On the Ergodicity, Bias and Asymptotic Normality of Randomized Midpoint Sampling Method

Ye He
University of California, Davis
leohe@ucdavis.edu

Krishnakumar Balasubramanian
University of California, Davis
kbala@ucdavis.edu

Murat A. Erdogdu
University of Toronto and Vector Institute
erdogdu@cs.toronto.edu

Abstract

The randomized midpoint method, proposed by [40], has emerged as an optimal discretization procedure for simulating the continuous time underdamped Langevin diffusion. In this paper, we analyze several probabilistic properties of the randomized midpoint discretization method, considering both overdamped and underdamped Langevin dynamics. We first characterize the stationary distribution of the discrete chain obtained with constant step-size discretization and show that it is biased away from the target distribution. Notably, the step-size needs to go to zero to obtain asymptotic unbiasedness. Next, we establish the asymptotic normality of numerical integration using the randomized midpoint method and highlight the relative advantages and disadvantages over other discretizations. Our results collectively provide several insights into the behavior of the randomized midpoint discretization method, including obtaining confidence intervals for numerical integrations.

1 Introduction

We consider the problem of computing the following expectation

$$\mathbb{E}_\pi[\varphi(x)] \quad \text{where} \quad \pi(x) = \tfrac{1}{Z_f} e^{-f(x)}, \tag{1}$$

for a potential function $f : \mathbb{R}^d \to \mathbb{R}$ and a test function $\varphi : \mathbb{R}^d \to \mathbb{R}$, when the normalization constant $Z_f = \int e^{-f(x)} dx$ is unknown. This problem frequently arises in statistics and machine learning with numerous applications to high-dimensional Bayesian inference [45, 24, 30, 10], numerical integration [21, 19], volume computation [43], optimization and learning [37, 17, 32], graphical models [20], and molecular dynamics [34, 22]. Markov chain Monte Carlo (MCMC) methods provide a powerful framework for computing the integral in (1), and have been successfully deployed in various scientific fields [26].

In particular, MCMC algorithms that are based on diffusion processes have received a lot of attention recently. The fundamental idea behind such algorithms is that a continuous-time diffusion with its invariant measure as the target π is approximately simulated via a numerical sampler. The intuition behind the success of these methods is that by appropriately selecting the step-size parameter, the discrete approximation resulting from the numerical sampler tracks the continuous-time diffusion. Thus, rapid convergence properties of the diffusion process (see, for example, [38, 23, 14, 15, 27, 12]) is inherited by the discrete algorithm with an invariant measure that

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

is close to that of the diffusion, which is the target π. While a variety of diffusion processes can lead to a rich class of MCMC samplers, algorithms that are based on discretizing Langevin dynamics have been the primary focus of research due to their simplicity, accuracy, and well-understood theoretical guarantees in high-dimensional settings [6, 3, 4, 9, 42, 29, 3, 11, 13, 16].

Although motivated by the problem of computing the integral in (1), much of the theoretical focus on analyzing sampling methods in the recent literature has been on providing guarantees for the sampling problem itself (see [41] for an exception), i.e., the number of iterations needed to reach ϵ-neighborhood of a d-dimensional target distribution in some probability metric. The choice of step-size of the sampler is crucial to obtain such theoretical guarantees. While the problem of estimating expectations such as in (1) is based on sampling from the target π itself, the theoretical guarantees established for the sampling problem can provide very little to no information on computing the expectation in (1) based on the sampler. The main reason for this is, the step-size choice of the sampler required to obtain optimal theoretical guarantees for numerical integration of (1) turns out to be different from that of sampling. Furthermore, if the ultimate task is to perform inference on the quantity $\mathbb{E}_\pi[\varphi(x)]$, confidence intervals are required. Thus, one needs central limit theorems (CLT) to quantify the fluctuations of the estimator of the expectation in (1), depending on a specific numerical integrator being used.

The randomized midpoint method, a numerical sampler proposed by [40], has emerged as an optimal algorithm for sampling from strongly log-concave densities, achieving the information theoretical lower bound for this problem in terms of both dimension and tolerance dependency [1]. In lieu of this optimality result, one anticipates a superior performance from the randomized midpoint method in other fundamental problems that relies on a MCMC sampler as the main computation tool, e.g. estimating expectations of the form (1). However, properties of this sampler for the purpose of numerical integration, in particular its inferential properties, are not well-studied. In this paper, we explore various probabilistic properties of the randomized midpoint discretization method, when used as a numerical integrator. Towards that, we examine several results for the randomized midpoint method considering both the overdamped and underdamped Langevin diffusions. Our first contribution is the explicit characterization of the bias of the randomized midpoint numerical scheme, namely the difference between its stationary distribution and the target distribution π. We show that asymptotic unbiasedness, a desired property in general, can be achieved under a decreasing step-size sequence. As our principal contribution, we establish the ergodicity of the randomized midpoint method and prove a central limit theorem which can be leveraged for inference on the expectation (1). We compute the bias and the variance of the asymptotic normal distribution for various step size choices, and show that different step-size sequences are suitable for making inference in different settings.

Our Contributions. We summarize our contributions as follows:

1. We show the ergodicity of constant step-size (denoted as h) randomized midpoint discretization of the overdamped and underdamped Langevin diffusions in Theorems 1 and 3, respectively. For both cases, the stationary distribution π_h of the resulting discretized Markov chain is unique and is biased away from the target distribution π.

2. The choice of a constant step-size for the randomized midpoint discretization causes bias in sampling. We characterize this bias explicitly in Propositions 2.2 and 3.1 for the overdamped and underdamped Langevin diffusions, respectively. We show that Wasserstein-2 distance between π_h and π is of order $\mathcal{O}(h^{0.5})$ and $\mathcal{O}(h^{1.5})$ respectively for the overdamped and underdamped Langevin diffusions.

3. The established order of bias points toward using particular choices of decreasing step-size sequence for the sake of inference. Specifically, we prove a CLT for numerical integration using the randomized midpoint discretization of the overdamped and underdamped Langevin diffusions in Theorems 2 and 4 respectively, for various choices of decreasing step-size. Depending on the specific choice of step-size sequence, the CLT is either unbiased or biased. When discretizing the overdamped Langevin diffusion with polynomially decreasing step-size choices, the rate of unbiased CLT turns out to be $\mathcal{O}(n^{(1/3)-\epsilon})$ for any $\epsilon > 0$. But the optimal rate turns out to be $\mathcal{O}(n^{1/3})$ for which one can only obtain a biased CLT. When discretizing underdamped Langevin diffusions with polynomially decreasing step-size choices, we show that the optimal rate can be improved to $\mathcal{O}(n^{5/8})$ under a certain condition, which is satisfied only by the class of constant test functions.

1.1 Notations and Preliminaries

We denote an ℓ-th order symmetric tensor of dimension d by $A \in \mathbb{R}^{d\otimes\ell}$. For a given vector $u \in \mathbb{R}^d$, we use $\|u\|$ to denote the Euclidean-norm of the vector. We define the ℓ-th order rank-1 tensor formed from $u \in \mathbb{R}^d$ as $u^{\otimes\ell}$. In addition, let A and B be two ℓ-th order tensors, we define the inner product between A and B as $\langle A, B\rangle = \sum_{j_1=1}^{d} \cdots \sum_{j_\ell=1}^{d} A_{j_1 j_2 \ldots j_\ell} \cdot B_{j_1 j_2 \ldots j_k}$. For a function $f : \mathbb{R}^d \to \mathbb{R}$, $\nabla f \in \mathbb{R}^d$ and $D^\ell \in \mathbb{R}^{d\otimes\ell}$ represents the gradient, and ℓ-th order derivative tensor (for $\ell > 1$). We let $(\Omega, \mathcal{F}, P)$ represent a probability space, and denote by $\mathcal{B}(\mathbb{R}^d)$, the Borel σ-field of $\mathbb{R}^d$. We use $\xrightarrow{d}$ and $\xrightarrow{p}$ to denote convergence in distribution and probability respectively. The set of all twice continuously differentiable functions $f : \mathbb{R}^d \to \mathbb{R}$ is denoted as $\mathcal{C}^2(\mathbb{R}^d)$. We use I_d to represent the $d \times d$ identity matrix. Let $x_0, x_1, \ldots$ be a d-dimensional Markov chain. The transition probability of the chain, at the k-th step is defined as $P^k(x, A) := P(x_k \in A | x_0 = x)$, for some $x \in \mathbb{R}^d$ and represents the probability that the chain is in set A at time n given the starting point was $x \in \mathbb{R}^d$. We use $\tilde{\mathcal{O}}$ to hide $\log$ factors. Finally, for a sequence γ_k and positive integer ℓ, we define $\Gamma_n^{(\ell)} := \sum_{k=1}^{n} \gamma_k^\ell$. We also make the following assumption on the potential function.

Assumption 1.1. *The potential function $f \in \mathcal{C}^2(\mathbb{R}^d)$ satisfies the following properties. For some $0 < m \leq M < \infty$: (a) f has a M-Lipschitz gradient; that is, $D^2 f \preceq M I_d$, and (b) f is m-strongly convex; that is, $m I_d \preceq D^2 f$. We also define the condition number as $\kappa := M/m$.*

2 Results for the Overdamped Langevin Diffusion

The overdamped Langevin diffusion is described by the following stochastic differential equation:

$$dx(t) = -\nabla f(x(t))dt + \sqrt{2}dW(t), \tag{2}$$

where $W(t)$ is a d-dimensional Brownian motion. It is well-known that this diffusion has $\pi(x) \propto e^{-f(x)}$ as its stationary distribution under mild regularity conditions. In general, simulating a continuous-time diffusion such as (2) is impractical; thus, a numerical integration scheme is needed.

We now describe the *randomized midpoint discretization* of the above diffusion in (2), which we denote as RLMC. Denoting the n-th iteration of the algorithm with x_n, the integral formulation of the diffusion with x_n as the initial value would then be $x_n^*(t) = x_n - \int_0^t \nabla f(x_n^*(s))ds + \sqrt{2}W(t)$. Let $h > 0$ be the choice of step size for the discretization and, let (α_n) be an i.i.d. sequence of random variables following uniform distribution on $[0, 1]$, i.e. $\alpha_n \sim U[0, 1]$. The fundamental idea behind the randomized midpoint technique is to use $h\nabla f(x_n^*(\alpha_{n+1}h))$ to approximate the integral $\int_0^h \nabla f(x_n^*(s))ds$. Indeed, notice that $\mathbb{E}[h\nabla f(x_n^*(\alpha_{n+1}h))] = h\int_0^1 \nabla f(x_n^*(\alpha h))d\alpha = \int_0^h \nabla f(x_n^*(s))ds$. RLMC proceeds by approximating $x_n^*(\alpha_{n+1}h)$ with the Euler discretization, which ultimately yields an explicit numerical integration step. Although [40] considered this discretization only for the constant step-size choice and the *underdamped* Langevin diffusion (which we discuss in Section 3), below we present a single iteration of the RLMC algorithm with the choice of variable step-size γ_{n+1} for the overdamped diffusion in (2):

$$\begin{aligned} x_{n+\frac{1}{2}} &= x_n - \alpha_{n+1}\gamma_{n+1}\nabla f(x_n) + \sqrt{2\alpha_{n+1}\gamma_{n+1}}U'_{n+1}, \\ x_{n+1} &= x_n - \gamma_{n+1}\nabla f(x_{n+\frac{1}{2}}) + \sqrt{2\gamma_{n+1}}U_{n+1}, \end{aligned} \tag{RLMC}$$

where (U_n) and (U'_n) are sequences of i.i.d d-dimensional standard Gaussian vectors independent of (α_n) and the initial point x_0. We briefly digress now to make the following remark. If instead of $\alpha_n \sim U[0, 1]$, one uses $\alpha_n = 1$ for all n deterministically, then the iterates of (RLMC) algorithm is reminiscent of the extra-gradient descent algorithm from the optimization literature [28], perturbed by Gaussian noise in each step. Furthermore, its noteworthy that with the deterministic choice of $\alpha_n = 1$, one cannot obtain the improved rates that the uniformly random α_n provides. Lastly, the filtration $(\mathcal{F}_n)$ is defined by $\mathcal{F}_n := \sigma(\alpha_k, U_k, U'_k; 1 \leq k \leq n)$, the smallest σ-algebra generated by the noise sequence and uniform random variables that are used in the first n iterations.

2.1 Wasserstein-2 Rates for Constant Step-size RLMC

Before, we state our main result, we investigate a few important characteristics of the (RLMC) algorithm that are not explored yet. We start with its rate of convergence in Wasserstein-2 distance

(see [44] for definition) for the (RLMC) algorithm. The proof of the proposition below essentially follows from a similar idea of the more general result for the underdamped Langevin dynamics in [40]. We include the result with its proof for the sake of completeness.

Proposition 2.1. *Suppose f satisfies Assumption 1.1. Set $x_0 = \arg\min_x f(x)$, $\gamma_n := h = \mathcal{O}(\epsilon^{2/3}/\kappa^{1/3}M)$ when $\kappa h M > 1$, and $\gamma_n := h = \mathcal{O}(\epsilon/M)$ when $\kappa h M \leq 1$ for some constant $C > 0$. After running the (RLMC) algorithm for*

$$K = \tilde{\mathcal{O}}\left(\frac{\kappa^{4/3}}{\epsilon^{2/3}} + \frac{\kappa}{\epsilon}\right) \quad \textit{steps,}$$

we have $W_2(\nu_K, \pi) \leq \epsilon\sqrt{d/m}$, where ν_K is the probability distribution of x_K.

When κ is of constant order, we see that W_2 rate is of order $\tilde{\mathcal{O}}(1/\epsilon)$. Notably, with the randomized midpoint technique, we obtain this particular ϵ-dependency by discretizing just the overdamped Langevin diffusion with only the Lipschitz gradient condition on the potential function f. Prior works require Euler-discretization of higher-order Langevin diffusions to obtain a W_2 rate of order $\tilde{\mathcal{O}}(1/\epsilon)$ [8, 35] or require higher-order smoothness assumption along with other specialized discretization methods [39, 25, 10, 8].

2.2 Analysis of the Markov Chain Generated by Constant Step-size RLMC

Using the randomized midpoint technique, we obtain an improved dependency on ϵ for the W_2 rate under weaker assumptions while discretizing the Langevin diffusion in (2). Although not explicit from the proof of Proposition 2.1, the rate improvement is obtained by a careful balancing of bias and variance through the choice of step-size parameter h. In this section, in Theorem 1, we first show that the (RLMC) Markov chain is ergodic and has a unique stationary distribution, denoted by π_h. Due to the choice of constant step-size h, it is not hard to see that the stationary distribution of the (RLMC) is different from the stationary distribution π of the Lanvegin diffusion in (2), i.e $\pi_h \neq \pi$. Hence, in Proposition 2.2, we characterize the Wasserstein-2 distance between π and π_h.

Firstly, if $f \in \mathcal{C}^2(\mathbb{R}^d)$ and f has a Lipschitz gradient with parameter M, then we can immediately see that the transition kernel of chain (x_n), $P(x, y) \in \mathcal{C}(\mathbb{R}^d \times \mathbb{R}^d)$ is positive everywhere. Therefore, it's easy to obtain that the chain (x_n) is μ^{Leb}-irreducible and aperiodic. Given all this information, we can give a sufficient condition to make sure that the chain has a unique invariant probability measure, and it is ergodic.

Theorem 1. *Let the potential function f satisfy part (a) of Assumption 1.1, and let $\gamma_n := h$ be small enough. Then the (RLMC) Markov chain (x_n) has a unique stationary probability measure π_h, and for every $x \in \mathbb{R}^d$, we have*

$$\sup_{A \in \mathcal{B}(\mathbb{R}^d)} |P^n(x, A) - \pi_h(A)| \to 0 \quad \textit{as} \quad n \to \infty.$$

We next address the question: how far is π_h from π? This question can be typically answered by a careful inspection on the proof of Proposition 2.1. However, for (RLMC), this is not the case, and requires using a different technique. Towards that, we derive an upper bound of $W_2(\pi, \pi_h)$ under the same assumptions in the previous theorem and the additional assumption that f is also strongly convex with parameter m.

Proposition 2.2. *Let the potential function satisfy Assumption 1.1, and let $\gamma_n := h \in (0, \frac{2}{m+M})$ in the (RLMC) algorithm. Then, we have*

$$W_2(\pi, \pi_h) \leq 3\sqrt{dh}\frac{(1 + 2Mh)^2}{\kappa^{-1} - Mh/\sqrt{3}}. \tag{3}$$

Remark 1. *The above proposition shows that the order of the bias between the stationary distribution of the Langevin diffusion and that of the (RLMC) chain is of the order $\mathcal{O}(\sqrt{h})$.*

2.3 Wasserstein-2 rates and CLT with Decreasing Step-size

In this part, we consider the (RLMC) algorithm with a fast decreasing time step sequence (γ_n) and establish a convergence rate in W_2 distance as well as a CLT for the numerical integration (1).

Proposition 2.3. *Suppose f satisfies Assumption 1.1. Let $x_0 := \arg\min_x f(x)$ and $\gamma_{n+1} \leq \frac{m}{m^2+M^2(33+n)}$. After running (RLMC) algorithm for $K = \mathcal{O}\left(\kappa^{1.5}/\epsilon\right)$ steps, we obtain $W_2(\nu_K, \pi) \leq \epsilon\sqrt{d/m}$, where ν_K is the probability distribution of x_K.*

Remark 2. *There are two aspects of this result. The first aspect is rather standard; there is no logarithmic factor in $1/\epsilon$ compared to the result in Proposition 2.1. Similar phenomenon has been previously observed for the LMC algorithm [8]. The second aspect is that we never obtain the $\mathcal{O}(1/\epsilon^{2/3})$ term as in Proposition 2.1, with the constant step-size choice. This is not an artifact of our analysis. This is due to the fact that with this choice of decreasing step-size, we reduce the bias much more at the expense of slightly increased variance. However, as we demonstrate next, this choice of decreasing step-size is crucial for obtaining an unbiased CLT for numerical integration.*

As the main contribution of this section, we characterize the fluctuations of (RLMC) when it is used for computing the integral $\int_{\mathbb{R}^d} \varphi d\pi$ for a π-integrable function φ. Choosing the Langevin diffusion in (2) with the stationary distribution π, we have by Theorem 1 that it is ergodic, and $\lim_{t\to+\infty} \frac{1}{t}\int_0^t \varphi(X(s))ds = \int_{\mathbb{R}^d} \varphi d\pi := \pi(\varphi)$, almost surely. Motivated by this, we first discretize the diffusion using (RLMC) and then compute a discrete analogue of the average. The procedure consists of two successive phases:

(a) **Discretization:** The (RLMC) algorithm is run with a step size sequence (γ_n) satisfying for all n, $\gamma_n > 0$, $\lim_{n\to+\infty} \gamma_n = 0$, and $\lim_{n\to+\infty} \Gamma_n = +\infty$, where $\Gamma_n := \sum_{k=1}^n \gamma_k$.

(b) **Averaging:** Using the (RLMC) iterates (x_n), construct a weighted empirical measure via the same weight sequence $\gamma := (\gamma_n)$: For every $n \geq 1$ and every $\omega \in \Omega$, set

$$\pi_n^\gamma(\omega, dx) := \frac{\gamma_1 \delta_{x_0(\omega)} + \cdots + \gamma_{k+1}\delta_{x_k(\omega)} + \cdots + \gamma_n \delta_{x_{n-1}(\omega)}}{\gamma_1 + \cdots + \gamma_n},$$

and use $\pi_n^\gamma(\omega, \varphi) := \int_{\mathbb{R}^d} \varphi \pi_n^\gamma(\omega, dx) = \frac{1}{\Gamma_n}\sum_{k=1}^n \gamma_k \varphi(x_{k-1}(\omega))$ to estimate the expectation (1).

For numerical purposes, for a fixed function φ, $\pi_n^\gamma(\omega, \varphi)$ can be recursively computed as follows:

$$\pi_{n+1}^\gamma(\omega, \varphi) = \pi_n^\gamma(\omega, \varphi) + \tilde{\gamma}_{n+1}\left(\varphi(x_n(\omega)) - \pi_n^\gamma(\omega, \varphi)\right) \quad \text{with } \tilde{\gamma}_{n+1} := \frac{\gamma_{n+1}}{\Gamma_{n+1}}.$$

We now provide the main result of this section, a central limit theorem for the algorithm (RLMC) when it is used to compute integrals of the form in (1).

Theorem 2. *Let π be such that its potential f satisfies Assumption 1.1. Consider a test function $\varphi : \mathbb{R}^d \to \mathbb{R}$ of the form $\varphi = \mathcal{A}\phi$ for some function $\phi : \mathbb{R}^d \to \mathbb{R}$, where $\mathcal{A}$ denotes the generator of the diffusion (2), i.e., $\mathcal{A}\phi := -\langle \nabla f, \nabla \phi\rangle + \Delta\phi$. Define $\hat{\gamma}_n := \frac{1}{\sqrt{\Gamma_n}}\sum_{k=1}^n \gamma_k^2$ and let $\hat{\gamma}_\infty = \lim_{n\to\infty} \hat{\gamma}_n$. Then for all $\phi \in \mathcal{C}^4(\mathbb{R}^d)$ with $D^2\phi$, $D^3\phi$ being bounded, and $D^4\phi$ being bounded and Lipschitz, and $\sup_{x\in\mathbb{R}^d} \|\nabla\phi(x)\|^2/(1+\|x\|^2) < +\infty$, we have the following central limit theorem for the numerical integration computed via (RLMC):*

(i) If $\hat{\gamma}_\infty = 0$, then $\sqrt{\Gamma_n}\pi_n^\gamma(\varphi) \xrightarrow{d} \mathcal{N}(0, 2\int_{\mathbb{R}^d} \|\nabla\phi(x)\|^2 \pi(dx))$,

(ii) If $\hat{\gamma}_\infty \in (0, +\infty)$, then $\sqrt{\Gamma_n}\pi_n^\gamma(\varphi) \xrightarrow{d} \mathcal{N}(\varrho\,\hat{\gamma}_\infty, 2\int_{\mathbb{R}^d} \|\nabla\phi(x)\|^2 \pi(dx))$,

(iii) If $\hat{\gamma}_\infty = +\infty$, then $\frac{\sqrt{\Gamma_n}}{\hat{\gamma}_n}\pi_n^\gamma(\varphi) \xrightarrow{p} \varrho$,

where the mean ϱ is given as

$$\begin{aligned}\varrho = &\int\int \langle D^3\phi(x), \nabla f(x) \otimes u \otimes u\rangle \mu(du)\pi(dx) - \tfrac{1}{2}\int \langle D^2 f(x), \nabla\phi(x) \otimes \nabla f(x)\rangle \pi(dx) \\ &+ \tfrac{1}{2}\int\int \langle D^3 f(x), \nabla\phi(x) \otimes u \otimes u\rangle \mu(du)\pi(dx) - \tfrac{1}{2}\int \langle D^2\phi(x), \nabla f(x) \otimes \nabla f(x)\rangle \pi(dx) \\ &- \tfrac{1}{6}\int\int \langle D^4\phi(x), u^{\otimes 4}\rangle \mu(du)\pi(dx),\end{aligned}$$

and μ is the distribution for a d-dimensional standard Gaussian measure.

Remark 3. *First note that a CLT for the Euler discretization of Langevin diffusion follows from [21, Thm. 10]. The rates of the CLT established in Theorem 2 are similar to that case, with only the bias term ρ being different. Specifically, following the same computation in [21], we see that the optimal*

rate with polynomially decaying step-size choice $\gamma_k = k^{-\alpha}$, for some $\alpha > 0$, is $\mathcal{O}(n^{1/3})$. But in this case, the established CLT is biased. However, for any $0 < \alpha < 1/3$, we obtain an unbiased CLT as well. Hence, although the (RLMC) chain provides rate improvements for sampling (with respect to W_2 distance), as demonstrated in [40] and in Proposition 2.1, it does not seem to provide any improvements for CLT. In retrospect, this is expected as the rate improvements for sampling is achieved by the choice of constant step-size for which it is not possible to establish even a nearly unbiased CLT.

The class of test functions that the above CLT can cover is intimately related to the solution of the *Stein equation* (or Poisson equation) $\varphi = \mathcal{A}\phi$. Given φ, there is an explicit characterization of ϕ that solves the Stein's equation, and various properties of φ are translated to ϕ [18, 17].

3 Results for the Underdamped Langevin Diffusion

The underdamped Langevin diffusion is given by

$$d\begin{bmatrix} x(t) \\ v(t) \end{bmatrix} = \begin{bmatrix} v(t) \\ -(\beta v(t) + u\nabla f(x(t))) \end{bmatrix} dt + \sqrt{2\beta u}\begin{bmatrix} 0_d \\ I_d \end{bmatrix} dW(t), \tag{4}$$

where $\beta > 0$ is the friction coefficient and $u > 0$ is the inverse mass. For simplicity, we will consider $\beta = 2$ in the later text. Under mild conditions, it is well-known that the continuous-time Markov process $(x(t), v(t))$ is positive recurrent, and its invariant distribution is given by $\nu(x, v) \propto \exp\left\{ - f(x) - \frac{1}{2u}\left\|v\right\|^2 \right\}$, $x \in \mathbb{R}^d$, $v \in \mathbb{R}^d$. This diffusion, with an additional Hamiltonian component, has gathered a lot of attention recently due to its improved convergence properties [7, 5, 40, 27, 12] and empirical performance [36, 2].

The *randomized midpoint discretization* of the underdamped Langevin diffusion (4) is given as:

$$\begin{aligned}
x_{n+\frac{1}{2}} &= x_n + \tfrac{1}{2}(1-e^{-2\alpha_{n+1}\gamma_{n+1}})v_n - \tfrac{u}{2}\left(\alpha_{n+1}\gamma_{n+1} - \tfrac{1}{2}(1-e^{-2\alpha_{n+1}\gamma_{n+1}})\right)\nabla f(x_n) + \sqrt{u}\sigma^{(1)}_{n+1}U^{(1)}_{n+1}, \\
x_{n+1} &= x_n + \tfrac{1}{2}(1-e^{-2\gamma_{n+1}})v_n - \tfrac{u}{2}\gamma_{n+1}(1-e^{-2(1-\alpha_{n+1})\gamma_{n+1}})\nabla f(x_{n+\frac{1}{2}}) + \sqrt{u}\sigma^{(2)}_{n+1}U^{(2)}_{n+1}, \\
v_{n+1} &= v_n e^{-2\gamma_{n+1}} - u\gamma_{n+1}e^{-2(1-\alpha_{n+1})\gamma_{n+1}}\nabla f(x_{n+\frac{1}{2}}) + 2\sqrt{u}\sigma^{(3)}_{n+1}U^{(3)}_{n+1}, \qquad \text{(RULMC)}
\end{aligned}$$

where (γ_n) is the sequence of time steps, $\sigma^{(1)}_n$, $\sigma^{(2)}_n$ and $\sigma^{(3)}_n$ are positive with $(\sigma^{(1)}_n)^2 = \alpha_n\gamma_n + \frac{1-e^{-4\alpha_n\gamma_n}}{4} - (1-e^{-2\alpha_n\gamma_n})$, $(\sigma^{(2)}_n)^2 = \gamma_n + \frac{1-e^{-4\gamma_n}}{4} - (1-e^{-2\gamma_n})$ and $(\sigma^{(3)}_n)^2 = \frac{1-e^{-4\gamma_n}}{4}$, and (α_n) is a sequence of identically distributed random variables following the distribution $\alpha_n \sim U[0,1]$. $(U^{(1)}_n, U^{(2)}_n, U^{(3)}_n)$ are independent centered Gaussian random vectors in $\mathbb{R}^{3d}$, also independent of (α_n) and initial point (x_0, v_0), having the following pairwise covariances:

$$\begin{aligned}
\mathrm{cov}(\sigma^{(1)}_n U^{(1)}_n, \sigma^{(2)}_n U^{(2)}_n) &= \left(\alpha_n\gamma_n - \left(e^{-\alpha_n\gamma_n} + e^{-2\gamma_n}\sinh(\alpha_n\gamma_n)\right)\sinh(\alpha_n\gamma_n)\right) I_{d\times d}, \\
\mathrm{cov}(\sigma^{(2)}_n U^{(2)}_n, \sigma^{(3)}_n U^{(3)}_n) &= \left(e^{-2\gamma_n}\sinh(\gamma_n)^2\right) I_{d\times d}, \\
\mathrm{cov}(\sigma^{(1)}_n U^{(1)}_n, \sigma^{(3)}_n U^{(3)}_n) &= \left(e^{-2\gamma_n}\sinh(\alpha_n\gamma_n)^2\right) I_{d\times d}.
\end{aligned}$$

The (RULMC) algorithm has emerged as an optimal sampling algorithm in the sense that it achieves the information theoretical lower bound in both tolerance ϵ and dimension d for sampling from a strongly log-concave densities [1, 40]. Therefore, it is interesting to examine if (RULMC) based numerical integrator have any benefits in other MCMC-based tasks such as (1). Towards that, we characterize the order of bias with a constant step-size choice for (RULMC) iterates as proposed in [40]. Compared to the bias result in Proposition 2.2 for the (RLMC) discretization, we note that order of bias is increased (i.e. smaller bias). Next, in Theorem 4 we provide a CLT for numerical integration with (RULMC). Our results show that when it comes to computing expectations of the form in (1) using (RULMC) and characterizing its fluctuations, the (RULMC) discretization obtains rate improvements only for a class of constant test functions (as described in Remark 6).

3.1 Analysis of the Markov Chain generated by Constant Step-size RULMC

Recall that $\pi(x)$ is the marginal density function of $\nu(x, v)$ with respect to x. Similarly $\nu_h(x, v)$ be the stationary density function of the Markov chain generated by (RULMC) chain and $\pi_h(x)$

be the marginal density function of $\nu_h(x,v)$, with respect to x. Furthermore, the filtration $(\mathcal{F}_n)$ is defined as $\mathcal{F}_n := \sigma(\alpha_k, U_k^{(i)}; 1 \le k \le n, i = 1,2,3)$. When $f \in \mathcal{C}^2(\mathbb{R}^d)$ and is gradient Lipschitz with parameter M, then we can immediately see that the transition kernel of chain (x_n, v_n): $P((x,v),(x',v')) \in \mathcal{C}(\mathbb{R}^{2d} \times \mathbb{R}^{2d})$ is positive everywhere. Therefore, it's easy to obtain that the chain (x_n, v_n) is μ^{Leb}-irreducible and aperiodic. Given all this information, we can give a sufficient condition to make sure that the chain has a unique invariant probability measure and is ergodic.

Theorem 3. *Let the potential function f satisfy part (a) of Assumption 1.1, and let $\gamma_n := h$ be small enough. Then if $u \in (0, \frac{4}{2M-m})$, the (RULMC) Markov chain (x_n, v_n) has a unique stationary probability measure ν_h and for every $(x,v) \in \mathbb{R}^{2d}$, we have*

$$\sup_{A \in \mathcal{B}(\mathbb{R}^{2d})} |P^n((x,v),A) - \nu_h(A)| \to 0 \quad \text{as} \quad n \to \infty.$$

We next derive an upper bound on the bias $W_2(\pi, \pi_h)$ of (RULMC) algorithm, under the additional strong convexity assumption on the potential function f.

