Discretisation of Langevin diffusion in the weak log-concave case

Marelys Crespo Navas MARELYS.CRESPO-NAVAS@UNIV-TLSE3.FR
Institut de Mathématiques de Toulouse
Université Toulouse III Paul Sabatier
118 Rte de Narbonne, 31062 Toulouse, France

Abstract

The Euler discretisation of Langevin diffusion, also known as Unadjusted Langevin Algorithm, is commonly used in machine learning for sampling from a given distribution $\mu \propto e^{-U}$. In this paper we investigate a potential $U : \mathbb{R}^d \longrightarrow \mathbb{R}$ which is a weakly convex function and has Lipschitz gradient. We parameterize the weak convexity with the help of the Kurdyka-Łojasiewicz (KL) inequality, that permits to handle a vanishing curvature settings, which is far less restrictive when compared to the simple strongly convex case. We prove that the final horizon of simulation to obtain an ε approximation (in terms of entropy) is of the order $\varepsilon^{-1} d^{1+2(1+r)^2} \mathrm{Poly}(\log(d), \log(\varepsilon^{-1}))$, where the parameter r is involved in the KL inequality and varies between 0 (strongly convex case) and 1 (limiting Laplace situation).

Keywords: Unadjusted Langevin Algorithm, Entropy, Weak convexity, Rate of convergence

1 Introduction

Motivation: This paper is devoted to the study of the Langevin Monte Carlo method for sampling a probability distribution over $\mathbb{R}^d$, that is absolutely continuous with respect to the Lebesgue measure and that may be written as a Gibbs field, for which the density μ is written as

$$\mu(\theta) = \frac{e^{-U(\theta)}}{\mathcal{Z}}, \quad \text{with} \qquad \mathcal{Z} = \int_{\mathbb{R}^d} e^{-U(\theta)} \mathrm{d}\theta, \tag{1}$$

where $U : \mathbb{R}^d \longrightarrow \mathbb{R}$ is a $\mathcal{C}^2$ convex function and ∇U is $L-$Lipschitz. Sampling a measure μ that may be written in the form (1) is a fundamental problem of applied mathematics, as it is involved in Bayesian estimation, Gelman et al. (2004), in high dimensional statistics, Dalalyan and Tsybakov (2012), in machine learning, Andrieu et al. (2003) or in partial differential equations, Lelièvre and Stoltz (2016), among others.

State of the art: To solve the sampling problem raised by (1), many methods use Markov kernels either in discrete or continuous settings as introduced by the physics community in the seminal contributions Metropolis et al. (1953); Hastings (1970) on the

Hastings-Metropolis method, and in Parisi (1981) on the over-damped Langevin diffusion associated to U

$$\mathrm{d}\vartheta_t = -\nabla U(\vartheta_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \quad t \geq 0, \tag{2}$$

where $(B_t)_{t\geq 0}$ is a $d-$dimensional standard Brownian Motion.

Standard results on Markov processes guarantee the strong existence and ergodicity of the continuous time process $(\vartheta_t)_{t\geq 0}$ that converges in many senses (Wasserstein, $L^2(\mu)$, entropy, total variation, ...) towards its invariant distribution μ (see Pavliotis, 2016). In what follows, we will use the standard generator $\mathcal{L}$ defined for any $f \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$ by

$$\mathcal{L}f(\theta) = -\langle \nabla U(\theta), \nabla f(\theta)\rangle + \Delta f(\theta), \tag{3}$$

which is associated to the diffusion (2).

Some probabilistic works are using coupling to obtain exponential mixing with Wasserstein or total variation metrics and we refer, among others, to Meyn and Tweedie (2012); Roberts and Tweedie (1996). Some other approaches are using the functional analysis point of view: they differentiate either the entropy or the L^2 distance between the law of $(\vartheta_t)_{t\geq 0}$ and μ, which involves the Fisher information, and then exploits spectral gaps or Log-Sobolev type inequalities to link this Fisher information with the energy itself. A huge literature exists and it is far impossible to be exhaustive, we simply refer to Bakry et al. (2007) for a link between dynamical repelling properties (Lyapunov) and functional inequalities (Poincaré and Sobolev ones), to Bolley et al. (2012) for a Wasserstein approach instead, and more generally to the book Bakry et al. (2014) and the references therein.

Even though of fundamental interest for the deep understanding of the mixing properties of $(\vartheta_t)_{t\geq 0}$, the previous works leave open the question of deriving some concrete algorithm to approximate μ, while being of major interest in the statistics and machine learning community. In the past ten years, a myriad of papers have studied some discretisation strategies of (2) and among them the most popular approach is certainly the first order Euler-Maruyama scheme, also referred to as the Unadjusted Langevin Algorithm (ULA for short). With a fixed step size $h > 0$, ULA is defined as

$$\theta_{(k+1)h} = \theta_{kh} - h\nabla U(\theta_{kh}) + \sqrt{2h}\xi_{k+1}, \quad k \geq 0, \tag{4}$$

where $(\xi_k)_{k\geq 1}$ are standard Gaussian random variables in $\mathbb{R}^d$, mutually independent and independent of θ_0. One of the first studies Dalalyan (2016) decomposes the distance between the distribution of θ_{kh} and μ into two terms, an approximation one that induces a mandatory choice of a small h, and then a mixing one that involves the ergodicity behaviour of the continuous time diffusion. Regarding this latter term, the strong convexity assumed in Dalalyan (2016) plays a crucial role for the understanding of ULA. To extend its study to the non-strongly log-concave situation, Dalalyan (2016) introduces then a penalized ULA and derives some trade-off to balance all the effects (bias, approximation and ergodicity).

Thereafter, a striking literature appeared, with the aim to understand the effect of the dimension d, to improve the obtained computational costs Durmus and Moulines (2017);

Durmus et al. (2019), to relax some structural hypotheses, and in particular the strongly convex one Dalalyan et al. (2022); Vempala and Wibisono (2019); Chewi et al. (2022); Erdogdu and Hosseinzadeh (2021), which is problematic for machine learning applications, or to finally extend ULA to other frameworks like noisy or online ones Welling and Teh (2011); Dalalyan and Karagulyan (2019); Crespo et al. (2023); Wibisono and Yang (2022). Again, sampling has gained a lot of popularity and it appears to be rather difficult to be exhaustive with this huge literature.

Our framework: In this paper, we will be interested in the ULA while trying to relax the strong convexity assumption. In this view, we introduce a weakly log-concave situation described by the family of Kurdyka-Łojasiewicz inequalities Kurdyka (1998); Lojasiewicz (1963). These inequalities have been intensively used in optimization theory Bolte et al. (2010) and have shown to be efficient for stochastic optimization Gadat and Panloup (2022) or for sampling Gadat et al. (2022). We will assess our convergence results with the help of the relative entropy (or Kullback-Leibler divergence) between the sampled measure and the target one μ. Our framework is therefore close to some extent to the recent contributions Chewi et al. (2022); Erdogdu and Hosseinzadeh (2021).

We consider the probability distribution of θ_{kh}, thanks to the Euler explicit scheme (4) that involves an elliptic Gaussian convolution, for any initial distribution m_0, the law of θ_{kh} is absolutely continuous with respect to the Lebesgue measure and we denote by m_k the associated density, which is indeed infinitely differentiable as soon as $k \geq 1$. We then define the relative entropy of m_k with respect to the target measure μ as follows

$$J_k = \int_{\mathbb{R}^d} \log\left(\frac{m_k(\theta)}{\mu(\theta)}\right) \mathrm{d}m_k(\theta). \tag{5}$$

At each iteration k, J_k measures the divergence between the instantaneous law of the process at time kh and the (presumably) invariant distribution μ: a small value of J_k induces the closeness of m_k to μ. Of course, since we are studying the *Unadjusted* Langevin Algorithm, for a fixed value of h, m_k corresponds to the distribution of a Markov chain with an invariant measure μ_h and $\mu_h \neq \mu$. Consequently, and oppositely to the MALA modification (see, for example Roberts and Rosenthal, 1998), we cannot expect that $J_k \longrightarrow 0$ when $k \longrightarrow \infty$ without any fine tuning of the parameter h, that must be chosen carefully to balance the bias and the ergodicity of the chain.

In what follows, for a given small ε, we will carefully address the choice of the step size h_ε, and of the horizon of simulation K_ε to guarantee an ε smallness of J_k within our framework of weak KL convexity. In particular, we will focus on the impact of d to the computational cost, as well as the one of r involved in $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ (see below).

Our contribution: In this work we prove that for any weakly log-concave measure whose potential has a Lipschitz gradient, the final horizon of simulation to obtain an ε approximation (in terms of entropy) is of order $\varepsilon^{-1} d^{1+2(1+r)^2}$ where the parameter r is involved in the Kurdyka-Łojasiewicz inequality. In most situations this is the state of the art result.

Structure of the paper: In Section 2, the hypotheses and main convergence results are presented and compared with those previously used in the literature. Section 3 is dedicated to proving Theorem 3, where Proposition 5 plays an important role. Section 4 has a more probabilistic approach. In this section, we modify the Weak Log-Sobolev inequality: we extend its validity to a general context and obtain quantitative upper bounds (in terms of the dimension d and the convexity parameter r). Numerical experiments are presented in Section 5 to support our theoretical findings.

2 Assumptions and main result

2.1 Main assumptions on U

We are interested in studying the convergence of the ULA in a weakly convex framework, that is, when U is a convex function but not necessarily strongly convex. In Vempala and Wibisono (2019), the order of convergence of the entropy in the strongly convex case is obtained using the Log-Sobolev and Talagrand inequalities, which are equivalent in this context. However, the Log-Sobolev inequality (LSI for short) generally requires the convexity (see, for example Bobkov, 1999; Bakry et al., 2014) to be reasonably dimension-dependent, and even the strong convexity to be dimension-free. As we said previously, we have chosen to parameterize this lack of strong convexity with the help of the Kurdyka-Łojasiewicz inequality. This hypothesis allows us to understand the influence of the lack of strong convexity and not simply assume a LSI.

For any twice differentiable function V, we denote the spectrum of the Hessian matrix of V as $Sp(\nabla^2 V(\theta))$. Furthermore, if V is convex, we denote

$$\underline{\lambda}_{\nabla^2 V}(\theta) = \inf Sp(\nabla^2 V(\theta)), \quad \forall \theta \in \mathbb{R}^d.$$

Hypothesis $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$: We say that a function $V : \mathbb{R}^d \to \mathbb{R}$ satisfies a $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$-condition if:

a) V is a $\mathcal{C}^2$-function.

b) V is a convex function and $\min_{\theta \in \mathbb{R}^d} V(\theta) = V(\theta^*) > 0$.

c) ∇V is L-Lipschitz.

d) There exist two constants $0 \leq r < 1$ and $\mathfrak{c} > 0$ such that

$$\mathfrak{c} V^{-r}(\theta) \leq \underline{\lambda}_{\nabla^2 V}(\theta), \quad \forall \theta \in \mathbb{R}^d.$$

In almost all the results of this study we assume that U satisfies a $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$-condition, so, let us comment this assumption.

- The hypothesis assumes a lower bound for the smallest eigenvalue of $\nabla^2 V(\theta)$, similarly we could consider the same type of upper bound for the the highest eigenvalue of $\nabla^2 V(\theta)$: $\overline{\lambda}_{\nabla^2 V}(\theta) = \sup Sp(\nabla^2 V(\theta))$, $\theta \in \mathbb{R}^d$, where $\overline{\lambda}_{\nabla^2 V}(\theta) \leq \tilde{c}V^{-q}(\theta)$, $\tilde{c} > 0$ and $0 \leq q \leq r$. This parameterization is considered in Gadat et al. (2022). When $q = r$, a global standard KL inequality is recovered (see Gadat and Panloup, 2022; Bolte et al., 2010). In our study, we chose to keep a simpler case, note that the L-Lipschitz continuity of ∇V implies that particularly $\tilde{c} = L$ and $q = 0$.

- A value $r = 1$ would correspond to the limiting Laplace case, while $r = 0$ represents the strongly convex situation where the curvature of the function is uniformly bounded by $\mathfrak{c}$.

- The function $V(\theta) = (1 + \|\theta\|^2)^p$, for $\theta \in \mathbb{R}^d$ and $p \in (1/2, 1]$ satisfies a $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$-condition with $r = \frac{1-p}{p}$, $\mathfrak{c} = 2p(1 - 2(1-p))$ and $L = 2p$ (see Remark 7 of Gadat et al., 2022).

- Our work is tightly related to Erdogdu and Hosseinzadeh (2021): in its Assumption 1, it is considered that the potential function U is degenerately convex at infinity, which means that there exists a function $\widetilde{U}$ such that for a constant $\epsilon \geq 0$, $\|U - \widetilde{U}\|_\infty \leq \epsilon$ where $\underline{\lambda}_{\nabla^2 \widetilde{U}}(\theta) \geq \kappa(1 + \frac{1}{4}\|\theta\|^2)^{-\tau/2}$ for some $\kappa > 0$ and $\tau \geq 0$. This parameterization have a close relation with that presented in hypothesis $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$. The previous example which satisfies a $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$-condition with $r = \frac{1-p}{p}$ is also degenerated at infinity since $\widetilde{U}(\theta) = U(\theta) = (1 + \|\theta\|^2)^p$ and $\tau = 2(1-p) = \frac{2r}{1+r}$.

  However, in the general case, the hypothesis $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ states a lower bound of the eigenvalues of $\nabla^2 U(\theta)$ by the function $U^r(\theta)$, while in Erdogdu and Hosseinzadeh (2021) it is compared to $\|\theta\|^\tau$. Below, we will compare our results with $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ with those of Erdogdu and Hosseinzadeh (2021) even though their quantitative hypothesis is not strictly equivalent to ours (see details in Section 2.2).

Assumption on $\min U$ and $\arg\min U$: One of the advantages of assuming $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ is that it is possible to obtain upper and lower bounds of $U(\theta)$ through $\min U$ and powers of $\|\theta - \arg\min U\|$, see Appendix A. Therefore, we need to specify the dependence of $\min U$ and $\arg\min U$ with respect to the dimension d in order to later understand the dimension dependence of the convergence result.

Similarly to Crespo et al. (2023), we will assume that the minimizer of the function U is contained in a ball of radius that only depends on d. Furthermore, we consider that $\min U$ is at most of the order d. For this purpose, we introduce $a \lesssim_{uc} b$ ($a \gtrsim_{uc} b$), which means $a \leq cb$ (respectively $a \geq cb$) where c is a universal constant that is a positive constant independent of d.

Hypothesis $\mathcal{H}_{\min}$:

$$\|\arg\min U\| \lesssim_{uc} \sqrt{d} \quad \text{and} \quad \min_{\theta \in \mathbb{R}^d} U(\theta) \lesssim_{uc} d.$$

Assumption $\mathcal{H}_{\min}$ is not restrictive and includes translations as simple as for example $U(\theta) = (1 + \|\theta - 1_d\|^2)^{\frac{1}{1+r}}$ where $1_d = (1, \dots, 1) \in \mathbb{R}^d$.

2.2 Entropic convergence result

We specify the initial distribution m_0, which is chosen by the user. We decided to use a Gaussian distribution which is a standard option and has the advantage of providing simpler computations. Implicitly, to fix m_0, we assume that L, $\mathfrak{c}$ and r are known. We denote by $0_d = (0, \dots, 0) \in \mathbb{R}^d$ and $I_d \in \mathbb{R}^d \times \mathbb{R}^d$ the identity matrix.

Hypothesis $\mathcal{H}^{\mathbf{r}}_{\mathbf{m_0}}(\mathfrak{c}, L)$: A positive constant σ^2 exists such that $m_0 = \mathcal{N}(0_d, \sigma^2 I_d)$. Moreover,

$$\sigma^2 \leq \min\left\{1/(2L), 4/\mathfrak{c}^{\frac{1}{1+r}}\right\}.$$

Conditions $\sigma^2 \leq 1/(2L)$ and $\sigma^2 \leq 4/\mathfrak{c}^{\frac{1}{1+r}}$ will be used in the proof of Proposition 1 and Lemma 15, respectively, to upper bound initial terms determined by m_0. A consequence of $\mathcal{H}^{\mathbf{r}}_{\mathbf{m_0}}(\mathfrak{c}, L)$ is an upper bound of the initial entropy J_0 as shown below.

