

# (Non)-penalized multilevel methods for non-uniformly log-concave distributions

Maxime Egéa\*

## Abstract

We study and develop multilevel methods for the numerical approximation of a log-concave probability  $\pi$  on  $\mathbb{R}^d$ , based on (over-damped) Langevin diffusion. In the continuity of [12] concentrated on the uniformly log-concave setting, we here study the procedure in the absence of the uniformity assumption. More precisely, we first adapt an idea of [8] by adding a penalization term to the potential to recover the uniformly convex setting. Such approach leads to an  $\varepsilon$ -complexity of the order  $\varepsilon^{-5} \pi(|\cdot|^2)^{9/2} d$  (up to logarithmic terms). Then, in the spirit of [14], we propose to explore the robustness of the method in a weakly convex parametric setting where the lowest eigenvalue of the Hessian of the potential  $U$  is controlled by the function  $U(x)^{-r}$  for  $r \in (0, 1)$ . In this intermediary framework between the strongly convex setting ( $r = 0$ ) and the “Laplace case” ( $r = 1$ ), we show that with the help of the control of exponential moments of the Euler scheme, we can adapt some fundamental properties for the efficiency of the method. In the “best” setting where  $U$  is  $C^3$  and  $U(x)^{-r}$  control the largest eigenvalue of the Hessian, we obtain an  $\varepsilon$ -complexity of the order  $c_{\rho, \delta} \varepsilon^{-2-\rho} d^{1+\frac{\rho}{2}+(4-\rho+\delta)r}$  for any  $\rho > 0$  (but with a constant  $c_{\rho, \delta}$  which increases when  $\rho$  and  $\delta$  go to 0).

**Keywords:** multilevel; ergodic diffusion; Langevin algorithm; Euler scheme; weakly convex.

**MSC2020 subject classifications:** Primary 65C05; 37M25, Secondary 65C40; 93E35.

Submitted to EJP on May 3, 2023, final version accepted on February 12, 2024.

Supersedes arXiv:2301.09471.

## 1 Introduction

In this paper, we are interested in the sampling of probability distribution named Gibbs measure whose density is  $\pi(dx) = \frac{1}{Z} e^{-\frac{U(x)}{2\sigma^2}} \lambda(dx)$  where  $\lambda$  is the Lebesgue measure,  $Z = \int_{\mathbb{R}^d} e^{-\frac{U(x)}{2\sigma^2}} \lambda(dx)$  and  $U : \mathbb{R}^d \rightarrow \mathbb{R}$  is a coercive function. Many applications require the computation of these measures in high dimension state space including for example machine learning, Bayesian estimation or statistical physics. Methods that are studied in

\*University of Angers, France. E-mail: maxime.egea@univ-angers.fr

this paper are based on the discretization of over-damped Langevin stochastic differential equation (**SDE**)

$$dX_t = -\nabla U(X_t)dt + \sigma dB_t, \quad (1.1)$$

where  $(B_t)_{t \geq 0}$  is a  $d$ -dimensional Brownian motion and  $\sigma \in \mathbb{R}_*^+$ . These methods received a lot of attention in the last few years, in particular when  $U$  is strongly convex (in the sense where, in the whole space, the smallest eigenvalue of its Hessian is lower bounded by a positive  $\alpha$ ). This assumption may be certainly constraining in view of applications. It is the reason why, in this paper, we suppose that  $U$  is not strongly convex but only weakly convex.<sup>1</sup>

More precisely, we will assume that the potential  $U$  is a convex twice differentiable function whose gradient is Lipschitz continuous. Under these assumptions, strong existence and uniqueness of a solution  $(X_t)_{t \geq 0}$  classically hold and the solution to (1.1) is an ergodic Markov process whose invariant distribution is exactly the Gibbs distribution  $\pi \propto e^{-U} d\lambda$  (for background, see e.g. [28], [22], [24], [19]).

We respectively denote by  $(P_t)_{t \geq 0}$  and  $\mathcal{L}$  the related semi-group and infinitesimal generator. We recall that for a twice differentiable function  $f : \mathbb{R}^d \mapsto \mathbb{R}$  by

$$\mathcal{L}f = -\langle \nabla U, \nabla f \rangle + \frac{\sigma^2}{2} \Delta f.$$

It is also well-known that in this log-concave setting, the distribution  $\pi$  satisfies the Poincaré inequality (see e.g. [2]) and that convergence holds to equilibrium in distribution and in “pathwise average”: for any starting point  $x \in \mathbb{R}^d$ , the occupation measure converges to  $\pi$  in the following sense: for all continuous function  $f \in L^2(\pi)$ ,

$$\frac{1}{t} \int_0^t f(X_s^x) ds \xrightarrow[t \rightarrow +\infty]{} \pi(f) \quad a.s. \quad (1.2)$$

In the continuity of [12], our multilevel methods will be based on discretized adaptations of (1.2). More precisely, we first choose to approximate the stochastic process  $(X_t)_{t \geq 0}$  by the classical Euler-Maruyama scheme. When the related step size  $\gamma$  is constant, this discretization scheme is defined by  $\bar{X}_0 = x \in \mathbb{R}^d$  and:

$$\forall n \geq 0, \quad \bar{X}_{(n+1)\gamma} = \bar{X}_{n\gamma} - \gamma \nabla U(\bar{X}_{n\gamma}) + \sigma \sqrt{\gamma} Z_{n+1},$$

where  $(Z_n)_{n \in \mathbb{N}}$  is an *i.i.d* sequence of  $d$ -dimensional standard Gaussian random variables. In the long-time setting, these schemes and their convergence properties to equilibrium were first studied in the nineties by [33] and [32]. This discretization of the Langevin diffusion is widely studied in the literature under uniform convexity of the potential  $U$  [5, 11, 10] and also under weaker assumptions [29, 34] (e.g. isoperimetry assumption). Then, some decreasing step Euler schemes were investigated by [25] and [26] in order to manage, in the same time, the discretization and long-time errors. Here, we choose to keep the constant step size point of view in order to avoid some additional technicalities but our ideas could be probably adapted to this setting.

We also introduced the continuous-time extension of  $(\bar{X}_{n\gamma})_{n \geq 0}$  given by: for all  $n \in \mathbb{N}$  and for all  $t \in [n\gamma, (n+1)\gamma)$ ,

$$\bar{X}_t^\gamma := \bar{X}_{n\gamma} - (t - n\gamma) \nabla U(\bar{X}_{n\gamma}) + \sigma (B_t - B_{(n+1)\gamma}). \quad (1.3)$$

<sup>1</sup>We use in the sequel the usual terminology where “strongly convex” and “weakly convex” respectively means “uniformly strongly convex” and “non-uniformly strongly convex”.

If we denote by  $t_\gamma$  the discretization time related to a positive number  $t$ , i.e.,

$$t_\gamma = \gamma \sup\{n \geq 1, n\gamma \leq t\}, \tag{1.4}$$

we remark that

$$\forall t \geq 0, \quad \bar{X}_t = x - \int_0^t \nabla U(\bar{X}_{s_\gamma}) ds + \sigma \int_0^t dB_s.$$

This ‘‘pseudo-diffusion’’ form is usually convenient for proofs but it is worth noting that the procedure is only based on the discrete-time Euler scheme. If no confusion arises, we will sometimes write  $t$  instead of  $t_\gamma$ , and  $\bar{X}_t$  or  $\bar{X}_t^\gamma$  instead of  $\bar{X}_t^{\gamma, x_0}$  to alleviate the notations. We now mimic (1.2) with the Euler scheme to approximate the target measure  $\pi$ . Thus, consider the following occupation measure (for background see [33]), for  $N \in \mathbb{N}$

$$\nu_N^\gamma(f) := \frac{1}{N} \sum_{i=0}^{N-1} f(\bar{X}_{i\gamma}).$$

### 1.1 Multilevel methods

Multilevel methods introduced by [20] and [15] (see also [23]) initially used for the approximation of  $\mathbb{E}[f(X_T)]$ , are now widely exploited in many other settings, for example in stochastic optimization (see [13, 9]) or to speed up MCMC methods (see e.g. [21, 18]). The rough idea is the following: assume that the target  $\mathbb{E}[X]$  is the expectation of a random variable that cannot be sampled (with a reasonable cost) and consider a family of random variables  $(X_j)_j$  approximating  $X$ , with a cost of simulation and a precision which typically increases with  $j$ . The principle of the multilevel is to stack correcting layers with a low variance to a coarse approximation  $X_0$  of the target. More precisely, writing

$$\mathbb{E}[X_J] = \underbrace{\mathbb{E}[X_0]}_{\text{Coarse}} + \sum_{j=1}^J \underbrace{\mathbb{E}[X_j - X_{j-1}]}_{\text{Correcting layer}}, \tag{1.5}$$

the multilevel method consists in building a procedure based on the addition of Monte-Carlo approximations of  $\mathbb{E}[X_0]$  and of  $\mathbb{E}[X_j - X_{j-1}]$ ,  $j = 1, \dots, J$ . Then, if the random variables  $X_j - X_{j-1}$  have low variance, the approximation of  $\mathbb{E}[X_j - X_{j-1}]$  requires few simulations and, in view of (1.5), we can obtain a procedure which has the bias related to  $X_J$  but with a cost which may be much less than the one generated by a standard Monte-Carlo method applied to estimate  $\mathbb{E}[X_J]$ .

In the discretization setting, the family of random variables  $(X_j)_j$  is a sequence of Euler schemes  $(\bar{X}^{\gamma_j})_j$  where  $(\gamma_j)_j$  is a family of decreasing time steps.<sup>2</sup> Following the heuristic (1.5), the (independent) correcting layers are built by coupling of Euler schemes with steps  $\gamma_{j-1}$  and  $\gamma_j$ . Note that in view of the simulation of the (synchronous) coupling, we need  $\gamma_{j-1}$  to be a multiple of  $\gamma_j$  (in this paper, we will assume that  $\gamma_j = \gamma_0 2^{-j}$ ).

Multilevel methods have been already studied in the literature for the approximation of the invariant distribution of the Langevin diffusion [16, 17, 27, 30, 35]. In [16], the authors take advantage of the convergence in distribution to equilibrium. Thus, the classical Monte-Carlo point of view is adopted: the approximation of  $\pi(f)$  is obtained by sampling a large number of Euler schemes for each level. In [30]<sup>3</sup> and [12], the point of view is to take advantage of the convergence of the occupation measure. Thus, each level is based on only one path of the Euler scheme or of the couple of Euler schemes

<sup>2</sup>Decreasing according the levels but constant for each layer.

<sup>3</sup>This paper is written in the multiplicative setting with a so-called Multilevel-Romberg point of view.

whose length decreases (since the variance of the correcting layers decreases) and discretization step increases. All these papers show that, in the uniformly strongly convex setting, the integral  $\pi(f)$  against the invariant distribution can be approximated (for a Lipschitz continuous functions  $f$ ) with a precision  $\varepsilon$  (in a  $L^2$ -sense) using a Multilevel procedure whose complexity is of order  $\varepsilon^{-2}$  or  $\varepsilon^{-2} \log^p(\varepsilon)$  with  $p \in [1, 3]$ . Moreover, in [12], a particular attention is paid to the dependency in the dimension. In this case, it is shown that one can build a multilevel procedure that produces an  $\varepsilon$ -approximation of the target for a complexity cost proportional to  $d\varepsilon^{-2}$  (with an explicit expression of the dependence in the Lipschitz constant of the gradient of  $U$  and the contraction parameter  $\alpha$ ).

The more involved weakly convex case seems to be less explored in the multilevel paradigm but, in view of applications (for instance for Bayesian Lasso), it is natural to ask about the robustness of these methods when one relaxes the contraction assumption.

## 1.2 Contributions and plan of the paper

The main goal of this paper is thus to extend the multilevel Langevin algorithm for the Gibbs sampling to the weakly convex setting, and if possible to obtain some quantitative bounds for the complexity related to the computation of an  $\varepsilon$ -approximation of the target (see Section 1.4 for a definition of  $\varepsilon$ -approximation).

We first investigate the *penalized multilevel* method: in the continuity of [6] and [8], we build a multilevel procedure based on the following observation: consider a new equation with another potential  $U_\alpha(x) := U(x) + \frac{\alpha}{2}|x|^2$ , this new equation has an invariant distribution named  $\pi_\alpha$  which converges to  $\pi$  when  $\alpha$  tends to 0 in Wasserstein metric. The idea is that this new invariant distribution is easier to sample because of the uniform convexity of the potential  $U_\alpha$ . In Section 2.1, Theorem 2.1 combines the benefits of the penalized approach and of the multilevel methods. For a Lipschitz continuous function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  and a  $C^2$ -potential  $U$ , the multilevel procedure performs an  $\varepsilon$ -approximation of  $\pi(f)$  with a complexity cost proportional to  $\pi(|\cdot|^2)^3 d\varepsilon^{-5}$ . As in [8], our result depends on the generally unknown constant  $\pi(|\cdot|^2)$  which is at least proportional to  $d$  (see Remark 2.2 for details and comparisons with [8]).

Because of the above remarks, we chose in a second part to try to develop some tools which tackle the weakly convex setting from a dynamic point of view and which can improve the complexity in terms of  $\varepsilon$ . More precisely, in the spirit of [14], we study an intermediary framework (called weakly parametric convex setting in the sequel). We assume that the eigenvalues of the Hessian matrix of  $U$  vanish when  $|x|$  goes to  $+\infty$ , but with a rate controlled by the function  $x \mapsto U^{-r}(x)$  with  $r \in [0, 1)$  (see Assumption  $(\mathbf{H}_r^1)$ ). The parameter  $r$  characterizes the “lack of strong convexity”, the case  $r = 1$  referring to the “Laplace case”<sup>4</sup> whereas  $r = 0$  corresponds to the uniformly convex setting. When one assumes such an assumption, one can get some bounds for the exponential moments of the Euler scheme (at the price of technicalities). One is also able to preserve some *confluence* properties, *i.e.* two paths of the Euler scheme have a tendency to stick at infinity. Finally, in this setting, it is also possible to control the distance between diffusion paths and the related Euler schemes. These three ingredients (obtained with a lower quality than in the strongly convex setting) allow us to tackle the multilevel procedure in this framework.

<sup>4</sup>The “Laplace case” will refer to the setting where the potential has a flat gradient. The simplest example is  $U(x) = |x|$ . In this case, the invariant distribution is a Laplace distribution. That is the reason why we use this terminology.

The related main contribution is Theorem 2.3. In this result, we provide a series of statements under different sets of assumptions: when  $U$  is only  $\mathcal{C}^2$  or when  $U$  is  $\mathcal{C}^3$ . Under  $(\mathbf{H}_r^1)$  only or under  $(\mathbf{H}_r^1)$  and  $(\mathbf{H}_r^2)$ , where  $(\mathbf{H}_r^2)$  denotes an additional assumption which requires the highest eigenvalue to be also controlled by the function  $x \mapsto U^{-r}(x)$  (we could roughly say that under  $(\mathbf{H}_r^1)$  and  $(\mathbf{H}_r^2)$ , the potential is *uniformly weakly convex* in the sense that “the decrease of the contraction is uniform”). In each statement, we provide a multilevel procedure adapted to the assumptions. The related complexity is exhibited in terms of  $d, \varepsilon$  but also in terms of the contraction parameter and the Lipschitz constant of  $\nabla U$ . Without going too much into the details, when  $U$  is only  $\mathcal{C}^2$ , the complexity is of the order  $\varepsilon^{-3}$  whereas when  $U$  is  $\mathcal{C}^3$ , we can obtain a rate of the order  $\varepsilon^{-2-\rho}$  for any  $\rho > 0$  and thus approach the “optimal” complexity  $\varepsilon^{-2}$ . Now, in terms of the dimension, when only  $(\mathbf{H}_r^1)$  holds the dependence in the dimension of the complexity is bounded by  $d^{\frac{\frac{3}{2}+(\frac{9}{2}+\delta)r}{1-r}}$  whereas when  $(\mathbf{H}_r^1)$  and  $(\mathbf{H}_r^2)$  hold, we obtain in  $d^{\frac{3}{2}+(\frac{9}{2}+\delta)r}$  when  $U$  is  $\mathcal{C}^2$  and  $d^{1+\frac{\rho}{2}+4r}$  for any  $\rho > 0$  when  $U$  is  $\mathcal{C}^3$ . This means that when  $U$  is  $\mathcal{C}^3$  and  $(\mathbf{H}_r^1)$  and  $(\mathbf{H}_r^2)$  hold, the complexity is of the order  $\varepsilon^{-2-\rho}d^{1+\frac{\rho}{2}+4r}$  for any  $\rho > 0$ . With respect to the paper [14], our multilevel procedure improves the dependence in  $\varepsilon$  and is most comparable in terms of the dimension. Note that when only  $(\mathbf{H}_r^1)$  holds, the dependency in the dimension dramatically increases on  $r$ . Whereas, when the potential is *uniformly weakly convex*, the dependence in the dimension does not explode when  $r \rightarrow 1$  (see Theorem 2.4 for more details).

**Plan of the paper** As detailed in the previous paragraphs, Sections 2.1 and 2.2 are respectively devoted to the statement of the main theorems for the penalized multilevel and in the parametric weakly convex case. Then, Section 3 is dedicated to the proof of the first main theorem (Theorem 2.1). In Proposition 3.2, we obtain a Wasserstein bound related to the bias induced by the penalization on the invariant distribution. The proof of Theorem 2.1 is then an adaptation of [12, Theo 2.2]. From Section 4, we focus on the proof of Theorem 2.3. In Section 4, we prove some preliminary results on the diffusion and its Euler scheme under  $(\mathbf{H}_r^1)$ : we begin with some controls of the exponential moments (Proposition 4.1 and Proposition 4.3) which in turn lead to some bounds of the polynomial moments (Proposition 4.5). In this section, we also show that the discretization error can be controlled in long time (Proposition 4.7) and finally obtain an integrable rate of convergence to equilibrium for the Euler scheme (Proposition 4.11). With the help of these fundamental bricks, in Section 5, we obtain some bounds on the bias (Proposition 4.17) and of the variance of the procedure (Proposition 5.5) which in turn allow us to finally provide the proof of Theorem 2.1 in Theorem 2.3.

### 1.3 Design of the algorithm

We now build the multilevel procedure. Let  $x \in \mathbb{R}^d$  be the initialization of the procedure,  $J \in \mathbb{N}$  be a number of levels,  $(\gamma_j)_{0 \leq j \leq J}$  be a sequence of time steps,  $(T_j)_{0 \leq j \leq J}$  be a sequence of final times. Define by  $\mathcal{Y}(J, (\gamma_j)_j, \tau, (T_j)_j, x, \cdot)$  the *multilevel occupation measure*: for all  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ ,

$$\begin{aligned} \mathcal{Y}(J, (\gamma_j)_j, \tau, (T_j)_j, x, \alpha, f) &:= \frac{1}{T_0 - \tau} \int_{\tau}^{T_0} f(\bar{X}_{\underline{s}_{\gamma_0}}^{\gamma_0, x}) ds \\ &+ \sum_{j=1}^J \frac{1}{T_j - \tau} \int_{\tau}^{T_j} f(\bar{X}_{\underline{s}_{\gamma_{j-1}}}^{\gamma_j, x}) - f(\bar{X}_{\underline{s}_{\gamma_{j-1}}}^{\gamma_{j-1}, x}) ds, \end{aligned} \tag{1.6}$$

where trajectories on each level are coupled with the same Brownian motion. To ease notation, we simplify  $\mathcal{Y}(J, (\gamma_j)_j, \tau, (T_j)_j, x, f)$  by  $\mathcal{Y}(f)$ . The parameter  $\tau \geq 0$  is the time

where we begin the average. Indeed, this “warm-start trick” can improve the precision of the estimation in the case of exponential convergence to equilibrium of the process toward its invariant distribution (see [12] for more information about this parameter). However, we don’t always have this rate of convergence, and without it, the trick becomes less effective. In this situation, we set  $\tau$  to 0 (as in the second part of this work).

**1.4 Notations**

The usual scalar product on  $\mathbb{R}^d$  and the induced Euclidean norm are respectively denoted by  $\langle \cdot, \cdot \rangle$  and  $|\cdot|$ . The set  $\mathcal{M}_{d,d}$  refers to the set of  $d \times d$ -real matrices, we denote by  $\|\cdot\|$  the operator norm associated with the Euclidean norm. For a symmetric matrix  $A$ , we denote respectively by  $\underline{\lambda}_A$  and  $\bar{\lambda}_A$  its lowest and highest eigenvalues. The Frobenius norm for  $A \in \mathcal{M}_{d,d}$  is denoted by  $\|A\|_F$ .

The Lipschitz constant of a given Lipschitz continuous function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is denoted by  $[f]_1$ . A function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is  $C^k$ ,  $k \in \mathbb{N}$ , if all its partial derivatives are well-defined and continuous up to order  $k$ . The gradient and Hessian matrix of  $f$  are respectively denoted by  $\nabla f$  and  $D^2 f$ . The probability space is denoted by  $(\Omega, \mathcal{F}, \mathbb{P})$ . The Laplace operator is denoted by  $\Delta$ :  $\Delta f = \sum_{i=1}^d \partial_{i,i}^2 f$ . The  $L^p$ -norm on  $(\Omega, \mathcal{F}, \mathbb{P})$  is denoted by  $\|\cdot\|_p$ . For two probability measures  $\mu$  and  $\nu$ , we define the Wasserstein distance of order  $p$  by

$$\mathcal{W}_p(\mu, \nu) = \inf_{\zeta \in \Pi(\mu, \nu)} \left( \int_{\mathbb{R}^d} |x - y|^p d\zeta(x, y) \right)^{\frac{1}{p}},$$

where  $\Pi(\mu, \nu)$  is the set of couplings of  $\mu$  and  $\nu$ .

- $\lesssim_P$  **and**  $\lesssim_{uc}$ : For two positive real numbers  $a$  and  $b$  and a set of parameters  $P$ , one writes  $a \lesssim_P b$  if  $a \leq C_P b$  where  $C_P$  is a positive constant which only depends on the parameters  $P$ . When  $a \leq Cb$  where  $C$  is a universal constant, we write  $a \lesssim_{uc} b$ .
- **$\varepsilon$ -approximation**: We say that  $\mathcal{Y}$  is an  $\varepsilon$ -approximation (or more precisely an  $\varepsilon$ -approximation of  $a$  for the  $L^2$ -norm) if  $\|\mathcal{Y} - a\|_2 = \mathbb{E}[|\mathcal{Y} - a|^2]^{\frac{1}{2}} \leq \varepsilon$ . Equivalently,  $\mathcal{Y}$  is said to be an  $\varepsilon$ -approximation of  $a$  if the related Mean-Squared Error (MSE) is lower than  $\varepsilon^2$ .
- **Complexity/ $\varepsilon$ -complexity**: For a random variable  $\mathcal{Y}$  built with some iterations of a standard Euler scheme, we denote by  $\mathcal{C}(\mathcal{Y})$ , the number of iterations of the Euler scheme which is needed to compute  $\mathcal{Y}$ . For instance,  $\mathcal{C}(\bar{X}_{n\gamma}^\gamma) = n$ . We sometimes call  $\varepsilon$ -complexity of the algorithm, the complexity related to the algorithm which produces an  $\varepsilon$ -approximation.

**2 Main results**

**2.1 The penalized approach**

In this section, we develop a *penalized multilevel method* to sample a non-strongly log-concave probability distribution  $\pi$ . The idea is based on [8] and [7] where the authors consider the potential  $U^\alpha(x) := U(x) + \frac{\alpha}{2}|x|^2$  with  $\alpha > 0$  which is called *penalized* version of  $U$ . We here assume that  $U$  satisfies the following assumption:

**WC<sub>L</sub>**:  $U$  is a non-zero  $C^2$ -function and there exists  $L > 0$  such that

$$\forall x \in \mathbb{R}^d, \quad 0 \preceq D^2 U(x) \preceq L \text{Id}_{\mathbb{R}^d}, \tag{2.1}$$

the inequalities being taken in the sense of symmetric matrices. Denote by  $\pi_\alpha$  the invariant measure of the diffusion process  $(X_t^\alpha)_{t \geq 0}$  solution of the stochastic differential equation

$$dX_t^\alpha = -\nabla U^\alpha(X_t^\alpha)dt + \sigma dB_t. \tag{2.2}$$

It appears that  $\pi_\alpha$  satisfies the Bakry-Émery criterion thus we can apply our multilevel method that requires strong convexity to approximate  $\pi_\alpha$ . But our target is  $\pi$  then we have to control the distance (in a Wasserstein sense) between  $\pi$  and  $\pi_\alpha$ . To this end, the results of [3] and [8] ensure the convergence of  $\pi_\alpha$  when  $\alpha$  goes to 0 with a bound of the Kullback-Leibler divergence. This leads to the following theorem:

**Theorem 2.1.** *Assume  $\text{WC}_L$ . Let  $\varepsilon > 0$ , let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a Lipschitz continuous function,  $x \in \mathbb{R}^d$  and define  $\mathcal{Y}^\alpha$  as the multilevel procedure described in (1.6) using the Euler schemes of the SDE (2.2). Then, there exist parameters such that,*

$$\|\mathcal{Y}^\alpha(f) - \pi(f)\|_2 \leq \varepsilon, \tag{2.3}$$

with a complexity of order  $\varepsilon^{-5} d \pi(|\cdot|^2)^{9/2}$  up to logarithmic terms. More precisely, the following parameters lead to the result.

$$\begin{aligned} J &= \left\lceil 2 \log_2(15\sqrt{2} d m_2^{3/2} \varepsilon^{-2}) \right\rceil, \quad \tau = \frac{15\sqrt{2}}{\sigma^2 m_2^{3/2}} \varepsilon^{-1} \log(15\sqrt{2} d m_2^{3/2} \varepsilon^{-1}), \\ \alpha &= \frac{\varepsilon \sigma^2}{15\sqrt{2} m_2^{3/2}}, \quad \gamma_0 = \frac{\varepsilon}{15\sqrt{2} m_2^{3/2} L^2}, \quad T_0 = \frac{550}{\sigma^2} d \log(\gamma_0^{-1}) m_2^3 \varepsilon^{-5} J^2, \\ \forall j \in \{0, \dots, J\}, \quad \gamma_j &= \gamma_0 2^{-j}, \quad T_j = T_0 2^{-j}, \end{aligned} \tag{2.4}$$

where  $m_2 = \mathbb{E}_\pi[|\cdot|^2]$ . Furthermore, the complexity satisfies

$$\mathcal{C}(\mathcal{Y}) \leq \frac{10 \times 15^3}{\sigma^4} d m_2^{9/2} \varepsilon^{-5} L^2 \left[ \log_2 \left( 15\sqrt{2} m_2^{3/2} d \varepsilon^{-3} \right) \right]^3 \log \left( 15\sqrt{2} m_2^{3/2} L^2 \varepsilon^{-1} \sigma^{-2} \right). \tag{2.5}$$

**Remark 2.2.**  $\triangleright$  Note that the dimension dependency is not really linear because there may be some dependency hidden in  $m_2$ . Indeed, when  $\pi$  is an isotropic probability,  $\pi(|\cdot|^2) \lesssim_{uc} d$  (see e.g. [1]<sup>5</sup>) and this implies that the complexity is of the order  $d^{11/2} \varepsilon^{-5}$ .

