COMPLEXITY OF RANDOMIZED ALGORITHMS FOR UNDERDAMPED LANGEVIN DYNAMICS[*]

YU CAO[†], JIANFENG LU[‡], AND LIHAN WANG[§]

Abstract. We establish an information complexity lower bound of randomized algorithms for simulating underdamped Langevin dynamics. More specifically, we prove that the worst L^2 strong error is of order $\Omega(\sqrt{d}N^{-3/2})$, for solving a family of d-dimensional underdamped Langevin dynamics, by any randomized algorithm with only N queries to ∇U, the driving Brownian motion and its weighted integration, respectively. The lower bound we establish matches the upper bound for the randomized midpoint method recently proposed by Shen and Lee [NIPS 2019], in terms of both parameters N and d.

1. Introduction

The underdamped Langevin dynamics have been widely used to sample high-dimensional probability distributions [38,43,57], as it could provide a faster convergence rate compared to the overdamped Langevin dynamics. The analysis of the sampling algorithms based on underdamped Langevin dynamics consists of two key aspects:

(i) the mixing time of continuous-time underdamped Langevin dynamics;

(ii) the time-discretization error for numerically integrating underdamped Langevin dynamics.

The first question has been widely studied for various metrics of convergence, see, e.g., [3,7,9,43,49].

Our focus in this work is the performance of discretization algorithms for underdamped Langevin dynamics. This has also been quite extensively studied, in terms of both asymptotic analysis [13,34,42,55] and non-asymptotic analysis [4,5,7,32,50]. The algorithm with the best rate up to date was proposed by Shen and Lee [50]. Their randomized midpoint method (RMM) for underdamped Langevin dynamics has a strong L^2 error $\mathcal{O}(\sqrt{d}N^{-3/2})$ using only N gradient queries, where d is the dimension. On the other hand, it is not clear yet from the literature what error rate an optimal algorithm can achieve. In other words, what the intrinsic difficulty of numerical integration of underdamped Langevin dynamics is. This paper provides an answer in the framework of *information-based complexity* (IBC). In particular, we show that the randomized midpoint method is order optimal with respect to both d and N.

Information-based complexity [41,53], which is closely related to the notion of the information-theoretic lower bound, studies the intrinsic complexity of a family of com-

putational problems, based on the type of queries that one has, rather than focusing on a particular algorithm for the task. Intuitively, the algorithmic performance would depend on the information one could acquire (for example, the gradients of the potential function U for Langevin dynamics). IBC aims to establish a lower bound of the accuracy of a family of algorithms, provided the amount and the type of information.

In this work, we adopt the framework of IBC to study randomized algorithms for approximating the strong solution of the underdamped Langevin dynamics with gradient queries to strongly convex potentials and also the driving Brownian motion.

1.1. Underdamped Langevin dynamics.   We consider the following underdamped Langevin dynamics $(X_t, V_t) \in \mathbb{R}^d \times \mathbb{R}^d$ (we adopt the parameter scaling used in [5, 50], which is slightly different from the usual physical model of underdamped Langevin dynamics)

$$\begin{aligned}
\mathrm{d}X_t &= V_t\,\mathrm{d}t, \\
\mathrm{d}V_t &= -2V_t\,\mathrm{d}t - \frac{1}{L}\nabla U(X_t)\,\mathrm{d}t + \frac{2}{\sqrt{L}}\,\mathrm{d}W_t,
\end{aligned} \tag{1.1}$$

on the time interval $[0, T]$, with the fixed initial condition $X_0 = x^\star$ and $V_0 = 0$, where $x^\star \equiv x^\star(U) \in \mathbb{R}^d$ is a local minimum of the potential function U; W_t is the d-dimensional standard Brownian motion; the parameter $L > 0$ has a physical meaning as the mass of the particle. The unique stationary distribution of (1.1) is $\rho_\infty(x, v) \propto \exp\left(-U(x) - \frac{|v|^2}{2/L}\right)$. As time $t \to \infty$, the distribution of (X_t, V_t) converges to the equilibrium exponentially fast under mild conditions; see, e.g., [3, 9, 10, 33, 43, 49, 54]. Generalization of our main result (Theorem 1.1) below to underdamped Langevin dynamics with general friction coefficient is straightforward, and we will not pursue such generality herein for simplicity.

ASSUMPTION 1.1.   *In this work, we shall only consider strongly convex U with Lipschitz gradient, i.e., we consider the following family of potential functions,*

$$\mathcal{F} \equiv \mathcal{F}(d, \ell, L) := \left\{ U \in C^2(\mathbb{R}^d) \;\middle|\; \ell\mathbf{I}_d \le \nabla^2 U(x) \le L\mathbf{I}_d, \;\forall x \in \mathbb{R}^d \right\} \tag{1.2}$$

for fixed parameters $0 < \ell < L < \infty$. The notation $\mathbf{I}_d$ herein means $d \times d$ identity matrix. Under the strong convexity assumption, we know that $x^\star$ is uniquely determined by U. The condition number $\varkappa$ is defined as $\varkappa := L/\ell$.

1.2. Main results.   Denote the probability space for underdamped Langevin dynamics (1.1) as $(\mathbb{M}, \Sigma, \mathbb{P})$. In the context of Langevin sampling, we assume that, besides the Brownian motion, the query to a weighted Brownian motion $\widetilde{W}_t^{(\theta)}$ is also admissible, where

$$\widetilde{W}_t^{(\theta)}(\omega) := \int_0^t e^{\theta s}\,\mathrm{d}W_s(\omega), \qquad \forall\, \omega \in \mathbb{M}. \tag{1.3}$$

When $\theta = 0$, $\widetilde{W}_t^{(\theta=0)}(\omega) \equiv W_t(\omega)$. In general, we define a correlated Gaussian process $\widetilde{W}_t^{\vec{\theta}}$ for $\vec{\theta} = (\theta_1, \theta_2, \cdots, \theta_\mathsf{J})$:

$$\widetilde{W}_t^{\vec{\theta}}(\omega) := \left(\widetilde{W}_t^{(\theta_1)}(\omega) \;\cdots\; \widetilde{W}_t^{(\theta_j)}(\omega) \;\cdots\; \widetilde{W}_t^{(\theta_\mathsf{J})}(\omega) \right),$$

as a short-hand notation. In particular, we shall use $\widetilde{W}_t^{(0,2)} \equiv (W_t, \widetilde{W}_t^{(2)})$ frequently below, as in our main theorem. The reason for such an assumption is that in the context

of Langevin sampling, generating correlated Gaussian random vectors is not computationally expensive, whereas computing ∇U is usually the computational bottleneck.

More formally, we assume that there is an oracle query $\Upsilon_{U,\omega} : \mathbb{R}^d \times [0,T] \to \mathbb{R}^d \times \mathbb{R}^{2d}$, defined as

$$\Upsilon_{U,\omega}(x,t) := \left(\nabla U(x), \widetilde{W}_t^{(0,2)}(\omega) \right), \ x \in \mathbb{R}^d, \ t \in [0,T],$$

for any $U \in \mathcal{F}$ and $\omega \in \mathbb{M}$, and the set of *admissible information operations* is defined by

$$\Lambda := \left\{ \Upsilon_{U,\omega}(x,t) \middle| \ x \in \mathbb{R}^d, \ t \in [0,T] \right\}. \tag{1.4}$$

Our main result is the following information-based complexity bound for solving the underdamped Langevin dynamics with $U \in \mathcal{F}$.

THEOREM 1.1 (Information-based complexity with queries to ∇U and weighted Brownian motions). *Consider the complexity problem $\mathcal{F}$ (1.2) with Λ as the set of admissible information operations. Then whenever $N \geq N_0$ for some integer N_0 (independent of the dimension d),*

$$C_{low}\sqrt{d}N^{-3/2} \lesssim \inf_{A \in \mathcal{A}_N^{Rand}} e_{\mathcal{F},\Lambda}(A) \lesssim C_{up}\sqrt{d}N^{-3/2}, \tag{1.5}$$

where the prefactor $C_{up} = \sqrt{\frac{T^3}{\ell} + \frac{T^4}{L}}$, and $C_{low} \equiv C_{low}(\ell, L, T)$ can be chosen as

$$C_{low} = \sup_{\substack{C_x > 0, \ C_v > 0, \\ \ell < u < u_R \leq L}} \sqrt{\mathsf{P}(C_x, C_v, u, T)} C_x^2 \min\{u - \ell, u_R - u\} \overline{C}(C_x, C_v, u_R, L, T), \tag{1.6}$$

where $\mathsf{P}(C_x, C_v, u, T)$ for $\ell \leq u \leq L$ is defined below in (2.9), and $\overline{C}(C_x, C_v, u_R, L, T)$ for $\ell \leq u_R \leq L$ is defined below in (3.5).

As a remark, the choice of $\widetilde{W}_t^{(0,2)}$ inside the oracle query $\Upsilon_{U,\omega}$ as well as the set of information operations Λ (in particular, the choice of the weighted Brownian motion $\widetilde{W}_t^{(2)}$) comes from the friction coefficient in the underdamped Langevin dynamics (1.1). In the above, $e_{\mathcal{F},\Lambda}(A)$ is the worst L^2 strong error for any algorithm A, defined later in (2.4); the notation $\mathcal{A}_N^{Rand}$ means the set of randomized algorithms that use N information operations $\{\Upsilon_{U,\omega}(Y_j, t_j)\}_{j=1,2,\cdots,N}$ (namely, N queries of ∇U and N queries of $\widetilde{W}_t^{(0,2)}$). The notion of randomized algorithms using only N queries will be elaborated further in Section 2.1 below. The proof of Theorem 1.1 will be given in Section 3. The proof of the lower bound estimate is based on a novel non-asymptotic perturbation result with respect to the potential U (see Proposition 3.1).

REMARK 1.1.

(i) As a corollary to the theorem, the randomized midpoint method (see (2.6) below for the algorithm) is order optimal.

(ii) The fundamental challenge for the computational problem $(\mathcal{F}, \Lambda)$ comes from the insufficient information of ∇U, instead of the path irregularity of the random process $\widetilde{W}_t^{(0,2)}$. However, if we replace $\widetilde{W}_t^{(0,2)}$ in the admissible information operations Λ (1.4) by W_t only (the Brownian motion itself), then the complexity lower bound becomes $\Omega(\sqrt{d}N^{-1})$, as the irregularity of the Brownian motion W_t becomes the complexity bottleneck. This follows from the classical result by Clark and Cameron [6] (see also the literature review in the next subsection).

(iii) If we replace $\mathcal{F}$ in (1.2) by the following larger set of potentials

$$\left\{ U \in C^1(\mathbb{R}^d) \;\middle|\; \ell \leq \frac{\|\nabla U(x) - \nabla U(y)\|}{\|x - y\|} \leq L, \; \forall x \neq y \in \mathbb{R}^d \right\},$$

the above lower bound in (1.5) also holds, by the definition of $e_{\mathcal{F}, \Lambda}$ (2.4).

(iv) In (1.6), the scaling of C_{low} with respect to the time T and the condition number $\varkappa = L/\ell$ is complicated. Providing a tight estimate of C_{low} appears to be rather challenging, and we shall leave it to future works.

(v) Though we focus on underdamped Langevin dynamics in this paper, the ideas and techniques herein are applicable to a wide range of SDEs.

1.3. Literature review.   Below is a brief literature review for underdamped Langevin algorithms and also the information-based complexity for differential equations.

Underdamped Langevin algorithms. The Euler-Maruyama method for the underdamped Langevin dynamics, which replaces $\nabla U(X_t)$ with $\nabla U(X_{kh})$ in (1.1) and solves the modified equation for a short time h at the kth time step, is the most widely studied algorithm. To the best of our knowledge, it was first proposed and studied by Ermak and Buckholz [11]. Cheng, Chatterji, Bartlett, and Jordan [5] and later Dalalyan and Riou-Durand [7] proved that the strong error of the Euler-Maruyama algorithm is $\mathcal{O}(\sqrt{d}N^{-1})$. Results were generalized to non-convex potentials in [4] with the same rate in d and N but with much worse prefactors. The sampling error of the Euler-Maruyama algorithm in Kullback-Leibler divergence was studied in [32]. Recently, Shen and Lee [50] proposed the randomized midpoint method, which reduces the error to $\mathcal{O}(\sqrt{d}N^{-3/2})$. There have been other algorithms, for example, the BBK scheme [2, 42], the Verlet-type scheme proposed in [13], and the Leimkuhler-Matthews scheme [31], but non-asymptotic analysis has not been studied for these yet.

Information-based complexity for differential equations. We now provide a concise review of related works on the information-based complexity for both ordinary and stochastic differential equations.

Information-based complexity analysis for ODEs was initially studied by Kacewicz [23–26] for deterministic algorithms, and later by Kacewicz [27, 28], Heinrich and Milla [15], and Daun [8] for randomized algorithms. For some particular classes of ODE systems, the matching complexity bounds and order optimal algorithms are well known, for both deterministic and randomized algorithms. A notable observation is that compared to deterministic algorithms, randomized algorithms may achieve order $1/2$ speed-up, which relates to the universal convergence rate of Monte Carlo methods. This phenomenon also occurs for solving SDE in our case: compared to the Euler-Maruyama method, the randomized midpoint method [50] achieves order $1/2$ improvement. On the other hand, we shall comment that it is still open whether such an improvement is non-trivial in our problem by employing randomized algorithms, namely, whether there exists a strong order $3/2$ deterministic algorithm with only gradient queries.

As for the information-based complexity result for SDEs, the lack of full information about both drift and diffusion terms might contribute to the overall computational complexity. It is a common practice to study the complexity from drift and diffusion separately. Therefore, most works in the literature focus on the complexity due to the diffusion term, since the complexity of the drift term (with trivial diffusion term) reduces to the ODE problem. However, in our problem, it appears unlikely to reduce the

problem into analyzing drift and diffusion terms separately, because both terms are non-trivial for our computational problem of simulating underdamped Langevin dynamics. As far as we know, our problem does not fit into known IBC problem formulations for SDEs.

The study of information-based complexity for the diffusion part dates back at least to the seminal work of Clark and Cameron [6]. They considered SDEs with C^3 drift term (with bounded derivatives up to the third-order) and constant diffusion term, and they proved that the strong one-point approximation error is asymptotically $\Omega(N^{-1})$, for algorithms with uniform mesh grids, where N is the number of queries to the Brownian motion W_t [6, Theorem 1]. The order optimal algorithm is simply the Euler-Maruyama method [16, 44]. Clark and Cameron also showed that for general SDEs with Lipschitz diffusion terms, a certain family of algorithms with uniform mesh grid will result in the one-point approximation error with order $\Omega(N^{-1/2})$ [6]. The minimum one-point approximation error for scalar SDEs could be found in [36]. Apart from the one-point approximation error, the error for trajectories is also considered, i.e., the global approximation error. A series of works of Hofmann, Müller-Gronbach, and Ritter, addressed this problem for L^2 error [17, 18, 20], and for L^∞ error [19, 35]. We refer readers to a survey paper by Müller-Gronbach and Ritter [37] for more details. Around the last decade, Przybyłowicz [45–47], Przybyłowicz and Morkisz [48] studied the time-irregular SDEs. More recently, Hefter, Herzwurm, and Müller-Gronbach provided probabilistic lower bound estimates [14].

Finally, we also point out other related works studying differential equations with inexact information, see, e.g., [1, 29] for ODEs and [39, 40] for SDEs.

