[1]

ON THE COST OF BAYESIAN POSTERIOR MEAN STRATEGY FOR LOG-CONCAVE MODELS

SÉBASTIEN GADAT, FABIEN PANLOUP AND CLÉMENT PELLEGRINI

Abstract. In this paper, we investigate the problem of computing Bayesian estimators using Langevin Monte-Carlo type approximation. The novelty of this paper is to consider together the statistical and numerical counterparts (in a general log-concave setting). More precisely, we address the following question: given n observations in $\mathbb{R}^q$ distributed under an unknown probability $\mathbb{P}_{\theta^\star}$, $\theta^\star \in \mathbb{R}^d$, what is the optimal numerical strategy and its cost for the approximation of $\theta^\star$ with the Bayesian posterior mean?

To answer this question, we establish some quantitative statistical bounds related to the underlying Poincaré constant of the model and establish new results about the numerical approximation of Gibbs measures by Cesaro averages of Euler schemes of (over-damped) Langevin diffusions. These last results include in particular some quantitative controls in the weakly convex case based on new bounds on the solution of the related *Poisson equation* of the diffusion.

1. Introduction

1.1. **Log-concave statistical models .** In this paper, we consider a statistical model $(\mathbb{P}_\theta)_{\theta\in\mathbb{R}^d}$ parametrized by a parameter $\theta \in \mathbb{R}^d$. We assume that each distribution $\mathbb{P}_\theta$ defines a probability measure on $(\mathbb{R}^q, \mathcal{B}(\mathbb{R}^q))$ and that all the distributions $\mathbb{P}_\theta$ are absolutely continuous with respect to the Lebesgue measure λ_q, we denote by π_θ the corresponding density:

$$\forall \xi \in \mathbb{R}^q \qquad \pi_\theta(\xi) := \frac{\mathrm{d}\mathbb{P}_\theta}{\mathrm{d}\lambda_q}(\xi).$$

We assume that we observe n i.i.d. realizations $(\xi_1, \ldots, \xi_n)$, sampled according to $\mathbb{P}_{\theta^\star}$ where $\theta^\star$ is an unknown parameter. We are then interested in Bayesian statistical procedures designed to recover $\theta^\star$. In all the paper, we restrict our study to the specific class of *log-concave models* where the distributions are described by:

$$\pi_\theta(\xi) := e^{-U(\xi,\theta)}, \tag{1}$$

where $(\xi, \theta) \longmapsto U(\xi, \theta) = -\log(\pi_\theta(\xi))$ is assumed to be a convex function. Note that implicitly, the normalizing constant $Z_\theta := \int_{\mathbb{R}^q} e^{-U(\xi,\theta)}\mathrm{d}\xi$ is assumed to be equal to 1, which is not restrictive up to a modification of U.

Besides the Gaussian toy model that trivially falls into our framework, log-concave statistical models have a longstanding history in a wide range of applied mathematics and it seems almost impossible to enumerate exhaustively the range of possible applications. For instance, the log-concave setting appears with exponential families thanks to the Pitman-Koopman-Darmois Theorem, in extreme value theory, tests (chi-square distributions), Bayesian statistics among others. Log-concave distributions also play a central role in probability and functional

[1] S. Gadat acknowledges funding from the French National Research Agency (ANR) under the Investments for the Future program (Investissements d'Avenir, grant ANR-17-EURE-0010) and for the grant MaSDOL - 19-CE23-0017-01.

Date: October 2, 2020.

analysis ([BBCG08, Bob99]), or geometry (see *e.g.* [KLS95]). The log-concave property is commonly used in economics (for example the density of customer's utility parameters is generally assumed to satisfy this property [BB05]), in game theory (see *e.g.* [CN91a] and [LT88]), in political science and social choice (see *e.g.* [CN91b]) or in econometrics (for example through the Roy model, see *e.g.* [HH90]).

An obvious and important example comes from all distributions that are built with a multivariate convex function $U : \mathbb{R}^d \times \mathbb{R}^q \longrightarrow \mathbb{R}$ and where the first d coordinates are considered as the hidden parameter θ while the q other ones are the observations. Starting from a convex function $U : \mathbb{R}^q \longrightarrow \mathbb{R}^q$, translation models $(\xi, \theta) \longmapsto e^{-U(\xi-\theta)}$ also generate typical examples of log-concave models in (ξ, θ). Many other distributions satisfy the log-concave property: Gumbel and Weibull distributions with a shape parameter larger than 1. In particular, this includes a large class of parametric probability distributions such as Gaussian or Laplace models, logistic regression models, Subbotin distributions, Gamma or Wishart distributions, Beta and uniform distributions on real intervals among others. We refer to [SW14] for a detailed survey on properties of log-concave distributions, to [Wal09] for a list of modeling and application issues and to [KS16] for non-parametric density estimation procedures of log-concave distributions.

1.2. **Bayesian estimation of $\theta^\star$.** We briefly sketch Bayesian strategies for estimating $\theta^\star$.

• **Bayesian paradigm.** Considering a prior distribution Π_0 on θ, we assume that Π_0 is absolutely continuous with respect to the Lebesgue measure λ_d on $\mathbb{R}^d$. Without any possible confusion with the familiy of densities $(\pi_\theta)_{\theta\in\mathbb{R}^d}$, the associated density of the prior is denoted by π_0:

$$\forall \theta \in \mathbb{R}^d \qquad \pi_0(\theta) := \frac{\mathrm{d}\Pi_0}{\mathrm{d}\lambda_d}(\theta).$$

We further assume that π_0 is also log-concave and we write π_0:

$$\forall \theta \in \mathbb{R}^d \qquad \pi_0(\theta) := e^{-V_0(\theta)}, \tag{2}$$

where V_0 encodes the prior knowledge on θ. We emphasize that this last assumption is not restrictive since the prior distribution is chosen by the user. We denote by π_n the density of the posterior distribution (that depends on the observations $\xi^{\mathbf{n}} := (\xi_1, \ldots, \xi_n)$) given by:

$$\pi_n(\theta) \propto \pi_0(\theta) \prod_{k=1}^{n} \pi_\theta(\xi_k).$$

The posterior distribution is a data-driven probability distribution that may be written as:

$$\forall \theta \in \mathbb{R}^d \qquad \pi_n(\theta) = \frac{e^{-W_n(\xi^{\mathbf{n}},\theta)}}{Z_n(\xi^{\mathbf{n}})} \pi_0(\theta) \quad \text{where} \quad W_n(\xi^{\mathbf{n}}, \theta) = \sum_{i=1}^{n} U(\xi_i, \theta). \tag{3}$$

The quantity $Z_n(\xi^{\mathbf{n}})$ corresponds to the normalizing constant and depends as well on $\xi^{\mathbf{n}}$:

$$Z_n(\xi^{\mathbf{n}}) = \int_{\mathbb{R}^d} \pi_0(\theta) e^{-W_n(\xi^{\mathbf{n}},\theta)} \mathrm{d}\theta.$$

It is well known that the posterior distribution enjoys consistency properties (see *e.g.* [Sch65, IH81]): under mild assumptions on the prior distribution and on the statistical model, the posterior distribution concentrates its mass around $\theta \in \mathbb{R}^d$ whose distribution is close to $\mathbb{P}_{\theta^\star}$.

• **Bayesian contraction rate** With additional metric and identifiability assumptions, some stronger results may be obtained in general parametric or non-parametric models. We refer to the seminal contribution of [GGvdV00] and the references therein, to the work of [CvdV12] for an extension to less standard situations of high-dimensional models and to [vdVvZ08] for infinite dimensional models.

The posterior distribution may be used to define Bayesian estimators, in particular, we shall introduce the popular posterior mean estimator of $\theta^\star$ defined by:

$$\widetilde{\theta}_n = \int_{\mathbb{R}^d} \theta \pi_n(\theta) \mathrm{d}\theta. \tag{4}$$

This estimator is usually consistent and a statistical issue for statisticians is to establish *rates of convergence* towards $\theta^\star$. In our paper, such rates of convergence will be investigated through a L^p-approach, by providing a sequence $\varepsilon_n \longrightarrow 0$ as $n \longrightarrow +\infty$ such that $(\widetilde{\theta}_n)_{n\geqslant 1}$ satisfies:

$$\mathbb{E}_{\theta^\star}\left(|\widetilde{\theta}_n - \theta^\star|^p\right) \leqslant \varepsilon_n^p, \tag{5}$$

for a given $p > 1$. Nevertheless, when a such bound is obtained, the story is not over. Actually, the theoretical posterior mean given by Equation (4) being generally not explicit, the practical use of the above statistical bounds certainly requires to provide computable algorithms that may approximate $\widetilde{\theta}_n$. In particular, it is legitimate to look for a tractable estimator $\widehat{\theta}_n$ that approaches an ε_n-neighborhood of $\widetilde{\theta}_n$ with as less operations as possible.

1.3. **Over-damped Langevin Monte Carlo (LMC) diffusion and continuous-time approximation.** To approximate $\widetilde{\theta}_n$, it is commonly used to write π_n as a Gibbs field:

$$\pi_n(\theta) \propto \exp(-\widetilde{W}_n(\xi^{\mathbf{n}}, \theta)) \quad \text{with} \quad \widetilde{W}_n(\theta) = \sum_{i=1}^n U(\xi_i, \theta) + \log\left(\frac{1}{\pi_0(\theta)}\right). \tag{6}$$

Under some mild assumptions on $\widetilde{W}_n$, it is well known that a such probability measure is the unique invariant distribution of the (over-damped) Langevin diffusion defined by:

$$dX_t^{(n)} = -\nabla \widetilde{W}_n(X_t^{(n)})dt + \sqrt{2}dB_t, \tag{7}$$

where (B_t) is a d-dimensional standard Brownian motion.

Thus, the probability π_n can be approximated using the long-time convergence of $(X_t^{(n)})_{t\geqslant 0}$ towards its invariant distribution π_n. One can mainly distinguish two types of convergences towards π_n: the convergence of the distribution of $X_t^{(n)}$ as $t \to +\infty$ or the *a.s.* convergence of the occupation measure of $(X_t^{(n)})_{t\geqslant 0}$. Here, we build our algorithm with the second type of convergence, which requires only one path of the diffusion.

• **Cesaro average** We are thus led to consider the occupation measure applied to the identity function denoted by $\mathbf{I}_d$. In this case, this is nothing but the Cesaro average related to the diffusion:

$$\forall n \in \mathbb{N}^\star \quad \forall t > 0 \qquad \widehat{\theta}_{n,t} := \frac{1}{t}\int_0^t \mathbf{I}_d(X_s^{(n)})\mathrm{d}s = \frac{1}{t}\int_0^t X_s^{(n)}\mathrm{d}s. \tag{8}$$

Under mild conditions, $(\widehat{\theta}_{n,t})_{t\geqslant 0}$ converges *a.s.* towards $\widetilde{\theta}_n$. One objective is thus to sharply estimate the related error in order to assess the complexity for the approximation of $\theta^\star$ (with the help of (5)).

• **Shifted Cesaro average** We will also introduce a second approximation, which is a *shifted* Cesaro average that omits the very first times involved in the simulation of the trajectory, before t_0. This threshold t_0 will be chosen carefully to optimize our approximation procedure:

$$\forall n \in \mathbb{N}^\star \quad \forall t > t_0 > 0 \qquad \widehat{\theta}_{n,t_0,t} := \frac{1}{t-t_0}\int_{t_0}^{t} X_s^{(n)} \mathrm{d}s. \tag{9}$$

1.4. **Langevin Monte Carlo discretization and practical estimator.** In (8) and (9), Cesaro averages are based on the "true" diffusion but to obtain a tractable algorithm, we need to introduce Cesaro averages of some discretization schemes of (8) and (9). For this purpose, we consider a step sequence $(\gamma_k)_{k\geqslant 1}$ of positive numbers such that $\sum_{k\geqslant 1}\gamma_k = +\infty$. Throughout the paper, this sequence will be constant or decreasing (typically, $\gamma_k = \gamma k^{-r}$ with $r \in [0,1]$). Denoting by $t_0 = 0$ and $t_k = \sum_{\ell=1}^{k}\gamma_\ell$, we introduce $(\bar{X}_t)_{t\geqslant 0}$ the stepwise constant explicit Euler-Maruyama scheme related to $X^{(n)}$ (omitting the index n for simplicity):

$$\bar{X}_{t_{k+1}} := \bar{X}_{t_k} - \gamma_{k+1}\nabla \widetilde{W}_n(\bar{X}_{t_k}) + \sqrt{2}\zeta_{k+1}, \tag{10}$$

where for all $k \geqslant 1$, $\zeta_k = B_{t_k} - B_{t_{k-1}}$ and $(B_t)_{t\geqslant 0}$ is a standard d-dimensional Brownian motion. It is possible to define a continuous affine interpolation of (10) but from a practical point of view, it will be more comfortable to consider some initialization and ending times in the discrete grid $(t_k)_{k\geqslant 0}$. For $s > 0$, we define $\underline{s}$ the largest grid point in $(t_k)_{k\geqslant 0}$ below s:

$$\underline{s} := \sup\{t_k \,:\, t_k \leqslant s\}. \tag{11}$$

Then, for any time-shift $t_{k_0} > 0$ and any final horizon $t_N > 0$, we denote by $\bar{\mu}_{t_{k_0},t_N}$, the occupation measure:

$$\bar{\mu}_{t_{k_0},t_N} = \frac{1}{t_N - t_{k_0}}\int_{t_{k_0}}^{t_N} \delta_{\bar{X}_{\underline{s}}}\mathrm{d}s = \frac{1}{t_N - t_{k_0}}\sum_{i=k_0}^{N-1}\gamma_{i+1}\delta_{\bar{X}_{t_i}}.$$

Then, the approximation of $\widehat{\theta}_{n,t_0,t}$ with a step-size sequence $(\gamma_k)_{k\geqslant 0}$ is given by:

$$\widehat{\theta}^{\gamma}_{n,t_{k_0},t_N} := \int_{\mathbb{R}^d} x\bar{\mu}_{t_{k_0},t_N}(\mathrm{d}x) = \frac{\sum_{j=k_0}^{N-1}\gamma_{j+1}\bar{X}_{t_j}}{\sum_{j=k_0}^{N-1}\gamma_{j+1}}, \tag{12}$$

which corresponds to the weighted Cesaro average of the discretized trajectory (10) from iteration k_0 to iteration N. This Cesaro construction first appeared in [Tal90] where some convergence properties of the empirical measure of the Euler scheme with constant step size were investigated. In a series of more recent papers (among others, see [LP02, LP03, PP12] or [PP18] for a *multilevel* extension), the decreasing-step setting has been deeply studied. Compared with these papers, the novelty of our work is that, we propose some non-asymptotic quantitative bounds (see Section 1.5 for details about the corresponding results). Note that for ease of presentation, we prefered to mainly consider the (less technical) constant step setting.

1.5. **Contributions and plan of the paper.** Before stating the main results, we provide a brief description of our main contributions. As mentioned before, our aim is to tackle the Bayesian estimation problem with a quantitative computational approach taking into account:

- The Bayesian consistency problem and the (as sharp as possible) control of the distance between the Bayesian posterior mean $\widetilde{\theta}_n$ and $\theta^\star$.

- The numerical question related to the approximation of $\widetilde{\theta}_n$ by a computable algorithm.

In short, for some given n and d, we first exhibit an ε_n that upper-bounds the (L^p-type) error between $\widetilde{\theta}_n$ and $\theta^\star$. Then, for this ε_n, we aim at tuning the procedure (12) in order to obtain an ε_n-approximation[1] of $\widetilde{\theta}_n$ with a minimal computational cost (in terms of n and d). This cost N (number of iterations of the Euler scheme) will be explicited as a $\mathcal{O}_{id}(n^a d^b)$ (for some positive a and b) where for some sequences (u_n) and (v_n),

$$u_n = \mathcal{O}_{id}(v_n) \Longleftrightarrow |u_n| \lesssim_{id} |v_n| \Longleftrightarrow \exists\, C \text{ independent of } d \text{ such that } \forall\, n,\ |u_n| \leqslant C|v_n|.$$

Our main contribution states that Bayesian learning can be optimally performed in $\mathcal{O}_{id}(nd \vee n^2 d^{-1})$ for strongly convex situations and remain polynomial in weakly convex cases.

Related to these objectives, we state our main results in Section 2 starting with Theorem 2.3 on the Bayesian consistency. We first provide a new bound of type (5) expressed in terms of the *Poincaré constant* of the model (which is assumed to satisfy a uniformity condition). Compared with the literature on this problem (see *e.g.* [MHW⁺19]), this result is written under general assumptions on the family of log-concave models and does not require specific assumptions related to a dynamical system.

Then, we turn out to the second objective by stating a series of results on the approximation of $\widetilde{\theta}_n$. Inspired by [CCG12], Theorem 2.4 (resp. Theorem 2.5) bounds the L^2-error between the Cesaro (resp. shifted) average of the LMC diffusion and $\widetilde{\theta}_n$. These results, written in a general log-concave setting and related to the "true" diffusion, which can not be simulated in general, represent a benchmark. All the more it gives some indications on the right rate of convergence and on the role of the time shift (say also *Warm start*).

Our main results about Cesaro-type LMC and optimal tuning of the parameters (in terms of n and d) are Theorems 2.8 and 2.9. In the first one, we investigate the strongly convex case where it is possible to control the distance between the "true" and discretized diffusions. Note that the strategy of proof of Theorem 2.8 is similar to [DM19][2], where the authors estimate the invariant distribution through the distribution of the Euler scheme (instead of a Cesaro average in our setting).

In the weakly-convex case, our argument is based on a "Poisson approach", *i.e.* on the inversion of the infinitesimal generator of the diffusion. Even though classical in the literature for the study of Cesaro averages (see *e.g.* [LP02, HMP20]), the difficulty here is to provide some quantitative controls on the solution of the *Poisson equation*. Here, with the help of a sharp study of the tangent process of the diffusion, we obtain some explicit bounds on the distance between the discretized Cesaro average and the posterior mean (see in particular Theorem 6.2). These bounds lead to Theorem 2.9, which provides our optimal tuning in terms of n and d in the weakly convex setting. Section 2 ends with an extended discussion and comparison with the historical and state of the art contributions on Bayesian learning theory (Section 2.5).

For an improved readability of the work, numerous proofs and technical results are deferred to the supplementary file [GPP20].

2. Main results and discussion

We provide here some notations and assumptions all along the statement of our results.

[1]By ε-approximation, we mean an approximation of the target with an L^p-error of the order $\mathcal{O}(\varepsilon)$.

[2]A more detailed comparison with the (huge) literature on this topic is given in Section 2.5.4.

2.1. **Functional inequality and Assumption ($\mathbf{PI_U}$).** For any measure μ and $f \in L^1(\mu)$, $\mu(f)$ refers to the mean value of f, and when $f \in L^2(\mu)$, $Var_\mu(f)$ is the variance of f:

$$\mu(f) := \int_{\mathbb{R}^q} f(\xi)\mathrm{d}\mu(\xi) \qquad \text{and} \qquad Var_\mu(f) := \int_{\mathbb{R}^q} [f(\xi) - \mu(f)]^2 \mathrm{d}\mu(\xi).$$

A crucial property of log-concave measures is that they satisfy a Poincaré inequality. This will be used extensively in the rest of the paper. We refer to [Led01, BGL14] for a complete presentation and some applications on concentration inequalities and Markov processes.

Definition 2.1 (Poincaré inequality)**.** μ satisfies a Poincaré inequality with $C_P(\mu)$ if

$$\forall f \in L^2(\mu) \qquad Var_\mu(f) \leqslant C_P(\mu)\mu(|\nabla f|^2).$$

We remind an important result obtained in [Bob99] (see also [BBCG08]) that establishes the existence of a Poincaré inequality for every log-concave probability distribution.

Theorem 2.2 ([Bob99, KLS95])**.** *Every log-concave measure with density μ satisfies a Poincaré inequality: a universal constant K exists such that:*

$$C_P(\mu) \leqslant 4K^2 \mu(\|\mathbf{I}_d - \mu(\mathbf{I}_d)\|^2).$$

Since $(\xi,\theta) \longmapsto U(\xi,\theta)$ is a convex function, Theorem 2.2 implies that for all $\theta \in \mathbb{R}^d$, $\mathbb{P}_\theta$ satisfies a Poincaré inequality of constant $C_P(\mathbb{P}_\theta)$. We introduce an assumption that stands for a uniform bound of the collection of Poincaré constants $C_P(\mathbb{P}_\theta)$.

Assumption 2.1 (Uniform Poincaré Inequality ($\mathbf{PI_U}$))**.** *A constant C_P^U exists such that*

$$\forall \theta \in \mathbb{R}^d \qquad C_P(\mathbb{P}_\theta) \leqslant C_P^U.$$

We emphasize that according to Theorem 2.2, a uniform bound on the variance of each distribution $\mathbb{P}_\theta$ over $\theta \in \mathbb{R}^d$ entails ($\mathbf{PI_U}$).

2.2. **Bayesian consistency.**

2.2.1. *Notations, Assumptions* ($\mathbf{A}_L$) *and* ($\mathbf{I_{W_1}(c)}$). For any integrable random variable $\psi(\Xi)$ in $L^1(\pi_\theta)$, $\mathbb{E}_\theta[\psi(\Xi)]$ will refer to the expectation of the random variable $\psi(\Xi)$ when Ξ is sampled according to $\mathbb{P}_\theta$:

$$\mathbb{E}_\theta[\psi(\Xi)] = \int_{\mathbb{R}^q} \psi(\xi) e^{-U(\xi,\theta)} \lambda_q(\mathrm{d}\xi).$$

In our work, two sources of randomness are considered. The first one is derived from the observations $\xi^{\mathbf{n}}$: $\mathbb{P}_\theta$ and $\mathbb{E}_\theta$ refer to the probability and expectation on the unknown distribution of the sampling process. The second source of randomness is related to the posterior distribution π_n over $\mathbb{R}^d$: for any Borelian $\mathcal{B}$ of $\mathbb{R}^d$, $\pi_n(\mathcal{B})$ is the probability of $\mathcal{B}$ when θ is sampled according to π_n, conditionnally to Ξ. Hence, $\mathbb{E}_{\pi_n}$ is the expectation when $\theta \sim \pi_n$, conditionnally to $\xi^{\mathbf{n}}$.

We introduce some mild assumptions necessary to obtain some consistency rates of $\widetilde{\theta}_n$. First, we handle smooth functions $(\xi,\theta) \longmapsto U(\xi,\theta)$ and assume that:

Assumption 2.2 (Assumption ($\mathbf{A}_L$))**.** *U satisfies the $\mathcal{C}^1_L$ hypothesis: i.e. the partial gradient of U with respect to θ is a L-Lipschitz function:*

$$\forall \xi \in \mathbb{R}^q \quad \forall \theta_1,\theta_2 \in \mathbb{R}^d : \qquad |\nabla_\theta U(\xi,\theta_1) - \nabla_\theta U(\xi,\theta_2)| \leqslant L|\theta_1 - \theta_2|.$$

To make the estimation problem feasible, we need to manipulate some statistically identifiable models. If statistical identifiability is a free result for any L^1 location model when $U(\xi,\theta) = U(\xi - \theta)$, this is no longer the case in general statistical models, even log-concave ones. We therefore introduce an identifiability assumption, which will be important below for our theoretical results.

Assumption 2.3 (Assumption ($\mathbf{I}_{\mathbf{W_1}}(\mathbf{c})$) - Wasserstein identifiability)**.** *A strictly increasing map $c : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ and a 1-Lipschitz function Ψ exist such that $c(0) = 0$ and :*

$$\forall \theta_1, \theta_2 \in \mathbb{R}^d, \quad |\pi_{\theta_1}(\Psi) - \pi_{\theta_2}(\Psi)| \geqslant c(|\theta_1 - \theta_2|).$$

Furthermore, we shall assume that a pair $(b_1, b_2) \in \{\mathbb{R}_+^\star\}^2$ and $\alpha_c>0$ exists such that:

$$\forall \Delta \geqslant 0 \qquad c(\Delta) \geqslant b_1 \Delta^{\alpha_c} \mathbf{1}_{\Delta \leqslant 1} + b_2[\log(\Delta) + 1]\mathbf{1}_{\Delta \geqslant 1}. \tag{13}$$

Assumption ($\mathbf{I}_{\mathbf{W_1}}(\mathbf{c})$) is a quantitative identifiability condition. It implies in particular that

$$\mathcal{W}_1(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2}) \geqslant c(|\theta_1 - \theta_2|),$$

where $\mathcal{W}_1(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2})$ refers to the 1-Wasserstein distance between $\mathbb{P}_{\theta_1}$ and $\mathbb{P}_{\theta_2}$. The constants b_1 and b_2 are not fundamental in our forthcoming analysis but the parameter α_c plays a central role: it asserts how the distributions $\pi_{\theta_1+\delta}$ moves from π_{θ_1} for small values of δ. We refer to Section 2.5 for an extended discussion on this assumption.

2.2.2. *Bayesian consistency* . We obtain the next result, which is still valid besides the log-concave settings. We have chosen to keep this setting for the sake of readability, even though ($\mathbf{PI}_\mathbf{U}$) is sufficient here to guarantee the result.

Theorem 2.3. *If $\pi_0 = e^{-V_0}$ is a $\mathcal{C}^1(\mathbb{R}^d, \mathbb{R})$ log-concave prior with $V_0 \in \mathcal{C}^1_1$, if ($\mathbf{PI}_\mathbf{U}$), ($\mathbf{A}_L$) and ($\mathbf{I}_{\mathbf{W_1}}(\mathbf{c})$) hold, then*

$$\forall p > 1 \qquad \left(\mathbb{E}_{\theta^\star}\left(\left|\widetilde{\theta}_n - \theta^\star\right|^p\right)\right)^{1/p} \lesssim_{id} K(U) \left(d \frac{\log(n)}{n}\right)^{\frac{1}{2\alpha_c}},$$

with $K(U) = \left(\sqrt{C_P^U} L\right)^{\frac{1}{\alpha_c}}$ and C_P^U is the Poincaré constant given in ($\mathbf{PI}_\mathbf{U}$) (Assumption 2.1).

This result states an upper bound on the L^p loss between the posterior mean $\widetilde{\theta}_n$ and $\theta^\star$:

$$\varepsilon_n^2 := \left(C_P^U L^2\right)^{1/\alpha_c} \mathcal{O}_{id}\left(\frac{d \log n}{n}\right)^{1/\alpha_c}, \tag{14}$$

and in particular, when $p = 2$, we recover the standard mean square error rate. If the separation provided by ($\mathbf{I}_{\mathbf{W_1}}(\mathbf{c})$) is sharp, *i.e.* when $\alpha_c = 1$, the L^2 loss is proportional to $\sqrt{\frac{d}{n}}$ (up to a log-term), which is the optimal loss in many statistical models. Our upper bound is deteriorated when α_c increases, *i.e.* when the separation of the distributions $\mathbb{P}_\theta$ near $\mathbb{P}_{\theta^\star}$ is "flat", *i.e.* when the Wasserstein distance $\mathcal{W}_1(\mathbb{P}_\theta, \mathbb{P}_{\theta^\star}) \sim b_1 \|\theta - \theta^\star\|^{1+\varepsilon}$ for $\varepsilon > 0$ near $\theta^\star$.

2.3. **Continuous-time Cesaro strategies .**

2.3.1. *Speed of approximation.* For a fixed sample $\xi^{\mathbf{n}} = (\xi_1, \ldots, \xi_n)$, the function $\widetilde{W}_n$ introduced in (6) is a convex function, so that Theorem 2.2 entails a Poincaré inequality for the posterior distribution π_n. Explicitely, π_n satisfies a Poincaré inequality with a (sample-dependent) constant $C_P(\pi_n)$ whereas the spectral gap $\lambda_{1,n}$ is related to $C_P(\pi_n)$ through:

$$C_P(\pi_n) := \frac{1}{\lambda_{1,n}}.$$

In the next result, we connect the approximation of $\widetilde{\theta}_n$ by a continuous-time LMC diffusion with the (sample-dependent) moments of π_n:

$$\mathbb{V}_{n,2} := \mathbb{E}_{\pi_n}\left(|\theta - \widetilde{\theta}_n|^2\right) \qquad \text{and} \qquad \mathbb{V}_{n,4} := \mathbb{E}_{\pi_n}\left(|\theta - \widetilde{\theta}_n|^4\right). \tag{15}$$

Again, the next result is still valid as soon as π_n enjoys a Poincaré inequality with constant $C_P(\pi_n)$. The log-concave setting is only a convenient way to guarantee this functional inequality.

Theorem 2.4 (Approximation of the posterior mean: Cesaro average with $(\widehat{\theta}_{n,t})_{t>0}$)**.** *For any μ such that $X_0^{(n)} \sim \mu$, for any horizon time $t > 0$ and any $\alpha > 0$:*

$$\mathbb{E}_\mu\left(|\widehat{\theta}_{n,t} - \widetilde{\theta}_n|^2\right) \leqslant \sqrt{\mathbb{V}_{n,4}}\left[2\sqrt{2}\frac{\alpha^2 \log(t)^2}{\lambda_{1,n}^2 t^2}\sqrt{1 + J_{\mu,0}} + \frac{4}{t\lambda_{1,n}} + \sqrt{J_{\mu,0}}t^{-\alpha}\right],$$

where $J_{\mu,0}$ stands for the initial $L^2(\pi_n)$-distance between $\mu\,\pi_n^{-1}$ and $\mathbf{1}$: $J_{\mu,0} = \|\mu\pi_n^{-1} - \mathbf{1}\|^2_{L^2(\pi_n)}$.

Remark 1. Theorem 2.4 shows that the efficiency of $\widehat{\theta}_{n,t}$ highly depends on $\mathbb{V}_{n,4}$, $\lambda_{1,n}$ and $J_{\mu,0}$. If $\mathbb{V}_{n,4}$ and $\frac{1}{\lambda_{1,n}^2}$ depends on the concentration of the posterior distribution π_n and will be shown to be small, we also observe that the distance between the initialization of the Langevin diffusion and the target measure π_n may crucially harm the error bound. In particular, we will prove that the typical *warm start* (or shift) strategies shall produce $J_{\mu,0}$ of the order $O(e^{d\log d})$ for our log-concave model, which leads to very poor approximation rate (in terms of the dimension d) in Theorem 2.4. In particular, the time t needed to "kill" the poor initialization of the process with μ will be exponential with the dimension if we consider the very first part of the trajectory $(\theta_s^{(n)})_{s\leqslant t_0}$ since we need to upper bound $\sqrt{J_{\mu,0}}t^{-2}$.

We now state that $\widehat{\theta}_{n,t_0,t}$ introduced in (9) significantly improves the approximation of $\widetilde{\theta}_n$.

Theorem 2.5 (Approximation of the posterior mean: *Shifted* Cesaro average with $\widehat{\theta}_{n,t_0,t}$)**.** *For any μ such that $X_0^{(n)} \sim \mu$:*

$$\forall t > 0 \qquad \mathbb{E}_\mu\left(|\widehat{\theta}_{n,t_0,t} - \widetilde{\theta}_n|^2\right) \leqslant \sqrt{\mathbb{V}_{n,4}}\left[\frac{4}{(t-t_0)\lambda_{1,n}} + \sqrt{J_{\mu,0}}e^{-\lambda_{1,n}t_0}\right].$$

2.3.2. *ε−computational cost.* Theorems 2.4 and 2.5 are useful to assess the speed of approximation of $\widetilde{\theta}_n$ with $(\widehat{\theta}_{n,t})_{t\geqslant 0}$. In particular, this speed highly depends on several sample dependent random variables related to π_n: $\mathbb{V}_{n,4}$, $\lambda_{1,n}$ and $J_{\mu,0}$. We denote by $\Sigma_{\mu,n}$ and $\Sigma'_{\mu,n}$:

$$\Sigma_{\mu,n} = \sqrt{2}(1 + J_{\mu,0})^{1/4}\frac{\mathbb{V}_{n,4}^{1/4}}{\lambda_{1,n}} \qquad \text{and} \qquad \Sigma'_{\mu,n} = 4\frac{\sqrt{J_{\mu,0}\mathbb{V}_{n,4}}}{\lambda_{1,n}}. \tag{16}$$

A straightforward application of Theorem 2.4 and Theorem 2.5 yields:

Corollary 2.6. *Assume that* $(\mathbf{I_{W_1}(c)})$, $(\mathbf{A}_L)$ *and* $(\mathbf{PI_U})$ *hold, then we have:*

i) For any $\varepsilon > 0$ *and* $\alpha > 0$*, we have:*

$$t \geqslant t^\star_{\varepsilon,n} := \frac{\alpha \Sigma_{\mu,n}}{\varepsilon} \log^2\left(\frac{\alpha\Sigma_{\mu,n}}{\varepsilon}\right) \vee \frac{\Sigma'_{\mu,n}}{\varepsilon^2} \vee \left(\frac{\sqrt{\mathbb{V}_{n,4}}}{\varepsilon^2}\right)^{1/\alpha} \Longrightarrow \mathbb{E}_\mu\left(|\widehat{\theta}_{n,t} - \widetilde{\theta}_n|^2\right) \leqslant 3\varepsilon^2.$$

ii) For any $\varepsilon > 0$*, we have:*

$$\underbrace{t_0 \geqslant \frac{1}{\lambda_{1,n}} \log\left(\frac{\sqrt{\mathbb{V}_{n,4} J_{\mu,0}}}{\varepsilon^2}\right)}_{:=t^{0,\star}_{\varepsilon,n}} \quad \text{and} \quad t - t_0 \geqslant \underbrace{\frac{\sqrt{\mathbb{V}_{n,4}}}{\lambda_{1,n}\varepsilon^2}}_{:=\{\Delta t\}^\star_{\varepsilon,n}} \Longrightarrow \mathbb{E}_\mu\left(|\widehat{\theta}_{n,t_0,t} - \widetilde{\theta}_n|^2\right) \leqslant 2\varepsilon^2.$$

The effect of $J_{\mu,0}$ on the computational time of the Langevin diffusion is therefore crucial. It is well known that the initialization of the Langevin diffusion must be tuned carefully. We refer to Section 4.1 of [Dal17] for an extended discussion, with a chi-square distance $\chi^2(\mu, \pi_n)$ (instead of a L^2 distance in our case). We introduce the (sample dependent) minimizer of $\widetilde{W}_n$:

$$\theta^{\widetilde{W}_n} = \arg\min_{\theta \in \mathbb{R}^d} \widetilde{W}_n(\theta). \tag{17}$$

and the following assumption:

Assumption 2.4 (Behaviour of $U(\xi, .)$ - $(\mathbf{C}^\beta)$)**.** *A pair* $(\beta, c_\beta) \in [1,2] \times \mathbb{R}^\star_+$ *exists such that:*

$$\forall \xi \qquad \limsup_{|\theta| \longrightarrow +\infty} U(\xi,\theta)|\theta|^{-\beta} \geqslant c_\beta.$$

Using the value ε_n introduced in Theorem 2.3 (see Equation (14)), we obtain the following corollary, whose proof is deferred to Section 3 of the supplementary materials [GPP20]. In particular the technical choice of the initial measure is explained in depth.

Corollary 2.7. *Assume that* $(\mathbf{I_{W_1}(c)})$, $(\mathbf{A}_L)$ *and* $(\mathbf{PI_U})$ *hold. If* μ *is the uniform distribution over* $\mathcal{B}(\theta^{\widetilde{W}_n}, a)$ *with* $a < n^{-1/2}$ *and* $(\mathbf{C}^\beta)$ *holds, we have:*

$$\mathbb{E}_{\theta^\star}[t^\star_{\varepsilon_n,n}] = \mathcal{O}_{id}\left(e^{[d\log(d) + \log(n)]}\right)$$

and

$$\mathbb{E}_{\theta^\star}[t^{0,\star}_{\varepsilon_n,n}] = \mathcal{O}_{id}\left(\varepsilon_n^2[d\log d + \log n]\right) \qquad \text{and} \qquad \mathbb{E}_{\theta^\star}[\{\Delta t\}^\star_{\varepsilon_n,n}] = \mathcal{O}_{id}\left(\varepsilon_n^2\right).$$

In this general setting, the shifted strategy is far better than the Cesaro averaging over a whole trajectory. Therefore, in this continuous-time setting, the essential cost of Bayesian learning is brought by $t^{0,\star}_{\varepsilon_n,n}$. Note that in some less general settings (including strongly convex diffusions), $J_{\mu,0}$ is significantly smaller. It implies that the above theoretical result is not explicitly used in the discretization part below.

At this stage, this result does not really quantify the cost of an *algorithm*. We will see in the next paragraph that the essential cost of the LMC is inherited from the discretization.

2.4. **Discretization cost.** We now assess the efficiency of the whole Bayesian approach when using $\widehat{\theta}^{\gamma}_{n,t_{k_0},t_N}$ introduced in (12). We present our statements into two separate paragraphs. These results highly depend on two general results on the discretization of LMC, that are deferred to Section 5 and Section 6. We still use ε_n^2 defined in (14) in this paragraph. At this stage, we introduce the expectation $\mathbf{E}[.]$ that is computed with respect to the sampling of $\xi^{\mathbf{n}}$ and the discrete LMC procedure.

2.4.1. *Strongly convex case.* This paragraph concerns the strongly convex and L-smooth case.

