



**HAL**  
open science

# Efficient stochastic optimisation by unadjusted Langevin Monte Carlo. Application to maximum marginal likelihood and empirical Bayesian estimation

Valentin de Bortoli, Alain Durmus, Marcelo Alejandro Pereyra, Ana Fernandez Vidal

## ► To cite this version:

Valentin de Bortoli, Alain Durmus, Marcelo Alejandro Pereyra, Ana Fernandez Vidal. Efficient stochastic optimisation by unadjusted Langevin Monte Carlo. Application to maximum marginal likelihood and empirical Bayesian estimation. *Statistics and Computing*, 2020, 31 (3), pp.29. 10.1007/s11222-020-09986-y . hal-01978999v1

**HAL Id: hal-01978999**

**<https://hal.science/hal-01978999v1>**

Submitted on 12 Jan 2019 (v1), last revised 2 Dec 2020 (v2)

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

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

# Stochastic Optimization with Unadjusted Kernel: the SOUK Algorithm

Valentin De Bortoli <sup>\*1</sup>, Alain Durmus <sup>† 1</sup>, Marcelo Pereyra <sup>‡ 2</sup>, and Ana Fernandez Vidal <sup>§2</sup>

<sup>1</sup>CMLA - École normale supérieure Paris-Saclay, CNRS, Université Paris-Saclay, 94235 Cachan, France.

<sup>2</sup>School of Mathematical and Computer Sciences, Heriot Watt University & Maxwell Institute for Mathematical Sciences.

December 18, 2018

## Abstract

In this paper we propose a stochastic approximation algorithm to minimize functions for which the gradient writes as the expectation of some integrable features. Considering the case where computing feature moments or sampling from the underlying probability distribution is not feasible we mix a Markov chain dynamic with the stochastic gradient descent dynamic, thus combining optimization tools with statistics controls. This approach is motivated with examples from Empirical Bayesian statistics. We assess the convergence as well as rates of convergence for various objective functions.

## 1 Introduction

Computational Bayesian statistics often rely on complex probabilistic modeling which involves latent variables and therefore the underlying a posteriori distribution of the parameters is expressed as an untractable parameter-dependent integral. For example, some characteristics of the prior distribution such as the mean or the variance are frequently unknown and are treated as hyperparameters which leads to consider hierarchical Bayesian models [42]. A first approach is to consider the joint a posteriori probability distribution and perform inference for the parameters of interest by marginalizing along the hyperparameters. This framework is referred to as a fully Bayesian setting. However, solving this problem can be computationally prohibitive since it requires to sample from the joint probability distribution. For instance this can be realized with a Gibbs sampling strategy but the convergence of the associated Markov chain can be very slow in some situations.

---

\*Email: valentin.debortoli@cmla.ens-cachan.fr

†Email: alain.durmus@cmla.ens-cachan.fr

‡Email: m.pereyra@hw.ac.uk

§Email: af69@hw.ac.uk

A second approach, which we consider in this paper, referred to as the Empirical Bayes (EB) setting [7, 8, 40], consists in first finding a single-point estimate for the hyperparameters, similarly to the Maximum A Posteriori (MAP) estimator, based on the conditional probability distribution of the hyperparameters given the observations. Second, this estimate is plugged in the a posteriori distribution of the parameters given the observations and the hyperparameters in order to make inference. This methodology is theoretically founded: [37, 28, 45, 11] show that fully Bayesian and EB approaches lead to similar inference conclusions when the sample size is large enough under additional appropriate conditions. However, EB inference remains a challenging problem since it relies on the optimization of a high-dimensional parameter-dependent integral. When the parameter space is small, numerical integration techniques can be considered, like the recently introduced nested Laplace approximations [35, 46], however they might introduce some bias which can be hard to quantify. If the dimension is large, the optimization can be achieved using a Robbins-Monro scheme [41, 29, 20] which requires to be able to sample from the probability distribution of the parameters given the observations and the hyperparameters. This former step is usually tackled through Monte Carlo or Monte Carlo Markov Chain (MCMC) algorithms. EB inference has found many applications in Bayesian inference problem [21, 6], large-scale inverse problem [50], biostatistics [47, 26, 17].

In contrast to previous work, in this contribution we propose and analyze the use of inexact MCMC algorithms to get approximate samples from the a posteriori distribution of the parameters given the observations and the hyperparameters. While this method introduces some bias in the estimation of the stochastic gradient used in the Robbins-Monro procedure, we show that it can be controlled explicitly and does not affect the convergence of the hyperparameter recursion. The use of inexact MCMC algorithms is justified because while existing results can be applied to stochastic approximation based on exact MCMC schemes it requires strong conditions on the associated Markov kernel (see e.g. [2, H6]). The basic condition of geometric ergodicity can be hard to check, for example for the Metropolis Adjusted Langevin Algorithm [43] or Hamiltonian Monte Carlo [13]. On the other hand, inexact MCMC algorithms such that the Unadjusted Langevin-based Monte Carlo methods are easier to analyze and can exhibit stronger convergence properties [5, 12, 9, 16]. In addition, they give in practice similar inference results than their exact counterparts at the cost of a speed and accuracy trade-off. As a result, there have been increasingly popular in the computational statistics and machine learning communities [49, 33, 18, 22, 36, 24].

The paper is organized as follows. In Section 2 we motivate and introduce our new methodology based on a Robbins-Monro procedure using inexact MCMC algorithms for Empirical Bayes estimation. In Section 3 we summarize our main theoretical contributions and check that they hold for Langevin-based methods. In particular, we show that under appropriate conditions the complexity of the proposed methodology is linear in the dimension. Rigorous theoretical treatment of the convergence of the algorithm is given in the appendix.

## Notations and convention

Denote by  $\mathcal{B}(\mathbb{R}^d)$  the Borel  $\sigma$ -field of  $\mathbb{R}^d$ ,  $\mathbb{F}(\mathbb{R}^d)$  the set of all Borel measurable functions on  $\mathbb{R}^d$  and for  $f \in \mathbb{F}(\mathbb{R}^d)$ ,  $\|f\|_\infty = \sup_{x \in \mathbb{R}^d} |f(x)|$ . For  $\mu$  a probability measure on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  and  $f \in \mathbb{F}(\mathbb{R}^d)$  a  $\mu$ -integrable function, denote by  $\mu(f)$  the integral of  $f$  w.r.t.  $\mu$ . For  $f \in \mathbb{F}(\mathbb{R}^d)$ , the  $V$ -norm of  $f$  is given by  $\|f\|_V = \sup_{x \in \mathbb{R}^d} |f(x)|/V(x)$ . Let  $\xi$  be a finite signed measure on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . The

$V$ -total variation distance of  $\xi$  is defined as

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

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

Let  $U$  an open set of  $\mathbb{R}^d$ . We denote by  $C^k(U, \mathbb{R}^p)$ , respectively  $C_c^k(U, \mathbb{R}^p)$ , the set of  $\mathbb{R}^p$ -valued  $k$ -differentiable functions, respectively the set of compactly supported  $\mathbb{R}^p$ -valued  $k$ -differentiable functions. Let  $f \in C^1(U, \mathbb{R})$ , we denote by  $\nabla_x f$ , the gradient of  $f$ .  $f$  is said to be  $m$ -convex with  $m \geq 0$  if for all  $x, y \in \mathbb{R}^d$

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

We recall that if  $f$  is twice differentiable at point  $a$ , its Laplacian is given by  $\Delta_x f(a) = \sum_{i=1}^d \frac{\partial^2 f}{\partial x_i^2}(a)$ . Let  $F \in C^1(U \times V, \mathbb{R})$  we denote by  $\nabla_x$ , the gradient of  $F$  along the first coordinate and  $\nabla_y$  the gradient of  $F$  along the second coordinate. For any  $A \subset \mathbb{R}^d$ , we denote by  $\partial A$  the boundary of  $A$ . Let  $A \in \mathcal{B}(\mathbb{R}^d)$ , we define  $\text{Vol}(A) = \text{Leb}(\mathbb{1}_A)$  where  $\text{Leb}$  is the Lebesgue measure over  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . Let  $B \subset \mathbb{R}^d$ , we introduce its convex hull  $\text{Conv}(B)$ , the intersection of all the convex sets containing  $B$ . Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space. We denote by  $L^2(\Omega, \mathcal{F}) = \{X : X \text{ is a random variable on } \Omega \text{ such } \mathbb{E}[X^2] < +\infty\}$ . Denote by  $\mu \ll \nu$  if  $\mu$  is absolutely continuous w.r.t.  $\nu$  and  $d\mu/d\nu$  an associated density. Let  $\mu, \nu$  be two probability measures on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . Define the Kullback-Leibler divergence of  $\mu$  from  $\nu$  by

$$\text{KL}(\mu|\nu) = \begin{cases} \int_{\mathbb{R}^d} \frac{d\mu}{d\nu}(x) \log \left( \frac{d\mu}{d\nu}(x) \right) d\nu(x) , & \text{if } \mu \ll \nu \\ +\infty & \text{otherwise .} \end{cases}$$

## 2 EB inference based on inexact MCMC methods, the Stochastic Optimization with Unadjusted Kernel (SOUK)

### 2.1 EB inference

Consider the hierarchical model based on the observations  $y$  in the subspace  $Y \subset \mathbb{R}^{d_y}$  specified by the conditional density with respect to the Lebesgue measure  $(y, x, \theta) \mapsto p(y|x, \theta)$ , and prior distributions  $(x, \theta) \mapsto p(x|\theta)$  and  $\theta \mapsto p(\theta)$  where  $x \in \mathbb{R}^d$  is the parameter of interest and  $\theta \in \Theta \subset \mathbb{R}^m$  is a hyperparameter. Applying Bayes formula, we get that the a posteriori distribution is given for any  $x \in \mathbb{R}^d$  by  $p(x|y) \propto \int_{\Theta} p(y|x, \theta) p(x|\theta) p(\theta) d\theta$ . The problem of sampling from this distribution is the subject of many works and remains a challenge. A common strategy relies on the Gibbs algorithm which consists in sampling from  $p(\theta|x, y)$  and  $p(x|y, \theta)$  alternatively and performing inference using the marginal distribution along the variable  $x$ . However, the two underlying problems are either computationally very expensive or even impossible to solve. In this paper, we consider the EB setting where the a posteriori distribution of  $x$  given  $y$  is approximated by  $p(y|x, \theta^*) p(x|\theta^*) p(\theta^*)$  up to a normalizing constant with

$$\theta^* \in \arg \max p(\theta|y) \text{ with } p(\theta|y) \propto \int p(y|\theta, x) p(x|\theta) p(\theta) dx = \int p(x|y, \theta) dx . \quad (1)$$

In order to realize this task, the main difficulty is to estimate  $\theta^*$ . To do so, we use a stochastic approximation approach defining the recursion  $(\theta_n)_{n \in \mathbb{N}}$  starting from  $\theta_0 \in \Theta$  for any  $n \in \mathbb{N}$  by

$$\theta_{n+1} = \Pi_{\Theta}(\theta_n - \delta_{n+1} \Delta_{\theta_n}) , \quad (2)$$

where  $\Pi_{\Theta}$  is the projection onto  $\Theta$  and, for any  $\theta \in \Theta$ ,  $\Delta_{\theta}$  is an estimator of  $\nabla_{\theta} \log p(\theta|y)$ , where  $\nabla_{\theta}$  is the gradient with respect to the  $\theta$  variable. This estimator can be written in the following form

$$\nabla_{\theta} \log p(\theta|y) = \int \frac{\nabla_{\theta} p(x, y, \theta)}{p(x, y, \theta)} \pi_{\theta}(x) dx , \quad (3)$$

where for any  $\theta \in \Theta$ ,  $\pi_{\theta}$  stands for the conditional distribution  $x \mapsto p(x|y, \theta)$ . This methodology was already studied and proposed in numerous papers [2, 1] in the case where  $\Delta_{\theta}$  is unbiased, directly sampling from  $\pi_{\theta}$ , or based on a Markov Chain Monte Carlo (MCMC) algorithm targeting  $\pi_{\theta}$ . In this framework, for any  $n \in \mathbb{N}$

$$\Delta_{\theta_n} = \frac{\nabla_{\theta} p(\theta_n)}{p(\theta_n)} + m_n^{-1} \sum_{k=1}^{m_n} \frac{\nabla_{\theta} p(X_k^n, y|\theta_n)}{p(X_k^n, y|\theta_n)} , \quad (4)$$

where  $(m_n)_{n \in \mathbb{N}}$  is a sequence of positive integers and for any  $n \in \mathbb{N}$ ,  $(X_k^n)_{k \in \{0, \dots, m_n\}}$  is either a sequence either i.i.d. random variables with distribution  $\pi_{\theta_n}$  or a Markov Chain targeting this distribution.

In this contribution, we propose to use inexact MCMC algorithms which do not target exactly  $\pi_{\theta}$  but a close distribution in a sense which will be precised below. Indeed, there have been recently several major results on this family of methodologies showing that while they are biased, they benefit from explicit convergence properties (and subsampling strategies can be applied in big-data setting) contrary to most Metropolis-Hastings type algorithms [16, 12, 9]. These approximate MCMC schemes are based on stochastic continuous dynamics  $(Y_t^{\theta})_{t \geq 0}$  for which the target distribution  $\pi_{\theta}$  is invariant. Two fundamental examples are the Langevin dynamics solution of the following Stochastic Differential Equation (SDE)

$$dY_t^{\theta} = \nabla_x U_{\theta}(Y_t^{\theta}) dt + \sqrt{2} dB_t , \quad (5)$$

or the kinetic Langevin dynamics solution of

$$dY_t^{\theta} = V_t^{\theta} , \quad dV_t^{\theta} = \nabla_x U_{\theta}(Y_t^{\theta}) dt - V_t^{\theta} dt + \sqrt{2} dB_t ,$$

where for any  $\theta \in \Theta$ ,  $U_{\theta} : x \mapsto \log \pi_{\theta}(x)$  and  $(B_t)_{t \geq 0}$  is a standard  $d$ -dimensional Brownian motion. Under mild assumptions on  $U_{\theta}$ , these two SDEs admit strong solutions for which  $\pi_{\theta}$  and  $\tilde{\pi}_{\theta} : (x, v) \mapsto \pi_{\theta}(x) \exp(-\|v\|^2/2)/(2\pi)^{d/2}$  are invariant probability measures. In addition, explicit convergence of  $(Y_t^{\theta})_{t \geq 0}$  (respectively  $(Y_t^{\theta}, V_t^{\theta})_{t \geq 0}$ ) to  $\pi_{\theta}$  (respectively to  $\tilde{\pi}_{\theta}$ ), in different metrics have been obtained [44, 15, 14]. However, sampling path solutions of these dynamics is in general not feasible. Therefore discretizations have to be used instead. In this paper we mainly focus on the Euler-Maruyama discretizations [44] of (5)

$$X_{k+1} = X_k - \gamma \nabla_x U_{\theta}(X_k) + \sqrt{2\gamma} Z_{k+1} . \quad (6)$$

where  $\gamma > 0$  is a stepsize and  $(Z_k)_{k \in \mathbb{N}^*}$  is a sequence of i.i.d  $d$ -dimensional zero-mean Gaussian random variables with covariance matrix identity. We have all the necessary tools to present our new algorithm. Our new methodology can be also applied to a broad class of stochastic optimization problems. Indeed, as mentioned previously, the EB estimation (1) can be casted in this framework, see (2). Therefore, we present it in full generality in the next section.

## 2.2 Stochastic Optimization with Unadjusted MCMC

The recursion (6) defines a Markov kernel  $R_{\gamma,\theta}$ . Under appropriate conditions, for any  $\gamma \in (0, \bar{\gamma})$ , and  $\theta \in \Theta$  with  $\bar{\gamma} > 0$ , we can show that  $R_{\gamma,\theta}$  admits an invariant distribution, see Lemma 16 in the appendix. This observation motivates the following stochastic approximation scheme.

Let  $\Theta$  be a convex closed set in  $\mathbb{R}^m$ . Consider an objective function  $f : \Theta \rightarrow \mathbb{R}$  which we want to minimize. We assume that  $f$  is continuously differentiable and that its gradient is given for any  $\theta \in \Theta$  by

$$\nabla f(\theta) = \int_{\mathbb{R}^d} H_\theta(x) \pi_\theta(dx), \quad (7)$$

where  $(\theta, x) \mapsto H_\theta(x) \in C(\Theta \times \mathbb{R}^d, \mathbb{R})$ , for any  $\theta \in \Theta$ ,  $H_\theta$  is  $\pi_\theta$ -integrable and  $(\pi_\theta)_{\theta \in \Theta}$  is a family of probability distributions over  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . Note that in the EB setting, we have from (3) that  $H_\theta(x) = \nabla_\theta p(x, y, \theta) / p(x, y, \theta)$ . As emphasized in the previous section, we propose a stochastic approximation scheme which relies on biased estimates of  $\nabla f(\theta)$  through a family of Markov kernels  $\{K_{\gamma,\theta}, \gamma \in (0, \bar{\gamma}) \text{ and } \theta \in \Theta\}$ , for  $\bar{\gamma} > 0$ , such that for any  $\theta \in \Theta$  and  $\gamma \in (0, \bar{\gamma})$ ,  $K_{\gamma,\theta}$  admits an invariant probability distribution  $\pi_{\gamma,\theta}$  on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . We assume in addition that the bias associated to the use of this family of Markov kernels can be controlled w.r.t. to  $\gamma$  uniformly in  $\theta$ , *i.e.* for example there exists  $C > 0$  such that for all  $\gamma \in (0, \bar{\gamma})$  with  $\bar{\gamma} > 0$  and  $\theta \in \Theta$ ,  $\|\pi_{\gamma,\theta} - \pi_\theta\|_{\text{TV}} \leq C\gamma^\alpha$  with  $\alpha > 0$ . In our applications  $K_{\gamma,\theta}$  stands for  $R_{\gamma,\theta}$  for any  $\gamma \in (0, \bar{\gamma})$  with  $\bar{\gamma} > 0$  and  $\theta \in \Theta$ , where  $R_{\gamma,\theta}$  is associated with (6).

Let now  $(\delta_n)_{n \in \mathbb{N}^*} \in (\mathbb{R}_+^*)^{\mathbb{N}^*}$ ,  $(m_n)_{n \in \mathbb{N}} \in (\mathbb{N}^*)^{\mathbb{N}}$  be sequences of stepsizes and batch sizes which will be used to define the sequence relatively to the variable  $\theta$  similarly to (2) and (4). Let  $(\gamma_n)_{n \in \mathbb{N}} \in (\mathbb{R}_+^*)^{\mathbb{N}}$  be a sequence of stepsizes which will be used to get approximate samples from  $\pi_{\theta_n}$ , similarly to (6). Starting from  $X_0^0 \in \mathbb{R}^d$  and  $\theta_0 \in \Theta$  we define on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ ,  $(\{X_k^n : k \in \{0, \dots, m_n\}\}, \theta_n)_{n \in \mathbb{N}}$  by the following recursion for  $n \in \mathbb{N}$  and  $k \in \{0, \dots, m_n - 1\}$

$$X_{k+1}^n \text{ is a Markov chain with kernel } K_{\gamma_n, \theta_n} \text{ and } X_0^n = X_{m_{n-1}}^{n-1} \text{ conditionally to } \mathcal{F}_{n-1} \text{ if } n \geq 1 \quad (8)$$

$$\theta_{n+1} = \Pi_\Theta \left[ \theta_n - \frac{\delta_{n+1}}{m_n} \sum_{k=1}^{m_n} H_{\theta_n}(X_k^n) \right], \quad (9)$$

where  $\Pi_\Theta$  is the projection onto  $\Theta$  and where  $\mathcal{F}_n$  is defined as follows for all  $n \in \mathbb{N}^*$

$$\mathcal{F}_n = \sigma(\theta_0, \{(X_k^\ell)_{k \in \{0, \dots, m_\ell\}} : \ell \in \{0, \dots, n\}\}) , \quad (10)$$

where  $\{(X_k^\ell)_{k \in \{0, \dots, m_\ell\}} : \ell \in \{0, \dots, n\}\}$  is given by (8). Note that such a construction is always possible by Kolmogorov extension theorem [27, Theorem 5.16] by (9), for any  $n \in \mathbb{N}$ ,  $\theta_{n+1}$  is  $\mathcal{F}_n$ -measurable. Then the sequence of approximate minimizers of  $f$  is given by  $(\hat{\theta}_N)_{N \in \mathbb{N}^*}$  where for any  $N \in \mathbb{N}^*$

$$\hat{\theta}_N = \left\{ \sum_{n=1}^N \delta_n \theta_n \right\} / \left\{ \sum_{n=1}^N \delta_n \right\} . \quad (11)$$

The pseudo-code associated with this method is given in Algorithm 1.

