Normalizing constants of log-concave densities

Nicolas Brosse, Alain Durmus, Éric Moulines

Normalizing constants of log-concave densities

Nicolas Brosse

Centre de Mathématiques Appliquées, UMR 7641, Ecole Polytechnique, France.
e-mail: nicolas.brosse@polytechnique.edu

Alain Durmus

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

and

Éric Moulines

Centre de Mathématiques Appliquées, UMR 7641, Ecole Polytechnique, France.
e-mail: eric.moulines@polytechnique.edu

Abstract: We derive explicit bounds for the computation of normalizing constants Z for log-concave densities $\pi = \mathrm{e}^{-U}/Z$ w.r.t. the Lebesgue measure on $\mathbb{R}^d$. Our approach relies on a Gaussian annealing combined with recent and precise bounds on the Unadjusted Langevin Algorithm [15]. Polynomial bounds in the dimension d are obtained with an exponent that depends on the assumptions made on U. The algorithm also provides a theoretically grounded choice of the annealing sequence of variances. A numerical experiment supports our findings. Results of independent interest on the mean squared error of the empirical average of locally Lipschitz functions are established.

MSC 2010 subject classifications: Primary 65C05, 60F25, 62L10; secondary 65C40, 60J05, 74G10, 74G15.
Keywords and phrases: Normalizing constants, Bayes Factor, Annealed Importance Sampling, Unadjusted Langevin Algorithm.

1. Introduction

Let $U : \mathbb{R}^d \to \mathbb{R}$ be a continuously differentiable convex function such that $Z = \int_{\mathbb{R}^d} \mathrm{e}^{-U(x)}\mathrm{d}x < +\infty$. Z is the normalizing constant of the probability density π associated with the potential U, defined for $x \in \mathbb{R}^d$ by $\pi(x) = Z^{-1}\mathrm{e}^{-U(x)}$. We discuss in this paper a method to estimate Z with polynomial complexity in the dimension d.

Computing the normalizing constant is a challenge which has applications in Bayesian inference and statistical physics in particular. In statistical physics, Z is better known under the name of partition function or free energy [3], [30]. Free energy differences allow to quantify the relative likelihood of different states (microscopic configurations) and are linked to thermodynamic work and heat exchanges. In Bayesian inference, the models can be compared by the computation of the Bayes factor which is the ratio of two normalizing constants (see e.g. [43, chapter 7]). This problem has consequently attracted a wealth of contribution; see for example [9, chapter 5], [31], [20], [2], [16], [29], [49] and, for a more specific molecular simulations flavor, [30]. Compared to the large number of proposed methods to estimate Z, few theoretical guarantees have been obtained on the ouput of these algorithms; see below for further references and comments. Our algorithm relies on a sequence of Gaussian densities with increasing variances, combined with the precise bounds of [15].

The paper is organized as follows. The outline of the algorithm is first described, followed by the assumptions made on U. Our main results are then stated and compared to previous works on the subject. The theoretical analysis of the algorithm is detailed in Section 2. In Section 3, a numerical experiment is provided to support our theoretical claims. Finally, the proofs are gathered in Section 5. In Section 4, a result of independent interest on the mean squared error of the empirical average of locally Lipschitz functions is established.

Notations and conventions

Denote by $\mathcal{B}(\mathbb{R}^d)$ the Borel σ-field of $\mathbb{R}^d$. For μ a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and f a μ-integrable function, denote by $\mu(f)$ the integral of f w.r.t. μ. We say that ζ is a transference plan of μ and ν if it is a probability measure on $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d))$ such that for all measurable sets A of $\mathbb{R}^d$, $\zeta(\mathrm{A} \times \mathbb{R}^d) = \mu(\mathrm{A})$ and $\zeta(\mathbb{R}^d \times \mathrm{A}) = \nu(\mathrm{A})$. We denote by $\Pi(\mu, \nu)$ the set of transference plans of μ and ν. Furthermore, we say that a couple of $\mathbb{R}^d$-random variables (X, Y) is a coupling of μ and ν if there exists $\zeta \in \Pi(\mu, \nu)$ such that (X, Y) are distributed according to ζ. For two probability measures μ and ν, we define the Wasserstein distance of order $p \geq 1$ as

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

By [46, Theorem 4.1], for all μ, ν probability measure on $\mathbb{R}^d$, there exists a transference plan $\zeta^\star \in \Pi(\mu, \nu)$ such that the infimum in (1) is reached in $\zeta^\star$. $\zeta^\star$ is called an optimal transference plan associated with W_p.

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

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

For $k \in \mathbb{N}$, $\mathcal{C}^k(\mathbb{R}^d)$ denotes the set of k-continuously differentiable functions $\mathbb{R}^d \to \mathbb{R}$, with the convention that $\mathcal{C}^0(\mathbb{R}^d)$ is the set of continuous functions. Let $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$, $n, m \in \mathbb{N}^*$ and $F : \mathbb{R}^n \to \mathbb{R}^m$ be a twice continuously differentiable function. Denote by ∇F and $\nabla^2 F$ the Jacobian and the Hessian of F respectively. For $m = 1$, the Laplacian is defined by $\Delta F = \operatorname{Tr} \nabla^2 F$ where Tr is the trace operator. In the sequel, we take the convention that for $n, p \in \mathbb{N}$,

$n < p$ then $\sum_p^n = 0$ and $\prod_p^n = 1$. By convention, $\inf\{\emptyset\} = +\infty$, $\sup\{\emptyset\} = -\infty$ and for $j > i$ in $\mathbb{Z}$, $\{j, \ldots, i\} = \emptyset$. For a finite set E, $|E|$ denotes the cardinality of E. For $a, b \in \mathbb{R}$, $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. Let $\psi, \phi : \mathbb{R}_+^\star \to \mathbb{R}_+^\star$. We write $\psi = \tilde{\mathcal{O}}(\phi)$ if there exists $t_0 > 0$, $C, c > 0$ such that $\psi(t) \leq C\phi(t) |\log t|^c$ for all $t \in (0, t_0]$. Denote by $\mathrm{B}(x, r) = \{y \in \mathbb{R}^d : \|y - x\| \leq r\}$.

Presentation of the algorithm

Since $Z < +\infty$ and U is convex, by [7, Lemma 2.2.1], there exist constants $\rho_1 > 0$ and $\rho_2 \in \mathbb{R}$ such that $U(x) \geq \rho_1 \|x\| - \rho_2$. Therefore, by continuity, U has a minimum $x^\star$. Without loss of generality, it is assumed in the sequel that $x^\star = 0$ and $U(x^\star) = 0$.

Let $M \in \mathbb{N}^\star$, $\{\sigma_i^2\}_{i=0}^M$ be a positive increasing sequence of real numbers and set $\sigma_M^2 = +\infty$. Consider the sequence of functions $\{U_i\}_{i=0}^M$ defined for all $i \in \{0, \ldots, M\}$ and $x \in \mathbb{R}^d$ by

$$U_i(x) = \frac{\|x\|^2}{2\sigma_i^2} + U(x) , \tag{2}$$

with the convention $1/\infty = 0$. We define a sequence of probability densities $\{\pi_i\}_{i=0}^M$ for $i \in \{0, \ldots, M\}$ and $x \in \mathbb{R}^d$ by

$$\pi_i(x) = Z_i^{-1} \mathrm{e}^{-U_i(x)} , \qquad Z_i = \int_{\mathbb{R}^d} \mathrm{e}^{-U_i(y)} \mathrm{d}y . \tag{3}$$

The dependence of Z_i in σ_i^2 is implicit. By definition, note that $U_M = U$, $Z_M = Z$ and $\pi_M = \pi$. As in the multistage sampling method [22, Section 3.3], we use the following decomposition

$$\frac{Z}{Z_0} = \prod_{i=0}^{M-1} \frac{Z_{i+1}}{Z_i} . \tag{4}$$

Z_0 is estimated by choosing σ_0^2 small enough so that π_0 is sufficiently close to a Gaussian distribution of mean 0 and covariance $\sigma_0^2 \operatorname{Id}$. For $i \in \{0, \ldots, M-1\}$, the ratio Z_{i+1}/Z_i may be expressed as

$$\frac{Z_{i+1}}{Z_i} = \int_{\mathbb{R}^d} g_i(x)\pi_i(x)\mathrm{d}x = \pi_i(g_i) , \tag{5}$$

where $g_i : \mathbb{R}^d \to \mathbb{R}_+$ is defined for all $x \in \mathbb{R}^d$ by

$$g_i(x) = \exp\left(a_i \|x\|^2\right) , \qquad a_i = \frac{1}{2}\left(\frac{1}{\sigma_i^2} - \frac{1}{\sigma_{i+1}^2}\right) . \tag{6}$$

The quantity $\pi_i(g_i)$ is estimated by the Unadjusted Langevin Algorithm (ULA) targeting π_i. Introduced in [18] and [38] (see also [44]), the ULA algorithm can be described as follows. For $i \in \{0,\ldots,M-1\}$, the (overdamped) Langevin stochastic differential equation (SDE) is given by

$$\mathrm{d}Y_{i,t} = -\nabla U_i(Y_{i,t})\mathrm{d}t + \sqrt{2}\mathrm{d}B_{i,t} \;, \quad Y_{i,0} = 0 \;, \tag{7}$$

where $\{(B_{i,t})_{t\geq 0}\}_{i=0}^{M-1}$ are independent d-dimensional Brownian motions. The sampling method is based on the Euler discretization of the Langevin diffusion, which defines a discrete-time Markov chain, for $i \in \{0,\ldots,M-1\}$ and $k \in \mathbb{N}$

$$X_{i,k+1} = X_{i,k} - \gamma_i \nabla U_i(X_{i,k}) + \sqrt{2\gamma_i} W_{i,k+1} \;, \quad X_{i,0} = 0 \;, \tag{8}$$

where $\{(W_{i,k})_{k\in\mathbb{N}^\star}\}_{i=0}^{M-1}$ are independent i.i.d. sequences of standard Gaussian random variables and $\gamma_i > 0$ is the stepsize. For $i \in \{0,\ldots,M-1\}$, consider the following estimator of Z_{i+1}/Z_i,

$$\hat{\pi}_i(g_i) = \frac{1}{n_i}\sum_{k=N_i+1}^{N_i+n_i} g_i(X_{i,k}) \;, \tag{9}$$

where $n_i \geq 1$ is the sample size and $N_i \geq 0$ the burn-in period. We introduce the following assumptions on U.

H 1. *$U : \mathbb{R}^d \to \mathbb{R}$ is continuously differentiable and L-gradient Lipschitz, i.e. there exists $L \geq 0$ such that for all $x, y \in \mathbb{R}^d$,*

$$\|\nabla U(x) - \nabla U(y)\| \leq L \|x - y\| \;. \tag{10}$$

H2 (m)**.** *$U : \mathbb{R}^d \to \mathbb{R}$ is continuously differentiable and satisfies for all $x, y \in \mathbb{R}^d$,*

$$U(y) \geq U(x) + \langle \nabla U(x), y - x\rangle + (m/2)\|x - y\|^2 \;. \tag{11}$$

H 3. *The function U is three times continuously differentiable and there exists $\tilde{L} \geq 0$ such that for all $x, y \in \mathbb{R}^d$*

$$\left\|\nabla^2 U(x) - \nabla^2 U(y)\right\| \leq \tilde{L}\|x - y\| \;. \tag{12}$$

The strongly convex case (**H**2(m) with $m > 0$) is considered in Section 2.1 and the convex case (**H**2(m) with $m = 0$) is dealt with in Section 2.2. Assuming **H**1 and **H**2(m) for $m \geq 0$, for $i \in \{0,\ldots,M\}$, U_i defined in (2) is L_i-gradient Lipschitz and m_i-strongly convex if $m_i > 0$ (and convex if $m_i = 0$) where

$$L_i = L + \frac{1}{\sigma_i^2} \;, \quad m_i = m + \frac{1}{\sigma_i^2} \;. \tag{13}$$

Define also the following useful quantities,

$$\kappa = \frac{2mL}{m+L} \;, \quad \kappa_i = \frac{2m_iL_i}{m_i + L_i} \;. \tag{14}$$

H3 enables to have tighter bounds on the mean squared error of $\hat{\pi}_i(g_i)$ defined in (9). Under **H**3, for all $i \in \{0,\ldots,M\}$, U_i satisfies (12) with $\tilde{L}$. Finally, since

$Z < +\infty$ and by [7, Lemma 2.2.1], there exist $\rho_1 > 0$ and $\rho_2 \in \mathbb{R}$ such that for all $x \in \mathbb{R}^d$,

$$U(x) \geq \rho_1 \|x\| - \rho_2 \; . \tag{15}$$

Denote by $\mathcal{S}$ the set of simulation parameters,

$$\mathcal{S} = \left\{M, \{\sigma_i^2\}_{i=0}^{M-1}, \{\gamma_i\}_{i=0}^{M-1}, \{n_i\}_{i=0}^{M-1}, \{N_i\}_{i=0}^{M-1}\right\} \;, \tag{16}$$

and by $\hat{Z}$ the following estimator of Z,

$$\hat{Z} = (2\pi\sigma_0^2)^{d/2}(1+\sigma_0^2 m)^{-d/2} \left\{\prod_{i=0}^{M-1} \hat{\pi}_i(g_i)\right\} \;, \tag{17}$$

where $\hat{\pi}_i(g_i)$ is defined in (9). The dependence of $\hat{Z}$ in $\mathcal{S}$ is implicit. Note that $\hat{Z}$ is a biased estimator of Z because Z_0 is approximated by $(2\pi\sigma_0^2)^{d/2}(1+\sigma_0^2 m)^{-d/2}$. We define the cost of the algorithm by the total number of iterations performed by the M Markov chains $(X_{i,n})_{n\geq 0}$ for $i \in \{0, \ldots, M-1\}$, *i.e.*

$$\text{cost} = \sum_{i=0}^{M-1} \{N_i + n_i\} \; . \tag{18}$$

Observe that each step of the Markov chain takes time linear in d. We state below a simplified version of our results; explicit bounds are given in Theorems 5, 6, 12 and 13.

Theorem 1. *Assume* $\boldsymbol{H}1$, $\boldsymbol{H}2(m)$ *for* $m \geq 0$. *Let* $\mu, \epsilon \in (0,1)$. *There exists an explicit choice of the simulation parameters* $\mathcal{S}$ *such that the estimator* $\hat{Z}$ *defined in* (17) *satisfies*

$$\mathbb{P}\left(\left|\hat{Z}/Z - 1\right| > \epsilon\right) \leq \mu \; . \tag{19}$$

Moreover, the cost *of the algorithm* (18) *is upper-bounded by,*

	$\boldsymbol{H}1$,$\boldsymbol{H}2(m)$ *for* $m > 0$
cost	$\frac{L^3}{\mu^2 m^3} \log(d) d^3 \times \tilde{\mathcal{O}}(\epsilon^{-4})$
	$\boldsymbol{H}1$,$\boldsymbol{H}2(m)$ *for* $m > 0$,$\boldsymbol{H}3$
cost	$\left(\frac{\tilde{L}}{\mu^{3/2} m^{3/2}} + \frac{L^2}{\mu^{3/2} m^2}\right) \log(d) d^{5/2} \times \tilde{\mathcal{O}}(\epsilon^{-3})$
	$\boldsymbol{H}1$,$\boldsymbol{H}2(m)$ *for* $m \geq 0$
cost	$\frac{L^2}{\mu^2 \rho_1^4}(d+\rho_2)^4 \log(d) d^3 \times \tilde{\mathcal{O}}(\epsilon^{-4})$
	$\boldsymbol{H}1$,$\boldsymbol{H}2(m)$ *for* $m \geq 0$,$\boldsymbol{H}3$
cost	$\left(\frac{L^2}{\mu^{3/2}\rho_1^4} + \frac{\tilde{L}}{\mu^{3/2}(d+\rho_2)\rho_1^3}\right)(d+\rho_2)^4 \log(d) d^{5/2} \times \tilde{\mathcal{O}}(\epsilon^{-3})$

By the median trick (see e.g. [27, Lemma 6.1] or [36]), the dependence in μ of the cost can be reduced to a logarithmic factor, see Corollaries 7 and 14.

It is interesting to compare these complexity bounds with previously reported results. In [33] and [5] (see also [12]), the authors propose to use sequential Monte Carlo (SMC) samplers to estimate the normalizing constant Z of a probability distribution π. In [5], π is supported on a compact set K included in $\mathbb{R}^d$ and satisfies for $x = (x_1, \ldots, x_d) \in \mathsf{K}$, $\pi(x) = Z^{-1} \prod_{i=1}^{d} \exp(g(x_i))$. [5, Theorem 3.2] states that there exists an estimator $\hat{Z}$ of Z such that $\lim_{d\to+\infty} \mathbb{E}[|\hat{Z}/Z - 1|^2] = C/N$ where N is the number of particles and C depends on g and on the parameters of the SMC (choice of the Markov kernel and of the annealing schedule). With our definition (18), the computational cost of the SMC algorithm is $\mathcal{O}(Nd)$ (there are d phases and N particles for each phase). To obtain an estimator $\hat{Z}$ satisfying (19) implies a cost of $d\mu^{-1}\mathcal{O}(\epsilon^{-2})$. However, the product form of the density π is restrictive, the result is only asymptotic in d and the state space is assumed to be compact. [13] combines SMC with a multilevel approach and [26] establishes results on a multilevel particle filter.

[23] deals with the case where $\pi(x) = \exp(-\beta H(x))/Z(\beta)$ where $x \in \Omega$, a finite state space, $\beta \geq 0$ and $H(x) \in \{0, \ldots, n\}$. These distributions known as Gibbs distributions include in particular the Ising model. To compute $Z(\beta)$, [23] relies on an annealing process on the parameter β, starting from $Z(0)$. Let $q = \log(Z(0))/\log(Z(\beta))$. [23, Theorem 1.1] states that there exists an estimator $\hat{Z}(\beta)$ of $Z(\beta)$ such that (19) is satisfied with $\mu = 1/4$ and $q \log(n)\tilde{\mathcal{O}}(\epsilon^{-2})$ draws from the Gibbs distribution.

Our complexity results can also be related to the computation of the volume of a convex body K (compact convex set with non-empty interior) on $\mathbb{R}^d$. This problem has attracted a lot of attention in the field of computer science, starting with the breakthrough of [17] until the most recent results of [10]. Define for $x \in \mathbb{R}^d$, $\pi(x) = \mathbb{1}_{\mathsf{K}}(x)/\operatorname{Vol}(\mathsf{K})$. Under the assumptions $\mathrm{B}(0,1) \subset \mathsf{K}$ and $\int_{\mathbb{R}^d} \|x\|^2 \pi(x)\mathrm{d}x = \mathcal{O}(d)$, [10, Theorem 1.1] states that there exists an estimator $\hat{Z}$ of $Z = \operatorname{Vol}(\mathsf{K})$ such that (19) is satisfied with $\mu = 1/4$ and a cost of $\log(d)d^3\tilde{\mathcal{O}}(\epsilon^{-2})$.

Nonequilibrium methods have been recently developed and studied in order to compute free energy differences or Bayes factors, see [24] and [30, Chapter 4]. They are based on an inhomogeneous diffusion evolving (for example) from $t = 0$ to $t = 1$ such that π_0 and π_1 are the stationary distributions respectively for $t = 0$ and $t = 1$. Recently, [1] provided an asymptotic and non-asymptotic analysis of the bias and variance for estimators associated with this methodology. The main aim of this paper is to obtain polynomial complexity and inspection of their results suggests a cost of order d^{15} at most to compute an estimator $\hat{Z}$ satisfying (19). However, this cost may be due to the strategy of proofs.

Multistage sampling type algorithms are widely used and known under different names: multistage sampling [45], (extended) bridge sampling [22], annealed importance sampling (AIS) [34], thermodynamic integration [37], power posterior [4]. For the stability and accuracy of the method, the choice of the parameters (in our case $\{\sigma_i^2\}_{i=0}^{M-1}$) is crucial and is known to be difficult. Indeed, the issue has been pointed out in several articles under the names of tuning tempered transitions [4], temperature placement [21], annealing sequence [5, Sections 3.2.1, 4.1], temperature ladder [37, Section 3.3.2], effects of grid size [16], cooling schedule [10]. In Sections 2.1 and 2.2, we explicitly define the sequence $\{\sigma_i^2\}_{i=0}^{M-1}$.