Assumption 2.5 (Assumption $(\mathbf{SC}_\rho)$)**.** *We assume that for any $\xi \in \mathbb{R}^q$, $U(\xi,.)$ is ρ-strongly convex: i.e., for any $\xi \in \mathbb{R}^q$:*

$$\forall z \in \mathbb{R}^d \quad \forall \theta \in \mathbb{R}^d : \qquad {}^t z \nabla^2_\theta U(\xi,\theta) z \geqslant \rho |z|^2.$$

The next theorem gives a quantitative setup to attain an ε_n^2 accuracy for a discrete LMC procedure with a constant step-size γ.

Theorem 2.8. *Assume that $(\mathbf{I}_{\mathbf{W_1}}(\mathbf{c}))$, $(\mathbf{A}_L)$ and $(\mathbf{PI_U})$. Let ε_n be defined by (14) and suppose that $(\mathbf{SC}_\rho)$ holds. Then, $\mathbf{E}[|\widehat{\theta}^{\gamma}_{n,t_{k_0},t_N} - \theta^\star|^2] \leqslant \varepsilon_n^2$ for the following tuning of the parameters:*

$$\gamma = \frac{\rho}{L^2}\mathcal{O}_{id}\left(\frac{\varepsilon_n^2}{nd}\right), \quad k_0 = \left(\frac{L}{\rho}\right)^2 \mathcal{O}_{id}\left(\varepsilon_n^{-2} d^2 \log(d)\right), \quad \textit{and} \quad N_0 := N - k_0 = \left(\frac{L}{\rho}\right)^2 \mathcal{O}_{id}\left(d\varepsilon_n^{-4}\right).$$

In particular, the number of iterations should be of the order

$$\left(\frac{L}{\rho}\right)^2 \left(\mathcal{O}_{id}\left(d^{1-\frac{2}{\alpha_c}} n^{\frac{2}{\alpha_c}}\right) \vee \mathcal{O}_{id}\left(d^{2-\frac{1}{\alpha_c}} \log(d) n^{\frac{1}{\alpha_c}}\right)\right).$$

Remark 2. This result deserves several comments.

- This theorem indicates that the step size should be chosen as $\gamma \propto \frac{\varepsilon_n^{-2}}{nd}$, which becomes smaller when n increases. This is due to the sharper statistical accuracy we can expect with the posterior mean when we have a large amount of observations.
- The computational cost is translated by k_0 (the size of the shift before we record the computations of the Cesaro average) and N_0 (the length of the Cesaro window). The main part of this computational cost is brought either by k_0 or N_0 and this balance depends on d, n and α_c. Straightforward computations (if we omit the log terms) show that:

$$\begin{aligned} k_0 \leqslant N_0 &\Longleftrightarrow d^2 \log(d) \varepsilon_n^{-2} \leqslant d\varepsilon_n^{-4} \\ &\Longleftrightarrow d\log(d) \leqslant \varepsilon_n^{-2} \\ &\Longleftrightarrow n \geqslant d^{\alpha_c+1}. \end{aligned}$$

- For "easy" estimation problem like location statistical models where $\alpha_c = 1$, we obtain an overall number of iterations of the order:

$$\frac{n^2}{d} \vee nd \quad \text{(up to log-terms)}.$$

- Another extreme case corresponds to $\alpha_c = 2$ that translates an important flatness of the application $\theta \longmapsto \mathbb{P}_\theta$ around $\theta^\star$ and a much more harder estimation problem. In that case, the number of iterations should be of the order:

$$n \vee d^{3/2}\sqrt{n} \quad \text{(up to log-terms)}.$$

- Finally, we observe that with $(\mathbf{SC}_\rho)$, besides the obvious curse of dimensionality for large d concerning the *statistical* accuracy of $\widehat{\theta}_n$, this phenomenon doesn't appear when looking at the *computational* cost, which is polynomial in terms of n and d

2.4.2. *Weakly convex case.* We still assume $(\mathbf{A}_L)$ and introduce for any convex function W:

$$\bar{\lambda}_{\nabla^2 W(x)} := \sup\{Sp(\nabla^2 W(x))\} \quad \text{and} \quad \underline{\lambda}_{\nabla^2 W}(x) = \inf\{Sp(\nabla^2 W(x))\}. \tag{18}$$

For some $\mathfrak{c}_1 > 0$, $\mathfrak{c}_2 > 0$ and $r \in [0,1]$, we also introduce the two following assumptions:

Assumption 2.6 (Assumption $(\mathbf{H}_{\mathfrak{c}_1})$). $\lim_{|x|\to+\infty} |\nabla W(x)|^2 = +\infty$ *and* $|\nabla W|^2 \leqslant \mathfrak{c}_1 W$.

Assumption 2.7 (Assumption $(\mathbf{H}_{\mathfrak{c}_2,r})$). $x \longmapsto \underline{\lambda}_{\nabla^2 W}(x)$ *is positive and*

$$\forall x \in \mathbb{R}^d \qquad \underline{\lambda}_{\nabla^2 W}(x) \geqslant \mathfrak{c}_2^{-1} W(x)^{-r}.$$

The parameter r can be viewed as a parametrization of the lack of curvature, varying from 0 (in the strongly convex case) to 1 in the "almost" Laplace case. In particular, the complexity of the procedure will increase with r.
For any positive c and any $\gamma_0 > 0$, let $\mathcal{C}_{c,\gamma_0}$ be defined by:

$$\mathcal{C}_{c,\gamma_0} := \{x \in \mathbb{R}^d, |\nabla W(x)|^2 - cd\bar{\lambda}_{\text{loc}}(\gamma_0, x) \leqslant 1\} \text{ where } \bar{\lambda}_{\text{loc}}(\gamma_0, x) := \sup_{|u-x|<\gamma_0|\nabla W(x)|+2} \bar{\lambda}_{\nabla^2 W}(u). \tag{19}$$

We observe that $(\mathbf{H}_{\mathfrak{c}_1})$ entails the compactness of $\mathcal{C}_{c,\gamma_0}$. For a positive λ, we set:

$$\beta(\lambda, c, \gamma_0) := \sup_{x\in\mathcal{C}_{c,\gamma_0}} e^{\lambda W(x)}(1 + cdL) \quad \text{and} \quad \mathfrak{b}_d = \log(\beta(1/8, 5, \gamma_0)) \vee 1. \tag{20}$$

Finally, when W is $\mathcal{C}^3$, we introduce the following notation:

$$\|D^3 W\|_\infty := \sup_i \sup_{x\in\mathbb{R}^d} \|\nabla^3_{i,.,.} W(x)\|_F, \tag{21}$$

where $\nabla^3_{i,.,.}$ refers to the squared matrix built with the third order partial derivatives of W when the variable i is kept fixed.

In the next result, we will have to use the above assumptions for the family of functions $(U_\xi := U(\xi, .))_{\xi\in\mathbb{R}^q}$. We will write $\mathfrak{b}_d^{(\xi)}$ instead of $\mathfrak{b}_d$ in order to recall that $\mathfrak{b}_d$ depends on U_ξ.

Theorem 2.9. *Assume* $(\mathbf{I}_{\mathbf{W_1}}(\mathbf{c}))$, $(\mathbf{A}_L)$ *and* $(\mathbf{PI_U})$. *Let* ε_n *be defined by* (14). *Suppose that for any* $\xi \in \mathbb{R}^q$, $U(\xi, .)$ *satisfies* $(\mathbf{H}_{\mathfrak{c}_1})$ *and* $(\mathbf{H}_{\mathfrak{c}_2,r})$, *with* $\mathfrak{c}_1$, $\mathfrak{c}_2$ *and* $r \in [0,1]$ *independent of* ξ, *and that* $\sup_{\xi\in\mathbb{R}^d} \mathfrak{b}_d^{(\xi)} \lesssim_{id} d$. *Then,* $\mathbf{E}\left[|\widehat{\theta}^\gamma_{n,0,t_N} - \theta^\star|^2\right] \lesssim_{id} \varepsilon_n^2$ *for the following tuning of the parameters:*

i) *For a given arbitrary small* $\mathfrak{e}$,

$$\gamma = \frac{\varepsilon_n}{Lnd^{1+r+\frac{\mathfrak{e}}{2}}} \wedge \frac{\varepsilon_n^2}{L^2 d^{2+2r+\mathfrak{e}}}. \quad \text{and} \quad N = \frac{L^2 d^{2+\frac{5}{2}r+2\mathfrak{e}}}{\varepsilon_n^2} \max\left(n^{1+r}, n^r d^{1+r}\varepsilon_n^{-1}, \frac{d^{1+\frac{3}{2}r}}{(n\varepsilon_n)^2}\right).$$

ii) *For a given arbitrary small* $\mathfrak{e}$, $\gamma = \varepsilon_n \left(\frac{1}{Lnd^{1+r+\frac{\mathfrak{e}}{2}}} \wedge \frac{1}{d^{\frac{5}{2}+2r+\mathfrak{e}} n^{r+\frac{\mathfrak{e}}{2}}}\right)$ *and*

$$N = \max\left(\frac{Ln^{1+r}d^{2+\frac{5}{2}r+\mathfrak{e}}}{\varepsilon_n^3}, \frac{n^{2r+\frac{\mathfrak{e}}{2}} d^{\frac{7}{2}(1+r)+\frac{3}{2}\mathfrak{e}}}{\varepsilon_n^3}, \frac{n^{r-2} d^{\frac{7}{2}+4r+2\mathfrak{e}}}{\varepsilon_n^3}, \frac{L^2 d^{2+2r+\mathfrak{e}}}{\varepsilon_n^2}\right),$$

under the additional assumption $\sup_{\xi\in\mathbb{R}^q}\|D^3U(\xi,.)\|_\infty = \mathcal{O}_{id}(1)$.

Note that in the above result, the constant behind "$\lesssim_{id}$" depends on $\mathfrak{e}$ (this explains that we do not state the result for $\mathfrak{e}=0$). In $ii)$, we assume that the third order derivatives are upper bounded uniformly and independently from d. Instead of this assumption, we could have equivalently written that the constant behind the "$\mathcal{O}_{id}$" in the choice of γ and N depends on $\sup_{\xi\in\mathbb{R}^q}\|D^3U(\xi,.)\|_\infty$ (Note that this dependency is made precise in Theorem 6.2 which is the cornerstone of the above theorem). The next result is a straightforward consequence of Theorem 2.9 in the standard situation where $\alpha_c=1$ (*i.e.* with a sharp separation of distributions around $\theta^\star$) and when, once again, $\mathfrak{b}_d\lesssim_{id} d$ (on this point, see Remark 4).

Corollary 2.10. *Under the assumptions of Theorem 2.9 with* $\alpha_c=1$*, the mean-squared error satisfies:* $\mathbf{E}[|\widehat{\theta}^{\gamma}_{n,0,t_N}-\theta^\star|^2]\lesssim_{id}\frac{d\log(n)}{n}$ *when, "up to*[3] *the parameters* L*,* $\mathfrak{e}$ *and* $\log n$*,"*

i) $\gamma = n^{-\frac{3}{2}}d^{-(r+\frac{1}{2})}\wedge n^{-1}d^{-(1+2r)}$ *and* $N = nd^{1+3r}\max\left(n^{1+\frac{r}{2}}d^{-\frac{1}{2}}, n^{\frac{1}{2}+r}d^{\frac{1+r}{2}}, d^r n^{-1}\right)$.

ii) If $\gamma = n^{-\frac{3}{2}}d^{-(\frac{1}{2}+r)}\wedge n^{-(\frac{1}{2}+r)}d^{-2(1+r)}$ *and*

$$N = \max\left(n^{\frac{3}{2}+2r}d^{2+\frac{7}{2}r}, n^{\frac{5}{2}+r}d^{\frac{1}{2}+\frac{5}{2}r}\right),$$

under the additional assumption $\sup_{\xi\in\mathbb{R}^q}\|D^3U(\xi,.)\|_\infty = \mathcal{O}_{id}(1)$.

Remark 3. The main tool for this result is Theorem 6.2, where we obtain for a given weakly convex potential W and its associated Gibbs distribution $\pi = e^{-W}$, some L^2-bounds for the distance between the Cesaro average of the Euler scheme between $\pi(\mathbf{I}_d)$. As mentioned in the introduction, this result relies on some bounds on the solution of the Poisson equation, derived from a sharp study of the tangent process of the diffusion.

Remark 4. Let us comment the assumptions with $W(x) = (1+|x|^2)^p$ with $p\in(1/2,1]$ (especially $\mathfrak{b}_d\lesssim_{id} d$ which may appear mysterious). First, we shall verify that

$$\nabla W(x) = 2pxW(x)^{(p-1)/p} \quad\text{and}\quad (\nabla^2W(x))_{ij} = 4p(p-1)x_ix_jW(x)^{(p-2)/p}+2p\delta_{ij}W(x)^{(p-1)/p}.$$

Thus, $(\mathbf{H}_{\mathfrak{c}_1})$ holds with $\mathfrak{c}_1 = 4p^2$. Moreover, for any vector u with $|u|=1$,

$$\langle\nabla^2W(x)u,u\rangle = 2p(1+|x|^2)^{p-1}\left(1-2(1-p)\frac{\langle x,u\rangle^2}{1+|x|^2}\right),$$

so that

$$\underline{\lambda}_{\nabla^2W}(x)\geqslant 2p(1-2(1-p))(1+|x|^2)^{p-1} = \mathfrak{c}_2(p)W(x)^{-\frac{1-p}{p}}.$$

This entails $(\mathbf{H}_{\mathfrak{c}_2,r})$ with $\mathfrak{c}_2 = 2p(1-2(1-p))$ and $r=(1-p)/p$. Finally, for $\gamma_0\leqslant(4p)^{-1}$,

$$\bar{\lambda}_{\text{loc}}(\gamma_0,x)\leqslant 2p(1+(\{\frac{1}{2}|x|-2\}\vee 0)^2)^{p-1}\leqslant c_pW(x)^{-\frac{1-p}{p}},$$

where c_p is a constant independent of the dimension. Easy computations lead to

$$\mathcal{C}_{c,\gamma_0}\subset\{x\in\mathbb{R}^d, W(x)\lesssim_{id} d\},$$

so that $\log(\beta(\lambda,c,\gamma_0))\lesssim_{id} d+\log(1+cdL)\lesssim_{id} d$. In particular, this implies that:

$$\mathfrak{b}_d\lesssim_{id} d.$$

[3] For the sake of readability, we choose to provide parameters γ and N without taking into account the parameters $\log n$-terms and setting $\mathfrak{e}=0$. This means that the expressions of γ and N are correct up to some multiplicative terms of the $\log n$ or $d^{\mathfrak{e}}$ order (with arbitrary small $\mathfrak{e}$).

The assumption $\mathfrak{b}_d \lesssim_{id} d$ is always satisfied for this class of (benchmark) potentials and it seems that $\mathfrak{b}_d$ does not depend on p. This suggests that this property may be more universal.

2.5. **Discussion and related results .** The contributions of our paper deserve several comments and have to be situated regarding the large state of the art on Bayesian consistency and Bayesian computation. First, we emphasize that our results apply in log-concave situations included the weakly convex case. and we only need some reasonnable assumptions on U that are discussed in the next paragraphs.

2.5.1. *Bayesian consistency.*

Nature of the result. Theorem 2.3 describes the behaviour of the posterior *mean* $\int \theta \mathrm{d}\pi_n(\theta)$ and not the entire posterior distribution. In a sense, such a result seems less informative than the knowledge of the behaviour of the entire distribution π_n. However, a good behaviour of the posterior mean requires a sharp control of the tails of the posterior distributions whereas a "contraction rate" relative to the Hellinger distance (or with other distances such as the Kullback-Leibler or total variation ones) sometimes blurs the tail behaviour of the posterior. We refer to [CvdV12] for a meaningful illustration in high dimensional linear models of the efforts needed to extend posterior concentration to posterior mean consistency.

Assumption $(\mathbf{A}_L)$. To obtain a Bayesian posterior contraction rate, we introduced the (smoothness) Assumption $(\mathbf{A}_L)$: the function $U(\xi, .)$ must have a Lipschitz gradient, uniformly in ξ. Such an assumption is standard in the optimization community (see [Nes04, Bub15]) and essentially enables to quantify the error made when using a first order Taylor expansion. In optimization theory, L-smooth functions are then used to produce descent lemma. For us, it is instead used for lower-bounding the normalizing constants of the Bayesian posterior distributions (see Equation (27)). This assumption also appears in the recent contribution [MHW$^+$19] but the L-smooth property is only assumed for the function $\theta \longmapsto \mathbb{E}_{\theta^\star}[U(\xi,\theta)]$ associated with either a strong or weak convex assumption on U on U (Assumptions (S.1) or (W.1) of [MHW$^+$19]).

Assumption $(\mathbf{I}_{\mathbf{W_1}}(\mathbf{c}))$. We naturally introduced a separation assumption that is related to the ability of hypothesis testing in the statistical model. Statistical test has a longstanding history in Bayesian literature (see *e.g.* [LC86, Bir83, GGvdV00, CvdV12] among others). In general, the former papers build some global statistical tests using metric considerations with covering arguments on the statistical models with the help of the Hellinger distance or the Kullback-Leibler divergence. Here, our assumption is related to a separation with the help of the Wasserstein 1 distance over $\mathbb{P}_\theta$ and the function Ψ involved in $(\mathbf{I}_{\mathbf{W_1}}(\mathbf{c}))$ is used to build a global test. It is straightforward to verify that $(\mathbf{I}_{\mathbf{W_1}}(\mathbf{c}))$ is satisfied in the location models with $\Psi = \mathbf{I}_d$ since in that case $|\pi_{\theta_1}(\mathbf{I}_d) - \pi_{\theta_2}(\mathbf{I}_d)| = |\theta_1 - \theta_2|$.

In a sense, a link exists between the conjonction of $(\mathbf{A}_L)+(\mathbf{I}_{\mathbf{W_1}}(\mathbf{c}))$ and a metric complexity (in terms of covering numbers) as used in the seminal contribution [GGvdV00][Equation (2.2)]. In particular, it is a straightforward exercise to prove that if $N(\varepsilon, \theta \cap K, d_{KL})$ is the covering number of the statistical model with the Kullback-Leibler divergence and if K is a compact subset of $\mathbb{R}^d$, then

$$\log N(\varepsilon, \theta \cap K, d_{KL}) \lesssim d \log(\sqrt{L}\varepsilon^{-1}).$$

Hence, $(\mathbf{I}_{\mathbf{W_1}}(\mathbf{c}))$ shall be thought of as a way to both "compactify" the space where θ is living and make a local link between $d(\mathbb{P}_\theta, \mathbb{P}_{\theta^\star})$ and $|\theta - \theta^\star|$. This is useful to avoid sieve considerations (see *i.e.* [GGvdV00, vdVvZ08, SW01] for example) and this allows to quantify the tail behaviour of the posterior distribution π_n far away from $\theta^\star$ (see *e.g.* [CvdV12]).

Finally, ($\mathbf{I_{W_1}(c)}$) also has a tight link with Assumption ($W.1$) of [MHW⁺19], stated below:

Assumption 2.8 (Assumption ($W.1$)-[MHW⁺19])**.** *If $F(\theta) := \mathbb{E}_{\theta^\star}[\log p_\theta(\xi)]$, we have:*

$$\langle \nabla F(\theta), \theta^\star - \theta \rangle \geqslant h(|\theta - \theta^\star|),$$

where h is a non-decreasing convex function such that $h(0) = 0$.

For any $\theta \in \mathbb{R}^d$, we introduce $\theta_t = \theta^\star + t(\theta - \theta^\star)$ and $f_\theta(t) = KL(\mathbb{P}_{\theta_t}, \mathbb{P}_{\theta^\star})$. A straightforward computations show that Assumption ($W.1$) implies that:

$$\begin{aligned} KL(\mathbb{P}_\theta, \mathbb{P}_{\theta^\star}) &= f_\theta(1) = f_\theta(0) + \int_0^1 f'_\theta(s)\mathrm{d}s \\ &= \int_0^1 \langle \nabla F(\theta_s), \theta^\star - \theta \rangle \mathrm{d}s = \int_0^1 \frac{\langle \nabla F(\theta_s), \theta^\star - \theta_s \rangle}{s}\mathrm{d}s \geqslant \int_0^1 \frac{h(s|\theta - \theta^\star|)}{s}\mathrm{d}s. \end{aligned}$$

Hence, if h is smooth and $h'(0) > 0$, then the convexity of h yields:

$$KL(\mathbb{P}_\theta, \mathbb{P}_{\theta^\star}) \gtrsim h'(0)|\theta - \theta^\star|.$$

Oppositely, if $h'(0) = 0$, it is reasonable to assume that h is β-Hölder around $\theta^\star$ with $\beta > 1$ and we obtain again a local inequality of the form $KL(\mathbb{P}_\theta, \mathbb{P}_{\theta^\star}) \gtrsim |\theta - \theta^\star|^\beta$ near $\theta^\star$. In the meantime, since h is non-decreasing, it is also possible to use $\int_0^1 \frac{h(s|\theta-\theta^\star|)}{s}\mathrm{d}s \geqslant \int_{1/2}^1 \frac{h(s|\theta-\theta^\star|)}{s}\mathrm{d}s \geqslant h\left(\frac{|\theta-\theta^\star|}{2}\right)$. Therefore, ($W.1$) implies that a $\beta > 1$ exists such that:

$$KL(\mathbb{P}_\theta, \mathbb{P}_{\theta^\star}) \gtrsim |\theta - \theta^\star|^\beta \wedge h\left(\frac{|\theta - \theta^\star|}{2}\right).$$

The link between ($\mathbf{I_{W_1}(c)}$) and ($W.1$) is then made with the help of functional inequalities: if $\mathbb{P}_{\theta^\star}$ is strongly log-concave, then $\mathbb{P}_{\theta^\star}$ satisfies the T_1 inequality (see *e.g.* [Led01]) and $W_1(\mathbb{P}_\theta, \mathbb{P}_{\theta^\star}) \lesssim \sqrt{KL((\mathbb{P}_\theta, \mathbb{P}_{\theta^\star})}$. More generally, if $P_{\theta^\star}$ satisfies a sub-Gaussian concentration inequality, [BG99] shows that this Talagrand inequality still holds. Hence, ($\mathbf{I_{W_1}(c)}$) also implies a lower bound on the KL-divergence in many log-concave reasonnable situations.

Assumption ($\mathbf{PI_U}$). At last, we also needed a uniform upper bound on the Poincaré constant involved in the family of distributions $(\pi_\theta)_{\theta\in\theta}$ to obtain a common constants in the concentration inequalities. Note that, among other settings, Assumption ($\mathbf{PI_U}$) holds true for any location model since, in this case, the Poincaré constant of each distribution π_θ is independent of θ, as indicated by the next proposition:

Proposition 2.11. *If $U(\xi, \theta) = U(\xi - \theta)$, for all ξ and θ, then $C_P(\pi_\theta) = C_P(\pi)$ where $C_P(\pi)$ stands for the Poincaré constant related to $\pi \propto e^{-U}$.*

We also point out that this assumption may be verified in a more general setting using the upper bound of [KLS95] (see the statement in Theorem 2.2). Finally, the Poincaré inequality is satisfied as soon as the log-concave distribution has a second order moment. We observe here that C_P^U may include a dimensional effect even though it is clear that it is not the case for strongly log-concave probability distributions with the help of the Bakry-Emery result (see [BE83]). If we believe in the *Kannan-Lovász-Simonovits conjecture* [KLS95], then the constant C_P^U may be considered in our model as independent from the dimension d, which entails a correct minimax dependency of the Bayesian strategy with respect to d.

Consistency rate. If we now pay attention to the convergence rate obtained in Theorem 2.3, we emphasize that when the separation is sharp, *i.e.* when $\alpha_c = 1$ in $(\mathbf{I_{W_1}(c)})$, we obtain the standard minimax convergence rate $\frac{d}{n}$ up to a $\log(n)$ term, and this result is in a sense non-asymptotic (a universal constant could be exhibited with a price of a huge technicalicity). In comparison with [MHW$^+$19], we also obtained a slightly better convergence rate of $(d/n)^{1/\alpha_c}$ (see Corollary 1 of [MHW$^+$19]). But in general situations, their value of β is equal to 1 so that the rates derived from Theorem 2.3 and Corollary 1 of [MHW$^+$19] are equivalent.

2.5.2. *Ergodicity and Cesaro averaging.*

Convex and strongly convex function. The simulation of the posterior distribution is, in our work, provided by the Langevin diffusion with the potential $\widetilde{W}_n$, which is a convenient alternative approach to Metropolis-Hastings Markov chain strategy. In Theorems 2.4 and 2.5, we only used a Poincaré inequality instead of a strongly log-concave settings, as it is dealt with in [Dal17, MCJ$^+$19] for the Langevin diffusion or in [DCWY18] for the MCMC algorithm. We emphasize that we still derive some explicit exponential bounds of the semi-group and we discuss below some key underlying ingredients.

L^2-distance. Our bounds are obtained under the L^2-norm, *i.e.* we manage exactly the mean-squared error related to the estimation of the posterior mean. Let us remark that, our results could be extended to a larger class of functions, for instance Lipschitz ones, which would lead us to Wasserstein 1 bounds between the occupation measure of the Euler scheme and the target measure π_n. In the weakly convex setting, this extension would imply other constants. More precisely, in Proposition 6.3, which is a crucial step of the proof in the weakly convex case, the fact that g is the solution of the Poisson equation related to $f = \mathbf{I}_d$ plays a central role in the constants. We emphasize that in our setting involving averages of Dirac measures, total variation distance or Kullback-Leibler divergence do not make sense. Furthemore, our objective being to integrate the identity function, such distances seem inapropriate to this problem.

Shifted Cesaro average. We finally point out that we introduce in our work a slight modification that consists in shifting from $[0, t]$ to $[t_0, t]$ the time origin for computing the Cesaro average. The main reason is that the approximation $\widehat{\theta}_{n,t}$ with a Cesaro averaging over $[0, t]$ forgets the (wrong) initialization $\sqrt{J_{\mu,0}}$ of the Langevin diffusion with a speed t^{-2}. In comparison, the shifted Cesaro average also forgets the initialization at the same speed but if we begin the averaging procedure at time t_0, then we benefit from the exponential ergodicity of the process, which entails an initial error in our approximation of the order $\sqrt{J_{\mu,0}}e^{-\lambda_{1,n}t_0}$.

2.5.3. *Warm start and continuous complexity cost.* Warm start has been reported to significantly improve the computational cost of Bayesian strategy in earlier works. We refer to [Dal17] for an extended discussion. In our paper, this is translated by $J_{\mu,0}$ introduced in Theorem 2.4 and we provide a strategy in Proposition 3.3 that gives an exponential scaling with d: we recover here the dependency described in Section 3.2, condition (6) of [DM19]. This is due to the very nature of the L^2 distance that poorly scales with d, contrary to the Kullback-Leibler divergence (see *e.g.* Lemma 7 of [MCJ$^+$19]).

Nevertheless, we emphasize that $J_{\mu,0}$ is only important for *continuous time* estimators, which hold under very general assumptions: Theorem 2.4 may be stated only under the existence of a Poincaré constant $\lambda_{1,n}^{-1}$ without the convex settings. The complexity, as reported in Corollary 2.7, is $\mathcal{O}(\varepsilon_n^2 d[\log(d) + \log(n)])$: it corresponds to a standard sampling procedure that needs $\mathcal{O}(\lambda^{-1} d \log(d/\varepsilon))$ to obtain an ε accuracy, but this formulation is misleading: the spectral gap λ seriously improves the amount of time needed to sample π_n and it is remarkable to observe that the number of observations positively influences the final horizon of simulation.

Regarding now the convex framework and especially the discretization, we point out that warm start is not so important as stated in Theorem 2.8 and Theorem 2.9. For example, in the strongly convex case, it is possible to completely anneal some "far" initialization of the discretization (see Theorem 5.1, *iii*) below) using exponential convergence of coupled trajectories.

2.5.4. *Discretization.* Overall, the leading take-home message when considering the discrete approximation and the concrete estimator is that in both strongly and weakly convex case, we obtain a complexity that evolves as a polynomial of n and d, the complexity being much lower in the strongly convex case. It is also important to notice that the continuous approach completely blurs the real complexity of an effective approximation (the statement of Corollary 2.7 has nothing in common with the ones of Theorem 2.8 or Theorem 2.9). All the more, we observe that in our results, the complexity of Bayesian learning is seriously damaged with the loss of strong convexity, both in terms of n and d.

Second, as an intermediary step, we obtained the complexity to compute an ε-approximation of the posterior mean such that the M.S.E. becomes smaller than ε^2. This is the purpose of Theorems 5.1 and 6.2 (see also Corollary 6.4), respectively in the (uniformly) strongly convex and weakly convex cases. The related orders of complexity are given in Table 1, where we also draw some comparisons with some state of the art results related to the complexity of Bayesian sampling, but up to a normalization factor: actually, in the recent papers [DMM19, DRDK19, MCC$^+$19] that we compare with, the complexity is defined in a slightly different way, omitting the Monte-Carlo factor. More precisely, oppositely to our paper based on a Cesaro average (involving only one path), these papers use a classical Monte-Carlo approach to approximate $\pi(f)$ (for a given function f) by $N_{MC}^{-1}(Z_1 + \ldots + Z_{N_{MC}})$ where $(Z_j)_{1\leqslant j\leqslant N_{MC}}$ denotes an *i.i.d.* sequence of N_{MC} i.i.d. random variables. Then, for a given ε, these papers define the complexity as the number n_ε of iterations of the Euler-scheme to compute Z_1. In order to draw some fair comparisons, we need to consider the "true" complexity, *i.e.* to multiply their complexity n_ε by $MC(\varepsilon) = \mathrm{Var}(Z_1)\varepsilon^{-2}$, *i.e.* by the number of Monte-Carlo simulations that are necessary to obtain a Monte-Carlo M.S.E lower than ε^2. Furthermore, since the involved function is $\mathbf{I}_d$, it is reasonable to assume that $\mathrm{Var}(Z_1)\propto d$, so that we assume that the true complexity of the compared papers is $n_\varepsilon d\varepsilon^{-2}$. Finally, these papers state results with different distances: Total Variation, Kullback-Leibler, W_1 or W_2. We only consider W_1 or W_2 results, which seem to be the only ones that can apply to the non-bounded (Lipschitz) function $\mathbf{I}_d$.

3. Bayesian posterior mean consistency

This paragraph is dedicated to the proof of Theorem 2.3.

3.1. **Poincare inequality and consequences.** We state a famous result for the family $(\mathbb{P}_\theta)_{\theta\in\mathbb{R}^d}$ of Bobkov and Ledoux (see *e.g.* [BL97]), borrowed in [Led01][4].

[4]In [BL97], the authors assume that the function f is bounded. However, when the concentration function $\frac{\delta^2}{4Ck} \wedge \frac{\delta}{2\sqrt{Ck}}$ goes to ∞ when $\delta \to \infty$, the boundedness assumption can be removed (see Proposition 1.7 in [Led01] for details).

	$(\mathbf{SC}_\rho)$ and $(\mathbf{A}_L)$			Weakly convex and $(\mathbf{A}_L)$		
	γ	n_ε	$N(\varepsilon)$	γ	n_ε	$N(\varepsilon)$
This work						
$M.S.E \leqslant \varepsilon^2$	$\mathcal{O}(d^{-1}\varepsilon^2)$	-	$\mathcal{O}(d\varepsilon^{-4})$	$\mathcal{O}(d^{-2(1+r)}\varepsilon^2)$	-	$\mathcal{O}_{id}(d^{3+4r}\varepsilon^{-4})$
				$\mathcal{O}(d^{-(5/2+2r)}\varepsilon)$	-	$\mathcal{O}_{id}(d^{9/2+4r}\varepsilon^{-3})$
[DMM19]						
$W_2^2 \leqslant \varepsilon^2$	$\mathcal{O}_{id}(d^{-1}\varepsilon^2)$	$\mathcal{O}_{id}(d\varepsilon^{-2})$	$\mathcal{O}_{id}(d^2\varepsilon^{-4})$			
$KL \leqslant \varepsilon^2$	$\mathcal{O}_{id}(d^{-1}\varepsilon^2)$	$\mathcal{O}_{id}(d\varepsilon^{-2})$	$\mathcal{O}_{id}(d^2\varepsilon^{-4})$	$\mathcal{O}(d^{-1}\varepsilon^{-2})$	$\mathcal{O}_{id}(d\varepsilon^{-4})$	-
[DRDK19]						
$W_1^2 \leqslant \varepsilon^2$	-	-	-	$\mathcal{O}_{id}(d^{-1}\varepsilon^3)$	$\mathcal{O}_{id}(d^2\varepsilon^{-4})$	$\mathcal{O}_{id}(d^3\varepsilon^{-6})$
$W_2^2 \leqslant \varepsilon^2$	-	-	-	$\mathcal{O}_{id}(d^{-1}\varepsilon^4)$	$\mathcal{O}_{id}(d^2\varepsilon^{-6})$	$\mathcal{O}_{id}(d^3\varepsilon^{-8})$
[MCC+19]						
$KL \leqslant \varepsilon^2$	$\mathcal{O}_{id}(d^{-1/2}\varepsilon)$	$\mathcal{O}_{id}(d^{1/2}\varepsilon^{-1})$	$\mathcal{O}_{id}(d^{3/2}\varepsilon^{-3})$			

TABLE 1. Complexity $N(\varepsilon)$ of an ε-approximation with a constant step-size γ of several methods. We skip the effet of ρ and L for the sake of readability.

Proposition 3.1. *Assume* $(\mathbf{PI_U})$, *then for any differentiable k-Lipschitz real function* f*:*

$$\forall\theta\in\theta\quad\forall n\in\mathbb{N}^*\qquad\mathbb{P}_\theta\left(\left|\frac{1}{n}\sum_{i=1}^n f(\xi_i)-\pi_\theta(f)\right|\geqslant\delta\right)\leqslant 2e^{-n\frac{\delta^2}{4k^2C_P^U}\wedge\frac{\delta}{2k\sqrt{C_P^U}}}.$$

We will apply this result for $f=\Psi$ involved in $(\mathbf{I_{W_1}(c)})$ and with $f=\nabla U_\theta$. In particular, using Proposition 3.1, we obtain the following result (see the proof in [GPP20]).

Corollary 3.2. *Let* $(\mathbf{I_{W_1}(c)})$ *hold and denote by* Ψ *the corresponding 1-Lipschitz function. Then,*

i)

$$\forall\theta\in\mathbb{R}^d\qquad\mathbb{P}_\theta\left(\left|\frac{1}{n}\sum_{i=1}^n\Psi(\xi_i)-\mathbb{E}_\theta\Psi(\xi_1)\right|\geqslant\delta\right)\leqslant 2e^{-n\frac{\delta^2}{4C_P^U}\wedge\frac{\delta}{2\sqrt{C_P^U}}}.$$

ii)

$$\forall\theta\in\mathbb{R}^d\qquad\mathbb{P}_\theta\left(\left|\frac{1}{n}\sum_{i=1}^n\nabla_\theta U(\xi_i,\theta)\right|\geqslant\delta\right)\leqslant 2de^{-n\frac{\delta^2}{4L^2C_P^U d}\wedge\frac{\delta}{2L\sqrt{C_P^U d}}}.$$

Corollary 3.2 will be an essential ingredient for the construction of some efficient statistical tests in the family of probability distributions $(\mathbb{P}_\theta)_{\theta\in\mathbb{R}^d}$. This key corollary is used in Section 3.2.1.

3.2. **Consistency rate of the posterior mean.** To study the behavior of $(\widetilde{\theta}_n)_{n\geqslant 0}$ introduced in Equation (4), we adopt the presentation of [CvdV12] and in particular the link between the posterior mean and the posterior distribution. As noticed in [CvdV12], there is an important need to upper bound the tail of the posterior distribution (far from $\theta^\star$). To this end, for a non-negative sequence $(\varepsilon_n)_{n\geqslant 1}$ fixed later on, we introduce the separation radius:

$$r_{a,n}=a\varepsilon_n+r,\tag{22}$$

where a will be a constant picked sufficiently large, and r will vary from 0 to $+\infty$.

3.2.1. *Statistical tests.* Statistical tests have a long standing history in Bayesian literature (see *e.g.* [LC86, GGvdV00]) to obtain consistency results as well as rates of convergence of Bayes procedures. We introduce an appropriate family of tests $(\phi_n^r)_{n\geqslant 1}$ parametrized by $r>0$ (see Equation (22)) and defined for $n\in\mathbb{N}^\star$ by:

$$\phi_n^r\left(\xi^{\mathbf{n}}\right)=\mathbf{1}_{\left|\frac{1}{n}\sum_{i=1}^n\Psi(\xi_i)-\mathbb{E}_{\theta^\star}[\Psi(\Xi)]\right|\geqslant\frac{c(r_{a,n})}{2}}. \tag{23}$$

It is expected that ϕ_n^r is equal to 0 with an overwhelming probability under the null hypothesis $\mathbb{P}_{\theta^\star}$ whereas ϕ_n^r it is equal to 1 w.o.p. under $\mathbb{P}_\theta$ when $|\theta-\theta^\star|$ is large enough thanks to $(\mathbf{I_{W_1}(c)})$. We prove the following estimations in [GPP20] of the first and second type error of $(\phi_n^r)_{n\geqslant 1}$.

Proposition 3.3. *The sequence of tests $(\phi_n^r)_{n\geqslant 1}$ satisfies*

i) $\mathbb{P}_{\theta^\star}\left(\phi_n^r\left(\xi^{\mathbf{n}}\right)=1\right)\leqslant 2e^{-n\frac{c(r_{a,n})^2}{16C_P^U}\wedge\frac{c(r_{a,n})}{4\sqrt{C_P^U}}}$,

ii) $\sup_{\theta\,:\,|\theta-\theta^\star|\geqslant r_{a,n}}\mathbb{P}_\theta\left(\phi_n^r\left(\xi^{\mathbf{n}}\right)=0\right)\leqslant 2e^{-n\frac{c(r_{a,n})^2}{16C_P^U}\wedge\frac{c(r_{a,n})}{4\sqrt{C_P^U}}}$.

