Submitted to the Annals of Statistics

DIFFUSION APPROXIMATIONS AND CONTROL VARIATES FOR MCMC

By Nicolas Brosse, Alain Durmus, Sean Meyn,
Éric Moulines, Anand Radhakrishnan

Abstract: A new methodology is presented for the construction of control variates to reduce the variance of additive functionals of Markov Chain Monte Carlo (MCMC) samplers. Our control variates are defined through the minimization of the asymptotic variance of the Langevin diffusion over a family of functions, which can be seen as a quadratic risk minimization procedure. The use of these control variates is theoretically justified. We show that the asymptotic variances of some well-known MCMC algorithms, including the Random Walk Metropolis and the (Metropolis) Unadjusted/Adjusted Langevin Algorithm, are close to the asymptotic variance of the Langevin diffusion. Several examples of Bayesian inference problems demonstrate that the corresponding reduction in the variance is significant.

1. Introduction. Let $U : \mathbb{R}^d \to \mathbb{R}$ be a measurable function on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that $\int_{\mathbb{R}^d} \mathrm{e}^{-U(x)}\mathrm{d}x < \infty$. This function is associated to a probability measure π on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ defined for all $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$ by $\pi(\mathsf{A}) := \int_{\mathsf{A}} \mathrm{e}^{-U(x)}\mathrm{d}x / \int_{\mathbb{R}^d} \mathrm{e}^{-U(x)}\mathrm{d}x$. We are interested in approximating $\pi(f) := \int_{\mathbb{R}^d} f(x)\pi(\mathrm{d}x)$, where f is a π-integrable function. The classical Monte Carlo solution to this problem is to simulate i.i.d. random variables $(X_k)_{k\in\mathbb{N}}$ with distribution π, and then to estimate $\pi(f)$ by the sample mean

$$\hat{\pi}_n(f) = n^{-1} \sum_{i=0}^{n-1} f(X_i) \; . \tag{1}$$

In most applications, sampling from π is not an option. Markov Chain Monte Carlo (MCMC) methods provide samples from a Markov chain $(X_k)_{k\in\mathbb{N}}$ with unique invariant probability π. Under mild conditions (Meyn and Tweedie, 2009, Chapter 17), the estimator $\hat{\pi}_n(f)$ defined by (1) satisfies for any initial distribution a Central Limit Theorem (CLT)

$$n^{-1/2} \sum_{k=0}^{n-1} \tilde{f}(X_k) \overset{\text{weakly}}{\underset{n\to+\infty}{\Longrightarrow}} \mathcal{N}(0, \sigma^2_{\infty,\mathrm{d}}(f)) \; , \tag{2}$$

where $\tilde{f} = f - \pi(f)$ and $\sigma^2_{\infty,\mathrm{d}}(f) \geq 0$ is referred to as the asymptotic variance associated to f and $\mathcal{N}(m, \sigma^2)$ denotes a Gaussian distribution with mean m and variance σ^2.

MSC 2010 subject classifications: Primary 65C05, 60F05, 62L10; secondary 65C40, 60J05, 74G10, 74G15

Keywords and phrases: Bayesian inference, Control variates, Langevin diffusion, Markov Chain Monte Carlo, Poisson equation, Variance reduction

The aim of the present paper is to propose a new methodology to reduce the asymptotic variance of a family of MCMC algorithms. This method consists in constructing suitable control variates, *i.e.* we consider a family of π-integrable functions $\mathcal{H} \subset \{h : \mathbb{R}^d \to \mathbb{R} \,:\, \pi(h) = 0\}$ and then choose $h \in \mathcal{H}$ such that $\sigma^2_{\infty,\mathrm{d}}(f+h) \leq \sigma^2_{\infty,\mathrm{d}}(f)$. Reducing the variance of Monte Carlo estimators is a very active research domain: see e.g. (Robert and Casella, 2004, Chapter 4), (Liu, 2008, Section 2.3), and (Rubinstein and Kroese, 2017, Chapter 5) for an overview of the main methods - see also Section 2.2.

Analysis and motivation are based on the Langevin diffusion defined by

$$\mathrm{d}Y_t = -\nabla U(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \,, \tag{3}$$

where $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion. In the sequel, we assume that the Stochastic Differential Equation (SDE) (3) has a unique strong solution $(Y_t)_{t\geq 0}$ for every initial condition $x \in \mathbb{R}^d$. Under appropriate conditions (see Bhattacharya (1982); Cattiaux, Chafaı and Guillin (2012)), π is invariant for the Markov process $(Y_t)_{t\geq 0}$ and the following CLT holds:

$$t^{-1/2}\int_0^t \tilde{f}(Y_s)\mathrm{d}s \overset{\text{weakly}}{\underset{t\to+\infty}{\Longrightarrow}} \mathcal{N}(0, \sigma^2_\infty(f)) \,. \tag{4}$$

The main contribution of this paper is the introduction of a new method to compute control variates based on the expression of the asymptotic variance $\sigma^2_\infty(f)$ given in (4). For any twice continuously differentiable function φ, the differential generator acting on φ is denoted by

$$\mathscr{L}\varphi = -\langle \nabla U, \nabla\varphi\rangle + \Delta\varphi \,. \tag{5}$$

Under appropriate conditions on φ and π, it may be shown that $\pi(\mathscr{L}\varphi) = 0$. This property suggests to consider the class of control functionals $\mathcal{H} = \{h = \mathscr{L}g \,:\, g \in \mathcal{G}\}$ for the Langevin diffusion, where $\mathcal{G}$ is a family of "smooth" functions, and minimize over $\mathcal{H}$, the criterion

$$h \mapsto \sigma^2_\infty(f+h) \,. \tag{6}$$

The use of control functionals $h \in \mathcal{H}$ has already been proposed in Assaraf and Caffarel (1999) with applications to quantum Monte Carlo calculations; improved schemes have been later considered in Mira, Solgi and Imparato (2013); Papamarkou, Mira and Girolami (2014) with applications to computational Bayesian inference. Although $\mathcal{H}$ is a class of control functionals for the Langevin diffusion, the choice of controls variates minimizing the criterion (6) for some MCMC algorithms is motivated by the fact the asymptotic variance $\sigma^2_{\infty,\mathrm{d}}(f)$, defined in (2) and associated to the Markov chains associated with these methods, is (up to a scaling factor) a good approximation of the asymptotic variance of the Langevin diffusion $\sigma^2_\infty(f)$ defined in (4).

The remainder of the paper is organized as follows. In Section 2, we present our methodology to minimize (6) and the construction of control variates for some MCMC algorithms. In Section 3, we state our main result which guarantees that the asymptotic variance $\sigma^2_{\infty,\mathrm{d}}(f)$ defined in (2) and associated with a given MCMC method is close (up to a scaling factor) to the asymptotic variance of the Langevin diffusion $\sigma^2_{\infty}(f)$ defined in (4). We show that under appropriate conditions on U, the Metropolis Adjusted/Unadjusted Langevin Algorithm (MALA and ULA) and the Random Walk Metropolis (RWM) algorithm fit the framework of our methodology. In Section 4, Monte Carlo experiments illustrating the performance of our method are presented. The proofs are postponed to Sections 5 and 6 and to a supplementary document Brosse et al. (2019).

Notation. Let $\mathcal{B}(\mathbb{R}^d)$ denote the Borel σ-field of $\mathbb{R}^d$. Moreover, let $\mathrm{L}^1(\mu)$ be the set of μ-integrable functions for μ a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Further, $\mu(f) = \int_{\mathbb{R}^d} f(x)\mathrm{d}\mu(x)$ for an $f \in \mathrm{L}^1(\mu)$. Given a Markov kernel R on $\mathbb{R}^d$, for all $x \in \mathbb{R}^d$ and f integrable under $R(x,\cdot)$, denote by $Rf(x) = \int_{\mathbb{R}^d} f(y)R(x,\mathrm{d}y)$. Let $V : \mathbb{R}^d \to [1,\infty)$ be a measurable function. The V-total variation distance between two probability measures μ and ν on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is defined as $\|\mu - \nu\|_V = \sup_{|f|\leq V} |\mu(f) - \nu(f)|$. If $V = 1$, then $\|\cdot\|_V$ is the total variation denoted by $\|\cdot\|_{\mathrm{TV}}$. For a measurable function $f : \mathbb{R}^d \to \mathbb{R}$, define $\|f\|_V = \sup_{x\in\mathbb{R}^d} |f(x)|/V(x)$.

For $u, v \in \mathbb{R}^d$, define the scalar product $\langle u, v\rangle = \sum_{i=1}^d u_i v_i$ and the Euclidian norm $\|u\| = \langle u, u\rangle^{1/2}$. Denote by $\mathbb{S}(\mathbb{R}^d) = \left\{u \in \mathbb{R}^d : \|u\| = 1\right\}$. For $a, b \in \mathbb{R}$, denote by $a \vee b = \max(a,b)$, $a \wedge b = \min(a,b)$ and $a_+ = a \vee 0$. For $a \in \mathbb{R}_+$, $\lfloor a \rfloor$ and $\lceil a \rceil$ denote respectively the floor and ceil functions evaluated in a. We take the convention that for $n, p \in \mathbb{N}$, $n < p$ then $\sum_p^n = 0$, $\prod_p^n = 1$ and $\{p,\ldots,n\} = \emptyset$. Define for $t \in \mathbb{R}$, $\Phi(t) = (2\pi)^{-1/2}\int_{-\infty}^t \mathrm{e}^{-r^2/2}\mathrm{d}r$ and $\bar{\Phi}(t) = 1 - \Phi(t)$. In addition, $\boldsymbol{\varphi}$ stands for the d-dimensional standard Gaussian density, *i.e.* $\boldsymbol{\varphi}(z) = (2\pi)^{-d/2}\mathrm{e}^{-\|z\|^2/2}$ for $z \in \mathbb{R}^d$.

For $k \in \mathbb{N}$, $m, m' \in \mathbb{N}^*$ and Ω, Ω' two open sets of $\mathbb{R}^m, \mathbb{R}^{m'}$ respectively, denote by $\mathrm{C}^k(\Omega,\Omega')$, the set of k-times continuously differentiable functions. For $f \in \mathrm{C}^2(\mathbb{R}^d,\mathbb{R})$, denote by ∇f the gradient of f and by Δf the Laplacian of f. For $k \in \mathbb{N}$ and $f \in \mathrm{C}^k(\mathbb{R}^d,\mathbb{R})$, denote by $\mathrm{D}^i f$ the i-th order differential of f for $i \in \{0,\ldots,k\}$. For $x \in \mathbb{R}^d$ and $i \in \{1,\ldots,k\}$, define $\left\|\mathrm{D}^0 f(x)\right\| = |f(x)|$, $\left\|\mathrm{D}^i f(x)\right\| = \sup_{u_1,\ldots,u_i\in\mathbb{S}(\mathbb{R}^d)} \mathrm{D}^i f(x)[u_1,\ldots,u_i]$. For $k, p \in \mathbb{N}$ and $f \in \mathrm{C}^k(\mathbb{R}^d,\mathbb{R})$, define the semi-norm

$$\|f\|_{k,p} = \sup_{x\in\mathbb{R}^d,\ i\in\{0,\ldots,k\}} \left\|\mathrm{D}^i f(x)\right\| / (1 + \|x\|^p) .$$

Define $\mathrm{C}^k_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R}) = \left\{f \in \mathrm{C}^k(\mathbb{R}^d,\mathbb{R}) : \inf_{p\in\mathbb{N}} \|f\|_{k,p} < +\infty\right\}$ and for any $f \in \mathrm{C}^k_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$, we consider the semi-norm

$$\|f\|_k = \|f\|_{k,p} \text{ where } p = \min\{q \in \mathbb{N} : \|f\|_{k,q} < +\infty\} .$$

Finally, define $\mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R}) = \cap_{k\in\mathbb{N}}\mathrm{C}^k_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$.

2. Langevin-based control variates for MCMC methods.

2.1. *Method.* We introduce in the following our methodology based on control variates for the Langevin diffusion. In order not to obscure the main ideas of this method, we present it informally. Results which justify rigorously the related derivations are postponed to Section 3.

We consider a family of control functionals $\mathcal{G} \subset \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$. There is a great flexibility in the choice of the family $\mathcal{G}$. We illustrate our methodology through a simple example

$$\mathcal{G}_{\mathrm{lin}} = \{g = \langle \theta, \psi \rangle \,:\, \theta \in \Theta\} \text{ where } \psi = \{\psi_i\}_{i=1}^p,\ \psi_i \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R}),\ i \in \{1, \ldots, p\}\ , \tag{7}$$

with $\Theta \subset \mathbb{R}^p$, but the method developed in this paper is by no means restricted to a linear parameterized family.

A key property of the Langevin diffusion which is the basis of our methodology is the following "carré du champ" property (see for example (Bakry, Gentil and Ledoux, 2014, Section 1.6.2, formula 1.6.3)): for all $g_1, g_2 \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$,

$$\pi\left(g_1 \mathscr{L} g_2\right) = \pi\left(g_2 \mathscr{L} g_1\right) = -\pi\left(\langle \nabla g_1, \nabla g_2 \rangle\right)\ , \tag{8}$$

which reflects in particular that $\mathscr{L}$ is a self-adjoint operator on a dense subspace of $\mathrm{L}^2(\pi)$, the Hilbert space of square integrable function w.r.t. π. A straightforward consequence of (8) (setting $g_1 = 1$) is that $\pi(\mathscr{L} g) = 0$ for any function $g \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$. This observation implies that f and $f + \mathscr{L} g$ have the same expectation with respect to π for any $f \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and $g \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$. Therefore, as emphasized in the introduction, if the CLT (4) holds, a relevant choice of control variate for the Langevin diffusion to estimate $f \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, is $h^\star = \mathscr{L} g^\star$, where $g^\star$ is a minimizer of

$$g \mapsto \sigma^2_\infty(f + \mathscr{L} g)\ . \tag{9}$$

In the following, we explain how this optimization problem can be practically solved.

It is shown in Bhattacharya (1982) (see also Glynn and Meyn (1996) and Cattiaux, Chafaı and Guillin (2012)) that under appropriate conditions on U and f, the solution $(Y_t)_{t \geq 0}$ of the Langevin diffusion (3) satisfies the CLT (4) where the asymptotic variance is given by

$$\sigma^2_\infty(f) = 2\pi(\hat{f}\{f - \pi(f)\})\ , \tag{10}$$

and $\hat{f} \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ satisfies Poisson's equation:

$$\mathscr{L} \hat{f} = -\tilde{f}\ , \quad \text{where } \tilde{f} = f - \pi(f)\ . \tag{11}$$

Another expression for $\sigma^2_\infty(f)$ is, using (8) and (11):

$$\sigma^2_\infty(f) = 2\pi(\hat{f}\tilde{f}) = -2\pi(\hat{f}\mathscr{L}\hat{f}) = 2\pi(\|\nabla \hat{f}\|^2)\ . \tag{12}$$

Based on (8), (10) and (12), we see now how the minimization of (9) can be computed in practice. First, by definition (11), for all $g \in \mathcal{G}$, $\hat{f} - g \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ is a solution to the Poisson equation

$$\mathscr{L}(\hat{f} - g) = \pi(f + \mathscr{L} g) - (f + \mathscr{L} g)\ .$$

Therefore, we get for all $g \in \mathcal{G}$, using $\pi(\mathscr{L}g) = 0$ and (10)

$$\sigma^2_\infty(f + \mathscr{L}g) = 2\pi\left((\hat{f} - g)\left\{\tilde{f} + \mathscr{L}g\right\}\right) . = 2\pi(\|\nabla\hat{f} - \nabla g\|^2) .$$

In addition, by (8) and (11), we get that $\pi(\hat{f}\mathscr{L}g) = -\pi(\tilde{f}g)$, and we obtain using (12) that

$$\begin{aligned} \sigma^2_\infty(f + \mathscr{L}g) &= 2\pi(\hat{f}\tilde{f}) - 2\pi(g\tilde{f}) + 2\pi(\hat{f}\mathscr{L}g) - 2\pi(g\mathscr{L}g) \\ &= 2\pi(\hat{f}\tilde{f}) - 4\pi(g\tilde{f}) + 2\pi(\|\nabla g\|^2) . \end{aligned} \tag{13}$$

Minimizing the map (9) is equivalent to minimization of $g \mapsto -4\pi(g\tilde{f}) + 2\pi(\|\nabla g\|^2)$. It means that we might actually minimize the function $g \mapsto \sigma^2_\infty(f + \mathscr{L}g)$ *without* computing the solution $\hat{f}$ of the Poisson equation, which is in general a computational bottleneck.

When $g_\theta = \langle\theta, \psi\rangle \in \mathcal{G}_{\mathrm{lin}}$, then (13) may be rewritten as:

$$\sigma^2_\infty(f + \mathscr{L}g_\theta) = 2\theta^{\mathrm{T}}H\theta - 4\langle\theta, b\rangle + \sigma^2_\infty(f) ,$$

where $H \in \mathbb{R}^{p\times p}$ and b are given for any $i, j \in \{1, \ldots, p\}$ by

$$H_{ij} = \pi(\langle\nabla\psi_i, \nabla\psi_j\rangle) \quad \text{and} \quad b_i = \pi(\psi_i\tilde{f}) . \tag{14}$$

Note that H is by definition a symmetric semi-positive definite matrix. If $(1, \psi_1, \ldots, \psi_p)$ are linearly independent in $\mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, then H is full rank and the minimizer of $\sigma^2_\infty(f + \mathscr{L}g_\theta)$ is given by

$$\theta^* = H^{-1}b . \tag{15}$$

In conclusion, in addition to its theoretical interest, the Langevin diffusion (3) is an attractive model because optimization of the asymptotic variance is greatly simplified. However, we are not advocating simulation of this diffusion in MCMC applications. The main contribution of this paper is to show that the optimal control variate for the diffusion remains nearly optimal for many standard MCMC algorithms.

One example is the Unadjusted Langevin Algorithm (ULA), the Euler discretization scheme associated to the Langevin SDE (3):

$$X_{k+1} = X_k - \gamma\nabla U(X_k) + \sqrt{2\gamma}Z_{k+1} , \tag{16}$$

where $\gamma > 0$ is the step size and $(Z_k)_{k\in\mathbb{N}}$ is an i.i.d. sequence of standard Gaussian d-dimensional random vectors. The idea of using the Markov chain $(X_k)_{k\in\mathbb{N}}$ to sample approximately from π has been first introduced in the physics literature by Parisi (1981) and popularized in the computational statistics community by Grenander (1983) and Grenander and Miller (1994). As shown below, other examples are the Metropolis Adjusted Langevin Algorithm (MALA) algorithm (for which an additional Metropolis-Hastings correction step is added) but also for MCMC algorithms which do not seem to be "directly" related to the Langevin diffusion, like the Random Walk Metropolis algorithm (RWM).

