ACCELERATING LANGEVIN MONTE CARLO VIA RANDOMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

As a fundamental algorithmic task, sampling from a high-dimensional probability distribution finds wide-ranging applications in scientific computing, computational statistics and machine learning. In particular, Langevin Monte Carlo (LMC) sampling algorithms are among the most widely used ones and have received increasing attention in recent years. By injecting additional randomness for every iteration, this paper aims to propose two novel higher-order LMC algorithms, including a Taylor scheme involved with the Hessian of the potential and a Hessian-free Runge–Kutta type variant. Under gradient Lipschitz and Hessian Lipschitz conditions, both algorithms are shown to enjoy non-asymptotic $\mathcal{W}_2$-error bounds of order $O(dh^{\frac{3}{2}})$ beyond log-concavity, which guarantees a mixing time complexity $\tilde{O}(d^{\frac{2}{3}}\epsilon^{-\frac{2}{3}})$. This improves the dimension-dependence of mixing time complexity $\tilde{O}(d\epsilon^{-\frac{2}{3}})$ for the existing high-order LMC algorithms. Moreover, the error analysis of the newly proposed algorithms do not require any Lipschitz condition on the 3rd-order derivative of the potential, which is, however, necessary in obtaining the desired convergence order 1.5 of existing high-order LMC methods. Some numerical experiments are reported to support the effectiveness of new sampling algorithms.

1 INTRODUCTION

Sampling from a high-dimensional target probability distribution in the form of

$$\pi(\mathrm{d}x) \propto e^{-U(x)}\mathrm{d}x,$$

is a fundamental problem in a variety of fields such as statistics, machine learning and computational science, where $U\colon \mathbb{R}^d \to \mathbb{R}$ for large $d \in \mathbb{N}$ is called to be a potential function. To efficiently sample from the target distribution, Langevin-based methods have emerged as a powerful approach. Among these, Langevin Monte Carlo (LMC) is the most widely used one, given by

$$\mathbb{Y}_{n+1} = \mathbb{Y}_n - \nabla U(\mathbb{Y}_n)h + \sqrt{2h}\zeta_{n+1}, \quad \mathbb{Y}_0 = \xi, \tag{1}$$

where h is the uniform stepsize and $\zeta_n := (\zeta_n^1, \zeta_n^2, \cdots, \zeta_n^d)^T$, $n \in \mathbb{N}$, are i.i.d. standard d-dimensional Gaussian random variables. It is well-known that LMC can be viewed as the Euler–Maruyama method for the following overdamped Langevin dynamics (OLD):

$$\mathrm{d}X(t) = -\nabla U(X(t))\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W(t), \quad X(0) = X_0 = \xi,\ t > 0, \tag{2}$$

where $W := \left(W^1, W^2, \cdots, W^d\right)^T : [0,\infty) \times \Omega \to \mathbb{R}^d$ is d-dimensional Brownian motion. Under mild conditions on the potential U, the Langevin dynamics (2) admits π as its unique invariant distribution (see, e.g., (Pavliotis, 2014)). One can thus expect the samples generated by LMC (1) after long-time iterations, to approximately follow the target distribution π.

In recent years, considerable progress has been made in investigating non-asymptotic error bounds of the classical sampling algorithm (1). Early works (Durmus & Moulines, 2017; Dalalyan, 2017b;a; Durmus et al., 2019; Cheng & Bartlett, 2018) obtained the non-asymptotic convergence of order $O(\sqrt{dh})$ in total variation or $\mathcal{W}_2$-distance for LMC (1) under a strongly log-concave condition ($m > 0$):

$$\langle x - y, \nabla U(x) - \nabla U(y)\rangle \geq m|x-y|^2, \ \ \forall x, y \in \mathbb{R}^d \tag{3}$$

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

and the gradient Lipschitz condition

$$\big|\nabla U(x) - \nabla U(y)\big| \le L_1|x-y|, \quad \forall\, x, y \in \mathbb{R}^d.$$

By imposing additional smoothness assumptions on U, one can derive improved convergence rates of order $O(dh)$ (Durmus & Moulines, 2019) under the Hessian Lipschitz condition

$$\|\nabla^2 U(x) - \nabla^2 U(y)\| \le L_2|x-y|, \quad \forall\, x, y \in \mathbb{R}^d,$$

and order $O(\sqrt{d}h)$ under further smoothness condition (Li et al., 2022; Altschuler & Chewi, 2026).

Beyond log-concavity, much progress has been made recently in analyzing non-asymptotic error bounds of the classical LMC. For example, the non-asymptotic error bounds in various metrics of order $O(\sqrt{dh})$ under the gradient Lipschitz condition and order $O(dh)$ or $O(\sqrt{d}h)$ under smoother conditions were established for LMC, in the setting of the convexity at infinity and the log-Sobolev inequality (Cheng et al., 2018; Li & Wang, 2025; Majka et al., 2020; Mou et al., 2022; Altschuler & Chewi, 2026; Yang & Wang, 2025; Li et al., 2025a; Pang et al., 2025; Chewi et al., 2024; Li et al., 2025b; Pagès & Panloup, 2023; Mousavi-Hosseini et al., 2023; Altschuler & Chewi, 2024).

As shown above, the gradient Lipschitz condition suffices to promise the convergence order $O(\sqrt{dh})$, while achieving the order-one convergence $O(h)$ requires smoother condition like the Hessian Lipschitz condition. This limitation can be circumvented by randomized methods (Altschuler & Chewi, 2026; Wang & Yang, 2026; Yu et al., 2024), where the convergence rate of order $O(\sqrt{d}h)$ was obtained under no additional smoothness conditions other than the gradient Lipschitz condition.