$\triangleright$  Let us now compare with [8]: note that in this paper, the cost is not explicitly written. As usual in the Monte-Carlo literature, the authors control the number of iterations of the Euler scheme which is necessary to draw a random variable whose distance to the target is lower than  $\varepsilon \sqrt{m_2}$  instead of giving the real cost. In the Langevin Monte-Carlo case, when  $U$  is  $\mathcal{C}^2$ , they then obtain a number of iterations which is, up to logarithmic terms, of order  $m_2 d \varepsilon^{-4}$ . Normalizing  $\varepsilon$  (i.e. replacing  $\varepsilon$  by  $\varepsilon/\sqrt{m_2}$ ) leads to a number of iterations of order  $m_2^3 d \varepsilon^{-4}$ . But, to compare with our work, we need to include the Monte-Carlo cost, i.e. to the number of simulations which is necessary to make the variance lower than  $\varepsilon^2$ . Then, we have to multiply the previous number of iterations by  $\text{Var}_{\pi_\alpha^\gamma}(f) \varepsilon^{-2}$ . The moment of order two of  $\pi_\alpha^\gamma$  studied in [12] can be reasonably bounded by  $d/\alpha$  (since  $f$  is Lipschitz continuous). This means that the complexity of the penalized Langevin Monte-Carlo in [8] is of the order  $m_2^{9/2} d^2 \varepsilon^{-7}$  ( $\lesssim d^{13/2} \varepsilon^{-7}$  if  $\pi$  is isotropic). In consequence, the multilevel method allows us to improve the result of [8]. Note that the authors also provide other algorithms such as the Kinetic-LMC where the bound in  $\varepsilon$  is improved.

$\triangleright$  In the above result, we propose a Multilevel strategy based on a fixed penalization. A natural question arises: could we take advantage of the Multilevel strategy by keeping the same penalization for the highest level and progressively reducing it on the lower layers? Indeed, this is precisely what we do with the discretization bias, thus we can wonder about the effect of such a strategy for penalization. Indeed, the difference in Wasserstein distance between two trajectories with different penalization parameters

<sup>5</sup>Note that the control of  $\pi(|\cdot|^2)$  is strongly linked to the so-called KLS-conjecture.

can be controlled. However, our computations show that the benefit of such a strategy does not improve the dependency of the interest quantity ( $d, \varepsilon, L$ , etc...) in the complexity cost. Therefore, to not overload the document with superficial technical issues, we have chosen not to present this strategy.

## 2.2 Parametric weakly convex setting

The purpose of this section is to study the non-penalized multilevel procedure in the weakly convex setting. Instead of penalizing the dynamics, it is actually natural to ask about the robustness of the “standard” multilevel method in this case. To answer this question, we have to prove a series of properties in the spirit of the assumptions  $\mathbf{H}_i$  ( $i \in \{1, 2, 3, 4\}$ ) in [12]. These assumptions include the convergence to equilibrium of the Euler scheme with a quantitative rate, the long-time control of the  $L^2$ -distance between the Euler scheme and the diffusion, the control of the Wasserstein distance between  $\pi^\gamma$  and  $\pi$  and the control of the moments. Some of these properties (especially the long-time control of the  $L^2$ -distance) seem hard to check in a general convex setting. We thus propose to work in the parametric weakly convex setting used in [12] by introducing  $(\mathbf{H}_r^1)$  (see below) where we assume that the contraction vanishes at infinity but with a rate controlled by  $U^{-r}$ .

Let us now introduce our assumptions depending on a parameter  $r \in [0, 1)$ :

$(\mathbf{H}_r^1)$ : The potential  $U$  is a positive  $C^2$ -convex function with a unique minimum  $x^*$  such that  $U(x^*) = 1$ .  $\nabla U$  is  $L$ -Lipschitz with  $L \geq 1$ .<sup>6</sup> The function  $x \mapsto \underline{\lambda}_{D^2U}(x)$  is positive and there exist  $L$  and  $\underline{c} > 0$  such that,

$$\forall x \in \mathbb{R}^d, \quad \underline{\lambda}_{D^2U}(x) \geq \underline{c}U^{-r}(x).$$

The lower-bound can be seen as the “lack of uniform strong convexity” for the potential. Indeed, if  $r = 0$  we recover the strong convexity and  $r = 1$  corresponds to the weakest convexity case where the gradient is flat at infinity. Note that the assumption on the unique minimum is restrictive in the context of sampling or optimization. It is a counterpart to the efficiency of the Multilevel method. Indeed, we need a (strong) confluence between two trajectories of the Euler scheme to control the variance of each layer. Although, weak confluence can be obtained when there are several minima (using for example the arguments of [34]), the strong confluence is difficult to achieve without this assumption.

The fact that  $\nabla U$  is  $L$ -Lipschitz implies that  $x \mapsto \bar{\lambda}_{D^2U}(x)$  is upper-bounded by  $L$ . In order to improve the dependence in the dimension, we also introduce an additional assumption that deals with the case where the largest eigenvalue decreases at infinity with an intensity that is of the same order as the lowest eigenvalue:

$(\mathbf{H}_r^2)$ : There exists a positive  $\bar{c}$  such that for all  $x \in \mathbb{R}^d$ ,  $\bar{\lambda}_{D^2U}(x) \leq \bar{c}U^{-r}(x)$ .

This second assumption is restrictive because we force the potential to be uniformly flat at the infinity. However, there are a lot of interesting cases where the assumption  $(\mathbf{H}_r^2)$  holds. As an example, the potential  $x \mapsto (1 + |x|^2)^{1/r+1}$  for  $r \in [0, 1)$  satisfies  $(\mathbf{H}_r^1)$  and  $(\mathbf{H}_r^2)$ . This type of potential arises, for example, in high-dimensional Bayesian learning (in the construction of the posterior distribution). More precisely, the ridge Bayesian method needs to compute the mean of a posterior distribution satisfying  $(\mathbf{H}_r^1)$  and  $(\mathbf{H}_r^2)$

<sup>6</sup>The fact that  $L$  is greater than 1 is clearly not fundamental but allows to simplify the usually technical expressions which appear in the sequel.

for  $r = 0$  while the Lasso Bayesian method needs to compute the mean of a posterior distribution satisfying  $(\mathbf{H}_r^1)$  and  $(\mathbf{H}_r^2)$  for  $r = 1$ . Although our study does not deal with the  $r = 1$  case, it is an intermediary between the simple case of the ridge and the more difficult case of the Lasso.

Finally, as in [12], we can reduce the complexity cost by assuming greater regularity for the potential. Let us therefore establish the following assumption:

$(\mathbf{H}^3)$ :  $U$  is a  $\mathcal{C}^3$ -function.

To facilitate the comprehension of the rest of this section, we define four settings corresponding to different set of assumptions. The assumptions required for each setting are summarized in Table 1.

Table 1: Descriptions of different sets of assumptions.

Setting 1	$(\mathbf{H}_r^1)$
Setting 2	$(\mathbf{H}_r^1)$ and $(\mathbf{H}_r^2)$
Setting 3	$(\mathbf{H}_r^1)$ and $(\mathbf{H}^3)$
Setting 4	$(\mathbf{H}_r^1)$ , $(\mathbf{H}_r^2)$ and $(\mathbf{H}^3)$

Now, let us define the couple  $(\gamma^*, \bar{\Psi})$  in function of the setting by:

$$(\gamma^*, \bar{\Psi}) = \begin{cases} \left( \frac{1}{4L}, (1 + \sigma^2) \left( dL + \left( 1 + \frac{dL}{\underline{c}} \right)^{\frac{1}{1-r}} \right) \right) & \text{in Setting 1 or 3} \\ \left( \frac{1-r}{4(\bar{c}vL)}, c_r \frac{d(1+\sigma^2)(\bar{c}vL)}{\underline{c}} \right) & \text{in Setting 2 or 4,} \end{cases} \tag{2.6}$$

where  $c_r$  is a constant which only depends on  $r$ .  $\gamma^*$  will denote the largest value for  $\gamma_0$  whereas,  $\bar{\Psi}$  controls the moments of  $U(\bar{X}_{n\gamma})$  (see Proposition 4.5 for details). It is important to note that the (2.6) constants are better under stronger assumptions in the sense that they depend less on dimension. More precisely,  $\gamma^*$  depends on  $1/d$  when only  $(\mathbf{H}_r^1)$  holds whereas it does not depend on  $d$  when both assumptions  $(\mathbf{H}_r^1)$  and  $(\mathbf{H}_r^2)$  hold. As a consequence, the time step of our schemes can be larger in the second case which automatically leads to a reduction of the complexity cost. Regarding  $\bar{\Psi}$ , it depends on  $d^{\frac{1}{1-r}}$  when only  $(\mathbf{H}_r^1)$  holds and on  $d$  when both assumptions  $(\mathbf{H}_r^1)$  and  $(\mathbf{H}_r^2)$  hold. The control of the moment  $U(\bar{X}_{n\gamma})$  is then better in the second case since  $r < 1$  which is also an advantage for our method. The following result covers the 4 different settings of Table 1. In particular, the reader have to keep in mind that the  $\gamma^*$  and  $\bar{\Psi}$  values depend on the setting we are in.

We are now ready to state our main theorem in this setting:

**Theorem 2.3.** *Assume the assumptions of one of the settings described in Table 1. Let  $\gamma^*$  and  $\bar{\Psi}$  be the parameters defined in (2.6). Let  $\varepsilon > 0$ , let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a Lipschitz continuous function,  $\gamma_0 \in (0, \gamma^*]$ ,  $\delta \in (0, 1/4]$  and  $x \in \mathbb{R}^d$  such that  $U(x) \lesssim_r \bar{\Psi}$ . Let  $\mathcal{Y}$  be the multilevel procedure described in (1.6) where  $\tau$  is set to 0.*

(i) *In setting 1 and 2, for  $\delta$  small enough, there exist parameters such that*

$$\|\mathcal{Y}(f) - \pi(f)\|_2 \lesssim_{r,\delta} \varepsilon. \tag{2.7}$$

*Moreover, the complexity satisfies*

$$\mathcal{C}(\mathcal{Y}) \leq \sqrt{L}(\underline{c}^{-\frac{5}{4}} \wedge \underline{c}^{-\frac{7}{2}-\delta})\bar{\Psi}^{\frac{3}{2}+(\frac{9}{2}+\delta)r}\varepsilon^{-3}, \tag{2.8}$$

(ii) In setting 3 and 4, assume in addition that  $\sup_{x \in \mathbb{R}^d} \sum_{i=1}^d |\Delta(\partial_i U)|^2 \lesssim_r \sigma^{-4} L^3 \bar{\Psi}$  and let  $\rho \in [1/2, 1)$ . Then, for  $\delta$  small enough, there exists parameters such that

$$\|\mathcal{Y}(f) - \pi(f)\|_2 \lesssim_{r,\delta} \varepsilon. \tag{2.9}$$

Moreover, the complexity satisfies

$$\mathcal{C}(\mathcal{Y}) \leq \gamma_0^{-\rho} L^{2\rho} \left( \underline{c}^{\frac{5}{4} \wedge (2\rho) + \delta} \vee \underline{c}^{-\frac{5}{2} - \delta} \right) \bar{\Psi}^{1 + \frac{\rho}{2} + (4 - \rho + \delta)r} \varepsilon^{-2 - \rho}. \tag{2.10}$$

In particular, the following parameters lead to the result,

$$J = \begin{cases} \left\lceil \log_2 \left( \frac{L}{\underline{c}^{\frac{1}{1-\delta}} \wedge \underline{c}} \bar{\Psi}^{1 + (3 + \delta)r} \gamma_0 \varepsilon^{-2} \right) \right\rceil & \text{in Setting 1 and 2} \\ \left\lceil \log_2 \left( \underline{c}^{-\frac{2}{1-\delta}} L^3 \bar{\Psi}^{1 + \frac{2r}{1-\delta}} \gamma_0 \varepsilon^{-1} \right) \right\rceil & \text{in Setting 3 and 4,} \end{cases}$$

$$T_0 = \begin{cases} (\underline{c}^{-\frac{3}{4}} \vee \underline{c}^{-\frac{5}{2} - \delta}) \bar{\Psi}^{\frac{3}{2} + (\frac{9}{2} + \delta)r} \varepsilon^{-2} & \text{in Setting 1 and 2} \\ L^{\frac{\rho}{2}} \left( \underline{c}^{-(\frac{5}{4} - \rho) \wedge (3\rho) + \delta} \vee \underline{c}^{-\frac{5}{2} - \delta} \right) \bar{\Psi}^{1 + (4 - 2\rho + \delta)r} \varepsilon^{-2} & \text{in Setting 3 and 4,} \end{cases}$$

for all  $j \in \{1, \dots, J\}$ ,  $\gamma_j = \gamma_0 2^{-j}$  and

$$T_j = \begin{cases} T_0 2^{-\frac{j}{2}} & \text{in Setting 1 and 2} \\ T_0 2^{-(1-\rho)j} & \text{in Setting 3 and 4.} \end{cases}$$

To facilitate the comprehension of this theorem, Table 2 summarizes the dependency on dimension and precision  $\varepsilon$  for each setting. Additionally, this technical result deserves several comments:

Table 2: Order in the dimension and  $\varepsilon$  of theoretical complexity for  $\varepsilon$ -approximation by the Multilevel procedure in each setting.

	Setting 1	Setting 2	Setting 3	Setting 4
$d$ -dependency	$d^{\frac{1 + \frac{3}{2} + (\frac{9}{2} + \delta)r}{1-r}}$	$d^{1 + \frac{3}{2} + (\frac{9}{2} + \delta)r}$	$d^{\frac{1 + \frac{\rho}{2} + (4 - \rho + \delta)r}{1-r}}$	$d^{1 + \frac{\rho}{2} + (4 - \rho + \delta)r}$
$\varepsilon$ -dependency	$\varepsilon^{-3}$	$\varepsilon^{-3}$	$\varepsilon^{-2 - \rho}$	$\varepsilon^{-2 - \rho}$

**Remark 2.4.**  $\triangleright$  *Complexity in terms of  $\varepsilon$ .* If we only consider the dependence in  $\varepsilon$ , we obtain  $\varepsilon^{-3}$  when  $U$  is only  $\mathcal{C}^2$  and  $\varepsilon^{-2 - \rho}$  for any  $\rho > 0$  when  $U$  is  $\mathcal{C}^3$  and an additional (but reasonable<sup>7</sup>) assumption on  $\Delta(\nabla U)$  is satisfied. We can thus theoretically approach the complexity in  $\varepsilon^{-2}$ . However, it is worth noting that the non-explicit constants depending on  $\rho$  and  $\delta$  go to  $\infty$  (independently of the other parameters) when  $\rho$  and  $\delta$  go to 0. The fact that we “only” obtain a complexity in  $\varepsilon^{-3}$  when  $U$  is  $\mathcal{C}^2$  is due to the fact that in this case, our bound of the 1-Wasserstein distance between  $\pi^\gamma$  and  $\pi$  is of the order  $\sqrt{\gamma}$ . When  $U$  is  $\mathcal{C}^3$ , the bound on the 1-Wasserstein distance between  $\pi^\gamma$  and  $\pi$  is of order  $\gamma$ . This allows us to clearly improve the complexity but it can be noted that we do not retrieve the  $\varepsilon^{-2}$ -bound of the uniformly convex case. This is due to the rate of convergence to

<sup>7</sup>See Remark 4.22 for details on this assumption.

equilibrium. Actually, our rate is polynomial and not exponential, which in turn, implies a “slight cost” on the dependence in  $\varepsilon$ . In fact, we could get some (sub)-exponential rates but without controlling the dependency in the dimension which is of first importance for applications.

▷ *Complexity in terms of the dimension.* The dependence in the dimension strongly varies with the assumptions. In the “worst” case in setting 1 (i.e. where  $U$  is only  $\mathcal{C}^2$  and only  $(\mathbf{H}_r^1)$  holds), the complexity is of the order  $d^{\frac{\frac{3}{2}+(\frac{3}{2}+\delta)r}{1-r}}$ . Unfortunately, when  $r$  is close to 1, this means that this dependence seriously worsens. We retrieve this same phenomenon in setting 3 (i.e. when  $U$  is  $\mathcal{C}^3$  and only  $(\mathbf{H}_r^1)$  holds) but with a better bound of the order  $d^{\frac{1+\frac{\rho}{2}+(4-\rho+\delta)r}{1-r}}$  for any positive  $\rho$  and  $\delta$ . This bad behavior when  $r$  goes to 1 is due to the fact that when only  $(\mathbf{H}_r^1)$  holds, the bounds on the exponential moments of  $\sup_{t \geq 0} \mathbb{E}[e^{U(\bar{X}_t)}]$  are of the order  $\exp(d^{\frac{1}{1-r}})$ . Introducing  $(\mathbf{H}_r^2)$  in settings 2 and 4 dramatically improves this exponential bound since in this case, we are able to prove that this is of the order  $e^d$  (this implies that  $\sup_{t \geq 0} \mathbb{E}[U^p(\bar{X}_t)]$  is of the order  $d^p$ , see Propositions 4.5 and 4.3 for details). It is worth noting that in this case, the dependence in the dimension does not explode when  $r$  goes to 1 being of the order  $d^{1+\frac{\rho}{2}+(4-\rho+\delta)r}$  for any  $\rho > 0$ . Remark that when  $r = 0$ , we formally approach the rate of the uniformly convex case in  $d\varepsilon^{-2}$  obtained in [12].

▷ *Comparison with the literature:* In this setting, the only paper which we may reasonably compare with is [14] since we use similar assumptions. Compared with this paper, our multilevel procedure certainly improves the dependence in  $\varepsilon$ , replacing  $\varepsilon^{-4}$  by  $\varepsilon^{-3}$  when  $U$  is  $\mathcal{C}^2$  and  $\varepsilon^{-3}$  by  $\varepsilon^{-2-\rho}$  when  $U$  is  $\mathcal{C}^3$ . In terms of the dimension, our approach slightly increases the dependence on the dimension. For instance, when  $(\mathbf{H}_r^1)$  and  $(\mathbf{H}_r^2)$  hold, [14] obtain a bound in  $d^{1+4r}$  when  $U$  is  $\mathcal{C}^2$  or  $\mathcal{C}^3$ . We here retrieve a dependence which is somewhat similar when  $U$  is  $\mathcal{C}^3$  but when  $U$  is  $\mathcal{C}^2$ , our bound in  $d^{\frac{3}{2}+(\frac{3}{2}+\delta)r}$  is clearly worse.

▷ *About the parameters.* In applications, the dependence in the parameters,  $L$ ,  $\underline{c}$ , and  $\bar{c}$  may be of importance (think for instance to applications to Bayesian estimation where these parameters can strongly depend on the number of observations). This is why here, we chose to keep all these dependencies in the main result even if it sometimes adds many technicalities in the proof.

### 3 Proof of Theorem 2.1

This section is devoted to the proof of the first main result. We first quantify the bias induced by the approximation of  $\pi$  by  $\pi_\alpha$ . To this end, we use the Talagrand concentration inequality that estimates the Wasserstein distance between these two measures by their Kullback-Leibler divergence.

**Proposition 3.1.** *Assume that  $\mathbf{WC}_L$  hold. Then for all  $\alpha \geq 0$ , there is a universal constant  $C$  such that*

$$\mathcal{W}_1(\pi, \pi_\alpha) \leq C\sqrt{D_{\text{KL}}(\pi|\pi_\alpha)}.$$

We refer to [3, Cor 2.4] to find a proof of this result. In addition, in [8] the authors show that  $C \leq \sqrt{2\mathbb{E}_\pi[|X|^2]}$  (page 24). It remains to compute the Kullback Leibler divergence of  $\pi$  from  $\pi_\alpha$ , to bound the bias induced by the penalization.

**Proposition 3.2.** *Assume that  $\mathbf{WC}_L$  hold. Then for all  $\alpha \geq 0$ ,*

$$\mathcal{W}_1(\pi, \pi_\alpha) \leq \frac{15\sqrt{2}\alpha}{\sigma^2} \mathbb{E}_\pi[|\cdot|^2]^{3/2}.$$

We refer to [8, Proposition 1] to find a proof of this result for  $\sigma = \sqrt{2}$ . We won't repeat it here since an easy adaptation leads to the result for all  $\sigma > 0$ .

Now we switch to the proof of the main theorem. With the two previous propositions, we control the bias induced by the penalization, then it remains to compute the error and the complexity of a multilevel procedure in a uniformly convex setting. To this end, we use [12, Theorem 2.2] which gives parameters to perform an  $\varepsilon$ -approximation of the invariant distribution with an explicit complexity in terms of the parameters, especially in terms of the *contraction parameter*. Here, this is exactly our penalization parameter  $\alpha$  and we will thus optimize its choice in the proof.

*Proof of Theorem 2.1.* Let  $\varepsilon$  be a positive number and  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a Lipschitz continuous function. By the bias/variance decomposition, triangular inequality and the Monge-Kantorovich duality, we have

$$\begin{aligned} \|\mathcal{Y}(f) - \pi(f)\|_2^2 &\leq \mathbb{E} [|\mathcal{Y}(f) - \pi(f)|]^2 + \text{Var}(\mathcal{Y}(f)) \\ &\leq 2|\pi_\alpha(f) - \pi(f)|^2 + 2\mathbb{E}[\mathcal{Y}(f) - \pi_\alpha(f)]^2 + \text{Var}(\mathcal{Y}(f)) \\ &\lesssim_{uc} \underbrace{\mathcal{W}_1^2(\pi, \pi_\alpha)}_{P_1} + \underbrace{\mathbb{E}[\mathcal{Y}(f) - \pi_\alpha(f)]^2 + \text{Var}(\mathcal{Y}(f))}_{P_2}. \end{aligned} \tag{3.1}$$

The second term denoted by  $P_2$  is the mean squared error of a Multilevel procedure for the approximation of  $\pi_\alpha$ . This penalized measure is invariant for the diffusion process defined with the potential  $U_\alpha$ . By assumption  $\mathbf{WC}_L$ ,  $U_\alpha$  satisfies the following property  $\mathbf{C}_\alpha$ : For all  $x, y \in \mathbb{R}^d$ ,

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

[12, Theorem 2.2] ensures that with  $\alpha/L^2 \leq 1$ ,  $\sigma^2\alpha^{-1}d \geq 1$ ,  $|U^\alpha(x_0)|^2 \lesssim_{uc} \sigma^2\alpha d$  and the following parameters<sup>8</sup>:

$$\begin{aligned} J &= \lceil 2 \log_2(\sigma^2 d \alpha^{-1} \varepsilon^{-1}) \rceil, \quad T_j = \frac{\sigma^2 d \log(\gamma_0^{-1})}{\alpha^2} \varepsilon^{-2} J^2 2^{-j}, \\ r &\in \{0, \dots, J\}, \quad \gamma_0 = \alpha/(2L^2), \quad \tau = \alpha^{-1} \log(\sigma^2 d \alpha^{-1}), \end{aligned} \tag{3.2}$$

we have

$$P_2 \leq \varepsilon,$$

with a complexity satisfying

$$\mathcal{C}(\mathcal{Y}) \leq 5 \log \left( \frac{2L^2}{\alpha} \right) \frac{L^2}{2\alpha^3} \sigma^2 d \varepsilon^{-2} \left[ \log_2 \left( \frac{\sigma^2 d}{\alpha} \varepsilon^{-2} \right) \right]^3. \tag{3.3}$$

It remains to calibrate the penalization parameter  $\alpha$ . Proposition 3.2 implies

$$P_1 \lesssim_{uc} \frac{550\alpha^2}{\sigma^4} \mathbb{E}_\pi[|\cdot|^2]^3,$$

so that  $P_1 \leq \varepsilon^2$  for

$$\alpha = \frac{\varepsilon \sigma^2}{15\sqrt{2} \mathbb{E}_\pi[|\cdot|^2]^{3/2}}.$$

Putting  $\alpha$  in (3.2) and (3.3),

$$\mathcal{C}(\mathcal{Y}) \leq \frac{10 \times 15^3}{\sigma^4} \varepsilon^{-5} d m_2^{9/2} L^2 \left[ \log_2 \left( 15\sqrt{2} m_2^{3/2} d \varepsilon^{-3} \right) \right]^3 \log \left( 15\sqrt{2} m_2^{3/2} L^2 \varepsilon^{-1} \sigma^{-2} \right),$$

where  $m_2 = \mathbb{E}_\pi[|\cdot|^2]$ . □

<sup>8</sup>To ease notation we have voluntarily omitted an assumption about  $\varepsilon$ : we have to consider  $\varepsilon$  small enough, we refer to [12] to get more precision.