Upper bound on the initial entropy:

Proposition 1 *We assume $\mathcal{H}^{\mathbf{r}}_{\mathbf{m_0}}(\mathfrak{c}, L)$ and that U satisfies $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ and $\mathcal{H}_{\min}$. Then*

$$J_0 = \int_{\mathbb{R}^d} \log\left(\frac{m_0(\theta)}{\mu(\theta)}\right) \mathrm{d}m_0(\theta) \lesssim_{uc} d(1 + r\log d).$$

The proof of Proposition 1 may be found in Appendix B.1. Although the term J_0 is commonly ignored to describe the order of convergence of J_k, even with a simple choice such as an initial Gaussian distribution, J_0 depends on the dimension d in the strongly convex situation, and also depends on d and r in our settings.

Density of the discrete explicit Euler-Maruyama scheme under $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$: The discrete Euler-Maruyama scheme (4) involves a Gaussian convolution kernel at each iteration, of variance $2h$. Then, as said before, the distribution of θ_{kh} has an infinitely differentiable density m_k with respect to the Lebesgue measure for any $k \geq 1$. The next result states a uniform (over the iterations) upper bound of the density m_k and will be of primary importance for our analysis. Up to our knowledge, such a result is new in numerical probability, and has never been investigated.

Proposition 2 *We assume $\mathcal{H}^{\mathbf{r}}_{\mathbf{m_0}}(\mathfrak{c}, L)$ and that U satisfies $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ and $\mathcal{H}_{\min}$. If $h \lesssim_{uc} \min\{1/L, 1/d\}$, a positive constant A_d exists (that could depend on d) such that*

$$\forall \theta \in \mathbb{R}^d, \quad \forall k \geq 0, \qquad m_k(\theta) \leq A_d e^{-\frac{1}{10}U(\theta)}.$$

Moreover $\log(A_d) \lesssim_{uc} d^{\frac{1}{1-r}}$.

For the sake of readability, the proof of Proposition 2 is deferred to Appendix B.2.

Entropic convergence: Let us present the principal result of our work, which proves a decrease in the relative entropy of m_k towards the correct measure μ, as the number of iterations increases and the step-size h of the scheme decreases.

Theorem 3 *We assume $\mathcal{H}^{\mathbf{r}}_{\mathbf{m_0}}(\mathfrak{c}, L)$ and that U satisfies $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ and $\mathcal{H}_{\min}$. Then*

i) If $r = 0$, we reach $J_k \le \varepsilon$ after

$$k \gtrsim_{uc} \varepsilon^{-1} d \log(d/\varepsilon) \quad \textit{and} \quad h \lesssim_{uc} \min\left\{\frac{1}{L}, \frac{\varepsilon}{dL^2 C_{LSI}}\right\},$$

where $C_{LSI} = C_{LSI}(\mu)$ is the Log-Sobolev constant of μ.

ii) If $r > 0$, we reach $J_k \le \varepsilon$ after

$$k \gtrsim_{uc} \varepsilon^{-1} d^{1+2(1+r)^2} \log^2(A_d/\varepsilon) \log(d/\varepsilon) \quad \textit{and} \quad h \lesssim_{uc} \min\left\{\frac{1}{L}, \frac{\varepsilon}{d^{1+(1+r)^2}\log(A_d/\varepsilon)}\right\}.$$

Let us briefly comment this result. First, our bound obtained in Proposition 2 yields $\log(A_d) \lesssim_{uc} d^{\frac{1}{1-r}}$. Even if this result is worthwhile and new, one could expect a polynomial dependency with d of A_d, even in the limiting case $r = 1$, that is $\log A_d \lesssim_{uc} \log(d)$. We have chosen to leave this important axis of research and of improvement open as it is not the main subject of this paper. Second, when compared to the result stated in Crespo et al. (2023) in the same setting ($\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ and entropic convergence) but that only studies the *continuous time* stochastic Langevin diffusion, we observe that the computational time becomes Poly(d) ε^{-1}, while it was Poly(d) $\log^2(\varepsilon^{-1})$ in Crespo et al. (2023). Again, this exemplifies the price to pay to turn the continuous time diffusion into a tractable numerical discrete algorithm. Thanks to Theorem 3 it follows that $J_k \le \varepsilon$ after approximately

$$k \gtrsim_{uc} \varepsilon^{-1} d^{1+2(1+r)^2} \mathrm{Poly}(\log(d), \log(\varepsilon^{-1}))$$

iterations, where A_d is considered proportional to a polynomial of d.

State of the art comparison: Below we start by comparing our main result (part *ii*) of Theorem 3) with those of Cheng and Bartlett (2018), Durmus et al. (2019) and Erdogdu and Hosseinzadeh (2021) that establishes some results for the ULA in a similar framework. The hypothesis of the previous papers were adapted to the our $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ context where $r \in (0, 1)$ and assuming $\mathcal{H}_{\min}$ and $\mathcal{H}^{\mathbf{r}}_{\mathbf{m_0}}(\mathfrak{c}, L)$. In particular, we pay special attention to the dependency with ε, the dimension d and the parameter r. Table 1 shows a summary of these results while omitting the log-terms $\log(d)$ and $\log(\varepsilon^{-1})$.

• When compared with the bounds obtained in Durmus et al. (2019) (and in Cheng and Bartlett, 2018), we obtain a far better dependency with ε and a slight degradation with the dimension d. When paying a specific attention to the approach of Durmus et al.

	Iteration cost for Entropy $J_k \leq \varepsilon$
Cheng and Bartlett (2018)	$\varepsilon^{-3} d^{1+2(1+r)}$
Durmus et al. (2019)	$\varepsilon^{-2} d^{1+2(1+r)}$
Erdogdu and Hosseinzadeh (2021)	$\varepsilon^{-1} d^{3.5+\frac{2r(2+r)}{1+r}}$
This paper	$\varepsilon^{-1} d^{1+2(1+r)^2}$

Table 1: Comparison of iteration complexity when U satisfies $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ and $\mathcal{H}_{\min}$ and the initial distribution is defined as in $\mathcal{H}^{\mathbf{r}}_{\mathbf{m_0}}(\mathfrak{c}, L)$, admitting that $\log(A_d) \lesssim_{uc} \log(d)$, of the Unadjusted Langevin Algorithm, in the weakly log-concave situation.

(2019, Corollary 7), the improvement from ε^{-2} to ε^{-1} comes from a direct study of the Entropy in our paper, and not from a general catch-all inequality between entropy and the Wasserstein distance. We refer to Villani et al. (2009) for a complete presentation of Wasserstein distance. That being said, the coupling strategy of Durmus et al. (2019) generates a better dependency with respect to d than ours and an obvious computation shows that the bound stated in Durmus et al. (2019) is better than the one of this paper for some very high dimensional settings: $\log(d) \geq (2r(1+r))^{-1} \log(\varepsilon^{-1})$. Oppositely, when ε is chosen smaller than $d^{-2r(1+r)}$, our bound is tighter.

Furthermore, the result of Durmus et al. (2019) does not involve the distribution of ULA at iteration n, but a Césaro average over all iterations of ULA, which is an additional numerical difficulty when compared to simply construct one trajectory and take the value of the procedure attained at the final iteration.

• In order to compare our result with the one that would be obtained in Erdogdu and Hosseinzadeh (2021), let us consider for example $U(\theta) = (1 + \|\theta - 1_d\|^2)^{\frac{1}{1+r}}$, where $r \in (0, 1)$ and adjust their α-dissipativity and ζ-growth of gradient assumption. They establish that there exist two constants $a > 0$ and $b > 0$ such that $\langle \nabla U(\theta), \theta \rangle \geq a\|\theta\|^{\alpha} - b$, for all $\theta \in \mathbb{R}^d$. If we take $\theta = \arg\min U$, then $\|\arg\min U\| \leq (b/a)^{\frac{1}{\alpha}}$, so the dependence on d is not explicit or doesn't exist. Then, we consider that a is a universal constant and $b = d^{\frac{\alpha}{2}} = d^{\frac{1}{1+r}}$ in the particular case been studied. Furthermore, they assume that there exist two positive constant ζ and M such that $\|\nabla U(\theta)\| \leq M(1 + \|\theta\|^{\zeta})$, for all $\theta \in \mathbb{R}^d$. Using the specific function U, we get that $M \propto \sqrt{d}$.

After specifying the dependence on d of the constants b and M, Corollary 4 of Erdogdu and Hosseinzadeh (2021) shows that an ε-error is reached after

$$k \gtrsim_{uc} \varepsilon^{-1} d^{3.5+\frac{2r(2+r)}{1+r}} \mathrm{Poly}(\log(d), \log(\varepsilon^{-1}))$$

iterations. Comparing both results, we can notice that for small values of r, our lower bound of k is better than the previously known one, while the bounds of Erdogdu and Hosseinzadeh (2021) are better for values of r close to 1 (the limiting case being $r \simeq 0.38$).

3 Proof of the main result

In a similar way to Vempala and Wibisono (2019) and with the objective of studying the order of convergence of J_k, for any $k \geq 0$ we define a continuous time process $(\theta_{kh+t})_{0\leq t\leq h}$ in the time interval $[kh,(k+1)h]$, such that the distributions at the ends of the interval coincide with m_k and m_{k+1} respectively. The entropy of the density of said continuous process with respect to the law μ will allow us to obtain the order of convergence of J_k.

We consider one fixed step k of (4), then the output $\theta_{(k+1)h}$ is in fact the output at time h of the stochastic differential equation

$$\mathrm{d}\theta_{kh+t} = -\nabla U(\theta_{kh})\mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \quad t \in [0,h], \tag{6}$$

where $(W_t)_{0\leq t\leq h}$ is a standard Brownian Motion in $\mathbb{R}^d$. Indeed, at time $t \in [0,h]$, the solution of (6) is

$$\theta_{kh+t} = \theta_{kh} - t\nabla U(\theta_{kh}) + \sqrt{2}W_t, \tag{7}$$

so if $t = h$, then $\theta_{kh} - h\nabla U(\theta_{kh}) + \sqrt{2}W_h \overset{d}{=} \theta_{(k+1)h}$.

The process $(\theta_{kh+t}|\theta_{kh})_{0\leq t\leq h}$ that evolves following equation (6) is a continuous time Markov process associated to the generator $\mathcal{L}^k$ of the Markov semigroup, where

$$\mathcal{L}^k f(\theta) = -\langle \nabla U(\theta_{kh}), \nabla f(\theta)\rangle + \Delta f(\theta),$$

for all $f \in \mathcal{C}^2(\mathbb{R}^d)$ and $\theta \in \mathbb{R}^d$.

We denote as $n_{k,t}$ and $n_{t|k}$ the probability distribution with respect to the Lebesgue measure of θ_{kh+t} and $\theta_{kh+t}|\theta_{kh}$ respectively. Using (7), we have that $n_{t|k}(\theta|\eta) = \mathcal{N}(\eta - t\nabla U(\eta), 2t)$ for all $\theta, \eta \in \mathbb{R}^d$, then $n_{t|k}$ is $\mathcal{C}^\infty(\mathbb{R}^d)$, $n_{k,t}$ is also $\mathcal{C}^\infty(\mathbb{R}^d)$ and $\partial_t n_{t|k}$ exists.

We recall that if $\mathcal{L}^{k*}$ denotes the adjoint operator of $\mathcal{L}^k$, the forward Kolmogorov equation yields

$$\partial_t n_{t|k}(\theta|\eta) = \mathcal{L}^{k*} n_{t|k}(\theta|\eta), \quad (\theta,\eta) \in \mathbb{R}^d \times \mathbb{R}^d.$$

This equality is used in the proof of Lemma 4.

Similarly to (5), for any $k \geq 0$ and $t \in [0,h]$, the relative entropy of $n_{k,t}$ with respect to the target measure μ is denoted as

$$J_{k,t} = \int_{\mathbb{R}^d} \log\left(\frac{n_{k,t}(\theta)}{\mu(\theta)}\right) \mathrm{d}n_{k,t}(\theta).$$

Observe that, in particular $n_{k,0} = m_k$ and $n_{k,h} = m_{k+1}$, then $J_{k,0} = J_k$ and $J_{k,h} = J_{k+1}$. We also define the Dirichlet form (or relative Fisher information) of $n_{k,t}$ with respect to μ

as

$$\mathcal{E}_{k,t} = \int_{\mathbb{R}^d} \left\| \nabla \log\left(\frac{n_{k,t}(\theta)}{\mu(\theta)}\right)\right\|_2^2 \mathrm{d}n_{k,t}(\theta) = 4\int_{\mathbb{R}^d} \left\| \nabla \left(\sqrt{\frac{n_{k,t}(\theta)}{\mu(\theta)}}\right)\right\|_2^2 \mathrm{d}\mu(\theta),$$

for any $k \geq 0$ and $t \in [0,h]$. We recall that in the proof of Lemma 3 of Vempala and Wibisono (2019) it was obtained the following link between $\partial_t J_{k,t}$ and $\mathcal{E}_{k,t}$

$$\partial_t J_{k,t} \leq -\frac{3}{4}\mathcal{E}_{k,t} + \mathbb{E}_{m_k n_{t|k}}[\|\nabla U(\theta_{kh+t}) - \nabla U(\theta_{kh})\|^2],$$

where the second term is the discretisation error which in the strongly log-concave context was controlled using Talagrand (or equivalently Log-Sobolev) inequality. Since we are generally not in the strongly log-concave context, we control differently the error term in Lemma 4, the proof may be found in Appendix A. Furthermore, we remark that the procedure used to prove Lemma 4 is equivalent to that of Vempala and Wibisono (2019) but is based on the infinitesimal generator $\mathcal{L}^k$ and Kolmogorov's forward equation.

Lemma 4 *We assume that ∇U is L-Lipschitz and $h \leq 1/4L$, then for any $k \geq 0$ and $t \in [0,h]$*

$$\partial_t J_{k,t} \leq -\frac{1}{2}\mathcal{E}_{k,t} + 5dL^2 t.$$

To construct a differential inequality from the previous result, we have to use a functional inequality that lower bounds the Dirichlet form $\mathcal{E}_{k,t}$ by the entropy $J_{k,t}$. In the strongly log-concave situation ($r = 0$) it is possible to directly apply LSI. While in the weakly log-concave situation ($r > 0$) it is not possible for functions like $U(\theta) = (1 + \|\theta\|^2)^p$, when $1/2 < p < 1$, $\theta \in \mathbb{R}^d$. An alternative is to consider the Weak Log-Sobolev inequality (WLSI) but one of its terms depends on $n_{k,t}/\mu$ being bounded and, in general we do not have such a result. So, we modify the WLSI proof in Section 4 to obtain an inequality that can be applied in our $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ context.

From now on we study the cases $r = 0$ and $r > 0$ separately.

- Case $r = 0$.

  Since μ is a strongly log-concave measure, we follow the strategy used in Vempala and Wibisono (2019), which is based on applying LSI to $f_{k,t}^2 = n_{k,t}/\mu$, see details about functional inequalities in Section 4. Then, there exists a constant $C_{LSI} > 0$ independent on d such that

  $$J_{k,t} \leq C_{LSI}\mathcal{E}_{k,t}, \tag{8}$$

  We combine Lemma 4 and inequality (8) to get that for any $k \geq 0$,

  $$\partial_t J_{k,t} \leq -\frac{1}{2C_{LSI}} J_{k,t} + 5dL^2 h, \quad t \in (0,h),$$

where we used $t \leq h$ in the second term of the right hand side. We recall that $J_{k,0} = J_k$ and $J_{k,h} = J_{k+1}$, for any k, then if we apply Gronwall's lemma to the differential inequality it results that for all $k \geq 0$,

$$J_{k+1} \leq J_k e^{-\frac{h}{2C_{LSI}}} + CdL^2 hC_{LSI} \left[1 - e^{-\frac{h}{2C_{LSI}}}\right],$$

where C is a universal constant. Using a recursive argument and Proposition 1 to upper bound J_0, we get

$$J_k \leq J_0 e^{-\frac{kh}{2C_{LSI}}} + CdL^2 hC_{LSI} \left[1 - e^{-\frac{kh}{2C_{LSI}}}\right] \leq Cd\left[e^{-\frac{kh}{2C_{LSI}}} + L^2 hC_{LSI}\right].$$

Given $\varepsilon > 0$ small, a sufficient condition for $J_k \leq \varepsilon$ is

$$h \lesssim_{uc} \varepsilon (dL^2 C_{LSI})^{-1} \quad \text{and} \quad k \gtrsim_{uc} \varepsilon^{-1} d \log(d/\varepsilon).$$

- Case $r > 0$.