2. Theoretical analysis of the algorithm

In this Section, we analyse the algorithm outlined in Section 1. The strongly convex and convex cases are considered in Sections 2.1 and 2.2, respectively. The choice of the simulation parameters $\mathcal{S}$ explicitly depends on the (strong) convexity of U. Throughout this Section, we assume that $L > m$; note that if $L = m$, π is a Gaussian density and Z is known. For $M \in \mathbb{N}^\star$ and $i \in \{0, \ldots, M-1\}$, we first provide an upper bound on the mean squared error MSE_i of $\hat{\pi}_i(g_i)$ defined by

$$\mathrm{MSE}_i = \mathbb{E}\left[\{\hat{\pi}_i(g_i) - \pi_i(g_i)\}^2\right] , \tag{20}$$

where $\pi_i(g_i)$ and $\hat{\pi}_i(g_i)$ are given by (5) and (9) respectively. The MSE_i can be decomposed as a sum of the squared bias and variance,

$$\mathrm{MSE}_i = \{\mathbb{E}[\hat{\pi}_i(g_i)] - \pi_i(g_i)\}^2 + \mathrm{Var}\left[\hat{\pi}_i(g_i)\right] . \tag{21}$$

Propositions 2 and 3 give upper bounds on the squared bias and Proposition 4 on the variance. The results are based on the non-asymptotic bounds of the Wasserstein distance for a strongly convex potential obtained in [15] (see also [11], [14]). We introduce the following conditions on the stepsize γ_i used in the Euler discretization and the variance σ_{i+1}^2

$$\gamma_i \in \left(0, \frac{1}{m + L + 2/\sigma_i^2}\right] , \quad \sigma_{i+1}^2 \leq 2(d+4)\left(\frac{2d+7}{\sigma_i^2} - m\right)_+^{-1} , \tag{22}$$

where by convention $1/0 = +\infty$. Note that the condition on σ_{i+1}^2 is equivalent to $a_i \in \left[0, m_i/\{4(d+4)\} \wedge (2\sigma_i^2)^{-1}\right]$ where a_i is defined in (6) and m_i in (13). Assuming that γ_i and σ_{i+1}^2 satisfy (22), we define the positive quantities

$$C_{i,0} = \exp\left(\frac{4a_i(d+2)}{\kappa_i - 8a_i}\right) , \quad C_{i,1} = 2d\frac{1 - 8a_i\gamma_i}{\kappa_i - 8a_i} , \quad C_{i,2} = 4\frac{d}{m_i} , \tag{23}$$

where m_i, L_i and κ_i are defined in (13) and (14), respectively. Denote by,

$$A_{i,0} = 2L_i^2\kappa_i^{-1}d , \tag{24}$$

$$A_{i,1} = 2dL_i^2 + dL_i^4(\kappa_i^{-1} + (m_i + L_i)^{-1})(m_i^{-1} + 6^{-1}(m_i + L_i)^{-1}) . \tag{25}$$

Proposition 2. *Assume **H**1 and **H**2(m) for some $m \geq 0$. For $N_i \in \mathbb{N}$, $n_i \in \mathbb{N}^*$ and γ_i, σ_{i+1}^2 satisfying (22), we have*

$$\{\mathbb{E}[\hat{\pi}_i(g_i)] - \pi_i(g_i)\}^2 \leq 4a_i^2(C_{i,2} + C_{i,0}C_{i,1}) \times \left\{\frac{4d}{n_i m_i \kappa_i \gamma_i}\exp\left(-N_i\frac{\kappa_i\gamma_i}{2}\right) + 2\kappa_i^{-1}(A_{i,0}\gamma_i + A_{i,1}\gamma_i^2)\right\} .$$

Proof. The proof is postponed to Section 5.1. □

The squared bias can thus be controlled by adjusting the parameters γ_i, n_i and N_i. If U satisfies **H**3, the bound on the squared bias can be improved. Define,

$$B_{i,0} = d\left(2L_i^2 + \kappa_i^{-1}\{(d\tilde{L}^2)/3 + 4L_i^4/(3m_i)\}\right) , \tag{26}$$

$$B_{i,1} = dL_i^4\left(\kappa_i^{-1} + \{6(m_i+L_i)\}^{-1} + m_i^{-1}\right) . \tag{27}$$

Proposition 3. *Assume **H**1, **H**2(m) for some $m \geq 0$, and **H**3. For $N_i \in \mathbb{N}$, $n_i \in \mathbb{N}^*$ and γ_i, σ_{i+1}^2 satisfying (22), we have*

$$\{\mathbb{E}[\hat{\pi}_i(g_i)] - \pi_i(g_i)\}^2 \leq 4a_i^2(C_{i,2} + C_{i,0}C_{i,1}) \times \left\{\frac{4d}{n_i m_i \kappa_i \gamma_i}\exp\left(-N_i\frac{\kappa_i\gamma_i}{2}\right) + 2\kappa_i^{-1}(B_{i,0}\gamma_i^2 + B_{i,1}\gamma_i^3)\right\} .$$

Proof. The proof is postponed to Section 5.1. □

Note that the leading term is of order γ_i^2 instead of γ_i. We consider now the variance term in (21).

Proposition 4. *Assume **H**1 and **H**2(m) for some $m \geq 0$. For $N_i \in \mathbb{N}$, $n_i \in \mathbb{N}^*$ and γ_i, σ_{i+1}^2 satisfying (22), we have*

$$\mathrm{Var}\left[\hat{\pi}_i(g_i)\right] \leq \frac{32a_i^2 C_{i,0}C_{i,1}}{\kappa_i^2 n_i \gamma_i}\left(1 + \frac{2}{\kappa_i n_i \gamma_i}\right) .$$

Proof. The proof is postponed to Section 5.1. □

2.1. Strongly convex potential U

Theorem 5. *Assume **H**1 and **H**2(m) for $m > 0$ and let $\mu, \epsilon \in (0,1)$. There exists an explicit choice of the simulation parameters $\mathcal{S}$ (16) such that the estimator $\hat{Z}$ defined in (17) satisfies with probability at least $1-\mu$*

$$(1-\epsilon)Z \leq \hat{Z} \leq (1+\epsilon)Z ,$$

and the cost (18) *of the algorithm is upper-bounded by*

$$\mathrm{cost} \leq \left(\frac{6272C}{\epsilon^2\mu} + \log(5Cd^2)\right)\frac{(1088C)^2 d^2(d+4)}{\epsilon^2\mu}\left(\frac{m+L}{2m}\right)^3(C+3) , \tag{28}$$

with

$$C = \left\lceil \frac{1}{\log(2)}\log\left(d\left(d+\frac{7}{2}\right)\left(\frac{L}{m}-1\right)\frac{1}{\log(1+\epsilon/3)}\right)\right\rceil . \tag{29}$$

Proof. The proof is postponed to Section 5.3.3. □

Theorem 6. *Assume* **H**1, **H**2(m) *for* $m > 0$, **H**3 *and let* $\mu, \epsilon \in (0,1)$. *There exists an explicit choice of the simulation parameters* $\mathcal{S}$ (16) *such that the estimator* $\hat{Z}$ *defined in* (17) *satisfies with probability at least* $1-\mu$

$$(1-\epsilon)Z \leq \hat{Z} \leq (1+\epsilon)Z \;,$$

and the cost (18) *of the algorithm is upper-bounded by*

$$\text{cost} \leq \left(\frac{6272C}{\epsilon^2\mu} + \log(5Cd^2)\right)\sqrt{\frac{7}{3}}\frac{512Cd^{3/2}}{\epsilon\sqrt{\mu}}(d+4)(C+3) \\ \times\left\{\tilde{L}\frac{2^{3/2}}{m^{3/2}} + \sqrt{10}\left(\frac{m+L}{2m}\right)^2\right\} \;, \tag{30}$$

with C *defined in* (29).

Proof. The proof is postponed to Section 5.3.3. □

Using the median trick (see e.g. [27, Lemma 6.1] or [36]), we have the following corollary,

Corollary 7. *Let* $\epsilon, \tilde{\mu} \in (0,1)$. *Repeat* $2\left\lceil 4\log(\tilde{\mu}^{-1})\right\rceil + 1$ *times the algorithm of Theorems 5 and 6 with* $\mu = 1/4$ *and denote by* $\hat{Z}$ *the median of the output values. We have with probability at least* $1-\tilde{\mu}$,

$$(1-\epsilon)Z \leq \hat{Z} \leq (1+\epsilon)Z \;.$$

Proof. The proof is postponed to Section 5.3.3. □

The proof of Theorems 5 and 6 and corollary 7 relies on several lemmas which are stated below. These lemmas explain how the simulation parameters $\mathcal{S}$ must be chosen. The details of the proofs are gathered in Section 5.3. Set

$$\sigma_0^2 = \{2\log(1+\epsilon/3)\}/\{d(L-m)\} \;. \tag{31}$$

This choice of σ_0^2 is justified by the following result,

Lemma 8. *Under* **H**1 *and* **H**2(m) *for* $m \geq 0$, *we have*

$$Z_0 \leq (2\pi\sigma_0^2)^{d/2}/(1+\sigma_0^2 m)^{d/2} \leq Z_0\left(1+\epsilon/3\right) \;. \tag{32}$$

Proof. The proof is postponed to Section 5.3.1. □

Given a choice of $\mathcal{S}$, define the event

$$\mathrm{A}_{\mathcal{S},\epsilon} = \left\{\left|\prod_{i=0}^{M-1}\hat{\pi}_i(g_i) - \prod_{i=0}^{M-1}\pi_i(g_i)\right| \leq \prod_{i=0}^{M-1}\pi_i(g_i)\frac{\epsilon}{2}\right\} \;. \tag{33}$$

On $\mathrm{A}_{\mathcal{S},\epsilon}$, using Lemma 8, (4) and (17), we have:

$$Z\left(1-\epsilon/2\right) \leq \hat{Z} \leq Z\left(1+\epsilon\right) \;.$$

It remains to choose $\mathcal{S}$ to minimize approximately the cost defined in (18) under the constraint $\mathbb{P}(\mathrm{A}_{\mathcal{S},\epsilon}) \geq 1-\mu$. We define the positive increasing sequence $\{\sigma_i^2\}_{i=0}^{M-1}$ recursively, starting from $i=0$. For $i \in \mathbb{N}$, set

$$\sigma_{i+1}^2 = \varsigma_s(\sigma_i^2) \;, \tag{34}$$

where $\varsigma_s : \mathbb{R}_+^\star \to \mathbb{R}$ is defined for $t \in (0, (2d+7)/m)$ by

$$\varsigma_s(t) = \left(\frac{1}{t} - \frac{m + (2^{k(t)+1}\sigma_0^2)^{-1}}{2(d+4)} \right)^{-1} \;, \quad k(t) = \left\lfloor \frac{\log(t/\sigma_0^2)}{\log(2)} \right\rfloor \tag{35}$$

and $\varsigma_s(t) = +\infty$ otherwise. The subscript s in ς_s stresses that this choice is valid for the strongly convex case and will be different for the convex case. With this choice of $(\sigma_i^2)_{i \geq 0}$, the number of phases M is defined by

$$M = \inf\left\{ i \geq 1 : \sigma_{i-1}^2 \geq (2d+7)/m \right\} \;. \tag{36}$$

By (35), for $t \in \left[\sigma_0^2, (2d+7)/m\right)$, $\varsigma_s(t) \geq t(4d+16)/(4d+15)$, which implies $M < +\infty$. With this definition of ς_s, for $i \in \{0, \ldots, M-2\}$, we have

$$a_i = \frac{1}{2}\left(\frac{1}{\sigma_i^2} - \frac{1}{\sigma_{i+1}^2} \right) = \frac{m + (2^{k+1}\sigma_0^2)^{-1}}{4(d+4)} \;, \quad \text{if} \quad 2^k\sigma_0^2 \leq \sigma_i^2 < 2^{k+1}\sigma_0^2 \;, \tag{37}$$

and $a_{M-1} = (2\sigma_{M-1}^2)^{-1}$. Define $\mathcal{I}_k \subset \mathbb{N}$ for $k \in \mathbb{N}$ and $K \in \mathbb{N}$ by,

$$\mathcal{I}_k = \left\{ i \in \{0, \ldots, M-2\} : 2^k\sigma_0^2 \leq \sigma_i^2 < 2^{k+1}\sigma_0^2 \right\} \;, \tag{38}$$

$$K = \inf\left\{ k \geq 0 : \mathcal{I}_k = \emptyset \right\} < +\infty \;. \tag{39}$$

The number of phases M and variances $\{\sigma_i^2\}_{i=0}^{M-1}$ being defined, we now proceed with the choice of the stepsize γ_i, the number of samples n_i and the burn-in period N_i for $i \in \{0, \ldots, M-1\}$.

Lemma 9. *Set $\eta = (\epsilon\sqrt{\mu})/8$. Assume that there exists a choice of the simulation parameters $\{N_i\}_{i=0}^{M-1}$, $\{n_i\}_{i=0}^{M-1}$ and $\{\gamma_i\}_{i=0}^{M-1}$ satisfying,*

i) For all $k \in \{0, \ldots, K-1\}$, $i \in \mathcal{I}_k$,

$$\left|\mathbb{E}\left[\hat{\pi}_i(g_i)\right] - \pi_i(g_i)\right| \leq \frac{\eta}{K\left|\mathcal{I}_k\right|}, \quad \operatorname{Var}\left[\hat{\pi}_i(g_i)\right] \leq \frac{\eta^2}{K\left|\mathcal{I}_k\right|} \;,$$

ii) $\left|\mathbb{E}\left[\hat{\pi}_{M-1}(g_{M-1})\right] - \pi_{M-1}(g_{M-1})\right| \leq \eta, \quad \operatorname{Var}\left[\hat{\pi}_{M-1}(g_{M-1})\right] \leq \eta^2$,

where $\pi_i(g_i)$ is defined in (5) and $\hat{\pi}_i(g_i)$ in (9). Then $\mathbb{P}(\mathrm{A}_{\mathcal{S},\epsilon}) \geq 1-\mu$, where $\mathrm{A}_{\mathcal{S},\epsilon}$ is defined in (33).

Proof. The proof is postponed to Section 5.2 □

To show the existence of γ_i, n_i, N_i satisfying the conditions of Lemma 9 , we apply Propositions 2 to 4 for each $i \in \{0, \ldots, M-1\}$. We then have the following lemmas,

Lemma 10. *Set $\eta = (\epsilon\sqrt{\mu})/8$. Assume* **H1**, **H2**(m) *for $m > 0$ and,*

i) for all $k \in \{0, \ldots, K-1\}$, $i \in \mathcal{I}_k$,

$$\gamma_i \leq \frac{1}{2285}\frac{\eta^2\kappa_i^2\sigma_i^4 m_i}{K^2 d^2 L_i^2} \leq \frac{1}{m_i + L_i} , \tag{40}$$

$$n_i \geq \frac{196K}{\eta^2}\frac{\sqrt{m_i}}{\kappa_i\sigma_i}\frac{1}{\kappa_i\gamma_i} , \tag{41}$$

$$N_i \geq 2(\kappa_i\gamma_i)^{-1}\log\left(5Kd^2\right) , \tag{42}$$

ii)

$$\gamma_{M-1} \leq 40^{-1}\eta^2 L_{M-1}^{-2} m_{M-1} \leq (m_{M-1} + L_{M-1})^{-1} , \tag{43}$$

$$n_{M-1} \geq 19(\kappa_{M-1}\gamma_{M-1})^{-1}\eta^{-2} , \tag{44}$$

$$N_{M-1} \geq (\kappa_{M-1}\gamma_{M-1})^{-1} . \tag{45}$$

Then, the conditions i)-ii) of Lemma 9 are satisfied.

Proof. The proof is postponed to Section 5.3.2. □

We have a similar result under the additional assumption **H3**.

Lemma 11. *Set $\eta = (\epsilon\sqrt{\mu})/8$. Assume* **H1**, **H2**(m) *for $m > 0$,* **H3** *and,*

i) for all $k \in \{0, \ldots, K-1\}$, $i \in \mathcal{I}_k$,

$$\gamma_i \leq \sqrt{\frac{3}{7}}\frac{\eta\kappa_i m_i^{1/2}\sigma_i^2}{8Kd}\left(d\tilde{L}^2 + 10L_i^4 m_i^{-1}\right)^{-1/2} \leq \frac{1}{m_i + L_i} , \tag{46}$$

and n_i, N_i as in (41), (42),

ii)

$$\gamma_{M-1} \leq \sqrt{\frac{3}{7}}\frac{\eta\kappa_{M-1} m_{M-1}^{-1/2}}{4}\left(d\tilde{L}^2 + 10L_{M-1}^4 m_{M-1}^{-1}\right)^{-1/2} \leq \frac{1}{m_{M-1} + L_{M-1}} , \tag{47}$$

and n_{M-1}, N_{M-1} as in (44), (45).

Then, the conditions i)-ii) of Lemma 9 are satisfied.

Proof. The proof is postponed to the supplementary material [8, Appendix A.2]. □

2.2. Convex potential U

We now consider the convex case. The annealing process on the variances $\{\sigma_i^2\}_{i=0}^{M-1}$ is different from the strongly convex case and is defined in (54). In particular, the stopping criteria for the annealing process is distinct from the case where U is strongly convex and relies on a truncation argument. More precisely, a concentration theorem for log-concave functions [39, Theorem 3.1] states that for $\alpha \in (0,1)$,

$$\int_{\mathbb{R}^d} \mathbb{1}_{\{U \geq d(\tau_\alpha+1)\}}(x)\pi(x)\mathrm{d}x \leq \alpha \;, \quad \tau_\alpha = \left(\frac{16\log(3/\alpha)}{d}\right)^{1/2} \;.$$

Let $\epsilon \in (0,1)$, $\tau = \tau_{\epsilon/2}$ and $D = \rho_1^{-1}\{d(\tau+1)+\rho_2\}$. By (15), we have

$$\int_{\mathbb{R}^d} \mathbb{1}_{\mathrm{B}(0,D)}(x)\pi(x)\mathrm{d}x \geq 1-\epsilon/2 \;. \tag{48}$$

Given a choice of M and σ^2_{M-1}, define $\bar{g}_{M-1} : \mathbb{R}^d \to \mathbb{R}_+$ for all $x \in \mathbb{R}^d$ by

$$\bar{g}_{M-1}(x) = \exp\left\{\frac{1}{2\sigma^2_{M-1}}(\|x\|^2 \wedge D^2)\right\} \;, \tag{49}$$

and J by,

$$J = \int_{\mathbb{R}^d} \mathrm{e}^{-U(x)}\mathrm{d}x \Big/ \int_{\mathbb{R}^d} \mathrm{e}^{-U(x)-(\|x\|^2-D^2)_+/(2\sigma^2_{M-1})}\mathrm{d}x \;.$$

Note that $Z/Z_{M-1} = J \times \pi_{M-1}(\bar{g}_{M-1})$ and by (48),

$$J(1-\epsilon/2) \leq 1 \leq J \;. \tag{50}$$

On the event $\mathrm{A}_{\mathcal{S},\epsilon}$ defined in (33) with g_{M-1} replaced by $\bar{g}_{M-1}$ and by (32) (with $m=0$), (50), we get

$$Z\left(1-\epsilon/2\right)^2 \leq \hat{Z} \leq Z\left(1+\epsilon\right) \;,$$

where $\hat{Z}$ is defined in (17) with g_{M-1} replaced by $\bar{g}_{M-1}$. We now state our results in the convex case.