To further improve the convergence rate, a natural approach is to apply the Itô Taylor expansion of the solution to the Langevin SDE and truncate the expansion at strong order 1.5 (see, e.g., (Kloeden & Platen, 1992; Milstein & Tretyakov, 2004)), yielding the strong Taylor scheme (TELMC)

$$\begin{aligned}\hat{\mathbb{Y}}_{n+1} =& \hat{\mathbb{Y}}_n - \nabla U(\hat{\mathbb{Y}}_n)h + \sqrt{2}\Delta W_{n+1} + \tfrac{1}{2}\nabla^2 U(\hat{\mathbb{Y}}_n)\nabla U(\hat{\mathbb{Y}}_n)h^2 \\ & - \tfrac{1}{2}\nabla(\Delta U(\hat{\mathbb{Y}}_n))h^2 - \sqrt{2}\nabla^2 U(\hat{\mathbb{Y}}_n)\Delta\Xi_{n+1},\end{aligned} \tag{4}$$

where the increments ΔW_{n+1} and $\Delta\Xi_{n+1}$ are both Gaussian, defined by

$$\Delta W_{n+1} := W(t_{n+1}) - W(t_n), \quad \Delta\Xi_{n+1} := \int_{t_n}^{t_{n+1}}\int_{t_n}^{t} \mathrm{d}W(s)\,\mathrm{d}t. \tag{5}$$

Such a Taylor scheme for (2) was proposed and analyzed in (Sabanis & Zhang, 2019), with a non-asymptotic $\mathcal{W}_2$-error bound of order $O(d^2h^{3/2})$ obtained under the strong convexity assumption and smoothness conditions on derivatives of U up to order 3. More recently, (Neufeld & Zhang, 2024) established a non-asymptotic error estimate for the Taylor scheme in $\mathcal{W}_1$-distance of order $O(e^{O(d)}h^{3/2})$ in a non-convex setting.

Evidently, implementing the Taylor scheme (4) is computationally expensive as the scheme involves several higher-order derivatives of U, which are difficult and expensive to compute in multidimensional settings. An interesting question thus arises:

(Q1). *Can one construct efficient higher-order LMC algorithms avoiding or substantially reducing the use of high-order derivatives?*

This question motivates the development of stochastic Runge–Kutta methods with high-order accuracy, while avoiding the use of higher-order derivatives (Kloeden & Platen, 1992; Milstein & Tretyakov, 2004). For the dynamics (2), various Hessian-free Runge–Kutta methods were developed in (Li et al., 2019) and (Yang & Wang, 2026), where the state-of-the-art order $O(d^{\frac{3}{2}}h^{\frac{3}{2}})$ of $\mathcal{W}_2$-convergence was derived under and beyond log-concavity, respectively. In additional to gradient Lipschitz and Hessian Lipschitz conditions for order 1 methods discussed above, a Lipschitz condition on the 3rd-order derivative of U (Yang & Wang, 2026, Assumption 2.4) is necessary to obtain order 1.5. Given the significant role of randomization in sharpening error bounds for the Euler scheme as mentioned above, it is natural to ask the following question:

(Q2). *Can one leverage randomization to develop efficient higher-order LMC algorithms with improved error bounds under minimal smoothness assumptions?*

Following this question, we develop in this paper a so-called randomized Taylor method of strong order 1.5, given by

$$\begin{cases} \bar{Y}_n^\tau = Y_n - \tau_n h \nabla U(Y_n) + \sqrt{2}\big(W(t_n^\tau) - W(t_n)\big), \\ Y_{n+1} = Y_n - h\nabla U(\bar{Y}_n^\tau) - \sqrt{2}\nabla^2 U(Y_n)\big[\Delta\Xi_{n+1} - h\big(W(t_n^\tau) - W(t_n)\big)\big] + \sqrt{2}\,\Delta W_{n+1}, \end{cases} \tag{6}$$

and its Runge–Kutta type variant without the Hessian:

$$\begin{cases} \bar{Z}_n^\tau = Z_n - \tau_n h \nabla U(Z_n) + \sqrt{2}\,\big(W(t_n^\tau) - W(t_n)\big), \\ Z_{n+1} = Z_n - h\nabla U(\bar{Z}_n^\tau) + \sqrt{2}\,\Delta W_{n+1} \\ \qquad\qquad - \sqrt{2}\,\nabla U\big(Z_n + \Delta\Xi_{n+1} - h\big(W(t_n^\tau) - W(t_n)\big)\big) + \sqrt{2}\,\nabla U(Z_n), \end{cases} \tag{7}$$

where ΔW_{n+1}, $\Delta\Xi_{n+1}$ are defined by (5) and $t_n^\tau := t_n + \tau_n h$ with $\{\tau_n\}_{n\geq 0}$ being a sequence of i.i.d. random variables uniformly distributed on $[0,1]$. Compared with the usual Taylor scheme (4), the randomized scheme (6) is structurally simpler, where the use of high-order derivatives is substantially reduced. The price to pay for the simplification is to generate one more Gaussian and a $\mathcal{U}(0,1)$-distributed random variable per iteration, which is computationally cheap. As shown by Theorem 3.9, just under gradient Lipschitz and Hessian Lipschitz conditions, the non-asymptotic $\mathcal{W}_2$-error bound $O(dh^{\frac{3}{2}})$ can be established for these two novel schemes (6) and (7) beyond log-concavity, implying a mixing time complexity $\tilde{O}(d^{\frac{2}{3}}\epsilon^{-\frac{2}{3}})$. This considerably improves the state-of-the-art mixing-time $\tilde{O}(d\epsilon^{-\frac{2}{3}})$ for existing order 1.5 LMC algorithms, as mentioned above.