In the weak log-concave case we need to modify WLSI as is shown in the following proposition. Further details and the proof of this functional inequality under a $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$-condition may be found in Section 4.

Proposition 5 *We assume $\mathcal{H}^{\mathbf{r}}_{\mathbf{m_0}}(\mathfrak{c}, L)$ and that the function U satisfies $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ and $\mathcal{H}_{\min}$. If $h \lesssim_{uc} \min\{1/L, 1/d\}$ then there exist a non-increasing and positive function β and two positive constants C and $q > 1$ such that for any $k \geq 0$ and $t \in [0, h]$,*

$$J_{k,t} \leq 16\beta(s)\mathcal{E}_{k,t} + C\left[s + A_d s^{\frac{1}{1+2q}}\right], \quad s > 0.$$

Moreover, the function β could be defined as

$$\beta(s) = \begin{cases} \mathfrak{a} d^{(1+r)^2} \log\left(\frac{1}{s}\right), & 0 < s \leq \frac{e^2}{2e^2+1} \\ \mathfrak{b} d^{(1+r)^2}, & s > \frac{e^2}{2e^2+1} \end{cases},$$

where $\mathfrak{a}$ and $\mathfrak{b}$ are two positive constants.

We are now ready to derive a good differential inequality for $J_{k,t}$ by combining Lemma 4 and Proposition 5, then for all $s > 0$,

$$\partial_t J_{k,t} \leq -\frac{J_{k,t}}{32\beta(s)} + \frac{C}{\beta(s)}\left[s + A_d s^{\frac{1}{1+2q}}\right] + 5dL^2 h.$$

where we used $t \le h$. If we consider that s is constant with respect to t, then we apply Gronwall's lemma and for any $k \ge 0$ we have

$$J_{k+1} \le J_k e^{-\frac{h}{32\beta(s)}} + C\left[s + A_d s^{\frac{1}{1+2q}} + dh\beta(s)\right]\left[1 - e^{-\frac{h}{32\beta(s)}}\right].$$

We recall one more time that $J_{k,0} = J_k$ and $J_{k,h} = J_{k+1}$, for any k. Using a recursive argument and the result $J_0 \lesssim_{uc} d(1 + r\log d)$ proved in Proposition 1, we get

$$\begin{aligned} J_k &\le J_0 e^{-\frac{kh}{32\beta(s)}} + C\left[s + A_d s^{\frac{1}{1+2q}} + dh\beta(s)\right]\left[1 - e^{-\frac{kh}{32\beta(s)}}\right] \\ &\le C\left[d\left(1 + r\log d\right)e^{-\frac{kh}{32\beta(s)}} + s + A_d s^{\frac{1}{1+2q}} + dh\beta(s)\right]. \end{aligned}$$

Given $\varepsilon > 0$ small, a sufficient condition for $J_k \le \varepsilon$ is

$$s = \left(\varepsilon A_d^{-1}\right)^{1+2q}, \qquad h \lesssim_{uc} \frac{\varepsilon}{d^{1+(1+r)^2}\log(A_d/\varepsilon)},$$

and

$$k \gtrsim_{uc} \varepsilon^{-1} d^{1+2(1+r)^2} \log^2(A_d/\varepsilon)\log((1 + r\log(d))d/\varepsilon),$$

which concludes the proof.

4 Modified Weak Log-Sobolev Inequality under $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$

This section is devoted to the study of some functional inequalities starting by Poincaré, Log-Sobolev and Weak Log-Sobolev inequalities, which were previously used. It is precisely in the previous context of a non-strongly log-concave measure where we needed to apply Weak Log-Sobolev inequality to upper bound the relative entropy of a presumable unbounded function. Therefore, we present in Proposition 9 a modification of WLSI that is suitable for a family of unbounded functions. Furthermore, in the $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ setting we obtain a particular result paying special attention to the dependence of the constants involved in the inequality in terms of the parameter r and the dimension d.

4.1 Functional inequalities

Functional inequalities as Poincaré, Log-Sobolev and Weak Log-Sobolev inequalities are defined in a Sobolev space and link the norm of a function or a related quantity to the norm of its derivative. This family of inequalities have a wide range of applications such as in partial differential equations, functional analysis, theory of Sobolev spaces, probability theory and statistics, etc. Particularly in these last two areas, they are powerful tools used to derive concentration inequalities and the convergence of Markov processes, see Ledoux (2001) and Bakry et al. (2014) for a more in-depth study.

In order to understand the behavior of these inequalities in a setting described by an $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ hypothesis, we have to present some definitions and recall some well-known facts.

Let m be a probability measure defined on $\mathbb{R}^d$ and $p \geq 1$, we define as $\left[\int_{\mathbb{R}^d} \|f\|^p \mathrm{d}m\right]^{1/p}$ the p-norm of a real-valued function f when this quantity is finite. The space of functions with finite p-norm is denoted as $L^p(\mathbb{R}^d)$. Moreover, $\mathcal{H}_1(\mathbb{R}^d) = \left\{f : \mathbb{R}^d \to \mathbb{R}; f \in L^2(\mathbb{R}^d), \nabla f \in L^2(\mathbb{R}^d)\right\}$ and $\mathcal{C}^1_b(\mathbb{R}^d)$ is the set of bounded and derivable functions on $\mathbb{R}^d$.

For any function f such that $f \in L^1(\mathbb{R}^d)$, we define the variance of f as

$$Var_m(f) = \int\limits_{\mathbb{R}^d} (f - m[f])^2 \mathrm{d}m$$

where $m[f] = \int_{\mathbb{R}^d} f \mathrm{d}m$ and when $f \in \mathcal{H}_1(\mathbb{R}^d)$, the Dirichlet form of f is defined as

$$\mathcal{E}_m(f) = \int\limits_{\mathbb{R}^d} \|\nabla f\|^2 \mathrm{d}m.$$

We briefly introduce the Poincaré (or spectral gap) inequality which links the variance of f to its Dirichlet form as follows

Definition 6 (Poincaré inequality) *The measure m satisfies a Poincaré inequality if there exists a positive constant $C_P(m)$ such that for any $f \in \mathcal{H}^1(\mathbb{R}^d)$,*

$$C_P(m) Var_m(f) \leq \mathcal{E}_m(f).$$

An important property of log-concave measures is that they satisfy a Poincaré inequality and in this situation a bound on the Poincaré constant may be found in Theorem 1.2 of Bobkov (1999).

Another powerful tool is the Log-Sobolev Inequality (LSI for short) which played an important role in the previous section. This inequality is commonly used to prove the convergence to equilibrium in Markov chain Monte Carlo methods and the study of log-concave distributions.

For any function $f : \mathbb{R}^d \to \mathbb{R}$ such that $\int_{\mathbb{R}^d} f^2 |\log(f)| \mathrm{d}m < \infty$, we define the entropy of f^2 as

$$Ent_m(f^2) = \int\limits_{\mathbb{R}^d} f^2 \log(f^2) \mathrm{d}m - \int\limits_{\mathbb{R}^d} f^2 \mathrm{d}m \log\left(\int\limits_{\mathbb{R}^d} f^2 \mathrm{d}m\right),$$

in this definition $0 \log 0$ is interpreted as 0. We are now able to introduce LSI and we refer to Bakry et al. (2014) for further details.

Definition 7 (Log-Sobolev Inequality) *The measure m satisfies a LSI if there exists a positive constant $C_{LSI}(m)$ such that for any $f \in \mathcal{H}^1(\mathbb{R}^d)$,*

$$Ent_m(f^2) \leq C_{LSI}(m)\mathcal{E}_m(f).$$

In the particular case when m is a strongly log-concave measure, a LSI is verified, see Bakry and Émery (1985), and the Log-Sobolev constant $C_{LSI}(m)$ is independent of the dimension d. Using this result and Lemma 4 a convergence rate bound is obtained in Section 3 under the hypothesis $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ when $r = 0$.

However, when $0 < r < 1$, we are in a weakly log-concave setting and a Weak Log-Sobolev Inequality (WLSI for short) would seem to be suitable to derive a convergence rate bound under a $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ assumption.

Definition 8 (Weak Log-Sobolev Inequality) *The measure m satisfies a WLSI if a non-increasing function $\beta : (0, +\infty) \mapsto \mathbb{R}_+$ exists such that for any $f \in \mathcal{C}^1_b(\mathbb{R}^d)$ and $s > 0$,*

$$Ent_m(f^2) \leq \beta(s)\mathcal{E}_m(f) + s\, Osc^2(f),$$

where $Osc(f) = \sup f - \inf f$.

This functional inequality was introduced in Cattiaux et al. (2007) and in said study they showed that if a probability measure m verifies a particular measure-capacity inequality (defined below) then m satisfies WLSI.

Modified WLSI: In the proof of Theorem 3, we needed to lower bound the Dirichlet form by the entropy of a function f with respect to a probability measure which is weakly log-concave. To apply WLSI the function f must be bounded which, in general, seems not to be true. The following result modifies WLSI and is suitable for a family of unbounded functions.

Proposition 9 (Modified Weak Log-Sobolev Inequality) *We assume that $m(\theta) \propto e^{-V(\theta)}$ where V is a convex function and there exists a positive constant R such that $V(\theta) \gtrsim_{uc} \|\theta\|$, for all $\|\theta\| \geq R$. Then, for any function $f \in \mathcal{H}^1(\mathbb{R}^d)$ such that $f^2(\theta) \leq Ae^{bV(\theta)}$ where $A > 0$ and $0 < b < 1$, there exist two positive constants C and $q > 1$ and a non-increasing function $\beta : (0, +\infty) \mapsto \mathbb{R}_+$ such that for all $s > 0$,*

$$Ent_m(f^2) \leq \beta(s)\mathcal{E}_m(f) + C\left[s + s^{\frac{1}{1+2q}}\right].$$

Moreover, the function β could be defined as

$$\beta(s) = \begin{cases} \frac{\mathfrak{a}}{C_P(m)} \log\left(\frac{1}{s}\right), & 0 < s \leq \frac{e^2}{2e^2+1} \\ \frac{\mathfrak{b}}{C_P(m)}, & s > \frac{e^2}{2e^2+1} \end{cases},$$

where $\mathfrak{a}$ and $\mathfrak{b}$ are two positive constants.

The proof is based in the proof of Proposition 2.2 in Cattiaux et al. (2007) and will be postponed to Section 4.3.

4.2 Preliminary results

A measure - capacity inequality: The measure - capacity inequalities are a class of inequalities introduced by Barthe and Roberto (2003) and allow some functional inequalities to be equivalently described, see Bakry et al. (2014) for a more detailed study. In order to prove Proposition 9 we need to specify the role of a particular measure-capacity inequality which is linked to WLSI. So, let us first define the capacity of a measurable set.

Definition 10 (Capacity) *Let $\mathcal{A}$ and Ω be two measurable sets of $\mathbb{R}^d$ such that $\mathcal{A} \subset \Omega$, the capacity $Cap_m(\mathcal{A}, \Omega)$ is defined as*

$$Cap_m(\mathcal{A}, \Omega) = \inf \left\{ \mathcal{E}_m(f), \mathbb{1}_{\mathcal{A}} \leq f \leq \mathbb{1}_{\Omega} \right\},$$

where f is a Lipschitz function on $\mathbb{R}^d$. If $m(\mathcal{A}) \leq \frac{1}{2}$, then we denote

$$Cap_m(\mathcal{A}) = \inf \left\{ Cap_m(\mathcal{A}, \Omega), \mathcal{A} \subset \Omega, m(\Omega) \leq \frac{1}{2} \right\}.$$

The following lemma presents the capacity-measure inequality which is a sufficient condition to prove WLSI.

Lemma 11 *Assume that the measure m satisfies a Poincaré Inequality of constant $C_P(m)$, then there exists a non-increasing and positive function β such that for any measurable subset $\mathcal{A}$ with $m(\mathcal{A}) \leq 1/2$ and for any $s > 0$,*

$$\frac{m(\mathcal{A}) \log\left(1 + \frac{e^2}{m(\mathcal{A})}\right) - s}{\beta(s)} \leq Cap_m(\mathcal{A}). \tag{9}$$

Moreover, the function β could be defined as

$$\beta(s) = \begin{cases} \frac{\mathfrak{a}}{C_P(m)} \log\left(\frac{1}{s}\right), & 0 < s \leq \frac{e^2}{2e^2+1} \\ \frac{\mathfrak{b}}{C_P(m)}, & s > \frac{e^2}{2e^2+1} \end{cases},$$

where $\mathfrak{a}$ and $\mathfrak{b}$ are two positive constants.

Although the existence of the β function in the previous inequality is known, we include the proof in Appendix C.1 to explicitly find the function.

A technical result to avoid $Osc(f)$ in WLSI: We were interested in establishing an inequality similar to WLSI but for a function f not necessarily bounded, so we had to replace the role of the term $Osc(f)$ in the proof of Proposition 2.2 in Cattiaux et al. (2007). Let's show how they used that f is bounded.

For any $f \in \mathcal{H}^1(\mathbb{R}^d)$, let M be the median of f with respect to the measure m and $\Omega \subset \{f > M\}$. For any non-negative function $H : \mathbb{R}^d \to \mathbb{R}$ such that $\int_{\mathbb{R}^d} e^{H(\theta)}\mathrm{d}m(\theta) \leq e^2 + 1$, the term $\int_\Omega (f - M)^2 H\mathrm{d}m$ is conveniently upper bounded as follows

$$\int_\Omega (f - M)^2 H\mathrm{d}m \leq Osc^2(f) \int_\Omega H\mathrm{d}m,$$

since the function f is assumed bounded.

In order to replace the role of $Osc(f)$, the following proposition shows an upper bound of $\int_\Omega (f - M)^2 H\mathrm{d}m$ by a power of $\int_\Omega H\mathrm{d}m$ in a particular setting.

Proposition 12 *We assume that $m(\theta) \propto e^{-V(\theta)}$ where V is a convex function and there exists a positive constant R such that $V(\theta) \gtrsim_{uc} \|\theta\|$, for all $\|\theta\| \geq R$. Then, for any function f such that $f^2(\theta) \leq Ae^{bV(\theta)}$ where $A > 0$ and $0 < b < 1$, there exist two positive constants $q > 1$ and $C_{A,b,q}$ such that for any $\Omega \subset \mathbb{R}^d$ and any function $H \geq 0$ such that $\int_{\mathbb{R}^d} e^{H(\theta)}\mathrm{d}\mu(\theta) \leq e^2 + 1$, we get*

$$\int_\Omega (f - M)^2 H\mathrm{d}m \leq C_{A,b,q} \left(\int_\Omega H\mathrm{d}m \right)^{1/q}.$$

The previous result is a key point to prove Proposition 9 and its proof is postponed to Appendix C.2.

4.3 Proof of the Modified Weak Log-Sobolev Inequality

The changes in the proof of Proposition 9 with respect to the one of the WLSI (Proposition 2.2 in Cattiaux et al., 2007) are based on Lemma 11 and Proposition 12. The rest of the proof remains quite similar.

Proof [Proposition 9] Let $M > 0$ denotes the median of f with respect to m and let $\Omega_+ = \{f > M\}$, $\Omega_- = \{f < M\}$,

$$F_+ = (f - M)\mathbb{1}_{\Omega_+} \quad \text{and} \quad F_- = (f - M)\mathbb{1}_{\Omega_-}.$$

Using the argument of Lemma 5 in Barthe and Roberto (2003), we obtain that

$$\begin{aligned} Ent_m(f^2) \;&\leq\; \sup\left\{ \int F_+^2 H\mathrm{d}m, H \geq 0, \int e^H \mathrm{d}m \leq e^2 + 1 \right\} \\ &+\; \sup\left\{ \int F_-^2 H\mathrm{d}m, H \geq 0, \int e^H \mathrm{d}m \leq e^2 + 1 \right\}. \end{aligned} \tag{10}$$

Let's study each term separately.