#### 4 Preliminary bounds under $(\mathbf{H}_r^1)$ and $(\mathbf{H}_r^2)$ :

From now, we switch to the proof of the second part of the main results *i.e.* we consider the weakly convex case under the parametric assumptions  $(\mathbf{H}_r^1)$  and  $(\mathbf{H}_r^2)$ . As mentioned before, these hypotheses deal with the behavior of the lowest and highest eigenvalues of the Hessian matrix of  $U$ . In some sense,  $(\mathbf{H}_r^1)$  quantifies the strict convexity of the potential which in turn implies the contraction of the dynamics. Note that such an assumption also appears in [4] where the authors obtain exponential rates to equilibrium under this parametric assumption.

In this preliminary section we state a series of results related to the diffusion and its Euler scheme under Assumption  $(\mathbf{H}_r^1)$ . For the upper-bounds of the eigenvalues of  $D^2U$ , we distinguish two cases: the first one where we assume that we have a uniform upper-bound by  $L$  (in others words that  $\nabla U$  is  $L$ -Lipschitz) and the second one, we add Assumption  $(\mathbf{H}_r^2)$  where the largest eigenvalues also decrease at infinity with a rate which is comparable to the one of the lowest eigenvalues. In fact, in the second case, we will see that we are able to preserve a dependency of the moments in the dimension which is linear, whereas, without this assumption, the dependency is  $O(d^{\frac{1}{1-r}})$ .

In the second part we state a result about the longtime pathwise control of the distance between the diffusion and its Euler scheme. Third, we study the convergence to equilibrium for the Euler scheme. Finally, we quantify the bias induced by the discretization with some results on the 1-Wasserstein distance between  $\pi$  and  $\pi^\gamma$  (the invariant measure of the Euler scheme).

##### 4.1 Bounds on the exponential moment

In order to study the confluence between the continuous time process and its Euler scheme, let us start this section by a control of the moment of the continuous time process and the discrete time when the potential  $U$  is supposed convex. We first state a result on the control of the exponential moment of the continuous time process.

**Proposition 4.1.** *For all  $x \in \mathbb{R}^d$  and  $t > 0$ ,*

$$\sup_{t \geq 0} \mathbb{E}_x \left[ e^{\frac{U(X_t)}{\sigma^2}} \right] \leq e^{\frac{U(x)}{\sigma^2}} + \begin{cases} e^{\frac{1}{\sigma^2} \left(1 + \frac{dL}{2\bar{c}}\right)^{\frac{1}{1-r}}} & \text{under } (\mathbf{H}_r^1) \\ \frac{dL}{2\bar{c}} e^{\frac{1}{\sigma^2} \left(4^{\frac{1}{1-r}} \vee \frac{d\bar{c}}{\bar{c}}\right)} & \text{under } (\mathbf{H}_r^1) \text{ and } (\mathbf{H}_r^2). \end{cases}$$

We preface the proof by a technical lemma.

**Lemma 4.2.** *Let  $\theta \in (0, \frac{2}{\sigma^2})$  and  $M > 0$ , then*

$$\mathcal{C}_M := \left\{ x \in \mathbb{R}^d; (1 - \theta) |\nabla U(x)|^2 - \frac{\sigma^2}{2} \theta \Delta U(x) \leq M \right\} \subset \{x \in \mathbb{R}^d, U(x) \leq K_M\},$$

where,

$$K_M = \begin{cases} \left(1 + \frac{(2M + \theta \sigma^2 dL)}{2\bar{c}(2 - \theta \sigma^2)}\right)^{\frac{1}{1-r}} & \text{under } (\mathbf{H}_r^1) \\ \max \left( \left( \frac{M(1-r)}{(4 - 2\theta \sigma^2)\bar{c}} \vee 4 \right)^{\frac{1}{1-r}}, \frac{2\sigma^2 d\theta \bar{c}}{\bar{c}(2 - \theta \sigma^2)} \right) & \text{under } (\mathbf{H}_r^1) \text{ and } (\mathbf{H}_r^2). \end{cases}$$

In particular,  $\mathcal{C}_M$  is a compact set (since it is included in a level set of a coercive function).

*Proof.* Denote by  $y$  the solution of the ordinary differential equation  $y'(t) = -\nabla U(y(t))$  starting from  $y(0) = x$ . Define the function  $f : t \mapsto |\nabla U(y(t))|^2$  we have by chain rule for

all  $t \in \mathbb{R}^+$ . By  $(\mathbf{H}_r^1)$ ,

$$\begin{aligned} \frac{d}{dt}f(t) &= 2 \langle y''(t), y'(t) \rangle \\ &= 2 \langle D^2U(y(t))y'(t), y'(t) \rangle \\ &\geq 2\underline{c} \langle U^{-r}(y(t))y'(t), y'(t) \rangle = \frac{2\underline{c}}{1-r} \frac{d}{dt}(U^{1-r}(y(t))), \end{aligned}$$

Since  $\lim_{t \rightarrow +\infty} y(t) = x^*$ , we get by integration

$$|\nabla U(x)|^2 \geq \frac{2\underline{c}}{1-r} (U^{1-r}(x) - U^{1-r}(x^*)). \tag{4.1}$$

Therefore,

$$\left(1 - \frac{\theta\sigma^2}{2}\right) |\nabla U(x)|^2 - \frac{\sigma^2\theta}{2} \Delta U(x) \geq \frac{2(1-\theta)\underline{c}}{1-r} (U^{1-r}(x) - U^{1-r}(x^*)) - \frac{\sigma^2\theta}{2} \Delta U(x).$$

Since  $\Delta U = \text{Tr}(D^2U) \leq d\bar{\lambda}_{D^2U} \leq dL$  (where  $\bar{\lambda}_A$  stands for the largest eigenvalue of symmetric matrix  $A$ ) and  $U(x^*) = 1$ , it follows that

$$\left(1 - \frac{\theta\sigma^2}{2}\right) |\nabla U(x)|^2 - \frac{\sigma^2\theta}{2} \Delta U(x) \geq \frac{2(1-\frac{\theta\sigma^2}{2})\underline{c}}{1-r} (U^{1-r}(x) - U^{1-r}(x^*)) - \frac{\theta\sigma^2 dL}{2},$$

so that

$$\mathcal{C}_M \subset \left\{ x \in \mathbb{R}^d; U(x) \leq \left(1 + \frac{2M + \theta\sigma^2 dL}{2\underline{c}(2 - \theta\sigma^2)}\right)^{\frac{1}{1-r}} \right\}. \tag{4.2}$$

If we now consider the case where  $(\mathbf{H}_r^2)$  also holds, we use that  $\Delta U \leq \underline{c}U^{-r}(x)$  to obtain:

$$\begin{aligned} \left(1 - \frac{\theta\sigma^2}{2}\right) |\nabla U(x)|^2 - \frac{\sigma^2\theta}{2} \Delta U(x) \\ \geq \frac{(2 - \theta\sigma^2)\underline{c}}{1-r} U^{1-r}(x) \left(1 - U^{r-1}(x) - \frac{(1-r)\sigma^2\theta\underline{c}d}{2\underline{c}(2 - \theta\sigma^2)} U(x)\right). \end{aligned}$$

To ensure that the right-hand member is lower-bounded by  $M$ , it is enough to ensure that

$$\frac{(2 - \theta\sigma^2)\underline{c}}{1-r} U^{1-r}(x) > \frac{M}{2}, \quad U^{r-1}(x) \leq \frac{1}{4} \quad \text{and} \quad \frac{\sigma^2\theta\underline{c}d}{\underline{c}(4 - 2\theta\sigma^2)} U(x) \leq \frac{1}{4}.$$

This concludes the proof. □

*Proof of Proposition 4.1.* Let  $\theta \in (0, 1)$ , (to be chosen latter) and for all  $x \in \mathbb{R}$  define,

$$f_\theta(x) := e^{\theta U(x)}, \tag{4.3}$$

show that  $f_\theta$  is a Lyapunov function for the dynamic  $\mathcal{L}$ :

$$\begin{aligned} \mathcal{L}f_\theta(x) &= -\theta f_\theta(x) \left( \left(1 - \frac{\theta\sigma^2}{2}\right) |\nabla U(x)|^2 - \frac{\theta\sigma^2}{2} \Delta U(x) \right) \\ &\leq -\theta M f_\theta(x) - \theta f_\theta(x) \left( \left(1 - \frac{\theta\sigma^2}{2}\right) |\nabla U(x)|^2 - \frac{\theta\sigma^2}{2} \Delta U(x) \right) 1_{\mathcal{C}_M}(x), \end{aligned}$$

where  $\mathcal{C}_M$  is defined in Lemma 4.2. In the proof of this lemma we showed that  $\mathcal{C}_M$  is included in a level set  $L_{K_M} = \{x \in \mathbb{R}^d \mid U(x) \leq K_M\}$ . Thus,

$$\begin{aligned} -\theta f_\theta(x) \left( \left(1 - \frac{\theta\sigma^2}{2}\right) |\nabla U(x)|^2 - \frac{\theta\sigma^2}{2} \Delta U(x) \right) 1_{\mathcal{C}_M}(x) &\leq \frac{\theta^2\sigma^2}{2} f_\theta(x) \Delta U(x) 1_{L_{K_M}} \\ &\leq \frac{\theta^2\sigma^2 dL}{2} e^{\theta K_M} 1_{L_{K_M}}. \end{aligned}$$

Finally  $f_\theta$  is a Lyapunov function for the dynamics: *i.e.* for all  $x \in \mathbb{R}^d$ ,

$$\mathcal{L}f_\theta(x) \leq -\theta M f_\theta(x) + \frac{\theta^2 \sigma^2 dL}{2} e^{\theta K_M} 1_{L_{K_M}}(x),$$

Hence by a Gronwall argument we get,

$$P_t f_\theta(x) \leq e^{-\theta M t} f_\theta(x) + \frac{\theta \sigma^2 dL}{2M} e^{\theta K_M}.$$

Choosing  $\theta = \frac{1}{\sigma^2}$ ,  $M = \frac{\theta \sigma^2 dL}{2}$  under  $(\mathbf{H}_r^1)$  and  $M = \frac{(4-2\theta \sigma^2)\underline{c}}{1-r}$  under  $(\mathbf{H}_r^1)$  and  $(\mathbf{H}_r^2)$  leads to the result.  $\square$

We now state an analogous result for the Euler scheme.

**Proposition 4.3.** (i) Assume  $(\mathbf{H}_r^1)$ . Then, if  $\gamma \in (0, \frac{1}{4L}]$  and  $\theta \in [0, \frac{1}{8\sigma^2}]$ ,

$$\sup_{n \geq 0} \mathbb{E}_x \left[ e^{\theta U(\bar{X}_{n\gamma})} \right] \leq e^{\theta U(x)} + e^{\theta(1 + \frac{8dL}{\underline{c}})^{\frac{1}{1-r}} + \theta dL} \cosh(\theta \gamma dL).$$

(ii) Assume  $(\mathbf{H}_r^1)$  and  $(\mathbf{H}_r^2)$ . If  $\gamma \in (0, \frac{1-r}{4\underline{c}VL}]$  and  $\theta \in [0, \frac{1}{8\sigma^2} \wedge 1]$ , then for all  $x \in \mathbb{R}^d$  and  $n \in \mathbb{N}$  we have

$$\sup_{n \geq 0} \mathbb{E}_x \left[ e^{\theta U(\bar{X}_{n\gamma})} \right] \leq e^{\theta U(x)} + c e^{c_r \frac{\theta d(1+\sigma^2)\underline{c}}{\underline{c}}},$$

where  $c$  denotes a constant independent of the parameters and  $c_r$  a constant which only depends on  $r$ .

The proof of this proposition is postponed in Section 6.

**Remark 4.4.** The reader will find more explicit (but more technical) bounds in the proof of the second case. It is worth noting that  $\gamma$  does not depend on  $d$  (as in the strongly convex setting). This is of first importance in our multilevel setting where it is much more efficient if the rough layers of the method can be implemented with step sequences with large sizes. The proof is very close to [14] but the bounds are refined. In particular, compared to this paper, we precisely do not require that the step size decreases with  $d$ .

Thanks to the two previous results we are now able to control the moment of the continuous time and the discrete time processes.

**Proposition 4.5.** Assume  $(\mathbf{H}_r^1)$ .

(i) For all  $x \in \mathbb{R}^d$ ,  $p \geq 0$ ,

$$\sup_{t \geq 0} \mathbb{E}_x [U^p(X_t)] \lesssim_p (U(x) + \Psi)^p.$$

where

$$\Psi = \begin{cases} (1 + \sigma^2) \left(1 + \frac{dL}{\underline{c}}\right)^{\frac{1}{1-r}} & \text{under } (\mathbf{H}_r^1) \text{ only} \\ 4^{\frac{1}{1-r}} \sqrt{\frac{(1+\sigma^2)(\underline{c}VL)}{\underline{c}}} d & \text{under } (\mathbf{H}_r^1) \text{ and } (\mathbf{H}_r^2), \end{cases}$$

(ii) Let  $\gamma \in (0, \gamma^*]$  with  $\gamma^*$  defined by (2.6). Then,

$$\sup_{n \geq 0} \mathbb{E}_x [U^p(\bar{X}_{n\gamma})] \lesssim_p (U(x) + \bar{\Psi})^p$$

where  $\bar{\Psi}$  is defined by (2.6)

(iii) In particular,

$$\max \left( \sup_{t \geq 0} \mathbb{E}_x [U^p(X_t)], \sup_{t \geq 0, \gamma \in (0, \gamma^*]} \mathbb{E}_x [U^p(\bar{X}_t)] \right) \lesssim_{r,p} (U(x) + \bar{\Psi})^p. \quad (4.4)$$

and

$$\max \left( \int U^p(x)\pi(dx), \sup_{\gamma \in (0, \gamma^*]} \int U^p(x)\pi^\gamma(dx) \right) \lesssim_{r,p} \bar{\Psi}^p. \tag{4.5}$$

*Proof.* The proof similar to [14, Prop B.4] is postponed in Section 7. □

**Remark 4.6.** In order to avoid the distinction between cases in all the proofs, we choose to adopt only one notation for  $\Psi$  and  $\bar{\Psi}$  but the reader has to keep in mind that the definition of these quantities depends on the fact that  $(\mathbf{H}_r^2)$  is satisfied or not. Let us also recall that the notation  $\lesssim_{r,p}$  means that the constant only depends on  $r$  and  $p$ . These constants are certainly locally bounded: for any compact subset  $K$  of  $[0, 1) \times [1, +\infty)$ , there exists a universal constant  $c$  such that for any  $(r, p) \in K$ , the underlying constant  $c_{r,p}$  related to  $\lesssim_{r,p}$  is bounded by  $c$ . Finally, note that we chose to keep all the dependencies in the other parameters.

**4.2 Longtime strong discretization error under  $(\mathbf{H}_r^1)$ :**

The following proposition studies the  $L^2$ -error induced by the discretization of the SDE under  $(\mathbf{H}_r^1)$ . The notations  $\gamma^*$  and  $\bar{\Psi}$  come from Proposition 4.5.

**Proposition 4.7.** Assume  $(\mathbf{H}_r^1)$  and let  $x \in \mathbb{R}^d$ ,  $\gamma \in (0, \gamma^*]$  and  $\delta \in (0, 1)$ . Then,

$$\sup_{t \geq 0} \mathbb{E}_x [|X_t - \bar{X}_t|^2] \lesssim_r \frac{L\gamma}{c\delta} (U(x) + \bar{\Psi})^{1+2r} \left( \Gamma\left(\frac{1}{\delta}\right) + \frac{(U(x) + \bar{\Psi})^{r + \frac{2\delta r}{1-\delta}}}{c^{1 + \frac{2\delta}{1-\delta}}} \right)$$

where  $\Gamma$  is the gamma function.

**Remark 4.8.** The control of the  $L^2$ -distance between the diffusion and its Euler scheme is a fundamental property for the efficiency of the multilevel method. Actually, it allows us to control the variance of each level. The fact that we are able to obtain such a property in this (semi)-weakly convex setting is new.

We start with two technical lemmas.

**Lemma 4.9.** Assume  $(\mathbf{H}_r^1)$  then for all  $x \in \mathbb{R}^d$  we have

$$|x - x^*|^2 \leq \frac{U^{1+r}(x) - U^{1+r}(x^*)}{c(1+r)}.$$

*Proof.* First, one can check that for all  $x \in \mathbb{R}^d$  and for all eigenvalue of the Hessian we have

$$\lambda_{D^2U^{1+r}(x)} \geq (1+r)U^r(x)\lambda_{D^2U(x)} \geq (1+r)c, \tag{4.6}$$

where in the last inequality we have used assumption  $(\mathbf{H}_r^1)$ . By the Taylor formula,

$$U^{1+r}(x) - U^{1+r}(x^*) = \langle \nabla U^{1+r}(x^*), x - x^* \rangle + \int_0^1 \langle D^2U^{1+r}(\xi_\theta)(x - x^*), x - x^* \rangle d\theta,$$

where  $\xi_\theta = \lambda x + (1 - \lambda)x^*$ . By (4.6) and the fact that  $\nabla U^{1+r}(x^*) = 0$ , we get

$$U^{1+r}(x) - U^{1+r}(x^*) \geq (1+r)c \int_0^1 |x - x^*|^2 d\theta.$$

This concludes the proof. □

The next lemma is a bound on the moments of the increment of the Euler scheme (with the notations  $\gamma^*$  and  $\bar{\Psi}$  introduced in Proposition 4.5).

**Lemma 4.10.** Assume  $(\mathbf{H}_r^1)$  with  $r \in [0, 1)$ . Let  $\gamma \in (0, \gamma^*)$ . Then for all  $t > 0$  and  $k \in \mathbb{N}^*$ ,

$$\mathbb{E}_x \left[ \left| \bar{X}_t - \bar{X}_{\underline{t}} \right|^{2k} \right]^{1/k} \lesssim_{r,k} L^2(t - \underline{t})^2 (U(x) + \bar{\Psi})^{1+r} + (t - \underline{t})d\sigma^2.$$

*Proof.* By the definition of the Euler scheme we have

$$\begin{aligned} \mathbb{E}_x \left[ \left| \bar{X}_t - \bar{X}_{\underline{t}} \right|^{2k} \right]^{1/k} &= \mathbb{E}_x \left[ \left( \left| -(t - \underline{t})\nabla U(\bar{X}_{\underline{t}}) + \sigma(B_t - B_{\underline{t}}) \right|^2 \right)^k \right]^{1/k} \\ &\leq \mathbb{E}_x \left[ \left( 2(t - \underline{t})^2 |\nabla U(\bar{X}_{\underline{t}})|^2 + 2\sigma^2 |B_t - B_{\underline{t}}|^2 \right)^k \right]^{1/k}. \end{aligned}$$

Then, by Minkowski inequality,

$$\begin{aligned} \mathbb{E}_x \left[ \left| \bar{X}_t - \bar{X}_{\underline{t}} \right|^{2k} \right]^{1/k} &\leq 2(t - \underline{t})^2 \mathbb{E}_x \left[ |\nabla U(\bar{X}_{\underline{t}})|^{2k} \right]^{1/k} + 2\sigma^2 \mathbb{E}_x \left[ |B_t - B_{\underline{t}}|^{2k} \right]^{1/k} \\ &\leq 2L^2(t - \underline{t})^2 \mathbb{E}_x \left[ |\bar{X}_{\underline{t}} - x^*|^{2k} \right]^{1/k} + 2(t - \underline{t})\sigma^2 \mathbb{E}_x \left[ |Z|^{2k} \right]^{1/k}, \end{aligned}$$

where in the last line we used the  $L$ -Lipschitz continuous property of  $\nabla U$  and the fact that  $B_t - B_{\underline{t}} \sim \sqrt{t - \underline{t}}Z$  with  $Z \sim \mathcal{N}(0, \text{Id})$ . Finally, by Lemma 4.9 and Proposition 4.5, we obtain

$$\begin{aligned} \mathbb{E}_x \left[ \left| \bar{X}_t - \bar{X}_{\underline{t}} \right|^{2k} \right]^{1/k} &\leq \frac{2L^2(t - \underline{t})^2}{1 + r} \sup_{n \geq 0} \mathbb{E} \left[ (U^{1+r}(\bar{X}_{n\gamma}))^{2k} \right]^{\frac{1}{2k}} + 2d\sigma^2 \left( \frac{(2k)!}{2^k k!} \right)^{1/k} (t - \underline{t}) \\ &\lesssim_{r,k} L^2(t - \underline{t})^2 (U(x) + \bar{\Psi})^{1+r} + (t - \underline{t})d\sigma^2. \quad \square \end{aligned}$$

We are now ready to prove Proposition 4.7.

*Proof of Proposition 4.7.* Let  $x \in \mathbb{R}^d$  and consider the following process in  $\mathbb{R}^d \times \mathbb{R}^d$ ,

$$\begin{cases} X_t = x - \int_0^t \nabla U(X_s) ds + \sigma B_t \\ \bar{X}_t = \bar{X}_{\underline{t}} - (t - \underline{t})\nabla U(\bar{X}_{\underline{t}}) + \sigma (B_t - B_{\underline{t}}), \end{cases}$$

with  $\bar{X}_0 = x$ . Denote by  $F_x(t) := |X_t - \bar{X}_t|^2$ . By the Lebesgue differentiability theorem,

$$\begin{aligned} F'_x(t) &= 2 \langle X_t - \bar{X}_t, \nabla U(\bar{X}_{\underline{t}}) - \nabla U(X_t) \rangle \\ &= 2 \underbrace{\langle X_t - \bar{X}_t, \nabla U(\bar{X}_{\underline{t}}) - \nabla U(X_t) \rangle}_{E_1} + 2 \underbrace{\langle X_t - \bar{X}_t, \nabla U(\bar{X}_{\underline{t}}) - \nabla U(\bar{X}_t) \rangle}_{E_2}. \end{aligned}$$

To control  $E_1$  we use a Taylor expansion and obtain,

$$\begin{aligned} E_1 &= -2 \left\langle X_t - \bar{X}_t, \int_0^1 D^2U(\lambda X_t + (1 - \lambda)\bar{X}_t) (X_t - \bar{X}_t) d\lambda \right\rangle \\ &= -2 \int_0^1 \langle X_t - \bar{X}_t, D^2U(\lambda X_t + (1 - \lambda)\bar{X}_t) (X_t - \bar{X}_t) \rangle d\lambda. \end{aligned}$$

Since  $D^2U$  is a symmetric matrix,

$$E_1 \leq -2\xi_t |X_t - \bar{X}_t|^2,$$

where  $\xi_t = \int_0^1 \lambda_{D^2U(\lambda X_t + (1 - \lambda)\bar{X}_t)} d\lambda$ . For the second term  $E_2$ , using the inequality  $\langle a, b \rangle \leq \frac{\xi_t}{2}|a|^2 + \frac{1}{2\xi_t}|b|^2$  and the fact that  $\nabla U$  is  $L$ -Lipschitz, we get

$$E_2 \leq \xi_t |X_t - \bar{X}_t|^2 + \frac{L^2}{\xi_t} |\bar{X}_t - \bar{X}_{\underline{t}}|^2.$$

Thus,

$$F'_x(t) \leq -\xi_t F_x(t) + \frac{L^2}{\xi_t} |\bar{X}_t - \bar{X}_{\underline{t}}|^2,$$

and a Gronwall argument leads to

$$F_x(t) \leq \int_0^t \frac{L^2}{\xi_s} |\bar{X}_s - \bar{X}_{\underline{s}}|^2 e^{-\int_s^t \xi_u du} ds.$$

Taking the expectation, the Fubini's theorem implies

$$\mathbb{E}_x [ |X_t - \bar{X}_t|^2 ] \leq \int_0^t \underbrace{\mathbb{E}_x \left[ \frac{1}{\xi_s} |\bar{X}_s - \bar{X}_{\underline{s}}|^2 e^{-\int_s^t \xi_u du} \right]}_{A_s} ds.$$

Now let  $\phi$  be a real non negative function,  $\phi : t \mapsto \phi(t)$

$$\begin{aligned} A_s &\leq \mathbb{E}_x \left[ \frac{1}{\xi_s} |\bar{X}_s - \bar{X}_{\underline{s}}|^2 e^{-\int_s^t \xi_u du} \mathbf{1}_{\{\int_s^t \xi_u du > \phi(s)\}} \right] \\ &+ \mathbb{E}_x \left[ \frac{1}{\xi_s} |\bar{X}_s - \bar{X}_{\underline{s}}|^2 e^{-\int_s^t \xi_u du} \mathbf{1}_{\{\int_s^t \xi_u du \leq \phi(s)\}} \right] \\ &\leq \underbrace{\mathbb{E}_x \left[ \frac{1}{\xi_s} |\bar{X}_s - \bar{X}_{\underline{s}}|^2 \right]}_{A_s^{(1)}} e^{-\phi(s)} + \underbrace{\mathbb{E}_x \left[ \frac{1}{\xi_s} |\bar{X}_s - \bar{X}_{\underline{s}}|^2 \mathbf{1}_{\{\int_s^t \xi_u du \leq \phi(s)\}} \right]}_{A_s^{(2)}}, \end{aligned} \tag{4.7}$$

Using  $(\mathbf{H}_r^1)$ , the convexity of  $x \mapsto U(x)$  and  $t \mapsto t^{-r}$  and the Jensen inequality, we have

$$\begin{aligned} \xi_t &= \int_0^1 \lambda_{D^2 U}(\lambda X_t + (1-\lambda)\bar{X}_t) d\lambda \geq \underline{c} \int_0^1 U^{-r}(\lambda X_t + (1-\lambda)\bar{X}_t) d\lambda \\ &\geq \underline{c} \int_0^1 (\lambda U(X_t) + (1-\lambda)U(\bar{X}_t))^{-r} d\lambda \\ &\geq \underline{c} \left( \int_0^1 \lambda U(X_t) + (1-\lambda)U(\bar{X}_t) d\lambda \right)^{-r} \\ &\geq 2^r \underline{c} (U(X_t) + U(\bar{X}_t))^{-r}. \end{aligned}$$

Thus by the inequality  $(a + b)^p \leq a^p + b^p$  for  $a, b \geq 0$  and  $p \in [0, 1]$ ,

$$\xi_t^{-1} \leq \frac{1}{2^r \underline{c}} (U(X_t) + U(\bar{X}_t))^r \leq \frac{1}{2^r \underline{c}} (U^r(X_t) + U^r(\bar{X}_t)). \tag{4.8}$$

By Cauchy-Schwarz inequality, Proposition 4.5(iii) and Lemma 4.10 we get

$$\begin{aligned} A_s^{(1)} &\leq \mathbb{E}[\xi_s^{-2}]^{\frac{1}{2}} \mathbb{E}[|\bar{X}_s - \bar{X}_{\underline{s}}|^4]^{\frac{1}{2}} e^{-\phi(s)} \\ &\lesssim_r e^{-\phi(s)} \underline{c}^{-1} (U(x) + \bar{\Psi})^r \left[ L^2(s - \underline{s})^2 (U(x) + \bar{\Psi})^{1+r} + (s - \underline{s})d\sigma^2 \right] \\ &\lesssim_r e^{-\phi(s)} \frac{L\gamma}{\underline{c}} (U(x) + \bar{\Psi})^{1+2r}, \end{aligned} \tag{4.9}$$

where in the last line, we used that  $d\sigma^2 \leq \bar{\Psi}$  and  $\gamma L \leq 1$ .