Proposition 3.1. *Suppose that f satisfies Assumption 1.1. If we run the (RULMC) algorithm with $u = 1/M$ and $\gamma_n := h$, for universal constants $C_1, C_2 > 0$, we have*

$$W_2^2(\pi, \pi_h) \le \frac{C_1 h^3 (\kappa h^3 + 1) d}{1 - \frac{h}{4\kappa} - C_2 h^3 \kappa (1 + \kappa h^3)}.$$

Remark 4. *Note that we have $W_2(\pi, \pi_h) \to 0$ as $h \to 0$. Furthermore, as $h \to 0$, $W_2(\pi, \pi_h) < \mathcal{O}(h^{\frac{3}{2}})$. Hence, the bias order is increased for the underdamped Langevin diffusion compared to the overdamped case (cf. Proposition 2.2), providing a smaller bias for the same step-size.*

3.2 Wasserstein-2 rates and CLT with Decreasing Step-size

We now provide the rate of convergence in Wasserstein-2 metric with decreasing step-size for (RULMC). The specific choice for the decreasing step-size that we consider below, also is satisfied for our CLT result in Remark 6.

Proposition 3.2. *Suppose f satisfies Assumption 1.1. Fix $u = 1/M$. Let $x_0 := \arg\min_x f(x)$ and choose $\gamma_n = \frac{16\kappa}{32\kappa^{\frac{5}{3}} + (n - K_1)^+}$, for a $K_1 \in (0, \infty)$ (where $(a)^+ := \max(0, a)$). After running (RULMC) for $K = \tilde{\mathcal{O}}\left(\kappa^{3/2}/\epsilon^{2/3}\right)$ steps, we obtain $W_2(\nu_K, \pi) \le \epsilon\sqrt{d/m}$, where ν_K is the probability distribution of x_K.*

Remark 5. *Similar to the result in Proposition 2.3, there are two aspects of this result. The first aspect is again removing the logarithmic factor in $1/\epsilon$ compared to the result in Theorem 3 in [40], which is quite standard in the literature. The second aspect is that we never obtain the $\mathcal{O}(1/\epsilon^{1/3})$ part, as in Theorem 3 in [40] with the constant step-size choice.*

Similar to the previous case, we now describe the numerical integration procedure using the (RULMC) discretization. We denote the n-th iterate as (x_n, v_n). The time-step we use is (γ_n) such that $\forall n \in \mathbb{N}^*, \gamma_n \ge 0, \lim_n \gamma_n = 0$ and $\lim_n \Gamma_n^{(1)} = +\infty$, where $\Gamma_n^{(\ell)} := \sum_{i=1}^n \gamma_i^\ell$. Our averaging is a weighted empirical measure with $Y_n = (x_n, v_n)$ using the step size sequence $\gamma := (\gamma_n)$ as the weights. Let δ_x denote the Dirac mass at x. Then for every $n \ge 1$, set

$$\nu_n^\gamma(\omega, dx) := \frac{\gamma_1 \delta_{Y_0(\omega)} + \cdots + \gamma_{k+1}\delta_{Y_k(\omega)} + \cdots + \gamma_n \delta_{Y_{n-1}(\omega)}}{\gamma_1 + \cdots + \gamma_n}$$

and we can use $\nu_n^\gamma(\omega, \varphi)$ to approximate $\nu(\varphi) = \mathbb{E}_\nu[\varphi'(Y)]$, where $\varphi' : \mathbb{R}^{2d} \to \mathbb{R}$.

If we assume $g : \mathbb{R}^{2d} \to \mathbb{R}$ such that $\mathcal{L}g = \varphi'$, we can establish the following theorem, in which we state only the unbiased CLT result for simplicity.

Theorem 4. *Let π be such that its potential function f satisfies Assumption 1.1. Assume $u \in (0, \frac{4}{2M-m})$. Consider a test function $\varphi' = \mathcal{L}g$, for some function $g : \mathbb{R}^{2d} \to \mathbb{R}$, where $\mathcal{L} = 2u\Delta_v - 2\langle v, \nabla_v \rangle - u\langle \nabla f(x), \nabla_v \rangle + \langle v, \nabla_x \rangle$ denotes the generator of the diffusion (4). Suppose the step-size (γ_k) is non-increasing, $\lim_{n\to+\infty}(1/\sqrt{\Gamma_n}) \sum_{k=1}^n \gamma_k^{3/2} = +\infty$. Then, if $\lim_{n\to+\infty}(1/\sqrt{\Gamma_n}) \sum_{k=1}^n \gamma_k^2 = 0$, for every $g \in \mathcal{C}^4(\mathbb{R}^{2d})$ function with $D^2 g$ bounded, $D^3 g$*

bounded and Lipschitz and $\sup_{(x,v)\in\mathbb{R}^{2d}} \|\nabla g(x,v)\|/(1+\|x\|^2+\|v\|^2) < +\infty$*, we have the following central limit theorem for the numerical integration computed using the* (RULMC) *iterates:*

$$\sqrt{\Gamma_n}\nu_n^\gamma(\mathcal{L}g) \xrightarrow{d} \mathcal{N}\big(0, 4u\int \|\nabla_v g(x,v)\|^2 \nu(dx,dv)\big).$$

The rate of convergence of the CLT in Theorem 4 follows exactly the same behavior in Theorem 2. Hence, for the class of general test functions, Theorem 4 does not exhibit a rate improvement. Towards that, we make the following remarks under a carefully constructed condition for the class of test functions.

Remark 6. *Let* π *be such that its potential function* f *satisfies Assumption 1.1. Assume* $u \in (0, \frac{4}{2M-m})$*. Consider a test function* $\varphi = \mathcal{L}g$ *which could be written as* $\mathcal{L}g(v,\phi(x)) = \langle v, \nabla\phi(x)\rangle$*, for some function* $\phi : \mathbb{R}^d \to \mathbb{R}$*, where* $\mathcal{L} = 2u\Delta_v - 2\langle v, \nabla_v\rangle - u\langle \nabla f(x), \nabla_v\rangle + \langle v, \nabla_x\rangle$ *denotes the generator of the diffusion* (4)*. Suppose the time step-size* (γ_k) *is non-increasing, and satisfies* $\lim_{n\to\infty}(\gamma_{n-1}-\gamma_n)/\gamma_n^4 = 0$ *and* $\lim_{n\to\infty}\Gamma_n^{(4)} = +\infty$*. Define* $\hat{\gamma}_n := \Gamma_n^{(4)}/\sqrt{\Gamma_n^{(3)}}$ *and let* $\hat{\gamma}_\infty = \lim_{n\to\infty}\hat{\gamma}_n$*. Then, for all* $\phi \in \mathcal{C}^4(\mathbb{R}^d)$ *with* $D^2\phi$*,* $D^3\phi$ *and* $D^4\phi$ *bounded and Lipschitz and* $\sup_{(x,v)\in\mathbb{R}^{2d}} \|\nabla\phi(x)\|^2/(1+\|x\|^2+\|v\|^2) < +\infty$*, we obtain the following central limit theorem for numerical integration computed using the* (RULMC) *algorithm:*

(i) If $\hat{\gamma}_\infty = 0$*, we have* $\frac{\Gamma_n}{\sqrt{\Gamma_n^{(3)}}}\nu_n^\gamma(\mathcal{L}\phi) \xrightarrow{d} \mathcal{N}(0, \frac{10}{3}u\int_{\mathbb{R}^d}\|\nabla\phi(x)\|\pi(dx))$*,*

(ii) If $\hat{\gamma}_\infty \in (0,+\infty)$*, we have* $\frac{\Gamma_n}{\Gamma_n^{(4)}}\nu_n^\gamma(\mathcal{L}\phi) \xrightarrow{d} \mathcal{N}(\rho, \frac{10}{3}u\hat{\gamma}_\infty^{-2}\int_{\mathbb{R}^d}\|\nabla\phi(x)\|\pi(dx))$*,*

(iii) If $\hat{\gamma}_\infty = +\infty$*, we have* $\frac{\Gamma_n}{\Gamma_n^{(4)}}\nu_n^\gamma(\mathcal{L}\phi) \xrightarrow{p} \rho$*,*

where,

$$\begin{aligned}\rho = {} & \tfrac{5u}{12}\int\int\langle D^3\phi(x), \nabla f(x)\otimes v\otimes v\rangle\nu(dx,dv) + \tfrac{u}{24}\int\int\langle D^3 f(x), \nabla\phi(x)\otimes v\otimes v\rangle\nu(dx,dv)\\ & + \tfrac{7u}{12}\int\int(D^2\phi D^2 f)(x)v^{\otimes 2}\nu(dx,dv) - \tfrac{u^2}{4}\int\langle D^2\phi(x), \nabla f(x)^{\otimes 2}\rangle\pi(dx)\\ & - \tfrac{u^2}{24}\int\langle D^2 f(x), \nabla\phi(x)\otimes\nabla f(x)\rangle\pi(dx).\end{aligned}$$

Remark 7. *For polynomial time steps* $\gamma_k := k^{-\alpha}$*, since we require that* $\Gamma_n^{(4)} \to +\infty$ *as* $n \to +\infty$*, we need* $0 < \alpha \le \frac{1}{4}$*. Using L'Hospitals rule, it is straightforward to check that the condition* $\lim_{n\to+\infty}\frac{\gamma_{n-1}-\gamma_n}{\gamma_n^4} = 0$ *is satisfied when* $\alpha \in (0, \frac{1}{4}]$*. We then have the following order estimates:*

$$\Gamma_n \sim \frac{n^{1-\alpha}}{1-\alpha}, \quad \sqrt{\Gamma_n^{(3)}} \sim \frac{n^{\frac{1}{2}-\frac{3}{2}\alpha}}{\sqrt{1-3\alpha}}, \quad \Gamma_n^{(4)} \sim \begin{cases}\dfrac{n^{1-4\alpha}}{1-4\alpha}, & \text{if } \alpha \in (0,\frac{1}{4}),\\ \sqrt{\ln n}, & \text{if } \alpha = \frac{1}{4}.\end{cases}$$

Hence, as $n \to +\infty$*,*

$$\frac{\Gamma_n^{(4)}}{\sqrt{\Gamma_n^{(3)}}} \to \hat{\gamma}_\infty = \begin{cases}0 & \text{if } \alpha \in (\frac{1}{5},\frac{1}{4}],\\ \sqrt{10} & \text{if } \alpha = \frac{1}{5},\\ +\infty & \text{if } \alpha \in (0,\frac{1}{5}).\end{cases}$$

If $\alpha \in (\frac{1}{5},\frac{1}{4}]$*, the unbiased CLT holds at rate* $\Gamma_n/\sqrt{\Gamma_n^{(3)}} = \mathcal{O}(n^{\frac{1}{2}(1+\alpha)}) \le \mathcal{O}(n^{\frac{5}{8}})$*. The optimal rate is achieved when* $\alpha = \frac{1}{4}$*. If* $\alpha = \frac{1}{5}$*, the biased CLT holds at rate* $\Gamma_n/\sqrt{\Gamma_n^{(3)}} = \mathcal{O}(n^{3\alpha}) = \mathcal{O}(n^{\frac{3}{5}})$*. If* $\alpha \in (0,\frac{1}{5})$*, the rate of the convergence in probability is* $\Gamma_n/\sqrt{\Gamma_n^{(3)}} = \mathcal{O}(n^{3\alpha}) < \mathcal{O}(n^{\frac{3}{5}})$*. Therefore the optimal convergence rate* $\mathcal{O}(n^{\frac{5}{8}})$ *is obtained when an unbiased CLT holds. While the rate of this CLT is better than that of Theorem 2, the test functions that satisfy this condition is severely restricted.*

4 Discussion

In this work, we present several probablistic properties of the randomized midpoint discretization technique, focussing our attention on the overdamped and underdamped Langevin diffusions. Our results could be summarized as follows: To obtain optimal rates for sampling (in W_2 distance), one needs to have a constant choice of step-size. With such a constant step-size choice, the Markov chain generated by the discretization process is biased. This suggests that a decreasing step-size choice is required when using the randomized midpoint method for numerical integration. For several decreasing choices of step-sizes, we establish CLTs and highlight the relative merits and disadvantages of using randomized midpoint technique for numerical integration. In particular, our results have interesting consequence for computing confidence interval for numerical integration.

Broader Impact

The paper predominantly concerns about theoretical results on numerical integration with MCMC based samplers. The presented results are of interest to researchers in machine learning community concerned with constructing confidence intervals for their numerical integration problems. Although not the main focus of this paper, the results might have positive consequence for building robust machine learning systems based on the obtained confidence intervals.

Funding Disclosure

The research of YH and KB were supported in parts by UC Davis CeDAR (Center for Data Science and Artificial Intelligence Research) Innovative Data Science Seed Funding Program. The research of YH was also supported in parts by the NSF Grant CCF-1934568. The research of MAE is supported by NSERC Grant [2019-06167], Connaught New Researcher Award, CIFAR AI Chairs program, and CIFAR AI Catalyst grant.

References

[1] Yu Cao, Jianfeng Lu, and Lihan Wang. Complexity of randomized algorithms for underdamped langevin dynamics. *arXiv preprint arXiv:2003.09906*, 2020.

[2] Tianqi Chen, Emily Fox, and Carlos Guestrin. Stochastic gradient hamiltonian monte carlo. In *International conference on machine learning*, pages 1683–1691, 2014.

[3] Xiang Cheng and Peter Bartlett. Convergence of langevin mcmc in kl-divergence. In *Algorithmic Learning Theory*, pages 186–211, 2018.

[4] Xiang Cheng, Niladri S Chatterji, Yasin Abbasi-Yadkori, Peter L Bartlett, and Michael I Jordan. Sharp convergence rates for langevin dynamics in the nonconvex setting. *arXiv preprint arXiv:1805.01648*, 2018.

[5] Xiang Cheng, Niladri S Chatterji, Peter L Bartlett, and Michael I Jordan. Underdamped Langevin MCMC: A non-asymptotic analysis. *arXiv preprint arXiv:1707.03663*, 2017.

[6] Arnak Dalalyan. Further and stronger analogy between sampling and optimization: Langevin monte carlo and gradient descent. In *Proceedings of the 2017 Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pages 678–689. PMLR, 07–10 Jul 2017.

[7] Arnak Dalalyan and Lionel Riou-Durand. On sampling from a log-concave density using kinetic Langevin diffusions. *Bernoulli*, 26(3):1956–1988, 2020.

[8] Arnak S Dalalyan and Avetik Karagulyan. User-friendly guarantees for the langevin monte carlo with inaccurate gradient. *Stochastic Processes and their Applications*, 129(12):5278–5311, 2019.

[9] Alain Durmus, Szymon Majewski, and Blazej Miasojedow. Analysis of langevin monte carlo via convex optimization. *Journal of Machine Learning Research*, 20(73):1–46, 2019.

[10] Alain Durmus and Eric Moulines. High-dimensional bayesian inference via the unadjusted langevin algorithm. *Bernoulli*, 25(4A):2854–2882, 2019.

[11] Alain Durmus, Eric Moulines, and Marcelo Pereyra. Efficient bayesian computation by proximal markov chain monte carlo: when langevin meets moreau. *SIAM Journal on Imaging Sciences*, 11(1):473–506, 2018.

[12] Alain Durmus, Eric Moulines, and Eero Saksman. On the convergence of hamiltonian monte carlo. *The Annals of Statistics (to appear)*, 2019+.

[13] Alain Durmus and Éric Moulines. Nonasymptotic convergence analysis for the unadjusted langevin algorithm. *The Annals of Applied Probability*, 27(3):1551–1587, 06 2017.

[14] Andreas Eberle. Reflection couplings and contraction rates for diffusions. *Probability theory and related fields*, 166(3-4):851–886, 2016.

[15] Andreas Eberle, Arnaud Guillin, Raphael Zimmer, et al. Couplings and quantitative contraction rates for langevin dynamics. *The Annals of Probability*, 47(4):1982–2010, 2019.

[16] Murat A Erdogdu and Rasa Hosseinzadeh. On the convergence of langevin monte carlo: The interplay between tail growth and smoothness. *arXiv preprint arXiv:2005.13097*, 2020.

[17] Murat A Erdogdu, Lester Mackey, and Ohad Shamir. Global non-convex optimization with discretized diffusions. In *Advances in Neural Information Processing Systems*, pages 9671–9680, 2018.

[18] Jackson Gorham, Andrew B Duncan, Sebastian J Vollmer, and Lester Mackey. Measuring sample quality with diffusions. *arXiv preprint arXiv:1611.06972*, 2016.

[19] Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations*, volume 31. Springer Science & Business Media, 2006.

[20] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[21] Damien Lamberton and Gilles Pages. Recursive computation of the invariant distribution of a diffusion. *Bernoulli*, 8(3):367–405, 2002.

[22] Ben Leimkuhler and Charles Matthews. *Molecular Dynamics: With Deterministic and Stochastic Numerical Methods*. Springer, 2016.

[23] Tony Lelievre and Gabriel Stoltz. Partial differential equations and stochastic methods in molecular dynamics. *Acta Numerica*, 25:681–880, 2016.

[24] Chunyuan Li, Changyou Chen, David Carlson, and Lawrence Carin. Preconditioned stochastic gradient langevin dynamics for deep neural networks. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[25] Xuechen Li, Yi Wu, Lester Mackey, and Murat A Erdogdu. Stochastic runge-kutta accelerates langevin monte carlo and beyond. In *Advances in Neural Information Processing Systems*, pages 7748–7760, 2019.

[26] Jun S Liu. *Monte Carlo strategies in scientific computing*. Springer Science & Business Media, 2008.

[27] Samuel Livingstone, Michael Betancourt, Simon Byrne, and Mark Girolami. On the geometric ergodicity of hamiltonian monte carlo. *Bernoulli*, 25(4A):3109–3138, 2019.

[28] Zhi-Quan Luo and Paul Tseng. Error bounds and convergence analysis of feasible descent methods: a general approach. *Annals of Operations Research*, 46(1):157–178, 1993.

[29] Yi-An Ma, Niladri Chatterji, Xiang Cheng, Nicolas Flammarion, Peter Bartlett, and Michael I Jordan. Is there an analog of nesterov acceleration for mcmc? *arXiv preprint arXiv:1902.00996*, 2019.

[30] Stephan Mandt, Matthew D Hoffman, and David M Blei. Stochastic gradient descent as approximate bayesian inference. *The Journal of Machine Learning Research*, 18(1):4873–4907, 2017.

[31] Jonathan C Mattingly, Andrew M Stuart, and Desmond J Higham. Ergodicity for sdes and approximations: locally lipschitz vector fields and degenerate noise. *Stochastic processes and their applications*, 101(2):185–232, 2002.

[32] Eric Mazumdar, Aldo Pacchiano, Yi-an Ma, Peter L Bartlett, and Michael I Jordan. On thompson sampling with langevin algorithms. *arXiv preprint arXiv:2002.10002*, 2020.

[33] Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.

[34] Grigori Noah Milstein and Michael V Tretyakov. *Stochastic numerics for mathematical physics*. Springer Science & Business Media, 2013.

[35] Wenlong Mou, Yi-An Ma, Martin J Wainwright, Peter L Bartlett, and Michael I Jordan. High-order langevin diffusion yields an accelerated mcmc algorithm. *arXiv preprint arXiv:1908.10859*, 2019.

[36] Radford M Neal. MCMC using hamiltonian dynamics. *Handbook of Markov chain Monte Carlo*, 2(11):2, 2011.

[37] Maxim Raginsky, Alexander Rakhlin, and Matus Telgarsky. Non-convex learning via stochastic gradient langevin dynamics: a nonasymptotic analysis. In *Proceedings of the 2017 Conference on Learning Theory*, volume 65, pages 1674–1703, 2017.

[38] Gareth O Roberts, Richard L Tweedie, et al. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

[39] Sotirios Sabanis and Ying Zhang. Higher order langevin monte carlo algorithm. *Electronic Journal of Statistics*, 13(2):3805–3850, 2019.

[40] Ruoqi Shen and Yin Tat Lee. The randomized midpoint method for log-concave sampling. In *Advances in Neural Information Processing Systems*, pages 2098–2109, 2019.

[41] Yee Whye Teh, Alexandre H Thiery, and Sebastian J Vollmer. Consistency and fluctuations for stochastic gradient langevin dynamics. *The Journal of Machine Learning Research*, 17(1):193–225, 2016.

[42] Santosh Vempala and Andre Wibisono. Rapid convergence of the unadjusted langevin algorithm: Isoperimetry suffices. In *Advances in Neural Information Processing Systems*, pages 8092–8104, 2019.

[43] Santosh S Vempala. Recent progress and open problems in algorithmic convex geometry. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2010)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2010.

[44] Cédric Villani. The wasserstein distances. In *Optimal Transport*, pages 93–111. Springer, 2009.

[45] Max Welling and Yee W Teh. Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 681–688, 2011.

5 Additional Notations

We also use the following notations for the proofs. Due to the ease of presentation, whenever it is clear in the proof, we refer to the inner product between two compatible vectors $\langle a, b\rangle$ simply by $a \cdot b$. For any random variable X, $\|X\|_{L^2} := \mathbb{E}[\|X\|]$ where the expectation is taken over all randomness of X.

6 Proofs for Section 2

We now define the following condition, which is a consequence of Assumption 1.1

Assumption 6.1. *There exists a twice differentiable function $V : \mathbb{R}^d \to [1, \infty)$ such that: (i) $\lim_{\|x\|\to\infty} V(x) = +\infty$, (ii) there exists $\alpha > 0$ and $\beta > 0$: $\langle \nabla V(x), \nabla f(x)\rangle \geq \alpha V(x) - \beta$ for every x, (iii) there exists $c_V > 0$: $\|\nabla V(x)\|^2 + \|\nabla f(x)\|^2 \leq c_V V(x)$ for every x, and (iv) $\left\|D^2 V\right\|_\infty := \sup_{x\in\mathbb{R}^d} \|D^2 V\|_{op} < \infty$ (where $\|\cdot\|_{op}$ denotes the operator norm).*

Lemma 1. *Assumption 1.1 implies Assumption 6.1.*

Proof. Since $f \in \mathcal{C}^2(\mathbb{R}^d)$ is strongly convex, $\lim_{|x|\to+\infty} f(x) = +\infty$ and f has a unique global minimizer $x^* \in \mathbb{R}^d$. It's easy to observe that $\nabla f(x^*) = 0$. We consider our $V(x) = f(x) - f(x^*) + 1$. Then it's easy to see (i) is satisfied. (iv) is also satisfied because f is gradient Lipschitz. (iii) is equivalent to that there exists a $C > 0$ such that

$$\frac{|\nabla f(x)|^2}{f(x) - f(x^*) + 1} \leq C \quad \text{for } \forall x \in \mathbb{R}^d$$

We Taylor expand the numerator and denominator:

$$\begin{aligned}
|\nabla f(x)|^2 &= \sum_{i=1}^d \left(f_i(x^*) + \nabla f_i(\xi)^T (x - x^*)\right)^2 \\
&\leq \sum_{i,j=1}^d |f_{ij}(\xi)|^2 |x - x^*|^2 = \left\|D^2 f(\xi)\right\|_F^2 |x - x^*|^2 \\
&\leq d^2 M^2 |x - x^*|^2 \\
f(x) - f(x^*) + 1 &= \nabla f(x^*)^T (x - x^*) + \frac{1}{2} D^2 f(\xi)(x - x^*)^{\otimes 2} + 1 \\
&= \frac{1}{2} D^2 f(\xi)(x - x^*)^{\otimes 2} + 1 \\
&\geq \frac{m}{2} |x - x^*|^2
\end{aligned}$$

Then

$$\frac{|\nabla f(x)|^2}{f(x) - f(x^*) + 1} \leq \frac{2d^2 M^2}{m} \quad \text{for } \forall x \in \mathbb{R}^d$$

(ii) is equivalent to that there exists $\alpha, \beta > 0$ such that

$$|\nabla f(x)|^2 \geq \alpha(f(x) - f(x^*) + 1) - \beta \quad \text{for } \forall x \in \mathbb{R}^d$$

According to the strongly convexity of f, we have

$$\begin{aligned}
f(x^*) - f(x) &\geq \nabla f(x)^T (x^* - x) + \frac{m}{2} |x^* - x|^2 \\
&= \frac{m}{2} |x^* - x + \frac{1}{m} \nabla f(x)|^2 - \frac{1}{2m} |\nabla f(x)|^2
\end{aligned}$$

which then implies

$$|\nabla f(x)|^2 \geq 2m \left(f(x) - f(x^*) + 1\right) - 2m \quad \text{for } \forall x \in \mathbb{R}^d$$

(ii) is satisfied by choosing $\alpha = \beta = 2m > 0$. □

Remark 8. *For the $V(x)$ we choose in the proof, under assumption 1.1, we can verify that: $V(x) = O(|x|^2)$ when $|x| \to +\infty$. We will use this fact later in the proof when we establish the CLT statement.*

6.1 Proofs for section 2.1

Lemma 2. *Let $x(t)$ be the solution to Langevin dynamics SDE with initial condition x_0 and $y(t)$ be the solution to Langevin dynamics SDE with initial condition y_0. Then we have the following estimates for Langevin dynamics when f satisfies Assumption 1.1 and $Mh < \frac{1}{2}$:*

$$\mathbb{E}[\sup_{t\in[0,h]} \|\nabla f(x(t))\|^2] \le 4\|\nabla f(x_0)\|^2 + 8M^2dh$$
$$\mathbb{E}[\sup_{t\in[0,h]} \|x(t)-x_0\|^2] \le O(h^2\|\nabla f(x_0)\|^2 + M^2h^3d + 2dh)$$
$$\mathbb{E}[\|x(t)-y(t)\|^2] \le e^{-2mt}\|x_0-y_0\|^2$$

Proof. By triangle inequality we have

$$\mathbb{E}[\sup_{t\in[0,h]} \|\nabla f(x(t))\|^2] \le 2\|\nabla f(x_0))\|^2 + 2M^2\mathbb{E}[\sup_{t\in[0,h]} \|x(t)-x_0\|^2]$$

Furthermore, we have

$$\begin{aligned}\mathbb{E}[\sup_{t\in[0,h]} \|x(t)-x_0\|^2] &= \mathbb{E}[\sup_{t\in[0,h]} \left\|-\int_0^t \nabla f(x(s))ds + \sqrt{2}W_t\right\|^2] \\ &\le h^2\mathbb{E}[\sup_{t\in[0,h]} \|\nabla f(x(t))\|^2] + 2dh\end{aligned}$$

Combining the two inequalities and $Mh < \frac{1}{2}$, we can obtain the first two estimates. The last estimate could be easily obtained by energy method. □

Proof of Theorem 2.1. We denote $x_n = x_n(0)$ to be the algorithm iterate points, y_n to be the n-th step of Langevin diffusion with $y_0 \sim \exp(-f(y))$, $x^*_{n+1} = x_n(h)$ to be one step solution of Langevin dynamics with initial values x_n. When $Mh < \frac{1}{2}$, apply lemma 2 and we get:

$$\mathbb{E}[\sup_{t\in[0,h]} \|x_{n-1}(\alpha_n h) - x_{n-1}(t)\|^2] \le O(h^2\|\nabla f(x_{n-1})\|_{L^2}^2 + M^2h^3d + 2dh)$$

$$\begin{aligned}\mathbb{E}[\left\|\nabla f(x_{n-\frac{1}{2}}) - \nabla f(x_{n-1}(\alpha_n h))\right\|^2] &\le M^2\mathbb{E}\left\|\int_0^{\alpha_n h} \nabla f(x_{n-1}(s)) - \nabla f(x_{n-1}(0))ds\right\|^2 \\ &\le M4h^2\mathbb{E}[\alpha_n^2 \sup_{t\in[0,\alpha_n h]} \|x_{n-1}(t) - x_{n-1}(0)\|^2] \\ &\le O(M^4h^4\|\nabla f(x_{n-1})\|_{L^2}^2 + dM^4h^3 + dM^6h^5)\end{aligned}$$

Consider the distance between our iterates and the continuous process:

$$\begin{aligned}\mathbb{E}_{\alpha_K}[\|x_K - y_K\|^2] &= \mathbb{E}_{\alpha_K}[\|x_K - x^*_K + x^*_K - y_K\|^2] \\ &\le \|y_K - x^*_K\|^2 + \mathbb{E}_{\alpha_K}[\|x_K - x^*_K\|^2] - 2(y_K - x^*_K)^T(\mathbb{E}_{\alpha_K}x_K - x^*_K) \\ &\le (1+hm)\|y_K - x^*_K\|^2 + \frac{1}{hm}\|\mathbb{E}_{\alpha_K}x_K - x^*_K\|^2 + \mathbb{E}_{\alpha_K}[\|x_K - x^*_K\|^2]\end{aligned}$$

Taking expectations over $\{\alpha_k, U_l, U'_l; 1 \le k \le K-1, 1 \le l \le K\}$, applying lemma 2 again and using induction, we have

$$\begin{aligned}\|x_K - y_K\|_{L^2}^2 &\le (1+hm)\|y_K - x^*_K\|_{L^2}^2 + \frac{1}{hm}\mathbb{E}\|\mathbb{E}_{\alpha_K}x_K - x^*_K\|^2 + \|x_K - x^*_K\|_{L^2}^2 \\ &\le (1+hm)e^{-2mh}\|x_{K-1} - y_{K-1}\|_{L^2}^2 + \frac{1}{hm}\mathbb{E}\|\mathbb{E}_{\alpha_K}x_K - x^*_K\|^2 + \|x_K - x^*_K\|_{L^2}^2 \\ &\le (1+hm)e^{-2mKh}\|x_0 - y_0\|_{L^2}^2 + \sum_{n=1}^K \frac{1}{hm}\mathbb{E}\|\mathbb{E}_{\alpha_n}x_n - x^*_n\|^2 + \sum_{n=1}^K \|x_n - x^*_n\|_{L^2}^2 \\ &\le e^{-mKh}\|x_0 - y_0\|_{L^2}^2 + A + B\end{aligned}$$

Next we bound part A and part B. For part A:

$$\begin{aligned}\|\mathbb{E}_{\alpha_n} x_n - x_n^*\|^2 &= \left\|\mathbb{E}_{\alpha_n}[h\nabla f(x_{n-\frac{1}{2}})] - \int_0^h \nabla f(x_{n-1}(s))ds\right\|^2 \\ &\le 2\mathbb{E}\left\|h\nabla f(x_{n-1}) - h\nabla f(x_{n-1}(\alpha_n h))\right\|^2 + 2\left\|\mathbb{E}_{\alpha_n}[h\nabla f(x_{n-1}(\alpha_n h))] - \int_0^h \nabla f(x_{n-1}(s))ds\right\|^2 \\ &\le 2h^2\mathbb{E}_{\alpha_n}\left\|\nabla f(x_{n-\frac{1}{2}}) - \nabla f(x_{n-1}(\alpha_n h))\right\|^2 + 0\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{E}\left\|\mathbb{E}_{\alpha_n} x_n - x_n^*\right\|^2 &\le 2h^2\mathbb{E}[\left\|\nabla f(x_{n-\frac{1}{2}}) - \nabla f(x_{n-1}(\alpha_n h))\right\|^2] \\ &\le O(M^4h^6\left\|\nabla f(x_{n-1})\right\|_{L^2}^2 + dM^4h^5)\end{aligned}$$