To deal with these different algorithms within the same theoretical framework, we consider a family of Markov kernels $\{R_\gamma : \gamma \in (0,\bar{\gamma}]\}$, parameterized by a scalar parameter $\gamma \in (0,\bar{\gamma}]$ where $\bar{\gamma} > 0$. For the ULA and MALA algorithm, γ is the stepsize in the Euler discretization of the diffusion; for the RWM this is the variance of the random walk proposal. For any initial probability ξ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $\gamma \in (0,\bar{\gamma}]$, denote by $\mathbb{P}_{\xi,\gamma}$ and $\mathbb{E}_{\xi,\gamma}$ the probability and the expectation respectively on the canonical space of the Markov chain with initial probability ξ and of transition kernel R_γ. By convention, we set $\mathbb{E}_{x,\gamma} = \mathbb{E}_{\delta_x,\gamma}$ for all $x \in \mathbb{R}^d$. We denote by $(X_k)_{k\geq 0}$ the canonical process. It is assumed below that $\{R_\gamma : \gamma \in (0,\bar{\gamma}]\}$, f and $\mathcal{G}$ satisfy the following assumptions. Roughly speaking, these conditions impose that for any $\gamma \in (0,\bar{\gamma}]$ and $g \in \mathcal{G}$, the discrete CLT (2) holds for the function $f + \mathscr{L}g$, and that the associated asymptotic variance $\sigma^2_{\infty,\gamma}(f+\mathscr{L}g)$ is sufficiently close to $\sigma_\infty(f+\mathscr{L}g)$ given by the continuous CLT (3), as $\gamma \downarrow 0^+$, so that control functionals for the Markov chain $(X_k)_{k\in\mathbb{N}}$ can be derived using the methodology we developed above for the Langevin diffusion.

(I) For each $\gamma \in (0,\bar{\gamma}]$, R_γ has an invariant probability distribution π_γ satisfying $\pi_\gamma(|f+\mathscr{L}g|) < \infty$ for any $g \in \mathcal{G}$.

(II) For any $g \in \mathcal{G}$ and $\gamma \in (0,\bar{\gamma}]$,

$$\sqrt{n}(\hat{\pi}_n(f+\mathscr{L}g) - \pi_\gamma(f+\mathscr{L}g)) \overset{\text{weakly}}{\underset{n\to+\infty}{\Longrightarrow}} \mathcal{N}(0, \sigma^2_{\infty,\gamma}(f+\mathscr{L}g)) \tag{17}$$

where $\hat{\pi}_n(f+\mathscr{L}g)$ is the sample mean (see (1)), and $\sigma^2_{\infty,\gamma}(f+\mathscr{L}g) \geq 0$ is the asymptotic variance (see (2)) relatively to R_γ.

(III) For any $g \in \mathcal{G}$, as $\gamma \downarrow 0^+$,

$$\gamma\sigma^2_{\infty,\gamma}(f+\mathscr{L}g) = \sigma^2_\infty(f+\mathscr{L}g) + o(1)\ , \tag{18}$$

$$\pi_\gamma(f+\mathscr{L}g) = \pi(f+\mathscr{L}g) + O(\gamma)\ , \tag{19}$$

where $\sigma^2_\infty(f+\mathscr{L}g)$ is defined in (10).

The verification that these assumptions are satisfied for the ULA, RWM and MALA algorithms (under appropriate technical conditions), in the case $f \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$ and $\mathcal{G} \subset \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$, is postponed to Section 3. The standard conditions (I)–(II) are in particular satisfied if, for any $\gamma \in (0,\bar{\gamma}]$, R_γ is V-uniformly geometrically ergodic for some measurable function $V : \mathbb{R}^d \to [1,+\infty)$, *i.e.* it admits an invariant probability measure π_γ such that $\pi_\gamma(V) < +\infty$ and there exist $C_\gamma \geq 0$ and $\rho_\gamma \in [0,1)$ such that for any probability measure ξ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $n \in \mathbb{N}$,

$$\|\xi R^n_\gamma - \pi_\gamma\|_V \leq C_\gamma \xi(V)\rho^n_\gamma\ ,$$

(see e.g. Meyn and Tweedie (2009) or Douc et al. (2018)). Condition (III) requires a specific form of the dependence of C_γ and ρ_γ on γ.

Based on (I)–(III) and (15), the estimator of $\pi(f)$ we suggest is given for $N, n, m \in \mathbb{N}^*$ by

$$\pi^{\mathrm{CV}}_{N,n,m}(f) = \frac{1}{n}\sum_{k=N}^{n+N-1} (f(X_k) + \mathscr{L}g^\star_m(X_k))\ , \tag{20}$$

where N is the length of the burn-in period and $g_m^\star \in \arg\min_{g\in\mathcal{G}} \mathrm{R}_m(g)$ is a minimizer of the structural risk associated with (13)

$$\mathrm{R}_m(g) = \frac{1}{m}\sum_{k=N}^{N+m-1}\left\{-2g(\tilde{X}_k)\tilde{f}_m(\tilde{X}_k) + \|\nabla g(\tilde{X}_k)\|^2\right\} \,, \tag{21}$$

where $\tilde{f}_m(x) = f(x) - m^{-1}\sum_{k=N}^{N+m-1} f(\tilde{X}_k)$. Here $(\tilde{X}_k)_{k\in\mathbb{N}}$ can be an independent copy of (or be identical to) the Markov chain $(X_k)_{k\in\mathbb{N}}$ and m is the length of the sequence used to estimate the control variate. In this article, we do not study to what extent minimizing the empirical asymptotic variance (21) leads to the minimization of the asymptotic variance of $\pi^{\mathrm{CV}}_{N,n,m}(f)$ (20) as $n \to +\infty$; such a problem has been tackled by Belomestny, Iosipoi and Zhivotovskiy (2018) in the i.i.d. case. To control the complexity of the class of functions $\mathcal{G}$, a penalty term may be added in (21). The use of a penalty term to control the excess risk in the estimation of the control variate has been proposed and discussed in South, Mira and Drovandi (2018). Concerning the choice of $\mathcal{G}$, the simplest case is $\mathcal{G}_{\mathrm{lin}}$ defined by (7), corresponding to the parametric case, and it is by far the most popular approach. It is possible to go one step further and adopt fully non-parametric approaches like kernel regression methods Oates, Girolami and Chopin (2016) or neural networks Zhu, Wan and Zhong (2018).

If the control function is a linear combination of functions, $g_\theta = \langle\theta,\psi\rangle$ where $\psi = \{\psi_i : 1 \le i \le p\}$, then the empirical risk (21) may be expressed as

$$\mathrm{R}_m(g_\theta) = -2\langle\theta, b_m\rangle + \langle\theta, H_m\theta\rangle \,, \tag{22}$$

where for $1 \le i,j \le p$,

$$[b_m]_i = \frac{1}{m}\sum_{k=N}^{N+m-1}\psi_i(\tilde{X}_k)\tilde{f}_m(\tilde{X}_k) \quad , \; [H_m]_{ij} = \frac{1}{m}\sum_{k=N}^{N+m-1}\left\langle\nabla\psi_i(\tilde{X}_k), \nabla\psi_j(\tilde{X}_k)\right\rangle . \tag{23}$$

In this simple case, an optimizer is obtained in closed form

$$\theta_m^* = H_m^+ b_m \,, \tag{24}$$

where H_m^+ is the Moore-Penrose pseudoinverse of H_m.

2.2. *Comparison with other control variate methods for Monte Carlo simulation.* The construction of control variates for MCMC and the related problem of approximating solutions of Poisson equations are very active fields of research. It is impossible to give credit for all the contributions undertaken in this area: see Dellaportas and Kontoyiannis (2012), Papamarkou, Mira and Girolami (2014), Oates, Girolami and Chopin (2016) and references therein for further background. We survey in this section only the methods which are closely connected to our approach. Henderson (1997) and (Meyn, 2008, Section 11.5) proposed control variates of the form $(R - \mathrm{Id})g_\theta$ where $g_\theta := \langle\theta,\psi\rangle$ and R is the Markov kernel associated to a Markov chain $(X_k)_{k\in\mathbb{N}}$ and $\psi = (\psi_1,\ldots,\psi_p)$ are known

π-integrable functions. The parameter $\theta \in \mathbb{R}^p$ is obtained by minimizing the asymptotic variance

$$\min_{\theta \in \mathbb{R}^p} \sigma^2_{\infty,\mathrm{d}}(f + (R - \mathrm{Id})g_\theta) = \min_{\theta \in \mathbb{R}^p} \pi\left(\left\{\hat{f}_\mathrm{d} - g_\theta\right\}^2 - \left\{R(\hat{f}_\mathrm{d} - g_\theta)\right\}^2\right) , \tag{25}$$

where $\hat{f}_\mathrm{d}$ is solution of the *discrete* Poisson equation $(R - \mathrm{Id})\hat{f}_\mathrm{d} = -\tilde{f}$. The method suggested in (Meyn, 2008, Section 11.5) to minimize (25) requires estimates of the solution $\hat{f}_\mathrm{d}$ of the Poisson equation. Temporal Difference learning is a possible candidate, but this method is complex to implement and suffers from high variance.

Dellaportas and Kontoyiannis (2012) noticed that if R is reversible w.r.t. π, it is possible to optimize the limiting variance (25) without computing explicitly the Poisson solution $\hat{f}_\mathrm{d}$. This approach is of course closely related with our proposed method: the reversibility of the Markov kernel is replaced here by the self-adjointness of the generator of the Langevin diffusion which implies the reversibility of the semi-group.

Each of the algorithms in the aforementioned literature requires computation of $R\psi_i$ for each $i \in \{1, \ldots, p\}$, which is in general difficult except in very specific examples. In Henderson (1997); Meyn (2008) this is addressed by restricting to kernels $R(x, \cdot)$ with finite support for each x. In Dellaportas and Kontoyiannis (2012) the authors consider mainly Gibbs samplers in their numerical examples.

Our methodology is also related to the Zero Variance method proposed by Mira, Solgi and Imparato (2013); Papamarkou, Mira and Girolami (2014); Oates, Girolami and Chopin (2016); South, Mira and Drovandi (2018), which uses $\mathscr{L}g$ as a control variate and chooses g by minimizing $\pi(\{\tilde{f} + \mathscr{L}g\}^2)$. A drawback of this method stems from the fact that the optimization criterion is theoretically justified if $(X_k)_{k\in\mathbb{N}}$ is i.i.d. and might significantly differ from the asymptotic variance $\sigma^2_{\infty,\gamma}(f + \mathscr{L}g)$ defined in (17). We compare the two approaches in Section 4.

3. Asymptotic expansion for the asymptotic variance of MCMC algorithms. In this Section, we provide conditions upon which the approximations (18)-(19) are satisfied for $f \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and $\mathcal{G} \subset \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$. We first assume that the gradient of the potential is smooth:

H1. *$U \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and ∇U is Lipschitz, i.e. there exists $L \geq 0$ such that*

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

Denote by $(P_t)_{t\geq 0}$ the semigroup associated to the SDE (3) defined by $P_t f(x) = \mathbb{E}[f(Y_t)]$ where f is bounded measurable and $(Y_t)_{t\geq 0}$ is a solution of (3) started at x. By construction, the target distribution π is invariant for $(P_t)_{t\geq 0}$.

The conditions we consider require that $\{R_\gamma,\ \gamma \in (0, \bar{\gamma}]\}$ is a family of Markov kernels such that for any $\gamma \in (0, \bar{\gamma}]$, R_γ approximates P_γ in a sense specified below. Let $V : \mathbb{R}^d \to [1, +\infty)$ be a measurable function.

H2. *(i) For any $\gamma \in (0, \bar{\gamma}]$, R_γ has a unique invariant distribution π_γ.*

(ii) *There exists* $c>0$ *such that* $\liminf_{\|x\|\to\infty}\{V(x)\exp(-c\|x\|)\}>0$, $\pi(V)<+\infty$ *and* $\sup_{\gamma\in(0,\bar{\gamma}]}\pi_\gamma(V)<+\infty$.

(iii) *There exist* $C>0$ *and* $\rho\in[0,1)$ *such that for all* $x\in\mathbb{R}^d$,

$$\text{for any } n\in\mathbb{N},\text{ and } \gamma\in(0,\bar{\gamma}]\,,\qquad \|\delta_x R_\gamma^n-\pi_\gamma\|_V\le C\rho^{n\gamma}V(x)\,,$$
$$\text{for any } t\ge 0\,,\qquad \|\delta_x P_t-\pi\|_V\le C\rho^t V(x)\,. \tag{26}$$

These conditions imply that the kernels R_γ are V-uniformly geometrically ergodic "uniformly" with respect to the parameter $\gamma\in(0,\bar{\gamma}]$ with a mixing time going to infinity as the inverse of the stepsize γ when $\gamma\downarrow 0^+$. Note that the mixing time of P_γ is also inversely proportional to γ when $\gamma\downarrow 0^+$.

Under **H**1 and **H**2, by (Kopec, 2015, Lemma 2.6), there exists a solution $\hat{f}\in\mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$ to Poisson's equation (11) for any $f\in\mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$ which is given for any $x\in\mathbb{R}^d$ by

$$\hat{f}(x)=\int_0^{+\infty}P_t\tilde{f}(x)\mathrm{d}t\,. \tag{27}$$

Moreover, (Cattiaux, Chafaı and Guillin, 2012, Theorem 3.1) shows that, for any $f\in\mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$, $t^{-1/2}\int_0^t\tilde{f}(Y_s)\mathrm{d}s$ where $(Y_t)_{t\ge0}$ is the solution of the Langevin SDE, converges weakly to $\mathcal{N}(0,\sigma^2_\infty(f))$ where $\sigma^2_\infty(f)$ is given by (10).

Note that the assumption **H**2 implies that for any $x\in\mathbb{R}^d$,

$$\text{for any } \gamma\in(0,\bar{\gamma}],\ n\in\mathbb{N}^*\,,\qquad R_\gamma^n V(x)\le C\rho^{n\gamma}V(x)+\sup_{\gamma\in(0,\bar{\gamma}]}\pi_\gamma(V)\,, \tag{28}$$

$$\text{for any } t\ge0\,,\qquad P_tV(x)\le C\rho^tV(x)+\pi(V)\,. \tag{29}$$

We now introduce an assumption guaranteeing that the limit $\gamma^{-1}(R_\gamma-\mathrm{Id})$ as $\gamma\downarrow0^+$ is equal to the infinitesimal generator of the Langevin diffusion defined, for a bounded measurable function f and $x\in\mathbb{R}^d$, as $\mathscr{L}f(x)=\lim_{t\to+\infty}\{(P_tf(x)-f(x))/t\}$, if the limit exists. This is a natural assumption if the semigroup of the Langevin diffusion evaluated at time $t=\gamma$, P_γ, and R_γ are close as $\gamma\downarrow0^+$.

H3. *There exist* $\alpha\ge3/2$ *and a family of operators* $(\mathscr{E}_\gamma)_{\gamma\in(0,\bar{\gamma}]}$ *with* $\mathscr{E}_\gamma:\mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})\to\mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$, *such that for all* $f\in\mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$ *and* $\gamma\in(0,\bar{\gamma}]$,

$$R_\gamma f=f+\gamma\mathscr{L}f+\gamma^\alpha\mathscr{E}_\gamma f\,. \tag{30}$$

In addition, there exist $k_e\in\mathbb{N}$, $k_e\ge2$ *such that for all* $p\in\mathbb{N}$ *there exist* $q\in\mathbb{N}$ *and* $C\ge0$ *(depending only on* k_e,p*) such that for any* $f\in\mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$,

$$\sup_{\gamma\in(0,\bar{\gamma}]}\|\mathscr{E}_\gamma f\|_{0,q}\le C\|f\|_{k_e,p}\,.$$

We show below that these conditions are satisfied for the Metropolis Adjusted / Unadjusted Langevin Algorithm (MALA and ULA) algorithms (in which case γ is the stepsize in the Euler discretization of the Langevin diffusion) and also by the Random Walk Metropolis algorithm (RWM) (in which case γ is the variance of the increment distribution). We next give an upper bound on the difference between π_γ and π which implies that (19) holds. The proofs are postponed to Section 5.

PROPOSITION 1. *Assume **H**1, **H**2 and **H**3 and let $p \in \mathbb{N}$. Then there exists $C < \infty$ such that for all $f \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and $\gamma \in (0, \bar{\gamma}]$,*

$$|\pi_\gamma(f) - \pi(f)| \leq C \|f\|_{k_e,p} \gamma^{\alpha-1} \; . \tag{31}$$

PROOF. The proof is postponed to Section 5.1. □

The next result which is the main theorem of this Section precisely formalizes (18).

THEOREM 2. *Assume **H**1, **H**2 and **H**3. Then, there exists $C \geq 0$ such that for all $f \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $\gamma \in (0, \bar{\gamma}]$, $x \in \mathbb{R}^d$, and $n \in \mathbb{N}^*$*

$$\begin{aligned} \left| \frac{\gamma}{n} \mathbb{E}_{x,\gamma} \left[\left(\sum_{k=0}^{n-1} \{ f(X_k) - \pi_\gamma(f) \} \right)^2 \right] - \sigma^2_\infty(f) \right| & \\ \leq C \|f\|^2_{k_e+2,p} \left\{ \gamma^{(\alpha-1)\wedge 1} + V(x)/(n^{1/2}\gamma^{1/2}) \right\} \; , & \end{aligned} \tag{32}$$

where $\sigma^2_\infty(f)$ is defined in (10).

PROOF. The proof is postponed to Section 5.2. □

We now consider the ULA algorithm. The Markov kernel R^{ULA}_γ associated to the ULA algorithm is given for $\gamma > 0$, $x \in \mathbb{R}^d$ and $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$ by

$$R^{\mathrm{ULA}}_\gamma(x, \mathsf{A}) = \int_{\mathbb{R}^d} \mathbb{1}_{\mathsf{A}} \left(x - \gamma \nabla U(x) + \sqrt{2\gamma} z \right) \boldsymbol{\varphi}(z) \mathrm{d}z \; , \tag{33}$$

where $\boldsymbol{\varphi}$ is the d-dimensional standard Gaussian density $\boldsymbol{\varphi} : z \mapsto (2\pi)^{-d/2} \mathrm{e}^{-\|z\|^2}$. Consider the following additional assumption.