For the second term use Cauchy-Schwarz inequality,

$$\begin{aligned} A_s^{(2)} &\leq \mathbb{E}_x \left[ \left( \frac{1}{\xi_s} |\bar{X}_s - \bar{X}_{\underline{s}}|^2 \right)^2 \right]^{1/2} \mathbb{P}_x \left( \int_s^t \xi_u du \leq \phi(s) \right)^{1/2} \\ &\leq \mathbb{E}_x \left[ \frac{1}{\xi_s^4} \right]^{1/4} \mathbb{E}_x [ |\bar{X}_s - \bar{X}_{\underline{s}}|^8 ]^{1/4} \mathbb{P}_x \left( \int_s^t \xi_u du \leq \phi(s) \right)^{1/2}. \end{aligned}$$

With the help of Inequality (4.8) and Proposition 4.5, we have

$$A_s^{(2)} \lesssim_r \frac{(U(x) + \bar{\Psi})^r}{\underline{c}} \left[ L^2(s - \underline{s})^2 (U(x) + \bar{\Psi})^{1+r} + (s - \underline{s})d\sigma^2 \right] \mathbb{P}_x \left( \int_s^t \xi_u du \leq \phi(s) \right)^{1/2}.$$

For the third term of this product, let  $\kappa$  be a positive number and use Markov inequality:

$$\begin{aligned} \mathbb{P}_x \left( \int_s^t \xi_u du \leq \phi(s) \right)^{1/2} &\leq \mathbb{P}_x \left( \left( \int_s^t \xi_u du \right)^{-\kappa} \geq \phi^\kappa(s) \right)^{1/2} \\ &\leq \phi^{\kappa/2}(s) \mathbb{E} \left[ \left( \int_s^t \xi_u du \right)^{-\kappa} \right]^{1/2}. \end{aligned}$$

The function  $x \mapsto x^{-\kappa}$  being convex on  $(0, +\infty)$  it follow from Jensen inequality that

$$\mathbb{P}_x \left( \int_s^t \xi_u du \leq \phi(s) \right)^{1/2} \leq \phi^{\kappa/2}(s) (t - s)^{-\kappa/2} \mathbb{E} \left[ \frac{1}{t - s} \int_s^t \xi_u^{-\kappa} du \right]^{1/2}.$$

Using again inequality (4.8) and Proposition 4.5,

$$\begin{aligned} \mathbb{P}_x \left( \int_s^t \xi_u du \leq \phi(s) \right)^{1/2} &\lesssim_r \underline{c}^{-\frac{\kappa}{2}} \left( \frac{\phi(s)}{t - s} \right)^{\kappa/2} \sup_{u \in [s, t]} \mathbb{E} [U^{\kappa r}(X_t) + U^{\kappa r}(\bar{X}_t)]^{1/2} \\ &\lesssim_r \underline{c}^{-\frac{\kappa}{2}} \left( \frac{\phi(s)}{t - s} \right)^{\kappa/2} (U(x) + \bar{\Psi})^{\frac{\kappa r}{2}}. \end{aligned}$$

Finally, we get

$$A_s^{(2)} \lesssim_r \underline{c}^{-1 - \frac{\kappa}{2}} (U(x) + \bar{\Psi})^{r + \frac{\kappa r}{2}} \left[ L^2(s - \underline{s})^2 (U(x) + \bar{\Psi})^{1+r} + (s - \underline{s})d\sigma^2 \right] \left( \frac{\phi(s)}{t - s} \right)^{\kappa/2}.$$

Now let  $\phi(s) = \frac{t-s}{(t+1-s)^{1-\delta}}$  with  $a > 0$ ,  $\delta \in (0, 1)$  and  $\kappa = \frac{2(1+\delta)}{1-\delta}$  we have

$$\begin{aligned} A_s^{(2)} &\lesssim_r \underline{c}^{-2 - \frac{2\delta}{1-\delta}} \left[ L^2(s - \underline{s})^2 (U(x) + \bar{\Psi})^{1+3r + \frac{2\delta r}{1-\delta}} \right. \\ &\quad \left. + (s - \underline{s})d\sigma^2 (U(x) + \bar{\Psi})^{2r + \frac{2\delta r}{1-\delta}} \right] \left( \frac{1}{t + 1 - s} \right)^{1+\delta}. \end{aligned}$$

As a consequence, since  $\gamma L \leq 1$  and  $d\sigma^2 \leq \bar{\Psi}$ , we obtain

$$A_s^{(2)} \lesssim_r \frac{L\gamma}{\underline{c}^{2 + \frac{2\delta}{1-\delta}}} (U(x) + \bar{\Psi})^{1+3r + \frac{2\delta r}{1-\delta}} \left( \frac{1}{t + 1 - s} \right)^{1+\delta}.$$

Back to (4.7), we deduce from (4.9) and from the above inequality that

$$\begin{aligned} \mathbb{E}_x \left[ |X_t - \bar{X}_t|^2 \right] &\leq \int_0^t A_s^{(1)} ds + \int_0^t A_s^{(2)} ds \\ &\lesssim_r \frac{L\gamma}{\underline{c}} (U(x) + \bar{\Psi})^{1+2r} \left( \int_0^t e^{-\frac{u}{(u+1)^{1-\delta}}} du + \frac{(U(x) + \bar{\Psi})^{r + \frac{2\delta r}{1-\delta}}}{\underline{c}^{1 + \frac{2\delta}{1-\delta}}} \int_0^t \left( \frac{1}{1+u} \right)^{1+\delta} du \right) \\ &\lesssim_r \frac{L\gamma}{\underline{c}} (U(x) + \bar{\Psi})^{1+2r} \left( \int_0^t e^{-u^\delta} du + \frac{(U(x) + \bar{\Psi})^{r + \frac{2\delta r}{1-\delta}}}{\delta \underline{c}^{1 + \frac{2\delta}{1-\delta}}} \right) \\ &\lesssim_r \frac{L\gamma}{\underline{c}\delta} (U(x) + \bar{\Psi})^{1+2r} \left( \Gamma \left( \frac{1}{\delta} \right) + \frac{(U(x) + \bar{\Psi})^{r + \frac{2\delta r}{1-\delta}}}{\underline{c}^{1 + \frac{2\delta}{1-\delta}}} \right) \end{aligned}$$

The result follows. □

**4.3 Convergence to equilibrium for the Euler Scheme under  $(H_r^1)$ :**

We now proceed to establish the weak error between the discrete semi group and its invariant measure denoted by  $\pi^\gamma$ . The proof of this result is based on the control of the so-called tangent process  $\bar{T}_t^x := \nabla X_t^x$ .

**Proposition 4.11.** *Assume  $(H_r^1)$  and let  $x \in \mathbb{R}^d$ ,  $\gamma \in (0, \gamma^*]$ . Let  $\kappa > 0$  and  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$  be a positive function. Assume that  $U(x) \lesssim_r \bar{\Psi}$ . Then, for any  $r \in [0, 1)$ , there exists a constant  $c_{r,\kappa}$  (depending only on  $r$  and  $\kappa$ ) such that for all  $n > 0$ , for all Lipschitz continuous function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ ,*

$$|\mathbb{E}_x [f(\bar{X}_{n\gamma})] - \pi^\gamma(f)| \leq c_{r,\kappa} [f]_1 h_{\phi,\kappa}(n),$$

where

$$h_{\phi,\kappa}(n) = \underline{c}^{-\frac{1}{2}} \bar{\Psi}^{\frac{1+r}{2}} e^{-\phi(n)} + \underline{c}^{-\frac{\kappa+1}{2}} \bar{\Psi}^{\frac{1+r(1+\kappa)}{2}} \left( \frac{\phi(n)}{n\gamma} \right)^{\frac{\kappa}{2}},$$

with  $\gamma^*$  and  $\bar{\Psi}$  defined in Proposition 4.5.

**Remark 4.12.**  $\triangleright$  In order to alleviate the purpose, the result is stated under the assumption that the initial condition  $x$  satisfies  $U(x) \lesssim_r \bar{\Psi}$  but the reader will find some bounds without this assumption in the proofs.

$\triangleright$  The function  $h_{\phi,\kappa}$  plays the role of convergence rate to equilibrium. In this setting where the Hessian is not lower-bounded, we adopt a strategy which consists in separating the space into two parts. In the first one, we assume that we have some good contraction properties parametrized by the function  $\phi$  and in the other one, we try simply to control the probability that such a good contraction does not occur. This leads to a balance between two terms depending on  $\phi$  and  $\kappa$ . In the following, we will choose  $\phi$  and  $\kappa$  in order that  $h_{\phi,\kappa}$  is summable with the smallest impact on the dependence in the dimension.

Note that in [4], some exponential rates are exhibited under similar assumptions in the continuous case (with the help of concentration inequalities). However, this exponential rate depends on some constants whose control seems to be difficult to obtain (typically, when the starting distribution is absolutely continuous with respect to the invariant distribution, the constants involve the  $L^2$ -moment of the related density). Probably, some ideas could be adapted to the Euler scheme (starting from a deterministic point) but with technicalities that seem to carry us too far for this paper.

We preface the proof of Proposition 4.11 by a lemma about the shape of the *first variation process* of the continuous time Euler scheme,  $\bar{T}_t^x = \nabla_x \bar{X}_t^x$ .

**Lemma 4.13.** *For all  $n \in \mathbb{N}$ ,  $x \in \mathbb{R}^d$  and  $\gamma \in [0, 1)$ ,*

$$\bar{T}_{(n+1)\gamma}^x = \prod_{i=0}^n (\text{Id}_{\mathbb{R}^d} - \gamma D^2 U(\bar{X}_{i\gamma}^x)).$$

*Proof of Lemma 4.13.* First, observe that for all  $n \in \mathbb{N}$ ,

$$\frac{\bar{T}_{(n+1)\gamma}^x - \bar{T}_{n\gamma}^x}{\gamma} = \nabla_x \left( \frac{\bar{X}_{(n+1)\gamma}^x - \bar{X}_{n\gamma}^x}{\gamma} \right),$$

and by the definition of the Euler scheme and the chain rule,

$$\begin{aligned} \frac{\bar{T}_{(n+1)\gamma}^x - \bar{T}_{n\gamma}^x}{\gamma} &= \nabla_x (-\nabla_x U(\bar{X}_{n\gamma}^x) + \sigma(B_{(n+1)\gamma} - B_{(n-1)\gamma})) \\ &= -D^2 U(\bar{X}_{n\gamma}^x) \bar{T}_{n\gamma}^x. \end{aligned}$$

Then we get,

$$\bar{T}_{(n+1)\gamma}^x = \bar{T}_{n\gamma}^x (\text{Id}_{\mathbb{R}^d} - \gamma D^2 U(\bar{X}_{n\gamma}^x)),$$

and the proof follows by a simple induction. □

Consider two paths defined with the same Brownian motion and different starting points:  $x, y \in \mathbb{R}^d$ . The following proposition shows that there is a pathwise confluence, i.e. the two trajectories get closer when  $n$  goes to infinity.

**Proposition 4.14.** Assume  $(\mathbf{H}_r^1)$  and let  $x, y \in \mathbb{R}^d$ ,  $\gamma \in (0, \gamma^*]$ ,  $\kappa > 0$ . Let  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$  be a positive function. Then,

$$\sup_{n \geq 0} \mathbb{E}_{(x,y)} \left[ |\bar{X}_{n\gamma}^x - \bar{X}_{n\gamma}^y|^2 \right] \lesssim_{r,\kappa} |x - y|^2 h_{\phi,\kappa,x,y}(n),$$

where,

$$h_{\phi,\kappa,x,y}(n) = e^{-2\phi(n)} + \left( \frac{\phi(n)}{n\gamma} \right)^\kappa \underline{c}^{-\kappa} (U(x) + U(y) + \bar{\Psi})^{\kappa r},$$

with  $\gamma^*$  and  $\bar{\Psi}$  given by Proposition 4.5.

**Remark 4.15.** In the sequel, this property is typically applied with a polynomial function  $\phi$  which leads to polynomial rates to equilibrium. It is worth noting that the proof could be adapted to provide exponential rates (the idea would be to consider an exponentially decreasing convex function instead of  $x \mapsto x^{-\kappa}$  in the proof below). However, with our method, such rates would lead to exponential dependence in the dimension. This is why we do not give such bounds here.

*Proof of Proposition 4.14.* For  $x, y \in \mathbb{R}^d$  and  $n \in \mathbb{N}$ , let us start by a Taylor expansion of the function  $x \mapsto \bar{X}_{n\gamma}^x$ ,

$$\begin{aligned} |\bar{X}_{n\gamma}^x - \bar{X}_{n\gamma}^y|^2 &= \left| \int_0^1 \bar{T}_{n\gamma}^{\lambda x + (1-\lambda)y} (x - y) d\lambda \right|^2 \\ &= |x - y|^2 \left| \int_0^1 \bar{T}_{n\gamma}^{\lambda x + (1-\lambda)y} d\lambda \frac{x - y}{|x - y|} \right|^2 \\ &\leq |x - y|^2 \left\| \int_0^1 \bar{T}_{n\gamma}^{\lambda x + (1-\lambda)y} d\lambda \right\|^2, \end{aligned}$$

where  $\|\cdot\|$  is the operator norm associated with the Euclidean norm. By Jensen inequality and Lemma 4.13,

$$\begin{aligned} |\bar{X}_{n\gamma}^x - \bar{X}_{n\gamma}^y|^2 &\leq |x - y|^2 \int_0^1 \left\| \prod_{i=0}^{n-1} \text{Id}_{\mathbb{R}^d} - \gamma D^2 U(\bar{X}_{i\gamma}^{\lambda x + (1-\lambda)y}) \right\|^2 d\lambda \\ &\leq |x - y|^2 \int_0^1 \prod_{i=0}^{n-1} \left\| \text{Id}_{\mathbb{R}^d} - \gamma D^2 U(\bar{X}_{i\gamma}^{\lambda x + (1-\lambda)y}) \right\|^2 d\lambda. \end{aligned}$$

The operator norm associated with the Euclidean norm of a symmetric matrix is equal to its spectral radius, so we get

$$\begin{aligned} |\bar{X}_{n\gamma}^x - \bar{X}_{n\gamma}^y|^2 &\leq |x - y|^2 \int_0^1 \prod_{i=0}^{n-1} \left( 1 - \gamma \lambda_{D^2 U(\bar{X}_{i\gamma}^{\lambda x - (1-\lambda)y})} \right)^2 d\lambda \\ &\leq |x - y|^2 \int_0^1 e^{-2\gamma \sum_{i=0}^{n-1} \lambda_{D^2 U(\bar{X}_{i\gamma}^{\lambda x - (1-\lambda)y})}} d\lambda, \end{aligned}$$

and

$$\mathbb{E}_{(x,y)} \left[ |\bar{X}_{n\gamma}^x - \bar{X}_{n\gamma}^y|^2 \right] \leq |x - y|^2 \int_0^1 \mathbb{E}_{(x,y)} \left[ e^{-2\xi_n(\lambda)} \right] d\lambda, \tag{4.10}$$

with,

$$\xi_n(\lambda) := \gamma \sum_{i=0}^{n-1} \lambda_{D^2U}(\bar{X}_{i\gamma}^{\lambda x - (1-\lambda)y}) = \int_0^{n\gamma} \lambda_{D^2U}(\bar{X}_{\underline{u}}^{\lambda x - (1-\lambda)y}) du.$$

For a given real non negative function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  we have

$$\begin{aligned} \mathbb{E}_{(x,y)} \left[ e^{-2\xi_n(\lambda)} \right] &\leq \mathbb{E}_{(x,y)} \left[ e^{-2\xi_n} 1_{\{\xi_n > \phi(n)\}} \right] + \mathbb{E}_{(x,y)} \left[ e^{-2\xi_n} 1_{\{\xi_n \leq \phi(n)\}} \right] \\ &\leq e^{-2\phi(n)} + \mathbb{E}_{(x,y)} \left[ 1_{\{\xi_n \leq \phi(n)\}} \right] \\ &\leq e^{-2\phi(n)} + \mathbb{P}_{(x,y)} \left( \int_0^{n\gamma} \lambda_{D^2U}(\bar{X}_{\underline{u}}^{\lambda x - (1-\lambda)y}) du \leq \phi(n) \right). \end{aligned}$$

For a positive number  $\kappa$  we have

$$\mathbb{E}_{(x,y)} \left[ e^{-2\xi_n(\lambda)} \right] \leq e^{-2\phi(n)} + \mathbb{P}_{(x,y)} \left( \left( \int_0^{n\gamma} \lambda_{D^2U}(\bar{X}_{\underline{u}}^{\lambda x - (1-\lambda)y}) du \right)^{-\kappa} \geq \phi^{-\kappa}(n) \right),$$

and using the Markov inequality,

$$\mathbb{E}_{(x,y)} \left[ e^{-2\xi_n(\lambda)} \right] \leq e^{-2\phi(n)} + \phi^\kappa(n) \mathbb{E}_{(x,y)} \left[ \left( \int_0^{n\gamma} \lambda_{D^2U}(\bar{X}_{\underline{u}}^{\lambda x - (1-\lambda)y}) du \right)^{-\kappa} \right].$$

The function  $x \mapsto x^{-\kappa}$  is convex on  $(0, +\infty)$  then by Jensen inequality,

$$\begin{aligned} \mathbb{E}_{(x,y)} \left[ e^{-2\xi_n(\lambda)} \right] &\leq e^{-2\phi(n)} + \phi^\kappa(n) (n\gamma)^{-\kappa} \mathbb{E}_{(x,y)} \left[ \frac{1}{n\gamma} \int_0^{n\gamma} \lambda_{D^2U}^{-\kappa}(\bar{X}_{\underline{u}}^{\lambda x - (1-\lambda)y}) du \right] \\ &\leq e^{-2\phi(n)} + \left( \frac{\phi(n)}{n\gamma} \right)^\kappa \sup_{k \in \{0, \dots, n-1\}} \mathbb{E} \left[ \lambda_{D^2U}^{-\kappa}(\bar{X}_{k\gamma}^{\lambda x - (1-\lambda)y}) \right]. \end{aligned}$$

Observe that assumption  $(\mathbf{H}_r^1)$  implies,

$$\mathbb{E}_{(x,y)} \left[ e^{-2\xi_n(\lambda)} \right] \leq e^{-2\phi(n)} + \left( \frac{\phi(n)}{n\gamma} \right)^\kappa \sup_{k \in \{0, \dots, n-1\}} \mathbb{E} \left[ \underline{c}^{-\kappa} U^{\kappa r} \left( \bar{X}_{k\gamma}^{\lambda x + (1-\lambda)y} \right) \right].$$

By Proposition 4.5 and the convexity of  $U$ , this implies that for any  $\gamma \in (0, \gamma^*]$ ,

$$\mathbb{E}_{(x,y)} \left[ e^{-2\xi_n(\lambda)} \right] \lesssim_{r,\kappa} e^{-2\phi(n)} + \left( \frac{\phi(n)}{n\gamma} \right)^\kappa \underline{c}^{-\kappa} (U(x) + U(y) + \bar{\Psi})^{\kappa r}. \quad \square$$

Thanks to this confluence property we are now able to prove the convergence to equilibrium of the Euler scheme and to give the rate of this convergence.

*Proof of Proposition 4.11.* Since  $\pi^\gamma$  is invariant for  $(\bar{X}_{n\gamma})_{n \in \mathbb{N}}$  we deduce from Fubini's Theorem and Jensen inequality that

$$\begin{aligned} |\mathbb{E}_x [f(\bar{X}_{n\gamma})] - \pi^\gamma(f)|^2 &= \left| \int_{\mathbb{R}^d} \mathbb{E}_{(x,y)} [f(\bar{X}_{n\gamma}^x) - f(\bar{X}_{n\gamma}^y)] \pi^\gamma(dy) \right|^2 \\ &\leq \int_{\mathbb{R}^d} \mathbb{E}_{(x,y)} \left[ |f(\bar{X}_{n\gamma}^x) - f(\bar{X}_{n\gamma}^y)|^2 \right] \pi^\gamma(dy). \end{aligned}$$

The Lipschitz property of  $f$  implies that

$$|\mathbb{E}_x [f(\bar{X}_{n\gamma})] - \pi^\gamma(f)|^2 \leq [f]_1^2 \int_{\mathbb{R}^d} \mathbb{E}_{(x,y)} [|\bar{X}_{n\gamma}^x - \bar{X}_{n\gamma}^y|^2] \pi^\gamma(dy),$$

where  $[f]_1$  is the Lipschitz constant of  $f$ . Proposition 4.14 implies

$$|\mathbb{E}_x [f(\bar{X}_{n\gamma})] - \pi^\gamma(f)|^2 \lesssim_{r,\kappa} [f]_1^2 \int_{\mathbb{R}^d} |x - y|^2 h_{\phi,\kappa,x,y}(n) \pi^\gamma(dy). \tag{4.11}$$

By Lemma 4.9,

$$|x - y|^2 \leq 2(|x - x^*|^2 + |y - y^*|^2) \leq \frac{2}{\underline{c}(1+r)} (U^{1+r}(x) + U^{1+r}(y)).$$

With the help of the Young inequality, we also have

$$|x - y|^2 (U(x) + U(y))^{r\kappa} \lesssim_{r,\kappa} \underline{c}^{-1} (U^{1+r(1+\kappa)}(x) + U^{1+r(1+\kappa)}(y)).$$

Plugging these controls into (4.11) yields

$$\begin{aligned} |\mathbb{E}_x [f(\bar{X}_{n\gamma})] - \pi^\gamma(f)|^2 &\lesssim_{r,\kappa} \frac{[f]_1^2}{\underline{c}} (U^{1+r}(x) + \pi^\gamma(U^{1+r})) e^{-2\phi(n)} \\ &\quad + [f]_1^2 \underline{c}^{-\kappa-1} (U^{1+r(1+\kappa)}(x) + \pi^\gamma(U^{1+r(1+\kappa)})) \left(\frac{\phi(n)}{n\gamma}\right)^\kappa. \\ &\quad + [f]_1^2 \underline{c}^{-\kappa-1} \bar{\Psi}^{r\kappa} (U^{1+r}(x) + \pi^\gamma(U^{1+r})) \left(\frac{\phi(n)}{n\gamma}\right)^\kappa. \end{aligned} \tag{4.12}$$

To conclude, we now use the bound (4.5) of Proposition 4.5(iii) and the assumption  $U(x) \lesssim_r \bar{\Psi}$ . □

#### 4.4 Bias induced by the discretization under $(\mathbf{H}_r^1)$ :

We now need to provide estimates of  $\mathcal{W}_1(\pi, \pi^\gamma)$ . We provide two results: Lemma 4.16 where we directly derive from Proposition 4.7 a bound in  $O(\sqrt{\gamma})$  which “only” requires the potential  $U$  to be  $\mathcal{C}^2$ . However, such a bound has a serious impact on the dependency in  $\varepsilon$  of the complexity. Thus, we propose a second result when  $U$  is  $\mathcal{C}^3$  where we recover a bound in  $O(\gamma)$ .

##### 4.4.1 A first bound in $O(\sqrt{\gamma})$

As mentioned before, a first estimate can be directly deduced from Proposition 4.7. Actually, since in this result, the  $L^2$ -error between the process and its discretization is controlled uniformly in time, this leads to a similar bound for  $\mathcal{W}_1(\pi, \pi^\gamma)$  by letting  $t$  go to  $\infty$ . More precisely,

**Lemma 4.16.** Assume  $(\mathbf{H}_r^1)$ . Let  $\gamma \in (0, \gamma^*]$ . Then,

$$\mathcal{W}_1(\pi, \pi^\gamma)^2 \lesssim_{r,\delta} \frac{L\gamma}{\underline{c}^{1-\delta} \wedge \underline{c}} \bar{\Psi}^{1+3r+\frac{2\delta r}{1-\delta}}. \tag{4.13}$$

*Proof.* Owing to the stationarity of  $\pi$ , we have for every  $n \geq 0$ ,

$$\mathcal{W}_1(\pi, \pi^\gamma) \leq \mathcal{W}_1(\pi P_{n\gamma}, \pi \bar{P}_{n\gamma}) + \mathcal{W}_1(\pi \bar{P}_{n\gamma}, \pi^\gamma),$$

so that

$$\mathcal{W}_1(\pi, \pi^\gamma) \leq \limsup_{n \rightarrow +\infty} \mathcal{W}_1(\pi P_{n\gamma}, \pi \bar{P}_{n\gamma}),$$

since  $\mathcal{W}_1(\pi \bar{P}_{n\gamma}, \pi^\gamma) \xrightarrow{n \rightarrow +\infty} 0$  by Proposition 4.11 (more precisely, this property can be deduced from an integration of (4.12) with respect to  $\pi$  and from the fact that  $\pi(U^p) < +\infty$  for any  $p$  by Proposition 4.5).