For part B, use our previous estimates:

$$\begin{aligned}\|x_n - x_n^*\|_{L^2}^2 &= \left\|h\nabla f(x_{n-\frac{1}{2}}) - \int_0^h \nabla f(x_{n-1}(s))ds\right\|_{L^2}^2 \\ &\le 2\left\|h\nabla f(x_{n-\frac{1}{2}}) - h\nabla f(x_{n-1}(\alpha_n h))\right\|_{L^2}^2 + 2\left\|\int_0^h \nabla f(x_{n-1}(s)) - \nabla f(x_{n-1}(\alpha_n h))ds\right\|_{L^2}^2 \\ &\le 2h^2\left\|\nabla f(x_{n-\frac{1}{2}}) - \nabla f(x_{n-1}(\alpha_n h))\right\|_{L^2}^2 + 2M^2h^2\mathbb{E}[\sup_{t\in[0,h]}\|x_{n-1}(\alpha_n h) - x_{n-1}(t)\|^2] \\ &\le O(M^2h^4\left\|\nabla f(x_{n-1})\right\|_{L^2}^2 + dM^2h^3)\end{aligned}$$

Plug the estimates on A and B into the inequality we have

$$\begin{aligned}\|x_K - y_K\|_{L^2}^2 \le e^{-mKh}\|x_0 - y_0\|_{L^2}^2 &+ O(m^{-1}M^4h^5\sum_{n=0}^{K-1}\|\nabla f(x_n)\|_{L^2}^2 + dm^{-1}M^4Kh^4) \\ &+ O(M^2h^4\sum_{n=0}^{K-1}\|\nabla f(x_n)\|_{L^2}^2 + dM^2Kh^3)\end{aligned}$$

Next we need to estimate $\sum_{n=0}^{K-1}\|\nabla f(x_n)\|_{L^2}^2$. Since

$$\begin{aligned}f(x_n(h)) &= f(x_n(0)) + \int_0^h df(x_n(t)) \\ &= f(x_n(0)) - \int_0^h |\nabla f(x_n(t))|^2 dt + \sqrt{2}\int_0^h \nabla f(x_n(t))dW(t) + \int_0^h \Delta f(x_n(t))dt\end{aligned}$$

we have

$$\mathbb{E}[f(x_{n+1}(0))] - \mathbb{E}[f(x_n(h))] = \mathbb{E}[f(x_{n+1}(0)) - f(x_n(0))] + \mathbb{E}[\int_0^h |\nabla f(x_n(t))|^2 dt] - \mathbb{E}[\int_0^t \Delta f(x_n(t))dt]$$

When $Mh < \frac{1}{4}$,

$$\begin{aligned}\mathbb{E}[\inf_{t\in[0,h]}\|\nabla f(x(t))\|^2] &\ge \frac{1}{2}\|\nabla f(x(0))\|_{L^2}^2 - \mathbb{E}[\sup_{t\in[0,h]}\|\nabla f(x(t)) - \nabla f(x(0))\|^2] \\ &\ge \frac{1}{2}\|\nabla f(x(0))\|_{L^2}^2 - M^2\mathbb{E}[\sup_{t\in[0,h]}\|x(t) - x(0)\|^2] \\ &\ge \frac{1}{4}\|\nabla f(x(0))\|_{L^2}^2 + O(dM^2h) \\ |\Delta f(x_n(t))| &\le d\left\|\nabla^2 f(x_n(t))\right\| \le Md\end{aligned}$$

Plug these two estimates into our previous identity and we obtain,

$$\mathbb{E}[f(x_{n+1}(0)) - f(x_n(h))] \geq \mathbb{E}[f(x_{n+1}) - f(x_n)] + \frac{h}{4}\left\|\nabla f(x_n)\right\|_{L^2}^2 - dMh + O(dM^2h^2)$$

Next we consider that

$$\begin{aligned}\mathbb{E}_{\alpha_{n+1}}[f(x_{n+1}(0))] &\leq f(x_n(h)) + \nabla f(x_n(h))^T(\mathbb{E}_{\alpha_{n+1}}[x_{n+1}(0)] - x_n(h)) + \frac{M}{2}\mathbb{E}_{\alpha_{n+1}}[\left\|x_{n+1}(0) - x_n(h)\right\|^2] \\ &\leq f(x_n(h)) + Mh^2\left\|\nabla f(x_n(h))\right\|_{L^2}^2 + M^{-1}h^{-2}\left\|\mathbb{E}_{\alpha_{n+1}}[x_{n+1}(0)] - x_n(h)\right\|^2 \\ &+ \frac{M}{2}\mathbb{E}_{\alpha_{n+1}}[\left\|x_{n+1}(0) - x_n(h)\right\|^2]\end{aligned}$$

where

$$\begin{aligned}Mh^2\mathbb{E}[\left\|\nabla f(x_n(h))\right\|^2] &\leq O(Mh^4\left\|\nabla f(x_n)\right\|_{L^2}^2 + dMh^3) \\ M^{-1}h^{-2}\left\|\mathbb{E}_{\alpha_{n+1}}[x_{n+1}(0)] - x_n(h)\right\|^2 &\leq O(M^3h^4\left\|\nabla f(x_n)\right\|_{L^2}^2 + dM^3h^3) \\ \frac{M}{2}\mathbb{E}[\left\|x_{n+1} - x_n(h)\right\|^2] &\leq O(M^3h^4\left\|\nabla f(x_n)\right\|_{L^2}^2 + dM^3h^3)\end{aligned}$$

Hence we have

$$\mathbb{E}[f(x_{n+1}(0)) - f(x_n(h))] \leq O(M^3h^4\left\|\nabla f(x_n)\right\|_{L^2}^2 + dM^3h^3)$$

and

$$O(M^3h^4\left\|\nabla f(x_n)\right\|_{L^2}^2 + dM^3h^3) \geq \mathbb{E}[f(x_{n+1}) - f(x_n)] + \frac{h}{4}\left\|\nabla f(x_n)\right\|_{L^2}^2 + O(dM^2h^2) - dMh$$

sum up over k from 0 to $K-1$:

$$O(M^3h^4\sum_{k=0}^{K-1}\left\|\nabla f(x_n)\right\|_{L^2}^2 + dM^3Kh^3) \geq \mathbb{E}[f(x_K) - f(x_0)] + \frac{h}{4}\sum_{k=1}^{K-1}\left\|\nabla f(x_n)\right\|_{L^2}^2 + O(dM^2Nh^2) - dMKh$$

Picking $x_0 = \text{argmin} f(x)$, we can ensure $\mathbb{E}[f(x_K) - f(x_0)] \geq 0$, when $Mh < \frac{1}{2}$, we have

$$\begin{aligned}&\frac{h}{8}\sum_{k=0}^{K-1}\left\|\nabla f(x_n)\right\|_{L^2}^2 \leq dKMh - O(dKM^2h^2) + O(dKM^3h^3) \\ \Longrightarrow\quad &\sum_{k=0}^{K-1}\left\|\nabla f(x_n)\right\|_{L^2}^2 \leq O(dKM)\end{aligned}$$

Therefore

$$\begin{aligned}\left\|x_K - y_K\right\|_{L^2}^2 &\leq e^{-mKh}\left\|x_0 - y_0\right\|_{L^2}^2 + O(m^{-1}M^5h^5Kd + m^{-1}M^4h^4Kd) + O(M^3h^4Kd + M^2h^3Kd) \\ &\leq e^{-mKh}\left\|x_0 - y_0\right\|_{L^2}^2 + O(\kappa M^3h^4Kd) + O(M^2h^3Kd)\end{aligned}$$

Therefore we have

$$W_2(\nu_K, \pi)^2 \leq e^{-mKh}\left\|x_0 - y_0\right\|_{L^2}^2 + O(M^3h^4Kd)\max\{\kappa, \frac{1}{Mh}\}$$

a) When $\kappa > \frac{1}{Mh}$, by choosing $h \sim O(\frac{\epsilon^{2/3}}{\kappa^{1/3}M})$, we can ensure $W_2(\nu_K, \pi)^2 \leq \epsilon^2 d/m$ after K steps when $K \sim \tilde{O}(\frac{\kappa^{4/3}}{\epsilon^{2/3}})$.

b) When $\kappa \leq \frac{1}{Mh}$, by choosing $h \sim O(\frac{\epsilon}{M})$, we can ensure $W_2(\nu_K, \pi)^2 \leq \epsilon^2 d/m$ after K steps when $K \sim \tilde{O}(\frac{\kappa}{\epsilon})$.

□

6.2 Proofs for Section 2.2

Proof of Theorem 1. Under the assumption 6.1, we can show that the following Lyapunov condition is satisfied for small h.
(Lyapunov Condition): There exists a function $V: \mathbb{R}^d \to [1, \infty)$ such that:

0) $\lim_{|x|\to\infty} V(x) = +\infty$,

1) There exists $\hat{\alpha} \in (0,1)$ and $\hat{\beta} \geq 0$: $\mathbb{E}[V(x_{n+1})|\mathcal{F}_n] \leq \hat{\alpha} V(x_n) + \hat{\beta}$.

Proof: To show that assumption 6.1 implies Lyapunov condition, we first do Taylor expansion of $V(x_{n+1})$ at x_n:

$$\begin{aligned}V(x_{n+1}) &= V(x_n) - h\langle \nabla V(x_n), \nabla f(x_n)\rangle + \alpha_{n+1}h^2\langle D^2 f(x_n); \nabla f(x_n), \nabla V(x_n)\rangle \\ &\quad - \sqrt{2\alpha_{n+1}}h^{\frac{3}{2}}\langle D^2 f(x_n); \nabla V(x_n), U'_{n+1}\rangle + \sqrt{2h}\nabla V(x_n)\cdot U_{n+1} \\ &\quad + \frac{1}{2}D^2V(\theta_n)(-h\nabla f(x_n) + \alpha_{n+1}h^2 D^2 f(x_n)\nabla f(x_n) - \sqrt{2\alpha_{n+1}}h^{\frac{3}{2}}U'_{n+1} + \sqrt{2h}U_{n+1})^{\otimes 2}\end{aligned}$$

where θ_n is a random point on the line segment joining x_n and x_{n+1}. Using the fact that f is M-gradient Lipschitz, we have:

$$\begin{aligned}\mathbb{E}[V(x_{n+1})|\mathcal{F}_n] &\leq V(x_n) - h\langle \nabla V(x_n), \nabla f(x_n)\rangle + \frac{1}{4}Mh^2(|\nabla f(x_n)|^2 + |\nabla V(x_n)|^2) + h^{\frac{3}{2}}Md + h^{\frac{3}{2}}M|\nabla V(x_n)|^2 \\ &\quad + 2\left\|D^2V\right\|_\infty (h^2|\nabla f(x_n)|^2 + \frac{1}{3}M^2h^4|\nabla f(x_n)|^2 + h^3 d + 2hd) \\ &\leq (1 - \alpha h + \frac{1}{4}Mh^2c_V + Mh^{\frac{3}{2}}c_V + 2\left\|D^2V\right\|_\infty h^2 c_V + \frac{2}{3}c_V\left\|D^2V\right\|_\infty M^2h^4c_V)V(x_n) \\ &\quad + \beta h + dMh^{\frac{3}{2}} + 2d\left\|D^2V\right\|_\infty h^3 + 4d\left\|D^2V\right\|_\infty h \\ &\leq \hat{\alpha}V(x_n) + \hat{\beta}\end{aligned}$$

for some $\hat{\alpha} \in (0,1)$ and $\hat{\beta} \geq 0$ when h is small. □

Once we have the Lyapunov condition, we can define the stopping time $\tau_C = \inf\{n > 0 : x_n \in C\}$ and show that $\sup_{x\in C}\mathbb{E}_x[\tau_C] \leq M_C < \infty$ for all small set C. Then uniqueness of stationary probability measure and ergodicity all follow by Theorem 1.3.1 in [33]. Next we prove that $\sup_{x\in C}\mathbb{E}_x[\tau_C] \leq M_C < \infty$ given Lyapunov condition. To do so, note that we have

$$\begin{aligned}\mathbb{E}_x[\tau_C] &= \sum_{k=1}^{\infty} k\mathbb{P}(\tau_C = k) \\ &= \sum_{k\geq 1}\mathbb{P}(\tau_C > k-1)\end{aligned}$$

Under Lyapunov condition, for any stopping time N, according to Lemma A.3 and Corollary A.4 in [31], we have

$$\begin{aligned}\mathbb{P}(\tau_C > k-1) &\leq \mathbb{E}[V(x_n)1_{\tau_C > k-1}] \\ &\leq \frac{\kappa[\gamma^{k-1}V(x_0) + 1]}{1-\gamma} \\ &\leq \kappa\gamma^{n-1}[V(x_0) + 1]\end{aligned}$$

for some $\gamma \in (\hat{\alpha}, 1)$ and constant κ. Therefore we have

$$\begin{aligned}\mathbb{E}_x[\tau_C] &\leq \sum_{k\geq 1}\kappa\gamma^{n-1}[V(x_0) + 1] \\ &= \frac{\kappa[V(x) + 1]}{1-\gamma}\end{aligned}$$

and

$$\sup_{x\in C}\mathbb{E}_x[\tau_C] \leq \frac{\kappa}{1-\gamma}\sup_{x\in C}V(x) + \frac{\kappa}{1-\gamma} \leq M_C < \infty$$

So as a conclusion, the statement of the theorem follows. □

Proof of Proposition 2.2. Consider that $x_n \sim \pi_h$ and $x_n^* \sim \pi$ are two independent random variables. Define x_{n+1} to be the one step RLMC result starting from x_n and $x_n^*(h)$ to be the solution of Langevin dynamics with initial value x_n^*. Therefore, $x_{n+1} \sim \pi_h$ and $x_n^*(h) \sim \pi$ are also independent and $\|x_n^* - x_n\|_{L^2} = \|x_n^*(h) - x_{n+1}\|_{L^2}$. We can compute the diffenrence between x_{n+1} and $x_n^*(h)$:

$$x_n^*(h) - x_{n+1} = (x_n^* - x_n) - \int_0^h \nabla f(x_n^*(s))ds + h\nabla f(x_n^*(\alpha_{n+1}h)) - h(-\nabla f(x_{n+\frac{1}{2}}) + \nabla f(x_n^*(\alpha_{n+1}h)))$$

It's easy to see that $\mathbb{E}_{\alpha_{n+1}}[\int_0^h \nabla f(x_n^*(s))ds - h\nabla f(x_n^*(\alpha_{n+1}h))] = 0$. And we can rewrite the last term as

$$\begin{aligned} h(-\nabla f(x_{n+\frac{1}{2}}) + \nabla f(x_n^*(\alpha_{n+1}h))) &= h(\nabla f(x_{n+\frac{1}{2}} + x_n^* - x_n) - \nabla f(x_{n+\frac{1}{2}})) \\ &\quad + h\nabla f(x_n^* - \int_0^{\alpha_{n+1}h} \nabla f(x_n^*(s))ds + \sqrt{2}W_{\alpha_{n+1}h}) \\ &\quad - h\nabla f(x_n^* - \alpha_{n+1}h\nabla f(x_n) + \sqrt{2\alpha_{n+1}h}U'_{n+1}) \end{aligned}$$

Take L_2-norm on other randomness:

$$\begin{aligned} \|x_n^*(h) - x_{n+1}\|_{L^2} \leq & \left\|(x_n^* - x_n) - h(\nabla f(x_{n+\frac{1}{2}} + x_n^* - x_n) - \nabla f(x_{n+\frac{1}{2}}))\right\|_{L^2} \\ & + h\left\|\nabla f(x_n^* - \int_0^{\alpha_{n+1}h} \nabla f(x_n^*(s))ds + \sqrt{2}W_{\alpha_{n+1}h}) - \nabla f(x_n^* - \alpha_{n+1}h\nabla f(x_n) + \sqrt{2\alpha_{n+1}h}U'_{n+1})\right\|_{L^2} \\ & + \left\|\int_0^h \nabla f(x_n^*(s))ds - h\nabla f(x_n^*(\alpha_{n+1}h))\right\|_{L^2} \end{aligned}$$

Since f is twice differentiable and f is also M-gradient Lipschitz and strongly convex with parameter m,

$$\left\|(x_n^* - x_n) - h(\nabla f(x_{n+\frac{1}{2}} + x_n^* - x_n) - \nabla f(x_{n+\frac{1}{2}}))\right\|_{L^2} \leq \rho \|x_n^* - x_n\|_{L^2}$$

where $\rho = \max(1 - mh, Mh - 1) = 1 - mh$.
For the second term:

$$\begin{aligned} & h\left\|\nabla f(x_n^* - \int_0^{\alpha_{n+1}h} \nabla f(x_n^*(s))ds + \sqrt{2}W_{\alpha_{n+1}h}) - \nabla f(x_n^* - \alpha_{n+1}h\nabla f(x_n) + \sqrt{2\alpha_{n+1}h}U'_{n+1})\right\|_{L^2} \\ & \leq Mh\left\|\int_0^{\alpha_{n+1}h} \nabla f(x_n^*(s)) - \nabla f(x_n)ds\right\|_{L^2} \\ & \leq \frac{\sqrt{3}}{3}M^2h^2\|x_n^* - x_n\|_{L^2} + \frac{\sqrt{3}}{3}M^2h^2 \sup_{0<s<h}\|x_n^*(s) - x_n^*\|_{L^2} \\ & \leq \frac{\sqrt{3}}{3}M^2h^2\|x_n^* - x_n\|_{L^2} + \frac{\sqrt{3}}{3}M^2h^2(4h^2\|\nabla f(x_n^*)\|^2 + 8M^2dh^3 + 2dh)^{\frac{1}{2}} \\ & \leq \frac{\sqrt{3}}{3}M^2h^2\|x_n^* - x_n\|_{L^2} + \frac{\sqrt{3}}{3}M^2h^2(2dh + 4Mdh^2 + 8M^2dh^3)^{\frac{1}{2}} \end{aligned}$$

For the third term:

$$\begin{aligned}
\left\|\int_0^h \nabla f(x_n^*(s))ds - h\nabla f(x_n^*(\alpha_{n+1}h))\right\|_{L^2} &= \{\mathbb{E}\mathbb{E}_{\alpha_{n+1}}[(\int_0^h \nabla f(x_n^*(s))ds - h\nabla f(x_n^*(\alpha_{n+1}h)))^2]\}^{\frac{1}{2}} \\
&= \{\mathbb{E}[h\int_0^h |\nabla f(x_n^*(s))|^2 ds - (\int_0^h \nabla f(x_n^*(s))ds)]\}^{\frac{1}{2}} \\
&= \{\mathbb{E}[h\int_0^h |\nabla f(x_n^*(s)) - \frac{1}{h}\int_0^h \nabla f(x_n^*(s'))ds'|^2 ds]\}^{1/2} \\
&\le \{\mathbb{E}[h\int_0^h \frac{1}{h}\int_0^h \|\nabla f(x_n^*(s)) - \nabla f(x_n^*(s'))\|^2 ds'ds]\}^{1/2} \\
&\le 2Mh\{\sup_{s\in(0,h)} \|x_n^*(s) - x_n^*\|^2\}^{1/2} \\
&\le 2Mh(4h^2\|\nabla f(x_n^*)\|^2 + 8M^2dh^3 + 2dh)^{1/2} \\
&\le 2Mh(2dh + 4Mdh^2 + 8M^2dh^3)^{\frac{1}{2}}
\end{aligned}$$

Combine all the bounds:

$$\|x_n^* - x_n\|_{L^2} \le \frac{\frac{\sqrt{3}}{3}M^2h^2(2dh + 4Mdh^2 + 8M^2dh^3)^{\frac{1}{2}} + 2Mh(2dh + 4Mdh^2 + 8M^2dh^3)^{\frac{1}{2}}}{mh - \frac{\sqrt{3}}{3}M^2h^2}$$

The final statement follows by the fact that $W_2(\pi, \pi_h) \le \|x_n - x_n^*\|_{L^2}$. □

6.3 Proofs for Section 2.3

Proof of Theorem 2.3. From previous analysis, if we keep track of the coefficients in all those bounds and assume that $M\gamma_n \le \frac{1}{2}$ for all n, we have:

$$\begin{aligned}
\mathbb{E}[\|x_{n+1} - y_{n+1}\|^2] &\le (1 + m\gamma_{n+1})\mathbb{E}[\left\|y_{n+1} - x_{n+1}^*\right\|^2] + \frac{1}{m\gamma_{n+1}}\mathbb{E}\left\|\mathbb{E}_{\alpha_{n+1}}x_{n+1} - x_{n+1}^*\right\|^2 + \mathbb{E}\left\|x_{n+1} - x_{n+1}^*\right\|^2 \\
&\le (1 + m\gamma_{n+1})e^{-2m\gamma_{n+1}}\mathbb{E}\|x_n - y_n\|^2 + \frac{1}{m\gamma_{n+1}}\mathbb{E}\left\|\mathbb{E}_{\alpha_{n+1}}x_{n+1} - x_{n+1}^*\right\|^2 + \mathbb{E}\left\|x_{n+1} - x_{n+1}^*\right\|^2 \\
&\le (1 + m\gamma_{n+1})e^{-2m\gamma_{n+1}}\mathbb{E}\|x_n - y_n\|^2 + \frac{2\gamma_{n+1}^2}{m\gamma_{n+1}}\mathbb{E}\left\|\nabla f(x_{n+\frac{1}{2}}) - \nabla f(x_n(\alpha_{n+1}\gamma_{n+1}))\right\|^2 \\
&\quad + 2\gamma_{n+1}^2\mathbb{E}\left\|\nabla f(x_{n+\frac{1}{2}}) - \nabla f(x_n(\alpha_{n+1}\gamma_{n+1}))\right\|^2 + 2M^2\gamma_{n+1}^2\mathbb{E}\sup_{t\in[0,\gamma_{n+1}]}\|x_n(\alpha_{n+1}\gamma_{n+1}) - x_n(t)\|^2 \\
&\le (1 + m\gamma_{n+1})e^{-2m\gamma_{n+1}}\mathbb{E}\|x_n - y_n\|^2 \\
&\quad + 2\gamma_{n+1}^4(1 + \frac{1}{m\gamma_{n+1}})M^4(\frac{1}{5}\gamma_{n+1}^2\|\nabla f(x_n)\|_{L^2}^2 + \frac{1}{6}M^2d\gamma_{n+1}^3 + \frac{2}{3}d\gamma_{n+1}) \\
&\quad + 4M^2\gamma_{n+1}^2\left(\frac{4\gamma_{n+1}^2}{1 - 2M^2\gamma_{n+1}}\|\nabla f(x_n)\|_{L^2}^2 + \frac{8M^2d\gamma_{n+1}^3}{1 - 2M^2\gamma_{n+1}^2} + 4d\gamma_{n+1}\right) \\
&\le (1 + m\gamma_{n+1})e^{-2m\gamma_{n+1}}\mathbb{E}\|x_n - y_n\|^2 + (33 + \kappa)M^2\gamma_{n+1}^4\|\nabla f(x_n)\|_{L^2}^2 + (33 + \kappa)M^2d\gamma_{n+1}^3
\end{aligned}$$

We can further bound $\|\nabla f(x_n)\|_{L^2}^2$:

$$\begin{aligned}
\|\nabla f(x_n)\|_{L^2}^2 &\le 2\|\nabla f(y_n)\|_{L^2}^2 + 2\|\nabla f(y_n) - \nabla f(x_n)\|_{L^2}^2 \\
&\le 2\|\nabla f(y_n)\|_{L^2}^2 + 2M^2\|x_n - y_n\|_{L^2}^2 \\
&\le 2Md + 2M^2\|x_n - y_n\|_{L^2}^2
\end{aligned}$$

Therefore we have the following iterative inequality:

$$\begin{aligned}
\mathbb{E}[\|x_{n+1} - y_{n+1}\|^2] &\le (1 + m\gamma_{n+1})e^{-2m\gamma_{n+1}}\mathbb{E}\|x_n - y_n\|^2 + 2(33 + \kappa)M^2d\gamma_{n+1}^3 + 2(33 + \kappa)M^4\gamma_{n+1}^4\mathbb{E}\|x_n - y_n\|^2 \\
&\le \left[1 - m\gamma_{n+1} + (\frac{m^2}{2} + \frac{M^2(33 + \kappa)}{2})\gamma_{n+1}^2\right]\mathbb{E}\|x_n - y_n\|^2 + 2(33 + \kappa)M^2d\gamma_{n+1}^3
\end{aligned}$$

Since (γ_n) is fast decreasing, we can assume that $\gamma_{n+1} \leq \frac{m}{m^2+M^2(33+\kappa)} \leq \frac{1}{m+34M}$ for large n, and for those n we have

$$\mathbb{E}[\|x_{n+1} - y_{n+1}\|^2] \leq (1 - \frac{1}{2}m\gamma_{n+1})\mathbb{E}\|x_n - y_n\|^2 + 2(33+\kappa)M^2 d\gamma_{n+1}^3$$

Our strategy of choosing (γ_n): for the first K_1 steps, we choose constant step size $h = \frac{1}{m+34M}$, K_1 is the first time so that $\mathbb{E}[\|x_{K_1} - y_{K_1}\|^2] \leq 5\kappa(\kappa+33)M(\frac{d^{\frac{1}{2}}}{m+34M})^2$. such K_1 exists because

$$\begin{aligned}\mathbb{E}[\|x_{K_1} - y_{K_1}\|^2] &\leq (1 - \frac{m}{2m+68M})^{K_1}\mathbb{E}[\|x_0 - y_0\|^2] + \frac{2M^2(\kappa+33)d}{(m+34M)^3}\frac{2(m+34M)}{m} \\ &= (1 - \frac{m}{2m+68M})^{K_1}\mathbb{E}[\|x_0 - y_0\|^2] + 4\kappa(\kappa+33)M(\frac{d^{\frac{1}{2}}}{m+34M})^2\end{aligned}$$

Claim: There exists $\lambda > 0$ such that if we choose $\gamma_{n+1} = \frac{1}{m+34M+\lambda(n-K_1)}$ for all $n \geq K_1$, we can ensure that $\mathbb{E}[\|x_k - y_k\|^2] \leq 5\kappa(\kappa+33)M(\frac{d^{\frac{1}{2}}}{m+34M+\lambda(n-K_1)})^2$ for all $n \geq K_1$.

Proof of Claim: Simply use induction:

$$\begin{aligned}\mathbb{E}[\|x_{n+1} - y_{n+1}\|^2] &\leq (1 - \frac{1}{2}m\gamma_{n+1})5\kappa(\kappa+33)Md\gamma_{n+1}^2 + 2M^2(\kappa+33)d\gamma_{n+1}^3 \\ &= 5\kappa(\kappa+33)Md\gamma_{n+1}^2(1 - \frac{m}{10}\gamma_{n+1})\end{aligned}$$

Our goal is to ensure $5\kappa(\kappa+33)Md\gamma_{n+1}^2(1-\frac{m}{10}\gamma_{n+1}) < 5\kappa(\kappa+33)M(\frac{d^{\frac{1}{2}}}{m+34M+\lambda(n+1-K_1)})^2$. It boils down to discuss the following polynomial inequality relates to λ:

$$G(\lambda) = (K - \frac{1}{10}m(K+1)^2)\lambda^2 + (X - \frac{1}{5}mX(K+1))\lambda - \frac{1}{10}mX^2 \leq 0$$

where $X = m + 34M$ and $K = n - K_1 > 0$. It's not hard to see that there's always positive λ satisfying the inequality.

At last to get small error, we require $\mathbb{E}\|x_n - y_n\|^2 \leq \frac{d\epsilon^2}{m}$, i.e

$$5\kappa(\kappa+33)M\frac{d}{(m+34M+\lambda(n-K_1))^2} \leq \frac{d\epsilon^2}{m}$$

Then we have

$$n \geq K_1 + \lambda^{-1}m^{\frac{1}{2}}M^{\frac{1}{2}}\kappa^{\frac{1}{2}}(\kappa+33)^{\frac{1}{2}}/\epsilon - \lambda^{-1}(m+34M) \sim O(\kappa^{\frac{3}{2}}/\epsilon)$$

□

6.3.1 Proof of Theorem 2

Before we prove Theorem 2, we need several intermediate results on the tightness of the (RLMC) chain.

Lemma 3. *Under assumption 6.1, for every continuous function φ satisfying $\varphi(x) = o(V^k(x))$ for some $k \in \mathbb{N}$, $\lim_n \pi_n^\gamma(\varphi) = \pi(\varphi)$.*

Proof of Lemma 3. The proof is divided into three steps:

1) For all $p \geq 1$, there exists $\tilde{\alpha} \in (0,1)$ and $\tilde{\beta}, n_0 \in \mathbb{N}$ such that $\mathbb{E}[V^p(x_{n+1})|\mathcal{F}_n] \leq V^p(x_n) + \gamma_{n+1}V^{p-1}(x_n)(\tilde{\beta} - \tilde{\alpha}V(x_n))$ for all $n \geq n_0$.
When $p = 1$, the statement follows from assumption 6.1.
When $p > 1$, first we Taylor expand $V^p(x_{n+1})$ at x_n:

$$\begin{aligned}V^p(x_{n+1}) &= V^p(x_n) + pV^{p-1}(x_n)\nabla V(x_n)\cdot(x_{n+1} - x_n) + \frac{1}{2}D^2(V^p)(\xi_{n+1})(x_{n+1} - x_n)^{\otimes 2} \\ &= V^p(x_n) - \gamma_{n+1}pV^{p-1}(x_n)\nabla V(x_n)\cdot\nabla f(x_{n+\frac{1}{2}}) + \sqrt{2\gamma_{n+1}}pV^{p-1}\nabla V(x_n)\cdot U_{n+1} \\ &\quad + \frac{1}{2}D^2(V^p)(\xi_{n+1})\left(-\gamma_{n+1}\nabla f(x_{n+\frac{1}{2}}) + \sqrt{2\gamma_{n+1}}U_{n+1}\right)^{\otimes 2} \\ &\leq V^p(x_n) - \gamma_{n+1}pV^{p-1}(x_n)\nabla V(x_n)\cdot\nabla f(x_{n+\frac{1}{2}}) + \sqrt{2\gamma_{n+1}}pV^{p-1}(x_n)\nabla V(x_n)\cdot U_{n+1} \\ &\quad + p\lambda_p V^{p-1}(\xi_{n+1})|-\gamma_{n+1}\nabla f(x_{n+\frac{1}{2}}) + \sqrt{2\gamma_{n+1}}U_{n+1}|^2\end{aligned}$$

where ξ_{n+1} is a point on the line segment joining x_n and x_{n+1} and $\lambda_p := \frac{1}{2}\lambda_{D^2V+(p-1)(\nabla V\otimes\nabla V)/V} < +\infty$.