H 4. *There exist $K_1 \geq 0$ and $m > 0$ such that for any $x \notin \mathrm{B}(0, K_1)$, and $y \in \mathbb{R}^d$, $\langle \mathrm{D}^2 U(x) y, y \rangle \geq m \|y\|^2$. Moreover, there exists $M \geq 0$ such that for any $x \in \mathbb{R}^d$, $\|\mathrm{D}^3 U(x)\| \leq M$.*

PROPOSITION 3. *Assume **H**1 and **H**4. There exist $\bar{\gamma} > 0$ and $V : \mathbb{R}^d \to [0, +\infty)$ such that **H**2 is satisfied for the family of Markov kernels $\{R^{\mathrm{ULA}}_\gamma : \gamma \in (0, \bar{\gamma}]\}$.*

PROOF. The proof is postponed to Section 6.1. □

REMARK 4. *Note that **H**2 holds under milder conditions on U using the results obtained in Eberle (2015); Eberle and Majka (2018); De Bortoli and Durmus (2019). For example, if **H**1 holds and there exist $a_1, a_2 > 0$ and $c \geq 0$ such that*

$$\langle \nabla U(x), x \rangle \geq a_1 \|x\| + a_2 \|\nabla U(x)\|^2 - c \,, \tag{34}$$

*(De Bortoli and Durmus, 2019, Theorem 14, Proposition 24) imply that **H**2 holds with $V(x) = \exp\{(a_1/8)(1 + \|x\|^2)^{1/2}\}$.*

We now establish **H**3. Let $\varphi \in \mathrm{C}^\infty_{\text{poly}}(\mathbb{R}^d, \mathbb{R})$, $\bar{\gamma} > 0$, $\gamma \in [0, \bar{\gamma}]$ and $x \in \mathbb{R}^d$. Using $X_1 = X_0 - \gamma \nabla U(X_0) + \sqrt{2\gamma} Z_1$ where Z_1 is an i.i.d. standard d-dimensional Gaussian vector. A Taylor expansion of $\varphi(X_1)$ at x and integration show that $R^{\text{ULA}}_\gamma \varphi(x) = \varphi(x) + \gamma \mathscr{L} \varphi(x) + \gamma^2 \mathscr{E}^{\text{ULA}}_\gamma \varphi(x)$ where,

$$\begin{aligned} \mathscr{E}^{\text{ULA}}_\gamma \varphi(x) &= \frac{1}{2} \mathrm{D}^2 \varphi(x)[\nabla U(x)^{\otimes 2}] - \frac{1}{6} \gamma \, \mathrm{D}^3 \varphi(x)[\nabla U(x)^{\otimes 3}] - \mathbb{E}\left[\mathrm{D}^3 \varphi(x)[\nabla U(x), Z^{\otimes 2}]\right] \\ &+ \frac{1}{6} \int_0^1 (1-t)^3 \mathbb{E}\left[\mathrm{D}^4 \varphi(x - t\gamma \nabla U(x) + t\sqrt{2\gamma} Z)[(-\sqrt{\gamma} \nabla U(x) + \sqrt{2} Z)^{\otimes 4}]\right] \mathrm{d}t \,. \end{aligned} \tag{35}$$

By the dominated convergence theorem, the last term in the RHS of the previous equation goes to zero and we have therefore established the following result:

LEMMA 5. *Assume **H**1. Then for any $\bar{\gamma} > 0$, $\left\{R^{\text{ULA}}_\gamma : \gamma \in (0, \bar{\gamma}]\right\}$ satisfies **H**3 with $\mathscr{E}_\gamma = \mathscr{E}^{\text{ULA}}_\gamma$ and $k_e = 4$.*

We now examine the MALA algorithm. The Markov kernel R^{MALA}_γ of the MALA algorithm, see Roberts and Tweedie (1996), is given for $\gamma > 0$, $x \in \mathbb{R}^d$, and $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$ by

$$\begin{aligned} R^{\text{MALA}}_\gamma(x, \mathsf{A}) = &\int_{\mathbb{R}^d} \mathbb{1}_{\mathsf{A}} \left(x - \gamma \nabla U(x) + \sqrt{2\gamma} z\right) \min(1, \mathrm{e}^{-\tau^{\text{MALA}}_\gamma(x,z)}) \boldsymbol{\varphi}(z) \mathrm{d}z \\ &+ \updelta_x(\mathsf{A}) \int_{\mathbb{R}^d} \left\{1 - \min(1, \mathrm{e}^{-\tau^{\text{MALA}}_\gamma(x,z)})\right\} \boldsymbol{\varphi}(z) \mathrm{d}z \,, \end{aligned} \tag{36}$$

$$\begin{aligned} \tau^{\text{MALA}}_\gamma(x, z) = &\, U(x - \gamma \nabla U(x) + \sqrt{2\gamma} z) - U(x) \\ &+ \frac{\left\|z - (\gamma/2)^{1/2} \left\{\nabla U(x) + \nabla U(x - \gamma \nabla U(x) + \sqrt{2\gamma} z)\right\}\right\|^2 - \|z\|^2}{2} \,. \end{aligned} \tag{37}$$

The analysis of the MALA algorithm is closely related to the study of the ULA algorithm. Indeed, the difference between the two Markov kernels can be expressed for any bounded measurable function $\phi : \mathbb{R}^d \to \mathbb{R}$ by

$$\begin{aligned} R^{\text{ULA}}_\gamma \phi(x) - R^{\text{MALA}}_\gamma \phi(x) = &\int_{\mathbb{R}^d} \{\phi(x - \gamma \nabla U(x) + \sqrt{2\gamma} z) - \phi(x)\} \\ &\times \{1 - \min(1, \mathrm{e}^{-\tau^{\text{MALA}}_\gamma(x,z)})\} \boldsymbol{\varphi}(z) \mathrm{d}z \,. \end{aligned} \tag{38}$$

Comparison of ULA and MALA relies on the analysis of the acceptance ratio, and more precisely on $\tau_\gamma^{\mathrm{MALA}}$. Combining Proposition 3 and a detailed study of this term enables us to show **H**2 for the MALA algorithm.

PROPOSITION 6. *Assume **H**1 and **H**4. There exist $\bar{\gamma} > 0$ and $V : \mathbb{R}^d \to [0, +\infty)$ such that **H**2 is satisfied for the family of Markov kernels $\{R_\gamma^{\mathrm{MALA}} : \gamma \in (0, \bar{\gamma}]\}$.*

PROOF. The proof is postponed to Section 6.2. □

We now check **H**3. Using (38) and the analysis of ULA (see (35) and Lemma 5), we get for any $\varphi \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $\bar{\gamma} > 0$, $\gamma \in [0, \bar{\gamma}]$ and $x \in \mathbb{R}^d$, $R_\gamma^{\mathrm{MALA}}\varphi(x) = \varphi(x) + \gamma \mathscr{L}\varphi(x) + \gamma^2 \mathscr{E}_\gamma^{\mathrm{MALA}}\varphi(x)$, with $\mathscr{E}_\gamma^{\mathrm{MALA}}\varphi(x) = \mathscr{E}_\gamma^{\mathrm{ULA}}\varphi(x) + \tilde{\mathscr{E}}_\gamma^{\mathrm{MALA}}\varphi(x)$ and

$$
\begin{aligned}
(39)\quad \tilde{\mathscr{E}}_\gamma^{\mathrm{MALA}}\varphi(x) = \mathbb{E}\Bigg[&\gamma^{-3/2}\left\{1 - \min(1, \mathrm{e}^{-\tau_\gamma^{\mathrm{MALA}}(x,Z)})\right\} \\
&\times \left\{\int_0^1 \left\langle \nabla\varphi(x - t\gamma\nabla U(x) + t\sqrt{2\gamma}Z), \sqrt{\gamma}\nabla U(x) - \sqrt{2}Z \right\rangle \mathrm{d}t \right\}\Bigg] .
\end{aligned}
$$

By expanding the acceptance ratio (see Section 6), it is straightforward to show that

LEMMA 7. *Assume **H**1 and **H**4. Then, for any $\bar{\gamma} > 0$, there exists $C_{1,\bar{\gamma}} < \infty$ such that for any $x, z \in \mathbb{R}^d$, $\gamma \in (0, \bar{\gamma}]$, it holds*

$$
\left|\tau_\gamma^{\mathrm{MALA}}(x, z)\right| \le C_{1,\bar{\gamma}}\gamma^{3/2}\{1 + \|z\|^4 + \|x\|^2\} .
$$

Combining this result with (39) shows that

LEMMA 8. *Assume **H**1 and **H**4. Then for any $\bar{\gamma} > 0$, $\{R_\gamma^{\mathrm{MALA}} : \gamma \in (0, \bar{\gamma}]\}$ satisfies **H**3 with $\mathscr{E}_\gamma = \mathscr{E}_\gamma^{\mathrm{MALA}}$ and $k_e = 4$.*

In (Brosse et al., 2019, Section **??**), we establish similar results for the RWM algorithm: In this case γ is the variance of the increment distribution.

4. Numerical experiments. In this Section, we compare numerically our methodology with the Zero Variance method suggested by Mira, Solgi and Imparato (2013), see Section 2.2, that consists in minimizing the marginal variance $\min_{g\in\mathcal{G}} \pi(\{\tilde{f} + \mathscr{L}g\}^2)$ instead of the asymptotic variance $\min_{g\in\mathcal{G}} \sigma_\infty^2(f + \mathscr{L}g)$. In Sections 4.1 and 4.2, we first focus on one and two dimensional examples where explicit calculations are possible. In Section 4.3, we study Bayesian logistic and probit regressions. The code used to run the experiments is available at `https://github.com/nbrosse/controlvariates`.

4.1. *One dimensional example.* We consider an equally weighted mixture of two Gaussian densities of means $(\mu_1, \mu_2) = (-1, 1)$ and variance $\sigma^2 = 1/2$, and a test function $f(x) = x + x^3/2 + 3\sin(x)$. The derivative of the Poisson equation (11) is in such case analytically known: $\hat{f}'(x) = -(1/\pi(x)) \int_{-\infty}^x \pi(t)\tilde{f}(t)\mathrm{d}t$, see (Brosse et al., 2019, Section S2.1) for a practical implementation.

We build a control variate $g_\theta \in \mathcal{G}_{\mathrm{lin}} = \{\langle \theta, \psi \rangle : \theta \in \mathbb{R}^p\}$ where $\psi = (\psi_1, \ldots, \psi_p)$ are p Gaussian kernels regularly spaced on $[-4, 4]$, *i.e.* for all $i \in \{1, \ldots, p\}$ and $x \in \mathbb{R}$

$$\psi_i(x) = (2\pi)^{-1/2} \mathrm{e}^{-(x-\mu_i)^2/2} \;, \quad \text{where } \mu_i \in [-4, 4] \;. \tag{40}$$

The optimal parameter $\theta^* \in \mathbb{R}^p$ minimizing the asymptotic variance $\sigma_\infty^2(f + \mathscr{L} g_\theta)$ can be explicitly computed according to (15). For the Zero Variance estimator, the optimal parameter is given by

$$\theta_{\mathrm{zv}}^* = -H_{\mathrm{zv}}^{-1} b_{\mathrm{zv}} \;, \tag{41}$$

where for $1 \leq i, j \leq p$, $[H_{\mathrm{zv}}]_{ij} = \pi(\langle \mathscr{L}\psi_i, \mathscr{L}\psi_j \rangle)$ and $[b_{\mathrm{zv}}]_i = \pi(\tilde{f} \mathscr{L} \psi_i)$. H_{zv} is invertible if $(\mathscr{L}\psi_1, \ldots, \mathscr{L}\psi_p)$ are linearly independent in $\mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$.

The asymptotic variance $\sigma_\infty^2(f + \mathscr{L} g_\theta)$ for the two different parameters, θ^* and θ_{zv}^* are compared against the number of Gaussian kernels $p \in \{4, \ldots, 10\}$ in Figure 1. Note that the asymptotic variance $\sigma_\infty^2(f)$ is 92.5. We observe that the variance reduction is better for an even number p of basis functions; when $p \geq 8$, the two methods achieve an almost identical large variance reduction. These results are supported by the plots of g_θ' and $\mathscr{L} g_\theta$ for $\theta \in \{\theta^*, \theta_{\mathrm{zv}}^*\}$ in Figure 1, see also (Brosse et al., 2019, Section S2.1).

We fix the number of basis functions $p = 4$ and we now turn to the application to MCMC algorithms. We first define the sample mean with a burn-in period $N \in \mathbb{N}^*$ by

$$\hat{\pi}_{N,n}(f) = \frac{1}{n} \sum_{k=N}^{N+n-1} f(X_k) \;, \tag{42}$$

where $n \in \mathbb{N}^*$ is the number of samples. In this Section, we consider the following estimators of $\pi(f)$: $\hat{\pi}_{N,n}(f + \mathscr{L}\langle \theta^*, \psi \rangle)$ and $\hat{\pi}_{N,n}(f + \mathscr{L}\langle \theta_{\mathrm{zv}}^*, \psi \rangle)$ where θ^* and θ_{zv}^* are given in (15) and (41) respectively. In this simple one dimensional example, the optimal parameters θ^* and θ_{zv}^* are explicitly computable; the problem of estimating them in higher dimensional models is addressed numerically in Section 4.3.

The sequence $(X_k)_{k \in \mathbb{N}}$ is generated by the ULA, MALA or RWM algorithms starting at 0, with a step size $\gamma = 10^{-2}$ for ULA and $\gamma = 5 \times 10^{-2}$ for RWM and MALA, a burn-in period $N = 10^5$ and a number of samples $n = 10^6$. For a test function $h : \mathbb{R} \to \mathbb{R}$ ($h = f + \mathscr{L}\langle \theta, \psi \rangle$, $\theta \in \{0, \theta^*, \theta_{\mathrm{zv}}^*\}$), we estimate the asymptotic variance $\sigma_{\infty,\gamma}^2(h)$ of $\hat{\pi}_{N,n}(h)$ by a spectral estimator $\hat{\sigma}_{N,n}^2(h)$ with a Tukey-Hanning window, see Flegal and Jones (2010), given by

$$\hat{\sigma}_{N,n}^2(h) = \sum_{k=-(\lfloor n^{1/2} \rfloor - 1)}^{\lfloor n^{1/2} \rfloor - 1} \left\{ \frac{1}{2} + \frac{1}{2} \cos\left(\frac{\pi \, |k|}{\lfloor n^{1/2} \rfloor} \right) \right\} \omega_{N,n}^h(|k|) \;, \tag{43}$$

$$\omega_{N,n}^h(k) = \frac{1}{n} \sum_{s=N}^{N+n-1-k} \{h(X_s) - \hat{\pi}_{N,n}(h)\} \{h(X_{s+k}) - \hat{\pi}_{N,n}(h)\} \;.$$

We compute the average of these estimators $\hat{\sigma}_{N,n}^2(f + \mathscr{L}\langle \theta, \psi \rangle)$, $\theta \in \{0, \theta^*, \theta_{\mathrm{zv}}^*\}$ over 10 independent runs of the Monte Carlo algorithm (ULA, RWM or MALA), see Table 1.

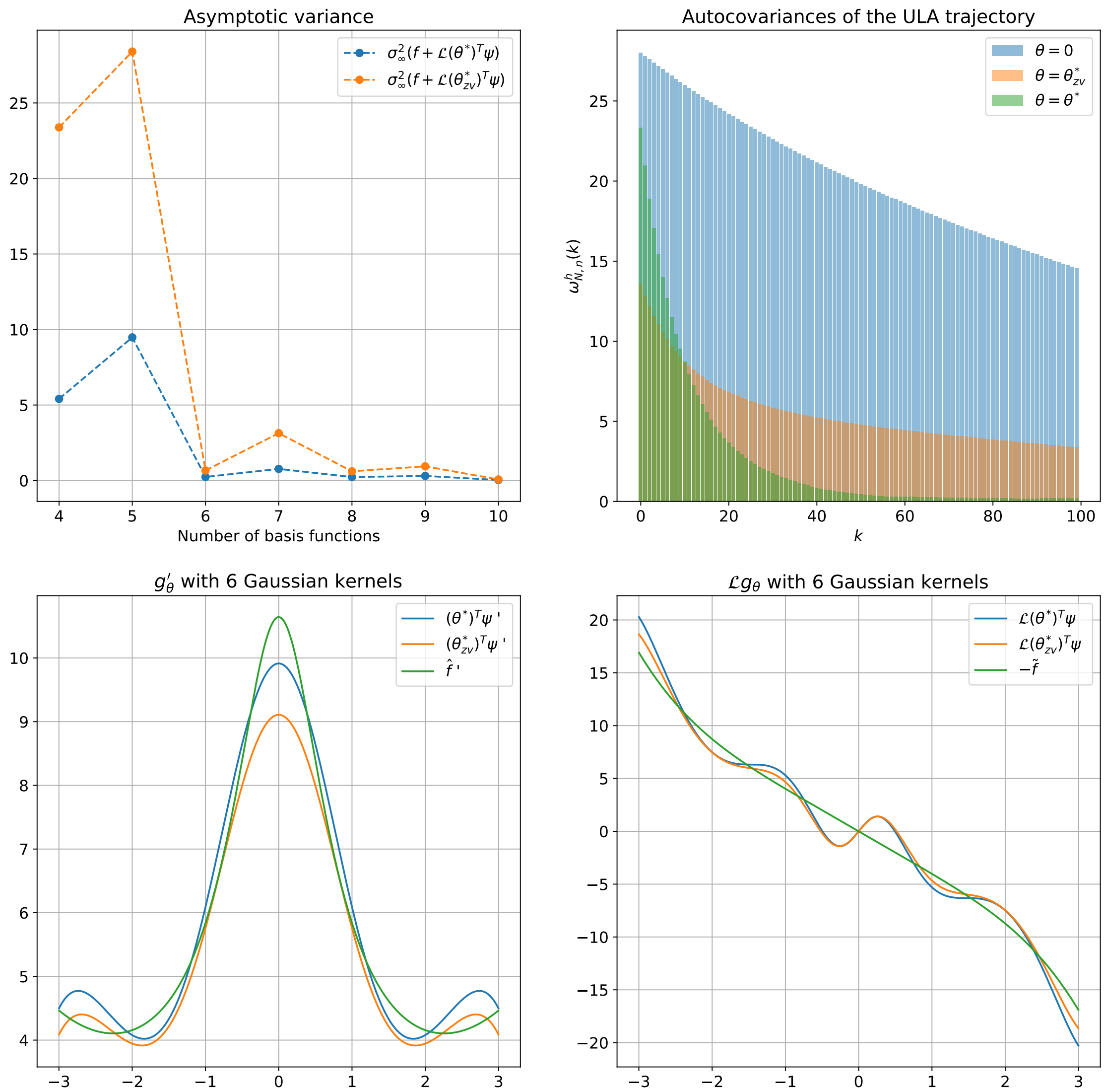

Figure 1: **Top Left.** Plot of the asymptotic variance $\sigma^2_\infty(f + \mathscr{L}g_\theta)$ for $\theta \in \{\theta^*, \theta^*_{\rm zv}\}$ and $p \in \{4, \ldots, 10\}$. **Top Right.** Autocovariances plot of ULA displaying $\omega^h_{N,n}(k)$ for $h = f + \mathscr{L}\langle\theta, \psi\rangle$, $\theta \in \{0, \theta^*, \theta^*_{\rm zv}\}$ and $0 \le k < 100$. **Bottom Left and Right.** Plots of g'_θ and $\mathscr{L}g_\theta$ for $\theta \in \{\theta^*, \theta^*_{\rm zv}\}$ and $p = 6$.