Now, integrating with respect to  $\pi$  the bound of Proposition 4.7 and using that  $\pi(U^p) \lesssim_r \bar{\Psi}^p$  by Proposition 4.5 leads to the result.  $\square$

#### 4.4.2 A second bound in $O(\gamma)$

Even if the above bound is quite explicit in terms of its dependency with respect to  $L$ ,  $\underline{c}$ ,  $\bar{c}$  and  $d$ , the fact that it is in  $O(\sqrt{\gamma})$  dramatically impacts the complexity in terms of  $\varepsilon$  (at least).

In fact, it is possible to get a 1-Wasserstein error of the order  $\gamma$  by using a combination of the control of the rate of convergence to equilibrium of the continuous process and of the finite-time weak error (between the process and its discretization). Such a strategy is used in several papers: in [31], this idea is developed in a multiplicative setting with a so-called “domino” approach for the control of the 1-Wasserstein and  $TV$  distances between the process and its discretization, uniformly in time. For the control of  $\mathcal{W}_1(\pi, \pi^\gamma)$  itself, our approach follows [11] which provides a series of bounds in many models and sets of assumptions which are mainly based on the following principle (see Lemma 1 of [11]). Taking advantage of the stationarity of  $\pi^\gamma$ , for any  $p \geq 1$ , for any  $t > 0$ ,

$$\mathcal{W}_p(\pi, \pi^\gamma) \leq \mathcal{W}_p(\pi, \pi^\gamma P_t) + \mathcal{W}_p(\pi^\gamma P_t, \pi^\gamma \bar{P}_t),$$

so that if we assume that

$$\mathcal{W}_p(\pi, \pi^\gamma P_t) \leq \varepsilon_1(t) \mathcal{W}_p(\pi, \pi^\gamma) \quad \text{and} \quad \mathcal{W}_p(\pi^\gamma P_t, \pi^\gamma \bar{P}_t) \leq \varepsilon_2(t),$$

then,

$$\mathcal{W}_p(\pi, \pi^\gamma) \leq \inf \left\{ \frac{\varepsilon_2(t)}{1 - \varepsilon_1(t)}, t > 0 \right\}. \tag{4.14}$$

We thus propose to estimate  $\varepsilon_1(t)$  and  $\varepsilon_2(t)$  under Assumption  $(\mathbf{H}_r^1)$  (with or without  $(\mathbf{H}_r^2)$ ). This is the purpose of Lemmas 4.19 and 4.21 respectively. These two estimates lead to the following proposition

**Proposition 4.17.** *Assume  $(\mathbf{H}_r^1)$  and let  $\delta \in (0, 1)$ . Assume that  $U$  is  $C^3$  and  $\|\Delta(\nabla U)\|_{2,\infty}^2 \lesssim_r \sigma^{-4} L^3 \bar{\Psi}$  (with  $\|\Delta(\nabla U)\|_{2,\infty}$  defined in Lemma 4.21). Then, a constant  $c_{r,\delta}$  (depending only on  $r$  and  $\delta$ ) exists such that for all  $\gamma \in (0, \gamma^*]$ ,*

$$\mathcal{W}_2(\pi, \pi^\gamma) \leq c_{r,\delta} \left( \underline{c}^{-\frac{1}{1-\delta}} L^{\frac{3}{2}} \bar{\Psi}^{\frac{1}{2} + \frac{r}{1-\delta}} \right) \gamma.$$

**Remark 4.18.**  $\triangleright$  Note that this result is clearly in the spirit of [11, Theorem 6]. However, there are several differences. First, we need here to adapt our proof to a setting where we only have polynomial convergence to equilibrium (instead of exponential convergence). Second, under our assumptions on  $U$  which are more restrictive than the one of [11, Theorem 6], we can improve the constants (and in particular avoid some exponential dependence in the Lipschitz constant  $L$  of  $\nabla U$ ).

$\triangleright$  Compared with Lemma 4.16, this result improves the dependence in  $\gamma$  but it is worth noting that the bound is also better with respect to  $\bar{\Psi}$  (and thus to the dimension).

*Proof.* With the notations of [11, Theorem 6], let  $t_{\text{rel}} = \inf\{t > 0, \varepsilon_1(t) \leq 1/\sqrt{2}\}$ . Using Lemma 4.19, one can upper-bound  $t_{\text{rel}}$  by asking the two right-hand terms of (4.15) to be bounded by  $1/4$ . With  $\phi(t) = t^\delta$  with  $\delta \in (0, 1)$ , this leads to:

$$t_{\text{rel}} \lesssim_r \max\{(\log 4)^\delta, (4\underline{c}^{-1})^{\frac{1}{1-\delta}} \bar{\Psi}^{\frac{r}{1-\delta}}\}.$$

By (4.14) and Lemma 4.21(ii), we get for any  $\lambda \in (0, 1]$ ,

$$\mathcal{W}_2(\pi, \pi^\gamma) \leq c_r \frac{\gamma}{\lambda} \sqrt{e^{\lambda t_{\text{rel}}} L^3 \bar{\Psi}}.$$

Let  $\lambda = t_{\text{rel}}^{-1}$ . In this case, we obtain

$$\mathcal{W}_2(\pi, \pi^\gamma) \leq c_r \gamma t_{\text{rel}} \sqrt{L^3 \bar{\Psi}} \lesssim_{r,\delta} \gamma \underline{c}^{-\frac{1}{1-\delta}} L^{\frac{3}{2}} \bar{\Psi}^{\frac{1}{2} + \frac{r}{1-\delta}}.$$

□

**Lemma 4.19.** Assume  $(\mathbf{H}_r^1)$ . Then, for any  $\gamma \in (0, \gamma^*]$  and for any positive function  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ ,

$$\mathcal{W}_2^2(\pi, \pi^\gamma P_t) \lesssim_r \mathcal{W}_2^2(\pi, \pi^\gamma) \left( e^{-2\phi(t)} + \frac{\phi(t)}{ct} \bar{\Psi}^r \right). \tag{4.15}$$

**Remark 4.20.** ▷ The proof of (i) is mostly a continuous-time version of the one of Proposition 4.14.

▷ As mentioned in Remark 4.15, the proof can be adapted to provide exponential rates but unfortunately our method would lead to exponential dependence in the dimension. For this section, the lack of exponential rate does not have a serious impact on the bounds.

Nevertheless, if we needed to improve our bounds, an idea would be to apply [4, Theorem 5.6]. In this result, the authors provide exponential rates under assumptions which are similar to ours. However the related constant depends on the density of the semi-group and it would be necessary to be able to control it with respect to the parameters of the model.

*Proof.* Denoting by  $T_t^x = \partial_x X_t^x$ , the first variation process related to  $(X_t^x)_{t \geq 0}$ , we have:

$$X_t^y - X_t^x = \int_0^1 T_t^{x+\lambda(y-x)}(y-x) d\lambda.$$

Thus,

$$|X_t^y - X_t^x|^2 \leq \int_0^1 \|T_t^{x+\lambda(y-x)}\|^2 |y-x|^2 d\lambda,$$

where  $\|\cdot\|$  stands for the operator norm associated with the Euclidean norm. Since  $T^x$  is the solution to  $dT_t^x = -D^2U(X_t^x)T_t^x dt$  with  $T_0^x = I_d$ , one easily checks that

$$\|T_t^{x+\lambda(y-x)}\|^2 \leq e^{-2 \int_0^t \lambda_U(X_s^{x+\lambda(y-x)}) ds}.$$

Thus,

$$|X_t^y - X_t^x|^2 \leq |x-y|^2 \int_0^1 e^{-2 \int_0^t \lambda_U(X_s^{x+\lambda(y-x)}) ds} d\lambda.$$

Following the arguments of Proposition 4.14, we get for any positive function  $\phi$ ,

$$\begin{aligned} \mathbb{E} \left[ e^{-2 \int_0^t \lambda_U(X_s^{x+\lambda(y-x)}) ds} \right] &\leq e^{-2\phi(t)} + \frac{\phi(t)}{t} \mathbb{E} \left[ \frac{1}{t} \int_0^t \lambda_{D^2U}^{-1}(X_u^{x+\lambda(y-x)}) du \right] \\ &\leq e^{-2\phi(t)} + \frac{\phi(t)}{t} \sup_{s \in [0,t]} \mathbb{E} \left[ \lambda_{D^2U}^{-1}(X_s^{x+\lambda(y-x)}) \right]. \end{aligned}$$

By  $(\mathbf{H}_r^1)$ , one deduces that

$$\begin{aligned} \mathbb{E}[|X_t^y - X_t^x|^2] &\leq |x - y|^2 \left( e^{-2\phi(t)} + \frac{\phi(t)}{\underline{c}t} \sup_{t \geq 0} \mathbb{E} \left[ U^r \left( X_t^{(1-\lambda)x + \lambda y} \right) \right] \right) \\ &\lesssim_{uc} |x - y|^2 \left( e^{-2\phi(t)} + \frac{\phi(t)}{\underline{c}t} (U^r(x) + U^r(y) + \Psi^r) \right), \end{aligned}$$

where in the second line, we used Proposition 4.5 and the convexity of  $U$ . Let now  $\nu$  be a coupling of  $\pi$  and  $\pi^\gamma$ . We have

$$\mathcal{W}_2^2(\pi, \pi^\gamma P_t) \leq \int |x - y|^2 (\nu(dx, dy) \left( e^{-2\phi(t)} + \frac{\phi(t)}{\underline{c}t} (\pi(U^r) + \pi^\gamma(U^r) + \Psi^r) \right)).$$

Taking the infimum over the set of couplings  $\nu$  of  $\pi$  and  $\pi^\gamma$  and using again Proposition 4.5, this yields

$$\mathcal{W}_2^2(\pi, \pi^\gamma P_t) \lesssim_r \mathcal{W}_2^2(\pi, \pi^\gamma) \left( e^{-2\phi(t)} + \frac{\phi(t)}{\underline{c}t} \bar{\Psi}^r \right). \quad \square$$

**Lemma 4.21.** (i) Let  $U$  be a  $\mathcal{C}^3$ -convex function. Then,

$$\mathbb{E}[|X_\gamma^x - \bar{X}_\gamma^y|^2] \leq |x - y|^2 e^{\lambda\gamma} + \mathbf{c}_{\gamma, \lambda}(x, y) \gamma^3.$$

with

$$\mathbf{c}_{\gamma, \lambda}(x, y) = \frac{e^{\lambda\gamma}}{6\lambda} (L^2 |\nabla U(y)|^2 + \sigma^4 \|\Delta(\nabla U)\|_{2, \infty}^2 + 2\lambda\sigma L \|D^2 U\|_{2, \infty} (\sqrt{\gamma}))$$

where

$$\|D^2 U\|_{2, \infty} = \sup_{x \in \mathbb{R}^d} \|D^2 U\|_F, \quad \|\Delta(\nabla U)\|_{2, \infty} = \sup_{x \in \mathbb{R}^d} \sum_{i=1}^d |\Delta \partial_i U|^2,$$

and  $S(x, \gamma) = \sup_{u \in [0, \gamma]} \mathbb{E}[|b(X_u^x)|^2]^{\frac{1}{2}}$ .

(ii) Let  $(\mathbf{H}_r^1)$  hold. Assume that  $U$  is  $\mathcal{C}^3$  with  $\|\Delta(\nabla U)\|_{2, \infty}^2 \lesssim_r \sigma^{-4} L^3 \bar{\Psi}$ . Then, a constant  $c_r$  exists such that for all  $\gamma \in (0, \gamma^*]$ , for all  $\lambda \in (0, 1]$ ,

$$\mathcal{W}_2^2(\pi^\gamma P_{n\gamma}, \pi^\gamma) \leq c_r \frac{\gamma^2}{\lambda^2} e^{\lambda(n+1)\gamma} L^3 \bar{\Psi}.$$

**Remark 4.22.** The assumption on  $\Delta(\nabla U)$  is calibrated to control its contribution by  $L^3 \bar{\Psi}$ . That simplifies the purpose and we could keep its specific contribution at the price of technicalities. However, this assumption is not really restrictive: denoting by  $A(x)$  the  $d \times d$ -matrix defined by  $A_{i,j}(x) = D_{i,j}^3 U(x)$ . One easily checks that

$$\|\Delta(\nabla U)\|_{2, \infty}^2 = \sup_{x \in \mathbb{R}^d} \|A(x)\|_F^2 \leq d \sup_{x \in \mathbb{R}^d} \bar{\lambda}_{A(x)},$$

the second inequality coming from a classical inequality related to the Frobenius norm. Since  $L \geq \sup_{x \in \mathbb{R}^d} \bar{\lambda}_{D^2 U(x)}$ , the assumption is for instance true if

$$\sup_{x \in \mathbb{R}^d} \bar{\lambda}_{A(x)} \leq \sigma^{-2} \sup_{x \in \mathbb{R}^d} (\bar{\lambda}_{D^2 U(x)})^4 \frac{\bar{\Psi}}{\sigma^2 d L}.$$

To conclude, note that  $\frac{\bar{\Psi}}{\sigma^2 d L}$  is well controlled: for instance, under  $(\mathbf{H}_r^1)$  and  $(\mathbf{H}_r^2)$ ,

$$\frac{\bar{\Psi}}{\sigma^2 d L} \lesssim_r \underline{c}^{-1} \left( \frac{\bar{c}}{L} \vee 1 \right).$$

**Remark 4.23.** The calibration of the parameter  $\lambda$  is of first importance in the proof of Proposition 4.17 in order to avoid exponential dependence in the dimension.

*Proof.* The proof is an adaptation of Lemma 5.2 and Proposition 5.3 of [12] but with the viewpoint that  $U$  is only convex. More precisely, we start with a one-step control of the error between the diffusion and its Euler scheme by setting

$$F_{x,y}(t) = \frac{1}{2} \mathbb{E}[|X_t^x - \bar{X}_t^y|^2], \quad t \in [0, \gamma].$$

Then, setting  $b = -\nabla U$ , we have

$$\begin{aligned} F'_{x,y}(t) &= \mathbb{E}[\langle X_t^x - \bar{X}_t^y, b(X_t^x) - b(y) \rangle] \\ &= \mathbb{E}[\langle X_t^x - \bar{X}_t^y, b(X_t^x) - b(\bar{X}_t^y) \rangle] + \mathbb{E}[\langle X_t^x - \bar{X}_t^y, b(\bar{X}_t^y) - b(y) \rangle] \\ &\leq \mathbb{E}[\langle X_t^x - \bar{X}_t^y, b(\bar{X}_t^y) - b(y) \rangle], \end{aligned}$$

where in the last line we used the convexity of  $U$  which involves that

$$\langle \nabla U(x) - \nabla U(y), x - y \rangle \geq 0.$$

We then write

$$\mathbb{E}[\langle X_t^x - \bar{X}_t^y, b(\bar{X}_t^y) - b(y) \rangle] = \mathbb{E}[\langle X_t^x - \bar{X}_t^y, b(\bar{X}_t^y) - b(y + \sigma B_t) \rangle] \tag{4.16}$$

$$+ \mathbb{E}[\langle X_t^x - \bar{X}_t^y, b(y + \sigma B_t) - b(y) \rangle]. \tag{4.17}$$

Let  $\lambda > 0$ . For the right-hand side of (4.16), we use the elementary inequality,  $|uv| \leq \frac{\lambda}{2}|u|^2 + \frac{1}{2\lambda}|v|^2$  to obtain

$$\mathbb{E}[\langle X_t^x - \bar{X}_t^y, b(\bar{X}_t^y) - b(y + \sigma B_t) \rangle] \leq \frac{\lambda}{2} F_{x,y}(t) + \frac{t^2}{2\lambda} L^2 |b(y)|^2. \tag{4.18}$$

For (4.17), the Itô formula applied to  $b_i = -\partial_i U$  leads to

$$b_i(y + \sigma B_t) - b_i(y) = \sigma^2 \int_0^t \Delta b_i(y + \sigma B_s) ds + \sigma \int_0^t \langle \nabla b_i(y + \sigma B_s), dB_s \rangle.$$

On the one hand, setting  $\Delta b = (\Delta b_i)_{i=1}^d$ ,

$$\mathbb{E} \left[ \langle X_t^x - \bar{X}_t^y, \sigma^2 \int_0^t \Delta b(y + \sigma B_s) ds \rangle \right] \leq \frac{\lambda}{2} F_{x,y}(t) + \frac{\sigma^4}{2\lambda} t^2 \|\Delta b\|_{2,\infty}^2,$$

where

$$\|\Delta b\|_{2,\infty}^2 = \sup_{x \in \mathbb{R}^d} \sum_{i=1}^d |\Delta b_i(x)|^2.$$

On the other hand, using the fact that  $\mathcal{M}$  defined by  $\mathcal{M}_t = \int_0^t \langle \nabla b(y + \sigma B_s), dB_s \rangle$  is a martingale (we refer to [12] for the details), we get

$$|\mathbb{E}[\langle X_t^x - \bar{X}_t^y, \sigma \int_0^t \langle \nabla b(y + \sigma B_s), dB_s \rangle \rangle]| \leq \sigma L \|\nabla b\|_{2,\infty} \left( s^{\frac{3}{2}} S(x, \gamma) + \sigma s \sqrt{d} \right),$$

where  $S(x, \gamma) = \sup_{u \in [0, \gamma]} \mathbb{E}[|b(X_u^x)|^2]^{\frac{1}{2}}$  and

$$\|\nabla b\|_{2,\infty} = \sup_{x \in \mathbb{R}^d} \|\nabla b(x)\|_F.$$

Finally, from what precedes, we deduce that

$$F'_{x,y}(t) \leq \lambda F_{x,y}(t) + \frac{t^2}{2\lambda} \left( L^2 |b(y)|^2 + \sigma^4 \|\Delta b\|_{2,\infty}^2 + 2\lambda L \|\nabla b\|_{2,\infty} \left( \sqrt{t} S(x, \gamma) + \sigma \sqrt{d} \right) \right).$$

A standard Gronwall argument then leads to

$$\begin{aligned} \mathbb{E}[|X_\gamma^x - \bar{X}_\gamma^y|^2] &\leq |x - y|^2 e^{\lambda\gamma} \\ &+ \int_0^\gamma s^2 e^{\lambda(\gamma-s)} ds \left( L^2 |b(y)|^2 + \sigma^4 \|\Delta b\|_{2,\infty}^2 + 2\lambda\sigma L \|\nabla b\|_{2,\infty} \left( \sqrt{t} S(x, \gamma) + \sigma \sqrt{d} \right) \right) \\ &\leq |x - y|^2 e^{\lambda\gamma} + \gamma^3 \mathfrak{c}_{\gamma,\lambda}(x, y), \end{aligned}$$

with

$$\mathfrak{c}_{\gamma,\lambda}(x, y) = \frac{e^{\lambda\gamma}}{6\lambda} \left( L^2 |b(y)|^2 + \sigma^4 \|\Delta b\|_{2,\infty}^2 + 2\lambda\sigma L \|\nabla b\|_{2,\infty} \left( \sqrt{\gamma} S(x, \gamma) + \sigma \sqrt{d} \right) \right).$$

(ii) Iterating the above inequality, we obtain for each  $n \geq 1$ ,

$$\mathbb{E}[|X_{n\gamma}^x - \bar{X}_{n\gamma}^x|^2] \leq \gamma^3 \sum_{k=0}^{n-1} \mathbb{E}[\mathfrak{c}_{\gamma,\lambda}(X_{k\gamma}^x, \bar{X}_{k\gamma}^x)] e^{\lambda(n-k)\gamma}$$

Integrating the initial condition with respect to  $\pi^\gamma$ , we get

$$\begin{aligned} \mathcal{W}_2(\pi^\gamma P_{n\gamma}, \pi^\gamma) &\leq \gamma^2 e^{\lambda n\gamma} \sup_{k \geq 0} \int \mathbb{E}[\mathfrak{c}_{\gamma,\lambda}(X_{k\gamma}^x, \bar{X}_{k\gamma}^x)] \pi^\gamma(dx) \\ &\leq \frac{\gamma^2}{\lambda^2} e^{\lambda(n+1)\gamma} \left( L^2 \pi^\gamma(|b|^2) + \sigma^4 \|\Delta b\|_{2,\infty}^2 \right. \\ &\quad \left. + 2\lambda\sigma L \|\nabla b\|_{2,\infty} \left( \sqrt{\gamma} \sup_{n \geq 0} \int \mathbb{E}[S(X_{n\gamma}^x, \gamma)] \pi^\gamma(dx) + \sigma \sqrt{d} \right) \right), \end{aligned}$$

where in the second line, we used the stationarity property of  $\pi^\gamma$ . Now, under  $(\mathbf{H}_r^1)$ ,  $|b|^2 = |\nabla U|^2 \leq 2LU$  (with the same idea than one which leads to (4.1)) so that by Proposition 4.5(iii),  $\pi^\gamma(|b|^2) \lesssim_r L\bar{\Psi}$ . On the other hand, by the Itô formula and the fact that  $\Delta U \leq dL$ ,

$$\mathbb{E}[U(X_t^x)] \leq U(x) + \sigma^2 \int_0^t \mathbb{E}[\Delta U(X_s^x)] ds \leq U(x) + \sigma^2 dL,$$

so that

$$S(x, \gamma) \leq \sqrt{2L} \sup_{u \in [0, \gamma]} \mathbb{E}[U(X_u^x)]^{\frac{1}{2}} \leq \sqrt{2LU(x)} + \sigma \sqrt{2dL}.$$

Again, with the help of Proposition 4.5(iii),

$$\sup_{n \geq 0} \int \mathbb{E}[S(X_{n\gamma}^x, \gamma)] \pi^\gamma(dx) \lesssim_r \sqrt{L\bar{\Psi}}^{\frac{1}{2}} + \sigma \sqrt{dL}.$$

Thus, using that  $\gamma \leq L^{-1}$ ,

$$\mathcal{W}_2(\pi^\gamma P_{n\gamma}, \pi^\gamma) \lesssim_r \frac{\gamma^2}{\lambda^2} e^{\lambda(n+1)\gamma} \left( L^3 \bar{\Psi} + \sigma^4 \|\Delta b\|_{2,\infty}^2 + \sigma\lambda \|\nabla b\|_{2,\infty} (\sqrt{\bar{\Psi}} + \sigma \sqrt{dL}) \right).$$

Since for a symmetric  $d \times d$ -matrix  $A$ ,  $\|A\|_F \leq \sqrt{d\lambda_A}$ , one deduces that  $\|\nabla b\|_{2,\infty} = \|D^2 U\|_{2,\infty} \leq \sqrt{dL}$ . It easily follows that  $\sigma\lambda L \|\nabla b\|_{2,\infty} (\sqrt{\bar{\Psi}} + \sigma \sqrt{dL}) \leq L^3 \bar{\Psi}$  (using that  $L \geq 1$  and  $\bar{\Psi} \geq d$ ). The result follows.  $\square$

### 5 Proof of Theorem 2.3

Following the bias-variance decomposition of the MSE:

$$\|\mathcal{Y}(f) - \pi(f)\|_2^2 \leq [\mathbb{E}[\mathcal{Y}(f)] - \pi(f)]^2 + \text{Var}(\mathcal{Y}(f)),$$

we successively study the bias and the variance contributions and end the section by the proof of Theorem 2.3.

#### 5.1 Step 1: Bias of the procedure

In the sequel,  $\mathcal{Y}(J, (\gamma_j)_j, (T_j)_j, f)$  is usually written  $\mathcal{Y}(f)$  for the sake of simplicity. We start with a telescopic-type decomposition:

$$\begin{aligned} \mathcal{Y}(f) - \pi(f) &= \frac{1}{T_0} \int_0^{T_0} f(\bar{X}_{\underline{s}_{\gamma_0}}^{\gamma_0, x_0}) - \pi^{\gamma_0}(f) ds \\ &+ \sum_{j=1}^J \left( \frac{1}{T_j} \int_0^{T_j} f(\bar{X}_{\underline{s}_{\gamma_j-1}}^{\gamma_j, x_0}) - \pi^{\gamma_j}(f) ds - \frac{1}{T_j} \int_0^{T_j} f(\bar{X}_{\underline{s}_{\gamma_j-1}}^{\gamma_{j-1}, x_0}) - \pi^{\gamma_{j-1}}(f) ds \right) \\ &+ \pi^{\gamma_J}(f) - \pi(f). \end{aligned} \tag{5.1}$$

Let us now study the bias generated by the first and second terms of the right-hand side of (5.1).

**Lemma 5.1.** Assume  $(\mathbf{H}_r^1)$  and  $\gamma_0 \in (0, \gamma^*)$ . Let  $x \in \mathbb{R}^d$  such that  $U(x) \lesssim_r \bar{\Psi}$ . Then, for any  $r \in [0, 1)$  and  $\delta \in (0, \frac{1}{2}]$ , there exists a constant  $c_{r,\delta}$  (depending only on  $r$  and  $\delta$ ) such that for all  $T \geq 1$ , for all Lipschitz continuous function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ ,

$$\left| \frac{1}{T} \int_0^T \mathbb{E}_x[f(\bar{X}_{\underline{s}}^{\gamma, x_0})] - \pi^\gamma(f) ds \right|^2 \leq c_{r,\delta} \frac{[f]_1^2 \mathfrak{C}_{\text{bias}}^{(1)}}{T^2},$$

with  $\mathfrak{C}_{\text{bias}}^{(1)} = \left( \underline{c}^{-\frac{1}{2}} \vee \underline{c}^{-3 - \frac{4\delta}{1-\delta}} \right) \bar{\Psi}^{1+3r + \frac{4\delta r}{1-\delta}}$ .