Due to $\nabla(\sqrt{V}) = \frac{\nabla V}{2V}$ and $|\nabla V|^2 \leq c_V V$, we have $\sqrt{V}$ is Lipschitz continuous and the Lipschitz constant $[\sqrt{V}]_1 = \frac{1}{4}c_V < +\infty$. Hence for a point ξ_{n+1} on the line segment between x_n and x_{n+1},

$$\begin{aligned}
V^{p-1}(\xi_{n+1}) = (\sqrt{V})^{2(p-1)}(\xi_{n+1}) &\leq \left(\sqrt{V}(x_n) + [\sqrt{V}]_1|x_{n+1}-x_n|\right)^{2(p-1)} \\
&\leq \begin{cases} V^{p-1}(x_n) + [\sqrt{V}]_1^{2(p-1)}|x_{n+1}-x_n|^{2(p-1)}, & 2(p-1)\leq 1 \\ V^{p-1}(x_n) + c\left(V^{(2p-3)/2}(x_n)|x_{n+1}-x_n| + |x_{n+1}-x_n|^{2(p-1)}\right), & 2(p-1)>1 \end{cases}
\end{aligned}$$

We can further bound

$$\begin{aligned}
|x_{n+1}-x_n| &= |-\gamma_{n+1}\nabla f(x_n) + \sqrt{2\gamma_{n+1}}U_{n+1} - \gamma_{n+1}\left(\nabla f(x_{n+\frac{1}{2}}) - \nabla f(x_n)\right)| \\
&\leq \gamma_{n+1}|\nabla f(x_n)| + \sqrt{2\gamma_{n+1}}|U_{n+1}| + M\gamma_{n+1}|-\tilde{\gamma}_{n+1}\nabla f(x_n) + \sqrt{2\alpha_{n+1}\gamma_{n+1}}U'_{n+1}| \\
&\leq \gamma_{n+1}(1+M\alpha_{n+1}\gamma_{n+1})|\nabla f(x_n)| + \sqrt{2\gamma_{n+1}}|U_{n+1}| + \sqrt{2}M\gamma_{n+1}\alpha_{n+1}^{\frac{1}{2}}\gamma_{n+1}^{\frac{1}{2}}|U'_{n+1}| \\
&\leq C\sqrt{V}(x_n)\gamma_{n+1}^{\frac{1}{2}}(1+|U_{n+1}|+|U'_{n+1}|)
\end{aligned}$$

Plug these results into the last term in the first inequality we obtained from Taylor expansion:

$$\begin{aligned}
p\lambda_p V^{p-1}(\xi_{n+1})|x_{n+1}-x_n|^2 &\leq p\lambda_p V^{p-1}(x_n)|x_{n+1}-x_n|^2 \\
&\quad + Cp\lambda_p \begin{cases} |x_{n+1}-x_n|^{2p}, & 2p\leq 3 \\ V^{(2p-3)/2}(x_n)|x_{n+1}-x_n|^3 + |x_{n+1}-x_n|^{2p}, & 2p>3 \end{cases} \\
&\leq p\lambda_p V^{p-1}(x_n)|x_{n+1}-x_n|^2 + C\gamma_{n+1}^{p\wedge\frac{3}{2}}V^p(x_n)(1+|U_{n+1}|^{2p}+|U'_{n+1}|^{2p})
\end{aligned}$$

We then take conditional expectation, there exists $\alpha>0$ and $\beta\geq 0$ such that for all $n\geq n_0$:

$$\begin{aligned}
\mathbb{E}[V^p(x_{n+1})|\mathcal{F}_n] &\leq V^p(x_n) - pV^{p-1}(x_n)(\alpha V(x_n)-\beta) \\
&\quad - p\gamma_{n+1}V^{p-1}(x_n)\mathbb{E}[\nabla V(x_n)\cdot\left(\nabla f(x_{n+\frac{1}{2}}) - \nabla f(x_n)\right)|\mathcal{F}_n] \\
&\quad + 2p\lambda_p V^{p-1}(x_n)\mathbb{E}[\gamma_{n+1}^2|\nabla f(x_{n+\frac{1}{2}})|^2 + 2\gamma_{n+1}|U_{n+1}|^2|\mathcal{F}_n] \\
&\quad + CV^p(x_n)(1+\mathbb{E}|U_{n+1}|^{2p}+\mathbb{E}|U'_{n+1}|^{2p})\gamma_{n+1}^{p\wedge\frac{3}{2}} \\
&\leq V^p(x_n) - pV^{p-1}(x_n)(\alpha V(x_n)-\beta) + 2p\lambda_p\mathbb{E}|U_{n+1}|^2\gamma_{n+1}V^{p-1}(x_n) \\
&\quad + CV^p(x_n)(1+\mathbb{E}|U_{n+1}|^{2p}+\mathbb{E}|U'_{n+1}|^{2p})\gamma_{n+1}^{p\wedge\frac{3}{2}} \\
&\quad + c_V Mp\gamma_{n+1}^2 V^p(X_n) + \sqrt{2}c_V Mp\gamma_{n+1}^{\frac{3}{2}}\mathbb{E}|U'_{n+1}|V^{p-1/2}(x_n) \\
&\quad + c_V p\lambda_p\gamma_{n+1}^2 V^{p-1}(x_n)\mathbb{E}[V(x_{n+\frac{1}{2}})|\mathcal{F}_n]
\end{aligned}$$

From x_n to $x_{n+\frac{1}{2}}$, it's simply the Euler discretization with time step $\alpha_{n+1}\gamma_{n+1}$, we could use the result in [21]: there exists a $\bar{\alpha}>0$ and $\bar{\beta}\in\mathbb{R}$ such that for all $n\geq n_0$:

$$\mathbb{E}[V(x_{n+\frac{1}{2}})|\mathcal{F}_n] \leq V(x_n)(1-\bar{\alpha}\tilde{\gamma}_{n+1}) + \bar{\beta}\tilde{\gamma}_{n+1}$$

Therefore we have

$$\mathbb{E}[V^{x_{n+1}}|\mathcal{F}_n] \leq (1-\alpha p\gamma_{n+1}+o(\gamma_{n+1}))V^p(x_n) + \gamma_{n+1}V^{p-1}(x_n)(p\beta+2p\lambda_p\mathbb{E}|U_{n+1}|^2 + c_V Mp\mathbb{E}|U'_{n+1}|^2)$$

There exists $\hat{\alpha}>0$ and $\hat{\beta}\in\mathbb{R}$ such that for all $n\geq n_0$:

$$\mathbb{E}[V^p(x_{n+1})|\mathcal{F}_n] \leq V^p(x_n) + \gamma_{n+1}V^{p-1}(x_n)\left(\hat{\beta}-\hat{\alpha}V(x_n)\right)$$

2) From step 1), we derive

$$\sup_{n\geq n_0}\mathbb{E}[V^p(x_n)] \leq (\frac{\hat{\beta}}{\hat{\alpha}})^p \vee \mathbb{E}[V^p(x_{n_0})]$$

Hence $\sup_n \mathbb{E}[V^p(x_n)] < +\infty$ for all $p\geq 1$. Therefore $\sup_n \pi_n^\gamma(\omega, V^p) < +\infty \ \ \mathbb{P}-a.s$ for all $p\geq 1$.

3) Identification of the weak limit: To identify the limit, we essentially follow the same steps in [21] and hence we omit the proof.
   (a) **(Echeverrría-Weiss Theorem)** Let E be a locally compact Polish space and A a linear operator satisfying the positive maximum principle. Assume that its domain $\mathcal{D}(A)$ is an algebra everywhere dense in $(\mathcal{C}_0(E), \|\ \|_\infty)$ containing a sequence $(f_n)_{n\in\mathbb{N}}$ satisfying
   $$\sup_{n\in\mathbb{N}} (\|f_n\|_\infty + \|\mathcal{L}f_n\|_\infty) < +\infty, \ \ \forall x \in E, \ \ f_n(x) \to 1 \ \text{ and } \ Af_n(x) \to 0.$$
   If a distribution on $(E, \mathcal{B}(E))$ satisfies $\int_E Af d\nu = 0$ for every $f \in \mathcal{D}(A)$, then there exists a stationary solution for the martingale problem (A, ν) (this means that there exists a stationary continuous-time homogeneous Markov process with infinitesimal generator A and invariant distribution ν).
   (b) The generator of the Langevin dynamics, $\mathcal{A}$, satisfies the assumptions of the Echeverrría-Weiss theorem.
   (c) Under assumption 6.1, for every bounded Lipschitz continuous function $\varphi : \mathbb{R}^d \to \mathbb{R}$, $\lim_n \frac{1}{\Gamma_n} \sum_{k=1}^n \mathbb{E}[\varphi(x_k) - \varphi(x_{k-1})|\mathcal{F}_{k-1}] = 0 \ \ \mathbb{P} - a.s.$
   (d) Under assumption 6.1, for every twice continuously differentiable function φ with compact support, $\lim_n \left(\frac{1}{\Gamma_n} \sum_{k=1}^n \mathbb{E}[\varphi(x_k) - \varphi(x_{k-1})|\mathcal{F}_{k-1}] - \pi_n^\gamma(\mathcal{A}\varphi) \right) = 0 \ \ \mathbb{P} - a.s.$

   $a), b), c), d)$ together imply that the weak limit of the empirical distribution π_n^γ is π, i.e the stationary distribution of the Langevin dynamics.

□

Proof of Theorem 2. Since f satisfies assumption 1.1, we can show that the Langevin dynamics satisfies assumption 6.1. Therefore lemma 3 is true. Then we may use the following method to discuss the CLT of (RLMC).

$$\begin{aligned}
x_k - x_{k-1} &= -\gamma_k \left(\nabla f(x_{k-1}) + D^2 f(x_{k-1})(x_{k-\frac{1}{2}} - x_{k-1}) + r_2(x_{k-\frac{1}{2}}, x_{k-1}) \right) + \sqrt{2\gamma_k} U_k \\
&= -\gamma_k \nabla f(x_{k-1}) + \sqrt{2\gamma_k} U_k - \gamma_k D^2 f(x_{k-1})(x_{k-\frac{1}{2}} - x_{k-1}) - \gamma_k r_2(x_{k-\frac{1}{2}}, x_{k-1}) \\
&= -\gamma_k \nabla f(x_{k-1}) + \sqrt{2\gamma_k} U_k + \alpha_k \gamma_k^2 D^2 f(x_{k-1}) \nabla f(x_{k-1}) - \sqrt{2\alpha_k} \gamma_k^{\frac{3}{2}} \nabla^2 f(x_{k-1}) U_k' - \gamma_k r_2(x_{k-\frac{1}{2}}, x_{k-1})
\end{aligned}$$

where

$$\begin{aligned}
r_2(x_{k-\frac{1}{2}}, x_{k-1}) &= \nabla f(x_{k-\frac{1}{2}}) - \nabla f(x_{k-1}) - D^2 f(x_{k-1})(x_{k-1} - x_{k-\frac{1}{2}}) \\
&= \frac{1}{2} D^3 f(x_{k-1})(x_{k-\frac{1}{2}} - x_{k-1})^{\otimes 2} + \frac{1}{6} D^4 f(x_{k-1})(x_{k-\frac{1}{2}} - x_{k-1})^{\otimes 3} + O(\gamma_k^2) \\
&= \alpha_k \gamma_k D^3 f(x_{k-1}) U_k'^{\otimes 2} - \sqrt{2} \alpha_k^{\frac{3}{2}} \gamma_k^{\frac{3}{2}} \langle D^3 f(x_{k-1}); \nabla f(x_{k-1}), U_k' \rangle \\
&\quad + \frac{\sqrt{2}}{3} \alpha_k^{\frac{3}{2}} D^4 f(x_{k-1}) U_k'^{\otimes 4} + O(\gamma_k^2)
\end{aligned}$$

Then

$$\begin{aligned}
x_k - x_{k-1} = &-\gamma_k \nabla f(x_{k-1}) + \sqrt{2\gamma_k} U_k - \sqrt{2\alpha_k} \gamma_k^{\frac{3}{2}} \nabla^2 f(x_{k-1}) U_k' \\
&+ \alpha_k \gamma_k^2 D^2 f(x_{k-1}) \nabla f(x_{k-1}) - \alpha_k \gamma_k^2 D^3 f(x_{k-1}) U_k'^{\otimes 2} + O(\gamma_k^{\frac{5}{2}})
\end{aligned}$$

We can decompose $\phi(x_k)$:

$$\begin{aligned}
\phi(x_k) - \phi(x_{k-1}) &= \nabla \phi(x_{k-1})(x_k - x_{k-1}) + \frac{1}{2} D^2 \phi(x_{k-1})(x_k - x_{k-1})^{\otimes 2} + \frac{1}{6} D^3 \phi(x_{k-1})(x_k - x_{k-1})^{\otimes 3} \\
&\quad + \frac{1}{24} D^4 \phi(x_{k-1})(x_k - x_{k-1})^{\otimes 4} + O(\gamma_k^{\frac{5}{2}}) \\
&= \nabla \phi(x_{k-1})(\sqrt{2} \gamma_k^{\frac{1}{2}} U_k - \gamma_k \nabla f(x_{k-1}) - \sqrt{2\alpha_k} \gamma_k^{\frac{3}{2}} D^2 f(x_{k-1}) U_k' \\
&\quad + \alpha_k \gamma_k^2 D^2 f(x_{k-1}) \nabla f(x_{k-1}) - \alpha_k \gamma_k^2 D^3 f(x_{k-1}) U_k'^{\otimes 2}) \\
&\quad + \frac{1}{2} D^2 \phi(x_{k-1}) \left(\sqrt{2} \gamma_k^{\frac{1}{2}} U_k - \gamma_k \nabla f(x_{k-1}) - \sqrt{2\alpha_k} \gamma_k^{\frac{3}{2}} D^2 f(x_{k-1}) U_k' \right)^{\otimes 2} \\
&\quad + \frac{1}{6} D^3 \phi(x_{k-1}) \left(\sqrt{2} \gamma_k^{\frac{1}{2}} U_k - \gamma_k \nabla f(x_{k-1}) \right)^{\otimes 3} + \frac{1}{24} D^4 \phi(x_{k-1}) (\sqrt{2} \gamma_k^{\frac{1}{2}} U_k)^{\otimes 4} + O(\gamma_k^{\frac{5}{2}})
\end{aligned}$$

If $\mathcal{A}$ is the generator of Langevin dynamics and summing up over k:

$$\begin{aligned}
\sum_{k=1}^{n}\gamma_k\mathcal{A}\phi(x_{k-1}) &= \phi(x_n)-\phi(x_0)-\sqrt{2}\sum_{k=1}^{n}\gamma_k^{\frac{1}{2}}\nabla\phi(x_{k-1})U_k-\sum_{k=1}^{n}\gamma_k\left(D^2\phi(x_{k-1})U_k^{\otimes 2}-\mathbb{E}[D^2\phi(x_{k-1})U_k^{\otimes 2}|\mathcal{F}_{k-1}]\right)\\
&+\sqrt{2}\sum_{k=1}^{n}\gamma_k^{\frac{3}{2}}\langle D^2\phi(x_{k-1});\nabla f(x_{k-1}),U_k\rangle-\frac{\sqrt{2}}{3}\sum_{k=1}^{n}\gamma_k^{\frac{3}{2}}D^3\phi(x_{k-1})U_k^{\otimes 3}\\
&+\sum_{k=1}^{n}\sqrt{2\alpha_k}\gamma_k^{\frac{3}{2}}\langle D^2 f(x_{k-1});\nabla\phi(x_{k-1}),U_k'\rangle+\sum_{k=1}^{n}\gamma_k^2\langle D^3\phi(x_{k-1});\nabla f(x_{k-1}),U_k^{\otimes 2}\rangle\\
&-\sum_{k=1}^{n}\alpha_k\gamma_k^2\langle D^2 f(x_{k-1});\nabla\phi(x_{k-1}),\nabla f(x_{k-1})\rangle+\sum_{k=1}^{n}\alpha_k\gamma_k^2\langle D^3 f(x_{k-1});\nabla\phi(x_{k-1}),U_k'^{\otimes 2}\rangle\\
&-\frac{1}{2}\sum_{k=1}^{n}\gamma_k^2 D^2\phi(x_{k-1})\nabla f(x_{k-1})^{\otimes 2}+\sum_{k=1}^{n}2\alpha_k^{\frac{1}{2}}\gamma_k^2\langle D^2\phi(x_{k-1});D^2\phi(x_{k-1})U_k',U_k\rangle\\
&-\frac{1}{6}\sum_{k=1}^{n}\gamma_k^2 D^4\phi(x_{k-1})U_k^{\otimes 4}+\sum_{k=1}^{n}O(\gamma_k^{\frac{5}{2}})\\
&:= N_n^{(0)}+N_n^{(\frac{1}{2})}+N_n^{(1)}+N_n^{(\frac{3}{2})}+N_n^{(2)}+N_n^{(\frac{5}{2})}
\end{aligned}$$

In the fast decreasing time step situation($\sum_{k=1}^{n}\gamma_k^2/\sqrt{\Gamma_n}\to 0$), the CLT for (RLMC) is the same as that of LMC. In the slowly decreasing time step situation, when $\sum_{k=1}^{n}\gamma_k^2/\sqrt{\Gamma_n}\to\hat{\gamma}\in(0,+\infty]$:

a) $\frac{\phi(x_n)-\phi(x_0)}{\Gamma_n^{(2)}}\to 0$ because (x_n) is tight and ϕ is continuous.

b) $\frac{-\sqrt{2}\sum_{k=1}^{n}\gamma_k^{\frac{1}{2}}\nabla\phi(x_{k-1})U_k}{\sqrt{\Gamma_n}}\implies\mathcal{N}(0,2\int_{\mathbb{R}^d}|\nabla\phi(x)|^2\pi(dx))$. Therefore,

$$\frac{-\sqrt{2}\sum_{k=1}^{n}\gamma_k^{\frac{1}{2}}\nabla\phi(x_{k-1})U_k}{\Gamma_n^{(2)}}\implies\begin{cases}\mathcal{N}(0,2\hat{\gamma}^{-2}\int_{\mathbb{R}^d}|\nabla\phi(x)|^2\pi(dx)), & \text{when } \hat{\gamma}<+\infty\\ 0\quad, & \text{when } \hat{\gamma}=+\infty\end{cases}$$

c) $\frac{-\sum_{k=1}^{n}\gamma_k\left(D^2\phi(x_{k-1})U_k^{\otimes 2}-\mathbb{E}[D^2\phi(x_{k-1})U_k^{\otimes 2}|\mathcal{F}_{k-1}]\right)}{\sqrt{\Gamma_n}}\to 0$ in L^2.

d) $\frac{\sqrt{2}\sum_{k=1}^{n}\gamma_k^{\frac{3}{2}}\langle D^2\phi(x_{k-1});\nabla f(x_{k-1}),U_k\rangle}{\sqrt{\Gamma_n}}\to 0$ in L^2.

$\frac{-\frac{\sqrt{2}}{3}\sum_{k=1}^{n}\gamma_k^{\frac{3}{2}}D^3\phi(x_{k-1})U_k^{\otimes 3}}{\sqrt{\Gamma_n}}\to 0$ in probability because $\mathbb{E}[U_k^{\otimes 3}]=0$.

$\frac{\sum_{k=1}^{n}\sqrt{2\alpha_k}\gamma_k^{\frac{3}{2}}\langle D^2 f(x_{k-1});\nabla\phi(x_{k-1}),U_k'\rangle}{\sqrt{\Gamma_n}}\to 0$ in L^2.

Therefore $\frac{N_n^{(\frac{3}{2})}}{\Gamma_n^{(2)}}\to 0$ in probability.

e) $\frac{\sum_{k=1}^{n}\gamma_k^2\langle D^3\phi(x_{k-1});\nabla f(x_{k-1}),U_k^{\otimes 2}\rangle}{\Gamma_n^{(2)}}\to\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\langle D^3\phi(x);\nabla f(x),u^{\otimes 2}\rangle\mu(du)\pi(dx)$ in probability.

$\frac{-\sum_{k=1}^{n}\alpha_k\gamma_k^2\langle D^2 f(x_{k-1});\nabla\phi(x_{k-1}),\nabla f(x_{k-1})\rangle}{\Gamma_n^{(2)}}\to-\frac{1}{2}\int_{\mathbb{R}^d}\langle D^2 f(x);\nabla\phi(x),\nabla f(x)\rangle\pi(dx)$ in probability.

$\frac{\sum_{k=1}^{n}\alpha_k\gamma_k^2\langle D^3 f(x_{k-1});\nabla\phi(x_{k-1}),U_k'^{\otimes 2}\rangle}{\Gamma_n^{(2)}}\to\frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^2}\langle D^3 f(x);\nabla\phi(x),u^{\otimes 2}\rangle\mu'(du)\pi(dx)$ in probability.

$\frac{-\frac{1}{2}\sum_{k=1}^{n}\gamma_k^2 D^2\phi(x_{k-1})\nabla f(x_{k-1})^{\otimes 2}}{\Gamma_n^{(2)}}\to-\frac{1}{2}\int_{\mathbb{R}^d}D^2\phi(x)\nabla f(x)^{\otimes 2}\pi(dx)$ in probability.

$\frac{\sum_{k=1}^{n}2\alpha_k^{\frac{1}{2}}\gamma_k^2\langle D^2\phi(x_{k-1});D^2\phi(x_{k-1})U_k',U_k\rangle}{\Gamma_n^{(2)}}\to 0$ in L^2.

$\frac{-\frac{1}{6}\sum_{k=1}^{n}\gamma_k^2 D^4\phi(x_{k-1})U_k^{\otimes 4}}{\Gamma_n^{(2)}}\to-\frac{1}{6}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}D^4\phi(x)u^{\otimes 4}\mu(du)\pi(dx)$ in probability.

Therefore

$$\frac{N_n^{(2)}}{\Gamma_n^{(2)}} \to \varrho \qquad \text{in probability}$$

where

$$\begin{aligned}\varrho = &\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \langle D^3\phi(x); \nabla f(x), u^{\otimes 2}\rangle \mu(du)\pi(dx) - \frac{1}{2}\int_{\mathbb{R}^d} \langle D^2 f(x); \nabla\phi(x), \nabla f(x)\rangle \pi(dx)\\ &+ \frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^2} \langle D^3 f(x); \nabla\phi(x), u^{\otimes 2}\rangle \mu(du)\pi(dx) - \frac{1}{2}\int_{\mathbb{R}^d} D^2\phi(x)\nabla f(x)^{\otimes 2}\pi(dx)\\ &- \frac{1}{6}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} D^4\phi(x)u^{\otimes 4}\mu(du)\pi(dx)\end{aligned}$$

and μ is the distribution for a d-dimensional standard Gaussian random variable.

f) $\frac{N_n^{\frac{5}{2}}}{\Gamma_n^{(2)}} \to 0$ in L^1.

As a conclusion, we obtain the proof of part (1) of the theorem:

$$\frac{\sum_{k=1}^n \gamma_k \mathcal{A}\phi(x_{k-1})}{\Gamma_n^{(2)}} \to \begin{cases}\mathcal{N}(\varrho, 2\hat{\gamma}^{-2}\int_{\mathbb{R}^d} |\nabla\phi(x)|^2\pi(dx)), & \text{when } \hat{\gamma} < +\infty\\ \varrho \quad , & \text{when } \hat{\gamma} = +\infty\end{cases}$$

For the fast decreasing step, i.e., part (2) of the theorem, the proof follows by the same arguments in the corresponding part of Theorem 10 in [21] and hence we omit it. □

7 Proofs for Section 3

In this section, we would denote the drift function that appears in 4 as $b(x, v)$, i.e.

$$b(x, v) = \begin{bmatrix} v \\ -2v - u\nabla f(x)\end{bmatrix}$$

Assumption 7.1. *There exists a twice differentiable function $V : \mathbb{R}^{2d} \to [1, \infty)$ such that: (0) $\lim_{\|(x,v)\|\to\infty} V(x, v) = +\infty$, (1) there exists $\alpha > 0$ and $\beta > 0$: $\langle \nabla V(x, v), b(x, v)\rangle \le -\alpha V(x, v) + \beta$ for every (x, v), (2) there exists $c_V > 0$: $\|\nabla V(x, v)\|^2 + \|b(x, v)\|^2 \le c_V V(x, v)$ for every (x, v), and (3) $\left\|D^2 V\right\|_\infty := \sup_{(x,v)\in\mathbb{R}^{2d}} \|D^2 V\|_{op} < \infty$.*

Lemma 4. *Assumption 1.1 implies Assumption 7.1 when $u \in (0, \frac{4}{2M-m})$.*

Proof of Lemma 4. For simplicity, We choose $V(x, v) = \|x - x_*\|^2 + \|x - x_* + v\|^1 + 1$ with $f(x_*) = \min f(x)$. Now we check conditions 0), 1), 2), 3) in $(\mathcal{L}_{V,\infty})$ are satisfied.

0) It's onvious that $\lim_{|(x,v)|\to+\infty} V(x, v) = +\infty$ and $V(x, v) \ge 1$ for all $(x, v) \in \mathbb{R}^d$.
3) The Hessian of V we choose is

$$D^2 V(x, v) = \begin{bmatrix} 4I_d & 2I_d \\ 2I_d & 2I_d\end{bmatrix}$$

For arbitrary $(x, v)^T, (y, w)^T \in \mathbb{R}^{2d}$:

$$\begin{aligned}\left\|D^2 V(x, v)(y, w)^T\right\|^2 &= \left\|\begin{bmatrix}4y + 2w\\ 2y + 2w\end{bmatrix}\right\|^2\\ &\le 40\left\|(y, w)^T\right\|^2\end{aligned}$$

Therefore $\left\|D^2 V\right\|_\infty < \infty$.

2) Take gradient of the V we choose:

$$\nabla V(x,v) = \begin{bmatrix} 2(x-x_*) + 2(x-x_*+v) \\ 2(x-x_*+v) \end{bmatrix}$$

Then for all $(x,v) \in \mathbb{R}^{2d}$,

$$\begin{aligned} |\nabla V(x,v)|^2 + |b(x,v)|^2 &\le 2(4\|x-x_*\|^2 + 4\|x-x_*+v\|^2) + 4\|x-x_*+v\|^2 \\ &\quad + \|v\|^2 + 2(4\|v\|^2 + u^2\|\nabla f(x)\|^2) \\ &\le 8\|x-x_*\|^2 + 12\|x-x_*+v\|^2 + 9\|v\|^2 + 2u^2M^2\|x-x_*\|^2 \\ &\le \max\{26 + 2u^2M^2, 30\} V(x,v) \end{aligned}$$

1) Last we consider

$$\begin{aligned} \langle \nabla V(x,v), b(x,v)\rangle &= 2(x-x_*)\cdot v + 2(x-x_*+v)\cdot v - 4(x-x_*+v)\cdot v \\ &\quad - 2u(x-x_*+v)\cdot \nabla f(x) \\ &\le -2\|v\|^2 - 2u\left[f(x) - f(x_*-v) + \frac{m}{2}\|x-x_*+v\|^2\right] \\ &\le -2\|v\|^2 - um\|x-x_*+v\|^2 - 2u\left(f(x_*) + \frac{m}{2}\|x-x_*\|^2\right) \\ &\quad + 2u\left(f(x_*) + \frac{M}{2}\|v\|^2\right) \\ &= -um\|x-x_*+v\|^2 - um\|x-x_*\|^2 - (2-uM)\|v\|^2 \end{aligned}$$

The second inequality follows from the fact that f is m-strongly convex.
When $u \in (0, \frac{2}{M}]$, $\langle \nabla V(x,v), b(x,v)\rangle \le -umV(x,v) + um$ for all $(x,v) \in \mathbb{R}^{2d}$. Therefore 1) is satisfied.
When $u > \frac{2}{M}$, we can use triangle inequality to further bound our result:

$$\begin{aligned} \langle \nabla V(x,v), b(x,v)\rangle &\le -um\|x-x_*+v\|^2 - um\|x-x_*\|^2 + (uM-2)\|v\|^2 \\ &\le [-um + 2(uM-2)](\|x-x_*+v\|^2 + \|x-x_*\|^2) \\ &\le -[4-u(2M-m)]V(x,v) - [4-u(2M-m)] \end{aligned}$$

When $u \in (\frac{2}{M}, \frac{4}{2M-m})$, 1) is satisfied because $4-u(2M-m) > 0$.
Therefore, 1) holds when $u \in (0, \frac{4}{2M-m})$.

□

Remark 9. *For the $V(x,v)$ we choose in the proof, under assumption 1.1, we can verify that: $V(x,v) = O(|x|^2 + |v|^2)$ when $|(x,v)| \to +\infty$. We will use this fact later in the proof when we establish the CLT statement.*

7.1 Proofs for Section 3.1

Proof of Theorem 3. Under the assumption 7.1, we can show that the following Lyapunov condition is satisfied for small h.
(Lyapunov Condition): There exists a function $V : \mathbb{R}^{2d} \to [1,\infty)$ such that:

0) $\lim_{|(x,v)|\to\infty} V(x,v) = +\infty$,
1) There exists $\hat{\alpha} \in (0,1)$ and $\hat{\beta} \ge 0$: $\mathbb{E}[V(x_{n+1},v_{n+1})|\mathcal{F}_n] \le \hat{\alpha}V(x_n,v_n) + \hat{\beta}$.