Notation. The Lebesgue measure on $\mathbb{R}^d$ (for any dimension d) is denoted by μ. The probability space for (1.1) is $(\mathbb{M}, \Sigma, \mathbb{P})$ and the probability space of the source of randomness for randomized algorithms is denoted by $(\widetilde{\mathbb{M}}, \widetilde{\Sigma}, \widetilde{\mathbb{P}})$. Suppose $\mathcal{X}, \mathcal{Y}$ are two arbitrary spaces, and let $\mathscr{X} : \mathbb{M} \times \mathcal{X} \to \mathcal{Y}$ be a function such that $\mathscr{X}(\cdot, \iota)$ is a random variable on $(\mathbb{M}, \Sigma, \mathbb{P})$ for any fixed $\iota \in \mathcal{X}$. Then,

$$\mathbb{E}_\omega[\mathscr{X}(\omega, \iota)] \equiv \mathbb{E}_{\omega \sim \mathbb{P}}[\mathscr{X}(\omega, \iota)] := \int \mathscr{X}(\omega, \iota) \ d\mathbb{P}(\omega).$$

Other notations like $\mathbb{E}_{\widetilde{\omega}} \equiv \mathbb{E}_{\widetilde{\omega} \sim \widetilde{\mathbb{P}}}$ and $\mathbb{E}_{(\omega, \widetilde{\omega})} \equiv \mathbb{E}_{(\omega, \widetilde{\omega}) \sim \mathbb{P} \times \widetilde{\mathbb{P}}}$ are similarly defined.

For two arbitrary real-valued functions f and g, when $cf \leq g$ for some universal constant c, we denote this by $f \lesssim g$ as a simplification; the notation $\gtrsim$ is similarly defined. The notations Ω and $\mathcal{O}$ follow the convention in complexity analysis, i.e., $f = \Omega(g)$ means $f \gtrsim g$, and $f = \mathcal{O}(g)$ means $f \lesssim g$. The binary relations $\prec$ and $\succ$ are partial order relations on the Boolean lattice $\{0, 1\}^N$.

2. Preliminaries

In this section, we will illustrate the setup of the IBC problem under consideration, the integral form of (1.1), the randomized midpoint method [50], and the exact solution of (1.1) for 1D quadratic potentials.

2.1. IBC problem setup.   We have explained the family of computational problems, characterized by the family $\mathcal{F}$ (1.2) in the introduction. Next, we shall explain more about the admissible information operations and the family of randomized algorithms. We refer readers to e.g., [37, Sec. 2] for a more abstract framework.

Solution map. For the computational problem under consideration, we want to approximate the following solution map, for a given $T > 0$:

$$X_T : (U, \omega) \in (\mathcal{F} \times \mathbb{M}) \mapsto X_T(U, \omega) \in \mathbb{R}^d,$$

where $X_T(U,\omega)$ denotes the strong solution of (1.1) at time T, for a given potential function U.

Admissible information operations. The set of admissible information operations Λ (1.4), recalled here, is

$$\Lambda = \Big\{ \Upsilon_{U,\omega}(x,t) \,\big|\, x \in \mathbb{R}^d, \ t \in [0,T] \Big\}.$$

We need to use a finite amount of information (operations) to predict the strong solution $X_T(U,\omega)$ in (1.1), and such a prediction is known as an algorithm.

Algorithms. We shall always choose

$$Y_1 = x^\star(U), \qquad t_1 = 0, \tag{2.1}$$

as these are the initial conditions in the dynamics (1.1). A deterministic algorithm A with N evaluations means there is a sequence of deterministic mappings $\varphi_j : \mathbb{R}^d \times (\mathbb{R}^d \times \mathbb{R}^{2d})^j \to \mathbb{R}^d \times \mathbb{R}$ $(1 \le j \le N-1)$ and a mapping $\phi : \mathbb{R}^d \times (\mathbb{R}^d \times \mathbb{R}^{2d})^N \to \mathbb{R}^d$ such that

$$(Y_{j+1}, t_{j+1}) := \varphi_j\big(Y_1, \Phi_j\big) \equiv \varphi_j\big(x^\star(U), \Phi_j\big), \qquad \text{for} \ \ 1 \le j \le N-1; \tag{2.2a}$$

$$\Phi_j := \big(\Upsilon_{U,\omega}(Y_1, t_1), \Upsilon_{U,\omega}(Y_2, t_2), \cdots, \Upsilon_{U,\omega}(Y_j, t_j)\big), \qquad \text{for} \ 1 \le j \le N; \tag{2.2b}$$

and the algorithm A is given by

$$\begin{aligned}
A(U,\omega) &= \phi(Y_1, \Phi_N) \\
&\equiv \phi\big(x^\star(U), \Upsilon_{U,\omega}(Y_1, t_1), \Upsilon_{U,\omega}(Y_2, t_2), \cdots, \Upsilon_{U,\omega}(Y_N, t_N)\big).
\end{aligned} \tag{2.3}$$

Ideally, we hope that $X_T(U,\omega) \approx A(U,\omega)$ under certain norms. The family of all such algorithms is denoted by $\mathcal{A}_N^{\mathrm{Det}}$.

The family of randomized algorithms with only N queries is denoted by $\mathcal{A}_N^{\mathrm{Rand}}$. Consider any other probability space $(\widetilde{\mathbb{M}}, \widetilde{\Sigma}, \widetilde{\mathbb{P}})$ as the source of randomness. We define randomized algorithms $A(U,\omega,\widetilde{\omega})$ as mappings from the space $\mathcal{F} \times \mathbb{M} \times \widetilde{\mathbb{M}}$ to the vector space $\mathbb{R}^d$:

$$(U,\omega,\widetilde{\omega}) \in \mathcal{F} \times \mathbb{M} \times \widetilde{\mathbb{M}} \to A(U,\omega,\widetilde{\omega}) \in \mathbb{R}^d,$$

such that

$$A(\cdot,\cdot,\widetilde{\omega}) \in \mathcal{A}_N^{\mathrm{Det}}, \qquad \forall \, \widetilde{\omega} \in \widetilde{\mathbb{M}}.$$

Apparently, $\mathcal{A}_N^{\mathrm{Det}} \subseteq \mathcal{A}_N^{\mathrm{Rand}}$ and moreover, $\mathcal{A}_{N_1}^{\mathrm{Det}} \subseteq \mathcal{A}_{N_2}^{\mathrm{Det}}$, $\mathcal{A}_{N_1}^{\mathrm{Rand}} \subseteq \mathcal{A}_{N_2}^{\mathrm{Rand}}$ whenever $N_1 \le N_2$.

The error of the randomized algorithm $A(\cdot,\cdot,\widetilde{\omega})$ is measured in the L^2 sense herein:

$$e_{\mathcal{F},\Lambda}(A) := \sup_{U \in \mathcal{F}} \Big(\mathbb{E}_{(\omega,\widetilde{\omega}) \sim \mathbb{P} \times \widetilde{\mathbb{P}}} \big[|X_T(U,\omega) - A(U,\omega,\widetilde{\omega})|^2 \big] \Big)^{1/2}. \tag{2.4}$$

REMARK 2.1.   For the underdamped Langevin dynamics used to sample log-concave probability distributions $e^{-U} / \int e^{-U}$, we are interested only in X_T, instead of the whole trajectory on the interval $[0,T]$.

2.2. The integral form and the randomized midpoint method. Several numerical algorithms for underdamped Langevin dynamics [5, 7, 11, 50] are based on its integral form

$$X_t = X_0 + \frac{1-e^{-2t}}{2}V_0 + \frac{1}{\sqrt{L}}\int_0^t (1-e^{2(s-t)})\mathrm{d}W_s - \frac{1}{2L}\int_0^t (1-e^{2(s-t)})\nabla U(X_s)\mathrm{d}s,$$

$$V_t = e^{-2t}V_0 + \frac{2}{\sqrt{L}}\int_0^t e^{2(s-t)}\mathrm{d}W_s - \frac{1}{L}\int_0^t e^{2(s-t)}\nabla U(X_s)\mathrm{d}s. \tag{2.5}$$

In [50], Shen and Lee considered a randomized midpoint method for simulating underdamped Langevin dynamics, in the context of sampling log-concave distributions. Let N_s be the total number of steps and $h := T/N_s$ is the time step size. Given $\hat{X}_k$ and $\hat{V}_k$ at time $s_k := kh$ for integer $0 \leq k \leq N_s - 1$, the randomized midpoint method approximates $\hat{X}_{k+1}$ and $\hat{V}_{k+1}$ in the following way:

$$\hat{X}_{k+1} = \hat{X}_k + \frac{1-e^{-2h}}{2}\hat{V}_k + \frac{1}{\sqrt{L}}\int_0^h (1-e^{2(s-h)})\mathrm{d}W_{s_k+s}$$

$$- \frac{1}{2L}h(1-e^{2(\eta_k h - h)})\nabla U(\hat{X}_{k+1/2}),$$

$$\hat{V}_{k+1} = e^{-2h}\hat{V}_k + \frac{2}{\sqrt{L}}\int_0^h e^{2(s-h)}\mathrm{d}W_{s_k+s}$$

$$- \frac{1}{L}he^{2(\eta_k h - h)}\nabla U(\hat{X}_{k+1/2}), \tag{2.6}$$

$$\hat{X}_{k+1/2} = \hat{X}_k + \frac{1-e^{-2\eta_k h}}{2}\hat{V}_k + \frac{1}{\sqrt{L}}\int_0^{\eta_k h} (1-e^{2(s-\eta_k h)})\mathrm{d}W_{s_k+s}$$

$$- \frac{1}{2L}\Big(\int_0^{\eta_k h} 1 - e^{2(s-\eta_k h)}\,\mathrm{d}s\Big)\nabla U(\hat{X}_k),$$

where *i.i.d.* random variables $\{\eta_k\}_{0 \leq k \leq N_s - 1}$ are uniformly distributed on the interval $[0,1]$, independent of W_t; the initial condition is $\hat{X}_0 = x^\star(U)$ and $\hat{V}_0 = 0$.

In the above, $\hat{X}_{k+1}$ and $\hat{V}_{k+1}$ are obtained by approximating the integral with respect to $\nabla U(X_s)$ in (2.5) by its evaluation at a single point $\hat{X}_{k+1/2}$; $\hat{X}_{k+1/2}$ is obtained using the Euler-Maruyama scheme for the integral form (2.5) at the random time $s_k + \eta_k h$.

REMARK 2.2. The RMM with N_s steps is a randomized algorithm with $N = 2N_s$ evaluations in the form of (2.2) and (2.3); please refer to Appendix A for more details.

Many previous works have proposed and analyzed various randomized algorithms to solve differential equations. For ODEs, the analogy of the randomized midpoint method can be found in, e.g., [1, 8, 15]. More randomized ODE solvers could be found in e.g., [22, 51, 52]. As for SDEs, the randomized Euler's method, which is in a very similar spirit as the randomized midpoint method, has been studied in, e.g., [39, 46–48]. A randomized Milstein method was studied by Kruse and Wu [30] for non-differentiable drift functions; a randomized derivative-free Milstein method was studied by Morkisz and Przybyłowicz [40] for scalar SDEs with inexact information.

REMARK 2.3. While we focus on the strong error in this paper, we would like to comment that for underdamped Langevin dynamics, the randomized midpoint method defined in (2.6) has a weak error of $\mathcal{O}(h^3)$. As the full analysis is tedious and it is

not the main focus of our paper, we only include a heuristic argument here. We shall consider a one-step error only and let $\eta \equiv \eta_0$ herein for simplicity.

Note that for a fixed realization of W_t, we can compare

$$
\begin{aligned}
\hat{X}_1 - X_h = & -\frac{1}{2L}\Big(h(1-e^{-2(h-\eta h)})\nabla U(\hat{X}_{\frac{1}{2}}) - \int_0^h (1-e^{-2(h-s)})\nabla U(X_s)\mathrm{d}s\Big) \\
= & -\frac{1}{2L}\Big(h(1-e^{-2(h-\eta h)})\nabla U(\hat{X}_{\frac{1}{2}}) - h(1-e^{-2(h-\eta h)})\nabla U(X_{\eta h}) \\
& + h(1-e^{-2(h-\eta h)})\nabla U(X_{\eta h}) - \int_0^h (1-e^{-2(h-s)})\nabla U(X_s)\mathrm{d}s\Big),
\end{aligned}
$$

$$
\begin{aligned}
\hat{V}_1 - V_h = & -\frac{1}{L}\Big(he^{-2(h-\eta h)}\nabla U(\hat{X}_{\frac{1}{2}}) - \int_0^h e^{-2(h-s)}\nabla U(X_s)\mathrm{d}s\Big) \\
= & -\frac{1}{L}\Big(he^{-2(h-\eta h)}\nabla U(\hat{X}_{\frac{1}{2}}) - he^{-2(h-\eta h)}\nabla U(X_{\eta h})\Big) \\
& -\frac{1}{L}\Big(he^{-2(h-\eta h)}\nabla U(X_{\eta h}) - \int_0^h e^{-2(h-s)}\nabla U(X_s)\mathrm{d}s\Big) \\
= & : I + II.
\end{aligned}
$$

We observe that $\hat{X}_1 - X_h$ is of higher order than $\hat{V}_1 - V_h$. Therefore, keeping only the error from V and by the Taylor expansion around (X_h, V_h), we can estimate, for a smooth enough test function f, that

$$
\big|\mathbb{E}\big[f(\hat{X}_1,\hat{V}_1) - f(X_h,V_h)\big]\big| \lesssim \big|\mathbb{E}\big[\nabla_v f(X_h,V_h)\cdot(I+II)\big]\big| \\
+ \big|\mathbb{E}\big[(I+II)\cdot\nabla_v^2 f(X_h,V_h)(I+II)\big]\big|.
$$

Now let $\mathbb{E}_\eta$ denote the expectation with respect to η only. Since η is uniformly distributed on $[0,1]$, we have

$$
\mathbb{E}_\eta\big[he^{-2(h-\eta h)}\nabla U(X_{\eta h})\big] = h\int_0^1 e^{-2(h-\eta h)}\nabla U(X_{\eta h})\mathrm{d}\eta = \int_0^h e^{-2(h-s)}\nabla U(X_s)\mathrm{d}s.
$$

This shows $\mathbb{E}_\eta\big[II\big] = 0$. Since (X_h,V_h) is independent of η, we have

$$
\mathbb{E}\big[\nabla_v f(X_h,V_h)\cdot II\big] = \mathbb{E}\big[\nabla_v f(X_h,V_h)\cdot \mathbb{E}_\eta\big[II\big]\big] = 0.
$$

Therefore,

$$
\big|\mathbb{E}\big[f(\hat{X}_1,\hat{V}_1) - f(X_h,V_h)\big]\big| \lesssim \frac{h}{L}\Big|\mathbb{E}\big[\nabla_v f(X_h,V_h)\cdot\big(\nabla U(\hat{X}_{\frac{1}{2}}) - \nabla U(X_{\eta h})\big)\big]\Big| \\
+ \mathbb{E}\big[\|\nabla_v^2 f(X_h,V_h)\|\cdot|\hat{V}_1 - V_h|^2\big].
$$

Finally, assuming that all derivatives of f are bounded, we use [50, Lemma 9] for the first term below, and [50, Lemma 2] for the second term below,

$$
\big|\mathbb{E}\big(f(\hat{X}_1,\hat{V}_1) - f(X_h,V_h)\big)\big| \lesssim \frac{h}{L}\Big(\mathbb{E}|\nabla U(\hat{X}_{\frac{1}{2}}) - \nabla U(X_{\eta h})|^2\Big)^{\frac{1}{2}} + \mathbb{E}|\hat{V}_1 - V_h|^2 = \mathcal{O}(h^4).
$$

This local truncation error gives $\mathcal{O}(h^3)$ weak error by Grönwall's inequality as usual.