Theorem 12. *Assume* ***H****1,* ***H****2(m) for $m \geq 0$. Let $\epsilon, \mu \in (0,1)$. There exists an explicit choice of the simulation parameters $\mathcal{S}$ (16) such that the estimator $\hat{Z}$ defined in (17) (with g_{M-1} replaced by $\bar{g}_{M-1}$ defined in (49)) satisfies with probability at least $1-\mu$*

$$(1-\epsilon)Z \leq \hat{Z} \leq (1+\epsilon)Z \;,$$

and the cost *(18) of the algorithm is upper-bounded by*

$$\begin{aligned}\mathrm{cost} \leq &\left(\frac{17728C}{\epsilon^2\mu} + \log(Cd^2)\right)\frac{(487C)^2d^2(d+4)}{\epsilon^2\mu} \\ &\times\left(C + \frac{6L\{d(\tau+1)+\rho_2\}^2}{\rho_1^2} + \frac{8L^2\{d(\tau+1)+\rho_2\}^4}{3\rho_1^4}\right) \;,\end{aligned} \tag{51}$$

where ρ_1, ρ_2 are defined in (15), $\tau = 4d^{-1/2}\{\log(6/\epsilon)\}^{1/2}$ and

$$C = \left\lceil \frac{1}{\log(2)}\log\left(\frac{dL\{d(\tau+1)+\rho_2\}^2}{2\rho_1^2\log(1+\epsilon/3)}\right)\right\rceil \;. \tag{52}$$

Theorem 13. *Assume* **H** *1,* **H** *2(m) for* $m \geq 0$ *and* **H** *3. Let* $\epsilon, \mu \in (0,1)$. *There exists an explicit choice of the simulation parameters* $\mathcal{S}$ (16) *such that the estimator* $\hat{Z}$ *defined in* (17) *(with* g_{M-1} *replaced by* $\bar{g}_{M-1}$ *defined in* (49)*) satisfies with probability at least* $1-\mu$

$$(1-\epsilon)Z \leq \hat{Z} \leq (1+\epsilon)Z \;,$$

and the cost (18) *of the algorithm is upper-bounded by*

$$\begin{aligned}\mathrm{cost} \leq 2474 \left(\frac{17728 C}{\epsilon^2 \mu} + \log(Cd^2) \right) \frac{(C+1)d(d+4)}{\epsilon\sqrt{\mu}} \Bigg\{ \frac{8L^2\{d(\tau+1)+\rho_2\}^4}{3\rho_1^4} \\ + \frac{d^{1/2}\tilde{L}\{d(\tau+1)+\rho_2\}^3}{\sqrt{10}\rho_1^3} \max\left(\frac{5\rho_1}{d(\tau+1)+\rho_2}, \left(\frac{5}{9} + \frac{\rho_1^2}{\{d(\tau+1)+\rho_2\}^2 L} \right)^2 \right) \\ + \frac{6L\{d(\tau+1)+\rho_2\}^2}{\rho_1^2} + C \Bigg\} \;, \end{aligned} \tag{53}$$

where ρ_1, ρ_2, C *are defined in* (15), (52) *respectively and* $\tau = 4d^{-1/2}\{\log(6/\epsilon)\}^{1/2}$.

Corollary 14. *Let* $\epsilon, \tilde{\mu} \in (0,1)$. *Repeat* $2\left\lceil 4\log(\tilde{\mu}^{-1}) \right\rceil + 1$ *times the algorithm of Theorems 12 and 13 with* $\mu = 1/4$ *and denote by* $\hat{Z}$ *the median of the output values. We have with probability at least* $1-\tilde{\mu}$,

$$(1-\epsilon)Z \leq \hat{Z} \leq (1+\epsilon)Z \;.$$

The proofs follow the same arguments as Theorems 5 and 6 and corollary 7 and are detailed in the supplementary material [8, Appendix B.3].

Note that $\bar{g}_{M-1}$ (49) is a $\|\bar{g}_{M-1}\|_{\mathrm{Lip}}$-Lipschitz function where,

$$\|\bar{g}_{M-1}\|_{\mathrm{Lip}} = \frac{D}{\sigma_{M-1}^2} \exp\left(\frac{D^2}{2\sigma_{M-1}^2} \right) \;.$$

The results of Section 4 give an upper bound on MSE_{M-1} which is polynomial in the parameters if σ_{M-1}^2 is approximately equal to D^2. For $i \in \mathbb{N}^\star$, we define $(\sigma_i^2)_{i\geq 0}$ recursively. Set σ_0^2 as in (31) and

$$\sigma_{i+1}^2 = \varsigma_c(\sigma_i^2) \;, \tag{54}$$

where $\varsigma_c : \mathbb{R}_+^\star \to \mathbb{R}$ is defined for $t \in \left(0, D^2\right)$ by,

$$\varsigma_c(t) = \left(\frac{1}{t} - \frac{1}{2(d+4)(2^{k(t)+1}\sigma_0^2)} \right)^{-1} \;, \quad k(t) = \left\lfloor \frac{\log(t/\sigma_0^2)}{\log(2)} \right\rfloor \;, \tag{55}$$

and $\varsigma_c(t) = +\infty$ otherwise. Define M in this Section by,

$$M = \inf\left\{ i \geq 1 : \sigma_{i-1}^2 \geq D^2 \right\} \;. \tag{56}$$

By (55), for $t \in \left[\sigma_0^2, D^2\right)$, $\varsigma_c(t) \geq \{(4d+16)/(4d+15)\}\, t$, which implies $M < +\infty$. The following lemmas are the counterparts of Lemmas 10 and 11. They specify the choice of $\{\gamma_i\}_{i=0}^{M-1}$, $\{n_i\}_{i=0}^{M-1}$, $\{N_i\}_{i=0}^{M-1}$ to satisfy the conditions of Lemma 9.

Lemma 15. *Set* $\eta = (\epsilon\sqrt{\mu})/8$. *Assume* ***H****1*, ***H****2(m) for* $m \geq 0$ *and,*

i) for all $k \in \{0, \ldots, K-1\}$, $i \in \mathcal{I}_k$,

$$\gamma_i \leq \frac{1}{462}\frac{\eta^2 L_i^{-2}\sigma_i^{-2}}{K^2 d^2} \leq \frac{1}{m_i + L_i} \;, \tag{57}$$

$$n_i \geq \frac{453K}{\eta^2}\frac{1}{\kappa_i\gamma_i} \;, \tag{58}$$

$$N_i \geq 2(\kappa_i\gamma_i)^{-1}\log\left(Kd^2\right) \;, \tag{59}$$

ii)

$$\gamma_{M-1} \leq (1/26)\eta^2 d^{-1} L_{M-1}^{-2}\kappa_{M-1} \leq (m_{M-1} + L_{M-1})^{-1} \;, \tag{60}$$

$$n_{M-1} \geq 29\eta^{-2}(\kappa_{M-1}\gamma_{M-1})^{-1} \;, \tag{61}$$

$$N_{M-1} \geq 2(\kappa_{M-1}\gamma_{M-1})^{-1}\log(d) \;. \tag{62}$$

Then, the conditions i)-ii) of Lemma 9 are satisfied, with g_{M-1} *replaced by* $\bar{g}_{M-1}$.

Proof. The proof is postponed to the supplementary material [8, Appendix B.1]. □

We have a similar result under the additional assumption **H**3.

Lemma 16. *Set* $\eta = (\epsilon\sqrt{\mu})/8$. *Assume* ***H****1*, ***H****2(m) for* $m \geq 0$, ***H****3 and,*

i) for all $k \in \{0, \ldots, K-1\}$, $i \in \mathcal{I}_k$,

$$\gamma_i \leq \sqrt{\frac{3}{7}}\frac{\eta\sigma_i^{-1}}{8Kd}\left(d\tilde{L}^2 + 10L_i^4\sigma_i^2\right)^{-1/2} \leq \frac{1}{m_i + L_i} \;, \tag{63}$$

n_i, N_i *as in* (58), (59) *and,*

ii)

$$\gamma_{M-1} \leq \sqrt{\frac{3}{8\mathrm{e}}}\frac{\eta\kappa_{M-1}\sigma_{M-1}}{\sqrt{d}}\left(d\tilde{L}^2 + 10L_{M-1}^4\sigma_{M-1}^2\right)^{-1/2} \leq \frac{1}{m_{M-1} + L_{M-1}} \;, \tag{64}$$

n_{M-1}, N_{M-1} *as in* (61), (62).

Then, the conditions i)-ii) of Lemma 9 are satisfied, with g_{M-1} *replaced by* $\bar{g}_{M-1}$.

Proof. The proof is postponed to the supplementary material [8, Appendix B.2]. □

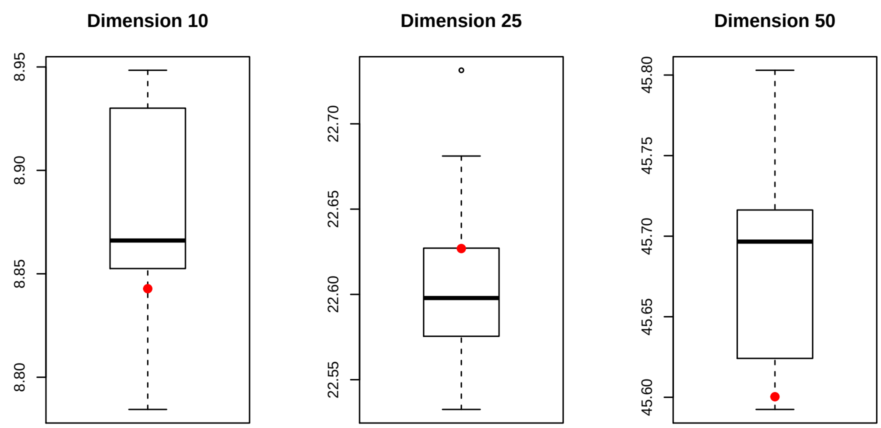

Fig 1. *Boxplots of the logarithm of the normalizing constants of a multivariate Gaussian distribution in dimension* $d \in \{10, 25, 50\}$.

3. Numerical experiments

For the following numerical experiments, the code and data are available at https://github.com/nbrosse/normalizingconstant. We first experiment our algorithm to compute the logarithm of the normalizing constant of a multivariate Gaussian distribution in dimension $d \in \{10, 25, 50\}$, of mean 0 and inverse covariance matrix $\mathrm{diag}(2, 1^{\otimes(d-1)})$. We set $\epsilon = \mu = 0.1$. The number of phases M of the algorithm and the variances $\{\sigma_i^2\}_{i=0}^{M-1}$ are chosen accordingly to the formulas (34) and (36). For each phase of the algorithm, the step size γ_i is set equal to $10^{-2}(m_i + L_i)^{-1}$, the burn-in period N_i to 10^4 and the number of samples n_i to 10^5 where m_i, L_i are defined in (13). We carry out 10 independent runs of the algorithm and compute the boxplots in Figure 1. The true values of the logarithm of the normalizing constants are known and displayed by the red points in Figure 1.

We illustrate then our methodology to compute Bayesian model evidence; see [20] and the references therein. Let $y \in \mathbb{R}^p$ be a vector of observations and $\mathcal{M}_1, \ldots, \mathcal{M}_l$ be a collection of competing models. Let $\{p(\mathcal{M}_i)\}_{i=1}^{l}$ be a prior distribution on the collection of models. For $i \in \{0, \ldots, l\}$, denote by $p(y|\theta^{(\mathcal{M}_i)}, \mathcal{M}_i)$ the likelihood of the model $\mathcal{M}_i$. The dominating measure is implicitly considered to be the Lebesgue measure on $\mathbb{R}^p$. Similarly, for $i \in \{0, \ldots, l\}$, denote by $p(\theta^{(\mathcal{M}_i)}|\mathcal{M}_i)$ the prior density on the parameters $\theta^{(\mathcal{M}_i)}$ under the model $\mathcal{M}_i$ where the dominating measure is implicitly considered to be the Lebesgue measure on $\mathbb{R}^{d^{(\mathcal{M}_i)}}$. The posterior distribution of interest is then for $i \in \{0, \ldots, l\}$,

$$p(\theta^{(\mathcal{M}_i)}, \mathcal{M}_i|y) \propto p(y|\theta^{(\mathcal{M}_i)}, \mathcal{M}_i)p(\theta^{(\mathcal{M}_i)}|\mathcal{M}_i)p(\mathcal{M}_i)$$

The posterior distribution conditional on model $\mathcal{M}_i$ can also be considered

$$p(\theta^{(\mathcal{M}_i)}|\mathcal{M}_i, y) \propto p(y|\theta^{(\mathcal{M}_i)}, \mathcal{M}_i)p(\theta^{(\mathcal{M}_i)}|\mathcal{M}_i) \tag{65}$$

For $i \in \{0, \ldots, l\}$, the evidence $p(y|\mathcal{M}_i)$ of the model $\mathcal{M}_i$ is defined by the normalizing constant for the posterior distribution (65)

$$p(y|\mathcal{M}_i) = \int_{\mathbb{R}^{d(\mathcal{M}_i)}} p(y|\theta^{(\mathcal{M}_i)}, \mathcal{M}_i)p(\theta^{(\mathcal{M}_i)}|\mathcal{M}_i)\mathrm{d}\theta^{(\mathcal{M}_i)} .$$

The Bayes factor BF_{12} between two models $\mathcal{M}_i$ and $\mathcal{M}_j$ is then defined by the ratio of evidences [43, Section 7.2.2], $BF_{ij} = p(y|\mathcal{M}_i)/p(y|\mathcal{M}_j)$. In the following experiments, we estimate the log evidence $\log(p(y|\mathcal{M}_i))$. For ease of notation, the dependence on the model $\mathcal{M}$ of the parameters θ and the dimension d of the state space is implicit in the sequel.

Define $\ell^{(\mathcal{M})} : \mathbb{R}^d \to \mathbb{R}$ by $\ell^{(\mathcal{M})}(\theta) = -\log(p(y|\theta, \mathcal{M})p(\theta|\mathcal{M}))$ for $\theta \in \mathbb{R}^d$. In the examples we consider, $\ell^{(\mathcal{M})}$ satisfies **H**1, **H**2, **H**3 and has a unique minimum $\theta_\star^{(\mathcal{M})}$. Define then $U^{(\mathcal{M})} : \mathbb{R}^d \to \mathbb{R}_+$ by $U^{(\mathcal{M})}(\theta) = \ell^{(\mathcal{M})}(\theta + \theta_\star^{(\mathcal{M})}) - \ell^{(\mathcal{M})}(\theta_\star^{(\mathcal{M})})$ for $\theta \in \mathbb{R}^d$. The algorithm described in Section 2 can be applied to $U^{(\mathcal{M})}$. For each example, two different models will be considered and $U^{(\mathcal{M})}$ will be written as $U^{(k)}$ for $k = 1, 2$.

The numerical experiments are carried out on a Gaussian linear and logistic regression following the experimental setup of [20, Section 4], which is now considered as a classical benchmark. The linear regression is conducted on $p = 42$ specimens of radiata pine [47]. The response variable $y \in \mathbb{R}^p$ is the maximum compression strength parallel to the grain. The explanatory variables are $x \in \mathbb{R}^p$ the density and $z \in \mathbb{R}^p$ the density adjusted for resin content. x and z are centered. The covariates of the first model $\mathcal{M}_1$, $X^{(1)} \in \mathbb{R}^{p\times 2}$, are composed of an intercept and x, while the covariates of the second model $\mathcal{M}_2$, $X^{(2)} \in \mathbb{R}^{p\times 2}$, are composed of an intercept and z. For $k = 1, 2$, the likelihood is defined by,

$$p(y|\theta, \mathcal{M}_k) = \left(\frac{\lambda}{2\pi}\right)^{d/2} \exp\left(-(\lambda/2)\left\|y - X^{(k)}\theta\right\|^2\right) ,$$

where $\lambda = 10^{-5}$. For the two models, the parameter θ follows the same Gaussian prior of mean $(3000, 185)$ and inverse covariance matrix $\lambda Q_0 = \lambda\,\mathrm{diag}(0.06, 6)$ where diag denotes a diagonal matrix. These values are taken from [20, section 4.1]. For $k = 1, 2$, $U^{(k)}$ is $m^{(k)}$-strictly convex and $L^{(k)}$-gradient Lipschitz, where $m^{(k)}$ (resp. $L^{(k)}$) is the minimal (resp. maximal) eigenvalue of $\lambda([X^{(k)}]^{\mathrm{T}}X^{(k)} + Q_0)$. We set $\epsilon = \mu = 0.1$. The number of phases M of the algorithm and the variances $\left\{\sigma_i^2\right\}_{i=0}^{M-1}$ are chosen accordingly to the formulas (34) and (36). For each phase, the step size γ_i is set equal to $10^{-2}(\kappa_i\sigma_i^2 m_i)/(dL_i^2)$, the burn-in period N_i to $10^3(\kappa_i\gamma_i)^{-1}$ and the number of samples n_i to $10^4 m_i^{1/2}/(\kappa_i^2\sigma_i\gamma_i)$

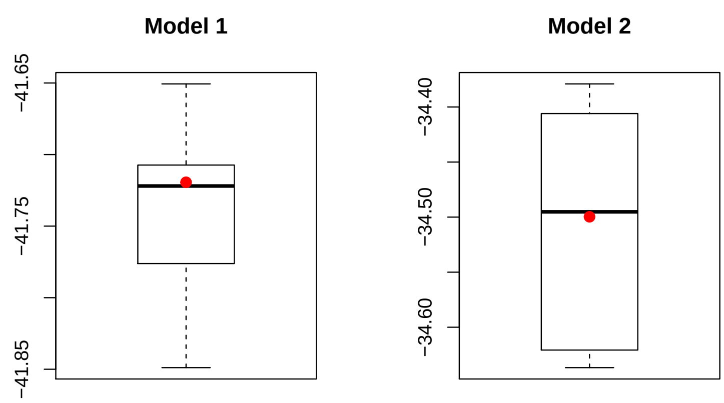

Fig 2. *Boxplots of the log evidence for the two models on the Gaussian regression.*

where m_i, L_i, κ_i are defined in (13) and (14). The experiments are repeated 10 times and the boxplots for each model $\mathcal{M}$ are plotted in Figure 2. Note that for this Gaussian model, the log evidence is known and displayed by the red points in Figure 2.

With the same parameters for the algorithm, we run 10 independent runs at each phase to measure the variability of each estimator $\hat{\pi}_i(g_i)$ defined in (9). The result is plotted in Figure 3 for the model $\mathcal{M}_1$. The last estimator $\hat{\pi}_{M-1}(g_{M-1})$ is much higher, which underlines the specificity of the last phase in the algorithm.

The logistic regression is performed on the Pima Indians dataset[1]. In this case, $y \in \{0,1\}^p$ is a vector of diabetes indicators for $p = 532$ Pima Indian women and the potential predictors for diabetes are: number of pregnancies $\mathrm{NP} \in \mathbb{R}^p$, plasma glucose concentration $\mathrm{PGC} \in \mathbb{R}^p$, diastolic blood pressure $\mathrm{BP} \in \mathbb{R}^p$, triceps skin fold thickness $\mathrm{TST} \in \mathbb{R}^p$, body mass index $\mathrm{BMI} \in \mathbb{R}^p$, diabetes pedigree function $\mathrm{DP} \in \mathbb{R}^p$ and age $\mathrm{AGE} \in \mathbb{R}^p$. These variates are centered and standardized. The covariates of the first model $\mathcal{M}_1$ are $X^{(1)} = (\mathrm{intercept}, \mathrm{NP}, \mathrm{PGC}, \mathrm{BMI}, \mathrm{DP}) \in \mathbb{R}^{p\times 5}$ and the covariates of the second model $\mathcal{M}_2$ are $X^{(2)} = (\mathrm{intercept}, \mathrm{NP}, \mathrm{PGC}, \mathrm{BMI}, \mathrm{DP}, \mathrm{AGE}) \in \mathbb{R}^{p\times 6}$, where intercept is the intercept of the regressions. The likelihood is defined for $k = 1, 2$ by,

$$p(y|\theta, \mathcal{M}_k) = \exp\left(\sum_{i=1}^{p}\left\{y_i \theta^{\mathrm{T}} X_i^{(k)} - \log\left(1 + \mathrm{e}^{\theta^{\mathrm{T}} X_i^{(k)}}\right)\right\}\right) ,$$

where $X_i^{(k)}$ denotes the i^{th} row of $X^{(k)}$. For the two models, the prior on θ

[1] http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes

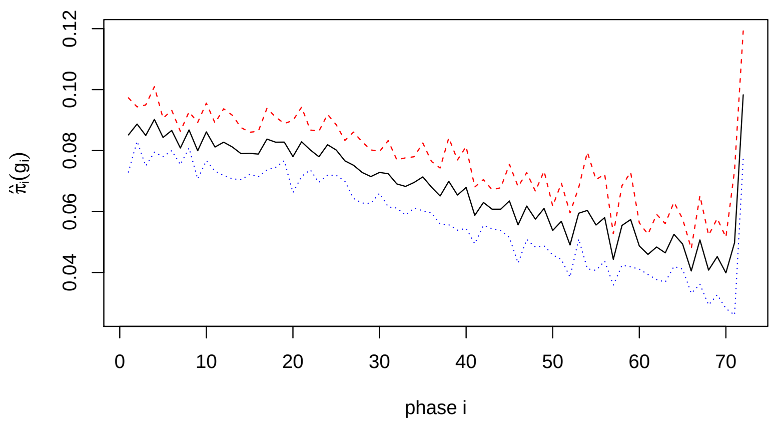

FIG 3. *Error plot of $\hat{\pi}_i(g_i)$ for $i \in \{0, \dots, M-1\}$ in the example of the Gaussian regression (model $\mathcal{M}_1$). The mean of $\hat{\pi}_i(g_i)$ is displayed in black and is spaced apart from the other two curves by the standard deviation of $\hat{\pi}_i(g_i)$.*

is Gaussian, of mean 0 and inverse covariance matrix $\tau \,\mathrm{Id}$ where $\tau = 0.01$. For $k = 1, 2$, $U^{(k)}$ is τ-strongly convex and $L^{(k)}$-gradient Lipschitz, where $L^{(k)} = \lambda_{\max}([X^{(k)}]^{\mathrm{T}} X^{(k)})/4 + \tau$ and $\lambda_{\max}([X^{(k)}]^{\mathrm{T}} X^{(k)})$ is the maximal eigenvalue of $[X^{(k)}]^{\mathrm{T}} X^{(k)}$. We set $\epsilon = \mu = 0.1$. The algorithm to estimate $\log(p(y|\mathcal{M}))$ described in Section 2.1 is applied with the following modifications. The number of phases is decreased and the recurrence for the variances $\{\sigma_i^2\}_{i=0}^{M-1}$ is thus redefined by $\sigma_{i+1}^2 = \varsigma_s^5(\sigma_i^2)$ as long as the stopping condition (36) is not fulfilled. For $i \in \{1, \dots, 30\}$, the burn-in period N_i is set equal to 10^4, the number of samples n_i to 10^6 and the step size γ_i to $10^{-2}(m_i + L_i)^{-1}$ where m_i, L_i are defined in (13); for $i > 30$, the number of samples n_i is set equal to 10^5 and the step size γ_i to $10^{-1}(m_i + L_i)^{-1}$. We compare our results with different methods reviewed in [20] and implemented in [48]. These are the Laplace method (L), Laplace at the Maximum a Posteriori (L-MAP), Chib's method (C) Annealed Importance Sampling (AIS) and Power Posterior (PP). The experiments are repeated 10 times and the boxplots for each model $\mathcal{M}$ and each method are plotted in Figure 4.