Proof: To show that assumption 7.1 implies Lyapunov condition, we first do Taylor expansion of $V(x_{n+1},v_{n+1})$ at (x_n,v_n):

$$V(x_{n+1},v_{n+1}) = V(x_n,v_n) + \nabla V(x_n,v_n)\cdot(x_{n+1}-x_n, v_{n+1}-v_n)^T + \frac{1}{2}D^2V(\theta_n)[(x_{n+1}-x_n, v_{n+1}-v_n)^T]^{\otimes 2}$$

where θ_n is a random point on the line segment joining (x_n, v_n) and (x_{n+1}, v_{n+1}). Use the RULMC algorithm and part (a) of Assumption 1.1:

$$\begin{aligned}\mathbb{E}[V(x_{n+1},v_{n+1})|\mathcal{F}_n] \le V(x_n,v_n) &+ \nabla V(x_n,v_n)\cdot\begin{bmatrix}\frac{1-e^{-2h}}{2}v_n - \frac{u}{2}(h-\frac{1-e^{-2h}}{2})\nabla f(x_n)\\ -2\frac{1-e^{-2h}}{2}v_n - u\frac{1-e^{-2h}}{2}\nabla f(x_n)\end{bmatrix}\\ &- \nabla V(x_n.v_n)\cdot\begin{bmatrix}\frac{u}{2}(h-\frac{1-e^{-2h}}{2})\mathbb{E}[\nabla f(x_{n+\frac12})-\nabla f(x_n)|\mathcal{F}_n]\\ u\frac{1-e^{-2h}}{2}\mathbb{E}[\nabla f(x_{n+\frac12})-\nabla f(x_n)|\mathcal{F}_n]\end{bmatrix}\\ &+\frac{3M}{2}\left[5(\frac{1-e^{-2h}}{2})^2|v_n|^2+u^2h^2|\nabla f(x_n)^2|+(\sigma_{n+1}^{(2)}{}^2+4\sigma_{n+1}^{(3)}{}^2)ud\right]\\ &+\frac{3M}{2}u^2h^2\mathbb{E}[|\nabla f(x_{n+\frac12})|^2-|\nabla f(x_n)|^2|\mathcal{F}_n]\end{aligned}$$

where we can further estimate

$$\begin{aligned}\mathbb{E}[\nabla f(x_{n+\frac12})-\nabla f(x_n)|\mathcal{F}_n] &\le M\mathbb{E}[x_{n+\frac12}-x_n|\mathcal{F}_n]\\ &\le M\frac{1}{2h}(h-\frac{1-e^{-2h}}{2})|v_n|+\sqrt{ud}M\sigma_{n+1}^{(1)}\\ &\quad+\frac{u}{2}(\frac{h}{2}-\frac{h-\frac{1-e^{-2h}}{2}}{2h})|\nabla f(x_n)|\end{aligned}$$

and there exists ξ_n such that $|\nabla f(x_{n+\frac12})|^2-|\nabla f(x_n)|^2 = 2(x_{n+\frac12}-x_n)^TD^2f(\xi_n)\nabla f(\xi_n)$ and ξ_n is on the line segment joining x_n and $x_{n+\frac12}$. Therefore $|\xi_n-x_n|\le|x_{n+\frac12}-x_n|$. then we have

$$\begin{aligned}\mathbb{E}[|\nabla f(x_{n+\frac12})|^2-|\nabla f(x_n)|^2|\mathcal{F}_n] &\le 2M\mathbb{E}[|\nabla f(\xi_n)||x_{n+\frac12}-x_n||\mathcal{F}_n]\\ &\le 2M|\nabla f(x_n)|\mathbb{E}[|x_{n+\frac12}-x_n||\mathcal{F}_n]+2M^2\mathbb{E}[|x_{n+\frac12}-x_n|^2|\mathcal{F}_n]\\ &\le |\nabla f(x_n)|^2+3M^2\mathbb{E}[|x_{n+\frac12}-x_n|^2|\mathcal{F}_n]\\ &\le |\nabla f(x_n)|^2+6M^2(\frac{h^2}{3}|v_n|^2+\frac{u^2h^4}{20}|\nabla f(x_n)|^2+ud\sigma_{n+1}^{(1)}{}^2)\end{aligned}$$

When h is small, we can use polynomials of h to bound those exponential coefficients. We can obtain that there exists $C>0$:

$$\mathbb{E}[V(x_{n+1},v_{n+1})|\mathcal{F}_n]\le V(x_n,v_n)+h\nabla V(x_n,v_n)\cdot b(x_n,v_n)^T+Ch^2(d+|v_n|^2+|\nabla f(x_n)|^2)$$

then assumption 7.1 imples that there exists $\alpha>0,\beta>0$ such that

$$\mathbb{E}[V(x_{n+1},v_{n+1})|\mathcal{F}_n]\le(1-\alpha h+Cc_Vh^2)V(x_n,v_n)+Ch^2d+\beta$$

When h is small, there exists $\hat\alpha = 1-\alpha h+Cc_Vh^2\in(0,1)$ and $\hat\beta = Ch^2d+\beta>0$ such that $\mathbb{E}[V(x_{n+1},v_{n+1})|\mathcal{F}_n]\le\hat\alpha V(x_n,v_n)+\hat\beta$. □

Once we have the Lyapunov condition, we can define the stopping time $\tau_C = \inf\{n>0 : (x_n,v_n)\in C\}$ and show that $\sup_{(x,v)\in C}\mathbb{E}_{(x,v)}[\tau_C]\le M_C<\infty$ for all small set C. Then uniqueness of stationary probability measure and ergodicity all follow by Theorem 1.3.1 in [33]. Next we prove that $\sup_{(x,v)\in C}\mathbb{E}_{(x,v)}[\tau_C]\le M_C<\infty$ given Lyapunov condition. To do so, note that we have

$$\begin{aligned}\mathbb{E}_{(x,v)}[\tau_C] &= \sum_{n=1}^{\infty}n\mathbb{P}(\tau_C=n)\\ &= \sum_{n\ge1}\mathbb{P}(\tau_C>n-1)\end{aligned}$$

Under Lyapunov condition, for any stopping time N, according to Lemma A.3 and Corollary A.4 in [31], we have

$$\begin{aligned}\mathbb{P}(\tau_C>n-1) &\le \mathbb{E}[V(x_n,v_n)1_{\tau_C>n-1}]\\ &\le \frac{\kappa[\gamma^{n-1}V(x_0,v_0)+1]}{1-\gamma}\\ &\le \kappa\gamma^{n-1}[V(x_0,v_0)+1]\end{aligned}$$

for some $\gamma \in (\hat{\alpha}, 1)$ and constant κ. Therefore, we have

$$\begin{aligned}\mathbb{E}_{(x,v)}[\tau_C] &\leq \sum_{k\geq 1} \kappa\gamma^{n-1}[V(x_0, v_0) + 1] \\ &= \frac{\kappa[V(x, v) + 1]}{1-\gamma}\end{aligned}$$

and

$$\sup_{(x,v)\in C} \mathbb{E}_{(x,v)}[\tau_C] \leq \frac{\kappa}{1-\gamma} \sup_{(x,v)\in C} V(x, v) + \frac{\kappa}{1-\gamma} \leq M_C < \infty$$

So as a conclusion, the statement of the theorem follows. □

Before proving Proposition 3.1, we require some preliminary estimtes from [40], that we present below. First, let (y_n, w_n) be the solution of Underdamped Langevin dynamics evaluated at $t = \sum_{k=1}^{n} \gamma_k$ with initial value (y_0, w_0). (x_n, v_n) is the nth iterates in the (RULMC) algorithm with initial value (x_0, v_0). $(x_n^*(t), v_n^*(t))$ is the solution of Underdamped Langevin dynamics with initial value (x_{n-1}, v_{n-1}) and $(x_n^*, v_n^*) = (x_{n-1}^*(\gamma_n), v_{n-1}^*(\gamma_n))$. Then, we have the following results from Lemma 2 in [40]. When $\gamma_{n+1} < \frac{1}{2}$ and $u = \frac{1}{M}$, we have:

$$\begin{aligned}\mathbb{E}\left\|\mathbb{E}_\alpha x_{n+1} - x_{n+1}^*\right\|^2 &\leq 45(\gamma_{n+1}^{10}\mathbb{E}\left\|v_n\right\|^2 + M^{-2}\gamma_{n+1}^{12}\mathbb{E}\left\|\nabla f(x_n)\right\|^2 + M^{-1}d\gamma_{n+1}^{11}) \\ \mathbb{E}\left\|x_{n+1} - x_{n+1}^*\right\|^2 &\leq 1800(\gamma_{n+1}^{6}\mathbb{E}\left\|v_n\right\|^2 + M^{-2}\gamma_{n+1}^{4}\mathbb{E}\left\|\nabla f(x_n)\right\|^2 + M^{-1}d\gamma_{n+1}^{7}) \\ \mathbb{E}\left\|\mathbb{E}_\alpha v_{n+1} - v_{n+1}^*\right\|^2 &\leq 45(\gamma_{n+1}^{8}\mathbb{E}\left\|v_n\right\|^2 + M^{-2}\gamma_{n+1}^{10}\mathbb{E}\left\|\nabla f(x_n)\right\|^2 + M^{-1}d\gamma_{n+1}^{9}) \\ \mathbb{E}\left\|v_{n+1} - v_{n+1}^*\right\|^2 &\leq 1300(\gamma_{n+1}^{4}\mathbb{E}\left\|v_n\right\|^2 + M^{-2}\gamma_{n+1}^{4}\mathbb{E}\left\|\nabla f(x_n)\right\|^2 + M^{-1}d\gamma_{n+1}^{5})\end{aligned}$$

Proof of Proposition 3.1 . Denote $A_n^2 = \mathbb{E}[\|x_n - y_n\|^2 + \|(x_n + v_n) - (y_n + w_n)\|^2]$. Using triangle inequality we have

$$\begin{aligned}\mathbb{E}_\alpha[\|x_n - y_n\|^2 + \|(x_n + v_n) - (y_n + w_n)\|^2] \leq (1 + \frac{h}{2\kappa})(\|x_k^* - y_n\|^2 + \|(x_k^* + v_k^*) - (y_n + w_n)\|^2) \\ + \frac{2\kappa}{h}(\|\mathbb{E}_\alpha[x_n] - x_k^*\|^2 + \|\mathbb{E}_\alpha[x_n + v_n] - (x_n^* + v_n^*)\|^2) \\ + \mathbb{E}_\alpha\left\|x_n - x_n^*\right\|^2 + \mathbb{E}_\alpha\left\|(x_n + v_n) - (x_n^* + v_n^*)\right\|^2\end{aligned}$$

Furthermore, we can take expectation on ω and use the contraction of Underdamped Langevin dynamics:

$$\begin{aligned}A_n^2 &\leq (1 + \frac{h}{2\kappa})e^{-\frac{h}{\kappa}}A_{n-1}^2 + \frac{2\kappa}{h}(\mathbb{E}\left\|\mathbb{E}_\alpha x_n - x_n^*\right\|^2 + \mathbb{E}\left\|\mathbb{E}_\alpha[x_n + v_n] - (x_n^* + v_n^*)\right\|^2) \\ &\quad + \mathbb{E}\left\|x_n^* - x_n\right\|^2 + \mathbb{E}\left\|(x_n + v_n) - (x_n^* + v_n^*)\right\|^2 \\ &\leq e^{-\frac{h}{2\kappa}}A_{n-1}^2 + \frac{2\kappa}{h}(3\mathbb{E}\left\|\mathbb{E}_\alpha x_n - x_n^*\right\|^2 + 2\mathbb{E}\left\|\mathbb{E}_\alpha v_n - v_n^*\right\|^2) \\ &\quad + 3\mathbb{E}\left\|x_n - x_n^*\right\|^2 + 2\mathbb{E}\left\|v_n - v_n^*\right\|^2\end{aligned}$$

When $h < \frac{1}{2}$, $u = \frac{1}{M}$ and $m = 1$:

$$A_n^2 \leq e^{-\frac{h}{2\kappa}}A_{n-1}^2 + 8250\left[(\kappa h^7 + h^4)\mathbb{E}\left\|v_{n-1}\right\|^2 + (\kappa^{-1}h^8 + \kappa^{-2}h^4)\mathbb{E}\left\|\nabla f(x_{n-1})\right\|^2 + (\kappa^{-1}h^5 + h^7)\right]$$

Our next step is to bound $\mathbb{E}\left\|v_{n-1}\right\|^2$ and $\mathbb{E}\left\|\nabla f(x_{n-1})\right\|^2$. First for Underdamped Langevin dynamics with f satisfying Assumption 1.1, it's easy to compute that:

$$\begin{aligned}\mathbb{E}\left\|w_{n-1}\right\|^2 &= d/M \\ \mathbb{E}\left\|\nabla f(y_{n-1})\right\|^2 &= \frac{1}{\int e^{-f(x)}dx}\int |\nabla f(x)|^2 e^{-f(x)}dx \\ &= -\frac{1}{\int e^{-f(x)}dx}\int (\nabla f(x))^T \nabla e^{-f(x)}dx \\ &= \frac{1}{\int e^{-f(x)}dx}\int \Delta f(x) e^{-f(x)}dx \\ &\leq \left\|\Delta f(x)\right\|_\infty \leq Md\end{aligned}$$

Therefore, we have

$$\mathbb{E}\left\|v_{n-1}\right\|^2 \leq 2d/M + 2\mathbb{E}\left\|v_{n-1} - w_{n-1}\right\|^2 \leq 2d/M + 4A_{n-1}^2$$
$$\mathbb{E}\left\|\nabla f(x_{n-1})\right\|^2 \leq 2Md + 2M^2\mathbb{E}\left\|x_{n-1} - y_{n-1}\right\|^2 \leq 2Md + 2M^2A_{n-1}^2$$

Plug the upper bounds into our previous result:

$$\begin{aligned}A_n^2 &\leq e^{-\frac{h}{2\kappa}}A_{n-1}^2 + 8250\left[(\kappa h^7 + h^4)(2d/M + 4A_{n-1}^2) + (\kappa^{-1}h^8 + \kappa^{-2}h^4)(2Md + 2M^2A_{n-1}^2) + (\kappa^{-1}h^5 + h^7)\right] \\ &\leq \left[1 - \frac{h}{2\kappa} + \frac{h^2}{8\kappa^2} + 49500(h^4 + \kappa h^7)\right]A_{n-1}^2 + 41250d(h^7 + \kappa^{-1}h^4)\end{aligned}$$

If we choose $(x_{n-1}, v_{n-1}) \sim \pi_h^*(x, v)$ and $(y_{n-1}, w_{n-1}) \sim \pi^*(x, v)$ such that

$$A_{n-1}^2 = \min_{X\sim\pi_h^*,\, Y\sim\pi^*} \mathbb{E}\left\|X - Y\right\|^2$$

Then we have

$$W_2(\pi, \pi_h)^2 \leq A_{n-1}^2 \leq \frac{82500h^3(\kappa h^3 + 1)d}{1 - \frac{h}{4\kappa} - 99000h^3\kappa(1 + \kappa h^3)}$$

We can see that $W_2(\pi, \pi_h) \to 0$ as $h \to 0$. Furthermore, as $h \to 0$, $W_2(\pi, \pi_h) < O(h^{\frac{3}{2}})$. □

7.2 Proofs for Section 3.2

Proof of Theorem 3.2. Define $A_n^2 = \mathbb{E}[\|x_n - y_n\|^2 + \|(x_n + v_n) - (y_n + w_n)\|^2]$. From the proof of proposition 3.1, we know that

$$A_n^2 \leq \left[1 - \frac{\gamma_n}{2\kappa} + \frac{\gamma_n^2}{8\kappa^2} + 49500(\gamma_n^4 + \kappa\gamma_n^7)\right]A_{n-1}^2 + 41250d(\gamma_n^7 + \kappa^{-1}\gamma_n^4)$$

When time step h is a constant, apply the inequality repeatedly to get

$$A_n^2 \leq \left[1 - \frac{h}{2\kappa} + \frac{h^2}{8\kappa^2} + 49500(h^4 + \kappa h^7)\right]^k A_0^2 + \frac{82500h^3(\kappa h^3 + 1)d}{1 - \frac{h}{4\kappa} - 99000h^3\kappa(1 + \kappa h^3)}$$

Denote ν_n to be the density function of x_n, then $W_2(\nu_n, \pi) \leq A_n$. By choosing $\gamma_n = h \sim O(\epsilon^{\frac{2}{3}})$, we can guarantee that $W_2(\nu_n, \pi) < \epsilon\sqrt{\frac{d}{m}}$ for all $n > K \sim \tilde{O}(\epsilon^{-\frac{2}{3}})$.
When the time step γ_n is variant, the inequality we correspondingly have

$$A_n^2 \leq \left[1 - \frac{\gamma_n}{2\kappa} + \frac{\gamma_n^2}{8\kappa^2} + 49500(\gamma_n^4 + \kappa\gamma_n^7)\right]A_{n-1}^2 + 41250d(\gamma_n^7 + \kappa^{-1}\gamma_n^4)$$

When $\gamma_n < 1$, $\frac{\gamma_n^2}{8\kappa^2} < \frac{\gamma_n}{8\kappa}$. When $\gamma_n < (\frac{99000}{8\kappa^2}) < 24\kappa^{-\frac{2}{3}}$, we have $49500(\gamma_n^4 + \kappa\gamma_n^7) < \frac{\gamma_n}{8\kappa}$. Similarly, when $\gamma_n < 1$, we have $41250d(\gamma_n^7 + \kappa^{-1}\gamma_n^4) < 82500d\gamma_n^4$. Therefore, when $\gamma_n < \min\{1/2, 24\kappa^{-\frac{2}{3}}\}$, we have

$$A_n^2 < (1 - \frac{\gamma_n}{4\kappa})A_{n-1}^2 + 82500d\gamma_n^4$$

If we choose $\gamma_n = \frac{16\kappa}{32\kappa^{\frac{5}{3}} + (n-K_1)^+}$, where K_1 is the smallest integer such that

$$A_{K_1}^2 < (1 - \frac{4}{\kappa^{\frac{5}{3}}})^{K_1}A_0^2 + (82500)d\frac{1}{2\kappa} < 2\frac{82500d}{\kappa}$$

Then we claim that for all $n \geq K_1$, we have

$$A_n^2 < \frac{82500(16)^4 d\kappa^4}{(32\kappa^{\frac{5}{3}} + n - K_1)^3}$$

The claim can be proved by induction: Assume that the claim hold for A_n^2 and denote $b = 32\kappa^{\frac{5}{3}} + n - K_1$, then

$$\begin{aligned} A_{n+1}^2 &< (1 - \frac{4}{1+b})\frac{82500(16)^4 d\kappa^4}{b^3} + \frac{82500d(16)^4\kappa^4}{(b+1)^4} \\ &= \frac{82500(16)^4 d\kappa^4}{(b+1)^3}\left[\frac{(b-3)(b+1)^2}{b^3} + \frac{1}{b+1}\right] \\ &< \frac{82500(16)^4 d\kappa^4}{(b+1)^3} \\ &= \frac{82500(16)^4 d\kappa^4}{(32\kappa^{\frac{5}{3}} + n + 1 - K_1)^3} \end{aligned}$$

Therefore, under our choice of time step (γ_n), we can guarantee $W_2(\nu_n, \pi) < \epsilon\sqrt{\frac{d}{m}}$ for all $n > K \sim O(\epsilon^{-\frac{2}{3}})$. Compared to the running time of constant step size RULMC, vanishing step size help reduce the factor $\log(\frac{1}{\epsilon})$ in the guarantees. □

Now we introduce the CLT statement for another sampling algorithm related to (RULMC) and give a complete proof of the statement. The proof of Remark 6 can be done in the same way. In the following theorem, we give a central limit result with specific choice of weights and time step-size. The Euler-discretization of the underdamped Langevin diffusion (which we call as KLMC, following [7]) is given by the following algorithm:

$$\begin{aligned} x_{n+1} &= x_n + \frac{1 - e^{-2\gamma_{n+1}}}{2} v_n - \frac{u}{2}(\gamma_{n+1} - \frac{1 - e^{-2\gamma_{n+1}}}{2})\nabla f(x_n) + \sqrt{u}\sigma_{n+1}^{(1)} U_{n+1}^{(1)} \\ v_{n+1} &= v_n e^{-2\gamma_{n+1}} - u\frac{1 - e^{-2\gamma_{n+1}}}{2}\nabla f(x_n) + 2\sqrt{u}\sigma_{n+1}^{(2)} U_{n+1}^{(2)} \end{aligned} \tag{KLMC}$$

where $\{\gamma_n\}$ are the time steps. $\sigma_n^{(1)}$ and $\sigma_n^{(2)}$ are positive with ${\sigma_n^{(1)}}^2 = \gamma_n + \frac{1-e^{-4\gamma_n}}{4} - (1 - e^{-2\gamma_n})$, ${\sigma_n^{(2)}}^2 = \frac{1-e^{-4\gamma_n}}{4}$. $\{(U_n^{(1)}, U_n^{(2)})\}_n$ are independent Centered Gaussian random vectors in $\mathbb{R}^{2d}$ with $(U_n^{(1)}, U_n^{(2)}) \sim \mathcal{N}(0, \sigma_n^2 I_d)$ and $\sigma_n^2 = \frac{1+e^{-4\gamma_n} - 2e^{-2\gamma_n}}{4\sigma_n^{(1)}\sigma_n^{(2)}}$. Numerical integration with the above sampler follows the same steps as described in Section 3.2. We now provide the following CLT.

Theorem 5. *Assume potential function f satisfies Assumption 1.1. Let $\{(x_k, v_k)\}$ and $\{(U_k^{(1)}, U_k^{(2)})\}$ be the same as what we have in the (KLMC) algorithm and the time step-size $\{\gamma_k\}$ is non-increasing and $\lim_k(\gamma_{k-1} - \gamma_k)/\gamma_k^4 = 0$. If $\lim_n(1/\sqrt{\Gamma_n^{(3)}})\sum_{k=1}^n \gamma_k^4 = \hat{\gamma} \in (0, +\infty]$ and $\lim_n \Gamma_n^{(4)} = +\infty$, then for all $\phi \in \mathcal{C}^3$ with $D^2\phi$, $D^3\phi$ and $D^4\phi$ bounded and Lipschitz and $\sup_{(x,v)\in\mathbb{R}^{2d}} |\nabla\phi(x)|^2/V(x,v) < +\infty$, we have*

$$\begin{aligned} &\frac{\Gamma_n}{\Gamma_n^{(4)}}\nu_n^\gamma(\mathcal{L}\phi) \to \mathcal{N}(\rho, \frac{10}{3}u\hat{\gamma}^{-2}\int_{\mathbb{R}^d} |\nabla\phi(x)|\pi(dx)) \qquad \text{if } \hat{\gamma} < +\infty \\ &\frac{\Gamma_n}{\Gamma_n^{(4)}}\nu_n^\gamma(\mathcal{L}\phi) \to \rho \qquad\qquad\qquad\qquad\qquad\qquad\quad \text{if } \hat{\gamma} = +\infty, \end{aligned}$$

where

$$\begin{aligned} \rho = &\tfrac{u}{6}\int\int\langle D^3\phi(x); \nabla f(x), v^{\otimes 2}\rangle\nu(dx,dv) + \tfrac{u}{24}\int\int\langle D^3 f(x); \nabla\phi(x), v^{\otimes 2}\rangle\nu(dx,dv) \\ &+ \tfrac{u}{12}\int\int(D^2\phi D^2 f)(x)v^{\otimes 2}\nu(dx,dv) - \tfrac{1}{12}\int\int D^4\phi(x)v^{\otimes 4}\nu(dx,dv) \\ &- \tfrac{u^2}{24}\int\langle D^2 f(x); \nabla\phi(x), \nabla f(x)\rangle\pi(dx). \end{aligned}$$

In the following context we'll discuss the weak convergence of empirical measure ν_n^η and build a central limit theorem under certain assumptions.

1) (Lyapunov Conditions) The underdamped Langevin dynamics can be rewritten as

$$dY_t = b(Y_t)dt + \sigma(Y_t)dW_t$$

where $Y_t = [X_t, V_t]^T$, $b(y) = b(x,v) = [v, -2v - u\nabla f(x)]^T$, $\sigma(y) = 2\sqrt{u}[0_d, I_d]^T$ for all $x, v \in \mathbb{R}^d$. $\{W_t\}$ is a 2d-dimensional Brownian motion.
The Lyapunov condition is similar to the one that's introduced in[21].

Assumption $(\mathcal{L}_{V,\infty})$: There's a $\mathcal{C}^2$ function $V : \mathbb{R}^{2d} \to [v_*, +\infty)$ for some $v_* > 0$ satisfying the following conditions:

a) $\left\|D^2V\right\|_\infty = \sup_{(x,v)^T \in \mathbb{R}^{2d}} \left\|D^2V(x,v)\right\|_{op} < +\infty$ and $\lim_{|(x,v)| \to +\infty} V(x,v) = +\infty$;
b) $|\nabla V(x,v)|^2 + |b(x,v)|^2 \le c_V V(x,v)$ for all $(x,v)^T \in \mathbb{R}^{2d}$ and some $c_V > 0$;
c) $\langle \nabla V(x,v), b(x,v) \rangle \le -\alpha V(x,v) + \beta$ for some $\alpha > 0$ and $\beta \in \mathbb{R}$.

Assumption $(\mathcal{L}_{V,p})$: There's a $\mathcal{C}^2$ function $V : \mathbb{R}^{2d} \to [v_*, +\infty)$ for some $v_* > 0$ satisfying for some $p \ge 1$:

a) $\left\|D^2V\right\|_\infty = \sup_{(x,v)^T \in \mathbb{R}^{2d}} \left\|D^2V(x,v)\right\|_{op} < +\infty$ and $\lim_{|(x,v)| \to +\infty} V(x,v) = +\infty$;
b) $|\nabla V(x,v)|^2 + |b(x,v)|^2 + \mathrm{Tr}(\sigma(x,v)\sigma(x,v)^T) \le c_V V(x,v)$ for all $(x,v)^T \in \mathbb{R}^{2d}$ and some $c_V > 0$;
c) $\langle \nabla V(x,v), b(x,v) \rangle + \lambda_p \mathrm{Tr}(\sigma(x,v)\sigma(x,v)^T) \le -\alpha V(x,v) + \beta$ for some $\alpha > 0$ and $\beta \in \mathbb{R}$, where $\lambda_p = \frac{1}{2}\lambda_{D^2V + (p-1)(\nabla V \otimes \nabla V)/V}$.

Remark 10. *1) We can show that:* $(\mathcal{L}_{V,p'}) \implies (\mathcal{L}_{V,p})$ *if* $p' \ge p \ge 1$. *Especially* $(\mathcal{L}_{V,\infty}) \implies (\mathcal{L}_{V,p})$ *for all* $p \ge 1$.

2) If we choose b and σ the same as those in the Underdamped Langevin dynamics, then $(\mathcal{L}_{V,\infty})$ is almost the same as assumption 7.1. We can instantly obtain that assumption 7.1 implies $(\mathcal{L}_{V,\infty})$. Therefore, according to lemma 4, assumption 1.1 implies $(\mathcal{L}_{V,\infty})$.

2) (Tightness Result) We now establish the almost sure tightness of the weighted empirical measures. The filtration $\{\mathcal{F}_n\}$ we consider is $\mathcal{F}_n = \sigma(Y_0, (U_1^{(1)}, U_1^{(2)}), \cdots, (U_n^{(1)}, U_n^{(2)}))$.

Lemma 5. *(a) If $(\mathcal{L}_{V,1})$ holds, then for every $a \ge \frac{1}{2}$,*

$$|V^a(Y_{n+1}) - V^a(Y_n)| \le c_a \sqrt{\gamma_{n+1}} V^a(Y_n)(1 + |U_{n+1}^{(1)}|^{2a} + |U_{n+1}^{(2)}|^{2a})$$

(b) If $(\mathcal{L}_{V,p})$ holds for some $p \ge 1$, then there exists real numbers $\tilde{\alpha} > 0$ and $\tilde{\beta}$ and $n_0 \in \mathbb{N}$ such that

$$\mathbb{E}[V^p(Y_{n+1})|\mathcal{F}_n] \le V^(Y_n) + \gamma_{n+1} V^{p-1}(Y_n)(\tilde{\beta} - \tilde{\alpha} V(Y_n)), \quad \forall\, n \ge n_0$$

and furthermore

$$\sup_{n \in \mathbb{N}} \mathbb{E}[V^p(Y_n)] < +\infty$$

Proof of Lemma 5. (a) Using mean value theorem and $(\mathcal{L}_{V,1})$:

$$\begin{aligned} |V^a(Y_{n+1}) - V^a(Y_n)| &= a|V^{a-1}(\xi_{n+1})\langle \nabla V(\xi_{n+1}), Y_{n+1} - Y_n \rangle| \\ &\le C V^{a-\frac{1}{2}}(\xi_{n+1})|Y_{n+1} - Y_n| \end{aligned}$$

From $(\mathcal{L}_{V,1})$-b) we get that $\nabla\sqrt{V}$ is bounded, i.e $\sqrt{V}$ is Lipschitz with parameter $[\sqrt{V}]_1$. Hence

$$\begin{aligned} V^{a-\frac{1}{2}}(\xi_{n+1}) &\le (\sqrt{V}(Y_n) + [\sqrt{V}]_1 |Y_{n+1} - Y_n|)^{2a-1} \\ &\le 2^{2a-1}\left(V^{a-\frac{1}{2}}(Y_n) + [\sqrt{V}]_1^{2a-1}|Y_{n+1} - Y_n|^{2a-1}\right) \end{aligned}$$

Meanwhile,

$$\begin{aligned} |Y_{n+1} - Y_n|^2 &= \left| \begin{bmatrix} \frac{1-e^{-2\gamma_{n+1}}}{2} v_n - \frac{u}{2}(\gamma_{n+1} - \frac{1-e^{-2\gamma_{n+1}}}{2})\nabla f(x_n) + \sqrt{u}\sigma_{n+1}^{(1)} U_{n+1}^{(1)} \\ -2\frac{1-e^{-2\gamma_{n+1}}}{2} v_n - u\frac{1-e^{-2\gamma_{n+1}}}{2}\nabla f(x_n) + 2\sqrt{u}\sigma_{n+1}^{(2)} U_{n+1}^{(2)} \end{bmatrix} \right|^2 \\ &\le 15(\frac{1-e^{-2\gamma_{n+1}}}{2})^2 |v_n^2| + [\frac{3u^2}{4}(\gamma_{n+1} - \frac{1-e^{-2\gamma_{n+1}}}{2})^2 + 3u^2(\frac{1-e^{-2\gamma_{n+1}}}{2})^2]|\nabla f(x_n)|^2 \\ &\quad + 3u{\sigma_{n+1}^{(1)}}^2 |U_{n+1}^{(1)}|^2 + 12u{\sigma_{n+1}^{(2)}}^2 |U_{n+1}^{(2)}|^2 \end{aligned}$$

Since $\gamma_n \to 0$ as $\nu \to \infty$ and $\frac{1-e^{-2\gamma_n}}{2} \sim O(\gamma_n)$, $\gamma_n - \frac{1-e^{-2\gamma_n}}{2} \sim O(\gamma_n^2)$, $\sigma_n^{(1)} \sim O(\gamma_n^{\frac{3}{2}})$ and $\sigma_n^{(2)} \sim O(\gamma_n^{\frac{1}{2}})$, there exist $C_1, C_2, C_3 > 0$ such that

$$\begin{aligned}
|Y_{n+1} - Y_n|^2 &\le C_1 \left[\gamma_{n+1}^2(|v_n|^2 + |\nabla f(x_n)|^2) + \gamma_{n+1}(|U_{n+1}^{(1)}|^2 + |U_{n+1}^{(2)}|^2)\right] \\
&\le C_2 \left[\gamma_{n+1}^2 V(Y_n) + \gamma_{n+1}(|U_{n+1}^{(1)}|^2 + |U_{n+1}^{(2)}|^2 + 1)\right] \\
\implies \quad |Y_{n+1} - Y_n| &\le C_3\sqrt{\gamma_{n+1}}\sqrt{V(Y_n)}(|U_{n+1}^{(1)}| + |U_{n+1}^{(2)}| + 1)
\end{aligned}$$

Combining our estimations, since $a \ge 1/2$. we get

$$\begin{aligned}
|V^a(Y_{n+1}) - V^a(Y_n)| &\le C2^{2a-1}\left(V^{a-\frac{1}{2}}(Y_n) + [\sqrt{V}]_1^{2a-1}|Y_{n+1} - Y_n|^{2a-1}\right)|Y_{n+1} - Y_n| \\
&\le c_a'\left(\sqrt{\gamma_{n+1}}V^a(Y_n)(|U_{n+1}^{(1)}| + |U_{n+1}^{(2)}| + 1) + \gamma_{n+1}^a V^a(Y_n)(|U_{n+1}^{(1)}| + |U_{n+1}^{(2)}| + 1)^{2a}\right) \\
&\le c_a\sqrt{\gamma_{n+1}}V^a(Y_n)(|U_{n+1}^{(1)}|^{2a} + |U_{n+1}^{(2)}|^{2a} + 1)
\end{aligned}$$

(b) We Taylor expand $V^p(Y_{n+1})$ at Y_n:

$$V^p(Y_{n+1}) = V^p(Y_n) + pV^{p-1}(Y_n)\langle\nabla V(Y_n), Y_{n+1} - Y_n\rangle + \frac{1}{2}D^2(V^p)(\xi_{n+1})(Y_{n+1} - Y_n)^{\otimes 2}$$

Since $D^2(V^p) = pV^{p-1}D^2V + p(p-1)V^{p-1}\nabla V\nabla V^T$, by the definition of λ_p:

$$D^2(V^p)(\xi_{n+1})(Y_{n+1} - Y_n)^{\otimes 2} \le 2p\lambda_p V^{p-1}(\xi_{n+1})|Y_{n+1} - Y_n|^2$$

Therefore

$$V^p(Y_{n+1}) \le V^p(Y_n) + pV^{p-1}(Y_n)\langle\nabla V(Y_n), Y_{n+1} - Y_n\rangle + p\lambda_p V^{p-1}(\xi_{n+1})|Y_{n+1} - Y_n|$$

When $p = 1$, take conditional expectation on $\mathcal{F}_n$:

$$\begin{aligned}
\mathbb{E}[V(Y_{n+1})|\mathcal{F}_n] \le V(Y_n) &+ \frac{1 - e^{-2\gamma_{n+1}}}{2}\langle\nabla V(x_n, v_n), b(x_n, v_n)\rangle \\
&- \frac{u}{2}(\gamma_{n+1} - \frac{1 - e^{-2\gamma_{n+1}}}{2})\nabla_x V(x_n, v_n)\cdot\nabla f(x_n) \\
&+ \lambda_1(\frac{1 - e^{-2\gamma_{n+1}}}{2})^2[5|v_n|^2 + u^2|\nabla f(x_n)|^2 + 4u\nabla f(x_n)\cdot v_n] \\
&- \frac{u}{2}\frac{1 - e^{-2\gamma_{n+1}}}{2}(\gamma_{n+1} - \frac{1 - e^{-2\gamma_{n+1}}}{2})\nabla f(x_n)\cdot v_n \\
&+ \frac{u^2}{4}(\gamma_{n+1} - \frac{1 - e^{-2\gamma_{n+1}}}{2})^2|\nabla f(x_n)|^2 + u({\sigma_{n+1}^{(1)}}^2 + 4{\sigma_{n+1}^{(2)}}^2)d
\end{aligned}$$

There exists $n_0 \in \mathbb{N}$ such that for all $n \ge n_0$

$\frac{1 - e^{-2\gamma_{n+1}}}{2}\langle\nabla V(x_n, v_n), b(x_n, v_n)\rangle \le \gamma_{n+1}(-\alpha V(Y_n) + \beta)$, for some $\alpha > 0, \beta \in \mathbb{R}$;

$-\frac{u}{2}(\gamma_{n+1} - \frac{1 - e^{-2\gamma_{n+1}}}{2})\nabla_x V(x_n, v_n)\cdot\nabla f(x_n) \le C\gamma_{n+1}^2(|\nabla V(Y_n)|^2 + |b(Y_n)|^2) \le C\gamma_{n+1}^2 V(Y_n)$;

$\lambda_1(\frac{1 - e^{-2\gamma_{n+1}}}{2})^2[5|v_n|^2 + u^2|\nabla f(x_n)|^2 + 4u\nabla f(x_n)\cdot v_n] \le C\gamma_{n+1}^2|b(Y_n)|^2 \le C\gamma_{n+1}^2 V(Y_n)$;

$-\frac{u}{2}\frac{1 - e^{-2\gamma_{n+1}}}{2}(\gamma_{n+1} - \frac{1 - e^{-2\gamma_{n+1}}}{2})\nabla f(x_n)\cdot v_n \le C\gamma_{n+1}^3|b(Y_n)|^2 \le C\gamma_{n+1}^3 V(Y_n)$;

$\frac{u^2}{4}(\gamma_{n+1} - \frac{1 - e^{-2\gamma_{n+1}}}{2})^2|\nabla f(x_n)|^2 \le C\gamma_{n+1}^4|b(Y_n)|^2 \le C\gamma_{n+1}^4 V(Y_n)$;

$u({\sigma_{n+1}^{(1)}}^2 + 4{\sigma_{n+1}^{(2)}}^2)d \le C\gamma_{n+1}$.