Under different sets of conditions on  $f, H, (\delta_n)_{n \in \mathbb{N}}, (\gamma_n)_{n \in \mathbb{N}}$  and  $(m_n)_{n \in \mathbb{N}}$  we obtain that  $(\theta_N)_{N \in \mathbb{N}^*}$  converges a.s. to an element of  $\arg \min_\Theta f$ . In particular we consider the case where  $f$  is assumed to be convex, in Section 3. To summarize our results, we establish that if  $(\gamma_n)_{n \in \mathbb{N}}$  and  $(\delta_n)_{n \in \mathbb{N}}$  go to 0 sufficiently fast,  $\mathbb{E}[f(\hat{\theta}_N)] - \min_\Theta f$  goes to 0 with a quantitative rate of convergence. In

the case where  $(\gamma_n)_{n \in \mathbb{N}}$  is held fixed, *i.e.* for all  $n \in \mathbb{N}$ ,  $\gamma_n = \gamma$ , we show that while  $\mathbb{E}[f(\hat{\theta}_N)]$  does not converge to 0, there exists  $C, \alpha > 0$  such that  $\limsup_{N \rightarrow +\infty} \mathbb{E}[f(\hat{\theta}_N)] - \min_{\Theta} f \leq C\gamma^\alpha$ . In the case where  $f$  is non-convex we show that we can apply some results of stochastic approximation which imply that the sequence  $(\theta_n)_{n \in \mathbb{N}}$  converges a.s. to a stationary point of the projected ordinary differential equation associated with  $\nabla f$  and  $\Theta$ . We postpone this result to the appendix, see Appendix A, since it involves a theoretical background which we think is out of the scope of the main document.

---

**Algorithm 1** Stochastic optimization with unadjusted Markov kernel (SOUK)

---

```

1: function SOUK
2:   Inputs:
    $\{K_{\gamma, \theta} : \gamma > 0, \theta \in \Theta\}, (\gamma_n)_{n \in \mathbb{N}}, (\delta_n)_{n \in \mathbb{N}}, (m_n)_{n \in \mathbb{N}}, N$ 
3:   Initialize:
    $X_0^0 \in \mathbb{R}^d$  and  $\theta_0 \in \Theta$ 
4:   for  $n \in \{0, \dots, N-1\}$  do
5:     if  $n \geq 1$  then
6:        $X_0^n = X_{m_{n-1}}^{n-1}$ 
7:     end if
8:     for  $k \in \{0, \dots, m_n - 1\}$  do
9:        $X_{k+1}^n \sim K_{\gamma_n, \theta_n}(X_k^n, \cdot)$ 
10:    end for
11:     $\theta_{n+1} = \Pi_{\Theta} \left[ \theta_n - \frac{\delta_{n+1}}{m_n} \sum_{k=1}^{m_n} H_{\theta_n}(X_k^n) \right]$ 
12:  end for
13:  Outputs:
    $\hat{\theta}_N = \left\{ \sum_{n=1}^N \delta_n \theta_n \right\} / \left\{ \sum_{n=1}^N \delta_n \right\}$ 
14: end function

```

---

## 3 Stochastic Optimization with Unadjusted MCMC

### 3.1 Convex objective function

We consider the following assumptions on  $\Theta$  and  $f$ , where  $\Theta \subset \mathbb{R}^m$ .

**A1.**  $\Theta$  is a convex compact set and  $\Theta \subset \bar{B}(0, M_{\Theta})$  with  $M_{\Theta} > 0$ .

**A2.** There exist an open set  $U \subset \mathbb{R}^m$  and  $L_f \geq 0$  such that  $\Theta \subset U$  and  $f \in C^1(U, \mathbb{R})$  is convex and for any  $\theta_1, \theta_2 \in \Theta$

$$\|\nabla f(\theta_1) - \nabla f(\theta_2)\| \leq L_f \|\theta_1 - \theta_2\|.$$

Note that under **A1** and **A2**,  $\arg \min_{\Theta} f$  is non-empty. The following assumption ensures that the gradient of the objective function can be written as the expectation of some integrable function.

**A3.** For any  $\theta \in \Theta$ , there exist  $H_{\theta} : \mathbb{R}^d \rightarrow \mathbb{R}^m$  and a probability distribution  $\pi_{\theta}$  on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  satisfying that  $(\theta, x) \mapsto H_{\theta}(x)$  is measurable,  $H_{\theta}$  is  $\pi_{\theta}$ -integrable and (7).

We impose some stability condition on the stochastic process  $\{(X_k^n)_{k \in \{0, \dots, m_n\}} : n \in \mathbb{N}\}$  defined by (8) and that for any  $\gamma \in (0, \bar{\gamma})$ , with  $\bar{\gamma} > 0$  and  $\theta \in \Theta$  the iterates of  $K_{\gamma, \theta}$  are close enough to  $\pi_\theta$  after a sufficiently large number of iterations.

**H1.** (i) *There exists a measurable function  $V : \mathbb{R}^d \rightarrow [1, +\infty)$  and  $A_1 \geq 0$  such that for any  $n \in \mathbb{N}$*

$$\mathbb{E} [V(X_{m_n}^n) | X_0^0] \leq A_1 V(X_0^0), \quad \mathbb{E} [V(X_0^0)] < +\infty. \quad (12)$$

(ii) *In addition, there exist  $\bar{\gamma} > 0$ ,  $A_2, A_3 \geq 0$ ,  $B : \mathbb{R}_+^* \rightarrow \mathbb{R}_+$  such that for any  $\gamma \in (0, \bar{\gamma})$ ,  $\theta \in \Theta$ ,  $x \in \mathbb{R}^d$  and  $n \in \mathbb{N}$*

$$\left| \int_{\mathbb{R}^d} K_{\gamma, \theta}^n(x, dy) H_\theta(y) - \pi_\theta(H_\theta) \right| \leq A_2 \rho^{n\gamma} V(x) + A_3 B(\gamma). \quad (13)$$

This assumption does not impose that the Markov kernel  $K_{\gamma, \theta}$ , with  $\gamma \in (0, \bar{\gamma})$  where  $\bar{\gamma} > 0$  and  $\theta \in \Theta$  is uniformly  $V$ -geometrically ergodic. However if an appropriate ergodicity condition holds for the family of kernels  $\{K_{\gamma, \theta}, \gamma \in (0, \bar{\gamma}), \theta \in \Theta\}$ , as in **H5** below, then (13) can be written as  $\|\pi_{\gamma, \theta}(H_\theta) - \pi_\theta(H_\theta)\| \leq A_3 B(\gamma)$  for any  $\gamma \in (0, \bar{\gamma})$  with  $\bar{\gamma} > 0$  and  $\theta \in \Theta$ , where  $\pi_{\gamma, \theta}$  is the unique invariant probability distribution of  $K_{\gamma, \theta}$ . We show that this condition holds in the case of the Langevin algorithm in Proposition 18.

**Theorem 1.** *Assume **A1**, **A2**, **A3**. Let  $\bar{\gamma} > 0$ ,  $(\gamma_n)_{n \in \mathbb{N}}$ ,  $(\delta_n)_{n \in \mathbb{N}^*}$  be sequences of non-increasing positive real numbers and  $(m_n)_{n \in \mathbb{N}}$  be a sequence of positive integers satisfying  $\sup_{n \in \mathbb{N}} \delta_n < 1/L_f$ ,  $\sup_{n \in \mathbb{N}} \gamma_n < \bar{\gamma}$  and*

$$\sum_{n=0}^{+\infty} \delta_{n+1} = +\infty, \quad \sum_{n=0}^{+\infty} \delta_{n+1} B(\gamma_n) < +\infty, \quad \sum_{n=0}^{+\infty} \delta_{n+1} / (m_n \gamma_n) < +\infty. \quad (14)$$

Let  $\{(X_k^n)_{k \in \{0, \dots, m_n\}} : n \in \mathbb{N}\}$  be given by (8). Assume in addition that **H1** holds. Then the following statements hold:

- (a)  $(\theta_n)_{n \in \mathbb{N}}$  defined by (9) converges a.s. to some  $\theta^* \in \arg \min_{\Theta} f$ ;
- (b) furthermore, a.s. there exists  $C \geq 0$  such that for any  $n \in \mathbb{N}^*$

$$\left\{ \frac{\sum_{k=1}^n \delta_k f(\theta_k)}{\sum_{k=1}^n \delta_k} \right\} - \min_{\Theta} f \leq C / \left( \sum_{k=1}^n \delta_k \right).$$

We start with the following technical lemma. Consider  $(\eta_n)_{n \in \mathbb{N}}$  the sequence defined for any  $n \in \mathbb{N}$  by

$$\eta_n = m_n^{-1} \sum_{k=1}^{m_n} \{H_{\theta_n}(X_k^n) - \pi_{\theta_n}(H_{\theta_n})\}, \quad (15)$$

**Lemma 2.** *Under the hypotheses of Theorem 1 we have for any  $n \in \mathbb{N}^*$*

$$\mathbb{E} [\|\eta_n\|] \leq (A_1 \vee 1) A_2 \frac{\rho^{-\bar{\gamma}} \mathbb{E} [V(X_0^0)]}{\log(1/\rho) \gamma_n m_n} + A_3 B(\gamma_n).$$



*Proof.* Using the definition of  $(\mathcal{F}_n)_{n \in \mathbb{N}}$ , see (10), the Markov property, Lemma 11 and **H1-(ii)** we have for any  $n \in \mathbb{N}$

$$\begin{aligned} \mathbb{E} [\|\eta_{n+1}\| | \mathcal{F}_n] &\leq (1/m_{n+1}) \sum_{k=1}^{m_{n+1}} |K_{\gamma_{n+1}, \theta_{n+1}}^k H_{\theta_{n+1}}(X_{m_n}^n) - \pi_{\theta_{n+1}}(H_{\theta_{n+1}})| \\ &\leq m_{n+1}^{-1} \sum_{k=1}^{m_{n+1}} \{A_2 \rho^{k\gamma_{n+1}} V(X_{m_n}^n) + A_3 B(\gamma_{n+1})\} \\ &\leq \frac{A_2 \rho^{-\bar{\gamma}} V(X_{m_n}^n)}{\log(1/\rho)\gamma_{n+1}m_{n+1}} + A_3 B(\gamma_{n+1}), \end{aligned}$$

where for the last inequality we use that  $\sum_{k=1}^{+\infty} \rho^k \leq \rho^{-\bar{\gamma}} / [\log(1/\rho)\gamma]$ . In a similar manner, we have

$$\mathbb{E} [\|\eta_0\| | X_0^0] \leq \frac{A_2 \rho^{-\bar{\gamma}} V(X_0^0)}{\log(1/\rho)\gamma_0 m_0} + A_3 B(\gamma_0).$$

We conclude using (12) in **H1-(i)**. □

We now turn to the proof of Theorem 1.

*Proof.* The proof is an application of [2, Theorem 2, Theorem 3]

(a) To apply [2, Theorem 2], it is enough to show that the following series converge a.s.

$$\sum_{n=0}^{+\infty} \delta_{n+1} \langle \Pi_{\Theta}(\theta_n - \delta_{n+1} \nabla f(\theta_n)), \eta_n \rangle, \quad \sum_{n=0}^{+\infty} \delta_{n+1} \|\eta_n\|, \quad \sum_{n=0}^{+\infty} \delta_{n+1}^2 \|\eta_n\|^2.$$

where the sequence  $(\eta_n)_{n \in \mathbb{N}}$  is defined for any  $n \in \mathbb{N}$  by (15). Since  $\Pi_{\Theta}(\theta_n - \delta_{n+1} \nabla f(\theta_n))$  is bounded, we are reduced to proving that a.s.  $\sum_{n=0}^{+\infty} \delta_{n+1} \|\eta_n\| < +\infty$ . As a result, using (14) and Lemma 2 we obtain that  $\sum_{n \in \mathbb{N}} \delta_{n+1} \mathbb{E}[\|\eta_n\|] < +\infty$ , which implies the stated convergence applying [2, Theorem 2].

(b) Combining [2, Theorem 3] with **A1** we obtain that a.s. for any  $n \in \mathbb{N}$

$$\sum_{k=1}^n \delta_k \left\{ f(\theta_k) - \min_{\Theta} f \right\} \tag{16}$$

$$\begin{aligned} &\leq \frac{\|\theta_0 - \theta^*\|^2}{2} - \sum_{k=0}^{n-1} \delta_{k+1} \langle \Pi_{\Theta}(\theta_k - \delta_{k+1} \nabla f(\theta_k)) - \theta^*, \eta_k \rangle + \sum_{k=0}^{n-1} \delta_{k+1}^2 \|\eta_k\|^2 \\ &\leq 2M_{\Theta}^2 + 2M_{\Theta} \sum_{k=0}^{n-1} \delta_{k+1} \|\eta_k\| + \sum_{k=0}^{n-1} \delta_{k+1}^2 \|\eta_k\|^2. \end{aligned} \tag{17}$$

Since  $\sum_{k \in \mathbb{N}} \delta_{k+1} \|\eta_k\|$  is finite a.s. so is  $\sum_{k \in \mathbb{N}} \delta_{k+1}^2 \|\eta_k\|^2$ . The proof is then completed upon dividing (16) by  $\sum_{k=1}^n \delta_k$ . □

In (8),  $X_0^n = X_{m_{n-1}}^{n-1}$  for  $n \in \mathbb{N}^*$ . This procedure is referred to as warm-start in the sequel. In the proof of Theorem 1,  $X_0^n$  could be a random variable independent from  $\mathcal{F}_{n-1}$  for any  $n \in \mathbb{N}$  with  $\sup_{n \in \mathbb{N}^*} \mathbb{E}[V(X_0^n)] < +\infty$ . This will not be the case in the fixed batch size case, see Theorem 4, where the warm-start procedure is crucial for the convergence to occur.

We extend this theorem to non convex objective function see Theorem 9 in Appendix A. Under the conditions of Theorem 1 with the additional assumption that  $\partial\Theta$  is a smooth manifold we obtain that  $(\hat{\theta}_n)_{n \in \mathbb{N}}$  converges a.s. to some point  $\theta^*$  such that  $\nabla f(\theta^*) + \mathbf{n} = 0$  with  $\mathbf{n} = 0$  if  $\theta^* \in \text{int}(\Theta)$  and  $\mathbf{n} \in T(\theta^*, \partial\Theta)^\perp$  if  $\theta^* \in \partial\Theta$ , where  $T(\theta, \partial\Theta)$  is the tangent space of  $\partial\Theta$  at point  $\theta \in \partial\Theta$ , see [3, Chapter 2].

In our applications  $B(\gamma) = \gamma^{1/2}$ , see Theorem 5. Assume then that there exist  $a, b, c > 0$  such that for any  $n \in \mathbb{N}^*$ ,  $\delta_n = n^{-a}$ ,  $\gamma_n = n^{-b}$  and  $m_n = n^c$  then (14) is equivalent to

$$a < 1, \quad a + b/2 > 1, \quad a - b + c > 1.$$

Suppose  $a \in [0, 1)$  is given, then the previous equation reads

$$b = 2(1 - a) + \varsigma_1, \quad c = 3(1 - a) + \varsigma_2, \quad \varsigma_2 > \varsigma_1. \quad (18)$$

Thus  $b$  and  $c$  are proportional to  $1 - a$ . This illustrates a trade-off between the intrinsic inaccuracy of our algorithm through the family of Markov kernels (8) which do not exactly target  $\pi_\theta$  and the minimization aim of our scheme. Note also that  $(\delta_n)_{n \in \mathbb{N}}$  is allowed to be constant. This situation yield  $\gamma_n = n^{-2-\varsigma_1}$  and  $m_n = n^{3+\varsigma_2}$  with  $\varsigma_2 > \varsigma_1 > 0$ .

With the additional following assumption we establish quantitative bounds on  $\mathbb{E}[f(\hat{\theta}_n) - \min_{\Theta} f]$ , where  $(\hat{\theta}_n)_{n \in \mathbb{N}}$  is given by (11).

**H2.** *There exists  $A_4 \geq 0$  such that for any  $n \in \mathbb{N}$  we have*

$$\mathbb{E} \left[ \left( \sum_{k=1}^{m_n} H_{\theta_n}(X_k^n) - \pi_{\theta_n}(H_{\theta_n}) \right)^2 \right] \leq A_4 D(\gamma_n, m_n) \mathbb{E}[V(X_0^0)]. \quad (19)$$

Note that this assumption strengthens **H1** and imposes a control on the mean square error associated with the family of Markov kernels  $\{K_{\gamma, \theta} : \gamma \in (0, \bar{\gamma}), \theta \in \Theta\}$  used to target  $(\pi_\theta)_{\theta \in \Theta}$  and the family of  $\{x \mapsto H_\theta(x) : \theta \in \Theta\}$ .

**Theorem 3.** *Assume **A1**, **A2**, **A3**. Let  $(\gamma_n)_{n \in \mathbb{N}}$ ,  $(\delta_n)_{n \in \mathbb{N}}$  be sequences of non-increasing positive real numbers and  $(m_n)_{n \in \mathbb{N}}$  be a sequence of positive integers satisfying  $\sup_{n \in \mathbb{N}} \delta_n < 1/L_f$ ,  $\sup_{n \in \mathbb{N}} \gamma_n < \bar{\gamma}$ . Let  $\{(X_k^n)_{k \in \{0, \dots, m_n\}} : n \in \mathbb{N}\}$  be given by (8). Assume in addition that **H1** and **H2** hold. Then, there exists  $(E_n)_{n \in \mathbb{N}}$  such that*

$$\mathbb{E} \left[ \left\{ \frac{\sum_{k=1}^n \delta_k f(\theta_k)}{\sum_{k=1}^n \delta_k} \right\} - \min_{\Theta} f \right] \leq E_n / \left( \sum_{k=1}^n \delta_k \right),$$

with for any  $n \in \mathbb{N}$ ,

$$\begin{aligned} E_n = & 2M_{\Theta}^2 + 2M_{\Theta} \mathbb{E}[V(X_0^0)] \left[ \frac{(A_1 \vee 1)A_2}{\rho^{\bar{\gamma}} \log(1/\rho)} \sum_{k=0}^n \{\delta_{k+1}/(\gamma_k m_k)\} + A_3 \sum_{k=0}^n \delta_{k+1} B(\gamma_k) \right] \\ & + A_4 \mathbb{E}[V(X_0^0)] \sum_{k=0}^n \delta_{k+1}^2 m_k^{-2} D(\gamma_k, m_k). \end{aligned} \quad (20)$$

*Proof.* Taking the expectation in (16) we get that for every  $n \in \mathbb{N}$

$$\mathbb{E} \left[ \sum_{k=0}^{n-1} \delta_{k+1} \left\{ f(\theta_k) - \min_{\Theta} f \right\} \right] \leq 2M_{\Theta}^2 + 2M_{\Theta} \sum_{k=0}^{n-1} \delta_{k+1} \mathbb{E} [\|\eta_k\|] + \sum_{k=0}^{n-1} \delta_{k+1}^2 \mathbb{E} [\|\eta_k\|^2] . \quad (21)$$

Using (15) and (19) we obtain that for any  $n \in \mathbb{N}$

$$\mathbb{E} [\|\eta_n\|^2] \leq A_4 m_n^{-2} D(\gamma_n, m_n) \mathbb{E} [V(X_0^0)] . \quad (22)$$

Combining (21), Lemma 2 and (22) we conclude the proof.  $\square$

In the case where  $K_{\gamma, \theta} = R_{\gamma, \theta}$  is the Markov kernel associated with the Langevin update (6) for any  $\theta \in \Theta$  and  $\gamma \in (0, \bar{\gamma})$  with  $\bar{\gamma} > 0$ ,  $B(\gamma) = \gamma^{1/2}$  and  $D(\gamma, m) = m^2$ , see Theorem 5. Therefore if (14) and  $\sum_{k=0}^{+\infty} \delta_{k+1}^2 < +\infty$  hold then  $(E_n)_{n \in \mathbb{N}}$  is bounded by  $E_{\infty}$ . Thus we obtain the L<sup>1</sup> convergence of the sequence  $(f(\hat{\theta}_n))_{n \in \mathbb{N}}$  to  $\min_{\Theta} f$  with rate  $(\sum_{k=1}^n \delta_k)^{-1}$  and explicit bound  $E_{\infty}$ .