With the same parameters for the algorithm, we run 10 independent runs at each phase to measure the variability of each estimator $\hat{\pi}_i(g_i)$ defined in (9) and display the result in Figure 5 for the model $\mathcal{M}_1$.

The final example we address is a Bayesian analysis of a finite mixture of Gaussian distributions, see [33, Section 4.2] and we aim at estimating the log evidence of the posterior distribution. Note that this model does not fit into our assumptions because the potential U is not continuously differentiable on $\mathbb{R}^d$

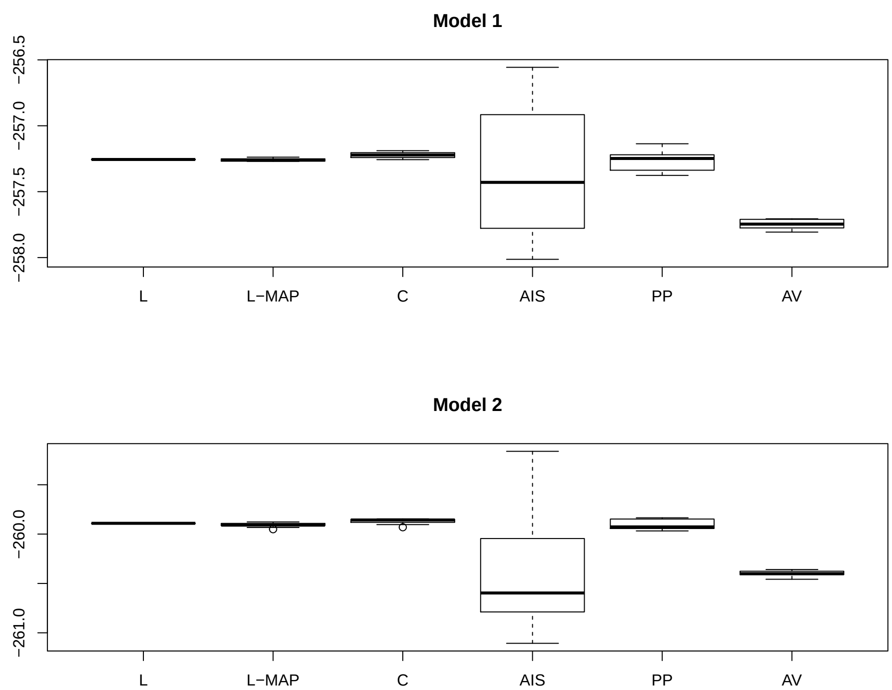

Fig 4. *Boxplots of the log evidence for the two models on the logistic regression. The methods are the Laplace method (L), Laplace at the Maximum a Posteriori (L-MAP), Chib's method (C), Annealed Importance Sampling (AIS), Power Posterior (PP) and our method (AV).*

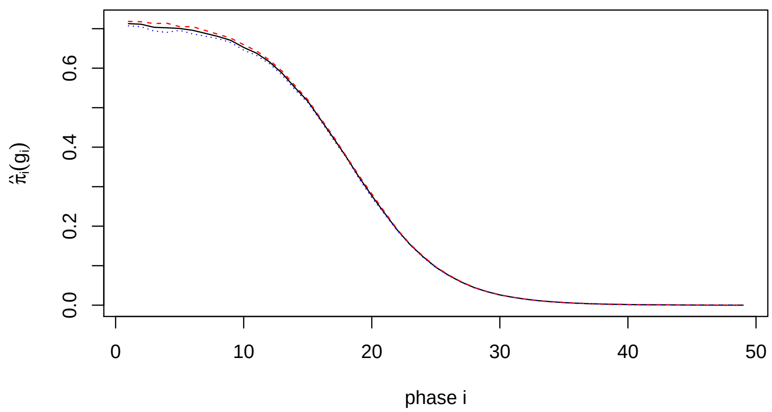

FIG 5. *Error plot of $\hat{\pi}_i(g_i)$ for $i \in \{0, \dots, M-1\}$ in the example of the logistic regression (model $\mathcal{M}_1$). The mean of $\hat{\pi}_i(g_i)$ is displayed in black and is spaced apart from the other two curves by the standard deviation of $\hat{\pi}_i(g_i)$.*

and neither convex. Nevertheless, we experiment heuristically our algorithm on a close model given by its likelihood

$$p(y|\{\theta_j\}_{j=1}^4) = \prod_{i=1}^p \left[\frac{1}{4} \left(\frac{\lambda}{2\pi} \right)^2 \left\{ \sum_{j=1}^4 \exp\left(-(\lambda/2)(y_i - \theta_j)^2\right) \right\} \right]$$

for $y = (y_1, \dots, y_p) \in \mathbb{R}^p$ a vector of observations. The prior distributions are set following the recommendations of [33, Section 4.2.1] and [42]. For $j \in \{1, \dots, 4\}$, θ_j is drawn from a Gaussian distribution of mean $\xi = 1.35$ and inverse variance $\varsigma = 7.6 \times 10^{-3}$. λ is set equal to 0.03. The observations $y \in \mathbb{R}^{100}$ are 100 simulated data points from an equally weighted mixture of four Gaussian densities with means $(-3, 0, 3, 6)$ and standard deviations 0.55, taken from [25]. Define for $\theta = (\theta_1, \dots, \theta_4) \in \mathbb{R}^4$, $\ell : \mathbb{R}^4 \to \mathbb{R}$ by $\ell(\theta) = -\log(p(y|\theta)p(\theta))$. The `optim` function of R [41] gives a local minimum at $\theta^* \approx (1.76562^{\otimes 4})$. Define then the potential $U : \mathbb{R}^4 \to \mathbb{R}$ for $\theta \in \mathbb{R}^4$ by $U(\theta) = \ell(\theta + \theta^*) - \ell(\theta^*)$. Set $\epsilon = \mu = 0.1$, $m = \varsigma$ and $L = 1$. Similarly to the logistic regression, to decrease the running time of the algorithm, the recurrence for the variances $\{\sigma_i^2\}_{i=0}^{M-1}$ is defined by $\sigma_{i+1}^2 = \varsigma_s^5(\sigma_i^2)$ as long as the stopping condition (36) is not fulfilled. For each phase, the step size γ_i is set equal to $10^{-1}(\kappa_i \sigma_i^2 m_i)/(dL_i^2)$, the burn-in period N_i to 10^4 and the number of samples n_i to 10^5 where m_i, L_i, κ_i are defined in (13) and (14). For comparison purposes, we run the same algorithm using the Metropolis Adjusted Langevin Algorithm (MALA) instead of ULA to estimate $\hat{\pi}_i(g_i)$ at each phase. The step size γ_i is set equal to $(\kappa_i \sigma_i^2 m_i)/(dL_i^2)$ and the number of samples n_i to

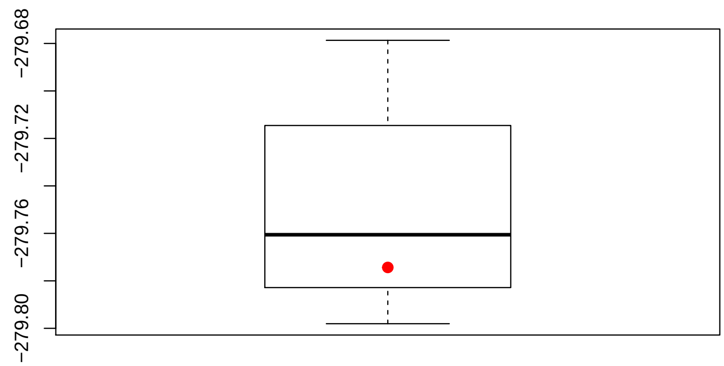

Fig 6. *Boxplot of the log evidence for the mixture of Gaussian distributions.*

10^6. The experiments are repeated 10 times. The boxplot is plotted in Figure 6 and the red point indicates the mean of our algorithm using MALA.

4. Mean squared error for locally Lipschitz functions

In this Section, we extend the results of [15, Section 3] to locally Lipschitz functions. This Section is of independent interest and only Propositions 17 and 20 are used in Section 5. Let $U : \mathbb{R}^d \to \mathbb{R}$ be a continuously differentiable function. Consider the target distribution π with density $x \mapsto \mathrm{e}^{-U(x)} / \int_{\mathbb{R}^d} \mathrm{e}^{-U(y)} \mathrm{d}y$ w.r.t. the Lebesgue measure. We deal with the problem of estimating $\int_{\mathbb{R}^d} f(x) \mathrm{d}\pi(x)$ for locally Lipschitz $f : \mathbb{R}^d \to \mathbb{R}$ by the ULA algorithm defined for $k \in \mathbb{N}$ by,

$$X_{k+1} = X_k - \gamma_{k+1} \nabla U(X_k) + \sqrt{2\gamma_{k+1}} Z_{k+1} , \tag{66}$$

where $(Z_k)_{k \geq 1}$ is an i.i.d. sequence of d-dimensional Gaussian vectors with zero mean, identity covariance and $(\gamma_k)_{k \geq 1}$ is a sequence of positive step sizes, which can either be held constant or be chosen to decrease to 0. For $n, p \in \mathbb{N}$, denote by

$$\Gamma_{n,p} \stackrel{\text{def}}{=} \sum_{k=n}^{p} \gamma_k , \qquad \Gamma_n = \Gamma_{1,n} , \tag{67}$$

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

$$R_\gamma(x, \mathrm{A}) = \int_{\mathrm{A}} (4\pi\gamma)^{-d/2} \exp\left(-(4\gamma)^{-1} \left\| y - x + \gamma \nabla U(x) \right\|^2 \right) \mathrm{d}y . \tag{68}$$

Define

$$Q_\gamma^{n,p} = R_{\gamma_n} \cdots R_{\gamma_p} \ , \qquad Q_\gamma^n = Q_\gamma^{1,n} \ , \tag{69}$$

with the convention that for $n, p \geq 0$, $n < p$, $Q_\gamma^{p,n}$ and $Q_\gamma^{0,0}$ are the identity operator.

For all initial distribution μ_0 on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $\mathbb{P}_{\mu_0}$ and $\mathbb{E}_{\mu_0}$ denote the probability and the expectation respectively associated with the sequence of Markov kernels (68) and the initial distribution μ_0 on the canonical space $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes \mathbb{N}})$ and $(X_k)_{k\in\mathbb{N}}$ denotes the canonical process. Let $f : \mathbb{R}^d \to \mathbb{R}$ and consider the following assumption,

L1. *1. There exists $L_f : \mathbb{R}^d \to [0, +\infty)$ a continuous function such that for all $x, y \in \mathbb{R}^d$, $|f(y) - f(x)| \leq \|y - x\| \max\{L_f(x), L_f(y)\}$.*

2. There exist $\varepsilon > 0$, $C_\pi > 0$ and continuous functions $C_Q, C_{Q,\epsilon} : \mathbb{R}^d \to [0, +\infty)$ such that for all $x \in \mathbb{R}^d$,

$$\pi(L_f^2) \leq C_\pi \ , \quad \sup_{p\geq n\geq 0} \delta_x Q_\gamma^{n,p}\left(L_f^2\right) \leq C_Q(x) \ , \tag{70}$$

$$\sup_{p\geq n\geq 0} \delta_x Q_\gamma^{n,p}\left(L_f^{2(1+\epsilon)}\right) \leq C_{Q,\epsilon}(x)$$

Under **L** 1, we study the approximation of $\int_{\mathbb{R}^d} f(y)\pi(\mathrm{d}y)$ by the weighted average estimator

$$\hat{\pi}_n^N(f) = \sum_{k=N+1}^{N+n} \omega_{k,n}^N f(X_k) \ , \quad \omega_{k,n}^N = \gamma_{k+1}\Gamma_{N+2,N+n+1}^{-1} \ , \tag{71}$$

where $N \geq 0$ is the length of the burn-in period and $n \geq 1$ is the number of samples. The Mean Squared Error (MSE) of $\hat{\pi}_n^N(f)$ is defined by:

$$\mathrm{MSE}_f(x, N, n) = \mathbb{E}_x\left[\left\{\hat{\pi}_n^N(f) - \pi(f)\right\}^2\right] \ , \tag{72}$$

and can be decomposed as,

$$\mathrm{MSE}_f(x, N, n) = \left\{\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right\}^2 + \mathrm{Var}_x\left[\hat{\pi}_n^N(f)\right] \ . \tag{73}$$

The analysis of $\mathrm{MSE}_f(x, N, n)$ is similar to [15, Section 3]. First, the squared bias in (73) is bounded. Denote by,

$$A_0 = 2L^2\kappa^{-1}d \ , \tag{74}$$

$$A_1 = 2dL^2 + dL^4(\kappa^{-1} + (m+L)^{-1})(m^{-1} + 6^{-1}(m+L)^{-1}) \ , \tag{75}$$

$$B_0 = d\left(2L^2 + \kappa^{-1}\{d\tilde{L}^2/3 + 4L^4/(3m)\}\right) \ , \tag{76}$$

$$B_1 = dL^4\left(\kappa^{-1} + \{6(m+L)\}^{-1} + m^{-1}\right) \ , \tag{77}$$

where κ is given by (14). Define then for $n \in \mathbb{N}^\star$,

$$u_n^{(1)}(\gamma) = \prod_{k=1}^{n} (1 - \kappa\gamma_k/2) , \tag{78}$$

$$u_n^{(2)}(\gamma) = \sum_{i=1}^{n} \left(A_0\gamma_i^2 + A_1\gamma_i^3\right) \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2) , \tag{79}$$

$$u_n^{(3)}(\gamma) = \sum_{i=1}^{n} \left(B_0\gamma_i^3 + B_1\gamma_i^4\right) \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2) . \tag{80}$$

Proposition 17. *Assume* ***H****1 and* ***H****2(m) for $m > 0$. Let $f : \mathbb{R}^d \to \mathbb{R}$ satisfying* ***L****1. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$. Let $x^\star$ be the unique minimizer of U. Let $(X_n)_{n\geq 0}$ be given by (66) and started at $x \in \mathbb{R}^d$. Then for all $N \geq 0$, $n \geq 1$:*

$$\left\{\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right\}^2 \leq \{C_\pi + C_Q(x)\} \times \sum_{k=N+1}^{N+n} \omega_{k,n}^N \left\{2(\|x - x^\star\|^2 + d/m)u_k^{(1)}(\gamma) + w_k(\gamma)\right\} , \tag{81}$$

where $u_n^{(1)}(\gamma)$ is given in (78) and $w_n(\gamma)$ is equal to $u_n^{(2)}(\gamma)$ defined by (79) and to $u_n^{(3)}(\gamma)$, defined by (80), if ***H****3 holds.*

Proof. For all $k \in \{N+1, \ldots, N+n\}$, let ξ_k be the optimal transference plan between $\delta_x Q_\gamma^k$ and π for W_2. By the Jensen and the Cauchy Schwarz inequalities, and **L**1, we have:

$$\begin{aligned}
\left(\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right)^2 &= \left(\sum_{k=N+1}^{N+n} \omega_{k,n}^N \int_{\mathbb{R}^d\times\mathbb{R}^d} \{f(z) - f(y)\}\xi_k(\mathrm{d}z, \mathrm{d}y)\right)^2 \\
&\leq \sum_{k=N+1}^{N+n} \omega_{k,n}^N \left(\int_{\mathbb{R}^d\times\mathbb{R}^d} \|z - y\| \max\{L_f(z), L_f(y)\}\, \xi_k(\mathrm{d}z, \mathrm{d}y)\right)^2 \\
&\leq \{C_\pi + C_Q(x)\} \sum_{k=N+1}^{N+n} \omega_{k,n}^N \int_{\mathbb{R}^d\times\mathbb{R}^d} \|z - y\|^2\, \xi_k(\mathrm{d}z, \mathrm{d}y) .
\end{aligned}$$

The proof follows from [15, Theorems 5 and 8]. □

To deal with the variance term in (73), we adapt the proof of [28, Theorem 2] to our setting, where f is only locally Lipschitz and the Markov chain (66) is inhomogeneous. It is based on the Gaussian Poincaré inequality [6, Theorem 3.20]. Let $Z = (Z_1, \ldots, Z_d)$ be a Gaussian vector with identity covariance matrix and $f : \mathbb{R}^d \to \mathbb{R}$ be a locally Lipschitz function. Recall that by Rademacher's Theorem [19, Theorem 3.2], a locally Lipschitz function is almost everywhere differentiable w.r.t. Lebesgue measure on $\mathbb{R}^d$. The Gaussian Poincaré inequality states that $\mathrm{Var}\left[f(Z)\right] \leq \mathbb{E}[\|\nabla f(Z)\|^2]$. Noticing that for all $x \in \mathbb{R}^d$, $R_\gamma(x, \cdot)$ defined in (68) is a Gaussian distribution with mean $x - \gamma \nabla U(x)$ and covariance matrix $2\gamma \,\mathrm{I}_d$, the Gaussian Poincaré inequality implies:

$$0 \leq \int R_\gamma(x, \mathrm{d}y) \left\{f(y) - R_\gamma f(x)\right\}^2 \leq 2\gamma \int R_\gamma(x, \mathrm{d}y) \left\|\nabla f(y)\right\|^2 \; . \tag{82}$$

First consider the following decomposition of $\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)]$ as the sum of martingale increments,

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

where $(\mathcal{G}_n)_{n\geq 0}$ is the natural filtration associated with the Markov chain $(X_n)_{n\geq 0}$. This implies that the variance may be decomposed as the following sum

$$\mathrm{Var}_x\left[\hat{\pi}_n^N(f)\right] = \sum_{k=N}^{N+n-1} \mathbb{E}_x\left[\left(\mathbb{E}_x^{\mathcal{G}_{k+1}}\left[\hat{\pi}_n^N(f)\right] - \mathbb{E}_x^{\mathcal{G}_k}\left[\hat{\pi}_n^N(f)\right]\right)^2\right] \\ + \mathbb{E}_x\left[\left(\mathbb{E}_x^{\mathcal{G}_N}\left[\hat{\pi}_n^N(f)\right] - \mathbb{E}_x[\hat{\pi}_n^N(f)]\right)^2\right] \; . \tag{83}$$

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

$$\Phi_{n,k}^N : x_k \mapsto \omega_{k,n}^N f(x_k) + R_{\gamma_{k+1}} \Phi_{n,k+1}^N(x_k) \; , \tag{84}$$

with the convention $\Phi_{n,N+n+1}^N = 0$. Denote finally

$$\Psi_n^N : x_N \mapsto R_{\gamma_{N+1}} \Phi_{n,N+1}^N(x_N) \; . \tag{85}$$

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

$$\Phi_{n,k+1}^N(X_{k+1}) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(X_k) = \mathbb{E}_x^{\mathcal{G}_{k+1}}\left[\hat{\pi}_n^N(f)\right] - \mathbb{E}_x^{\mathcal{G}_k}\left[\hat{\pi}_n^N(f)\right] \; , \tag{86}$$

and $\Psi_n^N(X_N) = \mathbb{E}_x^{\mathcal{G}_N}\left[\hat{\pi}_n^N(f)\right]$. With these notations, (83) may be equivalently expressed as

$$\mathrm{Var}_x\left[\hat{\pi}_n^N(f)\right] = \sum_{k=N}^{N+n-1} \mathbb{E}_x\left[R_{\gamma_{k+1}}\left\{\Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(X_k)\right\}^2(X_k)\right] \\ + \mathrm{Var}_x\left[\Psi_n^N(X_N)\right] \; . \tag{87}$$

Now for $k = N+n, \ldots, N+1$, we will use the Gaussian Poincaré inequality (82) to the sequence of function $\Phi_{n,k}^N$. It is required to prove that $\Phi_{n,k}^N$ is locally Lipschitz (see Lemma 18). For the variance of $\Psi_n^N(X_N)$, similar arguments apply using Lemma 19.