Therefore, for all $n \ge n_0$, there exist $\tilde{\alpha} > 0,\ \tilde{\beta} \in \mathbb{R}$ such that

$$\mathbb{E}[V(Y_{n+1})|\mathcal{F}_n] \le V(Y_n)(1 - \alpha\gamma_{n+1} + C(2\gamma_{n+1}^2 + \gamma_{n+1}^3 + \gamma_{n+1}^4)) + \gamma_{n+1}(\beta + C)$$

$$\leq V(Y_n)(1-\tilde{\alpha}\gamma_{n+1})+\tilde{\beta}\gamma_{n+1}$$

and $1-\tilde{\alpha}\gamma_{n+1}>0$. This leads to

$$\mathbb{E}[V(Y_{n+1})]\leq \mathbb{E}[V(Y_n)](1-\tilde{\alpha}\gamma_{n+1})+\tilde{\beta}\gamma_{n+1}$$

We could use induction to prove:

$$\sup_{n\geq n_0}\mathbb{E}[V(Y_n)]\leq \frac{\tilde{\beta}}{\tilde{\alpha}}\vee \mathbb{E}[V(Y_{n_0})]$$

Assume now $p>1$. Due to $(\mathcal{L}_{V,p})$-b), we derive that $\sqrt{V}$ is Lipschitz with parameter $[\sqrt{V}]_1$. Consequently,

$$\begin{aligned}V^{p-1}(\xi_{n+1})=\sqrt{V}^{2(p-1)}(\xi_{n+1})&\leq \left(\sqrt{V}(Y_n)+[\sqrt{V}]_1|Y_{n+1}-Y_n|\right)^{2(p-1)}\\&\leq\begin{cases}V^{p-1}(Y_n)+([\sqrt{V}]_1|Y_{n+1}-Y_n|)^{2(p-1)} & \text{if } 2(p-1)\leq 1,\\ V^{p-1}(Y_n)+C\left(V^{(2p-3)/2}(Y_n)|Y_{n+1}-Y_n|+|Y_{n+1}-Y_n|^{2(p-1)}\right) & \text{if } 2(p-1)>1.\end{cases}\end{aligned}$$

Using the fact we've proved in part a):

$$|Y_{n+1}-Y_n|^2\leq C_2\left[\gamma_{n+1}^2V(Y_n)+\gamma_{n+1}(|U_{n+1}^{(1)}|^2+|U_{n+1}^{(2)}|^2+1)\right]$$

We derive

$$V^{p-1}(\xi_{n+1})|Y_{n+1}-Y_n|^2\leq V^{p-1}(Y_n)|Y_{n+1}-Y_n|^2+C\gamma_{n+1}^{p\wedge\frac{3}{2}}V^p(Y_n)(1+|U_{n+1}^{(1)}|^{2p}+|U_{n+1}^{(2)}|^{2p})$$

Then we take conditional expectation:

$$\begin{aligned}\mathbb{E}[V^p(Y_{n+1})|\mathcal{F}_n]\leq V^p(Y_n)&+pV^{p-1}\frac{1-e^{-2\gamma_{n+1}}}{2}\langle\nabla V(x_n,v_n),b(x_n,v_n)\rangle\\&-pV^{p-1}(Y_n)\frac{u}{2}(\gamma_{n+1}-\frac{1-e^{-2\gamma_{n+1}}}{2})\nabla_xV(x_n,v_n)\cdot\nabla f(x_n)\\&+CV^{p-1}(Y_n)\left[\gamma_{n+1}^2V(Y_n)+\gamma_{n+1}(|U_{n+1}^{(1)}|^2+|U_{n+1}^{(2)}|^2+1)\right]\\&+C\gamma_{n+1}^{p\wedge\frac{3}{2}}V^p(Y_n)(1+|U_{n+1}^{(1)}|^{2p}+|U_{n+1}^{(2)}|^{2p})\end{aligned}$$

There exists $n_0\in\mathbb{N}$ such that for all $n\geq n_0$

$$\frac{1-e^{-2\gamma_{n+1}}}{2}\langle\nabla V(x_n,v_n),b(x_n,v_n)\rangle\leq\gamma_{n+1}(-\alpha V(Y_n)+\beta),\quad \text{for some } \alpha>0,\beta\in\mathbb{R};$$

$$-\frac{u}{2}(\gamma_{n+1}-\frac{1-e^{-2\gamma_{n+1}}}{2})\nabla_xV(x_n,v_n)\cdot\nabla f(x_n)\leq C\gamma_{n+1}^2(|\nabla V(Y_n)|^2+|b(Y_n)|^2)\leq C\gamma_{n+1}^2V(Y_n).$$

Since $\gamma_n^{p\wedge\frac{3}{2}},\gamma_n^2\sim o(\gamma_n)$, there exists $\tilde{\alpha}>0,\tilde{\beta}\in\mathbb{R}$, such that for all $n\geq n_0$:

$$\mathbb{E}[V^p(Y_n)|\mathcal{F}_n]\leq V^p(Y_n)+\gamma_{n+1}V^{p-1}(Y_n)(\tilde{\beta}-\tilde{\alpha}V(Y_n))$$

Same as the proof for $p=1$, we can show

$$\sup_{n\in\mathbb{N}}\mathbb{E}[V^p(Y_n)]<+\infty$$

□

Theorem 6. *Let $p\in[0,+\infty)$. Assume $(\mathcal{L}_{V,p})$, If there exists $s\in(0,1]$ such that*

$$\sum_{n\geq1}\frac{1}{H_n}(\Delta\frac{\eta_n}{\gamma_n})_+<+\infty\ \textit{and}\ \sum_{n\geq1}(\frac{\eta_n}{H_n\sqrt{\gamma_n}})^{1+s}<+\infty$$

then

$$\mathbb{P}(d\omega)-a.s\quad\sup_{n\in\mathbb{N}}\nu_n^\eta(\omega,V^{p/(1+s)})<+\infty$$

Based on Lemma 5, the proof of Theorem 6 immediately follows, by using the same steps in the proof of Theorem 4 in [21]. Hence we don't replicate the proof here.

3) (Identification of the limit)

Theorem 7 (Echeverrría-Weiss Theorem)**.** *Let E be a locally compact Polish space and $\mathcal{L}$ a linear operator satisfying the positive maximum principle. Assume that its domain $\mathcal{D}(A)$ is an algebra everywhere dense in $(\mathcal{C}_0(E), \|\ \|_\infty)$ containing a sequence $(f_n)_{n\in\mathbb{N}}$ satisfying*

$$\sup_{n\in\mathbb{N}} (\|f_n\|_\infty + \|\mathcal{L}f_n\|_\infty) < +\infty,\ \ \forall x \in E,\ \ f_n(x) \to 1 \ \text{ and } \ \mathcal{L}f_n(x) \to 0.$$

If a distribution on $(E, \mathcal{B}(E))$ satisfies $\int_E \mathcal{L}f d\nu = 0$ for every $f \in \mathcal{D}(A)$, then there exists a stationary solution for the martingale problem $(\mathcal{L}, \nu)$ (this means that there exists a stationary continuous-time homogeneous Markov process with infinitesimal generator $\mathcal{L}$ and invariant distribution ν).

Lemma 6. *If the potential function f is Gradient Lipschitz and strongly convex, then the generator of kinetic, $\mathcal{L}$, satisfies the assumptions of the Echeverrría-Weiss theorem.*

Proof of lemma 6. First it's well-known that the infinitesimal generator of a Fellerian semigroup satisfies the maximum principle. We can choose our $f_n(y) = \phi(y/n)$ for any $y \in \mathbb{R}^{2d}$ where ϕ is $\mathcal{C}^2$ with compact support and $\phi(0) = 1$. It's easy to check that $\forall y \in \mathbb{R}^{2d}$, $f_n(y) \to 0$ and $\mathcal{L}f_n(y) \to 0$. It's also straightforward that $\sup_{n\in\mathbb{N}} \|f_n\|_\infty < +\infty$. The last thing to check is $\sup_{n\in\mathbb{N}} \|\mathcal{L}f_n\|_\infty < +\infty$. Since $\mathcal{L}$ can also be written as $b(x, v) \cdot [\nabla_x,\ \nabla_v]^T + 2u\Delta_v$ and we've shown that under our assumptions on f, $(\mathcal{L}_{V,\infty})$ is satisfied, we have the Lyapunov function $V(y) = O(|y|^2)$ and $|b(x, v)| \le C(1 + |(x, v)|)$. Therefore we get $\sup_{n\in\mathbb{N}} \|\mathcal{L}f_n\|_\infty < +\infty$. □

Theorem 8. *Assume that f is gradient Lipschitz and strongly convex. Assume also*

$$\lim_n \frac{1}{H_n} \sum_{k=1}^n |\Delta \frac{\eta_n}{\gamma_n}| = 0 \ \text{ and } \ \sum_{n\ge1} (\frac{\eta_n}{\sqrt{\gamma_n} H_n})^2 < +\infty$$

Let $a \ge \frac{1}{2}$. Assume $\sup_n \nu_n^\eta(V^a) < +\infty$ $\mathbb{P} - a.s.$ If $a < 1$, assume also that $\sum_{n\ge1} \eta_n\gamma_n/H_n < +\infty$. Then $\mathbb{P}-a.s$, every limiting distribution $\nu_\infty(\omega, dx)$ of the sequence $(\nu_n^\eta(\omega, dx))$ is an invariant distribution of the underdamped Langevin dynamics introduced in the previous section.

The proof of theorem 8 follows immediately from Theorem 7, lemma 6, lemma 7 and lemma 8.

Lemma 7. *Under the assumptions in theorem 8, then for every bounded Lipschitz continuous function $g : \mathbb{R}^{2d} \to \mathbb{R}$,*

$$\mathbb{P} - a.s \ \ \lim_n \frac{1}{H_n} \sum_{k=1}^n \frac{\eta_k}{\gamma_k} \mathbb{E}[g(Y_k) - g(Y_{k-1})|\mathcal{F}_{k-1}] = 0$$

Proof of lemma 7. Setting $\eta_0/\gamma_0 = 0$ gives

$$\frac{1}{H_n} \sum_{k=1}^n \mathbb{E}[g(Y_k) - g(Y_{k-1})|\mathcal{F}_{k-1}] = \frac{1}{H_n} \sum_{k=1}^n \frac{\eta_k}{\gamma_k} (g(Y_k) - g(Y_{k-1})) - \frac{1}{H_n} \sum_{k=1}^n \frac{\eta_k}{\gamma_k} \left(g(Y_k) - \mathbb{E}[g(Y_k)|\mathcal{F}_{k-1}]\right).$$

As g is bounded, it follows by lemma 3-b) in[21] that

$$\mathbb{P} - a.s \ \ \lim_n \frac{1}{H_n} \sum_{k=1}^n (g(Y_k) - g(Y_{k-1})) = 0.$$

Then

$$\frac{1}{H_n} \sum_{k=1}^n \frac{\eta_k}{\gamma_k} \left(g(Y_k) - \mathbb{E}[g(Y_k)|\mathcal{F}_{k-1}]\right)$$

will converge to 0 once the martingale

$$M_n^g := \sum_{k=1}^n \frac{\eta_k}{\gamma_k H_k} \left(g(Y_k) - \mathbb{E}[g(Y_k)|\mathcal{F}_{k-1}]\right)$$

converge a.s in $\mathbb{R}$.

$$\mathbb{E}\langle M_n^g\rangle_\infty = \sum_{n\geq 1}(\frac{\eta_n}{\gamma_n H_n})^2 \left\|g(Y_n) - \mathbb{E}[g(Y_n)|\mathcal{F}_{n-1}]\right\|_2^2 \leq \sum_{n\geq 1}(\frac{\eta_n}{\gamma_n H_n})^2 \left\|g(Y_n) - g(Y_{n-1})\right\|_2^2$$
$$\leq [f]_1^2 \sum_{n\geq 1}(\frac{\eta_n}{\gamma_n H_n})^2 \left\|Y_n - Y_{n-1}\right\|_2^2$$

Since $(\mathcal{L}_{V,1})$ holds under our assumptions on f and by lemma 2-b)

$$\|Y_n - Y_{n-1}\|_2^2 \leq C'\mathbb{E}[\gamma_n^2 V(Y_{n-1}) + (2d+1)\gamma_n] \leq C\gamma_n$$

Therefore

$$\mathbb{E}\langle M_n^g\rangle_\infty \leq C\sum_{n\geq 1}(\frac{\eta_n}{\sqrt{\gamma_n}H_n})^2 < +\infty$$

□

Lemma 8. *Under the assumptions in theorem 8, then for every $g \in \mathcal{C}^2(\mathbb{R}^{2d})$ with compact support,*

$$\lim_n \left(\frac{1}{H_n}\sum_{k=1}^n \frac{\eta_k}{\gamma_k}\mathbb{E}[g(Y_k) - g(Y_{k-1})|\mathcal{F}_{k-1}] - \nu_n^\eta(\mathcal{L}g)\right) = 0 \ \ a.s$$

Proof of lemma 8. Setting $R_2(y_1,y_2) := g(y_2) - g(y_1) - \langle\nabla g(y_1), y_2 - y_1\rangle - \frac{1}{2}D^2 g(y_1)(y_2-y_1)^{\otimes 2}$, we obtain for every $k \in \mathbb{N}$,

$$\begin{aligned}
g(Y_k) - g(Y_{k-1}) &= \langle\nabla g(Y_{k-1}), Y_k - Y_{k-1}\rangle + \frac{1}{2}D^2 g(Y_{k-1})(Y_k - Y_{k-1})^{\otimes 2} + R_2(Y_{k-1}, Y_k)\\
&= \nabla_x g(x_{k-1}, v_{k-1})\cdot[\frac{1-e^{-2\gamma_k}}{2}v_{k-1} - \frac{u}{2}(\gamma_k - \frac{1-e^{-2\gamma_k}}{2})\nabla f(x_{k-1}) + \sqrt{u}\sigma_k^{(1)}U_k^{(1)}]\\
&+ \nabla_v g(x_{k-1}, v_{k-1})\cdot[-2\frac{1-e^{-2\gamma_k}}{2}v_{k-1} - u\frac{1-e^{-2\gamma_k}}{2}\nabla f(x_{k-1}) + 2\sqrt{u}\sigma_k^{(2)}U_k^{(2)}]\\
&+ \frac{1}{2}D_x^2 g(x_{k-1}, v_{k-1})[\frac{1-e^{-2\gamma_k}}{2}v_{k-1} - \frac{u}{2}(\gamma_k - \frac{1-e^{-2\gamma_k}}{2})\nabla f(x_{k-1}) + \sqrt{u}\sigma_k^{(1)}U_k^{(1)}]^{\otimes 2}\\
&+ \frac{1}{2}D_v^2 g(x_{k-1}, v_{k-1})[-2\frac{1-e^{-2\gamma_k}}{2}v_{k-1} - u\frac{1-e^{-2\gamma_k}}{2}\nabla f(x_{k-1}) + 2\sqrt{u}\sigma_k^{(2)}U_k^{(2)}]^{\otimes 2}\\
&+ \langle D_{xv} g(x_{k-1}, v_{k-1}); \frac{1-e^{-2\gamma_k}}{2}v_{k-1} - \frac{u}{2}(\gamma_k - \frac{1-e^{-2\gamma_k}}{2})\nabla f(x_{k-1}) + \sqrt{u}\sigma_k^{(1)}U_k^{(1)},\\
&\qquad - 2\frac{1-e^{-2\gamma_k}}{2}v_{k-1} - u\frac{1-e^{-2\gamma_k}}{2}\nabla f(x_{k-1}) + 2\sqrt{u}\sigma_k^{(2)}U_k^{(2)}\rangle\\
&+ R_2(Y_{k-1}, Y_k)\\
&= \gamma_k \mathcal{L}g(Y_{k-1}) - (\gamma_k - \frac{1-e^{-2\gamma_k}}{2})\nabla_x g(Y_{k-1})\cdot v_{k-1} - \frac{u}{2}(\gamma_k - \frac{1-e^{-2\gamma_k}}{2})\nabla_x g(Y_{k-1})\cdot\nabla f(x_{k-1})\\
&+ 2(\gamma_k - \frac{1-e^{-2\gamma_k}}{2})\nabla_v g(Y_{k-1})\cdot v_{k-1} + u(\gamma_k - \frac{1-e^{-2\gamma_k}}{2})\nabla_v g(Y_{k-1})\cdot\nabla f(x_{k-1})\\
&+ \sqrt{u}\sigma_k^{(1)}\nabla g(Y_{k-1})\cdot U_k^{(1)} + 2\sqrt{u}\sigma_k^{(2)}\nabla g(Y_{k-1})\cdot U_k^{(2)}\\
&+ \frac{1}{2}(\frac{1-e^{-2\gamma_k}}{2})^2 D_x^2 g(Y_{k-1})v_{k-1}^{\otimes 2} + \frac{u^2}{8}(\gamma_k - \frac{1-e^{-2\gamma_k}}{2})^2 D_x^2 g(Y_{k-1})\nabla f(x_{k-1})^{\otimes 2}\\
&+ \frac{u}{2}{\sigma_k^{(1)}}^2 D_x^2 g(Y_{k-1}){U_k^{(1)}}^{\otimes 2} - \frac{u}{2}\frac{1-e^{-2\gamma_k}}{2}(\gamma_k - \frac{1-e^{-2\gamma_k}}{2})\langle D_x^2 g(Y_{k-1}); v_{k-1}, \nabla f(x_{k-1})\rangle\\
&+ \sqrt{u}\frac{1-e^{-2\gamma_k}}{2}\sigma_k^{(1)}\langle D_x^2 g(Y_{k-1}); v_{k-1}, U_k^{(1)}\rangle\\
&- \frac{u^{3/2}}{2}(\gamma_k - \frac{1-e^{-2\gamma_k}}{2})\sigma_k^{(1)}\langle D_x^2 g(Y_{k-1}); \nabla f(x_{k-1}), U_k^{(1)}\rangle\\
&+ 2(\frac{1-e^{-2\gamma_k}}{2})^2 D_v^2 g(Y_{k-1})v_{k-1}^{\otimes 2} + \frac{u^2}{2}(\frac{1-e^{-2\gamma_k}}{2})^2 D_v^2 g(Y_{k-1})\nabla f(x_{k-1})^{\otimes 2}
\end{aligned}$$

$$
\begin{aligned}
&+ 2u\left({\sigma_k^{(2)}}^2 D_v^2 g(Y_{k-1}) {U_k^{(2)}}^{\otimes 2} - \gamma_k \mathbb{E}[D_v^2 g(Y_{k-1}) {U_k^{(2)}}^{\otimes 2} | \mathcal{F}_{k-1}]\right) \\
&+ 2u(\frac{1-e^{-2\gamma_k}}{2})^2 \langle D_v^2 g(Y_{k-1}); v_{k-1}, \nabla f(x_{k-1})\rangle - 4\sqrt{u}\frac{1-e^{-2\gamma_k}}{2}\sigma_k^{(2)}\langle D_v^2 g(Y_{k-1}); v_{k-1}, U_k^{(2)}\rangle \\
&- 2u^{3/2}\frac{1-e^{-2\gamma_k}}{2}\sigma_k^{(2)}\langle D_v^2 g(Y_{k-1}); \nabla f(x_{k-1}), U_k^{(2)}\rangle + R_2(Y_{k-1}, Y_k)
\end{aligned}
$$

Take conditional expectation:

$$
\begin{aligned}
\mathbb{E}[g(Y_k) - g(Y_{k-1})|\mathcal{F}_{k-1}] - \gamma_k \mathcal{L} g(Y_{k-1}) = &-(\gamma_k - \frac{1-e^{-2\gamma_k}}{2})\nabla_x g(Y_{k-1})\cdot v_{k-1} \\
&- \frac{u}{2}(\gamma_k - \frac{1-e^{-2\gamma_k}}{2})\nabla_x g(Y_{k-1})\cdot \nabla f(x_{k-1}) \\
&+ 2(\gamma_k - \frac{1-e^{-2\gamma_k}}{2})\nabla_v g(Y_{k-1})\cdot v_{k-1} \\
&+ u(\gamma_k - \frac{1-e^{-2\gamma_k}}{2})\nabla_v g(Y_{k-1})\cdot \nabla f(x_{k-1}) \\
&+ \frac{1}{2}(\frac{1-e^{-2\gamma_k}}{2})^2 D_x^2 g(Y_{k-1}) v_{k-1}^{\otimes 2} \\
&+ \frac{u^2}{8}(\gamma_k - \frac{1-e^{-2\gamma_k}}{2})^2 D_x^2 g(Y_{k-1}) \nabla f(x_{k-1})^{\otimes 2} \\
&+ \frac{u}{2}{\sigma_k^{(1)}}^2 \Delta_x g(Y_{k-1}) \\
&- \frac{u}{2}\frac{1-e^{-2\gamma_k}}{2}(\gamma_k - \frac{1-e^{-2\gamma_k}}{2})\langle D_x^2 g(Y_{k-1}); v_{k-1}, \nabla f(x_{k-1})\rangle \\
&+ 2(\frac{1-e^{-2\gamma_k}}{2})^2 D_v^2 g(Y_{k-1}) v_{k-1}^{\otimes 2} + \frac{u^2}{2}(\frac{1-e^{-2\gamma_k}}{2})^2 D_v^2 g(Y_{k-1}) \nabla f(x_{k-1})^{\otimes 2} \\
&+ 2u\left({\sigma_k^{(2)}}^2 - \gamma_k\right)\Delta_v g(Y_{k-1}) \\
&+ 2u(\frac{1-e^{-2\gamma_k}}{2})^2 \langle D_v^2 g(Y_{k-1}); v_{k-1}, \nabla f(x_{k-1})\rangle \\
&+ \mathbb{E}[R_2(Y_{k-1}, Y_k)|\mathcal{F}_{k-1}]
\end{aligned}
$$

Observe that for all the terms, except for $R_2(Y_{k-1}, Y_k)$, on the right hand side of the equation, their coefficients are of order $O(\gamma_k^2)$ or $o(\gamma_k^2)$. Furthermore, ∇g and $D^2 g$ are bounded because g is $\mathcal{C}^2$ and compact supported. Since $(\mathcal{L}_{V,\infty})$ is satisfied under our assumptions, $\sup_{n\in\mathbb{N}} \mathbb{E}[|v_n|^2 + |\nabla f(x_n)|^2] < C \sup_{n\in\mathbb{N}} \mathbb{E}[V(Y_n))] < +\infty$. Therefore, we obtain that as $n \to 0$,

$$
\frac{1}{H_n}\sum_{k=1}^n \frac{\eta_k}{\gamma_k}\mathbb{E}[g(Y_k) - g(Y_{k-1})|\mathcal{F}_{k-1}] - \eta_k \mathcal{L} g(Y_{k-1}) - \frac{\eta_k}{\gamma_k}\mathbb{E}[R_2(Y_{k-1}, Y_k)|\mathcal{F}_{k-1}] \to 0
$$

because $\frac{1}{H_n}\sum_{k=1}^n \eta_k\gamma_k \to 0$ as $n \to 0$.
Now we deal with $\mathbb{E}[R_2(Y_{k-1}, Y_k)|\mathcal{F}_{k-1}]$. For any $x, y \in \mathbb{R}^{2d}$, define

$$
r_2(x,y) := \frac{1}{2}\sup_{t\in(0,1)} \left\|D^2 g(x + t(y-x)) - D^2 g(x)\right\|
$$

It's easy to see that r_2 is a bounded continuous function on $\mathbb{R}^d \times \mathbb{R}^d$, $r_2(x,x) = 0$ and

$$
|R_2(x,y)| \le r_2(x,y)|x-y|^2
$$

Therefore we obtain

$$
\frac{\eta_k}{\gamma_k}|\mathbb{E}[R_2(Y_{k-1}, Y_k)|\mathcal{F}_{k-1}]| \le C\left(\eta_k\gamma_k \left\|r_2\right\|_\infty V(Y_{k-1}) + (2d+1)\eta_k \mathbb{E}[r_2(Y_{k-1}, Y_k)(|U_k^{(1)}|^2 + |U_k^{(2)}|^2)|\mathcal{F}_{k-1}]\right)
$$

If $a \ge 1$, $\mathbb{P} - a.s$

$$
\frac{1}{H_n}\sum_{k=1}^n C\eta_k\gamma_k \left\|r_2\right\|_\infty V(Y_{k-1}) \le C' \frac{1}{H_n}\sum_{k=1}^n \eta_k\gamma_k V(Y_{k-1}) \to 0 \quad \text{as } \sup_{n\in\mathbb{N}} \nu_n^\eta(V) < +\infty \text{ and } \gamma_n \to 0
$$

If $a \in [1/2, 1)$, the same limit follows from the Kronecker lemma mentioned in[21] and $\sum_{n\geq 1} \eta_n\gamma_n/H_n < +\infty$.
Meanwhile,

$$J(\gamma, x, v) = \int_{\mathbb{R}^d\times\mathbb{R}^d} r_2((x,v),(x',v'))(|r_1|^2 + |r_2|^2)\mu(dr_1, dr_2)$$

where

$$(x', v') = (x + \frac{1-e^{-2\gamma}}{2}v - \frac{u}{2}(\gamma - \frac{1-e^{-2\gamma}}{2})\nabla f(x) + \sqrt{u}\sigma^{(1)}r_1, e^{-2\gamma}v - u\frac{1-e^{-2\gamma}}{2}\nabla f(x) + 2\sqrt{u}\sigma^{(2)}r_2)$$

$$\sigma^{(1)} = \left(\gamma + \frac{1-e^{-4\gamma}}{4} - (1-e^{-2\gamma})\right)^{1/2}, \quad \sigma^{(2)} = \left(\frac{1-e^{-4\gamma}}{4}\right)^{1/2}$$

and $(U^{(1)}, U^{(2)}) \sim \mu = \mathcal{N}(0, \frac{1+e^{-4\gamma}-2e^{-2\gamma}}{4\sigma^{(1)}\sigma^{(2)}}I_{2d})$

We can see that J is a bounded continuous function on $\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d$ and $J(0, x, v) = 0$. Since $\lim_{|y|\to\infty} V(y) = +\infty$. We can also write

$$(2d+1)\eta_k\mathbb{E}[r_2(Y_{k-1}, Y_k)(|U_k^{(1)}|^2 + |U_k^{(2)}|^2)|\mathcal{F}_{k-1}] = \eta_k V^a((x_{k-1}, v_{k-1}))\theta((x_{k-1}, v_{k-1}))J(\gamma_k, x_{k-1}, v_{k-1})$$

where $\lim_{|(x_{k-1},v_{k-1})|\to\infty} \theta((x_{k-1}, v_{k-1})) = 0$ It remains to show that

$$\mathbb{P} - a.s \quad \lim_n \frac{1}{H_n}\sum_{k=1}^n \eta_k V^a((x_{k-1}, v_{k-1}))\theta((x_{k-1}, v_{k-1}))J(\gamma_k, x_{k-1}, v_{k-1}) = 0$$

For a fixed number $A > 0$, J is uniformly continuous on $[0, \sup_n \gamma_n] \times \bar{B}_{2d}(0, A)$, then

$$J(\gamma_k, x_{k-1}, v_{k-1})1_{|(x_{k-1},v_{k-1})|\leq A} \to 0 \ \ \mathbb{P} - a.s.$$

And $V^a((x_{k-1}, v_{k-1}))\theta((x_{k-1}, v_{k-1}))$ is bounded on $\bar{B}_{2d}(0, A)$. Therefore

$$\mathbb{P} - a, s \quad \lim_n \frac{1}{H_n}\sum_{k=1}^n \eta_k V^a((x_{k-1}, v_{k-1}))\theta((x_{k-1}, v_{k-1}))J(\gamma_k, x_{k-1}, v_{k-1})1_{|(x_{k-1},v_{k-1})|\leq A} = 0$$

On the other hand side

$$\begin{aligned}\limsup_n \frac{1}{H_n}\sum_{k=1}^n \eta_k V^a((x_{k-1}, v_{k-1}))\theta((x_{k-1}, v_{k-1}))J(\gamma_k, x_{k-1}, v_{k-1})1_{|(x_{k-1},v_{k-1})|>A} \\ \leq \sup_{|(x,v)|>A} |\theta(x,v)|\,\|J\|_\infty \sup_n \nu_n^\eta(V^a) \to 0 \quad \textbf{as } A \to +\infty\end{aligned}$$

So taking $A \to +\infty$ completes the proof. □

Theorem 9. *Let $p \in [1, +\infty)$. Assume $(\mathcal{L}_{V,p})$. Let $s \in (0, 1]$. Assume that*

$$\sum_{n\geq 1}\frac{1}{H_n}\left(\Delta\frac{\eta_n}{\gamma_n}\right)_+ < +\infty. \ \lim_n \frac{1}{H_n}\sum_{k=1}^n |\Delta\frac{\eta_k}{\gamma_k}| = 0 \text{ and } \sum_{n\geq 1}\left(\frac{\eta_n}{H_n\sqrt{\gamma_n}}\right)^{1+s} < +\infty$$

(a) Then

$$\mathbb{P} - a.s \quad \sup_{n\in\mathbb{N}} \nu_n^\eta(\omega, V^{p/(1+s)}) < +\infty.$$

(b) When $p \leq 1 + s$, assume also $\sum_{n\geq 1}\eta_n\gamma_n/H_n < +\infty$. Then with probability 1, any weak limit of the sequence (ν_m^η) is an invariant distribution of the underdamped Langevin dynamics.