- First term of (10). Let be $c > 0$ and $\rho \in (0,1)$. We introduce for any $i = 0, 1, 2, \ldots$, the sequence of measurable subsets

$$\Omega_i = \left\{F_+^2 \geq c\rho^i\right\},$$

which is increasing and $\bigcup_{i\geq 0} \Omega_i = \Omega_+$ so that, for every function $H \geq 0$,

$$\int F_+^2 H \mathrm{d}m = \int_{\Omega_0} F_+^2 H \mathrm{d}m + \sum_{k>0} \int_{\Omega_i \setminus \Omega_{i-1}} F_+^2 H \mathrm{d}m. \tag{11}$$

  - First term of (11). Thanks to Proposition 12, there exists two constants $C_{A,b,q} > 0$ (which could proportionally change from line to line) and $q > 1$ such that

$$\int_{\Omega_0} F_+^2 H \mathrm{d}m \leq C_{A,b,q} \left(\int_{\Omega_0} H \mathrm{d}m\right)^{\frac{1}{q}}.$$

Lemma 6 of Barthe and Roberto (2003) implies that

$$\sup\left\{\int_{\Omega_0} H \mathrm{d}m, H \geq 0, \int e^H \mathrm{d}m \leq e^2 + 1\right\} = m\left(\Omega_0\right) \log\left(1 + \frac{e^2}{m\left(\Omega_0\right)}\right).$$

Using that $\Omega_0 \subset \left\{f^2 \geq c\right\} \subset \left\{Ae^{bV} \geq c\right\}$, we verify that

$$m(\Omega_0) = \int_{\Omega_0} \mathrm{d}m \leq A \int_{Ae^{bV} \geq c} e^{-bV} \mathrm{d}e^{-(1-b)V} \leq \frac{A'}{c},$$

where A' is proportional to A^2 and the normalizing constant of the measure $e^{-(1-b)V}$. Moreover, the inequality $\log(1+x) \leq \frac{x}{\sqrt{1+x}}$, for all $x \geq 0$, guarantees that

$$m\left(\Omega_0\right) \log\left(1 + \frac{e^2}{m\left(\Omega_0\right)}\right) \leq \frac{e^2}{\sqrt{\frac{ce^2}{A'} + 1}} \leq e\sqrt{\frac{A'}{c}},$$

then

$$\sup\left\{\int_{\Omega_0} F_+^2 H \mathrm{d}m, H \geq 0, \int e^H \mathrm{d}m \leq e^2 + 1\right\} = C_{A,b,q} c^{-\frac{1}{2q}}. \tag{12}$$

– Second term of (11). We have for all $i > 0$, due to the fact that $c\rho^i \leq F_+^2 \leq c\rho^{i-1}$ on $\Omega_i \setminus \Omega_{i-1}$,

$$\int\limits_{\Omega_i \setminus \Omega_{i-1}} F_+^2 H \mathrm{d}m \leq c\rho^{i-1} \int\limits_{\Omega_i \setminus \Omega_{i-1}} H \mathrm{d}m.$$

Then using again Lemma 6 of Barthe and Roberto (2003), for any $i > 0$, we obtain that

$$\begin{aligned} & \sup\left\{ \int\limits_{\Omega_i \setminus \Omega_{i-1}} F_+^2 H \mathrm{d}m, H \geq 0, \int e^H \mathrm{d}m \leq e^2 + 1 \right\} \\ \leq \quad & c\rho^{i-1} m\left(\Omega_i \setminus \Omega_{i-1}\right) \log\left(1 + \frac{e^2}{m\left(\Omega_i \setminus \Omega_{i-1}\right)}\right) \\ \leq \quad & sc\rho^{i-1} + c\rho^{i-1}\beta(s) Cap_m\left(\Omega_i \setminus \Omega_{i-1}\right), \end{aligned}$$

where we used inequality (9) applied to μ in the second step.

Let set for any $i > 0$,

$$g_i = \begin{cases} 1 & \text{on } \Omega_i \\ \frac{F_+ - \sqrt{c\rho^{i+1}}}{\sqrt{c\rho^i} - \sqrt{c\rho^{i+1}}} & \text{on } \Omega_{i+1} \setminus \Omega_i \\ 0 & \text{on } \Omega_{i+1}^c \end{cases},$$

so that we have $\mathbb{1}_{\Omega_i} \leq g_i \leq \mathbb{1}_{\Omega_+}$, recall that $m(\Omega_+) = 1/2$. This implies, using the definition of $Cap_m\left(\Omega_i \setminus \Omega_{i-1}\right)$, that

$$c\rho^{i-1} Cap_m\left(\Omega_i \setminus \Omega_{i-1}\right) \leq \frac{1}{\rho(1-\sqrt{\rho})^2} \int\limits_{\Omega_{i+1} \setminus \Omega_i} \|\nabla F_+\|^2 \mathrm{d}m,$$

then

$$\begin{aligned} & \sup\left\{ \int\limits_{\Omega_i \setminus \Omega_{i-1}} F_+^2 H \mathrm{d}m, H \geq 0, \int e^H \mathrm{d}m \leq e^2 + 1 \right\} \\ \leq \quad & sc\rho^{i-1} + \frac{\beta(s)}{\rho(1-\sqrt{\rho})^2} \int\limits_{\Omega_{i+1} \setminus \Omega_i} \|\nabla F_+\|^2 \mathrm{d}m. \end{aligned} \tag{13}$$

We replace (12) and (13) in (11),

$$\begin{aligned}
&\sup\left\{\int F_+^2 H\mathrm{d}m, H\geq 0, \int e^H \mathrm{d}m \leq e^2+1\right\}\\
\leq\ & \sup\left\{\int_{\Omega_0} F_+^2 H\mathrm{d}m, H\geq 0, \int e^H \mathrm{d}m \leq e^2+1\right\}\\
&+\sum_{i>0}\sup\left\{\int_{\Omega_i\setminus\Omega_{i-1}} F_+^2 H\mathrm{d}m, H\geq 0, \int e^H \mathrm{d}m \leq e^2+1\right\}\\
\leq\ & C_{A,b,q}c^{-\frac{1}{2q}}+\frac{sc}{1-\rho}+\frac{\beta(s)}{\rho(1-\sqrt{\rho})^2}\int\|\nabla F_+\|^2\mathrm{d}m.
\end{aligned}$$

The optimal values of c and ρ are $s^{-\frac{1+2q}{2q}}$ and $1/4$ respectively, so we get

$$\begin{aligned}
&\sup\left\{\int F_+^2 H\mathrm{d}m, H\geq 0, \int e^H \mathrm{d}m \leq e^2+1\right\}\\
\leq\ & \left(C_{A,b,q}+\frac{4}{3}\right)s^{\frac{1}{1+2q}}+16\beta(s)\int_{\Omega_+}\|\nabla f_{k,t}\|^2\mathrm{d}m.
\end{aligned}$$

- Since $0\leq f\leq M$ on Ω_-, the second term of (10) will be treated exactly as in Cattiaux et al. (2007). Then

$$\sup\left\{\int F_-^2 H\mathrm{d}m, H\geq 0, \int e^H \mathrm{d}m \leq e^2+1\right\}\leq\frac{7M^2}{3}s+16\beta(s)\int\|\nabla F_-\|^2\mathrm{d}m.$$

We obtain that

$$Ent_m(f^2)\leq C_{A,b,q}\left(s+s^{\frac{1}{1+2q}}\right)+\beta(s)\int_{\mathbb{R}^d}\|\nabla f\|^2\mathrm{d}m,$$

which concludes the proof. ■

4.4 Modified WLSI under $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c},L)$

Let us recall that the measure m is in fact denoted as $\mu(\theta)=\frac{1}{Z}e^{-U(\theta)}$, where U satisfies a $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c},L)$-condition. In Lemma 4 we obtained the inequality

$$\forall k\geq 0,\quad \forall t\in[0,h],\quad \partial_t J_{k,t}\leq -\frac{1}{2}\mathcal{E}_{k,t}+5dL^2t,$$

and we were interested in an inequality that linked the relative entropy $J_{k,t}$ and the Dirichlet form $\mathcal{E}_{k,t}$ of $f_{k,t} = \sqrt{n_{k,t}/\mu}$ when $0 < r < 1$. Since the function $f_{k,t}$ is not necessarily bounded we can't apply WLSI so we are going to use Proposition 9.

To do this, we need first to lower bound the Poincaré constant $C_P(\mu)$ to determine the function β involved in the measure inequality (9) and obtain an upper bound of $n_{k,t}$ of the type

$$n_{k,t}(\theta) \leq A e^{-(1-b)U(\theta)}$$

which would imply that $f_{k,t}^2 \leq A\mathcal{Z} \exp\{bU(\theta)\}$.

4.4.1 Upper bound on the density of the ULA under $\mathcal{H}_{\mathbf{KL}}^{\mathbf{r}}(\mathfrak{c}, L)$

Proposition 2 states an upper bound of the density m_k which is proportional to $e^{-\frac{1}{10}U(\theta)}$. Using the same procedure but applied to the continuous process

$$\theta_{kh+t} = \theta_{kh} - t\nabla U(\theta_{kh}) + \sqrt{2}W_t,$$

for any $k \geq$ and any fixed $t \in (0, h)$ we prove the following result.

Proposition 13 *We assume $\mathcal{H}_{\mathbf{m_0}}^{\mathbf{r}}(\mathfrak{c}, L)$ and that the function U satisfies $\mathcal{H}_{\mathbf{KL}}^{\mathbf{r}}(\mathfrak{c}, L)$ and $\mathcal{H}_{\min}$. If $h \lesssim_{uc} \min\{1/L, 1/d\}$, a positive constant A_d exists (that could depend on d) such that*

$$\forall \theta \in \mathbb{R}^d, \quad \forall k \geq 0, \quad \forall t \in [0, h], \quad n_{k,t}(\theta) \leq A_d e^{-\frac{1}{10}U(\theta)}.$$

Moreover $\log(A_d) \lesssim_{uc} d^{\frac{1}{1-r}}$.

Performing a similar proof to that of Proposition 12 and using the bound of the $n_{k,t}$, we obtain the following corollary.

Corollary 14 *We assume $\mathcal{H}_{\mathbf{m_0}}^{\mathbf{r}}(\mathfrak{c}, L)$ and that the function U satisfies $\mathcal{H}_{\mathbf{KL}}^{\mathbf{r}}(\mathfrak{c}, L)$ and $\mathcal{H}_{\min}$. If $h \lesssim_{uc} \min\{1/L, 1/d\}$ then there exist two positive constants $q > 1$ and $C \geq 0$ such that for any $\Omega \subset \mathbb{R}^d$ and any function $H \geq 0$ such that $\int_{\mathbb{R}^d} e^{H(\theta)} \mathrm{d}\mu(\theta) \leq e^2 + 1$, we get*

$$\int_{\Omega} H \mathrm{d}n_{k,t} \leq A_d \left(\int_{\Omega} H \mathrm{d}\mu \right)^{1/q}.$$

The proof may be found in Appendix C.3.

4.4.2 Poincaré constant

Since μ is a log-concave measure, a Poincaré inequality is verified with constant $C_P(\mu)$. Using the Bobkov bound of the Poincaré constant in Bobkov (1999) and the procedure of Proposition 3.5 in Crespo et al. (2023), we formulate the following lemma, which proof is deferred to Appendix C.4.

Lemma 15 *We assume* $\mathcal{H}^{\mathbf{r}}_{\mathbf{m_0}}(\mathfrak{c}, L)$ *and that* U *satisfies hypothesis* $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ *and* $\mathcal{H}_{\min}$*, then*

$$C_P(\mu) \gtrsim_{uc} d^{-(1+r)^2}.$$

As a consequence of the lemma above, we recall that the measure-capacity (9) applied to μ is verified and the function β could be defined as

$$\beta(s) = \begin{cases} \mathfrak{a} d^{(1+r)^2} \log\left(\frac{1}{s}\right), & 0 < s \leq \frac{e^2}{2e^2+1} \\ \mathfrak{b} d^{(1+r)^2}, & s > \frac{e^2}{2e^2+1} \end{cases},$$

where $\mathfrak{a}$ and $\mathfrak{b}$ are two positive constants.

5 Numerical experiments

In this section, we study two numerical experiments: an application of the ULA to a Bayesian logistic regression problem and a simulation study in a synthetic situation where we can estimate numerically and verify the entropic convergence stated in Theorem 3.

5.1 Bayesian logistic regression

We are interested in applying ULA to the Bayesian logistic regression problem studied in Durmus et al. (2019) (see also Held and Holmes, 2006; Gramacy and Polson, 2012; Park and Hastie, 2007).

We consider $N \geq 1$ i.i.d. observations $(X_1, Y_1), \ldots, (X_N, Y_N)$ where $X_1, \ldots, X_N$ are d-dimensional input variables and $Y_1, \ldots, Y_N$ are binary output responses. Logistic regression framework amounts to assume the responses to be distributed as Bernoulli random variables such that

$$Y_n \sim \mathrm{Ber}\left(\phi(\theta^\top X_n)\right), \quad n \in \{1, \ldots, N\},$$

where ϕ is the logit function defined by $\phi(x) = (1 + e^{-x})^{-1}$, $x \in \mathbb{R}$ and θ is the parameter of interest. We consider a prior distribution given by the following density with respect to Lebesgue measure on $\mathbb{R}^d$,

$$p_r(\theta) \propto \exp\left\{-a_1 \sum_{i=1}^{d} |\theta^{(i)}| - a_2 \sum_{i=1}^{d} |\theta^{(i)}|^{2/(1+r)}\right\}, \quad \theta = (\theta^{(1)}, \ldots, \theta^{(d)}) \in \mathbb{R}^d,$$

where we particularly consider $a_1 = 0.1$, $a_2 = 0.9$ and $r \in [0, 1)$. The variations of p_r with r enable to address situations from strongly log-concave posterior distributions ($r = 0$) to weakly log-concave ones ($0 < r < 1$). This choice of prior leads to the log-concave posterior distribution of θ which is given by the density

$$p_r\left(\theta | (X_1, Y_1), \ldots, (X_N, Y_N)\right) \propto \exp\left\{-\sum_{n=1}^{N} \ell_n(\theta) - a_1 \sum_{i=1}^{d} |\theta^{(i)}| - a_2 \sum_{i=1}^{d} |\theta^{(i)}|^{2/(1+r)}\right\},$$

where the log-likelihoods are given by

$$\ell_n(\theta) = \log\left(1 + \exp\left((1-2Y_n)\theta^\top X_n\right)\right).$$

We now introduce the potential

$$U_r^N(\theta) = \sum_{n=1}^N \ell_n(\theta) + a_1 \sum_{i=1}^d |\theta^{(i)}| + a_2 \sum_{i=1}^d |\theta^{(i)}|^{2/(1+r)},$$

which verifies a $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$-condition. The associated gradient (sub-differential indeed) defined by $\nabla U_r^N(\theta) = (\partial_1 U_r^N(\theta), \ldots, \partial_d U_r^N(\theta))^\top$ is given by

$$\begin{aligned}\partial_j U_r^N(\theta) \quad &= \quad \sum_{n=1}^N (1-2Y_n)\phi\left((1-2Y_n)\theta^\top X_n\right) X_{n,j} + a_1 \mathrm{sign}(\theta^{(j)}) \\ &\quad + \begin{cases} 2a_2\theta^{(j)}, & \text{if } r = 0 \\ \frac{2a_2}{1+r}|\theta^{(j)}|^{(1-r)/(1+r)}, & \text{if } 0 < r < 1 \end{cases},\end{aligned}$$

for any $j \in \{1, \ldots, d\}$.