The next result is a technical estimation related to the denominator (normalizing constant) involved in the posterior distribution distribution. A simple application of Corollary 3.2 yields the next result.

Lemma 3.4. *For any $t>0$, define the sets $\mathcal{A}_n^t=\left\{\left|\frac{1}{n}\sum_{i=1}^n\nabla_\theta U(\xi_i,\theta^\star)\right|\leqslant t\right\}$, then:*

$$\forall n\in\mathbb{N}^\star\quad\forall t>0\qquad\mathbb{P}_{\theta^\star}(\{\mathcal{A}_n^t\}^c)\leqslant 2de^{-n\left(\frac{t^2}{4L^2C_P^Ud}\wedge\frac{t}{2L\sqrt{C_P^U}}\right)}.$$

3.2.2. *Proof of the posterior mean consistency .*

Proof of Theorem 2.3. Our proof adopts the strategy of [CvdV12].
Step 1: Decomposition of the quadratic risk. We remark that for all $n\in\mathbb{N}^\star$:

$$\begin{aligned}
\mathbb{E}_{\theta^\star}\left[\left|\widetilde{\theta}_n-\theta^\star\right|^p\right]&=\mathbb{E}_{\theta^\star}\left[\left|\int_{\mathbb{R}^d}(\theta-\theta^\star)\mathrm{d}\pi_n(\theta)\right|^p\right]\\
&\leqslant\mathbb{E}_{\theta^\star}\left[\int_{\mathbb{R}^d}|\theta-\theta^\star|^p\,\mathrm{d}\pi_n(\theta)\right]\\
&=p\mathbb{E}_{\theta^\star}\left[\int_0^\infty t^{p-1}\pi_n(|\theta-\theta^\star|\geqslant t)\mathrm{d}t\right]\\
&=p\mathbb{E}_{\theta^\star}\left[\int_0^{a\varepsilon_n}t^{p-1}\pi_n(|\theta-\theta^\star|\geqslant t)\mathrm{d}t+\int_{a\varepsilon_n}^\infty t^{p-1}\pi_n(|\theta-\theta^\star|\geqslant t)\mathrm{d}t\right]\\
&\leqslant a^p\varepsilon_n^p+p\int_0^{+\infty}r_{a,n}^{p-1}\mathbb{E}_{\theta^\star}\left[\pi_n(|\theta-\theta^\star|\geqslant r_{a,n})\right]\mathrm{d}r,
\end{aligned}\tag{24}$$

where we used the Jensen inequality in the second line, an integration by part in the third line, a direct integration $\int_0^{a\varepsilon_n}pt^{p-1}\mathrm{d}t=a^p\varepsilon_n^p$ in the last line associated with the Fubini relationship.
Step 2: Use of the tests $(\phi_n^r)_{n\geqslant 1}$. We now use the tests $(\phi_n^r)_{n\geqslant 1}$ and the sets $(\mathcal{A}_n^t)$, we can write:

$$\begin{aligned}
\mathbb{E}_{\theta^\star}\left[\pi_n(|\theta-\theta^\star|\geqslant r_{a,n})\right]=&\,\mathbb{E}_{\theta^\star}\left[\phi_n^r(\xi^{\mathbf{n}})\pi_n(|\theta-\theta^\star|\geqslant r_{a,n})\right]\\
&+\mathbb{E}_{\theta^\star}\left[(1-\phi_n^r(\xi^{\mathbf{n}}))\pi_n(|\theta-\theta^\star|\geqslant r_{a,n})\mathbf{1}_{\mathcal{A}_n^t}\right]\\
&+\mathbb{E}_{\theta^\star}\left[(1-\phi_n^r(\xi^{\mathbf{n}}))\pi_n(|\theta-\theta^\star|\geqslant r_{a,n})\mathbf{1}_{\{\mathcal{A}_n^t\}^c}\right].
\end{aligned}\tag{25}$$

From this expression we can deduce the following inequality:

$$\begin{aligned}
\mathbb{E}_{\theta^\star}\left[\pi_n(|\theta-\theta^\star|\geqslant r_{a,n})\right]\leqslant&\,\mathbb{E}_{\theta^\star}\left[\phi_n^r(\xi^{\mathbf{n}})\right]+\mathbb{E}_{\theta^\star}\left[(1-\phi_n^r(\xi^{\mathbf{n}}))\pi_n(|\theta-\theta^\star|\geqslant r_{a,n})\mathbf{1}_{\mathcal{A}_n^t}\right]\\
&+\mathbb{P}_{\theta^\star}(\{\mathcal{A}_n^t\}^c).
\end{aligned}$$

• Study of $\mathbb{E}_{\theta^\star}\left[(1-\phi_n^r(\xi^{\mathbf{n}}))\pi_n(|\theta-\theta^\star|\geqslant r_{a,n})\mathbf{1}_{\mathcal{A}_n^t}\right]$. We write that:

$$\pi_n(|\theta-\theta^\star|\geqslant r_{a,n})=\int_{\theta\,:\,|\theta-\theta^\star|\geqslant r_{a,n}}\mathrm{d}\pi_n(\theta)=\frac{\displaystyle\int_{\theta\,:\,|\theta-\theta^\star|\geqslant r_{a,n}}\frac{e^{-W_n(\xi^{\mathbf{n}},\theta)}}{e^{-W_n(\xi^{\mathbf{n}},\theta^\star)}}\mathrm{d}\pi_0(\theta)}{\displaystyle\int_{\mathbb{R}^d}\frac{e^{-W_n(\xi^{\mathbf{n}},\theta)}}{e^{-W_n(\xi^{\mathbf{n}},\theta^\star)}}\mathrm{d}\pi_0(\theta)}. \tag{26}$$

At this stage we control the denominator and the numerator separately. Let us denote by $Z_t=\pi_0(B(\theta^\star,t))$ the prior mass of the Euclidean ball centered at $\theta^\star$ and of radius t. We have

$$\begin{aligned}\log\left(\int_{\mathbb{R}^d}\frac{e^{-W_n(\xi^{\mathbf{n}},\theta)}}{e^{-W_n(\xi^{\mathbf{n}},\theta^\star)}}d\pi_0(\theta)\right)&\geqslant\log\left(\int_{B(\theta^\star,t)}\frac{e^{-W_n(\xi^{\mathbf{n}},\theta)}}{e^{-W_n(\xi^{\mathbf{n}},\theta^\star)}}\mathrm{d}\pi_0(\theta)\right)\\&\geqslant\log\left(\int_{B(\theta^\star,t)}\frac{e^{-W_n(\xi^{\mathbf{n}},\theta)}}{e^{-W_n(\xi^{\mathbf{n}},\theta^\star)}}\frac{\mathrm{d}\pi_0(\theta)}{Z_t}\right)+\log Z_t\\&\geqslant\int_{B(\theta^\star,t)}\sum_{i=1}^n[U(\xi_i,\theta^\star)-U(\xi_i,\theta)]\frac{\mathrm{d}\pi_0(\theta)}{Z_t}+\log Z_t,\end{aligned}$$

where we used the Jensen inequality in the last line with the concave function log and the normalized measure $\mathrm{d}\pi_0 Z_t^{-1}$ over $B(\theta^\star,t)$. Using that $\nabla_\theta U(\xi,.)$ is L-Lipschitz we get

$$\forall x\in\mathbb{R}^d\qquad U(\xi,\theta_1)-U(\xi,\theta_2)\leqslant\langle\theta_1-\theta_2,\nabla_\theta U(\xi,\theta_2)\rangle+\frac{L}{2}\|\theta_1-\theta_2\|^2,$$

for all $(\theta_1,\theta_2)\in\mathbb{R}^d$, which implies that:

$$\forall i\in\{1,\ldots,n\}\quad\forall\theta\in\mathbb{R}^q\qquad|U(\xi_i,\theta)-U(\xi_i,\theta^\star)-\langle\theta-\theta^\star,\nabla U(\xi_i,\theta^\star)\rangle|\leqslant\frac{L}{2}|\theta-\theta^\star|^2.$$

Using a sum over i and the triangle inequality, we then deduce that:

$$\forall\theta\in\mathbb{R}^d\quad\sum_{i=1}^n U(\xi_i,\theta^\star)-U(\xi_i,\theta)\geqslant-\left|\left\langle\theta-\theta^\star,\sum_{i=1}^n\nabla_\theta U(\xi_i,\theta^\star)\right\rangle\right|-n\frac{L}{2}|\theta-\theta^\star|^2.$$

The Cauchy-Schwarz inequality yields:

$$\left|\left\langle\theta-\theta^\star,\sum_{i=1}^n\nabla_\theta U(\xi_i,\theta^\star)\right\rangle\right|\leqslant|\theta-\theta^\star|\left|\sum_{i=1}^n\nabla_\theta U(\xi_i,\theta^\star)\right|.$$

An integration over $B(\theta^\star,t)$ with the normalized measure $\pi_0 Z_t^{-1}$ leads to:

$$\begin{aligned}\log\left(\int_{\mathbb{R}^d}\frac{e^{-W_n(\xi^{\mathbf{n}},\theta)}}{e^{-W_n(\xi^{\mathbf{n}},\theta^\star)}}\mathrm{d}\pi_0(\theta)\right)&\geqslant-n\frac{L}{2}t^2\frac{\pi_0(B(\theta^\star,t))}{Z_t}-t\left\|\sum_{i=1}^n\nabla_\theta U(\xi_i,\theta^\star)\right\|+\log(Z_t)\\&\geqslant-n\frac{L}{2}t^2-t\left\|\sum_{i=1}^n\nabla_\theta U(\xi_i,\theta^\star)\right\|+\log(Z_t).\end{aligned}$$

To lower bound the denominator, we use the set $\mathcal{A}_n^t$ and we have

$$\begin{aligned}\log\left(\int_{\mathbb{R}^d}\frac{e^{-W_n(\xi^{\mathbf{n}},\theta)}}{e^{-W_n(\xi^{\mathbf{n}},\theta^\star)}}\mathrm{d}\pi_0(\theta)\right)\mathbf{1}_{\mathcal{A}_n^t}&\geqslant\left(-n\frac{L}{2}t^2-t\left\|\sum_{i=1}^n\nabla_\theta U(\xi_i,\theta^\star)\right\|+\log(Z_t)\right)\mathbf{1}_{\mathcal{A}_n^t}\\&\geqslant\left(-nt^2\left(\frac{L}{2}+1\right)+\log(Z_t)\right)\mathbf{1}_{\mathcal{A}_n^t}.\end{aligned}\tag{27}$$

Using (27) with (26) and the Jensen Inequality, we have

$$\begin{aligned}
&\mathbb{E}_{\theta^\star}\left((1-\phi_n^r(\xi^{\mathbf{n}}))\pi_n(|\theta-\theta^\star|\geqslant r_{a,n})\mathbf{1}_{\mathcal{A}_n^t}\right)\\
&\leqslant \frac{\mathbb{E}_{\theta^\star}\left[(1-\phi_n^r(\xi^{\mathbf{n}}))\int_{\theta:|\theta-\theta^\star|\geqslant r_{a,n}}\frac{e^{-W_n(\xi^{\mathbf{n}},\theta)}}{e^{-W_n(\xi^{\mathbf{n}},\theta^\star)}}d\pi_0(\theta)\right]}{Z_t e^{-n\left(\frac{L}{2}+1\right)t^2}}\\
&\leqslant \int_{\theta:|\theta-\theta^\star|\geqslant r_{a,n}}\mathbb{E}_{\theta^\star}\left[(1-\phi_n^r(X))\frac{e^{-W_n(\xi^{\mathbf{n}},\theta)}}{e^{-W_n(\xi^{\mathbf{n}},\theta^\star)}}\right]d\pi_0(\theta)e^{nt^2\left(\frac{L}{2}+1\right)}(Z_t)^{-1}\\
&\leqslant (Z_t)^{-1}e^{nt^2\left(\frac{L}{2}+1\right)}\sup_{\{\theta\,:|\theta-\theta^\star|\geqslant r_{a,n}\}}\mathbb{E}_\theta\left[1-\phi_n^r(\xi^{\mathbf{n}})\right]\\
&\leqslant 2e^{nt^2\left(\frac{L}{2}+1\right)-\log\pi_0(B(\theta^\star,t))}e^{-n\frac{c(r_{a,n})^2}{16C_P^U}\wedge\frac{c(r_{a,n})}{4\sqrt{C_P^U}}},
\end{aligned}$$

where in the penultimate line we used a change of measure from $\mathbb{P}_{\theta^\star}$ to $\mathbb{P}_\theta$.

• Study of $\mathbb{E}_{\theta^\star}[\phi_n^r(\xi^{\mathbf{n}})]$. Using the first type error given by $i)$ of Proposition 3.3, we have:

$$\mathbb{E}_{\theta^\star}[\phi_n^r(\xi^{\mathbf{n}})]\leqslant 2e^{-n\frac{c(r_{a,n})^2}{16C_P^U}\wedge\frac{c(r_{a,n})}{4\sqrt{C_P^U}}}.$$

• Study of $\mathbb{E}_{\theta^\star}\left[(1-\phi_n^r(\xi^{\mathbf{n}}))\pi_n(|\theta-\theta^\star|\geqslant r_{a,n})\mathbf{1}_{\{\mathcal{A}_n^t\}^c}\right]$. We upper bound $\mathbb{P}_{\theta^\star}(\{\mathcal{A}_n^t\}^c)$ with $t=t_{r,n}$ and apply Lemma 3.4. We deduce that:

$$\mathbb{P}_{\theta^\star}(\{\mathcal{A}_n^t\}^c)\leqslant 2de^{-n\left(\frac{t_{r,n}^2}{4L^2C_P^Ud}\wedge\frac{t_{r,n}}{2L\sqrt{C_P^Ud}}\right)}.$$

We then obtain that:

$$\begin{aligned}
&\mathbb{E}_{\theta^\star}\left[\pi_n(|\theta-\theta^\star|\geqslant r_{a,n})\right]\\
&\leqslant 4e^{nt_{r,n}^2\left(\frac{L}{2}+1\right)-\log\pi_0(B(\theta^\star,t_{r,n}))}e^{-n\frac{c(r_{a,n})^2}{16C_P^U}\wedge\frac{c(r_{a,n})}{4\sqrt{C_P^U}}}+2de^{-n\left(\frac{t_{r,n}^2}{4L^2C_P^Ud}\wedge\frac{t_{r,n}}{2L\sqrt{C_P^Ud}}\right)}.
\end{aligned}$$

Step 3: Small ball calibration and prior mass We shall now adjust the different parameters in order to obtain the best possible rate for $(\widetilde{\theta}_n)_{n\geqslant 0}$. We choose $t_{r,n}$ according to

$$t_{r,n}=\frac{c(r_{a,n})\wedge\sqrt{c(r_{a,n})}}{A},$$

with A sufficiently large such that:

$$nt_{r,n}^2\left(\frac{L}{2}+1\right)-n\frac{c(r_{a,n})^2}{16C_P^U}\wedge\frac{c(r_{a,n})}{4\sqrt{C_P^U}}\leqslant -n\frac{c(r_{a,n})^2}{32C_P^U}\wedge\frac{c(r_{a,n})}{8\sqrt{C_P^U}}.$$

In the meantime, we get that:

$$-\log\pi_0(B(\theta^\star,t_{r,n}))\leqslant -\log\pi_0\left(B(\theta^\star,A^{-1}[c(a\varepsilon_n)\wedge\sqrt{c(a\varepsilon_n)}])\right).$$

ε_n will be chosen close to 0 and c is increasing so that:

$$-\log\pi_0(B(\theta^\star,t_{r,n}))\leqslant -\log\pi_0\left(B(\theta^\star,A^{-1}c(a\varepsilon_n))\right).$$

Since $\pi_0=e^{-V_0}$ with V_0 a $\mathcal{C}_1^1$ function, we then deduce that:

$$\forall\theta\in B(\theta^\star,\delta)\qquad |V_0(\theta)-V_0(\theta^\star)|\leqslant\delta\|\nabla V(\theta^\star)\|+\frac{1}{2}\delta^2,$$

which implies:

$$\begin{aligned}
&-\log \pi_0(B(\theta^\star, t_{r,n}) \\
&\leqslant -\log\left(\int_{B(\theta^\star, A^{-1}c(a\varepsilon_n))} e^{V_0(\theta^\star)-A^{-1}c(a\varepsilon_n)\|\nabla V(\theta^\star)\|-A^{-2}c(a\varepsilon_n)^2/2}\mathrm{d}\lambda_d(\theta)\right) \\
&= -V_0(\theta^\star) + A^{-1}c(a\varepsilon_n)\|\nabla V(\theta^\star)\| + \frac{A^{-2}c(a\varepsilon_n)^2}{2} + d\log(Ac(a\varepsilon_n)^{-1}) - \log\lambda_d(B(0,1)) \\
&\leqslant -V_0(\theta^\star) + A^{-1}c(a\varepsilon_2)\|\nabla V(\theta^\star)\| + \frac{A^{-2}c(a\varepsilon_2)^2}{2} + d\log(c(a\varepsilon_n)^{-1}).
\end{aligned}$$

Using the behaviour of c near 0 (see $(\mathbf{I_{W_1}}(\mathbf{c}))$), a constant $C_{\theta^\star}$ exists such that:

$$-\log\pi_0(B(\theta^\star, t_{r,n}) \leqslant \log(C_{\theta^\star}) + d\alpha_c\log(\varepsilon_n^{-1}).$$

We then obtain that a universal constant K exists such that:

$$\mathbb{E}_{\theta^\star}\left[\pi_n(|\theta-\theta^\star| \geqslant r_{a,n})\right] \lesssim de^{-\frac{n}{K}\left[\frac{c(r_{a,n})^2}{L^2 C_P^U d}\wedge\frac{c(r_{a,n})}{L\sqrt{C_P^U d}}\right]+d\alpha_c\log(\varepsilon_n^{-1})}.$$

Finally, Equation (13) yields for K large enough:

$$\mathbb{E}_{\theta^\star}\left[\pi_n(|\theta-\theta^\star| \geqslant r_{a,n})\right] \lesssim de^{-\frac{n}{K}\frac{\{r_{a,n}\}^{2\alpha_c}}{L^2C_P^U d}}\mathbf{1}_{r_{a,n}\leqslant 1} + de^{-\frac{n}{K}\frac{\log(r_{a,n})+1}{L\sqrt{C_P^U d}}}\mathbf{1}_{r_{a,n}\geqslant 1}. \tag{28}$$

Step 4: Convergence rate We use (24) and (28) and obtain that:

$$\begin{aligned}
\mathbb{E}_{\theta^\star}[\|\widetilde{\theta}_n-\theta^\star\|^p] &\lesssim (a\varepsilon_n)^p + d\int_0^{+\infty} r_{a,n}^{p-1}\left(e^{-\frac{n}{K}\frac{\{r_{a,n}\}^{2\alpha_c}}{L^2C_P^U d}}\mathbf{1}_{r_{a,n}\leqslant 1} + e^{-\frac{n}{K}\frac{\log(r_{a,n})+1}{L\sqrt{C_P^U d}}}\mathbf{1}_{r_{a,n}\geqslant 1}\right)\mathrm{d}r \\
&\lesssim \varepsilon_n^p + d\left[\int_{a\varepsilon_n}^1 r^{p-1}e^{-\frac{n}{K}\frac{r^{2\alpha_c}}{L^2C_P^U d}}\mathrm{d}r + \int_1^{+\infty} r^{p-1}e^{-\frac{n}{K}\frac{\log(r)+1}{L\sqrt{C_P^U d}}}\mathrm{d}r\right].
\end{aligned}$$

If we choose ε_n such that

$$\varepsilon_n = \left(L^2C_P^U d\frac{\log n}{n}\right)^{1/2\alpha_c},$$

we then observe that

$$\int_{a\varepsilon_n}^{+\infty} r^{p-1}e^{-\frac{n}{K}\frac{r^{2\alpha_c}}{L^2C_P^U d}}\mathrm{d}r = \left(\frac{KL^2C_P^U d}{n}\right)^{p/2\alpha_c}\Gamma\left(\frac{p}{2\alpha_c}; a^{2\alpha_c}\log(n)\right) = \mathcal{O}_{id}(\varepsilon_n^p).$$

The second integral may be made exponentially small (in terms of n). □

We end this paragraph with a rapid discussion on the (random) moments of the posterior distriubtion $\mathbb{M}_{n,2}$ and $\mathbb{M}_{n,4}$ (used later on for the mixing rate of the Langevin strategy):

$$\mathbb{M}_{n,2} = \mathbb{E}_{\pi_n}\left(|\mathbf{I}_d|^2\right) = \pi_n(|\mathbf{I}_d|^2) \qquad \mathbb{M}_{n,4} = \mathbb{E}_{\pi_n}\left(|\mathbf{I}_d|^4\right) = \pi_n(|\mathbf{I}_d|^4). \tag{29}$$

The expected values of these moments can be upper bounded using Theorem 2.3.

Corollary 3.5. *If π_0 is a $\mathcal{C}^1(\mathbb{R}^d,\mathbb{R})$ log-concave prior, if U is $\mathcal{C}^1_L$ and if $(\mathbf{I_{W_1}}(\mathbf{c}))$ holds, then*

$$\mathbf{E}[\mathbb{M}_{n,2}] \lesssim_{id} |\theta^\star|^2 + \{L^2C_P^U d\}^{1/\alpha_c}\left(\frac{\log n}{n}\right)^{1/\alpha_c}$$

and

$$\mathbf{E}[\mathbb{M}_{n,4}] \lesssim_{id} |\theta^\star|^4 + \{L^2C_P^U d\}^{2/\alpha_c}\left(\frac{\log(n)}{n}\right)^{2/\alpha_c}.$$

Proof. We observe that $|\theta|^p \leqslant 2^{p-1}(|\theta-\theta^\star|^p + |\theta^\star|^p)$ so that

$$\begin{aligned}\mathbf{E}[\mathbb{M}_{n,p}] &\leqslant 2^{p-1}|\theta^\star|^p + 2^{p-1}\mathbb{E}_{\theta^\star}\left(\mathbb{E}_{\pi_n}\left(|\theta-\theta^\star|^p\right)\right)\\ &\leqslant 2^{p-1}|\theta^\star|^p + 2^{p-1}\mathbb{E}_{\theta^\star}\left(\int_\theta |\theta-\theta^\star|^p \mathrm{d}\pi_n(\theta)\right).\end{aligned}$$

We then use the argument of Theorem 2.3, which concludes the proof. □

4. L^2-CONVERGENCE OF $(\hat{\theta}_t^{(n)})$ TOWARDS $\widetilde{\theta}_n$

This section is dedicated to the analysis of the (shifted) Cesaro averages (8) and (9) of the *continuous-time* Langevin diffusion defined in (7) for the computation of the posterior mean $(\widetilde{\theta}_n)$, whose dynamical is recalled below:

$$dX_t^{(n)} = -\nabla\widetilde{W}_n(X_t^{(n)})dt + \sqrt{2}dB_t \quad \text{with} \quad \forall x\in\mathbb{R}^d \qquad \widetilde{W}_n(x) = \underbrace{\log(\pi_0^{-1}(x))}_{:=V_0(x)} + \sum_{i=1}^n U(\xi_i, x). \tag{30}$$

In this Section, $\widetilde{W}_n$ is sample dependent and we work with a fixed sample $(\xi_1,\ldots,\xi_n)$, the randomness is then brought by the Brownian motion $(B_t)_{t\geqslant 0}$.

4.1. **Basics.**
Before going further, let us recall some basic properties and notations related to SDE (30) (familiar readers may skip Section 4.1). When $\widetilde{W}_n$ is Lipschitz continuous, existence and uniqueness classically holds for the solution of (30) and $X^{(n)}$ is a Markov process with infinitesimal generator $\mathcal{L}_n$ defined by:

$$\forall f\in\mathcal{C}^2(\mathbb{R}^d,\mathbb{R}) \qquad \mathcal{L}_n f(x) = -\langle\nabla\widetilde{W}_n(x), \nabla f(x)\rangle + \Delta f(x). \tag{31}$$

We denote its semi-group by $(P_t^n)_{t\geqslant 0}$. However, in our setting, $\widetilde{W}_n$ may be only locally Lipschitz continuous and in this case, one has to check that the solutions do not explode in finite time. Such a property is ensured by the convexity assumptions. More precisely, let $\mathcal{I}_2$ refers to $x \longmapsto \|x\|^2$, then a straightforward computation shows that:

$$\mathcal{L}_n\mathcal{I}_2(x) = -2\sum_{i=1}^n \langle x, \nabla_x U(\xi_i, x)\rangle - 2\langle x, \nabla V_0(x)\rangle + 2.$$

The function $U(\xi_i, .)$ being convex and coercive for any ξ_i, we have $\lim_{|x|\longmapsto+\infty}\langle x, \nabla_x U(\xi_i,x)\rangle = +\infty$ whereas since $V_0 = \log(\pi_0^{-1}(.))$ is strongly convex, we have

$$\liminf_{x\longrightarrow+\infty}\frac{\langle x, \nabla V_0(x)\rangle}{|x|^2} > 0.$$

Hence $\mathcal{I}_2$ satisfies a Lyapunov mean-reverting condition: a pair $(\alpha,\beta)\in\mathbb{R}_+^2$ exists such that:

$$\mathcal{L}_n\mathcal{I}_2 \leqslant \beta - \alpha\mathcal{I}_2.$$

Applying standard results (see [EK05, Has02]), we get the non-explosion of the solutions and the existence of an invariant distribution. Combined with the ellipticity of the diffusion, this yields the following result:

Proposition 4.1. *Let V_0 be strongly convex and $x\mapsto U(\xi,x)$ be convex and coercive for every ξ. Then, a.s. in $\xi^{\mathbf{n}}$, Equation (7) admits a unique strong solution $X^{(n)}$ which defines a uniformly elliptic, positive recurrent Markov process whose unique invariant distribution is the posterior distribution π_n.*

4.2. **Ergodicity.** We denote the law of the process at time s, initialized at any point $x \in \mathbb{R}^d$ by $\mathbb{P}_x^{X_s^{(n)}}$. From Proposition 4.1, this law is absolutely continuous with respect to the Lebesgue measure for any finite time $s > 0$, and therefore absolutely continuous with respect to π_n. We shall denote by $m_{\mu,s}^{(n)}$ the corresponding density (given by the Radon Nykodim Theorem) at time s when the process starts from a randomized initial state distributed according to μ:

$$\forall x \in \mathbb{R}^d \qquad m_{\mu,s}^{(n)}(x) = \frac{\mathrm{d}\mathbb{P}_\mu^{X_s^{(n)}}}{\mathrm{d}\pi_n}(x).$$

The asymptotic consistency of $\widehat{\theta}_{n,t}$ is related to the ergodic behavior of $(X_s^{(n)})_{s\geqslant 0}$, *i.e.* the convergence of $\mathbb{P}_x^{X_s^{(n)}}$ towards the unique equilibrium π_n. Such a long time convergence will be asserted in terms of some decrease towards0 of some L^2-norm. To make this discussion more precise we recall a straightforward consequence of Theorem 2.2.

Proposition 4.2. *The measure π_n is log-concave and satisfies a Poincaré inequality:*

$$\exists\, C_{P,n} > 0 \qquad \forall f \in L^2(\pi_n) \qquad Var_{\pi_n}(f) \leqslant C_{P,n}\pi_n(|\nabla f|^2).$$

Below, we will frequently use the following notation

$$\forall t \geqslant 0 \qquad J_{\mu,t} = \|m_{\mu,t}^{(n)} - \mathbf{1}\|^2_{L^2(\pi_n)}. \tag{32}$$

We state the quantitative consequence on the convergence of $\mathbb{P}_x^{X_s^{(n)}}$ borrowed from [BBCG08].

Theorem 4.3 (L^2-ergodicity of $(X_s^{(n)})_{s\geqslant 0}$)**.** *For any $t > 0$:*

$$\forall f \in L^2(\pi_n) \qquad \int_{x\in\mathbb{R}^d} (\mathbb{E}_x[f(X_t^{(n)})] - \pi_n(f))^2 \mathrm{d}\pi_n(x) \leqslant e^{-\frac{2}{C_{P,n}}t} \int_{x\in\mathbb{R}^d} (f(x) - \pi_n(f))^2 \mathrm{d}\pi_n(x).$$

For any $t \geqslant t_0 \geqslant 0$, we have

$$J_{\mu,t} \leqslant e^{-\frac{2}{C_{P,n}}(t-t_0)} J_{\mu,t_0}.$$

Theorem 4.3 assesses an exponential convergence of the distribution $\mathbb{P}_x^{X_s^{(n)}}$ towards π_n in terms of the variance introduced above. This result is translated into a convergence of $m_{\mu,s}^{(n)}$ towards the constant function $\mathbf{1}$ in $\mathbb{L}^2(\pi_n)$. The spectral gap $\lambda_{1,n}$ will refer to the inverse of the Poincaré constant:

$$\lambda_{1,n} = \frac{1}{C_{P,n}}. \tag{33}$$

Remark 5. The constant $C_{P,n}$ (or $\lambda_{1,n}$) will be fundamental to assess the efficiency of the Langevin diffusion for the approximation of $\widetilde{\theta}_n$. We will discuss on the size of the spectral gap below, in particular its dependency with the sample size n and with the dimension d.

4.3. **Convergence of $(\widehat{\theta}_{n,t})_{t\geqslant 0}$ towards $\widetilde{\theta}_n$.** We introduce the cumulative integral:

$$\forall t > 0 \qquad S_{n,t} = \int_0^t X_s^{(n)} \mathrm{d}s \qquad \text{so that} \qquad \widehat{\theta}_{n,t} = \frac{S_{n,t}}{t}.$$

Of course, if $X^{(n)}$ were initialized with the invariant distribution π_n, we would then have

$$\mathbb{E}_{\pi_n}[\widehat{\theta}_{n,t}] = \frac{1}{t}\mathbb{E}_{\pi_n}\left[\int_0^t X_s^{(n)}\mathrm{d}s\right] = \frac{1}{t}\int_0^t \mathbb{E}_{\pi_n}\left[X_s^{(n)}\right]\mathrm{d}s = \widetilde{\theta}_n,$$

where the last equality comes from the invariance property. The measure π_n being unknown, the process $(X_s^{(n)})$ is initialized differently and we need to translate the convergence of $\mathbb{P}_x^{X_s^{(n)}}$ towards π_n.

Below, we will use the Chasles decomposition and write that

$$S_{n,t} = S_{n,t_0} + S_{n,t_0,t} \qquad \text{where} \qquad S_{n,t_0,t} = \int_{t_0}^t X_s^{(n)} \mathrm{d}s.$$

$S_{n,t_0,t}$ will bring the main part of the *variance* of the estimation. The Young inequality yields:

$$\mathbb{E}_\mu\left(\left|\widehat{\theta}_{n,t}-\widetilde{\theta}_n\right|^2\right)\leqslant\frac{2\mathbb{E}_\mu\left(|S_{n,t_0}-t_0\widetilde{\theta}_n|^2\right)+2\mathbb{E}_\mu\left(|S_{n,t_0,t}-(t-t_0)\widetilde{\theta}_n|^2\right)}{t^2}. \tag{34}$$

The estimation of the two terms of the last inequality is based on the use of the convergence of the process towards its invariant distribution. This is the content of the two following proposition whose proof is deferred to the supplementary document [GPP20, Section 2].

Proposition 4.4. *For any $t_0>0$:*

$$\mathbb{E}_\mu\left(|S_{n,t_0}-t_0\widetilde{\theta}_n|^2\right)\leqslant t_0^2\mathbb{V}_{n,2}+t_0\frac{\sqrt{\mathbb{V}_{n,4}}\sqrt{J_{\mu,0}}}{\lambda_{1,n}}.$$

For any $t\geqslant t_0>0$, we have

$$\mathbb{E}_\mu\left(|S_{n,t_0,t}-(t-t_0)\widetilde{\theta}_n|^2\right)\leqslant\frac{2(t-t_0)\mathbb{V}_{n,2}}{\lambda_{1,n}}+\sqrt{\mathbb{V}_{n,4}}(t-t_0)^2e^{-\lambda_{1,n}t_0}\sqrt{J_{\mu,0}}.$$

4.4. **Proof of Theorem 2.4 and Theorem 2.5.** This paragraph is devoted to the computation of the rate of convergence of the Cesaro average $(\widehat{\theta}_{n,t})$ towards $\widetilde{\theta}_n$.

Proof of Theorem 2.4 . Proof of i): Our starting point is the Chasles decomposition associated to the Young inequality (34). For any $t_0>0$, we apply Proposition 4.4. We obtain that:

$$\mathbb{E}_\mu\left(|\widehat{\theta}_{n,t}-\widetilde{\theta}_n|^2\right)\leqslant 2\frac{t_0^2}{t^2}\mathbb{V}_{n,2}+2\frac{t_0\sqrt{\mathbb{V}_{n,4}}\sqrt{J_{\mu,0}}}{\lambda_{1,n}t^2}+\frac{4\mathbb{V}_{n,2}}{t\lambda_{1,n}}+\sqrt{\mathbb{V}_{n,4}}e^{-\lambda_{1,n}t_0}\sqrt{J_{\mu,0}}$$

We choose $t_0=\alpha\lambda_{1,n}^{-1}\log(t)$ and use the Cauchy-Schwarz inequality $\mathbb{V}_{n,2}\leqslant\sqrt{\mathbb{V}_{n,4}}$ and $2\sqrt{2}\sqrt{a+b}\geqslant\sqrt{a}+\sqrt{b}$. We then obtain that:

$$\mathbb{E}_\mu\left(|\widehat{\theta}_{n,t}-\widetilde{\theta}_n|^2\right)\leqslant\sqrt{\mathbb{V}_{n,4}}\left[2\sqrt{2}\frac{\alpha^2\log(t)^2}{\lambda_{1,n}^2t^2}\sqrt{1+J_{\mu,0}}+\frac{4}{t\lambda_{1,n}}+\sqrt{J_{\mu,0}}t^{-\alpha}\right]. \tag{35}$$

□

Proof of Theorem 2.5 . We remark that

$$\widehat{\theta}_{n,t_0,t}-\widetilde{\theta}_n=\frac{1}{t-t_0}\int_{t_0}^t[X_s^{(n)}-\widetilde{\theta}_n]\mathrm{d}s=\frac{S_{n,t_0,t}-(t-t_0)\widetilde{\theta}_n}{t-t_0},$$

so that:

$$\mathbb{E}_\mu\left(|\widehat{\theta}_{n,t_0,t}-\widetilde{\theta}_n|^2\right)=\frac{\mathbb{E}_\mu\left(\left|S_{n,t_0,t}-(t-t_0)\widetilde{\theta}_n\right|^2\right)}{(t-t_0)^2}\leqslant\frac{2\mathbb{V}_{n,2}}{(t-t_0)\lambda_{1,n}}+\sqrt{\mathbb{V}_{n,4}}\sqrt{e^{-2\lambda_{1,n}t_0}J_{\mu,0}},$$

where the last line is a straightforward application of Proposition 4.4. Using the Cauchy-Schwarz inequality, $\mathbb{V}_{n,2}\leqslant\sqrt{\mathbb{V}_{n,4}}$, we obtain the conclusion. □

The proof of Corollary 2.7 is deferred to the supplementary document [GPP20, Section 3].

5. Discretization of the Langevin procedure - Strongly convex case

After a brief reminder, we present a general result of approximation of the Cesaro mean of a trajectory in the strongly convex case. We specify this result in the Bayesian framework to assess the cost of an optimal learning with a discrete LMC and strongly convex models.

5.1. **LMC approximation with strongly convex functions .** In this section, we prepare the proof of Theorem 5.2 by establishing a bound related to LMC-approximation with discretized Cesaro averages for a given strong convex potential W. We thus introduce the following SDE:

$$dX_t = b(X_t)dt + \sqrt{2}dB_t,$$

where $b = -\nabla W$. The related approximation scheme is defined as follows: for a step-size sequence $(\gamma_k)_{k\geqslant 1}$ of positive numbers introduced in Section 1.4 and its associated clock-time sequence $(t_k)_{k\geqslant 1}$, we define the discretization scheme $(\bar{X}_{t_k})_{k\geqslant 0}$ by: $\bar{X}_0 = x \in \mathbb{R}^d$ and for all $k \geqslant 0$:

$$\bar{X}_{t_{k+1}} = \bar{X}_{t_k} + \gamma_{k+1} b(\bar{X}_{t_k}) + \sqrt{2}U_{k+1}, \tag{36}$$

where for all $k \geqslant 1$, $U_k = B_{t_{k+1}} - B_{t_k}$. For *one* path of the discrete LMC (36), we introduce the discrete and continuous Cesaro average *built with the same Brownian trajectory*:

$$\widehat{X}^{\gamma}_{t_{k_0},t_N} = \frac{1}{t_N - t_{k_0}} \sum_{j=k_0}^{N-1} \gamma_{j+1} \bar{X}_{t_j} \quad \text{and} \quad \widehat{X}_{t_{k_0},t_N} = \frac{1}{t_N - t_{k_0}} \int_{t_{k_0}}^{t_N} X_s \mathrm{d}s.$$

We assume in this paragraph that W is ρ-strongly convex (Assumption $(\mathbf{SC}_\rho)$) and L-smooth (Assumption $(\mathbf{A}_L)$). It entails the following lower bound:

$$\forall (x_1, x_2) \in \mathbb{R}^d \qquad \langle b(x_1) - b(x_2), x_1 - x_2 \rangle \leqslant -\frac{\rho}{2}|x_1 - x_2|^2. \tag{37}$$

Setting $x^\star = \arg\min W$, the main result of the paragraph is as follows.