Lemma 18. *Assume **H**1, **H**2(m) for $m > 0$ and let $f : \mathbb{R}^d \to \mathbb{R}$ satisfying **L** 1. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Then for all $\ell \geq n \geq 0$, $Q_\gamma^{n,\ell} f$ is locally Lipschitz and differentiable for almost all $x \in \mathbb{R}^d$. Its gradient is bounded by,*

$$\left\|\nabla Q_\gamma^{n,\ell} f(x)\right\| \leq \prod_{k=n}^{\ell} (1-\kappa\gamma_k)^{1/2} (\delta_x Q_\gamma^{n,\ell} L_f^2)^{1/2} . \tag{88}$$

Proof. Let $\xi_{x,y}$ be the optimal transference plan between $\delta_x Q_\gamma^{n,\ell}$ and $\delta_y Q_\gamma^{n,\ell}$ for W_2. By Rademacher's Theorem [19, Theorem 3.2], $\nabla Q_\gamma^{n,\ell} f(x)$ exists for almost all $x \in \mathbb{R}^d$. For such x, using Cauchy-Schwarz's inequality and [15, Theorem 4], we have

$$\begin{aligned}
&\left\|\nabla Q_\gamma^{n,\ell} f(x)\right\| = \sup_{\|u\|\leq 1} \lim_{t\to 0} \left| \left(Q_\gamma^{n,\ell} f(x+tu) - Q_\gamma^{n,\ell} f(x)\right)/t \right| \\
&= \sup_{\|u\|\leq 1} \lim_{t\to 0} \left| t^{-1} \int_{\mathbb{R}^d\times\mathbb{R}^d} \{f(z_2) - f(z_1)\}\, \xi_{x,x+tu}(\mathrm{d}z_1, \mathrm{d}z_2) \right| \\
&\leq \sup_{\|u\|\leq 1} \liminf_{t\to 0} t^{-1} W_2(\delta_x Q_\gamma^{n,\ell}, \delta_{x+tu} Q_\gamma^{n,\ell}) \\
&\qquad\qquad \times \left\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} (L_f^2(z_1) \vee L_f^2(z_2)) \xi_{x,x+tu}(\mathrm{d}z_1, \mathrm{d}z_2) \right\}^{1/2} \\
&\leq \sup_{\|u\|\leq 1} \liminf_{t\to 0} \prod_{k=n}^{\ell} (1-\kappa\gamma_k)^{1/2} \left\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} (L_f^2(z_1) \vee L_f^2(z_2)) \xi_{x,x+tu}(\mathrm{d}z_1, \mathrm{d}z_2) \right\}^{1/2} .
\end{aligned}$$

It is then sufficient to prove that,

$$\lim_{y\to x} \int_{\mathbb{R}^d\times\mathbb{R}^d} L_f^2(z_1) \vee L_f^2(z_2) \xi_{x,y}(\mathrm{d}z_1, \mathrm{d}z_2) = \int_{\mathbb{R}^d} L_f^2(z_1) \delta_x Q_\gamma^{n,\ell}(\mathrm{d}z_1) .$$

Let $\varepsilon, \eta, \mathsf{R} > 0$ and $y \in \mathbb{R}^d$. Since $a \vee b - a = (b-a)_+$, we have

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} (L_f^2(z_2) - L_f^2(z_1))_+ \xi_{x,y}(\mathrm{d}z_1, \mathrm{d}z_2) = E_1(y) + E_2(y) + E_3(y)$$

where,

$$\begin{aligned}
E_1(y) &= \int_{\mathbb{R}^d\times\mathbb{R}^d} \left(L_f^2(z_2) - L_f^2(z_1)\right)_+ \mathbb{1}_{\{\|z_1\|+\|z_2\|\geq 2\mathsf{R}\}} \xi_{x,y}(\mathrm{d}z_1, \mathrm{d}z_2) , \\
E_2(y) &= \int_{\mathbb{R}^d\times\mathbb{R}^d} \left(L_f^2(z_2) - L_f^2(z_1)\right)_+ \mathbb{1}_{\{\|z_1\|+\|z_2\|\leq 2\mathsf{R}\}} \mathbb{1}_{\{\|z_1-z_2\|\leq \eta\}} \xi_{x,y}(\mathrm{d}z_1, \mathrm{d}z_2) , \\
E_3(y) &= \int_{\mathbb{R}^d\times\mathbb{R}^d} \left(L_f^2(z_2) - L_f^2(z_1)\right)_+ \mathbb{1}_{\{\|z_1\|+\|z_2\|\leq 2\mathsf{R}\}} \mathbb{1}_{\{\|z_1-z_2\|\geq \eta\}} \xi_{x,y}(\mathrm{d}z_1, \mathrm{d}z_2) .
\end{aligned}$$

Hölder's inequality gives for $p,q>1$, $1/p+1/q=1$,

$$E_1(y) \leq \left(\int_{\mathbb{R}^d} L_f^{2q}(z_2)\delta_y Q_\gamma^{n,\ell}(\mathrm{d}z_2)\right)^{1/q} \times \left(\int_{\mathbb{R}^d} \mathbb{1}_{\{\|z_1\|\geq \mathsf{R}\}}\delta_x Q_\gamma^{n,\ell}(\mathrm{d}z_1) + \int_{\mathbb{R}^d} \mathbb{1}_{\{\|z_2\|\geq \mathsf{R}\}}\delta_y Q_\gamma^{n,\ell}(\mathrm{d}z_2)\right)^{1/p} .$$

Under **L**1-2, the first term on the right hand side is dominated by a constant for q small enough, and the second term tends to 0 for R large enough, uniformly for y in a compact neighborhood of x by [15, Theorem 3] and

$$\int_{\mathbb{R}^d} \mathbb{1}_{\{\|z_2\|\geq \mathsf{R}\}}\delta_y Q_\gamma^{n,\ell}(\mathrm{d}z_2) \leq \mathsf{R}^{-2}\int_{\mathbb{R}^d} \|z_2\|^2\,\delta_y Q_\gamma^{n,\ell}(\mathrm{d}z_2) .$$

We can then choose R such that $E_1(y)\leq \varepsilon/3$. We consider now $E_2(y)$. L_f^2 is a continuous function, uniformly continuous on a compact set and we can then choose η such that $E_2(y)\leq\varepsilon/3$. We finally consider $E_3(y)$. By Markov's inequality and $\lim_{y\to x} W_2^2(\delta_x Q_\gamma^{n,\ell},\delta_y Q_\gamma^{n,\ell})=0$, there exists a compact neighborhood $\mathcal{V}(x)$ of x such that $y\in\mathcal{V}(x)$ implies $E_3(y)\leq\varepsilon/3$.

□

Lemma 19. *Assume* ***H**1 and **H**2(m) for $m>0$. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1\leq 2/(m+L)$ and $N\geq 0$. Let $f:\mathbb{R}^d\to\mathbb{R}$ be such that $Q_\gamma^{k+1,N}f$ is locally Lipschitz for $k\in\{1,\ldots,N\}$. Then for all $x\in\mathbb{R}^d$,*

$$\int_{\mathbb{R}^d} Q_\gamma^N(x,\mathrm{d}y)\left\{f(y)-Q_\gamma^N f(x)\right\}^2 \leq 2\sum_{k=1}^N \gamma_k \int_{\mathbb{R}^d} Q_\gamma^k(x,\mathrm{d}y)\left\|\nabla Q_\gamma^{k+1,N}f(y)\right\|^2 .$$

Proof. Using $\mathbb{E}_x^{\mathcal{G}_k}[f(X_N)] = Q_\gamma^{k+1,N}f(X_k)$, we get

$$\begin{aligned}\mathrm{Var}_x[f(X_N)] &= \sum_{k=1}^N \mathbb{E}_x\left[\mathbb{E}_x^{\mathcal{G}_{k-1}}\left[\left(\mathbb{E}_x^{\mathcal{G}_k}[f(X_N)]-\mathbb{E}_x^{\mathcal{G}_{k-1}}[f(X_N)]\right)^2\right]\right] \\ &= \sum_{k=1}^N \mathbb{E}_x\left[R_{\gamma_k}\left\{Q_\gamma^{k+1,N}f(\cdot)-R_{\gamma_k}Q_\gamma^{k+1,N}f(X_{k-1})\right\}^2(X_{k-1})\right] .\end{aligned}$$

Eq. (82) implies that

$$\mathrm{Var}_x[f(X_N)] \leq 2\sum_{k=1}^N \gamma_k\int_{\mathbb{R}^d} Q_\gamma^k(x,\mathrm{d}y)\left\|\nabla Q_\gamma^{k+1,N}f(y)\right\|^2 .$$

□

Proposition 20. *Assume* ***H**1 and **H**2(m) for $m>0$. Let $f:\mathbb{R}^d\to\mathbb{R}$ satisfying **L**1 and $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1\leq 2/(m+L)$. Then for all $N\geq 0$, $n\geq 1$, we get*

$$\mathrm{Var}_x\left[\hat{\pi}_n^N(f)\right] \leq \frac{8C_Q(x)}{\kappa^2\Gamma_{N+2,N+n+1}}\left\{1+\Gamma_{N+2,N+n+1}^{-1}\left(\kappa^{-1}+\frac{2}{m+L}\right)\right\} . \quad (89)$$

Proof. For $k \in \{N, \dots, N+n-1\}$ and for all $y, x \in \mathbb{R}^d$, we have

$$\left|\Phi_{n,k+1}^N(y) - \Phi_{n,k+1}^N(x)\right| = \left|\omega_{k+1,n}^N \{f(y) - f(x)\} + \sum_{i=k+2}^{N+n} \omega_{i,n}^N \left\{Q_\gamma^{k+2,i} f(y) - Q_\gamma^{k+2,i} f(x)\right\}\right| . \quad (90)$$

By Lemma 18, $\Phi_{n,k+1}^N$ is locally Lipschitz and for almost all $x \in \mathbb{R}^d$,

$$\left\|\nabla \Phi_{n,k+1}^N(x)\right\| \le \sum_{i=k+1}^{N+n} \omega_{i,n}^N \left\{\prod_{\ell=k+2}^{i} (1-\kappa\gamma_\ell)^{1/2}\right\} (\delta_x Q_\gamma^{k+2,i} L_f^2)^{1/2} .$$

For $k \in \{N, \dots, N+n-1\}$ and $x \in \mathbb{R}^d$, we have by (82) and the Cauchy-Schwarz inequality,

$$R_{\gamma_{k+1}} \left\{\Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(x)\right\}^2 (x) \le 2\gamma_{k+1} \Omega_{k,n}^N \left\{\sum_{i=k+1}^{N+n} \omega_{i,n}^N \prod_{\ell=k+2}^{i} (1-\kappa\gamma_\ell)^{1/2} (\delta_x Q_\gamma^{k+1,i} L_f^2)\right\} ,$$

where,

$$\Omega_{k,n}^N = \sum_{i=k+1}^{N+n} \omega_{i,n}^N \prod_{\ell=k+2}^{i} (1-\kappa\gamma_\ell)^{1/2} . \quad (91)$$

By **L**1-2, we get for $k \in \{N, \dots, N+n-1\}$

$$\mathbb{E}_x \left[R_{\gamma_{k+1}} \left\{\Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(X_k)\right\}^2 (X_k)\right]$$
$$\le 2\gamma_{k+1} \Omega_{k,n}^N \left\{\sum_{i=k+1}^{N+n} \omega_{i,n}^N \prod_{\ell=k+2}^{i} (1-\kappa\gamma_\ell)^{1/2} (\delta_x Q_\gamma^{i} L_f^2)\right\} \le 2\gamma_{k+1} C_Q(x) (\Omega_{k,n}^N)^2 .$$

Using $(1-t)^{1/2} \le (1-t/2)$ for $t \in [0,1]$, we have

$$\Omega_{k,n}^N \le (\kappa \Gamma_{N+2,N+n+1}/2)^{-1} . \quad (92)$$

Using this inequality, we get

$$\sum_{k=N}^{N+n-1} \mathbb{E}_x \left[R_{\gamma_{k+1}} \left\{\Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(X_k)\right\}^2 (X_k)\right] \le 8 C_Q(x) \Gamma_{N+1,N+n} / (\kappa \Gamma_{N+2,N+n+1})^2 . \quad (93)$$

We now bound $\mathrm{Var}_x\left[\Psi_n^N(X_N)\right]$. Since for all $x \in \mathbb{R}^d$, we have

$$\Psi_n^N(x) = \sum_{i=N+1}^{N+n} \omega_{i,n}^N Q_\gamma^{N+1,i} f(x) ,$$

by Lemma 18, $Q_\gamma^{k+1,N}\Psi_n^N$ is locally Lipschitz for $k \in \{1,\dots,N\}$ with for almost all $x \in \mathbb{R}^d$,

$$\left\|\nabla Q_\gamma^{k+1,N}\Psi_n^N(x)\right\| \le \sum_{i=N+1}^{N+n} \omega_{i,n}^N \prod_{\ell=k+1}^{i} (1-\kappa\gamma_\ell)^{1/2} \left(\delta_x Q_\gamma^{k+1,i} L_f^2\right)^{1/2} .$$

Isolating the term $\prod_{\ell=k+1}^N (1-\kappa\gamma_\ell)^{1/2}$ and since $(1-\kappa\gamma_{N+1})^{1/2} \le 1$, the Cauchy-Schwarz inequality implies

$$\left\|\nabla Q_\gamma^{k+1,N}\Psi_n^N(x)\right\|^2 \le \left\{\prod_{\ell=k+1}^{N}(1-\kappa\gamma_\ell)\right\} \times \Omega_{N,n}^N \sum_{i=N+1}^{N+n} \omega_{i,n}^N \prod_{\ell=N+1}^{i} (1-\kappa\gamma_\ell)^{1/2} \delta_x Q_\gamma^{k+1,i} L_f^2 .$$

Plugging this inequality in Lemma 19, using **L**1-2, $\sum_{k=1}^N \gamma_k \prod_{i=k+1}^N (1-\kappa\gamma_i) \le \kappa^{-1}$ and (92), we get

$$\operatorname{Var}_x\left[\Psi_n^N(X_N)\right] \le 2\kappa^{-1} C_Q(x)(\kappa/2)^{-2}\Gamma_{N+2,N+n+1}^{-2} . \tag{94}$$

Combining (93) and (94) in (83) concludes the proof.

□

5. Proofs

5.1. Proofs of propositions 2, 3, 4

We assume in this Section that **H**1 and **H**2(m) for some $m \ge 0$ hold. The proofs rely on the results given in Section 4, Propositions 17 and 20 which establish bounds on the mean squared error for locally Lipschitz functions. For $i \in \{0,\dots,M-1\}$, $\sigma_i^2 > 0$ and $\gamma_i > 0$, consider the Markov chain $(X_{i,n})_{n\ge 0}$ (8) and its associated Markov kernel R_i defined for all $\mathrm{A} \in \mathcal{B}(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$ by

$$R_i(x,\mathrm{A}) = \int_{\mathrm{A}} (4\pi\gamma_i)^{-d/2} \exp\left(-(4\gamma_i)^{-1}\left\|y - x + \gamma_i\nabla U_i(x)\right\|^2\right) \mathrm{d}y . \tag{95}$$

Under **H**1 and **H**2(m) for $m \ge 0$, [35, Theorems 2.1.12, 2.1.9] show the following useful inequalities for all $x, y \in \mathbb{R}^d$,

$$\langle \nabla U_i(y) - \nabla U_i(x), y - x\rangle \ge \frac{\kappa_i}{2}\|y-x\|^2 + \frac{1}{m_i + L_i}\|\nabla U_i(y) - \nabla U_i(x)\|^2 , \tag{96}$$

$$\langle \nabla U_i(y) - \nabla U_i(x), y - x\rangle \ge m_i \|y-x\|^2 , \tag{97}$$

where L_i, m_i are defined in (13) and κ_i in (14). We then check **L**1 for g_i, where $g_i : \mathbb{R}^d \to \mathbb{R}$ is defined in (6). Note that g_i is continuously differentiable and for $x \in \mathbb{R}^d$, $\nabla g_i(x) = 2a_i x \mathrm{e}^{a_i\|x\|^2}$. Define $L_{g_i} : \mathbb{R}^d \to \mathbb{R}_+$ for $x \in \mathbb{R}^d$ by,

$$L_{g_i}(x) = 2a_i \|x\| \mathrm{e}^{a_i\|x\|^2} \tag{98}$$

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

$$\begin{aligned} |g_i(y) - g_i(x)| = \left| \int_0^1 \langle \nabla g_i(ty + (1-t)x), y - x \rangle \, \mathrm{d}t \right| \\ \leq \|y - x\| \max(L_{g_i}(x), L_{g_i}(y)) \ , \end{aligned} \tag{99}$$

which implies that **L**1-1 holds for g_i. The following Lemmas 21 and 22 enable to check **L**1-2 for g_i.

Lemma 21. *Assume* ***H**1* *and* ***H**2(m)* *for $m \geq 0$. For all $\sigma_i^2 \in (0, +\infty)$, $n \in \mathbb{N}$, $\gamma_i \in (0, 2/(m_i + L_i)]$, $a_i \in \left[0, \kappa_i/8 \wedge (2\sigma_i^2)^{-1}\right)$ and $x \in \mathbb{R}^d$, we have:*

$$\sup_{n \in \mathbb{N}} R_i^n \left(L_{g_i}^2\right)(x) \leq 4a_i^2 g_i^2(x) C_{i,0} \left\{ \|x\|^2 + C_{i,1} \right\} \ ,$$

where L_{g_i} is defined in (98) and $C_{i,0}, C_{i,1}$ in (23).