Contributions. To conclude, this paper answers questions (Q1) and (Q2) affirmatively and its contribution can be summarized as follows:

- We exploit randomization in a novel way to develop two novel higher-order LMC algorithms, including a Taylor scheme and a Hessian-free Runge–Kutta type variant.
- Under minimal smoothness conditions, namely, gradient Lipschitz and Hessian Lipschitz conditions, both algorithms are proved to enjoy non-asymptotic $\mathcal{W}_2$-error bounds $O(dh^{\frac{3}{2}})$ beyond log-concavity, which considerably improve the state-of-the-art error bound $O(d^{\frac{3}{2}}h^{\frac{3}{2}})$ obtained based on additional Lipschitz smoothness of the 3rd derivative.

The outline of this paper is as follows. The next section introduces the notation used throughout the paper and outlines the ideas underlying the construction of the proposed schemes. Section 3 presents the main results for the proposed algorithms. Proofs of main results are put in Appendix sections. Numerical experiments are provided in Section 4.

2 Accelerated Langevin Monte Carlo via Randomization

2.1 Notation.

Throughout this paper, we use $\mathbb{N}$ to denote the set of all positive integers. For any $n \in \mathbb{N}$, we denote $[n] := \{1, 2, \ldots, n\}$ and $[n]_0 := \{0, 1, \ldots, n\}$. Also, we use $\widetilde{O}(\cdot)$ to stand for $O(\cdot)\log^{O(1)}(\cdot)$. Let $\langle\cdot,\cdot\rangle$ denote the inner products of vectors in $\mathbb{R}^d$. By $|\cdot|$ we denote the Euclidean norm of vectors in $\mathbb{R}^{d_1}$ and by $\|\cdot\|$ the induced operator norm of matrices in $\mathbb{R}^{d_1\times d_2}$ with $d_1, d_2 \in \mathbb{N}$. Let $\mathbb{R}^{d_1\times\cdots\times d_n}$ be the space of n-th order tensors of dimensions $d_1, \ldots, d_n \in \mathbb{N}$. For any tensor $\mathcal{A} = (a_{i_1\cdots i_n}) \in \mathbb{R}^{d_1\times\cdots\times d_n}$ and vectors $x_k = (x_{k1}, \ldots, x_{kd_k}) \in \mathbb{R}^{d_k}$ for $1 \leq k \leq n$, define the associated multilinear form by

$$\mathcal{A}(x_1, \ldots, x_n) := \sum_{i_1=1}^{d_1} \cdots \sum_{i_n=1}^{d_n} a_{i_1\cdots i_n}\, x_{1i_1} \cdots x_{ni_n}.$$

The spectral norm of $\mathcal{A}$ is defined as

$$\|\mathcal{A}\| := \sup_{|x_k|=1,\ k\in[n]} |\mathcal{A}(x_1, \ldots, x_n)|. \tag{8}$$

For $f \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$, we denote the gradient and the Hessian of f by ∇f and $\nabla^2 f$, respectively. The Laplacian of f is denoted by Δf. Let $\mathcal{P}(\mathbb{R}^d)$ denote the space of probability distributions on $\mathbb{R}^d$. For any $\nu_1, \nu_2 \in \mathcal{P}(\mathbb{R}^d)$, their L^p-Wasserstein ($\mathcal{W}_p$ in short) distance is defined by

$$\mathcal{W}_p(\nu_1, \nu_2) := \inf_{\gamma \in \Gamma(\nu_1, \nu_2)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \mathrm{d}\gamma(x, y) \right)^{1/p},$$

where $\Gamma(\nu_1, \nu_2)$ denotes the set of probability distributions on $\mathbb{R}^d \times \mathbb{R}^d$ with marginal distributions ν_1 and ν_2. Moreover, we consider the product probability space

$$(\Omega, \mathcal{F}, \mathbb{P}) := \big(\Omega_W \times \Omega_\tau,\, \mathcal{F}^W \otimes \mathcal{F}^\tau,\, \mathbb{P}_W \otimes \mathbb{P}_\tau\big), \tag{9}$$

where $\big(\Omega_W, \mathcal{F}^W, \mathbb{P}_W\big)$ carries the Brownian motion in OLD (2) and $\big(\Omega_\tau, \mathcal{F}^\tau, \mathbb{P}_\tau\big)$ is another probability space carrying a sequence of i.i.d. random variables $\{\tau_n\}_{n\geq 0}$ uniformly distributed on $[0, 1]$, with $\mathcal{F}^\tau$ independent of $\mathcal{F}^W$. Let $\mathcal{L}(X)$ denote the law of the random variable X on the given probability space $(\Omega, \mathcal{F}, \mathbb{P})$. By $\mathbb{E}$, $\mathbb{E}_W$, and $\mathbb{E}_\tau$ we denote the expectations with respect to $\mathbb{P}$, $\mathbb{P}_W$, and $\mathbb{P}_\tau$, respectively. As a result,

$$\mathbb{E}[X] = \mathbb{E}_W \mathbb{E}_\tau[X] = \mathbb{E}_\tau \mathbb{E}_W[X].$$

2.2 Novel higher-order randomized methods: idea of method construction

Recall the overdamped Langevin dynamics (OLD):

$$\mathrm{d}X(t) = -\nabla U(X(t))\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W(t), \qquad X(0) = X_0, \quad t > 0. \tag{10}$$

Let us look at the one-step evolution of the dynamics

$$X(t_{n+1}) = X(t_n) + \int_{t_n}^{t_{n+1}} -\nabla U(X(s))\,\mathrm{d}s + \sqrt{2}\big[W(t_{n+1}) - W(t_n)\big]. \tag{11}$$