Theorem 5.1 (Discrete LMC $(\mathbf{SC}_\rho) - (\mathbf{A}_L)$)**.** *Assume that W satisfies $(\mathbf{A}_L)$ and $(\mathbf{SC}_\rho)$.*

i) If $(\gamma_k)_{k\geqslant 1}$ is a decreasing step-size such that $\gamma_1 L \leqslant 1/2$, then

$$\mathbb{E}\left(|\widehat{X}^{\gamma}_{t_{k_0},t_N} - \widehat{X}_{t_{k_0},t_N}|^2\right) \lesssim_{id} \frac{1}{\rho(t_N - t_{k_0})} + \frac{L^2 d}{\rho(t_N - t_{k_0})} \sum_{j=k_0}^{N} \gamma_j^2 + \frac{L^4 d}{\rho^3 (t_N - t_{k_0})} \sum_{j=k_0}^{N} \gamma_j^3.$$

ii) If $\gamma_k = \gamma$ with $\gamma L \leqslant 1/2$, then

$$\mathbb{E}\left(|\widehat{X}^{\gamma}_{t_{k_0},t_N} - \widehat{X}_{t_{k_0},t_N}|^2\right) \lesssim_{id} \frac{1}{\rho(N - k_0)\gamma} + \frac{L^2 d}{\rho}\gamma + \frac{L^4 d}{\rho^3}\gamma^2.$$

iii) For $\varepsilon \in (0,1]$, use $\gamma = \frac{\rho}{L^2} d^{-1}\varepsilon^2$. Then, $\mathbb{E}|\widehat{X}^{\gamma}_{t_{k_0},t_N} - \widetilde{\theta}_n|^2 \lesssim_{id} \varepsilon^2$ when:

$$k_0 \gtrsim_{id} \left(\frac{L}{\rho}\right)^2 d\epsilon^{-2} \log((d + |x - x^\star|)\rho^{-1}\epsilon^{-1}) \quad \text{and} \quad N - k_0 := \frac{L^2}{\rho^2} d\varepsilon^{-4}.$$

Remark 6. In the strongly convex case, the interesting point is that the process and its Euler discretization get closer when the time goes on (at least in L^2-sense). Thus, the corresponding Cesaro averages inherit this property. The two first statements of the above theorem quantify this convergence. Then, it remains to estimate the L^2-distance between the Cesaro average of the true process and the mean value of the invariant distribution to deduce the result. Note that the cost is proportional to $d\varepsilon^{-4}$. A sharper expansion of the L^2-error between the process and the Euler scheme could lead to ε^{-3} but would involve some damages on the dependency with d.

The next paragraphs are devoted to the proof of Theorem 5.1.

5.1.1. *One-Step control.* Let $\gamma > 0$. Consider two starting points x and y in $\mathbb{R}^d$ of two processes $(Z_t^x)_{t\geqslant 0}$ and $(\bar{Z}_t^y)_{t\geqslant 0}$ where $(Z_t^x)_{t\geqslant 0}$ solves the stochastic differential equation:

$$Z_t^x = x + \int_0^t b(Z_s^x)\mathrm{d}s + \sqrt{2}B_t \quad \forall t \in [0,\gamma],$$

whereas $(\bar{Z}_t^y)_{t\geqslant 0}$ is the discretized trajectory built with the same Brownian motion[5]:

$$\bar{Z}_t^y = y + tb(y) + \sqrt{2}B_t \quad \forall t \in [0,\gamma].$$

We first state an important technical proposition that compares the difference between $(\bar{Z}_t^y)_{t\geqslant 0}$ and $(Z_t^x)_{t\geqslant 0}$, along a step-size of length γ.

Proposition 5.2. *If* $(\mathbf{SC}_\rho)$ *and* $(\mathbf{A}_L)$ *hold, let* $\gamma > 0$*, then*

$$\forall (x,y) \in \mathbb{R}^d \quad \forall t \in [0,\gamma] \qquad \mathbb{E}[|Z_t^x - \bar{Z}_t^y|^2] \leqslant |x-y|^2 e^{-\frac{\rho}{4}t} + \frac{L^2}{\rho} dt^2 + \frac{4}{3}\frac{L^2}{\rho}|b(y)|^2 t^3.$$

The proof of Proposition 5.2 is postponed to [GPP20, Section 4].

5.1.2. *Conclusion of the proof.* In the final step of the proof, we shall need a precise control of $\sup_{k\geqslant 0}\mathbb{E}|b(\bar{X}_{t_k})|^2$ and $\sup_{t\geqslant 0}\mathbb{E}|b(Z_t^x)|^2$. Below, we intensively use that when $b = -\nabla W$, with W ρ-strictly convex and b L-Lipschitz, we have:

$$2\frac{\rho^2}{L}(W(x) - \min W) \leqslant |b(x)|^2 \leqslant \frac{2L^2}{\rho}(W(x) - \min W). \tag{38}$$

We have the following proposition whose proof is deferred to [GPP20, Section 4].

Proposition 5.3. *Assuming* $(\mathbf{SC}_\rho)$, $(\mathbf{A}_L)$ *and that* $\gamma_k = \gamma_1 k^{-\delta}$ *with* $\delta \in [0,1]$ *and* $\gamma_1 L \leqslant \frac{1}{2}$*, we have the following control*

$$\sup_{k\geqslant 0}\mathbb{E}|b(\bar{X}_{t_k})|^2 \leqslant \mathbb{E}|b(\bar{X}_0)|^2 + d\frac{L^2}{\rho^2} \quad \textit{and} \quad \sup_{t\geqslant 0}\mathbb{E}|b(Z_t^x)|^2 \leqslant \frac{2L^2}{\rho}\left[\widetilde{W}(x) + \frac{dL^2}{2\rho^2}\right].$$

We now prove the main result of Section 5.

Proof of Theorem 5.1. i). We consider the trajectory $(X_t)_{t\geqslant 0}$ of (45) and its discretized counterpart $(\bar{X}_{t_j})_{j\geqslant 1}$. For an initialization $t_{k_0} \geqslant 0$ and an ending discretization horizon $t_N > t_{k_0}$.We observe that

$$\begin{aligned}
\widehat{X}^\gamma_{t_{k_0},t_N} - \widehat{X}_{t_{k_0},t_N} &= \frac{1}{t_N - t_{k_0}}\left(\sum_{j=k_0}^N \gamma_j \bar{X}_{t_j} - \int_{t_{k_0}}^{t_N} X_t \mathrm{d}t\right)\\
&= \frac{1}{t_N - t_{k_0}}\left(\sum_{j=k_0}^N \int_{t_j}^{t_{j+1}} (\bar{X}_{t_j} - X_t)\mathrm{d}t\right)\\
&= \sum_{j=k_0}^N \omega_{j,N,k_0}\left[\frac{1}{\gamma_j}\int_{t_j}^{t_{j+1}} (\bar{X}_{t_j} - X_t)\mathrm{d}t\right],
\end{aligned}$$

where $\omega_{j,N,k_0} = \frac{\gamma_j}{t_N - t_{k_0}}$ satisfies $\sum_{j=k_0}^N \omega_{j,N,k_0} = 1$. The Jensen inequality yields:

$$|\widehat{X}^\gamma_{t_{k_0},t_N} - \widehat{X}_{t_{k_0},t_N}|^2 \leqslant \sum_{j=k_0}^N \omega_{j,k_0,N}\left|\frac{1}{\gamma_j}\int_{t_j}^{t_{j+1}} (\bar{X}_{t_j} - X_t)\mathrm{d}t\right|^2 \leqslant \sum_{j=k_0}^k \frac{\omega_{j,k_0,k}}{\gamma_j}\int_{t_j}^{t_{j+1}} \left|\bar{X}_{t_j} - X_t\right|^2 \mathrm{d}t.$$

[5]We therefore adopt a synchronous coupling point of view where we explicitly build a discretized trajectory with the same Brownian motion as the one used with the continuous solution.

We conclude that:

$$\mathbb{E}|\widehat{X}^{\gamma}_{t_{k_0},t_N} - \widehat{X}_{t_{k_0},t_N}|^2 \leqslant \frac{1}{t_N - t_{k_0}} \sum_{j=k_0}^{N} \int_{t_j}^{t_{j+1}} \mathbb{E}\left|\bar{X}_{t_j} - X_t\right|^2 \mathrm{d}t. \tag{39}$$

We are led to study $\mathbb{E}\left|\bar{X}_{t_j} - X_t\right|^2$ when $t \in [t_j, t_{j+1}]$. We first decompose this last quantity as

$$\mathbb{E}\left|\bar{X}_{t_j} - X_t\right|^2 = \mathbb{E}\left|\bar{X}_{t_j} - X_{t_j} + X_{t_j} - X_t\right|^2 \leqslant 2\mathbb{E}\left|\bar{X}_{t_j} - X_{t_j}\right|^2 + 2\mathbb{E}\left|X_{t_j} - X_t\right|^2.$$

- We let $\Upsilon_j := \mathbb{E}|\bar{X}_{t_j} - X_{t_j}|^2$, apply Proposition 5.2 and Proposition 5.3 and obtain that:

$$\Upsilon_{j+1} \leqslant \Upsilon_j e^{-\rho\frac{\gamma_{j+1}}{4}} + \underbrace{\frac{L^2 d}{\rho}}_{:=\kappa_1} \gamma_{j+1}^2 + \underbrace{\frac{4L^2}{\rho}\left[\mathbb{E}|b(\bar{X}_0)|^2 + d\frac{L^2}{\rho^2}\right]}_{:=\kappa_2} \gamma_{j+1}^3,$$

A direct recursion yields:

$$\begin{aligned}\forall j \geqslant k_0 \qquad \Upsilon_j &\leqslant \Upsilon_{k_0} \prod_{i=k_0+1}^{j} e^{-\frac{\rho}{4}\gamma_i} + \kappa_1 \sum_{i=k_0+1}^{j} \gamma_i^2 \prod_{\ell=i+1}^{j} e^{-\frac{\rho}{4}\gamma_\ell} + \kappa_2 \sum_{i=k_0+1}^{j} \gamma_i^3 \prod_{\ell=i+1}^{j} e^{-\frac{\rho}{4}\gamma_\ell} \\ &= \Upsilon_{k_0} e^{-\frac{\rho}{4}(t_j - t_{k_0})} + \sum_{i=k_0+1}^{j} [\kappa_1\gamma_i^2 + \kappa_2\gamma_i^3] e^{-\frac{\rho}{4}(t_j - t_i)}.\end{aligned}$$

We then use Lemmas 4.1, 4.2 and 4.3 of the supplementary document [GPP20, Section 4] and write that some universal constants (c_1, c_2) exist such that

$$\forall j \geqslant k_0 \qquad \Upsilon_j \leqslant e^{-\frac{\rho}{4}(t_j - t_{k_0})} + c_1\kappa_1\gamma_j + c_2\kappa_2\gamma_j^2. \tag{40}$$

- To study the second term, we observe that:

$$\forall t \in [t_j, t_{j+1}] \qquad X_t - X_{t_j} = \int_{t_j}^{t} b(X_t)\mathrm{d}s + \sqrt{2}(B_t - B_{t_j}),$$

which implies that

$$\mathbb{E}|X_t - X_{t_j}|^2 \leqslant 2\mathbb{E}\left|\int_{t_j}^{t} b(X_s)\mathrm{d}s\right|^2 + 4\mathbb{E}(|B_t - B_{t_j}|^2).$$

The triangle inequality and the Brownian increment $B_t - B_{t_j}$ yield:

$$\mathbb{E}|X_t - X_{t_j}|^2 \leqslant 2(t - t_j)^2 \sup_{s \geqslant 0} \mathbb{E}|b(X_s)|^2 + 4d(t - t_j).$$

Again, Proposition 5.3 leads to

$$\forall t \in [t_j, t_{j+1}] \qquad \mathbb{E}|X_t - X_{t_j}|^2 \leqslant \kappa_3(t - t_j) + \kappa_4(t - t_j)^2, \tag{41}$$

where

$$\kappa_3 = 4d \qquad \text{and} \qquad \kappa_4 = \frac{4L^2}{\rho}\left[\widetilde{W}(x) + \frac{dL^2}{2\rho^2}\right].$$

We now use (40) and (41) in (39) and obtain that

$$\begin{aligned}\mathbb{E}|\widehat{X}^{\gamma}_{t_{k_0},t_N} - \widehat{X}_{t_{k_0},t_N}|^2 &\leqslant \frac{2}{t_N - t_{k_0}} \sum_{j=k_0}^{k} \int_{t_j}^{t_{j+1}} \left[\Upsilon_j + \mathbb{E}|X_t - X_{t_j}|^2 \mathrm{d}t\right] \\ &\leqslant \frac{2}{t_N - t_{k_0}} \sum_{j=k_0}^{k} e^{-\frac{\rho}{4}(t_j - t_{k_0})}\gamma_j + \left(c_1\kappa_1 + \frac{\kappa_3}{2}\right)\gamma_j^2 + \left(c_2\kappa_2 + \frac{\kappa_4}{3}\right)\gamma_j^3\end{aligned}$$

Finally, Lemma 4.4 of the supplementary document [GPP20] implies

$$\mathbb{E}|\widehat{X}^{\gamma}_{t_{k_0},t_N}-\widehat{X}_{t_{k_0},t_N}|^2\leqslant \frac{4}{\rho(t_N-t_{k_0})}+\frac{c_1\kappa_1+\kappa_3/2+2\rho}{t_N-t_{k_0}}\sum_{j=k_0}^{N}\gamma_j^2+\frac{c_2\kappa_2+\kappa_4}{t_N-t_{k_0}}\sum_{j=k_0}^{N}\gamma_j^3.$$

Using that $\rho\leqslant L$, we then observe that

$$c_1\kappa_1+\kappa_3/2+2\rho\lesssim_{id}\frac{L^2}{\rho}d \qquad \text{and} \qquad c_2\kappa_2+\kappa_4\lesssim_{id}\frac{L^4}{\rho^3}d.$$

This concludes the proof of $i)$. $ii)$ is a straightforward computation.

Proof of Theorem 5.1, $iii)$. First, observe that our tuning of the parameters in our statement yields:

$$\mathbb{E}|\widehat{X}^{\gamma}_{t_{k_0},t_N}-\widehat{X}_{t_{k_0},t_N}|^2\lesssim_{id}\varepsilon^2.$$

Thus, it remains to prove that $\mathbb{E}|\widehat{X}_{t_{k_0},t_N}-\pi_n(\mathbf{I}_d)|^2\lesssim_{id}\varepsilon^2$ (for an appropriate choice of k_0). To this end, let us recall that, under $(\mathbf{SC}_\rho)$, a classical Gronwall-type argument leads to:

$$\forall(x,y,t)\in\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}_+: \qquad |X^x_t-X^y_t|^2\leqslant|x-y|^2e^{-2\rho t} \quad a.s. \tag{42}$$

$$\forall(x,t)\in\mathbb{R}^d\times\mathbb{R}_+: \qquad \mathbb{E}_x[|X_t-x^\star|^2]\leqslant|x-x^\star|^2e^{-2\rho t}+\frac{d}{2\rho}. \tag{43}$$

Then, setting $f=\mathbf{I}_d-\pi_n(\mathbf{I}_d)$, we remark that for some given $0\leqslant t<T$,

$$\mathbb{E}_x\left(\left|\frac{1}{T-t}\int_t^T X_s ds-\pi_n(\mathbf{I}_d)\right|^2\right)=\frac{2}{(T-t)^2}\int_t^T\int_u^T\mathbb{E}_x[f(X_s)f(X_u)]\mathrm{d}s\mathrm{d}u. \tag{44}$$

By the Markov property, $\mathbb{E}_x[f(X_s)f(X_u)]=\mathbb{E}_x[\Psi_{s-u}(X_u)f(X_u)]$ where for $(z,v)\in\mathbb{R}^d\times\mathbb{R}_+$, $\Psi_v(z)=\mathbb{E}_z[f(X_v)]=\mathbb{E}_z[X_v]-\pi_n(\mathbf{I}_d)$. By (42), one deduces that

$$|\Psi_v(z)|\leqslant\int_{\mathbb{R}^d}\mathbb{E}[|X^z_t-X^y_t|]\pi_n(\mathrm{d}y)\leqslant\int_{\mathbb{R}^d}|z-y|\pi_n(\mathrm{d}y)e^{-\rho v}\leqslant\left(|z-x^\star|+\int_{\mathbb{R}^d}|y-x^\star|\pi_n(\mathrm{d}y)e^{-\rho v}\right).$$

Thus,

$$\begin{aligned}\mathbb{E}_x[\Psi^2_{s-u}(X_u)]^{\frac12}&\leqslant e^{-\rho(s-u)}\left(\mathbb{E}_x[X_u-x^\star|^2]^{\frac12}+\left(\int_{\mathbb{R}^d}|y-x^\star|^2\pi_n(\mathrm{d}y)\right)^{\frac12}\right)\\&\leqslant 2e^{-\rho(s-u)}\sup_{t\geqslant0}\mathbb{E}_x[X_t-x^\star|^2]^{\frac12}\leqslant 2e^{-\rho(s-u)}\left(|x-x^\star|+\sqrt{\frac{d}{2\rho}}\right),\end{aligned}$$

where in the last line, we used (43) and the convergence of $\mathcal{L}(X^x_t)\longrightarrow\pi_n$, which implies that:

$$\left(\int_{\mathbb{R}^d}|y-x^\star|^2\pi_n(\mathrm{d}y)\right)^{\frac12}\leqslant\sup_{t\geqslant0}\mathbb{E}_x[X_t-x^\star|^2]^{\frac12}.$$

On the other hand, using again (43), we have:

$$\mathbb{E}_x[f^2(X_u)]^{\frac12}\leqslant\mathbb{E}_x[X_u-x^\star|^2]^{\frac12}+\left(\int_{\mathbb{R}^d}|y-x^\star|^2\pi_n(\mathrm{d}y)\right)^{\frac12}\leqslant 2|x-x^\star|+\sqrt{\frac{d}{2\rho}}.$$

From what precedes and the Cauchy-Schwarz inequality, we deduce that:

$$|\mathbb{E}_x[f(X_s)f(X_u)]\leqslant E_x[\Psi^2_{s-u}(X_u)]^{\frac12}\mathbb{E}_x[f^2(X_u)]^{\frac12}\leqslant\mathfrak{c}e^{-\rho(s-u)}\left(|x-x^\star|^2+\frac{d}{\rho}\right),$$

with $\mathfrak{c} = 8(|x-x^\star|^2 + d/\rho)$. Hence, plugging this inequality into (44), we obtain that:

$$\mathbb{E}_x\left(\left|\frac{1}{T-t}\int_t^T X_s \mathrm{d}s - \pi_n(\mathbf{I}_d)\right|^2\right) \leqslant \frac{2\mathfrak{c}}{(T-t)^2}\int_t^T\int_u^T e^{-\rho(s-u)}\mathrm{d}s\mathrm{d}u \leqslant \frac{2\mathfrak{c}e^{-\rho t}}{\rho^2(T-t)^2}.$$

Setting $t=\gamma k_0$ and $T=N\gamma$, our parameters (N,γ) ensures $\mathbb{E}|\widehat{X}_{t_{k_0},t_N} - \pi_n(\mathbf{I}_d)|^2 \leqslant \varepsilon^2$ as soon as:

$$\frac{\sqrt{2\mathfrak{c}}e^{-\frac{\rho}{2}k_0\gamma}}{\rho} \leqslant \varepsilon \Longleftarrow k_0 \geqslant \frac{2}{\rho\gamma}\log\left(\frac{\sqrt{2\mathfrak{c}}}{\rho}\varepsilon^{-1}\right).$$

$\square$

5.2. Bayesian learning with discrete LMC - strongly convex case - Theorem 2.8.

Proof of Theorem 2.8. We shall apply the above results in our Bayesian framework. Given the set of n observations $(\xi_1,\ldots,\xi_n)$, we have:

$$W_n(\xi^{\mathbf{n}},x) = \sum_{i=1}^n U(\xi_i,x).$$

Assuming that $U(\xi,.)$ satisfies $(\mathbf{A}_L)$, the triangle inequality shows that $W_n(\xi^{\mathbf{n}},.)$ satisfies $(\mathbf{A}_{nL})$ regarless the value of $\xi^{\mathbf{n}}$. In the meantime, assuming that $U(\xi,.)$ satisfies $(\mathbf{SC}_\rho)$ yields $W_n(\xi^{\mathbf{n}},.)$ is $n\rho$-strongly convex i.e satisfies $(\mathbf{SC}_{n\rho})$.

If we denote by $N_0 = N-k_0$ the number of iterates used for the Cesaro averaging all along our discrete trajectory used with a constant step-size $\gamma>0$, we shall remark that $\widehat{X}_{t_{k_0},t_N} = \widehat{\theta}_{n,\gamma k_0,\gamma N}$. We apply Theorems 2.3, 2.5, and 5.1: the symbol $\mathbf{E}[.]$ refers to the expectation with respect to the sampling and to the discretization procedures. We obtain that:

$$\begin{aligned}
\mathbf{E}\left[|\widehat{X}^\gamma_{t_{k_0},t_N} - \theta^\star|^2\right] &\lesssim_{id} \mathbf{E}\left[|\widehat{X}^\gamma_{t_{k_0},t_N} - \widehat{X}_{t_{k_0},t_N}|^2\right] + \mathbf{E}\left[|\widehat{\theta}_{n,\gamma k_0,\gamma N} - \tilde{\theta}_n|^2\right] + \mathbb{E}_{\theta^\star}\left[|\tilde{\theta}_n - \theta^\star|^2\right]\\
&\lesssim_{id} \left(\frac{1}{[n\rho]N_0\gamma} + \frac{[nL]^2 d}{[n\rho]}\gamma + \frac{[nL]^4 d}{[n\rho]^3}\gamma^2\right) + \mathbb{E}_{\theta^\star}\left(\frac{\sqrt{\mathbb{V}_{n,4}}}{N_0\gamma\lambda_{1,n}} + \sqrt{\mathbb{V}_{n,4}J_{\mu,0}}e^{-\lambda_{1,n}k_0\gamma}\right) + \varepsilon_n^2\\
&\lesssim_{id} \left(\frac{1}{n\rho N_0\gamma} + \frac{L^2}{\rho}nd\gamma + \frac{L^4}{\rho^3}nd\gamma^2\right) + \frac{\sqrt{\mathbb{E}_{\theta^\star}[\mathbb{V}_{n,4}]}}{n\rho N_0\gamma} + d^{d/(2\beta)}e^{-n\rho k_0\gamma}\sqrt{\mathbb{E}_{\theta^\star}[\mathbb{V}_{n,4}]} + \varepsilon_n^2,
\end{aligned}$$

where the last line comes from the Cauchy-Schwarz inequality, Proposition 3.3 of the supplementary document [GPP20] and the Bakry-Emery criterion [BE83] that entails $\lambda_{1,n} \geqslant n\rho$. We then use Proposition 3.1 of the supplementary document [GPP20] and obtain that:

$$\mathbf{E}\left[|\widehat{X}^\gamma_{t_{k_0},t_N} - \theta^\star|^2\right] \lesssim_{id} \left(\frac{1}{n\rho N_0\gamma} + \frac{L^2}{\rho}nd\gamma + \frac{L^4}{\rho^3}nd\gamma^2\right) + \frac{\varepsilon_n^2}{\rho n N_0\gamma} + d^{d/(2\beta)}\varepsilon_n^2 e^{-\rho n k_0\gamma} + \varepsilon_n^2.$$

To obtain an ε_n^2 M.S.E with $\varepsilon_n^2 = \left(\frac{d\log(n)}{n}\right)^{1/\alpha_c} \longrightarrow 0$ as $n \longrightarrow +\infty$, we are led to choose:

$$\frac{L^2}{\rho}nd\gamma \leqslant \varepsilon_n^2 \Longleftarrow \gamma = \frac{\rho}{L^2}\mathcal{O}_{id}\left(\frac{\varepsilon_n^2}{nd}\right).$$

We then observe the consequence of this choice on k_0 and N_0 that need to satisfy:

$$k_0 \geqslant \frac{d\log(d)}{2\beta\rho n\gamma} = \left(\frac{L}{\rho}\right)^2\mathcal{O}_{id}\left(\varepsilon_n^{-2}d^2\log(d)\right) \quad \text{and} \quad N_0 \geqslant \frac{\varepsilon_n^{-2}}{n\gamma\rho} = \left(\frac{L}{\rho}\right)^2\mathcal{O}_{id}\left(d\varepsilon_n^{-4}\right).$$

$\square$

6. Discretization of the Langevin procedure - Weakly convex case

The weakly convex (*i.e.* not uniformly strongly convex) case is tackled with a completely different approach. Actually, in the weakly convex case, a series of properties disappears. For instance, one can not easily control the pathwise distance between the process and its discretization. The problem is then significantly harder and we choose here to make use of the inversion of the Poisson equation, which leads to a relatively tractable formulation of the error between the discretized Cesaro average and the invariant distribution (applied to the identity function). In particular, this "Poisson equation approach" is in the continuity of [LP02, HMP20]) and has a long-standing history in the study of central limit theorem for Markov chains. We refer to [Mey71, Nev76, Rev84, GM96] for seminal contributions on additive functionals of Markov chains. We first stay at an informal level in this paragraph for the sake of readability. We sketch the general idea behind the use of this equation with an Euler scheme.

Again, we first state some general results with a diffusion process $(X_t)_{t\geqslant 0}$ solution of:

$$dX_t = -\nabla W(X_t)dt + \sqrt{2}dB_t. \tag{45}$$

6.1. **How to use the Poisson equation $f - \pi(f) = \mathcal{L}g$?** This approach is based on the inversion of the operator $\mathcal{L}$ of the diffusion. For a given function f, we recall that the solution of the Poisson equation is the function g such that $\pi(g) = 0$ and that satisfies:

$$f - \pi(f) = \mathcal{L}g,$$

where π denotes the invariant distribution of the diffusion (see below for background on existence and uniqueness of the solution). We consider g the solution of the Poisson equation.

• For such a solution, a first important ingredient is based on the following remark: if $(X_t)_{t\geqslant 0}$ is a Markov process with generator $\mathcal{L}$ and g belongs to the (extended) domain, then the Ito formula yields:

$$g(X_t) = g(X_0) + \int_0^t \mathcal{L}g(X_s)\mathrm{d}s + \mathcal{M}_t^g,$$

so that

$$\int_0^t f(X_s) - \pi(f)\mathrm{d}s = \int_0^t \mathcal{L}g(X_s)\mathrm{d}s - (g(X_t) - g(X_0)) \tag{46}$$

is a local martingale (and certainly a true martingale under appropriate conditions). Thus, the control of the distance between $(\frac{1}{t}\int_0^t f(\xi_s))_{t\geqslant 0}$ and $\pi(f)$ can be tackled from a martingale point of view.

• The second main interest of this approach is the possibility to specify that our estimator involves $f = \mathrm{Id}$, which is an important ingredient of the approximation of $\pi(f)$. Such a precision is untractable when we handle distances between probability distributions.

We first state that the Poisson equation is well-posed in our setting and recall a classical formulation of this solution. The proof is postponed to the supplementary document [GPP20, Section 4.2.4]. Note that this result is only stated under the assumptions of our main theorems but may be certainly extended to a more general setting (see [CCG12, Corollary 3.2] for a more general result).

Proposition 6.1 (Poisson equation)**.** *Assume* $(\mathbf{H}_{\mathfrak{c}_1})$ *and* $(\mathbf{H}_{\mathfrak{c}_2,r})$ *and suppose that* W *is* $\mathcal{C}^3$ *with bounded third derivatives. Then, Equation* (45) *admits a unique invariant distribution and for any* $\mathcal{C}^2$*-function* f *with bounded derivatives, the problem* $\mathcal{L}g = f - \pi(f)$ *is well-posed on the set of* $\mathcal{C}^2$*-functions such that* $\pi(g) = 0$ *and the unique solution is given by:*

$$g(x) = \int_0^{+\infty} [\pi(f) - P_s f(x)]\mathrm{d}s.$$

Note that in what follows, we will solve this equation d times for a multivariate function $f = (f^1, \ldots, f^d)$. More precisely we shall need to consider $f = \mathbf{I}_d$.

6.2. **Poisson equation and discretization.** In the discretized case, the aim is then to mimick the martingale property of Equation (46) but some additional error terms appear with the discretization approximation. Such ideas have been strongly studied in [LP02, HMP20] but since the solution to the Poisson equation is not explicit (in general), the previous works have usually made *ad hoc* assumptions on the function g and its derivatives. For our purpose, we identify the key properties satisfied by the solution when $f = \mathbf{I}_d$ in terms of the dimensional dependence.

Theorem 6.2. *Assume* $(\mathbf{H}_{\mathfrak{c}_1})$ *and* $(\mathbf{H}_{\mathfrak{c}_2,r})$ *with* $r \in [0,1]$*. Let* $\bar{X}^\gamma$ *denote the Euler scheme with constant step-size* γ *initialized at* x_0*. Assume that* $\gamma \leqslant \gamma_0 := \frac{1}{8}((dL)^{-1} \wedge \frac{1}{8})$*. Assume that* $W(x_0) \lesssim_{id} \mathfrak{b}_d$*.*

i) Then for any $\mathfrak{e} > 0$*,*

$$\mathbb{E}_{x_0}\left[\left|\frac{1}{N}\sum_{k=0}^{N-1}\bar{X}^\gamma_{t_k} - \pi(\mathbf{I}_d)\right|^2\right] \lesssim_{id} \mathfrak{c}_2^{2+\mathfrak{e}} d\mathfrak{b}_d^{2r+\mathfrak{e}}\left(\frac{1}{t_N} + L^2\left(\mathfrak{c}_1\gamma^2\mathfrak{b}_d + \gamma d\right)\right) + \mathfrak{c}_2^{1+\mathfrak{e}}\frac{d\mathfrak{b}_d^{1+3r+\mathfrak{e}}}{t_N^2},$$

where in the above inequality , the constant C hidden in "$\lesssim_{id}$" depends only on $\mathfrak{e}$*.*

ii) If furthermore, W *is* $\mathcal{C}^3$ *with* $\|D^3W\|_\infty < +\infty$ *(defined by (21)), then:*

$$\mathbb{E}_{x_0}\left[\left|\frac{1}{N}\sum_{k=0}^{N-1}\bar{X}^\gamma_{t_k} - \pi(\mathbf{I}_d)\right|^2\right] \lesssim_{id} \mathfrak{c}_2^{2+\mathfrak{e}} d\mathfrak{b}_d^{2r+\mathfrak{e}}\left(\tfrac{1}{t_N} + \gamma^2(\mathfrak{c}_1 L^2\mathfrak{b}_d + \|D^3W\|_\infty^2 d^4\mathfrak{b}_d^{2r+\mathfrak{e}})\right) + \mathfrak{c}_2^{1+\mathfrak{e}}\tfrac{d\mathfrak{b}_d^{1+3r+\mathfrak{e}}}{t_N^2} + \mathfrak{c}_2^{2(1+\mathfrak{e})}\tfrac{L^2d^2\mathfrak{b}_d^{2r+\mathfrak{e}}\gamma}{t_N}.$$

Proof. We observe that $(\bar{X}_{t_k})_{k\geqslant 1}$, computed through the recursion

$$\bar{X}_{t_k} = -\gamma_k\nabla W(\bar{X}_{t_{k-1}}) + \sqrt{2\gamma_k}\zeta_k,$$

where $(\zeta_k)_{k\geqslant 0}$ is an i.i.d. sequence of standard d-dimensional Gaussian random variables, is a sequence of discrete time observations of the continuous time process $(\bar{X}_t)_{t\geqslant 0}$ defined by:

$$\forall t \in [t_k, t_{k+1}] \qquad d\bar{X}_t = -\nabla W(\bar{X}_{t_k})dt + \sqrt{2}dB_t,$$

with $\sqrt{\gamma_k}\zeta_k = B_{t_k} - B_{t_{k-1}}$.

Considering a *multivariate* function $g = (g^1, \ldots, g^d) : \mathbb{R}^d \to \mathbb{R}^d$, we denote by Dg the multidimensional gradient, which corresponds to the matrix $Dg = [\nabla g^1, \ldots, \nabla g^d]$. Similarly, Δg refers to the vector built with $(\Delta g^1, \ldots, \Delta g^d)$. We then observe that:

$$g(\bar{X}_t) = g(x) + \int_0^t \bar{\mathcal{L}}g(\bar{X}_s, \bar{X}_{\underline{s}})\mathrm{d}s + \mathcal{M}_t^{(g)},$$

where $\underline{s}$ is defined in (11), $\bar{\mathcal{L}}$ is given by:

$$\bar{\mathcal{L}}g(x, \underline{x}) = -Dg(x)\nabla W(\underline{x}) + \Delta g(x), \tag{47}$$

and $\mathcal{M}^{(g)}$ is the $\mathbb{R}^d$-valued local martingale defined by:

$$\mathcal{M}_t^{(g)} = \sqrt{2}\int_0^t Dg(\bar{X}_s)\mathrm{d}B_s. \tag{48}$$

Similarly, the definition of $\mathcal{L}$ shall be extended to multivariate functions by

$$\mathcal{L}g(x) = (\mathcal{L}g^1(x), \ldots \mathcal{L}g^d(x)) = -Dg(x)\nabla W(x) + \Delta g(x). \tag{49}$$

The plan of the proof is the same for $i)$ and $ii)$ and is decomposed into three steps. Steps 1 and 2 are common whereas the last one is treated separately. Before going into the proof of these three steps, we first state some crucial bounds for the solution of the Poisson equation associated to $f = \mathbf{I}_d$.

The proof of this proposition is postponed to [GPP20, Section 4.2.4]. We also use the sharp analysis (in terms of the effect of the dimension d) of the exponential moments given in [GPP20, Lemma 4.6].

Proposition 6.3. *Assume* $(\mathbf{H}_{\mathfrak{c}_1})$ *and* $(\mathbf{H}_{\mathfrak{c}_2,r})$ *with* $\mathfrak{c}_1 > 0$, $\mathfrak{c}_2 > 0$ *and* $r \in [0,1]$, *and suppose that* W *is* $\mathcal{C}^3$ *with bounded third derivatives. Let* g *denote the solution of the Poisson equation with* $f = \mathrm{Id}$. *Then,* g *is twice-differentiable and for every* $\mathfrak{e} \in (0,1)$, *a constant* $c_{\mathfrak{e}}$ *exists (which only depends on* $\mathfrak{e}$ *and not on* d*), such that for any* x*:*

- $i-a)$ $\|Dg(x)\|_F^2 \leqslant c_{\mathfrak{e}} d\left(1 + \mathfrak{c}_2^{2(1+\mathfrak{e})}\left(W^{2r(1+\mathfrak{e})}(x) + \mathfrak{b}_d^{2r(1+\mathfrak{e})}\right)\right)$.
- $i-b)$ *If* $W(x_0) \lesssim_{id} \mathfrak{b}_d$, $\sup_{t\geqslant 0} \mathbb{E}_{x_0}[\|Dg(\bar{X}_t)\|_F^2] \leqslant c_{\mathfrak{e}}\mathfrak{c}_2^{2(1+\mathfrak{e})} d\mathfrak{b}_d^{2r(1+\mathfrak{e})}$.
- $ii-a)$ $|g(x) - g(x^\star)| \leqslant c_{\mathfrak{e}}\sqrt{d}\left(\mathfrak{c}_2^{\frac{1}{2}+\mathfrak{e}} W^{\frac{1+3r}{2}+r\mathfrak{e}}(x) + \mathfrak{b}_d^{r(1+\mathfrak{e})} W^{\frac{1+r}{2}}(x)\right)$.
- $ii-b)$ *If* $W(x_0) \lesssim_{id} \mathfrak{b}_d$, *then* $\sup_{t\geqslant 0} \mathbb{E}_{x_0}[|g(\bar{X}_t) - g(x_0)|^2] \leqslant c_{\mathfrak{e}}\mathfrak{c}_2^{1+\mathfrak{e}} d\mathfrak{b}_d^{1+3r+\mathfrak{e}}$.
- $iii-a)$ $\|D^2 g(x)\|_F^2 := \sum_{i,j,k} |D^2_{jk} g_i(x)|^2 \leqslant c_{\mathfrak{e}} d^3 \|D^3 W\|_\infty^2 \mathfrak{c}_2^{2(1+\mathfrak{e})}\left(W^{4r(1+\mathfrak{e})}(x) + \mathfrak{b}_d^{4r(1+\mathfrak{e})}\right)$.
- $iii-b)$ *If* $W(x_0) \lesssim_{id} \mathfrak{b}_d$, $\sup_{t\geqslant 0} \mathbb{E}_{x_0}[\|D^2 g(\bar{X}_t)\|_F^4]^{\frac{1}{2}} \leqslant c_{\mathfrak{e}} d^3 \|D^3 W\|_\infty^2 \mathfrak{c}_2^{2(1+\mathfrak{e})} \mathfrak{b}_d^{4r(1+\mathfrak{e})}$.