In the case of the Langevin update (6), if there exist  $a, b, c > 0$  such that for any  $n \in \mathbb{N}^*$ ,  $\delta_n = n^{-a}$ ,  $\gamma_n = n^{-b}$  and  $m_n = n^c$ , (14) is equivalent to (18) and the accuracy, respectively the complexity, of the algorithm are of orders  $(\sum_{k=1}^n \delta_k)^{-1} = \mathcal{O}(n^{a-1})$ , respectively  $\sum_{k=0}^n m_k = \mathcal{O}(n^{3(1-a)+\varsigma_2+1})$  for  $\varsigma_2 > 0$ . Thus for a fix target precision  $\varepsilon > 0$  the complexity reads  $\varepsilon^{-3} (\log(1/\varepsilon)/(1-a))^{1+\varsigma_2}$ . On the other hand if we fix the complexity budget to  $N$  the accuracy is of order  $N^{-(3+(1+\varsigma_2)/(1-a))^{-1}}$ . These two considerations suggest to set  $a$  close to 0.

A case of interest is the fix stepsize setting, *i.e.* for all  $n \in \mathbb{N}$ ,  $\gamma_n = \gamma$ . In that case if  $(\delta_n)_{n \in \mathbb{N}}$  is non-increasing,  $\sum_{n \in \mathbb{N}^*} \delta_n = +\infty$ ,  $\lim_{n \rightarrow +\infty} m_n = +\infty$  and  $\lim_{n \rightarrow +\infty} D(\gamma_0, m_n) m_n^{-2} \leq A_0(\gamma_0)$ , using [38, Problem 80] we obtain that

$$\limsup_n \mathbb{E} \left[ \left\{ \frac{\sum_{k=1}^n \delta_k f(\theta_k)}{\sum_{k=1}^n \delta_k} \right\} - \min f \right] \leq \mathbb{E} [V(X_0^0)] (M_{\Theta} A_3 B(\gamma_0) + A_4 \delta_0 A_0(\gamma_0)) .$$

Similar bounds can be obtained if  $(m_n)_{n \in \mathbb{N}}$  and  $(\gamma_n)_{n \in \mathbb{N}}$  are constant. However if  $(m_n)_{n \in \mathbb{N}}$  is constant the convergence cannot be obtained using Theorem 1. Indeed in (14), the condition  $\sum_{n \in \mathbb{N}} \delta_{n+1} \gamma_n^{-1} < +\infty$  is in contradiction with the condition  $\sum_{n \in \mathbb{N}^*} \delta_n = +\infty$ .

### 3.2 Convex objective function – Fixed batch size

In this section, we consider the case where the batch size is fixed, *i.e.*  $m_n = m_0$  for all  $n \in \mathbb{N}$ . For ease of exposition we only consider the case where  $m_0 = 1$ . However the general case can be adapted from the proof of the result stated below. For simplicity of notation we let  $\tilde{X}_{n+1} = X_1^n$  for any  $n \in \mathbb{N}$ . More precisely the recursion (8) and (9) can be written for any  $n \in \mathbb{N}$  as

$$\tilde{X}_{n+1} \text{ is a sample from } K_{\gamma_n, \tilde{\theta}_n}(\tilde{X}_n, \cdot) \text{ conditionally to } \tilde{\mathcal{F}}_n , \quad (23)$$

$$\tilde{\theta}_{n+1} = \Pi_{\Theta} [\tilde{\theta}_n - \delta_{n+1} H_{\tilde{\theta}_n}(\tilde{X}_{n+1})] , \quad (24)$$

starting from  $\theta_0 \in \mathbb{R}^d$ ,  $\tilde{X}_0$  a random variable and where  $\tilde{\mathcal{F}}_n$  is given by

$$\tilde{\mathcal{F}}_n = \sigma(\tilde{\theta}_0, (\tilde{X}_{\ell})_{\ell \in \{0, \dots, n\}}) . \quad (25)$$

We consider the following additional assumptions. We impose some Lipschitz-regularity over  $(\theta, x) \mapsto H_{\theta}(x)$ .

**A4.** *There exists a measurable function  $V : \mathbb{R}^d \rightarrow [1, +\infty)$  and  $M_H, L_H \geq 0$  such that for any  $x \in \mathbb{R}^d$  and  $\theta_1, \theta_2 \in \Theta$  we have the following inequalities*

$$\|H_{\theta_1}(x) - H_{\theta_2}(x)\| \leq L_H \|\theta_1 - \theta_2\| V(x)^{1/2}, \quad \|H_{\theta_1}(x)\| \leq M_H V^{1/4}(x).$$

**H3.** *There exists a measurable function  $V : \mathbb{R}^d \rightarrow [1, +\infty)$  and  $A_5 \geq 0$  such that for any  $n \in \mathbb{N}$*

$$\mathbb{E} [V(\tilde{X}_n) | \tilde{X}_0] \leq A_5 V(\tilde{X}_0), \quad \mathbb{E} [V(\tilde{X}_0)] < +\infty.$$

We consider a similar property as **A4** on the family of kernels  $\{K_{\gamma, \theta}, \gamma \in (0, \bar{\gamma}), \theta \in \Theta\}$ , with  $\bar{\gamma} > 0$ , which weakens the assumption [2, H6].

**H4.** *There exist  $A_6 \geq 0$ ,  $\Lambda_1 : (\mathbb{R}_+^*)^2 \rightarrow \mathbb{R}_+$  and  $\Lambda_2 : (\mathbb{R}_+^*)^2 \rightarrow \mathbb{R}_+$  such that for any  $\gamma_1, \gamma_2 \in (0, \bar{\gamma})$  with  $\bar{\gamma} > 0$ ,  $\theta_1, \theta_2 \in \Theta$ ,  $x \in \mathbb{R}^d$  and  $a \in [1/4, 1/2]$*

$$\|\delta_x K_{\gamma_1, \theta_1} - \delta_x K_{\gamma_2, \theta_2}\|_{V^a} \leq A_6 [\Lambda_1(\gamma_1, \gamma_2) + \Lambda_2(\gamma_1, \gamma_2) \|\theta_1 - \theta_2\|] V^{2a}(x).$$

Finally, we suppose that each kernel  $K_{\gamma, \theta}$  admits an invariant probability measure and converges geometrically to this probability distribution in  $V$ -norm with constants not depending on  $\gamma$  and  $\theta$ . We show in [10] that this condition holds for the Langevin Monte Carlo algorithm.

**H5.** *For all  $\theta \in \Theta$  and  $\gamma \in (0, \bar{\gamma})$  with  $\bar{\gamma} > 0$ ,  $K_{\gamma, \theta}$  is uniformly  $V$ -geometrically ergodic, i.e. there exists a probability distribution  $\pi_{\gamma, \theta}$  on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ ,  $A_7 \geq 0$  and  $\kappa \in (0, 1)$  such that for any  $n \in \mathbb{N}$  and  $a \in [1/4, 1]$*

$$\|K_{\gamma, \theta}^n - \pi_{\gamma, \theta}\|_{V^a} \leq A_7 \kappa^{n\gamma} V^a(x).$$

The following theorem ensures convergence properties for the algorithm similar to Theorem 1. The proof of this result is based on a generalization of [20, Lemma 4.2], which only holds in the case of exact MCMC schemes.

**Theorem 4.** *Assume **A1**, **A2**, **A3** and **A4**. Let  $\bar{\gamma} > 0$ ,  $(\gamma_n)_{n \in \mathbb{N}}$  and  $(\delta_n)_{n \in \mathbb{N}^*}$  be sequences of non-increasing positive real numbers satisfying  $\sup_{n \in \mathbb{N}} \delta_n < 1/L_f$ ,  $\sup_{n \in \mathbb{N}} \gamma_n < \bar{\gamma}$ ,  $\sup_{n \in \mathbb{N}} |\delta_{n+1} - \delta_n| \delta_n^{-2} < +\infty$ ,  $\sum_{n=0}^{+\infty} \delta_{n+1} = +\infty$  and*

$$\sum_{n=0}^{+\infty} \delta_{n+1} B(\gamma_n) < +\infty, \quad \sum_{n=0}^{+\infty} \frac{\delta_{n+1}^2}{\gamma_n^2} < +\infty, \quad \sum_{n=0}^{+\infty} \frac{\delta_{n+1}}{\gamma_{n+1}} [\Lambda_1(\gamma_n, \gamma_{n+1}) + \delta_{n+1} \Lambda_2(\gamma_n, \gamma_{n+1})] < +\infty. \quad (26)$$

Let  $(\tilde{X}_n)_{n \in \mathbb{N}}$  be given by (23). Assume in addition that **H1-(ii)**, **H3**, **H4**, **H5** hold. Then the following statements hold:

- (a)  $(\tilde{\theta}_n)_{n \in \mathbb{N}}$  defined by (24) converges a.s. to some  $\theta^* \in \arg \min_{\Theta} f$ ;
- (b) furthermore, a.s. there exists  $C \geq 0$  such that for any  $n \in \mathbb{N}^*$

$$\left\{ \frac{\sum_{k=1}^n \delta_k f(\tilde{\theta}_k)}{\sum_{k=1}^n \delta_k} \right\} - \min_{\Theta} f \leq C / \left( \sum_{k=1}^n \delta_k \right).$$

*Proof.* The proof of this theorem is postponed to Appendix B.1. □

In our case of application  $B(\gamma) = \gamma^{1/2}$ ,  $\Lambda_1(\gamma_1, \gamma_2) = (\gamma_2^{-1} - \gamma_1^{-1})^{1/2} + \gamma_2^{-1/2} |\gamma_1 - \gamma_2|$  and  $\Lambda_2(\gamma_1, \gamma_2) = \gamma_2^{1/2}$ . Thus we obtain that the following series should be convergent

$$\sum_{n=0}^{+\infty} \delta_{n+1} \gamma_n^{1/2} < +\infty, \quad \sum_{n=0}^{+\infty} \delta_{n+1}^2 / \gamma_n^2 < +\infty, \quad \sum_{n=0}^{+\infty} \delta_{n+1} (\gamma_{n+1}^{-1} - \gamma_n^{-1})^{1/2} / \gamma_n^2 < +\infty.$$

A possible choice is given  $\delta_n = n^{-1}$  and  $\gamma_n = \log(n+1)^{-4}$ . In this setting it should be noted that the sequence of  $(\tilde{\theta}_n)_{n \in \mathbb{N}}$  moves slowly compared to the sequence  $(\tilde{X}_n)_{n \in \mathbb{N}}$ . In the limit case, if we suppose that  $\tilde{\theta}_n$  is constant we obtain that averaging the sequence  $(\tilde{X}_n)_{n \in \mathbb{N}}$  is an estimator of  $\nabla f(\tilde{\theta}_n)$ . This is coherent with the increasing batch size approach.

### 3.3 SOUL method

We consider the following assumption on the family of probability distributions  $(\pi_\theta)_{\theta \in \Theta}$ .

**L1.** For any  $\theta \in \Theta$ , there exists  $U_\theta : \mathbb{R}^d \rightarrow \mathbb{R}$  such that  $\pi_\theta$  admits a probability density function w.r.t. to the Lebesgue measure proportional to  $x \mapsto \exp(-U_\theta(x))$ . In addition  $(\theta, x) \mapsto U_\theta(x)$  is continuous,  $x \mapsto U_\theta(x)$  is differentiable for all  $\theta \in \Theta$  and there exists  $L \geq 0$  such that for any  $x, y \in \mathbb{R}^d$ ,

$$\sup_{\theta \in \Theta} \|\nabla_x U_\theta(x) - \nabla_x U_\theta(y)\| \leq L \|x - y\|,$$

and  $\{\|\nabla_x U_\theta(0)\| : \theta \in \Theta\}$  is bounded.

Under **L1**, the Langevin diffusion defined by (5) admits a unique strong solution for any  $\theta \in \Theta$ . Therefore we consider the family of Markov kernels  $\{R_{\gamma, \theta}, \gamma \in (0, \bar{\gamma}), \theta \in \Theta\}$  with  $\bar{\gamma} > 0$ , induced by the recursion (6). The recursion step in (8) can be rewritten for any  $n \in \mathbb{N}$  and  $k \in \{0, \dots, m_n - 1\}$

$$X_{k+1}^n = X_k^n - \gamma_n \nabla_x U_{\theta_n}(X_k^n) + \sqrt{2\gamma_n} Z_{k+1}^n, \quad \text{with } X_0^n = X_{m_n-1}^{n-1} \text{ if } n \geq 1, \quad (27)$$

given a sequence of stepsizes  $(\gamma_n)_{n \in \mathbb{N}}$ ,  $(m_n)_{n \in \mathbb{N}} \in (\mathbb{N}^*)^{\mathbb{N}}$  and  $(Z_k^n)_{n \in \mathbb{N}, k \in \{1, \dots, m_n\}}$  a family of i.i.d  $d$ -dimensional zero-mean Gaussian random variables with covariance matrix identity. Consider one of the following additional tail conditions on  $U_\theta$  to ensure geometric ergodicity of  $R_{\gamma, \theta}$  for any  $\gamma \in (0, \bar{\gamma})$  and  $\theta \in \Theta$ , with  $\bar{\gamma} > 0$  which will be specified below.

**L2.** There exist  $\eta > 0$  and  $m, c, M_\eta \geq 0$  such that for any  $\theta \in \Theta$  and  $x \in \mathbb{R}^d$ ,

$$\langle \nabla_x U_\theta(x), x \rangle \geq \eta \|x\| \mathbb{1}_{\overline{B}(0, M_\eta)^c}(x) + m \|\nabla_x U_\theta(x)\|^2 - c.$$

**L3.** There exist  $\eta > 0$  and  $c, M_\eta \geq 0$  such that for any  $\theta \in \Theta$  and  $x \in \mathbb{R}^d$

$$\langle \nabla_x U_\theta(x), x \rangle \geq \eta \|x\|^2 \mathbb{1}_{\overline{B}(0, M_\eta)^c}(x) - c,$$

The next theorem asserts the convergence of the scheme defined by (9) and (27) referred to as the Stochastic Optimization with Unadjusted Langevin (SOUL) algorithm.

**Theorem 5.** Assume **A1**, **A2**, **A3** and there exists  $M_H \geq 0$  such that for any  $\theta \in \Theta$  and  $x \in \mathbb{R}^d$ ,  $\|H_\theta(x)\| \leq M_H V^{1/2}(x)$ . In addition, assume **L1** and **L2** or **L3**. There exists  $\bar{\gamma} > 0$  such that for all sequences of non-increasing positive real numbers  $(\gamma_n)_{n \in \mathbb{N}}$ ,  $(\delta_n)_{n \in \mathbb{N}^*}$  and sequence of positive integers  $(m_n)_{n \in \mathbb{N}}$ , satisfying  $\sup_{n \in \mathbb{N}} \gamma_n < \bar{\gamma}$ , **H1** and **H2** hold with  $B(\gamma) = \gamma^{1/2}$  and  $D(\gamma, m) = m^2$ .

*Proof.* The proof is postponed to Appendix B.2.4.  $\square$

Consider the following assumption for  $m \geq 0$ .

**L4** ( $m$ ). For any  $\theta \in \Theta$ ,  $U_\theta$  is  $m$ -convex, continuously differentiable and satisfies

$$\sup_{\theta \in \Theta} \int_{\mathbb{R}^d} \exp(-U_\theta(x)) dx < +\infty.$$

In addition,  $\{y \in \mathbb{R}^d : \text{there exists } \theta \in \Theta, \nabla_x U_\theta(y) = 0\} \subset B(0, M_*)$  with  $M_* \geq 0$ .

Note that  $\{y \in \mathbb{R}^d : \text{there exists } \theta \in \Theta, \nabla_x U_\theta(y) = 0\}$  is the set of minimizers of the family of functions  $(U_\theta)_{\theta \in \Theta}$ .

**Proposition 6.** Assume **A1**, **L1** and **L4**(0). Then **L2** holds.

*Proof.* The proof is postponed to Appendix B.3.  $\square$

While Proposition 6 shows that **L4**(0) and additional assumptions implies **L2**, it does not give an explicit expression for the constants appearing in this latter condition. In the case where **L4**( $m$ ) with  $m > 0$  holds, we get the following result which implies that the bound provided by Theorem 3 is linear in the dimension.

**Corollary 7.** Assume **A1**, **A2**, **A3** and there exists  $M_H \geq 0$  such that for any  $\theta \in \Theta$  and  $x \in \mathbb{R}^d$ ,  $\|H_\theta(x)\| \leq M_H V^{1/2}(x)$ . In addition assume **L1** and **L4**( $m$ ), with  $m > 0$ . There exists  $\bar{\gamma} > 0$  such that for all sequences of non-increasing positive real numbers  $(\gamma_n)_{n \in \mathbb{N}}$ ,  $(\delta_n)_{n \in \mathbb{N}^*}$  and sequence of positive integers  $(m_n)_{n \in \mathbb{N}}$ , satisfying  $\sup_{n \in \mathbb{N}} \delta_n < 1/L_f$ ,  $\sup_{n \in \mathbb{N}} \gamma_n < \bar{\gamma}$  and (14) with  $B(\gamma) = \gamma^{1/2}$ , and  $D(\gamma, m) = m^2$ , the conclusions of Theorem 3 hold with  $(E_n)_{n \in \mathbb{N}}$  such that for any  $n \in \mathbb{N}$ ,  $|E_n| \leq Cd$ , where  $C \geq 0$  is a constant independent of the dimension  $d$ .

*Proof.* The proof is postponed to Appendix B.4.  $\square$

**L5.** There exists  $L_U \geq 0$  such that for any  $x \in \mathbb{R}^d$  and  $\theta_1, \theta_2 \in \Theta$

$$\|\nabla_x U_{\theta_1}(x) - \nabla_x U_{\theta_2}(x)\| \leq L_U \|\theta_1 - \theta_2\| V(x)^{1/2}.$$

**Theorem 8.** Assume **A1**, **A2**, **A3** and **A4**. In addition assume **L1**, **L5** and **L2** or **L3**. There exists  $\bar{\gamma} > 0$  such that for all sequences of non-increasing positive real numbers  $(\gamma_n)_{n \in \mathbb{N}}$ ,  $(\delta_n)_{n \in \mathbb{N}^*}$  and sequence of positive integers  $(m_n)_{n \in \mathbb{N}}$ , satisfying  $\sup_{n \in \mathbb{N}} \gamma_n < \bar{\gamma}$ , **H4** and **H5** hold with

$$\Lambda_1(\gamma_1, \gamma_2) = (\gamma_2^{-1} - \gamma_1^{-1})^{1/2} + \gamma_2^{-1/2} |\gamma_1 - \gamma_2|, \quad \Lambda_2(\gamma_1, \gamma_2) \leq \gamma_2^{1/2},$$

*Proof.* The proof is postponed to Appendix B.5.  $\square$

## References

- [1] S. Allasonnière and E. Kuhn. Convergent stochastic expectation maximization algorithm with efficient sampling in high dimension. Application to deformable template model estimation. *Comput. Statist. Data Anal.*, 91:4–19, 2015.