Proof. In the proof, the subscript i is not specified for ease of notation. Let $\gamma \in (0, 2/(m+L)]$. Note that for all $\alpha \in \left[0, (4\gamma)^{-1}\right)$, we have

$$\begin{aligned} R_\gamma(\mathrm{e}^{\alpha\|\cdot\|^2})(x) &= \frac{\mathrm{e}^{-(4\gamma)^{-1}\|x-\gamma\nabla U(x)\|^2}}{(4\pi\gamma)^{d/2}} \int_{\mathbb{R}^d} \mathrm{e}^{(\alpha-(4\gamma)^{-1})\|y\|^2+(2\gamma)^{-1}\langle y, x-\gamma\nabla U(x)\rangle} \mathrm{d}y \\ &= \phi(x) \ , \end{aligned}$$

where $\phi(x) = (1-4\gamma\alpha)^{-d/2} \exp\{(\alpha/(1-4\alpha\gamma)) \|x - \gamma\nabla U(x)\|^2\}$. By the Leibniz integral rule and (96), we obtain:

$$\begin{aligned} &R_\gamma(\|\cdot\|^2 \mathrm{e}^{\alpha\|\cdot\|^2})(x) = \partial_\alpha R_\gamma(\mathrm{e}^{\alpha\|\cdot\|^2})(x) \\ &= (1-4\gamma\alpha)^{-d/2-1} \left\{ 2\gamma d + \frac{\|x - \gamma\nabla U(x)\|^2}{1-4\alpha\gamma} \right\} \exp\left(\frac{\alpha}{1-4\alpha\gamma} \|x - \gamma\nabla U(x)\|^2 \right) \\ &\leq (1-4\gamma\alpha)^{-d/2-1} \left\{ 2\gamma d + \frac{1-\kappa\gamma}{1-4\alpha\gamma} \|x\|^2 \right\} \exp\left(\frac{\alpha(1-\kappa\gamma)}{1-4\alpha\gamma} \|x\|^2 \right) \ . \end{aligned}$$

Let $a \in [0, \kappa/8)$. Since $a < (4\gamma)^{-1}$, by a straightforward induction we have

$$\begin{aligned}\delta_x R_\gamma^p(\|\cdot\|^2 \mathrm{e}^{2a\|\cdot\|^2}) &\leq (1-4\gamma\alpha_0)^{-d/2-1} \exp\left(\alpha_p \|x\|^2\right) \\ &\times \sum_{\ell=0}^{p-1} 2\gamma d \alpha_\ell \alpha_0^{-1} \left\{\prod_{k=1}^{\ell}(1-4\gamma\alpha_k)^{-d/2-1}\right\}\left\{\prod_{k=\ell+1}^{p-1}(1-4\gamma\alpha_k)^{-d/2}\right\} \\ &+ (1-4\gamma\alpha_0)^{-d/2-1}\left\{\prod_{k=1}^{p-1}(1-4\gamma\alpha_k)^{-d/2-1}\right\}\alpha_p\alpha_0^{-1}\|x\|^2 \exp\left(\alpha_p\|x\|^2\right) \\ &\leq \frac{1}{\alpha_0}\exp\left(\alpha_p\|x\|^2\right)\left\{\prod_{k=0}^{p-1}(1-4\gamma\alpha_k)^{-d/2-1}\right\}\left\{\alpha_p\|x\|^2 + 2d\gamma\sum_{\ell=0}^{p-1}\alpha_\ell\right\}, \quad (100)\end{aligned}$$

where $(\alpha_\ell)_{\ell\in\mathbb{N}}$ is the decreasing sequence defined for $\ell \geq 1$ by:

$$\alpha_0 = 2a, \quad \alpha_\ell = \alpha_{\ell-1}\frac{(1-\kappa\gamma)}{1-4\alpha_{\ell-1}\gamma}. \quad (101)$$

We now bound the right-hand-side of (100). First, by using the following inequality,

$$\log(1-4\gamma\alpha) = -4\alpha\int_0^\gamma (1-4\alpha t)^{-1}\mathrm{d}t \geq -4\alpha\gamma(1-4\alpha\gamma)^{-1},$$

we have:

$$\begin{aligned}\prod_{k=0}^{p-1}(1-4\gamma\alpha_k)^{-d/2-1} &= \exp\left(-\left(\frac{d}{2}+1\right)\sum_{k=0}^{p-1}\log(1-4\alpha_k\gamma)\right) \\ &\leq \exp\left(\left(\frac{d}{2}+1\right)\frac{4\gamma}{1-\kappa\gamma}\sum_{k=0}^{p-1}\alpha_k\frac{1-\kappa\gamma}{1-4\alpha_k\gamma}\right). \quad (102)\end{aligned}$$

Second, by a straightforward induction we get for all $\ell \geq 0$, $\alpha_\ell \leq 2a\{(1-\kappa\gamma)(1-8a\gamma)^{-1}\}^\ell$. Using (101) and this result implies:

$$\sum_{k=0}^{p-1}\alpha_k\frac{1-\kappa\gamma}{1-4\alpha_k\gamma} = \sum_{k=1}^{p}\alpha_k \leq 2a\frac{1-\kappa\gamma}{\kappa\gamma-8a\gamma}, \qquad \sum_{\ell=0}^{p-1}\alpha_\ell \leq 2a\frac{1-8a\gamma}{\kappa\gamma-8a\gamma}.$$

Combining these inequalities and (102) in (100) concludes the proof. □

Lemma 22. *Assume* **H**1 *and* **H**2(m) *for* $m \geq 0$. *For all* $\sigma_i^2 \in (0, +\infty)$ *and* $a_i \in \left[0, m_i/\{4(d+4)\} \wedge (2\sigma_i^2)^{-1}\right]$, *we have*

$$\pi(L_{g_i}^2) \leq 4a_i^2 C_{i,2},$$

where $C_{i,2}$ *is defined in* (23).

Proof. In the proof, the subscript i is not specified for ease of notations. Recall that the generator of the Langevin diffusion (7) associated to U is defined for any f in $\mathcal{C}^2(\mathbb{R}^d)$ by

$$\mathscr{A} f = -\langle \nabla U, \nabla f \rangle + \Delta f \ .$$

In particular, for $f(x) = \left\|x\right\|^2 \mathrm{e}^{2a\left\|x\right\|^2}$ and $x \in \mathbb{R}^d$, we have

$$\begin{aligned}
\nabla f(x) &= 2(1+2a\left\|x\right\|^2)x\mathrm{e}^{2a\left\|x\right\|^2} \ , \\
\Delta f(x) &= \mathrm{e}^{2a\left\|x\right\|^2}\left\{16a^2\left\|x\right\|^4 + 4a(d+4)\left\|x\right\|^2 + 2d\right\} \ .
\end{aligned}$$

Using (97) and $\nabla U(0) = 0$, we get

$$\mathscr{A}(\left\|\cdot\right\|^2 \mathrm{e}^{2a\left\|\cdot\right\|^2})(x) \leq \mathrm{e}^{2a\left\|x\right\|^2}\left\{2d + 2\left(2a(d+4) - m\right)\left\|x\right\|^2 + 4a\left(4a-m\right)\left\|x\right\|^4\right\} .$$

Using that $a \in [0, m/(4(d+4))]$, we have $2a(4a-m) \leq -(8/5)am$. Then an elementary study of $t \mapsto \mathrm{e}^{2at}\left\{2d + 4a\left(4a-m\right)t^2\right\}$ on $\mathbb{R}_+$ shows that:

$$\sup_{x \in \mathbb{R}^d} \mathrm{e}^{2a\left\|x\right\|^2}\left\{2d + 4a\left(4a-m\right)\left\|x\right\|^4\right\} \leq 4d \ .$$

Therefore we get using $2(2a(d+4)-m) \leq -m$,

$$\mathscr{A}(\left\|\cdot\right\|^2 \mathrm{e}^{2a\left\|\cdot\right\|^2})(x) \leq -m\left\|x\right\|^2 \mathrm{e}^{2a\left\|x\right\|^2} + 4d \ .$$

Finally applying [32, Theorem 4.3-(ii)] shows the result. □

Proofs of Propositions 2 and 3. Lemmas 21 and 22 and proposition 17 prove the result. □

Proof of Proposition 4. The proof follows from Lemma 21 and proposition 20. □

5.2. *Proof of Lemma 9*

The case $K = 0$ being straightforward, assume $K \in \mathbb{N}^\star$. Using Markov's inequality, we have

$$\mathbb{P}(\mathrm{A}^{\mathrm{c}}_{\mathcal{S},\epsilon}) \leq \frac{4}{\epsilon^2}\frac{\mathbb{E}\left[\left(\prod_{i=0}^{M-1}\hat{\pi}_i(g_i) - \prod_{i=0}^{M-1}\pi_i(g_i)\right)^2\right]}{\left(\prod_{i=0}^{M-1}\pi_i(g_i)\right)^2} \ . \tag{103}$$

Since $\hat{\pi}_i(g_i)$ for $i \in \{0, \ldots, M-1\}$ are independent, we get

$$\frac{\mathbb{E}\left[\left(\prod_{i=0}^{M-1}\hat{\pi}_i(g_i) - \prod_{i=0}^{M-1}\pi_i(g_i)\right)^2\right]}{\left(\prod_{i=0}^{M-1}\pi_i(g_i)\right)^2} = F_1^2(F_2-1) + (F_1-1)^2 \ , \tag{104}$$

where

$$F_1 = \prod_{i=0}^{M-1} \mathbb{E}\left[\hat{\pi}_i(g_i)\right]/\pi_i(g_i) \ , \quad F_2 = \prod_{i=0}^{M-1} \mathbb{E}\left[\{\hat{\pi}_i(g_i)\}^2\right]/\mathbb{E}^2\left[\hat{\pi}_i(g_i)\right] \ .$$

In addition, since $\{0,\ldots,M-2\} = \cup_{k=0}^{K-1}\mathcal{I}_k$, we can consider the following decomposition

$$F_1 = \prod_{k=0}^{K-1}\prod_{i\in\mathcal{I}_k}\left(1+\frac{\mathbb{E}\left[\hat{\pi}_i(g_i)\right]-\pi_i(g_i)}{\pi_i(g_i)}\right) \times\left(1+\frac{\mathbb{E}\left[\hat{\pi}_{M-1}(g_{M-1})\right]-\pi_{M-1}(g_{M-1})}{\pi_{M-1}(g_{M-1})}\right) \ ,$$

$$F_2 = \prod_{k=0}^{K-1}\prod_{i\in\mathcal{I}_k}\left(1+\frac{\mathrm{Var}\left[\hat{\pi}_i(g_i)\right]}{\mathbb{E}\left[\hat{\pi}_i(g_i)\right]^2}\right)\left(1+\frac{\mathrm{Var}\left[\hat{\pi}_{M-1}(g_{M-1})\right]}{\mathbb{E}\left[\hat{\pi}_{M-1}(g_{M-1})\right]^2}\right) \ .$$

We now bound F_1, F_2 separately. Using $1+t \leq \exp(t)$ for $t \in \mathbb{R}$ with $t = \eta/(K\left|\mathcal{I}_k\right|)$ and leaving the term $i = M-1$ out, we get by conditions i)-ii)

$$F_1 \leq (1+\eta)\exp(\eta) \ . \tag{105}$$

Since $\hat{\pi}_i(g_i) \geq 1$, we have $\mathrm{Var}\left[\hat{\pi}_i(g_i)\right]/\mathbb{E}\left[\hat{\pi}_i(g_i)\right]^2 \leq \eta^2/K\left|\mathcal{I}_k\right|$. Therefore using $1+t \leq \exp(t)$ for $t\in\mathbb{R}$ with $t = \eta^2/(K\left|\mathcal{I}_k\right|)$ leaving the term $i = M-1$ out, we obtain by conditions i)-ii)

$$F_2 \leq \left(1+\eta^2\right)\exp(\eta^2) \ . \tag{106}$$

By combining (103), (104), (105) and (106), we get:

$$(\epsilon^2/4)\mathbb{P}(\mathrm{A}_{\mathcal{S},\epsilon}^{\mathrm{c}}) \leq (1+\eta)^2\mathrm{e}^{2\eta}\left((1+\eta^2)\mathrm{e}^{\eta^2}-1\right) + \left((1+\eta)\mathrm{e}^{\eta}-1\right)^2 \ .$$

With $\eta \leq 1/8$ and $\mathrm{e}^t - 1 \leq t\mathrm{e}^t$ for $t \geq 0$, we have $(\epsilon^2/4)\mathbb{P}(\mathrm{A}_{\mathcal{S},\epsilon}^{\mathrm{c}}) \leq 9\eta^2$.

5.3. Proofs of Section 2.1

We preface the proofs by a technical lemma which gathers useful bounds and inequalities. We recall that in this Section the number of phases M is defined by (36)

$$M = \inf\left\{i \geq 1 : \sigma_{i-1}^2 \geq (2d+7)/m\right\} \ .$$

Lemma 23. *Assume* ***H****1 and* ***H****2(m) for $m > 0$. Let $\{\sigma_i^2\}_{i=0}^{M-1}$ defined by (34) for σ_0^2 given in (31) and M in (36).*

1. *$K \leq \left\lceil (1/\log(2))\log\{(2d+7)/(m\sigma_0^2)\}\right\rceil$ where K is defined in (39).*
2. *For all $k \in \{0,\ldots,K-1\}$ and $i \in \mathcal{I}_k$, $2^{k+1}\sigma_0^2 a_i\left|\mathcal{I}_k\right| \leq 1$, where a_i is defined in (37) and $\mathcal{I}_k$ in (38).*

3. *For all* $i \in \{0, \ldots, M-1\}$ *and* $\gamma_i \le 1/(m_i + L_i)$*, there exist* $\alpha_i \in [4, 14]$ *and* $\beta_i \in [1, 10]$ *such that* $C_{i,2} + C_{i,0}C_{i,1} = \alpha_i d m_i^{-1}$ *and* $C_{i,0}C_{i,1} = \beta_i d \kappa_i^{-1}$ *where* $C_{i,0}, C_{i,1}, C_{i,2}$ *and* κ_i *are given in* (23) *and* (14) *respectively.*
4. *For all* $i \in \{0, \ldots, M-1\}$, $0 < A_{i,1} \le 4dL_i^4 \kappa_i^{-1} m_i^{-1}$*, where* L_i, m_i *and* κ_i *are given in* (13) *and* (14) *respectively.*
5. *For all* $i \in \{0, \ldots, M-2\}$, $\kappa_i \sigma_i^2 \le 4d + 16$.
6. *For all* $i \in \{0, \ldots, M-1\}$, $\sqrt{m_i}/(\kappa_i \sigma_i) \le 1$
7. *For all* $i \in \{0, \ldots, M-1\}$,

$$\frac{m_i + L_i}{2m_i} \le \frac{m+L}{2m} \; , \quad \frac{L_i^2}{\kappa_i^3 \sigma_i^4 m_i} \le \left(\frac{m+L}{2m} \right)^3 \; .$$

8. *For all* $k \in \{0, \ldots, K-1\}$ *and* $i \in \mathcal{I}_k$,

$$\kappa_i^{-2} m_i^{-1/2} \sigma_i^{-2} \le \frac{(2^{k+1}\sigma_0^2)^{3/2}}{(1 + m 2^k \sigma_0^2)^{5/2}} \; , \quad \frac{L_i^2 m_i^{-1/2}}{\kappa_i^2 \sigma_i^2 m_i^{1/2}} \le \left(\frac{m+L}{2m} \right)^2 \frac{1}{1 + m2^k\sigma_0^2} \; .$$

Proof.

1. By (36) and (38),

$$K \le \inf \left\{ k \ge 0 : 2^k \sigma_0^2 \ge \frac{2d+7}{m} \right\} = \left\lceil \frac{1}{\log(2)} \log\left(\frac{2d+7}{m\sigma_0^2} \right) \right\rceil \; .$$

2. Let $k \in \{0, \ldots, K-1\}$. Denote by $i_0 = \inf \mathcal{I}_k$. By (37) and (38),

$$|\mathcal{I}_k| \, a_i = \sum_{i \in \mathcal{I}_k} a_i \le \frac{1}{2\sigma_{i_0}^2} \le \frac{1}{2^{k+1}\sigma_0^2} \; ,$$

   and the proof follows.
3. Let $k \in \{0, \ldots, K-1\}$ and $i \in \mathcal{I}_k$, Since $m_i \ge m + (2^{k+1}\sigma_0^2)^{-1}$, $a_i \le m_i/\{4(d+4)\}$. Therefore using in addition that $\gamma_i \le 1/(m_i + L_i)$, we have $\kappa_i - 8a_i \ge \kappa_i(d+2)/(d+4)$ and $1 - 8a_i\gamma_i \ge (d+3)/(d+4)$. The definition of $C_{i,0}, C_{i,1}, C_{i,2}$ (23) completes the proof.
4. The upper bound is a straightforward consequence of $(1+\kappa_i(m_i+L_i)^{-1})(1+6^{-1}m_i(m_i+L_i)^{-1}) \le 2$.
5. The bound follows using that $\sigma_{M-2}^2 \le (2d+7)/m$ by (36) and the sequence $\{\kappa_i\sigma_i^2\}_{i=0}^{M-2}$ is non-decreasing since

$$\kappa_i \sigma_i^2 = 2 \left\{ 1 + \frac{mL\sigma_i^2 - 1/\sigma_i^2}{m + L + 2/\sigma_i^2} \right\} \; ,$$

   and $\{\sigma_i^2\}_{i=0}^{M-2}$ is non-decreasing.
6. The proof is a direct consequence of the fact that the sequence $i \mapsto \sqrt{m_i}/(\kappa_i\sigma_i)$ is non-increasing since $m < L$, $\{\sigma_i^2\}_{i=0}^{M-2}$ is non-decreasing and

$$\frac{\sqrt{m_i}}{\kappa_i \sigma_i} = \frac{1}{2} \frac{1}{\sqrt{1 + m\sigma_i^2}} \left\{ 1 + \frac{1 + m\sigma_i^2}{1 + L\sigma_i^2} \right\} \; .$$

7. Using that $\{\sigma_i^2\}_{i=0}^{M-2}$ is non-decreasing and

$$\frac{m_i+L_i}{2m_i} = \frac{2+(m+L)\sigma_i^2}{2+2m\sigma_i^2} \;, \quad \frac{L_i^2}{\kappa_i^3\sigma_i^4 m_i} \leq \left(\frac{(m+L)\sigma_i^2+2}{(2m)\sigma_i^2+2}\right)^3 \;,$$

concludes the proof.

8. Let $k \in \{0,\ldots,K-1\}$ and $i \in \mathcal{I}_k$. Since $\kappa_i \geq m_i$ and $2^k\sigma_0^2 \leq \sigma_i^2 \leq 2^{k+1}\sigma_0^2$, we have

$$\kappa_i^{-2} m_i^{-1/2}\sigma_i^{-2} \leq \frac{\sigma_i^3}{(m_i\sigma_i^2)^{5/2}} \leq \frac{(2^{k+1}\sigma_0^2)^{3/2}}{(1+m2^k\sigma_0^2)^{5/2}} \;,$$

and

$$\frac{L_i^2}{\kappa_i^2} \leq \left(\frac{m+L}{2m}\right)^2 \;, \quad \frac{1}{m_i\sigma_i^2} \leq \frac{1}{1+m2^k\sigma_0^2} \;.$$

□

5.3.1. *Proof of Lemma 8*

Because U satisfies **H**1, **H**2(m) for $m \geq 0$ and $U(0)=0$, $\nabla U(0)=0$, we have:

$$\exp(-(L/2)\left\|x\right\|^2) \leq \exp(-U(x)) \leq \exp(-(m/2)\left\|x\right\|^2) \;,$$

which implies by integration that,

$$(2\pi\sigma_0^2)^{d/2}/(1+\sigma_0^2 L)^{d/2} \leq Z_0 \leq (2\pi\sigma_0^2)^{d/2}/(1+\sigma_0^2 m)^{d/2} \;,$$

where $Z_0 = \int_{\mathbb{R}^d} \mathrm{e}^{-U_0}$ and U_0 is defined in (2). The proof follows from the expression of σ_0^2 and the bound,

$$\left(\frac{1+L\sigma_0^2}{1+m\sigma_0^2}\right)^{d/2} \leq \exp\left(\frac{d}{2}\sigma_0^2(L-m)\right) \;.$$

5.3.2. *Proof of Lemma 10*

Let $k \in \{0,\ldots,K-1\}$ and $i \in \mathcal{I}_k$. Assume that $\gamma_i \leq (m_i+L_i)^{-1}$. By Proposition 2, Proposition 4, Lemma 23-2 and $\sigma_i^2 \leq 2^{k+1}\sigma_0^2$, to check condition-i) of Lemma 9, it is then sufficient for γ_i, n_i, N_i to satisfy,

$$\frac{4d}{n_i m_i \kappa_i\gamma_i}\exp\left(-N_i\frac{\kappa_i\gamma_i}{2}\right) + 2\kappa_i^{-1}\left(A_{i,0}\gamma_i + A_{i,1}\gamma_i^2\right) \leq \frac{\eta^2}{4K^2}\frac{\sigma_i^4}{C_{i,2}+C_{i,0}C_{i,1}} \;, \tag{107}$$

$$\frac{32 a_i C_{i,0}C_{i,1}}{\kappa_i^2 n_i\gamma_i}\left(1+\frac{2}{\kappa_i n_i\gamma_i}\right) \leq \frac{\sigma_i^2\eta^2}{K} \;. \tag{108}$$

By (24), Lemma 23-3 and Lemma 23-4, there exist $\alpha_i \in [4, 14]$ and $\beta_i \in [1, 10]$ such that these two inequalities hold if γ_i, n_i, N_i satisfy

$$2L_i^2\kappa_i^{-1}d\gamma_i + 4dL_i^4\kappa_i^{-1}m_i^{-1}\gamma_i^2 \le \frac{\eta^2}{16K^2}\frac{\kappa_i m_i \sigma_i^4}{\alpha_i d} , \tag{109}$$

$$\frac{1}{n_i}\left(1 + \frac{2}{\kappa_i n_i \gamma_i}\right) \le \frac{\eta^2\sigma_i^2\kappa_i^3\gamma_i}{32K a_i \beta_i d} , \tag{110}$$

$$N_i \ge \bar{N}_i = -2(\kappa_i\gamma_i)^{-1}\log\left(\frac{\eta^2\sigma_i^4 m_i^2 n_i \kappa_i \gamma_i}{32K^2\alpha_i d^2}\right) . \tag{111}$$

These inequalities are shown to be true successively for γ_i, n_i and N_i chosen as in the statement of the Lemma. Denote by $\bar{\gamma}_i$ and $\bar{n}_i^{-1}$ the positive roots associated to (109) and (110) seen as equalities and given by

$$\bar{\gamma}_i = 4^{-1}L_i^{-2}m_i\left(-1 + \sqrt{1 + \frac{\eta^2\kappa_i^2\sigma_i^4}{4\alpha_i K^2 d^2}}\right) , \tag{112}$$

$$\bar{n}_i^{-1} = 4^{-1}\kappa_i\gamma_i\left(-1 + \sqrt{1 + \frac{\eta^2\kappa_i^2\sigma_i^2}{4K a_i \beta_i d}}\right) . \tag{113}$$

Note that for (109) and (110) to hold, it suffices that $\gamma_i \le \bar{\gamma}_i$ and $n_i \ge \bar{n}_i$. We now lower bound $\bar{\gamma}_i$ and upper bound $\bar{n}_i$.

