Bayesian inverse problems and aggregation of estimators. All these problems boil down to sample a target distribution π having a density w.r.t. the Lebesgue measure on $\mathbb{R}^d$, known up to a normalisation factor $x \mapsto \mathrm{e}^{-U(x)} / \int_{\mathbb{R}^d} \mathrm{e}^{-U(y)} \mathrm{d}y$ where U is continuously differentiable. We consider a sampling method based on the Euler discretization of the overdamped Langevin stochastic differential equation (SDE)

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

where $(B_t^d)_{t\geq 0}$ is a d-dimensional Brownian motion. It is well-known that the Markov semi-group associated with the Langevin diffusion $(Y_t)_{t\geq 0}$ is reversible w.r.t. π. Under suitable conditions, the convergence to π takes place at geometric rate. Precise quantitative estimates of the rate of convergence with explicit

dependency on the dimension d of the state space have been recently obtained using either functional inequalities such as Poincaré and log-Sobolev inequalities (see [3, 9] [4]) or by coupling techniques (see [15]). The Euler-Maruyama discretization scheme associated to the Langevin diffusion yields the discrete time-Markov chain given by

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

where $(Z_k)_{k\geq 1}$ is an i.i.d. sequence of standard Gaussian d-dimensional random vectors 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. The idea of using the Markov chain $(X_k)_{k\geq 0}$ to sample approximately from the target π has been first introduced in the physics literature by [34] and popularised in the computational statistics community by [17] and [18]. It has been studied in depth by [35], which proposed to use a Metropolis-Hastings step at each iteration to enforce reversibility w.r.t. π leading to the Metropolis Adjusted Langevin Algorithm (MALA). They coin the term *unadjusted* Langevin algorithm (ULA) when the Metropolis-Hastings step is skipped.

The purpose of this paper is to study the convergence of the ULA algorithm. The emphasis is put on non-asymptotic computable bounds; we pay a particular attention to the way these bounds scale with the dimension d and constants characterizing the smoothness and curvature of the potential U. Our study covers both constant and decreasing step sizes and we analyse both the "finite horizon" (where the total number of simulations is specified before running the algorithm) and "any-time" settings (where the algorithm can be stopped after any iteration).

When the step size $\gamma_k = \gamma$ is constant, under appropriate conditions (see [35]), the Markov chain $(X_n)_{n\geq 0}$ is V-uniformly geometrically ergodic with a stationary distribution π_γ. With few exceptions, the stationary distribution π_γ is different from the target π. If the step size γ is small enough, then the stationary distribution of this chain is in some sense close to π. We provide non-asymptotic bounds of the V-total variation distance between π_γ and π, with explicit dependence on the step size γ and the dimension d. Our results complete and extend the recent works by [13] and [12].

When $(\gamma_k)_{k\geq 1}$ decreases to zero, then $(X_k)_{k\geq 0}$ is a non-homogeneous Markov chain. If in addition $\sum_{k=1}^{\infty}\gamma_k = \infty$, we show that the marginal distribution of this non-homogeneous chain converges, under some mild additional conditions, to the target distribution π, and provide explicit bounds for the convergence. Compared to the related works by [23], [24], [25] and [26], we establish not only the weak convergence of the weighted empirical measure of the path to the target distribution but a much stronger convergence in total variation, similarly to [12], where the strongly log-concave case is considered.

The paper is organized as follows. In Section 2, the main convergence results are stated under abstract assumptions. We then specialize in Section 3 these results to different classes of densities. The proofs are gathered in Section 4. Some general convergence results for diffusions based on reflection coupling, which are of independent interest, are stated in Section 5.

Notations and conventions

$\mathcal{B}(\mathbb{R}^d)$ denotes the Borel σ-field of $\mathbb{R}^d$ and $\mathbb{F}(\mathbb{R}^d)$ the set of all Borel measurable functions on $\mathbb{R}^d$. For $f \in \mathbb{F}(\mathbb{R}^d)$ set $\|f\|_\infty = \sup_{x\in\mathbb{R}^d} |f(x)|$. Denote by $\mathbb{M}(\mathbb{R}^d)$ the space of finite signed measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $\mathbb{M}_0(\mathbb{R}^d) = \{\mu \in \mathbb{M}(\mathbb{R}^d) \mid \mu(\mathbb{R}^d) = 0\}$. For $\mu \in \mathbb{M}(\mathbb{R}^d)$ and $f \in \mathbb{F}(\mathbb{R}^d)$ a μ-integrable function, denote by $\mu(f)$ the integral of f w.r.t. μ. Let $V : \mathbb{R}^d \to [1, \infty)$ be a measurable function. For $f \in \mathbb{F}(\mathbb{R}^d)$, the V-norm of f is given by $\|f\|_V = \sup_{x\in\mathbb{R}^d} |f(x)|/V(x)$. For $\mu \in \mathbb{M}(\mathbb{R}^d)$, the V-total variation distance of μ is defined as

$$\|\mu\|_V = \sup_{f\in\mathbb{F}(\mathbb{R}^d), \|f\|_V \le 1} \left| \int_{\mathbb{R}^d} f(x) \mathrm{d}\mu(x) \right| .$$

If $V \equiv 1$, then $\|\cdot\|_V$ is the total variation denoted by $\|\cdot\|_{\mathrm{TV}}$.

For $p \ge 1$, denote by $\mathrm{L}^p(\pi)$ the set of measurable functions such that $\pi(|f|^p) < \infty$. For $f \in \mathrm{L}^2(\pi)$, the variance of f under π is denoted by $\mathrm{Var}_\pi\{f\}$. For all functions f such that $f \log(f) \in \mathrm{L}^1(\pi)$, the entropy of f with respect to π is defined by

$$\mathrm{Ent}_\pi(f) = \int_{\mathbb{R}^d} f(x) \log(f(x)) \mathrm{d}\pi(x) .$$

Let μ and ν be two probability measures on $\mathbb{R}^d$. If $\mu \ll \nu$, we denote by $\mathrm{d}\mu/\mathrm{d}\nu$ the Radon-Nikodym derivative of μ w.r.t. ν. Denote for all $x, y \in \mathbb{R}^d$ by $\langle x, y\rangle$ the scalar product of x and y and $\|x\|$ the Euclidean norm of x. For $k \ge 0$, denote by $C^k(\mathbb{R}^d)$, the set of k-times continuously differentiable functions $f : \mathbb{R}^d \to \mathbb{R}$. For $f \in C^2(\mathbb{R}^d)$, denote by ∇f the gradient of f and Δf the Laplacian of f. For all $x \in \mathbb{R}^d$ and $M > 0$, we denote by $\mathrm{B}(x, M)$, the ball centered at x of radius M. Denote for $K \ge 0$, the oscillation of a function $f \in C^0(\mathbb{R}^d)$ in the ball $\mathrm{B}(0, K)$ by $\mathrm{osc}_K(f) = \sup_{\mathrm{B}(0,K)}(f) - \inf_{\mathrm{B}(0,K)}(f)$. Denote the oscillation of a bounded function $f \in C^0(\mathbb{R}^d)$ on $\mathbb{R}^d$ by $\mathrm{osc}_{\mathbb{R}^d}(f) = \sup_{\mathbb{R}^d}(f) - \inf_{\mathbb{R}^d}(f)$. 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. General conditions for the convergence of ULA

In this section, we derive a bound on the convergence of the ULA to the target distribution π when the Langevin diffusion is geometrically ergodic and the Markov kernel associated with the EM discretization satisfies a Foster-Lyapunov drift inequality.

Consider the following assumption on the potential U:

L1. *The function U is continuously differentiable on $\mathbb{R}^d$ and gradient Lipschitz, i.e. there exists $L \ge 0$ such that for all $x, y \in \mathbb{R}^d$,*

$$\|\nabla U(x) - \nabla U(y)\| \le L \|x - y\| .$$

Under **L**1, by [20, Theorem 2.4-3.1] for every initial point $x \in \mathbb{R}^d$, there exists a unique strong solution $(Y_t(x))_{t\ge0}$ to the Langevin SDE (1). Define for all

$t \geq 0$, $x \in \mathbb{R}^d$ and $\mathrm{A} \in \mathcal{B}(\mathbb{R}^d)$, $P_t(x, \mathrm{A}) = \mathbb{P}\left(Y_t(x) \in \mathrm{A}\right)$. The semigroup $(P_t)_{t \geq 0}$ is reversible w.r.t. π, and hence admits π as its (unique) invariant distribution. In this section, we consider the case where $(P_t)_{t \geq 0}$ is geometrically ergodic, i.e. there exists $\kappa \in [0, 1)$ such that for any initial distribution μ_0 and $t > 0$,

$$\left\| \mu_0 P_t - \pi \right\|_{\mathrm{TV}} \leq C(\mu_0) \kappa^t \, , \tag{3}$$

for some constant $C(\mu_0) \in [0, +\infty]$. Denote by $\mathscr{A}^{\mathrm{L}}$ the generator associated with the semigroup $(P_t)_{t \geq 0}$, given for all $f \in C^2(\mathbb{R}^d)$ by

$$\mathscr{A}^{\mathrm{L}} f = - \left\langle \nabla U, \nabla f \right\rangle + \Delta f \, .$$

A twice continuously differentiable function $V : \mathbb{R}^d \to [1, \infty)$ is a *Lyapunov function* for the generator $\mathscr{A}^{\mathrm{L}}$ if there exist $\theta > 0$, $\beta \geq 0$ and $\mathcal{E} \subset \mathcal{B}$ such that,

$$\mathscr{A}^{\mathrm{L}} V \leq -\theta V + \beta 1_{\mathcal{E}} \, . \tag{4}$$

By [35, Theorem 2.2], if $\mathcal{E}$ in (4) is a non-empty compact set, then the Langevin diffusion is geometrically ergodic.

Consider now the EM discretization of the diffusion (2). Let $(\gamma_k)_{k \geq 1}$ be a sequence of positive and nonincreasing step sizes and for $0 \leq n \leq p$, denote by

$$\Gamma_{n,p} = \sum_{k=n}^{p} \gamma_k \, , \qquad \Gamma_n = \Gamma_{1,n} \, . \tag{5}$$

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

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

The discretized Langevin diffusion $(X_n)_{n \geq 0}$ given in (2) is a time-inhomogeneous Markov chain, for $p \geq n \geq 1$ and $f \in \mathbb{F}_+(\mathbb{R}^d)$, $\mathbb{E}^{\mathcal{F}_n}\left[f(X_p)\right] = Q_\gamma^{n,p} f(X_n)$ where $\mathcal{F}_n = \sigma(X_\ell, 0 \leq \ell \leq n)$ and

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

with the convention that for $n, p \geq 0$, $n < p$, $Q_\gamma^{p,n}$ is the identity operator. Under **L**1, the Markov kernel R_γ is strongly Feller, irreducible and strongly aperiodic. We will say that a function $V : \mathbb{R}^d \to [1, \infty)$ satisfies a Foster-Lyapunov drift condition for R_γ if there exist constants $\bar{\gamma} > 0$, $\lambda \in [0, 1)$ and $c > 0$ such that, for all $\gamma \in (0, \bar{\gamma}]$

$$R_\gamma V \leq \lambda^\gamma V + \gamma c \, . \tag{6}$$

The particular form of (6) reflects how the mixing rate of the Markov chain depends upon the step size $\gamma > 0$. If $\gamma = 0$, then $R_0(x, \mathrm{A}) = \delta_x(\mathrm{A})$ for $x \in \mathbb{R}^d$ and $\mathrm{A} \in \mathcal{B}(\mathbb{R}^d)$. A Markov chain with transition kernel R_0 is not mixing. Intuitively, as γ gets larger, then it is expected that the mixing of R_γ increases. If for some $\gamma > 0$, R_γ satisfies (6), then R_γ admits a unique stationary distribution π_γ. We use (6) to control quantitatively the moments of the time-inhomogeneous chain. The types of bounds which are needed, are summarised in the following elementary Lemma.

Lemma 1. *Let $\bar{\gamma} > 0$. Assume that for all $x \in \mathbb{R}^d$ and $\gamma \in (0, \bar{\gamma}]$, (6) holds for some constants $\lambda \in (0,1)$ and $c > 0$. Let $(\gamma_k)_{k\geq 1}$ be a sequence of nonincreasing step sizes such that $\gamma_k \in (0, \bar{\gamma}]$ for all $k \in \mathbb{N}^*$. Then for all $n \geq 0$ and $x \in \mathbb{R}^d$, $Q_\gamma^n V(x) \leq F(\lambda, \Gamma_n, c, \gamma_1, V(x))$ where*

$$F(\lambda, a, c, \gamma, w) = \lambda^a w + c(-\lambda^\gamma \log(\lambda))^{-1} . \tag{7}$$

Proof. The proof is postponed to Section 4.1. □

Note that Lemma 1 implies that $\sup_{k\geq 0}\{Q_\gamma^k V(x)\} \leq G(\lambda, c, \gamma_1, V(x))$ where

$$G(\lambda, c, \gamma, w) = w + c(-\lambda^\gamma \log(\lambda))^{-1} . \tag{8}$$

We give below the main ingredients which are needed to obtain a quantitative bound for $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}}$ for all $x \in \mathbb{R}^d$. This quantity is decomposed as follows: for all $0 \leq n < p$,

$$\begin{aligned}\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \\ \leq \|\delta_x Q_\gamma^n Q_\gamma^{n+1,p} - \delta_x Q_\gamma^n P_{\Gamma_{n+1,p}}\|_{\mathrm{TV}} + \|\delta_x Q_\gamma^n P_{\Gamma_{n+1,p}} - \pi\|_{\mathrm{TV}} .\end{aligned} \tag{9}$$

To control the first term on the right hand side, we use a method introduced in [13] and elaborated in [12]. The second term is bounded using the convergence of the semi-group to π, see (3).

Proposition 2. *Assume that **L** 1 and (3) hold. Let $(\gamma_k)_{k\geq 0}$ be a sequence of nonnegative step sizes. Then for all $x \in \mathbb{R}^d$, $n \geq 0$, $p \geq 1$, $n < p$,*

$$\begin{aligned}\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \\ \leq 2^{-1/2} L \left(\sum_{k=n}^{p-1} \left\{ (\gamma_{k+1}^3/3) A(\gamma, x) + d\gamma_{k+1}^2 \right\} \right)^{1/2} + C(\delta_x Q_\gamma^n) \kappa^{\Gamma_{n+1,p}} ,\end{aligned} \tag{10}$$

where $\kappa, C(\delta_x Q_\gamma^n)$ are defined in (3) and

$$A(\gamma, x) = \sup_{k\geq 0} \int_{\mathbb{R}^d} \|\nabla U(z)\|^2 \, Q_\gamma^k(y, \mathrm{d}z) . \tag{11}$$

Proof. The proof follows the same lines as [12, Lemma 2] but is given for completeness. For $0 \leq s \leq t$, let $\mathrm{C}([s,t], \mathbb{R}^d)$ be the space of continuous functions on $[s,t]$ taking values in $\mathbb{R}^d$. For all $y \in \mathbb{R}^d$, denote by $\mu_{n,p}^y$ and $\bar{\mu}_{n,p}^y$ the laws on $\mathrm{C}([\Gamma_n, \Gamma_p], \mathbb{R}^d)$ of the Langevin diffusion $(Y_t(y))_{\Gamma_n \leq t \leq \Gamma_p}$ and of the continuously-interpolated Euler discretization $(\bar{Y}_t(y))_{\Gamma_n \leq t \leq \Gamma_p}$, both started at y at time Γ_n. Denote by $(Y_t(y), \overline{Y}_t(y))_{t\geq \Gamma_n}$ the unique strong solution started at (y,y) at time $t = \Gamma_n$ of the time-inhomogeneous diffusion defined for $t \geq \Gamma_n$, by

$$\begin{cases} \mathrm{d}Y_t = -\nabla U(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t^d \\ \mathrm{d}\bar{Y}_t = -\overline{\nabla U}(\bar{Y}, t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t^d , \end{cases} \tag{12}$$

where for any continuous function $\mathrm{w} : \mathbb{R}_+ \to \mathbb{R}^d$ and $t \geq \Gamma_n$

$$\overline{\nabla U}(\mathrm{w}, t) = \sum_{k=n}^{\infty} \nabla U(\mathrm{w}_{\Gamma_k}) 1_{[\Gamma_k, \Gamma_{k+1})}(t) \ . \tag{13}$$

Girsanov's Theorem [21, Theorem 5.1, Corollary 5.16, Chapter 3] shows that $\mu^y_{n,p}$ and $\bar{\mu}^y_{n,p}$ are mutually absolutely continuous and in addition, $\bar{\mu}^y_{n,p}$-almost surely

$$\frac{\mathrm{d}\mu^y_{n,p}}{\mathrm{d}\bar{\mu}^y_{n,p}} = \exp\left(\frac{1}{2} \int_{\Gamma_n}^{\Gamma_p} \left\langle \nabla U(\bar{Y}_s(y)) - \overline{\nabla U}(\bar{Y}(y), s), \mathrm{d}\bar{Y}_s(y) \right\rangle \right.$$
$$\left. - \frac{1}{4} \int_{\Gamma_n}^{\Gamma_p} \left\{ \left\| \nabla U(\bar{Y}_s(y)) \right\|^2 - \left\| \overline{\nabla U}(\bar{Y}(y), s) \right\|^2 \right\} \mathrm{d}s \right) \ . \tag{14}$$

Under **L**1, (14) implies for all $y \in \mathbb{R}^d$:

$$\begin{aligned} \mathrm{KL}(\mu^y_{n,p} | \bar{\mu}^y_{n,p}) &\leq 4^{-1} \int_{\Gamma_n}^{\Gamma_p} \mathbb{E}\left[\left\| \nabla U(\bar{Y}_s(y)) - \overline{\nabla U}(\bar{Y}(y), s) \right\|^2 \right] \mathrm{d}s \\ &\leq 4^{-1} \sum_{k=n}^{p-1} \int_{\Gamma_k}^{\Gamma_{k+1}} \mathbb{E}\left[\left\| \nabla U(\bar{Y}_s(y)) - \nabla U(\bar{Y}_{\Gamma_k}(y)) \right\|^2 \right] \mathrm{d}s \\ &\leq 4^{-1} L^2 \sum_{k=n}^{p-1} \left\{ (\gamma_{k+1}^3/3) \int_{\mathbb{R}^d} \|\nabla U(z)\|^2 \, Q_\gamma^{n+1,k}(y, \mathrm{d}z) + d\gamma_{k+1}^2 \right\} \ . \end{aligned} \tag{15}$$

By the Pinsker inequality, $\|\delta_y Q_\gamma^{n+1,p} - \delta_y P_{\Gamma_{n+1,p}}\|_{\mathrm{TV}} \leq \sqrt{2}\{\mathrm{KL}(\mu^y_{n,p} | \bar{\mu}^y_{n,p})\}^{1/2}$. The proof is concluded by combining this inequality, (15) and (3) in (9). □

In the sequel, depending on the conditions on the potential U and the techniques of proof, for any given $x \in \mathbb{R}^d$, $C(\delta_x Q_\gamma^n)$ can have two kinds of upper bounds, either of the form $-\log(\gamma_n) W(x)$, or $\exp(a\Gamma_n) W(x)$, for some function $W : \mathbb{R}^d \to \mathbb{R}$ and $a > 0$. In both cases, as shown in Proposition 3, it is possible to choose n as a function of p, so that $\lim_{p \to +\infty} \|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} = 0$ under appropriate conditions on the sequence of step sizes $(\gamma_k)_{k \geq 1}$.