- [2] Y. F. Atchadé, G. Fort, and E. Moulines. On perturbed proximal gradient algorithms. *J. Mach. Learn. Res.*, 18(1):310–342, 2017.
- [3] T. Aubin. *A course in differential geometry*. Graduate Studies in Mathematics. AMS, 2000.
- [4] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.
- [5] N. Brosse, A. Durmus, S. Meyn, and E. Moulines. Diffusion approximations and control variates for mcmc. *arXiv preprint arXiv:1808.01665*, 2018.
- [6] E. Buta and H. Doss. Computational approaches for empirical bayes methods and bayesian sensitivity analysis. *Ann. Statist.*, 39(5):2658–2685, 10 2011.
- [7] B. P. Carlin and T. A. Louis. Empirical Bayes: past, present and future. *J. Amer. Statist. Assoc.*, 95(452):1286–1289, 2000.
- [8] G. Casella. An introduction to empirical Bayes data analysis. *Amer. Statist.*, 39(2):83–87, 1985.
- [9] A. S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(3):651–676, 2017.
- [10] V. De Bortoli, A. Durmus, and E. Moulines. From continuous to discrete processes and back. *In preparation*, 2018.
- [11] S. Donnet, V. Rivoirard, J. Rousseau, and C. Scricciolo. Posterior concentration rates for empirical bayes procedures with applications to dirichlet process mixtures. *Bernoulli*, 24(1):231–256, 2018.
- [12] A. Durmus and E. Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.*, 27(3):1551–1587, 2017.
- [13] A. Durmus, E. Moulines, and E. Saksman. On the convergence of hamiltonian monte carlo. *arXiv preprint arXiv:1705.00166*, 2017.
- [14] A. Eberle. Reflection coupling and Wasserstein contractivity without convexity. *C. R. Math. Acad. Sci. Paris*, 349(19-20):1101–1104, 2011.
- [15] A. Eberle. Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields*, 166(3-4):851–886, 2016.
- [16] A. Eberle and M. B. Majka. Quantitative contraction rates for markov chains on general state spaces. *arXiv preprint arXiv:1808.07033*, 2018.
- [17] B. Efron and R. Tibshirani. Empirical bayes methods and false discovery rates for microarrays. *Genetic epidemiology*, 23(1):70–86, 2002.
- [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] S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.
- [20] G. Fort, E. Moulines, and P. Priouret. Convergence of adaptive and interacting Markov chain Monte Carlo algorithms. *Ann. Statist.*, 39(6):3262–3289, 2011.
- [21] E. I. George and D. P. Foster. Calibration and empirical bayes variable selection. *Biometrika*, 87(4):731–747, 2000.
- [22] U. Grenander and M. I. Miller. Representations of knowledge in complex systems. *J. Roy. Statist. Soc. Ser. B*, 56(4):549–603, 1994. With discussion and a reply by the authors.
- [23] P. L. C. H. H. Bauschke. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics. Springer-Verlag New York, 1 edition, 2011.
- [24] Y.-P. Hsieh, A. Kavis, P. Rolland, and V. Cevher. Mirrored langevin dynamics. In *Advances in Neural Information Processing Systems*, pages 2880–2889, 2018.
- [25] L. Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1-2):134–139, 1918.
- [26] W. E. Johnson, C. Li, and A. Rabinovic. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics*, 8(1):118–127, 2007.
- [27] O. Kallenberg. *Foundations of modern probability*. Springer Science & Business Media, 2006.
- [28] B. Knapik, B. Szabó, A. van der Vaart, and J. van Zanten. Bayes procedures for adaptive inference in inverse problems for the white noise model. *Probability Theory and Related Fields*, 164(3-4):771–813, 2016.
- [29] H. J. Kushner and G. G. Yin. *Stochastic approximation and recursive algorithms and applications*, volume 35 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.
- [30] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes. I. Criteria for discrete-time chains. *Adv. in Appl. Probab.*, 24(3):542–574, 1992.
- [31] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag London, Ltd., London, 1993.
- [32] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.*, 25(3):518–548, 1993.
- [33] R. M. Neal. Bayesian learning via stochastic dynamics. In *Advances in Neural Information Processing Systems 5, [NIPS Conference]*, pages 475–482, San Francisco, CA, USA, 1993. Morgan Kaufmann Publishers Inc.
- [34] Y. Nesterov. *Introductory lectures on convex optimization*, volume 87 of *Applied Optimization*. Kluwer Academic Publishers, Boston, MA, 2004. A basic course.



- [35] H. E. Ogden. A sequential reduction method for inference in generalized linear mixed models. *Electron. J. Stat.*, 9(1):135–152, 2015.
- [36] G. Parisi. Correlation functions and computer simulations. *Nuclear Physics B*, 180:378–384, 1981.
- [37] S. Petrone, J. Rousseau, and C. Scricciolo. Bayes and empirical Bayes: do they merge? *Biometrika*, 101(2):285–302, 2014.
- [38] G. Pólya and G. Szegő. *Problems and theorems in analysis. I*. Classics in Mathematics. Springer-Verlag, Berlin, 1998. Series, integral calculus, theory of functions, Translated from the German by Dorothee Aepli, Reprint of the 1978 English translation.
- [39] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.
- [40] H. Robbins. An empirical Bayes approach to statistics. In *Herbert Robbins Selected Papers*, pages 41–47. Springer, 1985.
- [41] H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Statistics*, 22:400–407, 1951.
- [42] C. Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, 2007.
- [43] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.
- [44] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.
- [45] J. Rousseau and B. Szabo. Asymptotic behaviour of the empirical Bayes posteriors associated to maximum marginal likelihood estimator. *Ann. Statist.*, 45(2):833–865, 2017.
- [46] J. Schelldorfer, L. Meier, and P. Bühlmann. GLMMLasso: an algorithm for high-dimensional generalized linear mixed models using  $\ell_1$ -penalization. *J. Comput. Graph. Statist.*, 23(2):460–477, 2014.
- [47] G. K. Smyth. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology*, 3(1):1–25, 2004.
- [48] D. W. Stroock and S. R. S. Varadhan. *Multidimensional diffusion processes*. Classics in Mathematics. Springer-Verlag, Berlin, 2006. Reprint of the 1997 edition.
- [49] M. Welling and Y. W. Teh. Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the International Conference on Machine Learning*, pages 681–688, 2011.
- [50] Z. Zhou, R. Leahy, and J. Qi. Approximate maximum likelihood hyperparameter estimation for Gibbs prior. *IEEE Trans. Image Process.*, 6(6):844–861, June 1997.

## A Non-convex objective function

We now turn to the case where  $f$  is non-convex. We recall that the normal space of a sub-manifold  $\mathcal{M} \subset \mathbb{R}^m$  at point  $x$  is given by

$$\mathbf{N}(x, \mathcal{M}) = \begin{cases} \mathbf{T}(x, \mathcal{M})^\perp & \text{if } x \in \mathcal{M}; \\ \{0\} & \text{otherwise,} \end{cases}$$

where  $\mathbf{T}(x, \mathcal{M})$  is the tangent space of the sub-manifold  $\mathcal{M}$  at point  $x$ , see [3].

**Theorem 9.** *Assume **A1** and that  $\partial\Theta$  is a  $\mathbb{R}^{m-1}$  connected manifold with continuously differentiable outer normal. In addition, assume that  $f \in C^1(\mathbb{R}^p, \mathbb{R})$ , **A3**. Let  $\bar{\gamma} > 0$ ,  $(\gamma_n)_{n \in \mathbb{N}}$ ,  $(\delta_n)_{n \in \mathbb{N}^*}$  be sequences of non-increasing positive real numbers and  $(m_n)_{n \in \mathbb{N}}$  be a sequence of positive integers such that  $\sup_{n \in \mathbb{N}} \delta_n < 1/L_f$ ,  $\sup_{n \in \mathbb{N}} \gamma_n < \bar{\gamma}$  and (14) are satisfied. Let  $\{(X_k^n)_{k \in \{0, \dots, m_n\}} : n \in \mathbb{N}\}$  be given by (8). Assume in addition **H1** holds. Then  $(\theta_n)_{n \in \mathbb{N}}$  defined by (9) converges a.s. to some  $\theta^* \in \{\theta \in \Theta : \nabla f(\theta) + \mathbf{n} = 0, \mathbf{n} \in \mathbf{N}(\theta, \partial\Theta)\}$ , where  $\mathbf{N}(\theta, \partial\Theta)$  is the normal space of  $\partial\Theta$  at point  $\theta$ .*

**Theorem 10.** *Assume **A1** and that  $\partial\Theta$  is a  $\mathbb{R}^{m-1}$  connected manifold with continuously differentiable outer normal. In addition, assume that  $f \in C^1(\mathbb{R}^p, \mathbb{R})$ , **A3** and **A4**. Let  $\bar{\gamma} > 0$ ,  $(\gamma_n)_{n \in \mathbb{N}}$ ,  $(\delta_n)_{n \in \mathbb{N}^*}$  be sequences of non-increasing positive real numbers and  $(m_n)_{n \in \mathbb{N}}$  be a sequence of positive integers such that  $\sup_{n \in \mathbb{N}} \delta_n < 1/L_f$ ,  $\sup_{n \in \mathbb{N}} \gamma_n < \bar{\gamma}$ ,  $\sup_{n \in \mathbb{N}} |\delta_{n+1} - \delta_n| \delta_n^{-2} < +\infty$ ,  $\sum_{n=0}^{+\infty} \delta_{n+1} = +\infty$  and (26) are satisfied. Let  $(\tilde{X}_n)_{n \in \mathbb{N}}$  be given by (23). Assume in addition that **H1-(ii)**, **H3**, **H4** and **H5** hold. Then  $(\tilde{\theta}_n)_{n \in \mathbb{N}}$  defined by (24) converges a.s. to some  $\theta^* \in \{\theta \in \Theta : \nabla f(\theta) + \mathbf{n} = 0, \mathbf{n} \in \mathbf{N}(\theta, \partial\Theta)\}$ , where  $\mathbf{N}(\theta, \partial\Theta)$  is the normal space of  $\partial\Theta$  at point  $\theta$ .*

*Proof.* The proof is a direct application of [29, Chapter 5, Theorem 2.3] using the decomposition of the error term considered in the proof of Theorem 1 and Theorem 4. Indeed we decompose the error term  $\eta_n$  defined by (15) as  $\eta_n = \delta M_n + B_n$ , where  $\delta M_n$  is a martingale increment. Then, we only need to show that the following sums converge

$$\sum_{k=0}^n \delta_{k+1}^2 \mathbb{E} [\|\delta M_k\|^2], \quad \sum_{k=0}^n \delta_{k+1} \mathbb{E} [\|B_k\|].$$

This is exactly what we show in Theorem 1 with  $\delta M_n = 0$  and  $B_n = \eta_n$  and in Theorem 4 with  $\delta M_n = \tilde{\eta}_n^a$  and  $B_n = \eta_n^b + \eta_n^c + \eta_n^d$  where  $\{\tilde{\eta}_n^i : n \in \mathbb{N}, i \in \{a, b, c, d\}\}$  is defined by (34).  $\square$

## B Postponed proofs

### B.1 Proof of Theorem 4

In all the proof we use the following lemma

**Lemma 11.** *Let  $t \in (0, 1)$  and  $\gamma \in (0, \bar{\gamma})$  with  $\bar{\gamma} > 0$  then  $\sum_{n \in \mathbb{N}} t^{n\gamma} \leq t^{-\bar{\gamma}} \log^{-1}(1/t) \gamma^{-1}$  and  $\sum_{n \in \mathbb{N}} n t^{n\gamma} \leq t^{-\bar{\gamma}} \log^{-2}(1/t) \gamma^{-2}$ .*

*Proof.* Let  $t \in (0, 1)$  and  $\gamma \in (0, \bar{\gamma})$  with  $\bar{\gamma} > 0$ . Using that  $e^u - 1 \leq ue^u$  for all  $u \geq 0$  we have

$$\sum_{n \in \mathbb{N}} t^{n\gamma} = -(t^\gamma - 1)^{-1} \leq -\gamma^{-1} \log^{-1}(t) \exp(-\log(t)\gamma) \leq t^{-\bar{\gamma}} \log^{-1}(1/t)\gamma^{-1},$$

and

$$\sum_{n \in \mathbb{N}} nt^{n\gamma} = t^\gamma (t^\gamma - 1)^{-2} \leq t^\gamma \{\gamma^{-1} \log^{-1}(t) \exp(-\log(t)\gamma)\}^2 \leq t^{-\bar{\gamma}} \log^{-2}(1/t)\gamma^{-2},$$

which completes the proof.  $\square$

Consider  $(\tilde{X}_k, \tilde{\theta}_k)_{k \in \mathbb{N}}$  defined by (23) and (24) and define for any  $n \in \mathbb{N}$ ,

$$\tilde{\eta}_n = H_{\tilde{\theta}_n}(\tilde{X}_{n+1}) - \nabla f(\tilde{\theta}_n) = H_{\tilde{\theta}_n}(\tilde{X}_{n+1}) - \pi_{\tilde{\theta}_n}(H_{\tilde{\theta}_n}).$$

In order to apply [2, Theorem 2] we show that the following sums converge a.s.

$$\sum_{k=0}^n \delta_{k+1} \langle \Pi_\Theta [\tilde{\theta}_k - \delta_k \nabla f(\tilde{\theta}_k)], \tilde{\eta}_k \rangle, \quad \sum_{k=0}^n \delta_{k+1} \tilde{\eta}_k, \quad \sum_{k=0}^n \delta_{k+1}^2 \|\tilde{\eta}_k\|^2. \quad (28)$$

Using  $\|x + y\|^2 \leq 2(\|x\|^2 + \|y\|^2)$  for any  $x, y \in \mathbb{R}^d$ , **A1**, **A2**, **A3**, **H3** and **A4**, we get for any  $k \in \mathbb{N}$ ,  $\mathbb{E}[\|\tilde{\eta}_k\|^2] \leq 2M_H^2 A_6 \mathbb{E}[V(\tilde{X}_0)] + 2 \sup_{\theta \in \Theta} \|\nabla f(\theta)\|^2 < +\infty$ , which implies that for any  $n \in \mathbb{N}$ ,

$$\mathbb{E} \left[ \sum_{k=0}^n \delta_{k+1}^2 \|\tilde{\eta}_k\|^2 \right] \leq \sup_{k \in \mathbb{N}} \mathbb{E} [\|\tilde{\eta}_k\|^2] \sum_{k=1}^n \delta_k^2 \leq \sup_{k \in \mathbb{N}} \mathbb{E} [\|\tilde{\eta}_k\|^2] \sum_{k=1}^{+\infty} \delta_k^2 < +\infty. \quad (29)$$

By **A4** and **H5**, for any  $\theta \in \Theta$  and  $\gamma \in (0, \bar{\gamma})$ , there exists a function  $\hat{H}_{\gamma, \theta} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  solution of the *Poisson equation*,

$$(\text{Id} - K_{\gamma, \theta}) \hat{H}_{\gamma, \theta} = H_\theta - \pi_{\gamma, \theta}(H_\theta) \quad (30)$$

defined for any  $x \in \mathbb{R}^d$  by

$$\hat{H}_{\gamma, \theta}(x) = \sum_{j \in \mathbb{N}} \{K_{\gamma, \theta}^j H_\theta(x) - \pi_{\gamma, \theta}(H_\theta)\}, \quad (31)$$

which satisfies using for all  $x \in \mathbb{R}^d$

$$\|\hat{H}_\theta(x)\| \leq C_P V^{1/4}(x), \quad (32)$$

where  $C_P = A_7 \log(1/\kappa)^{-1} \kappa^{-\bar{\gamma}} \gamma^{-1}$  by Lemma 11. Using (30), we consider the following decomposition: for any  $k \in \mathbb{N}$

$$\begin{aligned} \tilde{\eta}_k &= \hat{H}_{\tilde{\theta}_k}(\tilde{X}_{k+1}) - K_{\gamma_k, \tilde{\theta}_k} \hat{H}_{\tilde{\theta}_k}(\tilde{X}_{k+1}) + \pi_{\gamma_k, \tilde{\theta}_k}(H_{\tilde{\theta}_k}) - \pi_{\tilde{\theta}_k}(H_{\tilde{\theta}_k}) \\ &= \tilde{\eta}_k^a + \tilde{\eta}_k^b + \tilde{\eta}_k^c + \tilde{\eta}_k^d, \end{aligned} \quad (33)$$

where

$$\begin{cases} \tilde{\eta}_k^a = \hat{H}_{\tilde{\theta}_k}(\tilde{X}_{k+1}) - K_{\gamma_k, \tilde{\theta}_k} \hat{H}_{\tilde{\theta}_k}(\tilde{X}_k); \\ \tilde{\eta}_k^b = K_{\gamma_k, \tilde{\theta}_k} \hat{H}_{\tilde{\theta}_k}(\tilde{X}_k) - K_{\gamma_{k+1}, \tilde{\theta}_{k+1}} \hat{H}_{\tilde{\theta}_{k+1}}(\tilde{X}_{k+1}); \\ \tilde{\eta}_k^c = K_{\gamma_{k+1}, \tilde{\theta}_{k+1}} \hat{H}_{\tilde{\theta}_{k+1}}(\tilde{X}_{k+1}) - K_{\gamma_k, \tilde{\theta}_k} \hat{H}_{\tilde{\theta}_k}(\tilde{X}_{k+1}); \\ \tilde{\eta}_k^d = \pi_{\gamma_k, \tilde{\theta}_k}(H_{\tilde{\theta}_k}) - \pi_{\tilde{\theta}_k}(H_{\tilde{\theta}_k}). \end{cases} \quad (34)$$

Note that if a.s. the following sums converge

$$\sum_{k=0}^n \delta_{k+1} \langle \Pi_{\Theta} [\tilde{\theta}_k - \delta_k \nabla f(\tilde{\theta}_k)], \tilde{\eta}_k^i \rangle, \quad \sum_{k=0}^n \delta_{k+1} \tilde{\eta}_k^i,$$

where  $i \in \{a, b, c, d\}$ , then combining this result with (33) and (29) shows that (28) holds which proves the theorem.

(a) By definition (23) and (34), we have that  $(\tilde{\eta}_k^a)_{k \in \mathbb{N}}$  is a martingale increment with respect to  $(\tilde{\mathcal{F}}_j)_{j \in \mathbb{N}}$  defined by (25). Then, using (32), **A4** and **H3**, we obtain for any  $n \in \mathbb{N}$

$$\begin{aligned} \mathbb{E} \left[ \left\| \sum_{k=0}^n \delta_{k+1} \tilde{\eta}_k^a \right\|^2 \right] &= \sum_{k=0}^n \delta_{k+1}^2 \mathbb{E} [\|\tilde{\eta}_k^a\|^2] \\ &\leq \sum_{k=0}^n \delta_{k+1}^2 \mathbb{E} \left[ \mathbb{E} \left[ \|\hat{H}_{\tilde{\theta}_k}\|^2 (\tilde{X}_{k+1}) \middle| \mathcal{F}_k \right] \right] \leq C_a \sum_{k=0}^n \delta_{k+1}^2 \gamma_k^{-2}, \end{aligned}$$

with  $C_a = C_P A_5 \mathbb{E}[V(\tilde{X}_0)]$ . The martingale  $(\sum_{k=0}^n \delta_{k+1} \tilde{\eta}_k^a)_{n \in \mathbb{N}}$  is then bounded in  $L^2(\Omega, \mathbb{P})$  and thus a.s. converges. Similarly, the martingale  $(\sum_{k=0}^n \delta_{k+1} \langle \Pi_{\Theta} [\tilde{\theta}_k - \delta_k \nabla f(\tilde{\theta}_k)], \tilde{\eta}_k^a \rangle)_{n \in \mathbb{N}}$  is also bounded in  $L^2(\Omega, \mathbb{P})$  and converges a.s..