Taking the approximation

$$\int_{t_n}^{t_{n+1}} -\nabla U(X(s))\,\mathrm{d}s \approx \int_{t_n}^{t_{n+1}} -\nabla U(X(t_n))\,\mathrm{d}s = -\nabla U(X(t_n))\,h \tag{12}$$

gives the classical LMC algorithm given by

$$\mathbb{Y}_{n+1} = \mathbb{Y}_n - \nabla U(\mathbb{Y}_n)h + \sqrt{2h}\zeta_{n+1}, \qquad \mathbb{Y}_0 = X_0, \tag{13}$$

where $\{\zeta_n\}_{n\geq 1}$ are i.i.d. standard d-dimensional Gaussian random variables. For simplicity, in what follows we denote

$$f(x) := -\nabla U(x). \tag{14}$$

One approach to improve the approximation error produced by (12) is to use the following improved approximation:

$$\int_{t_n}^{t_{n+1}} f(X(s))\,\mathrm{d}s \approx \int_{t_n}^{t_{n+1}} \Big[f(X(t_n)) + \sqrt{2} f'(X(t_n))\big(W(s) - W(t_n)\big)\Big]\mathrm{d}s. \tag{15}$$

This way, the mean-square discrepancy of order $O(\sqrt{h})$ for (12) is improved to order $O(h)$ for (15). Next, we aim to further improve the discrepancy for (15):

$$R_{n+1} := \int_{t_n}^{t_{n+1}} f(X(s))\,\mathrm{d}s - \int_{t_n}^{t_{n+1}} \Big[f(X(t_n)) + \sqrt{2} f'(X(t_n))\big(W(s) - W(t_n)\big)\Big]\mathrm{d}s.$$

Given $\tau \sim \mathcal{U}(0, 1)$, we first note that

$$\begin{aligned}
R_{n+1} &= h \int_0^1 f\big(X(t_n + \tau h)\big)\,d\tau - h \int_0^1 \Big[f(X(t_n)) + \sqrt{2} f'(X(t_n))\big(W(t_n + \tau h) - W(t_n)\big)\Big] d\tau \\
&= h\,\mathbb{E}_\tau\big[f\big(X(t_n + \tau h)\big)\big] - h\,\mathbb{E}_\tau\Big[f(X(t_n)) + \sqrt{2} f'(X(t_n))\big(W(t_n + \tau h) - W(t_n)\big)\Big] \\
&= h\,\mathbb{E}_\tau\Big[f\big(X(t_n + \tau h)\big) - f(X(t_n)) - \sqrt{2} f'(X(t_n))\big(W(t_n + \tau h) - W(t_n)\big)\Big].
\end{aligned}$$

To improve the convergence rate of R_{n+1}, we subtract the term R^τ_{n+1} given by

$$R^\tau_{n+1} := hf\big(X(t_n+\tau h)\big) - hf(X(t_n)) - \sqrt{2}hf'(X(t_n))\big(W(t_n+\tau h) - W(t_n)\big),$$

and obtain a new local truncation error term

$$\begin{aligned}
\widehat{R}_{n+1} &:= R_{n+1} - R^\tau_{n+1} \\
&= \int_{t_n}^{t_{n+1}} f(X(s))\,\mathrm{d}s - hf(X(t_n)) - \sqrt{2}f'(X(t_n))\int_{t_n}^{t_{n+1}}\big(W(s)-W(t_n)\big)\,\mathrm{d}s \\
&\quad - \Big[hf\big(X(t_n+\tau h)\big) - hf(X(t_n)) - \sqrt{2}hf'(X(t_n))\big(W(t_n+\tau h)-W(t_n)\big)\Big] \\
&= \int_{t_n}^{t_{n+1}} f(X(s))\,\mathrm{d}s - \sqrt{2}f'(X(t_n))\int_{t_n}^{t_{n+1}}\big(W(s)-W(t_n+\tau h)\big)\,\mathrm{d}s \\
&\quad - hf\big(X(t_n+\tau h)\big). \qquad (16)
\end{aligned}$$

Clearly,

$$\mathbb{E}_\tau\big[R^\tau_{n+1}\big] = R_{n+1},$$

and thus

$$\mathbb{E}_\tau\big[\widehat{R}_{n+1}\big] = 0.$$

This property is crucial to build a martingale via summation of $\widehat{R}_{n+1}$, which can result in an additional order $\frac{1}{2}$ in the mean-square sense. In order to design a time-stepping scheme whose local truncation error corresponds to (16), we still need to approximate the unknown value $X(t_n+\tau h)$. A naive approximation of $X(t_n+\tau h)$ is

$$X(t_n+\tau h) \approx X(t_n) + \tau h f(X(t_n)) + \sqrt{2}\big(W(t_n+\tau h) - W(t_n)\big).$$

• Randomized Taylor LMC (RT-LMC).

Taking the above knowledge into account, we propose the following Randomized Taylor LMC (RT-LMC) algorithm:

$$\begin{cases} \bar{Y}^\tau_n = Y_n + \tau_n h f(Y_n) + \sqrt{2}\big(W(t^\tau_n) - W(t_n)\big), \\ Y_{n+1} = Y_n + hf(\bar{Y}^\tau_n) + \sqrt{2}f'(Y_n)\displaystyle\int_{t_n}^{t_{n+1}}\big[W(s) - W(t^\tau_n)\big]\mathrm{d}s + \sqrt{2}\,\Delta W_{n+1},\ Y_0 = X_0, \end{cases} \qquad (17)$$

where we denote

$$t^\tau_n := t_n + \tau_n h, \qquad \Delta W_{n+1} := W(t_{n+1}) - W(t_n).$$

In the notation of $\Delta\Xi_n$ defined by (5) and noting

$$\int_{t_n}^{t_{n+1}}\big[W(s) - W(t^\tau_n)\big]\mathrm{d}s = \Delta\Xi_{n+1} - h\big(W(t^\tau_n) - W(t_n)\big),$$

one can equivalently rewrite (17) as

$$\begin{cases} \bar{Y}^\tau_n = Y_n - \tau_n h\nabla U(Y_n) + \sqrt{2}\big(W(t^\tau_n) - W(t_n)\big), \\ Y_{n+1} = Y_n - h\nabla U(\bar{Y}^\tau_n) - \sqrt{2}\nabla^2 U(Y_n)\big[\Delta\Xi_{n+1} - h\big(W(t^\tau_n) - W(t_n)\big)\big] + \sqrt{2}\,\Delta W_{n+1}. \end{cases} \qquad (18)$$

Such a scheme can be regarded as a predictor-corrector method, where the three random variables involved are all Gaussian and their samples can be easily generated.