2.3. Exact solution for quadratic potentials in 1D. Our estimate of the prefactor C_{low} in Theorem 1.1 relies on the behavior of $X_t(U_u,\omega)$ and $V_t(U_u,\omega)$, where the potential has the quadratic form $U_u(x) := ux^2/2$ (thus $x^\star(U_u) = 0$ for this case). Therefore, in this subsection, we shall first review the exact solution of the underdamped Langevin dynamics with quadratic potentials and then define a quantity in (2.9) to be used in the next section.

It is easy to rewrite (1.1) as

$$
\mathrm{d}\begin{bmatrix} X_t \\ V_t \end{bmatrix} = H \begin{bmatrix} X_t \\ V_t \end{bmatrix} \mathrm{d}t + \begin{bmatrix} 0 \\ \frac{2}{\sqrt{L}} \end{bmatrix} \mathrm{d}W_t, \qquad H = \begin{bmatrix} 0 & 1 \\ -\frac{u}{L} & -2 \end{bmatrix}.
$$

Then its integral form can be immediately obtained as follows

$$
\begin{bmatrix} X_t \\ V_t \end{bmatrix} = \int_0^t e^{H(t-s)} \begin{bmatrix} 0 \\ \frac{2}{\sqrt{L}} \end{bmatrix} \mathrm{d}W_s.
$$

The matrix exponential of H can be explicitly computed for $u < L$:

$$
e^{Ht} = \frac{1}{2\sqrt{1-u/L}} \begin{bmatrix} \left(\lambda_+ e^{-t\lambda_-} - \lambda_- e^{-t\lambda_+}\right) & \left(e^{-t\lambda_-} - e^{-t\lambda_+}\right) \\ \frac{u}{L}\left(e^{-t\lambda_+} - e^{-t\lambda_-}\right) & \left(\lambda_+ e^{-t\lambda_+} - \lambda_- e^{-t\lambda_-}\right) \end{bmatrix},
$$

where

$$
\lambda_\pm = 1 \pm \sqrt{1-u/L}. \tag{2.7}
$$

This immediately leads to the following result.

LEMMA 2.1 (Exact solution). *When $d = 1$ and $U_u(x) = ux^2/2$, we have for $u < L$*

$$
\begin{aligned}
X_t(U_u,\cdot) &= \frac{1}{\sqrt{L-u}} \int_0^t \left(e^{-(t-s)\lambda_-} - e^{-(t-s)\lambda_+}\right) \mathrm{d}W_s, \\
V_t(U_u,\cdot) &= \frac{1}{\sqrt{L-u}} \int_0^t \left(-\lambda_- e^{-(t-s)\lambda_-} + \lambda_+ e^{-(t-s)\lambda_+}\right) \mathrm{d}W_s.
\end{aligned} \tag{2.8}
$$

Next, let us introduce the following quantity for $C_x, C_v > 0$, $\ell \le u \le L$,

$$
\mathsf{P}(C_x,C_v,u,T) := \mathbb{P}\Big(\omega : \sup_{0 \le t \le T} X_t(U_u,\omega) \ge 2C_x,
$$
$$
\inf_{0 \le t \le T} X_t(U_u,\omega) \le -2C_x, \ \sup_{0 \le t \le T} |V_t(U_u,\omega)| \le C_v/2\Big). \tag{2.9}
$$

The event under consideration requires $X_t(U_u,\omega)$ to cross $-2C_x$ and $2C_x$, whereas the velocity $V_t(U_u,\omega)$ is uniformly bounded on the whole interval $[0,T]$. Typically, we should expect to choose a small C_x and a large C_v in order to have $\mathsf{P}(C_x,C_v,u,T) = \mathcal{O}(1)$. Indeed, if C_x/C_v is too large, the probability $\mathsf{P}(C_x,C_v,u,T)$ equals to 0, as stated in the following lemma.

LEMMA 2.2. *When $12C_x/C_v > T$, then $\mathsf{P}(C_x,C_v,u,T) = 0$.*

Proof. For any ω satisfying the condition in (2.9), the travel distance of X_t must be at least $6C_x$ (since it starts from 0 and has to cross levels $2C_x$ and $-2C_x$) with velocity at most $C_v/2$, then the total time must be at least $12C_x/C_v$. $\qquad\square$

However, note that for large enough C_v, $\mathsf{P}(C_x, C_v, u, T) > 0$.

LEMMA 2.3.   *For fixed $C_x > 0$, $u \in (\ell, L)$, and $T > 0$, and for sufficiently large C_v, we have*

$$\mathsf{P}(C_x, C_v, u, T) \geq \frac{1}{2}\mathbb{P}\Big(\omega: \sup_{0 \leq t \leq T} X_t(U_u, \omega) \geq 2C_x, \inf_{0 \leq t \leq T} X_t(U_u, \omega) \leq -2C_x\Big) > 0.$$

The proof of this lemma is postponed to Appendix B.

Therefore, the prefactor C_{low} in (1.6) is non-zero. The precise dependence of C_{low} on parameters ℓ, L and T, however, appears to be a challenging problem and will be left for future investigations.

3. Proof of Theorem 1.1.

The proof of the lower bound estimate relies on the (non-asymptotic) perturbation analysis in Section 3.1, in particular, the lower bound estimate in Proposition 3.1. The overall strategy, from the information-based complexity perspective, is similar to the lower bound estimate for randomized algorithms for integration problems, see, e.g., [41]. The new ingredient is the perturbation-type analysis for the particular problem under consideration. The proof of the upper bound is known from [50]. Thus, we shall only provide a sketch of the main steps to prove the upper bound, for completeness.

3.1. Non-asymptotic perturbation analysis with respect to U.   We consider the case $d = 1$, which is assumed throughout this subsection. We postpone proofs for all results in this subsection to Section 3.3 for clarity.

Let us consider

$$U_u(x) := \frac{ux^2}{2}, \ \forall u \in [\ell, L],$$

and let us also introduce a set parameterized by $u \in (\ell, L)$ and $\epsilon > 0$

$$\mathcal{F}_{u,\epsilon} := \{U \in \mathcal{F}: \|\nabla U(x) - ux\|_\infty \leq \epsilon, \ x^\star(U) = 0\}. \tag{3.1}$$

Recall from Section 1.1 that $x^\star(U)$ is the local minimum of potential function $U \in \mathcal{F}$; thus for any $U \in \mathcal{F}_{u,\epsilon}$, we have the initial condition $X_0(U, \omega) = 0$.

First, we show that when the potential function is slightly perturbed away from a quadratic function U_u, the strong solutions of $X_t(U, \omega)$ and $V_t(U, \omega)$ are at most perturbed by an order of $\mathcal{O}(\epsilon)$.

LEMMA 3.1 (Upper bound). *Consider any $u \in (\ell, L)$. For any $U \in \mathcal{F}_{u,\epsilon}$, any $\omega \in \mathbb{M}$, and $t \in [0, T]$, we have*

$$\begin{aligned}
|X_t(U, \omega) - X_t(U_u, \omega)| &\leq \frac{\epsilon}{2L(1 - \sqrt{1 - \frac{u}{L}})\sqrt{1 - \frac{u}{L}}}, \\
|V_t(U, \omega) - V_t(U_u, \omega)| &\leq \frac{\epsilon}{L\sqrt{1 - \frac{u}{L}}}.
\end{aligned} \tag{3.2}$$

For any $u \in (\ell, L)$, $C_x > 0$ and $C_v > 0$, let us define

$$\bar{\epsilon} \equiv \bar{\epsilon}(C_x, C_v, u, L)$$
$$:= \min\Big\{2L\Big(1 - \sqrt{1 - \frac{u}{L}}\Big)\sqrt{1 - \frac{u}{L}}C_x, \ L\sqrt{1 - \frac{u}{L}}\frac{C_v}{2}\Big\},$$

which is strictly positive. Moreover, for $\epsilon > 0$, let us define a set $\mathcal{E} \equiv \mathcal{E}(C_x, C_v, u, T, \epsilon)$

$$
\mathcal{E} := \Big\{ \omega : \sup_{0 \leq t \leq T} X_t(U, \omega) \geq C_x, \ \inf_{0 \leq t \leq T} X_t(U, \omega) \leq -C_x,
$$
$$
\sup_{0 \leq t \leq T} |V_t(U, \omega)| \leq C_v, \ \forall U \in \mathcal{F}_{u, \epsilon} \Big\}. \tag{3.3}
$$

In the following lemma, we shall bound $\mathbb{P}(\mathcal{E})$ by $\mathsf{P}(C_x, C_v, u, T)$ from below, for small enough ϵ.

LEMMA 3.2. *Consider any $u \in (\ell, L)$, and any $C_x, C_v > 0$. For arbitrary $0 < \epsilon < \bar{\epsilon}$, $\mathbb{P}(\mathcal{E})$ is uniformly bounded from below by $\mathsf{P}(C_x, C_v, u, T)$ defined in* (2.9).

In general, for two fixed potential functions $U_1 \neq U_2$, the distance $|X_T(U_1, \omega) - X_T(U_2, \omega)|$ highly depends on the realization of the Brownian motion $W_t(\omega)$, and it is unlikely to establish a uniform non-trivial lower bound of $|X_T(U_1, \omega) - X_T(U_2, \omega)|$ for arbitrary ω. However, if we restrict the outcome ω to a "nice" set, i.e., $\omega \in \mathcal{E}$ defined in (3.3), then we can provide a lower bound estimate of $|X_T(U_1, \omega) - X_T(U_2, \omega)|$ as in the following proposition. This lower bound estimate is the key to prove Theorem 1.1.

PROPOSITION 3.1. *Consider any $u \in (\ell, L)$, $C_x, C_v, \epsilon > 0$, and let $\mathcal{E} \equiv \mathcal{E}(C_x, C_v, u, T, \epsilon)$ be the set defined in* (3.3). *Consider two potential functions $U_1, U_2 \in \mathcal{F}_{u, \epsilon}$. Assume that*

(i) the continuous function $g := \nabla U_1 - \nabla U_2$ is non-negative on $\mathbb{R}$;

(ii) there exists $\mathcal{I} \subseteq [-C_x/2, C_x/2]$, a finite union of closed bounded intervals, such that

$$
g(x) \geq \frac{\epsilon}{2} \mathbb{1}_{\mathcal{I}}(x), \qquad \forall x \in \mathbb{R}, \tag{3.4}
$$

where $\mathbb{1}_{\mathcal{I}}$ is the indicator function of set $\mathcal{I}$ on $\mathbb{R}$.

Let us introduce $u_R \in [u, L]$ as a constant such that

$$
\nabla^2 U_2(x) \leq u_R, \qquad \forall x \in \mathbb{R}.
$$

Then for any $\omega \in \mathcal{E}$, we have

$$
|X_T(U_1, \omega) - X_T(U_2, \omega)| \geq \overline{C} \, \epsilon \mu(\mathcal{I}),
$$

where μ is the Lebesgue measure and $\overline{C} \equiv \overline{C}(C_x, C_v, u_R, L, T)$ is given by

$$
\overline{C} := \begin{cases} \dfrac{e^{(\frac{3C_x}{2C_v} - T)(1 - \sqrt{1 - \frac{u_R}{L}})} \left(1 - e^{-\frac{C_x}{C_v}\sqrt{1 - \frac{u_R}{L}}} \right)}{4LC_v\sqrt{1 - \frac{u_R}{L}}}, & \text{if } u_R < L; \\[20pt] \dfrac{C_x e^{\frac{3C_x}{2C_v} - T}}{4LC_v^2}, & \text{if } u_R = L. \end{cases} \tag{3.5}
$$

3.2. Proof of the lower bound estimate for Theorem 1.1.
We shall proceed to prove the lower bound estimate, based on the results in Section 3.1.

Case (I): $d = 1$. We shall first consider the case $d = 1$.

Step (1): Setup and notations. Since we only access ∇U at N points, we will not be able to gain the full information of ∇U based on the local queries. In this step, we shall consider a family of U (see (3.7) and (3.8) below) as small perturbations of a quadratic

potential with the mode (Hessian) u, and we shall estimate how the deviation of ∇U contributes to the error $e_{\mathcal{F},\Lambda}(A)$ (2.4) for any randomized algorithm $A \in \mathcal{A}_N^{\mathrm{Rand}}$.

Without loss of generality, we assume that N is an even integer. We shall pick $\ell < u < u_R \leq L$. We could then fix C_x and C_v satisfying $\mathsf{P}(C_x, C_v, u, T) > 0$. Let us define

$$\xi := \min\{u - \ell, u_R - u\} > 0, \qquad \epsilon := \frac{C_x \xi}{8N}. \tag{3.6}$$

We will pick N sufficiently large such that $\epsilon < \bar{\epsilon}(C_x, C_v, u, L)$. Then we could also determine the set $\mathcal{E}$ defined in (3.3).

We divide the interval $I := \left[-\frac{C_x}{2}, \frac{C_x}{2}\right]$ into $2N$ equally spaced sub-intervals. Let $x_j = \frac{C_x}{2N} j$ for integers $-N \leq j \leq N$; let $I_j = [x_j, x_{j+1})$ for $-N \leq j \leq N-2$ and let $I_{N-1} = [x_{N-1}, x_N]$. Define a non-negative function g supported on $\left[0, \frac{C_x}{2N}\right]$ by

$$g(x) := \begin{cases} ax^2, & x \in \left[0, \frac{C_x}{8N}\right]; \\ -a(x - \frac{C_x}{4N})^2 + 2a\left(\frac{C_x}{8N}\right)^2, & x \in \left[\frac{C_x}{8N}, \frac{3C_x}{8N}\right]; \\ a(x - \frac{C_x}{2N})^2, & x \in \left[\frac{3C_x}{8N}, \frac{C_x}{2N}\right]; \\ 0, & x \notin \left[0, \frac{C_x}{2N}\right], \end{cases}$$

where $a := \xi \frac{4N}{C_x}$. It is easy to verify that $g \in C^1(\mathbb{R})$ and moreover,

$$\|g'\|_\infty = \xi, \qquad \|g\|_\infty = \epsilon.$$

Moreover, we choose $\mathscr{I} = \left[\frac{C_x}{8N}, \frac{3C_x}{8N}\right]$ (thus, the length $\mu(\mathscr{I}) = \frac{C_x}{4N}$), and we know that $g(x) \geq \frac{\epsilon}{2} \mathbb{1}_{\mathscr{I}}(x)$ for any $x \in \mathbb{R}$.

For index $\beta = (\beta_{-N}, \beta_{-N+1}, \cdots, \beta_{N-1}) \in \{0,1\}^{2N}$, we define U_β with $U_\beta(0) = 0$ by its derivative

$$\nabla U_\beta(x) := ux + \sum_{j=-N}^{N-1} \beta_j g(x - x_j) \geq ux + \sum_{j=-N}^{N-1} \beta_j \frac{\epsilon}{2} \mathbb{1}_{\mathscr{I}}(x - x_j). \tag{3.7}$$

Apparently, U_β is well-defined, $U_\beta \in \mathcal{F}_{u,\epsilon}$, and $\nabla^2 U_\beta(x) \leq u_R$, $\forall \beta$, $\forall x \in \mathbb{R}$. Define a space

$$\mathcal{G} \equiv \mathcal{G}_{u,\epsilon} := \{U_\beta : \beta \in \{0,1\}^{2N}\} \subseteq \mathcal{F}_{u,\epsilon}, \tag{3.8}$$

and let $\mu_\mathcal{G}$ be a uniform probability distribution on the set $\mathcal{G}$, i.e., $\mu_\mathcal{G}(U_\beta) = \frac{1}{2^{2N}}$ for any β.