(b) Regarding  $(\tilde{\eta}_k^b)_{k \in \mathbb{N}}$ , using (34), and a discrete integration formula, we obtain for any  $n \in \mathbb{N}$ ,

$$\begin{aligned} \sum_{k=0}^n \delta_{k+1} \tilde{\eta}_k^b &= \sum_{k=0}^n \delta_{k+1} \left( K_{\gamma_k, \tilde{\theta}_k} \hat{H}_{\tilde{\theta}_k}(\tilde{X}_k) - K_{\gamma_{k+1}, \tilde{\theta}_{k+1}} \hat{H}_{\tilde{\theta}_{k+1}}(\tilde{X}_{k+1}) \right) \\ &= \sum_{k=1}^n (\delta_{k+1} - \delta_k) K_{\gamma_k, \tilde{\theta}_k} \hat{H}_{\tilde{\theta}_k}(\tilde{X}_k) - \delta_{n+1} K_{\gamma_{n+1}, \tilde{\theta}_{n+1}} \hat{H}_{\tilde{\theta}_{n+1}}(\tilde{X}_{n+1}) + \delta_1 K_{\gamma_0, \tilde{\theta}_0} \hat{H}_{\tilde{\theta}_0}(\tilde{X}_0). \end{aligned}$$

We have using (32), **A4** and **H3** for any  $n \in \mathbb{N}$

$$\sum_{k=1}^n |\delta_{k+1} - \delta_k| \mathbb{E} \left[ K_{\gamma_k, \tilde{\theta}_k} \|\hat{H}_{\tilde{\theta}_k}\|(\tilde{X}_k) \right] \leq C_b \sum_{k=1}^n |\delta_{k+1} - \delta_k| \gamma_k^{-1},$$

where  $C_b = C_P A_5 \mathbb{E}[V(\tilde{X}_0)]$ . Since by (26) and  $\sup_{k \in \mathbb{N}} \gamma_k < \bar{\gamma}$ ,  $\sum_{k=0}^{+\infty} \delta_k^2 / \gamma_k < +\infty$ , then the condition  $\sup_{k \in \mathbb{N}} |\delta_{k+1} - \delta_k| \delta_k^{-2} < +\infty$  implies

$$\sum_{k=1}^{+\infty} |\delta_{k+1} - \delta_k| \mathbb{E} \left[ K_{\gamma_k, \tilde{\theta}_k} \|\hat{H}_{\tilde{\theta}_k}\|(\tilde{X}_k) \right] \leq C_b \sup_{k \in \mathbb{N}} \{|\delta_{k+1} - \delta_k| \delta_k^{-2}\} \sum_{k \in \mathbb{N}^*} \delta_k^2 \gamma_k^{-1} < +\infty$$

Therefore,  $\sum_{k \in \mathbb{N}^*} (\delta_k - \delta_{k+1}) K_{\gamma_k, \tilde{\theta}_k} \hat{H}_{\tilde{\theta}_k}(\tilde{X}_k)$  is a.s. absolutely convergent. On the other hand, we have that for any  $n \in \mathbb{N}$ ,  $K_{\gamma_{n+1}, \tilde{\theta}_{n+1}} \|\hat{H}_{\tilde{\theta}_{n+1}}\|(\tilde{X}_{n+1})$  is bounded by  $\gamma_{n+1}^{-1} M_H K_{\gamma_{n+1}, \tilde{\theta}_{n+1}} V^{1/2}(\tilde{X}_{n+1})$ . It follows that for any  $\varepsilon > 0$ , using the Markov inequality and **H3**, we get

$$\begin{aligned} \sum_{n \in \mathbb{N}} \mathbb{P} \left( \delta_n \gamma_n^{-1} K_{\gamma_{n+1}, \tilde{\theta}_{n+1}} \|\hat{H}_{\tilde{\theta}_{n+1}}\|(\tilde{X}_{n+1}) \geq \varepsilon \right) &\leq \sum_{n \in \mathbb{N}} \mathbb{P} \left( \delta_n \gamma_n^{-1} K_{\gamma_{n+1}, \tilde{\theta}_n} V^{1/2}(\tilde{X}_{n+1}) \geq \varepsilon \right) \\ &\leq \varepsilon^{-2} \sum_{n \in \mathbb{N}} \delta_n^2 \gamma_n^{-2} M_H A_5 \mathbb{E} [V(\tilde{X}_0)] < +\infty, \end{aligned}$$

by (26). Using the Borel-Cantelli lemma we get that a.s.  $\lim_{n \rightarrow +\infty} \delta_n K_{\gamma_n, \tilde{\theta}_n} \hat{H}_{\tilde{\theta}_n}(\tilde{X}_n) = 0$ . This completes the proof of the convergence of  $\sum_{k \in \mathbb{N}} \delta_{k+1} \tilde{\eta}_k^b$ .

We conclude similarly for the convergence of  $\sum_{k \in \mathbb{N}} \delta_{k+1} \langle \Pi_{\Theta} [\tilde{\theta}_k - \delta_{k+1} \nabla f(\tilde{\theta}_k)], \tilde{\eta}_k^b \rangle$ , upon verifying that for any  $k \in \mathbb{N}$ ,  $a_{k+1} = \delta_{k+1} \Pi_{\Theta} [\vartheta_{k+1}]$  where  $\vartheta_{k+1} = \tilde{\theta}_k - \delta_{k+1} \nabla f(\tilde{\theta}_k)$  satisfies that there exists  $A \geq 0$  such that for any  $k \in \mathbb{N}$ ,

$$\|a_{k+1} - a_k\| \leq A\{|\delta_{k+1} - \delta_k|(1 + \delta_k) + \delta_k^2\}.$$

Indeed, using **A1-A2** and that  $\Pi_{\Theta}$  is non-expansive by **A1** and [23, Proposition 4.8], we get for any  $k \in \mathbb{N}$ ,

$$\begin{aligned} \|a_{k+1} - a_k\| &\leq |\delta_{k+1} - \delta_k| \|\Pi_{\Theta}(\vartheta_{k+1})\| + \delta_k \|\Pi_{\Theta}(\vartheta_{k+1}) - \Pi_{\Theta}(\tilde{\theta}_k)\| \\ &\leq M_{\Theta} |\delta_{k+1} - \delta_k| + \delta_k \|\vartheta_{k+1} - \tilde{\theta}_k\| \leq M_{\Theta} |\delta_{k+1} - \delta_k| + \delta_k \delta_{k+1} \sup_{\Theta} \|\nabla f\|. \end{aligned}$$

(c) We now turn to  $(\tilde{\eta}_k^{\ell})_{k \in \mathbb{N}}$ . We start by giving an upper-bound on  $\|\pi_{\gamma_1, \theta_1} - \pi_{\gamma_2, \theta_2}\|_{V^{1/2}}$  for  $\gamma_1, \gamma_2 \in (0, \bar{\gamma})$  with  $\gamma_1 > \gamma_2$  and,  $\theta_1, \theta_2 \in \Theta$ . Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a measurable function satisfying  $\sup_{x \in \mathbb{R}^d} \{|f|/V^{1/2}\} \leq 1$ . Using **H5**, **H4**, **H3** and Lemma **11**, we get that for any  $\gamma_1, \gamma_2 \in (0, \bar{\gamma})$  with  $\gamma_1 > \gamma_2$ ,  $\theta_1, \theta_2 \in \Theta$ ,  $x \in \mathbb{R}^d$  and  $\ell \in \mathbb{N}^*$

$$\begin{aligned} |K_{\gamma_1, \theta_1}^{\ell} f(x) - K_{\gamma_2, \theta_2}^{\ell} f(x)| &= \left| \sum_{j=0}^{\ell} K_{\gamma_1, \theta_1}^j (K_{\gamma_1, \theta_1} - K_{\gamma_2, \theta_2}) \left\{ K_{\gamma_2, \theta_2}^{(\ell-j)} f(x) - \pi_{\gamma_2, \theta_2}(f) \right\} \right| \\ &\leq A_7 \sum_{j=0}^{\ell} \kappa^{(\ell-j)\gamma_2} \left| K_{\gamma_1, \theta_1}^j (K_{\gamma_1, \theta_1} - K_{\gamma_2, \theta_2}) V^{1/2}(x) \right| \\ &\leq A_7 A_6 \sum_{j=0}^{\ell} \kappa^{(\ell-j)\gamma_2} [\Lambda_1(\gamma_1, \gamma_2) + \Lambda_2(\gamma_1, \gamma_2) \|\theta_1 - \theta_2\|] \sup_{k \in \mathbb{N}} K_{\gamma_1, \theta_1}^k V(x) \\ &\leq A_7 A_6 A_5 \log(1/\kappa)^{-1} \kappa^{-\bar{\gamma}} \gamma_2^{-1} [\Lambda_1(\gamma_1, \gamma_2) + \Lambda_2(\gamma_1, \gamma_2) \|\theta_1 - \theta_2\|] V(x). \end{aligned}$$

Taking  $\ell \rightarrow +\infty$  and using **H5**, we obtain that for any  $\theta_1, \theta_2 \in \Theta$  and  $\gamma_1, \gamma_2 \in (0, \bar{\gamma})$  with  $\gamma_1 > \gamma_2$ ,

$$\|\pi_{\gamma_1, \theta_1} - \pi_{\gamma_2, \theta_2}\|_V \leq C_{\pi} \gamma_2^{-1} [\Lambda_1(\gamma_1, \gamma_2) + \Lambda_2(\gamma_1, \gamma_2) \|\theta_1 - \theta_2\|], \quad (35)$$

with  $C_{\pi} = A_7 A_6 A_5 \log(1/\kappa)^{-1} \kappa^{-\bar{\gamma}} V(0)$ . By (31), (32) and **H5**, we have for any  $\theta_1, \theta_2 \in \Theta$ ,  $\gamma_1, \gamma_2 \in (0, \bar{\gamma})$  with  $\gamma_1 > \gamma_2$  and  $x \in \mathbb{R}^d$ ,

$$\begin{aligned} &\left\| K_{\gamma_1, \theta_1} \hat{H}_{\gamma, \theta_1}(x) - K_{\gamma_2, \theta_2} \hat{H}_{\gamma, \theta_2}(x) \right\| \\ &= \left\| \sum_{\ell \in \mathbb{N}^*} \{K_{\gamma_1, \theta_1}^{\ell} H_{\theta_1}(x) - \pi_{\gamma_1, \theta_1}(H_{\theta_1})\} - \sum_{\ell \in \mathbb{N}^*} \{K_{\gamma_2, \theta_2}^{\ell} H_{\theta_2}(x) - \pi_{\gamma_2, \theta_2}(H_{\theta_2})\} \right\| \\ &\leq \sum_{\ell \in \mathbb{N}^*} \left\| \{K_{\gamma_1, \theta_1}^{\ell} H_{\theta_1}(x) - \pi_{\gamma_1, \theta_1}(H_{\theta_1})\} - \{K_{\gamma_2, \theta_2}^{\ell} H_{\theta_2}(x) - \pi_{\gamma_2, \theta_2}(H_{\theta_2})\} \right\|. \end{aligned}$$

We bound now each term of the sum in the right hand side. Note first that for any bounded measurable functions  $f_1, f_2$  from  $\mathbb{R}^d$  to  $\mathbb{R}$ ,  $\theta_1, \theta_2 \in \Theta$ ,  $\gamma_1, \gamma_2 \in (0, \bar{\gamma})$  with  $\gamma_1 > \gamma_2$ ,  $x \in \mathbb{R}^d$  and

$\ell \in \mathbb{N}^*$ , it holds that

$$\begin{aligned}
K_{\gamma_1, \theta_1}^\ell f_1(x) - K_{\gamma_2, \theta_2}^\ell f_2(x) &= K_{\gamma_1, \theta_1}^\ell f_1(x) - K_{\gamma_2, \theta_2}^\ell f_1(x) + K_{\gamma_2, \theta_2}^\ell (f_1(x) - f_2(x)) \\
&= \sum_{j=0}^{\ell-1} \left\{ K_{\gamma_1, \theta_1}^j - \pi_{\gamma_1, \theta_1} \right\} (K_{\gamma_1, \theta_1} - K_{\gamma_2, \theta_2}) \left\{ K_{\gamma_2, \theta_2}^{\ell-1-j} f_1(x) - \pi_{\gamma_2, \theta_2}(f_1) \right\} \\
&\quad + \sum_{j=0}^{\ell-1} \pi_{\gamma_1, \theta_1} \left\{ K_{\gamma_2, \theta_2}^{\ell-1-j} f_1(x) - K_{\gamma_2, \theta_2}^{\ell-j} f_1(x) \right\} + K_{\gamma_2, \theta_2}^\ell (f_1(x) - f_2(x)) \\
&= \sum_{j=0}^{\ell-1} \left\{ K_{\gamma_1, \theta_1}^j - \pi_{\gamma_1, \theta_1} \right\} (K_{\gamma_1, \theta_1} - K_{\gamma_2, \theta_2}) \left\{ K_{\gamma_2, \theta_2}^{\ell-1-j} f_1(x) - \pi_{\gamma_2, \theta_2}(f_1) \right\} \\
&\quad - \pi_{\gamma_1, \theta_1} (K_{\gamma_2, \theta_2}^\ell f_1(x) - f_1(x)) + K_{\gamma_2, \theta_2}^\ell (f_1(x) - f_2(x)).
\end{aligned}$$

Setting  $\tilde{H}_{\theta_1} = H_{\theta_1} - \pi_{\gamma_1, \theta_1}(H_{\theta_1})$  and  $\tilde{H}_{\theta_2} = H_{\theta_2} - \pi_{\gamma_2, \theta_2}(H_{\theta_2})$ , we obtain that

$$\begin{aligned}
K_{\gamma_1, \theta_1}^\ell \tilde{H}_{\theta_1}(x) - K_{\gamma_2, \theta_2}^\ell \tilde{H}_{\theta_2}(x) \\
= \sum_{j=0}^{\ell-1} \left\{ K_{\gamma_1, \theta_1}^j - \pi_{\gamma_1, \theta_1} \right\} (K_{\gamma_1, \theta_1} - K_{\gamma_2, \theta_2}) \left\{ K_{\gamma_2, \theta_2}^{\ell-1-j} \tilde{H}_{\theta_1}(x) - \pi_{\gamma_2, \theta_2}(\tilde{H}_{\theta_1}) \right\} + \Xi, \quad (36)
\end{aligned}$$

where

$$\begin{aligned}
\Xi &= -\pi_{\gamma_1, \theta_1} (K_{\gamma_2, \theta_2}^\ell H_{\theta_1}(x) - H_{\theta_1}(x)) + K_{\gamma_2, \theta_2}^\ell [H_{\theta_1}(x) - H_{\theta_2}(x) + \pi_{\gamma_2, \theta_2}(H_{\theta_2}) - \pi_{\gamma_1, \theta_1}(H_{\theta_1})] \\
&= -\pi_{\gamma_1, \theta_1} K_{\gamma_2, \theta_2}^\ell H_{\theta_1}(x) + K_{\gamma_2, \theta_2}^\ell [H_{\theta_1}(x) - H_{\theta_2}(x) + \pi_{\gamma_2, \theta_2}(H_{\theta_2})] \\
&= (\pi_{\gamma_2, \theta_2} - \pi_{\gamma_1, \theta_1}) (K_{\gamma_2, \theta_2}^\ell H_{\theta_1}(x) - \pi_{\gamma_2, \theta_2}(H_{\theta_1})) - \pi_{\gamma_2, \theta_2}(H_{\theta_1}) \\
&\quad + K_{\gamma_2, \theta_2}^\ell [H_{\theta_1}(x) - H_{\theta_2}(x) + \pi_{\gamma_2, \theta_2}(H_{\theta_2})] \\
&= (\pi_{\gamma_2, \theta_2} - \pi_{\gamma_1, \theta_1}) (K_{\gamma_2, \theta_2}^\ell H_{\theta_1}(x) - \pi_{\gamma_2, \theta_2}(H_{\theta_1})) + K_{\gamma_2, \theta_2}^\ell (H_{\theta_1} - H_{\theta_2})(x) - \pi_{\gamma_2, \theta_2}(H_{\theta_1} - H_{\theta_2}). \quad (37)
\end{aligned}$$

For the first term in the decomposition (36), using **A4**, **H5** and **H4**, we obtain for any  $\theta_1, \theta_2 \in \Theta$ ,  $\gamma_1, \gamma_2 \in (0, \bar{\gamma})$  with  $\gamma_1 > \gamma_2$ ,  $x \in \mathbb{R}^d$  and  $\ell \in \mathbb{N}^*$

$$\begin{aligned}
&\left\| \sum_{j=0}^{\ell-1} \left\{ K_{\gamma_1, \theta_1}^j - \pi_{\gamma_1, \theta_1} \right\} (K_{\gamma_1, \theta_1} - K_{\gamma_2, \theta_2}) \left\{ K_{\gamma_2, \theta_2}^{\ell-1-j} H_{\theta_1}(x) - \pi_{\gamma_2, \theta_2}(H_{\theta_1}) \right\} \right\| \\
&\leq A_7 M_H \sum_{j=0}^{\ell-1} \left| \left\{ K_{\gamma_1, \theta_1}^j - \pi_{\gamma_1, \theta_1} \right\} (K_{\gamma_1, \theta_1} - K_{\gamma_2, \theta_2}) V^{1/4}(x) \right| \\
&\leq A_7^2 A_6 M_H [\Lambda_1(\gamma_1, \gamma_2) + \Lambda_2(\gamma_1, \gamma_2) \|\theta_1 - \theta_2\|] \sum_{j=0}^{\ell-1} \kappa^j \gamma \kappa^{(\ell-1-j)\gamma_2} V^{1/2}(x) \\
&\leq A_7^2 A_6 M_H [\Lambda_1(\gamma_1, \gamma_2) + \Lambda_2(\gamma_1, \gamma_2) \|\theta_1 - \theta_2\|] \ell \kappa^{\ell\gamma_2} V^{1/2}(x). \quad (38)
\end{aligned}$$

For the first term in (37), using **A4**, **H5** and (35) we obtain for any  $\theta_1, \theta_2 \in \Theta$ ,  $\gamma_1, \gamma_2 \in (0, \bar{\gamma})$  with  $\gamma_1 > \gamma_2$ ,  $x \in \mathbb{R}^d$  and  $\ell \in \mathbb{N}^*$

$$\begin{aligned} \left\| (\pi_{\gamma_1, \theta_1} - \pi_{\gamma_2, \theta_2})(K_{\gamma_2, \theta_2}^\ell H_{\theta_1}(x) - \pi_{\gamma_2, \theta_2}(H_{\theta_1})) \right\| &\leq A_7 M_H \kappa^{\ell \gamma_2} \|\pi_{\gamma_1, \theta_1} - \pi_{\gamma_2, \theta_2}\|_{V^{1/2}} \\ &\leq A_7 M_H C_\pi \kappa^{\ell \gamma_2} \gamma_2^{-1} [\Lambda_1(\gamma_1, \gamma_2) + \Lambda_2(\gamma_1, \gamma_2) \|\theta_1 - \theta_2\|] . \end{aligned} \quad (39)$$

For the second term in (37), using **A4** and **H5**, we obtain for any  $\theta_1, \theta_2 \in \Theta$ ,  $\gamma_1, \gamma_2 \in (0, \bar{\gamma})$  with  $\gamma_1 > \gamma_2$ ,  $x \in \mathbb{R}^d$  and  $\ell \in \mathbb{N}^*$

$$\left\| K_{\gamma_2, \theta_2}^\ell (H_{\theta_1} - H_{\theta_2})(x) - \pi_{\gamma_2, \theta_2}(H_{\theta_1} - H_{\theta_2}) \right\| \leq A_7 L_H \kappa^{\ell \gamma_2} \|\theta_1 - \theta_2\| V^{1/2}(x) . \quad (40)$$

Combining (38), (39), (40), (37) in (36), and using Lemma 11, we obtain that for any  $\theta_1, \theta_2 \in \Theta$ ,  $\gamma_1, \gamma_2 \in (0, \bar{\gamma})$  with  $\gamma_1 > \gamma_2$ ,  $x \in \mathbb{R}^d$  that

$$\begin{aligned} \left\| K_{\gamma_1, \theta_1} \hat{H}_{\gamma_1, \theta_1}(x) - K_{\gamma_2, \theta_2} \hat{H}_{\gamma_2, \theta_2}(x) \right\| \\ \leq C_c \gamma_2^{-1} [\gamma_2^{-1} \{\Lambda_1(\gamma_1, \gamma_2) + \Lambda_2(\gamma_1, \gamma_2) \|\theta_1 - \theta_2\|\} + \|\theta_1 - \theta_2\|] V^{1/2}(x) , \end{aligned}$$

with

$$C_c = A_7(L_H + C_\pi M_H) \log(1/\kappa)^{-1} \kappa^{-\bar{\gamma}} + A_7^2 A_6 M_H \log(1/\kappa)^{-2} \kappa^{-\bar{\gamma}} .$$

Combining this result and that for any  $k \in \mathbb{N}$ ,  $\|\tilde{\theta}_{k+1} - \tilde{\theta}_k\| \leq \delta_{k+1} M_H V^{1/2}(\tilde{X}_{k+1})$  using (9), **A4** and  $\Pi_\Theta$  is non-expansive, we get that for any  $k \in \mathbb{N}$ ,

$$\begin{aligned} \left\| K_{\gamma_k, \tilde{\theta}_k} \hat{H}_{\gamma_k, \tilde{\theta}_k}(x) - K_{\gamma_{k+1}, \tilde{\theta}_{k+1}} \hat{H}_{\gamma_{k+1}, \tilde{\theta}_{k+1}}(x) \right\| \\ \leq C_c (1 + M_H) \gamma_{k+1}^{-1} [\gamma_{k+1}^{-1} \{\Lambda_1(\gamma_k, \gamma_{k+1}) + \Lambda_2(\gamma_k, \gamma_{k+1}) \delta_{k+1}\} + \delta_{k+1}] V(\tilde{X}_{k+1}) . \end{aligned}$$

Then since  $\sup_{n \in \mathbb{N}} \mathbb{E}[V(\tilde{X}_n)] < +\infty$  by **H3**, (34) and (26), we get

$$\sum_{k \in \mathbb{N}} \delta_{k+1} \mathbb{E}[\|\tilde{\eta}_k^c\|] < +\infty .$$

(d) By a straightforward application of **H1-(ii)**, **H5** and (26), we conclude that  $\sum_{k \in \mathbb{N}} \delta_{k+1} \|\tilde{\eta}_k^d\|$  converges.