Using that $\sqrt{1+t} \ge 1 + 2^{-1}t(1+t)^{-1/2}$ for $t = \eta^2\kappa_i^2\sigma_i^4/(4\alpha_i K^2 d^2)$ and $(\eta^2\kappa_i^2\sigma_i^4)/(4\alpha_i K^2 d^2) \le 25$ by $\alpha_i \ge 4$ and Lemma 23-5, concludes that if (40) holds then $\gamma_i \le \bar{\gamma}_i$. The fact that $\gamma_i \le (m_i + L_i)^{-1}$ can be checked by simple algebra.

First, by (37) and the definition of $\mathcal{I}_k$, $a_i \le m_i/\{4(d+4)\}$,

$$\bar{n}_i^{-1} \ge 4^{-1}\kappa_i\gamma_i\left(-1 + \sqrt{1 + \frac{\eta^2\kappa_i^2\sigma_i^2(d+4)}{K m_i \beta_i d}}\right) .$$

Then using that $\sqrt{1+t} \ge 1 + 2^{-1}t(1+t)^{-1/2}$ for $t = \eta^2\kappa_i^2\sigma_i^2(d+4)/(K m_i \beta_i d)$ and $\beta_i \ge 1$ concludes that if (41) holds then $n_i \ge \bar{n}_i$. Finally, we have by (41), $(n_i\kappa_i\gamma_i)^{-1} \le \eta^2\kappa_i\sigma_i/(196\sqrt{m_i}K)$, which gives with $\kappa_i \ge m_i$,

$$\begin{aligned}\bar{N}_i &\le 2(\kappa_i\gamma_i)^{-1}\log\left\{\frac{64\alpha_i}{196}Kd^2(1 + m\sigma_i^2)^{-3/2}\right\} \\ &\le 2(\kappa_i\gamma_i)^{-1}\log\left(5Kd^2\right) ,\end{aligned}$$

which concludes that (42) implies (111).

The same reasoning applies to check condition-ii) of Lemma 9. The details are gathered in the supplementary material [8, Appendix A.1].

5.3.3. Proof of Theorems 5 and 6 and corollary 7

For $i \in \{0, \dots, M-1\}$, set γ_i, n_i, N_i such that (40), (41), (42), (43), (44) and (45) are equalities. By (18), we consider the following decomposition for the cost $= A + B$ where $A = \sum_{i=0}^{M-2}\{N_i + n_i\}$ and $B = n_{M-1} + N_{M-1}$. We bound A and B separately.

First Lemma 23-6 implies that for all $i \in \{0, \dots, M-2\}$, $n_i \leq (196K)/(\eta^2 \kappa_i \gamma_i)$ and therefore using Lemma 23-7

$$A \leq \left(\frac{196K}{\eta^2} + 2\log(5Kd^2)\right) \frac{2285K^2d^2}{\eta^2} \left(\frac{m+L}{2m}\right)^3 (M-1) . \tag{114}$$

We now give a bound on $M-1$. Define

$$K_{\text{int}} = \sup\left\{k \geq 1 : m2^k\sigma_0^2 \leq 1\right\} \wedge K \leq \left\lfloor -\frac{\log(m\sigma_0^2)}{\log(2)} \right\rfloor . \tag{115}$$

By Lemma 23-2 and (37), we have

$$\frac{M-1}{4(d+4)} = \sum_{k=0}^{K-1} \frac{|\mathcal{I}_k|}{4(d+4)} \leq K_{\text{int}} + 2 . \tag{116}$$

Note that $K_{\text{int}} \leq K \leq C$ by (115) and Lemma 23-1. Combining (114), Lemma 23-7 and (116), we get

$$\sum_{i=0}^{M-2} N_i + n_i \leq \left(\frac{98K}{\eta^2} + \log(5Kd^2)\right) \frac{4570K^2d^2}{\eta^2} \left(\frac{m+L}{2m}\right)^3 4(d+4)(C+2) . \tag{117}$$

Regarding the term $i = M-1$, we have

$$n_{M-1} + N_{M-1} \leq \left(\frac{19}{\eta^2} + 1\right) \frac{40}{\eta^2} \frac{m+L}{2m} \frac{L}{m} . \tag{118}$$

Replacing η by $(\epsilon\sqrt{\mu})/8$ and combining (117) and (118) gives (28).

Assume **H**3. We now prove Theorem 6 and use Lemma 11 instead of Lemma 10. For $i \in \{0, \dots, M-1\}$, set γ_i, n_i, N_i such that (46), (41), (42), (47), (44) and (45) are equalities. By (18), we have the decomposition cost $= A + B$ where $A = \sum_{i=0}^{M-2}\{N_i + n_i\}$ and $B = n_{M-1} + N_{M-1}$. Lemma 23-6 implies that for all $i \in \{0, \dots, M-2\}$, $n_i \leq (196K)/(\eta^2\kappa_i\gamma_i)$, and using that for $a, b \geq 0$, $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, we have

$$A \leq \left(\frac{196K}{\eta^2} + 2\log(5Kd^2)\right) \sqrt{\frac{7}{3}} \frac{8Kd}{\eta} \sum_{i=0}^{M-2} \frac{d^{1/2}\tilde{L} + \sqrt{10}L_i^2 m_i^{-1/2}}{\kappa_i^2 \sigma_i^2 m_i^{1/2}} .$$

Then, by Lemma 23-2 and Lemma 23-8, and splitting the sum in two parts

$k \le K_{\mathrm{int}}$ and $k > K_{\mathrm{int}}$,

$$\begin{aligned}\sum_{i=0}^{M-2} \frac{1}{\kappa_i^2 \sigma_i^2 m_i^{1/2}} &\le \sum_{k=0}^{K-1} \sum_{i \in \mathcal{I}_k} \frac{(2^{k+1}\sigma_0^2)^{3/2}}{(1+m2^k\sigma_0^2)^{5/2}} \\ &\le 4(d+4)\frac{2^{3/2}}{m^{3/2}} \sum_{k=0}^{K-1} \frac{(m2^k\sigma_0^2)^{3/2}}{(1+m2^k\sigma_0^2)^{7/2}} \\ &\le 4(d+4)\frac{2^{3/2}}{m^{3/2}} \left(K_{\mathrm{int}} + \sum_{k=K_{\mathrm{int}}+1}^{K-1} (m2^k\sigma_0^2)^{-2} \right) \\ &\le 4(d+4)\frac{2^{3/2}}{m^{3/2}} \left(K_{\mathrm{int}} + \frac{4}{3} \right) .\end{aligned}$$

We have similarly by Lemma 23-2 and Lemma 23-8,

$$\sum_{i=0}^{M-2} \frac{L_i^2 m_i^{-1/2}}{\kappa_i^2 \sigma_i^2 m_i^{1/2}} \le 4(d+4) \left(\frac{m+L}{2m} \right)^2 (K_{\mathrm{int}} + 2) .$$

Combining these inequalities with

$$n_{M-1} + N_{M-1} \le \left(\frac{19}{\eta^2} + 1 \right) \sqrt{\frac{7}{3}} \frac{4}{\eta} \left\{ \frac{d^{1/2}\tilde{L}}{m^{3/2}} + \sqrt{10} \left(\frac{m+L}{2m} \right)^2 \right\} , \tag{119}$$

and replacing η by $(\epsilon\sqrt{\mu})/8$ establish (30).

Proof of Corollary 7. Let $N = \lceil 4\log(\tilde{\mu}^{-1}) \rceil$ and $(\hat{Z}_i)_{i\in\{1,\ldots,2N+1\}}$ be $2N+1$ independent outputs of the algorithms of Theorems 5 and 6 with $\mu = 1/4$, sorted by increasing order. Denote by $\hat{Z} = \hat{Z}_{N+1}$ the median of $(\hat{Z}_i)_{i\in\{1,\ldots,2N+1\}}$. In addition, define the independent Bernoulli random variables $(W_i)_{i\in\{1,\ldots,2N+1\}}$ by

$$W_i = \mathbb{1}_{\mathsf{A}_i} , \text{ where } \mathsf{A}_i = \left\{ \left| \hat{Z}_i/Z - 1 \right| \ge \epsilon \right\} .$$

Since $\hat{Z}$ is the median of $(\hat{Z}_i)_{i\in\{1,\ldots,2N+1\}}$, we have

$$\mathbb{P}\left(\left| \hat{Z}/Z - 1 \right| > \epsilon \right) \le \mathbb{P}\left(\sum_{i=1}^{2N+1} W_i \ge N+1 \right) .$$

In addition since $\mathbb{P}(W_i = 1) \le 1/4$, we have by [40, Corollary 5.2]

$$\mathbb{P}\left(\sum_{i=1}^{2N+1} W_i \ge N+1 \right) \le \mathbb{P}\left(\sum_{i=1}^{2N+1} \tilde{W}_i \ge N+1 \right) ,$$

where $(\tilde{W}_i)_{i\in\{1,\ldots,2N+1\}}$ are i.i.d. Bernoulli random variables with parameter $1/4$. Then by Hoeffding's inequality [6, Theorem 2.8] and using for all $t \ge 1$,

$8(t/2+3/4)^2/\{t(2t+1)\} \geq 1$, we get

$$\mathbb{P}\left(\sum_{i=1}^{2N+1} \tilde{W}_i \geq N+1\right) \leq \mathbb{P}\left(\sum_{i=1}^{2N+1} \tilde{W}_i - (1/4)(2N+1) \geq N/2+3/4\right)$$
$$\leq \exp\left(\frac{-2(N/2+3/4)^2}{2N+1}\right)$$
$$\leq \exp(-N/4) \;,$$

which concludes the proof. □

Acknowledgements

This work was supported by the Ecole Polytechnique Data Science Initiative. The authors thank the anonymous referees for useful suggestions which improved the manuscript.

Appendix A: Additional proofs of Section 2.1

A.1. Proof of Lemma 10

In this Section, the proof for the case $i = M-1$ of Lemma 10 is dealt with. Note that $a_{M-1} = (2\sigma_{M-1}^2)^{-1}$. By Propositions 2 and 4, to check condition-ii) of Lemma 9, it is then sufficient for $\gamma_{M-1}, n_{M-1}, N_{M-1}$ to satisfy,

$$\frac{4d}{n_{M-1} m_{M-1} \kappa_{M-1} \gamma_{M-1}} \exp\left(-N_{M-1}\frac{\kappa_{M-1}\gamma_{M-1}}{2}\right)$$
$$+ 2\kappa_{M-1}^{-1}\left(A_{M-1,0}\gamma_{M-1} + A_{M-1,1}\gamma_{M-1}^2\right) \leq \frac{\eta^2\sigma_{M-1}^4}{C_{M-1,2} + C_{M-1,0}C_{M-1,1}} \;, \tag{120}$$
$$\frac{8C_{M-1,0}C_{M-1,1}}{\kappa_{M-1}^2 n_{M-1}\gamma_{M-1}}\left(1 + \frac{2}{\kappa_{M-1}n_{M-1}\gamma_{M-1}}\right) \leq \sigma_{M-1}^4\eta^2 \;. \tag{121}$$

Then (120) and (121) are satisfied if,

$$2L_{M-1}^2\kappa_{M-1}^{-1} d\gamma_{M-1} + 4dL_{M-1}^4\kappa_{M-1}^{-1}m_{M-1}^{-1}\gamma_{M-1}^2 \leq \frac{\eta^2}{4}\frac{\kappa_{M-1}m_{M-1}\sigma_{M-1}^4}{\alpha_{M-1}d} \;, \tag{122}$$
$$\frac{1}{n_{M-1}}\left(1 + \frac{2}{\kappa_{M-1}n_{M-1}\gamma_{M-1}}\right) \leq \frac{\eta^2\sigma_{M-1}^4\kappa_{M-1}^3\gamma_{M-1}}{8\beta_{M-1}d} \;, \tag{123}$$
$$-2(\kappa_{M-1}\gamma_{M-1})^{-1}\log\left(\frac{\eta^2\sigma_{M-1}^4 m_{M-1}^2 n_{M-1}\kappa_{M-1}\gamma_{M-1}}{8\alpha_{M-1}d^2}\right) = \bar{N}_{M-1} \leq N_{M-1} \;. \tag{124}$$

Denote by $\bar{\gamma}_{M-1}$ and $\bar{n}_{M-1}^{-1}$ the positive roots associated to (122) and (123) seen as equalities. We have:

$$\bar{\gamma}_{M-1} = 4^{-1} L_{M-1}^{-2} m_{M-1} \left(-1 + \sqrt{1 + \frac{\eta^2 \kappa_{M-1}^2 \sigma_{M-1}^4}{\alpha_{M-1} d^2}} \right) , \tag{125}$$

$$\bar{n}_{M-1}^{-1} = 4^{-1} \kappa_{M-1} \gamma_{M-1} \left(-1 + \sqrt{1 + \frac{\eta^2 \kappa_{M-1}^2 \sigma_{M-1}^4}{\beta_{M-1} d}} \right) . \tag{126}$$

Note that for (122) and (123) to hold, it suffices that $\gamma_{M-1} \leq \bar{\gamma}_{M-1}$ and $n_{M-1} \geq \bar{n}_{M-1}$. We now lower bound $\bar{\gamma}_{M-1}$ and upper bound $\bar{n}_{M-1}$.

Using that $t \geq 0$, $\sqrt{1+t} \geq 1+2^{-1}t(1+t)^{-1/2}$ for $t = (\eta^2 \kappa_{M-1}^2 \sigma_{M-1}^4)/(\alpha_{M-1} d^2)$ and $\kappa_{M-1}\sigma_{M-1}^2 d^{-1} \geq 2$, $\alpha_{M-1} \geq 4$ concludes that if (43) holds then $\gamma_{M-1} \leq \bar{\gamma}_{M-1}$. The fact that $\gamma_{M-1} \leq (m_{M-1} + L_{M-1})^{-1}$ can be checked by simple algebra.

Then using that $\sqrt{1+t} \geq 1+2^{-1}t(1+t)^{-1/2}$ for $t = (\eta^2 \kappa_{M-1}^2 \sigma_{M-1}^4)/(\beta_{M-1} d)$ and $\kappa_{M-1}\sigma_{M-1}^2 \geq 10$, $\beta_{M-1} \geq 1$ concludes that if (44) holds then $n_{M-1} \geq \bar{n}_{M-1}$. Finally, by (44), we get

$$\bar{N}_{M-1} \leq (\kappa_{M-1}\gamma_{M-1})^{-1} \log(7/3) ,$$

which concludes that (45) implies (124).

A.2. Proof of Lemma 11

Let $k \in \{0, \ldots, K-1\}$ and $i \in \mathcal{I}_k$. Assume that $\gamma_i \leq (m_i + L_i)^{-1}$. The proof of Lemma 10 only needs to be slightly adapted. More precisely, Proposition 3 is applied instead of Proposition 2. By (26) and (27), we have

$$B_{i,0} \leq 3^{-1} d \kappa_i^{-1} (d\tilde{L}^2 + 10 L_i^4 m_i^{-1}) , \quad B_{i,1} \leq (25/12) d L_i^4 m_i^{-1} . \tag{127}$$

It is sufficient for γ_i, n_i, N_i to satisfy (107) and (108) with $A_{i,0}\gamma_i + A_{i,1}\gamma_i^2$ replaced by $B_{i,0}\gamma_i^2 + B_{i,1}\gamma_i^3$. The counterpart of (109) is then

$$\frac{1}{3\kappa_i} \left(d\tilde{L}^2 + 10 L_i^4 m_i^{-1} \right) \gamma_i^2 + \frac{25}{12} L_i^4 m_i^{-1} \gamma_i^3 \leq \frac{\eta^2 \kappa_i m_i \sigma_i^4}{16 K^2 \alpha_i d^2} . \tag{128}$$

Since $\gamma_i \leq 1/(m_i + L_i)$ and $\kappa_i \leq L_i$, we have

$$(3\kappa_i)^{-1} \left(d\tilde{L}^2 + 10 L_i^4 m_i^{-1} \right) \geq (25/12) L_i^4 m_i^{-1} \gamma_i ,$$

which establishes that if (46) holds, then (128) is satisfied. $\gamma_i \leq (m_i + L_i)^{-1}$ can be checked by simple algebra. For $i = M-1$, the conclusion follows from $m_{M-1}\sigma_{M-1}^2 d^{-1} \geq 2$ because $\sigma_{M-1}^2 \geq (2d+7)/m$.

Appendix B: Additional proofs of Section 2.2

First, we state a technical lemma that gathers useful bounds. We recall that M is defined in this Section by (56),

$$M = \inf\left\{i \geq 1 : \sigma_{i-1}^2 \geq D^2\right\} .$$

Lemma 24. *Assume **H**1 and **H**2(m) for $m \geq 0$. Let $\{\sigma_i^2\}_{i=0}^{M-1}$ defined by (54) for σ_0^2 given in (31) and M in (56).*

1. *$K \leq \left\lceil (1/\log(2)) \log\left(\sigma_0^{-2}\rho^{-2}d^2(\tau+1)^2\right)\right\rceil$ where K is defined in (39).*
2. *For $k \in \{0,\ldots,K-1\}$ and $i \in \mathcal{I}_k$, $2^{k+1}\sigma_0^2 a_i \left|\mathcal{I}_k\right| \leq 1$ where a_i is defined in (37) (with $m = 0$) and $\mathcal{I}_k$ in (38). As a consequence, $|\mathcal{I}_k| \leq 4(d+4)$.*
3. *For $i \in \{0,\ldots,M-1\}$, $\kappa_i\sigma_i^2 \in [1,2]$.*
4. *$\sigma_{M-1}^2 \in \left[D^2, (10/9)D^2\right]$.*
5. *For all $i \in \{0,\ldots,M-1\}$ and $\gamma_i \leq 1/(m_i+L_i)$, there exist $\alpha_i \in [4,14]$ and $\beta_i \in [1,10]$ such that $C_{i,2}+C_{i,0}C_{i,1} = \alpha_i d m_i^{-1}$ and $C_{i,0}C_{i,1} = \beta_i d\kappa_i^{-1}$ where $C_{i,0}, C_{i,1}, C_{i,2}$ and κ_i are given in (23) and (14) respectively.*
6. *For all $i \in \{0,\ldots,M-1\}$, $0 < A_{i,1} \leq 4dL_i^4\kappa_i^{-1}m_i^{-1}$, where L_i, m_i and κ_i are given in (13) and (14) respectively.*

Proof. The proofs of 1,2,5,6 are identical to the ones of Lemma 23.

3. $\kappa_i\sigma_i^2 = (2L_i)/(m_i+L_i)$.
4. By definition of M, $\sigma_{M-2}^2 \leq D^2$ and $a_{M-2} \leq \sigma_{M-2}^{-2}/\{4(d+4)\}$. By (6), we get:
$$\sigma_{M-1}^{-2} = \sigma_{M-2}^{-2} - 2a_{M-2} \geq \sigma_{M-2}^{-2}\left(1 - \frac{1}{2(d+4)}\right)$$
that is $\sigma_{M-1}^2 \leq (10/9)\sigma_{M-2}^2 \leq (10/9)D^2$.

□

B.1. Proof of Lemma 15

Let $k \in \{0,\ldots,K-1\}$ and $i \in \mathcal{I}_k$. Assume that $\gamma_i \leq (m_i+L_i)^{-1}$. The proof follows the same lines as the one in Section 5.3.2. By Lemma 24-5 and Lemma 24-6, to check condition-i) of Lemma 9, it suffices that $\gamma_i \leq \bar{\gamma}_i$, $n_i \geq \bar{n}_i$ and N_i satisfies (111), where $\bar{\gamma}_i$ is defined in (112) and $\bar{n}_i$ in (113).