Then by Definition (2.4),

$$e_{\mathcal{F},\Lambda}(A)^2 \geq \mathbb{E}_{U_\beta \sim \mu_\mathcal{G}} \mathbb{E}_{(\omega,\widetilde{\omega}) \sim \mathbb{P} \times \widetilde{\mathbb{P}}}\left[|X_T(U_\beta, \omega) - A(U_\beta, \omega, \widetilde{\omega})|^2\right]$$

$$= \mathbb{E}_{(\omega,\widetilde{\omega}) \sim \mathbb{P} \times \widetilde{\mathbb{P}}}\left[\mathbb{E}_{U_\beta \sim \mu_\mathcal{G}}\left[|X_T(U_\beta, \omega) - A(U_\beta, \omega, \widetilde{\omega})|^2\right]\right]$$

$$\geq \mathbb{P}(\mathcal{E}) \, \mathbb{E}_{(\omega,\widetilde{\omega}) \sim \mathbb{P}|_\mathcal{E} \times \widetilde{\mathbb{P}}}\left[\mathbb{E}_{U_\beta \sim \mu_\mathcal{G}}\left[|X_T(U_\beta, \omega) - A(U_\beta, \omega, \widetilde{\omega})|^2\right]\right], \tag{3.9}$$

where the probability measure $\mathbb{P}|_\mathcal{E}$ is the restriction of $\mathbb{P}$ to the event $\mathcal{E}$; more specifically, $\mathbb{P}|_\mathcal{E}(B) = \mathbb{P}(B \cap \mathcal{E})/\mathbb{P}(\mathcal{E})$ for any event $B \in \Sigma$. Besides, $\mathbb{E}_{U_\beta \sim \mu_\mathcal{G}}$ means expectation with respect to U_β, where U_β is treated as a uniformly distributed random variable on the set $\mathcal{G}$; to avoid introducing too many notations, we slightly abuse the notation of U_β

to represent both a single potential function (or say an element in $\mathcal{G}$) and a (uniformly distributed) random variable on the set $\mathcal{G}$.

We know by Lemma 3.2 that $\mathbb{P}(\mathcal{E})$ is uniformly bounded from below by a positive value $\mathsf{P}(C_x, C_v, u, T)$, and we claim (to be proved below in the *Step (2)*)

$$\mathbb{E}_{U_\beta \sim \mu_\mathcal{G}} \left[|X_T(U_\beta, \omega) - A(U_\beta, \omega, \widetilde{\omega})|^2 \right] \gtrsim \frac{C_x^4 \overline{C}^2 \xi^2}{N^3}. \tag{3.10}$$

Therefore,

$$e_{\mathcal{F}, \Lambda}(A) \gtrsim C_{\mathrm{low}} N^{-3/2},$$

where

$$C_{\mathrm{low}} = \sqrt{\mathsf{P}(C_x, C_v, u, T)} C_x^2 \overline{C} \xi$$
$$\overset{(3.6)}{=} \sqrt{\mathsf{P}(C_x, C_v, u, T)} C_x^2 \overline{C} \min\{u - \ell, u_R - u\}.$$

Since C_x, C_v, u, and u_R are parameters to be tuned, we arrive at (1.6) by optimizing.

Step (2): Proof of the perturbation bound (3.10). From now on, we fix both $\widetilde{\omega} \in \widetilde{\mathbb{M}}$ and $\omega \in \mathcal{E}$. The main task is to estimate the fluctuation of the exact solution $X_T(U_\beta, \omega)$ for those U_β with the same algorithmic output $A(U_\beta, \omega, \widetilde{\omega})$ by Proposition 3.1; this is done by a quantitative perturbation analysis.

The first task is to characterize the set of U_β with the same algorithmic output $A(U_\beta, \omega, \widetilde{\omega})$. This is given by Lemma 3.3 below. To state the result, let us define some notations. The access points for ∇U_β are denoted by $Y_1^\beta, Y_2^\beta, \cdots, Y_N^\beta$, which only depend on the choice of β. Let us denote the union of sub-intervals that Y_j^β belong to as $\mathcal{J}^\beta$:

$$\mathcal{J}^\beta := \bigcup_{\substack{-N \le j \le N-1, \\ Y_k^\beta \in I_j \text{ for some } 1 \le k \le N}} I_j.$$

If $\mu(\mathcal{J}^\beta) < C_x/2$ (e.g., there exist two indices $j_1 < j_2$ such that $Y_{j_1}^\beta$ and $Y_{j_2}^\beta$ belong to the same sub-interval), then we add sub-intervals with the largest indices to complete $\mathcal{J}^\beta$. More specifically, for each fixed β, let us define

$$\mathcal{L} = \left\{ j \mid -N \le j \le N-1, \ Y_k^\beta \notin I_j, \ \forall k \in \{1, 2, 3, \cdots, N\} \right\}.$$

When the cardinality of $\mathcal{L}$ is larger than N (i.e., $|\mathcal{L}| > N$), or equivalently $\mu(\mathcal{J}^\beta) < C_x/2$, we add $|\mathcal{L}| - N$ sub-intervals in a procedure described by the following pseudocode:

Pseudocode: The completion procedure of $\mathcal{J}^\beta$
---
   while $\mu(\mathcal{J}^\beta) < C_x/2$ **do**
      $j \leftarrow$ largest value in $\mathcal{L}$
      $\mathcal{L} \leftarrow \mathcal{L} \setminus \{j\}$
      $\mathcal{J}^\beta \leftarrow \mathcal{J}^\beta \cup I_j$
   end while

EXAMPLE 3.1. If $N = 3$, then there are six sub-intervals $I_{-3}, I_{-2}, I_{-1}, I_0, I_1, I_2$. Let us consider the following examples:

- if all $Y_1^\beta, Y_2^\beta, Y_3^\beta \in I_0$, then we set $\mathcal{J}^\beta := I_0 \cup (I_1 \cup I_2)$;

- if all $Y_1^\beta, Y_2^\beta, Y_3^\beta \in I_{-1}$, then we set $\mathcal{J}^\beta := I_{-1} \cup (I_1 \cup I_2)$;
- if $Y_1^\beta, Y_2^\beta \in I_0$, and $Y_3^\beta \in I_1$, then we set $\mathcal{J}^\beta := I_0 \cup I_1 \cup (I_2)$;
- if $Y_1^\beta \in I_0$, $Y_2^\beta \in I_1$, and $Y_3^\beta \in I_2$, then we set $\mathcal{J}^\beta := I_0 \cup I_1 \cup I_2$.

The interval within the parenthesis is the additional sub-interval(s) that we add to complete $\mathcal{J}^\beta$.

Note that such a procedure is always valid, since $Y_1^\beta, Y_2^\beta, \cdots, Y_N^\beta$ reside in at most N sub-intervals; for each β, $\mathcal{J}^\beta$ is always uniquely defined. From now on, when we use the notation $\mathcal{J}^\beta$, we refer to the "completed" version. Likewise, time points for queries, denoted by t_j^β, also depend on β only.

LEMMA 3.3.   *For any fixed index β, if for some index β', $\nabla U_{\beta'} = \nabla U_\beta$ on the domain $\mathcal{J}^\beta$, then*

(i) $Y_j^\beta = Y_j^{\beta'}$, $t_j^\beta = t_j^{\beta'}$ for all $1 \le j \le N$;

(ii) $\mathcal{J}^\beta = \mathcal{J}^{\beta'}$ and moreover, $A(U_\beta, \omega, \widetilde{\omega}) = A(U_{\beta'}, \omega, \widetilde{\omega})$.

Proof.   Part (ii) trivially follows from the form of algorithms in (2.3) and part (i). Then it suffices to prove part (i). This comes from induction: if $Y_j^\beta = Y_j^{\beta'}$, $t_j^\beta = t_j^{\beta'}$ for all $j \le k$, then we know the information up to k queries are the same, i.e., $\Phi_k^\beta = \Phi_k^{\beta'}$ where Φ_k is defined in (2.2b), and superscripts are used to indicate the dependence; to verify $\Phi_k^\beta = \Phi_k^{\beta'}$, we need to check the following two relations for any $j \le k$:

$$\widetilde{W}_{t_j^\beta}^{(0,2)}(\omega) = \widetilde{W}_{t_j^{\beta'}}^{(0,2)}(\omega), \ \nabla U_{\beta'}(Y_j^{\beta'}) = \nabla U_\beta(Y_j^\beta).$$

Since $t_j^\beta = t_j^{\beta'}$ by the assumption of induction, obviously, $\widetilde{W}_{t_j^\beta}^{(0,2)}(\omega) = \widetilde{W}_{t_j^{\beta'}}^{(0,2)}(\omega)$. Next, since $Y_j^\beta = Y_j^{\beta'}$, we have $\nabla U_{\beta'}(Y_j^{\beta'}) = \nabla U_{\beta'}(Y_j^\beta)$. It is then enough to verify $\nabla U_{\beta'}(Y_j^\beta) = \nabla U_\beta(Y_j^\beta)$.

- If $Y_j^\beta \notin [-C_x/2, C_x/2]$, then $\nabla U_{\beta'}(Y_j^\beta) \overset{(3.7)}{=} uY_j^\beta \overset{(3.7)}{=} \nabla U_\beta(Y_j^\beta)$.
- If $Y_j^\beta \in [-C_x/2, C_x/2]$, then $Y_j^\beta \in \mathcal{J}^\beta$ by definition. By the assumption that $\nabla U_{\beta'} = \nabla U_\beta$ on the domain $\mathcal{J}^\beta$, we know $\nabla U_{\beta'}(Y_j^\beta) = \nabla U_\beta(Y_j^\beta)$.

Then $(Y_{k+1}^\beta, t_{k+1}^\beta) = \varphi_k(Y_1^\beta, \Phi_k^\beta) = \varphi_k(Y_1^{\beta'}, \Phi_k^{\beta'}) = (Y_{k+1}^{\beta'}, t_{k+1}^{\beta'})$. As for the base case, note that for an arbitrary β, $U_\beta \in \mathcal{F}_{u,\epsilon}$, and thus $Y_1^\beta \overset{(2.1)}{=} x^\star(U_\beta) \overset{(3.1)}{=} 0$.   □

DEFINITION 3.1.   *For two arbitrary indices β and β', we define a binary relation $\beta \sim \beta'$ if $\beta'_j = \beta_j$ whenever $I_j \subseteq \mathcal{J}^\beta$. By the above Lemma 3.3, we know $\mathcal{J}^\beta = \mathcal{J}^{\beta'}$ and $A(U_\beta, \omega, \widetilde{\omega}) = A(U_{\beta'}, \omega, \widetilde{\omega})$. It is easy to verify that such a relation is an equivalence relation.*

For any index β, apparently, there are exactly $2^N - 1$ other indices belonging to the same equivalence class (since we could freely choose $\beta_j \in \{0, 1\}$ whenever $I_j \nsubseteq \mathcal{J}^\beta$), and there are exactly 2^N such equivalence classes. Let us enumerate these equivalence classes by $\mathcal{K}_1, \mathcal{K}_2, \cdots, \mathcal{K}_{2^N}$.

EXAMPLE 3.2.   When $N = 3$, for an index $\beta = (0, 0, 0, 0, 0, 0)$, we assume that $\mathcal{J}^\beta = I_0 \cup I_1 \cup I_2$ as an example. Then we could freely choose the first three indices, and the equivalence class containing β is exactly $\left\{ (\beta_{-3}, \beta_{-2}, \beta_{-1}, 0, 0, 0) \mid \beta_{-3}, \beta_{-2}, \beta_{-1} \in \{0, 1\} \right\}$.

We now consider how much the actual solution $X_T(U_\beta, \omega)$ can fluctuate within the same class. Consider an arbitrary equivalence class $\mathcal{K}$, and suppose $\beta \in \mathcal{K}$. For any index β', recall that $\mathcal{J}^{\beta'}$ are the same for all $\beta' \in \mathcal{K}$ (i.e., $\mathcal{J}^{\beta'}$ only depends on the equivalence class $\mathcal{K}$ that we consider), and β'_j are the same if $I_j \subseteq \mathcal{J}^\beta$. This motivates us to define the reduced index below.

DEFINITION 3.2 (Reduced index). *For an equivalence class $\mathcal{K}$, let us denote $\mathcal{J} \equiv \mathcal{J}^\beta$ for an arbitrary $\beta \in \mathcal{K}$; from the explanation above, the set $\mathcal{J}$ is well-defined. We introduce the* reduced index $\widetilde{\beta} := (\beta_j)_{I_j \not\subseteq \mathcal{J}} \in \{0,1\}^N$. *For each class $\mathcal{K}$, there is a one-to-one correspondence between $\beta \in \mathcal{K}$ and a reduced index $\widetilde{\beta} \in \{0,1\}^N$. Thus, we slightly abuse the notation and denote $U_{\widetilde{\beta}} \equiv U_\beta$, whenever the equivalence class $\mathcal{K}$ is clear from the context. For the reduced index, we define a* partial order $\succ$ *as follows: if $\widetilde{\beta}'_j \geq \widetilde{\beta}_j$ for all $1 \leq j \leq N$ (namely, $\nabla U_{\beta'}(x) \geq \nabla U_\beta(x)$ for all $x \in \mathbb{R}$), then we denote $\widetilde{\beta}' \succ \widetilde{\beta}$ (or $\widetilde{\beta} \prec \widetilde{\beta}'$).*

Recall from the Step (1) above that $\mathscr{I} = \left[\frac{C_x}{8N}, \frac{3C_x}{8N}\right]$ with length $\mu(\mathscr{I}) = \frac{C_x}{4N}$. By Proposition 3.1 with $\mathcal{I} = \bigcup_{j:\ \widetilde{\beta}'_j > \widetilde{\beta}_j} (x_j + \mathscr{I})$, we immediately have the following result.