## B.2 Proofs of Theorem 5

In this section, we give the proof of Theorem 5 by showing that **H1** and **H2** hold. However, first of all, we establish stability results uniform in the parameter  $\theta \in \Theta$  for Langevin diffusion (5) and the associated Euler-Maruyama discretization (6) based on Foster-Lyapunov drift condition with constants independent of  $\theta$ .

Under **L1**, for any  $\theta \in \Theta$ , (5) defines a Markov semi-group  $(P_{t, \theta})_{t \geq 0}$  for any  $x \in \mathbb{R}^d$  and  $A \in \mathcal{B}(\mathbb{R}^d)$  by  $P_{t, \theta}(x, A) = \mathbb{P}(Y_t^\theta \in A)$  where  $(Y_t^\theta)_{t \geq 0}$  is the solution of (5) with  $Y_0^\theta = x$ . Consider now the generator of  $(P_{t, \theta})_{t \geq 0}$  for any  $\theta \in \Theta$ , defined for any  $f \in C^2(\mathbb{R}^d)$  by

$$\mathcal{A}_\theta f = \langle \nabla_x f, \nabla_x U_\theta(x) \rangle + \Delta_x f .$$

We say that a Markov kernel  $R$  on  $\mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d)$  satisfies a discrete Foster-Lyapunov drift condition  $\mathbf{D}_d(V, \lambda, b)$  if there exist  $\lambda \in (0, 1)$ ,  $b \geq 0$  and a measurable  $V : \mathbb{R}^d \rightarrow [1, +\infty)$  such that for all  $x \in \mathbb{R}^d$

$$RV(x) \leq \lambda V(x) + b.$$

We say that a Markov semi-group  $(P_t)_{t \geq 0}$  on  $\mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d)$  with extended infinitesimal generator  $(\mathcal{A}, D(\mathcal{A}))$  (see e.g. [32] for the definition of  $(\mathcal{A}, D(\mathcal{A}))$ ) satisfies a continuous drift condition  $\mathbf{D}_c(V, \zeta, \beta)$  if there exist  $\zeta > 0$ ,  $\beta \geq 0$  and a measurable function  $V : \mathbb{R}^d \rightarrow [1, +\infty)$  with  $V \in D(\mathcal{A})$  such that for all  $x \in \mathbb{R}^d$

$$\mathcal{A}V(x) \leq -\zeta V(x) + \beta.$$

### B.2.1 Foster-Lyapunov drift conditions uniform on $\theta$

**Proposition 12.** *Assume L2. Let  $\bar{\gamma} < \min(1, 2m)$  and define  $V : \mathbb{R}^d \rightarrow [1, +\infty)$  such that for all  $x \in \mathbb{R}^d$ ,  $V(x) = \exp(\tilde{\kappa}\phi(x)/4)$  with  $\phi(x) = \sqrt{\|x\|^2 + 1}$  and  $\tilde{\kappa} = \eta/4$ . Then there exist  $\lambda \in (0, 1)$  and  $b \geq 0$  such that for all  $\gamma \in (0, \bar{\gamma})$  and  $\theta \in \Theta$  the Markov kernel  $R_{\gamma, \theta}$  satisfies the discrete drift condition  $\mathbf{D}_d(V, \lambda^\gamma, b\gamma)$ , i.e. for all  $x \in \mathbb{R}^d$*

$$R_{\gamma, \theta}V(x) \leq \lambda^\gamma V(x) + b\gamma,$$

with

$$\lambda = e^{-2^{-4}\eta^2(2^{1/2}-1)}, \quad b = \tilde{\kappa}(d+c+2^{1/2}\tilde{\kappa}) \exp[\tilde{\kappa}\{(d+c+\tilde{\kappa})\bar{\gamma} + \phi(\max(1, 2(d+c)/\eta, M_\eta))\}].$$

*Proof.* Since  $\phi$  is 1-Lipschitz, by the log-Sobolev inequality [4, Proposition 5.4.1] and the concavity of  $x \mapsto x^{1/2}$ , we have for any  $x \in \mathbb{R}^d$

$$R_{\gamma, \theta}V(x) \leq e^{\tilde{\kappa}R_{\gamma, \theta}\phi(x) + \tilde{\kappa}^2\gamma} \leq e^{\tilde{\kappa}\sqrt{\|x - \gamma\nabla_x U_\theta(x)\|^2 + 2\gamma d + 1} + \tilde{\kappa}^2\gamma}.$$

Using L2 and  $\gamma < 2m$  we obtain that for any  $x \in \mathbb{R}^d$ ,

$$\begin{aligned} \|x - \gamma\nabla_x U_\theta(x)\|^2 &\leq \|x\|^2 - 2\gamma\langle x, \nabla_x U_\theta(x) \rangle + \gamma^2\|\nabla_x U_\theta(x)\|^2 \\ &\leq \|x\|^2 - 2\eta\gamma\|x\|\mathbb{1}_{\|x\| \geq M_\eta} + \gamma(\gamma - 2m)\|\nabla_x U_\theta(x)\|^2 + 2\gamma c \\ &\leq \|x\|^2 - 2\eta\gamma\|x\|\mathbb{1}_{\|x\| \geq M_\eta} + 2\gamma c. \end{aligned}$$

Therefore, for any  $x \in \mathbb{R}^d$ ,  $\|x\| \geq \max(1, 2(d+c)/\eta, M_\eta)$ , using for any  $a > 0$ ,  $\sqrt{1+a} - 1 \leq a/2$ , we get for any  $\theta \in \Theta$

$$\begin{aligned} &\sqrt{\|x - \gamma\nabla_x U_\theta(x)\|^2 + 2\gamma d + 1} - \phi(x) \\ &\leq \phi(x) \left\{ \sqrt{\phi^{-2}(x)(\|x\|^2 + 1 - 2\eta\gamma\|x\| + 2\gamma(d+c))} - 1 \right\} \\ &\leq \phi(x) \left\{ \sqrt{1 + 2\gamma\phi^{-2}(x)(d+c - \eta\|x\|)} - 1 \right\} \\ &\leq \gamma\phi^{-1}(x)(d+c - \eta\|x\|) \leq -\eta\gamma\phi^{-1}(x)\|x\|/2 \leq -2^{-3/2}\eta\gamma. \end{aligned} \quad (41)$$

This gives  $R_{\gamma, \theta}V(x) \leq e^{-2^{-4}\eta^2(2^{1/2}-1)\gamma} = \lambda^\gamma$ . For all  $x \in \mathbb{R}^d$  we obtain with the same arguments

$$\sqrt{\|x - \gamma\nabla_x U_\theta(x)\|^2 + 2\gamma d + 1} - \phi(x) \leq \gamma(d+c), \quad (42)$$



and consequently, combining (42) and (41)

$$R_{\gamma,\theta}V(x) \leq \lambda^\gamma V(x) + \left( e^{\tilde{\kappa}(d+c+\tilde{\kappa})\gamma} - \lambda^\gamma \right) e^{\tilde{\kappa}\phi(\max(1,2(d+c)/\eta, M_\eta))} \mathbb{1}_{\|x\| \leq \max(1,2(d+c)/\eta, M_\eta)} .$$

Using  $e^t - 1 \leq te^t$  for  $t \geq 0$  we obtain the result.  $\square$

**Proposition 13.** *Let  $\bar{\gamma} < \min(1, \eta/L^2)$  and define  $V : \mathbb{R}^d \rightarrow [1, +\infty)$  such that for all  $x \in \mathbb{R}^d$ ,  $V(x) = 1 + \|x\|^2$ . Assume **L1** and **L3**. Then there exist  $\lambda \in (0, 1)$  and  $b \geq 0$  such that for all  $\gamma \in (0, \bar{\gamma})$  and  $\theta \in \Theta$  the Markov kernel  $R_{\gamma,\theta}$  satisfies the discrete drift condition  $\mathbf{D}_d(V, \lambda^\gamma, b\gamma)$ , i.e. for all  $x \in \mathbb{R}^d$*

$$R_{\gamma,\theta}V(x) \leq \lambda^\gamma V(x) + b\gamma ,$$

with  $\lambda = \exp(-2(\eta - \bar{\gamma}L^2))$  and  $b = 2(c + d + \eta(M_\eta^2 + 1) + \bar{\gamma} \sup_{\theta \in \Theta} \|\nabla_x U_\theta(0)\|^2)$ .

*Proof.* Set  $Z$  a standard Gaussian random variable over  $\mathbb{R}^d$  we have using **L1** and  $(a+b)^2 \leq 2(a^2+b^2)$

$$\begin{aligned} R_{\gamma,\theta}V(x) &= 1 + \mathbb{E}(\|x - \gamma \nabla_x U_\theta(x) + \sqrt{2\gamma}Z\|^2) \\ &\leq 1 + \|x\|^2 - 2\gamma \langle x, \nabla_x U_\theta(x) \rangle + 2\gamma d + \gamma^2 \|\nabla_x U_\theta(x)\|^2 \\ &\leq V(x) + 2\gamma c - 2\gamma \eta \|x\|^2 \mathbb{1}_{\mathbb{B}(0, M_\eta)^c}(x) + 2\gamma d + 2\gamma^2 L^2 \|x\|^2 + 2\gamma^2 \sup_{\theta \in \Theta} \|\nabla_x U_\theta(0)\|^2 \\ &\leq (1 - 2\gamma(\eta - \gamma L^2))V(x) + 2\gamma(c + d + \eta(M_\eta^2 + 1) + \gamma \sup_{\theta \in \Theta} \|\nabla_x U_\theta(0)\|^2) , \end{aligned}$$

where  $Z$  is a zero mean  $d$ -dimensional Gaussian random variable with covariance identity.  $\square$

**Proposition 14.** *Let  $\bar{\gamma} < \min(1, \eta/L^2)$  and define  $V : \mathbb{R}^d \rightarrow [1, +\infty)$  such that for all  $x \in \mathbb{R}^d$ ,  $V(x) = 1 + \|x\|^4$ . Assume **L1** and **L3**. Then there exist  $\lambda \in (0, 1)$  and  $b \geq 0$  such that for all  $\gamma \in (0, \bar{\gamma})$  and  $\theta \in \Theta$  the Markov kernel  $R_{\gamma,\theta}$  satisfies the discrete drift condition  $\mathbf{D}_d(V, \lambda^\gamma, b\gamma)$ , i.e. for all  $x \in \mathbb{R}^d$*

$$R_{\gamma,\theta}V(x) \leq \lambda^\gamma V(x) + b\gamma ,$$

*Proof.* Let  $a, b \in \mathbb{R}^d$ . We have the following equality

$$\|a + b\|^2 = \|a\|^4 + \|b\|^4 + 4\langle a, b \rangle^2 + 2\|a\|^2\|b\|^2 + 4\langle a, b \rangle\|a\|^2 + 4\langle a, b \rangle\|b\|^2 . \quad (43)$$

Let  $a = x$  and  $b = -\gamma \nabla_x U_\theta(x) + \sqrt{2\gamma}Z$  where  $Z$  is a zero mean  $d$ -dimensional Gaussian random variable with covariance identity. Using (43) we get for any  $\theta \in \Theta$  and  $\gamma > 0$

$$\begin{aligned} R_{\gamma,\theta}V(x) &= 1 + \|x\|^4 + \mathbb{E} \left[ \left\| -\gamma \nabla_x U_\theta(x) + \sqrt{2\gamma}Z \right\|^4 \right] + 4\mathbb{E} \left[ \langle x, -\gamma \nabla_x U_\theta(x) + \sqrt{2\gamma}Z \rangle^2 \right] \\ &\quad + 2\|x\|^2 \mathbb{E} \left[ \left\| -\gamma \nabla_x U_\theta(x) + \sqrt{2\gamma}Z \right\|^2 \right] + 4\mathbb{E} \left[ \langle x, -\gamma \nabla_x U_\theta(x) + \sqrt{2\gamma}Z \rangle \left\| -\gamma \nabla_x U_\theta(x) + \sqrt{2\gamma}Z \right\|^2 \right] \\ &\quad + 4\mathbb{E} \left[ \langle x, -\gamma \nabla_x U_\theta(x) + \sqrt{2\gamma}Z \rangle \|x\|^2 \right] . \end{aligned}$$

Let  $n(x) = \|\nabla_x U_\theta(x)\|$  and  $r(x) = \langle \nabla_x U_\theta(x), x \rangle$ . Using the Isserlis formula [25] we obtain that

$$\begin{aligned} R_{\gamma,\theta}V(x) &= V(x) + \gamma^4 n^4(x) + 4\gamma^2 d(d+2) + 4\gamma^2 n^2(x) + 4\gamma^2 n^2(x)d + \gamma^2 r^2(x) + 2\gamma \|x\|^2 - 4\gamma r(x)\|x\|^2 \\ &\quad - 4\gamma^3 n^2(x)r(x) - 8\gamma^2 dr(x) - 16\gamma^2 r(x) + 2\gamma^2 \|x\|^2 n^2(x) + 4\gamma d\|x\|^2 . \end{aligned} \quad (44)$$

Let  $\|x\| \geq M_\eta$  then  $r(x) \geq \|x\|^2 - c$ ,  $n(x) \leq L\|x\| + \sup_{\theta \in \Theta} \|\nabla_x U_\theta(0)\|$  and using the Cauchy-Schwarz inequality we have  $r(x) \leq L\|x\|^2 + \sup_{\theta \in \Theta} \|\nabla_x U_\theta(0)\|$ . Therefore, using (44), there exists a polynomial  $P$  of order strictly less than 4 such that

$$R_{\gamma, \theta} V(x) \leq V(x) + \gamma(\gamma^3 L^4 + 3\gamma L^2 - 4\eta)V(x) + \gamma P(\|x\|).$$

Since  $\bar{\gamma} = \min(1, \eta/L^2)$ , for all  $\gamma \in (0, \bar{\gamma})$ ,  $\theta \in \Theta$  and  $\|x\| \geq M_\eta$ , we have that

$$R_{\gamma, \theta} V(x) \leq \tilde{\lambda}^\gamma V(x) + \gamma P(V(x)^{1/4}),$$

with  $\tilde{\lambda} = \exp(\bar{\gamma}^3 L^4 + 3\bar{\gamma} L^2 - 4\eta) \in (0, 1)$ . Let  $1 > \lambda > \tilde{\lambda}$  and  $b_1$  defined by

$$b_1 = \sup_{\gamma \in (0, \bar{\gamma})} \sup_{x \in \mathbb{R}^d} \gamma^{-1} (\tilde{\lambda}^\gamma - \lambda^\gamma) V(x) + P(V(x)^{1/4}) < +\infty.$$

We obtain that for any  $\gamma \in (0, \bar{\gamma})$ ,  $\theta \in \Theta$  and  $\|x\| \geq M_\eta$  we have

$$R_{\gamma, \theta} V(x) \leq \lambda^\gamma V(x) + b_1 \gamma.$$

In addition there exists  $b_2 \geq 0$  such that for any  $\gamma \in (0, \bar{\gamma})$ ,  $\theta \in \Theta$  and  $\|x\| \leq M_\eta$  we have

$$R_{\gamma, \theta} V(x) \leq b_2 \gamma.$$

Let  $b = \max(b_1, b_2)$  we obtain that for any  $\gamma \in (0, \bar{\gamma})$ ,  $\theta \in \Theta$  and  $x \in \mathbb{R}^d$

$$R_{\gamma, \theta} V(x) \leq \lambda^\gamma V(x) + b \gamma.$$

□

**Proposition 15.** *The following properties hold*

1. Assume **L2** then there exist  $\zeta > 0$  and  $\beta \geq 0$  such that for all  $\theta \in \Theta$ ,  $(P_{t, \theta})_{t>0}$  satisfies the continuous drift condition  $\mathcal{D}(V, \zeta, \beta)$  for  $V$  defined in Proposition 12.
2. Assume **L1** and **L3** then there exist  $\zeta > 0$  and  $\beta \geq 0$  such that for all  $\theta \in \Theta$ ,  $(P_{t, \theta})_{t>0}$  satisfies the continuous drift condition  $\mathcal{D}(V, \zeta, \beta)$  for  $V$  defined in Proposition 13. In addition  $\zeta$  and  $\beta$  are defined by

$$\zeta = 2\eta, \quad \beta = 2(d + c + \eta(M_\eta^2 + 1)).$$

3. Assume **L1** and **L3** then there exist  $\zeta > 0$  and  $\beta \geq 0$  such that for all  $\theta \in \Theta$ ,  $(P_{t, \theta})_{t>0}$  satisfies the continuous drift condition  $\mathcal{D}(V, \zeta, \beta)$  for  $V$  defined in Proposition 14.

*Proof.* (a) Assuming **L2** the drift condition is a direct consequence of [12, Proposition 16] for  $V(x) = \exp(\eta\phi(x)/4)$  as in Proposition 12.