Given a step size $h > 0$ and an initial standard deviation factor $\sigma > 0$, we introduce the sequence $(\theta_k)_{k\geq 0}$ defined by the following recursion,

$$\theta_0 = \sigma\xi_0 \quad \text{and} \quad \theta_{k+1} = \theta_k - h\nabla U_r^N(\theta_k) + \sqrt{2h}\xi_{k+1},\ k \geq 0,$$

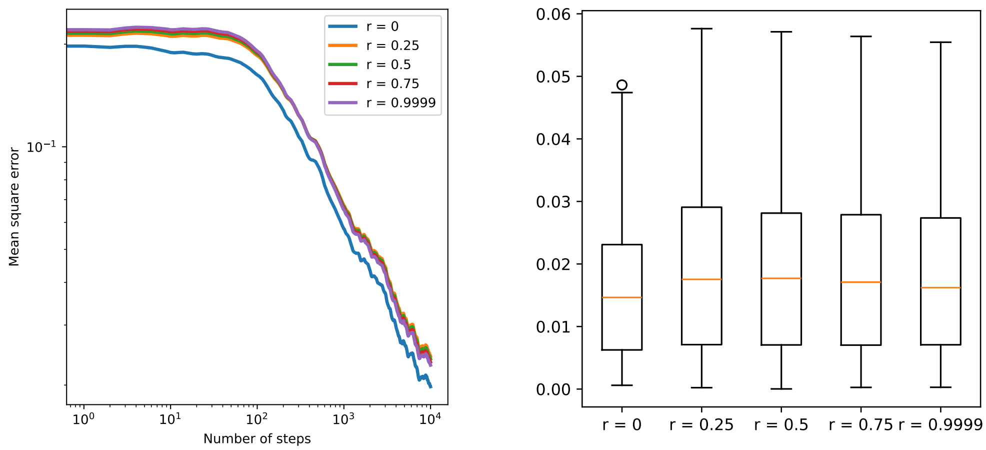
where $(\xi_k)_{k\geq 0}$ is a sequence of i.i.d. standard Gaussian random variables in $\mathbb{R}^d$.

In order to verify the convergence of the ULA, we define the Césaro average of the j-th component of a path of length K starting at K_0 as

$$\widehat{\theta^{(j)}_{K_0,K}} = \frac{1}{K-K_0} \sum_{k=K_0}^{K-1} \theta_k^{(j)}, \qquad j \in \{1, \ldots, d\}.$$

When $K_0 = 0$, we only write $\widehat{\theta^{(j)}_K}$.

We consider two datasets from the UCI repository: the Heart disease dataset with dimensions $N = 214$ and $d = 13$ (see Janosi et al., 1988) and the Musk dataset with dimensions $N = 6598$ and $d = 166$ (see Chapman and Jain, 1994). In both cases, we standardize the data and run 100 independent trajectories of ULA for different values of r and initial standard deviation factor $\sigma = 0.2$. For each dataset and for each r, we compared the results with a *true value* obtained by performing $K = 10^6$ steps with $h = 10^{-4}$ for Heart disease dataset ($K = 10^5$ and $h = 10^{-5}$ for Musk dataset) and taking the Césaro average over the last 10^5 (respectively 10^4) values. The results are presented on Figures 1 and 2.



(a) (b)

Figure 1: (a) Mean square error of the first component estimator at time k, $\widehat{\theta_k^{(1)}}$, for Heart disease dataset where we run 100 trajectories with step size $h = 10^{-3}$ and initial standard deviation factor $\sigma = 0.2$. (b) Boxplot of the absolute error of the estimator at time $K = 10^4$.

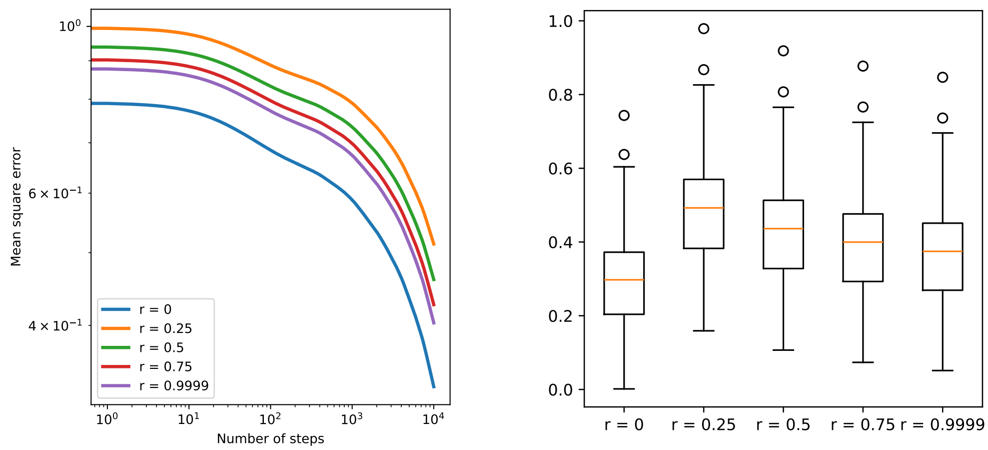
We observe that in both cases the mean square error has a tendency to decrease when the number of ULA steps increases. The convergence of the error for the Musk dataset is slower also due to the large size of the dimension d. In the strongly convex situation the convergence is faster and the curve is apparently different from the rest of the curves. Likewise, the absolute error is much lower in the case $r = 0$. These numerical results show the complexity of the study when dealing with a weakly log-concave distribution compared to a strongly log-concave one.

5.2 Synthetic example and entropy convergence

We are interested in verifying the order of convergence of the entropy obtained in Theorem 3 through simulation. It is important to mention that so far we have not found a numerical study in the related literature in which entropy is estimated.

We consider $0 \leq r < 1$ and we define the potential function $U_r : \mathbb{R}^d \to \mathbb{R}$ as

$$U_r(\theta) = (1 + \|\theta\|^2)^{\frac{1}{1+r}},$$



(a) (b)

Figure 2: (a) Mean square error of the first component estimator at time k, $\widehat{\theta_k^{(1)}}$, for Musk dataset where we run 100 trajectories with step size $h = 10^{-5}$ and initial standard deviation factor $\sigma = 0.2$. (b) Boxplot of the absolute error of the estimator at time $K = 10^4$.

which verifies a $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$-condition. Then its associated gradient is given by $\nabla U_r(\theta) = \frac{2}{1+r} U_r^{-r}(\theta)\theta$. This choice of potential leads to the log-concave density function

$$\mu_r(\theta) = \frac{e^{-U_r(\theta)}}{\mathcal{Z}},$$

where the normalizing constant could be computed using a change to d-dimensional spherical coordinates and, in the case $0 < r < 1$, we approximated the integral for the radial coordinate x,

$$\mathcal{Z} = \int_{\mathbb{R}^d} e^{-U_r(\theta)} \mathrm{d}\theta = \begin{cases} \pi^{d/2} e^{-1}, & \text{if } r = 0, \\ \frac{d\pi^{d/2}}{\Gamma(d/2+1)} \int_0^\infty x^{d-1} e^{-(1+x^2)^{\frac{1}{1+r}}} \mathrm{d}x, & \text{if } 0 < r < 1. \end{cases}$$

Once again, the value of r allows describing from strongly log-concave distributions ($r = 0$) to weakly log-concave ones ($0 < r < 1$).

Given a step size $h > 0$ and an initial standard deviation factor $\sigma > 0$, we introduce the sequence $(\theta_k)_{k \geq 0}$ defined by the following recursion,

$$\theta_0 = \sigma \xi_0 \quad \text{and} \quad \theta_{k+1} = \theta_k - h \nabla U_r(\theta_k) + \sqrt{2h}\xi_{k+1}, \ k \geq 0,$$

where $(\xi_k)_{k \geq 0}$ is a sequence of i.i.d. standard Gaussian random variables in $\mathbb{R}^d$.

We recall that the entropy at any step $k \geq 0$ is defined in (5) as

$$J_k = \int_{\mathbb{R}^d} \log\left(\frac{m_k(\theta)}{\mu_r(\theta)}\right) \mathrm{d}m_k(\theta),$$

where m_k is the density function of θ_k. Then, if we observe $N \geq 1$ independent trajectories $(\theta_k^1)_{k\geq 0}, (\theta_k^2)_{k\geq 0}, \ldots, (\theta_k^N)_{k\geq 0}$, we are able to estimate at any step k, the density function m_k as follows

$$\widehat{m}_k^N(\theta) = \frac{1}{N\delta} \sum_{i=1}^{N} \gamma_1\left(\frac{\theta - \theta_k^i}{\delta}\right),$$

where γ_1 is the standard normal density function and the bandwidth δ is chosen such that it minimizes the mean integrated square error (for a more in-depth study see Scott, 2015; Silverman, 2018).

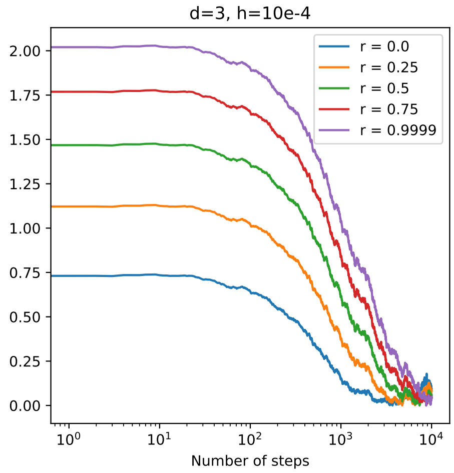
We estimate the entropy using

$$\widehat{J}_k^N = \frac{1}{N} \sum_{i=1}^{N} \log\left[\frac{\widehat{m}_k^N(\theta_k^i)}{\mu_r(\theta_k^i)}\right] = \frac{1}{N} \sum_{i=1}^{N} \left[\log(\widehat{m}_k^N(\theta_k^i)) + U_r(\theta_k^i)\right] + \log(\mathcal{Z}).$$

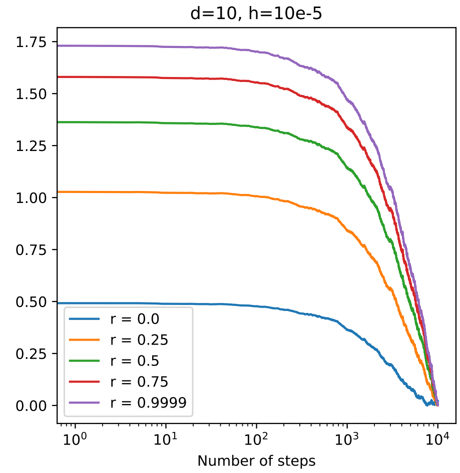
Of course we expect $\widehat{J}_k^N$ to be biased, since for fixed (even if small) h, m_k is the distribution of a Markov chain with invariant measure $\mu_h \neq \mu$. Therefore, in the simulations we will focus on the order of convergence of $\widehat{J}_k^N$ when the number of iterations k increases, which should be approximately proportional to $k^{-1}d^{1+2(1+r)^2}$ when $0 < r < 1$ and $k^{-1}d$ when $r = 0$.

We consider three space dimension $d = 3$, $d = 10$ and $d = 30$ with step sizes $h = 10^{-4}$ $h = 10^{-5}$ and $h = 10^{-7}$ respectively, initial standard deviation factor $\sigma = 0.2$ and fives values of the r parameter. We simulate $N = 100$ trajectories of $(\theta_k)_{k\geq 0}$ and for each k we estimate $\widehat{J}_k^N$. The results are presented on Figure 3.

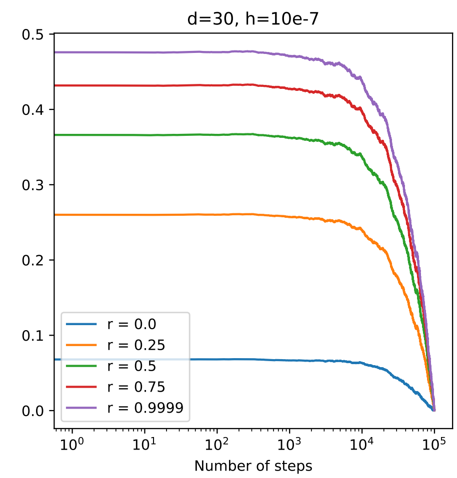
We observe that the entropy estimator decreases as the number of steps increases, with the convergence of the curve apparently being faster for $r = 0$. The curves are arranged according to the value of r, as expected while the increasing dimension d is reflected in the need to use a decreasing value of h and in the slope of the curve.



(a)



(b)



(c)

Figure 3: Convergence of $\widehat{J}_k^N$ where the initial standard deviation factor is $\sigma = 0.2$ and the number of trajectories is $N = 100$.

Acknowledgments and Disclosure of Funding

The author would like to acknowledge Sébastien Gadat and Xavier Gendre for the invaluable guidance and insightful comments all along this work, which is a part of her PhD thesis.

This work is supported from the University Research School EUR-MINT (State support managed by the National Research Agency for Future Investments program bearing the reference ANR-18-EURE-0023)

References

C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50(1–2):5–43, 2003.

D. Bakry and M. Émery. Séminaire de probabilités xix 1983/84, 1985.

D. Bakry, P. Cattiaux, and A. Guillin. Rate of convergence for ergodic continuous Markov processes : Lyapunov versus Poincaré. *Journal of Functional Analysis*, 254:727–759, 2007. URL `https://api.semanticscholar.org/CorpusID:12104713`.

D. Bakry, I. Gentil, M. Ledoux, et al. *Analysis and geometry of Markov diffusion operators*, volume 103. Springer, 2014.

K. Balasubramanian, S. Chewi, M. A. Erdogdu, A. Salim, and S. Zhang. Towards a theory of non-log-concave sampling: first-order stationarity guarantees for Langevin Monte Carlo. In *Conference on Learning Theory*, pages 2896–2923. PMLR, 2022.

F. Barthe and C. Roberto. Sobolev inequalities for probability measures on the real line. *Studia Mathematica*, 159:481–497, 2003.

S. G. Bobkov. Isoperimetric and analytic inequalities for log-concave probability measures. *The Annals of Probability*, 27(4):1903–1921, 1999.

F. Bolley, I. Gentil, and A. Guillin. Convergence to equilibrium in Wasserstein distance for Fokker–Planck equations. *Journal of Functional Analysis*, 263(8):2430–2457, 2012.

J. Bolte, A. Daniilidis, O. Ley, and L. Mazet. Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity. *Transactions of the American Mathematical Society*, 362(6):3319–3363, 2010.

P. Cattiaux, I. Gentil, and A. Guillin. Weak logarithmic sobolev inequalities and entropic convergence. *Probability theory and related fields*, 139:563–603, 2007.

D. Chapman and A. Jain. Musk (Version 2). UCI Machine Learning Repository, 1994. DOI: https://doi.org/10.24432/C51608.

X. Cheng and P. Bartlett. Convergence of Langevin MCMC in KL-divergence. In *Algorithmic Learning Theory*, pages 186–211. PMLR, 2018.

S. Chewi, M. A. Erdogdu, M. Li, R. Shen, and S. Zhang. Analysis of Langevin Monte Carlo from Poincare to log-Sobolev. In *Proceedings of Thirty Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 1–2, 02–05 Jul 2022.

M. Crespo, S. Gadat, and X. Gendre. Stochastic Langevin Monte Carlo for (weakly) log-concave posterior distributions. *arXiv preprint arXiv:2301.03077*, 2023.

A. Dalalyan and A. Tsybakov. Sparse regression learning by aggregation and Langevin Monte-Carlo. *Journal of Computer and System Sciences*, 78(5):1423–1443, 2012.

A. S. Dalalyan. Theoretical Guarantees for Approximate Sampling from Smooth and Log-Concave Densities. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(3):651–676, 2016.

A. S. Dalalyan and A. Karagulyan. User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient. *Stochastic Processes and their Applications*, 129(12): 5278–5311, 2019.

A. S. Dalalyan, A. G. Karagulyan, and L. Riou-Durand. Bounding the error of discretized Langevin algorithms for non-strongly log-concave targets. *J. Mach. Learn. Res.*, 23: 235:1–235:38, 2022.

A. Durmus and É. Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *The Annals of Applied Probability*, 27(3):1551 – 1587, 2017. doi: 10.1214/16-AAP1238. URL `https://doi.org/10.1214/16-AAP1238`.

A. Durmus, S. Majewski, and B. Miasojedow. Analysis of Langevin Monte Carlo via convex optimization. *The Journal of Machine Learning Research*, 20(1):2666–2711, 2019.

M. A. Erdogdu and R. Hosseinzadeh. On the Convergence of Langevin Monte Carlo: The Interplay between Tail Growth and Smoothness. In *Conference on Learning Theory, COLT 2021, 15-19 August 2021, Boulder, Colorado, USA*, volume 134 of *Proceedings of Machine Learning Research*, pages 1776–1822, 2021.

S. Gadat and F. Panloup. Optimal non-asymptotic bound of the Ruppert-Polyak averaging without strong convexity. *Stochastic Processes and their Applications, to appear*, 2022.

S. Gadat, F. Panloup, and C. Pellegrini. On the cost of Bayesian posterior mean strategy for log-concave models. *Preprint*, 2022.

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2nd ed. edition, 2004.