• Randomized Runge–Kutta LMC (RRK-LMC).

For the previous Taylor scheme, one needs to compute the Hessian of the potential, which might be very expensive in high-dimensions. To remedy it, we replace the derivative $f'(Y_n)$ by a difference and propose the following Runge–Kutta variant:

$$\begin{cases} \bar{Z}^\tau_n = Z_n + \tau_n h f(Z_n) + \sqrt{2}\,\big(W(t^\tau_n) - W(t_n)\big), \\ Z_{n+1} = Z_n + hf(\bar{Z}^\tau_n) + \sqrt{2}\,\Delta W_{n+1} \\ \qquad\qquad + \sqrt{2}\,f\left(Z_n + \displaystyle\int_{t_n}^{t_{n+1}}\big[W(s) - W(t^\tau_n)\big]\,\mathrm{d}s\right) - \sqrt{2}\,f(Z_n), \quad Z_0 = X_0, \end{cases} \qquad (19)$$

which can be also equivalently rewritten as

$$\begin{cases} \bar{Z}_n^\tau = Z_n + \tau_n h \nabla U(Z_n) + \sqrt{2}\,\big(W(t_n^\tau) - W(t_n)\big), \\ Z_{n+1} = Z_n + h\nabla U(\bar{Z}_n^\tau) + \sqrt{2}\,\Delta W_{n+1} \\ \qquad\quad + \sqrt{2}\,\nabla U\big(Z_n + \Delta\Xi_{n+1} - h\big(W(t_n^\tau) - W(t_n)\big)\big) - \sqrt{2}\,\nabla U(Z_n), \quad Z_0 = X_0. \end{cases} \tag{20}$$

3 Main results

We first make two smoothness assumptions that will be essentially used to establish the finite-time error bounds.

Assumption 3.1 (Gradient Lipschitz condition)**.** Assume that the potential $U \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$ and there exists a constant $L_1 > 0$ such that

$$\big|\nabla U(x) - \nabla U(y)\big| \le L_1 |x - y|, \quad \forall\, x, y \in \mathbb{R}^d. \tag{21}$$

Consequently, assuming $|\nabla U(0)| \le L_1' \sqrt{d}$ for some constant $L_1' > 0$ independent of d, we obtain the linear growth bound

$$|\nabla U(x)| \le |\nabla U(0)| + L_1 |x| \le L_1' \sqrt{d} + L_1 |x|. \tag{22}$$

Assumption 3.2 (Hessian Lipschitz condition)**.** Assume the potential $U \in \mathcal{C}^3(\mathbb{R}^d, \mathbb{R})$ and there exists a constant $L_2 > 0$ such that

$$\|\nabla^2 U(x) - \nabla^2 U(y)\| \le L_2 |x - y|, \quad \forall\, x, y \in \mathbb{R}^d. \tag{23}$$

Under these two assumptions, we obtain the following proposition on finite-time error bounds, whose proof can be found in Appendix A.

Proposition 3.3 (Finite-time error bounds)**.** *Let Assumptions 3.1 and 3.2 hold. Then for fixed $T = Nh$, $N \in \mathbb{N}$, the numerical approximations (18) and (20) obey*

$$\begin{aligned} \sup_{0 \le n \le N} \big\|X(t_n) - Y_n\big\|_{L^2(\Omega;\mathbb{R}^d)} &\le C(T)\big(d + \|X_0\|^2_{L^4(\Omega;\mathbb{R}^d)}\big) h^{\frac{3}{2}}, \\ \sup_{0 \le n \le N} \big\|X(t_n) - Z_n\big\|_{L^2(\Omega;\mathbb{R}^d)} &\le C(T)\big(d + \|X_0\|^2_{L^4(\Omega;\mathbb{R}^d)}\big) h^{\frac{3}{2}}, \end{aligned} \tag{24}$$

where $C(T)$ is nondecreasing in T and depends on L_1 and L_2.

Next, we impose a dissipativity condition to establish uniform-in-time moment bounds of numerical solutions.

Assumption 3.4 (Dissipativity condition)**.** There exist two constants μ, $\mu' > 0$ such that

$$\langle x, \nabla U(x) \rangle \ge \mu |x|^2 - \mu' d, \quad \forall\, x \in \mathbb{R}^d. \tag{25}$$

Proposition 3.5 (Uniform-in-time moment bounds of OLD)**.** *Let Assumption 3.4 hold. Then there exists a constant $\widehat{C}$, independent of d and h, such that*

$$\big\|X(t)\big\|_{L^4(\Omega;\mathbb{R}^d)} \le \widehat{C}\sqrt{d} + \big\|X_0\big\|_{L^4(\Omega;\mathbb{R}^d)}, \quad \forall\, t \ge 0. \tag{26}$$

Related estimates can be found in Lemma 2.4 of Yang & Wang (2025) and Lemma 3.1 of Pang et al. (2025). The following result establishes uniform-in-time moment bounds for RT-LMC and RRK-LM. Its proof is deferred to Appendix B.