Follow-up of the proof of Theorem 6.2. Step 1: Decomposition of $\pi_N(f) - \pi(f)$. We observe that $\forall N \geqslant 1$:

$$
\begin{aligned}
\pi_N(f) - \pi(f) &:= \frac{1}{t_N}\sum_{k=1}^N \gamma_k f(\bar{X}_{t_{k-1}}) - \pi(f) = \frac{1}{t_N}\sum_{k=1}^N \int_{t_{k-1}}^{t_k} f(\bar{X}_{\underline{s}})\mathrm{d}s - \pi(f) \\
&= \frac{1}{t_N}\int_0^{t_N} [f(\bar{X}_{\underline{s}}) - \pi(f)]\mathrm{d}s \\
&= \frac{1}{t_N}\int_0^{t_N} [f(\bar{X}_s) - \pi(f)]\mathrm{d}s + \frac{1}{t_N}\int_0^{t_N} [f(\bar{X}_{\underline{s}}) - f(\bar{X}_s)]\mathrm{d}s
\end{aligned}
$$

Now, we may use the Poisson equation $f - \pi(f) = \mathcal{L}g$ with $f = \mathbf{I}_d$ and deduce that:

$$
\begin{aligned}
\pi_N(\mathbf{I}_d) - \pi(\mathbf{I}_d) &= \frac{1}{t_N}\int_0^{t_N} \mathcal{L}g(\bar{X}_s)\mathrm{d}s + \frac{1}{t_N}\int_0^{t_N} [\bar{X}_{\underline{s}} - \bar{X}_s]\mathrm{d}s \\
&= \frac{1}{t_N}\int_0^{t_N} \bar{\mathcal{L}}g(\bar{X}_s, \bar{X}_{\underline{s}})\mathrm{d}s + \frac{1}{t_N}\int_0^{t_N} [\mathcal{L}g(\bar{X}_s) - \bar{\mathcal{L}}g(\bar{X}_s, \bar{X}_{\underline{s}})]\mathrm{d}s \\
&+ \frac{1}{t_N}\int_0^{t_N} [\bar{X}_{\underline{s}} - \bar{X}_s]\mathrm{d}s.
\end{aligned}
$$

To handle the first term, we use the Ito formula to obtain:

$$
g(\bar{X}_{t_N}) = g(x_0) + \int_0^{t_N} \bar{\mathcal{L}}g(\bar{X}_s, \bar{X}_{\underline{s}})\mathrm{d}s + \mathcal{M}_{t_N}^{(g)} \quad \text{with} \quad \mathcal{M}_{t_N}^{(g)} = 2\int_0^{t_N} Dg(\bar{X}_s)\mathrm{d}B_s.
$$

By (47), we remark that

$$
\mathcal{L}g(x) - \bar{\mathcal{L}}g(x, \underline{x}) = Dg(x)[\nabla W(\underline{x}) - \nabla W(x)].
$$

We then obtain that:

$$\begin{aligned}\pi_N(\mathbf{I}_d)-\pi(\mathbf{I}_d) &= \frac{g(\bar X_{t_N})-g(x_0)}{t_N}-\frac{\mathcal{M}^{(g)}_{t_N}}{t_N}+\frac{1}{t_N}\int_0^{t_N} Dg(\bar X_s)[\nabla W(\bar X_{\underline{s}})-\nabla W(\bar X_s)]\mathrm{d}s\\ &+\frac{1}{t_N}\int_0^{t_N}[\bar X_{\underline{s}}-\bar X_s]\mathrm{d}s\\ &= \overbrace{\frac{g(\bar X_{t_N})-g(x_0)}{t_N}}^{:=A^{(0)}_{t_N}}-\overbrace{\frac{\mathcal{M}^{(g)}_{t_N}}{t_N}}^{:=A^{(1)}_{t_N}}+\overbrace{\frac{1}{t_N}\int_0^{t_N} Dg(\bar X_s)[\nabla W(\bar X_{\underline{s}})-\nabla W(\bar X_s)]\mathrm{d}s}^{:=A^{(2)}_{t_N}}\\ &+\underbrace{\frac{1}{t_N}\int_0^{t_N}[\bar X_{\underline{s}}-\bar X_s]\mathrm{d}s}_{:=A^{(3)}_{t_N}}\,.\end{aligned}\tag{50}$$

The rest of the proof consists in studying the mean-squared error related to each term of the above righ-hand side and to deduce the result the upper-bound for $\mathbb{E}|\pi_N(f)-\pi(f)|^2$.

<u>*Step 2: Mean squared error related to $A^{(0)}_{t_N}$, $A^{(1)}_{t_N}$ and $A^{(3)}_{t_N}$:*</u>

- By Proposition 6.3,

$$\mathbb{E}_{x_0}\left|A^{(0)}_{t_N}\right|^2=\mathbb{E}_{x_0}\left|\frac{g(\bar X_{t_N})-g(\bar X_0)}{t_N}\right|^2\leqslant c_{\mathfrak{e}}\mathfrak{c}_2^{1+\mathfrak{e}}\frac{d\mathfrak{b}_d^{1+3r+\mathfrak{e}}}{t_N^2}.$$

- Let us consider the martingale term $A^{(1)}_{t_N}$:

$$\mathbb{E}_{x_0}[|\mathcal{M}^{(g)}_{t_N}|^2]=2\int_0^{t_N}\mathbb{E}\|Dg(\bar X_s)\|_F^2\mathrm{d}s,$$

where $\|.\|_F$ refers to the Frobenius norm. Then, Proposition 6.3 implies:

$$\mathbb{E}_{x_0}\left|A^{(1)}_{t_N}\right|^2\lesssim_{id} c_{\mathfrak{e}}\mathfrak{c}_2^{2(1+\mathfrak{e})}\frac{d\mathfrak{b}_d^{2r(1+\mathfrak{e})}}{t_N}.$$

- Let us now consider $A^{(3)}_{t_N}$. On $[t_{k-1},t_k)$:

$$\bar X_s-\bar X_{\underline{s}}=\bar X_s-\bar X_{t_{k-1}}=-(s-t_{k-1})\nabla W(\bar X_{t_{k-1}})+\sqrt{2}\left(B_s-B_{t_{k-1}}\right),$$

so that

$$\int_{t_{k-1}}^{t_k}(\bar X_s-\bar X_{t_{k-1}})\mathrm{d}s=-\frac{\gamma^2}{2}\nabla W(\bar X_{t_{k-1}})+\sqrt{2}\underbrace{\int_{t_{k-1}}^{t_k}(B_s-B_{t_{k-1}})\mathrm{d}s}_{:=\Delta N_k}.$$

On the first hand, using that $(\Delta N_k)_{k\geqslant1}$ is a sequence of independent and centered variables:

$$\begin{aligned}\mathbb{E}_{x_0}\left[\left(\frac{1}{t_N}\int_0^{t_N}(B_s-B_{\underline{s}})\mathrm{d}s\right)^2\right]&=\frac{1}{t_N^2}\sum_{k=1}^N\mathbb{E}_{x_0}|\Delta N_k|^2=\frac{1}{t_N^2}\sum_{k=1}^N\mathbb{E}_{x_0}\left|\int_{t_{k-1}}^{t_k}(B_s-B_{t_{k-1}})\mathrm{d}s\right|^2\\ &=\frac{N}{t_N^2}\mathbb{E}_{x_0}\left|\int_0^\gamma B_s\mathrm{d}s\right|^2=\frac{1}{N\gamma^2}\mathbb{E}_{x_0}\left|\int_0^\gamma(\gamma-s)\mathrm{d}B_s\right|^2=\frac{d\gamma}{3N}.\end{aligned}$$

On the other hand, the Jensen inequality yields:

$$\begin{aligned}\mathbb{E}_{x_0}\left|\frac{1}{t_N}\sum_{k=1}^{N}\frac{\gamma^2}{2}\nabla W(\bar{X}_{t_{k-1}})\right|^2 &= \mathbb{E}_{x_0}\left|\frac{1}{N\gamma}\sum_{k=1}^{N}\gamma\left(\frac{\gamma}{2}\nabla W(\bar{X}_{t_{k-1}})\right)\right|^2\\ &\leqslant \frac{1}{N\gamma}\sum_{k=1}^{N}\gamma\mathbb{E}_{x_0}\left[\frac{\gamma^2}{4}\left|\nabla W(\bar{X}_{t_{k-1}})\right|^2\right]\\ &\leqslant \frac{\mathfrak{c}_1\gamma^2}{4}\sup_{k\geqslant 1}\mathbb{E}_{x_0}W(\bar{X}_{t_{k-1}}) \lesssim_{id} \gamma^2(W(x_0)+\mathfrak{b}_d),\end{aligned}$$

where in the last inequality, we used Proposition 4.7 of the supplementary document [GPP20]. Using that $W(x_0)\lesssim_{id}\mathfrak{b}_d$, we deduce from what precedes that:

$$\mathbb{E}_{x_0}\left|A^{(3)}_{t_N}\right|^2 \lesssim_{id} \frac{d\gamma}{3N}+\gamma^2\mathfrak{b}_d.$$

Step 3: Mean squared error related to $A^{(2)}_{t_N}$ The study of this term is isolated not only because its study is more involved, but also because this term differentiates the bound of i) and ii).
We separate the drift and the diffusion components and define $\Delta_{\underline{s}s}=\sqrt{2}(B_s-B_{\underline{s}})$. We have:

$$\begin{aligned}A^{(2)}_{t_N} = &-\overbrace{\frac{1}{t_N}\int_0^{t_N} Dg(\bar{X}_s)[\nabla W(\bar{X}_s)-\nabla W(\bar{X}_{\underline{s}}+\Delta_{\underline{s}s})]\mathrm{d}s}^{:=A^{(2,1)}_{t_N}}\\ &\underbrace{-\frac{1}{t_N}\int_0^{t_N} Dg(\bar{X}_s)[\nabla W(\bar{X}_{\underline{s}}+\Delta_{\underline{s}s})-\nabla W(\bar{X}_{\underline{s}})]\mathrm{d}s}_{:=A^{(2,2)}_{t_N}}\,.\end{aligned}$$

Since $|\nabla W(\bar{X}_s)-\nabla W(\bar{X}_{\underline{s}}+\Delta_{\underline{s}s})|\leqslant L|s-\underline{s}|.|\nabla W(\bar{X}_{\underline{s}})|$,

$$\mathbb{E}_{x_0}[\|A^{(2,1)}_{t_N}\|^2]\leqslant \frac{L^2}{t_N}\int_0^{t_N}(s-\underline{s})^2\mathbb{E}_{x_0}[\|Dg(\bar{X}_s)\|^2]\mathbb{E}_{x_0}[|\nabla W(\bar{X}_{\underline{s}})|^2]\mathrm{d}s.$$

Using $(\mathbf{H}_{\mathfrak{c}_1})$, $(\mathbf{H}_{\mathfrak{c}_2,r})$, Proposition 4.7 of the supplementary document [GPP20], Proposition 6.3 and $W(x_0)\lesssim_{id}\mathfrak{b}_d$, we have:

$$\mathbb{E}_{x_0}[\|Dg(\bar{X}_s)\|^2]\mathbb{E}_{x_0}[|\nabla W(\bar{X}_{\underline{s}})|^2]\lesssim_{id} c_{\mathfrak{e}}\mathfrak{c}_2^{2(1+\mathfrak{e})}d\mathfrak{b}_d^{2r(1+\mathfrak{e})}\mathfrak{c}_1\mathfrak{b}_d,$$

so that:

$$\mathbb{E}_{x_0}[|A^{(2,1)}_{t_N}|^2]\lesssim_{id} c_{\mathfrak{e}}L^2\gamma^2\mathfrak{c}_1\mathfrak{c}_2^{2(1+\mathfrak{e})}d\mathfrak{b}_d^{1+2r(1+\mathfrak{e})}.$$

We finally separate the study of $A^{(2,2)}_{t_N}$ into two cases, respectively for i) and ii).

Step 4a: End of Proof of Theorem 6.2 i): The Cauchy-Schwarz inequality yields

$$\mathbb{E}_{x_0}[|A^{(2,2)}_{t_N}|^2]\leqslant \frac{L^2}{t_N}\int_0^{t_N}\mathbb{E}_{x_0}[\|Dg(\bar{X}_s)\|^2]\mathbb{E}_{x_0}[|\Delta_{\underline{s}s}|^2]\mathrm{d}s.$$

Again Proposition 6.3 implies that:

$$\mathbb{E}_{x_0}[|A^{(2,2)}_{t_N}|^2]\lesssim_{id}\frac{L^2}{t_N}\int_0^{t_N}c_{\mathfrak{e}}\mathfrak{c}_2^{2(1+\mathfrak{e})}d\mathfrak{b}_d^{2r(1+\mathfrak{e})}\times d(s-\underline{s})\mathrm{d}s\lesssim_{\mathrm{id}}\mathrm{L}^2\mathrm{c}_{\mathfrak{e}}\mathfrak{c}_2^{2(1+\mathfrak{e})}\mathrm{d}^2\gamma.$$

The result follows by collecting the bounds obtained for $A^{(0)}_{t_N}$, $A^{(1)}_{t_N}$, $A^{(2,1)}_{t_N}$, $A^{(2,2)}_{t_N}$ and $A^{(3)}_{t_N}$.

Step 4b: End of Proof of Theorem 6.2 ii): It is possible to exploit the centering of $\Delta_{\underline{s}s}$. We decompose into two parts: we write

$$A^{(2,2)}_{t_N} = \overbrace{\frac{1}{t_N}\int_0^{t_N}(Dg(\bar{X}_s) - Dg(\tilde{X}_{\underline{s}}))[\nabla W(\bar{X}_{\underline{s}} + \Delta_{\underline{s}s}) - \nabla W(\bar{X}_{\underline{s}})]ds}^{:=①}$$
$$+ \underbrace{\frac{1}{t_N}\int_0^{t_N} Dg(\tilde{X}_{\underline{s}})[\nabla W(\bar{X}_{\underline{s}} + \Delta_{\underline{s}s}) - \nabla W(\bar{X}_{\underline{s}})]ds}_{:=②},$$

where $\tilde{X}_{\underline{s}} = \bar{X}_{\underline{s}} - (s-\underline{s})\nabla W(\bar{X}_{\underline{s}})$. For the first term, we use the Jensen inequality so that:

$$\mathbb{E}_{x_0}[|①|^2] \leqslant \frac{1}{t_N}\int_0^{t_N}\mathbb{E}[|\mathfrak{G}_s\delta_s|^2]\mathrm{d}s,$$

where $\mathfrak{G}_s$ is the $d\times d$-matrix $\mathfrak{G}_s := Dg(\bar{X}_s) - Dg(\bar{X}_{\underline{s}})$ and $\delta_s := \nabla W(\bar{X}_{\underline{s}} + \Delta_{\underline{s}s}) - \nabla W(\bar{X}_{\underline{s}})]$. Using $|Ax| \leqslant \|A\|_F|x|$ and since ∇W is L-Lispchitz, we obtain:

$$\mathbb{E}_{x_0}[|\mathfrak{G}_s\delta_s|^2] \leqslant L^2\mathbb{E}_{x_0}[\|\mathfrak{G}_s\|_F^2|\Delta_{\underline{s}s}|^2] \leqslant L^2\sum_{i,j}\mathbb{E}_{x_0}[|\mathfrak{G}_s^{i,j}|^2|\Delta_{\underline{s}s}|^2]. \tag{51}$$

Setting $\tilde{X}^{(\eta)}_s = \bar{X}_{\underline{s}} - (s-\underline{s})\nabla W(\bar{X}_{\underline{s}}) + \eta\Delta_{\underline{s}s}$, we deduce from the Taylor formula that

$$\mathfrak{G}_s^{i,j} = \int_0^1 \langle\nabla(D_jg_i)(\tilde{X}^{(\eta)}_s), \Delta_{\underline{s}s}\rangle\mathrm{d}\eta.$$

Again, the Jensen and Cauchy-Schwarz inequalities lead to:

$$\begin{aligned}\sum_{i,j}\mathbb{E}_{x_0}[|\mathfrak{G}_s^{i,j}|^2|\Delta_{\underline{s}s}|^2] &= \sum_{i,j}\mathbb{E}_{x_0}\left[\left|\int_0^1\langle\nabla(D_jg_i)(\tilde{X}^{(\eta)}_s), \Delta_{\underline{s}s}\rangle\mathrm{d}\eta\right|^2|\Delta_{\underline{s}s}|^2\right]\\
&\leqslant 2\mathbb{E}_{x_0}\left[|\Delta_{\underline{s}s}|^4\left(\int_0^1\sum_{i,j}|\nabla(D_jg_i)(\tilde{X}^{(\eta)}_s)|^2\mathrm{d}\eta\right)\right]\\
&= 2\int_0^1\mathbb{E}_{x_0}\left[|\Delta_{\underline{s}s}|^4 |||D^2g(\tilde{X}^{(\eta)}_s)|||^2\right]\mathrm{d}\eta,\end{aligned}$$

where

$$|||D^2g(x)|||^2 := \sum_{i,j,k}|D^2_{j,k}g_i(x)|^2 = \sum_{i=1}^d\|D^2g_i(x)\|_F^2.$$

Thus, by Cauchy-Schwarz inequality,

$$\begin{aligned}\mathbb{E}_{x_0}[|\mathfrak{G}_s\delta_s|^2] &\leqslant L^2\mathbb{E}[|\Delta_{\underline{s}s}|^8]^{\frac12}\sup_{\theta\in[0,1]}\mathbb{E}[|||D^2g(\tilde{X}^{(\theta)}_s)|||^4]^{\frac12}\\
&\lesssim_{id} L^2(s-\underline{s})^2d^2\sup_{\theta\in[0,1]}\mathbb{E}[|||D^2g(\tilde{X}^{(\eta)}_s)|||^4]^{\frac12}.\\
&\lesssim_{id} L^2c_{\mathfrak{e}}(s-\underline{s})^2d^5\|\nabla^3W\|_\infty^2\mathfrak{c}_2^{2(1+\mathfrak{e})}\left(\sup_{\eta\in[0,1]}\mathbb{E}[W^{4r(1+\mathfrak{e})}(\tilde{X}^{(\eta)}_s)] + \mathfrak{b}_d^{4r(1+\mathfrak{e})}\right),\end{aligned}$$

using Lemma 4.8 of the supplementary document [GPP20]. Then, by a slight adaptation of the proof of Lemma 4.6 of the supplementary document [GPP20], we get that:

$$\sup_{t\geqslant 0,\eta\in[0,1]}\mathbb{E}_{x_0}[e^{\frac18 W(\tilde{X}^{(\eta)}_t)}] \lesssim_{id} e^{\frac18 W(x_0)} + \mathfrak{b}_d.$$

Following the proof of Proposition 4.7 and using the fact that $W(x_0) \lesssim_{id} \mathfrak{b}_d$, we get:

$$\sup_{t\geqslant 0,\eta\in[0,1]} \mathbb{E}_{x_0}[W^{4r(1+\mathfrak{e})}(\tilde{X}_t^{(\eta)})] \lesssim_{id} c_{\mathfrak{e}} \mathfrak{b}_d^{4r(1+\mathfrak{e})}.$$

It follows that:

$$\mathbb{E}[|①|^2] \lesssim_{id} c_{\mathfrak{e}} L^2 \gamma^2 d^5 \|\nabla^3 W\|_\infty^2 \mathfrak{c}_2^{2(1+\mathfrak{e})} \mathfrak{b}_d^{4r(1+\mathfrak{e})}. \tag{52}$$

Let us finally consider ②. Using the Taylor formula, we obtain that:

$$Dg(\tilde{X}_{\underline{s}})[\nabla W(\bar{X}_{\underline{s}} + \Delta_{\underline{s}s}) - \nabla W(\bar{X}_{\underline{s}})] = Dg(\tilde{X}_{\underline{s}}) \nabla^2 W(\bar{X}_{\underline{s}}) \Delta_{\underline{s}s} + \int_0^1 \Delta_{\underline{s}s}^T \nabla^3 W(\bar{X}_{\underline{s}} + \eta \Delta_{\underline{s}s}) \Delta_{\underline{s}s} \mathrm{d}\eta.$$

Using that the first term is a martingale, we obtain that:

$$\begin{aligned} \mathbb{E}[|②|^2] \lesssim_{id} & \frac{1}{t_N^2} \int_0^{t_N} \mathbb{E}[\|Dg(\tilde{X}_{\underline{s}}) \nabla^2 W(\bar{X}_{\underline{s}})\|_F^2](s - \underline{s}) \mathrm{d}s \\ & + \frac{1}{t_N} \int_0^{t_N} \|\nabla^3 W\|_\infty^2 d\mathbb{E}[\|Dg(\tilde{X}_{\underline{s}})\|_F^2 |\Delta_{\underline{s}s}|^4] \mathrm{d}s. \end{aligned} \tag{53}$$

The Frobenius norm being sub-multiplicative and $\|\nabla^2 W\|_F \leqslant \sqrt{d} L$, we have:

$$\|Dg(\tilde{X}_{\underline{s}}) \nabla^2 W(\bar{X}_{\underline{s}})\|_F^2 \leqslant dL^2 \|Dg(\tilde{X}_{\underline{s}})\|^2.$$

Once again, by a slight adaptation of Lemma 4.6 of the supplementary document [GPP20], we get the same bounds for $\mathbb{E}[\|Dg(\tilde{X}_{\underline{s}})\|^2]$ as for $\mathbb{E}[\|Dg(\bar{X}_{\underline{s}})\|^2]$ in Proposition 6.3 so that:

$$\begin{aligned} \frac{1}{t_N^2} \int_0^{t_N} \mathbb{E}[\|Dg(\tilde{X}_{\underline{s}}) \nabla^2 W(\bar{X}_{\underline{s}})\|_F^2](s - \underline{s}) \mathrm{d}s & \lesssim_{id} c_{\mathfrak{e}} \mathfrak{c}_2^{2(1+\mathfrak{e})} d^2 \mathfrak{b}_d^{2r(1+\mathfrak{e})} L^2 \frac{1}{t_N^2} \int_0^{t_N} (s - \underline{s}) \mathrm{d}s \\ & \lesssim_{id} c_{\mathfrak{e}} \mathfrak{c}_2^{2(1+\mathfrak{e})} d^2 \mathfrak{b}_d^{2r(1+\mathfrak{e})} L^2 \frac{\gamma}{t_N}. \end{aligned} \tag{54}$$

Finally, for (53), the Cauchy-Schwarz inequality yields:

$$|(53)| \lesssim_{id} \|\nabla^3 W\|_\infty^2 d^3 \frac{1}{t_N} \int_0^{t_N} (s - \underline{s})^2 \mathbb{E}[\|Dg(\tilde{X}_{\underline{s}})\|_F^4]^{\frac{1}{2}} \mathrm{d}s.$$

Once again, a slight adaptation of the proof of Proposition 6.3(ii) yields the same inequality for $\sup_{s\geqslant 0} \mathbb{E}[\|Dg(\tilde{X}_{\underline{s}})\|_F^4]^{\frac{1}{2}}$ as the one of $\sup_{s\geqslant 0} \mathbb{E}[\|Dg(\bar{X}_{\underline{s}})\|^2]$:

$$|(53)| \lesssim_{id} c_{\mathfrak{e}} \mathfrak{c}_2^{2(1+\mathfrak{e})} \|\nabla^3 W\|_\infty^2 d^5 \mathfrak{b}_d^{2r(1+\mathfrak{e})} \gamma^2.$$

Thus, by the above inequality, (54) and (52), we conclude that:

$$\mathbb{E}[|A_{t_N}^{(2,2)}|^2] \lesssim_{id} c_{\mathfrak{e}} \mathfrak{c}_2^{2(1+\mathfrak{e})} d^2 \mathfrak{b}_d^{2r(1+\mathfrak{e})} \left(\|\nabla^3 W\|_\infty^2 d^3 \mathfrak{b}_d^{2r(1+\mathfrak{e})} \gamma^2 + \frac{L^2 \gamma}{t_N} \right).$$

The result follows by collecting the previous bounds and the new one established for $A_{t_N}^{(2,2)}$. ☐

We conclude this section by a result related to the complexity of our algorithm. By complexity, we mean here the number of iterations of the scheme which is necessary to obtain a given L^2-error. We thus denote by N_η the number of iterations which is necessary to guarantee that

$$\mathbb{E}_{x_0} \left[\left| \frac{1}{N_\varepsilon} \sum_{k=1}^{N_\varepsilon} \bar{X}_{k\gamma}^\gamma - \pi(\mathbf{I}_d) \right|^2 \right] \lesssim_{id} \varepsilon^2. \tag{55}$$

Corollary 6.4. *Let the assumptions of Theorem 6.2 be in force and assume that $\mathfrak{b}_d \lesssim_{id} d$. Let $\varepsilon > 0$. There exists a constant c depending only on $\mathfrak{e}$, $\mathfrak{c}_1$, $\mathfrak{c}_2$, L and $\|\nabla^3 W\|_\infty$ such that*

(i) *If $\gamma = c\varepsilon^2 d^{-2-2r(1+\mathfrak{e})}$, then $N_\varepsilon = \varepsilon^{-4} d^{3+4r(1+\mathfrak{e})}$, then Condition (55) is satisfied.*

(ii) If $\gamma = c\varepsilon d^{-\frac{5}{2}-2r-\mathfrak{e}}$, then $N_\varepsilon = \varepsilon^{-3} d^{\frac{7}{2}+4r+2\mathfrak{e}}$, then Condition (55) is satisfied.

Remark 7. The above results may be considered as very bad with respect to the existing literature. However, we have to recall that this result is really obtained in the non-uniformly convex case, which is significantly harder than the uniformly convex one. The second remark is that, as mentioned in Section 2.5.4, the literature using a "standard" Monte-Carlo approach (*i.e.* based on the average of M paths of a given discretization scheme and not on a pathwise average) usually calls complexity, the number of iterations needed for the simulation of one path. In other words, the Monte-Carlo cost is not taken into account. Hence, each result of such literature needs a multiplication by $\varepsilon^{-2}\sigma_d^2$ (where σ_d^2 denotes the variance of the considered function) to really be compared with our results. Actually, here, we consider occupation measures so that, we only compute one path. The number of iterations is thus exactly the complexity (when the complexity unit is one iteration of the scheme).

Proof. (i) Owing to the bound obtained in Theorem 6.2(i), we remark that in order to guarantee Condition (55) (up to parameters $\mathfrak{c}_1$, $\mathfrak{c}_2$, L and $\mathfrak{e}$), we need that:

$$\frac{d^{1+2r+\mathfrak{e}}}{N_\varepsilon \gamma} \leqslant \varepsilon^2 \quad \text{and} \quad \gamma d^{2+2r+\mathfrak{e}} \leqslant \varepsilon^2,$$

which involves that

$$N_\varepsilon \geqslant \varepsilon^{-2}\gamma^{-1} d^{1+2r+\mathfrak{e}} \quad \text{and} \quad \gamma^{-1} \geqslant d^{2+2r(1+\mathfrak{e})}\varepsilon^{-2}.$$

Then, one checks that if we set $\gamma = c\varepsilon^2 d^{-2-2r(1+\mathfrak{e})}$, then $N_\varepsilon = \varepsilon^{-4} d^{3+4r(1+\mathfrak{e})}$, the other terms are controlled by ε^2.

(ii) In this case, we deduce from a careful inspection of each term of the right-hand member of Theorem 6.2(ii) that Condition (55) is satisfied if (up to a constant depending on $\mathfrak{c}_1$, $\mathfrak{c}_2$, L, $\mathfrak{e}$ and $\|\nabla^3 W\|_\infty$)

$$\frac{d^{1+2r+\mathfrak{e}}}{N_\varepsilon \gamma} \leqslant \varepsilon^2 \quad \text{and} \quad \gamma d^{\frac{5}{2}+2r+\mathfrak{e}} \leqslant \varepsilon,$$

which involves that

$$N_\varepsilon \geqslant \varepsilon^{-2}\gamma^{-1} d^{1+2r+\mathfrak{e}} \quad \text{and} \quad \gamma^{-1} \geqslant d^{\frac{5}{2}+2r+\mathfrak{e}}\varepsilon^{-1}.$$

The result follows by setting $\gamma = d^{-\frac{5}{2}-2r-\mathfrak{e}}\varepsilon$ and by plugging it in the first inequality. $\square$

6.3. Bayesian learning with discrete LMC - weakly convex case - Theorem 2.9.

Proof of Theorem 2.9-i). We know that for any ξ, $U(\xi,.)$ satifies $(\mathbf{A}_L)$. It implies that W_n is nL-smooth (see Section 5.2). Since $U(\xi,.)$ satisfies $(\mathbf{H}_{\mathfrak{c}_1})$, a direct computation shows that W_n satisfies $(\mathbf{H}_{n\mathfrak{c}_1})$. In the meantime, Assumption $(\mathbf{H}_{\mathfrak{c}_2,r})$ on each $U(\xi,.)$ implies that:

$$\begin{aligned}
\underline{\lambda}(\nabla^2 W_n) &= \underline{\lambda}\left(\sum_{i=1}^n \nabla^2 U(\xi_i,.)\right) \\
&\geqslant \sum_{i=1}^n \underline{\lambda}(\nabla^2 U(\xi_i,.)) \\
&\geqslant \sum_{i=1}^n \mathfrak{c}_2^{-1} U(\xi_i,.)^{-r} = n\left(\frac{1}{n}\sum_{i=1}^n \mathfrak{c}_2^{-1} U(\xi_i,.)^{-r}\right) \\
&\geqslant \{\mathfrak{c}_2 n^{-(1+r)}\}^{-1} W_n^{-r},
\end{aligned}$$

where we applied the Jensen inequality to the convex function $u \mapsto u^{-r}$. Therefore, W_n satisfies $(\mathbf{H})_{\mathfrak{c}_2 n^{-(1+r)},r}$. Finally, we observe that $\mathfrak{b}_d^{(n)}$ associated to W_n is upper bounded by $\mathfrak{b}_d^{(n)} \lesssim_{id} n\mathfrak{b}_d$. By

Theorem 6.2-i) and Theorem 2.3, we deduce that for any $\mathfrak{e} > 0$:

$$\begin{aligned}\mathbf{E}\left[|\widehat{X}^{\gamma}_{t_{k_0},t_N} - \theta^{\star}|^2\right] &\leqslant 2\mathbf{E}\left[|\widehat{X}^{\gamma}_{t_{k_0},t_N} - \tilde{\theta}_n|^2\right] + 2\mathbb{E}_{\theta^{\star}}\left[|\tilde{\theta}_n - \theta^{\star}|^2\right] \\ &\lesssim_{id} \Bigg((n^{-(1+r)}\mathfrak{c}_2)^{2+\mathfrak{e}} d(n\mathfrak{b}_d)^{2r+\mathfrak{e}}\left(\frac{1}{N\gamma} + (nL)^2\left((n\mathfrak{c}_1)\gamma^2(n\mathfrak{b}_d) + \gamma d\right)\right) \\ &\qquad + (n^{-(1+r)}\mathfrak{c}_2)^{1+\mathfrak{e}}\frac{d(n\mathfrak{b}_d)^{1+3r+\mathfrak{e}}}{N^2\gamma^2}\Bigg) + \varepsilon_n^2.\end{aligned}$$

We now study each terms separately: the first one is defined by

$$① := (n^{-(1+r)}\mathfrak{c}_2)^{2+\mathfrak{e}} d(n\mathfrak{b}_d)^{2r+\mathfrak{e}}\frac{1}{N\gamma} \lesssim_{id} \frac{d\mathfrak{b}_d^{2r+\mathfrak{e}}}{n^{2+r\mathfrak{e}}N\gamma} \lesssim_{id} \frac{d^{1+2r+\mathfrak{e}}}{n^2N\gamma},$$

owing to the assumption $\mathfrak{b}_d \lesssim_{id} d$. The second and third terms are respectively:

$$② := (n^{-(1+r)}\mathfrak{c}_2)^{2+\mathfrak{e}} d(n\mathfrak{b}_d)^{2r+\mathfrak{e}}(nL)^2 n\gamma^2 n\mathfrak{b}_d \lesssim_{id} L^2n^2d^{2+2r+\mathfrak{e}}\gamma^2,$$

and

$$③ := (n^{-(1+r)}\mathfrak{c}_2)^{2+\mathfrak{e}} d(n\mathfrak{b}_d)^{2r+\mathfrak{e}}(nL)^2\gamma d \lesssim_{id} L^2\gamma d^{2+2r+\mathfrak{e}}.$$

Finally, the last term is

$$④ := (n^{-(1+r)}\mathfrak{c}_2)^{1+\mathfrak{e}}\frac{d(n\mathfrak{b}_d)^{1+3r+\mathfrak{e}}}{N^2\gamma^2} \lesssim_{id} \frac{n^{2r}d^{2+3r+\mathfrak{e}}}{N^2\gamma^2}.$$

We first derive an appropriate choice for γ to attain an ε_n^2 accuracy, the constraints are brought by $② \vee ③ \lesssim_{id} \varepsilon_n^2$, which leads to:

$$\gamma \lesssim_{id} \frac{\varepsilon_n}{Lnd^{1+r+\frac{\mathfrak{e}}{2}}} \wedge \frac{\varepsilon_n^2}{L^2d^{2+2r+\mathfrak{e}}}.$$

Plugging this constraint on γ in ① and ④, we get:

$$① \lesssim_{id} \varepsilon_n^2 \quad \text{if} \quad N \gtrsim_{id} \quad \frac{Ld^{2+3r+\frac{3}{2}\mathfrak{e}}}{n\varepsilon_n^3} \vee \frac{L^2d^{3+4r+2\mathfrak{e}}}{n^2\varepsilon_n^4}$$

and

$$④ \lesssim_{id} \varepsilon_n^2 \quad \text{if} \quad N \gtrsim_{id} \frac{Ln^{1+r}d^{2+\frac{5}{2}r+\mathfrak{e}}}{\varepsilon_n^2} \vee \frac{L^2n^rd^{3+\frac{7}{2}r+\frac{3}{2}\mathfrak{e}}}{\varepsilon_n^3}.$$

As a conclusion of the two above statements, we get

$$① \vee ④ \lesssim_{id} \varepsilon_n^2 \quad \text{if} \quad N \gtrsim_{id} \frac{Ln^{1+r}d^{2+\frac{5}{2}r+\mathfrak{e}}}{\varepsilon_n^2} \vee \frac{L^2n^rd^{3+\frac{7}{2}r+\frac{3}{2}\mathfrak{e}}}{\varepsilon_n^3} \vee \frac{L^2d^{3+4r+2\mathfrak{e}}}{n^2\varepsilon_n^4},$$

which concludes the proof. Using that $\mathfrak{b}_d \lesssim_{id} d$ and $r \in [0,1]$, the result follows. □

Proof of Theorem 2.9-ii). We repeat similar arguments and use Theorem 6.2-ii): a straightforward computation shows that:

$$\|\nabla^3 W_n\|_\infty \lesssim_{id} n\|\nabla^3 U(\xi,.)\|_\infty.$$

The bounds ①, ② and ④ derived above still hold. We then observe that the two other terms involved in Theorem 6.2-ii) are respectively:

$$③' = (n^{-(1+r)}\mathfrak{c}_2)^{2+\mathfrak{e}} d(n\mathfrak{b}_d)^{4r+2\mathfrak{e}}n^2\|\nabla^3 U(\xi,.)\|_\infty^2 d^4\gamma^2 \lesssim_{id} d^{5+4r+2\mathfrak{e}}n^{2r+\mathfrak{e}}\gamma^2,$$

and

$$⑤ = (n^{-(1+r)}\mathfrak{c}_2)^{2(1+\mathfrak{e})}\frac{n^2L^2d^2(n\mathfrak{b}_d)^{2r+\mathfrak{e}}\gamma}{t_N} \lesssim_{id} L^2d^{2+2r+\mathfrak{e}}N^{-1}.$$

We choose γ small enough to obtain an ε_n^2-approximation: under the condition $\mathfrak{b}_d \lesssim_{id} d$, we observe that

$$② \vee ③' \lesssim_{id} \varepsilon_n^2 \quad \text{if} \quad \gamma \lesssim_{id} \varepsilon_n \min\left(\frac{1}{Lnd^{1+r+\frac{\mathfrak{e}}{2}}}, \frac{1}{d^{\frac{5}{2}+2r+\mathfrak{e}} n^{r+\frac{\mathfrak{e}}{2}}}\right).$$

The constraints on N driven by ①, ④ and ⑤ then lead to (after several computations):

$$N \gtrsim_{id} \max\left(\frac{Ln^{1+r} d^{2+\frac{5}{2}r+\mathfrak{e}}}{\varepsilon_n^3}, \frac{n^{2r+\frac{\mathfrak{e}}{2}} d^{\frac{7}{2}(1+r)+\frac{3}{2}\mathfrak{e}}}{\varepsilon_n^3}, \frac{n^{r-2} d^{\frac{7}{2}+4r+2\mathfrak{e}}}{\varepsilon_n^3}, \frac{L^2 d^{2+2r+\mathfrak{e}}}{\varepsilon_n^2}\right).$$

$\square$

References

[BB05] M. Bagnoli and T. Bergstrom. Log-concave probability and its applications. *Economic theory*, 26(2):445–469, 2005. 2

[BBCG08] D. Bakry, F. Barthe, P. Cattiaux, and A. Guillin. A simple proof of the poincare inequality for a large class of probability measures including the log-concave case. *Electronic communications in probability*, (13):60–66, 2008. 2, 6, 23

[BE83] D. Bakry and M. Emery. Diffusions hypercontractives. *Séminaire de probabilités, XIX, Univ. Strasbourg, Springer*, 84:177–206, 1983. 14, 29

[BG99] S. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic sobolev inequalities. *J. Funct. Anal.*, 163(1), 1999. 14