R. B. Gramacy and N. G. Polson. Simulation-based Regularized Logistic Regression. *Bayesian Analysis*, 7(3):567 – 590, 2012. doi: 10.1214/12-BA719. URL `https://doi.org/10.1214/12-BA719`.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

L. Held and C. C. Holmes. Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, 1(1):145 – 168, 2006. doi: 10.1214/06-BA105. URL `https://doi.org/10.1214/06-BA105`.

A. Janosi, W. Steinbrunn, M. Pfisterer, and R. Detrano. Heart Disease. UCI Machine Learning Repository, 1988. DOI: https://doi.org/10.24432/C52P4X.

K. Kurdyka. On gradients of functions definable in o-minimal structures. *Ann. Inst. Fourier (Grenoble)*, 48(3):769–783, 1998. ISSN 0373-0956. URL `http://www.numdam.org/item?id=AIF_1998__48_3_769_0`.

M. Ledoux. The concentration of measure phenomenon. *AMS Surveys and Monographs*, 89, 01 2001.

T. Lelièvre and G. Stoltz. Partial differential equations and stochastic methods in molecular dynamics. *Acta Numerica*, 25:681–880, 2016.

S. Lojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. *Editions du centre National de la Recherche Scientifique, Paris, Les Équations aux Dérivées Partielles*, pages 87–89, 1963.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21 (6):1087–1092, 1953.

S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Springer Science & Business Media, 2012.

L. Miclo. Recuit simulé sur $\mathbb{R}^n$. Étude de l'évolution de l'énergie libre. *Annales de l'I.H.P. Probabilités et statistiques*, 28(2):235–266, 1992. URL `http://www.numdam.org/item/AIHPB_1992__28_2_235_0/`.

G. Parisi. Correlation functions and computer simulations. *Nuclear Physics B*, 180(3): 378–384, 1981.

M. Y. Park and T. Hastie. L 1-regularization path algorithm for generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(4):659–677, 2007.

G. A. Pavliotis. *Stochastic processes and applications*. Springer, 2016.

G. O. Roberts and J. S. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60(1):255–268, 1998.

G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4), 1996.

D. W. Scott. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015.

B. W. Silverman. *Density estimation for statistics and data analysis*. Routledge, 2018.

S. Vempala and A. Wibisono. Rapid convergence of the Unadjusted Langevin Algorithm: Isoperimetry suffices. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL `https://proceedings.neurips.cc/paper_files/paper/2019/file/65a99bb7a3115fdede20da98b08a370f-Paper.pdf`.

C. Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

M. Welling and Y. W. Teh. Bayesian Learning via Stochastic Gradient Langevin Dynamics. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, page 681–688, 2011.

A. Wibisono and K. Y. Yang. Convergence in KL Divergence of the Inexact Langevin Algorithm with Application to Score-based Generative Models. *ArXiv*, abs/2211.01512, 2022.

Appendix A. Some basic properties under a $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$-condition

We recall some important consequences of the $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ hypothesis that implies several relations between the function and the norm of its gradient. These results appear in Lemma 15 of Gadat et al. (2022) (a factor 2 appears in the statement that has already been corrected by Crespo et al., 2023).

Proposition 16 *Assume that a function V satisfies $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$, then for all $\theta \in \mathbb{R}^d$,*

$$\frac{2\mathfrak{c}}{1-r}\left[V^{1-r}(\theta) - \min V^{1-r}\right] \leq \|\nabla V(\theta)\|^2 \leq 2L\left[V(\theta) - \min V\right].$$

Furthermore, it is possible to obtain lower and upper bounds of any function that satisfies $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ by a positive power of the distance to its minimizer.

Proposition 17 *Assume that a function V satisfies $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$, then for all $\theta \in \mathbb{R}^d$,*

$$V^{1+r}(\theta) - \min(V)^{1+r} \geq \frac{(1+r)\mathfrak{c}}{2}\|\theta - \arg\min V\|^2$$

and

$$V(\theta) - \min(V) \leq \frac{L}{2}\|\theta - \arg\min V\|^2.$$

We state the following result which upper bounds the normalizing constant $\mathcal{Z}$ of μ by using Proposition 17. We will use it several times.

Proposition 18 *Assume that U satisfies a $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$-condition, then the normalizing constant of μ verifies the following inequality*

$$\mathcal{Z} \leq 2\left(2\pi/\mathfrak{c}\right)^{\frac{d}{2}} d^{\frac{dr}{2}}.$$

Proof We compute an upper bound of $\mathcal{Z}$ using the lower bound of U induced by Proposition 17,

$$U(\theta) \geq c_r\|\theta - \theta^*\|^{\frac{2}{1+r}}, \tag{14}$$

where $c_r = \left(\frac{(1+r)\mathfrak{c}}{2}\right)^{\frac{1}{1+r}}$ and $\theta^* = \arg\min U$. Then,

$$\mathcal{Z} = \int_{\mathbb{R}^d} e^{-U(\theta)}\mathrm{d}\theta \leq \int_{\mathbb{R}^d} e^{-c_r\|\theta-\theta^*\|^{\frac{2}{1+r}}}\mathrm{d}\theta.$$

Using the well known equality

$$\forall a > 0, \quad \forall \ell > 0, \quad \int_{\mathbb{R}^d} e^{-a|\theta|^\ell}\mathrm{d}\theta = \frac{d\pi^{d/2}\Gamma(d/\ell)}{\ell a^{d/\ell}\Gamma(d/2+1)}, \tag{15}$$

we then deduce that

$$\mathcal{Z} \leq \frac{d(1+r)}{2}\left(\frac{2\pi}{(1+r)\mathfrak{c}}\right)^{\frac{d}{2}}\frac{\Gamma\left(d(1+r)/2\right)}{\Gamma\left(d/2+1\right)} \leq 2\left(\frac{2\pi}{\mathfrak{c}}\right)^{\frac{d}{2}} d^{\frac{dr}{2}},$$

where we used standard relations on the Gamma function. ■

Appendix B. Proofs of the results in Sections 2 and 3

B.1 Proof of Proposition 1

The initial relative entropy is defined as $J_0 = \int_{\mathbb{R}^d} \log\left(\frac{m_0(\theta)}{\mu(\theta)}\right) \mathrm{d}m_0(\theta)$. Replacing $m_0 = \mathcal{N}(0, \sigma^2 I_d)$ and $\mu(\theta) = \frac{1}{\mathcal{Z}}e^{-U(\theta)}$, we observe that

$$J_0 = \log\left(\frac{\mathcal{Z}}{(2\pi\sigma^2)^{\frac{d}{2}}}\right) + \int_{\mathbb{R}^d}\left(U(\theta) - \frac{1}{2\sigma^2}\|\theta\|^2\right)\mathrm{d}m_0(\theta).$$

Let's study each term separately.

- From Proposition 18, we deduce that the first term is upper bounded as follows

$$\log\left(\frac{\mathcal{Z}}{(2\pi\sigma^2)^{\frac{d}{2}}}\right) \leq \log\left(\frac{2d^{\frac{dr}{2}}}{(\mathfrak{c}\sigma^2)^{\frac{d}{2}}}\right) \lesssim_{uc} d(1 + r\log d).$$

- Let be $\theta^* = \arg\min U$, applying Proposition 17, we have that

$$U(\theta) - \frac{1}{2\sigma^2}\|\theta\|^2 \leq \frac{L}{2}\|\theta - \theta^*\|^2 - \frac{1}{2\sigma^2}\|\theta\|^2 + \min U.$$

  If $\sigma^2 < \frac{1}{L}$, then the function $\theta \mapsto \frac{L}{2}\|\theta - \theta^*\|^2 - \frac{1}{2\sigma^2}\|\theta\|^2 + \min U$ attains its global maximum at $\frac{\sigma^2 L}{\sigma^2 L - 1}\theta^*$ and

$$U(\theta) - \frac{1}{2\sigma^2}\|\theta\|^2 \leq \frac{L}{2(1-\sigma^2 L)}\|\theta^*\|^2 + \min U \lesssim_{uc} d,$$

  where we used hypothesis $\mathcal{H}_{\min}$ in the last step. Therefore

$$\int_{\mathbb{R}^d}\left(U(\theta) - \frac{1}{2\sigma^2}\|\theta\|^2\right)\mathrm{d}m_0(\theta) \lesssim_{uc} d.$$

Putting both parts together, the result is proven.

B.2 Proof of Proposition 2

A key part in our analysis is the Proposition 2 which bounds the density m_k proportional to $e^{-\frac{1}{10}U(\theta)}$, where the constant of proportionality is denoted as A_d and could depend on d.
Proof The goal of the proof is to prove the following inequality

$$\exists A_d > 0 \qquad \exists b > 0 : \qquad \forall k \geq 0, \qquad \forall \theta \in \mathbb{R}^d \qquad m_k(\theta) \leq A_d e^{-b[U(\theta) \vee 1]}, \qquad (16)$$

where $x \vee y = \max\{x, y\}$. After obtaining this inequality with $b = 1/10$, the result in the statement is immediate.

We recall the ULA defined in (4) as

$$\theta_{(k+1)h} = \theta_{kh} - h\nabla U(\theta_{kh}) + \sqrt{2h}\xi_{k+1}, \quad \forall k \geq 0.$$

The structure of the proof is divided in three parts. During the first part, we establish several preliminary considerations that will be used later. Secondly, we study the drift of the ULA. Finally, we define two compact sets $\mathcal{B}_1$ and $\mathcal{B}_2$ and we upper bound the density $m_k(\theta)$ by using two different techniques: one for $\theta \in \mathcal{B}_1 \cup \mathcal{B}_2$ and another for $\theta \notin \mathcal{B}_1 \cup \mathcal{B}_2$.

We assume, without loss of generality, that $\min U = U(0)$.

Preliminary considerations: We introduce the key application φ_h, defined by

$$\varphi_h : \theta \longmapsto \theta - h\nabla U(\theta)$$

We shall observe that φ_h is one to one for h small enough.
Injectivity: Consider (x_1, x_2) such that $\varphi_h(x_1) = \varphi_h(x_2)$, we then get

$$x_1 - x_2 = h[\nabla U(x_1) - \nabla U(x_2)].$$

We then take the norm and use the L-smoothness condition on U to derive

$$\|x_1 - x_2\| \leq Lh\|x_1 - x_2\|.$$

We conclude that φ_h is injective when $h < L^{-1}$.
Surjectivity: The proof use the Fenchel-Legendre associated to $\psi_h : x \longmapsto \frac{\|x\|^2}{2} - hU(x)$. When $hL < 1$, we verify that ψ_h is a strictly convex function and we shall define

$$\{\psi_h\}^*(y) = \max_{x \in \mathbb{R}^d} \left\{ \langle y, x \rangle - \psi_h(x) \right\}.$$

The maximum is attained at a x solution of

$$y = x - h\nabla U(x) \Longleftrightarrow \varphi_h(x) = y.$$

We then deduce that φ_h is a one to one mapping, and we define $Z_k = \varphi_h(\theta_{kh})$, so that

$$\theta_{(k+1)h} = Z_k + \sqrt{2h}\xi_{k+1},$$

which results into the convolution decomposition

$$m_{k+1}(\theta) = (\rho_k \star \gamma_{2h})(\theta),$$

where γ_{2h} refers to the density function of a $\mathcal{N}(0_d, 2hI_d)$ random variable and ρ_k is the density function of Z_k. Using the push-forward operator with φ_h, we observe that

$$\rho_k(z) = \frac{m_k(\varphi_h^{-1}(z))}{|\nabla\varphi_h(\varphi_h^{-1}(z))|},$$

where $|B|$ denotes the determinant of the quadratic matrix B. An immediate computation shows that $|\nabla\varphi_h(\varphi_h^{-1}(z))|$ is lower bounded when $h < L^{-1}$ with

$$|\nabla\varphi_h(\varphi_h^{-1}(z))| \geq (1-hL)^d.$$

this inequality being sharp when $U(\theta) = \frac{L\|\theta\|^2}{2}$.

Drift analysis: Our starting point is the L-smoothness property of U

$$U(y) \leq U(x) + \langle y-x, \nabla U(x)\rangle + \frac{L}{2}\|y-x\|^2, \qquad \forall x, y \in \mathbb{R}^d.$$

We apply this inequality with $y = \varphi_h(x)$ and obtain that

$$U(\varphi_h(x)) \leq U(x) - h\left(1 - \frac{Lh}{2}\right)\|\nabla U(x)\|^2.$$

Then, we deduce that

$$U(\varphi_h^{-1}(z)) \geq U(z) + h\left(1 - \frac{Lh}{2}\right)\|\nabla U(\varphi_h^{-1}(z))\|^2.$$

To lower bound $\|\nabla U(\varphi_h^{-1}(z))\|^2$, we use the triangular inequality and the fact that $\|\nabla U\|$ is L-Lipschitz as follows

$$\begin{aligned}\|\nabla U(\varphi_h(x))\| &\leq \|\nabla U(\varphi_h(x)) - \nabla U(x)\| + \|\nabla U(x)\| \\ &\leq L\|\varphi_h(x) - x\| + \|\nabla U(x)\| \\ &\leq (1+hL)\|\nabla U(x)\|.\end{aligned}$$

We rewrite this inequality with $z = \varphi_h(x)$ and deduce that

$$\|\nabla U(\varphi_h^{-1}(z))\|^2 \geq \frac{\|\nabla U(z)\|^2}{(1+Lh)^2}.$$

We finally conclude the key inequality

$$\forall z \in \mathbb{R}^d, \qquad U(\varphi_h^{-1}(z)) \geq U(z) + c_h\|\nabla U(z)\|^2, \tag{17}$$

where $c_h = \frac{h\left(1-\frac{Lh}{2}\right)}{(1+Lh)^2}$. We also state some inequalities that will be useful later on. First the $\mathcal{C}_1^L$ inequality

$$\|\nabla U(\theta - x)\|^2 \geq \frac{1}{2}\|\nabla U(\theta)\|^2 - L^2\|x\|^2, \tag{18}$$

and second the convexity inequality

$$U(\theta - x) \geq U(\theta) - \langle x, \nabla U(\theta)\rangle. \tag{19}$$

Upper bound: We will use all along our bound, a value of h small enough, and in particular $h \leq 1/4L$. In the meantime, we will also use

$$\forall t \in [0, 1/2], \qquad \log(1-t) \geq -2t,$$

which is a standard consequence of the first order Taylor formula for $t \longmapsto \log(1-t)$.

Below, we study two situations according to the size of $\|\theta\|$. For this purpose, we will need to define a specific value for b, that should not be too large. Even though mysterious at the first glance, we will need to choose

$$b = \frac{1}{10},$$

and this choice will be made clearer in what follows.

We also introduce two compact sets. The first one is defined by

$$\mathcal{B}_1 = B\left(0_d, \frac{2\sqrt{2}}{\sqrt{\mathfrak{c}(1+r)}}\right).$$

We then observe that Proposition 17 yields

$$\forall \theta \notin \mathcal{B}_1 \qquad \|z\| \geq \frac{\|\theta\|}{2} \Longrightarrow U(z) \geq 1.$$

The second compact set is defined through the constraint set

$$\mathcal{C} = \left\{\theta \,:\, \frac{e^{-\frac{bh}{8}\|\nabla U(\theta)\|^2}}{[(1-hL)^2(1-h^2L^2/2)]^{d/2}} \geq 1 - Lh\right\}.$$

We shall observe that

$$\begin{aligned}
\theta \in \mathcal{C} &\Longleftrightarrow \frac{bh}{8}\|\nabla U(\theta)\|^2 \leq -(d+1)\log(1-hL) - d/2\log(1-h^2L^2/2) \\
&\Longrightarrow \frac{bh}{8}\|\nabla U(\theta)\|^2 \leq 2(d+1)hL + dh^2L^2/2 \\
&\Longrightarrow \|\nabla U(\theta)\|^2 \leq 32Lb^{-1}(1+hL)d \Longrightarrow \|\theta\| \leq Rd^{\frac{1+r}{2(1-r)}},
\end{aligned}$$

where R is a constant independent of d and was obtained using the hypothesis $\mathcal{H}_{\min}$ and Propositions 16 and 17. We are led to define

$$\mathcal{B}_2 = B\left(0_d, Rd^{\frac{1+r}{2(1-r)}}\right),$$

and we then observe from our construction that

$$\forall \theta \notin \mathcal{B}_2, \qquad \frac{e^{-\frac{bh}{8}\|\nabla U(\theta)\|^2}}{[(1-hL)^2(1-h^2L^2/2)]^{d/2}} \leq 1 - Lh.$$

Case $\theta \in \mathcal{B}_1 \cup \mathcal{B}_2$: This situation is certainly the easiest to consider. Indeed, m_k refers to the density at iteration k of an homogeneous Markov chain. This Markov chain is recurrent (we refer to the standard contributions of Meyn and Tweedie (2012) to verify the mean reverting effect conditions). The Markov chain is also irreducible thanks to the Gaussian noise at each iteration. Hence, the Markov chain is ergodic and converges towards its invariant distribution, that possesses a density μ_h with respect to the Lebesgue measure.