*Proof.* Let us apply Proposition 4.11 with  $\phi(n) = \frac{n\gamma}{(n\gamma+1)^{1-\delta}}$  and  $\kappa = \frac{2(1+\delta)}{1-\delta} = 2 + \frac{4\delta}{1-\delta}$  for  $\delta \in (0, 1/2]$ . Using that  $\phi(n) \geq (n\gamma)^\delta$ , we have

$$\begin{aligned} \left| \frac{1}{T} \int_0^T \mathbb{E}_x[f(\bar{X}_{\underline{s}}^{\gamma, x_0})] - \pi^\gamma(f) ds \right| &\lesssim_{r,\kappa} \frac{[f]_1}{T} \int_0^T h_{\phi,\kappa}(\underline{s}) ds \\ &\lesssim_r \frac{[f]_1}{T} \left( \underline{c}^{-\frac{1}{2}} \bar{\Psi}^{\frac{1+r}{2}} \int_0^T e^{-\underline{t}^\delta} dt + \underline{c}^{-\frac{3}{2} - \frac{2\delta}{1-\delta}} \bar{\Psi}^{\frac{1+3r}{2} + \frac{2\delta r}{1-\delta}} \int_0^T (1+\underline{t})^{-1-\delta} dt \right) \\ &\lesssim_r \frac{[f]_1}{\delta T} \left( \Gamma\left(\frac{1}{\delta}\right) \underline{c}^{-\frac{1}{2}} \bar{\Psi}^{\frac{1+r}{2}} + \underline{c}^{-\frac{3}{2} - \frac{2\delta}{1-\delta}} \bar{\Psi}^{\frac{1+3r}{2} + \frac{2\delta r}{1-\delta}} \right), \end{aligned}$$

where in the last line, we used standard arguments of comparisons between series and integrals. The result follows.  $\square$

We are now ready to state a proposition about the control of the bias of the procedure.

**Proposition 5.2.** Assume  $(\mathbf{H}_r^1)$  and  $\gamma_0 \in (0, \gamma^*)$ . Let  $x \in \mathbb{R}^d$  such that  $U(x) \lesssim_r \bar{\Psi}$ . Let  $\delta \in (0, \frac{1}{2}]$  and let  $f$  be a Lipschitz continuous function with  $[f]_1 = 1$ . Let  $J \in \mathbb{N}^*$ . Then, (i)

$$|\mathbb{E}_x[\mathcal{Y}(f)] - \pi(f)|^2 \lesssim_{r,\delta} \mathfrak{C}_{\text{bias}}^{(1)} \sum_{j=0}^J \frac{1}{T_j^2} + \mathfrak{C}_{\text{bias}}^{(2,1)} \gamma_J \quad \text{with} \quad \mathfrak{C}_{\text{bias}}^{(2,1)} = \frac{L}{\underline{c}^{\frac{2}{1-\delta}} \wedge \underline{c}} \bar{\Psi}^{1+3r + \frac{2\delta r}{1-\delta}}.$$

(ii) If the assumptions of Proposition 4.17 are fulfilled,

$$|\mathbb{E}_x[\mathcal{Y}(f)] - \pi(f)|^2 \lesssim_{r,\delta} \mathfrak{C}_{\text{bias}}^{(1)} \sum_{j=0}^J \frac{1}{T_j^2} + \mathfrak{C}_{\text{bias}}^{(2,2)} \gamma_J^2 \quad \text{with} \quad \mathfrak{C}_{\text{bias}}^{(2,2)} = \underline{c}^{-\frac{2}{1-\delta}} L^3 \bar{\Psi}^{1+\frac{2r}{1-\delta}}.$$

*Proof.* (i) Taking the expectation in (5.1), we obtain:

$$\begin{aligned} |\mathbb{E}_x[\mathcal{Y}(f) - \pi(f)]| &\leq \left| \frac{1}{T_0} \int_0^{T_0} \mathbb{E}_{x_0} \left[ f(\bar{X}_{\underline{s}_{\gamma_0}}^{\gamma_0, x_0}) \right] - \pi^{\gamma_0}(f) ds \right| \\ &\quad + \sum_{j=1}^J \frac{1}{T_j} \left| \int_0^{T_j} \mathbb{E} \left[ f(\bar{X}_{\underline{s}_{\gamma_{j-1}}}^{\gamma_j, x_0}) \right] - \pi^{\gamma_j}(f) ds \right| \\ &\quad + \sum_{j=1}^J \frac{1}{T_j} \left| \int_0^{T_j} \mathbb{E} \left[ f(\bar{X}_{\underline{s}_{\gamma_{j-1}}}^{\gamma_{j-1}, x_0}) \right] - \pi^{\gamma_{j-1}}(f) ds \right| + |\pi^{\gamma_J}(f) - \pi(f)|. \end{aligned}$$

For the three first terms, we apply Lemma 5.1 and for the last one, Lemma 4.16. The result follows.

(ii) It is the same proof using Proposition 4.17 to control the last term (instead of Lemma 4.16).  $\square$

### 5.2 Step 2: Control of the variance

Now we have to control the variance of our estimator. Owing to the independency between the layers,

$$\text{Var}(\mathcal{Y}(J, (\gamma_j)_j, (T_j)_j, f)) = \text{Var} \left( \frac{1}{T_0} \int_0^{T_0} f(\bar{X}_{\underline{s}_{\gamma_0}}^{\gamma_0, x}) ds \right) + \sum_{j=1}^J \text{Var} \left( \frac{1}{T_j} \int_0^{T_j} G_s^{\gamma_j} ds \right), \quad (5.2)$$

where for some given  $\gamma > 0$  and  $s > 0$ ,

$$G_s^\gamma = f \left( \bar{X}_{\underline{s}_\gamma}^{\frac{\gamma}{2}, x} \right) - f \left( \bar{X}_{\underline{s}_\gamma}^{\gamma, x} \right).$$

Before going further, let us recall that in order that the multilevel method be efficient, the correcting layers must have a small variance. In the long-time setting, this involves to be able to control the  $L^2$ -distance between couplings of Euler schemes with steps  $\gamma$  and  $\gamma/2$ . By Proposition 4.7, this is still possible under  $(\mathbf{H}_r^1)$ , and such a property allows to obtain the following result:

**Lemma 5.3.** Assume  $(\mathbf{H}_r^1)$  and  $\gamma_0 \in (0, \gamma^*]$ . Let  $x \in \mathbb{R}^d$  such that  $U(x) \lesssim_r \bar{\Psi}$ . Let  $\delta \in (0, \frac{1}{2}]$  and  $\kappa > \frac{2}{1-\delta}$ . Let  $f$  be a Lipschitz continuous function. Then, for all  $T > 0$ ,

$$\text{Var} \left( \frac{1}{T} \int_0^T G_s^\gamma ds \right) \lesssim_{r,\delta,\kappa} [f]_1^2 \frac{\mathfrak{C}_{\text{var}}^{(1)}(\kappa, \delta) \gamma^{1-\frac{1}{\kappa(1-\delta)}}}{T},$$

where

$$\begin{aligned} \mathfrak{C}_{\text{var}}^{(1)}(\kappa, \delta) &= L^{1-\frac{1}{\kappa(1-\delta)}} \left( \underline{c}^{-\frac{3-\delta}{\kappa(1-\delta)^2}} \vee \underline{c}^{-\frac{3+\kappa-(\kappa+1)\delta}{\kappa(1-\delta)^2}} \right) \bar{\Psi}^{1+(4-\frac{2}{\kappa})r+\varepsilon(\delta,\kappa)r} \quad \text{with} \\ \varepsilon(\delta, \kappa) &= \left( 3 - \frac{2}{\kappa} - \frac{2}{\kappa(1-\delta)} \right) \frac{\delta}{1-\delta}. \end{aligned}$$

**Remark 5.4.** In the uniformly convex case, the variance is controlled by  $\frac{\gamma \log(1/\gamma)}{T}$  whereas, here, we are only able to obtain  $\frac{\gamma^{1-\frac{1}{\kappa(1-\delta)}}}{T}$ . This difference is due to the

lack of exponential convergence to equilibrium under our assumptions. Note that if we leave  $\kappa$  go to  $\infty$ , we are moving ever closer to the uniformly convex bound. However, the constant depends on  $\kappa$  and explodes when  $\kappa \rightarrow +\infty$ . The interesting point is that the exponent of  $\bar{\Psi}$  remains bounded when  $\kappa \rightarrow +\infty$ , which means that the dependence in the dimension is slightly impacted by the choice of  $\kappa$ .

*Proof.* A standard computation shows that

$$\text{Var} \left( \frac{1}{T} \int_0^T G_s^\gamma ds \right) \leq \frac{2}{T^2} \int_0^T \int_u^T \text{Cov} (G_s^\gamma, G_u^\gamma) dsdu.$$

First, at the price of replacing  $f$  by  $f/[f]_1$ , we can assume in the sequel that  $[f]_1 \leq 1$ . Then,

$$|G_s^\gamma| \leq |\bar{X}_{\underline{s}_\gamma}^{\frac{\gamma}{2},x} - X_{\underline{s}_\gamma}| + |\bar{X}_{\underline{s}_\gamma}^{\gamma,x} - X_{\underline{s}_\gamma}|.$$

By Proposition 4.7 and the fact that  $U(x) \lesssim_{uc} \bar{\Psi}$ , we deduce that for every  $\delta \in (0, 1]$ ,

$$\mathbb{E}[|G_u^\gamma|^2] \lesssim_{r,\delta} \frac{L\gamma}{c^{\frac{2}{1-\delta}}} \bar{\Psi}^{1+3r+\frac{2\delta r}{1-\delta}}. \tag{5.3}$$

This yields a first bound for  $\text{Cov} (G_s^\gamma, G_u^\gamma)$ :

$$\text{Cov} (G_s^\gamma, G_u^\gamma) \leq \mathbb{E}[|G_s^\gamma|^2]^{\frac{1}{2}} \mathbb{E}[|G_u^\gamma|^2]^{\frac{1}{2}} \lesssim_{r,\delta} \frac{L\gamma}{c^{\frac{2}{1-\delta}}} \bar{\Psi}^{1+3r+\frac{2\delta r}{1-\delta}}.$$

Hence, for any  $t_0 > 0$ ,

$$\int_u^{u+t_0} \text{Cov} (G_s^\gamma, G_u^\gamma) ds \lesssim_{r,\delta} \frac{L\gamma t_0}{c^{\frac{2}{1-\delta}}} \bar{\Psi}^{1+3r+\frac{2\delta r}{1-\delta}}. \tag{5.4}$$

We now want to take advantage of the convergence to equilibrium to get a second bound when  $s - u \geq t_0$ : since  $G_u^\gamma$  is  $\mathcal{F}_{\underline{u}_\gamma}$ -measurable, we have for any  $s \geq \underline{u}_\gamma$ ,

$$\mathbb{E} [G_s^\gamma G_u^\gamma] = \mathbb{E}[\mathbb{E}[G_s^\gamma | \mathcal{F}_{\underline{u}_\gamma}] G_u^\gamma].$$

Setting  $\mathfrak{F}(\gamma, t, x) = \mathbb{E}[f(\bar{X}_t^{\gamma,x})] - \pi^\gamma(f)$ , we deduce from the Markov property that

$$\mathbb{E}[G_s^\gamma | \mathcal{F}_{\underline{u}_\gamma}] = \mathfrak{F} \left( \frac{\gamma}{2}, \underline{s}_\gamma - \underline{u}_\gamma, \bar{X}_{\underline{u}_\gamma}^{\frac{\gamma}{2},x} \right) - \mathfrak{F} \left( \gamma, \underline{s}_\gamma - \underline{u}_\gamma, \bar{X}_{\underline{u}_\gamma}^{\gamma,x} \right) + (\pi^\gamma - \pi^{\frac{\gamma}{2}})(f),$$

and hence,

$$\mathbb{E} [G_s^\gamma G_u^\gamma] = \mathbb{E} \left[ \left( \mathfrak{F} \left( \frac{\gamma}{2}, \underline{s}_\gamma - \underline{u}_\gamma, \bar{X}_{\underline{u}_\gamma}^{\frac{\gamma}{2},x} \right) - \mathfrak{F} \left( \gamma, \underline{s}_\gamma - \underline{u}_\gamma, \bar{X}_{\underline{u}_\gamma}^{\gamma,x} \right) \right) G_u^\gamma \right] + (\pi^\gamma - \pi^{\frac{\gamma}{2}})(f) \mathbb{E}[G_u^\gamma].$$

On the other hand,

$$\mathbb{E}[G_s^\gamma] \mathbb{E}[G_u^\gamma] = \mathbb{E} \left[ \left( \mathfrak{F} \left( \frac{\gamma}{2}, \underline{s}_\gamma, \bar{X}_{\underline{u}_\gamma}^{\frac{\gamma}{2},x} \right) - \mathfrak{F} \left( \gamma, \underline{s}_\gamma, \bar{X}_{\underline{u}_\gamma}^{\gamma,x} \right) \right) \right] \mathbb{E} [G_u^\gamma] + (\pi^\gamma - \pi^{\frac{\gamma}{2}})(f) \mathbb{E}[G_u^\gamma].$$

As a consequence,

$$\text{Cov} (G_s^\gamma, G_u^\gamma) = \mathbb{E} \left[ \left( \mathfrak{F} \left( \frac{\gamma}{2}, \underline{s}_\gamma - \underline{u}_\gamma, \bar{X}_{\underline{u}_\gamma}^{\frac{\gamma}{2},x} \right) - \mathfrak{F} \left( \gamma, \underline{s}_\gamma - \underline{u}_\gamma, \bar{X}_{\underline{u}_\gamma}^{\gamma,x} \right) \right) G_u^\gamma \right] \tag{5.5}$$

$$- \mathbb{E} \left[ \left( \mathfrak{F} \left( \frac{\gamma}{2}, \underline{s}_\gamma, x \right) - \mathfrak{F} \left( \gamma, \underline{s}_\gamma, x \right) \right) \right] \mathbb{E} [G_u^\gamma]. \tag{5.6}$$

Let us study the two right-hand members successively. For (5.5), the Cauchy-Schwarz inequality and (5.3) yield:

$$\begin{aligned} & \mathbb{E} \left[ \left( \mathfrak{F} \left( \frac{\gamma}{2}, \underline{s}_\gamma - \underline{u}_\gamma, \bar{X}_{\underline{u}_\gamma}^{\frac{\gamma}{2}, x} \right) - \mathfrak{F} \left( \gamma, \underline{s}_\gamma - \underline{u}_\gamma, \bar{X}_{\underline{u}_\gamma}^{\gamma, x} \right) \right) G_u^\gamma \right] \\ & \lesssim_{r, \delta} \left( \left\| \mathfrak{F} \left( \frac{\gamma}{2}, \underline{s}_\gamma - \underline{u}_\gamma, \bar{X}_{\underline{u}_\gamma}^{\frac{\gamma}{2}, x} \right) \right\|_2 + \left\| \mathfrak{F} \left( \gamma, \underline{s}_\gamma - \underline{u}_\gamma, \bar{X}_{\underline{u}_\gamma}^{\gamma, x} \right) \right\|_2 \right) \sqrt{\frac{L\gamma}{\underline{c}^{\frac{2}{1-\delta}}} \bar{\Psi}^{1+3r+\frac{2\delta r}{1-\delta}}}. \end{aligned}$$

By Proposition 4.11 or more precisely by (4.12) combined with Proposition 4.5(iii)<sup>9</sup> applied with  $\phi(n) = (n\gamma)^\delta$ ,

$$\begin{aligned} & \left\| \mathfrak{F} \left( \frac{\gamma}{2}, \underline{s}_\gamma - \underline{u}_\gamma, \bar{X}_{\underline{u}_\gamma}^{\frac{\gamma}{2}, x} \right) \right\|_2 + \left\| \mathfrak{F} \left( \gamma, \underline{s}_\gamma - \underline{u}_\gamma, \bar{X}_{\underline{u}_\gamma}^{\gamma, x} \right) \right\|_2 \\ & \lesssim_{r, \delta, \kappa} \underline{c}^{-\frac{1}{2}} \bar{\Psi}^{\frac{1+r}{2}} e^{-(\underline{s}_\gamma - \underline{u}_\gamma)^\delta} + \underline{c}^{-\frac{\kappa+1}{2}} \bar{\Psi}^{\frac{1+r(1+\kappa)}{2}} (\underline{s}_\gamma - \underline{u}_\gamma)^{-\frac{\kappa(1-\delta)}{2}}. \end{aligned}$$

Now, let us remark that if  $s - u \geq t_0$ ,  $t_0 \geq 2$  and  $\gamma \in (0, \gamma^*]$ , then  $\underline{s}_\gamma - \underline{u}_\gamma \geq 1$  (since  $\gamma^* \leq 1$ ). Noting that for any  $t \geq 1$ ,  $e^{-t^\delta} \lesssim_\delta t^{-\frac{\kappa(1-\delta)}{2}}$  and that for any  $\kappa > \frac{2}{1-\delta}$ ,

$$\begin{aligned} & \int_{(u+t_0) \wedge T}^T (\underline{s}_\gamma - \underline{u}_\gamma)^{-\frac{\kappa(1-\delta)}{2}} ds \leq \int_{u+t_0}^{+\infty} (s - u - \gamma)^{-\frac{\kappa(1-\delta)}{2}} ds \\ & \leq \frac{(t_0 - \gamma)^{1 - \frac{\kappa(1-\delta)}{2}}}{\frac{\kappa(1-\delta)}{2} - 1} \\ & \lesssim_{\kappa, \delta} \left( \frac{t_0}{2} \right)^{1 - \frac{\kappa(1-\delta)}{2}}. \end{aligned}$$

We deduce that for any  $t_0 \geq 2$  and for any  $\gamma \in (0, \gamma^*]$

$$\begin{aligned} & \int_{(u+t_0) \wedge T}^T \mathbb{E} \left[ \left( \mathfrak{F} \left( \frac{\gamma}{2}, \underline{s}_\gamma - \underline{u}_\gamma, \bar{X}_{\underline{u}_\gamma}^{\frac{\gamma}{2}, x} \right) - \mathfrak{F} \left( \gamma, \underline{s}_\gamma - \underline{u}_\gamma, \bar{X}_{\underline{u}_\gamma}^{\gamma, x} \right) \right) G_u^\gamma \right] ds \\ & \lesssim_{r, \delta, \kappa} \frac{(\gamma L)^{\frac{1}{2}} \bar{\Psi}^{1+2r+\frac{\delta r}{1-\delta}+\frac{\kappa r}{2}}}{\underline{c}^{\frac{3}{2}+\frac{\delta}{1-\delta}} \wedge \underline{c}^{\frac{3+\kappa}{2}+\frac{\delta}{1-\delta}}} \left( \frac{t_0}{2} \right)^{1 - \frac{\kappa(1-\delta)}{2}}. \end{aligned}$$

For (5.6), using that  $\underline{s}_\gamma \geq \underline{s}_\gamma - \underline{u}_\gamma$ , we remark that we can obtain the same bound so that:

$$\int_{(u+t_0) \wedge T}^T \text{Cov} (G_s^\gamma, G_u^\gamma) ds \lesssim_{r, \delta, \kappa} \frac{(\gamma L)^{\frac{1}{2}}}{\underline{c}^{\frac{3}{2}+\frac{\delta}{1-\delta}} \wedge \underline{c}^{\frac{3+\kappa}{2}+\frac{\delta}{1-\delta}}} \bar{\Psi}^{1+2r+\frac{\delta r}{1-\delta}+\frac{\kappa r}{2}} \left( \frac{t_0}{2} \right)^{1 - \frac{\kappa(1-\delta)}{2}}.$$

In view of the above bound and of the one obtained in (5.4), we now optimize the choice of  $t_0$  by taking  $t_0$  solution to:

$$\frac{(\gamma L)^{\frac{1}{2}}}{\underline{c}^{\frac{3}{2}+\frac{\delta}{1-\delta}} \wedge \underline{c}^{\frac{3+\kappa}{2}+\frac{\delta}{1-\delta}}} \bar{\Psi}^{1+2r+\frac{\delta r}{1-\delta}+\frac{\kappa r}{2}} \left( \frac{t_0}{2} \right)^{1 - \frac{\kappa(1-\delta)}{2}} = \frac{L\gamma t_0}{\underline{c}^{\frac{2}{1-\delta}}} \bar{\Psi}^{1+3r+\frac{2\delta r}{1-\delta}}.$$

i.e.,

$$t_0 = 2^{1 - \frac{2}{\kappa(1-\delta)}} \left( \underline{c}^{\frac{1}{2}+\frac{\delta}{1-\delta}} \vee \underline{c}^{\frac{1-\kappa}{2}+\frac{\delta}{1-\delta}} \right)^{\frac{2}{\kappa(1-\delta)}} (L\gamma)^{-\frac{1}{\kappa(1-\delta)}} \bar{\Psi}^{(1-\frac{2}{\kappa})\frac{r}{1-\delta} - \frac{2\delta}{\kappa(1-\delta)}r}.$$

Plugging this value of  $t_0$  into (5.4), this leads to: for any  $\kappa > \frac{2}{1-\delta}$ ,

$$\int_u^T \text{Cov} (G_s^\gamma, G_u^\gamma) ds \lesssim_{r, \delta, \kappa} (\gamma L)^{1 - \frac{1}{\kappa(1-\delta)}} \left( \underline{c}^{-\frac{3}{2} - \frac{\delta}{1-\delta}} \vee \underline{c}^{-\frac{3+\kappa}{2} - \frac{\delta}{1-\delta}} \right)^{\frac{2}{\kappa(1-\delta)}} \bar{\Psi}^{1+(4-\frac{2}{\kappa})r+\varepsilon(\delta, \kappa)r},$$

where  $\varepsilon(\delta, \kappa) = \left( 3 - \frac{2}{\kappa} - \frac{2}{\kappa(1-\delta)} \right) \frac{\delta}{1-\delta}$ . The result follows.  $\square$

<sup>9</sup> In fact, Proposition 4.11 is written under the assumption  $U(x) \lesssim \bar{\Psi}$  but here, we need to integrate with respect to the initial condition. To extend to this setting, the idea is to start from (4.12) and to use the bounds of Proposition 4.5(iii). This allows us to retrieve controls which are similar to Proposition 4.11.

In the next proposition, we are now able to work on the variance of the multilevel procedure.

**Proposition 5.5.** Assume  $(\mathbf{H}_r^1)$  and  $\gamma_0 \in (0, \gamma^*)$ . Let  $x \in \mathbb{R}^d$  such that  $U(x) \lesssim_r \bar{\Psi}$ . Let  $\delta \in (0, \frac{1}{2}]$  and  $\kappa > \frac{2}{1-\delta}$ . Let  $f$  be a Lipschitz continuous function. Then,

$$\text{Var}(\mathcal{Y}(J, (\gamma_j)_j, (T_j)_j, f)) \lesssim_{r, \delta, \kappa} [f]_1^2 \left( \mathfrak{C}_{\text{var}}^{(1)}(\kappa, \delta) \sum_{j=1}^J \frac{\gamma_j^{1-\frac{1}{\kappa(1-\delta)}}}{T_j} + \frac{\mathfrak{C}_{\text{var}}^{(2)}}{T_0} \right),$$

where  $\mathfrak{C}_{\text{var}}^{(1)}(\kappa, \delta)$  is defined in Lemma 5.3 and  $\mathfrak{C}_{\text{var}}^{(2)} = \left( \underline{c}^{-\frac{3}{2}} \vee \underline{c}^{-\frac{5}{2} - \frac{2\delta}{1-\delta}} \right) \bar{\Psi}^{1+2r+\frac{2\delta r}{1-\delta}}$ .