[BGL14] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014. 6

[Bir83] L. Birgé. Approximation dans les espaces métriques et théorie de l'estimation. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, pages 181–237, 1983. 13

[BL97] S. Bobkov and M. Ledoux. Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields*, 107(3):383–400, 1997. 16

[Bob99] S. G. Bobkov. Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.*, 27(4):1903–1921, 1999. 2, 6

[Bub15] S. Bubeck. Convex optimization: Algorithms and complexity. 8(3,4):232, 2015. 13

[CCG12] P. Cattiaux, D. Chafai, and A. Guillin. Central limit theorems for additive functionals of ergodic markov diffusions. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 9:337–382, 2012. 5, 30

[CN91a] A. Caplin and B. Naelbuff. Aggregation and imperfect competition: On the existence of equilibrium. *Econometrica*, 59(1):25–59, 1991. 2

[CN91b] A. Caplin and B. Naelbuff. Aggregation and social choice: A mean voter theorem. *Econometrica*, 59(1):1–23, 1991. 2

[CvdV12] I. Castillo and A. van der Vaart. Needles and straw in a haystack: posterior concentration for possibly sparse sequences. *Ann. Stat.*, 40(4), 2012. 2, 13, 17, 18

[Dal17] A Dalalyan. Theoretical guarantees for approximate sampling from a smooth and log-concave density. *Journal of the Royal Statistical Society*, 79:651–676, 2017. 9, 15

[DCWY18] R. Diwivedi, Y. Chen, M. J. Wainwright, and B. Yu. Log-concave sampling: Metropolis-hastings algorithms are fast! *Preprint*, 2018. 15

[DM19] A. Durmus and E. Moulines. High-dimensional Bayesian inference via the unadjusted langevin algorithm. *Bernoulli*, 25(4A):2854, 2019. 5, 15

[DMM19] A. Durmus, S. Majewski, and B. Miasojedow. Analysis of langevin monte carlo via convex optimization. *Journal of Machine Learning Research*, 20:1, 2019. 16, 17

[DRDK19] A Dalalyan, L. Riou-Durand, and A. Karagulyan. Bounding the error of discretized langevin algorithms for non-strongly log-concave targets. *Preprint*, 2019. 16, 17

[EK05] N. Ethier and T. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, Series in Probability and Statistics, 2005. 22

[GGvdV00] S. Ghosal, J. K. Ghosh, and A. W. van der Vaart. Convergence rates of posterior distributions. *Ann. Stat.*, 28(2):500–531, 2000. 2, 13, 18

[GM96] P. Glynn and S. Meyn. A Liapounov bound for solutions of the Poisson equation. *Ann. of Probab.*, 24(2):916–931, 1996. 30

[GPP20] S. Gadat, F. Panloup, and P. Pellegrini. Supplementary materials: On the cost of Bayesian posterior mean strategy for log-concave models. 2020. 5, 9, 17, 18, 24, 26, 27, 28, 29, 30, 32, 34, 35, 36

[Has02] R. Has'minskii. *Stochastic Stability of Differential Equations*, volume 66 of *Stochastic Modelling and Applied Probability*. Springer, 2002. 22

[HH90] J.J. Heckman and B.E. Honore. The empirical content of the roy model. *Econometrica*, 58(5):1121–1149, 1990. 2

[HMP20] I. Honoré, S. Menozzi, and G. Pagès. Non-asymptotic gaussian estimates for the recursive approximation of the invariant measure of a diffusion. *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques, to appear*, 2020. 5, 30, 31

[IH81] I. Ibragimov and R. Has'minskii. *Statistical Estimation, Asymptotic Theory*. Stochastic Modelling and Applied Probability. Springer, 1981. 2

[KLS95] R. Kannan, L. Lovasz, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete Comput. Geom.*, 13:541–559, 1995. 2, 6, 14

[KS16] A. K. H. Kim and R. J. Samworth. Global rates of convergence in log-concave density estimation. *Ann. of Stat.*, 44(6):2756–2779, 2016. 2

[LC86] L. M. Le Cam. *Asymptotic Methods in Statistical Decision Theory*. Springer, New York, 1986. 13, 18

[Led01] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001. 6, 14, 16

[LP02] D. Lamberton and G. Pagès. Recursive computation of the invariant distribution of a diffusion. *Bernoulli*, 8(3):367–405, 2002. 4, 5, 30, 31

[LP03] D. Lamberton and G. Pagès. Recursive computation of the invariant distribution of a diffusion: the case of a weakly mean reverting drift. *Stoch. Dyn.*, 3(4):435–451, 2003. 4

[LT88] J.J. Laffont and J. Tirole. Dynamics of incentive contracts. *Econometrica*, 56(5):1153–1175, 1988. 2

[MCC+19] Y. Ma, N. Chatterji, X. Cheng, N. Flammarion, P. Bartlett, and M. Jordan. Is there an analog of nesterov acceleration for mcmc? *Preprint*, 2019. 16, 17

[MCJ+19] Y. Ma, Y. Chen, C. Jin, N. Flammarion, and M. Jordan. Sampling can be faster than optimization. *Proceedings of the National Academy of Sciences (PNAS)*, 116:20881–20885, 2019. 15

[Mey71] P.A. Meyer. Solutions de l'équation de poisson dans le cas récurrent. *Séminaire de probabilités de Strasbourg*, 5:251–269, 1971. 30

[MHW+19] W. Mou, N. Ho, M.J. Wainwright, P. Bartlett, and M.I. Jordan. A diffusion process perspective on posterior contraction rates for parameters. *Preprint*, 2019. 5, 13, 14, 15

[Nes04] Y. Nesterov. *Introductory Lectures on Convex Optimization. A basic course.* Applied Optimization. Kluwer Academic Publishers, Boston, MA, 2004. 13

[Nev76] J. Neveu. Potentiel markovien recurrent des chaines de harris. *Ann. Inst. Fourier (Grenoble)*, 22:7–130, 1976. 30

[PP12] Gilles Pagès and Fabien Panloup. Ergodic approximation of the distribution of a stationary diffusion: Rate of convergence. *Ann. Appl. Probab.*, 22(3):1059–1100, 06 2012. 4

[PP18] Gilles Pagès and Fabien Panloup. Weighted multilevel Langevin simulation of invariant measures. *Ann. Appl. Probab.*, 28(6):3358–3417, 2018. 4

[Rev84] D. Revuz. *Markov Chains.* 2nd ed. North-Holland, Amsterdam, 1984. 30

[Sch65] L. Schwartz. On Bayes procedures. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 4(1), 1965. 2

[SW01] X. Shen and L. Wasserman. Rates of convergence of posterior distributions. *Ann. Stat.*, 29, 2001. 13

[SW14] A. Saumard and J.A. Wellner. Log-concavity and strong log-concavity: a review. *Statistics Surveys*, 8:45–113, 2014. 2

[Tal90] D. Talay. Second order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stoch. Stoch. Rep.*, 29(1):13–36, 1990. 4

[vdVvZ08] A. van der Vaart and J.H. van Zanten. Rates of contraction of posterior distributions based on gaussian process priors. *Ann. Stat.*, 36(3), 2008. 3, 13

[Wal09] G. Walther. Inference and modeling with log-concave distributions. *Statistical Science*, 24(3):319–327, 2009. 2

S. Gadat: Toulouse School of Economics, CNRS UMR 5314, Université Toulouse 1 Capitole, Esplanade de l'Université, Toulouse, France., ‡ Institut Universitaire de France.

F. Panloup: LAREMA-UMR CNRS 6093, Université d'Angers, 2, Bd Lavoisier, 49045 ANGERS Cedex 01.

C. Pellegrini: Institut de Mathématiques de Toulouse; UMR5219, UPS IMT, F-31062 Toulouse Cedex 9, France.

Email address: sebastien.gadat(at)tse-fr.eu, fabien.panloup(at)univ-angers.fr, clement.pellegrini(at)math.univ-toulouse.fr

SUPPLEMENTARY MATERIALS: ON THE COST OF BAYESIAN POSTERIOR MEAN STRATEGY FOR LOG-CONCAVE MODELS

Abstract. This article is a companion paper of [GPP20] that contains a series of (usually technical) proofs of the main document.

1. Concentration results

In this paragraph, we provide the technical proofs related to the posterior mean concentration, stated in Section 3 of the paper [GPP20].

Proof of Proposition 2.11. We consider $\theta \in \mathbb{R}^d$ and $f \in L^2(\pi_\theta)$. For any $\varepsilon > 0$, a density argument proves that a $f_\varepsilon \in \mathcal{C}_K^\infty(\mathbb{R}^d, \mathbb{R})$, *e.g.*, a compactly supported and infinitely differentiable function, exists such that $\pi_\theta((f - f_\varepsilon)^2) \leqslant \varepsilon^2$ and $\pi_\theta(|\nabla f - \nabla f_\varepsilon|^2) \leqslant \varepsilon^2$.

We shall remark that if $f_\varepsilon^{-\theta} : x \longmapsto f_\varepsilon(x + \theta)$, then $\pi_\theta(f_\varepsilon) = \pi(f_\varepsilon^{-\theta})$. The function $f_\varepsilon^{-\theta}$ is infinitely differentiable and compactly supported, we shall apply the Poincaré inequality with the measure π:

$$Var_\pi(f_\varepsilon^{-\theta}) \leqslant C_P(\pi)\pi(|\nabla f_\varepsilon^{-\theta}|^2). \tag{1}$$

Now, a straigthforward change of variable yields:

$$Var_{\pi_\theta}(f_\varepsilon) = Var_\pi(f_\varepsilon^{-\theta}) \quad \text{and} \quad \pi(|\nabla f_\varepsilon^{-\theta}|^2) = \pi_\theta(|\nabla f_\varepsilon|^2).$$

We then deduce from the previous equalities and from (1) that

$$Var_{\pi_\theta}(f_\varepsilon) \leqslant C_P(\pi)\pi_\theta(|\nabla f_\varepsilon|^2). \tag{2}$$

Now, we end the proof with a density argument: the Cauchy-Schwarz inequality shows that

$$|\pi_\theta(f) - \pi_\theta(f_\varepsilon)| \leqslant \sqrt{\pi_\theta((f - f_\varepsilon)^2)} \leqslant \varepsilon, \quad |\pi_\theta(f^2) - \pi_\theta(f_\varepsilon^2)| \leqslant 2[\pi_\theta(f^2) + \pi_\theta(f_\varepsilon^2)]\varepsilon.$$

Finally, we can prove that $|Var_{\pi_\theta}(f_\varepsilon) - Var_{\pi_\theta}(f)| \leqslant 5\varepsilon[\pi_\theta(f^2) + \pi_\theta(f_\varepsilon^2)]$, and in the meantime $|\pi_\theta(|\nabla f_\varepsilon|^2) - \pi_\theta(|\nabla f|^2)| \leqslant \varepsilon\sqrt{2\pi_\theta(|\nabla f|^2 + |\nabla f_\varepsilon|^2)}$. We use these last upper bounds in (2), we obtain since ε may be chosen arbitrarily small, that:

$$Var_{\pi_\theta}(f) \leqslant C_P(\pi)\pi_\theta(|\nabla f|^2),$$

which ends the proof of the proposition. □

Proof of Proposition 3.1. The proof is straightforward as soon as we remark that Assumption $(\mathbf{PI_U})$ implies that each π_θ satisfies a Poincaré inequality with constant C_P^U. Since f satisfies $\|\nabla f\|_\infty \leqslant k$ and for any θ, U_θ is a convex coercive function, then $U_\theta(x)$ has a linear growth for large values of x. Hence, $f \in L^2(\pi_\theta)$ and we then simply apply the concentration inequality stated in Corollary 3.2 of [BL97]. This ends the proof of the proposition. □

Date: October 2, 2020.

Proof of Corollary 3.2. The proof of $i)$ is a straightforward application of Proposition 3.1 of [GPP20] with $f = \Psi$, which is a 1-Lipschitz function. The proof of $ii)$ is similar. Since $Z_\theta = 1$ for any θ, a direct integration yields:

$$\mathbb{E}_\theta[\nabla_\theta U(\xi,\theta)] = \int_{\mathbb{R}^d} \nabla_\theta U(\xi,\theta) e^{-U(\xi,\theta)} \mathrm{d}\xi = \nabla_\theta \int_{\mathbb{R}^d} e^{-U(\xi,\theta)} \mathrm{d}\xi = \nabla_\theta Z_\theta = 0.$$

We then use a union bound deduced by the triangle inequality: if $Z_i = \nabla_\theta U(\xi_i,\theta)$, then:

$$\left\{ \left| \frac{1}{n} \sum_{i=1}^n Z_i \right| \geqslant \delta \right\} \subset \bigcup_{j=1}^d \left\{ \frac{1}{n} \left| \sum_{i=1}^n Z_i^j \right| \geqslant \frac{\delta}{\sqrt{d}} \right\}. \tag{3}$$

We now apply Proposition 3.1 to each term in the union bound and we get that:

$$\mathbb{P}_\theta \left(\left| \frac{1}{n} \sum_{i=1}^n Z_i^j \right| \geqslant \frac{\delta}{\sqrt{d}} \right) \leqslant 2 e^{-n \frac{\delta^2}{4dL^2 C_P^U} \wedge \frac{\delta}{2L\sqrt{dC_P^U}}}.$$

The bound being independent of j, the result follows by summing over j in (3) $\square$

Finally, we derive the proof of the upper bounds related to the first and second type error of the family of tests (ϕ_n^r).

Proof of Proposition 3.3. The first upper bound $i)$ follows directly from Lemma 3.2 applied with $\theta^\star$. For the second estimation, we consider θ such that $|\theta - \theta^\star| \geqslant r_{a,n}$ and we get:

$$\begin{aligned}
&\left\{ \left| \frac{\sum_{i=1}^n \Psi(\xi_i)}{n} - \mathbb{E}_{\theta^\star}[\Psi(X)] \right| \leqslant \frac{c(r_{a,n})}{2} \right\} \\
&= \left\{ \left| \frac{\sum_{i=1}^n \Psi(\xi_i)}{n} - \mathbb{E}_{\theta}[\Psi(X)] + \mathbb{E}_{\theta}[\Psi(X)] - \mathbb{E}_{\theta^\star}[\Psi(X)] \right| \leqslant \frac{c(r_{a,n})}{2} \right\} \\
&\subset \left\{ |\mathbb{E}_{\theta}[\Psi(X)] - \mathbb{E}_{\theta^\star}[\Psi(X)]| - \left| \frac{\sum_{i=1}^n \Psi(\xi_i)}{n} - \mathbb{E}_{\theta}[\Psi(X)] \right| \leqslant \frac{c(r_{a,n})}{2} \right\} \\
&\subset \left\{ \left| \frac{\sum_{i=1}^n \Psi(\xi_i)}{n} - \mathbb{E}_{\theta}[\Psi(X)] \right| \geqslant c(|\theta - \theta^\star|) - \frac{c(r_{a,n})}{2} \right\} \\
&\subset \left\{ \left| \frac{\sum_{i=1}^n \xi_i}{n} - \mathbb{E}_{\theta}[X] \right| \geqslant \frac{c(r_{a,n})}{2} \right\}.
\end{aligned}$$

In the previous lines, we used the triangle inequality $|a+b| \geqslant |a| - |b|$ in the third line, the identifiability property $\mathbf{I_{W_1}(c)}$ and the fact that c is an increasing map. Applying again $i)$, Corollary 3.2, we obtain an upper bound of the probability of deviations uniform regarding the condition $|\theta - \theta^\star| \geqslant r_{a,n}$. Taking the supremum over θ, we then obtain the proof of $ii)$. $\square$

2. Continuous-time Langevin

This section is devoted to the proof of Proposition 4.4 of [GPP20].

Proof of Proposition 4.4. We start from $S_{n,t_0} = \int_0^{t_0} X_s^{(n)} \mathrm{d}s$ and compute as follows:

$$\begin{aligned}
\mathbb{E}_\mu\left(|S_{n,t_0} - t_0\widetilde{\theta}_n|^2\right) &\leqslant t_0 \int_0^{t_0} \mathbb{E}_\mu |X_s^{(n)} - \widetilde{\theta}_n|^2 \mathrm{d}s \quad \text{(Fubini and Cauchy-Schwarz inequality)} \\
&\leqslant t_0 \int_0^{t_0} \mathbb{E}_{\pi_n} |X_s^{(n)} - \widetilde{\theta}_n|^2 \mathrm{d}s + t_0 \int_0^{t_0} \left(\mathbb{E}_\mu |X_s^{(n)} - \widetilde{\theta}_n|^2 - \mathbb{E}_{\pi_n} |X_s^{(n)} - \widetilde{\theta}_n|^2\right) \mathrm{d}s \\
&\leqslant t_0^2 \mathbb{V}_{n,2} + t_0 \int_0^{t_0} \left(\mathbb{E}_\mu |X_s^{(n)} - \widetilde{\theta}_n|^2 - \mathbb{E}_{\pi_n} |X_s^{(n)} - \widetilde{\theta}_n|^2\right) \mathrm{d}s. \qquad (4)
\end{aligned}$$

In the last equality, we have used that π_n is invariant, *i.e.* $\mathcal{L}(X_s^{(n)} \,|X_0^{(n)} \sim \pi_n) = \pi_n$, which implies that $\mathbb{E}_{\pi_n} |X_s^{(n)} - \widetilde{\theta}_n|^2 = \mathbb{E}_{\pi_n} |X_0^{(n)} - \widetilde{\theta}_n|^2 = \pi_n(|I - \widetilde{\theta}_n|^2) = \mathbb{V}_{n,2}$.

We now study the second term of the right hand side of (4). We can write that:

$$\begin{aligned}
&\mathbb{E}_\mu\left(|X_s^{(n)} - \widetilde{\theta}_n|^2\right) - \mathbb{E}_{\pi_n}\left(|X_s^{(n)} - \widetilde{\theta}_n|^2\right) \\
&= \int |I - \widetilde{\theta}_n|^2(\theta) m_{\mu,s}^{(n)}(\theta) \mathrm{d}\pi_n(\theta) - \pi_n(|I - \widetilde{\theta}_n|^2) \\
&= \pi_n\left[|I - \widetilde{\theta}_n|^2 (m_{\mu,s}^{(n)} - \mathbf{1})\right] \\
&\leqslant \sqrt{\pi_n[|I - \widetilde{\theta}_n|^4]} \sqrt{\pi_n[(m_{\mu,s}^{(n)} - \mathbf{1})^2]} \quad \text{(Cauchy-Schwarz inequality)} \\
&\leqslant \sqrt{\mathbb{V}_{n,4}} e^{-\lambda_{1,n} s} \| m_{\mu,0}^{(n)} - 1 \|_{L^2(\pi_n)} \\
&\leqslant \sqrt{\mathbb{V}_{n,4}} e^{-\lambda_{1,n} s} \sqrt{J_{\mu,0}}. \qquad (5)
\end{aligned}$$

Introducing (5) in (4), and integrating between 0 and t_0, we deduce that

$$\begin{aligned}
\mathbb{E}_\mu\left(|S_{n,t_0} - t_0\widetilde{\theta}_n|^2\right) &\leqslant t_0^2 \mathbb{V}_{n,2} + t_0 \sqrt{\mathbb{V}_{n,4}} \sqrt{J_{\mu,0}} \int_0^{t_0} e^{-\lambda_{1,n} s} \mathrm{d}s \\
&\leqslant t_0^2 \mathbb{V}_{n,2} + t_0 \frac{\sqrt{\mathbb{V}_{n,4}} \sqrt{J_{\mu,0}}}{\lambda_{1,n}},
\end{aligned}$$

which yields the first part of the proposition.

The proof of the second part of the proposition is divided into three steps.

• *Step 1: Bias/Variance decomposition of $S_{n,t_0,t}$.* Using the Markov property we have

$$\mathbb{E}_\mu\left(\left|S_{n,t_0,t} - (t - t_0)\widetilde{\theta}_n\right|^2\right) = \mathbb{E}_{\mathbb{P}_\mu^{t_0}}\left(\left|\int_0^{t-t_0} (X_s^{(n)} - \widetilde{\theta}_n) \mathrm{d}s\right|^2\right)$$

We use some computations close to the ones of [CCG12], except that we need to handle an initialization of the process with a measure μ instead of π_n. We define $f : \theta \longmapsto \theta - \widetilde{\theta}_n$ and we remark that

$$\begin{aligned}\mathbb{E}_\mu\left(\left|\int_{t_0}^t (X_s^{(n)} - \widetilde{\theta}_n)\mathrm{d}s\right|^2\right) &= \mathbb{E}_\mu\left(\int_{t_0}^t f(X_s^{(n)})f(X_u^{(n)})\mathrm{d}u\mathrm{d}s\right)\\ &= \mathbb{E}_{\mathbb{P}_\mu^{t_0}}\left[\int_0^{t-t_0}\int_0^{t-t_0} \langle f(X_u^{(n)}), f(X_s^{(n)})\rangle \mathrm{d}u\mathrm{d}s\right],\end{aligned}$$

where we apply in the last line the Markov property. The Fubini relationship yields:

$$\begin{aligned}\mathbb{E}_\mu\left(\left|\int_{t_0}^t (X_s^{(n)} - \widetilde{\theta}_n)\mathrm{d}s\right|^2\right) &= 2\int_0^{t-t_0}\int_0^s \mathbb{E}_{\mathbb{P}_\mu^{t_0}}\langle f(X_u^{(n)}), f(X_s^{(n)})\rangle \mathrm{d}u\mathrm{d}s\\ &= 2\int_0^{t-t_0}\int_0^s \mathbb{E}_{\mathbb{P}_\mu^{t_0}}\langle f(X_u^{(n)}), P_{s-u}f(X_u^{(n)})\rangle \mathrm{d}u\mathrm{d}s\\ &= 2\int_0^{t-t_0}\int_0^{t-t_0-u} \mathbb{E}_{\mathbb{P}_\mu^{t_0}}\langle f(X_u^{(n)}), P_v f(X_u^{(n)})\rangle \mathrm{d}v\mathrm{d}u,\end{aligned}$$

by choosing $v = s - u$ as a change of variable. At this stage for all v and all u, we denote by g_v and ϕ_u the functions defined by:

$$g_v(y) = \langle f(y), P_v f(y)\rangle \qquad \text{and} \qquad \phi_u(y) = \int_0^{t-t_0-u} g_v(y)\mathrm{d}v.$$

We then remark that:

$$\begin{aligned}\mathbb{E}_\mu\left(\left|S_{n,t_0,t} - (t-t_0)\widetilde{\theta}_n\right|^2\right) &= 2\int_0^{t-t_0}\int_0^{t-t_0-u} \mathbb{E}_{\mathbb{P}_\mu^{t_0}} g_v(X_u^{(n)})\mathrm{d}v\mathrm{d}u\\ &=2\int_0^{t-t_0}\int_0^{t-t_0-u} \mathbb{E}_{\mathbb{P}_\mu^{t_0}}(P_u g_v)\mathrm{d}v\mathrm{d}u = \mathbb{E}_{\mathbb{P}_\mu^{t_0}}\left[2\int_0^{t-t_0}\int_0^{t-t_0-u} P_u g_v \mathrm{d}v\mathrm{d}u\right]\\ &=\mathbb{E}_{\mathbb{P}_\mu^{t_0}}\left[2\int_0^{t-t_0} P_u\left(\int_0^{t-t_0-u} g_v\mathrm{d}v\right)\mathrm{d}u\right] = \mathbb{E}_{\mathbb{P}_\mu^{t_0}}\left[2\int_0^{t-t_0} P_u\phi_u\mathrm{d}u\right].\end{aligned}$$

Again, let us define ψ by $\psi_t = 2\int_0^t P_u\phi_u\mathrm{d}u$ for all $t \geqslant 0$, we remark that:

$$\mathbb{E}_\mu\left(\left|S_{n,t_0,t} - (t-t_0)\widetilde{\theta}_n\right|^2\right) = \mathbb{E}_{\mathbb{P}_\mu^{t_0}}\left[\psi_{t-t_0}\right].$$

Now, we shall use the closeness of the measure $\mathbb{P}_\mu^{t_0}$ to the invariant distribution π_n when t_0 is sufficiently large, *i.e.*, use the closeness of $m_{\mu,s}^{(n)}$ to $\mathbf{1}$ for large values of s. We have:

$$
\begin{aligned}
\mathbb{E}_\mu\left(\left|S_{n,t_0,t}-(t-t_0)\widetilde{\theta}_n\right|^2\right) &= \pi_n(\psi_{t-t_0}) + \mathbb{E}_{\mathbb{P}_\mu^{t_0}}\left[\psi_{t-t_0}\right] - \pi_n(\psi_{t-t_0}) \\
&= \pi_n(\psi_{t-t_0}) + \int_{\mathbb{R}^d} \psi_{t-t_0}(y)(m_{\mu,t_0}^{(n)}(y)-\mathbf{1})\mathrm{d}\pi_n(y) \\
&\leqslant \pi_n(\psi_{t-t_0}) + \left[\pi_n(\psi_{t-t_0}^2)\right]^{\frac{1}{2}}\left[\int_{\mathbb{R}^d}(m_{\mu,t_0}^{(n)}(y)-\mathbf{1})^2\mathrm{d}\pi_n(y)\right]^{\frac{1}{2}} \\
&= \pi_n(\psi_{t-t_0}) + \left[\pi_n(\psi_{t-t_0}^2)\right]^{\frac{1}{2}}\sqrt{J_{\mu,t_0}} \\
&\leqslant \pi_n(\psi_{t-t_0}) + \left[\pi_n(\psi_{t-t_0}^2)\right]^{\frac{1}{2}}\sqrt{e^{-2\lambda_{1,n}t_0}J_{\mu,0}}, \qquad (6)
\end{aligned}
$$

where in the last line we used the contraction of the L^2 semi-group given by Theorem 4.3.
• *Step 2: First moment of ψ_{t-t_0}*. We upper bound $\pi_n(\psi_{t-t_0})$ using that $\mathcal{L}_n$ is self-adjoint:

$$
\begin{aligned}
\pi_n(\psi_{t-t_0}) &= 2\int_0^{t-t_0}\int_0^{t-t_0-u}\pi_n\left(\langle f(X_u^{(n)}), P_v f(X_u^{(n)})\rangle\right)\mathrm{d}v\mathrm{d}u \\
&= 2\int_0^{t-t_0}\int_0^{t-t_0-u}\pi_n\left(\langle f(X_u^{(n)}), P_{\frac{v}{2}}P_{\frac{v}{2}}f(X_u^{(n)})\rangle\right)\mathrm{d}v\mathrm{d}u \\
&= 2\int_0^{t-t_0}\int_0^{t-t_0-u}\pi_n\left(\mid P_{\frac{v}{2}}f(X_u^{(n)})\mid^2\right)\mathrm{d}v\mathrm{d}u \quad \text{because } \mathcal{L} \text{ is self-adjoint in} \quad L^2(\pi_n) \\
&= 2\int_0^{t-t_0}\int_0^{t-t_0-u}\pi_n\left(\mid P_{\frac{v}{2}}f(X_u^{(n)})\mid^2\right)\mathrm{d}v\mathrm{d}u.
\end{aligned}
$$

In order to control the term $\mid P_{v/2}f(X_u^{(n)})\mid^2$, we use the convergence to equilibrium of $P_{v/2}f$ and the fact that $\pi_n(f(X_u^{(n)})=0$. More precisely, using Theorem 4.3, we obtain that:

$$
\pi_n\left(\mid P_{v/2}f(X_u^{(n)})\mid^2\right) \leqslant e^{-v\lambda_{1,n}}\pi_n(|f|^2) = e^{-v\lambda_{1,n}}\mathbb{V}_{n,2}.
$$

Integrating this last inequality on the domain associated to (u,v), we obtain that

$$
\begin{aligned}
\pi_n(\psi_{t-t_0}) &\leqslant 4(t-t_0)\mathbb{V}_{n,2}J_{\delta_{\{x\}},0}\int_0^{\frac{1}{2}(t-t_0)} e^{-2\lambda_{1,n}s}\mathrm{d}s \\
&\leqslant \frac{2(t-t_0)\mathbb{V}_{n,2}}{\lambda_{1,n}}. \qquad (7)
\end{aligned}
$$

• *Step 3: Second order moment of ψ_{t-t_0}*. We now study $\pi_n(\psi_{t-t_0}^2)$, we can write that:

$$
\begin{aligned}
\pi_n(\psi_{t-t_0}^2) &= 4\int_0^{t-t_0}\int_0^{t-t_0} \pi_n\left(P_u\phi_u(y)P_s\phi_s(y)\right)\mathrm{d}u\mathrm{d}s \\
&\leqslant 4\int_0^{t-t_0}\int_0^{t-t_0} \sqrt{\pi_n[(P_u\phi_u)^2]}\sqrt{\pi_n[(P_s\phi_s)^2]}\mathrm{d}u\mathrm{d}s \quad \text{(using the Cauchy-Schwarz inequality)} \\
&\leqslant 4\int_0^{t-t_0}\int_0^{t-t_0} \sqrt{\pi_n[(P_u\phi_u^2)]}\sqrt{\pi_n[(P_s\phi_s^2)]}\mathrm{d}u\mathrm{d}s, \quad \text{(using the Jensen inequality)} \\
&= 4\int_0^{t-t_0}\int_0^{t-t_0} \sqrt{\pi_n(\phi_u^2)}\sqrt{\pi_n(\phi_s^2)}\mathrm{d}u\mathrm{d}s \quad \text{(because } \pi_n \text{ is an invariant distribution)} \\
&= 4\left[\int_0^{t-t_0}\sqrt{\pi_n(\phi_s^2)}\mathrm{d}s\right]^2.
\end{aligned}
$$

Let us control the term $\pi_n(\phi_s^2)$ following the same guideline as above:

$$
\begin{aligned}
\pi_n(\phi_u^2) &= \pi_n\left(\left[\int_0^{t-t_0-u} g_s\mathrm{d}s\right]^2\right) \\
&= \pi_n\left(\int_0^{t-t_0-u}\int_0^{t-t_0-u} g_{s_1}g_{s_2}\mathrm{d}s_1\mathrm{d}s_2\right) \\
&\leqslant \left[\int_0^{t-t_0-u}\sqrt{\pi_n(g_s^2)}\mathrm{d}s\right]^2.
\end{aligned}
$$

We are turned to upper bound $\pi_n(g_s^2)$. We compute that

$$
\begin{aligned}
\forall s\in[0,t-t_0-u] \qquad \pi_n(g_s^2) &= \pi_n(\langle f, P_s f\rangle^2) \\
&\leqslant \pi_n\left(|f|^2|P_s f|^2\right) \quad \text{(using the Cauchy-Schwarz inequality)} \\
&\leqslant \pi_n\left(|f|^2 P_s(|f|^2)\right) \quad \text{(using the Jensen inequality)} \\
&= \pi_n\left(P_{s/2}|f|^2 P_{s/2}(|f|^2)\right) \quad (\mathcal{L} \text{ and } P_s \text{ are self-adjoint in } \mathbb{L}^2(\pi_n) \\
&= \pi_n\left(\left(P_{s/2}(|f|^2)\right)^2\right) \\
&\leqslant \pi_n\left(P_{s/2}(|f|^4)\right) \quad \text{(using the Cauchy-Schwarz inequality).}
\end{aligned}
$$

Finally, rescaling the integral between 0 and $(t-t_0-u)/2$, we obtain that:

$$
\begin{aligned}
\pi_n(\phi_u^2) &\leqslant 4\left[\int_0^{\frac{t-t_0-u}{2}}\sqrt{\pi_n\left(P_s(|f|^4)\right)}\mathrm{d}s\right]^2 \\
&= 4\left(\sqrt{\pi_n(|f|^4)}\frac{t-t_0-u}{2}\right)^2 \quad \text{(because } \pi_n \text{ is invariant)} \\
&= \mathbb{V}_{n,4}(t-t_0-u)^2.
\end{aligned}
$$

Hence, we deduce that

$$\begin{aligned}\pi_n(\psi_{t-t_0}^2) &\leqslant 4\mathbb{V}_{n,4}\left[\int_0^{t-t_0}(t-t_0-u)\mathrm{d}u\right]^2 \\ &= \mathbb{V}_{n,4}(t-t_0)^4. \end{aligned}\tag{8}$$

Finally, using (7) and (8) in (6), we deduce that:

$$\mathbb{E}_\mu\left(\left|S_{n,t_0,t}-(t-t_0)\widetilde{\theta}_n\right|^2\right) \leqslant \frac{2(t-t_0)\mathbb{V}_{n,2}}{\lambda_{1,n}} + \sqrt{\mathbb{V}_{n,4}}(t-t_0)^2\sqrt{e^{-2\lambda_{1,n}t_0}J_{\mu,0}},\tag{9}$$

which ends the proof of the proposition. □

3. Proof of Corollary 2.7 - Cost of continuous time Langevin diffusion

The proof of the first implications of $i)$ and $ii)$ of Corollary 2.7 are straightforward. The non trivial technical ingredients are related to some upper bounds of the expectation with respect to the samples $\xi^{\mathbf{n}}=(\xi_1,\ldots,\xi_n)$ of several quantities. These controls are given below.

Proposition 3.1. *A constant $C>0$ exists such that*

$$\forall n\geqslant 1 \quad \mathbb{E}_{\theta^\star}[\mathbb{V}_{n,4}]\leqslant C\varepsilon_n^4.$$

Proof. We recall that $\mathbb{V}_{n,4}=\mathbb{E}_{\pi_n}|\theta-\widetilde{\theta}_n|^4=\int_{\mathbb{R}^d}|\theta-\widetilde{\theta}_n|^4\mathrm{d}\pi_n(\theta)$. We then use $(a+b)^4\leqslant 8(a^4+b^4)$ and remark that:

$$\mathbb{V}_{n,4}\leqslant 8\left(\mathbb{E}_{\pi_n}|\theta-\theta^\star|^4+\mathbb{E}_{\pi_n}|\theta^\star-\widetilde{\theta}_n|^4\right)=8\left(\mathbb{E}_{\pi_n}|\theta-\theta^\star|^4+|\theta^\star-\widetilde{\theta}_n|^4\right).$$

The Jensen inequality (see Section 3.2.2) yields: $\mathbb{E}_{\theta^\star}|\theta^\star-\widetilde{\theta}_n|^4\leqslant\mathbb{E}_{\theta^\star}\left[\mathbb{E}_{\pi_n}|\theta-\theta^\star|^4\right]$. Finally,

$$\mathbb{E}_{\theta^\star}[\mathbb{V}_{n,4}]\leqslant 16\mathbb{E}_{\theta^\star}\left[\mathbb{E}_{\pi_n}|\theta-\theta^\star|^4\right].$$

An integration by parts (see Section 3.2.2, proof of Theorem 2.3 of [GPP20]) leads to the desired result. □

In order to deal with the Poincaré constant $\lambda_{1,n}^{-1}$, we will use the result borrowed from Inequality (1.8) of [Bob99] on log-concave measures with the help of the Cheeger inequality. We then deduce the next proposition.

Proposition 3.2. *A constant K exists such that:*

$$\mathbb{E}_{\theta^\star}\left[\frac{1}{\lambda_{1,n}^2}\right]\leqslant 16K^4\mathbb{E}_{\theta^\star}\left[Var(\pi_n)^2\right]\lesssim\varepsilon_n^4,$$

and

$$\mathbb{E}_{\theta^\star}\left[\frac{1}{\lambda_{1,n}^4}\right]\leqslant 64K^8\mathbb{E}_{\theta^\star}\left[Var(\pi_n)^2\right]\lesssim\varepsilon_n^8.$$

Proof. We apply Inequality (1.8) of [Bob99] to the log-concave distribution π_n and obtain the sample-dependent inequality:

$$\frac{1}{\lambda_{1,n}}\leqslant 4K^2Var(\pi_n).$$

We then verify that

$$
\begin{aligned}
\mathbb{E}_{\theta^\star}\left[\frac{1}{\lambda_{1,n}^2}\right] &\leqslant 16K^4\mathbb{E}_{\theta^\star}\left[Var(\pi_n)^2\right] \\
&= 16K^4\mathbb{E}_{\theta^\star}\left(\mathbb{E}_{\pi_n}|\theta-\widetilde{\theta}_n|^2\right)^2 \\
&\leqslant 16K^4\mathbb{E}_{\theta^\star}\left(\mathbb{E}_{\pi_n}|\theta-\widetilde{\theta}_n|^4\right) \\
&= 16K^4\mathbb{E}_{\theta^\star}[\mathbb{V}_{n,4}].
\end{aligned}
$$

We then apply Proposition 3.1 and obtain the desired result. The second inequality proceeds along the same lines:

$$
\begin{aligned}
\mathbb{E}_{\theta^\star}\left[\frac{1}{\lambda_{1,n}^4}\right] &\leqslant 64K^8\mathbb{E}_{\theta^\star}\left[Var(\pi_n)^4\right] \\
&\lesssim \mathbb{E}_{\theta^\star}\left(\mathbb{E}_{\pi_n}|\theta-\widetilde{\theta}_n|^2\right)^4 \\
&\lesssim \mathbb{E}_{\theta^\star}\left(\mathbb{E}_{\pi_n}|\theta-\widetilde{\theta}_n|^8\right) \lesssim \varepsilon_n^8.
\end{aligned}
$$

□

Now we aim to bound the quantity $\tilde{W}_n(\theta) - \min \tilde{W}_n$ where we recall that

$$
\theta^{\widetilde{W}_n} = \arg\min_{\theta\in\mathbb{R}^d} \widetilde{W}_n(\theta). \tag{10}
$$

Using the convexity of U and $(\mathbf{C}^\beta)$, a constant $c_- > 0$ exists such that $\mathbb{P}_{\theta^\star} - a.s.$:

$$
c_-\left(\|\theta-\theta^{\widetilde{W}_n}\|^2 \wedge \|\theta-\theta^{\widetilde{W}_n}\|^\beta\right) \leqslant \frac{\widetilde{W}_n(\theta)-\min\widetilde{W}_n}{n}.
$$

Furthermore using that U is Lipschitz

$$
\begin{aligned}
\|\nabla\tilde{W}_n(\theta)\| &= \|\nabla\tilde{W}_n(\theta) - \nabla W_n(\theta^{\tilde{W}_n})\| \\
&= \|\sum U(X_i,\theta) - U(X_i,\theta^{\tilde{W}_n})\| \\
&\leqslant \sum \|U(X_i,\theta) - U(X_i,\theta^{\tilde{W}_n})\| \\
&\leqslant nL\|\theta-\theta^{\tilde{W}_n}\|,
\end{aligned}
$$

we then deduce that:

$$
\tilde{W}_n(\theta) - \min\tilde{W}_n \leqslant \frac{nL}{2}\|\theta-\theta^{\tilde{W}_n}\|^2.
$$

Finally we have

$$
c_-\left(\|\theta-\theta^{\widetilde{W}_n}\|^2 \wedge \|\theta-\theta^{\widetilde{W}_n}\|^\beta\right) \leqslant \frac{\widetilde{W}_n(\theta)-\min\widetilde{W}_n}{n} \leqslant \frac{L}{2}\|\theta-\theta^{\tilde{W}_n}\|^2, \tag{11}
$$

which will be used to prove the next Proposition.