Proposition 3. *Assume that* ***L****1 and* (3) *hold. Let* $(\gamma_k)_{k \geq 1}$ *be a nonincreasing sequence satisfying* $\lim_{k \to +\infty} \Gamma_k = +\infty$ *and* $\lim_{k \to \infty} \gamma_k = 0$. *Then,* $\lim_{n \to \infty} \|\delta_x Q_\gamma^n - \pi\|_{\mathrm{TV}} = 0$ *for any* $x \in \mathbb{R}^d$ *for which one of the two following conditions holds:*

(i) $A(\gamma, x) < \infty$ *and* $\limsup_{n \to +\infty} C(\delta_x Q_\gamma^n)/(-\log(\gamma_n)) < +\infty$, *where* $A(\gamma, x)$ *is defined in* (11).

(ii) $\sum_{k=1}^{\infty} \gamma_k^2 < +\infty$, $A(\gamma, x) < \infty$ *and* $\limsup_{n \to +\infty} \log\{C(\delta_x Q_\gamma^n)\}/\Gamma_n < +\infty$.

Proof. (i) There exists $p_0 \geq 1$ such that for all $p \geq p_0$, $\kappa^{\gamma_p} > \gamma_p$ and $\kappa^{\Gamma_p} \leq \gamma_1$. Therefore, we can define for all $p \geq p_0$,

$$n(p) \stackrel{\mathrm{def}}{=} \min\left\{ k \in \{0, \cdots, p-1\} \, | \kappa^{\Gamma_{k+1,p}} > \gamma_{k+1} \right\} \ . \tag{16}$$

and $n(p) \geq 1$. We first show that $\liminf_{p\to\infty} n(p) = \infty$. The proof goes by contradiction. If $\liminf_{p\to\infty} n(p) < \infty$ we could extract a bounded subsequence $(n(p_k))_{k\geq 1}$. For such sequence, $(\gamma_{n(p_k)+1})_{k\geq 1}$ is bounded away from 0, but $\lim_{k\to+\infty} \kappa^{\Gamma_{n(p_k)+1,p_k}} = 0$ which yields to a contradiction. The definition of $n(p)$ implies that $\kappa^{\Gamma_{n(p),p}} \leq \gamma_{n(p)}$, showing that

$$\limsup_{p\to+\infty} C(\delta_x Q_\gamma^{n(p)}) \kappa^{\Gamma_{n(p),p}} \leq \limsup_{p\to+\infty} \frac{C(\delta_x Q_\gamma^{n(p)})}{-\log(\gamma_{n(p)})} \limsup_{p\to+\infty} \left\{\gamma_{n(p)}(-\log(\gamma_{n(p)}))\right\} = 0 \, .$$

On the other hand, since $(\gamma_k)_{k\geq 1}$ is nonincreasing, for any $\ell \geq 2$,

$$\sum_{k=n(p)+1}^{p} \gamma_k^\ell \leq \gamma_{n(p)+1}^{\ell-1} \Gamma_{n(p)+1,p} \leq \gamma_{n(p)+1}^{\ell-1} \log(\gamma_{n(p)+1}) / \log(\kappa) \, .$$

The proof follows from (10) using $\lim_{p\to\infty} \gamma_{n(p)} = 0$.

(ii) For all $p \geq 1$, define $n(p) = \max(0, \lfloor \log(\Gamma_p) \rfloor)$. Note that since $\lim_{k\to+\infty} \Gamma_k = +\infty$, we have $\lim_{p\to+\infty} n(p) = +\infty$. Using $\sum_{k=1}^{+\infty} \gamma_k^2 < +\infty$ and $(\gamma_k)_{k\geq 1}$ is a nonincreasing sequence, we get for all $\ell \geq 2$,

$$\lim_{p\to+\infty} \sum_{k=n(p)}^{p} \gamma_k^\ell = 0 \, ,$$

which shows that the first term in the right side of (10) goes to 0 as p goes to infinity. As for the second term, since $\limsup_{n\to+\infty} \log\{C(\delta_x Q_\gamma^n)\}/\Gamma_n < +\infty$, we get using that $(\gamma_k)_{k\geq 1}$ is nonincreasing and $n(p) \leq \log(\Gamma_p)$,

$$\begin{aligned} &C(\delta_x Q_\gamma^{n(p)}) \kappa^{\Gamma_{n(p),p}} \\ &\leq \exp\left(\log(\kappa)\Gamma_p + \left[\{\log(C(\delta_x Q_\gamma^{n(p)}))/\Gamma_{n(p)}\}_+ - \log(\kappa)\right] \Gamma_{n(p)}\right) \\ &\leq \exp\left(\log(\kappa)\Gamma_p + \left[\sup_{k\geq 1}\{\log(C(\delta_x Q_\gamma^k))/\Gamma_k\}_+ - \log(\kappa)\right] \gamma_1 \log(\Gamma_p)\right) \, . \end{aligned}$$

Using $\kappa < 1$ and $\lim_{k\to+\infty} \Gamma_k = +\infty$, we have $\lim_{p\to+\infty} C(\delta_x Q_\gamma^{n(p)}) \kappa^{\Gamma_{n(p),p}} = 0$, which concludes the proof.

□

Using (10), we can also assess the convergence of the algorithm for constant step sizes $\gamma_k = \gamma$ for all $k \geq 1$. Two different kinds of results can be derived. First, for a given precision $\varepsilon > 0$, we can try to optimize the step size γ to minimize the number of iterations p required to achieve $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$. Second if the total number of iterations is fixed $p \geq 1$, we may determine the step size $\gamma > 0$ which minimizes $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}}$.

Lemma 4. *Assume that (10) holds. Assume that there exists $\bar\gamma > 0$ such that $\bar C(x) = \sup_{\gamma\in(0,\bar\gamma]} \sup_{n\geq 1} C(\delta_x R_\gamma^n) < +\infty$ and $\sup_{\gamma\in(0,\bar\gamma]} A(\gamma, x) \leq \bar A(x)$, where $C(\delta_x R_\gamma^n)$ and $A(\gamma, x)$ are defined in (3) and (11) respectively. Then for all $\varepsilon > 0$, we get $\|\delta_x R_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ if*

$$p > T\gamma^{-1} \quad \text{and} \quad \gamma \leq \frac{-d + \sqrt{d^2 + (2/3)\bar A(x)\varepsilon^2 (L^2 T)^{-1}}}{2\bar A(x)/3} \wedge \bar\gamma \,, \tag{17}$$

where

$$T = \big(\log\{\bar C(x)\} - \log(\varepsilon/2)\big) \Big/ (-\log(\kappa)) \,.$$

Proof. For $p > T\gamma^{-1}$, set $n = p - \lfloor T\gamma^{-1} \rfloor$. Then using the stated expressions of γ and T in (10) concludes the proof. □

Note that an upper bound for γ defined in (17) is $\epsilon^2 (L^2 T d)^{-1}$. The dependency of T on the dimension d will be addressed in Section 3.

Lemma 5. *Assume that **L**1 and (3) hold. In addition, assume that there exist $\bar\gamma > 0$ and $n \in \mathbb{N}$, $n > 0$, such that $\bar C_n(x) = \sup_{\gamma\in(0,\bar\gamma]} C(\delta_x R_\gamma^n) < +\infty$ and $\sup_{\gamma\in(0,\bar\gamma]} A(\gamma, x) \leq \bar A(x)$. For all $p > n$ and all $x \in \mathbb{R}^d$, if $\gamma = \log(p-n)\{(p-n)(-\log(\kappa))\}^{-1} \leq \bar\gamma$, then*

$$\begin{aligned} &\|\delta_x R_\gamma^p - \pi\|_{\mathrm{TV}} \\ &\leq (p-n)^{-1/2}\{\bar C_n(x)(p-n)^{-1/2} + \log(p-n)(d + \bar A(x)\log(p-n)(p-n)^{-1})^{1/2}\} \,. \end{aligned}$$

Proof. The proof is a straightforward calculation using (10). □

To get quantitative bounds for the total variation distance $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}}$ it is therefore required to get bounds on κ, $A(\gamma, x)$ and to control $C(\delta_x Q_\gamma^n)$. We will consider in the sequel two different approaches to get (3), one based on functional inequalities, the other on coupling techniques. We will consider also increasingly stringent assumptions for the potential U. Whereas we will always obtain the same type of exponential bounds, the dependency of the constants on the dimension will be markedly different. In the worst case, the dependency is exponential. It is polynomial when U is convex.

3. Practical conditions for geometric ergodicity of the Langevin diffusion and their consequences for ULA

3.1. Superexponential densities

Assume first that the potential is superexponential outside a ball. This is a rather weak assumption (we do not assume convexity here).

H1. *The potential U is twice continuously differentiable and there exist $\rho > 0$, $\alpha \in (1, 2]$ and $M_\rho \geq 0$ such that for all $x \in \mathbb{R}^d$, $\|x - x^\star\| \geq M_\rho$, $\langle \nabla U(x), x - x^\star\rangle \geq \rho \|x - x^\star\|^\alpha$.*

The price to pay will be constants which are exponential in the dimension. Under **H**1, the potential U is unbounded off compact set. Since U is continuous, it has a global minimizer $x^\star$, which is a point at which $\nabla U(x^\star) = 0$. Without loss of generality, it is assumed that $U(x^\star) = 0$.

Lemma 6. *Assume **L**1 and **H**1. Then for all $x \in \mathbb{R}^d$,*

$$U(x) \geq \rho \left\|x - x^\star\right\|^\alpha / (\alpha + 1) - a_\alpha \quad \text{with} \quad a_\alpha = \rho M_\rho^\alpha / (\alpha + 1) + M_\rho^2 L/2 \; . \tag{18}$$

Proof. The elementary proof is postponed to Section 4.2. □

Following [35, Theorem 2.3], we first establish a drift condition for the diffusion.

Proposition 7. *Assume **L**1 and **H**1. For any $\varsigma \in (0,1)$, the drift condition (4) is satisfied with the Lyapunov function $V_\varsigma(x) = \exp(\varsigma U(x))$, $\theta_\varsigma = \varsigma dL$, $\mathcal{E}_\varsigma = \mathrm{B}(x^\star, K_\varsigma)$, $K_\varsigma = \max(\{2dL/(\rho(1-\varsigma))\}^{1/(2(\alpha-1))}, M_\rho)$ and $\beta_\varsigma = \varsigma dL \sup_{\{y \in \mathcal{E}_\varsigma\}} \{V_\varsigma(y)\}$. Moreover, there exist constants $C_\varsigma < \infty$ and $\upsilon_\varsigma > 0$ such that for all $t \in \mathbb{R}_+$ and probability measures μ_0 and ν_0 on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, satisfying $\mu_0(V_\varsigma) + \nu_0(V_\varsigma) < +\infty$,*

$$\left\|\mu_0 P_t - \nu_0 P_t\right\|_{V_\varsigma} \leq C_\varsigma \mathrm{e}^{-\upsilon_\varsigma t} \left\|\mu_0 - \nu_0\right\|_{V_\varsigma} \; , \quad \left\|\mu_0 P_t - \pi\right\|_{V_\varsigma} \leq C_\varsigma \mathrm{e}^{-\upsilon_\varsigma t} \mu_0(V_\varsigma) \; .$$

Proof. The proof, adapted from [35, Theorem 2.3] and [31, Theorem 6.1], is postponed to Section 4.3. □

Under **H**1, explicit expressions for C_ς and υ_ς have been developed in the literature but these estimates are in general very conservative. We now turn to establish (6) for the Euler discretization.

Proposition 8. *Assume **L**1 and **H**1. Let $\bar\gamma \in \left(0, L^{-1}\right)$. For all $\gamma \in (0, \bar\gamma]$ and $x \in \mathbb{R}^d$, R_γ satisfies the drift condition (6) with $V(x) = \exp(U(x)/2)$, $K = \max(M_\rho, (8\log(\lambda)/\rho^2)^{1/(2(\alpha-1))})$, $c = -2\log(\lambda)\lambda^{-\bar\gamma} \sup_{\{y \in \mathrm{B}(x^\star, K)\}} V(y)$ and $\lambda = \mathrm{e}^{-dL/\{2(1-L\bar\gamma)\}}$.*

Proof. The proof is postponed to Section 4.4. □

Theorem 9. *Assume **L**1 and **H**1. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 < \bar\gamma$, $\bar\gamma \in \left(0, L^{-1}\right)$. Then, for all $n \geq 0$, $p \geq 1$, $n < p$, and $x \in \mathbb{R}^d$, (10) holds with $\log(\kappa) = -\upsilon_{1/2}$ and*

$$A(\gamma, x) \leq L^2 \left(\frac{\alpha + 1}{\rho} \left[a_\alpha + \frac{4(2-\alpha)(\alpha+1)}{\alpha\rho} + 2\log\left\{G(\lambda, c, \gamma_1, V(x))\right\} \right] \right)^{2/\alpha}$$

$$C(\delta_x Q_\gamma^n) \leq C_{1/2} F(\lambda, \Gamma_{1,n}, c, \gamma_1, V(x)) \; , \tag{19}$$

where $C_{1/2}$, $\upsilon_{1/2}$ are given by Proposition 7, F by (7), V, λ, c in Proposition 8, G by (8), a_α in (18).

Proof. The proof is postponed to Section 4.5. □

Equation (19) implies that for all $x \in \mathbb{R}^d$, we have $\sup_{n\geq 0} C(\delta_x Q_\gamma^n) \leq G(\lambda, c, \gamma_1, V(x))$, so Proposition 3-(i) shows that $\lim_{p\to+\infty} \|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} = 0$ for all $x \in \mathbb{R}^d$ provided that $\lim_{k\to+\infty} \gamma_k = 0$ and $\lim_{k\to+\infty} \Gamma_k = +\infty$. In addition, for the case of constant step size $\gamma_k = \gamma$ for all $k \geq 1$, Lemma 4 and Lemma 5 can be applied.

Let $V : \mathbb{R}^d \to \mathbb{R}$, defined for all $x \in \mathbb{R}^d$ by $V(x) = \exp(U(x)/2)$. By Proposition 7, $(P_t)_{t\geq 0}$ is a contraction operator on the space of finite signed measure $\mu \in \mathbb{M}_0$, $\mu(V^{1/2}) < +\infty$, endowed with the norm $\|\cdot\|_{V^{1/2}}$. It is therefore possible to control $\|\delta_x Q_\gamma^p - \pi\|_{V^{1/2}}$. To simplify the notations, we limit our discussion to constant step sizes.