*Proof.* We assume (without loss of generality) that  $[f]_1 = 1$  and  $f(x^*) = 0$ . In view of the decomposition (5.2), we apply Lemma 5.3 for each level  $j \in \{1, \dots, J\}$  with  $T = T_j$  and  $\gamma = \gamma_{j-1}$ . We obtain for any  $\delta \in (0, 1/2]$  and  $\kappa > 2/(1 - \delta)$ ,

$$\text{Var}(\mathcal{Y}(J, (\gamma_j)_j, (T_j)_j, f)) \leq \text{Var} \left( \frac{1}{T_0} \int_0^{T_0} f(\bar{X}_{\underline{s}_{\gamma_0}}^{\gamma_0, x_0}) ds \right) + \mathfrak{C}_{\text{var}}^{(1)}(\kappa, \delta) \sum_{j=1}^J \frac{\gamma_j^{1-\frac{1}{\kappa(1-\delta)}}}{T_j}.$$

It remains to control the first term, i.e. the variance related to the first level. We use similar arguments as in the proof of Lemma 5.3 (see in particular (5.5) and what follows). First, one can check that for every  $0 \leq u \leq s \leq T$ ,

$$\begin{aligned} & \text{Cov} \left( f(\bar{X}_{\underline{u}_{\gamma_0}}^{\gamma_0, x_0}), f(\bar{X}_{\underline{s}_{\gamma_0}}^{\gamma_0, x_0}) \right) \\ &= \mathbb{E} \left[ \mathfrak{F} \left( \gamma_0, \underline{s}_{\gamma_0} - \underline{u}_{\gamma_0}, \bar{X}_{\underline{u}_{\gamma_0}}^{\gamma_0, x} \right) f(\bar{X}_{\underline{u}_{\gamma_0}}^{\gamma_0, x_0}) \right] - \mathbb{E} \left[ \mathfrak{F} \left( \gamma_0, \underline{s}_{\gamma_0}, x \right) \right] \mathbb{E} \left[ f(\bar{X}_{\underline{u}_{\gamma_0}}^{\gamma_0, x_0}) \right] \\ &\leq \left( \left\| \mathfrak{F} \left( \gamma_0, \underline{s}_{\gamma_0} - \underline{u}_{\gamma_0}, \bar{X}_{\underline{u}_{\gamma_0}}^{\gamma_0, x} \right) \right\|_2 + \left\| \mathfrak{F} \left( \gamma_0, \underline{s}_{\gamma_0}, x \right) \right\|_2 \right) \left\| f \left( \bar{X}_{\underline{u}_{\gamma_0}}^{\gamma_0, x_0} \right) \right\|_2, \end{aligned}$$

by Cauchy-Schwarz inequality. Then, by Proposition 4.11 (and footnote 9) applied with  $\phi(n) = n\gamma$  and  $\kappa = \frac{2(1+\delta)}{1-\delta}$ , one deduces that (we leave the details to the reader),

$$\begin{aligned} \int_u^T \text{Cov} \left( f(\bar{X}_{\underline{u}_{\gamma_0}}^{\gamma_0, x_0}), f(\bar{X}_{\underline{s}_{\gamma_0}}^{\gamma_0, x_0}) \right) ds &\lesssim_{r, \delta} \left( \underline{c}^{-\frac{1}{2}} \vee \underline{c}^{-\frac{3}{2} - \frac{2\delta}{1-\delta}} \right) \bar{\Psi}^{\frac{1+3r}{2} + \frac{2\delta r}{1-\delta}} \times \sup_{u \geq 0} \left\| f \left( \bar{X}_{\underline{u}_{\gamma_0}}^{\gamma_0, x_0} \right) \right\|_2 \\ &\lesssim_{r, \delta} \left( \underline{c}^{-\frac{3}{2}} \vee \underline{c}^{-\frac{5}{2} - \frac{2\delta}{1-\delta}} \right) \bar{\Psi}^{1+2r+\frac{2\delta r}{1-\delta}}, \end{aligned}$$

where in the second line, we used Proposition 4.5(iii) and the fact that (by Lemma 4.9)

$$|f(x)|^2 = |f(x) - f(x^*)|^2 \leq |x - x^*|^2 \leq \frac{U^{1+r}(x)}{\underline{c}}. \quad \square$$

### 5.3 Step 3: Proof of Theorem 2.3

Back to the bias-variance decomposition, we deduce from Proposition 5.2 and Proposition 5.5 that, up to a constant depending on  $\kappa, \delta$  and  $r$ , the MSE is lower than  $\varepsilon^2$  if the following conditions are satisfied (with  $\delta \in (0, 1/2]$  and  $\kappa > 2/(1 - \delta)$ ):

$$\left\{ \begin{array}{l} \text{(i.a): } \mathfrak{C}_{\text{bias}}^{(1)} \sum_{j=0}^J \frac{1}{T_j^2} \leq \varepsilon^2, \quad \text{(i.b): } \mathfrak{C}_{\text{bias}}^{(2,1)} \gamma_J \leq \varepsilon^2 \quad \text{or} \quad \text{(i.b)': } \mathfrak{C}_{\text{bias}}^{(2,2)} \gamma_J^2 \leq \varepsilon^2, \\ \text{(ii.a): } \frac{\mathfrak{C}_{\text{var}}^{(2)}}{T_0} \leq \varepsilon^2, \quad \text{(ii.b): } \mathfrak{C}_{\text{var}}^{(1)}(\kappa, \delta) \sum_{j=1}^J \frac{\gamma_j^{1-\frac{1}{\kappa(1-\delta)}}}{T_j} \leq \varepsilon^2. \end{array} \right. \quad (5.7)$$

Note that (i.b) corresponds to Proposition 5.2(i) whereas (i.b)' corresponds to Proposition 5.2(ii). Let us assume that

$$\gamma_j = \gamma_0 2^{-j} \quad \text{and} \quad T_j = T_0 2^{-(1-\rho)j},$$

where  $\rho \geq 1/2$ ,  $\gamma_0$  and  $T_0$  are positive numbers which will be calibrated further. With these choices, the above conditions read (up to universal constants):

$$\begin{cases} \text{(i.a)} : & T_0 \geq 2^{(1-\rho)J} \sqrt{\mathfrak{C}_{\text{bias}}^{(1)}} \varepsilon^{-1} & \text{(i.b)} : & 2^J \geq \mathfrak{C}_{\text{bias}}^{(2,1)} \gamma_0 \varepsilon^{-2} \quad \text{or} \quad \text{(i.b)' } : & 2^J \geq \sqrt{\mathfrak{C}_{\text{bias}}^{(2,2)}} \gamma_0 \varepsilon^{-1}, \\ \text{(ii.a)} : & T_0 \geq \mathfrak{C}_{\text{var}}^{(2)} \varepsilon^{-2}, & \text{(ii.b)} : & T_0 \geq \mathfrak{C}_{\text{var}}^{(1)}(\kappa, \delta) \gamma_0^{1-\frac{1}{\kappa(1-\delta)}} \sum_{j=1}^J 2^{(-\rho+\frac{1}{\kappa(1-\delta)})j} \varepsilon^{-2}. \end{cases} \tag{5.8}$$

*Proof of Theorem 2.3 (i)* In this case, we have to calibrate the parameters according to (i.a), (i.b), (ii.a) and (ii.b) are satisfied. First, for (i.b), we need  $2^J \geq \mathfrak{C}_{\text{bias}}^{(2,1)} \gamma_0 \varepsilon^{-2}$  so we can set:

$$J = \left\lceil \log_2 \left( \mathfrak{C}_{\text{bias}}^{(2,1)} \gamma_0 \varepsilon^{-2} \right) \right\rceil.$$

Then, set  $\rho = 1/2$ . With the above value of  $J$  and the condition  $\kappa > 2/(1-\delta)$ ,

$$2^{\frac{J}{2}} \leq \sqrt{2\gamma_0 \mathfrak{C}_{\text{bias}}^{(2,1)}} \varepsilon^{-1} \quad \text{and} \quad \sum_{j=1}^J 2^{(-\frac{1}{2}+\frac{1}{\kappa(1-\delta)})j} \lesssim_{\kappa, \delta} 1.$$

Hence, (i.a), (ii.a), (ii.b) are satisfied (up to a constant depending on  $\kappa, \delta$  and  $r$  only) if

$$T_0 \geq \mathfrak{C} \varepsilon^{-2} \quad \text{with} \quad \mathfrak{C} = \max \left( \sqrt{\gamma_0 \mathfrak{C}_{\text{bias}}^{(1)} \mathfrak{C}_{\text{bias}}^{(2,1)}}, \mathfrak{C}_{\text{var}}^{(2)}, \mathfrak{C}_{\text{var}}^{(1)}(\kappa, \delta) \gamma_0^{1-\frac{1}{\kappa(1-\delta)}} \right).$$

The complexity of the procedure is then:

$$\mathcal{C}(\mathcal{Y}) = 2 \left( \frac{T_0}{\gamma_0} + \sum_{j=1}^J \frac{T_j}{\gamma_j} \right) \lesssim \frac{T_0}{\gamma_0} 2^{\frac{J}{2}} \lesssim_{r, \delta, \kappa} \frac{\mathfrak{C} \sqrt{\mathfrak{C}_{\text{bias}}^{(2,1)}} \varepsilon^{-3}}{\sqrt{\gamma_0}}.$$

To deduce the result, it is now enough to remark that with  $\kappa = \frac{2(1+\delta)}{1-\delta}$ ,  $\gamma_0 \leq 1/(4L)$ ,

$$\mathfrak{C} \leq (\underline{c}^{-\frac{3}{4}} \vee \underline{c}^{-\frac{5}{2}-\frac{2\delta}{1-\delta}}) \bar{\Psi}^{1+3r+\frac{3\delta r}{1-\delta}} \quad \text{so that} \quad \mathfrak{C} \sqrt{\mathfrak{C}_{\text{bias}}^{(2,1)}} \leq \sqrt{L} (\underline{c}^{-\frac{5}{4}} \wedge \underline{c}^{-\frac{7}{2}-\frac{3\delta}{1-\delta}}) \bar{\Psi}^{\frac{3}{2}+\frac{3}{2}r+\frac{4\delta r}{1-\delta}},$$

as soon as  $\delta < 1/3$ . The main result then follows from a change of variable replacing  $\delta$  by  $\tilde{\delta} = a\delta$  with  $a$  small enough.

(ii) In this case, we have to calibrate the parameters according to (i.a), (i.b)', (ii.a) and (ii.b). First, for (i.b)', we need  $2^J \geq \sqrt{\mathfrak{C}_{\text{bias}}^{(2,2)}} \gamma_0 \varepsilon^{-1}$  so we can set:

$$J = \left\lceil \log_2 \left( \sqrt{\mathfrak{C}_{\text{bias}}^{(2,2)}} \gamma_0 \varepsilon^{-1} \right) \right\rceil.$$

Then, if  $\rho > \frac{1}{\kappa(1-\delta)}$ , (i.a), (ii.a), (ii.b) are satisfied (up to a constant depending on  $\kappa, \delta$  and  $r$  only) if

$$T_0 \geq \tilde{\mathfrak{C}} \varepsilon^{-2} \quad \text{with} \quad \tilde{\mathfrak{C}} = \max \left( \sqrt{\mathfrak{C}_{\text{bias}}^{(1)}} \left( \gamma_0^2 \mathfrak{C}_{\text{bias}}^{(2,2)} \right)^{\frac{1-\rho}{2}}, \mathfrak{C}_{\text{var}}^{(2)}, \mathfrak{C}_{\text{var}}^{(1)}(\kappa, \delta) \gamma_0^{1-\frac{1}{\kappa(1-\delta)}} \right),$$

and the complexity of the procedure satisfies:

$$\mathcal{C}(\mathcal{Y}) \lesssim \frac{T_0}{\gamma_0} 2^{\rho J} \lesssim_{r, \delta, \kappa, \rho} \frac{\tilde{\mathcal{C}}(\mathfrak{C}_{\text{bias}}^{(2,2)})^{\frac{\rho}{2}} \varepsilon^{-2-\rho}}{\gamma_0^\rho}.$$

Set  $\kappa = \frac{1+\delta}{\rho(1-\delta)} = \frac{1}{\rho} + \frac{2\delta}{\rho(1-\delta)}$ . For  $\delta$  small enough and  $\gamma_0 \leq 1/L$ ,

$$\gamma_0^{1 - \frac{1}{\kappa(1-\delta)}} \mathfrak{C}_{\text{var}}^{(1)}(\kappa, \delta) \leq (\underline{c}^{-3\rho(1-\frac{\delta}{1-\delta})} \vee \underline{c}^{-(1+3\rho)-5\delta}) \bar{\Psi}^{1+(4-2\rho)r + \frac{4\delta r}{1-\delta}}.$$

Using that  $\gamma_0 \leq 1/(4L)$ ,

$$\sqrt{\mathfrak{C}_{\text{bias}}^{(1)}} \left( \gamma_0^2 \mathfrak{C}_{\text{bias}}^{(2,2)} \right)^{\frac{1-\rho}{2}} \leq L^{\frac{\rho}{2}} (\underline{c}^{-\frac{1}{4} - \frac{1-\rho}{1-\delta}} \vee \underline{c}^{-\frac{3}{2} - \frac{1-\rho}{1-\delta} - \frac{2\delta}{1-\delta}}) \bar{\Psi}^{1-\frac{\rho}{2} + \frac{5-2\rho}{2}r + (3-\rho)\frac{\delta r}{1-\delta}}.$$

Using that  $\rho \leq 1/2$ , this yields (at the price of replacing  $\delta$  by  $a\delta$  for  $a$  small enough)

$$\tilde{\mathcal{C}} \leq L^{\frac{\rho}{2}} \left( \underline{c}^{-(\frac{5}{4}-\rho)\wedge(3\rho)+\delta} \vee \underline{c}^{-\frac{5}{2}-\delta} \right) \bar{\Psi}^{1+(4-2\rho+\delta)r}.$$

The result follows.

## Supplementary Material

### 6 Proof of Proposition 4.3

Consider the process  $(\bar{X}_t)_{t>0}$  defined by  $\bar{X}_t = x - t\nabla U(x) + \sigma\sqrt{t}Z$  where  $Z = (Z_1, \dots, Z_d) \sim \mathcal{N}(0, \text{Id}_{\mathbb{R}^d})$ . By a Taylor expansion with integral remainder of the function  $U$  we have

$$\begin{aligned} U(\bar{X}_t) &= U(x) + \langle \nabla U(x), \bar{X}_t - x \rangle + \int_0^1 \langle D^2U(\xi_\lambda)(\bar{X}_t - x), \bar{X}_t - x \rangle d\lambda \\ &\leq U(x) + \langle \nabla U(x), \bar{X}_t - x \rangle + |\bar{X}_t - x|^2 \int_0^1 \bar{\lambda}_{D^2U(\xi_\lambda)} d\lambda, \end{aligned} \tag{6.1}$$

where  $\xi_\lambda = \lambda\bar{X}_t + (1-\lambda)x$ .

(i) In this first part, we only assume that  $\nabla U$  is  $L$ -Lipschitz so that  $\bar{\lambda}_{D^2U(\xi_\lambda)} \leq L$ . Thus, since

$$|\bar{X}_t - x|^2 \leq 2t^2|\nabla U(x)|^2 + 2t\sigma^2|Z|^2, \tag{6.2}$$

we get

$$U(\bar{X}_t) \leq U(x) - t|\nabla U(x)|^2(1-2Lt) + \sigma\sqrt{t}\langle \nabla U(x), Z \rangle + 2L\sigma^2t|Z|^2.$$

Let  $\theta > 0$  and define,

$$f_\theta : x \mapsto e^{\theta U(x)} \tag{6.3}$$

We deduce from the previous inequality that

$$\begin{aligned} \mathbb{E}[f_\theta(\bar{X}_t)] &\leq f_\theta(x) e^{-\theta t|\nabla U(x)|^2(1-2Lt)} \mathbb{E} \left[ e^{\theta\sigma\sqrt{t}\langle \nabla U(x), Z \rangle + 2\theta L\sigma^2t|Z|^2} \right] \\ &\leq f_\theta(x) e^{-\theta t|\nabla U(x)|^2(1-2Lt)} \prod_{i=1}^d \mathbb{E}[e^{\alpha_i Z_i + \beta_i |Z_i|^2}], \end{aligned}$$

where  $\alpha_i = \theta\sigma\sqrt{t}\partial_i U(x)$  and  $\beta_i = 2\theta L\sigma^2t$ . A standard computation shows that for any  $u \in \mathbb{R}$  and any  $v < 1/2$ ,

$$\mathbb{E}_{Z_1 \sim \mathcal{N}(0,1)}[e^{uZ_1 + vZ_1^2}] = \frac{1}{\sqrt{1-2v}} e^{\frac{u^2}{2(1-2v)}}. \tag{6.4}$$

Thus,

$$\prod_{i=1}^d \mathbb{E}[e^{\alpha_i Z_i + \beta_i |Z_i|^2}] = \left( \frac{1}{1 - 2\theta L \sigma^2 t} \right)^{\frac{d}{2}} e^{\frac{\theta^2 \sigma^2 t |\nabla U(x)|^2}{2(1 - 2\theta L \sigma^2 t)}}$$

This yields

$$\mathbb{E}[f_\theta(\bar{X}_t)] \leq f_\theta(x) e^{-\theta t |\nabla U(x)|^2 (1 - 2Lt)} e^{-\frac{d}{2} \log(1 - 2\theta L \sigma^2 t)} \exp\left(\frac{\theta^2 \sigma^2 t |\nabla U(x)|^2}{2(1 - 2\theta L \sigma^2 t)}\right).$$

Since,  $\gamma L \leq 1/4$  and  $\theta \sigma^2 \leq 1/8$ ,  $2\theta L \sigma^2 t \in [0, 1/2]$  and we can use the elementary inequality  $\log(1 - x) \geq -2x$  for all  $x \in [0, 1/2]$  to obtain:

$$\mathbb{E}[f_\theta(\bar{X}_t)] \leq f_\theta(x) e^{-c_{t,x} \theta |\nabla U(x)|^2 + 2d\theta L \sigma^2 t} \quad \text{with} \quad c_{t,x} = \theta t (1 - 2Lt - \theta \sigma^2) \geq \frac{\theta t}{4}.$$

For  $t = \gamma$ , this yields

$$\mathbb{E}[f_\theta(\bar{X}_t)] \leq f_\theta(x) e^{-\frac{\theta \gamma}{4} (|\nabla U(x)|^2 - 8dL)}.$$

Let

$$\mathcal{C}_M := \{x \in \mathbb{R}^d, |\nabla U(x)|^2 - 8dL \leq M\}.$$

We get

$$\begin{aligned} \mathbb{E}[f_\theta(\bar{X}_t)] &\leq f_\theta(x) e^{2\theta \gamma dL} 1_{\{x \in \mathcal{C}_M\}} + f_\theta(x) e^{-\frac{\theta \gamma M}{4}} 1_{\{x \in \mathcal{C}_M^c\}} \\ &\leq f_\theta(x) e^{-\frac{\theta \gamma M}{4}} + \sup_{x \in \mathcal{C}_M} f_\theta(x) (e^{2\theta \gamma dL} - e^{-\frac{\theta \gamma M}{4}}) \end{aligned}$$

In order to control  $\sup_{x \in \mathcal{C}_M} f_\theta(x)$ , one needs to include  $\mathcal{C}_M$  in a level set of  $U$ . By (4.1) and the fact that  $U(x^*) = 1$ , one checks that

$$\mathcal{C}_M \subset \left\{ x \in \mathbb{R}^d, U(x) \leq \left( 1 + \frac{M + 8dL}{2\underline{c}} \right)^{\frac{1}{1-r}} \right\}$$

so that with  $t = \gamma$ ,

$$\mathbb{E}[f_\theta(\bar{X}_\gamma)] \leq e^{-\frac{\theta \gamma M}{4}} f_\theta(x) + \mathfrak{c}(M, \gamma, \theta) \quad \text{with} \quad \mathfrak{c}(M, \gamma, \theta) = e^{\theta \left( 1 + \frac{M + 8dL}{2\underline{c}} \right)^{\frac{1}{1-r}}} (e^{2\theta \gamma dL} - e^{-\frac{\theta \gamma M}{4}}).$$

An induction leads to

$$\sup_{n \geq 0} \mathbb{E}[f_\theta(\bar{X}_{n\gamma})] \leq f_\theta(x) + \frac{\mathfrak{c}(M, \gamma, \theta)}{1 - e^{-\frac{\theta \gamma M}{4}}}.$$

Setting  $M = 8dL$ ,

$$\frac{\mathfrak{c}(M, \gamma, \theta)}{1 - e^{-\frac{\theta \gamma M}{4}}} = \frac{e^{\theta \left( 1 + \frac{8dL}{\underline{c}} \right)^{\frac{1}{1-r}} \sinh(2\theta \gamma dL)}}{e^{-\theta \gamma dL} \sinh(\theta \gamma dL)} = e^{\theta \left( 1 + \frac{8dL}{\underline{c}} \right)^{\frac{1}{1-r}} + \theta dL} \cosh(\theta \gamma dL).$$

(ii) We now deal with the additional assumption  $(\mathbf{H}_r^2)$ . The idea is now to refine the controls by taking into account that the largest eigenvalue also decreases at infinity. Unfortunately, this refinement will require additional technicalities and concentration arguments. By (6.1), (6.2) and  $(\mathbf{H}_r^2)$ ,

$$U(\bar{X}_t) \leq U(x) - t |\nabla U(x)|^2 (1 - 2tL) + \sigma \sqrt{t} \langle \nabla U(x), Z \rangle + 2\bar{c} \sigma^2 t |Z|^2 \int_0^1 U^{-r}(\xi_\lambda) d\lambda.$$

Keeping in mind that  $f_\theta(x) = e^{\theta U(x)}$ , this leads to

$$\begin{aligned} & \mathbb{E} [f_\theta(\bar{X}_t)] \\ & \leq f_\theta(x) e^{-\theta t |\nabla U(x)|^2 (1-2tL)} \underbrace{\mathbb{E} \left[ \exp \left( \theta \sigma \sqrt{t} \langle \nabla U(x), Z \rangle + 2\theta \bar{c} \sigma^2 t |Z|^2 \int_0^1 U^{-r}(\xi_\lambda) d\lambda \right) \right]}_{=: E}. \end{aligned} \tag{6.5}$$

In order to get a sharp bound of the expectation  $E$ , we choose to divide it into two parts depending on a parameter  $K$  which will be calibrated below:

$$\begin{aligned} E &= \underbrace{\mathbb{E} \left[ \exp \left( \theta \sigma \sqrt{t} \langle \nabla U(x), Z \rangle + 2\theta \bar{c} \sigma^2 t |Z|^2 \right) 1_{\{|Z|^2 \geq K\}} \right]}_{=: E_1} \\ &+ \underbrace{\mathbb{E} \left[ \exp \left( \theta \sigma \sqrt{t} \langle \nabla U(x), Z \rangle + 2\theta \bar{c} \sigma^2 t |Z|^2 \int_0^1 U^{-r}(\xi_\lambda) d\lambda \right) 1_{\{|Z|^2 < K\}} \right]}_{=: E_2}, \end{aligned}$$

**Bound for  $E_1$ :** By Cauchy-Schwarz inequality we have

$$\begin{aligned} E_1 &\leq \mathbb{E} \left[ \exp \left( 2\theta \sigma \sqrt{t} \langle \nabla U(x), Z \rangle + 4\theta \bar{c} \sigma^2 t |Z|^2 \right) \right]^{1/2} \mathbb{P}(|Z|^2 \geq K)^{1/2} \\ &\leq \prod_{i=1}^d \mathbb{E}_{Z_i \sim \mathcal{N}(0,1)} \left[ \exp \left( 2\theta \sigma \sqrt{t} \partial_i U(x) Z_i + 4\theta \bar{c} \sigma^2 t Z_i^2 \right) \right]^{1/2} \mathbb{P}(|Z|^2 \geq K)^{1/2}. \end{aligned}$$

By (6.4), this yields

$$\begin{aligned} E_1 &\leq \left( \left( \frac{1}{1 - 8\theta \bar{c} \sigma^2 t} \right)^{\frac{d}{2}} \exp \left( \frac{2\theta^2 \sigma^2 t |\nabla U(x)|^2}{1 - 8\theta \bar{c} \sigma^2 t} \right) \right)^{1/2} \mathbb{P}(|Z|^2 \geq K)^{1/2} \\ &\leq \exp \left( -\frac{d}{4} \log(1 - 8\theta \bar{c} \sigma^2 t) + \frac{2\theta^2 t \sigma^2 |\nabla U(x)|^2}{1 - 8\theta \bar{c} \sigma^2 t} \right) \mathbb{P}(|Z|^2 \geq K)^{1/2}. \end{aligned}$$

By exponential Markov inequality and (6.4) applied with  $u = 0$  and  $v = 1/4$ ,

$$\begin{aligned} E_1 &\leq \exp \left( -\frac{d}{4} \log(1 - 8\theta \bar{c} \sigma^2 t) + \frac{2\theta^2 t \sigma^2 |\nabla U(x)|^2}{1 - 8\theta \bar{c} \sigma^2 t} - \frac{K}{8} \right) \mathbb{E} \left[ e^{\frac{|Z|^2}{4}} \right]^{1/2} \\ &\leq \exp \left( -\frac{d}{4} \log(1 - 8\theta \bar{c} \sigma^2 t) + \frac{2\theta^2 t \sigma^2 |\nabla U(x)|^2}{1 - 8\theta \bar{c} \sigma^2 t} - \frac{K}{8} + \frac{d \log 2}{4} \right). \end{aligned}$$

Under the assumptions on  $\theta$  and  $\bar{c}$ ,  $8\theta \bar{c} \sigma^2 t \leq \frac{1}{4}$  for any  $t \in [0, \gamma]$ . Using the elementary inequality  $\log(1 - x) \geq -2x$  for all  $x \in [0, 1/2]$ ,

$$\begin{aligned} E_1 &\leq \exp \left( 4d\theta \bar{c} \sigma^2 t + \frac{2\theta^2 t \sigma^2 |\nabla U(x)|^2}{1 - 8\theta \bar{c} \sigma^2 t} - \frac{K}{8} + \frac{d \log 2}{4} \right) \\ &\leq \exp \left( \frac{2\theta^2 t \sigma^2 |\nabla U(x)|^2}{1 - 8\theta \bar{c} \sigma^2 t} - \frac{K}{8} + \frac{d}{8} + \frac{d \log 2}{4} \right). \end{aligned}$$

Using again that  $\theta \sigma^2 \leq 1/8$  and  $8\theta \bar{c} \sigma^2 t \leq \frac{1}{4}$ , we deduce that

$$E_1 \leq \exp \left( \frac{\theta t}{3} |\nabla U(x)|^2 - c_K \right), \quad \text{with } c_K = \frac{1}{8} (K - d(1 + 2 \log 2)). \tag{6.6}$$

**Bound for  $E_2$ :** First, we have

$$\sup_{\xi \in [x, \bar{X}_t]} 1_{\{|Z|^2 \leq K\}} U^{-r}(\xi) \leq \sup_{\xi \in B(x, t|\nabla U(x)| + \sqrt{tK}\sigma^2)} U^{-r}(\xi) =: C_{U,K}(x),$$

that gives

$$E_2 \leq \mathbb{E} \left[ \exp \left( \sigma \theta \sqrt{t} \langle \nabla U(x), Z \rangle + 2\sigma^2 \theta \bar{c} t |Z|^2 C_{U,K}(x) \right) \right].$$