Proposition 3.6 (Uniform-in-time moment bounds of RT-LMC and RRK-LMC)**.** *Let Assumptions 3.1 and 3.4 hold. Assume that the uniform timestep h satisfies $0 < h \le 1 \wedge \frac{1}{L_1} \wedge \frac{1}{4\mu} \wedge \frac{\mu}{64 L_1^2}$, then there exists a constant $\bar{C} > 0$, depending only on μ, μ', L_1, L_1', such that*

$$\begin{aligned} \big\|Y_n\big\|_{L^4(\Omega;\mathbb{R}^d)} &\le \bar{C}\sqrt{d} + \big\|X_0\big\|_{L^4(\Omega;\mathbb{R}^d)} \qquad \forall\, n \ge 0, \\ \big\|Z_n\big\|_{L^4(\Omega;\mathbb{R}^d)} &\le \bar{C}\sqrt{d} + \big\|X_0\big\|_{L^4(\Omega;\mathbb{R}^d)}, \qquad \forall\, n \ge 0. \end{aligned} \tag{27}$$

Let $\{p_t\}_{t\geq 0}$ be the Markov semigroup associated to the solution $\{X_t\}_{t\geq 0}$ of SDE equation 2. Then for any $\nu \in \mathcal{P}(\mathbb{R}^d)$, νp_t denotes the distribution law of X_t starting from $X_0 = x_0 \sim \nu$, i.e., $\nu p_t := \mathcal{L}(X_t)$, $t \geq 0$.

Assumption 3.7 (Exponential ergodicity in $\mathcal{W}_2$-distance)**.** There exist two constants $\mathcal{K}$, $\eta > 0$, independent of d, t, such that

$$\mathcal{W}_2(\nu p_t, \pi) \leq \mathcal{K} e^{-\eta t} \mathcal{W}_2(\nu, \pi), \tag{28}$$

for any $t \geq 0$ and any initial distribution $\nu := \mathcal{L}(X_0)$.

To obtain the exponential ergodicity in $\mathcal{W}_2$-distance of $\{p_t\}_{t\geq 0}$, one can rely on the following LSI condition (see Proposition 3.8).

Assumption 3.7' (Log-Sobolev inequality) Let the target distribution $\pi(\mathrm{d}x) \propto e^{-U(x)}\mathrm{d}x$ satisfy the log-Sobolev inequality with constant ρ:

$$\pi(\phi^2 \log \phi^2) \leq \rho \pi(|\nabla \phi|^2), \quad \forall \phi \in C_b^1(\mathbb{R}^d), \quad \pi(\phi^2) = 1, \tag{29}$$

where the constant ρ does not depend on d.

Proposition 3.8 (Exponential ergodicity in $\mathcal{W}_2$-distance)**.** *Let Assumptions 3.1 and 3.7' hold. Then the exponential ergodicity equation 28 holds true, i.e., there exist constants $\mathcal{K}$, $\eta > 0$, independent of d, t, such that*

$$\mathcal{W}_2(\nu p_t, \pi) \leq \mathcal{K} e^{-\eta t} \mathcal{W}_2(\nu, \pi), \tag{30}$$

for any $t \geq 0$ and any initial distribution $\nu := \mathcal{L}(X_0)$.

This proposition can be found in Theorem 2.1 (2) and 2.6 (2) of Wang (2020) and Proposition 2.5 of Yang & Wang (2025). Armed with finite-time error bounds, uniform-in-time moment bounds and the exponential ergodicity established above, one can straightforwardly apply the framework of Yang & Wang (2025); Wang & Yang (2026); Yang & Wang (2026) to obtain the following non-asymptotic error bounds for RT-LMC and RRK-LMC. The proof is straightforward and thus omitted.

Theorem 3.9 (Main results for RT-LMC and RRK-LMC)**.** *Let Assumptions 3.1, 3.2, 3.4, 3.7 be fulfilled and let $\{q_n\}_{n\geq 0}$ and $\{\bar{q}_n\}_{n\geq 0}$ denote the Markov semigroups associated with the RT-LMC (6) and the RRK-LMC (7), respectively. Assume that there exists a dimension-independent constant $\sigma > 0$ such that the initial value obeys $\|X_0\|_{L^4(\Omega;\mathbb{R}^d)} \leq \sigma\sqrt{d}$. If the uniform stepsize h satisfies condition in Proposition 3.6, then there exist dimension-independent constants $C_i, i = 1, 2, 3, 4$, such that, for any $n \in \mathbb{N}$ and any initial distribution $\nu := \mathcal{L}(X_0)$*

$$\begin{aligned} \mathcal{W}_2(\nu q_n, \pi) \leq & C_1 d h^{\frac{3}{2}} + C_2 d^{\frac{1}{2}} e^{-\lambda n h}, \\ \mathcal{W}_2(\nu \bar{q}_n, \pi) \leq & C_3 d h^{\frac{3}{2}} + C_4 d^{\frac{1}{2}} e^{-\lambda n h}, \end{aligned} \tag{31}$$

where λ is a dimension-independent constant, depending only on $\eta, \mathcal{K}, L_1$.

We mention that the above results remain true with Assumption 3.7 replaced by Assumption 3.7', in light of Proposition 3.8. As a direct consequence of this theorem, we also have the following result on the mixing time.