Theorem 10. *Assume* ***L*** *1 and* ***H*** *1. Then, for all $p \geq 1$, $x \in \mathbb{R}^d$ and $\gamma \in \left(0, L^{-1}\right)$, we have*

$$\|\delta_x R_\gamma^p - \pi\|_{V^{1/2}} \leq C_{1/4} \kappa^{\gamma p} V^{1/2}(x) + B(\gamma, V(x)) \,, \tag{20}$$

where $\log(\kappa) = -\upsilon_{1/4}$, $C_{1/4}, \upsilon_{1/4}, \theta_{1/2}, \beta_{1/2}$ are defined in Proposition 7, V, λ, c in Proposition 8, G in (8) and

$$B^2(\gamma, v) = L^2 \max(1, C_{1/4}^2)(1+\gamma)(1-\kappa)^{-2} \left(2G(\lambda, c, \gamma, v) + \beta_{1/2}/\theta_{1/2}\right) \\ \times \left(\gamma d + 3^{-1}\gamma^2 \|\nabla U\|_{V^{1/2}}^2 G(\lambda, c, \gamma, v)\right) \,.$$

Moreover, R_γ has a unique invariant distribution π_γ and

$$\|\pi - \pi_\gamma\|_{V^{1/2}} \leq B(\gamma, 1) \,.$$

Proof. The proof of (20) is postponed to Section 4.6. The bound for $\|\pi - \pi_\gamma\|_{V^{1/2}}$ is an easy consequence of (20): by Proposition 13 and [30, Theorem 16.0.1], R_γ is $V^{1/2}$-uniformly ergodic: $\lim_{p\to+\infty} \|\delta_x R_\gamma^p - \pi_\gamma\|_{V^{1/2}} = 0$ for all $x \in \mathbb{R}^d$. Finally, (20) shows that for all $x \in \mathbb{R}^d$,

$$\|\pi - \pi_\gamma\|_{V^{1/2}} \leq \lim_{p\to+\infty} \left\{\|\delta_x R_\gamma^p - \pi\|_{V^{1/2}} + \|\delta_x R_\gamma^p - \pi_\gamma\|_{V^{1/2}}\right\} \leq B(\gamma, V(x)) \,.$$

Taking the minimum over $x \in \mathbb{R}^d$ concludes the proof. □

Note that Theorem 10 implies that there exists a constant $C \geq 0$ which does not depend on γ such that $\|\pi - \pi_\gamma\|_{V^{1/2}} \leq C\gamma^{1/2}$.

Remark 11. *It is shown in [37, Theorem 4] that for $\phi \in C^\infty(\mathbb{R}^d)$ with polynomial growth, $\pi_\gamma(\phi) - \pi(\phi) = b(\phi)\gamma + \mathcal{O}(\gamma^2)$, for some constant $b(\phi) \in \mathbb{R}$, provided that $U \in C^\infty(\mathbb{R}^d)$ satisfies* ***L*** *1 and* ***H*** *1. Our result does not match this bound since $B(\gamma, 1) = \mathcal{O}(\gamma^{1/2})$. However the bound $B(\gamma, 1)$ is uniform over the class of measurable functions ϕ satisfying for all $x \in \mathbb{R}^d$, $|\phi(x)| \leq V^{1/2}(x)$. Obtaining such uniform bounds in total variation is important in Bayesian inference, for example to compute high posterior density credible regions. Our result also strengthens and completes [29, Corollary 7.5], which states that under* ***H*** *1 with $\alpha = 2$, for any measurable functions $\phi : \mathbb{R}^d \to \mathbb{R}$ satisfying for all $x, y \in \mathbb{R}^d$,*

$$|\phi(x) - \phi(y)| \leq C \|x - y\| \left\{1 + \|x\|^k + \|y\|^k\right\} \,,$$

for some $C \geq 0$, $k \geq 1$, $|\pi_\gamma(\phi) - \pi(\phi)| \leq C\gamma^\chi$ for some constants $C \geq 0$ and $\chi \in (0, 1/2)$, which does not depend on ϕ.

The bounds in Theorem 9 and Theorem 10 depend upon the constants appearing in Proposition 7 which are computable but are known to be pessimistic in general; see [36]. More explicit rates of convergence for the semigroup can be obtained using Poincaré inequality; see [3], [9] and [4, Chapter 4] and the references therein. The probability measure π is said to satisfy a Poincaré inequality with the constant C_{P} if for every locally Lipschitz function h,

$$\mathrm{Var}_\pi \{h\} \leq C_{\mathrm{P}} \int_{\mathbb{R}^d} \|\nabla h(x)\|^2 \, \pi(\mathrm{d}x) \; . \tag{21}$$

This inequality implies by [9, Theorem 2.1] that for all $t \geq 0$ and any initial distribution μ_0, such that $\mu_0 \ll \pi$,

$$\|\mu_0 P_t - \pi\|_{\mathrm{TV}} \leq \exp(-t/C_{\mathrm{P}}) \left(\mathrm{Var}_\pi \{\mathrm{d}\mu_0/\mathrm{d}\pi\}\right)^{1/2} \; . \tag{22}$$

[2, Theorem 1.4] shows that if the Lyapunov condition (4) is satisfied, then the Poincaré inequality (21) holds with an explicit constant. Denote by

$$D_n(\gamma) \stackrel{\mathrm{def}}{=} \left(4\pi \left\{ \prod_{k=1}^n (1 - L\gamma_k) \right\}^2 \sum_{i=1}^n \gamma_i (1 - L\gamma_i)^{-1} \right)^{-d/2} \; . \tag{23}$$

Theorem 12. *Assume **L**1 and **H**1. Let $(\gamma_k)_{k \geq 1}$ be a non increasing sequence. Then for all $n \geq 1$ and $x \in \mathbb{R}^d$, Equation (3) holds with*

$$\log(\kappa) = \left(-\theta_{1/2}^{-1} \left\{ 1 + (4\beta_{1/2} K_{1/2}^2 / \pi^2) \mathrm{e}^{\mathrm{osc}_{K_{1/2}}(U)} \right\} \right)^{-1} \; ,$$

$$C(\delta_x Q_\gamma^n) \leq \frac{(\alpha+1)^d (2\pi)^{(d+1)/2} (d-1)!}{\rho^d \mathbf{\Gamma}((d+1)/2)} D_n(\gamma) \mathrm{e}^{a_\alpha} \, \mathrm{e}^{U(x)} \; ,$$

where $\mathbf{\Gamma}$ is the Gamma function and the constants $\beta_{1/2}, \theta_{1/2}, K_{1/2}, a_\alpha$ are given in Proposition 7 and (18) respectively.

Proof. The proof is postponed to Section 4.7. □

Note that for all $x \in \mathbb{R}^d$, $C(\delta_x Q_\gamma^n)$ satisfies the conditions of Proposition 3-(ii). Therefore using in addition the bound on $A(\gamma, x)$ for all $x \in \mathbb{R}^d$ and $\gamma \in \left(0, L^{-1}\right)$ given in Theorem 9, we get $\lim_{k \to +\infty} \|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} = 0$ if $\lim_{n \to +\infty} \Gamma_n = +\infty$ and $\lim_{n \to +\infty} \sum_{k=1}^n \gamma_k^2 < +\infty$.

3.2. Log-concave densities

We now consider the following additional assumption.

H2. *U is convex and admits a minimizer $x^\star$ for U. Moreover there exist $\eta > 0$ and $M_\eta \geq 0$ such that for all $x \in \mathbb{R}^d$, $\|x - x^\star\| \geq M_\eta$,*

$$U(x) - U(x^\star) \geq \eta \|x - x^\star\| \; . \tag{24}$$

It is shown in [2, Lemma 2.2] that if U satisfies **L**1 and is convex, then (24) holds for some constants η, M_η which depend in an intricate way on U. Since the constants η, M_η appear explicitly in the bounds we derive, we must assume that these constants are explicitly computable. We still assume in this section that $U(x^\star) = 0$. Define the function $W_{\mathrm{c}} : \mathbb{R}^d \to [1, +\infty)$ for all $x \in \mathbb{R}^d$ by

$$W_{\mathrm{c}}(x) = \exp((\eta/4)(\|x - x^\star\|^2 + 1)^{1/2}) \, . \tag{25}$$

We now derive a drift inequality for R_γ under **H**2.

Proposition 13. *Assume **L**1 and **H**2. Let $\bar\gamma \in \left(0, L^{-1}\right]$. Then for all $\gamma \in (0, \bar\gamma]$, W_{c} satisfies (6) with $\lambda = \mathrm{e}^{-2^{-4}\eta^2(2^{1/2}-1)}$, $R_{\mathrm{c}} = \max(1, 2d/\eta, M_\eta)$,*

$$c = \{(\eta/4)(d + (\eta\bar\gamma/4)) - \log(\lambda)\} \, \mathrm{e}^{\eta(R_{\mathrm{c}}^2+1)^{1/2}/4 + (\eta\bar\gamma/4)(d+(\eta\bar\gamma/4))} \, . \tag{26}$$

Proof. The proof is postponed to Section 4.8 □

Corollary 14. *Assume **L**1 and **H**2. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq \bar\gamma$, $\bar\gamma \in \left(0, L^{-1}\right]$. Then, for all $n \geq 0$, $p \geq 1$, $n < p$, and $x \in \mathbb{R}^d$,*

$$A(\gamma, x) = L^2 \left(4\eta^{-1} \left[1 + \log\left\{G(\lambda, c, \gamma_1, W_{\mathrm{c}}(x))\right\}\right]\right)^2 \, , \tag{27}$$

where $A(\gamma, x)$ is defined by (11) and G, W_{c}, λ, c, are given in (8), (25), Proposition 13 respectively.

Proof. The proof is postponed to Section 4.9. □

If U is convex, [5, Theorem 1.2] shows that π satisfies a Poincaré inequality with a constant depending only on the variance of π.

Theorem 15. *Assume **L**1 and **H**2. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq \bar\gamma$, $\bar\gamma \in \left(0, L^{-1}\right]$. Then, for all $n \geq 0$, $p \geq 1$, $n < p$, and $x \in \mathbb{R}^d$, (10) holds with $A(\gamma, x)$ given in (27),*

$$\log(\kappa) = \left(-432 \int_{\mathbb{R}^d} \left\| x - \int_{\mathbb{R}^d} y\pi(\mathrm{d}y) \right\|^2 \pi(\mathrm{d}x)\right)^{-1} \tag{28a}$$

$$C(\delta_x Q_\gamma^n) = \left(\frac{(2\pi)^{(d+1)/2}(d-1)!}{\eta^d \mathbf{\Gamma}((d+1)/2)} + \frac{\pi^{d/2} M_\eta^d}{\mathbf{\Gamma}(d/2+1)}\right) D_n(\gamma) \exp(U(x)) \, , \tag{28b}$$

where $D_n(\gamma)$ is given in (23).

Proof. The proof is postponed to Section 4.10. □

For all $x \in \mathbb{R}^d$, $C(\delta_x Q_\gamma^n)$ satisfies the conditions of Proposition 3-(ii). Therefore, if $\lim_{n\to+\infty} \Gamma_n = +\infty$ and $\lim_{n\to+\infty} \sum_{k=1}^n \gamma_k^2 < +\infty$, we get $\lim_{k\to+\infty} \|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} = 0$.

There are two difficulties when applying Theorem 15. First the Poincaré constant (28a) is in closed form but is not computable, although it can be bounded by a $\mathcal{O}(d^{-2})$. Second, the bound of $\mathrm{Var}_\pi\{\mathrm{d}\delta_x Q_\gamma^n/\mathrm{d}\pi\}$ is likely to

be suboptimal. To circumvent these two issues, we now give new quantitative results on the convergence of $(P_t)_{t\geq 0}$ to π in total variation. Instead of using functional inequality, we use in the proof the coupling by reflection, introduced in [27]. Define the function $\omega : (0,1)\times\mathbb{R}_+^* \to \mathbb{R}_+$ for all $\epsilon \in (0,1)$ and $R \geq 0$, by

$$\omega(\epsilon, R) = R^2 / \left\{2\mathbf{\Phi}^{-1}(1-\epsilon/2)\right\}^2 , \tag{29}$$

where $\mathbf{\Phi}$ is the cumulative distribution function of the standard Gaussian distribution and $\mathbf{\Phi}^{-1}$ is the associated quantile function. Before stating the theorem, we first show that (4) holds and provide explicit expressions for the constants which come into play. These constants will be used to obtain the explicit convergence rate of the semigroup $(P_t)_{t\geq 0}$ to π which is derived in Theorem 17.

Proposition 16. *Assume **L**1 and **H**2. Then W_{c} satisfies the drift condition (4) with $\theta = \eta^2/8$, $\mathcal{E} = \mathrm{B}(x^\star, K)$, $K = \max(1, M_\eta, 4d/\eta)$ and*

$$\beta = (\eta/4)\left((\eta/4)K + d\right)\max\left\{1, (K^2+1)^{-1/2}\exp(\eta(K^2+1)^{1/2}/4)\right\} .$$

Proof. The proof is adapted from [2, Corollary 1.6] and is postponed to Section 4.11. □

Theorem 17. *Assume **L**1 and **H**2. Then for all $x \in \mathbb{R}^d$, $\|\delta_x P_t - \pi\|_{\mathrm{TV}} \leq 2\Lambda(x)\mathrm{e}^{-\theta t/4} + 4\varpi^t$, where*

$$\log(\varpi) = -\log(2)(\theta/4) \times \left[\log\left\{\theta^{-1}\beta\left(3 + 4\mathrm{e}^{4^{-1}\theta\omega\left(2^{-1},(8/\eta)\log(4\theta^{-1}\beta)\right)}\right)\right\} + \log(2)\right]^{-1} , \tag{30a}$$

$$\Lambda(x) = (1/2)(W_{\mathrm{c}}(x) + \theta^{-1}\beta) + 2\theta^{-1}\beta\mathrm{e}^{4^{-1}\theta\omega(2^{-1},(8/\eta)\log(4\theta^{-1}\beta))} , \tag{30b}$$

the function W_{c} is defined in (25), the constants θ, β in Proposition 16.

Proof. The proof is postponed to Section 5.1. □

Note that the bound, we obtain is a little different from (3). The initial condition is isolated on purpose to get a better bound. A consequence of this result is the following bound on the convergence of the sequence $(\delta_x Q_\gamma^n)_{n\geq 0}$ to π.

Corollary 18. *Assume **L**1 and **H**2. Let $(\gamma_k)_{k\geq 0}$ be a sequence of nonnegative step sizes. Then for all $x \in \mathbb{R}^d$, $n \geq 0$, $p \geq 1$, $n < p$,*

$$\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \leq 2^{-1/2} L \left(\sum_{k=n}^{p-1}\left\{(\gamma_{k+1}^3/3)A(\gamma, x) + d\gamma_{k+1}^2\right\}\right)^{1/2} + 2\Lambda(\delta_x Q_\gamma^n)\mathrm{e}^{-\theta\Gamma_{n+1,p}/4} + 4\varpi^{\Gamma_{n+1,p}} ,$$

where $A(\gamma, x)$, ϖ are given by (27) and (30a) respectively and

$$\Lambda(\delta_x Q_\gamma^n) = (1/2)(F(\lambda, \Gamma_n, \gamma_1, c, W_{\mathrm{c}}(x)) + \theta^{-1}\beta) + 2\theta^{-1}\beta\mathrm{e}^{4^{-1}\theta\omega(2^{-1},(8/\eta)\log(4\theta^{-1}\beta))} , \tag{31}$$

	d	ε	L
γ	$\mathcal{O}(d^{-3})$	$\mathcal{O}(\varepsilon^2/\log(\varepsilon^{-1}))$	$\mathcal{O}(L^{-2})$
p	$\mathcal{O}(d^5)$	$\mathcal{O}(\varepsilon^{-2}\log^2(\varepsilon^{-1}))$	$\mathcal{O}(L^2)$

TABLE 1
For constant step sizes, dependency of γ and p in d, ε and parameters of U to get $\|\delta_x R_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ using Corollary 19

the functions F and W_{c} are defined in (7) and (25), the constants $\lambda, c, \theta, \beta$ in Proposition 13 and Proposition 16 respectively.

Proof. By Theorem 17, Proposition 13 and Lemma 1, we have for all $x \in \mathbb{R}^d$,

$$\|\delta_x Q_\gamma^n P_{\Gamma_{n+1,p}} - \pi\|_{\mathrm{TV}} \leq 2\Lambda(\delta_x Q_\gamma^n)\mathrm{e}^{-\theta\Gamma_{n+1,p}/4} + 4\varpi^{\Gamma_{n+1,p}} .$$

Finally the proof follows the same line as the one of Proposition 2. □

Contrary to (28b), (31) is uniformly bounded in n. By Corollary 18 and (27), we can apply Proposition 3-(i), which implies the convergence to 0 of $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}}$ as p goes to infinity, if $\lim_{k\to+\infty}\gamma_k = 0$ and $\lim_{k\to+\infty}\Gamma_k = +\infty$. Since $\log(\beta)$ in Proposition 16 is of order d, we get that the rate of convergence $\log(\varpi)$ is of order d^{-2} as d goes to infinity (note indeed that the leading term when d is large is $\theta\omega\left(2^{-1}, (8/\eta)\log(4\theta^{-1}\beta)\right)$ which is of order d^2). In the case of constant step sizes $\gamma_k = \gamma$ for all $k \geq 0$, we adapt Lemma 4 to the bound given by Corollary 18.