LEMMA 3.4. *If $\widetilde{\beta}' \succ \widetilde{\beta}$, then*

$$\begin{aligned}
|X_T(U_{\beta'}, \omega) - X_T(U_\beta, \omega)| &\geq \overline{C}\epsilon \frac{C_x}{4N} \#\{j:\ \widetilde{\beta}'_j > \widetilde{\beta}_j\} \\
&= \frac{C_x^2 \overline{C}\xi}{32N^2} \#\{j:\ \widetilde{\beta}'_j > \widetilde{\beta}_j\}.
\end{aligned} \tag{3.11}$$

Let us introduce the following set, for any integer $0 \leq k \leq N$,

$$\mathcal{M}_k := \left\{\widetilde{\beta} \in \{0,1\}^N: \sum_{j=1}^N \widetilde{\beta}_j = k\right\}.$$

LEMMA 3.5. *Consider even integer $N \geq 2$. For any integer $0 \leq k \leq N/2$, there exists a bijective map $\Upsilon: \mathcal{M}_k \to \mathcal{M}_{N-k}$ such that for any $\widetilde{\beta} \in \mathcal{M}_k$, we have $\widetilde{\beta} \prec \Upsilon(\widetilde{\beta})$. In particular, when $k = N/2$, $\Upsilon(\widetilde{\beta}) = \widetilde{\beta}$.*

Proof. This lemma follows immediately from the symmetric chain decomposition (SCD) for Boolean lattices; see, e.g., [12, 56] for an introduction, as well as proofs. A symmetric chain is a sequence $\gamma^{(n)} \prec \gamma^{(n+1)} \prec \cdots \prec \gamma^{(N-n)}$ where $\gamma^{(j)} \in \mathcal{M}_j$ for $n \leq j \leq N-n$. SCD states that the set $\{0,1\}^N$ can be decomposed into disjoint symmetric chains. Therefore, for any $\widetilde{\beta} \in \mathcal{M}_k$, it must belong to a particular chain, say $\gamma^{(n)} \prec \gamma^{(n+1)} \prec \cdots \prec \gamma^{(N-n)}$. By the definition of symmetric chains, we have $\gamma^{(k)} = \widetilde{\beta}$, and then we can simply define $\Upsilon(\widetilde{\beta}) := \gamma^{(N-k)} \succ \gamma^{(k)} \equiv \widetilde{\beta}$. Such a procedure is always valid, and since all symmetric chains are disjoint, Υ is a bijective map. $\square$

With these preparations, we can now continue to finish the proof of (3.10), and thus the lower bound in Theorem 1.1 for the case $d = 1$. Within any equivalence class $\mathcal{K}$, we have

$$\begin{aligned}
\sum_{\beta \in \mathcal{K}} |X_T(U_\beta, \omega) - A(U_\beta, \omega, \widetilde{\omega})|^2 &= \sum_{\widetilde{\beta} \in \{0,1\}^N} |X_T(U_{\widetilde{\beta}}, \omega) - A(U_{\widetilde{\beta}}, \omega, \widetilde{\omega})|^2 \\
&\geq \sum_{k=0}^{\frac{N}{2}-1} \sum_{\widetilde{\beta} \in M_k} |X_T(U_{\widetilde{\beta}}, \omega) - A(U_{\widetilde{\beta}}, \omega, \widetilde{\omega})|^2 + |X_T(U_{\Upsilon(\widetilde{\beta})}, \omega) - A(U_{\Upsilon(\widetilde{\beta})}, \omega, \widetilde{\omega})|^2
\end{aligned}$$

$$\gtrsim \sum_{k=0}^{\frac{N}{2}-1} \sum_{\widetilde{\beta} \in M_k} |X_T(U_{\widetilde{\beta}},\omega) - X_T(U_{\Upsilon(\widetilde{\beta})},\omega)|^2 \overset{(3.11)}{\gtrsim} \sum_{k=0}^{\frac{N}{2}-1} \binom{N}{k} \frac{C_x^4 \overline{C}^2 \xi^2}{N^4} (N-2k)^2$$

$$\gtrsim \frac{C_x^4 \overline{C}^2 \xi^2}{N^4} \sum_{k=0}^{\frac{N}{2}} \binom{N}{k} \left(\frac{N}{2}-k\right)^2 \gtrsim \frac{C_x^4 \overline{C}^2 \xi^2}{N^4} \sum_{k=0}^{N} \binom{N}{k} \left(\frac{N}{2}-k\right)^2 \gtrsim \frac{C_x^4 \overline{C}^2 \xi^2}{N^3} 2^N.$$

Finally, we have

$$\mathbb{E}_{U_\beta \sim \mu_{\mathcal{G}}}\left[|X_T(U_\beta,\omega) - A(U_\beta,\omega,\widetilde{\omega})|^2\right] = \frac{1}{2^{2N}} \sum_{\mathcal{K}_j} \sum_{\beta \in \mathcal{K}_j} |X_T(U_\beta,\omega) - A(U_\beta,\omega,\widetilde{\omega})|^2$$

$$\gtrsim \frac{1}{2^{2N}} \times 2^N \times \frac{C_x^4 \overline{C}^2 \xi^2}{N^3} 2^N \gtrsim \frac{C_x^4 \overline{C}^2 \xi^2}{N^3}.$$

Thus we complete the proof of the lower bound in Theorem 1.1 for the case $d=1$.

Case (II): General dimension d. For a general dimension d, we can choose $U(x) = U(x_1, x_2, \cdots, x_d) = \sum_{j=1}^{d} U^{(j)}(x_j)$, where $U^{(j)}: \mathbb{R} \to \mathbb{R}$, $x = (x_1, x_2, \cdots, x_d)$, and $x_j \in \mathbb{R}$. If $U^{(j)} \in \mathcal{F}(1,\ell,L)$, then $U \in \mathcal{F}(d,\ell,L)$. Note that if U takes this particular form, then each component of the underdamped Langevin dynamics (1.1) is evolving independently. Then we immediately know

$$e_{\mathcal{F},\Lambda}(A)^2 \geq \sup_{U:\ U(x) = \sum_{j=1}^{d} U^{(j)}(x_j)} \mathbb{E}_{(\omega,\widetilde{\omega}) \sim \mathbb{P} \times \widetilde{\mathbb{P}}}\left[|X_T(U,\omega) - A(U,\omega,\widetilde{\omega})|^2\right]$$

$$= \sup_{U:\ U(x) = \sum_{j=1}^{d} U^{(j)}(x_j)} \sum_{j=1}^{d} \mathbb{E}_{(\omega,\widetilde{\omega}) \sim \mathbb{P} \times \widetilde{\mathbb{P}}}\left[|X_T^{(j)}(U,\omega) - A^{(j)}(U,\omega,\widetilde{\omega})|^2\right]$$

$$\gtrsim\ dC_{\text{low}}^2 N^{-3},$$

where $X_T^{(j)}(U,\omega) \in \mathbb{R}$ is the jth component of $X_T(U,\omega)$, and $A^{(j)}(U,\omega,\widetilde{\omega})$ is the jth component of the algorithmic prediction $A(U,\omega,\widetilde{\omega})$. Therefore, we obtain $e_{\mathcal{F},\Lambda}(A) \gtrsim C_{\text{low}}\sqrt{d}N^{-3/2}$. The last inequality above is intuitively reasonable, since $A^{(j)}(U,\omega,\widetilde{\omega})$ could be regarded as an algorithm of the jth component, and having queries to the information from independent components, like $\nabla_{x_k} U^{(k)}(x_k)$ ($k \neq j$), would not improve the algorithmic prediction for the jth component.

More rigorously, one could directly generalize the proof of the Case (I). Below is a sketch of the only few technical differences. First, similar to (3.9),

$$e_{\mathcal{F},\Lambda}(A)^2 \geq \sum_{j=1}^{d} \mathbb{E}_{(\omega,\widetilde{\omega}) \sim \mathbb{P} \times \widetilde{\mathbb{P}}}\ \mathbb{E}_{U \sim \mu_{\mathcal{G}_1} \times \mu_{\mathcal{G}_2} \times \cdots \times \mu_{\mathcal{G}_d}}\left[|X_T^{(j)}(U,\omega) - A^{(j)}(U,\omega,\widetilde{\omega})|^2\right],$$

where $\mu_{\mathcal{G}_j}$ is a uniform measure of $U^{(j)} \in \mathcal{G}_j$, similar to (3.8), for $1 \leq j \leq d$. Without loss of generality, we only need to consider the component $j=1$, and we need to show that

$$\mathbb{E}_{(\omega,\widetilde{\omega}) \sim \mathbb{P} \times \widetilde{\mathbb{P}}}\ \mathbb{E}_{U \sim \mu_{\mathcal{G}_1} \times \mu_{\mathcal{G}_2} \times \cdots \times \mu_{\mathcal{G}_d}}\left[|X_T^{(1)}(U,\omega) - A^{(1)}(U,\omega,\widetilde{\omega})|^2\right] \gtrsim C_{\text{low}}^2 N^{-3}.$$

We shall fix $\omega \in \mathcal{E}$, where $\mathcal{E}$ is now defined in the same way as (3.3) by considering the components $X_t^{(1)}(U,\omega)$ and $V_t^{(1)}(U,\omega)$ only. We shall also fix $\widetilde{\omega} \in \widetilde{\mathbb{M}}$ and $U^{(j)}$ for $j \geq 2$. Then it suffices to prove that (cf. (3.10))

$$\mathbb{E}_{U^{(1)} \sim \mu_{\mathcal{G}_1}}\left[|X_T^{(1)}(U^{(1)},\omega) - A^{(1)}(U^{(1)},\omega,\widetilde{\omega})|^2\right] \gtrsim \frac{C_x^4 \overline{C}^2 \xi^2}{N^3}.$$

The proof of this inequality is essentially the same as the Step (2) in the Case (I). The only minor difference is that $\mathcal{J}^\beta$ is now defined as the completed union of sub-intervals, wherein the first components of $Y_1^\beta, Y_2^\beta, \cdots, Y_N^\beta$ reside.

3.3. Proof of results in Section 3.1.

Proof. (**Proof of Lemma 3.1.**)   Consider any fixed $U \in \mathcal{F}_{u,\epsilon}$ and any fixed $\omega \in \mathbb{M}$. Let us introduce $g(x) := \nabla U(x) - ux$, $\Delta_{X,t} := X_t(U,\omega) - X_t(U_u,\omega)$, and $\Delta_{V,t} := V_t(U,\omega) - V_t(U_u,\omega)$. By assumption, $\|g\|_\infty \leq \epsilon$ and $\Delta_{X,0} = \Delta_{V,0} = 0$. By (1.1), it is straightforward to derive that

$$\mathrm{d}\begin{bmatrix}\Delta_{X,t}\\\Delta_{V,t}\end{bmatrix} = H\begin{bmatrix}\Delta_{X,t}\\\Delta_{V,t}\end{bmatrix}\mathrm{d}t + \begin{bmatrix}0\\-\frac{1}{L}g(X_t(U,\omega))\end{bmatrix}\mathrm{d}t, \qquad H = \begin{bmatrix}0 & 1\\-\frac{u}{L} & -2\end{bmatrix}.$$

Then we could rewrite the above equation in the integral form,

$$\begin{bmatrix}\Delta_{X,t}\\\Delta_{V,t}\end{bmatrix} = \int_0^t e^{H(t-s)}\begin{bmatrix}0\\-\frac{g(X_s(U,\omega))}{L}\end{bmatrix}\mathrm{d}s.$$

Hence, by introducing $g_s \equiv g(X_s(U,\omega))$, and recalling $\lambda_\pm$ from (2.7), we have

$$\Delta_{X,t} = -\frac{1}{2\sqrt{L(L-u)}}\int_0^t g_s\left(e^{-(t-s)\lambda_-} - e^{-(t-s)\lambda_+}\right)\mathrm{d}s,$$

$$\Delta_{V,t} = -\frac{1}{2\sqrt{L(L-u)}}\int_0^t g_s\left(\lambda_+ e^{-(t-s)\lambda_+} - \lambda_- e^{-(t-s)\lambda_-}\right)\mathrm{d}s. \tag{3.12}$$

Since $|g_s| \leq \epsilon$, it is straightforward to obtain that

$$|\Delta_{X,t}| \leq \frac{\epsilon}{2\sqrt{L(L-u)}}\int_0^t |e^{-(t-s)\lambda_-} - e^{-(t-s)\lambda_+}|\mathrm{d}s$$

$$= \frac{\epsilon}{2\sqrt{L(L-u)}}\left(\frac{1 - e^{-\lambda_- t}}{\lambda_-} - \frac{1 - e^{-\lambda_+ t}}{\lambda_+}\right)$$

$$\leq \frac{\epsilon}{2\lambda_-\sqrt{L(L-u)}} = \frac{\epsilon}{2L(1-\sqrt{1-\frac{u}{L}})\sqrt{1-\frac{u}{L}}}.$$

Similarly, for $|\Delta_{V,t}|$, we have

$$|\Delta_{V,t}| \leq \frac{\epsilon}{2\sqrt{L(L-u)}}\int_0^t |\lambda_+ e^{-(t-s)\lambda_+} - \lambda_- e^{-(t-s)\lambda_-}|\mathrm{d}s$$

$$\leq \frac{\epsilon}{2\sqrt{L(L-u)}}\int_0^t \lambda_+ e^{-(t-s)\lambda_+} + \lambda_- e^{-(t-s)\lambda_-}\,\mathrm{d}s$$

$$= \frac{\epsilon}{2\sqrt{L(L-u)}}(1 - e^{-\lambda_+ t} + 1 - e^{-\lambda_- t}) \leq \frac{\epsilon}{L\sqrt{1-\frac{u}{L}}}.$$

$\square$

Proof. (**Proof of Lemma 3.2.**)   For any $C_x, C_v > 0$, let us pick any $0 < \epsilon < \bar{\epsilon}$. Notice that

$$\sup_{0 \leq t \leq T} X_t(U_u,\omega) \geq 2C_x \geq C_x + \frac{\epsilon}{2L(1-\sqrt{1-\frac{u}{L}})\sqrt{1-\frac{u}{L}}}$$

implies that $\sup_{0 \le t \le T} X_t(U, \omega) \ge C_x$ by (3.2), and likewise for the other two cases. We have

$$\mathbb{P}\Big(\omega \colon \sup_{0 \le t \le T} X_t(U, \omega) \ge C_x, \ \inf_{0 \le t \le T} X_t(U, \omega) \le -C_x,$$

$$\sup_{0 \le t \le T} |V_t(U, \omega)| \le C_v, \ \forall U \in \mathcal{F}_{u, \epsilon}\Big)$$

$$\ge \mathbb{P}\Big(\omega \colon \sup_{0 \le t \le T} X_t(U_u, \omega) \ge 2C_x, \ \inf_{0 \le t \le T} X_t(U_u, \omega) \le -2C_x,$$

$$\sup_{0 \le t \le T} |V_t(U_u, \omega)| \le C_v/2\Big).$$

Finally, recall the expression of $\mathsf{P}(C_x, C_v, u, T)$ from (2.9). □

Proof. **(Proof of Proposition 3.1.)** We shall fix $\omega \in \mathcal{E}$ throughout this proof. Let $\Delta_{X,t} := X_t(U_1, \omega) - X_t(U_2, \omega)$ and $\Delta_{V,t} := V_t(U_1, \omega) - V_t(U_2, \omega)$. Then by (1.1), we have

$$\mathrm{d}\begin{bmatrix} \Delta_{X,t} \\ \Delta_{V,t} \end{bmatrix} = H_t \begin{bmatrix} \Delta_{X,t} \\ \Delta_{V,t} \end{bmatrix} \mathrm{d}t + \begin{bmatrix} 0 \\ -\frac{1}{L} g(X_t(U_1, \omega)) \end{bmatrix} \mathrm{d}t, \qquad H_t = \begin{bmatrix} 0 & 1 \\ -\frac{u_t}{L} & -2 \end{bmatrix}, \tag{3.13}$$

with initial conditions $\Delta_{X,0} = \Delta_{V,0} = 0$, where $u. \colon t \to u_t$ is a continuous function of time such that

$$\nabla U_2(X_t(U_1, \omega)) - \nabla U_2(X_t(U_2, \omega)) = u_t \Delta_{X,t}.$$

To see why u_t is well-defined, notice that $U_2 \in \mathcal{F}$ is a $C^2(\mathbb{R})$ function (recall $\mathcal{F}$ in (1.2)). Then by the first-order Taylor's expansion,

$$\nabla U_2(X_t(U_1, \omega)) - \nabla U_2(X_t(U_2, \omega)) = \nabla^2 U_2(\vartheta)(X_t(U_1, \omega) - X_t(U_2, \omega)),$$

for some value ϑ between $X_t(U_1, \omega)$ and $X_t(U_2, \omega)$; we simply let $u_t = \nabla^2 U_2(\vartheta)$. Moreover, we could easily observe that $\ell \le u_t \le u_R \le L$. For simplicity, we shall again denote

$$g_t \equiv g(X_t(U_1, \omega)),$$

here and below.