We also know that $\mu_h \longrightarrow \mathcal{Z}^{-1}e^{-U}$ when $h \longrightarrow 0$, which proves that for h small enough μ_h is bounded on compact sets. Using that m_k evolves through a smooth convolution operator with Gaussian kernels, we then deduce that for any k, m_k is a $\mathcal{C}^\infty$ density function over $\mathbb{R}^d$.

Finally, for h small enough, a constant M exists such that

$$\forall \theta \in \mathcal{B}_1 \cup \mathcal{B}_2, \quad \forall k \geq 0, \qquad m_k(\theta) \leq M.$$

U being upper bounded inside the compact as follows

$$Cd^{\frac{1}{1-r}} = \max_{\theta \in \mathcal{B}_1 \cup \mathcal{B}_2} U(\theta),$$

where C is a constant independent of d (and could change proportionally from line to line). We finally deduce that $\forall \theta \in \mathcal{B}_1 \cup \mathcal{B}_2$, $\forall k \geq 0$,

$$m_k(\theta) \leq A_d e^{-b[U(\theta)\vee 1]} \qquad \text{with} \qquad A_d = Me^{b[Cd^{\frac{1}{1-r}}\vee 1]} = e^{Cd^{\frac{1}{1-r}}},$$

where we assumed $\log M \lesssim_{uc} d^{\frac{1}{1-r}}$.

Case $\theta \notin \mathcal{B}_1 \cup \mathcal{B}_2$: We assume that m_k satisfies (16). We decompose m_{k+1} as follows

$$\begin{aligned} m_{k+1}(\theta) &= (\rho_k \star \gamma_{2h})(\theta) \\ &= \int_{\mathbb{R}^d} \gamma_{2h}(x)\rho_k(\theta - x)\mathrm{d}x \\ &= \int_{\|x-\theta\|\geq \frac{\|\theta\|}{2}} \gamma_{2h}(x)\rho_k(\theta - x)\mathrm{d}x + \int_{\|x-\theta\|\leq \frac{\|\theta\|}{2}} \gamma_{2h}(x)\rho_k(\theta - x)\mathrm{d}x. \end{aligned}$$

• Let us first consider x such that $\|\theta - x\| \geq \|\theta\|/2$, we apply the recursive hypothesis to m_k

$$\int_{\|x-\theta\|\geq \frac{\|\theta\|}{2}} \gamma_{2h}(x)\rho_k(\theta - x)\mathrm{d}x \leq \int_{\|x-\theta\|\geq \frac{\|\theta\|}{2}} \gamma_{2h}(x)\frac{A_d e^{-b\left[U(\varphi_h^{-1}(\theta-x))\vee 1\right]}}{(1-hL)^d}\mathrm{d}x.$$

Furthermore, we observe from (17) that $U(\varphi_h^{-1}(\theta - x)) \geq U(\theta - x)$. Using that $\theta \notin \mathcal{B}_1$ and $\|\theta - x\| \geq \|\theta\|/2$, we deduce that $U(\varphi_h^{-1}(\theta - x)) \geq 1$, which entails that the $\vee 1$ may be removed in this case, and leads to

$$\int_{\|x-\theta\|\geq \frac{\|\theta\|}{2}} \gamma_{2h}(x)\rho_k(\theta - x)\mathrm{d}x \leq \int_{\|x-\theta\|\geq \frac{\|\theta\|}{2}} \gamma_{2h}(x)\frac{A_d e^{-bU(\varphi_h^{-1}(\theta-x))}}{(1-hL)^d}\mathrm{d}x.$$

We now use (17) and then (18) and (19) to obtain

$$\begin{aligned}
&\int_{\|x-\theta\|\geq \frac{\|\theta\|}{2}} \gamma_{2h}(x)\rho_k(\theta - x)\mathrm{d}x \\
&\leq \frac{A_d}{(1-hL)^d} \int_{\|x-\theta\|\geq \frac{\|\theta\|}{2}} \gamma_{2h}(x) e^{-b\left[U(x-\theta)+c_h\|\nabla U(x-\theta)\|^2\right]}\mathrm{d}x \\
&\leq \frac{A_d}{(1-hL)^d} \int_{\|x-\theta\|\geq \frac{|\theta|}{2}} \gamma_{2h}(x) e^{-bU(\theta)+b\langle x,\nabla U(\theta)\rangle-\frac{bc_h}{2}\|\nabla U(\theta)\|^2+bc_hL^2\|x\|^2}\mathrm{d}x \\
&\leq \frac{A_d e^{-bU(\theta)-\frac{bc_h}{2}\|\nabla U(\theta)\|^2}}{(1-hL)^d} \int_{\|x-\theta\|\geq \frac{\|\theta\|}{2}} \gamma_{2h}(x) e^{b\langle x,\nabla U(\theta)\rangle+bc_hL^2\|x\|^2}\mathrm{d}x \\
&\leq \frac{A_d e^{-bU(\theta)-\frac{bc_h}{2}\|\nabla U(\theta)\|^2}}{(1-hL)^d} \int_{\|x-\theta\|\geq \frac{\|\theta\|}{2}} \frac{e^{-\|x\|^2(1/4h-bL^2c_h)+b\langle x,\nabla U(\theta)\rangle}}{(4\pi h)^{d/2}}\mathrm{d}x.
\end{aligned}$$

Some almost straightforward computations show that

$$\int_{\|x-\theta\|\geq \frac{\|\theta\|}{2}} \gamma_{2h}(x)\rho_k(\theta - x)\mathrm{d}x \leq \frac{A_d e^{-bU(\theta)-\left(\frac{bc_h}{2}-\frac{b^2h}{1-4hbc_hL^2}\right)\|\nabla U(\theta)\|^2}}{(1-hL)^d(1-4hbc_hL^2)^{d/2}}.$$

Using now $h \leq \frac{1}{4L}$ and $b = \frac{1}{10}$, we then get

$$\int_{\|x-\theta\| \geq \frac{\|\theta\|}{2}} \gamma_{2h}(x)\rho_k(\theta - x)\mathrm{d}x \leq \frac{A_d e^{-bU(\theta) - \frac{bc_h}{4}\|\nabla U(\theta)\|^2}}{[(1-hL)^2(1-h^2L^2/2)]^{d/2}} \leq \frac{A_d e^{-bU(\theta) - \frac{bh}{8}\|\nabla U(\theta)\|^2}}{[(1-hL)^2(1-h^2L^2/2)]^{d/2}}.$$

We are led to use that $\theta \notin \mathcal{B}_2$, which entails

$$\frac{A_d e^{-bU(\theta) - \frac{bh}{8}\|\nabla U(\theta)\|^2}}{[(1-hL)^2(1-h^2L^2/2)]^{d/2}} \leq (1 - Lh)A_d e^{-bU(\theta)}.$$

Consequently, we have shown that when $\theta \notin \mathcal{B}_1 \cup \mathcal{B}_2$

$$\int_{\|x-\theta\| \geq \frac{\|\theta\|}{2}} \gamma_{2h}(x)\rho_k(\theta - x)\mathrm{d}x \leq (1 - Lh)A_d e^{-bU(\theta)}. \tag{20}$$

• We finally consider the complementary set of integration. For this purpose, we compute a rough upper bound of the d dimensional integral and get in our setting

$$\int_{\|x-\theta\| \leq \frac{\|\theta\|}{2}} \gamma_{2h}(x)\rho_k(\theta - x)\mathrm{d}x \leq \frac{A_d}{(1-hL)^d} \int_{\|x\| \geq \frac{\|\theta\|}{2}} \gamma_{2h}(x)\mathrm{d}x \leq \frac{A_d}{(1-hL)^d} 2^{\frac{d}{2}} e^{-\frac{\|\theta\|^2}{32h}}.$$

The growth property of U that is upper bounded as $U(\theta) \leq \min U + L\|\theta\|^2/2$ shows that for $hL \leq 1/4$ and $\theta \notin \mathcal{B}_2$

$$\frac{A_d}{(1-hL)^d} 2^{\frac{d}{2}} e^{-\frac{\|\theta\|^2}{32h}} \leq LhA_d e^{-bU(\theta)} \tag{21}$$

• Gathering the two previous upper bounds (20) and (21) and taking into account that $U(\theta) > 1$ if $\theta \notin \mathcal{B}_1 \cup \mathcal{B}_2$, we therefore deduce that

$$\forall \theta \notin \mathcal{B}_1 \cup \mathcal{B}_2 \qquad m_{k+1}(\theta) \leq A_d e^{-b[U(\theta) \vee 1]}.$$

■

B.3 Proof of Lemma 4

We begin by formulating the following lemma to control the expectation of $\|\nabla U\|^2$ under a change of measure. This result will be useful to prove Lemma 4. Its proof is taken from Chewi et al. (2022) and we include it for the sake of completeness.

Lemma 19 *Assume that ∇U is L-Lipschitz. For any probability measure m, it holds that*

$$\mathbb{E}_m\left[\|\nabla U(\theta)\|^2\right] \leq 4\mathbb{E}_\mu\left[\left\|\nabla\left(\sqrt{\frac{m(\theta)}{\mu(\theta)}}\right)\right\|^2\right] + 2dL.$$

Proof We recall that $\mathcal{L}$ is the infinitesimal generator of the Langevin diffusion in (2), then applying (3) to U, we observe that $\mathcal{L}U = \|\nabla U\|^2 - \Delta U$. We use integration by parts then,

$$\begin{aligned}
\mathbb{E}_m\left[\|\nabla U(\theta)\|^2\right] &= \mathbb{E}_m[\mathcal{L}U(\theta)] + \mathbb{E}_m[\Delta U(\theta)] \leq \int_{\mathbb{R}^d} \mathcal{L}U(\theta)\frac{m(\theta)}{\mu(\theta)}\mathrm{d}\mu(\theta) + dL \\
&= \int \langle \nabla U(\theta), \nabla\left(\frac{m(\theta)}{\mu(\theta)}\right)\rangle \mathrm{d}\mu(\theta) + dL \\
&\leq 2\int \langle \sqrt{\frac{m(\theta)}{\mu(\theta)}}\nabla U(\theta), \nabla\left(\sqrt{\frac{m(\theta)}{\mu(\theta)}}\right)\rangle \mathrm{d}\mu(\theta) + dL \\
&\leq \frac{1}{2}\mathbb{E}_m\left[\|\nabla U(\theta)\|^2\right] + 2\mathbb{E}_\mu\left[\left\|\nabla\left(\sqrt{\frac{m(\theta)}{\mu(\theta)}}\right)\right\|^2\right] + dL.
\end{aligned}$$

Rearrange this inequality to obtain the desired result. ■

Lemma 4 recalls the link obtained in Vempala and Wibisono (2019) between the derivative of the relative entropy of $n_{k,t}$ with respect to μ and its Dirichlet form. However, we use Lemma 19 to control the discretization error as it appears in Balasubramanian et al. (2022).

Proof [Lemma 4] Let k be fixed, the existence of $\partial_t J_{k,t}$ is due to Proposition 3 and hypothesis 3 of Miclo (1992). Moreover

$$\partial_t J_{k,t} = \int_{\mathbb{R}^d} \partial_t\left[\log\left(\frac{n_{k,t}(\theta)}{\mu(\theta)}\right) n_{k,t}(\theta)\right]\mathrm{d}\theta.$$

Now, we follow an argument equivalent to the one used in Lemma 3 of Vempala and Wibisono (2019). This argument is based on the infinitesimal generator $\mathcal{L}^k$ and

Kolmogorov's forward equation. The time derivative of $J_{k,t}$ satisfies that

$$\begin{aligned}
\partial_t J_{k,t} &= \int_{\mathbb{R}^d} \partial_t \left[\log\left(\frac{n_{k,t}(\theta)}{\mu(\theta)}\right) n_{k,t}(\theta)\right] \mathrm{d}\theta \\
&= \int_{\mathbb{R}^d} \partial_t n_{k,t}(\theta)\mathrm{d}\theta + \int_{\mathbb{R}^d} \log\left(\frac{n_{k,t}(\theta)}{\mu(\theta)}\right) \partial_t n_{k,t}(\theta)\mathrm{d}\theta \\
&= \int_{\mathbb{R}^d} \log\left(\frac{n_{k,t}(\theta)}{\mu(\theta)}\right) \partial_t n_{k,t}(\theta)\mathrm{d}\theta,
\end{aligned}$$

where it's easy to check that the first term vanishes

$$\int_{\mathbb{R}^d} \partial_t n_{k,t}(\theta)\mathrm{d}\theta = \partial_t \left[\int_{\mathbb{R}^d} n_{k,t}(\theta)\mathrm{d}\theta\right] = \partial_t[1] = 0.$$

We recall that $n_{t|k}$ and m_k are the distributions of θ_{kh+t} conditional to θ_{kh} and θ_{kh}, respectively, then we replace $n_{k,t}(\theta) = \int_{\mathbb{R}^d} n_{t|k}(\theta|\eta)\mathrm{d}m_k(\eta)$ in the equation for $\partial_t J_{k,t}$

$$\begin{aligned}
\partial_t J_{k,t} &= \int_{\mathbb{R}^d} \log\left(\frac{n_{k,t}(\theta)}{\mu(\theta)}\right) \partial_t \left[\int_{\mathbb{R}^d} n_{t|k}(\theta|\eta)\mathrm{d}m_k(\eta)\right] \mathrm{d}\theta \\
&= \int_{\mathbb{R}^d\times\mathbb{R}^d} \log\left(\frac{n_{k,t}(\theta)}{\mu(\theta)}\right) \partial_t n_{t|k}(\theta|\eta)\mathrm{d}m_k(\eta)\mathrm{d}\theta \\
&= \int_{\mathbb{R}^d\times\mathbb{R}^d} \log\left(\frac{n_{k,t}(\theta)}{\mu(\theta)}\right) \partial_t n_{t|k}(\theta|\eta)\mathrm{d}\theta\mathrm{d}m_k(\eta) \\
&= \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{L}^k \log\left(\frac{n_{k,t}(\theta)}{\mu(\theta)}\right) \mathrm{d}n_{t|k}(\theta|\eta)\mathrm{d}m_k(\eta),
\end{aligned}$$

where we used Fubini's theorem and Kolmogorov's forward equation in the last two steps, respectively.