Theorem 9 follows directly from theorem 6 and theorem 8.

Proof of Theorem 5. First we try to decompose $\sum_{k=1}^{n}\gamma_k\mathcal{L}\phi(x_{k-1})$ using Taylor expansion.

$$\phi(x_k)=\phi(x_{k-1})+\nabla\phi(x_{k-1})\cdot(x_k-x_{k-1})+\frac{1}{2}D^2\phi(x_{k-1})(x_k-x_{k-1})^{\otimes 2}+R_2^{(k)}$$

where $R_2^{(k)}=\phi(x_k)-\phi(x_{k-1})-\nabla\phi(x_{k-1})\cdot(x_k-x_{k-1})-\frac{1}{2}D^2\phi(x_{k-1})(x_k-x_{k-1})^{\otimes 2}$. We can plug our discretization into the equation and obtain:

$$\begin{aligned}
\phi(x_k)-\phi(x_{k-1})=&\gamma_k\mathcal{L}\phi(x_{k-1})-(\gamma_k-\frac{1-e^{-2\gamma_k}}{2})v_{k-1}\cdot\nabla\phi(x_{k-1})\\
&-\frac{u}{2}(\gamma_k-\frac{1-e^{-2\gamma_k}}{2})\nabla f(x_{k-1})\cdot\nabla\phi(x_{k-1})+\sqrt{u}\sigma_k^{(1)}\nabla\phi(x_{k-1})\cdot U_k^{(1)}\\
&+\frac{1}{2}(\frac{1-e^{-2\gamma_k}}{2})^2D^2\phi(x_{k-1})v_{k-1}^{\otimes 2}+\frac{u}{2}{\sigma_k^{(1)}}^2D^2\phi(x_{k-1}){U_k^{(1)}}^{\otimes 2}\\
&+\frac{u^2}{8}(\gamma_k-\frac{1-e^{-2\gamma_k}}{2})^2D^2\phi(x_{k-1})\nabla f(x_{k-1})^{\otimes 2}\\
&-\frac{u}{2}\frac{1-e^{-2\gamma_k}}{2}(\gamma_k-\frac{1-e^{-2\gamma_k}}{2})\langle D^2\phi(x_{k-1});v_{k-1},\nabla f(x_{k-1})\rangle\\
&+\sqrt{u}\sigma_k^{(1)}\frac{1-e^{-2\gamma_k}}{2}\langle D^2\phi(x_{k-1});v_{k-1},U_k^{(1)}\rangle\\
&-\frac{u^{3/2}}{2}\sigma_k^{(1)}(\gamma_k-\frac{1-e^{-2\gamma_k}}{2})\langle D^2\phi(x_{k-1});\nabla f(x_{k-1}),U_k^{(1)}\rangle\\
&+R_2^{(k)}
\end{aligned}$$

where

$$\begin{aligned}
R_2^{(k)}=&\frac{1}{6}(\frac{1-e^{-2\gamma_k}}{2})^3D^3\phi(x_{k-1})v_{k-1}^{\otimes 3}-\frac{u}{4}(\frac{1-e^{-2\gamma_k}}{2})^2(\gamma_k-\frac{1-e^{-2\gamma_k}}{2})\langle D^3\phi(x_{k-1});v_{k-1}^{\otimes 2},\nabla f(x_{k-1})\rangle\\
&+\frac{\sqrt{u}}{2}\sigma_k^{(1)}(\frac{1-e^{-2\gamma_k}}{2})^2\langle D^3\phi(x_{k-1});v_{k-1}^{\otimes 2},U_k^{(1)}\rangle+\frac{u}{2}{\sigma_k^{(1)}}^2\frac{1-e^{-2\gamma_k}}{2}\langle D^3\phi(x_{k-1});v_{k-1},{U_k^{(1)}}^{\otimes 2}\rangle\\
&+\frac{1}{24}(\frac{1-e^{-2\gamma_k}}{2})^4D^4\phi(x_{k-1})v_{k-1}^{\otimes 4}+r^{(k)}
\end{aligned}$$

Since f is gradient Lipschitz and strongly convex, we've shown $(\mathcal{L}_{V,\infty})$ holds. Using $(\mathcal{L}_{V,\infty})$ the fact that $D^4\phi$ is bounded and Lipschitz, we can show there exists a constant $C>0$ such that

$$|r_k|\le C\gamma_k^{9/2}V^2(x_{k-1},v_{k-1})$$

Apply theorem 6 for $p=4$ and $s=1$, we have $\sup_n\nu_n^\gamma(V^2)<+\infty\ \ \mathbb{P}-a.s.$ Therefore

$$\frac{1}{\Gamma_n^{(4)}}\sum_{k=1}^{n}r^{(k)}\to 0\qquad\text{in }\mathbb{L}^1$$

In the following proof, we will use $o(\gamma_k^4)$ to denote the sum of those terms b_k such that $\frac{1}{\Gamma_n^{(4)}}\sum_{k=1}^{n}b_k\to 0\ \ \mathbb{P}-a.s.$ According to our decomposition, we can pull out polynomials of γ_k from factors $\frac{1-e^{-2\gamma_k}}{2}$, $\gamma_k-\frac{1-e^{-2\gamma_k}}{2}$ and $\sigma_k^{(1)}$ so that the terms left could be included in $o(\gamma_k^4)$. Then we obtain

$$\begin{aligned}
\sum_{k=1}^{n}\gamma_k\mathcal{L}\phi(x_{k-1})=\sum_{k=1}^{n}\Big\{&[\phi(x_k)-\phi(x_{k-1})]+(\gamma_k^2-\frac{2}{3}\gamma_k^3+\frac{1}{3}\gamma_k^4)v_{k-1}\cdot\nabla\phi(x_{k-1})\\
&+\frac{u}{2}(\gamma_k^2-\frac{2}{3}\gamma_k^3+\frac{1}{3}\gamma_k^4)\nabla f(x_{k-1})\cdot\nabla\phi(x_{k-1})\\
&-\frac{2\sqrt{3u}}{3}\gamma_k^{\frac{3}{2}}\nabla\phi(x_{k-1})\cdot U_k^{(1)}\\
&-\frac{1}{2}(\gamma_k^2-2\gamma_k^3+\frac{7}{3}\gamma_k^4)D^2\phi(x_{k-1})v_{k-1}^{\otimes 2}
\end{aligned}$$

$$
\begin{aligned}
&- \frac{u^2}{8}\gamma_k^4 D^2\phi(x_{k-1})\nabla f(x_{k-1})^{\otimes 2} \\
&- \frac{u}{2}(\frac{4}{3}\gamma_k^3 - 2\gamma_k^4)D^2\phi(x_{k-1})U_k^{(1)^{\otimes 2}} \\
&+ \frac{u}{2}(\gamma_k^3 - \frac{5}{3}\gamma_k^4)\langle D^2\phi(x_{k-1}); v_{k-1}, \nabla f(x_{k-1})\rangle \\
&- \frac{1}{6}(\gamma_k^3 - 3\gamma_k^4)D^3\phi(x_{k-1})v_{k-1}^{\otimes 3} \\
&+ \frac{u}{4}\gamma_k^4\langle D^3\phi(x_{k-1}); v_{k-1}^{\otimes 2}, \nabla f(x_{k-1})\rangle \\
&- \frac{2u}{3}\gamma_k^4\langle D^3\phi(x_{k-1}); v_{k-1}, U_k^{(1)^{\otimes 2}}\rangle \\
&- \frac{1}{24}\gamma_k^4 D^4\phi(x_{k-1})v_{k-1}^{\otimes 4} + o(\gamma_k^4)\Big\} \\
:= &\, Z_n^{(0)} + Z_n^{(2)} + Z_n^{(3)} + Z_n^{(4)} + N_n + r_n
\end{aligned}
$$

where

$$
\begin{aligned}
Z_n^{(0)} &= \phi(x_n) - \phi(x_0) \\
Z_n^{(2)} &= \sum_{k=1}^n \gamma_k^2 \left[v_{k-1}\cdot\nabla\phi(x_{k-1}) + \frac{u}{2}\nabla f(x_{k-1})\cdot\nabla\phi(x_{k-1}) - \frac{1}{2}D^2\phi(x_{k-1})v_{k-1}^{\otimes 2}\right] \\
&:= \sum_{k=1}^n \gamma_k^2 z_{k-1}^{(2)} \\
Z_n^{(3)} &= \sum_{k=1}^n \gamma_k^3 \left[-\frac{2}{3}v_{k-1}\cdot\nabla\phi(x_{k-1}) - \frac{u}{3}\nabla f(x_{k-1})\cdot\nabla\phi(x_{k-1}) + D^2\phi(x_{k-1})v_{k-1}^2 \right. \\
&\quad \left. -\frac{2u}{3}D^2\phi(x_{k-1})U_k^{(1)^{\otimes 2}} + \frac{u}{2}\langle D^2\phi(x_{k-1}); v_{k-1}, \nabla f(x_{k-1})\rangle - \frac{1}{6}D^3\phi(x_{k-1})v_{k-1}^{\otimes 3}\right] \\
&:= -\sum_{k=1}^n \gamma_k^3 z_{k-1}^{(3)} \\
Z_n^{(4)} &= \sum_{k=1}^n \gamma_k^4 \left[\frac{1}{3}v_{k-1}\cdot\nabla\phi(x_{k-1}) + \frac{u}{6}\nabla f(x_{k-1})\cdot\nabla\phi(x_{k-1}) - \frac{7}{6}D^2\phi(x_{k-1})v_{k-1}^{\otimes 2}\right. \\
&\quad -\frac{u^2}{8}D^2\phi(x_{k-1})\nabla f(x_{k-1})^{\otimes 2} + uD^2\phi(x_{k-1})U_k^{(1)^{\otimes 2}} - \frac{5u}{6}\langle D^2\phi(x_{k-1}); v_{k-1}, \nabla f(x_{k-1})\rangle \\
&\quad +\frac{1}{2}D^3\phi(x_{k-1})v_{k-1}^{\otimes 3} + +\frac{u}{4}\langle D^3\phi(x_{k-1}); v_{k-1}^{\otimes 2}, \nabla f(x_{k-1})\rangle - \frac{2u}{3}\langle D^3\phi(x_{k-1}); v_{k-1}, U_k^{(1)^{\otimes 2}} \\
&\quad \left. -\frac{1}{24}D^4\phi(x_{k-1})v_{k-1}^{\otimes 4}\right] := \sum_{k=1}^n \gamma_k^4 z_{k-1}^{(4)} \\
N_n &= \sum_{k=1}^n \frac{2\sqrt{3u}}{3}\gamma_k^{\frac{3}{2}}\nabla\phi(x_{k-1})\cdot U_k^{(1)} \\
r_n &= \sum_{k=1}^n o(\gamma_k^4)
\end{aligned}
$$

First, it's easy to see that $r_n/\Gamma_n^{(4)} \to 0 \ \ \mathbb{P}-a.s$ as $n \to +\infty$. Apply lemma 5 and we obtain $\sup_n \mathbb{E}[V(x_n, v_n)] < +\infty$. Therefore we can further obtain the tightness of sequence $\{x_n\}$ and it follows from the continuity of ϕ that $\{\phi(x_n)\}$ is also tight. According to the tightness, $Z_n^{(0)}/\Gamma_n^{(4)} \to 0 \ \ \mathbb{P}-a.s$. For $Z_n^{(4)}$, under our assumptions on ϕ and f, we can show that

$$
\lim_{|(x_n,v_n)|\to+\infty} z_n^{(4)}/V^4(x_n, v_n) = 0
$$

Therefor apply theorem 9 with $p = 8,\ s = 1$ and we obtain:

$$\mathbb{P}-a.s \quad Z_n^{(4)}/\Gamma_n^{(4)} \to \frac{u}{4}\int_{\mathbb{R}^{2d}}\langle D^3\phi(x);\nabla f(x), v^{\otimes 2}\rangle\nu(dx,dv) - \frac{u^2}{8}\int_{\mathbb{R}^d} D^2\phi(x)\nabla f(x)^{\otimes 2}\pi(dx) - \frac{1}{24}\int_{\mathbb{R}^{2d}} D^4\phi(x)v^{\otimes 4}\nu(dx,dv)$$

To consider the limit of $Z_n^{(i)}/\Gamma_n^{(4)}$ for $i = 2, 3$, We first Taylor expand $\mathcal{L}\phi(x_{k-1})$ at x_{k-2}:

$$\begin{aligned}\mathcal{L}\phi(x_{k-1}) &= v_{k-2}\cdot\nabla\phi(x_{k-2}) + \langle D^2\phi(x_{k-2}); v_{k-2}, x_{k-1}-x_{k-2}\rangle + \nabla\phi(x_{k-2})\cdot(v_{k-1}-v_{k-2})\\ &+ \frac{1}{2}\langle D^3\phi(x_{k-2}); v_{k-2}, (x_{k-1}-x_{k-2})^{\otimes 2}\rangle + \langle D^2\phi(x_{k-2}); v_{k-1}-v_{k-2}, x_{k-1}-x_{k-2}\rangle\\ &+ \frac{1}{6}\langle D^4\phi(x_{k-2}); v_{k-2}, (x_{k-1}-x_{k-2})^{\otimes 3}\rangle\\ &+ \frac{1}{2}\langle D^3\phi(x_{k-2}); v_{k-1}-v_{k-2}, (x_{k-1}-x_{k-2})^{\otimes 2}\rangle\\ &+ o(\gamma_k^3)\end{aligned}$$

Plug the discretization into the Taylor expansions and preserve the "large" terms, then we obtain:

$$\begin{aligned}\mathcal{L}\phi(x_{k-1}) &= \mathcal{L}\phi(x_{k-2}) + (\gamma_{k-1}-\gamma_{k-1}^2+\frac{2}{3}\gamma_{k-1}^3)D^2\phi(x_{k-2})v_{k-2}^{\otimes 2}\\ &- \frac{u}{2}(\gamma_{k-1}^2-\frac{2}{3}\gamma_{k-1}^3)\langle D^2\phi(x_{k-2}); v_{k-2}, \nabla f(x_{k-2})\rangle\\ &+ \frac{2\sqrt{3u}}{3}\gamma_{k-1}^{\frac{3}{2}}\langle D^2\phi(x_{k-2}); v_{k-2}, U_{k-1}^{(1)}\rangle\\ &- (2\gamma_{k-1}-2\gamma_{k-1}^2+\frac{4}{3}\gamma_{k-1}^3)v_{k-2}\cdot\nabla\phi(x_{k-2})\\ &- u(\gamma_{k-1}-\gamma_{k-1}^2+\frac{2}{3}\gamma_{k-1}^3)\nabla f(x_{k-2})\cdot\nabla\phi(x_{k-2})\\ &+ 2\sqrt{u}\gamma_{k-1}^{\frac{1}{2}}\nabla\phi(x_{k-2})\cdot U_{k-1}^{(2)}\\ &+ \frac{1}{2}(\gamma_{k-1}^2-2\gamma_{k-1}^3)D^3\phi(x_{k-2})v_{k-2}^{\otimes 3}\\ &+ \frac{2u}{3}\gamma_{k-1}^3\langle D^3\phi(x_{k-2}); v_{k-2}, U_{k-1}^{(1)\ \otimes 2}\rangle\\ &+ \sqrt{u}\gamma_{k-1}^{\frac{5}{2}}\langle D^3\phi(x_{k-2}); v_{k-2}^{\otimes 2}, U_{k-1}^{(1)}\rangle\\ &- \frac{u}{2}\gamma_{k-1}^3\langle D^3\phi(x_{k-2}); v_{k-2}^{\otimes 2}, \nabla f(x_{k-2})\rangle\\ &- 2(\gamma_{k-2}^2-2\gamma_{k-1}^3)D^2\phi(x_{k-2})v_{k-2}^{\otimes 2}\\ &- u(\gamma_{k-1}^2-2\gamma_{k-1}^3)\langle D^2\phi(x_{k-2}); v_{k-2}, \nabla f(x_{k-2})\rangle\\ &+ 2\sqrt{u}\gamma_{k-1}^{\frac{3}{2}}\langle D^2\phi(x_{k-2}); v_{k-2}, U_{k-1}^{(2)}\rangle\\ &+ u\gamma_{k-1}^3\langle D^2\phi(x_{k-2}); v_{k-2}, \nabla f(x_{k-2})\rangle\\ &+ \frac{u^2}{2}\gamma_{k-1}^3 D^2\phi(x_{k-2})\nabla f(x_{k-2})^{\otimes 2}\end{aligned}$$

$$\begin{aligned}&- u^{\frac{3}{2}}\gamma_{k-1}^{\frac{5}{2}}\langle D^2\phi(x_{k-2}); \nabla f(x_{k-2}), U_{k-1}^{(2)}\rangle\\ &- 2\sqrt{u}\gamma_{k-1}^{\frac{5}{2}}\langle D^2\phi(x_{k-2}); v_{k-2}, U_{k-1}^{(1)}\rangle\\ &- u^{\frac{3}{2}}\gamma_{k-1}^{\frac{5}{2}}\langle D^2\phi(x_{k-2}); \nabla f(x_{k-2}), U_{k-1}^{(1)}\rangle\\ &+ \langle D^2\phi(x_{k-2}); \sqrt{u}\sigma_{k-1}^{(1)}U_{k-1}^{(1)}, 2\sqrt{u}\sigma_{k-1}^{(2)}U_{k-1}^{(2)}\rangle + o(\gamma_{k-1}^3)\end{aligned}$$

Apply theorem 9 with $p=4, s=1$ to the terms of order $o(V^2(x_{k-2}, v_{k-2}))$ in the decomposition. We obtain

$$\begin{aligned}
\lim_n \frac{\sum_{k=2}^n \gamma_k \mathcal{L}\phi(x_{k-1})}{\Gamma_n^{(4)}} = \lim_n \frac{1}{\Gamma_n^{(4)}} \Bigg[& \sum_{k=2}^n \gamma_k \mathcal{L}\phi(x_{k-2}) + \sum_{k=2}^n \gamma_k(\gamma_{k-1} - 3\gamma_{k-1}^2) D^2\phi(x_{k-2}) v_{k-2}^{\otimes 2} \\
& - \sum_{k=2}^n \frac{3u}{2} \gamma_k \gamma_{k-1}^2 \langle D^2\phi(x_{k-2}); \nabla f(x_{k-2}), v_{k-2} \rangle \\
& - \sum_{k=2}^n \gamma_k(2\gamma_{k-1} - 2\gamma_{k-1}^2) \nabla\phi(x_{k-2}) \cdot v_{k-2} \\
& - \sum_{k=2}^n u\gamma_k(\gamma_{k-1} - \gamma_{k-1}^2) \nabla f(x_{k-2}) \cdot \nabla\phi(x_{k-2}) \\
& + \sum_{k=2}^n \frac{1}{2} \gamma_k \gamma_{k-1}^2 D^3\phi(x_{k-2}) v_{k-2}^{\otimes 3} \\
& + \sum_{k=2}^n 2\sqrt{u}\gamma_k \gamma_{k-1}^{\frac{1}{2}} \nabla\phi(x_{k-2}) \cdot U_{k-1}^{(2)} \\
& + \sum_{k=2}^n \gamma_k \langle D^2\phi(x_{k-2}); \sqrt{u}\sigma_{k-1}^{(1)} U_{k-1}^{(1)}, 2\sqrt{u}\sigma_{k-1}^{(2)} U_{k-1}^{(2)} \rangle \Bigg] \\
& + 4u \int_{\mathbb{R}^d} \Delta\phi(x)\pi(dx) - \frac{u}{2} \int_{\mathbb{R}^{2d}} \langle D^3\phi(x); v^{\otimes 2}, \nabla f(x) \rangle \nu(dx, dv) \\
& + \frac{u^2}{2} \int_{\mathbb{R}^d} D^2\phi(x) \nabla f(x)^{\otimes 2} \pi(dx)
\end{aligned}$$

Since $\gamma_{k-1} - \gamma_k = o(\gamma_k^4)$, we can substitute all the γ_k on the right hand side with γ_{k-1} and it won't change the limits. For the last term inside the square bracket, notice that $Var(\sqrt{u}\sigma_{k-1}^{(1)} U_{k-1}^{(1)}, 2\sqrt{u}\sigma_{k-1}^{(2)} U_{k-1}^{(2)}) = \frac{u}{2}(1 + e^{-4\gamma_{k-1}} - 2e^{-2\gamma_{k-1}}) I_d \sim u(2\gamma_{k-1}^2 - 4\gamma_{k-1}^3) I_d$. Therefore

$$\begin{aligned}
\lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=2}^n \gamma_k \langle D^2\phi(x_{k-2}); \sqrt{u}\sigma_{k-1}^{(1)} U_{k-1}^{(1)}, 2\sqrt{u}\sigma_{k-1}^{(2)} U_{k-1}^{(2)} \rangle = & \lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=2}^n 2u\gamma_{k-1}^3 \Delta\phi(x_{k-2}) \\
& - 4u \int_{\mathbb{R}^d} \Delta\phi(x)\pi(dx)
\end{aligned}$$

We can rewrite the equation as

$$\begin{aligned}
\lim_n \frac{\sum_{k=2}^n \gamma_k \mathcal{L}\phi(x_{k-1})}{\Gamma_n^{(4)}} = & \lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=2}^n \gamma_k \mathcal{L}\phi(x_{k-2}) + \lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=2}^n 2\sqrt{u}\gamma_{k-1}^{\frac{3}{2}} \nabla\phi(x_{k-2}) \cdot U_{k-1}^{(2)} \\
& + \lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=2}^n \gamma_{k-1}^2 [D^2\phi(x_{k-2}) v_{k-2}^{\otimes 2} - 2\nabla\phi(x_{k-2}) \cdot v_{k-2} - u\nabla f(x_{k-2}) \cdot \nabla\phi(x_{k-2})] \\
& + \lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=2}^n \gamma_{k-1}^3 [-3D^{\phi}(x_{k-2}) v_{k-2}^{\otimes 2} - \frac{3u}{2} \langle D^2\phi(x_{k-2}); \nabla f(x_{k-2}), v_{k-2} \rangle \\
& + 2\nabla\phi(x_{k-2}) \cdot v_{k-2} + u\nabla f(x_{k-2}) \cdot \nabla\phi(x_{k-2}) + \frac{1}{2} D^3\phi(x_{k-2}) v_{k-2}^{\otimes 3} + 2u\Delta\phi(x_{k-2})] \\
& - \frac{u}{2} \int_{\mathbb{R}^{2d}} \langle D^3\phi(x); v^{\otimes 2}, \nabla f(x) \rangle \nu(dx, dv) + \frac{u^2}{2} \int_{\mathbb{R}^d} D^2\phi(x) \nabla f(x)^{\otimes 2} \pi(dx) \\
= & \lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=2}^n \gamma_k \mathcal{L}\phi(x_{k-2}) + \lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=2}^n 2\sqrt{u}\gamma_{k-1}^{\frac{3}{2}} \nabla\phi(x_{k-2}) \cdot U_{k-1}^{(2)} \\
& + \lim_n \frac{1}{\Gamma_n^{(4)}} (-2Z_n^{(2)} - 3Z_n^{(3)})
\end{aligned}$$

$$- \frac{u}{2}\int_{\mathbb{R}^{2d}} \langle D^3\phi(x); v^{\otimes 2}, \nabla f(x)\rangle \nu(dx,dv) + \frac{u^2}{2}\int_{\mathbb{R}^d} D^2\phi(x)\nabla f(x)^{\otimes 2}\pi(dx)$$

We can instantly get that

$$\begin{aligned}\lim_n \frac{1}{\Gamma_n^{(4)}}(2Z_2^{(n)} + 3Z_n^{(3)}) = \lim_n \frac{1}{\Gamma_n^{(4)}}\sum_{k=2}^n 2\sqrt{u}\gamma_{k-1}^{\frac{3}{2}}\nabla\phi(x_{k-2})\cdot U_{k-1}^{(2)} \\ + \frac{u^2}{2}\int_{\mathbb{R}^d} D^2\phi(x)\nabla f(x)^{\otimes 2}\pi(dx) \\ - \frac{u}{2}\int_{\mathbb{R}^{2d}} \langle D^3\phi(x); \nabla f(x), v^{\otimes 2}\rangle \nu(dx,dv)\end{aligned}$$

Similarly, apply Taylor expansion to $z_{k-1}^{(2)}$ at x_{k-2}, we achieve:

$$\begin{aligned}\nabla f(x_{k-1})\cdot\nabla\phi(x_{k-1}) = \nabla f(x_{k-2})\cdot\nabla\phi(x_{k-2}) + \langle D^2 f(x_{k-2}); \nabla\phi(x_{k-2}), x_{k-1}-x_{k-2}\rangle \\ + \langle D^2\phi(x_{k-2}); \nabla f(x_{k-2}), x_{k-1}-x_{k-2}\rangle \\ + \frac{1}{2}D^3(\nabla f\cdot\nabla\phi)(x_{k-2})(x_{k-1}-x_{k-2})^{\otimes 2} + o(\gamma_k^2)\end{aligned}$$

$$\begin{aligned}\frac{1}{2}D^2\phi(x_{k-1})v_{k-1}^{\otimes 2} = \frac{1}{2}D^2\phi(x_{k-2})v_{k-2}^{\otimes 2} + \frac{1}{2}\langle D^3\phi(x_{k-1}); v_{k-2}^{\otimes 2}, x_{k-1}-x_{k-2}\rangle \\ + \langle D^2\phi(x_{k-2}); v_{k-2}, v_{k-1}-v_{k-2}\rangle + \frac{1}{2}D^2\phi(x_{k-2})(v_{k-1}-v_{k-2})^{\otimes 2} \\ + \frac{1}{4}(\frac{1-e^{-2\gamma_{k-1}}}{2})^2 D^4\phi(x_{k-2})v_{k-2}^{\otimes 4} \\ + \frac{1}{2}\langle D^3\phi(x_{k-2}); v_{k-2}, x_{k-1}-x_{k-2}, v_{k-1}-v_{k-2}\rangle \\ + \frac{1}{6}\langle D^3\phi(x_{k-2}); x_{k-1}-x_{k-2}, (v_{k-1}-v_{k-2})^{\otimes 2}\rangle + o(\gamma_k^2)\end{aligned}$$