Corollary 19. *Assume **L**1 and **H**2. Let $(\gamma_k)_{k\geq 0}$ be a sequence of nonnegative step sizes. Then for all $\varepsilon > 0$, we get $\|\delta_x R_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ if p and γ satisfy (17) with*

$$T = \max\left\{4\theta^{-1}\log\left(8\varepsilon^{-1}\tilde{\Lambda}(x)\right), \log(16\varepsilon^{-1})\Big/(-\log(\varpi))\right\}$$
$$\tilde{\Lambda}(x) = (1/2)(G(\lambda, \gamma_1, c, W_{\mathrm{c}}(x)) + \theta^{-1}\beta) + 2\theta^{-1}\beta\mathrm{e}^{4^{-1}\theta\omega(2^{-1},(8/\eta)\log(4\theta^{-1}\beta))} ,$$

where $A(\gamma, x)$, ϖ are given by (27), (30a) respectively, the functions G and W_{c} are defined in (8) and (25), the constants $\lambda, c, \theta, \beta$ in Proposition 13 and Proposition 16 respectively.

Proof. The proof follows the same line as the one of Lemma 4 using Corollary 18 and that $\sup_{n\geq 0}\Lambda(\delta Q_\gamma^n) < \tilde{\Lambda}(x)$ for all $x \in \mathbb{R}^d$. □

In particular, with the notation of Corollary 19, since $\max(\log(\beta), \log(c))$ and $-(\log(\varpi))^{-1}$ are of order d and d^2 as d goes to infinity respectively, T is of order d^2. Therefore, γ defined by (17) is of order d^{-3} which implies a number of iteration p of order d^5 to get $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ for $\varepsilon > 0$; see also Table 1.

Corollary 19 can be compared with the results which establishes the dependency on the dimension for two kinds of Metropolis-Hastings algorithms to sample from a log-concave density: the random walk Metropolis algorithm (RWM) and the hit-and-run algorithm. It has been shown in [28, Theorem 2.1] that

for $\varepsilon > 0$, the hit-and-run and the RWM reach a ball centered at π, of radius ε for the total variation distance, in a number of iteration p of order d^4 as d goes to infinity. It should be stressed that [28, Theorem 2.1] does not assume any kind of smoothness about the density π contrary to Theorem 17. However, this result assumes that the target distribution is near-isotropic, i.e. there exists $C \geq 0$ which does not depend on the dimension such that for all $x \in \mathbb{R}^d$,

$$C^{-1} \|x\|^2 \leq \int_{\mathbb{R}^d} \langle x, y \rangle^2 \, \pi(\mathrm{d}y) \leq C \|x\|^2 \ .$$

Note that this condition implies that the variance of π is upper bounded by Cd.

To conclude our study on convex potential, we also mention [8] which studies the sampling of the uniform distribution over a convex subset $\mathsf{K} \subset \mathbb{R}^d$ using coupling techniques. Let $C > 0$. A convex set $\mathsf{K} \subset \mathbb{R}^d$ is C-well rounded if $\mathrm{B}(0,1) \subset \mathsf{K} \subset \mathrm{B}(0, Cd)$. [8] shows that a number of iteration of order d^9 as d goes to infinity is sufficient to sample uniformly over any C-well rounded convex set. Comparison with our result is difficult since we assume that π is positive on $\mathbb{R}^d$, continuously differentiable, while [8] studies the case of uniform distributions over a convex body. An adaptation of our result to non continuously differentiable potentials will appear in a forthcoming paper [14].

3.3. Strongly log-concave densities

More precise bounds can be obtained in the case where U is assumed to be strongly convex outside some ball; this assumption has been considered by [15] for convergence in the Wasserstein distance; see also [6].

H 3 (M_{s}). *U is convex and there exist $M_{\mathrm{s}} \geq 0$ and $m > 0$, such that for all $x, y \in \mathbb{R}^d$ satisfying $\|x - y\| \geq M_{\mathrm{s}}$,*

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

We will see in the sequel that under this assumption the convergence rate in (3) does not depend on the dimension d but only on the constants m and M_{s}.

Proposition 20. *Assume **L** 1 and **H** 3(M_{s}). Let $\bar{\gamma} \in \left(0, 2mL^{-2}\right)$. For all $\gamma \in (0, \bar{\gamma}]$, $V(x) = \|x - x^\star\|^2$ satisfies (6) with $\lambda = \mathrm{e}^{-2m + \bar{\gamma} L^2}$ and $\mathrm{c} = 2(d + mM_{\mathrm{s}}^2)$.*

Proof. The proof is postponed to Section 4.12. □

Theorem 21. *Assume **L** 1 and **H** 3(M_{s}). Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq \bar{\gamma}$, $\bar{\gamma} \in \left(0, 2mL^{-2}\right)$. Then, for all $n \geq 0$, $p \geq 1$, $n < p$, and $x \in \mathbb{R}^d$, (10) holds with*

$$\begin{aligned}
\log(\kappa) &= -(m/2)\log(2) \\
&\quad \times \left[\log\left\{\left(1 + \mathrm{e}^{m\omega(2^{-1}, \max(1, M_{\mathrm{s}}))/4}\right)(1 + \max(1, M_{\mathrm{s}}))\right\} + \log(2)\right]^{-1} \\
C(\delta_x Q_\gamma^n) &\leq 6 + 2\left(d/m + M_{\mathrm{s}}^2\right)^{1/2} + 2F^{1/2}(\lambda, \Gamma_{1,n}, \mathrm{c}, \gamma_1, \|x - x^\star\|^2) \\
A(\gamma, x) &\leq L^2 \, G(\lambda, \mathrm{c}, \gamma_1, \|x - x^\star\|^2) \ ,
\end{aligned}$$

	d	ε	L	m	M_{s}
γ	$\mathcal{O}(d^{-1})$	$\mathcal{O}(\varepsilon^2/\log(\varepsilon^{-1}))$	$\mathcal{O}(L^{-2})$	$\mathcal{O}(m)$	$\mathcal{O}(M_{\mathrm{s}}^{-4})$
p	$\mathcal{O}(d\log(d))$	$\mathcal{O}(\varepsilon^{-2}\log^2(\varepsilon^{-1}))$	$\mathcal{O}(L^2)$	$\mathcal{O}(m^{-2})$	$\mathcal{O}(M_{\mathrm{s}}^{8})$

TABLE 2
For constant step sizes, dependency of γ and p in d, ε and parameters of U to get $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ using Theorem 21

where F, G, ω are defined by (7), (8), (29) respectively, and λ, c are given in Proposition 20.

Proof. The proof is postponed to Section 5.1. □

Note that the conditions of Proposition 3-(i) are fulfilled. For constant step sizes $\gamma_k = \gamma$ for all $k \geq 1$, Lemma 4 and Lemma 5 can be applied. We give in Table 2 the dependency of the step size $\gamma > 0$ and the minimum number of iterations $p \geq 0$, provided in Lemma 4, on the dimension d and the other constants related to U, to get $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$, for a target precision $\varepsilon > 0$. We can see that the dependency on the dimension is milder than for the convex case. The number of iteration requires to reach a target precision ε is just of order $\mathcal{O}(d\log(d))$.

Consider the case where π is the d-dimensional standard Gaussian distribution. Then for all $p \in \mathbb{N}$, $\gamma \in (0,1)$ and $x \in \mathbb{R}^d$, $\delta_x R_\gamma^p$ is the d-dimensional Gaussian distribution with mean $(1-\gamma)^p x$ and covariance matrix $\sigma_\gamma \mathrm{I}_d$, with $\sigma_\gamma = (1-(1-\gamma)^{2(p+1)})(1-\gamma/2)^{-1}$. Therefore using the Pinsker inequality, we get:

$$\begin{aligned}\|\delta_x R_\gamma^p - \pi\|_{\mathrm{TV}}^2 &\leq 2\,\mathrm{KL}\left(\delta_x R_\gamma^p \middle| \pi\right) \\ &\leq d\left[\log\left(\sigma_\gamma\right) - 1 + \sigma_\gamma^{-1}\left\{1 + (1-\gamma)^{2p}\|x\|^2 d^{-1}\right\}\right] .\end{aligned}$$

Using the inequalities for all $t \in (0,1)$, $(1-t)^{-1} \leq 1 + t(1-t)^{-2}$ and for all $s \in (0,1/2)$, $-\log(1-s) \leq s + 2s^2$, we have:

$$\begin{aligned}\|\delta_x R_\gamma^p - \pi\|_{\mathrm{TV}}^2 \leq d\Big\{\gamma^2/2 + (1-\gamma)^{2(p+1)}(1-\gamma/2)(1-(1-\gamma)^{2(p+1)})^{-2}\Big\} \\ + \sigma_\gamma^{-1}(1-\gamma)^{2p}\|x\|^2 .\end{aligned}$$

This inequality implies that in order to have $\|\delta_x R_\gamma - \pi\|_{\mathrm{TV}} \leq \varepsilon$ for $\varepsilon > 0$, the step size γ has to be of order $d^{-1/2}$ and p of order $d^{1/2}\log(d)$. Therefore, the dependency on the dimension reported in Table 2 does not match this particular example. However it does not imply that this dependency can be improved.

3.4. Bounded perturbation of strongly log-concave densities

We now consider the case where U is a bounded perturbation of a strongly convex potential.

H4. *The potential U may be expressed as $U = U_1 + U_2$, where*

*(a) $U_1 : \mathbb{R}^d \to \mathbb{R}$ satisfies **H**3(0) (i.e. is strongly convex) and there exists $L_1 \geq 0$ such that for all $x, y \in \mathbb{R}^d$,*

$$\|\nabla U_1(x) - \nabla U_1(y)\| \leq L_1 \|x - y\| \ .$$

(b) $U_2 : \mathbb{R}^d \to \mathbb{R}$ is continuously differentiable and $\|U_2\|_\infty + \|\nabla U_2\|_\infty < +\infty$.

The probability measure π is said to satisfy a log-Sobolev inequality with constant $C_{\mathrm{LS}} > 0$ if for all locally Lipschitz function $h : \mathbb{R}^d \to \mathbb{R}$, we have

$$\mathrm{Ent}_\pi\left(h^2\right) \leq 2C_{\mathrm{LS}} \int \|\nabla h\|^2 \, \mathrm{d}\pi \ .$$

Then [9, Theorem 2.7] shows that for all $t \geq 0$ and any probability measure $\mu_0 \ll \pi$ satisfying $\mathrm{d}\mu_0/\mathrm{d}\pi \log(\mathrm{d}\mu_0/\mathrm{d}\pi) \in \mathrm{L}^1(\pi)$, we have

$$\|\mu_0 P_t - \pi\|_{\mathrm{TV}} \leq \mathrm{e}^{-t/C_{\mathrm{LS}}} \left\{ 2 \, \mathrm{Ent}_\pi \left(\frac{\mathrm{d}\mu_0}{\mathrm{d}\pi} \right) \right\}^{1/2} \ . \tag{32}$$

Under **H**4, [4, Corollary 5.7.2] and the Holley-Stroock perturbation principle [19, p. 1184], π satisfies a log-Sobolev inequality with a constant which only depends on the strong convexity constant m of U_1 and $\mathrm{osc}_{\mathbb{R}^d}(U_2)$. Define

$$\varpi = \frac{2mL_1}{m + L_1} \ . \tag{33}$$

Denote by $x_1^\star$ the minimizer of U_1.

Proposition 22. *Assume **H**4. Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m + L_1)$. Then for all $p \geq 1$ and $x \in \mathbb{R}^d$,*

$$\int_{\mathbb{R}^d} \|y - x_1^\star\|^2 \, Q_\gamma^p(x, \mathrm{d}y) \leq \prod_{k=1}^{p} (1 - \varpi\gamma_k/2) \, \|x - x_1^\star\|^2 + 2\varpi^{-1}(2d + (\gamma_1 + 2\varpi^{-1}) \, \|\nabla U_2\|_\infty^2) \ .$$

Proof. The proof is postponed to Section 4.13. □

Theorem 23. *Assume **L**1 and **H**4. Let $(\gamma_k)_{k \in \mathbb{N}^*}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m + L_1)$. Then, for all $n, p \geq 1$, $n < p$, and $x \in \mathbb{R}^d$, (10) holds with $-\log(\kappa) = m \exp\{-\mathrm{osc}_{\mathbb{R}^d}(U_2)\}$ and*

$$C^2(\delta_x Q_\gamma^n) \leq L_1 \mathrm{e}^{-\varpi\Gamma_n/2} \, \|x - x_1^\star\|^2 + L_1\gamma_n(\gamma_n + 2\varpi^{-1}) \, \|\nabla U_2\|_\infty^2 + 2\mathrm{osc}_{\mathbb{R}^d}(U_2) + 2L_1\varpi^{-1}(1 - \varpi\gamma_n)(2d + (\gamma_1 + 2\varpi^{-1}) \, \|\nabla U_2\|_\infty^2) - d(1 + \log(2\gamma_n m) - 2L_1\gamma_n)$$

$$A(\gamma, x) \leq 2L_1^2 \left\{ \|x_1^\star - x^\star\|^2 + 2\varpi^{-1}(2d + (\gamma_1 + 2\varpi^{-1}) \, \|\nabla U_2\|_\infty^2) \right\} + 2 \, \|\nabla U_2\|_\infty^2 \ ,$$

where ϖ is defined in (33).

Proof. The proof is postponed to Section 4.14. □

Note that $\sup_{n\geq 1}\{C(\delta_x Q_\gamma^n)/(-\log(\gamma_n))\} < +\infty$, therefore Proposition 3-(i) can be applied and $\lim_{p\to+\infty} \|\delta_\gamma Q_\gamma^p - \pi\|_{\mathrm{TV}} = 0$ if $\lim_{k\to+\infty}\gamma_k = 0$ and $\lim_{k\to+\infty}\Gamma_k = +\infty$.

4. Proofs

4.1. Proof of Lemma 1

By a straightforward induction, we get for all $n \geq 0$ and $x \in \mathbb{R}^d$,

$$Q_\gamma^n V(x) \leq \lambda^{\Gamma_n} V(x) + c\sum_{i=1}^n \gamma_i \lambda^{\Gamma_{i+1,n}} \,. \tag{34}$$

Note that for all $n \geq 1$, we have since $(\gamma_k)_{k\geq 1}$ is nonincreasing and for all $t \geq 0$, $\lambda^t = 1 + \int_0^t \lambda^s \log(\lambda)\mathrm{d}s$,

$$\begin{aligned}
&\sum_{i=1}^n \gamma_i \lambda^{\Gamma_{i+1,n}} \leq \sum_{i=1}^n \gamma_i \prod_{j=i+1}^n (1+\lambda^{\gamma_1}\log(\lambda)\gamma_j) \\
&\leq (-\lambda^{\gamma_1}\log(\lambda))^{-1} \sum_{i=1}^n \gamma_i \left\{ \prod_{j=i+1}^n (1+\lambda^{\gamma_1}\log(\lambda)\gamma_j) - \prod_{j=i}^n (1+\lambda^{\gamma_1}\log(\lambda)\gamma_j) \right\} \\
&\leq (-\lambda^{\gamma_1}\log(\lambda))^{-1} \,.
\end{aligned}$$

The proof is then completed using this inequality in (34).

4.2. Proof of Lemma 6

By **L**1, **H**1, the Cauchy-Schwarz inequality and $\nabla U(x^\star) = 0$, for all $x \in \mathbb{R}^d$, $\|x\| \geq M_\rho$, we have

$$\begin{aligned}
U(x) - U(x^\star) &= \int_0^1 \langle \nabla U(x^\star + t(x-x^\star)), x - x^\star\rangle \,\mathrm{d}t \\
&\geq \int_0^{\frac{M_\rho}{\|x-x^\star\|}} \langle \nabla U(x^\star + t(x-x^\star)), x - x^\star\rangle \,\mathrm{d}t \\
&\quad + \int_{\frac{M_\rho}{\|x-x^\star\|}}^1 \langle \nabla U(x^\star + t(x-x^\star)), t(x - x^\star)\rangle \,\mathrm{d}t \\
&\geq -M_\rho^2 L/2 + \rho \|x-x^\star\|^\alpha (\alpha+1)^{-1} \left\{1 - (M_\rho/\|x-x^\star\|)^{\alpha+1}\right\} \,.
\end{aligned}$$

On the other hand using again **L**1, the Cauchy-Schwarz inequality and $\nabla U(x^\star) = 0$, for all $x \in \mathrm{B}(x^\star, M_\rho)$,

$$U(x) - U(x^\star) = \int_0^1 \langle \nabla U(x^\star + t(x-x^\star)), x - x^\star\rangle \,\mathrm{d}t \geq -M_\rho^2 L/2 \,,$$

which concludes the proof.