	$\gamma\hat{\sigma}^2_{N,n}(f)$	$\gamma\hat{\sigma}^2_{N,n}(f+\mathscr{L}\langle\theta^*_{\mathrm{zv}},\psi\rangle)$	$\gamma\hat{\sigma}^2_{N,n}(f+\mathscr{L}\langle\theta^*,\psi\rangle)$
ULA	82.06	20.74	5.33
RWM	105.2	28.19	8.41
MALA	93.27	23.40	5.00

TABLE 1

Values of $\hat{\sigma}^2_{N,n}(f+\mathscr{L}\langle\theta,\psi\rangle)$, $\theta\in\{0,\theta^*,\theta^*_{\mathrm{zv}}\}$ rescaled by the step size γ.

We observe that minimizing the asymptotic variance improves upon the Zero Variance estimator.

A more detailed analysis can be carried out using the autocovariances plots that consist in displaying $\omega^h_{N,n}(k)$ for $h=f+\mathscr{L}\langle\theta,\psi\rangle$, $\theta\in\{0,\theta^*,\theta^*_{\mathrm{zv}}\}$ and $0\leq k<100$, see Figure 1. The autocovariances plots for RWM and MALA are similar. By (Douc et al., 2018, Theorem 21.2.11), the asymptotic variance $\sigma^2_{\infty,\gamma}(h)$ is the sum of the autocovariances:

$$\sigma^2_{\infty,\gamma}(h)=\pi_\gamma(\tilde{h}^2_\gamma)+2\sum_{k=1}^{+\infty}\pi_\gamma(\tilde{h}_\gamma R^k_\gamma\tilde{h}_\gamma)\ ,\quad \text{where } \tilde{h}_\gamma=h-\pi_\gamma(h)\ .$$

The two methods are effective at reducing the autocovariances compared to the case without control variate. The zero variance estimator decreases more the autocovariance at $k=0$ compared to our method, which is indeed the objective of θ^*_{zv}, the minimizer of $\theta\mapsto\pi((\tilde{f}+\mathscr{L}\langle\theta,\psi\rangle)^2)$. Using $\theta=\theta^*$ lowers more effectively the tail of the autocovariances (for k large enough) compared to $\theta=\theta^*_{\mathrm{zv}}$. This effect is predominant and explains the results of Table 1.

4.2. *Two dimensional example.* We conduct a similar study in $\mathbb{R}^2$ where we consider an equally weighted mixture of two Gaussian densities with means $(\mu_1,\mu_2)=([-1,0],[1,0])$, diagonal covariance $\sigma^2=0.3$ and a test function $f(x)=x_1+x_2^3/2+\sin(x_1)+\cos(x_2)$ for $x=(x_1,x_2)\in\mathbb{R}^2$. The solution of the Poisson equation $\hat{f}$ defined in (11) is computed numerically and $\sigma^2_\infty(f)=13.52$, see (Brosse et al., 2019, Section S2.1) for a practical implementation. We consider a control variate $g_\theta\in\mathcal{G}_{\mathrm{lin}}=\left\{\langle\theta,\psi\rangle:\theta\in\mathbb{R}^{100}\right\}$ where $\psi=(\psi_1,\ldots,\psi_{100})$ are 100 Gaussian kernels regularly spaced on $[-3,3]\times[-2,2]$ of covariance Id, *i.e.* for all $i\in\{1,\ldots,100\}$ and $x\in\mathbb{R}^2$

$$\psi_i(x)=(2\pi)^{-1}\mathrm{e}^{-\|x-\mu_i\|^2/2}\ ,\quad \text{where } \mu_i\in[-3,3]\times[-2,2]\ . \tag{44}$$

We compute the optimal parameter θ^* and the zero variance parameter θ^*_{zv} by solving (15) and (41) respectively, truncating the smallest eigenvalues. The asymptotic variances of the Langevin diffusion corresponding to the different methodologies are reported in Table 2; $\theta=\theta^*$ achieves a better variance reduction. These results are corroborated by the heatmaps of $\sqrt{\pi}\nabla\hat{f}$, $\sqrt{\pi}\nabla\langle\theta^*,\psi\rangle$ and $\sqrt{\pi}\nabla\langle\theta^*_{\mathrm{zv}},\psi\rangle$ in Figure 2.

Turning to the MCMC application, we use the same setting as the one dimensional case. The asymptotic variance of $\hat{\pi}_{N,n}(f+\mathscr{L}\langle\theta,\psi\rangle)$ defined in (42) for $\theta\in\{0,\theta^*,\theta^*_{\mathrm{zv}}\}$ is estimated by $\hat{\sigma}^2_{N,n}(f+\mathscr{L}\langle\theta,\psi\rangle)$ given in (43). We compute the average of $\hat{\sigma}^2_{N,n}(f+\mathscr{L}\langle\theta,\psi\rangle)$, $\theta\in\{0,\theta^*,\theta^*_{\mathrm{zv}}\}$ over 10 independent runs of ULA, RWM and MALA algorithms, see Table 2. We observe that the two control variates reduce the asymptotic

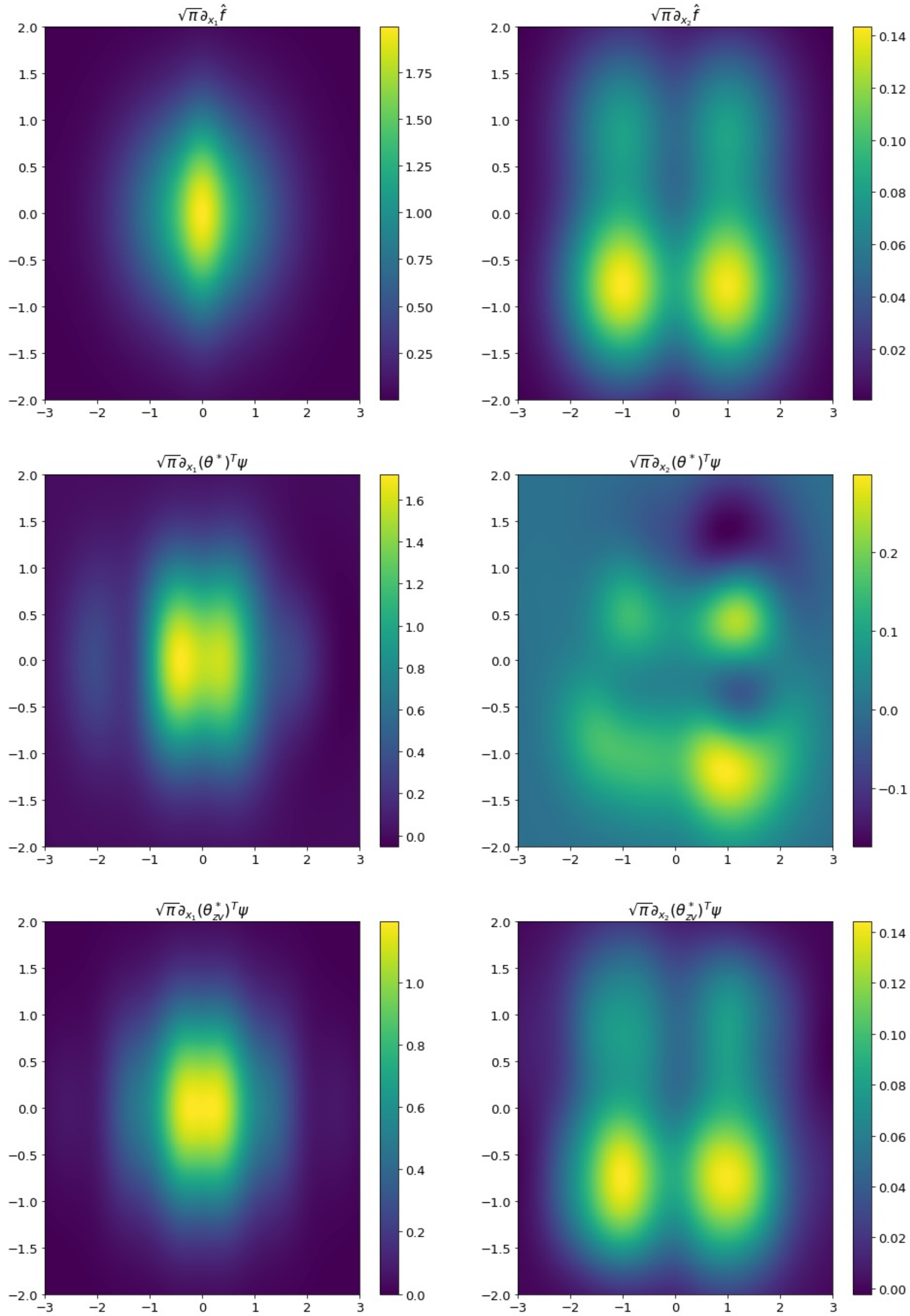

Figure 2: Heatmaps of $\sqrt{\pi}\nabla\hat{f}$, $\sqrt{\pi}\nabla\langle\theta^*,\psi\rangle$ and $\sqrt{\pi}\nabla\langle\theta^*_{\mathrm{zv}},\psi\rangle$.

	$\sigma^2_\infty(f)$	$\sigma^2_\infty(f + \mathscr{L}\langle\theta^*_{\mathrm{zv}}, \psi\rangle)$	$\sigma^2_\infty(f + \mathscr{L}\langle\theta^*, \psi\rangle)$
	13.52	1.33	0.70
	$\gamma\hat{\sigma}^2_{N,n}(f)$	$\gamma\hat{\sigma}^2_{N,n}(f + \mathscr{L}\langle\theta^*_{\mathrm{zv}}, \psi\rangle)$	$\gamma\hat{\sigma}^2_{N,n}(f + \mathscr{L}\langle\theta^*, \psi\rangle)$
ULA	11.95	1.24	0.73
RWM	18.38	1.77	1.40
MALA	14.22	1.55	0.77

TABLE 2

Values of $\sigma^2_\infty(f + \mathscr{L}\langle\theta, \psi\rangle)$ and $\hat{\sigma}^2_{N,n}(f + \mathscr{L}\langle\theta, \psi\rangle)$, $\theta \in \{0, \theta^*, \theta^*_{\mathrm{zv}}\}$ rescaled by the step size γ.

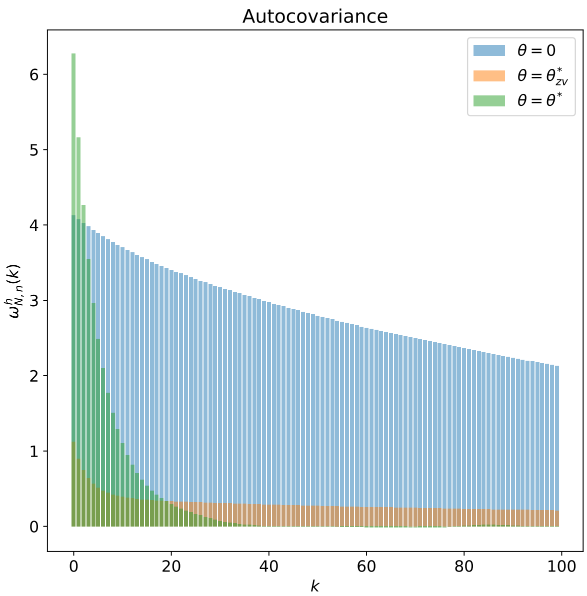

Figure 3: Autocovariances plot of the ULA trajectory for the two dimensional example displaying $\omega^h_{N,n}(k)$ for $h = f + \mathscr{L}\langle\theta, \psi\rangle$, $\theta \in \{0, \theta^*, \theta^*_{\mathrm{zv}}\}$ and $0 \le k < 100$.

variance and our method performs better. The autocovariances plot for ULA is displayed in Figure 3 (similar graphs are obtained by the RWM and MALA algorithms). The zero variance estimator decreases more the autocovariance at $k = 0$ compared to our method. Using $\theta = \theta^*$ lowers more effectively the tail of the autocovariances (for k large enough) compared to $\theta = \theta^*_{\mathrm{zv}}$.

4.3. *Bayesian logistic and probit regressions.* We illustrate the proposed control variates method on Bayesian logistic and probit regressions, see (Gelman et al., 2014, Chapter 16), (Marin and Robert, 2007, Chapter 4). The examples and the data sets are taken from Papamarkou, Mira and Girolami (2014). Let $\mathsf{Y} = (\mathsf{Y}_1, \ldots, \mathsf{Y}_n) \in \{0,1\}^N$ be a vector of binary response variables, $x \in \mathbb{R}^d$ be the regression coefficients, and $\mathsf{X} \in \mathbb{R}^{N \times d}$ be a design matrix. The log-likelihood for the logistic and probit regressions are given respectively by

$$\ell_{\log}(\mathsf{Y}|x,\mathsf{X}) = \sum_{i=1}^{N} \left\{ \mathsf{Y}_i \mathsf{X}_i^{\mathrm{T}} x - \ln\left(1 + \mathrm{e}^{\mathsf{X}_i^{\mathrm{T}} x}\right) \right\} ,$$

$$\ell_{\mathrm{pro}}(\mathsf{Y}|x,\mathsf{X}) = \sum_{i=1}^{N} \left\{ \mathsf{Y}_i \ln(\Phi(\mathsf{X}_i^{\mathrm{T}} x)) + (1-\mathsf{Y}_i) \ln(\Phi(-\mathsf{X}_i^{\mathrm{T}} x)) \right\} ,$$

where $\mathsf{X}_i^{\mathrm{T}}$ is the i^{th} row of X for $i \in \{1,\ldots,N\}$. For both models, a Gaussian prior of mean 0 and variance $\varsigma^2\,\mathrm{Id}$ is assumed for x where $\varsigma^2 = 100$. The unnormalized posterior probability distributions $\pi_{\log}$ and π_{pro} for the logistic and probit regression models are defined for all $x \in \mathbb{R}^d$ by

$$\pi_{\log}(x|\mathsf{Y},\mathsf{X}) \propto \exp\left(-U_{\log}(x)\right) \quad \text{with} \quad U_{\log}(x) = -\ell_{\log}(\mathsf{Y}|x,\mathsf{X}) + (2\varsigma^2)^{-1} \|x\|^2 ,$$

$$\pi_{\mathrm{pro}}(x|\mathsf{Y},\mathsf{X}) \propto \exp\left(-U_{\mathrm{pro}}(x)\right) \quad \text{with} \quad U_{\mathrm{pro}}(x) = -\ell_{\mathrm{pro}}(\mathsf{Y}|x,\mathsf{X}) + (2\varsigma^2)^{-1} \|x\|^2 .$$

The following lemma enables to check the assumptions on $U_{\log}$ and U_{pro} required to apply Theorem 2 for the ULA, MALA and RWM algorithms.

LEMMA 9. *$U_{\log}$ and U_{pro} satisfy* ***H1****,* (34)*,* ***H4*** *and* ***S1****.*

PROOF. The proof is postponed to (Brosse et al., 2019, Section S2.2). □

Following (Papamarkou, Mira and Girolami, 2014, Section 2.1), we compare two bases for the construction of a control variate, based on first and second degree polynomials and denoted by $\psi^{1\mathrm{st}} = (\psi_1^{1\mathrm{st}},\ldots,\psi_d^{1\mathrm{st}})$ and $\psi^{2\mathrm{nd}} = (\psi_1^{2\mathrm{nd}},\ldots,\psi_{d(d+3)/2}^{2\mathrm{nd}})$ respectively, see (Brosse et al., 2019, Section S2.3) for their definitions. The estimators associated to $\psi^{1\mathrm{st}}$ and $\psi^{2\mathrm{nd}}$ are referred to as CV-1 and CV-2, respectively.

For the ULA, MALA and RWM algorithms, we make a run of $n = 10^6$ samples with a burn-in period of 10^5 samples, started at the mode of the posterior. The step size is set equal to 10^{-2} for ULA and to 5×10^{-2} for MALA and RWM: with these step sizes, the average acceptance ratio in the stationary regime is equal to 0.23 for RWM and 0.57 for MALA, see Roberts, Gelman and Gilks (1997); Roberts and Rosenthal (1998). We consider $2d$ scalar test functions $\{f_k\}_{k=1}^{2d}$ defined for all $x \in \mathbb{R}^d$ and $k \in \{1,\ldots,d\}$ by $f_k(x) = x_k$ and $f_{k+d}(x) = x_k^2$.

Contrary to the one and two dimensional cases handled in Section 4.1, the optimal parameters θ^* and θ^*_{zv} corresponding to our method and to the zero variance estimator can not be computed in closed form and must be estimated. We consider then the control variate estimator $\pi_{N,n,n}^{\mathrm{CV}}(f)$ defined in (20) where $m = n$ and $(\tilde{X}_k)_{k\in\mathbb{N}}$ is equal to $(X_k)_{k\in\mathbb{N}}$; θ^* is approximated by θ_n^* given in (24). For $k \in \{1,\ldots,2d\}$, we compute the empirical average $\hat{\pi}_{N,n}(f_k)$ defined in (42) and confront it to $\pi_{N,n,n}^{\mathrm{CV}}(f_k)$. For comparison purposes, the zero-variance estimators of Papamarkou, Mira and Girolami (2014) using the same bases of functions $\psi^{1\mathrm{st}}$, $\psi^{2\mathrm{nd}}$ are also computed and are referred to as ZV-1 for $\psi^{1\mathrm{st}}$ and ZV-2 for $\psi^{2\mathrm{nd}}$.

We run 100 independent Markov chains for ULA, MALA, RWM algorithms. The boxplots for the logistic example are displayed in Figure 4 for x_1 and x_1^2. Note the impressive decrease in the variance using the control variates for each algorithm ULA, MALA and RWM. It is worthwhile to note that for ULA, the bias $|\pi(f) - \pi_\gamma(f)|$ is reduced dramatically using the CV-2 estimator. It can be explained by the fact that for n large enough, $g_{\theta_n^*} = \langle \theta_n^*, \psi^{\mathrm{2nd}} \rangle$ approximates well the solution $\hat{f}$ of the Poisson equation $\mathscr{L}\hat{f} = -\tilde{f}$. We then get

$$\pi_\gamma(f) + \pi_\gamma\left(\mathscr{L} g_{\theta_n^*}\right) \approx \pi_\gamma(f) - \pi_\gamma\left(\tilde{f}\right) = \pi(f) \, .$$

To have a more quantitative estimate of the variance reduction, we compute for each algorithm and test function $h \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, the spectral estimator $\hat{\sigma}^2_{N,n}(h)$ defined in (43) of the asymptotic variance. The average of these estimators $\hat{\sigma}^2_{N,n}(f + \mathscr{L}\langle \theta, \psi \rangle)$ for $\theta \in \{0, \theta_n^*, [\theta_{\mathrm{zv}}^*]_n\}$ over the 100 independent runs of the Markov chains for the logistic regression are reported in Table 3. $[\theta_{\mathrm{zv}}^*]_n$ is an empirical estimator of θ_{zv}^*, see Papamarkou, Mira and Girolami (2014) for its construction. The Variance Reduction Factor (VRF) is defined as the ratio of the asymptotic variances obtained by the ordinary empirical average and the control variate (or zero-variance) estimator. We again observe the considerable decrease of the asymptotic variances using control variates. In this example, our approach produces slightly larger VRFs compared to the zero-variance estimators. We obtain similar results for the probit regression; see the supplementary document (Brosse et al., 2019, Section S2.3).