Simplifying the coefficients lead us to

$$\begin{aligned}\nabla f(x_{k-1})\cdot\nabla\phi(x_{k-1}) = \nabla f(x_{k-2})\cdot\nabla\phi(x_{k-2}) + (\gamma_{k-1}-\gamma_{k-1}^2)\langle D^2 f(x_{k-2}); \nabla\phi(x_{k-2}), v_{k-2}\rangle \\ - \frac{u}{2}\gamma_{k-1}^2\langle D^2 f(x_{k-2}); \nabla\phi(x_{k-2}), \nabla f(x_{k-2})\rangle \\ + \frac{2\sqrt{3u}}{3}\gamma_{k-1}^{\frac{3}{2}}\langle D^2 f(x_{k-2}); \nabla\phi(x_{k-2}), U_{k-1}^{(1)}\rangle \\ + (\gamma_{k-1}-\gamma_{k-1}^2)\langle D^2\phi(x_{k-2}); \nabla f(x_{k-2}), v_{k-2}\rangle \\ - \frac{u}{2}\gamma_{k-1}^2 D^2\phi(x_{k-2})\nabla f(x_{k-2})^{\otimes 2} + \sqrt{u}\gamma_{k-1}^{\frac{3}{2}}\langle D^2\phi(x_{k-2}); \nabla f(x_{k-2}), U_{k-1}^{(1)}\rangle \\ + \frac{1}{2}\gamma_{k-1}^2(D^3 f\nabla\phi + 2D^2\phi D^2 f + D^3\phi\nabla f)(x_{k-2})v_{k-2}^{\otimes 2} + o(\gamma_{k-1}^2)\end{aligned}$$

$$\begin{aligned}\frac{1}{2}D^2\phi(x_{k-1})v_{k-1}^{\otimes 2} = \frac{1}{2}D^2\phi(x_{k-2})v_{k-2}^{\otimes 2} + \frac{1}{2}(\gamma_{k-1}-\gamma_{k-1}^2)D^3\phi(x_{k-2})v_{k-2}^{\otimes 3} \\ - \frac{u}{4}\gamma_{k-1}^2\langle D^3\phi(x_{k-2}); v_{k-2}^{\otimes 2}, \nabla f(x_{k-2})\rangle + \frac{\sqrt{u}}{2}\gamma_{k-1}^{\frac{3}{2}}\langle D^3\phi(x_{k-2}); v_{k-2}^{\otimes 2}, U_{k-1}^{(1)}\rangle \\ - 2(\gamma_{k-1}-\gamma_{k-1}^2)D^2\phi(x_{k-2})v_{k-2}^{\otimes 2} - u(\gamma_{k-1}-\gamma_{k-1}^2)\langle D^2\phi(x_{k-2}); \nabla f(x_{k-2}), v_{k-2}\rangle \\ + 2\sqrt{u}\gamma_{k-1}^{\frac{1}{2}}\langle D^2\phi(x_{k-2}); v_{k-2}, U_{k-1}^{(2)}\rangle + 2\gamma_{k-1}^2 D^2\phi(x_{k-2})v_{k-2}^{\otimes 2} \\ + \frac{u^2}{2}\gamma_{k-1}^2 D^2\phi(x_{k-2})\nabla f(x_{k-2})^{\otimes 2} + 2u(\gamma_{k-1}-2\gamma_{k-1}^2)D^2\phi(x_{k-2}){U_{k-1}^{(2)}}^{\otimes 2} \\ + 2u\gamma_{k-1}^2\langle D^2\phi(x_{k-2}); \nabla f(x_{k-2}), v_{k-2}\rangle - 4\sqrt{u}\gamma_{k-1}^{\frac{3}{2}}\langle D^2\phi(x_{k-2}); v_{k-2}, U_{k-1}^{(2)}\rangle\end{aligned}$$

$$
\begin{aligned}
&- 2u^{\frac{3}{2}}\gamma_{k-1}^{\frac{3}{2}}\langle D^2\phi(x_{k-2});\nabla f(x_{k-2}),U_{k-1}^{(2)}\rangle + \frac{1}{4}\gamma_{k-1}^2 D^4\phi(x_{k-2})v_{k-2}^{\otimes 4}\\
&- \gamma_{k-1}^2 D^3\phi(x_{k-2})v_{k-2}^{\otimes 3} - \frac{u}{2}\gamma_{k-1}^2\langle D^3\phi(x_{k-2});v_{k-2}^{\otimes 2},\nabla f(x_{k-2})\rangle\\
&+ \frac{1}{2}\langle D^3\phi(x_{k-2});v_{k-2},\sqrt{u}\sigma_{k-1}^{(1)}U_{k-1}^{(1)},2\sqrt{u}\sigma_{k-1}^{(2)}U_{k-1}^{(12}\rangle\\
&+ \frac{2u}{3}\gamma_{k-1}^2\langle D^3\phi(x_{k-1});v_{k-2};{U_{k-1}^{(2)}}^{\otimes 2}\rangle + o(\gamma_k^2)
\end{aligned}
$$

Take the limits and we obtain:

$$
\begin{aligned}
\lim_n \frac{\sum_{k=2}^n \gamma_k^2\nabla f(x_{k-1})\cdot\nabla\phi(x_{k-1})}{\Gamma_n^{(4)}} = \lim_n \frac{1}{\Gamma_n^{(4)}}\Bigg[&\sum_{k=2}^n \gamma_{k-1}^2\nabla f(x_{k-2})\cdot\nabla\phi(x_{k-2})\\
&+\sum_{k=2}^n \gamma_{k-1}^3\langle D^2 f(x_{k-2});\nabla\phi(x_{k-2}),v_{k-2}\rangle\\
&+\sum_{k=2}^n \gamma_{k-1}^3\langle D^2\phi(x_{k-2});\nabla f(x_{k-2}),v_{k-2}\rangle\Bigg]\\
&-\frac{u}{2}\int_{\mathbb{R}^d}\langle D^2 f(x);\nabla\phi(x),\nabla f(x)\rangle\pi(dx)\\
&-\frac{u}{2}\int_{\mathbb{R}^d} D^2\phi(x)\nabla f(x)^{\otimes 2}\pi(dx)\\
+\frac{1}{2}\int_{\mathbb{R}^d}&\left(D^3 f(x)\nabla\phi(x) + 2D^2 f(x)D^2\phi(x) + D^3\phi(x)\nabla f(x)\right)v^{\otimes 2}\nu(dx,dv)
\end{aligned}
$$

$$
\begin{aligned}
\lim_n \frac{1}{2}\frac{\sum_{k=2}^n \gamma_k^2 D^2\phi(x_{k-1})v_{k-1}^{\otimes 2}}{\Gamma_n^{(4)}} = \lim_n \Bigg\{&\frac{1}{\Gamma_n^{(4)}}\sum_{k=2}^n \frac{1}{2}\gamma_{k-1}^2 D^2\phi(x_{k-2})v_{k-2}^{\otimes 2}\\
&+\frac{1}{\Gamma_n^{(4)}}\sum_{k=2}^n \gamma_{k-1}^3\Bigg[\frac{1}{2}D^3\phi(x_{k-2})v_{k-2}^{\otimes 3} - 2D^2\phi(x_{k-2})v_{k-2}^{\otimes 2}\\
&-u\langle D^2\phi(x_{k-2});\nabla f(x_{k-2}),v_{k-2}\rangle + 2uD^2\phi(x_{k-2}){U_{k-1}^{(2)}}^{\otimes 2}\Bigg]\Bigg\}\\
&-\frac{3u}{4}\int_{\mathbb{R}^{2d}}\langle D^3\phi(x);\nabla f(x),v^{\otimes 2}\rangle\nu(dx,dv) + \frac{1}{4}\int_{\mathbb{R}^{2d}} D^4\phi(x)v^{\otimes 4}\pi(dx)\\
&+\frac{u^2}{2}\int_{\mathbb{R}^d} D^2\phi(x)\nabla f(x)^{\otimes 2}\pi(dx)
\end{aligned}
$$

Claim:

a) $\lim_n \frac{1}{\Gamma_n^{(4)}}\sum_{k=1}^n \gamma_k^2\nabla\phi(x_{k-1})\cdot v_{k-1} = 0.$
b) $\lim_n \frac{1}{\Gamma_n^{(4)}}\sum_{k=1}^n \gamma_k^3(\frac{u}{2}\nabla\phi(x_{k-1})\cdot\nabla f(x_{k-1}) - \frac{1}{2}D^2\phi(x_{k-1})v_{k-1}^{\otimes 2}) = 0.$
c) $\lim_n \frac{1}{\Gamma_n^{(4)}}\sum_{k=1}^n \gamma_k^3\nabla\phi(x_{k-1})\cdot v_{k-1} = 0.$

We'll prove the **Claim** at the end of our proof. We can use the **Claim** and our expansion of $Z_n^{(2)}$ to find the following relation:

$$
\begin{aligned}
\lim_n \frac{1}{\Gamma_n^{(4)}}Z_n^{(2)} = &\lim_n \frac{1}{\Gamma_n^{(4)}}\sum_{k=2}^n \gamma_k^2\nabla\phi(x_{k-1})\cdot v_{k-1}\\
&+\lim_n \frac{1}{\Gamma_n^{(4)}}\sum_{k=2}^n \gamma_k^2\left[\frac{u}{2}\nabla f(x_{k-1})\cdot\nabla\phi(x_{k-1}) - \frac{1}{2}D^2\phi(x_{k-1})v_{k-1}^{\otimes 2}\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=2}^n \gamma_{k-1}^2 \left[\frac{u}{2} \nabla f(x_{k-2}) \cdot \nabla \phi(x_{k-2}) - \frac{1}{2} D^2 \phi(x_{k-2}) v_{k-2}^{\otimes 2} \right] \\
&+ \lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=2}^n \gamma_{k-1}^3 \left[\frac{u}{2} \langle D^2 f(x_{k-2}); \nabla \phi(x_{k-2}), v_{k-2} \rangle + \frac{3u}{2} \langle D^2 \phi(x_{k-2}); \nabla f(x_{k-2}), v_{k-2} \rangle \right. \\
&\qquad \left. - \frac{1}{2} D^3 \phi(x_{k-2}) v_{k-2}^{\otimes 3} + 2 D^2 \phi(x_{k-2}) v_{k-2}^{\otimes 2} - 2u \Delta \phi(x_{k-2}) \right] \\
&- \frac{u^2}{4} \int_{\mathbb{R}^d} \langle D^2 f(x); \nabla \phi(x), \nabla f(x) \rangle \pi(dx) - \frac{u^2}{4} \int_{\mathbb{R}^d} D^2 \phi(x) \nabla f(x)^{\otimes 2} \pi(dx) \\
&+ \frac{u}{4} \int_{\mathbb{R}^d} \left(D^3 f(x) \nabla \phi(x) + 2 D^2 f(x) D^2 \phi(x) + D^3 \phi(x) \nabla f(x) \right) v^{\otimes 2} \nu(dx, dv) \\
&+ \frac{3u}{4} \int_{\mathbb{R}^{2d}} \langle D^3 \phi(x); \nabla f(x), v^{\otimes 2} \rangle \nu(dx, dv) - \frac{1}{4} \int_{\mathbb{R}^{2d}} D^4 \phi(x) v^{\otimes 4} \pi(dx) \\
&- \frac{u^2}{2} \int_{\mathbb{R}^d} D^2 \phi(x) \nabla f(x)^{\otimes 2} \pi(dx) \\
&= \lim_n \frac{1}{\Gamma_n^{(4)}} [Z_n^{(2)} + 3 Z_n^{(3)}] - \frac{u^2}{4} \int_{\mathbb{R}^d} \langle D^2 f(x); \nabla \phi(x), \nabla f(x) \rangle \pi(dx) \\
&- \frac{3u^2}{4} \int_{\mathbb{R}^d} D^2 \phi(x) \nabla f(x)^{\otimes 2} \pi(dx) - \frac{1}{4} \int_{\mathbb{R}^{2d}} D^4 \phi(x) v^{\otimes 4} \pi(dx) \\
&+ \frac{u}{4} \int_{\mathbb{R}^d} \left(D^3 f(x) \nabla \phi(x) + 2 D^2 f(x) D^2 \phi(x) + 4 D^3 \phi(x) \nabla f(x) \right) v^{\otimes 2} \nu(dx, dv)
\end{aligned}
$$

The last identity follows from **Claim**-a),b) and the fact that $\lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=1}^n \gamma_k^3 \langle D^2 f(x_{k-1}); \nabla \phi(x_{k-1}), v_{k-1} \rangle = 0$. To prove $\lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=1}^n \gamma_k^3 \langle D^2 f(x_{k-1}); \nabla \phi(x_{k-1}), v_{k-1} \rangle = 0$, we can assume ψ is a new test function satisfying $\nabla \psi(x) = D^2 f(x) \nabla \phi(x)$. Then the statement follows from **Claim**-c). This could be done because ψ satisfies the all assumptions on ϕ stated in the theorem. Therefore we obtain

$$
\begin{aligned}
\lim_n \frac{1}{\Gamma_n^{(4)}} Z_n^{(3)} &= \frac{u^2}{12} \int_{\mathbb{R}^d} \langle D^2 f(x); \nabla \phi(x), \nabla f(x) \rangle \pi(dx) + \frac{u^2}{4} \int_{\mathbb{R}^d} D^2 \phi(x) \nabla f(x)^{\otimes 2} \pi(dx) \\
&- \frac{u}{12} \int_{\mathbb{R}^d} \left(D^3 f(x) \nabla \phi(x) + 2 D^2 f(x) D^2 \phi(x) + 4 D^3 \phi(x) \nabla f(x) \right) v^{\otimes 2} \nu(dx, dv) \\
&+ \frac{1}{12} \int_{\mathbb{R}^{2d}} D^4 \phi(x) v^{\otimes 4} \pi(dx)
\end{aligned}
$$

Combine with our previous results on $2Z_n^{(2)} + 3Z_n^{(3)}$ and we obtain

$$
\begin{aligned}
\lim_n \frac{1}{\Gamma_n^{(4)}} [Z_n^{(2)} + Z_n^{(3)}] &= \lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=1}^n \sqrt{u} \gamma_k^{\frac{3}{2}} \nabla \phi(x_{k-1}) \cdot U_k^{(2)} + \frac{u^2}{8} \int_{\mathbb{R}^d} D^2 \phi(x) \nabla f(x)^{\otimes 2} \pi(dx) \\
&- \frac{u}{12} \int_{\mathbb{R}^{2d}} \langle D^3 \phi(x); \nabla f(x), v^{\otimes 2} \rangle \nu(dx, dv) + \frac{u}{24} \int_{\mathbb{R}^{2d}} \langle D^3 f(x); \nabla \phi(x), v^{\otimes 2} \rangle \nu(dx, dv) \\
&+ \frac{u}{12} \int_{\mathbb{R}^{2d}} (D^2 f D^2 \phi)(x) v^{\otimes 2} \nu(dx, dv) - \frac{u^2}{24} \int_{\mathbb{R}^d} \langle D^2 f(x); \nabla \phi(x), \nabla f(x) \rangle \pi(dx) \\
&- \frac{1}{24} \int_{\mathbb{R}^{2d}} D^4 \phi(x) v^{\otimes 4} \pi(dx)
\end{aligned}
$$

Then we plug this result in our original decomposition:

$$
\lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=1}^n \gamma_k \mathcal{L} \phi(x_{k-1}) = \lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=1}^n \gamma_k^{\frac{3}{2}} \frac{2\sqrt{3}}{3} \nabla \phi(x_{k-1}) \cdot (\sqrt{u} U_k^{(1)})
$$

$$
\begin{aligned}
&+ \lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=1}^n \sqrt{u}\gamma_k^{\frac{3}{2}} \nabla\phi(x_{k-1}) \cdot U_k^{(2)} + \frac{u^2}{8} \int_{\mathbb{R}^d} D^2\phi(x) \nabla f(x)^{\otimes 2} \pi(dx) \\
&- \frac{u}{12} \int_{\mathbb{R}^{2d}} \langle D^3\phi(x); \nabla f(x), v^{\otimes 2} \rangle \nu(dx, dv) + \frac{u}{24} \int_{\mathbb{R}^{2d}} \langle D^3 f(x); \nabla\phi(x), v^{\otimes 2} \rangle \nu(dx, dv) \\
&+ \frac{u}{12} \int_{\mathbb{R}^{2d}} (D^2 f D^2 \phi)(x) v^{\otimes 2} \nu(dx, dv) - \frac{u^2}{24} \int_{\mathbb{R}^d} \langle D^2 f(x); \nabla\phi(x), \nabla f(x) \rangle \pi(dx) \\
&- \frac{1}{24} \int_{\mathbb{R}^{2d}} D^4\phi(x) v^{\otimes 4} \pi(dx) + \frac{u}{4} \int_{\mathbb{R}^{2d}} \langle D^3\phi(x); \nabla f(x), v^{\otimes 2} \rangle \nu(dx, dv) \\
&- \frac{u^2}{8} \int_{\mathbb{R}^d} D^2\phi(x) \nabla f(x)^{\otimes 2} \pi(dx) - \frac{1}{24} \int_{\mathbb{R}^{2d}} D^4\phi(x) v^{\otimes 4} \nu(dx, dv) \\
&= \lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=1}^n \gamma_k^{\frac{3}{2}} \nabla\phi(x_{k-1}) \cdot (\frac{2\sqrt{3}}{3} \sqrt{u} U_k^{(1)} + \frac{1}{2} 2\sqrt{u} U_k^{(2)}) \\
&+ \frac{u}{6} \int_{\mathbb{R}^{2d}} \langle D^3\phi(x); \nabla f(x), v^{\otimes 2} \rangle \nu(dx, dv) + \frac{u}{24} \int_{\mathbb{R}^{2d}} \langle D^3 f(x); \nabla\phi(x), v^{\otimes 2} \rangle \nu(dx, dv) \\
&+ \frac{u}{12} \int_{\mathbb{R}^{2d}} (D^2\phi D^2 f)(x) v^{\otimes 2} \nu(dx, dv) - \frac{1}{12} \int_{\mathbb{R}^{2d}} D^4\phi(x) v^{\otimes 4} \nu(dx, dv) \\
&- \frac{u^2}{24} \int_{\mathbb{R}^d} \langle D^2 f(x); \nabla\phi(x), \nabla f(x) \rangle \pi(dx)
\end{aligned}
$$

It remains to determine the normal limit. Since $(U_k^{(1)}, U_k^{(2)})$ is Gaussian in $\mathbb{R}^{2d}$ with mean zero and covariance matrix $\frac{1+e^{-4\gamma_k}-2e^{-2\gamma_k}}{4\sigma_k^{(1)}\sigma_k^{(2)}} I_d$, we can find the distribution of $U_k := (\frac{2\sqrt{3}}{3}\sqrt{u}U_k^{(1)} + \frac{1}{2}2\sqrt{u}U_k^{(2)})$. $\{U_k\}$ are independent $2d$-Gaussian Random vectors with $U_k \sim \mathcal{N}(0, \Sigma_k)$, where

$$
\begin{aligned}
\Sigma_k &= \mathbb{E}[(\frac{2\sqrt{3}}{3}\sqrt{u}U_k^{(1)} + \frac{1}{2}2\sqrt{u}U_k^{(2)})^T (\frac{2\sqrt{3}}{3}\sqrt{u}U_k^{(1)} + \frac{1}{2}2\sqrt{u}U_k^{(2)})] \\
&= \frac{4u}{3} I_d + \frac{4u\sqrt{3}}{3} \frac{1 + e^{-4\gamma_k} - 2e^{-2\gamma_k}}{4\sigma_k^{(1)}\sigma_k^{(2)}} I_d + uI_d \\
&\sim \frac{10}{3} u I_d + O(\gamma_k) I_d
\end{aligned}
$$

Apply our weak convergence result and CLT for arrays of square-integrable martingale increments, we have that when $0 < \hat{\gamma} < +\infty$:

$$
\frac{1}{\Gamma_n^{(4)}} \sum_{k=1}^n \gamma_k^{\frac{3}{2}} \nabla\phi(x_{k-1}) \cdot U_k \implies \mathcal{N}(0, \sigma^2)
$$

where

$$
\sigma^2 = \lim_n \frac{1}{{\Gamma_n^{(4)}}^2} \sum_{k=1}^n \gamma_k^3 |\nabla\phi(x_{k-1})|^2 (\frac{10}{3}u + O(\gamma_k)) = \frac{10}{3} u \hat{\gamma}^{-2} \int_{\mathbb{R}^d} |\nabla\phi(x)|^2 \pi(dx)
$$

In conclusion, when $\hat{\gamma} \in (0, +\infty)$:

$$
\frac{\Gamma_n}{\Gamma_n^{(4)}} \nu_n^\gamma(\mathcal{L}\phi) \implies \mathcal{N}(\rho, \frac{10}{3} u \hat{\gamma}^{-2} \int_{\mathbb{R}^d} |\nabla\phi(x)|^2 \pi(dx))
$$

where

$$
\begin{aligned}
\rho = &\frac{u}{6} \int_{\mathbb{R}^{2d}} \langle D^3\phi(x); \nabla f(x), v^{\otimes 2} \rangle \nu(dx, dv) + \frac{u}{24} \int_{\mathbb{R}^{2d}} \langle D^3 f(x); \nabla\phi(x), v^{\otimes 2} \rangle \nu(dx, dv) \\
&+ \frac{u}{12} \int_{\mathbb{R}^{2d}} (D^2\phi D^2 f)(x) v^{\otimes 2} \nu(dx, dv) - \frac{1}{12} \int_{\mathbb{R}^{2d}} D^4\phi(x) v^{\otimes 4} \nu(dx, dv)
\end{aligned}
$$

$$-\frac{u^2}{24}\int_{\mathbb{R}^d}\langle D^2 f(x);\nabla\phi(x),\nabla f(x)\rangle\pi(dx)$$

When $\hat{\gamma}=0$,

$$\frac{\Gamma_n}{\sqrt{\Gamma_n^{(3)}}}\nu_n^\gamma(\mathcal{L}\phi)\implies\mathcal{N}(0,\frac{10}{3}u\int_{\mathbb{R}^d}|\nabla\phi(x)|^2\pi(dx))$$

When $\hat{\gamma}=+\infty$,

$$\frac{1}{\Gamma_n^{(4)}}\sum_{k=1}^n\gamma_k^{\frac{3}{2}}\nabla\phi(x_{k-1})\cdot(\frac{2\sqrt{3}}{3}\sqrt{u}U_k^{(1)}+\frac{1}{2}2\sqrt{u}U_k^{(2)})\to 0\quad\text{ in probability}$$

Therefore when $\hat{\gamma}=+\infty$,

$$\frac{\Gamma_n}{\Gamma_n^{(4)}}\nu_n^\gamma(\mathcal{L}\phi)\to\rho\qquad\text{in probability}$$

Proof of the claim: First we'll show that $\frac{1}{\Gamma_n^{(3)}}\sum_{k=1}^n\gamma_k^2\mathcal{L}\phi(x_{k-1})\to 0$. We can use our decomposition of $\mathcal{L}\phi(x_{k-1})$ and obtain:

$$\begin{aligned}\sum_{k=1}^n\gamma_k^2\mathcal{L}\phi(x_{k-1})=\sum_{k=1}^n\Big\{&\gamma_k\left(\phi(x_k)-\phi(x_{k-1})\right)+\gamma_k^3 v_{k-1}\cdot\nabla\phi(x_{k-1})\\&+\frac{u}{2}\gamma_k^3\nabla f(x_{k-1})\cdot\nabla\phi(x_{k-1})-\frac{1}{2}\gamma_k^3 D^2\phi(x_{k-1})v_{k-1}^{\otimes 2}\Big\}\end{aligned}$$

Since $\gamma_{k-1}-\gamma_k\sim o(\gamma_k^4)$ and $\{\phi(x_n)\}$ is tight, $\frac{1}{\Gamma_n^{(3)}}\sum_{k=1}^n\gamma_k\left(\phi(x_k)-\phi(x_{k-1})\right)\to 0$. Then we can apply theorem 9 with $p=6,s=1$ and obtain

$$\begin{aligned}\frac{1}{\Gamma_n^{(3)}}\sum_{k=1}^n\gamma_k^2\mathcal{L}\phi(x_{k-1})\to&\int_{\mathbb{R}^{2d}}v\cdot\nabla\phi(x)\nu(dx,dv)+\frac{u}{2}\int_{\mathbb{R}^d}\nabla\phi(x)\cdot\nabla f(x)\pi(dx)\\&-\frac{1}{2}\int_{\mathbb{R}^{2d}}D^2\phi(x)v^{\otimes 2}\nu(dx,dv)\\&=0\end{aligned}$$

The last identity follows from integration by parts and Fubini theorem. In the same way, we can also prove $\frac{1}{\Gamma_n^{(4)}}\sum_{k=1}^n\gamma_k^3\mathcal{L}\phi(x_{k-1})\to 0$.

Next, we'll show $\lim_n\frac{1}{\Gamma_n^{(3)}}\sum_{k=1}^n\gamma_k^2(\frac{u}{2}\nabla\phi(x_{k-1})\cdot\nabla f(x_{k-1})-\frac{1}{2}D^2\phi(x_{k-1})v_{k-1}^{\otimes 2})=0$, we'll use the same trick as we did in the proof of theorem 5. We Taylor expand $\mathcal{L}\phi(x_{k-1})$ at (x_{k-2},v_{k-2}):

$$\begin{aligned}\gamma_k^2\mathcal{L}\phi(x_{k-1})=&\gamma_k^2\mathcal{L}\phi(x_{k-2})+\gamma_k^2(\gamma_{k-1}-\gamma_{k-1}^2)D^2\phi(x_{k-2})v_{k-2}^{\otimes 2}\\&-\frac{u}{2}\gamma_k^2\gamma_{k-1}^2\langle D^2\phi(x_{k-2});v_{k-2},\nabla f(x_{k-2})\rangle\\&-\gamma_k^2(2\gamma_{k-1}-2\gamma_{k-1}^2)v_{k-2}\cdot\nabla\phi(x_{k-2})\\&-u\gamma_k^2(\gamma_{k-1}-\gamma_{k-1}^2)\nabla f(x_{k-2})\cdot\nabla\phi(x_{k-2})\\&+\frac{1}{2}\gamma_k^2\gamma_{k-1}^2D^3\phi(x_{k-2})v_{k-2}^{\otimes 3}-2\gamma_k^2\gamma_{k-2}^2D^2\phi(x_{k-2})v_{k-2}^{\otimes 2}\\&-u\gamma_k^2\gamma_{k-1}^2\langle D^2\phi(x_{k-2});v_{k-2},\nabla f(x_{k-2})\rangle\\&+\gamma_k^2\langle D^2\phi(x_{k-2});\sqrt{u}\sigma_{k-1}^{(1)}U_{k-1}^{(1)},2\sqrt{u}\sigma_{k-1}^{(2)}U_{k-1}^{(2)}\rangle+o(\gamma_{k-1}^3)\end{aligned}$$

Since $\gamma_{k-1}-\gamma_k=o(\gamma_k^4)$, we can change γ_k on the left hand side to γ_{k-1} when we take limits with scale $\Gamma_n^{(4)}$. Apply theorem 9 with $p=8,s=1$ to terms with order $o(\gamma_k^3)$-coefficients.

$$\lim_n\frac{1}{\Gamma_n^{(4)}}\sum_{k=2}^n\gamma_k^2\mathcal{L}\phi(x_{k-1})=\lim_n\frac{1}{\Gamma_n^{(4)}}\sum_{k=2}^n\gamma_{k-1}^2\mathcal{L}\phi(x_{k-2})-2\lim_n\frac{1}{\Gamma_n^{(4)}}\sum_{k=2}^n\gamma_{k-1}^3\mathcal{L}\phi(x_{k-2})$$

$$
\begin{aligned}
&- 2\lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=2}^n \gamma_{k-1}^3 (\frac{u}{2} \nabla\phi(x_{k-2}) \cdot \nabla f(x_{k-2}) - \frac{1}{2} D^2\phi(x_{k-2}) v_{k-2}^{\otimes 2}) \\
&- 3\int_{\mathbb{R}^{2d}} D^2\phi(x) v^{\otimes 2} \nu(dx, dv) + u \int_{\mathbb{R}^d} \nabla\phi(x) \cdot \nabla f(x) \pi(dx) \\
&+ \lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=2}^n \gamma_{k-1}^2 \langle D^2\phi(x_{k-2}); \sqrt{u}\sigma_{k-1}^{(1)} U_{k-1}^{(1)}, 2\sqrt{u}\sigma_{k-1}^{(2)} U_{k-1}^{(2)} \rangle
\end{aligned}
$$

Since we proved $\frac{1}{\Gamma_n^{(4)}} \sum_{k=1}^n \gamma_k^3 \mathcal{L}\phi(x_{k-1}) \to 0$ and from Theorem 5, we've shown that

$$
\lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=2}^n \gamma_{k-1}^2 \langle D^2\phi(x_{k-2}); \sqrt{u}\sigma_{k-1}^{(1)} U_{k-1}^{(1)}, 2\sqrt{u}\sigma_{k-1}^{(2)} U_{k-1}^{(2)} \rangle = 2u \int_{\mathbb{R}^d} \Delta\phi(x) \pi(dx)
$$

We obtain

$$
\begin{aligned}
&\lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=2}^n \gamma_{k-1}^2 (\frac{u}{2} \nabla\phi(x_{k-2}) \cdot \nabla f(x_{k-2}) - \frac{1}{2} D^2\phi(x_{k-2}) v_{k-2}^{\otimes 2}) \\
&= \frac{1}{2} \left[\lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=2}^n \gamma_k^2 \mathcal{L}\phi(x_{k-1}) - \lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=2}^n \gamma_{k-1}^2 \mathcal{L}\phi(x_{k-2}) \right] \\
&= 0
\end{aligned}
$$

Therefore, $\lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=1}^n \gamma_k^3 (\frac{u}{2} \nabla\phi(x_{k-1}) \cdot \nabla f(x_{k-1}) - \frac{1}{2} D^2\phi(x_{k-1}) v_{k-1}^{\otimes 2}) = 0$.
To prove the **Claim**, we need to use the decomposition again:

$$
\begin{aligned}
\sum_{k=1}^n \gamma_k^2 \mathcal{L}\phi(x_{k-1}) = \sum_{k=1}^n \Big\{ &\gamma_k [\phi(x_k) - \phi(x_{k-1})] + (\gamma_k^3 - \frac{2}{3}\gamma_k^4) v_{k-1} \cdot \nabla\phi(x_{k-1}) \\
&+ \frac{u}{2} (\gamma_k^3 - \frac{2}{3}\gamma_k^4) \nabla f(x_{k-1}) \cdot \nabla\phi(x_{k-1}) \\
&- \frac{1}{2} (\gamma_k^3 - 2\gamma_k^4) D^2\phi(x_{k-1}) v_{k-1}^{\otimes 2} \\
&- \frac{2u}{3} \gamma_k^4 D^2\phi(x_{k-1}) U_k^{(1)\otimes 2} \\
&+ \frac{u}{2} \gamma_k^4 \langle D^2\phi(x_{k-1}); v_{k-1}, \nabla f(x_{k-1}) \rangle \\
&- \frac{1}{6} \gamma_k^4 D^3\phi(x_{k-1}) v_{k-1}^{\otimes 3} + o(\gamma_k^4) \Big\}
\end{aligned}
$$

Since $\{\phi(x_n)\}$ is tight and $\gamma_{k-1} - \gamma_k = o(\gamma_k^4)$, we have $\frac{1}{\Gamma_n^{(4)}} \sum_{k=1}^n \gamma_k (\phi(x_k) - \phi(x_{k-1})) \to 0$. For the terms with coefficients of order γ_k^3, we can apply theorem 9 with $p = 8, s = 1$. Then we obtain:

$$
\begin{aligned}
\lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=1}^n \gamma_k^2 \mathcal{L}\phi(x_{k-1}) &= \lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=1}^n \gamma_k^3 (\mathcal{L}\phi(x_{k-1}) + \frac{u}{2} \nabla\phi(x_{k-1}) \cdot \nabla f(x_{k-1}) - \frac{1}{2} D^2\phi(x_{k-1}) v_{k-1}^{\otimes 2}) \\
&- \frac{u}{3} \int_{\mathbb{R}^d} \nabla\phi(x) \cdot \nabla f(x) \pi(dx) + \int_{\mathbb{R}^{2d}} D^2\phi(x) v^{\otimes 2} \nu(dx, dv) \\
&- \frac{2u}{3} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} D^2\phi(x) z^{\otimes 2} \mu(dz) \pi(dx) \\
&= \lim_n \frac{1}{\Gamma_n^{(4)}} \sum_{k=1}^n \gamma_k^3 (\mathcal{L}\phi(x_{k-1}) + \frac{u}{2} \nabla\phi(x_{k-1}) \cdot \nabla f(x_{k-1}) - \frac{1}{2} D^2\phi(x_{k-1}) v_{k-1}^{\otimes 2}) \\
&= 0
\end{aligned}
$$

The second identity follows from integration by parts and Fubini theorem. The last identity follows from the two statements we just proved. □