4.3. Proof of Proposition 7

For all $x \in \mathbb{R}^d$, we have

$$\mathscr{A}^{\mathrm{L}} V_\varsigma(x) = \varsigma(1-\varsigma)\left\{-\|\nabla U(x)\|^2 + (1-\varsigma)^{-1}\Delta U(x)\right\} V_\varsigma(x) \; .$$

If $\alpha > 1$, by the Cauchy-Schwarz inequality, under **L** 1-**H** 1 for all $x \in \mathbb{R}^d$, $\Delta U(x) \leq dL$ and $\|\nabla U(x)\| \geq \rho \|x - x^\star\|^{\alpha-1}$ for $\|x - x^\star\| \geq M_\rho$. Then, for all $x \notin \mathcal{E}_\varsigma$,

$$\mathscr{A}^{\mathrm{L}} V_\varsigma(x) \leq \varsigma(1-\varsigma)\left\{-\rho\|x - x^\star\|^{2(\alpha-1)} + (1-\varsigma)^{-1} dL\right\} V_\varsigma(x) \leq -\varsigma dL V_\varsigma(x) \; ,$$

and $\sup_{\{x \in \mathcal{E}_\varsigma\}} \mathscr{A}^{\mathrm{L}} V_\varsigma(x) \leq \varsigma dL \sup_{\{y \in \mathcal{E}_\varsigma\}} \{V_\varsigma(y)\}$.

4.4. Proof of Proposition 8

By **H**1, for all $x \notin \mathrm{B}(x^\star, M_\rho)$,

$$\|\nabla U(x)\| \geq \rho \|x - x^\star\|^{\alpha-1} \; . \tag{35}$$

Since under **L**1, for all $x, y \in \mathbb{R}^d$, $U(y) \leq U(x) + \langle \nabla U(x), y - x\rangle + (L/2)\|y - x\|^2$, we have for all $\gamma \in (0, \bar{\gamma})$ and $x \in \mathbb{R}^d$,

$$\begin{aligned}
&R_\gamma V(x)/V(x) \\
&= (4\pi\gamma)^{-d/2} \int_{\mathbb{R}^d} \exp\left(\{U(y) - U(x)\}/2 - (4\gamma)^{-1}\|y - x + \gamma\nabla U(x)\|^2\right) \mathrm{d}y \\
&\leq (4\pi\gamma)^{-d/2} \int_{\mathbb{R}^d} \exp\left(-4^{-1}\gamma\|\nabla U(x)\|^2 - (4\gamma)^{-1}(1 - \gamma L)\|y - x\|^2\right) \mathrm{d}y \\
&\leq (1 - \gamma L)^{-d/2} \exp(-4^{-1}\gamma\|\nabla U(x)\|^2) \; ,
\end{aligned}$$

where we used in the last line that $\gamma < L^{-1}$. Since $\log(1 - L\gamma) = -L\int_0^\gamma (1 - Lt)^{-1}\mathrm{d}t$, for all $\gamma \in (0, \bar{\gamma}]$, $\log(1 - L\gamma) \geq -L\gamma(1 - L\bar{\gamma})^{-1}$. Using this inequality, we get

$$R_\gamma V(x)/V(x) \leq \lambda^{-\gamma} \exp\left(-4^{-1}\gamma\|\nabla U(x)\|^2\right) \; . \tag{36}$$

By (35), for all $x \in \mathbb{R}^d$, $\|x - x^\star\| \geq K$, we have

$$R_\gamma V(x) \leq \lambda^\gamma V(x) \; . \tag{37}$$

Also by (36) and since for all $t \geq 0$, $\mathrm{e}^t - 1 \leq t\mathrm{e}^t$, we get for all $x \in \mathbb{R}^d$

$$R_\gamma V(x) - \lambda^\gamma V(x) \leq \lambda^\gamma(\lambda^{-2\gamma} - 1)V(x) \leq -2\gamma\log(\lambda)\lambda^{-\bar{\gamma}} V(x) \; .$$

The proof is completed combining the last inequality and (37).

4.5. Proof of Theorem 9

We first bound $A(\gamma, x)$ for all $x \in \mathbb{R}^d$. Let $x \in \mathbb{R}^d$. By **L**1, we have $\mathbb{E}_x[\|\nabla U(X_k)\|^2] \leq L^2 \mathbb{E}_x[\|X_k - x^\star\|^2]$. Consider now the function $\phi_\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ defined for all $t \geq 0$ by $\phi_\alpha(t) = \exp(A_\alpha(t + B_\alpha)^{\alpha/2})$ where $A_\alpha = \rho/(2(\alpha+1))$ and $B_\alpha = \{(2-\alpha)/(\alpha A_\alpha)\}^{2/\alpha}$. Since ϕ_α is convex and invertible on $\mathbb{R}_+$, we get using the Jensen inequality and Lemma 6 for all $k \geq 0$:

$$\mathbb{E}_x[\|X_k - x^\star\|^2] \leq \phi_\alpha^{-1}\left(\mathbb{E}_x[\phi_\alpha\left(\|X_k - x^\star\|^2\right)]\right) \leq \phi_\alpha^{-1}\left(\mathrm{e}^{a_\alpha/2 + B_\alpha^{\alpha/2}} \mathbb{E}_x[V(X_k)]\right) ,$$

where $V(x) = \exp(U(x)/2)$. Using that for all $t \geq 0$, $\phi_\alpha^{-1}(t) \leq (A_\alpha^{-1}\log(t))^{2/\alpha}$ and Lemma 1, we get

$$\sup_{k\geq 0} \mathbb{E}_x[\|X_k - x^\star\|^2] \leq \left(A_\alpha^{-1}\left[a_\alpha/2 + B_\alpha^{\alpha/2} + \log\left\{G(\lambda, c(\gamma_1), V(x))\right\}\right]\right)^{2/\alpha} .$$

Eq. (19) follows from Proposition 7, Proposition 8 and Lemma 1.

4.6. Proof of Theorem 10

Lemma 24. *Let μ and ν be two probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $V : \mathbb{R}^d \to [1, \infty)$ be a measurable function. Then*

$$\|\mu - \nu\|_V \leq \sqrt{2}\left\{\nu(V^2) + \mu(V^2)\right\}^{1/2} \mathrm{KL}^{1/2}(\mu|\nu) .$$

Proof. Without losing any generality, we assume that $\mu \ll \nu$. For all $t \in [0,1]$, $t\log(t) - t + 1 = \int_t^1 (u-t)u^{-1}\mathrm{d}u \geq 2^{-1}(1-t)^2$, and on $[1, +\infty)$, $t \mapsto 2(1+t)(t\log(t) - t + 1) - (1-t)^2$ is nonincreasing. Therefore, for all $t \geq 0$,

$$|1 - t| \leq (2(1+t)(t\log(t) - t + 1))^{1/2} . \tag{38}$$

Then, we have:

$$\begin{aligned}\|\mu - \nu\|_V &= \sup_{f \in \mathbb{F}(\mathbb{R}^d), \|f\|_V \leq 1} \left| \int_{\mathbb{R}^d} f(x)\mathrm{d}\mu(x) - \int_{\mathbb{R}^d} f(x)\mathrm{d}\nu(x) \right| \\ &= \sup_{f \in \mathbb{F}(\mathbb{R}^d), \|f\|_V \leq 1} \left| \int_{\mathbb{R}^d} f(x) \left\{\frac{\mathrm{d}\mu}{\mathrm{d}\nu} - 1\right\} \mathrm{d}\nu(x) \right| \leq \int_{\mathbb{R}^d} V(x) \left| \frac{\mathrm{d}\mu}{\mathrm{d}\nu} - 1 \right| \mathrm{d}\nu(x) .\end{aligned}$$

Using (38) and the Cauchy-Schwarz inequality in the previous inequality concludes the proof. □

Proof of Theorem 10. First note that by the triangle inequality and Proposition 7, for all $p \geq 1$

$$\|\pi - \delta_x Q_\gamma^p\|_{V^{1/2}} \leq C_{1/4}\kappa^{p\gamma} V^{1/2}(x) + \|\delta_x P_{\Gamma_p} - \delta_x Q_\gamma^p\|_{V^{1/2}} . \tag{39}$$

We now bound the second term of the right hand side. Let $\mathrm{k}_\gamma = \lceil \gamma^{-1} \rceil$ and q_γ and r_γ be respectively the quotient and the remainder of the Euclidean division of p by k_γ. The triangle inequality implies $\|\delta_x P_{\Gamma_p} - \delta_x Q_\gamma^p\|_{V^{1/2}} \leq A + B$ with

$$A = \left\| \delta_x Q_\gamma^{(\mathrm{q}_\gamma - 1)\mathrm{k}_\gamma} P_{\Gamma_{(\mathrm{q}_\gamma-1)\mathrm{k}_\gamma, p}} - \delta_x Q_\gamma^{(\mathrm{q}_\gamma-1)\mathrm{k}_\gamma} Q_\gamma^{(\mathrm{q}_\gamma-1)\mathrm{k}_\gamma+1,p} \right\|_{V^{1/2}}$$
$$B = \sum_{i=1}^{\mathrm{q}_\gamma} \left\| \delta_x Q_\gamma^{(i-1)\mathrm{k}_\gamma} P_{\Gamma_{(i-1)\mathrm{k}_\gamma+1,p}} - \delta_x Q_\gamma^{i\mathrm{k}_\gamma} P_{\Gamma_{i\mathrm{k}_\gamma+1,p}} \right\|_{V^{1/2}} .$$

It follows from Proposition 7 and $\mathrm{k}_\gamma \geq \gamma^{-1}$ that

$$B \leq \sum_{i=1}^{\mathrm{q}_\gamma} C_{1/4} \kappa^{\mathrm{q}_\gamma - i} \left\| \delta_x Q_\gamma^{(i-1)\mathrm{k}_\gamma} P_{\Gamma_{(i-1)\mathrm{k}_\gamma+1,i\mathrm{k}_\gamma}} - \delta_x Q_\gamma^{i\mathrm{k}_\gamma} \right\|_{V^{1/2}} . \tag{40}$$

We now bound each term of the sum in the right hand side. For all initial distribution ν_0 on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $i, j \geq 1$, $i < j$, it follows from Lemma 24, [22, Theorem 4.1, Chapter 2] and (15):

$$\begin{aligned}
&\|\nu_0 Q_\gamma^{i,j} - \nu_0 P_{\Gamma_{i,j}}\|_{V^{1/2}}^2 \leq 2 \left(\nu_0 Q_\gamma^{i,j}(V) + \nu_0 P_{\Gamma_{i,j}}(V)\right) \mathrm{KL}(\nu_0 Q_\gamma^{i,j} | \nu_0 P_{\Gamma_{i,j}}) \\
&\leq 2L^2 \left(\nu_0 Q_\gamma^{i,j}(V) + \nu_0 P_{\Gamma_{i,j}}(V)\right) \\
&\qquad \times (j-i) \left(\gamma^2 d + (\gamma^3/3) \sup_{k \in \{i, \cdots, j\}} \nu_0 Q_\gamma^{i,k-1}(\|\nabla U\|^2) \right) .
\end{aligned}$$

Proposition 7 implies by the proof of [31, Theorem 6.1] that for all $y \in \mathbb{R}^d$ and $t \geq 0$: $P_t V(y) \leq V(y) + \beta_{1/2}/\theta_{1/2}$. Then, using Proposition 8, Lemma 1 and $\mathrm{k}_\gamma \geq \gamma^{-1}$ in (40), we get

$$\begin{aligned}
&\sup_{i \in \{1, \cdots, \mathrm{q}_\gamma\}} \left\| \delta_x Q_\gamma^{(i-1)\mathrm{k}_\gamma} P_{\Gamma_{(i-1)\mathrm{k}_\gamma+1,i\mathrm{k}_\gamma}} - \delta_x Q_\gamma^{i\mathrm{k}_\gamma} \right\|_{V^{1/2}}^2 \\
&\qquad \leq 2^{-1}(1+\gamma) L^2 \left\{ 2G(\lambda, c, V(x)) + \beta_{1/2}/\theta_{1/2} \right\} \\
&\qquad\qquad \times \left\{ \gamma d + 3^{-1} \gamma^2 \|\nabla U\|_{V^{1/2}}^2 \, G(\lambda, c, V(x)) \right\} .
\end{aligned}$$

Finally, A can be bounded along the same lines. □

4.7. *Proof of Theorem 12*

Denote for $\gamma > 0$, $r_\gamma : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ the transition density of R_γ defined for $x, y \in \mathbb{R}^d$ by

$$r_\gamma(x, y) = (4\pi\gamma)^{-1} \exp(-(4\gamma)^{-1} \|y - x + \gamma \nabla U(x)\|^2) . \tag{41}$$

For all $n \geq 1$, denote by $q_\gamma^n : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ the transition density associated with Q_γ^n defined by induction by: for all $x, y \in \mathbb{R}^d$

$$q_\gamma^1(x, y) = r_{\gamma_1}(x, y) , \qquad q_\gamma^{n+1}(x, y) = \int_{\mathbb{R}^d} q_\gamma^n(x, z) r_{\gamma_{n+1}}(z, y) \mathrm{d}z \text{ for } n \geq 1 . \tag{42}$$

Lemma 25. *Assume* ***L****1. Let* $(\gamma_k)_{k\geq 1}$ *be a nonincreasing sequence with* $\gamma_1 < L$. *Then for all* $n \geq 1$ *and* $x, y \in \mathbb{R}^d$,

$$q_\gamma^n(x,y) \leq \frac{\exp\left(2^{-1}(U(x)-U(y)) - (2\sigma_{\gamma,n})^{-1}\|y-x\|^2\right)}{(2\pi\sigma_{\gamma,n}\prod_{i=1}^n(1-L\gamma_i))^{d/2}} ,$$

where $\sigma_{\gamma,n} = \sum_{i=1}^n 2\gamma_i(1-L\gamma_i)^{-1}$.

Proof. Under **L**1, we have for all $x, y \in \mathbb{R}^d$, $U(y) \leq U(x) + \langle \nabla U(x), y-x\rangle + (L/2)\|y-x\|^2$, which implies that for all $\gamma \in \left(0, L^{-1}\right)$

$$r_\gamma(x,y) \leq (4\pi\gamma)^{-d/2}\exp\left(2^{-1}(U(x)-U(y)) - (1-L\gamma)(4\gamma)^{-1}\|y-x\|^2\right) . \tag{43}$$

Then, the proof of the claimed inequality is by induction. By (43), the inequality holds for $n = 1$. Now assume that it holds for $n \geq 1$. By induction hypothesis and (43) applied for $\gamma = \gamma_{n+1}$, we have

$$\begin{aligned}
q_\gamma^{n+1}(x,y) &\leq (4\pi\gamma_{n+1})^{-d/2}\left\{2\pi\sigma_{\gamma,n}\prod_{i=1}^n(1-L\gamma_i)\right\}^{-d/2}\exp\left(2^{-1}(U(x)-U(y))\right) \\
&\quad\times\int_{\mathbb{R}^d}\exp\left(-(2\sigma_{\gamma,n})^{-1}\|z-x\|^2 - (1-L\gamma_{n+1})(4\gamma_{n+1})^{-1}\|z-y\|^2\right)\mathrm{d}z \\
&\leq (4\pi\gamma_{n+1})^{-d/2}\left\{2\pi\sigma_{\gamma,n}\prod_{i=1}^n(1-L\gamma_i)\right\}^{-d/2}(\sigma_{\gamma,n}^{-1} + (1-L\gamma_{n+1})/(2\gamma_{n+1}))^{-d/2} \\
&\quad\times(2\pi)^{d/2}\exp\left(2^{-1}(U(x)-U(y)) - (2\sigma_{\gamma,n+1})^{-1}\|y-x\|^2\right) .
\end{aligned}$$

Rearranging terms in the last inequality concludes the proof. □

Lemma 26. *Assume* ***L****1 and* ***H****1. Then* $\int_{\mathbb{R}^d}\mathrm{e}^{-U(y)}\mathrm{d}y \leq \vartheta_U$ *where*

$$\vartheta_U \overset{def}{=} \mathrm{e}^{a_\alpha}\frac{(2\pi)^{(d+1)/2}(d-1)!}{\eta^d\boldsymbol{\Gamma}((d+1)/2)} , \tag{44}$$

and a_α *is given in* (18).

Proof. By Lemma 6, for all $x \in \mathbb{R}^d$, $U(x) \geq \rho\|x-x^\star\|/(\alpha+1) - a_\alpha$. Using the spherical coordinates, we get

$$\int_{\mathbb{R}^d}\mathrm{e}^{-U(y)}\mathrm{d}y \leq \mathrm{e}^{a_\alpha}\left\{(2\pi)^{(d+1)/2}/\boldsymbol{\Gamma}((d+1)/2)\right\}\int_0^{+\infty}\mathrm{e}^{-\rho t/(\alpha+1)}t^{d-1}\mathrm{d}t .$$

Then the proof is concluded by a straightforward calculation. □

Corollary 27. *Assume* **L**1 *and* **H**1. *Let* $(\gamma_k)_{k\geq 1}$ *be a nonincreasing sequence with* $\gamma_1 < L$. *Then for all* $n \geq 1$ *and* $x \in \mathbb{R}^d$,

$$\mathrm{Var}_\pi \left\{ \frac{\mathrm{d}\delta_x Q_\gamma^n}{\mathrm{d}\pi} \right\} \leq (\vartheta_U \exp(U(x))) \left(4\pi \left\{ \prod_{k=1}^n (1 - L\gamma_k) \right\}^2 \sum_{i=1}^n \frac{\gamma_i}{1 - L\gamma_i} \right)^{-d/2} ,$$

where ϑ_U *is given by* (44).