5. Proofs of Proposition 1 and Theorem 2. In the proof the notation $A(\gamma, n, x, f) \lesssim B(\gamma, n, x, f)$ means that there exist $\bar{\gamma} > 0$, and $C < \infty$ such that for all $f \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $\gamma \in (0, \bar{\gamma}]$, $x \in \mathbb{R}^d$, $n \in \mathbb{N}$, $A(\gamma, n, x, f) \leq C B(\gamma, n, x, f)$.

We preface the proofs by a technical result which follows from (Kopec, 2015, Lemma 2.6, Proposition 2.7) and (27) establishing the regularity of solutions of Poisson's equation.

PROPOSITION 10. *Assume **H**1 and **H**2 and let $k \in \mathbb{N}^*$. For all $f \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, there exists $\hat{f} \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ such that $\mathscr{L}\hat{f} = -\tilde{f}$, where $\tilde{f} = f - \pi(f)$, $\mathscr{L}$ is the generator of the Langevin diffusion defined in (5). In addition, for all $p \in \mathbb{N}$, there exist $C \geq 0$, $q \in \mathbb{N}$ such that for all $f \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $\|\hat{f}\|_{k,q} \leq C\|f\|_{k,p}$.*

5.1. *Proof of Proposition 1.* Let $p \in \mathbb{N}$. Under **H**1 and **H**2, by Proposition 10, there exists $q_1 \in \mathbb{N}$ such that for all $f \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $\|\hat{f}\|_{k_e,q_1} \leq C\|f\|_{k_e,p}$, where $\mathscr{L}\hat{f} = -\tilde{f}$, $\tilde{f} = f - \pi(f)$. Under **H**3, we have for all $\gamma \in (0, \bar{\gamma}]$,

$$R_\gamma \hat{f} = \hat{f} + \gamma \mathscr{L}\hat{f} + \gamma^\alpha \mathscr{E}_\gamma \hat{f} = \hat{f} - \gamma\{f - \pi(f)\} + \gamma^\alpha \mathscr{E}_\gamma \hat{f} \, . \tag{45}$$

Integrating (45) w.r.t. π_γ, we obtain that $\pi_\gamma(f) - \pi(f) = \gamma^{\alpha-1}\pi_\gamma(\mathscr{E}_\gamma \hat{f})$. Under **H**3, there exists $q_2 \in \mathbb{N}$ such that $\|\mathscr{E}_\gamma \hat{f}\|_{0,q_2} \lesssim \|\hat{f}\|_{k_e,q_1}$. By **H**2, we get $|\pi_\gamma(\mathscr{E}_\gamma \hat{f})| \leq \pi_\gamma(|\mathscr{E}_\gamma \hat{f}|) \lesssim \|\mathscr{E}_\gamma \hat{f}\|_{0,q_2}$, which concludes the proof.

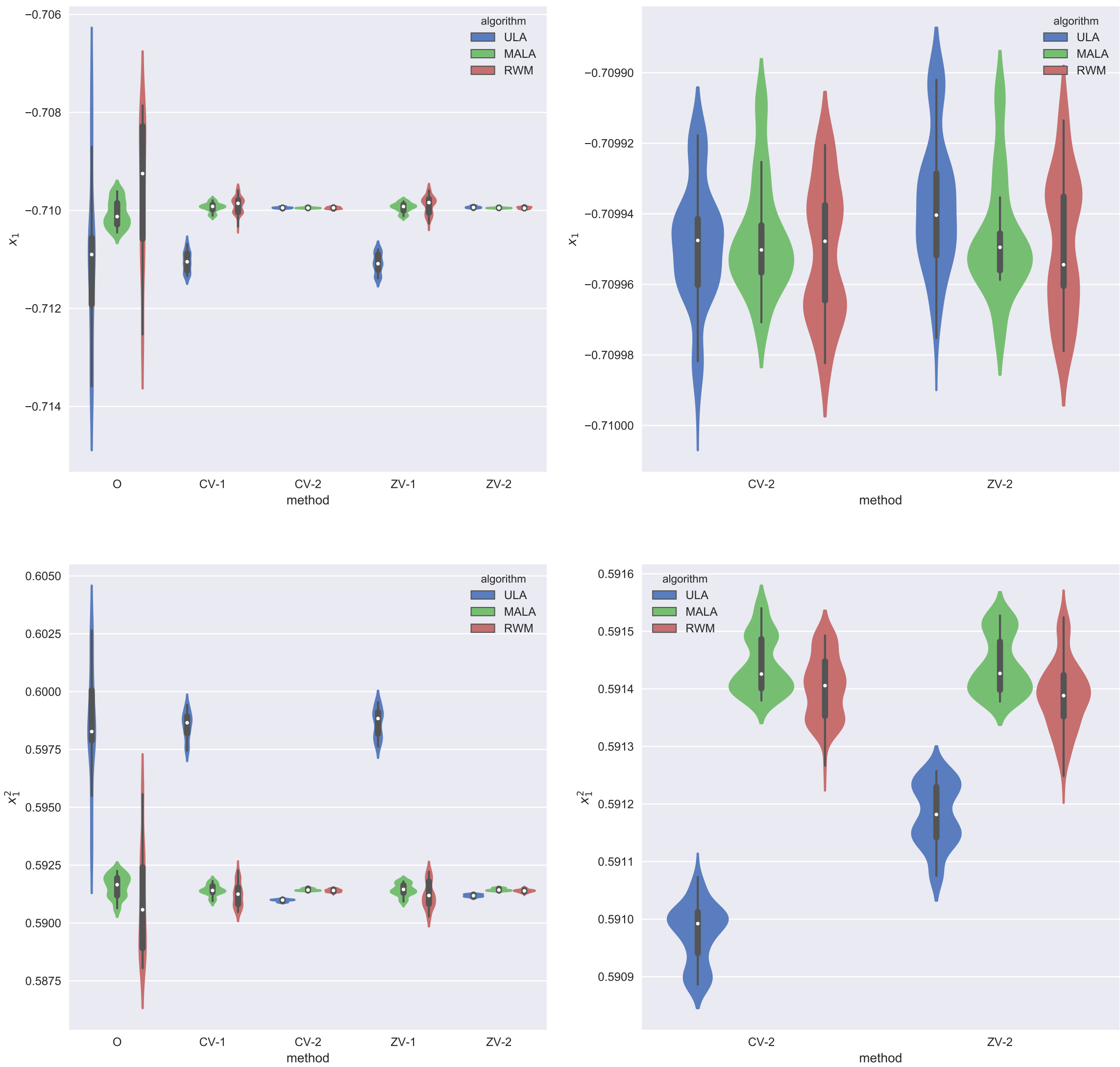

Figure 4: Boxplots of x_1, x_1^2 using the ULA, MALA and RWM algorithms for the logistic regression. The compared estimators are the ordinary empirical average (O), our estimator with a control variate (20) using first (CV-1) or second (CV-2) order polynomials for ψ, and the zero-variance estimators of Papamarkou, Mira and Girolami (2014) using a first (ZV-1) or second (ZV-2) order polynomial bases. The plots in the second column are close-ups for CV-2 and ZV-2.

		MCMC	CV-1-MCMC		CV-2-MCMC		ZV-1-MCMC		ZV-2-MCMC	
		Var.	VRF	Var.	VRF	Var.	VRF	Var.	VRF	Var.
x_1	ULA	2	33	0.061	3.2e+03	6.2e-4	33	0.061	3e+03	6.6e-4
	MALA	0.41	33	0.012	2.6e+03	1.6e-4	30	0.014	2.5e+03	1.7e-4
	RWM	1.3	33	0.039	2.6e+03	4.9e-4	32	0.04	2.7e+03	4.8e-4
x_2	ULA	10	57	0.18	8.1e+03	1.3e-3	53	0.19	7.4e+03	1.4e-3
	MALA	2.5	59	0.042	7.7e+03	3.2e-4	54	0.046	7.3e+03	3.4e-4
	RWM	5.6	52	0.11	5.6e+03	1.0e-3	50	0.11	5.6e+03	1.0e-3
x_2	ULA	10	56	0.18	7.3e+03	1.4e-3	52	0.19	6.7e+03	1.0e-35
	MALA	2.4	58	0.041	6.8e+03	3.5e-4	52	0.045	6.5e+03	3.7e-4
	RWM	5.6	45	0.13	5.1e+03	1.0e-31	42	0.13	5.1e+03	1.0e-31
x_4	ULA	13	26	0.5	3.9e+03	3.3e-3	22	0.59	3.4e+03	3.8e-3
	MALA	3.1	25	0.12	3.6e+03	8.7e-4	21	0.14	3.3e+03	9.5e-4
	RWM	7.5	19	0.4	2.5e+03	3.0e-3	18	0.43	2.4e+03	3.0e-31
x_1^2	ULA	4.6	10	0.46	5.5e+02	8.4e-3	9.3	0.49	4.8e+02	9.5e-3
	MALA	0.98	9.6	0.1	4.6e+02	2.1e-3	8.6	0.11	4.2e+02	2.3e-3
	RWM	3	8.3	0.36	4.3e+02	6.9e-3	8	0.37	4.3e+02	6.9e-3
x_2^2	ULA	29	11	2.6	5.2e+02	0.055	10	2.8	4.7e+02	0.062
	MALA	7	11	0.64	5.2e+02	0.013	10	0.68	4.8e+02	0.014
	RWM	16	9.1	1.8	4.4e+02	0.037	8.8	1.8	4.3e+02	0.037
x_3^2	ULA	46	11	4.1	6.7e+02	0.069	10	4.5	5.9e+02	0.079
	MALA	11	11	0.97	6e+02	0.018	10	1	5.6e+02	0.019
	RWM	26	9	2.9	4.3e+02	0.061	8.6	3.1	4.2e+02	0.062
x_4^2	ULA	5.1e+02	14	37	8.2e+02	0.62	12	43	6.9e+02	0.73
	MALA	1.2e+02	14	9	7.9e+02	0.15	12	10	7.1e+02	0.17
	RWM	2.9e+02	11	27	5.8e+02	0.51	10	29	5.6e+02	0.53

TABLE 3

Estimates of the asymptotic variances for ULA, MALA and RWM and each parameter x_i, x_i^2 for $i \in \{1, \dots, d\}$, and of the variance reduction factor (VRF) on the example of the logistic regression.

5.2. *Proof of Theorem 2.* The proof is divided into two parts. In the first part which gathers Lemma 11, Lemma 12 and Lemma 13, we establish preliminary and technical results. In particular, we derive in Lemma 11 an elementary bound on the second order moment of the estimator $\hat{\pi}_n(f)$ defined in (1), where $(X_k)_{k\in\mathbb{N}}$ is a Markov chain of kernel R_γ. The arguments are based solely on the study of R_γ and rely on **H**2. In a second part, using our preliminary results, the proof of Theorem 2 is then derived.

LEMMA 11. *Assume **H**1 and **H**2. Let $f : \mathbb{R}^d \to \mathbb{R}$ be such that $\|f\|_{V^{1/2}} < +\infty$. For all $n \in \mathbb{N}^*$,*

$$\mathbb{E}_{x,\gamma}\left[\left(\sum\nolimits_{k=0}^{n-1}\{f(X_k)-\pi_\gamma(f)\}\right)^2\right] \lesssim \gamma^{-1}\|f\|_{V^{1/2}}^2\left\{n+\gamma^{-1}V(x)\right\} .$$

PROOF. Note that under **H**2-(iii), by (Douc et al., 2018, Definition D.3.1-(i)) and Jensen inequality,

$$\|\delta_x R_\gamma^n - \pi_\gamma\|_{V^{1/2}} \lesssim \rho^{n\gamma/2}V^{1/2}(x) . \tag{46}$$

We have for all $n \in \mathbb{N}^*$

$$\begin{aligned}\mathbb{E}_{x,\gamma}&\left[\left(\sum\nolimits_{k=0}^{n-1}\{f(X_k)-\pi_\gamma(f)\}\right)^2\right] \\ &\lesssim \sum\nolimits_{k=0}^{n-1}\sum\nolimits_{s=0}^{n-1-k}\mathbb{E}_{x,\gamma}\left[(f(X_k)-\pi_\gamma(f))\,(f(X_{k+s})-\pi_\gamma(f))\right] .\end{aligned} \tag{47}$$

For $k \in \{0,\ldots,n-1\}$ and $s \in \{0,\ldots,n-1-k\}$,

$$\mathbb{E}_{x,\gamma}\left[(f(X_k)-\pi_\gamma(f))\,(f(X_{k+s})-\pi_\gamma(f))\right] = \mathbb{E}_{x,\gamma}\left[(f(X_k)-\pi_\gamma(f))\left(R_\gamma^s f(X_k)-\pi_\gamma(f)\right)\right] .$$

By (46), we obtain

$$\begin{aligned}&\left|\mathbb{E}_{x,\gamma}\left[(f(X_k)-\pi_\gamma(f))\,(f(X_{k+s})-\pi_\gamma(f))\right]\right| \\ &\qquad\lesssim \|f\|_{V^{1/2}}\,\rho^{\gamma s/2}\mathbb{E}_{x,\gamma}\left[|f(X_k)-\pi_\gamma(f)|\,V^{1/2}(X_k)\right] \\ &\qquad\lesssim \|f\|_{V^{1/2}}^2\,\rho^{\gamma s/2}\mathbb{E}_{x,\gamma}\left[V(X_k)\right] ,\end{aligned}$$

using that $V \geq 1$ and $|f(x)-\pi_\gamma(f)| \leq \|f\|_{V^{1/2}}(V^{1/2}(x)+\bar{\pi})$ where $\bar{\pi} = \sup_{\gamma\in(0,\bar{\gamma}]}\pi_\gamma(V) \lesssim 1$. By (28), we get

$$\left|\mathbb{E}_{x,\gamma}\left[(f(X_k)-\pi_\gamma(f))\,(f(X_{k+s})-\pi_\gamma(f))\right]\right| \lesssim \|f\|_{V^{1/2}}^2\,\rho^{\gamma s/2}\left\{\rho^{k\gamma}V(x)+\bar{\pi}\right\} .$$

Combining it with (47), we have

$$\mathbb{E}_{x,\gamma}\left[\left(\sum\nolimits_{k=0}^{n-1}\{f(X_k)-\pi_\gamma(f)\}\right)^2\right] \lesssim \frac{\|f\|_{V^{1/2}}^2}{1-\rho^{\gamma/2}}\left\{\frac{V(x)}{1-\rho^\gamma}+n\bar{\pi}\right\} .$$

Using that $1-\rho^{\beta\gamma} \geq \beta\gamma\log(1/\rho)\rho^{\beta\gamma}$ for all $\beta \in (0,1]$ concludes the proof. □

Define for any $f : \mathbb{R}^d \to \mathbb{R}$, $x \in \mathbb{R}^d$ and $\gamma \in (0, \bar{\gamma}]$, such that $R_\gamma f^2(x) < +\infty$,

$$m_\gamma(x) = \mathbb{E}_{x,\gamma}\left[\{f(X_1) - R_\gamma f(x)\}^2\right] \ . \tag{48}$$

LEMMA 12. *Assume **H**1 and **H**3. For all $\gamma \in (0, \bar{\gamma}]$ and $f \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $m_\gamma \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and in addition for all $p \in \mathbb{N}$ there exists $q \in \mathbb{N}$ such that for all $\gamma \in (0, \bar{\gamma}]$, $\|m_\gamma\|_{0,q} \lesssim \gamma \|f\|^2_{k_e,p}$.*

PROOF. Let $p \in \mathbb{N}$ and $f \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$. By **H**3, for all $\gamma \in (0, \bar{\gamma}]$ and $x \in \mathbb{R}^d$,

$$\begin{aligned}
0 \le m_\gamma(x) &= \mathbb{E}_{x,\gamma}\left[\{f(X_1) - f(x) - \gamma \mathscr{L} f(x) - \gamma^\alpha \mathscr{E}_\gamma f(x)\}^2\right] \\
&= \mathbb{E}_{x,\gamma}\left[\{f(X_1) - f(x)\}^2\right] - \gamma^2 \left\{\mathscr{L} f(x) + \gamma^{\alpha-1} \mathscr{E}_\gamma f(x)\right\}^2 && (49) \\
&\le \mathbb{E}_{x,\gamma}\left[\{f(X_1) - f(x)\}^2\right] \ . && (50)
\end{aligned}$$

Besides, for all $\gamma \in (0, \bar{\gamma}]$ and $x \in \mathbb{R}^d$,

$$\begin{aligned}
\mathbb{E}_{x,\gamma}\left[\{f(X_1) - f(x)\}^2\right] &= \mathbb{E}_{x,\gamma}\left[f^2(X_1)\right] + f^2(x) - 2f(x)\mathbb{E}_{x,\gamma}\left[f(X_1)\right] \\
&= \gamma \mathscr{L}(f^2)(x) + \gamma^\alpha \mathscr{E}_\gamma(f^2)(x) - 2\gamma f(x) \mathscr{L} f(x) - 2\gamma^\alpha f(x) \mathscr{E}_\gamma f(x) \\
&= \gamma \left\{2 \|\nabla f(x)\|^2 + \gamma^{\alpha-1}\left(\mathscr{E}_\gamma(f^2)(x) - 2f(x)\mathscr{E}_\gamma f(x)\right)\right\} \ . && (51)
\end{aligned}$$

Then, combining this result and (49), under **H**3, $m_\gamma \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and since $k_e \ge 2$, there exists $q \in \mathbb{N}$ such that $\|m_\gamma\|_{0,q} \lesssim \gamma \|f\|^2_{k_e,p}$. □

LEMMA 13. *Assume **H**1, **H**2 and **H**3. Then for any $p \in \mathbb{N}$,*

$$\left|\pi_\gamma(\hat{f} \mathscr{L} \hat{f}) - \pi(\hat{f} \mathscr{L} \hat{f})\right| \lesssim \|f\|^2_{k_e+2,p}\, \gamma^{\alpha-1} \ , \tag{52}$$

$$\sigma^2_\infty(f) = -2\pi(\hat{f} \mathscr{L} \hat{f}) \lesssim \|f\|^2_{2,p} \ , \tag{53}$$

where for any $f \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $\hat{f}$ is the solution of Poisson's equation (11) (see Proposition 10).