We introduce the infinitesimal generator $\mathcal{L}$ given in (3), to obtain the Fisher information $\mathcal{E}_{k,t}$

$$
\begin{aligned}
\partial_t J_{k,t} &= \int_{\mathbb{R}^d} \mathcal{L}\log\left(\frac{n_{k,t}(\theta)}{\mu(\theta)}\right)\mathrm{d}n_{k,t}(\theta) + \int_{\mathbb{R}^d\times\mathbb{R}^d}\left(\mathcal{L}^k-\mathcal{L}\right)\log\left(\frac{n_{k,t}(\theta)}{\mu(\theta)}\right)\mathrm{d}n_{t|k}(\theta|\eta)\mathrm{d}m_k(\eta)\\
&= \int_{\mathbb{R}^d} \mathcal{L}\log\left(\frac{n_{k,t}(\theta)}{\mu(\theta)}\right)\mathrm{d}n_{k,t}(\theta) + \int_{\mathbb{R}^d\times\mathbb{R}^d}\left\langle\nabla U(\theta)-\nabla U(\eta),\nabla\log\left(\frac{n_{k,t}(\theta)}{\mu(\theta)}\right)\right\rangle\mathrm{d}n_{t|k}(\theta|\eta)\mathrm{d}m_k(\eta)\\
&= -\mathcal{E}_{k,t} + \int_{\mathbb{R}^d\times\mathbb{R}^d}\left\langle\nabla U(\theta)-\nabla U(\eta),\nabla\log\left(\frac{n_{k,t}(\theta)}{\mu(\theta)}\right)\right\rangle\mathrm{d}n_{t|k}(\theta|\eta)\mathrm{d}m_k(\eta). \qquad (22)
\end{aligned}
$$

Let's study the second term. Using Young's inequality $\langle a,b\rangle \leq \frac{\epsilon}{2}\|a\|^2+\frac{1}{2\epsilon}\|b\|^2$, where $\epsilon>0$ and will be fixed later on, we observe that

$$
\begin{aligned}
&\int_{\mathbb{R}^d\times\mathbb{R}^d}\left\langle\nabla U(\theta)-\nabla U(\eta),\nabla\log\left(\frac{n_{k,t}(\theta)}{\mu(\theta)}\right)\right\rangle\mathrm{d}n_{t|k}(\theta|\eta)\mathrm{d}m_k(\eta)\\
\leq\ & \frac{\epsilon}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}\|\nabla U(\theta)-\nabla U(\eta)\|^2\,\mathrm{d}n_{t|k}(\theta|\eta)\mathrm{d}m_k(\eta)+\frac{1}{2\epsilon}\int_{\mathbb{R}^d}\left\|\nabla\log\left(\frac{n_{k,t}(\theta)}{\mu(\theta)}\right)\right\|^2\mathrm{d}n_{k,t}(\theta)\\
\leq\ & \frac{L^2\epsilon}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}\|\theta-\eta\|^2\,\mathrm{d}n_{t|k}(\theta|\eta)\mathrm{d}m_k(\eta)+\frac{1}{2\epsilon}\mathcal{E}_{k,t}, \qquad (23)
\end{aligned}
$$

where we used that ∇U is $L-$Lipschitz at the end.

Note that the error term can be written as follows

$$
\int_{\mathbb{R}^d\times\mathbb{R}^d}\|\theta-\eta\|^2\,\mathrm{d}n_{t|k}(\theta|\eta)\mathrm{d}m_k(\eta)=\mathbb{E}_{m_k n_{t|k}}\left[\|\theta_{kh+t}-\theta_{kh}\|^2\right].
$$

Moreover, the solution of (6) at time t is such that $\theta_{kh+t}-\theta_{kh}\overset{d}{=}-t\nabla U(\theta_{kh})+\sqrt{2t}\xi$ where ξ is a standard Gaussian in $\mathbb{R}^d$ independent of θ_{kh}. If γ_1 denotes the standard Gaussian density function, then the error term is

$$
\begin{aligned}
\mathbb{E}_{m_k n_{t|k}}\left[\|\theta_{kh+t}-\theta_{kh}\|^2\right] &= \mathbb{E}_{m_k,\gamma_1}\left[\left\|-t\nabla U(\theta_{kh})+\sqrt{2t}\xi\right\|^2\right]\\
&= t^2\mathbb{E}_{m_k}\left[\|\nabla U(\theta_{kh})\|^2\right]+2dt. \qquad (24)
\end{aligned}
$$

Next, since ∇U is L-Lipschitz,

$$
\begin{aligned}
\|\nabla U(\theta_{kh})\| &\leq \|\nabla U(\theta_{kh+t})\|+L\|\theta_{kh+t}-\theta_{kh}\|\\
&\leq \|\nabla U(\theta_{kh+t})\|+Lt\|\nabla U(\theta_{kh})\|+\sqrt{2t}L\|\xi\|,
\end{aligned}
$$

for $h < 1/L$, we can rearrange this inequality to obtain an upper bound for $\|\nabla U(\theta_{kh})\|$

$$\|\nabla U(\theta_{kh})\| \leq \frac{1}{1-Lt}\|\nabla U(\theta_{kh+t})\| + \frac{\sqrt{2t}L}{1-Lt}\|\xi\|.$$

Taking squares and expectations above, then using $(a+b)^2 \leq 2(a^2+b^2)$, it results

$$\begin{aligned} \mathbb{E}_{m_k}\left[\|\nabla U(\theta_{kh})\|^2\right] &\leq \frac{2}{(1-Lt)^2}\mathbb{E}_{n_{k,t}}\left[\|\nabla U(\theta_{kh+t})\|^2\right] + \frac{4dL^2 t}{(1-Lt)^2} \\ &\leq \frac{2}{(1-Lt)^2}\mathcal{E}_{k,t} + \frac{4dL(1+Lt)}{(1-Lt)^2}, \end{aligned} \tag{25}$$

where we used Lemma 19 with probability measure $m = n_{k,t}$.

Plugging (23), (24) and (25) into (22), we observe that

$$\partial_t J_{k,t} \leq -\left(1 - \frac{1}{2\epsilon} - \frac{\epsilon L^2 t^2}{(1-Lt)^2}\right)\mathcal{E}_{k,t} + \epsilon d L^2 t \frac{(1+3L^2t^2)}{(1-Lt)^2}.$$

If we assume $h \leq \frac{1}{4L}$ and optimize ϵ such that $1 - \frac{1}{2\epsilon} - \frac{\epsilon L^2 t^2}{(1-Lt)^2} \geq \frac{1}{2}$, we get $\epsilon = \frac{\sqrt{18}}{9}$ and finally conclude that

$$\partial_t J_{k,t} \leq -\frac{1}{2}\mathcal{E}_{k,t} + 5dL^2 t.$$

■

Appendix C. Proofs of the results in Section 4

C.1 Proof of Lemma 11

Poincaré inequality implies the capacity inequality

$$\frac{m(\mathcal{A})C_P(m)}{2} \leq Cap_m(\mathcal{A}),$$

(see Proposition 8.3.1 of Bakry et al., 2014). Then, we obtain inequality (9) from the previous result if we find a positive function β such that

$$\frac{2}{C_P(m)}\left[\log\left(1 + \frac{e^2}{m(\mathcal{A})}\right) - \frac{s}{m(\mathcal{A})}\right] \leq \beta(s),$$

where $0 < m(\mathcal{A}) \leq 1/2$.

Let us fix $s > 0$ small and define the function $g_s(x) = \log\left(1 + \frac{e^2}{x}\right) - \frac{s}{x}$, for $x \in (0, 1/2]$. Using that g_s reaches its maximum at $x_{\max} = \frac{se^2}{e^2 - s}$, then we have

$$g_s(x) \leq g_s = \begin{cases} \log\left(\frac{1}{s}\right) + \frac{s}{e^2} + 1, & 0 < s \leq \frac{e^2}{2e^2+1} \\ \log(1+2e^2) - 2s, & s > \frac{e^2}{2e^2+1} \end{cases}.$$

After some simplifications, we define β as in the statement.

C.2 Proof of Proposition 12

On $\Omega = \{f > M\}$, we simply upper bound $f - M$ by f, then

$$\int_\Omega (f(\theta) - M)^2 H(\theta) \mathrm{d}m(\theta) \leq \int_\Omega f^2(\theta) H(\theta) \mathrm{d}m(\theta).$$

We apply Holder's inequality where $p, q > 1$, $1/p + 1/q = 1$ and could be fixed later on

$$\int_\Omega f^2(\theta) H(\theta) \mathrm{d}m(\theta) \leq \left[\int_\Omega H(\theta) \mathrm{d}m(\theta)\right]^{\frac{1}{q}} \left[\int_\Omega f^{2p}(\theta) H(\theta) \mathrm{d}m(\theta)\right]^{\frac{1}{p}}.$$

Let's study the second integral. Consider another constant $q' > 1$ to fix later on. Since $\theta \leq q' e^{\frac{1}{q'}\theta - 1}$ (or equivalently $1 + x \leq e^x$ where $x = \frac{1}{q'}\theta - 1$), then

$$\begin{aligned}
\int_\Omega f^{2p}(\theta) H(\theta) \mathrm{d}m(\theta) &\leq \frac{q'}{e} \int_\Omega f^{2p}(\theta) e^{\frac{1}{q'} H(\theta)} \mathrm{d}m(\theta) \\
&\leq \frac{q'}{e} \left[\int_\Omega e^{H(\theta)} \mathrm{d}m(\theta)\right]^{\frac{1}{q'}} \left[\int_\Omega f^{2p'p}(\theta) \mathrm{d}m(\theta)\right]^{\frac{1}{p'}} \\
&\leq \frac{q'}{e} \left(e^2 + 1\right)^{\frac{1}{q'}} \left[\int_\Omega f^{2p'p}(\theta) \mathrm{d}m(\theta)\right]^{\frac{1}{p'}},
\end{aligned}$$

where we used one more time Holder's inequality with constants $p', q' > 1$ such that $1/p' + 1/q' = 1$ and the hypothesis $\int_{\mathbb{R}^d} e^{H(\theta)} \mathrm{d}\mu(\theta) \leq e^2 + 1$.

We have proven that

$$\int_\Omega f^2(\theta) H(\theta) \mathrm{d}m(\theta) \leq \left(\frac{q'}{e} \left(e^2 + 1\right)^{\frac{1}{q'}}\right)^{\frac{1}{p}} \left[\int_\Omega H(\theta) \mathrm{d}m(\theta)\right]^{\frac{1}{q}} \left[\int_\Omega f^{2p'p}(\theta) \mathrm{d}m(\theta)\right]^{\frac{1}{p'p}}. \quad (26)$$

Let be $m(\theta) = C e^{-V(\theta)}$ and $f^2(\theta) \leq A e^{bV(\theta)}$, then the expression inside the second integral in (26) is

$$f^{2p'p}(\theta) m(\theta) \leq A^{p'p} C e^{-(1 - bp'p) V(\theta)}.$$

If we choose $p > 1$ and $p' > 1$ such that $p'p < \frac{1}{b}$ then the exponent is negative and the integral converges since $V(\theta) \gtrsim_{uc} \|\theta\|$ when $\|\theta\|$ is large.

C.3 Proof of Corollary 14

Let k and t be fixed. We change of measure and then we apply Holder's inequality two times as in the proof of Proposition 12, see Appendix C.2 where $m = \mu$ and $f^2 = n_{k,t}/\mu$, then

$$\begin{aligned}\int_{\Omega} H(\theta)\mathrm{d}n_{k,t}(\theta) &= \mathcal{Z}\int_{\Omega} H(\theta)e^{U(\theta)}n_{k,t}(\theta)\mathrm{d}\mu(\theta)\\ &\leq \mathcal{Z}^{1-\frac{1}{p'p}}\left(\frac{q'}{e}\left(e^2+1\right)^{\frac{1}{q'}}\right)^{\frac{1}{p}}\left[\int_{\Omega} H(\theta)\mathrm{d}\mu(\theta)\right]^{\frac{1}{q}}\left[\int_{\Omega} e^{(p'p-1)U(\theta)}\ n_{k,t}^{p'p}(\theta)\ \mathrm{d}\theta\right]^{\frac{1}{p'p}}.\end{aligned}$$

where $p, q, p', q' > 1$, $1/p + 1/q = 1$ and $1/p' + 1/q' = 1$ and we used that $\int_{\mathbb{R}^d} e^{H(\theta)}\mathrm{d}\mu(\theta) \leq e^2 + 1$.

From Proposition 18, we get that

$$\int_{\Omega} H(\theta)\mathrm{d}n_{k,t}(\theta) \leq C_{d,r,p,p'}\left[\int_{\Omega} H(\theta)\mathrm{d}\mu(\theta)\right]^{\frac{1}{q}}\left[\int_{\Omega} e^{(p'p-1)U(\theta)}\ n_{k,t}^{p'p}(\theta)\ \mathrm{d}\theta\right]^{\frac{1}{p'p}}.$$

where

$$C_{d,r,p,p'} = \left(2\left(\frac{2\pi}{\mathfrak{c}}\right)^{\frac{d}{2}} d^{\frac{dr}{2}}\right)^{1-\frac{1}{p'p}}\left(\frac{q'}{e}\left(e^2+1\right)^{\frac{1}{q'}}\right)^{\frac{1}{p}}.$$

Now we study the second integral. The expression inside the integral is upper bounded using Proposition 13 as follows

$$e^{(p'p-1)U(\theta)}\ n_{k,t}^{p'p}(\theta) \leq A_d^{p'p}\exp\left\{\left(\frac{9p'p}{10}-1\right)U(\theta)\right\},$$

If we choose $p > 1$ and $p' > 1$ such that $p'p < \frac{10}{9}$ then the exponent is negative. Let's consider particularly the midpoint of $[1, 10/9]$, then $p'p = 19/18$ and

$$e^{(p'p-1)U(\theta)}\ n_{k,t}^{p'p}(\theta) \leq A_d^{p'p}e^{-\frac{1}{20}U(\theta)}.$$

Using (14) and (15) one more time, we get

$$\begin{aligned}\int_{\Omega} e^{(p'p-1)U(\theta)}\ n_{k,t}^{p'p}(\theta)\ \mathrm{d}\theta &\leq A_d^{p'p}\int_{\mathbb{R}^d} e^{-\frac{1}{20}U(\theta)}\mathrm{d}\theta \leq A_d^{p'p}\int_{\mathbb{R}^d} e^{-\frac{\mathfrak{c}r}{20}\|\theta\|^{\frac{2}{1+r}}}\mathrm{d}\theta\\ &\leq 2d^{\frac{dr}{2}}A_d^{p'p}\left(\frac{20^{1+r}2\pi}{\mathfrak{c}}\right)^{\frac{d}{2}}.\end{aligned}$$

The results above imply the inequality

$$\int_{\Omega} H(\theta)\mathrm{d}n_{k,t}(\theta) \leq A_d \left[\int_{\Omega} H(\theta)\mathrm{d}\mu(\theta)\right]^{\frac{1}{q}},$$

where the constant A_d changes proportionally. If we choose $p > 1$ and $p' > 1$ such that $p'p = 19/18$, we would obtain the value of q.

C.4 Proof of Lemma 15

We recall that the Bobkov bound on the Poincaré constant $C_P(m)$ for the log-concave probability measure m is, see Bobkov (1999)

$$C_P(m) \geq \frac{1}{4K^2 Var_m(id)}, \tag{27}$$

where K is a universal constant and id is the identity function on $\mathbb{R}^d$, that is $id(\theta) = \theta$ for any $\theta \in \mathbb{R}^d$. From now on, we simply denote $Var_m(id)$ as $Var_m(\theta)$, where $\theta \sim m$.

We prove Lemma 15 using the Bobkov bound applied to μ and following the same procedure as in Proposition 3.5 of Crespo et al. (2023).

Proof Since U satisfies hypothesis $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$, Proposition 17 implies that

$$\|\theta - \theta^*\|^2 \leq \frac{2}{(1+r)\mathfrak{c}} U^{1+r}(\theta), \quad \forall \theta \in \mathbb{R}^d,$$

where $\theta^* = \arg\min U$. We use the fact that for any distribution m, $Var_m(\theta) \leq \mathbb{E}_m[\|\theta - a\|^2]$, for any $a \in \mathbb{R}^d$, then

$$Var_\mu(\theta) \leq \int_{\mathbb{R}^d} \|\theta - \theta^*\|^2 \mathrm{d}\mu \leq \frac{2}{(1+r)\mathfrak{c}} \mathbb{E}_\mu[U^{1+r}(\theta)].$$

We formulate the following lemma to control the moments of U along the Langevin dynamics (2). This lemma is taken from Crespo et al. (2023) but the proof will be omitted.

Lemma 20 *Assume that U satisfies $\mathcal{H}^{\mathbf{r}}_{\mathbf{KL}}(\mathfrak{c}, L)$ and $\mathcal{H}_{\min}$, then for any $\alpha \geq 1$ and any $t > 0$*

$$\mathbb{E}_{p_t}[U^\alpha(\vartheta_t)] \lesssim_{uc} d^{\alpha(1+r)},$$

where $\vartheta_t \sim p_t$ is the Langevin dynamics in (2) with initial distribution $\mathcal{N}(0, \sigma^2 I_d)$.

The ergodic behaviour of $(\vartheta_t)_{t\geq 0}$ yields

$$Var_\mu(\theta) \leq \frac{2}{(1+r)\mathfrak{c}} \limsup_{t\geq 0} \mathbb{E}_{p_t}[U^{1+r}(\vartheta_t)] \leq C d^{(1+r)^2}.$$

We conclude by using the Bobkov bound in (27). ∎