Proof of Theorem 12. We bound the two terms of the right hand side of (10). The first term is dealt with the same reasoning as for the proof of Theorem 9. Regarding the second term, by [2, Theorem 1.4], π satisfies a Poincaré inequality with constant $\log^{-1}(\kappa)$. Then, the claimed bound follows from (22) and Corollary 27. □

4.8. Proof of Proposition 13

Set $\chi = \eta/4$ and for all $x \in \mathbb{R}^d$, $\phi(x) = (\|x - x^\star\|^2 + 1)^{1/2}$. Since ϕ is 1-Lipschitz, we have by the log-Sobolev inequality [7, Theorem 5.5] for all $x \in \mathbb{R}^d$,

$$R_\gamma W_{\mathrm{c}}(x) \leq \mathrm{e}^{\chi R_\gamma \phi(x) + \chi^2 \gamma} \leq \mathrm{e}^{\chi \sqrt{\|x - \gamma \nabla U(x) - x^\star\|^2 + 2\gamma d + 1} + \chi^2 \gamma} . \tag{45}$$

Under **L**1 since U is convex and $x^\star$ is a minimizer of U, [33, Theorem 2.1.5 Equation (2.1.7)] shows that for all $x \in \mathbb{R}^d$,

$$\langle \nabla U(x), x - x^\star \rangle \geq (2L)^{-1} \|\nabla U(x)\|^2 + \eta \|x - x^\star\| 1_{\{\|x - x^\star\| \geq M_\eta\}} ,$$

which implies that for all $x \in \mathbb{R}^d$ and $\gamma \in \left(0, L^{-1}\right]$, we have

$$\|x - \gamma \nabla U(x) - x^\star\|^2 \leq \|x - x^\star\|^2 - 2\gamma\eta \|x - x^\star\| 1_{\{\|x - x^\star\| \geq M_\eta\}} . \tag{46}$$

Using this inequality and for all $u \in [0, 1]$, $(1 - u)^{1/2} - 1 \leq -u/2$, we have for all $x \in \mathbb{R}^d$, satisfying $\|x - x^\star\| \geq R_{\mathrm{c}} = \max(1, 2d\eta^{-1}, M_\eta)$,

$$\begin{aligned}
&\left(\|x - \gamma \nabla U(x) - x^\star\|^2 + 2\gamma d + 1 \right)^{1/2} - \phi(x) \\
&\quad \leq \phi(x) \left\{ \left(1 - 2\gamma \phi^{-2}(x)(\eta \|x - x^\star\| - d)\right)^{1/2} - 1 \right\} \\
&\quad \leq -\gamma \phi^{-1}(x)(\eta \|x - x^\star\| - d) \leq -(\eta\gamma/2) \|x - x^\star\| \phi^{-1}(x) \leq -2^{-3/2} \eta\gamma .
\end{aligned}$$

Combining this inequality and (45), we get for all $x \in \mathbb{R}^d$, $\|x - x^\star\| \geq R_{\mathrm{c}}$,

$$R_\gamma W_{\mathrm{c}}(x)/W_{\mathrm{c}}(x) \leq \mathrm{e}^{\gamma\chi(\chi - 2^{-3/2}\eta)} = \lambda^\gamma .$$

By (46) and the inequality for all $a, b \geq 0$, $\sqrt{a + 1 + b} - \sqrt{1 + b} \leq a/2$, we get for all $x \in \mathbb{R}^d$,

$$\sqrt{\|x - \gamma \nabla U(x) - x^\star\|^2 + 2\gamma d + 1} - \phi(x) \leq \gamma d .$$

Then using this inequality in (45), we have for all $x \in \mathbb{R}^d$,

$$R_\gamma W_{\mathrm{c}}(x) \leq \lambda^\gamma W_{\mathrm{c}}(x) + \left(\mathrm{e}^{\chi\gamma(d+\chi)} - \lambda^\gamma\right) \mathrm{e}^{\eta(R_{\mathrm{c}}^2+1)^{1/2}/4} 1_{\mathrm{B}(x^\star,R_{\mathrm{c}})}(x) \; .$$

Using the inequality for all $t \geq 0$, $\mathrm{e}^t - 1 \leq t\mathrm{e}^t$ concludes the proof.

4.9. Proof of Corollary 14

We preface the proof by a Lemma.

Lemma 28. *Assume* **L**1 *and that* U *is convex. Let* $(\gamma_k)_{k\in\mathbb{N}^*}$ *be a nonincreasing sequence with* $\gamma_1 \leq L^{-1}$. *For all* $n \geq 0$ *and* $x \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} \|y - x^\star\|^2 \, Q_\gamma^n(x, \mathrm{d}y) \leq \left\{4\eta^{-1}\left[1 + \log\left\{G(\lambda, c, \gamma_1, W_{\mathrm{c}}(x))\right\}\right]\right\}^2 \; ,$$

where $W_{\mathrm{c}}, \lambda, c$ *are given in* (25) *and Proposition 13 respectively.*

Proof. Let $n \geq 0$ and $x \in \mathbb{R}^d$. Consider the function $\phi : \mathbb{R} \to \mathbb{R}$ defined by for all $t \in \mathbb{R}$, $\phi(t) = \exp\left\{(\eta/4)(t + (4/\eta)^2)^{1/2}\right\}$. Since this function is convex on $\mathbb{R}_+$, we have by the Jensen inequality and the inequality for all $t \geq 0$, $\phi(t) \leq \mathrm{e}^{1+(\eta/4)(t+1)^{1/2}}$,

$$\phi\left(\int_{\mathbb{R}^d} \|y - x^\star\|^2 \, Q_\gamma^n(x, \mathrm{d}y)\right) \leq \mathrm{e}^1 Q_\gamma^n W_{\mathrm{c}}(x) \; .$$

The proof is then completed using Proposition 13, Lemma 1 and that ϕ is one-to-one with for all $t \geq 1$, $\phi^{-1}(t) \leq \left(4\eta^{-1}\log(t)\right)^2$. □

Proof of Corollary 14. Using $\nabla U(x^\star) = 0$, **L**1 and Lemma 28, we have for all $k \geq 0$,

$$\int_{\mathbb{R}^d} \|\nabla U(y)\|^2 \, Q_\gamma^k(x, \mathrm{d}y) \leq L^2 \left(4\eta^{-1}\left\{1 + \log\left\{G(\lambda, c, \gamma_1, W_{\mathrm{c}}(x))\right\}\right\}\right)^2 \; .$$

□

4.10. Proof of Theorem 15

We preface the proof by a Lemma.

Lemma 29. *Assume* **L**1 *and that* U *is convex. Then*

$$\int_{\mathbb{R}^d} \mathrm{e}^{-U(y)} \mathrm{d}y \leq \left(\frac{(2\pi)^{(d+1)/2}(d-1)!}{\eta^d \boldsymbol{\Gamma}((d+1)/2)} + \frac{\pi^{d/2} M_\eta^d}{\boldsymbol{\Gamma}(d/2+1)}\right) \; . \tag{47}$$

Proof. By (24) and $U(x^\star) = 0$, we have

$$\int_{\mathbb{R}^d} \mathrm{e}^{-U(y)} \mathrm{d}y \leq \int_{\mathbb{R}^d} \mathrm{e}^{-\eta\|y-x^\star\|} \mathrm{d}y + \int_{\mathbb{R}^d} 1_{\{\|y-x^\star\| \leq M_\eta\}} \mathrm{d}y \; .$$

Then the proof is concluded using the spherical coordinates. □

Proof of Theorem 15. By [5, Theorem 1.2], π satisfies a Poincaré inequality with constant $\log^{-1}(\kappa)$. Therefore, the second term in (10) is dealt as in the proof of Theorem 12 using (22), Lemma 29 and Lemma 26. □

4.11. Proof of Proposition 16

For all $x \in \mathbb{R}^d$, we have

$$\mathscr{A}^{\mathrm{L}} W_{\mathrm{c}}(x) = \frac{\eta W_{\mathrm{c}}(x)}{4(\|x-x^\star\|^2+1)^{1/2}} \Big\{ (\eta/4)(\|x-x^\star\|^2+1)^{-1/2} \|x-x^\star\|^2 - \langle \nabla U(x), x-x^\star \rangle - (\|x-x^\star\|^2+1)^{-1} \|x-x^\star\|^2 + d \Big\} .$$

By (24), $\langle \nabla U(x), x-x^\star \rangle \geq \eta \|x-x^\star\|$ for all $x \in \mathbb{R}^d$, $\|x-x^\star\| \geq M_\eta$. Then, for all x, $\|x-x^\star\| \geq K = \max(M_\eta, 4d/\eta, 1)$, $\mathscr{A}^{\mathrm{L}} W_{\mathrm{c}}(x) \leq -(\eta^2/8) W_{\mathrm{c}}(x)$. In addition, since U is convex and $\nabla U(x^\star) = 0$, for all $x \in \mathbb{R}^d$, $\langle \nabla U(x), x-x^\star \rangle \geq 0$ and we get $\sup_{\{x \in \mathcal{E}\}} \mathscr{A}^{\mathrm{L}} W_{\mathrm{c}}(x) \leq \beta$.

4.12. Proof of Proposition 20

Under **L**1, using that $\nabla U(x^\star) = 0$, we get for all $x \in \mathbb{R}^d$,

$$\begin{aligned} \int_{\mathbb{R}^d} \|y-x^\star\|^2 R_\gamma(x, \mathrm{d}y) &= \|x-x^\star + \gamma(\nabla U(x^\star) - \nabla U(x))\|^2 + 2\gamma d \\ &\leq (1+(L\gamma)^2) \|x-x^\star\|^2 - 2\gamma \langle \nabla U(x) - \nabla U(x^\star), x-x^\star \rangle + 2\gamma d . \end{aligned} \tag{48}$$

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

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

Using again (48) and the convexity of U, it yields for all $x \in \mathbb{R}^d$, $\|x-x^\star\| \leq M_{\mathrm{s}}$,

$$\int_{\mathbb{R}^d} \|y-x^\star\|^2 R_\gamma(x, \mathrm{d}y) \leq \gamma \mathrm{c} ,$$

which concludes the proof.

4.13. Proof of Proposition 22

We preface the proof by a lemma.

Lemma 30. *Assume **H**4. Then, for all $x \in \mathbb{R}^d$,*

$$\|x - \gamma \nabla U(x) - x_1^\star\|^2 \leq (1 - \varpi\gamma/2) \|x-x_1^\star\|^2 + \gamma(\gamma + 2\varpi^{-1}) \|\nabla U_2\|_\infty^2 .$$

Proof. Using that for all $y, z \in \mathbb{R}^d$, $\|y+z\|^2 \leq (1+\varpi\gamma/2)\|y\|^2 + (1+2(\varpi\gamma)^{-1})\|z\|^2$, we get under **H**4-(b):

$$\|x - \gamma\nabla U(x) - x_1^\star\|^2 \leq (1+\varpi\gamma/2)\|x - \gamma\nabla U_1(x) - x_1^\star\|^2 + \gamma(\gamma + 2\varpi^{-1})\|\nabla U_2\|_\infty^2 \ . \quad (49)$$

By [33, Theorem 2.1.12, Theorem 2.1.9], **H**4-(b) implies that for all $x, y \in \mathbb{R}^d$:

$$\langle \nabla U_1(y) - \nabla U_1(x), y - x\rangle \geq (\varpi/2)\|y-x\|^2 + \frac{1}{m+L_1}\|\nabla U_1(y) - \nabla U_1(x)\|^2 \ ,$$

Using this inequality and $\nabla U_1(x_1^\star) = 0$ in (49) concludes the proof. □

Proof of Proposition 22. For any $\gamma \in (0, 2/(m+L_1))$, we have for all $x \in \mathbb{R}^d$:

$$\int_{\mathbb{R}^d} \|y - x_1^\star\|^2 R_\gamma(x, \mathrm{d}y) = \|x - \gamma\nabla U(x) - x_1^\star\|^2 + 2\gamma d \leq (1 - \varpi\gamma/2)\|x - x_1^\star\|^2 + \gamma\left\{(\gamma + 2\varpi^{-1})\|\nabla U_2\|_\infty^2 + 2d\right\} \ ,$$

where we have used Lemma 30 for the last inequality. Since $\gamma_1 \leq 2/(m+L_1)$ and $(\gamma_k)_{k\geq 1}$ is nonincreasing, by a straightforward induction, for $p \geq 1$ and $x \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} \|y - x_1^\star\|^2 Q_\gamma^p(x, \mathrm{d}y) \leq \prod_{k=1}^p (1 - \varpi\gamma_k/2)\|x - x_1^\star\|^2 + ((\gamma_1 + 2\varpi^{-1})\|\nabla U_2\|_\infty^2 + 2d)\sum_{i=n}^p \prod_{k=i+1}^p (1 - \varpi\gamma_k/2)\gamma_i \ . \quad (50)$$

Consider the second term in the right hand side of (50). Since $\gamma_1 \leq 2/(m+L_1)$, $m \leq L_1$ and $(\gamma_k)_{k\geq 1}$ is nonincreasing, $\max_{k\geq 1}\gamma_k \leq \varpi^{-1}$ and therefore:

$$\sum_{i=n}^p \prod_{k=i+1}^p (1 - \varpi\gamma_k/2)\gamma_i \leq \varpi^{-1}\sum_{i=n}^p \left\{\prod_{k=i+1}^p (1 - \varpi\gamma_k/2) - \prod_{k=i}^p (1 - \varpi\gamma_k/2)\right\} \leq 2\varpi^{-1} \ .$$

□

4.14. Proof of Theorem 23

We preface the proof of the Theorem by a preliminary lemma.

Lemma 31. *Assume **H**4. Let $\gamma \in (0, 2/(m+L_1))$, then for all $x \in \mathbb{R}^d$,*

$$\mathrm{Ent}_\pi\left(\frac{\mathrm{d}\delta_x R_\gamma}{\mathrm{d}\pi}\right) \leq (L_1/2)\left\{(1 - \varpi\gamma/2)\|x - x_1^\star\|^2 + \gamma(\gamma + 2\varpi^{-1})\|\nabla U_2\|_\infty^2\right\} + \mathrm{osc}_{\mathbb{R}^d}(U_2) - (d/2)(1 + \log(2\gamma m) - 2L_1\gamma) \ .$$

Proof. Let $\gamma \in (0, 2/(m + L_1))$ and r_γ be the transition density of R_γ given by (41). Under **H**4-(a) by [33, Theorems 2.1.8-2.1.9], we have for all $x \in \mathbb{R}^d$, $U_1(x) \leq U_1(x_1^\star) + (L_1/2) \|x - x_1^\star\|^2$ Therefore we have for all $x \in \mathbb{R}^d$

$$\mathrm{Ent}_\pi \left(\frac{\mathrm{d}\delta_x R_\gamma}{\mathrm{d}\pi} \right) = \int_{\mathbb{R}^d} \log(r_\gamma(x, y)/\pi(y)) r_\gamma(x, y) \mathrm{d}x \leq R_\gamma \psi(x) - (d/2)(1 + \log(4\pi\gamma)) \,, \quad (51)$$

where $\psi : \mathbb{R}^d \to \mathbb{R}$ is the function defined for all $y \in \mathbb{R}^d$ by

$$\psi(y) = U_2(y) + U_1(x_1^\star) + (L_1/2) \|y - x_1^\star\|^2 + \log \left(\int_{\mathbb{R}^d} \mathrm{e}^{-U(z)} \mathrm{d}z \right) .$$

By **H**4-(b) and Lemma 30, we get for all $x \in \mathbb{R}^d$:

$$\begin{aligned} R_\gamma \psi(x) &\leq (L_1/2) \|x - \gamma \nabla U(x) - x_1^\star\|^2 + \log \left(\int_{\mathbb{R}^d} \mathrm{e}^{-U_1(z) + U_1(x_1^\star)} \mathrm{d}z \right) \\ &\quad + \mathrm{osc}_{\mathbb{R}^d}(U_2) + dL_1\gamma \\ &\leq (L_1/2) \left\{ (1 - \varpi\gamma/2) \|x - x_1^\star\|^2 + \gamma(\gamma + 2\varpi^{-1}) \|\nabla U_2\|_\infty^2 \right\} \\ &\quad + \mathrm{osc}_{\mathbb{R}^d}(U_2) + dL_1\gamma \,. \end{aligned}$$

Plugging this bound in (51) gives the desired result. □

Proof of Theorem 23. We first deal with the second term in the right hand side of (10). Under **H**4, [4, Corollary 5.7.2] and the Holley-Stroock perturbation principle [19, p. 1184] show that π satisfies a log-Sobolev inequality with constant $C_{\mathrm{LS}} = -\log^{-1}(\kappa)$. So by (32) we have

$$\|\delta_x Q_\gamma^n P_t - \pi\|_{\mathrm{TV}} \leq \kappa^t \left\{ 2 \, \mathrm{Ent}_\pi \left(\frac{\mathrm{d}\delta_x Q_\gamma^n}{\mathrm{d}\pi} \right) \right\}^{1/2} .$$

We now bound $\mathrm{Ent}_\pi \left(\mathrm{d}\delta_x Q_\gamma^n / \mathrm{d}\pi \right)$ which will imply the upper bound of $C(\delta_x Q_\gamma^n)$. We proceed by induction. For $n = 1$, it is Lemma 31. For $n \geq 2$, by (42) and the Jensen inequality applied to the convex function $t \mapsto t \log(t)$, we have for all $x \in \mathbb{R}^d$ and $n \geq 1$,

$$\begin{aligned} &\mathrm{Ent}_\pi \left(\mathrm{d}\delta_x Q_\gamma^n / \mathrm{d}\pi \right) \\ &= \int_{\mathbb{R}^d} \log \left\{ \pi^{-1}(y) \int_{\mathbb{R}^d} q_\gamma^{n-1}(x, z) r_{\gamma_n}(z, y) \mathrm{d}z \right\} \int_{\mathbb{R}^d} q_\gamma^{n-1}(x, z) r_{\gamma_n}(z, y) \mathrm{d}z\mathrm{d}y \\ &\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \log \left\{ r_{\gamma_n}(z, y) \pi^{-1}(y) \right\} q_\gamma^{n-1}(x, z) r_{\gamma_n}(z, y) \mathrm{d}z\mathrm{d}y \,. \end{aligned} \quad (52)$$

Using Fubini's theorem, Lemma 31, Proposition 22, and the inequality $t \geq 0$, $1 - t \leq \mathrm{e}^{-t}$ in (52) implies the bound of $C(\delta_x Q_\gamma^n)$.