PROOF. Let $p \in \mathbb{N}$. By Proposition 10 and **H**1, there exists $q \in \mathbb{N}$ satisfying

$$\|\hat{f}\|_{k_e+2,q} \lesssim \|f\|_{k_e+2,p} \text{ and } \|U\|_{k_e+1,q} \lesssim 1 \ . \tag{54}$$

In addition, using Proposition 1, we have

$$|\pi_\gamma(\hat{f} \mathscr{L} \hat{f}) - \pi(\hat{f} \mathscr{L} \hat{f})| \lesssim \gamma^{\alpha-1} \|\hat{f} \mathscr{L} \hat{f}\|_{k_e,3q} \ .$$

Using that for any $k \in \mathbb{N}$ and $p_1, p_2 \in \mathbb{N}$, there exists $C_{k,p_1,p_2} \ge 0$ such that for any $g_1, g_2 \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $\|fg\|_{k,p_1+p_2} \le C_{k,p_1,p_2} \|f\|_{k,p_1} \|g\|_{k,p_2}$ by the general Leibniz rule, we get by definition of $\mathscr{L}$ (5),

$$\left|\pi_\gamma(\hat{f} \mathscr{L} \hat{f}) - \pi(\hat{f} \mathscr{L} \hat{f})\right| \lesssim \gamma^{\alpha-1} \|\hat{f}\|_{k_e,q} \|\mathscr{L} \hat{f}\|_{k_e,2q} \lesssim \gamma^{\alpha-1} \|\hat{f}\|^2_{k_e+2,q} \|U\|_{k_e+1,q} \ .$$

The proof of (52) then follows from (54). Similarly, by **H**2,

$$\sigma_\infty^2(f) = -2\pi(\hat{f}\mathscr{L}\hat{f}) \lesssim \|\hat{f}\mathscr{L}\hat{f}\|_{0,3q} \lesssim \|\hat{f}\|_{0,q}\|\mathscr{L}\hat{f}\|_{0,2q} \lesssim \|\hat{f}\|_{2,q}^2\|U\|_{1,q} ,$$

since $\|U\|_{1,q} \leq \|U\|_{k_e+1,q} \lesssim 1$. Using that $\|\hat{f}\|_{2,q} \leq \|f\|_{2,p}$ concludes the proof of (53). □

Proof of Theorem 2. Let $p \in \mathbb{N}$. For any $f \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$, let $\hat{f} \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$ be the solution of Poisson's equation $\mathscr{L}\hat{f} = -\tilde{f}$ (see Proposition 10). Using **H**3, we get for all $\gamma \in (0,\bar{\gamma}]$,

$$R_\gamma\hat{f} = \hat{f} + \gamma\mathscr{L}\hat{f} + \gamma^\alpha\mathscr{E}_\gamma\hat{f} = \hat{f} - \gamma\{f - \pi_\gamma(f)\} + \gamma^\alpha\mathscr{E}_\gamma\hat{f} - \gamma\{\pi_\gamma(f) - \pi(f)\} , \tag{55}$$

which implies that

$$\begin{aligned}\sum_{k=0}^{n-1}\{f(X_k) - \pi_\gamma(f)\} = \frac{\hat{f}(X_0) - \hat{f}(X_n)}{\gamma} + \frac{1}{\gamma}\sum_{k=0}^{n-1}\left\{\hat{f}(X_{k+1}) - R_\gamma\hat{f}(X_k)\right\} \\ + \gamma^{\alpha-1}\sum_{k=0}^{n-1}\left\{\mathscr{E}_\gamma\hat{f}(X_k) - \gamma^{1-\alpha}\left(\pi_\gamma(f) - \pi(f)\right)\right\} .\end{aligned} \tag{56}$$

Consider the following decomposition based on (56),

$$n^{-1}\mathbb{E}_{x,\gamma}\left[\left(\sum_{k=0}^{n-1}\{f(X_k) - \pi_\gamma(f)\}\right)^2\right] = \sum_{i=1}^4 A_i^f(x,n,\gamma) , \tag{57}$$

where,

$$A_1^f(x,n,\gamma) = \frac{\gamma^{2(\alpha-1)}}{n}\mathbb{E}_{x,\gamma}\left[\left(\sum_{k=0}^{n-1}\left\{\mathscr{E}_\gamma\hat{f}(X_k) - \gamma^{1-\alpha}\left(\pi_\gamma(f) - \pi(f)\right)\right\}\right)^2\right] , \tag{58}$$

$$A_2^f(x,n,\gamma) = (n\gamma^2)^{-1}\mathbb{E}_{x,\gamma}\left[(\hat{f}(X_0) - \hat{f}(X_n))^2\right] , \tag{59}$$

$$A_3^f(x,n,\gamma) = (n\gamma^2)^{-1}\mathbb{E}_{x,\gamma}\left[\left(\sum_{k=0}^{n-1}\hat{f}(X_{k+1}) - R_\gamma\hat{f}(X_k)\right)^2\right] ,$$

and by Cauchy-Schwarz inequality,

$$(1/2)\left|A_4^f(x,n,\gamma)\right| \leq \sum_{1\leq i<j\leq 3} A_i^f(x,n,\gamma)^{1/2}A_j^f(x,n,\gamma)^{1/2} . \tag{60}$$

We bound below $\left|A_i^f(x,n,\gamma)\right|$ for any $i \in \{1,\ldots,4\}$. By Proposition 10, there exists $q_1 \in \mathbb{N}$ such that

$$\|\hat{f}\|_{k_e,q_1} \lesssim \|f\|_{k_e,p} , \tag{61}$$

which combined with **H**2-(iii) and (28) yield for all $n \in \mathbb{N}^*$,

$$A_2^f(x,n,\gamma) \lesssim \|\hat{f}^2\|_V V(x)/(n\gamma^2) \lesssim \|f\|_{k_e,p}^2 V(x)/(n\gamma^2) . \tag{62}$$

For any $\gamma \in (0,\bar{\gamma}]$, by (55) and since $\mathscr{L}\hat{f} = -\tilde{f}$, $\pi_\gamma(\mathscr{E}_\gamma \hat{f}) = \gamma^{1-\alpha}\{\pi_\gamma(f) - \pi(f)\}$. Under **H3**, there exists $q_3 \in \mathbb{N}$ such that for all $\gamma \in (0,\bar{\gamma}]$, $\|\mathscr{E}_\gamma \hat{f}\|_{V^{1/2}} \lesssim \|\mathscr{E}_\gamma \hat{f}\|_{0,q_3} \lesssim \|\hat{f}\|_{k_e,q_1} \lesssim \|f\|_{k_e,p}$ by (61). Hence, applying Lemma 11 and using $\alpha \geq 3/2$ yield

$$A_1^f(x,n,\gamma) \lesssim \frac{\gamma^{2(\alpha-1)}}{n}\frac{\|f\|_{k_e,p}^2}{\gamma}\left(n + \frac{V(x)}{\gamma}\right) \lesssim \|f\|_{k_e,p}^2 \{1 + V(x)/(n\gamma)\} \ . \tag{63}$$

Since $(\sum_{k=0}^{n-1} \hat{f}(X_{k+1}) - R_\gamma \hat{f}(X_k))_{k\in\mathbb{N}}$ is a $\mathbb{P}_{x,\gamma}$-square integrable martingale, we get that for all $n \in \mathbb{N}$,

$$A_3^f(x,n,\gamma) = \gamma^{-2}\mathbb{E}_{x,\gamma}\left[n^{-1}\sum\nolimits_{k=0}^{n-1} g_\gamma(X_k)\right] \ , \tag{64}$$

where

$$g_\gamma(x) = \mathbb{E}_{x,\gamma}\left[\{\hat{f}(X_1) - R_\gamma \hat{f}(x)\}^2\right] \ . \tag{65}$$

Lemma 12 shows that $g_\gamma \in \mathrm{C}^\infty_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$ and that there exists $q_2 \in \mathbb{N}$ such that $\|g_\gamma\|_V \lesssim \|g_\gamma\|_{0,q_2} \lesssim \gamma\|\hat{f}\|_{k_e,q_1}^2 \lesssim \gamma\|f\|_{k_e,p}^2$. Applying (26), we get that for all $n \in \mathbb{N}^*$,

$$\left|\mathbb{E}_{x,\gamma}\left[n^{-1}\sum\nolimits_{k=0}^{n-1} g_\gamma(X_k)\right] - \pi_\gamma(g_\gamma)\right| \lesssim \|g_\gamma\|_V \,(n\gamma)^{-1} V(x) \lesssim n^{-1}\,\|f\|_{k_e,p}^2\, V(x) \ . \tag{66}$$

We now show that $\pi_\gamma(g_\gamma)$ is approximately equal to $\gamma\sigma_\infty^2(f)$. Observe that by (65) and since π_γ is invariant for R_γ, for any $\gamma \in (0,\bar{\gamma}]$,

$$\begin{aligned}\pi_\gamma(g_\gamma) &= \mathbb{E}_{\pi_\gamma,\gamma}\left[\{\hat{f}(X_1) - R_\gamma \hat{f}(X_0)\}^2\right] \\ &= \mathbb{E}_{\pi_\gamma,\gamma}\left[\{\hat{f}(X_1) - \hat{f}(X_0)\}^2\right] - \mathbb{E}_{\pi_\gamma,\gamma}\left[\{\hat{f}(X_0) - R_\gamma \hat{f}(X_0)\}^2\right] \ .\end{aligned} \tag{67}$$

Using that π_γ is the invariant distribution for R_γ again and (55), we have for any $\gamma \in (0,\bar{\gamma}]$,

$$\begin{aligned}\mathbb{E}_{\pi_\gamma,\gamma}\left[\{\hat{f}(X_1) - \hat{f}(X_0)\}^2\right] &= 2\mathbb{E}_{\pi_\gamma,\gamma}\left[\hat{f}(X_0)\{\hat{f}(X_0) - R_\gamma \hat{f}(X_0)\}\right] \\ &= -2\gamma\pi_\gamma(\hat{f}\mathscr{L}\hat{f}) - 2\gamma^\alpha \pi_\gamma(\hat{f}\mathscr{E}_\gamma \hat{f}) \ .\end{aligned} \tag{68}$$

In the next step, we consider separately the cases $\pi_\gamma = \pi$ and $\pi_\gamma \neq \pi$. If $\pi = \pi_\gamma$, then

$$-\pi_\gamma(\hat{f}\mathscr{L}\hat{f}) = (1/2)\sigma_\infty^2(f) \ . \tag{69}$$

If $\pi_\gamma \neq \pi$, Lemma 13 shows that

$$\left|\pi_\gamma(\hat{f}\mathscr{L}\hat{f}) + (1/2)\sigma_\infty^2(f)\right| = \left|\pi_\gamma(\hat{f}\mathscr{L}\hat{f}) - \pi(\hat{f}\mathscr{L}\hat{f})\right| \lesssim \|f\|_{k_e+2,p}^2\, \gamma^{\alpha-1} \ . \tag{70}$$

Using **H3**, (28) and $\left|\pi_\gamma(\hat{f}\mathscr{E}_\gamma \hat{f})\right| \lesssim \|f\|_{k_e,p}^2$ in (68), we obtain that

$$\left|\mathbb{E}_{\pi_\gamma,\gamma}\left[\{\hat{f}(X_1) - \hat{f}(X_0)\}^2\right] + 2\gamma\pi_\gamma(\hat{f}\mathscr{L}\hat{f})\right| = 2\gamma^\alpha\left|\pi_\gamma(\hat{f}\mathscr{E}_\gamma \hat{f})\right| \lesssim \|f\|_{k_e,p}^2\, \gamma^\alpha \ . \tag{71}$$

Similarly, using **H**2-(ii), (28), (55), (5), **H**3 and (61), it holds since $k_e \geq 2$ that

$$\mathbb{E}_{\pi_\gamma,\gamma}\left[\{\hat{f}(X_0) - R_\gamma \hat{f}(X_0)\}^2\right] \lesssim \|\hat{f}\|^2_{k_e,q_1}\gamma^2 \lesssim \|f\|^2_{k_e,p}\gamma^2 \,.$$

Combining this result with (69) or (70) and (71) in (67) and using that $\|f\|_{k_e,p} \leq \|f\|_{k_e+2,p}$, we obtain

$$\left|\pi_\gamma(g_\gamma) - \gamma\sigma^2_\infty(f)\right| \lesssim \|f\|^2_{k_e+2,p}\,\gamma^{\alpha\wedge 2} \,.$$

Plugging this inequality and (66) in (64), we obtain for all $n \in \mathbb{N}^*$,

$$\left|A_3^f(x,n,\gamma) - \gamma^{-1}\sigma^2_\infty(f)\right| \lesssim \|f\|^2_{k_e+2,p}\left\{\gamma^{(\alpha-2)\wedge 0} + (n\gamma^2)^{-1}V(x)\right\} \,. \tag{72}$$

Note that since $\alpha \geq 3/2$, by (53) and (72),

$$A_3^f(x,n,\gamma) \lesssim \|f\|^2_{k_e+2,p}\left\{\gamma^{-1} + (n\gamma^2)^{-1}V(x)\right\} \,.$$

Combining it with (60), (62) and (63) conclude the proof.

□

6. Geometric ergodicity for the ULA and MALA algorithms.

In this Section, we show that **H**2 is satisfied for the family of Markov kernel $\{R_\gamma^{\mathrm{ULA}} : \gamma \in (0,\bar{\gamma}]\}$ and $\{R_\gamma^{\mathrm{MALA}} : \gamma \in (0,\bar{\gamma}]\}$, with $\bar{\gamma} > 0$, associated to the ULA and MALA algorithms (see (33) and (36)).

To check (26) in **H**2, we establish minorization and drift conditions on $R_\gamma = R_\gamma^{\mathrm{ULA}}$ and $R_\gamma = R_\gamma^{\mathrm{MALA}}$, see e.g. (Douc et al., 2018, Chapter 19) with an explicit dependence with respect to the parameter γ. More precisely, assume that there exist $\lambda \in (0,1)$, $b < +\infty$ and $c > 0$ such that

$$R_\gamma V \leq \lambda^\gamma V + \gamma b \,, \quad \text{for all } \gamma \in (0,\bar{\gamma}] \,, \tag{73}$$

and if there exists $\varepsilon \in (0,1]$ such that for all $\gamma \in (0,\bar{\gamma}]$ and $x, x' \in \left\{V \leq \widetilde{M}\right\}$,

$$\|R_\gamma^{\lceil 1/\gamma\rceil}(x,\cdot) - R_\gamma^{\lceil 1/\gamma\rceil}(x',\cdot)\|_{\mathrm{TV}} \leq 2(1-\varepsilon) \,, \tag{74}$$

where

$$\widetilde{M} > \left(\frac{4b\lambda^{-\bar{\gamma}}}{\log(1/\lambda)} - 1\right) \vee 1 \,. \tag{75}$$

Then, (Douc et al., 2018, Theorem 19.4.1) shows that(26) holds. It suffices therefore to show (73) and (75).

For ease of notations, we denote in this Section R_γ^{MALA} by R_γ and R_γ^{ULA} by Q_γ for any $\gamma > 0$. We also assume without loss of generality that $\nabla U(0) = 0$. Note that under **H**1 and **H**4, $m \leq L$.

To establish **H**2, we show that the Markov kernels of ULA and MALA satisfy a minorization condition (74) in Propositions 16 and 21 and a drift condition (73) in Propositions 17 and 23. The geometric ergodicity result then follows from (Douc et al., 2018, Theorem 18.4.3).

We begin the proof by two technical lemmas, Lemmas 14 and 15 which are used repeatedly throughout this Section. The geometric ergodicity of ULA is proved in Section 6.1, and relying on these results and the analysis of ULA, the case of the MALA algorithm is examined in Section 6.2.

LEMMA 14. *Assume **H**1 and **H**4. Then there exists $K_2 \geq 0$ such that for any $x \notin \mathrm{B}(0, K_2)$, $\langle \nabla U(x), x \rangle \geq (m/2) \|x\|^2$ and in particular $\|\nabla U(x)\| \geq (m/2)\|x\|$.*

PROOF. Using **H**1 and **H**4, we have for any $x \in \mathbb{R}^d$, $\|x\| \geq K_1$,

$$\begin{aligned}\langle \nabla U(x), x \rangle &= \int_0^{K_1/\|x\|} \mathrm{D}^2 U(tx)[x^{\otimes 2}]\mathrm{d}t + \int_{K_1/\|x\|}^1 \mathrm{D}^2 U(tx)[x^{\otimes 2}]\mathrm{d}t \\ &\geq m\|x\|^2 \{1 - K_1(1 + L/m)/\|x\|\} ,\end{aligned}$$

which proves the first statement. The second statement is obvious. □

LEMMA 15. *Assume **H**1 and **H**4. Then, for any $t \in [0,1]$, $\gamma \in (0, 1/(4L)]$ and $x, z \in \mathbb{R}^d$, $\|z\| \leq \|x\|/(4\sqrt{2\gamma})$, it holds*

$$\left\| x + t\{-\gamma \nabla U(x) + \sqrt{2\gamma} z\} \right\| \geq \|x\|/2 .$$

PROOF. Let $t \in [0,1]$, $\gamma \in (0, 1/(4L)]$ and $x, z \in \mathbb{R}^d$, $\|z\| \leq \|x\|/(4\sqrt{2\gamma})$. Using the triangle inequality and **H**1, we have since $t \in [0,1]$

$$\left\| x + t\{-\gamma \nabla U(x) + \sqrt{2\gamma} z\} \right\| \geq (1 - \gamma L)\|x\| - \sqrt{2\gamma}\|z\| .$$

The conclusion then follows from $\gamma \leq 1/(4L)$ and $\|z\| \leq \|x\|/(4\sqrt{2\gamma})$. □

6.1. *Geometric ergodicity for the ULA algorithm.*

PROPOSITION 16. *Assume **H**1. Then for any $\mathsf{K} \geq 0$ there exists $\varepsilon > 0$ such that for any $x, y \in \mathbb{R}^d$, $\|x\| \vee \|y\| \leq \mathsf{K}$, and $\gamma \in (0, 1/L]$ we have*

$$\|\delta_x Q_\gamma^{\lceil 1/\gamma \rceil} - \delta_y Q_\gamma^{\lceil 1/\gamma \rceil}\|_{\mathrm{TV}} \leq 2(1 - \varepsilon) . \tag{76}$$

PROOF. By (De Bortoli and Durmus, 2019, Theorem 1, Lemma 2), since by **H**1 for any $x, y \in \mathbb{R}^d$,