Intuitively, the ODE dynamics (3.13) consists of two parts: the contraction part and the source part. Suppose $u_t \equiv u$ is independent of time, then it is easy to observe that e^{Ht} is a contraction operator for large enough time t, and the ODE dynamics (3.13) with $g \equiv 0$ will converge to the origin exponentially fast, for any initial condition; the source part $\begin{bmatrix} 0 \\ -\frac{g_t}{L} \end{bmatrix}$ will drag the velocity term (i.e., $\Delta_{V,t}$) towards the negative direction. Under the assumption that $\omega \in \mathcal{E}$, the term g_t takes non-zero value at least for a period of $\mu(\mathcal{I})/C_v$.

When $u_t = u$ for all t, (3.13) has an explicit solution shown below, similar to (2.8) and (3.12) above.

LEMMA 3.6.   *Suppose $u_t = u$ for all $t \in [0, T]$ in (3.13), then $\Delta_{X,T} = \mathcal{S}(u, T)$, where*

$$\mathcal{S}(u, T) := \begin{cases} \displaystyle\int_0^T \frac{e^{(t-T)(1+\sqrt{1-u/L})} - e^{(t-T)(1-\sqrt{1-u/L})}}{2\sqrt{L(L-u)}} g_t \, \mathrm{d}t, & \textit{if } \ell \le u < L; \\[4mm] \displaystyle\int_0^T \frac{(t-T)e^{t-T}}{L} g_t \, \mathrm{d}t, & \textit{if } u = L. \end{cases}$$

Observe that $\Delta_{X,T} \leq 0$ for both cases, which inspires us to propose the following general result.

LEMMA 3.7. *The region characterized by $\Delta_{V,t} \leq -\Delta_{X,t}$ and $\Delta_{X,t} \leq 0$ forms a trapping region for the dynamics (3.13). Therefore, the quantity $\Delta_{X,t} \equiv X_t(U_1, \omega) - X_t(U_2, \omega) \leq 0$ for any $t \in [0, T]$.*

Proof. To prove that the region formed by $\Delta_{V,t} \leq -\Delta_{X,t}$ and $\Delta_{X,t} \leq 0$ is a trapping region for the ODE dynamics (3.13), we consider the following three cases at the boundary:

- ($\Delta_{X,t} = 0$ and $\Delta_{V,t} = 0$). We know $\frac{\mathrm{d}}{\mathrm{d}t}\Delta_{X,t} = 0$ and $\frac{\mathrm{d}}{\mathrm{d}t}\Delta_{V,t} = -\frac{1}{L}g_t \leq 0$. Thus, the solution of $\Delta_{X,t}$ and $\Delta_{V,t}$ will not escape the trapping region.

- ($\Delta_{X,t} = 0$ and $\Delta_{V,t} < 0$, i.e., the negative half-line of the velocity-axis). We know

$$(\frac{\mathrm{d}}{\mathrm{d}t}\Delta_{X,t}, \frac{\mathrm{d}}{\mathrm{d}t}\Delta_{V,t}) \cdot (-1,0) = (\Delta_{V,t}, -2\Delta_{V,t} - \frac{g_t}{L}) \cdot (-1,0) = -\Delta_{V,t} > 0.$$

Therefore, the solution of $\Delta_{X,t}$ and $\Delta_{V,t}$ will not escape the trapping region from the negative half-line of the velocity-axis.

- ($\Delta_{V,t} = -\Delta_{X,t}$ for $\Delta_{X,t} < 0$).

$$(\frac{\mathrm{d}}{\mathrm{d}t}\Delta_{X,t}, \frac{\mathrm{d}}{\mathrm{d}t}\Delta_{V,t}) \cdot (-1,-1) = (\Delta_{V,t}, -\frac{u_t}{L}\Delta_{X,t} - 2\Delta_{V,t} - \frac{g_t}{L}) \cdot (-1,-1)$$
$$= (u_t/L - 1)\Delta_{X,t} + \frac{g_t}{L} \geq 0.$$

By summarizing the above three cases, we conclude that the region formed by $\Delta_{V,t} \leq -\Delta_{X,t}$ and $\Delta_{X,t} \leq 0$ is indeed a trapping region. Since $\Delta_{X,0} = \Delta_{V,0} = 0$, we know that $\Delta_{X,t} \leq 0$ for any time $t \geq 0$. □

By the above lemma (i.e., $\Delta_{X,t} \leq 0$ for any time $t \in [0, T]$), we know that

$$\frac{\mathrm{d}}{\mathrm{d}t}\Delta_{X,t} = \Delta_{V,t},$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\Delta_{V,t} = -\frac{u_t}{L}\Delta_{X,t} - 2\Delta_{V,t} - \frac{g_t}{L} \leq -\frac{u_R}{L}\Delta_{X,t} - 2\Delta_{V,t} - \frac{g_t}{L}. \tag{3.14}$$

LEMMA 3.8. *We claim that in general,*

$$\Delta_{X,T} \leq \mathcal{S}(u_R, T). \tag{3.15}$$

Proof. Let $(\Delta_{X,t}^{(1)}, \Delta_{V,t}^{(1)})$ be the solution of (3.13), and let $(\Delta_{X,t}^{(2)}, \Delta_{V,t}^{(2)})$ be the solution of (3.13) for $u_t \equiv u_R$. Then let us introduce $\Gamma_{X,t} := \Delta_{X,t}^{(1)} - \Delta_{X,t}^{(2)}$ and $\Gamma_{V,t} := \Delta_{V,t}^{(1)} - \Delta_{V,t}^{(2)}$. By (3.14), we immediately have

$$\frac{\mathrm{d}}{\mathrm{d}t}\Gamma_{X,t} = \Gamma_{V,t}, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\Gamma_{V,t} \leq -\frac{u_R}{L}\Gamma_{X,t} - 2\Gamma_{V,t}.$$

Since $\Gamma_{X,0} = \Gamma_{V,0} = 0$, by the same argument as in Lemma 3.7, we know that $\Gamma_{X,t} \leq 0$ for any $t \geq 0$. Therefore, $\Delta_{X,T}^{(1)} \leq \Delta_{X,T}^{(2)} \equiv \mathcal{S}(u_R, T)$. □

Let us consider the case $u_R < L$. Then we have

$$
\begin{aligned}
|\Delta_{X,T}| &= -\Delta_{X,T} \\
&\overset{(3.15)}{\geq} \int_0^T \frac{e^{(t-T)(1-\sqrt{1-u_R/L})} - e^{(t-T)(1+\sqrt{1-u_R/L})}}{2\sqrt{L(L-u_R)}} g_t \, dt \\
&\overset{(3.4)}{\geq} \frac{\epsilon}{4L\sqrt{1-\frac{u_R}{L}}} \int_0^T e^{(t-T)(1-\sqrt{1-u_R/L})} \times \\
&\qquad\qquad \left(1 - e^{2(t-T)\sqrt{1-u_R/L}}\right) \mathbb{1}_{\mathcal{I}}(X_t(U_1,\omega)) \, dt.
\end{aligned}
$$

Without loss of generality, we assume $\tau_1(\omega) < \tau_2(\omega)$, where τ_1 is the first hitting time of $X_t(U_1,\omega)$ to $-C_x$ and τ_2 is the first hitting time to C_x. Since we assume $\omega \in \mathcal{E}$, both τ_1 and τ_2 are well-defined and $0 \leq \tau_1, \tau_2 \leq T$. Then

$$
\begin{aligned}
&|\Delta_{X,T}| \\
&\geq \frac{\epsilon}{4L\sqrt{1-\frac{u_R}{L}}} \int_{\tau_1}^{\tau_2} e^{(t-T)(1-\sqrt{1-\frac{u_R}{L}})} \left(1 - e^{2(t-T)\sqrt{1-\frac{u_R}{L}}}\right) \mathbb{1}_{\mathcal{I}}(X_t(U_1,\omega)) \, dt \\
&= \frac{\epsilon}{4L\sqrt{1-\frac{u_R}{L}}} \int_{\tau_1+\frac{C_x}{2C_v}}^{\tau_2-\frac{C_x}{2C_v}} e^{(t-T)(1-\sqrt{1-\frac{u_R}{L}})} \left(1 - e^{2(t-T)\sqrt{1-\frac{u_R}{L}}}\right) \mathbb{1}_{\mathcal{I}}(X_t(U_1,\omega)) \, dt \\
&\geq \frac{\epsilon\left(1 - e^{2(\tau_2-\frac{C_x}{2C_v}-T)\sqrt{1-\frac{u_R}{L}}}\right)}{4L\sqrt{1-\frac{u_R}{L}}} \int_{\tau_1+\frac{C_x}{2C_v}}^{\tau_2-\frac{C_x}{2C_v}} e^{(t-T)(1-\sqrt{1-\frac{u_R}{L}})} \mathbb{1}_{\mathcal{I}}(X_t(U_1,\omega)) \, dt \\
&\geq \frac{\epsilon\left(1 - e^{-\frac{C_x}{C_v}\sqrt{1-\frac{u_R}{L}}}\right)}{4L\sqrt{1-\frac{u_R}{L}}} \int_{\tau_1+\frac{C_x}{2C_v}}^{\tau_2-\frac{C_x}{2C_v}} e^{(t-T)(1-\sqrt{1-\frac{u_R}{L}})} \mathbb{1}_{\mathcal{I}}(X_t(U_1,\omega)) \, dt,
\end{aligned}
$$

where we use the following observation in the second equality: since the velocity $V_t(U_1,\omega)$ is bounded by C_v, starting from the time τ_1 (note that $X_{\tau_1}(U_1,\omega) = -C_x$), it takes at least $C_x/(2C_v)$ amount of time to reach $-C_x/2$. Thus, we have $\mathbb{1}_{\mathcal{I}}(X_t(U_1,\omega)) = 0$ for $t \in \left[\tau_1, \tau_1 + \frac{C_x}{2C_v}\right]$ by the assumption (ii) that $\mathcal{I} \subseteq [-C_x/2, C_x/2]$ (similarly for the time period $\left[\tau_2 - \frac{C_x}{2C_v}, \tau_2\right]$).

By the fact that $\mu(t: X_t(U_1,\omega) \in \mathcal{I}) \geq \frac{\mu(\mathcal{I})}{C_v}$ (namely, $\mathbb{1}_{\mathcal{I}}(X_t(U_1,\omega)) = 1$ for at least $\mu(\mathcal{I})/C_v$ period of time), we have

$$
\begin{aligned}
&|\Delta_{X,T}| \\
&\geq \frac{\epsilon\left(1 - e^{-\frac{C_x}{C_v}\sqrt{1-\frac{u_R}{L}}}\right)}{4L\sqrt{1-\frac{u_R}{L}}} \int_{[\tau_1+\frac{C_x}{2C_v}, \tau_2-\frac{C_x}{2C_v}] \cap \{t:\, X_t(U_1,\omega) \in \mathcal{I}\}} e^{(t-T)(1-\sqrt{1-\frac{u_R}{L}})} \, dt \\
&\geq \frac{\epsilon\left(1 - e^{-\frac{C_x}{C_v}\sqrt{1-\frac{u_R}{L}}}\right)}{4L\sqrt{1-\frac{u_R}{L}}} \int_{\tau_1+\frac{C_x}{2C_v}}^{\tau_1+\frac{C_x}{2C_v}+\frac{\mu(\mathcal{I})}{C_v}} e^{(t-T)(1-\sqrt{1-\frac{u_R}{L}})} \, dt \\
&= \frac{\epsilon\left(1 - e^{-\frac{C_x}{C_v}\sqrt{1-\frac{u_R}{L}}}\right)}{4L\sqrt{1-\frac{u_R}{L}}} e^{(\tau_1+\frac{C_x}{2C_v}-T)(1-\sqrt{1-\frac{u_R}{L}})} \frac{e^{\frac{\mu(\mathcal{I})}{C_v}(1-\sqrt{1-\frac{u_R}{L}})} - 1}{1 - \sqrt{1-\frac{u_R}{L}}} \\
&\geq \frac{\epsilon\left(1 - e^{-\frac{C_x}{C_v}\sqrt{1-\frac{u_R}{L}}}\right)}{4L\sqrt{1-\frac{u_R}{L}}} e^{(\tau_1+\frac{C_x}{2C_v}-T)(1-\sqrt{1-\frac{u_R}{L}})} \frac{\mu(\mathcal{I})}{C_v}
\end{aligned}
$$

$$\geq \frac{\epsilon\mu(\mathcal{I})\big(1-e^{-\frac{C_x}{C_v}\sqrt{1-\frac{u_R}{L}}}\big)}{4LC_v\sqrt{1-\frac{u_R}{L}}}e^{(\frac{3C_x}{2C_v}-T)(1-\sqrt{1-\frac{u_R}{L}})},$$

where to get the final inequality, we use the fact that $\tau_1 \geq \frac{C_x}{C_v}$ by the same velocity-type argument. Then we finish the proof for the case $u_R < L$. As for $u_R = L$, one could either follow a similar approach like above, or simply pass the limit $u_R \to L$ to obtain the expression of $\overline{C}$.   □

3.4. Proof of the upper bound estimate for Theorem 1.1. The upper bound estimate is based on Ref. [50]. In the following, we shall provide a sketch only. We consider a non-adaptive mesh grid, i.e., $s_k = kh$ with $h = T/N_s$ as in [50].

Suppose $\hat{X}_k$, $\hat{X}_{k+1/2}$ and $\hat{V}_k$ are given by the randomized midpoint method as in (2.6), and suppose $\widetilde{X}_{k+1}$ and $\widetilde{V}_{k+1}$ are solutions of (1.1) at time s_{k+1}, conditioned on $X_{s_k} = \hat{X}_k$ and $V_{s_k} = \hat{V}_k$ at time s_k. By [50, Appendix E], we have

$$\mathbb{E}\big[\|\hat{X}_{N_s} - X_T\|^2 + \|\hat{X}_{N_s} + \hat{V}_{N_s} - X_T - V_T\|^2\big]$$
$$\leq e^{-N_s\frac{h}{2\varkappa}}\mathbb{E}\big[\|\hat{X}_0 - X_0\|^2 + \|\hat{X}_0 + \hat{V}_0 - X_0 - V_0\|^2\big]$$
$$+ \frac{2\varkappa}{h}\sum_{k=0}^{N_s-1}\Big(3\mathbb{E}\big\|\mathbb{E}_{\eta_k}[\hat{X}_{k+1} - \widetilde{X}_{k+1}]\big\|^2 + 2\mathbb{E}\big\|\mathbb{E}_{\eta_k}[\hat{V}_{k+1} - \widetilde{V}_{k+1}]\big\|^2\Big)$$
$$+ \sum_{k=0}^{N_s-1}\Big(3\mathbb{E}\|\hat{X}_{k+1} - \widetilde{X}_{k+1}\|^2 + 2\mathbb{E}\|\hat{V}_{k+1} - \widetilde{V}_{k+1}\|^2\Big)$$
$$= \frac{2\varkappa}{h}\sum_{k=0}^{N_s-1}\Big(3\mathbb{E}\big\|\mathbb{E}_{\eta_k}[\hat{X}_{k+1} - \widetilde{X}_{k+1}]\big\|^2 + 2\mathbb{E}\big\|\mathbb{E}_{\eta_k}[\hat{V}_{k+1} - \widetilde{V}_{k+1}]\big\|^2\Big)$$
$$+ \sum_{k=0}^{N_s-1}\Big(3\mathbb{E}\|\hat{X}_{k+1} - \widetilde{X}_{k+1}\|^2 + 2\mathbb{E}\|\hat{V}_{k+1} - \widetilde{V}_{k+1}\|^2\Big),$$

where in the last step, we use the fact that $\hat{X}_0 = X_0$ and $\hat{V}_0 = V_0$. In the above, $\mathbb{E}_{\eta_k}$ means expectation with respect to η_k for $0 \leq k \leq N_s - 1$.