Proposition 3.3 (Warm start - Size of $J_{\mu,0}$). *Assume that* $(\mathbf{C}^\beta)$ *holds with* $\beta \in [1,2]$ *and that* μ *is the uniform distribution over* $\mathcal{B}(\theta^{\widetilde{W}_n}, a)$ *with* $a < Cn^{-1/2}$*, then:*

$$J_{\mu,0} = \left\| \frac{d\mu}{d\pi_n} - \mathbf{1} \right\|^2_{L^2(\pi_n)} = \mathcal{O}_{id}(d^{d/\beta})$$

Proof. We observe that if μ is the uniform distribution on the ball centered at $\theta^{\widetilde{W}_n}$ and of radius a, denoted by $\mathcal{B}(\theta^{\widetilde{W}_n}, a)$, then $\mu(\theta) = \frac{1}{a^d V_d}\mathbf{1}_{\mathcal{B}(\theta^{\widetilde{W}_n},a)}$ where V_d refers to the Lebesgue measure of the unit Euclidean ball: $V_d = \frac{\pi^{d/2}}{\Gamma(d/2+1)}$. We then decompose the L^2 loss as:

$$\begin{aligned} J_{\mu,0} &= \int \left(m^{(n)}_{\mu,0}(\theta) - 1\right)^2 \pi_n(\theta)\mathrm{d}\theta \\ &= Z_n \int \mu^2(\theta) e^{\widetilde{W}_n(\theta)}\mathrm{d}\theta - 1 \\ &= \frac{Z_n e^{\min(\widetilde{W}_n)}}{a^{2d}V_d^2} \int_{\mathcal{B}(\theta^{\widetilde{W}_n},a)} e^{\widetilde{W}_n(\theta) - \widetilde{W}_n(\theta^{\widetilde{W}_n})}\mathrm{d}\theta - 1. \end{aligned}$$

- Upper bound on $Z_n e^{\min(\widetilde{W}_n)}$. The lower bound of (11) implies that:

$$\begin{aligned} Z_n e^{\min(\widetilde{W}_n)} &= \int e^{-(\widetilde{W}_n(\theta) - \widetilde{W}_n(\theta^{\widetilde{W}_n})}\mathrm{d}\theta \\ &\leqslant \int e^{-c_- n[\|\theta-\theta^{\widetilde{W}_n}\|^2 \wedge \|\theta-\theta^{\widetilde{W}_n}\|^\beta]}\mathrm{d}\theta \\ &\leqslant \int e^{-c_- n\|\theta\|^2} + e^{-c_- n\|\theta\|^\beta}\mathrm{d}\theta. \end{aligned}$$

The polar change of variables $\theta \longmapsto \left(\|\theta\|, \frac{\theta}{\|\theta\|}\right) \in \mathbb{R}_+ \times S^{d-1}$ leads to:

$$Z_n e^{\min(\widetilde{W}_n)} \leqslant (c_- n)^{-d/2} V_d \int_0^{+\infty} d e^{-r^2} r^{d-1}\mathrm{d}r + (c_- n)^{-d/\beta} V_d \int_0^{+\infty} d e^{-r^\beta} r^{d-1}\mathrm{d}r.$$

Using a change of variable $u = r^\beta$ and the moments of the exponential distribution, we get:

$$\int_0^{+\infty} d r^{d-1} e^{-r^\beta}\mathrm{d}r = \int_0^{+\infty} d u^{(d+\beta-2)/\beta} e^{-u}\mathrm{d}u = d\Gamma((d+\beta-2)/\beta + 1)$$

Therefore, a constant κ (that could be made explicit in terms of c_- and β) exists such that:

$$Z_n \leqslant \kappa V_d \left[n^{-d/2}\Gamma((d+1)/2) + n^{-d/\beta}\Gamma((d+1)/\beta)\right].$$

- Upper bound on $\int_{\mathcal{B}(\theta^{\widetilde{W}_n},a)} e^{\widetilde{W}_n(\theta) - \widetilde{W}_n(\theta^{\widetilde{W}_n})}\mathrm{d}\theta$. Using the upper bound of (11), we have:

$$\begin{aligned} \int_{\mathcal{B}(\theta^{\widetilde{W}_n},a)} e^{\widetilde{W}_n(\theta) - \widetilde{W}_n(\theta^{\widetilde{W}_n})}\mathrm{d}\theta &\leqslant \int_{\mathcal{B}(\theta^{\widetilde{W}_n},a)} e^{\frac{L}{2}n\|\theta-\theta^{\widetilde{W}_n}\|^2}\mathrm{d}\theta \\ &\leqslant [e^{\frac{L}{2}na^2}]a^d V_d. \end{aligned}$$

• Conclusion. We gather the previous inequalities and obtain that:

$$
\begin{aligned}
J_{\mu,0} &\leqslant \frac{\kappa V_d\left[n^{-d/2}\Gamma((d+1)/2)+n^{-d/\beta}\Gamma((d+1)/\beta)\right]}{a^{2d}V_d^2} a^d V_d[e^{\frac{L}{2}na^2}] \\
&\leqslant \kappa a^{-d}\left[n^{-d/2}\Gamma((d+1)/2)+n^{-d/\beta}\Gamma((d+1)/\beta)\right][e^{\frac{L}{2}na^2}] \\
&\leqslant 2\kappa a^{-d}\left[n^{-d/2}\Gamma((d+1)/\beta)\right][e^{\frac{L}{2}na^2}].
\end{aligned}
$$

Choosing $a = n^{-1/2} < 1$, the Stirling formula leads to:

$$
J_{\mu,0} \lesssim_{id} a^{-d}n^{-d/2}(d)^{d/\beta}[e^{\frac{L}{2}na^2}] = \mathcal{O}_{id}(d^{d/\beta}).
$$

□

The proof of Corollary 2.7 then follows from the previous bounds and the Cauchy-Schwarz inequality (for $\mathbb{E}_{\theta^\star}[t^\star_{\varepsilon_n,n}]$). Concerning the second assertion, we verify that:

$$
\mathbb{E}_{\theta^\star}\left[t^{0,\star}_{\varepsilon_n,n}\right] \leqslant \mathbb{E}_{\theta^\star}\left[\frac{1}{2\lambda_{1,n}}\left[\log(\mathbb{V}_{n,4})+4\log(\varepsilon_n^{-1})+\log(J_{\mu,0})\right]\right]
$$

Now, remark that $\lambda_{1,n}^{-1} = \mathcal{O}_{id}(\sqrt{\mathbb{V}_{n,4}})$, which entails

$$
\mathbb{E}_{\theta^\star}\left[t^{0,\star}_{\varepsilon_n,n}\right] = \mathcal{O}_{id}\left(\mathbb{E}_{\theta^\star}\left[\mathbb{V}_{n,4}+\sqrt{\mathbb{V}_{n,4}}\right]+(\log(\varepsilon_n^{-1})+\log(J_{\mu,0}))\mathbb{E}_{\theta^\star}\left[\lambda_{1,n}^{-1}\right]\right),
$$

where we used that $\sqrt{x}\log(x) \leqslant x+\sqrt{x}$ for non-negative x. The Cauchy-Schwarz inequality associated with our previous intermediary results leads to:

$$
\mathbb{E}_{\theta^\star}[t^{0,\star}_{\varepsilon_n,n}] = \mathcal{O}_{id}\left(\varepsilon_n^2\left[\log(\varepsilon_n^{-1})+d\log d\right]\right) = \mathcal{O}_{id}\left(\varepsilon_n^2[d\log d+\log n]\right).
$$

4. Discretization tools

4.1. **Technical lemmas - Strongly convex discretization.**

Lemma 4.1. *Consider a constant sequence $\gamma_i = \gamma > 0$ for any integer i and associated cumulative sum sequence $(t_i)_{i\geqslant 0}$ given by $t_i = \sum_{j=k_0}^{i}\gamma_j = (i-k_0)\gamma$ with $\gamma a < 1$. Then:*

$$
\sum_{i=k_0+1}^{j}\gamma_i^m e^{-a(t_j-t_i)} \leqslant \frac{\gamma^{m-1}}{a}.
$$

Proof. The proof is straightforward using the explicit expression of the sequence $t_j - t_i = (j-i)\gamma$:

$$
\sum_{i=k_0+1}^{j}\gamma_i^m e^{-a(t_j-t_i)} = \gamma^m\sum_{i=k_0+1}^{j}e^{-a(j-i)\gamma} \leqslant \gamma^m\sum_{i=0}^{+\infty}e^{-a\gamma i} \leqslant \frac{\gamma^m}{1-e^{-\gamma a}}.
$$

If we choose now γ such that $\gamma a < 1$ and use $1-e^{-u} \geqslant u/2$ when $u \in [0,1]$, we obtain that:

$$
\sum_{i=k_0+1}^{j}\gamma_i^m e^{-a(t_j-t_i)} \leqslant \frac{\gamma^m}{a\gamma} \leqslant \frac{\gamma^{m-1}}{a}.
$$

□

The next lemma is useful for decreasing step-size sequences $\gamma_j = \gamma j^{-b}$ when $b \in (1/2, 1]$.

Lemma 4.2. *Consider a decreasing sequence $(\gamma_i)_{i\geqslant 0}$ $\gamma_j = \gamma j^{-b}$ when $b \in (1/2,1)$ and the associated cumulative sum sequence $(t_i)_{i\geqslant 0}$ given by $t_i = \sum_{k=0}^{i} \gamma_k$. Then for any $a>0$:*

$$\left| \sum_{i=k_0}^{n} \gamma_i^m e^{a(t_i-t_n)} - \frac{\gamma_n^{m-1}}{a} \right| \leqslant C(a,b,\gamma)(\gamma_n^m \wedge \gamma_n^{(m-1)+(1/b-1)}),$$

where $C(a,b,\gamma)$ only depends on a,b and γ.

Proof. We introduce $S_{n,m} = \sum_{i=k_0}^{n} \gamma_i^m e^{at_i}$ and observe that

$$\forall i \geqslant 0 \qquad e^{at_i} - e^{at_{i-1}} = e^{at_i}[1-e^{-a\gamma_i}] = a\gamma_i e^{at_i} + \delta_i e^{at_i},$$

where $\delta_i = 1 - e^{-a\gamma_i} - a\gamma_i$. We then use an Abel transform:

$$\begin{aligned} S_{n,m} = \sum_{i=k_0}^{n} \gamma_i^m e^{at_i} &= \sum_{i=k_0}^{n} \gamma_i^{m-1} \frac{1}{a}[a\gamma_i e^{at_i}] \\ &= \sum_{i=k_0}^{n} \left(\gamma_i^{m-1} \frac{1}{a}[e^{at_i} - e^{at_{i-1}}] - \gamma_i^{m-1}\delta_i e^{at_i} \right) \\ &= \frac{1}{a} e^{at_n}\gamma_n^{m-1} - \frac{1}{a} e^{at_{k_0}} \gamma_{k_0}^{m-1} + \sum_{i=k_0}^{n} e^{at_i}[\gamma_i^{m-1} - \gamma_{i+1}^{m-1}] - e^{at_i}\gamma_i^{m-1}\delta_i \end{aligned}$$

Now, observe that when $\gamma_i = \gamma_1 i^{-b}$, then

$$\gamma_i^{m-1} - \gamma_{i+1}^{m-1} \leqslant \gamma_i^{m-1} \frac{b(m-1)}{i},$$

whereas $\delta_i \leqslant 0$ and:

$$|\delta_i| \leqslant a^2 \gamma_i^2.$$

We obtain:

$$\begin{aligned} S_{n,m} = \sum_{i=k_0}^{n} \gamma_i^m e^{at_i} &\leqslant \frac{e^{at_n}\gamma_n^{m-1}}{a} + b(m-1) \sum_{i=k_0}^{n} \frac{\gamma_i^{m-1}}{i} e^{at_i} + \sum_{i=k_0}^{n} \gamma_i^{m+1} e^{at_i} \\ &= \frac{e^{at_n}\gamma_n^{m-1}}{a} + b(m-1)\gamma_1^{-1/b} S_{n,m+1/b-1} + S_{n,m+1}. \end{aligned}$$

Using that when $m' > m$, we have:

$$\lim_{n \longrightarrow +\infty} \frac{S_{n,m'}}{S_{n,m}} = 0,$$

we deduce that:

$$S_{n,m} = \frac{e^{at_n}\gamma_n^{m-1}}{a} + o\left(\frac{e^{at_n}\gamma_n^{m-1}}{a} \right).$$

Then, a comparison argument leads to the conclusion of the lemma. □

Lemma 4.3. *Consider a decreasing sequence $(\gamma_i)_{i\geqslant 0}$ $\gamma_j = \gamma j^{-1}$ and the associated cumulative sum sequence $(t_i)_{i\geqslant 0}$ given by $t_i = \sum_{k=0}^{i} \gamma_k$. Then for any $a>0$, a constant $C(a,\gamma)$ exists such that*

$$\forall m \in \mathbb{N}^\star \qquad \left| \sum_{i=k_0}^{n} \gamma_i^m e^{a(t_i-t_n)} - \frac{\gamma_n^{m-1}}{a} \right| \leqslant C(a,\gamma)\gamma_n^m.$$

Proof. We use that a constant c_0 exists such that:

$$t_i = \sum_{k=0}^{i} \gamma_k = \gamma \log(i) + c_0 + \gamma\varepsilon(i),$$

where $\varepsilon(i) = \mathcal{O}(1/i)$. Therefore, we have:

$$\begin{aligned}
\sum_{i=k_0}^{n} \gamma_i^m e^{a(t_i - t_n)} &= \gamma^m e^{-at_n} e^{ac_0} \sum_{i=k_0}^{n} i^{-m} e^{a\gamma \log(i)} e^{\gamma\varepsilon(i)} \\
&= \gamma^m e^{-at_n} e^{ac_0} \sum_{i=k_0}^{n} i^{-m+a\gamma} e^{\gamma\varepsilon(i)} = \gamma_n^{m-1} \frac{\gamma}{a\gamma + m - 1} + \mathcal{O}(\gamma_n^m).
\end{aligned}$$

□

Lemma 4.4. *Consider a positive decreasing sequence $(\gamma_i)_{i\geqslant 0}$ and the associated cumulative sum sequence $(t_i)_{i\geqslant 0}$ given by $t_i = \sum_{k=0}^{i} \gamma_k$. Then for any $a > 0$ such that $\gamma_1 a < 1$:*

$$\left| \sum_{i=k_0}^{n} \gamma_i e^{-a(t_i - t_{k_0})} \right| \leqslant \frac{1}{a} + a \sum_{i=k_0}^{n} \gamma_i^2.$$

Proof. We write:

$$e^{-at_i} = e^{-at_{i-1}} e^{-a\gamma_i},$$

which implies that:

$$e^{-at_i} - e^{-at_{i-1}} = e^{-at_i}[1 - e^{a\gamma_i}] = -a\gamma_i e^{-at_i} + \delta_i e^{-at_i} \qquad \text{with} \qquad \epsilon_i = 1 + a\gamma_i - e^{a\gamma_i}.$$

Now, observe that when $a\gamma_1 \leqslant 1$, then $\epsilon_i \leqslant a^2\gamma_i^2$. Then, a telescopic sum argument yields:

$$\begin{aligned}
\sum_{i=k_0}^{n} \gamma_i e^{-a(t_i - t_{k_0})} &= e^{at_{k_0}} \sum_{i=k_0}^{n} \frac{1}{a}[e^{-at_{i-1}} - e^{-at_i}] + \sum_{i=k_0}^{n} \frac{\epsilon_i}{a} e^{-a(t_i - t_{k_0})} \\
&\leqslant \frac{1}{a} + \sum_{i=k_0}^{n} \frac{|\epsilon_i|}{a} \\
&\leqslant \frac{1}{a} + a \sum_{i=k_0}^{n} \gamma_i^2.
\end{aligned}$$

□

Proof of Proposition 5.2. We define $F_{x,y}$ by $F_{x,y}(t) = \frac{1}{2}\mathbb{E}[|Z_t^x - \bar{Z}_t^y|^2]$ where the expectation is computed with respect to the Brownian motion trajectory $(B_t)_{t\geqslant 0}$. The Lebesgue and Ito theorems yield:

$$\begin{aligned}
F'_{x,y}(t) &= \mathbb{E}\left[\langle Z_t^x - \bar{Z}_t^y, b(Z_t^x) - b(y)\rangle\right] \\
&= \mathbb{E}\left[\langle Z_t^x - \bar{Z}_t^y, b(Z_t^x) - b(\bar{Z}_t^y)\rangle\right] + \mathbb{E}\left[\langle Z_t^x - \bar{Z}_t^y, b(\bar{Z}_t^y) - b(y)\rangle\right] \\
&\leqslant -\frac{\rho}{2} F_{x,y}(t) + \mathbb{E}\left[\langle Z_t^x - \bar{Z}_t^y, b(\bar{Z}_t^y) - b(y)\rangle\right], \qquad (12)
\end{aligned}$$

where in the last line we used (37). The Young inequality yields:

$$\mathbb{E}\left[\langle Z_t^x - \bar{Z}_t^y, b(\bar{Z}_t^y) - b(y)\rangle\right] \leqslant \frac{\rho}{4}\mathbb{E}|Z_t^x - \bar{Z}_t^y|^2 + \frac{2}{\rho}\mathbb{E}|b(\bar{Z}_t^y) - b(y)|^2.$$

Since b is L-Lipschitz and $\bar{Z}_t^y = y + tb(y) + \sqrt{2}B_t$, we obtain that:

$$\mathbb{E}|b(\bar{Z}_t^y) - b(y)|^2 \leqslant L^2 \mathbb{E}|\bar{Z}_t^y - y|^2 = L^2(t^2|b(y)|^2 + 2dt).$$

We obtain the differential inequality:

$$F'_{x,y}(t) \leqslant -\frac{\rho}{4} F_{x,y}(t) + \frac{L^2}{\rho} t^2 |b(y)|^2 + 2\frac{L^2}{\rho} dt.$$

A straightforward integration yields:

$$\begin{aligned}
F_{x,y}(t) &\leqslant F_{x,y}(0)e^{-\frac{\rho}{4}t} + \frac{L^2}{\rho}|b(y)|^2 \int_0^t e^{\frac{\rho}{4}(s-t)} s^2 \mathrm{d}s + 2\frac{L^2}{\rho} d \int_0^t e^{\frac{\rho}{4}(s-t)} s \mathrm{d}s \\
&\leqslant F_{x,y}(0)e^{-\frac{\rho}{4}t} + \frac{L^2}{\rho}|b(y)|^2 \int_0^t s^2 \mathrm{d}s + 2\frac{L^2}{\rho} d \int_0^t s \mathrm{d}s \\
&\leqslant |x-y|^2 e^{-\frac{\rho}{4}t} + \frac{L^2}{\rho} dt^2 + \frac{4}{3}\frac{L^2}{\rho}|b(y)|^2 t^3,
\end{aligned}$$

which concludes the proof. □

Proof of Proposition 5.3. Recall that we shall use

$$2\frac{\rho^2}{L}(W(x) - \min W) \leqslant |b(x)|^2 \leqslant \frac{2L^2}{\rho}(W(x) - \min W). \tag{13}$$

In the sequel for a vector x the notation x^t holds for the transpose of the vector x. Let us start with $\sup_{k\geqslant 0} \mathbb{E}|b(\bar{X}_{t_k})|^2$. Using a conditional expectation and a Taylor formula, the notations $\widetilde{W} = W - \min W$ and $\zeta_{k+1} = U_{k+1}/\sqrt{2\gamma_{k+1}} \sim \mathcal{N}(0,1)$ such that $\sqrt{2\gamma_{k+1}}\zeta_{k+1} = B_{t_{k+1}} - B_{t_k}$, we observe that:

$$\begin{aligned}
\mathbb{E}[\widetilde{W}(\bar{X}_{t_{k+1}})] &= \mathbb{E}\left[\widetilde{W}(\bar{X}_{t_k}) + \langle \nabla \widetilde{W}(\bar{X}_{t_k}), \gamma_{k+1} b(\bar{X}_{t_k}) + \sqrt{2\gamma_{k+1}}\zeta_{k+1}\rangle\right] \\
&\quad + \mathbb{E}\left[\frac{1}{2}[\gamma_{k+1} b(\bar{X}_{t_k}) + \sqrt{2\gamma_{k+1}}\zeta_{k+1}]^t \nabla_2 \widetilde{W}(\upsilon_{k+1})[\gamma_{k+1} b(\bar{X}_{t_k}) + \sqrt{2\gamma_{k+1}}\zeta_{k+1}]\right] \\
&\leqslant \mathbb{E}[\widetilde{W}(\bar{X}_{t_k})] - \gamma_{k+1}\mathbb{E}[|b(\bar{X}_{t_k})|^2] + \gamma_{k+1}^2 L \mathbb{E}[|b(\bar{X}_{t_k})|^2] + 2L\gamma_{k+1} d \\
&\leqslant \mathbb{E}[\widetilde{W}(\bar{X}_{t_k})]\left(1 - \gamma_{k+1}(1-\gamma_{k+1}L)\frac{2\rho^2}{L}\right) + 2L\gamma_{k+1} d \\
&\leqslant \mathbb{E}[\widetilde{W}(\bar{X}_{t_k})]\left(1 - \gamma_{k+1}\frac{\rho^2}{L}\right) + 2L\gamma_{k+1} d,
\end{aligned}$$

where in the last line we used that $2L\gamma_{k+1} \leqslant 1$. Using the definition of $\widetilde{W}$ and a straigthforward recursion, we obtain that:

$$\forall k \geqslant 0 \qquad \mathbb{E}[W(\bar{X}_{t_k})] \leqslant W(y) + 2d\frac{L^2}{\rho^2}. \tag{14}$$

Now we treat $\sup_{t\geqslant 0} \mathbb{E}|b(Z_t^x)|^2$. We apply the Ito formula to $u_x(t) = \mathbb{E}[\widetilde{W}(Z_t^x)] = \mathbb{E}[W(Z_t^x) - \min W]$ and obtain:

$$\begin{aligned}
\forall x \in \mathbb{R}^d \quad \forall t \geqslant 0: \qquad u'_x(t) &= \mathbb{E}[\langle b(Z_t^x), \nabla W(Z_t^x)\rangle + \Delta W(Z_t^x)] \\
&= -\mathbb{E}|\nabla W(Z_t^x)|^2 + \mathbb{E}\Delta W(Z_t^x) \leqslant -2\frac{\rho^2}{L} u_x(t) + dL,
\end{aligned}$$

where in the last line we used Inequality (13). A straightforward integration yields:

$$\forall t \geqslant 0 \qquad u_x(t) \leqslant u_x(0)e^{-2\frac{\rho^2}{L}t} + dLe^{-2\frac{\rho^2}{L}t}\int_0^t e^{2\frac{\rho^2}{L}s}\mathrm{d}s \leqslant \widetilde{W}(x) + \frac{dL^2}{2\rho^2}.$$

Inequality (13) implies that

$$\sup_{s\geqslant 0}\mathbb{E}|b(Z_s^x)|^2 \leqslant \frac{2L^2}{\rho}\sup_{s\geqslant 0}\mathbb{E}[\widetilde{W}(Z_s^x)] \leqslant \frac{2L^2}{\rho}\left[\widetilde{W}(x) + \frac{dL^2}{2\rho^2}\right].$$

□

4.2. **Technical lemmas - Weakly convex discretization.**

4.2.1. *Exponential bounds for the continuous-time process.*

Lemma 4.5. *Assume that* $(\mathbf{A})_{\mathbf{L}}$ *and* $(\mathbf{H}_{\mathfrak{c}_1})$ *hold and let* $\mathfrak{a} \in (0,1)$. *Then,*

$$\mathbb{E}_x[e^{\mathfrak{a}W(X_t)}] \leqslant e^{\mathfrak{a}W(x)}e^{-\tilde{\mathfrak{a}}t} + \beta_d(\mathfrak{a}),$$

where $\tilde{\mathfrak{a}} = \mathfrak{a}(1-\mathfrak{a})^{-1}$ *and* $\beta_d(\mathfrak{a}) := \beta(\mathfrak{a},(1-\mathfrak{a})^{-1},0)$ *is defined by* (20).

Proof. We apply the Ito formula to $f_{\mathfrak{a}}$ defined by $f_{\mathfrak{a}}(x) = e^{\mathfrak{a}W(x)}$ ($\mathfrak{a}$ has to be chosen):

$$\begin{aligned}\frac{Lf_{\mathfrak{a}}(x)}{f_{\mathfrak{a}}(x)} &= -\mathfrak{a}|\nabla W(x)|^2 + \mathfrak{a}^2|\nabla W(x)|^2 + \mathfrak{a}\Delta W(x)\\ &\leqslant \frac{\mathfrak{a}}{1-\mathfrak{a}}\left(-|\nabla W(x)|^2 + d(1-\mathfrak{a})^{-1}\bar{\lambda}_{\nabla^2 W}(x)\right).\end{aligned}$$

Setting $\tilde{\mathfrak{a}} = \mathfrak{a}(1-\mathfrak{a})^{-1}$ and $c = (1-\mathfrak{a})^{-1}$, we deduce from $(\mathbf{H}_{\mathfrak{c}_1})$ and $(\mathbf{A})_{\mathbf{L}}$ that:

$$\begin{aligned}Lf_{\mathfrak{a}}(x) &\leqslant -\tilde{\mathfrak{a}}f_{\mathfrak{a}}(x)1_{\{x\in\mathcal{C}^c_{c,\gamma_0}\}} + \tilde{\mathfrak{a}}f_{\mathfrak{a}}(x)\left(-|\nabla W(x)|^2 + d(1-\mathfrak{a})^{-1}\bar{\lambda}_{\nabla^2 W}(x)\right)1_{\{x\in\mathcal{C}_{c,\gamma_0}\}} && (15)\\ &\leqslant -\tilde{\mathfrak{a}}f_{\mathfrak{a}}(x) + \tilde{\mathfrak{a}}f_{\mathfrak{a}}(x)\left(1 + d(1-\mathfrak{a})^{-1}\bar{\lambda}_{\nabla^2 W}(x)\right)1_{\{x\in\mathcal{C}_{c,\gamma_0}\}} && (16)\\ &\leqslant -\tilde{\mathfrak{a}}f_{\mathfrak{a}}(x) + \tilde{\mathfrak{a}}\beta(\mathfrak{a},(1-\mathfrak{a})^{-1},0).\end{aligned}$$

It follows that $(M_t)_{t\geqslant 0}$ is a true martingale and the Gronwall lemma leads to:

$$\mathbb{E}_x[f_{\mathfrak{a}}(X_t)] \leqslant f_{\mathfrak{a}}(x)e^{-\tilde{\mathfrak{a}}t} + \beta(\mathfrak{a},(1-\mathfrak{a})^{-1},0)\int_0^s \tilde{\mathfrak{a}}e^{\tilde{\mathfrak{a}}(s-t)}ds \leqslant f_{\mathfrak{a}}(x)e^{-\tilde{\mathfrak{a}}t} + \beta(\mathfrak{a},(1-\mathfrak{a})^{-1},0).$$

□

4.2.2. *Exponential bounds for the continuous-time Euler scheme.*

Lemma 4.6. *Assume* $(\mathbf{H}_{\mathfrak{c}_1})$. *Assume that* $\gamma \leqslant \gamma_0 := (4dL+512)^{-1}$, *then*

$$\sup_{t\geqslant 0}\mathbb{E}_x[e^{\frac{1}{8}W(\bar{X}_t)}] \lesssim_{id} e^{\frac{1}{8}W(x)} + \bar{\beta}_d,$$

where $\bar{\beta}_d = \beta(1/8, 5, \gamma_0)$.

Remark 1. The proof of this result is rather technical and the important thing is to pay a specific attention to the dependency with respect to d and L (the weakest dependency with respect to d). As indicated in our statement, we are led to choose γ proportional to $(Ld)^{-1}$.

Proof. Using the Taylor formula, we get:

$$\begin{aligned} W(\bar{X}_{t_{k+1}}) &\leqslant W(\bar{X}_{t_k}) - \gamma|\nabla W(\bar{X}_{t_k})|^2 + \langle \nabla W(\bar{X}_{t_k}), \Delta_{k+1}\rangle \\ &+ \left(\int_0^1 \bar{\lambda}_{\nabla^2 W}(\bar{X}_{t_k}^{(\theta)})\mathrm{d}\theta\right)(\gamma^2|\nabla W(\bar{X}_{t_k})|^2 + |\Delta_{k+1}|^2), \end{aligned}$$

where $\Delta_{k+1} = \sqrt{2}(B_{t_{k+1}} - B_{t_k})$ and $\bar{X}_{t_k}^{(\theta)} = \bar{X}_{t_k} + \theta(-\gamma\nabla W(\bar{X}_{t_k}) + \Delta_{k+1})$. Let $\mathfrak{a} \in (0, 1/8]$. Setting $f_{\mathfrak{a}}(x) = e^{\mathfrak{a}W(x)}$ and using $(\mathbf{A})_{\mathbf{L}}$, we deduce that:

$$\mathbb{E}[f_{\mathfrak{a}}(\bar{X}_{t_{k+1}})|\mathcal{F}_{t_k}] \leqslant f_{\mathfrak{a}}(\bar{X}_{t_k})e^{(-\mathfrak{a}\gamma + L\mathfrak{a}\gamma^2)|\nabla W(\bar{X}_{t_k})|^2}\Psi_\gamma(\bar{X}_{t_k}), \tag{17}$$

where

$$\Psi_\gamma : x \longmapsto \mathbb{E}\exp\left(\mathfrak{a}\sqrt{2\gamma}\langle\nabla W(x), Z\rangle + 2\mathfrak{a}\gamma\left(\int_0^1 \bar{\lambda}_{\nabla^2 W}(x(\theta,\gamma,Z))\mathrm{d}\theta\right)|Z|^2\right),$$

with $Z \sim \mathcal{N}(0, I_d)$ and $x(\theta,\gamma,z) := x + \theta(-\gamma\nabla W(x) + \sqrt{\gamma}z)$. We decompose Ψ_γ into two parts:

$$\begin{aligned} \Psi_\gamma(x) &= \underbrace{\mathbb{E}\left[\exp\left(\mathfrak{a}\sqrt{2\gamma}\langle\nabla W(x), Z\rangle + 2\mathfrak{a}\gamma\left(\int_0^1 \bar{\lambda}_{\nabla^2 W}(x(\theta,\gamma,Z))\mathrm{d}\theta\right)|Z|^2\right)1_{\{|Z|^2 \leqslant 2\gamma^{-1}\}}\right]}_{:=\Psi_\gamma^{(1)}(x)} \\ &+ \underbrace{\mathbb{E}\left[\exp\left(\mathfrak{a}\sqrt{2\gamma}\langle\nabla W(x), Z\rangle + 2\mathfrak{a}\gamma\left(\int_0^1 \bar{\lambda}_{\nabla^2 W}(x(\theta,\gamma,Z))\mathrm{d}\theta\right)|Z|^2\right)1_{\{|Z|^2 > 2\gamma^{-1}\}}\right]}_{:=\Psi_\gamma^{(2)}(x)}. \end{aligned}$$

- Upper bound of $\Psi_\gamma^{(1)}(x)$. On $\{|Z|^2 \leqslant 2\gamma^{-1}\}$, $|x(\theta,\gamma,Z) - x| \leqslant \gamma|\nabla W(x)| + 2$, so that:

$$\begin{aligned} \Psi_\gamma^{(1)}(x) &\leqslant \mathbb{E}\left[\exp\left(\mathfrak{a}\sqrt{2\gamma}\langle\nabla W(x), Z\rangle + 2\mathfrak{a}\gamma\bar{\lambda}_{\mathrm{loc}}(\gamma,x)|Z|^2\right)\right] \\ &\leqslant \prod_{i=1}^d \mathbb{E}_{Z_1\sim\mathcal{N}(0,1)}[e^{\sqrt{2\gamma}\mathfrak{a}\partial_i W(x)Z_1 + 2\mathfrak{a}\gamma\bar{\lambda}_{\mathrm{loc}}(\gamma,x)Z_1^2}], \end{aligned}$$

where we used $\bar{\lambda}_{\mathrm{loc}}(\gamma,x) := \sup_{u\in B(x,\gamma|\nabla W(x)|+2)} \bar{\lambda}_{\nabla^2 W}(u)$. Using that

$$\forall \alpha_1 \in \mathbb{R} \quad \forall \alpha_2 < 1/2 \qquad \mathbb{E}_{Z_1\sim\mathcal{N}(0,1)}[e^{\alpha_1 Z_1 + \alpha_2 Z_1^2}] = \frac{1}{\sqrt{1-2\alpha_2}}e^{\frac{\alpha_1^2}{2(1-2\alpha_2)}}. \tag{18}$$

Since $4\gamma\bar{\lambda}_{\mathrm{loc}}(\gamma,x)\mathfrak{a} \leqslant \gamma L/2 < 1$, we deduce that:

$$\Psi_\gamma^{(1)}(x) = \left(\frac{1}{1-4\gamma\mathfrak{a}\bar{\lambda}_{\mathrm{loc}}(\gamma,x)}\right)^{\frac{d}{2}} e^{\frac{\gamma\mathfrak{a}^2|\nabla W(x)|^2}{1-4\mathfrak{a}\gamma\bar{\lambda}_{\mathrm{loc}}(\gamma,x)}} = \exp\left(-\frac{d}{2}\log(1-4\mathfrak{a}\gamma\bar{\lambda}_{\mathrm{loc}}) + \frac{\gamma\mathfrak{a}^2|\nabla W(x)|^2}{1-4\mathfrak{a}\gamma\bar{\lambda}_{\mathrm{loc}}}\right).$$

Choosing $\mathfrak{a} \leqslant \frac{1}{8}$ and using that $4\gamma(dL + 128) \leqslant 1$, we verify that $4\mathfrak{a}\gamma\bar{\lambda}_{\mathrm{loc}} \leqslant 1/4$. Since $\log(1-u) \geqslant -5/4u$ when $u \in [0, 1/4]$, we obtain that:

$$\Psi_\gamma^{(1)}(x) \leqslant \exp\left(\frac{5}{2}\mathfrak{a}\gamma d\bar{\lambda}_{\mathrm{loc}}(\gamma,x) + \frac{\gamma\mathfrak{a}|\nabla W(x)|^2}{7}\right). \tag{19}$$

- Upper bound of $\Psi_\gamma^{(2)}(x)$. We use the Cauchy-Schwarz inequality and obtain that:

$$
\begin{aligned}
\Psi_\gamma^{(2)}(x) &\leqslant \mathbb{E}\left[\exp\left(2\mathfrak{a}\sqrt{2\gamma}\langle\nabla W(x), Z\rangle + 4\mathfrak{a}\gamma L|Z|^2\right)\right]^{\frac{1}{2}}\left[\mathbb{P}\left(|Z|^2 \geqslant 2\gamma^{-1}\right)\right]^{\frac{1}{2}} \\
&= \exp\left(-\frac{d}{4}\log(1-8\mathfrak{a}\gamma L) + \frac{2\gamma\mathfrak{a}^2|\nabla W(x)|^2}{(1-8\mathfrak{a}\gamma L)}\right)\left[\mathbb{P}\left(\frac{|Z|^2}{d} - 1 \geqslant \frac{2}{\gamma d} - 1\right)\right]^{\frac{1}{2}} \\
&\leqslant \exp\left(\frac{d}{4}2(8\mathfrak{a}\gamma L) + \frac{\gamma\mathfrak{a}|\nabla W(x)|^2}{6}\exp\left(-\frac{d}{8}\left(\frac{2}{\gamma d} - 1\right)\right)\right]^{\frac{1}{2}} \\
&\leqslant \exp\left(\frac{\gamma d L}{2} + \frac{\gamma\mathfrak{a}|\nabla W(x)|^2}{6}\right)\exp\left(-\frac{1}{8\gamma} + \frac{d}{16}\right) \\
&\leqslant \exp\left(-\frac{\gamma^{-1}}{8}\left[1 - 4\gamma^2 dL - \frac{\gamma d}{2}\right] + \frac{\gamma\mathfrak{a}|\nabla W(x)|^2}{6}\right),
\end{aligned}
$$

where the last inequality comes from the Bennett inequality applied with $u = 2(\gamma d)^{-1} - 1 \geqslant 1$. Using that $4\gamma d(L+4) \leqslant 1$, we then obtain that:

$$
\Psi_\gamma^{(2)}(x) \leqslant \exp\left(-\frac{\gamma^{-1}}{32} + \frac{\gamma\mathfrak{a}|\nabla W(x)|^2}{6}\right). \tag{20}
$$

We then use (20) and (19) in (17) and obtain that:

$$
\begin{aligned}
&\mathbb{E}[f_\mathfrak{a}(\bar{X}_{t_{k+1}})|\mathcal{F}_{t_k}] \leqslant \mathbb{E}[f_\mathfrak{a}(\bar{X}_{t_k})]e^{(-\mathfrak{a}\gamma + L\mathfrak{a}\gamma^2)|\nabla W(\bar{X}_{t_k})|^2} \\
&\quad\times\left[\exp\left(\frac{5}{2}\mathfrak{a}\gamma d\bar{\lambda}_{\text{loc}}(\gamma, \bar{X}_{t_k}) + \frac{\gamma\mathfrak{a}|\nabla W(\bar{X}_{t_k})|^2}{7}\right) + \exp\left(-\frac{\gamma^{-1}}{32} + \frac{\gamma\mathfrak{a}|\nabla W(\bar{X}_{t_k})|^2}{6}\right)\right] \\
&\quad\leqslant \mathbb{E}[f_\mathfrak{a}(\bar{X}_{t_k})]e^{-\mathfrak{a}\gamma\left(1 - L\gamma - \frac{1}{6}\right)|\nabla W(\bar{X}_{t_k})|^2}\left(\exp\left(\frac{5}{2}\mathfrak{a}\gamma d\bar{\lambda}_{\text{loc}}(\gamma, \bar{X}_{t_k})\right) + \exp\left(-\frac{1}{32\gamma}\right)\right) \\
&\quad\leqslant \mathbb{E}[f_\mathfrak{a}(\bar{X}_{t_k})]\left[\exp\left(-\frac{1}{32\gamma}\right) + e^{\frac{\mathfrak{a}\gamma}{2}\left[-|\nabla W(\bar{X}_{t_k})|^2 + 5d\bar{\lambda}_{\text{loc}}(\gamma, \bar{X}_{t_k})\right]}\right],
\end{aligned}
$$

because $1 - \gamma L - \frac{1}{6} \geqslant 1 - \frac{1}{4} - \frac{1}{6} \geqslant \frac{1}{2}$.