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

where $\kappa(\gamma) = (2L + L^2\gamma)$, we have for any $x, y \in \mathbb{R}^d$, $\|x\| \vee \|y\| \leq \mathsf{K}$, $\gamma \in (0, 1/L]$

$$\|\delta_x Q_\gamma^{\lceil 1/\gamma \rceil} - \delta_y Q_\gamma^{\lceil 1/\gamma \rceil}\|_{\mathrm{TV}} \leq 2\left[1 - 2\boldsymbol{\Phi}\left(-\frac{\kappa^{1/2}(\gamma)\,\|x-y\|}{(1-\exp\{-\kappa(\gamma)(1+\gamma\kappa(\gamma))^{-1}\})^{1/2}}\right)\right]$$
$$\leq 2(1-\varepsilon_1)\,, \quad \text{with} \quad \varepsilon_1 = 2\boldsymbol{\Phi}\left(-\frac{2\mathsf{K}(3L)^{1/2}}{\{1-\exp(-2L)\}^{1/2}}\right)\,, \tag{77}$$

where we have used for the last inequality that

$$\frac{\kappa^{1/2}(\gamma)\,\|x-y\|}{(1-\exp\{-\kappa(\gamma)(1+\gamma\kappa(\gamma))^{-1}\})^{1/2}} \leq \frac{2\mathsf{K}(3L)^{1/2}}{\{1-\exp(-2L)\}^{1/2}}$$

and that $t \mapsto 1 - 2\boldsymbol{\Phi}(-t)$ is increasing on $\mathbb{R}_+$. □

For any $\eta > 0$ define $V_\eta : \mathbb{R}^d \to [1, +\infty)$, for any $x \in \mathbb{R}^d$ by,

$$V_\eta(x) = \exp(\eta \|x\|^2)\,. \tag{78}$$

PROPOSITION 17. *Assume* **H1** *and* **H4** *and let $\bar\gamma \in \big(0, m/(4L^2)\big]$. Then, for any $\gamma \in (0, \bar\gamma]$,*

$$Q_\gamma V_{\bar\eta}(x) \leq \exp\left(-\bar\eta m\gamma \|x\|^2/4\right) V_{\bar\eta}(x) + b_{\bar\eta}\gamma \mathbb{1}_{\mathrm{B}(0,K_3)}(x)\,,$$

where $\bar\eta = \min(m/16, (8\bar\gamma)^{-1})$, $K_3 = \max(K_2, 4\sqrt{d/m})$, and

$$b_{\bar\eta} = \left[\bar\eta\left\{m/4 + (1+16\bar\eta\bar\gamma)(4\bar\eta + 2L + \bar\gamma L^2)\right\}K_3^2 + 4\bar\eta d\right]$$
$$\times \exp\left[\bar\gamma\bar\eta\left\{m/4 + (1+16\bar\eta\bar\gamma)(4\bar\eta + 2L + \bar\gamma L^2)\right\}K_3^2 + (d/2)\log(2)\right]\,. \tag{79}$$

PROOF. Let $\gamma \in (0, \bar\gamma]$. First since for any $x \in \mathbb{R}^d$, we have

$$\bar\eta\left\|x - \gamma\nabla U(x) + \sqrt{2\gamma}z\right\|^2 - \|z\|^2/2$$
$$= -2^{-1}(1-4\bar\eta\gamma)\left\|z - \frac{2(2\gamma)^{1/2}\bar\eta}{1-4\bar\eta\gamma}\{x - \gamma\nabla U(x)\}\right\|^2 + \frac{\bar\eta}{1-4\bar\eta\gamma}\|x - \gamma\nabla U(x)\|^2\,,$$

which implies since $1 - 4\bar\eta\gamma > 0$ that

$$Q_\gamma V_{\bar\eta}(x) = (2\pi)^{-d/2}\int_{\mathbb{R}^d}\exp\left(\bar\eta\left\|x - \gamma\nabla U(x) + \sqrt{2\gamma}z\right\|^2 - \|z\|^2/2\right)\mathrm{d}z$$
$$= (1-4\bar\eta\gamma)^{-d/2}\exp\left(\bar\eta(1-4\bar\eta\gamma)^{-1}\|x - \gamma\nabla U(x)\|^2\right)\,. \tag{80}$$

We now distinguish the case when $\|x\| \geq K_3$ and $\|x\| < K_3$.

By **H4** and Lemma 14, for any $x \in \mathbb{R}^d$, $\|x\| \geq K_3 \geq K_2$, using that $\bar\eta \leq m/16$ and $\gamma \leq \bar\gamma \leq m/(4L^2)$, we have

$$(1-4\bar\eta\gamma)^{-1}\|x - \gamma\nabla U(x)\|^2 - \|x\|^2$$
$$\leq \gamma\|x\|^2(1-4\bar\eta\gamma)^{-1}\left(4\bar\eta - m + \gamma L^2\right) \leq -\gamma(m/2)\|x\|^2(1-4\bar\eta\gamma)^{-1}\,.$$

Therefore, (80) becomes

$$Q_\gamma V_{\bar\eta}(x) \le \exp\left(-\gamma\bar\eta(m/2)(1-4\bar\eta\gamma)^{-1}\|x\|^2 - (d/2)\log(1-4\bar\eta\gamma)\right) V_{\bar\eta}(x)$$
$$\le \exp\left(\gamma\bar\eta\{-(m/2)\|x\|^2 + 4d\}\right) V_{\bar\eta}(x)\,,$$

where we have used for the last inequality that $-\log(1-t) \le 2t$ for $t \in [0,1/2]$ and $4\bar\eta\gamma \le 1/2$. The proof of the statement then follows since $\|x\| \ge K_3 \ge 4\sqrt{d/m}$.

In the case $\|x\| < K_3$, by (80), **H**1 and since $(1-t)^{-1} \le 1+4t$ for $t \in [0,1/2]$, we obtain

$$(1-4\bar\eta\gamma)^{-1}\|x-\gamma\nabla U(x)\|^2 - \|x\|^2 \le \gamma(1-4\bar\eta\gamma)^{-1}\{4\bar\eta + 2L + \gamma L^2\}\|x\|^2$$
$$\le \gamma(1+16\bar\eta\gamma)\{4\bar\eta + 2L + \gamma L^2\}\|x\|^2\,,$$

which implies that

$$Q_\gamma V_{\bar\eta}(x)/V_{\bar\eta}(x) \le \mathrm{e}^{-\bar\eta m\gamma\|x\|^2/4}$$
$$+ \exp\left[\gamma\bar\eta\left\{m/4 + (1+16\bar\eta\gamma)(4\bar\eta+2L+\gamma L^2)\right\}\|x\|^2 - (d/2)\log(1-4\bar\eta\gamma)\right] - 1\,.$$

The proof is then completed using that for any $t \ge 0$, $\mathrm{e}^t - 1 \le t\mathrm{e}^t$, for any $s \in [0,1/2]$, $-\log(1-s) \le 2s$ and $4\bar\eta\gamma \le 1/2$. □

6.2. *Geometric ergodicity for the MALA algorithm.* We first provide a decomposition in γ of $\tau_\gamma^{\mathrm{MALA}}$ defined in (37). For any $x, z \in \mathbb{R}^d$, a Taylor expansion shows that

$$\tau_\gamma^{\mathrm{MALA}}(x,z) = \sum_{k=2}^{6} \gamma^{k/2} A_{k,\gamma}(x,z) \tag{81}$$

where, setting $x_t = x + t\{-\gamma\nabla U(x) + \sqrt{2\gamma}z\}$,

$$A_{2,\gamma}(x,z) = 2\int_0^1 \mathrm{D}^2 U(x_t)\{z\}^{\otimes 2}(1/2-t)\mathrm{d}t \tag{82}$$

$$A_{3,\gamma}(x,z) = 2^{3/2}\int_0^1 \mathrm{D}^2 U(x_t)\{z\otimes\nabla U(x)\}(t-1/4)\mathrm{d}t\,, \tag{83}$$

$$A_{4,\gamma}(x,z) = -\int_0^1 \mathrm{D}^2 U(x_t)\{\nabla U(x)\}^{\otimes 2} t\mathrm{d}t + (1/2)\left\|\int_0^1 \mathrm{D}^2 U(x_t) z\mathrm{d}t\right\|^2 \tag{84}$$

$$A_{5,\gamma}(x,z) = -(1/2)^{1/2}\left\langle \int_0^1 \mathrm{D}^2 U(x_t)\nabla U(x)\mathrm{d}t, \int_0^1 \mathrm{D}^2 U(x_t) z\mathrm{d}t\right\rangle \tag{85}$$

$$A_{6,\gamma}(x,z) = (1/4)\left\|\int_0^1 \mathrm{D}^2 U(x_t)\nabla U(x)\mathrm{d}t\right\|^2\,. \tag{86}$$

Since $\int_0^1 \mathrm{D}^2 U(x)[z^{\otimes 2}](1/2-t)\mathrm{d}t = 0$, we get setting $x_t = x + t\{-\gamma\nabla U(x) + \sqrt{2\gamma}z\}$,

$$A_{2,\gamma}(x,z) = \sqrt{\gamma}\iint_0^1 \mathrm{D}^3 U(sx_t + (1-s)x)\left[z^{\otimes 2}\otimes\{-\gamma^{1/2}\nabla U(x) + \sqrt{2}z\}\right](1/2-t)t\mathrm{d}s\mathrm{d}t\,. \tag{87}$$

LEMMA 18. *Assume* **H**1 *and* **H**4. *Then, for any* $\bar{\gamma} \in \left(0, m^3/(4L^4)\right]$ *there exists* $C_{2,\bar{\gamma}} < \infty$ *such that for any* $\gamma \in (0, \bar{\gamma}]$, $x, z \in \mathbb{R}^d$ *satisfying* $\|x\| \geq \max(2K_1, K_2)$ *and* $\|z\| \leq \|x\| / (4\sqrt{2\gamma})$, *where* K_2 *is defined in Lemma 14, it holds*

$$\tau_\gamma^{\mathrm{MALA}}(x, z) \leq C_{2,\bar{\gamma}} \gamma \|z\|^2 \left\{1 + \|z\|^2\right\} .$$

PROOF. Let $\gamma \in (0, \bar{\gamma}]$, $x, z \in \mathbb{R}^d$ satisfying $\|x\| \geq \max(2K_1, K_2)$ and $\|z\| \leq \|x\| / (4\sqrt{2\gamma})$. Using (81), we get setting $A_{4,0,\gamma}(x, z) = \int_0^1 \mathrm{D}^2\, U(x_t)[\nabla U(x)^{\otimes 2}] t \mathrm{d}t$,

$$\begin{aligned}(88)\quad \tau_\gamma^{\mathrm{MALA}}(x, z) &\leq 2\gamma A_{2,\gamma}(x, z) - \gamma^2 A_{4,0,\gamma}(x, z) \\ &+ (2\gamma)^{3/2} L^2 \|z\| \|x\| + (\gamma^2/2) L^2 \|z\|^2 + (\gamma^5/2)^{1/2} L^3 \|z\| \|x\| + (\gamma^3/4) L^4 \|x\|^2 \ ,\end{aligned}$$

By **H**4, Lemma 14 and Lemma 15, we get for any $x \in \mathbb{R}^d$, $\|x\| \geq \max(2K_1, K_2)$,

$$A_{4,0,\gamma}(x, z) \geq (m/2)^3 \|x\|^2 \ . \tag{89}$$

Combining this result with (87), (89) in (88), we obtain using $\gamma \leq \bar{\gamma} \leq m^3/(4L^4)$

$$\begin{aligned}\tau_\gamma^{\mathrm{MALA}}(x, z) &\leq 2\gamma M \left\{ \sqrt{2\gamma} \|z\|^3 + \gamma L \|z\|^2 \|x\| \right\} - \gamma^2 (m^3/2^4) \|x\|^2 \\ &+ (2\gamma)^{3/2} L^2 \|z\| \|x\| + (\gamma^2/2) L^2 \|z\|^2 + (\gamma^5/2)^{1/2} L^3 \|z\| \|x\| \ ,\end{aligned}$$

Since for any $a, b \in \mathbb{R}^+$ and $\varepsilon > 0$, $ab \leq (\epsilon/2) a^2 + 1/(2\epsilon) b^2$, we obtain

$$\begin{aligned}&\tau_\gamma^{\mathrm{MALA}}(x, z) \\ &\leq \gamma \|z\|^2 \left\{ 2^{1/2} L^2 \varepsilon^{-1} + (\gamma/2) L^2 + 2^{-3/2} \gamma^{3/2} L^3 \varepsilon^{-1} + (2^3\gamma)^{1/2} M \|z\| + \gamma M L \varepsilon^{-1} \|z\|^2 \right\} \\ &\quad + \|x\|^2 \gamma^2 \left[\varepsilon \left\{ LM + 2^{1/2} L^2 + 2^{-3/2} \bar{\gamma}^{1/2} L^3 \right\} - m^3/2^4 \right] \ .\end{aligned}$$

Choosing $\varepsilon = (m^3/2^4)\{LM + 2^{1/2} L^2 + 2^{-3/2} \bar{\gamma}^{1/2} L^3\}^{-1}$ concludes the proof. □

LEMMA 19. *Assume* **H**1, **H**4 *and let* $\bar{\gamma} \in \left(0, m/(4L^2)\right]$. *Then, for any* $\gamma \in (0, \bar{\gamma}]$ *and* $x \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} \|y\|^2 Q_\gamma(x, \mathrm{d}y) \leq \{1 - (m\gamma)/2\} \|x\|^2 + \tilde{b}\gamma \mathbb{1}_{\mathrm{B}(0,K_4)}(x) \ ,$$

where Q_γ *is the Markov kernel of ULA defined in* (33),

$$K_4 = \max\left(K_2, 2\sqrt{(2d)/m}\right) \ , \quad \tilde{b} = 2d + K_4^2 \left(\bar{\gamma} L^2 + 2L + m/2\right) \ .$$

PROOF. Let $\gamma \in (0, \bar{\gamma}]$ and $x \in \mathbb{R}^d$. By **H**1, we have

$$\int_{\mathbb{R}^d} \|y\|^2 Q_\gamma(x, \mathrm{d}y) \leq 2\gamma d + \|x\|^2 (1 + \gamma^2 L^2) - 2\gamma \langle \nabla U(x), x \rangle \ .$$

We distinguish the case when $\|x\| \geq K_4$ and $\|x\| < K_4$. If $\|x\| \geq K_4 \geq K_2$, by Lemma 14, and since $\gamma \leq \bar{\gamma} \leq m/(4L^2)$, $\|x\| \geq K_4 \geq 2\sqrt{(2d)/m}$,

$$\int_{\mathbb{R}^d} \|y\|^2 Q_\gamma(x, \mathrm{d}y) \leq \|x\|^2 \left[1 - \gamma\left\{m - \gamma L^2 - (2d)/\|x\|^2\right\}\right] \leq \|x\|^2 \{1 - \gamma m/2\} \ .$$

If $\|x\| < K_4$, we obtain

$$\int_{\mathbb{R}^d} \|y\|^2 Q_\gamma(x, \mathrm{d}y) \leq \|x\|^2 \{1 - \gamma m/2\} + \gamma \|x\|^2 \left(\gamma L^2 + 2L + m/2\right) + 2\gamma d \ ,$$

which concludes the proof. $\square$

LEMMA 20. *Assume **H**1 and **H**4 and let $\bar{\gamma} \in \left(0, m/(4L^2)\right]$. Then, there exist $C_{3,\bar{\gamma}}, C_{4,\bar{\gamma}} \geq 0$ such that for any $x \in \mathbb{R}^d$ and $\gamma \in (0, \bar{\gamma}]$, we have*

$$\|\delta_x Q_\gamma - \delta_x R_\gamma\|_{\mathrm{TV}} \leq C_{3,\bar{\gamma}} \gamma^{3/2}(1 + \|x\|^2) \ , \tag{90}$$

$$\|\delta_x Q_\gamma^{\lceil 1/\gamma \rceil} - \delta_x R_\gamma^{\lceil 1/\gamma \rceil}\|_{\mathrm{TV}} \leq C_{4,\bar{\gamma}} \gamma^{1/2}(1 + \|x\|^2) \ . \tag{91}$$

PROOF. Let $x \in \mathbb{R}^d$ and $\gamma \in (0, \bar{\gamma}]$. We first show that (90) holds and then use this result to prove (91). Let $f : \mathbb{R}^d \to \mathbb{R}$ be a bounded and measurable function. Then, by (33) and (36), we have

$$\begin{aligned}
&|Q_\gamma f(x) - R_\gamma f(x)| \\
&\quad = \left| \int_{\mathbb{R}^d} \{f(x - \gamma \nabla U(x) + \sqrt{2\gamma} z) - f(x)\}\{1 - \min(1, \mathrm{e}^{-\tau_\gamma^{\mathrm{MALA}}(x,z)})\} \boldsymbol{\varphi}(z) \mathrm{d}z \right| \\
&\quad \leq 2\|f\|_\infty \int_{\mathbb{R}^d} \left|1 - \min(1, \mathrm{e}^{-\tau_\gamma^{\mathrm{MALA}}(x,z)})\right| \boldsymbol{\varphi}(z) \mathrm{d}z \leq 2\|f\|_\infty \int_{\mathbb{R}^d} \left|\tau_\gamma^{\mathrm{MALA}}(x,z)\right| \boldsymbol{\varphi}(z) \mathrm{d}z \ .
\end{aligned}$$

The conclusion of (90) then follows from an application of Lemma 7.