Using again (6.4), this involves that

$$\begin{aligned} E_2 &\leq \exp \left( -\frac{d}{2} \log(1 - 4\theta \bar{c} \sigma^2 t C_{U,K}(x)) + \frac{\sigma^2 \theta^2 t |\nabla U(x)|^2}{2(1 - 4\theta \bar{c} t \sigma^2)} \right) \\ &\leq \exp \left( 4d\theta \bar{c} \sigma^2 t C_{U,K}(x) + \frac{\sigma^2 \theta^2 t |\nabla U(x)|^2}{2(1 - 4\theta \bar{c} t \sigma^2)} \right), \end{aligned}$$

where in the last line we used the inequality  $\log(1 - x) \geq -2x$  for all  $x \in [0, 1/2]$ . Once again, under  $\theta \sigma^2 \leq 1/8$  and  $t\bar{c} \leq 1/4$ , one checks that  $\sigma^2 \theta (2(1 - 4\theta \bar{c} t \sigma^2))^{-1} \leq 2/7 \leq 1/3$ . Thus,

$$E_2 \leq \exp \left( 4d\theta \bar{c} \sigma^2 t C_{U,K}(x) + \frac{t\theta}{3} |\nabla U(x)|^2 \right). \tag{6.7}$$

Plugging (6.6) and (6.7) into (6.5), we obtain

$$\begin{aligned} \mathbb{E} [f_\theta(\bar{X}_t)] &\leq f_\theta(x) \exp \left( -\theta t |\nabla U(x)|^2 \left( 1 - 2tL - \frac{1}{3} \right) \right) \left( e^{-c_K} + e^{4d\sigma^2 \theta \bar{c} t C_{U,K}(x)} \right) \\ &\leq f_\theta(x) e^{-c_K} + f_\theta(x) \exp \left( -t \left( \frac{\theta}{6} |\nabla U(x)|^2 - 4d\sigma^2 \theta \bar{c} C_{U,K}(x) \right) \right), \end{aligned}$$

where in the second line, we used that  $2tL \leq 1/2$ . Now, let us follow the same strategy as in the first case by setting

$$\bar{C}_M = \left\{ x \in \mathbb{R}^d; \frac{\theta}{6} |\nabla U(x)|^2 - 4d\sigma^2 \theta \bar{c} C_{U,K}(x) \leq M \right\}.$$

Then

$$\begin{aligned} \mathbb{E} [f_\theta(\bar{X}_t)] &\leq f_\theta(x) e^{-c_K} + f_\theta(x) e^{-tM} 1_{\bar{C}_M}(x) \\ &\quad + f_\theta(x) \exp \left( 4td\sigma^2 \theta \bar{c} C_{U,K}(x) - \frac{\theta t}{6} |\nabla U(x)|^2 \right) 1_{\bar{C}_M}(x) \\ &\leq f_\theta(x) (e^{-c_K} + e^{-tM}) \\ &\quad + f_\theta(x) \left( \exp \left( 4td\sigma^2 \theta \bar{c} C_{U,K}(x) - \frac{\theta t}{6} |\nabla U(x)|^2 \right) - e^{-tM} \right) 1_{\bar{C}_M}, \end{aligned} \tag{6.8}$$

Now, let us show that  $\bar{C}_M$  is included in a level set of  $U$ . First, following the arguments which lead to (4.1)

$$|\nabla U(x)|^2 \geq \frac{2\bar{c}}{1-r} (U^{1-r}(x) - U^{1-r}(x^*)). \tag{6.9}$$

in the main document, one is also able to show that

$$|\nabla U(x)|^2 \leq \frac{2\bar{c}}{1-r} U^{1-r}(x), \tag{6.10}$$

which in turn implies that  $|\nabla \sqrt{U}|$  is bounded by  $\sqrt{\frac{\bar{c}}{2(1-r)}}$ . Thus,  $\sqrt{U}$  is  $\sqrt{\frac{\bar{c}}{2(1-r)}}$ -Lipschitz<sup>10</sup> and

$$\forall \xi \in B \left( x, t|\nabla U(x)| + \sqrt{tK\sigma^2} \right), \quad U^{1/2}(\xi) \geq U^{1/2}(x) - \sqrt{\frac{\bar{c}}{2(1-r)}} |\xi - x|.$$

<sup>10</sup> $\sqrt{U}$  is also  $\sqrt{L/2}$ -Lipschitz. This could be alternatively used in this proof.

By the triangular inequality we get

$$C_{U,K}(x) \leq U^{-r}(x) \left( 1 - t \frac{|\nabla U(x)|}{U^{1/2}(x)} \sqrt{\frac{\bar{c}}{2(1-r)}} - U^{-1/2}(x) \sqrt{\frac{tK\sigma^2\bar{c}}{2(1-r)}} \right)^{-2r},$$

$$\leq U^{-r}(x) \left( 1 - \frac{t\bar{c}}{(1-r)} - U^{-1/2}(x) \sqrt{\frac{\gamma K\sigma^2\bar{c}}{2(1-r)}} \right)^{-2r},$$

where in the second line, we used (6.10) and the fact that  $t \in [0, \gamma]$ . Since  $\frac{t\bar{c}}{1-r} \leq \frac{1}{4}$ , one can check here that if  $U(x) \geq \frac{8\gamma K\sigma^2\bar{c}}{1-r}$ , we have for every  $t \in [0, \gamma]$ ,

$$1 - t \frac{\bar{c}}{(1-r)} - U^{-1/2}(x) \sqrt{\frac{\gamma K\sigma^2\bar{c}}{2(1-r)}} \geq \frac{1}{2} \implies C_{U,K}(x) \leq 2U^{-r}(x).$$

With the help of (4.1), this implies that

$$\left\{ x \in \mathbb{R}^d; \frac{\theta \underline{c}}{3(1-r)}(U^{1-r}(x) - U^{1-r}(x^*)) - 8d\sigma^2\theta\bar{c}U^{-r}(x) > M \right\} \tag{6.11}$$

$$\cap \left\{ x \in \mathbb{R}^d; U(x) \geq \frac{8\gamma K\sigma^2\bar{c}}{1-r} \right\} \subset \mathcal{C}_M^c.$$

Keeping in mind that  $U(x^*) = 1$ , one can write

$$\frac{\theta \underline{c}}{3(1-r)}(U^{1-r}(x) - U^{1-r}(x^*)) - 8d\sigma^2\theta\bar{c}U^{-r}(x)$$

$$= \frac{\theta \underline{c}}{3(1-r)}U^{1-r}(x) \left( 1 - U^{r-1}(x) - \frac{24(1-r)d\sigma^2\bar{c}}{\underline{c}U(x)} \right),$$

and one can deduce that this term is greater than  $M$  if

$$\frac{\theta \underline{c}}{3(1-r)}U^{1-r}(x) > 2M, \quad U^{r-1}(x) \leq \frac{1}{4} \quad \text{and} \quad \frac{24(1-r)d\sigma^2\bar{c}}{\underline{c}U(x)} \leq \frac{1}{4}.$$

From these conditions and (6.11), we finally get

$$\mathcal{C}_M \subset \left\{ x \in \mathbb{R}^d; U(x) \leq \max \left( \underbrace{\frac{96d\sigma^2\bar{c}}{\underline{c}}, \left( \frac{2M(1-r)}{\theta\bar{c}} \vee 4 \right)^{\frac{1}{1-r}}, \frac{8\gamma K\sigma^2\bar{c}}{1-r}}_{\mathfrak{m}_K} \right) \right\}.$$

Going back to (6.8) (and using that  $C_{U,K}(x) \leq 1$  for all  $x$ ), we obtain

$$\forall t \in [0, \gamma], \quad \mathbb{E} [f_\theta(\bar{X}_t)] \leq f_\theta(x) (e^{-c_K} + e^{-tM}) + e^{\theta \mathfrak{m}_K} (e^{4t\bar{c}\sigma^2\theta d} - e^{-tM}).$$

If we now assume that the parameters are chosen in such a way that

$$e^{-c_K} + e^{-\gamma M} < 1, \tag{6.12}$$

an induction leads to

$$\sup_{n \geq 0} \mathbb{E}_x [f_\theta(\bar{X}_{n\gamma})] \leq f_\theta(x) + e^{\theta \mathfrak{m}_K} \frac{e^{4\gamma\bar{c}\sigma^2\theta d} - e^{-\gamma M}}{1 - e^{-c_K} - e^{-\gamma M}}. \tag{6.13}$$

From now on, assume that

$$K = d(1 + 2 \log 2) + \frac{4(1-r)}{\theta\gamma\bar{c}} \quad \text{and} \quad M = \frac{4\theta\bar{c}}{1-r}.$$

In this case, denoting by  $c_r$  a positive constant which only depends on  $r$  (and which may change from line to line), we have

$$m_K \leq c_r \frac{d(1 + \sigma^2)\bar{c}}{\underline{c}} \quad \text{and} \quad e^{-c_K} \leq \frac{\theta\gamma\bar{c}}{1 - r},$$

where for the second inequality, we used that  $e^{-x} \leq 1/(2x)$  for  $x \geq 1$  and  $\frac{1-r}{2\theta\gamma\underline{c}} \geq 1$  ( $\theta \leq 1$ ). One also checks that  $\gamma M \leq 1$  since  $\theta \leq 1$ . By the inequality  $e^{-x} \leq 1 - \frac{1}{2}x$  for  $x \in [0, 1]$ , this implies that

$$e^{-\gamma M} \leq 1 - \frac{2\gamma\theta\bar{c}}{1 - r} \implies e^{-c_K} + e^{-\gamma M} \leq 1 - \frac{\gamma\theta\bar{c}}{1 - r} < 1.$$

Plugging into (6.13) and using the inequality  $e^{-x} \geq 1 - x$  for  $x \geq 0$ , this yields

$$\sup_{n \geq 0} \mathbb{E}_x [f_\theta(\bar{X}_{n\gamma})] \leq f_\theta(x) + e^{c_r \frac{\theta d(1+\sigma^2)\bar{c}}{\underline{c}}} \frac{(1-r)(e^{4\gamma\bar{c}\sigma^2\theta d} - 1) + 2\gamma\theta\bar{c}}{\gamma\theta\bar{c}}.$$

To conclude, we need to separate two situations. If  $\gamma\theta\bar{c}d \leq 1$ , then the inequality  $e^x \leq 1 + 4x$  for  $x \in [0, 1]$  leads to

$$\sup_{n \geq 0} \mathbb{E}_x [f_\theta(\bar{X}_{n\gamma})] \leq f_\theta(x) + e^{c_r \frac{\theta d(1+\sigma^2)\bar{c}}{\underline{c}}} (16\sigma^2 d + 2).$$

If  $\theta\gamma\bar{c}d \geq 1$  (so that  $\theta\gamma\bar{c} \geq 1/d$ ), we obtain the following bound:

$$\sup_{n \geq 0} \mathbb{E}_x [f_\theta(\bar{X}_{n\gamma})] \leq f_\theta(x) + e^{c_r \frac{\theta d(1+\sigma^2)\bar{c}}{\underline{c}}} (de^{4\gamma\bar{c}\sigma^2\theta d} + 2) \leq f_\theta(x) + 3de^{c_r \frac{\theta d(1+\sigma^2)\bar{c}}{\underline{c}}}$$

where in the last inequality, we used that  $4\gamma\bar{c} \leq 1 \leq \frac{\underline{c}}{\bar{c}}$ . This concludes the proof.

(iii) Noting that  $\Psi \lesssim_r \bar{\Psi}$ , the first bound is obvious. For the second one, it is enough to note that under  $(\mathbf{H}_r^1)$ ,  $(X_t)_{t \geq 0}$  and  $(\bar{X}_{n\gamma})_{n \geq 0}$  converge in distribution to  $\pi$  and  $\pi^\gamma$  respectively so that with a uniform integrability argument combined with the first bound of (iii), the convergence holds for along functions  $U^p$  for any  $p > 0$ .

## 7 Proof of Proposition 4.5

The idea is to use Jensen inequality to derive controls of the polynomial moments from exponential moments. To this end, we begin with the following lemma:

**Lemma 7.1.** *Let  $V$  denote a non negative random variable which satisfies*

$$\mathbb{E}[e^{\theta V}] < e^a + \rho e^b \quad \text{for positive } \theta, a, \rho \text{ and } b.$$

Then, for any  $p \geq 1$ ,

$$\mathbb{E}[V^p] \leq \theta^{-p} (p - 1 + a + b + \log(2\rho))^p. \tag{7.1}$$

*Proof.* Let  $p \geq 1$  and remark that

$$\mathbb{E}[V^p] = \theta^{-p} \mathbb{E}[\log^p(e^{\theta V})] \leq \theta^{-p} \mathbb{E}[\log^p(e^{p-1+\theta V})].$$

The function  $x \mapsto \log^p(x)$  being concave on  $[e^{p-1}, +\infty)$ , we deduce from the Jensen inequality that

$$\mathbb{E}[V^p] \leq \theta^{-p} (p - 1 + \log(\mathbb{E}[e^{\theta V}]))^p \leq \theta^{-p} (p - 1 + \log(e^a + \rho e^b))^p.$$

The lemma then follows from the following inequality: for  $a, b > 0$  and  $\rho \geq 1$ ,  $\log(e^a + \rho e^b) \leq a + b + \log(2\rho)$  (since  $e^a + \rho e^b \leq 2\rho e^{a+b}$ ).  $\square$

*Proof of Proposition 4.5.* Let us consider separately the continuous and discrete cases. Let us also remark that it is enough to prove the result for  $p \geq 1$  (when  $p \leq 1$ , one can use the bound obtained for  $p = 1$  combined with the Jensen inequality).

(i) Owing to Proposition 4.1, we can apply Lemma 7.1 with  $\theta = \sigma^{-2}$ ,  $V = U(X_t)$ ,  $a = \frac{U(x)}{\sigma^2}$  and

$$(\rho, b) = \begin{cases} \left(1, \frac{1}{\sigma^2} \left(1 + \frac{dL}{2\underline{c}}\right)^{\frac{1}{1-r}}\right) & \text{under } (\mathbf{H}_r^1) \\ \left(\frac{dL}{2\underline{c}}, \frac{1}{\sigma^2} (4^{\frac{1}{1-r}} \vee \frac{d\bar{c}}{\underline{c}})\right) & \text{under } (\mathbf{H}_r^1) \text{ and } (\mathbf{H}_r^2). \end{cases}$$

When only  $(\mathbf{H}_r^1)$  holds, this yields

$$\begin{aligned} \mathbb{E}_x[U^p(X_t)] &\leq \left(\sigma^2(p-1) + U(x) + \left(1 + \frac{dL}{2\underline{c}}\right)^{\frac{1}{1-r}}\right)^p \\ &\leq c_p \left(U^p(x) + \left((1 + \sigma^2) \left(1 + \frac{dL}{2\underline{c}}\right)^{\frac{1}{1-r}}\right)^p\right). \end{aligned}$$

Under  $(\mathbf{H}_r^1)$  and  $(\mathbf{H}_r^2)$ , we get

$$\mathbb{E}_x[U^p(X_t)] \leq \left(\sigma^2(p-1) + \sigma^2 \log\left(\frac{dL}{\underline{c}}\right) + U(x) + 4^{\frac{1}{1-r}} \vee \frac{d\bar{c}}{\underline{c}}\right)^p.$$

Using that  $\log(x) \leq x$  and that  $L \leq \bar{c}$  (since  $U(x) \geq 1$ ), the second bound follows.

(ii) Assume  $(\mathbf{H}_r^1)$ . Owing to Proposition 4.1 (i) applied with  $\theta = 1/(8\sigma^2)$ , we can use Lemma 7.1 with  $V = U(\bar{X}_{n\gamma})$ ,  $a = e^{\theta U(x)}$ ,  $\rho = 1$  and  $b = \theta \left(1 + \frac{8dL}{\underline{c}}\right)^{\frac{1}{1-r}} + 2\theta dL$  (we used that  $\cosh(u) \leq e^u$  for  $u \geq 0$ ). This yields

$$\begin{aligned} \mathbb{E}_x[U^p(\bar{X}_{n\gamma})] &\leq \theta^{-p} \left(p-1 + \theta U(x) + \theta \left(1 + \frac{8dL}{\underline{c}}\right)^{\frac{1}{1-r}} + 2\theta dL\right)^p \\ &\leq \left(\frac{p-1}{\theta} + U(x) + \left(1 + \frac{8dL}{\underline{c}}\right)^{\frac{1}{1-r}} + 2dL\right)^p \\ &\leq \left(U(x) + c_p(1 + \sigma^2) \left(dL + \left(1 + \frac{dL}{\underline{c}}\right)^{\frac{1}{1-r}}\right)\right)^p. \end{aligned}$$

This yields the bound under  $(\mathbf{H}_r^1)$  only. For the bound under  $(\mathbf{H}_r^1)$  and  $(\mathbf{H}_r^2)$ , Proposition 4.2 (ii) applied with  $\theta = 1 \wedge 1/(8\sigma^2)$  allows us to use Lemma 7.1 with  $V = U(\bar{X}_{n\gamma})$ ,  $a = e^{\theta U(x)}$ ,  $\rho = c$  ( $c$  denoting a universal constant) and  $b = c_r \frac{\theta d(1+\sigma^2)\bar{c}}{\underline{c}}$ . This yields

$$\mathbb{E}_x[U^p(\bar{X}_{n\gamma})] \leq \left(\frac{p-1 + \log(2c)}{\theta} + U(x) + c_r \frac{d(1 + \sigma^2)\bar{c}}{\underline{c}}\right)^p,$$

and the last bound easily follows.

## References

- [1] David Alonso-Gutiérrez and Jesús Bastero, *Approaching the Kannan-Lovász-Simonovits and variance conjectures*, Lecture Notes in Mathematics, vol. 2131, Springer, Cham, 2015. MR3308635
- [2] Dominique Bakry, Franck Barthe, Patrick Cattiaux, and Arnaud Guillin, *A simple proof of the Poincaré inequality for a large class of probability measures*, Electronic Communications in Probability **13** (2008), 60–66. MR2386063

- [3] François Bolley and Cédric Villani, *Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities*, Ann. Fac. Sci. Toulouse Math. (6) **14** (2005), no. 3, 331–352. MR2172583
- [4] Patrick Cattiaux, Max Fathi, and Arnaud Guillin, *Self-improvement of the Bakry-Emery criterion for Poincaré inequalities and Wasserstein contraction using variable curvature bounds*, J. Math. Pures Appl. (9) **166** (2022), 1–29. MR4488132
- [5] Arnak S. Dalalyan, *Theoretical guarantees for approximate sampling from smooth and log-concave densities*, J. R. Stat. Soc. Ser. B. Stat. Methodol. **79** (2017), no. 3, 651–676. MR3641401
- [6] Arnak S. Dalalyan and Avetik Karagulyan, *User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient*, Stochastic Process. Appl. **129** (2019), no. 12, 5278–5311. MR4025705
- [7] Arnak S Dalalyan and Avetik Karagulyan, *Penalized langevin dynamics with vanishing penalty for smooth and log-concave targets*, Advances in Neural Information Processing Systems **33** (2020), 17594–17604.
- [8] Arnak S Dalalyan, Avetik Karagulyan, and Lionel Riou-Durand, *Bounding the error of discretized langevin algorithms for non-strongly log-concave targets*, Journal of Machine Learning Research **23** (2022), no. 235, 1–38. MR4577674
- [9] Steffen Dereich and Thomas Müller-Gronbach, *General multilevel adaptations for stochastic approximation algorithms of Robbins-Monro and Polyak-Ruppert type*, Numer. Math. **142** (2019), no. 2, 279–328. MR3941932
- [10] Alain Durmus and Éric Moulines, *High-dimensional Bayesian inference via the unadjusted Langevin algorithm*, Bernoulli **25** (2019), no. 4A, 2854–2882. MR4003567
- [11] Alain Oliviero Durmus and Andreas Eberle, *Asymptotic bias of inexact markov chain monte carlo methods in high dimension*, 2023.
- [12] Maxime Egea and Fabien Panloup, *Multilevel-langevin pathwise average for gibbs approximation*, 2021.
- [13] Noufel Frikha, *Multi-level stochastic approximation algorithms*, Ann. Appl. Probab. **26** (2016), no. 2, 933–985. MR3476630
- [14] Sébastien Gadat, Fabien Panloup, and Clément Pellegrini, *On the cost of bayesian posterior mean strategy for log-concave models*, 2022.
- [15] Michael B. Giles, *Multilevel Monte Carlo path simulation*, Oper. Res. **56** (2008), no. 3, 607–617. MR2436856
- [16] Michael B. Giles, Mateusz B. Majka, Lukasz Szpruch, Sebastian J. Vollmer, and Konstantinos C. Zygalakis, *Multi-level Monte Carlo methods for the approximation of invariant measures of stochastic differential equations*, Stat. Comput. **30** (2020), no. 3, 507–524. MR4065216
- [17] Peter W. Glynn and Chang-Han Rhee, *Exact estimation for Markov chain equilibrium expectations*, J. Appl. Probab. **51A** (2014), no. Celebrating 50 Years of The Applied Probability Trust, 377–389. MR3317370
- [18] Daniel Gruhlke, *Convergence of multilevel mcmc methods on path spaces*, Ph.D. thesis, Universitäts-und Landesbibliothek Bonn, 2014.
- [19] Martin Hairer, *Convergence of Markov processes*, Lecture notes (2010), .
- [20] Stefan Heinrich, *Multilevel monte carlo methods*, Large-Scale Scientific Computing: Third International Conference, LSSC 2001 Sozopol, Bulgaria, June 6–10, 2001 Revised Papers 3, Springer, 2001, pp. 58–67.
- [21] Viet Ha Hoang, Christoph Schwab, and Andrew M. Stuart, *Complexity analysis of accelerated MCMC methods for Bayesian inversion*, Inverse Problems **29** (2013), no. 8, 085010, 37. MR3084684
- [22] Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991. MR1121940
- [23] Ahmed Kebaier, *Statistical romberg extrapolation: a new variance reduction method and applications to option pricing*, The annals of applied probability **15** (2005), no. 4, 2681–2705. MR2187308

- [24] Rafail Khasminskii, *Stochastic stability of differential equations*, second ed., Stochastic Modelling and Applied Probability, vol. 66, Springer, Heidelberg, 2012, With contributions by G. N. Milstein and M. B. Nevelson. MR2894052
- [25] Damien Lambertson and Gilles Pagès, *Recursive computation of the invariant distribution of a diffusion*, Bernoulli **8** (2002), no. 3, 367–405. MR1913112
- [26] Damien Lambertson and Gilles Pagès, *Recursive computation of the invariant distribution of a diffusion: the case of a weakly mean reverting drift*, Stoch. Dyn. **3** (2003), no. 4, 435–451. MR2030742
- [27] Don McLeish, *A general method for debiasing a Monte Carlo estimator*, Monte Carlo Methods Appl. **17** (2011), no. 4, 301–315. MR2890424
- [28] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*, Communications and Control Engineering Series, Springer-Verlag London, Ltd., London, 1993. MR1287609
- [29] Wenlong Mou, Nicolas Flammarion, Martin J. Wainwright, and Peter L. Bartlett, *Improved bounds for discretization of Langevin diffusions: near-optimal rates without convexity*, Bernoulli **28** (2022), no. 3, 1577–1601. MR4411503
- [30] Gilles Pagès and Fabien Panloup, *Weighted multilevel Langevin simulation of invariant measures*, Ann. Appl. Probab. **28** (2018), no. 6, 3358–3417. MR3861816
- [31] Gilles Pagès and Fabien Panloup, *Unadjusted Langevin algorithm with multiplicative noise: total variation and Wasserstein bounds*, Ann. Appl. Probab. **33** (2023), no. 1, 726–779. MR4551562
- [32] Gareth O. Roberts and Richard L. Tweedie, *Exponential convergence of Langevin distributions and their discrete approximations*, Bernoulli **2** (1996), no. 4, 341–363. MR1440273
- [33] Denis Talay, *Second-order discretization schemes of stochastic differential systems for the computation of the invariant law*, Stochastics: An International Journal of Probability and Stochastic Processes **29** (1990), no. 1, 13–36.
- [34] Santosh S. Vempala and Andre Wibisono, *Rapid convergence of the unadjusted Langevin algorithm: isoperimetry suffices*, Geometric aspects of functional analysis, Lecture Notes in Math., vol. 2327, Springer, Cham, [2023] ©2023, pp. 381–438. MR4651229
- [35] Matti Vihola, *Unbiased estimators and multilevel Monte Carlo*, Oper. Res. **66** (2018), no. 2, 448–462. MR3782809

**Acknowledgments.** The present author is deeply grateful to F.Panloup for his numerous suggestions and discussions that improved the quality of this paper. The author also thanks the SIRIC ILIAD Nantes-Angers program supported by the French National Cancer Institute (INCA-DGOS-Inserm 12558 grant) for funding M. Egéa’s Ph.D. thesis.

---

# Electronic Journal of Probability

## Electronic Communications in Probability

---

### Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS<sup>1</sup>)
- Easy interface (EJMS<sup>2</sup>)

### Economical model of EJP-ECP

- Non profit, sponsored by IMS<sup>3</sup>, BS<sup>4</sup>, ProjectEuclid<sup>5</sup>
- Purely electronic

### Help keep the journal free and vigorous

- Donate to the IMS open access fund<sup>6</sup> (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

---

<sup>1</sup>LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

<sup>2</sup>EJMS: Electronic Journal Management System: <https://vtex.lt/services/ejms-peer-review/>

<sup>3</sup>IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

<sup>4</sup>BS: Bernoulli Society <http://www.bernoulli-society.org/>

<sup>5</sup>Project Euclid: <https://projecteuclid.org/>

<sup>6</sup>IMS Open Access Fund: <https://imstat.org/shop/donation/>