By [50, Lemma 2], we have

$$\mathbb{E}\big[\|\hat{X}_{N_s} - X_T\|^2 + \|\hat{X}_{N_s} + \hat{V}_{N_s} - X_T - V_T\|^2\big]$$
$$\lesssim \frac{2\varkappa}{h}\big(h^8\sum_{k=0}^{N_s-1}\mathbb{E}\|\hat{V}_k\|^2 + \frac{h^{10}}{L^2}\sum_{k=0}^{N_s-1}\mathbb{E}\|\nabla U(\hat{X}_k)\|^2 + \frac{N_s dh^9}{L}\big)$$
$$+ h^4\sum_{k=0}^{N_s-1}\mathbb{E}\|\hat{V}_k\|^2 + \frac{h^4}{L^2}\sum_{k=0}^{N_s-1}\mathbb{E}\|\nabla U(\hat{X}_k)\|^2 + \frac{N_s dh^5}{L}$$
$$\lesssim (h^4 + h^7\varkappa)\sum_{k=0}^{N_s-1}\mathbb{E}\|\hat{V}_k\|^2 + \big(\frac{h^9}{\ell L} + \frac{h^4}{L^2}\big)\sum_{k=0}^{N_s-1}\mathbb{E}\|\nabla U(\hat{X}_k)\|^2 + \big(\frac{N_s dh^8}{\ell} + \frac{N_s dh^5}{L}\big).$$

Next, we use [50, Lemma 12], and obtain that

$$\mathbb{E}\big[\|\hat{X}_{N_s} - X_T\|^2 + \|\hat{X}_{N_s} + \hat{V}_{N_s} - X_T - V_T\|^2\big]$$
$$\lesssim (h^4 + h^7\varkappa)\Big(\frac{N_s d}{L} + \frac{1}{L}\big|\mathbb{E}\big[\langle\nabla U(\hat{X}_{N_s}), \hat{V}_{N_s}\rangle\big]\big|\Big)$$

$$+\left(\frac{h^9}{\ell L}+\frac{h^4}{L^2}\right)\left(N_sLd+\frac{L}{h}\Big|\mathbb{E}\big[\langle\nabla U(\hat{X}_{N_s}),\hat{V}_{N_s}\rangle\big]\Big|\right)+\left(\frac{N_sdh^8}{\ell}+\frac{N_sdh^5}{L}\right)$$
$$\lesssim\frac{h^3}{L}\Big|\mathbb{E}\big[\langle\nabla U(\hat{X}_{N_s}),\hat{V}_{N_s}\rangle\big]\Big|+\frac{h^4N_sd}{L},$$

where in the last step, we use the fact that we are working on the L^2-strong error estimate and h is the small parameter herein. Then we need to estimate $\Big|\mathbb{E}\big[\langle\nabla U(\hat{X}_{N_s}),\hat{V}_{N_s}\rangle\big]\Big|$. Similar to [50, Appendix E],

$$\Big|\mathbb{E}\big[\langle\nabla U(\hat{X}_{N_s}),\hat{V}_{N_s}\rangle\big]\Big|$$
$$\lesssim\frac{1}{L}\mathbb{E}\big[\|\nabla U(\hat{X}_{N_s})\|^2\big]+L\mathbb{E}\big[\|\hat{V}_{N_s}\|^2\big]$$
$$\lesssim L\mathbb{E}\big[\|\hat{V}_{N_s}-V_T\|^2+\|\hat{X}_{N_s}-X_T\|^2\big]+L\mathbb{E}\big[\|V_T\|^2\big]+\frac{1}{L}\mathbb{E}\big[\|\nabla U(X_T)\|^2\big]$$
$$\lesssim L\mathbb{E}\big[\|\hat{X}_{N_s}-X_T\|^2+\|\hat{X}_{N_s}+\hat{V}_{N_s}-X_T-V_T\|^2\big]$$
$$\quad+L\mathbb{E}\big[\|V_T\|^2\big]+\frac{1}{L}\mathbb{E}\big[\|\nabla U(X_T)\|^2\big].$$

By combining the last two equations,

$$\mathbb{E}\big[\|\hat{X}_{N_s}-X_T\|^2\big]\leq\mathbb{E}\big[\|\hat{X}_{N_s}-X_T\|^2+\|\hat{X}_{N_s}+\hat{V}_{N_s}-X_T-V_T\|^2\big]$$
$$\lesssim\frac{h^3}{L}\left(L\mathbb{E}\big[\|V_T\|^2\big]+\frac{1}{L}\mathbb{E}\big[\|\nabla U(X_T)\|^2\big]\right)+\frac{h^4N_sd}{L}$$
$$\lesssim h^3\left(\mathbb{E}\big[\|V_T\|^2\big]+\mathbb{E}\big[\|X_T-x^\star\|^2\big]\right)+\frac{h^4N_sd}{L}.$$

Suppose (Y_t,Z_t) is another solution of the underdamped Langevin dynamics (1.1) with the initial distribution as ρ_∞. Then similar to [50, Appendix E],

$$\mathbb{E}\big[\|V_T\|^2\big]+\mathbb{E}\big[\|X_T-x^\star\|^2\big]$$
$$\lesssim\mathbb{E}\big[\|V_T-Z_T\|^2+\|X_T-Y_T\|^2\big]+\mathbb{E}\big[\|Z_T\|^2\big]+\mathbb{E}\big[\|Y_T-x^\star\|^2\big]$$
$$\lesssim\mathbb{E}\big[\|V_T-Z_T+X_T-Y_T\|^2+\|X_T-Y_T\|^2\big]+\frac{d}{L}+\frac{d}{\ell}$$
$$\lesssim e^{-\frac{T}{\varkappa}}\mathbb{E}\big[\|V_0-Z_0+X_0-Y_0\|^2+\|X_0-Y_0\|^2\big]+\frac{d}{\ell}\lesssim(e^{-\frac{T}{\varkappa}}+1)\frac{d}{\ell}\lesssim\frac{d}{\ell}.$$

Finally, we have

$$\mathbb{E}\big[\|\hat{X}_{N_s}-X_T\|^2\big]\lesssim\frac{dh^3}{\ell}+\frac{h^4N_sd}{L}\lesssim\frac{d}{N_s^3}\left(\frac{T^3}{\ell}+\frac{T^4}{L}\right).$$

Thus, if the algorithm A is the randomized midpoint method (2.6),

$$e_{\mathcal{F},\Lambda}(A)\lesssim C_{\mathrm{up}}\sqrt{d}N_s^{-3/2}\lesssim C_{\mathrm{up}}\sqrt{d}N^{-3/2},$$

where $C_{\mathrm{up}}=\sqrt{\frac{T^3}{\ell}+\frac{T^4}{L}}$, and recall that the number of queries is $N=2N_s$.

to thank Sam Hopkins (University of Minnesota) for introducing the symmetric chain decomposition [21].

Appendix A. RMM in the framework of randomized algorithms.

Rewriting the RMM (2.6) in the framework of randomized algorithms introduced in Section 2.1 is not technically challenging but many details require much attention, especially notations. Therefore, we would like to present in detail how the RMM fits into the framework of general randomized algorithms for readers' convenience, and also for the purpose of a rigorous proof and completeness.

Let us consider the RMM for N_s steps (time step $h = T/N_s$), which means $N = 2N_s$ evaluations of ∇U and $\widetilde{W}_t^{(0,2)}$. The probability space $\widetilde{\mathbb{M}} = [0,1]^{N_s}$ and $\vec{\eta} = (\eta_0, \eta_2, \cdots, \eta_{N_s-1}) \in \widetilde{\mathbb{M}}$ determines random time steps in (2.6); the σ-algebra $\widetilde{\Sigma}$ can be simply chosen as the Borel σ-algebra of $\widetilde{\mathbb{M}}$; the probability measure $\widetilde{\mathbb{P}}$ corresponds to a uniform distribution on $[0,1]^{N_s}$. Next, we need to rewrite the RMM with fixed $\vec{\eta} \in \widetilde{\mathbb{M}}$ as a deterministic algorithm; see (2.2) and (2.3). With fixed $\vec{\eta}$, the time step prediction is already complete. Therefore, we just need to explain those terms Φ_j and Y_j in (2.2), as well as the forms of φ_j and ϕ.

Step 1: Corresponding Y_j, Φ_j in (2.2).

Let us denote $s_{k+1/2} = s_k + \eta_k h \equiv (k + \eta_k)h$ for integer $0 \le k \le N_s - 1$, and introduce artificial terms $\hat{V}_{k+1/2} := \hat{V}_k$ for convenience of notations below. Let us introduce $t_j = s_{j/2}$ and $Y_j = \hat{X}_{(j-1)/2}$ for integer $1 \le j \le 2N_s$. Recall the notation Φ_j from (2.2b) that

$$\Phi_j := \Big(\Upsilon_{U,\omega}(Y_1, t_1), \Upsilon_{U,\omega}(Y_2, t_2), \cdots, \Upsilon_{U,\omega}(Y_j, t_j) \Big).$$

In the Table A.1 below, we summarize the notations discussed above for clarity.

General framework in Section 2.1	RMM (2.6)
$N = 2N_s$ evaluations	N_s steps
t_j	$s_{j/2}$
Y_j	$\hat{X}_{(j-1)/2}$

TABLE A.1. *This table summarizes the relationship between two sets of notations for the general framework and the RMM.*

Step 2: Choice of φ_j and ϕ. In this part, we will show that $Y_{j+1} \equiv \hat{X}_{j/2}$ is a function of $Y_1 \equiv x^\star(U)$ and Φ_j (i.e., to determine φ_j), and we will also show that $\hat{X}_{N_s} = \phi(Y_1, \Phi_{2N_s})$ for some deterministic function ϕ.

PROPOSITION A.1.   *We can find deterministic mappings $\psi_j^{(1)}$ and $\psi_j^{(2)}$, for integer $1 \le j \le 2N_s$, such that*

$$\hat{X}_{j/2} = \psi_j^{(1)}(Y_1, \Phi_j);$$
$$\hat{V}_{j/2} = \psi_j^{(2)}(Y_1, \Phi_j).$$

As a consequence, the choices of φ_j and ϕ are straightforward:

$$\varphi_j(Y_1, \Phi_j) := (\psi_j^{(1)}(Y_1, \Phi_j), t_{j+1}), \quad \text{for } 1 \le j \le 2N_s - 1 \equiv N - 1;$$

$$\phi(Y_1, \Phi_{2N_s}) := \psi_{2N_s}^{(1)}(Y_1, \Phi_{2N_s}) \equiv \hat{X}_{N_s}.$$

Proof. We prove it by mathematical induction. Suppose the result is true for any integer $1 \le j \le 2k$, where k is some integer $(1 \le k \le N_s - 1)$, then we show the forms of $\psi_j^{(1)}$ and $\psi_j^{(2)}$ for $j = 2k+1$ and $2k+2$. That is, both $\hat{X}_{k+1/2}$ and $\hat{V}_{k+1/2}$ are functions of $Y_1 \equiv x^\star(U)$ and Φ_{2k+1}; both $\hat{X}_{k+1}$ and $\hat{V}_{k+1}$ are functions of Y_1 and Φ_{2k+2}.

- **(Case $j = 2k+1$).** By definition, $\hat{V}_{k+1/2} \equiv \hat{V}_k = \psi_{2k}^{(2)}(Y_1, \Phi_{2k})$. As Φ_{2k+1} contains more information than Φ_{2k}, obviously, $\hat{V}_{k+1/2}$ is a deterministic function of Y_1 and Φ_{2k+1}. Next, we consider $\hat{X}_{k+1/2}$. By (2.6), (1.3) and by recalling $t_j \equiv s_{j/2}$,

$$\hat{X}_{k+1/2} = \hat{X}_k + \frac{1 - e^{-2\eta_k h}}{2}\hat{V}_k + \frac{1}{\sqrt{L}}\left(W_{t_{2k+1}} - W_{t_{2k}} - e^{-2t_{2k+1}}\left(\widetilde{W}_{t_{2k+1}}^{(2)} - \widetilde{W}_{t_{2k}}^{(2)}\right)\right)$$
$$- \frac{1}{2L}\left(\int_0^{\eta_k h} 1 - e^{2(s - \eta_k h)}\, ds\right)\nabla U(\hat{X}_k).$$

  Note that $\hat{X}_k = \psi_{2k}^{(1)}(Y_1, \Phi_{2k})$, $\hat{V}_k = \psi_{2k}^{(2)}(Y_1, \Phi_{2k})$ by assumption, and also note that

$$\Phi_{2k+1} = \left(\Phi_{2k}, \Upsilon_{U,\omega}(Y_{2k+1}, t_{2k+1})\right),$$
$$\Upsilon_{U,\omega}(Y_{2k+1}, t_{2k+1}) = \left(\nabla U(Y_{2k+1}), \widetilde{W}_{t_{2k+1}}^{(0,2)}\right) \equiv \left(\nabla U(\hat{X}_k), \widetilde{W}_{t_{2k+1}}^{(0,2)}\right),$$

  we immediately know that

$$\hat{X}_{k+1/2} = \psi_{2k}^{(1)}(Y_1, \Phi_{2k}) + \frac{1 - e^{-2\eta_k h}}{2}\psi_{2k}^{(2)}(Y_1, \Phi_{2k}) + f_{2k+1}(\Phi_{2k+1}),$$

  for some linear function f_{2k+1} on Φ_{2k+1}; the expression of f_{2k+1} should be clear from above. Then, we can conclude that $\hat{X}_{k+1/2} = \psi_{2k+1}^{(1)}(Y_1, \Phi_{2k+1})$ for some mapping $\psi_{2k+1}^{(1)}$, which is defined iteratively based on $\psi_{2k}^{(1)}$, $\psi_{2k}^{(2)}$, and f_{2k+1}.

- **(Case $j = 2k+2$).** Similarly, both $\hat{X}_{k+1}$ and $\hat{V}_{k+1}$ are functions of $\hat{X}_k$, $\hat{V}_k$, $\nabla U(\hat{X}_{k+1/2})$, $\widetilde{W}_{s_k}^{(0,2)}$ and $\widetilde{W}_{s_{k+1}}^{(0,2)}$, and the latter three terms are included in the information Φ_{2k+2}. By the same argument, we know that both $\hat{X}_{k+1}$ and $\hat{V}_{k+1}$ are functions of the initial condition $Y_1 \equiv x^\star(U)$ and Φ_{2k+2}.

- The base cases for $j = 1, 2$ can be similarly verified.
$\square$

Appendix B. Additional proofs.