We now turn to the proof of (91). Consider the following decomposition

$$\delta_x Q_\gamma^{\lceil 1/\gamma \rceil} - \delta_x R_\gamma^{\lceil 1/\gamma \rceil} = \sum_{k=0}^{\lceil 1/\gamma \rceil - 1} \delta_x Q_\gamma^k \{Q_\gamma - R_\gamma\} R_\gamma^{\lceil 1/\gamma \rceil - k - 1} \ .$$

Therefore using the triangle inequality, we obtain that

$$\|\delta_x Q_\gamma^{\lceil 1/\gamma \rceil} - \delta_x R_\gamma^{\lceil 1/\gamma \rceil}\|_{\mathrm{TV}} \leq \sum_{k=0}^{\lceil 1/\gamma \rceil - 1} \|\delta_x Q_\gamma^k \{R_\gamma - Q_\gamma\} R_\gamma^{\lceil 1/\gamma \rceil - k - 1}\|_{\mathrm{TV}} \ . \tag{92}$$

We now bound each term in the sum. Let $k \in \{0, \ldots, \lceil 1/\gamma \rceil - 1\}$ and $f : \mathbb{R}^d \to \mathbb{R}$ be a bounded and measurable function. By (90), we obtain that

$$\left|\delta_x \{R_\gamma - Q_\gamma\} R_\gamma^{\lceil 1/\gamma \rceil - k - 1} f\right| \leq C_{3,\bar{\gamma}} \|f\|_\infty \gamma^{3/2} \{1 + \|x\|^2\}$$

and therefore using Lemma 19, we get

$$\left|\updelta_x Q_\gamma^k \{R_\gamma - Q_\gamma\} R_\gamma^{\lceil 1/\gamma \rceil - k - 1} f\right| \leq C_{3,\bar{\gamma}} \left\|f\right\|_\infty \gamma^{3/2} \{1 + (1 - m\gamma/2)^k \left\|x\right\|^2 + 2\tilde{b}/m\} \ .$$

Plugging this result in (92), we obtain

$$\begin{aligned} \|\updelta_x Q_\gamma^{\lceil 1/\gamma \rceil} - \updelta_x R_\gamma^{\lceil 1/\gamma \rceil}\|_{\mathrm{TV}} &\leq C_{3,\bar{\gamma}} \gamma^{3/2} \sum_{k=0}^{\lceil 1/\gamma \rceil - 1} \{1 + (1 - m\gamma/2)^k \left\|x\right\|^2 + 2\tilde{b}/m\} \\ &\leq C_{3,\bar{\gamma}} \gamma^{1/2} \{1 + 2(\left\|x\right\|^2 + \tilde{b})/m\} \ , \end{aligned}$$

which concludes the proof. □

PROPOSITION 21. *Assume **H**1 and **H**4. Then for any $\mathsf{K} \geq 0$ there exist $\bar{\gamma} > 0$ and $\varepsilon > 0$, such that for any $x, y \in \mathbb{R}^d$, $\|x\| \vee \|y\| \leq \mathsf{K}$, and $\gamma \in (0, \bar{\gamma}]$ we have*

$$\|\updelta_x R_\gamma^{\lceil 1/\gamma \rceil} - \updelta_y R_\gamma^{\lceil 1/\gamma \rceil}\|_{\mathrm{TV}} \leq 2(1 - \varepsilon) \ . \tag{93}$$

PROOF. First note that for any $x, y \in \mathbb{R}^d$, $\gamma > 0$, by the triangle inequality, we obtain

$$\begin{aligned} \|\updelta_x R_\gamma^{\lceil 1/\gamma \rceil} - \updelta_y R_\gamma^{\lceil 1/\gamma \rceil}\|_{\mathrm{TV}} &\leq \|\updelta_x R_\gamma^{\lceil 1/\gamma \rceil} - \updelta_x Q_\gamma^{\lceil 1/\gamma \rceil}\|_{\mathrm{TV}} \\ &+ \|\updelta_x Q_\gamma^{\lceil 1/\gamma \rceil} - \updelta_y Q_\gamma^{\lceil 1/\gamma \rceil}\|_{\mathrm{TV}} + \|\updelta_y R_\gamma^{\lceil 1/\gamma \rceil} - \updelta_y Q_\gamma^{\lceil 1/\gamma \rceil}\|_{\mathrm{TV}} \ . \end{aligned} \tag{94}$$

We now give some bounds for each term on the right hand side for any $x, y \in \mathbb{R}^d$, $\|x\| \vee \|y\| \leq \mathsf{K}$ for a fixed $\mathsf{K} \geq 0$ and $\gamma \leq 1/L$. By Proposition 16, there exists $\varepsilon_1 > 0$ such that for any $x, y \in \mathbb{R}^d$, $\|x\| \vee \|y\| \leq \mathsf{K}$ and $\gamma \leq 1/L$,

$$\|\updelta_x Q_\gamma^{\lceil 1/\gamma \rceil} - \updelta_y Q_\gamma^{\lceil 1/\gamma \rceil}\|_{\mathrm{TV}} \leq 2(1 - \varepsilon_1) \ . \tag{95}$$

In addition, by Lemma 20, there exists $C \geq 0$ such that for any $\gamma \in \big(0, m/(4L^2)\big]$, and $z \in \mathbb{R}^d$, $\|z\| \leq \mathsf{K}$,

$$\|\updelta_z Q_\gamma^{\lceil 1/\gamma \rceil} - \updelta_z R_\gamma^{\lceil 1/\gamma \rceil}\|_{\mathrm{TV}} \leq C\gamma^{1/2}(1 + \mathsf{K}^2) \ .$$

Combining this result with (95) in (94), we obtain that for any $x, y \in \mathbb{R}^d$, $\|x\| \vee \|y\| \leq \mathsf{K}$, $\gamma \in \big(0, m/(4L^2)\big]$,

$$\left\|\updelta_x R_\gamma^{\lceil 1/\gamma \rceil} - \updelta_y R_\gamma^{\lceil 1/\gamma \rceil}\right\| \leq 2(1 - \varepsilon_1) + 2C\gamma^{1/2}(1 + \mathsf{K}^2) \ .$$

Therefore, we obtain that for any $x, y \in \mathbb{R}^d$, $\|x\| \vee \|y\| \leq \mathsf{K}$, $\gamma \in (0, \bar{\gamma}]$, (93) holds with $\varepsilon \leftarrow \varepsilon_1/2$ taking

$$\bar{\gamma} = m/(4L^2) \wedge \left[\varepsilon_1^2 \left(2C(1 + \mathsf{K}^2)\right)^{-2}\right] \ .$$

□

LEMMA 22. *Let $\bar{\gamma} > 0$ and $\gamma \in (0, \bar{\gamma}]$. Then, for any $x \in \mathbb{R}^d$, $\|x\| \geq 20\sqrt{2\bar{\gamma}d}$,*

$$\int_{\mathbb{R}^d \setminus \mathrm{B}(0, \|x\|/(4\sqrt{2\gamma}))} \boldsymbol{\varphi}(z) \mathrm{d}z \leq \exp(-\left\|x\right\|^2 / (128\gamma)) \ .$$

PROOF. Let $x > 0$. By (Laurent and Massart, 2000, Lemma 1),

$$\mathbb{P}(\|Z\|^2 \geq 2\{\sqrt{d} + \sqrt{x}\}^2) \leq \mathbb{P}(\|Z\|^2 \geq d + 2\sqrt{dx} + 2x) \leq \mathrm{e}^{-x} ,$$

where Z is a d-dimensional standard Gaussian vector. Setting $t = 2\{\sqrt{d} + \sqrt{x}\}^2$, we obtain

$$\mathbb{P}(\|Z\|^2 \geq t) \leq \exp\left(-\left\{d + t/2 - \sqrt{2td}\right\}\right) ,$$

and for $\sqrt{t} \geq 5\sqrt{d}$, we get $\mathbb{P}(\|Z\| \geq \sqrt{t}) \leq \mathrm{e}^{-t/4}$ which gives the result. □

PROPOSITION 23. *Assume* ***H1*** *and* ***H4****. There exist $\bar{\gamma} > 0$, $\varpi > 0$, and $K_5, \bar{b} \geq 0$ such that for any $\gamma \in (0, \bar{\gamma}]$ and $x \in \mathbb{R}^d$,*

$$R_\gamma V_{\bar{\eta}}(x) \leq (1 - \varpi\gamma)V_{\bar{\eta}}(x) + \bar{b}\gamma \mathbb{1}_{\mathrm{B}(0,K_5)}(x) ,$$

where R_γ is the Markov kernel of MALA defined by (36) and $\bar{\eta}$ is given by (79).

PROOF. Let $\bar{\gamma}_1 = m/(4L^2)$. By Proposition 17, for any $\gamma \in (0, \bar{\gamma}_1]$ and $x \in \mathbb{R}^d$,

$$\begin{aligned}
&R_\gamma V_{\bar{\eta}}(x) \\
&= Q_\gamma V_{\bar{\eta}}(x) + \int_{\mathbb{R}^d} \{1 - \min(1, \mathrm{e}^{-\tau_\gamma^{\mathrm{MALA}}(x,z)}\}\{V_{\bar{\eta}}(x) - V_{\bar{\eta}}(x - \gamma\nabla U(x) + \sqrt{2\gamma}z)\}\boldsymbol{\varphi}(z)\mathrm{d}z \\
&\leq Q_\gamma V_{\bar{\eta}}(x) + V_{\bar{\eta}}(x) \int_{\mathbb{R}^d} \{1 - \min(1, \mathrm{e}^{-\tau_\gamma^{\mathrm{MALA}}(x,z)}\}\boldsymbol{\varphi}(z)\mathrm{d}z \\
&\leq \mathrm{e}^{-\bar{\eta}m\gamma\|x\|^2/4}V_{\bar{\eta}}(x) + b_{\bar{\eta}}\gamma\mathbb{1}_{\mathrm{B}(0,K_3)}(x) + V_{\bar{\eta}}(x) \int_{\mathbb{R}^d} \{1 - \min(1, \mathrm{e}^{-\tau_\gamma^{\mathrm{MALA}}(x,z)}\}\boldsymbol{\varphi}(z)\mathrm{d}z ,
\end{aligned}$$

where K_3 and $b_{\bar{\eta}}$ are given in (79). Let

$$\bar{\gamma}_2 = \min\left(1, \bar{\gamma}_1, m^3/(4L^4)\right) , \quad \mathsf{K}_1 = \max\left(1, 2K_1, K_2, K_3, 20\sqrt{2\bar{\gamma}_2 d}\right) .$$

Then, by Lemma 18 and Lemma 22, there exist $C_1 \geq 0$ such that for any $x \in \mathbb{R}^d$, $\|x\| \geq \mathsf{K}_1$ and $\gamma \in (0, \bar{\gamma}_2]$,

$$\begin{aligned}
R_\gamma V_{\bar{\eta}}(x) &\leq \mathrm{e}^{-\bar{\eta}m\gamma\|x\|^2/4}V_{\bar{\eta}}(x) + V_{\bar{\eta}}(x)\left\{C_1\gamma + \exp(-\|x\|^2/(128\gamma))\right\} \\
&\leq \mathrm{e}^{-\bar{\eta}m\gamma\|x\|^2/4}V_{\bar{\eta}}(x) + V_{\bar{\eta}}(x)\left\{C_1\gamma + \exp(-1/(128\gamma))\right\} .
\end{aligned}$$

Using that there exists $C_2 \geq 0$ such that $\sup_{t\in(0,1)}\{t^{-1}\exp(-1/(128t))\} \leq C_2$ we get for any $x \in \mathbb{R}^d$, $\|x\| \geq \mathsf{K}_1$, $\gamma \in (0, \bar{\gamma}_2]$,

$$R_\gamma V_{\bar{\eta}}(x) \leq \mathrm{e}^{-\bar{\eta}m\gamma\|x\|^2/4}V_{\bar{\eta}}(x) + V_{\bar{\eta}}(x)\gamma\left\{C_1 + C_2\right\} .$$

Let

$$\mathsf{K}_2 = \max\left(\mathsf{K}_1, 4(C_1 + C_2)^{1/2}(\bar{\eta}m)^{-1/2}\right) , \quad \bar{\gamma}_3 = \min\left(\bar{\gamma}_2, 4\left\{m\bar{\eta}\mathsf{K}_2^2\right\}^{-1}\right) .$$

Then, since for any $t \in [0,1]$, $\mathrm{e}^{-t} \leq 1 - t/2$, we get for any $x \in \mathbb{R}^d$, $\|x\| \geq \mathsf{K}_2$, $\gamma \in (0, \bar{\gamma}_3]$,

$$\begin{aligned} R_\gamma V_{\bar{\eta}}(x) &\leq \mathrm{e}^{-\bar{\eta} m \gamma \mathsf{K}_2^2/4} V_{\bar{\eta}}(x) + V_{\bar{\eta}}(x)\gamma \{C_1 + C_2\} \\ &\leq \left[1 - \gamma\left\{\bar{\eta} m \mathsf{K}_2^2/8 - C_1 - C_2\right\}\right] V_{\bar{\eta}}(x) \\ &\leq \left\{1 - \gamma \bar{\eta} m \mathsf{K}_2^2/16\right\} V_{\bar{\eta}}(x) \; . \end{aligned} \tag{96}$$

In addition, by Lemma 7, using that for any $t \in \mathbb{R}$, $1 - \min(1, \mathrm{e}^{-t}) \leq |t|$, there exists $C_3 \geq 0$ such that for any $x \in \mathbb{R}^d$, $\|x\| \leq \mathsf{K}_2$ and $\gamma \in (0, \bar{\gamma}_3]$,

$$\begin{aligned} R_\gamma V_{\bar{\eta}}(x) &\leq V_{\bar{\eta}}(x) + b_{\bar{\eta}}\gamma \mathbb{1}_{\mathrm{B}(0,K_3)}(x) + C_3 \gamma^{3/2} \int_{\mathbb{R}^d} \{1 + \|x\|^2 + \|z\|^4\} \boldsymbol{\varphi}(z) \mathrm{d}z \\ &\leq (1 - \gamma \bar{\eta} m \mathsf{K}_2^2/16) V_{\bar{\eta}}(x) + \gamma \bar{\eta} m \mathsf{K}_2^2 \mathrm{e}^{\bar{\eta}\mathsf{K}_2^2}/16 + \gamma b_{\bar{\eta}} + C_3 \gamma \bar{\gamma}_3^{1/2} \left\{1 + \mathsf{K}_2^2 + C_4\right\} \;, \end{aligned}$$

where $C_4 = \int_{\mathbb{R}^d} \|z\|^4 \boldsymbol{\varphi}(z) \mathrm{d}z$. Combining this result and (96) completes the proof.

□

SUPPLEMENTARY MATERIAL

" **Supplement A: Supplement to: Diffusion approximations and control variates for MCMC**
(doi: 10.1214/00-AOASXXXXSUPP; .pdf). Supplementary document

References.

Assaraf, R. and Caffarel, M. (1999). Zero-variance principle for Monte Carlo algorithms. *Physical review letters* **83** 4682.

Bakry, D., Gentil, I. and Ledoux, M. (2014). *Analysis and geometry of Markov diffusion operators. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **348**. Springer, Cham. MR3155209

Belomestny, D. V., Iosipoi, L. S. and Zhivotovskiy, N. K. (2018). Variance Reduction in Monte Carlo Estimators via Empirical Variance Minimization. *Doklady Mathematics* **98** 494–497.

Bhattacharya, R. N. (1982). On Classical Limit Theorems for Diffusions. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* **44** 47-71.

Brosse, N., Durmus, A., Meyn, S., Moulines, É. and Radhakrishnan, A. (2019). Supplementary document: "Diffusion approximations and control variates for MCMC". working paper or preprint.

Cattiaux, P., Chafai, D. and Guillin, A. (2012). Central limit theorems for additive functionals of ergodic Markov diffusions processes. *ALEA* **9** 337–382.

De Bortoli, V. and Durmus, A. (2019). Convergence of diffusions and their discretizations:from continuous to discrete processes and back. *arXiv preprint arXiv:1904.09808*.

Dellaportas, P. and Kontoyiannis, I. (2012). Control variates for estimation based on reversible Markov chain Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74**.

Douc, R., Moulines, E., Priouret, P. and Soulier, P. (2018). *Markov chains. Springer Series in Operations Research and Financial Engineering*. Springer.

Eberle, A. (2015). Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields* 1-36.

Eberle, A. and Majka, M. B. (2018). Quantitative contraction rates for Markov chains on general state spaces. *arXiv preprint arXiv:1808.07033*.

Flegal, J. M. and Jones, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann. Statist.* **38** 1034–1070.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2014). *Bayesian data analysis* **2**. CRC press Boca Raton, FL.

Glynn, P. W. and Meyn, S. P. (1996). A Liapunov bound for solutions of the Poisson equation. *Ann. Probab.* **24** 916–931.

Grenander, U. (1983). Tutorial in pattern theory. Division of Applied Mathematics, Brown University, Providence.

Grenander, U. and Miller, M. I. (1994). Representations of knowledge in complex systems. *J. Roy. Statist. Soc. Ser. B* **56** 549–603. With discussion and a reply by the authors. MR1293234 (96e:62003)

Henderson, S. G. (1997). Variance reduction via an approximating markov process, PhD thesis, Department of Operations Research, Stanford University Available at http://people.orie.cornell.edu/shane/pubs/thesis.pdf.

Kopec, M. (2015). Weak backward error analysis for overdamped Langevin processes. *IMA J. Numer. Anal.* **35** 583–614. MR3335217

Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** 1302–1338.

Liu, J. S. (2008). *Monte Carlo strategies in scientific computing*. Springer Science & Business Media.

Marin, J.-M. and Robert, C. (2007). *Bayesian core: a practical approach to computational Bayesian statistics*. Springer Science & Business Media.

Meyn, S. (2008). *Control techniques for complex networks*. Cambridge University Press.

Meyn, S. and Tweedie, R. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge University Press, New York, NY, USA.

Mira, A., Solgi, R. and Imparato, D. (2013). Zero variance Markov chain Monte Carlo for Bayesian estimators. *Statistics and Computing* **23** 653–662.

Oates, C. J., Girolami, M. and Chopin, N. (2016). Control functionals for Monte Carlo integration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* n/a–n/a.

Papamarkou, T., Mira, A. and Girolami, M. (2014). Zero Variance Differential Geometric Markov Chain Monte Carlo Algorithms. *Bayesian Anal.* **9** 97–128.

Parisi, G. (1981). Correlation functions and computer simulations. *Nuclear Physics B* **180** 378–384.

Robert, C. P. and Casella, G. (2004). *Monte Carlo statistical methods*, second ed. *Springer Texts in Statistics*. Springer-Verlag, New York. MR2080278 (2005d:62006)

Roberts, G. O., Gelman, A. and Gilks, W. R. (1997). Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. *The Annals of Applied Probability* **7** 110-120.

Roberts, G. O. and Rosenthal, J. S. (1998). Optimal Scaling of Discrete Approximations to Langevin Diffusions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **60** 255-268.

Roberts, G. O. and Tweedie, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273 (98j:62014)

Rubinstein, R. Y. and Kroese, D. P. (2017). *Simulation and the Monte Carlo method*. *Wiley Series in Probability and Statistics*. John Wiley & Sons, Inc., Hoboken, NJ Third edition [of MR0624270]. MR3617204

South, L. F., Mira, A. and Drovandi, C. (2018). Regularised Zero-Variance Control Variates. *arXiv preprint arXiv:1811.05073*.

Zhu, Z., Wan, R. and Zhong, M. (2018). Neural Control Variates for Variance Reduction. *arXiv e-prints* arXiv:1806.00159.

Centre de Mathématiques Appliquées, UMR 7641,
Ecole Polytechnique, Palaiseau, France.
E-mail: nicolas.brosse@polytechnique.edu
eric.moulines@polytechnique.edu

Ecole Normale Supérieure CMLA,
61 Av. du Président Wilson,
94235 Cachan Cedex, France.
E-mail: alain.durmus@cmla.ens-cachan.fr

University of Florida,
Department of Electrical and Computer Engineering,
Gainesville, Florida.
E-mail: meyn@ece.ufl.edu
a4anandr@ufl.edu