(b) Assuming **L3** and setting  $V(x) = 1 + \|x\|^2$  we get for any  $x \in \mathbb{R}^d$  and  $\theta \in \Theta$

$$\begin{aligned} -\langle \nabla_x U_\theta(x), \nabla V(x) \rangle + \Delta_x V(x) &= -2\langle \nabla_x U_\theta(x), x \rangle + 2d \\ &\leq -2\eta \mathbb{1}_{\overline{B}(0, M_\eta)^c} V(x) + 2\eta + 2d + 2c \\ &\leq -2\eta V(x) + 2(d + c + \eta(M_\eta^2 + 1)). \end{aligned}$$

(c) Assuming **L3** and setting  $V(x) = 1 + \|x\|^4$  there exists a polynom P of order stricly less than 4 such that for any  $x \in \mathbb{R}^d$  and  $\theta \in \Theta$

$$\begin{aligned} -\langle \nabla_x U_\theta(x), \nabla V(x) \rangle + \Delta_x V(x) &= -4\|x\|^2 \langle \nabla_x U_\theta(x), x \rangle + 4\|x\|^2 \sum_{i=1}^d x_i \\ &\leq -4\eta \mathbb{1}_{\overline{\mathbb{B}}(0, M_\eta)^c} V(x) + 4\|x\|^2 \sum_{i=1}^d x_i + 4c + 4\eta\|x\|^2 \\ &\leq -4\eta V(x) + P(V(x)^{1/4}). \end{aligned}$$

Since P is of order stricly less than 4 there exists  $B \geq$  such that for all  $x \in \mathbb{R}^d$  and  $\theta \in \Theta$

$$-\langle \nabla_x U_\theta(x), \nabla V(x) \rangle + \Delta_x V(x) \leq -3\eta V(x) + B.$$

□

### B.2.2 Checking H1

**Lemma 16.** *Let  $V : \mathbb{R}^d \rightarrow [1, +\infty)$  satisfying  $\lim_{\|x\| \rightarrow +\infty} V(x) = +\infty$  and  $V \in D(\mathcal{A})$ .*

(a) *Assume there exist  $\lambda \in (0, 1)$ ,  $b \geq 0$  and  $\bar{\gamma} > 0$  such that for any  $\theta \in \Theta$  and  $\gamma \in (0, \bar{\gamma})$ ,  $R_{\gamma, \theta}$  satisfies  $\mathbf{D}_d(V, \lambda^\gamma, b\gamma)$ . Then for any  $\theta \in \Theta$  and  $\gamma \in (0, \bar{\gamma})$ ,  $R_{\gamma, \theta}$  admits an invariant probability measure  $\pi_{\gamma, \theta}$  on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  and there exists  $A_0 \geq 0$  such that for any  $x \in \mathbb{R}^d$  and  $k \in \mathbb{N}$*

$$\delta_x R_{\gamma, \theta}^k V \leq A_0 + V(x), \quad \pi_{\gamma, \theta}(V) \leq A_0, \quad A_0 = b\lambda^{-\bar{\gamma}} / \log(1/\lambda).$$

*In addition, for all  $\theta \in \Theta$  and  $x \in \mathbb{R}^d$ ,  $\lim_{k \rightarrow +\infty} \|\delta_x R_{\gamma, \theta}^k - \pi_{\gamma, \theta}\|_V = 0$ .*

(b) *Assume there exist  $\zeta > 0$  and  $\beta \geq 0$  such that for any  $\theta \in \Theta$ ,  $\mathcal{A}_\theta$  satisfies  $\mathcal{D}(V, \zeta, \beta)$ . Then for any  $\theta \in \Theta$ , the diffusion is non-explosive,  $\mathcal{A}_\theta$  admits  $\pi_\theta$  as an invariant probability measure and*

$$\pi_\theta(V) \leq \beta / \zeta.$$

*In addition, for all  $\theta \in \Theta$  and  $x \in \mathbb{R}^d$ ,  $\lim_{t \rightarrow +\infty} \|\delta_x P_{\theta, t} - \pi_\theta\|_V = 0$ .*

*Proof.* (a) for any  $\gamma \in (0, \bar{\gamma})$  and  $\theta \in \Theta$ ,  $R_{\gamma, \theta}$  is irreducible with respect to the Lebesgue measure on  $\mathbb{R}^d$ , has the Feller property and satisfies  $\mathbf{D}_d(V, \lambda^\gamma, b\gamma)$  then [30, Section 4.4] applies and  $R_{\gamma, \theta}$  admits an invariant probability measure  $\pi_{\gamma, \theta}$ . The discrete drift condition and [12, Lemma 1] give that for any  $\gamma \in (0, \bar{\gamma})$  and  $\theta \in \Theta$

$$R_{\gamma, \theta}^k V(x) \leq V(x) + b\lambda^{-\bar{\gamma}} / \log(1/\lambda), \quad \pi_{\gamma, \theta}(V) \leq b\lambda^{-\bar{\gamma}} / \log(1/\lambda).$$

We obtain that for all  $\theta \in \Theta$  and  $x \in \mathbb{R}^d$ ,  $\lim_{k \rightarrow +\infty} \|\delta_x P_{t, \theta} - \pi_{\gamma, \theta}\|_V = 0$  using [31, Theorem 16.0.1].

(b) Using  $\mathcal{D}(V, \zeta, \beta)$  and [32, Theorem 2.1] we get that the diffusion process is non-explosive and thus  $(P_{t, \theta})_{t \geq 0}$  is defined for any  $\theta \in \Theta$  and  $t \geq 0$ . Using [48, Corollary 10.1.4] for any  $\theta \in \Theta$ ,  $(P_{t, \theta})_{t \geq 0}$  is strongly Feller continuous, therefore any compact sets is petite for the Markov kernel  $P_{h, \theta}$ , for any  $h > 0$  and  $\theta \in \Theta$ , by [31, Theorem 6.0.1]. Using [39, Chapter 7, Proposition 1.5], [19, Chapter 4, Theorem 9.17], and the fact that  $\pi_\theta(\mathcal{A}_\theta f) = 0$  for any  $\theta \in \Theta$  and  $f \in C_c^2(\mathbb{R}^d)$ , we obtain

that for any  $\theta \in \Theta$ ,  $\pi_\theta$  is an invariant measure for  $(P_{t,\theta})_{t \geq 0}$ . Using  $\mathcal{D}(V, \zeta, \beta)$  and [32, Theorem 4.5, Theorem 6.1] we get that for all  $\theta \in \Theta$ ,  $\pi_\theta(V) \leq \beta/\zeta$  and the uniform  $V$ -geometric ergodicity for  $(P_{t,\theta})_{t \geq 0}$ .  $\square$

As an immediate corollary we obtain that under the conditions of Lemma 16 for any  $\theta \in \Theta$ ,  $\gamma \in (0, \bar{\gamma})$  and  $k \in \mathbb{N}$ ,  $\pi_\theta R_{\gamma,\theta}^k V \leq \beta/\zeta + b\lambda^{-\bar{\gamma}}/\log(1/\lambda)$ .

**Lemma 17.** *Let  $V : \mathbb{R}^d \rightarrow [1, +\infty)$ . Assume there exist  $\lambda \in (0, 1)$ ,  $b \geq 0$  and  $\bar{\gamma} > 0$  such that for any  $\theta \in \Theta$  and  $\gamma \in (0, \bar{\gamma})$   $R_{\gamma,\theta}$  satisfies  $\mathbf{D}_d(V, \lambda^\gamma, b\gamma)$  then there exists  $B_1 \geq 0$  such that  $(X_k^n)_{n \in \mathbb{N}, k \in \{0, \dots, m_n\}}$  given by (9) satisfies for all  $n \in \mathbb{N}$  and  $k \in \{0, \dots, m_n\}$*

$$\mathbb{E} [V(X_k^n) | X_0^0] \leq B_1 V(X_0^0), \quad B_1 = 1 + b\lambda^{-\bar{\gamma}}/\log(1/\lambda).$$

*Proof.* Let  $n \in \mathbb{N}$  and  $k \in \{0, \dots, m_{n+1}\}$ . By induction we obtain that

$$\mathbb{E} [V(X_k^{n+1}) | \mathcal{F}_n] = R_{\theta_{n+1}, \gamma_{n+1}}^k V(X_0^{n+1}) \leq \lambda^{k\gamma_{n+1}} V(X_0^{n+1}) + b\gamma_{n+1} \sum_{i=1}^k \lambda^{\gamma_{n+1}(k-i)}.$$

In the same manner we obtain for any  $k \in \{0, \dots, m_0\}$

$$\mathbb{E} [V(X_k^0) | X_0^0] = R_{\theta_0, \gamma_0}^k V(X_0^0) \leq \lambda^{k\gamma_0} V(X_0^0) + b\gamma_0 \sum_{i=1}^k \lambda^{\gamma_0(k-i)}.$$

Let  $\vartheta_{n,k} = \sum_{j=0}^{n-1} m_j + k$ ,  $\vartheta_n = \vartheta_{n,0}$  and  $\tilde{\gamma}_i = \sum_{j=0}^{+\infty} \gamma_j \mathbb{1}_{(\vartheta_j, \vartheta_{j+1}]}(i)$ . Let also  $\Gamma_{p,q} = \sum_{i=p}^q \tilde{\gamma}_i$  and  $\Gamma_p = \Gamma_{1,p}$ . By induction we get for any  $n \in \mathbb{N}$  and  $k \in \{0, \dots, m_n\}$

$$\mathbb{E} [V(X_k^n) | X_0^0] \leq \lambda^{\Gamma_{\vartheta_{n,k}}} V(X_0^0) + b \sum_{i=1}^{\vartheta_{n,k}} \tilde{\gamma}_i \lambda^{\Gamma_{i+1, \vartheta_{n,k}}}.$$

Since  $(\tilde{\gamma}_i)_{i \in \mathbb{N}}$  is nonincreasing and for all  $t \geq 0$ ,  $1 - \lambda^t \geq -t\lambda^t \log(\lambda)$ , we have for all  $n \in \mathbb{N}$  and  $k \in \{0, \dots, m_n\}$  that

$$\begin{aligned} \sum_{i=1}^n \tilde{\gamma}_i \lambda^{\Gamma_{i+1, n}} &\leq \sum_{i=1}^n \tilde{\gamma}_i \prod_{j=i+1}^n (1 + \lambda^{\tilde{\gamma}_j} \log(\lambda) \tilde{\gamma}_j) \\ &\leq (-\lambda^{\tilde{\gamma}_1} \log(\lambda))^{-1} \sum_{i=1}^n \left\{ \prod_{j=i+1}^n (1 + \lambda^{\tilde{\gamma}_j} \log(\lambda) \tilde{\gamma}_j) - \prod_{j=i}^n (1 + \lambda^{\tilde{\gamma}_j} \log(\lambda) \tilde{\gamma}_j) \right\} \\ &\leq (-\lambda^{\tilde{\gamma}_1} \log(\lambda))^{-1}. \end{aligned}$$

$\square$

**Proposition 18.** *Let  $V : \mathbb{R}^d \rightarrow [1, +\infty)$  and  $M_{V,2} \geq 0$  such that for any  $x \in \mathbb{R}^d$ ,  $\sup_{x \in \mathbb{R}^d} (1 + \|x\|)^2/V(x) \leq M_{V,2}$ . Assume L1 and there exist  $\lambda \in (0, 1)$ ,  $b \geq 0$  and  $\bar{\gamma} > 0$  such that for any  $\theta \in \Theta$  and  $\gamma \in (0, \bar{\gamma})$   $R_{\gamma,\theta}$  satisfies  $\mathbf{D}_d(V, \lambda^\gamma, b\gamma)$ . Assume there exist  $\zeta > 0$  and  $\beta \geq 0$  such that for any  $\theta \in \Theta$ ,  $\mathcal{A}_\theta$  satisfies  $\mathcal{D}(V, \zeta, \beta)$ . We obtain that there exists  $B_3 \geq 0$  such that for any  $\theta \in \Theta$  and  $\gamma \in (0, \bar{\gamma})$*

$$\|\pi_{\gamma,\theta} - \pi_\theta\|_{V^{1/2}} \leq B_3 \gamma^{1/2}.$$

*Proof.* Since for all  $\theta \in \Theta$  and  $\gamma \in (0, \bar{\gamma})$  we have that  $R_{\gamma, \theta}$  satisfies  $\mathbf{D}_d(V, \lambda, \gamma)$  we also have that  $R_{\gamma, \theta}$  satisfies  $D(V^{1/2}, \lambda^{1/2}, b\lambda^{-\bar{\gamma}/2}/2)$ . Using Lemma 16 we obtain that for any  $x \in \mathbb{R}^d$  and  $\theta \in \Theta$

$$\lim_{k \rightarrow +\infty} \|\pi_\theta R_{\gamma, \theta}^k - \pi_\theta P_{\gamma k, \theta}\|_{V^{1/2}} = \|\pi_{\gamma, \theta} - \pi_\theta\|_{V^{1/2}}.$$

Following the lines of [12, Theorem 10] we write  $k = q_\gamma m_\gamma + r_\gamma$  with  $m_\gamma = \lceil 1/\gamma \rceil$  and  $q_\gamma, r_\gamma \in \mathbb{N}$  and  $0 \leq r_\gamma < m_\gamma$ . Using the uniform ergodicity in [10] and that  $\pi_\theta$  is invariant for  $P_{t, \theta}$  with  $t \geq 0$ , see Lemma 16, we obtain for all  $\theta \in \Theta$ ,  $\gamma \in (0, \bar{\gamma})$  and  $k \in \mathbb{N}$

$$\begin{aligned} \|\pi_\theta R_{\gamma, \theta}^k - \pi_\theta P_{\gamma k, \theta}\|_{V^{1/2}} &\leq \sum_{\ell=0}^{q_\gamma-1} \|\pi_\theta P_{\gamma(\ell+1)m_\gamma, \theta} R_{\gamma, \theta}^{(q_\gamma - (\ell+1))m_\gamma + r_\gamma} - \pi_\theta P_{\gamma \ell m_\gamma, \theta} R_{\gamma, \theta}^{(q_\gamma - \ell)m_\gamma + r_\gamma}\|_{V^{1/2}} \\ &\quad + \|\pi_\theta P_{\gamma(q_\gamma m_\gamma + r_\gamma), \theta} - \pi_\theta P_{\gamma q_\gamma m_\gamma, \theta} R_{\gamma, \theta}^{r_\gamma}\|_{V^{1/2}} \\ &\leq \sum_{\ell=0}^{q_\gamma-1} C \xi^{\gamma m_\gamma (q_\gamma - (\ell+1))} \|\pi_\theta P_{\gamma \ell m_\gamma, \theta} P_{m_\gamma \gamma, \theta} - \pi_\theta P_{\gamma \ell m_\gamma, \theta} R_{\gamma, \theta}^{m_\gamma}\|_{V^{1/2}} \\ &\quad + \|\pi_\theta P_{\gamma(q_\gamma m_\gamma + r_\gamma), \theta} - \pi_\theta P_{\gamma q_\gamma m_\gamma, \theta} R_{\gamma, \theta}^{r_\gamma}\|_{V^{1/2}} \\ &\leq \|\pi_\theta P_{m_\gamma \gamma, \theta} - \pi_\theta R_{\gamma, \theta}^{m_\gamma}\|_{V^{1/2}} \sum_{\ell=1}^{q_\gamma} C \xi^{\ell \gamma m_\gamma} + \|\pi_\theta P_{\gamma r_\gamma, \theta} - \pi_\theta R_{\gamma, \theta}^{r_\gamma}\|_{V^{1/2}}, \end{aligned} \quad (45)$$

where  $C \geq 0, \xi \in (0, 1)$  are the constants given by [10] with drift condition  $D(V^{1/2}, \lambda^{1/2}, b/2)$ . Using L1 and  $(a+b)^2 \leq 2(a^2 + b^2)$ , we obtain that for any  $x \in \mathbb{R}^d$  and  $\theta \in \Theta$

$$\|\nabla_x U_\theta(x)\|^2 \leq 2L^2 \|x\|^2 + 2 \sup_{\theta \in \Theta} \|\nabla_x U_\theta(0)\|^2 \leq 2 \left[ L^2 M_{V,2} + \sup_{\theta \in \Theta} \|\nabla_x U_\theta(0)\|^2 \right] V(x). \quad (46)$$

Since  $\mathcal{A}_\theta$  satisfies a  $\mathcal{D}(V, \zeta, \beta)$  and  $R_{\gamma, \theta}$  satisfies  $\mathbf{D}_d(V, \lambda, b)$  for any  $\theta \in \Theta$ , we obtain that for any  $\theta \in \Theta$  and  $\gamma \in (0, \bar{\gamma})$

$$\pi_\theta P_{\gamma m_\gamma, \theta}(V) \leq \tilde{A}_1, \quad \pi_\theta R_{\gamma, \theta}^{m_\gamma}(V) \leq \tilde{A}_1, \quad \tilde{A}_1 = \beta/\zeta + b\lambda^{-\bar{\gamma}} \log(1/\lambda)^{-1}. \quad (47)$$

Combining (46) and (47) we can apply [10] and we get for any  $\theta \in \Theta$  and  $\gamma \in (0, \bar{\gamma})$

$$\|\pi_\theta P_{\gamma m_\gamma, \theta} - \pi_\theta R_{\gamma, \theta}^{m_\gamma}\|_{V^{1/2}} \leq B'_3 \gamma^{1/2}, \quad \|\pi_\theta P_{\gamma r_\gamma, \theta} - \pi_\theta R_{\gamma, \theta}^{r_\gamma}\|_{V^{1/2}} \leq B'_3 \gamma^{1/2}, \quad (48)$$

with

$$B'_3 = \tilde{A}_1^{1/2} (1 + \bar{\gamma})^{1/2} \left\{ d + 2\bar{\gamma} (L^2 + \sup_{\theta \in \Theta} \|\nabla_x U_\theta(0)\|) \tilde{A}_1 \right\}^{1/2} L.$$

Combining (45) and (48) we get for any  $k \in \mathbb{N}$ ,  $\theta \in \Theta$  and  $\gamma \in (0, \bar{\gamma})$

$$\|\pi_\theta R_{\gamma, \theta}^k - \pi_\theta P_{\gamma k, \theta}\|_{V^{1/2}} \leq C B'_3 \left( \sum_{\ell=1}^{q_\gamma} \xi^{\gamma m_\gamma \ell} + 1 \right) \gamma^{1/2} \leq C B'_3 \{1 + 1/(1 - \xi)\} \gamma^{1/2},$$

where we used that  $\xi^{\gamma m_\gamma} \leq \xi$ . We prove the result upon taking the limit in  $k$ .  $\square$

### B.2.3 Checking H2

**Proposition 19.** Assume **L1** and there exist a measurable function  $V : \mathbb{R}^d \rightarrow (1, +\infty)$ ,  $\lambda \in (0, 1)$ ,  $b \geq 0$  and  $\bar{\gamma} > 0$  such that  $\lim_{\|x\| \rightarrow +\infty} V(x) = +\infty$  and for any  $\theta \in \Theta$  and  $\gamma \in (0, \bar{\gamma})$ ,  $R_{\gamma, \theta}$  satisfies for any  $x \in \mathbb{R}^d$ ,

$$R_{\gamma, \theta} V(x) \leq \lambda^\gamma V(x) + b\gamma.$$

Then for any  $p \in \mathbb{N}^*$ , there exists  $B_{4,p} \geq 0$  such that for any  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ , with  $\sup_{\mathbb{R}^d} |f/V^{1/p}| \leq 1$ ,  $\theta \in \Theta$ ,  $\gamma \in (0, \bar{\gamma})$ ,  $x \in \mathbb{R}^d$  and  $n \in \mathbb{N}$

$$n^{-p} \mathbb{E} \left[ \left| \sum_{k=1}^n f(\bar{X}_k) - \pi_\theta(f) \right|^p \right] \leq B_{4,p} V(x),$$

where  $(\bar{X}_k)_{k \in \mathbb{N}}$  is the Markov chain starting from  $x$  defined by (6). In addition,  $B_{4,p} = CS_b(b)S_\lambda(\lambda)$  with  $C$  a universal constant and  $S_b, S_\lambda$  are polynomials such that

$$\sup_{b \in \mathbb{R}} |(1+b)^{-(p+1)} S_b(b)| \leq 1, \quad \sup_{[0,1]} |(1-\lambda)^{p+1} S_\lambda(\lambda)| \leq 1.$$