Using that $\sqrt{1+t} \geq 1 + 2^{-1}t(1+t)^{-1/2}$ for $t = (\eta^2\kappa_i^2\sigma_i^4)/(4\alpha_i K^2 d^2)$ and by Lemma 24-3, concludes that if (57) holds then $\gamma_i \leq \bar{\gamma}_i$. $\gamma_i \leq (m_i+L_i)^{-1}$ can be checked by simple algebra.

By (37) (with $m = 0$) and the definition of $\mathcal{I}_k$, $a_i \leq \sigma_i^{-2}/\{4(d+4)\}$,

$$\bar{n}_i^{-1} \geq 4^{-1}\kappa_i\gamma_i\left(-1+\sqrt{1+\frac{\eta^2\kappa_i^2\sigma_i^4(d+4)}{K\beta_i d}}\right) .$$

Using that $\sqrt{1+t} \geq 1 + 2^{-1}t(1+t)^{-1/2}$ for $t = (\eta^2\kappa_i^2\sigma_i^4)/(4\alpha_i K^2 d^2)$ and by Lemma 24-3, concludes that if (58) holds then $n_i \geq \bar{n}_i$. Finally, by (58), if (59) holds, (111) is satisfied.

The case $i = M-1$ is different because $\bar{g}_{M-1}$ is Lipschitz. Assume $\gamma_{M-1} \leq (m_{M-1} + L_{M-1})^{-1}$. [15, section 2.1] entails that condition-ii) of Lemma 9 is satisfied if

$$\|\bar{g}_{M-1}\|^2_{\mathrm{Lip}} \left\{ \frac{4d}{n_{M-1}m_{M-1}\kappa_{M-1}\gamma_{M-1}} \exp\left(-N_{M-1}\frac{\kappa_{M-1}\gamma_{M-1}}{2}\right) \right.$$
$$\left. + 2\kappa_{M-1}^{-1}\left(A_{M-1,0}\gamma_{M-1} + A_{M-1,1}\gamma_{M-1}^2\right) \right\} \leq \eta^2 \;, \quad (129)$$
$$\frac{8\|\bar{g}_{M-1}\|^2_{\mathrm{Lip}}}{\kappa_{M-1}^2 n_{M-1}\gamma_{M-1}} \left\{1 + \frac{2}{n_{M-1}\kappa_{M-1}\gamma_{M-1}}\right\} \leq \eta^2 \;.$$

Using $\|\bar{g}_{M-1}\|^2_{\mathrm{Lip}} \leq (\sigma_{M-1}^{-2}\mathrm{e})$ and (24), Lemma 24-6 for $i = M-1$, it is sufficient for $\gamma_{M-1}, n_{M-1}, n_{M-1}$ to satisfy

$$2L_{M-1}^2\kappa_{M-1}^{-1}d\gamma_{M-1} + 4dL_{M-1}^4\kappa_{M-1}^{-1}m_{M-1}^{-1}\gamma_{M-1}^2 \leq (4\mathrm{e})^{-1}\kappa_{M-1}\eta^2\sigma_{M-1}^2 \;, \quad (130)$$

$$n_{M-1}^{-1}\left(1 + 2(\kappa_{M-1}\gamma_{M-1}n_{M-1})^{-1}\right) \leq \frac{\eta^2\kappa_{M-1}^2\sigma_{Mc-1}^2\gamma_{M-1}}{8\mathrm{e}} \;, \quad (131)$$

$$-2\frac{\log\left((8\mathrm{e}d)^{-1}n_{M-1}\kappa_{M-1}\gamma_{M-1}\eta^2\right)}{\kappa_{M-1}\gamma_{M-1}} = \bar{N}_{M-1} \leq N_{M-1} \;. \quad (132)$$

Denote by $\bar{\gamma}_{M-1}, \bar{n}_{M-1}^{-1}$ the roots of (130), (131) seen as equalities. We have

$$\bar{\gamma}_{M-1} = 4^{-1}L_{M-1}^{-2}\sigma_{Mc-1}^{-2}\left\{-1 + \sqrt{1 + \frac{\eta^2\kappa_{M-1}^2\sigma_{M-1}^4}{\mathrm{e}d}}\right\} \;, \quad (133)$$

$$\bar{n}_{M-1}^{-1} = 4^{-1}\kappa_{M-1}\gamma_{M-1}\left\{-1 + \sqrt{1 + \mathrm{e}^{-1}\eta^2\kappa_{M-1}\sigma_{M-1}^2}\right\} \;. \quad (134)$$

Using that $\sqrt{1+t} \geq 1 + 2^{-1}t(1+t)^{-1/2}$ for $t = (\eta^2\kappa_{M-1}^2\sigma_{M-1}^4)/(\mathrm{e}d)$ and by Lemma 24-3, concludes that if (60) holds then $\gamma_{M-1} \leq \bar{\gamma}_{M-1}$. $\gamma_{M-1} \leq (m_{M-1} + L_{M-1})^{-1}$ can be checked by simple algebra.

Using that $\sqrt{1+t} \geq 1 + 2^{-1}t(1+t)^{-1/2}$ for $t = \mathrm{e}^{-1}\eta^2\kappa_{M-1}\sigma_{M-1}^2$ and by Lemma 24-3, concludes that if (61) holds then $n_{M-1} \geq \bar{n}_{M-1}$.

Finally by (61), if (62) holds, (132) is satisfied.

B.2. Proof of Lemma 16

The proof is identical to the one of Lemma 11. For $k \in \{0, \ldots, K-1\}$ and $i \in \mathcal{I}_k$, it is sufficient for γ_i to satisfy (63) by (46) and Lemma 24-3.

Regarding the case $i = M-1$, assuming that $\gamma_{M-1} \leq (m_{M-1} + L_{M-1})^{-1}$, it is sufficient for $\gamma_{M-1}, n_{M-1}, N_{M-1}$ to satisfy (129) with $A_{M-1,0}\gamma_{M-1} + A_{M-1,1}\gamma_{M-1}^2$ replaced by $B_{M-1,0}\gamma_{M-1}^2 + B_{M-1,1}\gamma_{M-1}^3$. The counterpart of (130) is then,

$$\frac{1}{3\kappa_{M-1}}\left(d\tilde{L}^2 + 10L_{M-1}^4 m_{M-1}^{-1}\right)\gamma_{M-1}^2 + \frac{25}{12}\frac{L_{M-1}^4}{m_{M-1}}\gamma_{M-1}^3 \leq \frac{\eta^2\kappa_{M-1}\sigma_{M-1}^2}{4\mathrm{e}d} .$$

This concludes the proof with the same argument as in Appendix A.2.

B.3. Proof of Theorems 12 and 13 and corollary 14

For $i \in \{0, \ldots, M-1\}$, set γ_i, n_i, N_i such that (57), (58), (59), (60), (61) and (62) are equalities. By (18), we have

$$\mathrm{cost} = \left(\frac{453K}{\eta^2} + 2\log(Kd^2)\right)\frac{462K^2d^2}{\eta^2}\sum_{i=0}^{M-2}\kappa_i^{-1}L_i^2\sigma_i^2 + n_{M-1} + N_{M-1} .$$

Note that for $i \in \{0, \ldots, M-2\}$,

$$\kappa_i^{-1}L_i^2\sigma_i^2 = 1 + (3/2)L\sigma_i^2 + (L^2/2)\sigma_i^4 .$$

By Lemma 24-2, for $k \in \{0, \ldots, K-1\}$, $|\mathcal{I}_k| \leq 4(d+4)$ and for $i \in \mathcal{I}_k$, $\sigma_i^2 \leq 2^{k+1}\sigma_0^2$. We then have

$$\begin{aligned}\sum_{i=0}^{M-2}\frac{L_i^2\sigma_i^2}{\kappa_i} &\leq 4(d+4)\sum_{k=0}^{K-1}\left\{1 + \frac{3L}{2}2^{k+1}\sigma_0^2 + \frac{L^2}{2}(2^{k+1}\sigma_0^2)^2\right\} \\ &\leq 4(d+4)\left\{K + 3L(2^K\sigma_0^2) + \frac{2L^2}{3}(2^K\sigma_0^2)^2\right\} .\end{aligned}$$

By (56) and the definition of K, (39), $2^K\sigma_0^2 \leq 2D^2$. The expressions of $\gamma_{M-1}, n_{M-1}, N_{M-1}$ give

$$n_{M-1} + N_{M-1} = \left(\frac{29}{\eta^2} + 2\log(d)\right)\frac{26d}{\eta^2}\kappa_{M-1}^{-2}L_{M-1}^2 ,$$

with $\kappa_{M-1}^{-2}L_{M-1}^2 = (1 + 2^{-1}L\sigma_{M-1}^2)^2$. By Lemma 24-4, we then have

$$n_{M-1} + N_{M-1} \leq \left(\frac{29}{\eta^2} + 2\log(d)\right)\frac{26d}{\eta^2}\left(1 + \frac{5L}{9}D^2\right)^2 , \tag{135}$$

and (51) is established.

Assume **H** 3. We now prove Theorem 13 and use Lemma 11 instead of Lemma 10. For $i \in \{0, \ldots, M-1\}$, set γ_i, n_i, N_i such that (63), (58), (59), (64), (61) and (62) are equalities. By (18) and using that for $a, b \geq 0$, $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, we have

$$\mathrm{cost} \leq \left(\frac{453K}{\eta^2} + 2\log(Kd^2)\right)\sqrt{\frac{7}{3}}\frac{8Kd}{\eta}\sum_{i=0}^{M-2}\frac{\sigma_i}{\kappa_i}\left(d^{1/2}\tilde{L} + \sqrt{10}L_i^2\sigma_i\right) + n_{M-1} + N_{M-1} .$$

For $k \in \{0, \ldots, K-1\}$ and $i \in \mathcal{I}_k$, note that

$$\kappa_i^{-1}\sigma_i \leq \sigma_i^3 \ , \quad \kappa_i^{-1}\sigma_i^2 L_i^2 = 1 + \frac{3L}{2}\sigma_i^2 + \frac{L^2}{2}\sigma_i^4 \ .$$

Using for $k \in \{0, \ldots, K-1\}$, $|\mathcal{I}_k| \leq 4(d+4)$ by Lemma 24-2 and for $i \in \mathcal{I}_k$, $\sigma_i^2 \leq 2^{k+1}\sigma_0^2$, we get

$$\begin{aligned}
\sum_{i=0}^{M-2} \frac{\sigma_i}{\kappa_i}\left(d^{1/2}\tilde{L} + \sqrt{10}L_i^2\sigma_i\right) &\leq 4(d+4)\sum_{k=0}^{K-1}\left\{d^{1/2}\tilde{L}(2^{k+1}\sigma_0^2)^{3/2}\right. \\
&\quad \left. + \sqrt{10}\left(1 + \frac{3L}{2}(2^{k+1}\sigma_0^2) + \frac{L^2}{2}(2^{k+1}\sigma_0^2)^2\right)\right\} \\
&\leq 4(d+4)\left\{5d^{1/2}\tilde{L}D^3 + \sqrt{10}\left(K + 6LD^2 + \frac{8L^2}{3}D^4\right)\right\} \ ,
\end{aligned}$$

with $2^K\sigma_0^2 \leq 2D^2$. The expressions of $\gamma_{M-1}, n_{M-1}, N_{M-1}$ give

$$n_{M-1} + N_{M-1} \leq \left(2\log(d) + \frac{29}{\eta^2}\right)\sqrt{\frac{8\mathrm{e}}{3}}\frac{\sqrt{d}}{\eta}\frac{d^{1/2}\tilde{L} + \sqrt{10}L_{M-1}^2\sigma_{M-1}}{\kappa_{M-1}^2\sigma_{M-1}} \ .$$

By Lemma 24-4, $\sigma_{M-1}^2 \in \left[D^2, (10/9)D^2\right]$. We get then

$$\kappa_{M-1}^{-2}\sigma_{M-1}^{-1} = \frac{\left(1 + (L/2)\sigma_{M-1}^2\right)^2}{L^2\sigma_{M-1}} \leq \frac{1}{DL^2}\left(1 + \frac{5L}{9}D^2\right)^2 \ ,$$

and,

$$\kappa_{M-1}^{-2}L_{M-1}^2 = \left(1 + \frac{L}{2}\sigma_{M-1}^2\right)^2 \leq \left(1 + \frac{5L}{9}D^2\right)^2 \ ,$$

which gives,

$$n_{M-1} + N_{M-1} \leq \left(2\log(d) + \frac{29}{\eta^2}\right)\sqrt{\frac{8\mathrm{e}}{3}}\frac{\sqrt{d}}{\eta}\left(1 + \frac{5L}{9}D^2\right)^2\left(\frac{d^{1/2}\tilde{L}}{DL^2} + \sqrt{10}\right) \ . \tag{136}$$

(53) is established. The proof of Corollary 14 is the same as the one of Corollary 7.

References

[1] C. Andrieu, J. Ridgway, and N. Whiteley. Sampling normalizing constants in high dimensions using inhomogeneous diffusions. *ArXiv e-prints*, Dec. 2016.

[2] D. Ardia, N. Baştürk, L. Hoogerheide, and H. K. Van Dijk. A comparative study of Monte Carlo methods for efficient evaluation of marginal likelihood. *Computational Statistics & Data Analysis*, 56(11):3398–3414, 2012.

[3] R. Balian. *From microphysics to macrophysics: methods and applications of statistical physics*, volume 1. Springer Science & Business Media, 2007.

[4] G. Behrens, N. Friel, and M. Hurn. Tuning tempered transitions. *Statistics and Computing*, 22(1):65–78, 2012. . URL http://dx.doi.org/10.1007/s11222-010-9206-z.

[5] A. Beskos, D. O. Crisan, A. Jasra, and N. Whiteley. Error bounds and normalising constants for sequential Monte Carlo samplers in high dimensions. *Adv. in Appl. Probab.*, 46(1):279–306, 03 2014. . URL http://dx.doi.org/10.1239/aap/1396360114.

[6] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: a nonasymptotic theory of independence*. Oxford university press, 2013.

[7] S. Brazitikos, A. Giannopoulos, P. Valettas, and B.-H. Vritsiou. *Geometry of isotropic convex bodies*, volume 196. American Mathematical Society Providence, 2014.

[8] N. Brosse, A. Durmus, and E. Moulines. Supplement to "Normalizing constants of log-concave densities". working paper or preprint, 2017.

[9] M. Chen, Q. Shao, and J. Ibrahim. *Monte Carlo methods in Bayesian computation*. Springer, New York, 2000.

[10] B. Cousins and S. Vempala. Bypassing KLS: Gaussian cooling and an $\mathrm{O}^*(n^3)$ volume algorithm. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pages 539–548. ACM, 2015.

[11] A. S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2016.

[12] P. Del Moral. *Feynman-Kac formulae*. Probability and its Applications (New York). Springer-Verlag, New York, 2004. ISBN 0-387-20268-4. URL https://doi.org/10.1007/978-1-4684-9393-1. Genealogical and interacting particle systems with applications.

[13] P. Del Moral, A. Jasra, K. Law, and Y. Zhou. Multilevel sequential Monte Carlo samplers for normalizing constants. *ArXiv e-prints*, Mar. 2016.

[14] A. Durmus and E. Moulines. Non-asymptotic convergence analysis for the unadjusted Langevin algorithm. July 2015.

[15] A. Durmus and E. Moulines. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. May 2016.

[16] R. Dutta, J. K. Ghosh, et al. Bayes model selection with path sampling: factor models and other examples. *Statistical Science*, 28(1):95–115, 2013.

[17] M. Dyer and A. Frieze. Computing the volume of convex bodies: a case where randomness provably helps. *Probabilistic combinatorics and its applications*, 44:123–170, 1991.

[18] D. L. Ermak. A computer simulation of charged particles in solution. i. technique and equilibrium properties. *The Journal of Chemical Physics*, 62 (10):4189–4196, 1975.

[19] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. CRC press, 2015.

[20] N. Friel and J. Wyse. Estimating the evidence–a review. *Statistica Neerlandica*, 66(3):288–308, 2012.

[21] N. Friel, M. Hurn, and J. Wyse. Improving power posterior estimation of statistical evidence. *Statistics and Computing*, 24(5):709–723, 2014. ISSN 1573-1375. . URL http://dx.doi.org/10.1007/s11222-013-9397-1.

[22] A. Gelman and X.-L. Meng. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical science*, pages 163–185, 1998.

[23] M. Huber. Approximation algorithms for the normalizing constant of Gibbs distributions. *Ann. Appl. Probab.*, 25(2):974–985, 04 2015. . URL http://dx.doi.org/10.1214/14-AAP1015.

[24] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Physical Review E*, 56(5):5018, 1997.

[25] A. Jasra, C. C. Holmes, and D. A. Stephens. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statist. Sci.*, 20(1):50–67, 02 2005. . URL https://doi.org/10.1214/088342305000000016.

[26] A. Jasra, K. Kamatani, P. P. Osei, and Y. Zhou. Multilevel particle filters: normalizing constant estimation. *Statistics and Computing*, pages 1–14, 2016. ISSN 1573-1375. . URL http://dx.doi.org/10.1007/s11222-016-9715-5.

[27] M. R. Jerrum, L. G. Valiant, and V. V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169 – 188, 1986. ISSN 0304-3975. . URL http://www.sciencedirect.com/science/article/pii/030439758690174X.

[28] A. Joulin and Y. Ollivier. Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.*, 38(6):2418–2442, 11 2010. . URL http://dx.doi.org/10.1214/10-AOP541.

[29] K. H. Knuth, M. Habeck, N. K. Malakar, A. M. Mubeen, and B. Placek. Bayesian evidence and model selection. *Digital Signal Processing*, 47:50 – 67, 2015. ISSN 1051-2004. . URL http://www.sciencedirect.com/science/article/pii/S1051200415001980. Special Issue in Honour of William J. (Bill) Fitzgerald.

[30] T. Lelièvre, G. Stoltz, and M. Rousset. *Free energy computations: A mathematical perspective*. World Scientific, 2010.

[31] J.-M. Marin and C. P. Robert. Importance sampling methods for Bayesian discrimination between embedded models. *arXiv preprint arXiv:0910.2325*, 2009.

[32] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes iii: Foster-Lyapunov criteria for continuous-time processes. *Advances in Applied Probability*, pages 518–548, 1993.

[33] P. D. Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 68(3):411–436, 2006. ISSN 13697412, 14679868. URL http://www.jstor.org/stable/3879283.

[34] R. M. Neal. Annealed importance sampling. *Statistics and Computing*, 11 (2):125–139, 2001. ISSN 1573-1375. . URL http://dx.doi.org/10.1023/

A:1008923215028.

[35] Y. Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

[36] W. Niemiro and P. Pokarowski. Fixed precision MCMC estimation by median of products of averages. *J. Appl. Probab.*, 46(2):309–329, 06 2009. . URL http://dx.doi.org/10.1239/jap/1245676089.

[37] C. J. Oates, T. Papamarkou, and M. Girolami. The controlled thermodynamic integral for Bayesian model evidence evaluation. *Journal of the American Statistical Association*, 111(514):634–645, 2016. . URL http://dx.doi.org/10.1080/01621459.2015.1021006.

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

[39] M. Pereyra. Maximum-a-posteriori estimation with Bayesian confidence regions. *arXiv preprint arXiv:1602.08590*, 2016.

[40] F. Proschan and J. Sethuraman. Stochastic comparisons of order statistics from heterogeneous populations, with applications in reliability. *Journal of Multivariate Analysis*, 6(4):608 – 616, 1976. ISSN 0047-259X. . URL http://www.sciencedirect.com/science/article/pii/0047259X76900087.

[41] R Core Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL https://www.R-project.org/.

[42] S. Richardson and P. J. Green. On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(4):731–792, 1997. ISSN 1467-9868. . URL http://dx.doi.org/10.1111/1467-9868.00095.

[43] C. Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, 2007.

[44] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996. ISSN 1350-7265. . URL http://dx.doi.org/10.2307/3318418.

[45] J. P. Valleau and D. N. Card. Monte Carlo estimation of the free energy by multistage sampling. *The Journal of Chemical Physics*, 57(12):5457–5462, 1972. . URL http://dx.doi.org/10.1063/1.1678245.

[46] C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[47] E. J. Williams and E. Williams. *Regression analysis*, volume 14. wiley New York, 1959.

[48] J. Wyse. Estimating the statistical evidence - a review, 2011. URL https://sites.google.com/site/jsnwyse/code.

[49] Y. Zhou, A. M. Johansen, and J. A. Aston. Towards automatic model comparison: an adaptive sequential Monte Carlo approach. *Journal of Computational and Graphical Statistics*, (just-accepted), 2015.