Proposition 3.10 (Mixing time for RT-LMC and RRK-LMC)**.** *Let all conditions in Theorem 3.9 be satisfied. Then, to achieve a prescribed accuracy level $\epsilon > 0$ in $\mathcal{W}_2$-distance, the number of iterations required for RT-LMC (6) and RRK-LMC (7) is of order $\widetilde{O}\big(d^{\frac{2}{3}}\epsilon^{-\frac{2}{3}}\big)$.*

4 Numerical experiments

In this section, we use K-component Gaussian mixture models (GMM) as target distributions to numerically validate the previous theoretical finding. Let us consider the target density taking the form

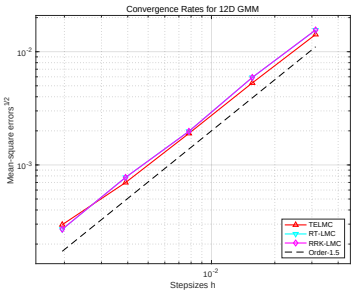
$$\pi(x) = \sum_{k=1}^{K} p_k \frac{1}{(2\pi s_k^2)^{d/2}} \exp\left(-\frac{|x-\mu_k|^2}{2s_k^2}\right),$$

where $p_k > 0$, $\sum_{k=1}^{K} p_k = 1$ and each component has mean $\mu_k \in \mathbb{R}^d$ and covariance matrix $s_k^2 I_d$ with $s_k > 0$. The associated potential is defined by $U(x) = -\log \pi(x)$.

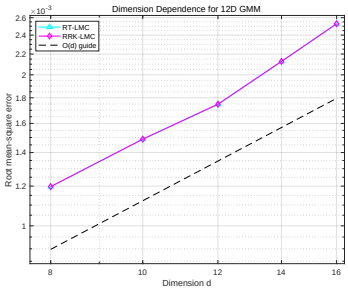
Strong convergence.

We compare TELMC, RT-LMC and RRK-LMC using a two-component Gaussian mixture in dimension $d = 12$, with $p_1 = 0.7$, $p_2 = 0.3$, $\mu_1 = \frac{1.8}{\sqrt{12}}\mathbf{1}_{12}$, $\mu_2 = -\mu_1$. Both components have covariance matrix I_{12} and their means satisfy $\|\mu_1\| = \|\mu_2\| = 1.8$.

All methods start from $X_0 = 0$ and are simulated until $T = 2$, using $M = 3000$ independent trajectories. A common reference solution is computed using TELMC with step size $h_{\mathrm{ref}} = 2^{-14}$. The five methods are evaluated at $h \in \{2^{-8}, 2^{-7}, 2^{-6}, 2^{-5}, 2^{-4}\}$. To couple the numerical solutions with the reference solution, each coarse Brownian increment is obtained by summing the corresponding fine-grid increments. The integrated Brownian increments are approximated by weighted sums of the same fine-grid increments. For the randomized methods, intermediate Brownian values are generated using a Brownian bridge within the fine-grid interval containing the random evaluation time. For each method, the terminal root mean-square error (RMSE) is estimated by $\widehat{e}(h) = \left[\frac{1}{M}\sum_{m=1}^{M}\left\|Y_T^{h,(m)} - X_T^{\mathrm{ref},(m)}\right\|^2\right]^{1/2}$ for large M. The errors are plotted on a log–log scale together with a reference line of slope 1.5. From Figure 1(a), we observe that both methods achieve a strong convergence order of 1.5.



(a) Convergence rates for 12D GMM.



(b) Dimension dependence for GMM.

Dimension dependence.

We compare RT-LMC with RRK-LMC for $d \in \{8, 10, 12, 14, 16\}$ and consider the target density

$$\pi_d(x) = \tfrac{1}{2}\mathcal{N}(x; 0.375\mathbf{1}_d, I_d) + \tfrac{1}{2}\mathcal{N}(x; -0.3\mathbf{1}_d, I_d).$$

where $\mathcal{N}(x; \mu, \Sigma)$ denotes the Gaussian density with mean μ and covariance matrix Σ. All methods start from $X_0 = 0$ and are simulated until $T = 2$, using $M = 3000$ independent trajectories. The reference solution is computed using (4) with $h_{\mathrm{ref}} = 2^{-14}$, while all four methods use the fixed step size $h = 2^{-5}$. The dimension dependence is plotted on a log–log scale together with a guide line proportional to d. Figure 1(b) indicates both methods exhibit an error constant that scales as $O(d)$.

Density histograms.

We examine the first-coordinate distributions generated by RT-LMC and RRK-LMC for the two-dimensional target

$$\pi(x) = \tfrac{1}{2}\mathcal{N}(x; \mu_1, I_2) + \tfrac{1}{2}\mathcal{N}(x; \mu_2, I_2), \qquad \mu_1 = (\sqrt{2}, \sqrt{2}), \qquad \mu_2 = -\mu_1.$$

Both methods start from $X_0 = 0$. We simulate $M = 5000$ independent trajectories until $T = 5$, using $h = 2^{-14}$.

The first coordinates of the terminal samples are displayed as probability-density-normalized histograms with 30 common bins. These histograms are compared with the exact first-coordinate marginal of the target,

$$\pi_1(z) = \tfrac{1}{2}\varphi(z; \sqrt{2}, 1) + \tfrac{1}{2}\varphi(z; -\sqrt{2}, 1),$$

where $\varphi(\,\cdot\,; \mu, \sigma^2)$ denotes the univariate Gaussian density with mean μ and variance σ^2. Figure 1 shows that the empirical histograms produced by the proposed methods closely match the target one-dimensional marginal density.

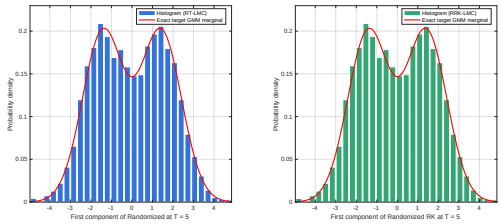


Figure 1: Histogram of the first component for two-mode GMM.

Scatter plots.