*Proof.* Let  $p \in \mathbb{N}^*$  and  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ , with  $\sup_{\mathbb{R}^d} |f/V^{1/p}| \leq 1$ ,  $\theta \in \Theta$  and  $\gamma \in (0, \bar{\gamma})$ . Using the Hölder inequality, setting  $q = p/(p-1)$ , and the triangle inequality we get that

$$\begin{aligned} \left| \left( \sum_{k=1}^n f(\bar{X}_k) - \pi_\theta(f) \right)^p \right| &\leq n^{p/q} \sum_{k=1}^n |f(\bar{X}_k) - \pi_\theta(f)|^p \leq n^{p/q} \sum_{k=1}^n \{|f|(\bar{X}_k) + |\pi_\theta(f)|\}^p \\ &\leq n^{p/q} \left( \sum_{j=0}^p \binom{p}{j} \pi_\theta(|f|)^{p-j} \sum_{k=1}^n |f|^j(\bar{X}_k) \right), \end{aligned} \quad (49)$$

Using Lemma 16 and  $\sup_{\mathbb{R}^d} |f/V^{1/p}| \leq 1$  we obtain that

$$\pi_\theta(|f|) \leq CA_0^{1/p}, \quad \mathbb{E}_x \left[ \sum_{j=1}^n |f|^j(\bar{X}_k) \right] \leq nCA_0^{j/p} V(x), \quad A_0 \leq (1+b)(1+\lambda^{-\bar{\gamma}}/\log(1/\lambda)) \quad (50)$$

where  $C$  is a universal constant. Combining (49) and (50) we get that

$$\sup_{|f|_{V^{1/2}} \leq 1} \mathbb{E}_x \left[ n^{-p} \left\{ \sum_{k=1}^n f(\bar{X}_k) - \pi_{\gamma, \theta}(f) \right\}^p \right] \leq B_{4,p} n^{1+p/q-p} V(x) \leq B_{4,p} V(x),$$

where  $B_{4,p} = CS_b(b)S_\lambda(\lambda)$  with  $C$  a universal constant and  $S_b, S_\lambda$  are polynomials such that

$$\sup_{b \in \mathbb{R}} |(1+|b|)^{-1} S_b(b)| \leq 1, \quad \sup_{\lambda \in [0,1]} |(1-\lambda) S_\lambda(\lambda)| \leq 1.$$

□

### B.2.4 Proof of Theorem 5

**L1** and **L2** or **L3** ensure a uniform drift condition on  $R_{\gamma,\theta}$ , see Proposition 12 and Proposition 13. Note that the Lyapunov functions  $V$  defined by Proposition 12 and Proposition 13 satisfy  $\sup_{x \in \mathbb{R}^d} (1 + \|x\|^2)/V(x) < +\infty$ . Assumption **H1** is then implied combining Lemma 17, Proposition 18, the result of uniform geometric ergodicity with constants independent from  $\theta \in \Theta$  in [10] and that for any  $\theta \in \Theta$  and  $x \in \mathbb{R}^d$ ,  $\|H_\theta(x)\| \leq M_H V^{1/2}(x)$ . Note that we have the following equalities in **H1**

$$A_1 = B_1, \quad A_2 = B_2 M_H, \quad A_3 = B_3 M_H,$$

where  $B_2$  is given in [10] and depends only on  $b$  and  $\lambda$  if **L4** holds. Using [10] we show that  $B_3$  depends only on  $b, \lambda, \zeta, \beta$  and  $d$  and satisfies  $\sup_{d \in \mathbb{R}} B_3(1 + |d|)^{-1/2} < +\infty$ .

We conclude that **H2** holds using Proposition 19 for  $p = 2$  and that for any  $\theta \in \Theta$  and  $x \in \mathbb{R}^d$ ,  $\|H_\theta(x)\| \leq M_H V(x)^{1/2}$ . We obtain that  $A_4 = B_{4,2} M_H^2$ .

### B.3 Proof of Proposition 6

*Proof.* The proof is divided in two parts. First we show there exist  $\eta > 0$  and  $R \geq 0$  such that for any  $\theta \in \Theta$  and  $x \in \mathbb{R}^d$ ,  $\|x\| > R$ ,

$$U_\theta(x) - U_\theta(0) \geq \eta \|x\|. \quad (51)$$

Let  $M$  defined by

$$M = \sup_{\theta \in \Theta} U_\theta(0) + \sup_{\theta \in \Theta, x \in \overline{B}(0,1)} U_\theta(x).$$

Note that by **L1** and since  $\Theta$  is compact we have that  $M < +\infty$ . By contradiction, we show that there exists  $R \geq 0$  such that for any  $\theta \in \Theta$   $\{x, U_\theta(x) \leq M + 1\} \subset B(0, R)$ . Assume that for any  $R \geq 0$ , there exist  $\theta_R \in \Theta$ ,  $x_R \in \{x, U_{\theta_R}(x) \leq M + 1\}$  and  $\|x_R\| \geq R$ . Then  $\text{Vol}(\text{Conv}[\overline{B}(0, 1) \cup \{x_R\}])$  grows at least linearly in  $R$  independently from  $\theta$ . On the other hand by definition of  $M$  and  $x_R$  and since **L4(0)** holds, we have for any  $R \geq 0$

$$\text{Vol}(\text{Conv}[\overline{B}(0, 1) \cup \{x_R\}]) \leq e^{M+1} \int_{\{U_{\theta_R}(x) \leq M+1\}} e^{-U_{\theta_R}(y)} dy \leq e^{M+1} \sup_{\theta \in \Theta} \int_{\mathbb{R}^d} e^{-U_\theta(y)} dy.$$

As a consequence, we conclude that there exists  $R \geq 0$ , such that for any  $\theta \in \Theta$ ,  $\{x, U_\theta(x) \leq M + 1\} \subset B(0, R)$ . Set  $x \in \overline{B}(0, R)^c$  and  $y = Rx/\|x\|$ . By **L4(0)** we get for any  $\theta \in \Theta$

$$M + 1 \leq U_\theta(y) \leq RU_\theta(x)/\|x\| + (1 - R/\|x\|)U_\theta(0),$$

which gives (51) with  $\eta = \{M + 1 - \sup_{\theta \in \Theta} U_\theta(0)\}/R > 0$  by definition of  $M$ . From (51) and **L1**, for any  $\theta \in \Theta$  there exists  $x_\theta^*$  minimizer of  $U_\theta$ . Thus we get for any  $x \in \overline{B}(0, R)^c$

$$U_\theta(x) - U_\theta(x_\theta^*) \geq U_\theta(x) - U_\theta(0) \geq \eta \|x\|.$$

Furthermore we have for any  $x \in \mathbb{R}^d$  we have  $U_\theta(x) - U_\theta(x_\theta^*) \geq 0$ . From these results, [34, Theorem 2.1.5, Equation (2.1.7)], **L1**, **L4(0)** and (51) we obtain that for any  $\theta \in \Theta$  and  $x \in \mathbb{R}^d$

$$\langle \nabla_x U_\theta(x), x - x_\theta^* \rangle \geq (2L)^{-1} \|\nabla_x U_\theta(x)\|^2 + \eta \|x\| \mathbb{1}_{\overline{B}(0, R)^c}(x). \quad (52)$$

Furthermore, using  $ab \leq 2^{-1}(\varepsilon a^2 + (b^2/\varepsilon))$ , the Cauchy-Schwarz inequality and **L4(0)** we get for any  $\theta \in \Theta$ ,  $x \in \mathbb{R}^d$  and  $\varepsilon > 0$

$$|\langle \nabla_x U_\theta(x), x_\theta^* \rangle| \leq \|\nabla_x U_\theta(x)\| \|x_\theta^*\| \leq \varepsilon \|\nabla_x U_\theta(x)\|^2/2 + M_*^2/2\varepsilon. \quad (53)$$

Combining (52) and (53), we obtain for any  $\theta \in \Theta$  and  $x \in \mathbb{R}^d$  that

$$\begin{aligned} \langle \nabla_x U_\theta(x), x \rangle &= \langle \nabla_x U_\theta(x), x - x_\theta^* \rangle + \langle \nabla_x U_\theta(x), x_\theta^* \rangle \\ &\geq \langle \nabla_x U_\theta(x), x - x_\theta^* \rangle - |\langle \nabla_x U_\theta(x), x_\theta^* \rangle| \\ &\geq (1/2)(\mathbf{L}^{-1} - \varepsilon) \|\nabla_x U_\theta(x)\|^2 + \eta \|x\| \mathbb{1}_{\overline{B}(0,R)^c}(x) - M_*^2/2\varepsilon, \end{aligned}$$

which concludes the proof upon taking  $\varepsilon = (2\mathbf{L})^{-1}$ .  $\square$

## B.4 Proof of Corollary 7

*Proof.* Note that applying Proposition 13 and Proposition 15 with  $\tilde{V}(x) = (\|x\| + 1)^2/d$  we obtain that  $b, \lambda, \zeta$  and  $\beta$  do not depend on the dimension  $d$ . Theorem 5 still holds for this Lyapunov function upon noticing that for any  $\theta \in \Theta$  and  $x \in \mathbb{R}^d$ ,  $\|H_\theta(x)\| \leq d^{1/2} M_H \tilde{V}(x)$ . Let  $M'_H = M_H d^{1/2}$

Theorem 3 holds using Theorem 5. We recall that the sequence  $(E_n)_{n \in \mathbb{N}}$  is given by for any  $n \in \mathbb{N}$  by

$$\begin{aligned} E_n &= 2M_\Theta^2 + 2M_\Theta \mathbb{E} [V(X_0^0)] \left[ \frac{(A_1 \vee 1)A_2}{\xi^{\tilde{\gamma}} \log(1/\xi)} \sum_{k=0}^n \{\delta_{k+1}/(\gamma_k m_k)\} + A_3 \sum_{k=0}^n \delta_{k+1} B(\gamma_k) \right] \\ &\quad + A_4 \mathbb{E} [V(X_0^0)] \sum_{k=0}^n \delta_{k+1}^2 m_k^{-2} D(\gamma_k, m_k). \end{aligned} \quad (54)$$

Using that  $A_1 = B_1$ ,  $A_2 = B_2 M'_H$ ,  $A_3 = B_3 M'_H$  and  $A_4 = B_{4,2} M_H'^2$  we have

$$\begin{aligned} E_n &= 2M_\Theta^2 + 2d^{1/2} M_\Theta M_H \mathbb{E} [V(X_0^0)] \left[ \frac{(B_1 \vee 1)B_2}{\xi^{\tilde{\gamma}} \log(1/\xi)} \sum_{k=0}^n \{\delta_{k+1}/(\gamma_k m_k)\} + B_3 \sum_{k=0}^n \delta_{k+1} B(\gamma_k) \right] \\ &\quad + dB_{4,2} M_H'^2 \mathbb{E} [V(X_0^0)] \sum_{k=0}^n \delta_{k+1}^2 m_k^{-2} D(\gamma_k, m_k). \end{aligned} \quad (55)$$

Since  $b, \lambda, \zeta$  and  $\beta$  do not depend on the dimension  $d$ ,  $B_1, B_2, B_{4,2}$  do not depend on the dimension as well. We have  $B_3 \leq Cd^{1/2}$ , where  $C$  is a constant independent of the dimension. Finally we obtain that  $E_n$  is bounded by  $Cd$  where  $C$  is a constant independent of the dimension.  $\square$

## B.5 Proof of Theorem 8

We preface the proof by a technical lemma.

**Lemma 20.** *Let  $V : \mathbb{R}^d \rightarrow [1, +\infty)$  and  $M_{V,4} \geq 0$  such that  $\sup_{x \in \mathbb{R}^d} (1 + \|x\|^4)/V(x) \leq M_{V,4}$ . Let  $M \geq 0$  such that for any  $\theta \in \Theta$ ,  $\gamma \in (0, \bar{\gamma})$ , with  $\bar{\gamma} > 0$  and  $x \in \mathbb{R}^d$ ,  $R_{\gamma,\theta} V(x) \leq MV(x)$ . Assume **L1** and **L5**, then we have for any  $\theta_1, \theta_2 \in \Theta$ ,  $\gamma_1, \gamma_2 \in (0, \bar{\gamma})$  with  $\gamma_1 > \gamma_2$ ,  $a \in [1/4, 1/2]$  and  $x \in \mathbb{R}^d$*

$$\|\delta_x R_{\gamma_1, \theta_1} - \delta_x R_{\gamma_2, \theta_2}\|_{V^a} \leq B_6 \left[ (\gamma_2^{-1} - \gamma_1^{-1})^{1/2} + \gamma_2^{-1/2} |\gamma_1 - \gamma_2| + \gamma_2^{1/2} \|\theta_1 - \theta_2\| \right] V(x)^{2a},$$



where  $\{R_{\gamma,\theta}, \gamma \in (0, \bar{\gamma}), \theta \in \Theta\}$  is the sequence of Markov kernels associated with the recursion (6) and

$$B_6 = \max \left( 2M^{1/2} \left[ d/4 + \sup_{\theta \in \Theta} \|\nabla_x U_\theta(0)\|^2 + L^2 M_{4,V}^{1/2} \right]^{1/2}, (2M)^{1/2} L_U \right).$$

*Proof.* Using [12, Lemma 24] we have that for any  $\theta_1, \theta_2 \in \Theta$  and  $\gamma_1, \gamma_2 \in (0, \bar{\gamma})$

$$\begin{aligned} \|\delta_x R_{\gamma_1, \theta_1} - \delta_x R_{\gamma_2, \theta_2}\|_{V^a} &\leq \sqrt{2} (R_{\gamma_1, \theta_1} V^{2a}(x) + R_{\gamma_2, \theta_2} V^{2a}(x))^{1/2} \text{KL}(\delta_x R_{\gamma_1, \theta_1} | \delta_x R_{\gamma_2, \theta_2})^{1/2} \\ &\leq 2M^a V^a(x) \text{KL}(\delta_x R_{\gamma_1, \theta_1} | \delta_x R_{\gamma_2, \theta_2})^{1/2} \end{aligned} \quad (56)$$

Denote for any  $z \in \mathbb{R}^d$  and  $\sigma^2 > 0$ ,  $\gamma_{z, \sigma^2}$  the  $d$ -dimensional Gaussian distribution with mean  $z$  and covariance matrix  $\sigma^2 \text{Id}$ . Using that for any  $z_1, z_2 \in \mathbb{R}^d$  and  $\sigma_1, \sigma_2 > 0$ ,

$$\text{KL}(\gamma_{z, \sigma^2} | \gamma_{z, \sigma^2}) = \ln(\sigma_2^2 / \sigma_1^2) + d\sigma_2^{-2} \{1 - \sigma_2^2 / \sigma_1^2\} / 2 + \sigma_2^{-2} \|\mu_1 - \mu_2\|^2 / 2,$$

we obtain that for any  $\theta_1, \theta_2 \in \Theta$ ,  $\gamma_1, \gamma_2 \in (0, \bar{\gamma})$  and  $x \in \mathbb{R}^d$ ,

$$\text{KL}(\delta_x R_{\gamma_1, \theta_1} | \delta_x R_{\gamma_2, \theta_2}) = \ln(\gamma_2 / \gamma_1) + d\gamma_2^{-1} (1 - \gamma_2 / \gamma_1) / 4 + \gamma_2^{-1} \Xi / 4, \quad (57)$$

where  $\Xi$  satisfies the following inequalities

$$\begin{aligned} \Xi &= \|\gamma_1 \nabla_x U_{\theta_1}(x) - \gamma_2 \nabla_x U_{\theta_2}(x)\|^2 \\ &= \|\gamma_1 \nabla_x U_{\theta_1}(x) - \gamma_2 \nabla_x U_{\theta_1}(x) + \gamma_2 \nabla_x U_{\theta_1}(x) - \gamma_2 \nabla_x U_{\theta_2}(x)\|^2 \\ &\leq 2\|\gamma_1 \nabla_x U_{\theta_1}(x) - \gamma_2 \nabla_x U_{\theta_1}(x)\|^2 + 2\|\gamma_2 \nabla_x U_{\theta_1}(x) - \gamma_2 \nabla_x U_{\theta_2}(x)\|^2 \\ &\leq 2(\gamma_1 - \gamma_2)^2 \|\nabla_x U_{\theta_1}(x)\|^2 + 2\gamma_2^2 \|\nabla_x U_{\theta_1}(x) - \nabla_x U_{\theta_2}(x)\|^2 \\ &\leq 2(\gamma_1 - \gamma_2)^2 \|\nabla_x U_{\theta_1}(x)\|^2 + 2\gamma_2^2 L_U^2 \|\theta_1 - \theta_2\|^2 V^{2a}(x), \end{aligned} \quad (58)$$

where we used **L5** in the last line. Using **L5** again,  $\sup_{\theta \in \Theta} \|\nabla_x U_\theta(0)\| < +\infty$  by **L1**, we obtain that for any  $\theta \in \Theta$ ,  $x \in \mathbb{R}^d$  and  $a \in [1/4, 1/2]$  we have

$$\|\nabla_x U_\theta(x)\|^2 \leq 2(\|\nabla_x U_\theta(x) - \nabla_x U_\theta(0)\|^2 + \sup_{\theta \in \Theta} \|\nabla_x U_\theta(0)\|^2) \leq C_\Theta V^{2a}(x),$$

with  $C_\Theta = 2\|\sup_{\theta \in \Theta} \nabla_x U_\theta(0)\|^2 + 2L^2 M_{4,V}^{1/2}$ . Combining this result and (58) in (57), we have that for any  $\theta_1, \theta_2 \in \Theta$ ,  $\gamma_1, \gamma_2 \in (0, \bar{\gamma})$  with  $\gamma_1 > \gamma_2$  and  $x \in \mathbb{R}^d$

$$\begin{aligned} &\text{KL}(\delta_x R_{\gamma_1, \theta_1} | \delta_x R_{\gamma_2, \theta_2}) \\ &\leq d\gamma_2^{-1} (1 - \gamma_2 / \gamma_1) / 4 + \gamma_2^{-1} (\gamma_1 - \gamma_2)^2 \|\nabla_x U(\theta_1, x)\|^2 / 2 + \gamma_2 L_U^2 \|\theta_1 - \theta_2\|^2 V^{2a}(x) / 2 \\ &\leq [d\gamma_2^{-1} (1 - \gamma_2 / \gamma_1) / 4 + \gamma_2^{-1} (\gamma_1 - \gamma_2)^2 C_\Theta / 2 + \gamma_2 L_U^2 \|\theta_1 - \theta_2\|^2 / 2] V^{2a}(x) \end{aligned}$$

This result substituted in (56) completes the proof with the inequality for any  $a, b \in \mathbb{R}_+$ ,  $(a+b)^{1/2} \leq a^{1/2} + b^{1/2}$ .  $\square$

*Proof of Theorem 8.* **L1** and **L2** or **L3** ensure a uniform drift condition on  $R_{\gamma,\theta}$ , see Proposition 12 and Proposition 14. Note that the Lyapunov functions  $V$  defined by Proposition 12 and Proposition 14 satisfy  $\sup_{x \in \mathbb{R}^d} (1 + \|x\|^4) / V(x) < +\infty$ . Assumption **H3** is obtained using Lemma 17, . Combining **H3**, the result of uniform geometric ergodicity with constants independent from  $\theta \in \Theta$  in [10] and that for any  $\theta \in \Theta$  and  $x \in \mathbb{R}^d$ ,  $\|H_\theta(x)\| \leq M_H V^{1/2}(x)$  we obtain **H5**. Using Proposition 18 and **H5** we obtain **H1-(ii)**. **H4** is a direct consequence of Lemma 20  $\square$