Finally, $A(\gamma, x)$ is bounded using the inequality for all $y, z \in \mathbb{R}^d$, $\|y + z\|^2 \leq 2(\|y\|^2 + \|z\|^2)$, **H**4 and Proposition 22. □

5. Quantitative convergence bounds in total variation for diffusions

In this part, we derived quantitative convergence results in total variation norm for d-dimensional SDEs of the form

$$\mathrm{d}\mathbf{X}_t = b(\mathbf{X}_t)\mathrm{d}t + \mathrm{d}B_t^d \,, \tag{53}$$

started at $\mathbf{X}_0$, where $(B_t^d)_{t\geq 0}$ is a d-dimensional standard Brownian motion and $b : \mathbb{R}^d \to \mathbb{R}^d$ satisfies the following assumptions.

G 1. *b is Lipschitz and for all $x, y \in \mathbb{R}^d$, $\langle b(x) - b(y), x - y\rangle \leq 0$.*

Under **G** 1, [20, Theorems 2.4-3.1-6.1, Chapter IV] imply that there exists a unique solution $(\mathbf{X}_t)_{t\geq 0}$ to (53) for all initial point $x \in \mathbb{R}^d$, which is strongly Markovian. Denote by $(\mathbf{P}_t)_{t\geq 0}$ the transition semigroup associated with (53). To derive explicit bound for $\|\mathbf{P}_t(x,\cdot) - \mathbf{P}_t(y,\cdot)\|_{\mathrm{TV}}$, we use the coupling by reflection, introduced in [27] to show convergence in total variation norm for solution of SDE, and recently used by [15] to obtain exponential convergence in the Wasserstein distance of order 1. This coupling is defined as (see [10, Example 3.7]) the unique strong Markovian process $(\mathbf{X}_t, \mathbf{Y}_t)_{t\geq 0}$ on $\mathbb{R}^{2d}$, solving the SDE:

$$\begin{cases} \mathrm{d}\mathbf{X}_t &= b(\mathbf{X}_t)\mathrm{d}t + \mathrm{d}B_t^d \\ \mathrm{d}\mathbf{Y}_t &= b(\mathbf{Y}_t)\mathrm{d}t + (\mathrm{Id} - 2e_t e_t^T)\mathrm{d}B_t^d \,, \end{cases} \quad \text{where } e_t = e(\mathbf{X}_t - \mathbf{Y}_t) \tag{54}$$

with $e(z) = z/\|z\|$ for $z \neq 0$ and $e(0) = 0$ otherwise. Define the coupling time

$$\tau_c = \inf\{s \geq 0 \mid \mathbf{X}_s \neq \mathbf{Y}_s\} \,. \tag{55}$$

By construction $\mathbf{X}_t = \mathbf{Y}_t$ for $t \geq \tau_c$. We denote in the sequel by $\tilde{\mathbb{P}}_{(x,y)}$ and $\tilde{\mathbb{E}}_{(x,y)}$ the probability and the expectation associated with the SDE (54) started at $(x, y) \in \mathbb{R}^{2d}$ on the canonical space of continuous function from $\mathbb{R}_+$ to $\mathbb{R}^{2d}$. We denote by $(\widetilde{\mathcal{F}}_t)_{t\geq 0}$ the canonical filtration. Since $\bar{B}_t^d = \int_0^t (\mathrm{Id} - 2e_s e_s^T)\mathrm{d}B_s^d$ is a d-dimensional Brownian motion, the marginal processes $(\mathbf{X}_t)_{t\geq 0}$ and $(\mathbf{Y}_t)_{t\geq 0}$ are under $\tilde{\mathbb{P}}_{(x,y)}$ weak solutions to (53) started at x and y respectively. The results in [27] are derived under less stringent conditions than **G** 1, but do not provide quantitative estimates.

Proposition 32 ([27, Example 5])**.** *Assume **G** 1 and let $(\mathbf{X}_t, \mathbf{Y}_t)_{t\geq 0}$ be the solution of (54). Then for all $t \geq 0$ and $x, y \in \mathbb{R}^d$, we have*

$$\tilde{\mathbb{P}}_{(x,y)}\left(\tau_c > t\right) = \tilde{\mathbb{P}}_{(x,y)}\left(\mathbf{X}_t \neq \mathbf{Y}_t\right) \leq 2\left(\mathbf{\Phi}\left\{\left(2t^{1/2}\right)^{-1}\|x - y\|\right\} - 1/2\right) \,.$$

Proof. For $t < \tau_c$, $\mathbf{X}_t - \mathbf{Y}_t$ is the solution of the SDE

$$\mathrm{d}\{\mathbf{X}_t - \mathbf{Y}_t\} = \{b(\mathbf{X}_t) - b(\mathbf{Y}_t)\}\,\mathrm{d}t + 2e_t \mathrm{d}B_t^1 \,,$$

where $B_t^1 = \int_0^t 1_{\{s<\tau_c\}} e_s^T \mathrm{d}B_s^d$. Using the Itô's formula and **G** 1, we have for all $t < \tau_c$,

$$\|\mathbf{X}_t - \mathbf{Y}_t\| = \|x - y\| + \int_0^t \langle b(\mathbf{X}_s) - b(\mathbf{Y}_s), e_s\rangle\,\mathrm{d}s + 2B_t^1 \leq \|x - y\| + 2B_t^1 \,.$$

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

$$\tilde{\mathbb{P}}_{(x,y)}\left(\tau_c > t\right) \leq \tilde{\mathbb{P}}_{(x,y)}\left(\min_{0\leq s\leq t} B_s^1 \geq \|x-y\|/2\right)$$
$$= \tilde{\mathbb{P}}_{(x,y)}\left(\max_{0\leq s\leq t} B_s^1 \leq \|x-y\|/2\right) = \tilde{\mathbb{P}}_{(x,y)}\left(|B_t^1| \leq \|x-y\|/2\right) ,$$

where we have used the reflection principle in the last identity. □

Define for $R > 0$ the set $\Delta_R = \{x, y \in \mathbb{R}^d \mid \|x-y\| \leq R\}$. Proposition 32 and Lindvall's inequality give that, for all $\epsilon \in (0,1)$ and $t \geq \omega(\epsilon, R)$,

$$\sup_{(x,y)\in\Delta_R} \|\mathbf{P}_t(x,\cdot) - \mathbf{P}_t(y,\cdot)\|_{\mathrm{TV}} \leq 2(1-\epsilon) , \tag{56}$$

where ω is defined in (29). To obtain quantitative exponential bounds in total variation for any $x, y \in \mathbb{R}^d$, it is required to control some exponential moments of the successive return times to Δ_R. This is first achieved by using a drift condition for the generator $\mathscr{A}$ associated with the SDE (53) defined for all $f \in C^2(\mathbb{R}^d)$ by

$$\mathscr{A}f = \langle b, \nabla f\rangle + (1/2)\Delta f .$$

Consider the following assumption:

G 2. *(i) There exist a twice continuously differentiable function $V : \mathbb{R}^d \mapsto [1,\infty)$ and constants $\theta > 0$, $\beta \geq 0$ such that*

$$\mathscr{A}V \leq -\theta V + \beta . \tag{57}$$

(ii) There exists $\delta > 0$ and $\mathrm{R} > 0$ such that $\Theta \subset \Delta_{\mathrm{R}}$ where

$$\Theta = \{(x,y) \in \mathbb{R}^{2d} \mid V(x) + V(y) \leq 2\theta^{-1}\beta + \delta\} . \tag{58}$$

For $t > 0$, and G a closed subset of $\mathbb{R}^{2d}$, define by $\mathrm{T}_1^{G,t}$ the first return time to G delayed by t:

$$\mathrm{T}_1^{G,t} = \inf\{s \geq t \mid (\mathbf{X}_s, \mathbf{Y}_s) \in G\} .$$

For $j \geq 2$, define recursively the j-th return time to G delayed by t by

$$\mathrm{T}_j^{G,t} = \inf\{s \geq \mathrm{T}_{j-1}^{G,t} + t \mid (\mathbf{X}_s, \mathbf{Y}_s) \in G\} = \mathrm{T}_{j-1}^{G,t} + \mathrm{T}_1^{G,t} \circ \mathrm{S}_{\mathrm{T}_{j-1}^{G,t}} , \tag{59}$$

where S is the shift operator on the canonical space. By [16, Proposition 1.5 Chapter 2], the sequence $(\mathrm{T}_j^{G,t})_{j\geq 1}$ is a sequence of stopping time with respect to $(\widetilde{\mathcal{F}}_t)_{t\geq 0}$.

Proposition 33. *Assume **G**1 and **G**2. For all $x, y \in \mathbb{R}^d$, $\epsilon \in (0,1)$ and $j \geq 1$, we have*

$$\tilde{\mathbb{E}}_{(x,y)}\left[\mathrm{e}^{\tilde{\theta}\mathrm{T}_j^{\Theta,\omega(\epsilon,\mathrm{R})}}\right] \leq \{\mathrm{K}(\epsilon)\}^{j-1}\left\{(1/2)(V(x)+V(y)) + \mathrm{e}^{\tilde{\theta}\omega(\epsilon,\mathrm{R})}\tilde{\theta}^{-1}\beta\right\} ,$$
$$\tilde{\theta} = \theta^2\delta(2\beta+\theta\delta)^{-1} , \ \mathrm{K}(\epsilon) = \tilde{\theta}^{-1}\beta\left(1+\mathrm{e}^{\tilde{\theta}\omega(\epsilon,\mathrm{R})}\right) + \delta/2 , \tag{60}$$

where ω is defined in (29).

Proof. For notational simplicity, set $\mathrm{T}_j = \mathrm{T}_j^{\Theta,\omega(\epsilon,\mathrm{R})}$. Note that for all $x, y \in \mathbb{R}^d$,

$$\mathscr{A}V(x) + \mathscr{A}V(y) \leq -\tilde{\theta}(V(x) + V(y)) + 2\beta 1_{\Theta}(x, y) \; .$$

Then by the Dynkin formula (see e.g. [32, Eq. (8)]) the process

$$t \mapsto (1/2)\mathrm{e}^{\tilde{\theta}(\mathrm{T}_1 \wedge t)} \left\{ V\left(\mathbf{X}_{\mathrm{T}_1 \wedge t}\right) + V\left(\mathbf{Y}_{\mathrm{T}_1 \wedge t}\right) \right\} \; , \quad t \geq \omega(\epsilon, \mathrm{R}) \; ,$$

is a positive supermartingale. Using the optional stopping theorem and the Markov property, we have, using that for all $t \geq 0$ $\tilde{\mathbb{E}}_{(x,y)}\left[\mathrm{e}^{\tilde{\theta}t} V(X_t)\right] \leq V(x) + \beta\tilde{\theta}^{-1}\mathrm{e}^{\tilde{\theta}t}$,

$$\tilde{\mathbb{E}}_{(x,y)}\left[\mathrm{e}^{\tilde{\theta}\mathrm{T}_1}\right] \leq (1/2)(V(x) + V(y)) + \mathrm{e}^{\tilde{\theta}\omega(\epsilon,\mathrm{R})}\tilde{\theta}^{-1}\beta \; .$$

The result then follows from this inequality and the strong Markov property. □

Theorem 34. *Assume **G**1 and **G**2. Then for all $\epsilon \in (0,1)$, $t \geq 0$ and $x, y \in \mathbb{R}^d$,*

$$\|\mathbf{P}_t(x,\cdot) - \mathbf{P}_t(y,\cdot)\|_{\mathrm{TV}} \leq 2\mathrm{e}^{-\tilde{\theta}t/2}\left\{(1/2)(V(x) + V(y)) + \mathrm{e}^{\tilde{\theta}\omega(\epsilon,\mathrm{R})}\tilde{\theta}^{-1}\beta\right\} + 4\kappa^t \; ,$$

where ω is defined in (29), $\tilde{\theta}, \mathrm{K}(\epsilon)$ in (60) and

$$\log(\kappa) = (\tilde{\theta}/2)\log(1-\epsilon)\{\log(\mathrm{K}(\epsilon)) - \log(1-\epsilon)\}^{-1} \; .$$

Proof. Let $x, y \in \mathbb{R}^d$ and $t \geq 0$. For all $\ell \geq 1$ and $\epsilon \in (0,1)$,

$$\tilde{\mathbb{P}}_{(x,y)}(\tau_c > t) \leq \tilde{\mathbb{P}}_{(x,y)}(\tau_c > t, \mathrm{T}_\ell \leq t) + \tilde{\mathbb{P}}_{(x,y)}(\mathrm{T}_\ell > t) \; , \tag{61}$$

where $\mathrm{T}_\ell = \mathrm{T}_\ell^{\Theta,\omega(\epsilon,\mathrm{R})}$. We now bound the two terms in the right hand side of this equation. For the first term, since $\Theta \subset \Delta_{\mathrm{R}}$, by (56), we have conditioning successively on $\widetilde{\mathcal{F}}_{\mathrm{T}_j}$, for $j = \ell, \ldots, 1$, and using the strong Markov property,

$$\tilde{\mathbb{P}}_{(x,y)}(\tau_c > t, \mathrm{T}_\ell \leq t) \leq (1-\epsilon)^\ell \; . \tag{62}$$

For the second term, using Proposition 33 and the Markov inequality, we get

$$\begin{aligned}\tilde{\mathbb{P}}_{(x,y)}(\mathrm{T}_\ell > t) &\leq \tilde{\mathbb{P}}_{(x,y)}(\mathrm{T}_1 > t/2) + \tilde{\mathbb{P}}_{(x,y)}(\mathrm{T}_\ell - \mathrm{T}_1 > t/2) \\ &\leq \mathrm{e}^{-\tilde{\theta}t/2}\left\{(1/2)(V(x) + V(y)) + \mathrm{e}^{\tilde{\theta}\omega(\epsilon,\mathrm{R})}\tilde{\theta}^{-1}\beta\right\} + \mathrm{e}^{-\tilde{\theta}t/2}\{\mathrm{K}(\epsilon)\}^{\ell-1} \; .\end{aligned}$$

The proof is completed combining this inequality and (62) in (61) and taking $\ell = \left\lceil 2^{-1}\tilde{\theta}t/(\log(\mathrm{K}(\epsilon)) - \log(1-\epsilon))\right\rceil$. □

More precise bounds can be obtained under more stringent assumption on the drift b; see [6] and [15].

G3. *There exist $\tilde{M}_{\mathrm{s}} \geq 1$ and $\tilde{m}_{\mathrm{s}} > 0$, such that for all $x, y \in \mathbb{R}^d$, $\|x - y\| \geq \tilde{M}_{\mathrm{s}}$,*

$$\langle b(x) - b(y), x - y\rangle \leq -\tilde{m}_{\mathrm{s}}\|x - y\|^2 \; .$$

Proposition 35. *Assume* ***G****1 and* ***G****3.*

(a) For all $x, y \in \mathbb{R}^d$ *and* $\epsilon \in (0,1)$

$$\tilde{\mathbb{E}}_{(x,y)}\left[\exp\left(\frac{\tilde{m}_{\mathrm{s}}}{2}\left(\tau_c \wedge \mathrm{T}_1^{\Delta_{\tilde{M}_{\mathrm{s}}},\omega(\epsilon,\tilde{M}_{\mathrm{s}})}\right)\right)\right] \leq 1 + \|x-y\| + (1+\tilde{M}_{\mathrm{s}})\mathrm{e}^{\tilde{m}_{\mathrm{s}}\omega(\epsilon,\tilde{M}_{\mathrm{s}})/2} .$$

(b) For all $x, y \in \mathbb{R}^d$, $\epsilon \in (0,1)$ *and* $j \geq 1$

$$\tilde{\mathbb{E}}_{(x,y)}\left[\exp\left((\tilde{m}_{\mathrm{s}}/2)\left(\tau_c \wedge \mathrm{T}_j^{\Delta_{\tilde{M}_{\mathrm{s}}},\omega(\epsilon,\tilde{M}_{\mathrm{s}})}\right)\right)\right] \leq \{\mathrm{D}(\epsilon)\}^{j-1}\left\{1 + \|x-y\| + (1+\tilde{M}_{\mathrm{s}})\mathrm{e}^{\tilde{m}_{\mathrm{s}}\omega(\epsilon,\tilde{M}_{\mathrm{s}})/2}\right\} ,$$

$$\mathrm{D}(\epsilon) = (1+\mathrm{e}^{\tilde{m}_{\mathrm{s}}\omega(\epsilon,\tilde{M}_{\mathrm{s}})/2})(1+\tilde{M}_{\mathrm{s}}) , \tag{63}$$

where ω *is given in* (29).