Using our assumption $(\mathbf{H}_{\mathfrak{c}_1})$ and the notations introduced in (19), we know that:

$$
-|\nabla W(x)|^2 + 5d\bar{\lambda}_{\text{loc}}(\gamma, \bar{X}_{t_k})) \leqslant -1 \quad \text{on } \mathcal{C}^c_{5,\gamma_0}.
$$

We introduce $\rho_\gamma = e^{-\frac{\gamma\mathfrak{a}}{2}} + e^{-\frac{1}{32\gamma}}$. Thereore, we have:

$$
\begin{aligned}
&\mathbb{E}[f_\mathfrak{a}(\bar{X}_{t_{k+1}})|\mathcal{F}_{t_k}] \leqslant f_\mathfrak{a}(\bar{X}_{t_k})\left(e^{-\frac{1}{32\gamma}} + e^{\frac{\mathfrak{a}\gamma}{2}\left[-|\nabla W(\bar{X}_{t_k})|^2 + 5d\bar{\lambda}_{\text{loc}}(\gamma, \bar{X}_{t_k})\right]}[1_{\{\bar{X}_{t_k}\in\mathcal{C}^c_{5,\gamma_0}\}} + 1_{\{\bar{X}_{t_k}\in\mathcal{C}_{5,\gamma_0}\}}]\right) \\
&\quad\leqslant f_\mathfrak{a}(\bar{X}_{t_k})\left(e^{-\frac{1}{32\gamma}} + e^{-\frac{\gamma\mathfrak{a}}{2}}1_{\{\bar{X}_{t_k}\in\mathcal{C}^c_{5,\gamma_0}\}} + e^{\frac{\mathfrak{a}\gamma}{2}(-|\nabla W((\bar{X}_{t_k})|^2 + 5dL)}1_{\{\bar{X}_{t_k}\in\mathcal{C}_{5,\gamma_0}\}}\right) \\
&\quad\leqslant f_\mathfrak{a}(\bar{X}_{t_k})\left(e^{-\frac{1}{32\gamma}} + e^{-\frac{\gamma\mathfrak{a}}{2}} - e^{-\frac{\gamma\mathfrak{a}}{2}}1_{\{\bar{X}_{t_k}\in\mathcal{C}_{5,\gamma_0}\}} + e^{\frac{\mathfrak{a}\gamma}{2}(-|\nabla W((\bar{X}_{t_k})|^2 + 5dL)}1_{\{\bar{X}_{t_k}\in\mathcal{C}_{5,\gamma_0}\}}\right) \\
&\quad\leqslant \rho_\gamma f_\mathfrak{a}(\bar{X}_{t_k}) + f_\mathfrak{a}(\bar{X}_{t_k})\left(e^{\frac{5}{2}\mathfrak{a}\gamma dL} - e^{-\frac{\gamma\mathfrak{a}}{2}}\right)1_{\{\bar{X}_{t_k}\in\mathcal{C}_{5,\gamma_0}\}}. \\
&\quad\leqslant \rho_\gamma f_\mathfrak{a}(\bar{X}_{t_k}) + e^{-\frac{\gamma\mathfrak{a}}{2}}f_\mathfrak{a}(\bar{X}_{t_k})\left(e^{\frac{\gamma\mathfrak{a}}{2}(1+5dL)} - 1\right)1_{\{\bar{X}_{t_k}\in\mathcal{C}_{5,\gamma_0}\}}.
\end{aligned}
$$

On the first hand, using that $\mathfrak{a} = 1/8$, $\gamma \leqslant 1/512$ and the elementary inequalities, $\exp(-x) \leqslant 1 - \frac{3}{4}x$ on $[0, 1/5]$ and $\exp(-x) \leqslant 1/(800x)$ on $[16, +\infty)$, we get:

$$\rho_\gamma \leqslant 1 - \frac{3\gamma}{64} + \frac{32\gamma}{800} \leqslant 1 - c\gamma,$$

where c is a positive numerical constant independent of d and L. On the other hand, we observe that $\frac{\gamma\mathfrak{a}}{2}(1 + 5dL) = \frac{\gamma}{16}(1 + 5dL) \leqslant 1$. Using $e^x \leqslant 1 + 2x$ on $[0, 1]$, we get:

$$e^{\frac{\gamma}{16}(1+5dL)} - 1 \leqslant \frac{\gamma}{8}(1 + 5dL).$$

Thus, setting $v_k = \mathbb{E}[f_\mathfrak{a}(\bar{X}_{t_k})]$, we obtain that:

$$\forall k \geqslant 0 \qquad v_{k+1} \leqslant (1 - c\gamma)v_k + \gamma\bar{\beta}_d,$$

where $\bar{\beta}_d = \frac{\beta(1/8,5,\gamma_0)}{8}$. An induction leads to

$$\forall k \geqslant 1 \qquad v_k \leqslant \gamma\bar{\beta}_d \sum_{j=0}^{k-1}(1 - c\gamma)^j + (1 - c\gamma)^k v_0 = \gamma\bar{\beta}_d \frac{1 - (1 - c\gamma)^k}{1 - (1 - c\gamma)} + (1 - c\gamma)^k v_0.$$

We finally deduce that

$$\sup_{k \geqslant 0} \mathbb{E}_x[e^{\frac{1}{8}W(\bar{X}_{t_k})}] \lesssim_{id} e^{\frac{1}{8}W(x)} + \bar{\beta}_d.$$

To extend to any time $t \geqslant 0$, it is enough to write for any $t \in [t_k, t_{k+1}]$:

$$\mathbb{E}[f_\mathfrak{a}(\bar{X}_t)] = \mathbb{E}[\mathbb{E}[f_\mathfrak{a}(\bar{X}_t)|\mathcal{F}_{t_k}]],$$

and then to adapt the beginning of the proof. The details are left to the reader. □

We conclude this section by a useful technical result for our purpose.

Proposition 4.7. *Assume* $(\mathbf{H}_{\mathfrak{c}_1})$. *Assume that* $\gamma \leqslant (4dL + 512)^{-1}$ *and consider* $p > 0$. *If* $\min W > 0$, *then:*

$$\sup_{t \geqslant 0} \mathbb{E}_x[W^p(X_t)] + \sup_{t \geqslant 0} \mathbb{E}_x[W^p(\bar{X}_t)] \leqslant c2^p \left(W^p(x) + \mathfrak{b}_d^p\right),$$

where $\mathfrak{b}_d = \log(\bar{\beta}_d)$ *is defined in Equation* (20) *and* c *does not depend on* d *and* p.

Proof. We first assume that $p > 1$: We denote by $T = \exp(W(\bar{X}_t)/8)$ and observe that $\{W(\bar{X}_t)\}^p = 8^p \log^p(T)$. We then compute:

$$\begin{aligned}
\mathbb{E}_x[W^p(X_t)] &= 8^p\mathbb{E}_x[\log^p(T)] \\
&\leqslant 8^p\mathbb{E}_x[\log^p(T)\mathbf{1}_{T \geqslant e^{p-1}}] + 8^p\mathbb{E}_x[\log^p(T)\mathbf{1}_{T \leqslant e^{p-1}}] \\
&\leqslant 8^p\mathbb{E}_x[\log^p(T)\mathbf{1}_{T \geqslant e^{p-1}}] + 8^p(p - 1)^p \\
&\leqslant 8^p \int_{e^{p-1}}^{+\infty} \log^p(t)q(t)\mathrm{d}t + 8^p(p - 1)^p,
\end{aligned}$$

where q is the probabiliy density function of T. We introduce $Q_p = \int_{e^{p-1}}^{+\infty} q(u)\mathrm{d}u \in (0, 1)$ and observe that $\tilde{q} := q/Q_p$ is a probability density function on $[e^{p-1}, +\infty[$.

$$
\begin{aligned}
\mathbb{E}_x[W^p(X_t)] &\leqslant 8^p Q_p \int_{e^{p-1}}^{+\infty} \log^p(t)\tilde{q}(t)\mathrm{d}t + 8^p(p-1)^p \\
&\leqslant 8^p Q_p \log^p \left[\int_{e^{p-1}}^{+\infty} t\tilde{q}(t)\mathrm{d}t\right] + 8^p(p-1)^p
\end{aligned}
$$

where in the last line we applied the Jensen inequality to the concave function $x \mapsto (\log x)^p$ on $[e^{p-1}, +\infty)$ when $p > 1$. We then deduce that

$$
\begin{aligned}
\mathbb{E}_x[W^p(X_t)] &\leqslant 8^p Q_p \log^p \left[\int_{e^{p-1}}^{+\infty} tq(t)\mathrm{d}t\right] + 8^p Q_p \log^p[Q_p^{-1}] + 8^p(p-1)^p \\
&\leqslant 8^p \log^p \left(\mathbb{E}_x\left[\exp(W(\bar{X}_t)/8)\right]\right) + 8^p \max_{x\in(0,1)} x\log^p(x^{-1}) + 8^p(p-1)^p \\
&= 8^p \log^p \left(\mathbb{E}_x\left[\exp(W(\bar{X}_t)/8)\right]\right) + 8^p p^p e^{-p} + 8^p(p-1)^p \\
&\lesssim_{id} \log^p \left[e^{\frac{1}{8}W(x)} + \bar{\beta}_d\right] + 1,
\end{aligned}
$$

where we applied Lemma 4.6 in the last line and use that the other terms are independent from d. We then observe that

$$
\log^p \left[e^{\frac{1}{8}W(x)} + \bar{\beta}_d\right] \leqslant \left(\log\left[e^{\frac{1}{8}W(x)} + \bar{\beta}_d\right]\right)^p \leqslant \left(\log 2e^{\frac{1}{8}W(x)} + \log 2\bar{\beta}_d\right)^p \lesssim_{id} 1 + W^p(x) + \mathfrak{b}_d^p.
$$

It then implies the conclusion of the lemma. The case $p \in (0;1]$ is dealt with in the same way, the function $\log^p$ being concave on $(0, +\infty)$, and a similar argument permits to conclude. $\square$

4.2.3. *Analysis of the first and second variation processes.* We introduce the *first variation process* $Y^{\cdot\cdot} = (Y^{ij})_{1\leqslant i,j\leqslant d}$ defined for all $(i,j) \in \{1,\ldots,d\}^2$ by $Y_s^{ij} = \partial_{x_j}\{X_s^x\}^i$ where $\{X_s^x\}^i$ denotes the i^{th} component of X_s^x. $Y^{\cdot\cdot}$ is thus a matrix-valued process solution of the ordinary differential equation:

$$
Y_0^{\cdot\cdot} = I_d \qquad \text{and} \qquad \frac{dY_s^{\cdot\cdot}}{ds} = -\nabla^2 W(X_t)Y_t^{\cdot\cdot}. \tag{21}
$$

We also need to introduce the second variation process $Y^{\cdot\cdot\cdot} = (Y^{ijk})_{1\leqslant i,j,k\leqslant d}$ defined for all $(i,j,k) \in \{1,\ldots,d\}^3$ by $Y_s^{ijk} = \partial_{x_j \xi_k}\{X_s^x\}^i$. For all $(j,k) \in \{1,\ldots,d\}^2$, $Y^{\cdot jk} = (Y^{ijk})_{1\leqslant i\leqslant d}$ is a solution of

$$
Y_0^{\cdot jk} = 0 \quad \text{and} \quad dY_t^{\cdot jk} = -\left((Y_t^{\cdot j})^T \nabla^3 W(X_t) Y_t^{\cdot k} + \nabla^2 W(X_t) Y_t^{\cdot jk}\right) dt. \tag{22}
$$

Lemma 4.8. *(i) For all $j \in \{1,\ldots,d\}$, $|Y_t^{\cdot j}|^2 \leqslant e^{-2\int_0^t \lambda_W(X_s)\mathrm{d}s}$. In particular,*

$$
\|Y_t^{\cdot\cdot}\|_F^2 \leqslant de^{-2\int_0^t \lambda_W(X_s)\mathrm{d}s}.
$$

(ii) For all $(j,k) \in \{1,\ldots,d\}^2$,

$$
|Y_t^{\cdot jk}|^2 \lesssim_{id} d\|\nabla^3 W\|_\infty^2 (1+t^2) e^{-\int_0^t \lambda_W(X_u)\mathrm{d}u},
$$

where $\|\nabla^3 W\|_\infty = 1 \vee \sup_{1\leqslant i\leqslant d, x\in\mathbb{R}^d} \bar{\lambda}_{\nabla^3_{i\cdot\cdot}W(x)}$ ($\nabla^3_{i\cdot\cdot}W(x)$ is the symetric matrix defined for all $i \in \{1,\ldots,d\}$ by $\nabla^3_{i\cdot\cdot}W(x) = (D^3_{ijk}(x))_{1\leqslant j,k\leqslant d}$).

Proof. (i) By (21), for all $j \in \{1, \dots, d\}$, $dY_t^{\cdot j} = \nabla^2 W(X_t) Y_t^{\cdot j} dt$. Thus,

$$\begin{aligned}\frac{d|Y_t^{\cdot j}|^2}{dt} &= -2\langle Y_t^{\cdot j}, \nabla^2 W(X_t) Y_t^{\cdot j}\rangle \\ &\leqslant -2\underline{\lambda}_W(X_t)|Y_t^{\cdot j}|^2 dt.\end{aligned}$$

The first assertion easily follows from a Gronwall-type argument.

(ii) From (22), we deduce that:

$$\begin{aligned}\frac{d|Y_t^{\cdot jk}|^2}{dt} &\leqslant -2\langle Y_t^{\cdot jk}, (Y_t^{\cdot j})^T \nabla^3 W(X_t) Y_t^{\cdot k}\rangle - 2\underline{\lambda}_W(X_t)|Y_t^{\cdot jk}|^2 \\ &\leqslant -2\underline{\lambda}_W(X_t)|Y_t^{\cdot jk}|^2 + 2|(Y_t^{\cdot j})^T \nabla^3 W(X_t) Y_t^{\cdot k}| \times |Y_t^{\cdot jk}|. \end{aligned} \tag{23}$$

Remark that:

$$\begin{aligned}|(Y_t^{\cdot j})^T \nabla^3 W(X_t) Y_t^{\cdot k}| &= \sqrt{\sum_{i=1}^d \langle Y_t^{\cdot,j}, \nabla^3 W_{i,.,.}(X_t) Y_t^{\cdot,k}\rangle^2} \\ &\leqslant \sqrt{d}\|\nabla^3 W\|_\infty |Y_t^{\cdot j}| \times |Y_t^{\cdot k}| \\ &\leqslant \underbrace{\sqrt{d}\|\nabla^3 W\|_\infty e^{-\int_0^t \underline{\lambda}_W(X_s)\mathrm{d}s}}_{=:A_t},\end{aligned}$$

by (i). Thus, setting $\varphi_t = |Y_t^{\cdot jk}|^2$ and plugging the previous control into inequality (23) yields:

$$\varphi_t' \leqslant -2\underline{\lambda}_W(X_t)\varphi_t + A_t\sqrt{\varphi_t}.$$

Set $\Lambda_t = \int_0^t \underline{\lambda}_W(X_s)ds$. By Lemma 4.9, it follows that:

$$\begin{aligned}|Y_t^{\cdot jk}|^2 &\leqslant 2\sqrt{d}\|\nabla^3 W\|_\infty e^{-\Lambda_t}\int_0^t e^{-\Lambda_s} ds + d\|\nabla^3 W\|_\infty^2 e^{-2\Lambda_t}\left[\int_0^t e^{-\Lambda_s} ds\right]^2 \\ &\leqslant 2\sqrt{d}\|\nabla^3 W\|_\infty t e^{-\Lambda_t} + d\|\nabla^3 W\|_\infty^2 t^2 e^{-2\Lambda_t},\end{aligned}$$

by using that $\Lambda_s \geqslant 0$ on $[0, t]$. The result follows. ☐

Lemma 4.9. *Define $\Lambda_t = \int_0^t \lambda_s \mathrm{d}s$ and assume that $t \longmapsto \varphi_t$ satisfies $\varphi_0 = 0$ and*

$$\varphi_t' \leqslant -2\lambda_t\varphi_t + 2A_t\sqrt{\varphi_t}, \quad t \geqslant 0,$$

where $A_t \geqslant 0$ for any t, then:

$$\varphi_t \leqslant 2e^{-\Lambda_t}\int_0^t A_s e^{\Lambda_s}\mathrm{d}s + \mathrm{e}^{-2\Lambda_t}\left[\int_0^{\mathrm{t}} \mathrm{A_s e^{\Lambda_s} ds}\right]^2.$$

Proof. We consider $t \longmapsto \sqrt{1+\varphi_t}e^{\Lambda_t}$ and observe that:

$$\begin{aligned}\left(\sqrt{1+\varphi_t}e^{\Lambda_t}\right)' &= e^{\Lambda_t}\left[\frac{\varphi_t'}{2\sqrt{1+\varphi_t}} + \lambda_t\sqrt{1+\varphi_t}\right] \\ &\leqslant \frac{e^{\Lambda_t}}{\sqrt{1+\varphi_t}}\left[-\lambda_t\varphi_t + \sqrt{\varphi_t}A_t + \lambda_t + \lambda_t\varphi_t\right] \\ &\leqslant \lambda_t e^{\Lambda_t} + A_t e^{\Lambda_t}.\end{aligned}$$

A straitghtforward integration yields:

$$\sqrt{1+\varphi_t}e^{\Lambda_t}-1 \leqslant e^{\Lambda_t}-1+\int_0^t A_s e^{\Lambda_s}\mathrm{d}s,$$

which implies that $\sqrt{1+\varphi_t} \leqslant 1+e^{-\Lambda_t}\int_0^t A_s e^{\Lambda_s}\mathrm{d}s$. We then obtain the desired result. □

4.2.4. *Solution of Poisson equation. Bounds on the solution of the Poisson equation and its derivatives .* We remind that g is solution of the Poisson equation $\mathrm{Id}-\pi(\mathrm{Id})=\mathcal{L}g$ and $x^\star$ is a minimizer of W.

Proof of Proposition 6.1. Uniqueness: Consider two $\mathcal{C}^2$ solutions g_1 and g_2. Then, $\mathcal{L}(g_1-g_2)=0$ and:

$$\int (g_1-g_2)\mathcal{L}(g_1-g_2)\mathrm{d}\pi = -\int |\nabla(g_1-g_2)|^2\mathrm{d}\pi.$$

Since the operator $\mathcal{L}$ is elliptic, we know that the density of π is $a.s.$ positive so that g_1-g_2 is constant. The constraint $\pi(g_1)=\pi(g_2)=0$ implies that $g_1=g_2$.
Existence: Let $g_t(x)=\int_0^t \nu(f)-P_sf(x)\mathrm{d}s$. Following the arguments of Proposition 6.3 and its proof below (mainly the fact that the first and second variation processes go to 0 in L^1, sufficiently fast and locally uniformly in x), g is well-defined, of class $\mathcal{C}^2$ and, (g_t), Dg_t and D^2g_t converge locally uniformly to g, Dg and D^2g respectively. In particular, $\mathcal{L}g=\lim_{t\to+\infty}\mathcal{L}g_t$. Now, using that $\mathcal{L}$ is a linear operator (null on constant functions) and the Dynkin formula, we get:

$$\mathcal{L}g_t(x)=\int_0^t \mathcal{L}(\nu(f)-P_sf)(x)\mathrm{d}s = P_0f(x)-P_tf(x)=f(x)-P_t(f)(x)\xrightarrow{t\to+\infty} f(x)-\pi(f).$$

Then, $\mathcal{L}g(x)=\lim_{t\to+\infty}\mathcal{L}g_t(x)=f(x)-\pi(f)$ for every $x\in\mathbb{R}^d$ (see Proposition A.8 of [PP14] for a similar but more detailed proof).

□

Proof of Proposition 6.3. The fact that g is twice-differentiable is proved along the proof.
Proof of i). If the conditions of the Lebesgue differentiability are met (checked later on), then:

$$Dg(x)=\int_0^{+\infty}\mathbb{E}[\ddot{Y}_t]\mathrm{d}t.$$

By Lemma 4.8,

$$\mathbb{E}_x[\|\ddot{Y}_t\|_F^2] \leqslant d\mathbb{E}_x[e^{-2\int_0^t \underline{\lambda}_W(X_s)\mathrm{d}s}].$$

Thus, for every positive δ_1, we have

$$\mathbb{E}_x[\|\ddot{Y}_t\|_F^2] \leqslant de^{-2(1+t)^{\delta_1}}+d\mathbb{P}\left(\int_0^t \underline{\lambda}_{\nabla^2 W}(X_u)\mathrm{d}u \leqslant (1+t)^{\delta_1}\right).$$

Then, using $(\mathbf{H}_{\mathfrak{c}_2,r})$ and the Markov inequality, we get for every positive $\mathfrak{e}$,

$$\begin{aligned}\mathbb{E}_x[\|Y_t^{\cdot\cdot}\|_F^2] &\leqslant de^{-2(1+t)^{\delta_1}} + d\mathbb{P}\left(\int_0^t W^{-r}(X_u)\mathrm{d}u \leqslant \mathfrak{c}_2(1+t)^{\delta_1}\right)\\ &\leqslant de^{-2t^{\delta_1}} + d\mathbb{P}\left(\left(\int_0^t W^{-r}(X_u)\mathrm{d}u\right)^{-2(1+\mathfrak{e})} \geqslant (\mathfrak{c}_2(1+t)^{\delta_1})^{-2(1+\mathfrak{e})}\right)\\ &\leqslant de^{-2t^{\delta_1}} + d(\mathfrak{c}_2(1+t)^{\delta_1})^{2(1+\mathfrak{e})}\mathbb{E}\left[\left(\int_0^t W^{-r}(X_u)\mathrm{d}u\right)^{-2(1+\mathfrak{e})}\right].\end{aligned}$$

Since $x \mapsto x^{-2(1+\mathfrak{e})}$ is convex on $(0, +\infty)$, it follows from the Jensen inequality that

$$\mathbb{E}_x[\|Y_t^{\cdot\cdot}\|_F^2] \leqslant de^{-2t^{\delta_1}} + d(\mathfrak{c}_2(1+t)^{\delta_1})^{2(1+\mathfrak{e})}(1+t)^{-2(1+\mathfrak{e})}\sup_{u\geqslant 0}\mathbb{E}_x[W^{2r(1+\mathfrak{e})}(X_u)]. \tag{24}$$

Setting $\delta_1 = \mathfrak{e}/(2(1+\mathfrak{e}))$ and using again the Jensen inequality, we obtain

$$\mathbb{E}_x[\|Y_t^{\cdot\cdot}\|_F] \leqslant \sqrt{d}e^{-2t^{\frac{\mathfrak{e}}{1+\mathfrak{e}}}} + \sqrt{d}\mathfrak{c}_2^{1+\mathfrak{e}}(1+t)^{-1-\frac{\mathfrak{e}}{2}}\sup_{u\geqslant 0}\mathbb{E}_x[W^{r(1+\mathfrak{e})}(X_u)].$$

Using Corollary 4.7, we have

$$\sup_{u\geqslant 0}\mathbb{E}_x[W^{r(1+\mathfrak{e})}(X_u)] \lesssim_{id} W^{r(1+\mathfrak{e})}(x) + \mathfrak{b}_d^{r(1+\mathfrak{e})}.$$

so that

$$\mathbb{E}_x[\|Y_t^{\cdot\cdot}\|_F] \lesssim_{id} \sqrt{d}e^{-2t^{\frac{\mathfrak{e}}{1+\mathfrak{e}}}} + \sqrt{d}\mathfrak{c}_2^{1+\mathfrak{e}}(1+t)^{-1-\frac{\mathfrak{e}}{2}}\left(W^{r(1+\mathfrak{e})}(x) + \mathfrak{b}_d^{r(1+\mathfrak{e})}\right).$$

The above property has a series of consequences. First, it implies that g is well-defined. Actually,

$$\begin{aligned}|P_tf(x) - \pi(f)| &\leqslant \int |P_tf(x) - P_tf(y)|\pi(\mathrm{d}y) \leqslant \|\nabla \mathrm{f}\|_\infty \int \sup_{\mathrm{z}\in[\mathrm{x},\mathrm{y}]}\mathbb{E}_{\mathrm{z}}[\|\mathrm{Y}_{\mathrm{t}}^{\cdot\cdot}\|_{\mathrm{F}}]|\mathrm{y}-\mathrm{x}|\pi(\mathrm{dy})\\ &\leqslant C\max(e^{-2t^{\frac{\mathfrak{e}}{1+\mathfrak{e}}}}(1+t)^{-1-\frac{\mathfrak{e}}{2}})\int (W^{r(1+\mathfrak{e})}(x) + W^{r(1+\mathfrak{e})}(y)|y-x|\pi(\mathrm{d}y)\\ &\leqslant C_x\max(e^{-2t^{\frac{\mathfrak{e}}{1+\mathfrak{e}}}}(1+t)^{-1-\frac{\mathfrak{e}}{2}}).\end{aligned} \tag{25}$$

In the above inequalities, we used the convexity of W and the fact that π integrates functions with polynomial growth (simple consequence of Lemma 4.5). It also implies that the Lebesgue differentiability theorem applies. $Dg(x)$ is thus well-defined on $\mathbb{R}^d$ and:

$$\|Dg(x)\|_F^2 \leqslant \left(\int_0^{+\infty}\mathbb{E}_x[\|Y_t^{\cdot\cdot}\|_F]\mathrm{dt}\right)^2 \leqslant c_{\mathfrak{e}}d\left(1 + \mathfrak{c}_2^{2(1+\mathfrak{e})}\left(W^{2r(1+\mathfrak{e})}(x) + \mathfrak{b}_d^{2r(1+\mathfrak{e})}\right)\right).$$

For the second inequality of this assertion, we use Lemma 4.6. More precisely:

$$\mathbb{E}_x[\|Dg(\bar{X}_t)\|_F^2] \leqslant c_{\mathfrak{e}}d\left(1 + \mathfrak{c}_2^{2(1+\mathfrak{e})}\left(\mathbb{E}_x[W^{2r(1+\mathfrak{e})}(\bar{X}_t)] + \mathfrak{b}_d^{2r(1+\mathfrak{e})}\right)\right). \tag{26}$$

Then, we conclude using Corollary 4.7.

Proof of $ii-a)$. Write $x_s = x^\star + s(x-x^\star)$, the Taylor and the Cauchy-Schwarz formulas yield:

$$
\begin{aligned}
|g(x)-g(x^\star)| &\leqslant \sqrt{\sum_{i=1}^{d}\left(\int_0^1 \langle \nabla g_i(x_s), x-x^\star\rangle \mathrm{d}s\right)^2} \leqslant \left(\int_0^1 \|Dg(x_s)\|_F \mathrm{d}s\right)|x-x^\star| \\
&\leqslant c_{\mathfrak{e}}\sqrt{d}|x-x^\star|\left(1+\mathfrak{c}_2^{1+\mathfrak{e}}\left(\int_0^1 W^{r(1+\mathfrak{e})}(x_s)\mathrm{d}s + \mathfrak{b}_d^{r(1+\mathfrak{e})}\right)\right).
\end{aligned} \tag{27}
$$

Using that $\nabla W(x^\star)=0$, assumption $(\mathbf{H}_{\mathfrak{c}_2,r})$ and a Taylor formula on $(W(x_s))_{s\in[0,1]}$:

$$
W(x) \geqslant W(x^\star) + \mathfrak{c}_2|x-x^\star|^2\int_0^1 W^{-r}(x_s)\mathrm{d}s.
$$

The application $x\mapsto x^{-\frac{1}{1+\mathfrak{e}}}$ is convex and the Jensen inequality leads to:

$$
\mathfrak{c}_2|x-x^\star|^2\left(\int_0^1 W^{r(1+\mathfrak{e})}(x_s)\mathrm{d}s\right)^{-\frac{1}{1+\mathfrak{e}}} \leqslant W(x).
$$

Thus, on the one hand,

$$
\begin{aligned}
\mathfrak{c}_2^{1+\mathfrak{e}}|x-x^\star|\left(\int_0^1 W^{r(1+\mathfrak{e})}(x_s)\mathrm{d}s\right) &\leqslant W^{\frac{1}{2}}(x)\mathfrak{c}_2^{\frac{1}{2}+\mathfrak{e}}\left(\int_0^1 W^{r(1+\mathfrak{e})}(x_s)\mathrm{d}s\right)^{1+\frac{1}{2(1+\mathfrak{e})}} \\
&\lesssim_{id} \mathfrak{c}_2^{\frac{1}{2}+\mathfrak{e}}W^{\frac{1}{2}}(x)\left(W^{r(1+\mathfrak{e})}(x)\right)^{1+\frac{1}{2(1+\mathfrak{e})}} \lesssim_{id} \mathfrak{c}_2^{\frac{1}{2}+\mathfrak{e}}W^{\frac{1+3r}{2}+r\mathfrak{e}}(x)
\end{aligned}
$$

where in the second line, we used that $s\longmapsto W(x_s)$ is increasing because W is convex. On the other hand, with similar arguments,

$$
|x-x^\star| \leqslant \mathfrak{c}_2^{-\frac{1}{2}}W^{\frac{1}{2}}(x)\left(\int_0^1 W^{r(1+\mathfrak{e})}(x_s)\mathrm{d}s\right)^{\frac{1}{2(1+\mathfrak{e})}} \lesssim_{id} \mathfrak{c}_2^{-\frac{1}{2}}W^{\frac{1+r}{2}}(x).
$$

Proof of $ii-b)$. We apply the result of $ii-a)$ and obtain that:

$$
\mathbb{E}_{x_0}[|g(\bar{X}_t)-g(x^\star)|^2] \leqslant c_{\mathfrak{e}}d\left(\mathfrak{c}_2^{1+2\mathfrak{e}}\mathbb{E}_{x^\star}[W^{1+3r+r\mathfrak{e}}(\bar{X}_t)] + \mathfrak{b}_d^{r(1+\mathfrak{e})}\mathbb{E}_{x^\star}[W^{\frac{1+r}{2}}(\bar{X}_t)]\right).
$$

Then, by Corollary (4.7) and the fact that $W(x_0)\lesssim_{id}\mathfrak{b}_d$, we deduce that:

$$
\mathbb{E}_{x_0}[|g(\bar{X}_t)-g(x^\star)|^2] \leqslant c_{\mathfrak{e}}d\left(\mathfrak{c}_2^{1+2\mathfrak{e}}\mathfrak{b}_d^{1+3r+2r\mathfrak{e}} + \mathfrak{b}_d^{r(1+\mathfrak{e})}\mathfrak{b}_d^{\frac{1+r}{2}}\right) \leqslant c_{\mathfrak{e}}\mathfrak{c}_2^{1+2\mathfrak{e}}d\mathfrak{b}_d^{1+3r+2r\mathfrak{e}}.
$$

To deduce the result, we upper bound $|g(x_0)-g(x^\star)|$ by (27) and use that $W(x_0)\lesssim_{id}\mathfrak{b}_d$.
Proof of $iii)$. We use Lemma 4.8, which implies that:

$$
\mathbb{E}[|Y_t^{\cdot jk}|] \lesssim_{id} \sqrt{d}\|\nabla^3 W\|_\infty(1+t)\mathbb{E}[e^{-\frac{1}{2}\int_0^t \underline{\lambda}_W(X_u)\mathrm{d}u}].
$$

The rest of the proof follows the lines of $ii)$. More precisely: for every positive δ_1,

$$
\mathbb{E}[|Y_t^{\cdot jk}|] \lesssim_{id} \sqrt{d}\|\nabla^3 W\|_\infty(1+t)\left(e^{-t^{\delta_1}} + \mathbb{P}\left(\int_0^t \underline{\lambda}_{\nabla^2 W}(X_u)\mathrm{d}u \leqslant 2(1+\mathrm{t})^{\delta_1}\right)\right).
$$

Then, similarly as in (24), we deduce from $(\mathbf{H}_{\mathfrak{c}_2,r})$, from the Markov inequality and from the convexity of $x\mapsto x^{-2(1+\mathfrak{e})}$ that for every positive $\mathfrak{e}>0$,

$$
\mathbb{E}[|Y_t^{\cdot jk}|] \lesssim_{id} \sqrt{d}\|\nabla^3 W\|_\infty(1+t)\left(e^{-t^{\delta_1}}(\mathfrak{c}_2(1+t)^{\delta_1})^{2(1+\mathfrak{e})}(1+t)^{-2(1+\mathfrak{e})}\sup_{u\geqslant 0}\mathbb{E}[W^{2r(1+\mathfrak{e})}(X_u)]\right).
$$

Then, with $\delta_1 = \mathfrak{c}(2(1+\mathfrak{c}))^{-1}$, it yields:

$$\mathbb{E}[|Y_t^{\cdot jk}|] \lesssim_{id} \sqrt{d}\|\nabla^3 W\|_\infty \left((1+t)e^{-t^{\delta_1}}\mathfrak{c}_2^{2(1+\mathfrak{c})}(1+t)^{-1-\mathfrak{c}}\sup_{u\geqslant 0}\mathbb{E}[W^{2r(1+\mathfrak{c})}(X_u)]\right).$$

Thus, similarly as in (ii), we deduce from Corollary 4.7 that a constant C, which does not depend on d, j and k such that:

$$\mathbb{E}_x[|Y_t^{\cdot jk}|] \leqslant C\sqrt{d}\|\nabla^3 W\|_\infty \left((1+t)e^{-t^{\delta_1}} + \mathfrak{c}_2^{2(1+\mathfrak{c})}(1+t)^{-1-\mathfrak{c}}\left(W^{2r(1+\mathfrak{c})}(x) + \mathfrak{b}_d^{2r(1+\mathfrak{c})}\right)\right).$$

Then, applying the Lebesgue differentiability theorem, it yields:

$$|D^2_{jk}g(x)|^2 \leqslant c_{\mathfrak{c}} d\|\nabla^3 W\|_\infty^2 \left(1 + \mathfrak{c}_2^{2(1+\mathfrak{c})}\left(W^{4r(1+\mathfrak{c})}(x) + \mathfrak{b}_d^{4r(1+\mathfrak{c})}\right)\right),$$

and $iii-a$) follows. The last point comes from Corollary 4.7 (details are left to the reader). □

References

[BL97] S. Bobkov and M. Ledoux. Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields*, 107(3):383–400, 1997. 1

[Bob99] S. G. Bobkov. Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.*, 27(4):1903–1921, 1999. 7

[CCG12] P. Cattiaux, D. Chafai, and A. Guillin. Central limit theorems for additive functionals of ergodic markov diffusions. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 9:337–382, 2012. 3

[GPP20] S. Gadat, F. Panloup, and P. Pellegrini. On the cost of bayesian posterior mean strategy for log-concave models. 2020. 1, 2, 7

[KLS95] R. Kannan, L. Lovasz, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete Comput. Geom.*, 13:541–559, 1995.

[PP14] G. Pagès and F. Panloup. A mixed-step algorithm for the approximation of the stationary regime of a diffusion. *Stochastic Process. Appl.*, 124(1):522–565, 2014. 20