Proof. **(Proof of Lemma 2.3.)** In the limit $C_v \to \infty$,

$$\lim_{C_v \to \infty} \mathsf{P}(C_x, C_v, u, T)$$
$$= \lim_{C_v \to \infty} \mathbb{P}\left(\omega: \sup_{0 \le t \le T} X_t(U_u, \omega) \ge 2C_x, \inf_{0 \le t \le T} X_t(U_u, \omega) \le -2C_x,\right.$$
$$\left.\sup_{0 \le t \le T}|V_t(U_u, \omega)| \le C_v/2\right)$$
$$= \mathbb{P}\left(\omega: \sup_{0 \le t \le T} X_t(U_u, \omega) \ge 2C_x, \inf_{0 \le t \le T} X_t(U_u, \omega) \le -2C_x\right).$$

Note that $\mathsf{P}(C_x, C_v, u, T)$ is a monotonically increasing function with respect to C_v. Therefore, for sufficiently large C_v, we must have

$$\mathsf{P}(C_x, C_v, u, T) \geq \frac{1}{2} \mathbb{P}\Big(\omega \colon \sup_{0 \leq t \leq T} X_t(U_u, \omega) \geq 2C_x, \ \inf_{0 \leq t \leq T} X_t(U_u, \omega) \leq -2C_x \Big).$$

For finite C_x, due to the fluctuation of Brownian motion, the probability that the trajectory $\{X_t(U_u, \omega)\}_{0 \leq t \leq T}$ crosses both levels $2C_x$ and $-2C_x$ is expected to be strictly positive. More rigorously,

$$\mathbb{P}\Big(\omega \colon \sup_{0 \leq t \leq T} X_t(U_u, \omega) \geq 2C_x, \ \inf_{0 \leq t \leq T} X_t(U_u, \omega) \leq -2C_x \Big)$$

$$\geq \mathbb{P}\Big(\omega \colon X_T(U_u, \omega) \geq 2C_x, \ X_{T/2}(U_u, \omega) \leq -2C_x \Big).$$

By (2.8), we know $\Big(X_T(U_u, \cdot), X_{T/2}(U_u, \cdot)\Big)$ are two correlated Gaussian random variables with mean zero. As long as we can show that $|\mathrm{corr}(X_T(U_u, \cdot), X_{T/2}(U_u, \cdot))| < 1$, then the above probability must be strictly positive. Let us denote

$$f_t(s) := \frac{1}{\sqrt{L-u}} \Big(e^{-(t-s)\lambda_-} - e^{-(t-s)\lambda_+} \Big).$$

Then

$$\mathbb{E}_{\omega \sim \mathbb{P}}[X_{T/2}(U_u, \omega)^2] = \int_0^{T/2} \big(f_{T/2}(s)\big)^2 \ \mathrm{d}s;$$

$$\mathbb{E}_{\omega \sim \mathbb{P}}[X_T(U_u, \omega)^2] = \int_0^T \big(f_T(s)\big)^2 \ \mathrm{d}s;$$

$$\mathbb{E}_{\omega \sim \mathbb{P}}[X_{T/2}(U_u, \omega) X_T(U_u, \omega)] = \int_0^{T/2} f_{T/2}(s) f_T(s) \ \mathrm{d}s.$$

By Cauchy-Schwarz inequality,

$$\Big| \mathbb{E}_{\omega \sim \mathbb{P}}\big[X_{T/2}(U_u, \omega) X_T(U_u, \omega)\big] \Big| = \Big| \int_0^{T/2} f_{T/2}(s) f_T(s) \ \mathrm{d}s \Big|$$

$$\leq \sqrt{\int_0^{T/2} \big(f_{T/2}(s)\big)^2 \ \mathrm{d}s \int_0^{T/2} \big(f_T(s)\big)^2 \ \mathrm{d}s}$$

$$< \sqrt{\int_0^{T/2} \big(f_{T/2}(s)\big)^2 \ \mathrm{d}s \int_0^T \big(f_T(s)\big)^2 \ \mathrm{d}s}$$

$$= \sqrt{\mathbb{E}_{\omega \sim \mathbb{P}}\big[X_{T/2}(U_u, \omega)^2\big] \ \mathbb{E}_{\omega \sim \mathbb{P}}\big[X_T(U_u, \omega)^2\big]}.$$

Therefore, $|\mathrm{corr}(X_{T/2}, X_T)| < 1$, and this completes the proof. $\square$

REFERENCES

[1] T. Bochacik, M. Goćwin, P.M. Morkisz, and P. Przybyłowicz, *Randomized Runge-Kutta method – Stability and convergence under inexact information*, J. Complex., 65:101554, 2020. 1.3, 2.2

[2] A. Brünger, C.L. Brooks III, and M. Karplus, *Stochastic boundary conditions for molecular dynamics simulations of ST2 water*, Chem. Phys. Lett., 105(5):495–500, 1984. 1.3

[3] Y. Cao, J. Lu, and L. Wang, *On explicit L^2-convergence rate estimate for underdamped Langevin dynamics*, arXiv preprint, arXiv:1908.04746, 2019. 1, 1.1

[4] X. Cheng, N.S. Chatterji, Y. Abbasi-Yadkori, P.L. Bartlett, and M.I. Jordan, *Sharp convergence rates for Langevin dynamics in the nonconvex setting*, arXiv preprint, arXiv:1805.01648, 2018. 1, 1.3

[5] X. Cheng, N.S. Chatterji, P.L. Bartlett, and M.I. Jordan, *Underdamped Langevin MCMC: A non-asymptotic analysis*, in S. Bubeck, V. Perchet, and P. Rigollet (eds.), Proceedings of the 31st Conference on Learning Theory, PMLR, 75:300–323, 2018. 1, 1.1, 1.3, 2.2

[6] J.M.C. Clark and R.J. Cameron, *The maximum rate of convergence of discrete approximations for stochastic differential equations*, in B. Grigelionis (ed.), Stochastic Differential Systems Filtering and Control, Lecture Notes in Control and Information Sciences, Springer, Berlin, Heidelberg, 162–171, 1980. ii, 1.3

[7] A.S. Dalalyan and L. Riou-Durand, *On sampling from a log-concave density using kinetic Langevin diffusions*, Bernoulli, 26(3):1956–1988, 2020. 1, 1.3, 2.2

[8] T. Daun, *On the randomized solution of initial value problems*, J. Complex., 27(3):300–311, 2011. 1.3, 2.2

[9] J. Dolbeault, C. Mouhot, and C. Schmeiser, *Hypocoercivity for linear kinetic equations conserving mass*, Trans. Amer. Math. Soc., 367(6):3807–3828, 2015. 1, 1.1

[10] A. Eberle, A. Guillin, and R. Zimmer, *Couplings and quantitative contraction rates for Langevin dynamics*, Ann. Probab., 47(4):1982–2010, 2019. 1.1

[11] D.L. Ermak and H. Buckholz, *Numerical integration of the Langevin equation: Monte Carlo simulation*, J. Comput. Phys., 35(2):169–182, 1980. 1.3, 2.2

[12] P. Gregor, *Lecture 17: Hypercube problems*, Lecture Notes, Hypercube Structures NTIN097, Charles University, 2017. 3.2

[13] N. Grønbech-Jensen and O. Farago, *A simple and effective Verlet-type algorithm for simulating Langevin dynamics*, Mol. Phys., 111(8):983–991, 2013. 1, 1.3

[14] M. Hefter, A. Herzwurm, and T. Müller-Gronbach, *Lower error bounds for strong approximation of scalar SDEs with non-Lipschitzian coefficients*, Ann. Appl. Probab., 29(1):178–216, 2019. 1.3

[15] S. Heinrich and B. Milla, *The randomized complexity of initial value problems*, J. Complex., 24(2):77–88, 2008. 1.3, 2.2

[16] D. Higham, *An algorithmic introduction to numerical simulation of stochastic differential equations*, SIAM Rev., 43(3):525–546, 2001. 1.3

[17] N. Hofmann and T. Müller-Gronbach, *On the global error of Itô-Taylor schemes for strong approximation of scalar stochastic differential equations*, J. Complex., 20(5):732–752, 2004. 1.3

[18] N. Hofmann, T. Müller-Gronbach, and K. Ritter, *Optimal approximation of stochastic differential equations by adaptive step-size control*, Math. Comput., 69(231):1017–1034, 2000. 1.3

[19] N. Hofmann, T. Müller-Gronbach, and K. Ritter, *Step size control for the uniform approximation of systems of stochastic differential equations with additive noise*, Ann. Appl. Probab., 10(2):616–633, 2000. 1.3

[20] N. Hofmann, T. Müller-Gronbach, and K. Ritter, *The optimal discretization of stochastic differential equations*, J. Complex., 17(1):117–153, 2001. 1.3

[21] S. Hopkins, *Construct a bijective map between subsets of binary sequence*, MathOverflow, 2020. 3.4

[22] A. Jentzen and A. Neuenkirch, *A random Euler scheme for Carathéodory differential equations*, J. Comput. Appl. Math., 224(1):346–359, 2009. 2.2

[23] B. Kacewicz, *On the optimal error of algorithms for solving a scalar autonomous ODE*, BIT Numer. Math., 22:503–518, 1982. 1.3

[24] B. Kacewicz, *Optimality of Euler-integral information for solving a scalar autonomous ODE*, BIT Numer. Math., 23:217–230, 1983. 1.3

[25] B. Kacewicz, *How to increase the order to get minimal-error algorithms for systems of ODE*, Numer. Math., 45:93–104, 1984. 1.3

[26] B. Kacewicz, *Optimal solution of ordinary differential equations*, J. Complex., 3(4):451–465, 1987. 1.3

[27] B. Kacewicz, *Randomized and quantum algorithms yield a speed-up for initial-value problems*, J. Complex., 20(6):821–834, 2004. 1.3

[28] B. Kacewicz, *Almost optimal solution of initial-value problems by randomized and quantum algorithms*, J. Complex., 22(5):676–690, 2006. 1.3

[29] B. Kacewicz and P. Przybyłowicz, *On the optimal robust solution of IVPs with noisy information*, Numer. Algor., 71:505–518, 2016. 1.3

[30] R. Kruse and Y. Wu, *A randomized Milstein method for stochastic differential equations with non-differentiable drift coefficients*, Discrete Contin. Dyn. Syst. Ser. B, 24(8):3475–3502, 2019. 2.2

[31] B. Leimkuhler and C. Matthews, *Rational construction of stochastic numerical methods for molecular sampling*, Appl. Math. Res., 2013(1):34–56, 2013. 1.3

[32] Y.-A. Ma, N. Chatterji, X. Cheng, N. Flammarion, P. Bartlett, and M.I. Jordan, *Is there an analog of Nesterov acceleration for MCMC?*, arXiv preprint, arXiv:1902.00996, 2019. 1, 1.3

[33] J.C. Mattingly, A.M. Stuart, and D.J. Higham, *Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise*, Stoch. Process. Appl., 101(2):185–232, 2002. 1.1

[34] B. Mishra and T. Schlick, *The notion of error in Langevin dynamics. I. Linear analysis*, J. Chem. Phys., 105(1):299–318, 1996. 1

[35] T. Müller-Gronbach, *The optimal uniform approximation of systems of stochastic differential equations*, Ann. Appl. Probab., 12(2):664–690, 2002. 1.3

[36] T. Müller-Gronbach, *Optimal pointwise approximation of SDEs based on Brownian motion at discrete points*, Ann. Appl. Probab., 14(4):1605–1642, 2004. 1.3

[37] T. Müller-Gronbach and K. Ritter, *Minimal errors for strong and weak approximation of stochastic differential equations*, in A. Keller, S. Heinrich, and H. Niederreiter (eds.), Monte Carlo and Quasi-Monte Carlo Methods 2006, Springer Berlin Heidelberg, 53–82, 2008. 1.3, 2.1

[38] H. Mori, *A continued-fraction representation of the time-correlation functions*, Prog. Theor. Phys., 34(3):399–416, 1965. 1

[39] P. Morkisz and P. Przybyłowicz, *Optimal pointwise approximation of SDEs from inexact information*, J. Comput. Appl. Math., 324:85–100, 2017. 1.3, 2.2

[40] P.M. Morkisz and P. Przybyłowicz, *Randomized derivative-free Milstein algorithm for efficient approximation of solutions of SDEs under noisy information*, J. Comput. Appl. Math., 383:113112, 2021. 1.3, 2.2

[41] E. Novak, *Deterministic and Stochastic Error Bounds in Numerical Analysis*, Lecture Notes in Mathematics, Springer-Verlag, Berlin Heidelberg, 1988. 1, 3

[42] R.W. Pastor, B.R. Brooks, and A. Szabo, *An analysis of the accuracy of Langevin and molecular dynamics algorithms*, Mol. Phys., 65(6):1409–1419, 1988. 1, 1.3

[43] G.A. Pavliotis, *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*, Texts in Appl. Math., Springer-Verlag New York, 2014. 1, 1, 1.1

[44] E. Platen, *An introduction to numerical methods for stochastic differential equations*, Acta Numer., 8:197–246, 1999. 1.3

[45] P. Przybyłowicz, *Adaptive Itô-Taylor algorithm can optimally approximate the Itô integrals of singular functions*, J. Comput. Appl. Math., 235(1):203–217, 2010. 1.3

[46] P. Przybyłowicz, *Minimal asymptotic error for one-point approximation of SDEs with time-irregular coefficients*, J. Comput. Appl. Math., 282:98–110, 2015. 1.3, 2.2

[47] P. Przybyłowicz, *Optimal global approximation of SDEs with time-irregular coefficients in asymptotic setting*, Appl. Math. Comput., 270:441–457, 2015. 1.3, 2.2

[48] P. Przybyłowicz and P. Morkisz, *Strong approximation of solutions of stochastic differential equations with time-irregular coefficients via randomized Euler algorithm*, Appl. Numer. Math., 78:80–94, 2014. 1.3, 2.2

[49] J. Roussel and G. Stoltz, *Spectral methods for Langevin dynamics and associated error estimates*, ESAIM Math. Model. Numer. Anal., 52(3):1051–1083, 2018. 1, 1.1

[50] R. Shen and Y.T. Lee, *The randomized midpoint method for log-concave sampling*, in H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), Advances in Neural Information Processing Systems 32, Curran Associates, Inc., 2019. 1, 1.1, 1.3, 1.3, 2, 2.2, 2.2, 2.3, 3, 3.4

[51] G. Stengle, *Numerical methods for systems with measurable coefficients*, Appl. Math. Lett., 3(4):25–29, 1990. 2.2

[52] G. Stengle, *Error analysis of a randomized numerical method*, Numer. Math., 70(1):119–128, 1995. 2.2

[53] J. Traub, *A brief history of information-based complexity*, in E. Novak, I. Sloan, J. Traub, and H. Woźniakowski (eds.), Essays on the Complexity of Continuous Problems, European Mathematical Society Publishing House, Zuerich, Switzerland, 2009. 1

[54] C. Villani, *Hypocoercivity*, Memoirs of the American Mathematical Society, 202(950), 2009. 1.1

[55] W. Wang and R.D. Skeel, *Analysis of a few numerical integration methods for the Langevin equation*, Mol. Phys., 101(14):2149–2156, 2003. 1

[56] Y. Zhu, *Boolean lattice and symmetric chain decompositions*, Combinatorics Seminar 2014, Shanghai, 2014. 3.2

[57] R. Zwanzig, *Nonlinear generalized Langevin equations*, J. Stat. Phys., 9:215–220, 1973. 1