Proof. In the proof, we set $\mathrm{T}_j = \mathrm{T}_j^{\Delta_{\tilde{M}_{\mathrm{s}}},\omega(\epsilon,\tilde{M}_{\mathrm{s}})}$.

(a) Consider the sequence of increasing stopping time

$$\tau_k = \inf\{t > 0 \mid \|\mathbf{X}_t - \mathbf{Y}_t\| \notin \left[k^{-1}, k\right]\} , \quad k \geq 1 ,$$

and set $\zeta_k = \tau_k \wedge \mathrm{T}_1$. We derive a bound on $\tilde{\mathbb{E}}_{(x,y)}[\exp\{(\tilde{m}_{\mathrm{s}}/2)\zeta_k\}]$ independent on k. Since $\lim_{k\to+\infty} \uparrow \tau_k = \tau_c$ almost surely, the monotone convergence theorem implies that the same bound holds for $\tilde{\mathbb{E}}_{(x,y)}[\exp\{(\tilde{m}_{\mathrm{s}}/2)(\tau_c \wedge \mathrm{T}_1)\}]$. Set now $\mathrm{W}_{\mathrm{s}}(x,y) = 1 + \|x-y\|$. Since $\mathrm{W}_{\mathrm{s}} \geq 1$ and $\tau_c < \infty$ a.s by Proposition 32, it suffices to give a bound on $\tilde{\mathbb{E}}_{(x,y)}[\exp\{(\tilde{m}_{\mathrm{s}}/2)\zeta_k\}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_{\zeta_k}, \mathbf{Y}_{\zeta_k})]$. By Itô's formula, we have for all $v, t \leq \tau_c$, $v \leq t$

$$\mathrm{e}^{\tilde{m}_{\mathrm{s}}t/2}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_t, \mathbf{Y}_t) = \mathrm{e}^{\tilde{m}_{\mathrm{s}}v/2}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_v, \mathbf{Y}_v) + (\tilde{m}_{\mathrm{s}}/2)\int_v^t \mathrm{e}^{\tilde{m}_{\mathrm{s}}u/2}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_u, \mathbf{Y}_u)\mathrm{d}u + \int_v^t \mathrm{e}^{\tilde{m}_{\mathrm{s}}u/2}\left\langle b(\mathbf{X}_u) - b(\mathbf{Y}_u), e_u\right\rangle \mathrm{d}u + 2\int_v^t \mathrm{e}^{\tilde{m}_{\mathrm{s}}u/2}\mathrm{d}B_u^1 . \tag{64}$$

Using $\mathbf{G}3(b)$, we have for all $k \geq 1$ and $t_{\mathrm{s}} = \omega(\epsilon, \tilde{M}_{\mathrm{s}}) \leq v \leq t$

$$\mathrm{e}^{(\tilde{m}_{\mathrm{s}}/2)(\zeta_k\wedge t)}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_{\zeta_k\wedge t}, \mathbf{Y}_{\zeta_k\wedge t}) \leq \mathrm{e}^{(\tilde{m}_{\mathrm{s}}/2)(\zeta_k\wedge v)}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_{\zeta_k\wedge v}, \mathbf{Y}_{\zeta_k\wedge v}) + 2\int_{\zeta_k\wedge v}^{\zeta_k\wedge t} \mathrm{e}^{\tilde{m}_{\mathrm{s}}u/2}\mathrm{d}B_u^1 .$$

So the process

$$\left\{\exp\left((\tilde{m}_{\mathrm{s}}/2)(\zeta_k \wedge t)\right)\mathrm{W}_{\mathrm{s}}(\mathbf{X}_{\zeta_k\wedge t}, \mathbf{Y}_{\zeta_k\wedge t})\right\}_{t\geq t_{\mathrm{s}}} ,$$

is a positive supermartingale and by the optional stopping theorem, we get

$$\tilde{\mathbb{E}}_{(x,y)}\left[\mathrm{e}^{(\tilde{m}_{\mathrm{s}}/2)\zeta_k}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_{\zeta_k}, \mathbf{Y}_{\zeta_k})\right] \leq \tilde{\mathbb{E}}_{(x,y)}\left[\mathrm{e}^{(\tilde{m}_{\mathrm{s}}/2)(\tau_k\wedge t_{\mathrm{s}})}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_{\tau_k\wedge t_{\mathrm{s}}}, \mathbf{Y}_{\tau_k\wedge t_{\mathrm{s}}})\right] , \tag{65}$$

where we used that $\zeta_k \wedge t_s = \tau_k \wedge t_s$. By (64), **G**1 and **G**3, we have

$$\tilde{\mathbb{E}}_{(x,y)}\left[\mathrm{e}^{(\tilde{m}_s/2)(\tau_k \wedge t_s)} \mathrm{W}_s(\mathbf{X}_{\tau_k \wedge t_s}, \mathbf{Y}_{\tau_k \wedge t_s})\right] \leq \mathrm{W}_s(x,y) + (1+\tilde{M}_s)\mathrm{e}^{\tilde{m}_s t_s/2} ,$$

and (65) becomes

$$\tilde{\mathbb{E}}_{(x,y)}\left[\mathrm{e}^{(\tilde{m}_s/2)\zeta_k} \mathrm{W}_s(\mathbf{X}_{\zeta_k}, \mathbf{Y}_{\zeta_k})\right] \leq \mathrm{W}_s(x,y) + (1+\tilde{M}_s)\mathrm{e}^{\tilde{m}_s t_s/2} .$$

(b) The proof is by induction. The case $j = 1$ has been established above. Now let $j \geq 2$. Since on the event $\{\tau_c > \mathrm{T}_{j-1}\}$, we have

$$\tau_c \wedge \mathrm{T}_j = \mathrm{T}_{j-1} + (\tau_c \wedge \mathrm{T}_1) \circ \mathrm{S}_{\mathrm{T}_{j-1}} ,$$

where S is the shift operator, we have conditioning on $\widetilde{\mathcal{F}}_{\mathrm{T}_{j-1}}$, using the strong Markov property, Proposition 32 and the first part,

$$\tilde{\mathbb{E}}_{(x,y)}\left[1_{\{\tau_c > \mathrm{T}_{j-1}\}} \mathrm{e}^{(\tilde{m}_s/2)(\tau_c \wedge \mathrm{T}_j)}\right] \leq \mathrm{D}(\epsilon)\, \tilde{\mathbb{E}}_{(x,y)}\left[1_{\{\tau_c > \mathrm{T}_{j-1}\}} \mathrm{e}^{(\tilde{m}_s/2)\mathrm{T}_{j-1}}\right] .$$

Then the proof follows since $\mathrm{D}(\epsilon) \geq 1$.

□

Theorem 36. *Assume **G**1 and **G**3. Then for all $\epsilon \in (0,1)$, $t \geq 0$ and $x, y \in \mathbb{R}^d$,*

$$\|\mathbf{P}_t(x,\cdot) - \mathbf{P}_t(y,\cdot)\|_{\mathrm{TV}} \leq 2\left\{(1-\epsilon)^{-1} + 1 + \|x-y\|\right\}\kappa^t$$

$$\log(\kappa) = (\tilde{m}_s/2)\log(1-\epsilon)(\log(\mathrm{D}(\epsilon)) - \log(1-\epsilon))^{-1} ,$$

where $\mathrm{D}(\epsilon)$ is defined in (63).

Proof. The proof is along the same lines as Theorem 34. Set $\mathrm{T}_j = \mathrm{T}_j^{\Delta_{\tilde{M}_s}, \omega(\epsilon, \tilde{M}_s)}$ for $j \geq 1$. Let $x, y \in \mathbb{R}^d$ and $t \geq 0$. For all $\ell \geq 1$ and $\epsilon \in (0,1)$,

$$\tilde{\mathbb{P}}_{(x,y)}(\tau_c > t) \leq \tilde{\mathbb{P}}_{(x,y)}(\tau_c > t, \mathrm{T}_\ell \leq t) + \tilde{\mathbb{P}}_{(x,y)}(\mathrm{T}_\ell \wedge \tau_c > t) . \tag{66}$$

For the first term, by (56) we have conditioning successively on $\widetilde{\mathcal{F}}_{\mathrm{T}_j}$, for $j = \ell, \cdots, 1$, and using the strong Markov property,

$$\tilde{\mathbb{P}}_{(x,y)}(\tau_c > t, \mathrm{T}_\ell \leq t) \leq (1-\epsilon)^\ell . \tag{67}$$

For the second term, using Proposition 35-(b) and the Markov inequality, we get

$$\tilde{\mathbb{P}}_{(x,y)}(\mathrm{T}_\ell \wedge \tau_c > t) \leq \mathrm{e}^{-\frac{\tilde{m}_s t}{2}}\{\mathrm{D}(\epsilon)\}^{\ell-1}\left\{1 + \|x-y\| + (1+\tilde{M}_s)\mathrm{e}^{\frac{\tilde{m}_s \omega(\epsilon,\tilde{M}_s)}{2}}\right\} . \tag{68}$$

Taking $\ell = \left\lfloor (\tilde{m}_s t/2)/(\log(\mathrm{D}(\epsilon)) - \log(1-\epsilon))\right\rfloor$ and combining (67)-(68) in (66) concludes the proof. □

5.1. Proof of Theorem 17 and Theorem 21

Recall that $(P_t)_{t\geq 0}$ is the Markov semigroup of the Langevin equation associated with U and let $\mathscr{A}^{\mathrm{L}}$ be the corresponding generator. Since $(P_t)_{t\geq 0}$ is reversible with respect to π, we deduce from Theorem 34 and Theorem 36 quantitative bounds for the exponential convergence of $(P_t)_{t\geq 0}$ to π in total variation noting that if $(Y_t)_{t\geq 0}$ is a solution of (1), then $(Y_{t/2})_{t\geq 0}$ is a weak solution of the rescaled Langevin diffusion:

$$\mathrm{d}\tilde{Y}_t = -(1/2)\nabla U(\tilde{Y}_t)\mathrm{d}t + \mathrm{d}B_t^d \,. \tag{69}$$

Proof of Theorem 17. Since the generator associated with the SDE (69) is $(1/2)\mathscr{A}^{\mathrm{L}}$, Proposition 16 shows that (57) holds for W_{c} with constants $\theta/2$ and $\beta/2$. Using that for all $a_1, a_2 \in \mathbb{R}$, $\mathrm{e}^{(a_1+a_2)/2} \leq (1/2)(\mathrm{e}^{a_1}+\mathrm{e}^{a_2})$, **G**2-(ii) holds for $\delta = 2\theta^{-1}\beta$ and $\mathrm{R} = (8/\eta)\log(4\theta^{-1}\beta)$. By Theorem 34 with $\epsilon = 1/2$, we get for all $x, y \in \mathbb{R}^d$ and $t \geq 0$

$$\begin{aligned}\|P_t(x,\cdot) - P_t(y,\cdot)\|_{\mathrm{TV}} &\leq 4\varpi^t \\ &+ 2\mathrm{e}^{-\theta t/4}\left\{(1/2)(W_{\mathrm{c}}(x)+W_{\mathrm{c}}(y)) + 2\theta^{-1}\beta\mathrm{e}^{4\theta^{-1}\omega(2^{-1},(8/\eta)\log(4\theta^{-1}\beta))}\right\} \,,\end{aligned} \tag{70}$$

where ϖ is defined in (30a). By [32, Theorem 4.3-(ii)], (57) implies that $\int_{\mathbb{R}^d} W_{\mathrm{c}}(y)\pi(\mathrm{d}y) \leq \beta\theta^{-1}$. The proof is then concluded using this bound, (70) and that π is invariant for $(P_t)_{t\geq 0}$ □

Proof of Theorem 21. By applying Theorem 36 with $\epsilon = 1/2$, the triangle inequality and using that π is invariant for $(P_t)_{t\geq 0}$, we have

$$\|P_t(x,\cdot) - \pi\|_{\mathrm{TV}} \leq 2\left\{3 + \|x - x^\star\| + \int_{\mathbb{R}^d} \|y - x^\star\|\,\mathrm{d}\pi(y)\right\}\kappa^t \,.$$

It remains to show that $\int_{\mathbb{R}^d} \|y - x^\star\|\,\mathrm{d}\pi(y) \leq (d/m + M_{\mathrm{s}}^2)^{1/2}$. For this, we establish a drift inequality for the generator $\mathscr{A}^{\mathrm{L}}$ of the Langevin SDE associated with U. Consider the function $W_{\mathrm{s}}(x) = \|x - x^\star\|^2$. For all $x \in \mathbb{R}^d$, we have using $\nabla U(x^\star) = 0$,

$$\mathscr{A}^{\mathrm{L}} W_{\mathrm{s}}(x) \leq 2(d - \langle \nabla U(x) - \nabla U(x^\star), x - x^\star\rangle) \,.$$

Therefore by **G**3, for all $x \in \mathbb{R}^d$, $\|x - x^\star\| \geq M_{\mathrm{s}}$, we get

$$\mathscr{A}^{\mathrm{L}} W_{\mathrm{s}}(x) \leq -2mW_{\mathrm{s}}(x) + 2d \,,$$

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

$$\mathscr{A}^{\mathrm{L}} W_{\mathrm{s}}(x) \leq -2mW_{\mathrm{s}}(x) + 2(d + mM_{\mathrm{s}}^2) \,.$$

By [32, Theorem 4.3-(ii)], we get $\int_{\mathbb{R}^d} W_{\mathrm{s}}(y)\mathrm{d}\pi(y) \leq d/m + M_{\mathrm{s}}^2$. The bound on $C(\delta_x Q_\gamma^n)$ is a consequence of the Cauchy-Schwarz inequality, Proposition 20 and Lemma 1. The bound for $A(\gamma, x)$ similarly follows from **L**1, Proposition 20 and Lemma 1.

□

Acknowledgements

The authors are indebted to Arnaud Guillin for sharing his knowledge of Poincaré and log-Sobolev inequalities. The authors are grateful to Andreas Eberle for very careful readings and many useful comments. The author thank the anonymous referees for their constructive feedback. 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] C. Andrieu, N. De Freitas, A. Doucet, and M. I Jordan. An introduction to MCMC for machine learning. *Machine learning*, 50(1-2):5–43, 2003.

[2] D. Bakry, F. Barthe, P. Cattiaux, and A. Guillin. A simple proof of the Poincaré inequality for a large class of probability measures. *Electronic Communications in Probability [electronic only]*, 13:60–66, 2008.

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

[4] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.

[5] S. G. Bobkov. Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.*, 27(4):1903–1921, 10 1999.

[6] 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.

[7] 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.

[8] 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.

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

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

[11] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. *Statist. Sci.*, 28(3):424–446, 2013.

[12] 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.

[13] A. S. Dalalyan and A. B. Tsybakov. Sparse regression learning by aggregation and Langevin Monte-Carlo. *J. Comput. System Sci.*, 78(5):1423–1443, 2012.
[14] A. Durmus, É. Moulines, and M. Pereyra. Sampling from convex non continuously differentiable functions, when Moreau meets Langevin. In preparation.
[15] A. Eberle. Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields*, pages 1–36, 2015.
[16] S.N. Ethier and T.G. Kurtz. *Markov processes: characterization and convergence*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley, 1986.
[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] R. Holley and D. Stroock. Logarithmic sobolev inequalities and stochastic ising models. *Journal of statistical physics*, 46(5):1159–1194, 1987.
[20] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North-Holland Mathematical Library. Elsevier Science, 1989.
[21] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer New York, 1991.
[22] S. Kullback. *Information theory and statistics*. Dover Publications, Inc., Mineola, NY, 1997. Reprint of the second (1968) edition.
[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] V. Lemaire and S. Menozzi. On some non asymptotic bounds for the Euler scheme. *Electron. J. Probab.*, 15:no. 53, 1645–1681, 2010.
[27] T. Lindvall and L. C. G. Rogers. Coupling of multidimensional diffusions by reflection. *Ann. Probab.*, 14(3):860–872, 1986.
[28] L. Lovász and S. Vempala. The geometry of logconcave functions and sampling algorithms. *Random Structures Algorithms*, 30(3):307–358, 2007.
[29] 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.
[30] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, New York, NY, USA, 2nd edition, 2009.
[31] S. P. Meyn and R. L. Tweedie. Stability of Markovian Processes III: Foster-Lyapunov Criteria for Continuous-Time Processes. *Advances in Applied Probability*, 25(3), 1993.
[32] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes. III. Foster-

Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.*, 25(3):518–548, 1993.

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

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

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

[36] G. O. Roberts and R. L. Tweedie. Rates of convergence of stochastically monotone and continuous time Markov models. *J. Appl. Probab.*, 37(2):359–373, 2000.

[37] 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.