We compare five methods (TELMC, RKLMC-2G in Yang & Wang (2026), SRK-LD in Li et al. (2019), RT-LMC and RRK-LMC) using a two-dimensional Gaussian mixture with ten equally weighted components. Each component has covariance matrix $0.7I_2$, and its mean is given by

$$\mu_k = 10\left(\cos\frac{2\pi(k-1)}{10}, \sin\frac{2\pi(k-1)}{10}\right), \qquad k = 1, \ldots, 10.$$

Thus, the ten component centers are equally spaced on a circle of radius 10.

We simulate $n = 256$ particles, initialized independently from $\mathcal{N}(0, I_2)$, until $T = 6$. All methods use $h = 0.02$, corresponding to 300 time steps and share the same initial particles. The terminal samples are shown in scatter plots with red contours of the target density. Figure 2 shows that RT-LMC and RRK-LMC, along with the other methods of order 1.5, capture the overall ten-mode structure of the target distribution.

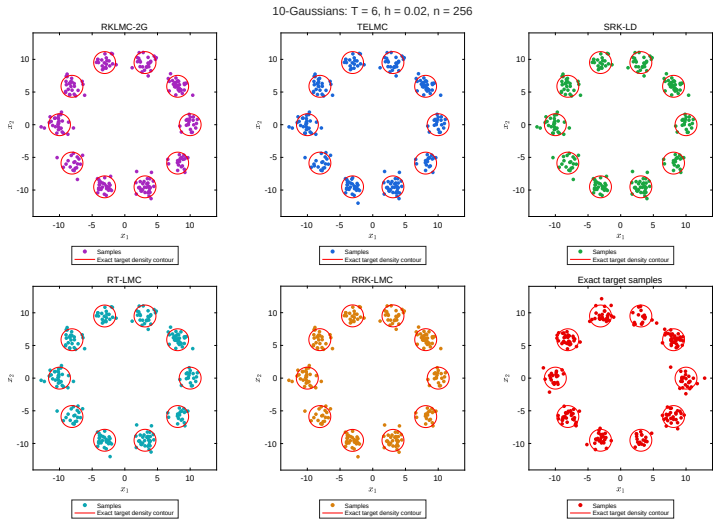


Figure 2: Scatter plots for ten-mode GMM.

AI USE STATEMENT

We did not use any generative AI tools at all.

ETHICS STATEMENT

Our paper submission does not raise any questions regarding the Code of Ethics.

REPRODUCIBILITY STATEMENT

In the section on numerical experiments, we present all settings of parameters. Further, all our theoretical results state the precise assumptions and contain full proofs.

ACKNOWLEDGMENTS

We would like to thank reviewers.

REFERENCES

Jason M Altschuler and Sinho Chewi. Shifted composition I: Harnack and reverse transport inequalities. *IEEE Transactions on Information Theory*, 2024.

Jason M Altschuler and Sinho Chewi. Shifted composition III: local error framework for KL divergence. *Foundations of Computational Mathematics*, pp. 1–61, 2026.

Xiang Cheng and Peter Bartlett. Convergence of Langevin MCMC in KL-divergence. In *Algorithmic Learning Theory*, pp. 186–211. PMLR, 2018.

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.

Sinho Chewi, Murat A Erdogdu, Mufan Li, Ruoqi Shen, and Matthew S Zhang. Analysis of Langevin Monte Carlo from Poincare to Log-Sobolev. *Foundations of Computational Mathematics*, pp. 1–51, 2024.

Arnak Dalalyan. Further and stronger analogy between sampling and optimization: Langevin Monte Carlo and gradient descent. In *Conference on Learning Theory*, pp. 678–689. PMLR, 2017a.

Arnak S Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(3):651–676, 2017b.

Alain Durmus and Éric Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *The Annals of Applied Probability*, 27(3):1551, 2017.

Alain Durmus and Éric Moulines. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli*, 25(4A):2854–2882, 2019.

Alain Durmus, Szymon Majewski, and Błażej Miasojedow. Analysis of Langevin Monte Carlo via convex optimization. *Journal of Machine Learning Research*, 20(73):1–46, 2019.

Peter E Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23. Springer Science & Business Media, Berlin, 1992.

Lei Li and Yuliang Wang. A sharp uniform-in-time error estimate for Stochastic Gradient Langevin Dynamics. *arXiv preprint arXiv:2207.09304, to appear in CSIAM Trans. Appl. Math.*, 2025.

Lei Li, Chen Wang, and Mengchao Wang. Ergodicity and error estimate of laws for a random splitting Langevin Monte Carlo. *arXiv preprint arXiv:2510.07676*, 2025a.

Ruilin Li, Hongyuan Zha, and Molei Tao. Sqrt (d) Dimension Dependence of Langevin Monte Carlo. In *The International Conference on Learning Representations*, 2022.

Xiang Li, Feng-Yu Wang, and Lihu Xu. Unadjusted Langevin Algorithms for SDEs with Hölder Drift. *SCIENCE CHINA Mathematics*, 68:3017–3042, 2025b.

Xuechen Li, Yi Wu, Lester Mackey, and Murat A Erdogdu. Stochastic Runge–Kutta accelerates Langevin Monte Carlo and beyond. *Advances in Neural Information Processing Systems*, 32, 2019.

Mateusz B Majka, Aleksandar Mijatović, and Lukasz Szpruch. Non-asymptotic bounds for sampling algorithms without log-concavity. *Annals of Applied Probability*, 30(4):1534–1581, 2020.

Grigori N Milstein and Michael V Tretyakov. *Stochastic Numerics for Mathematical Physics*, volume 39. Springer, Berlin, 2004.

Wenlong Mou, Nicolas Flammarion, Martin J. Wainwright, and Peter L. Bartlett. Improved bounds for discretization of Langevin diffusions: Near-optimal rates without convexity. *Bernoulli*, 28(3): 1577 – 1601, 2022.

Alireza Mousavi-Hosseini, Tyler K Farghly, Ye He, Krishna Balasubramanian, and Murat A Erdogdu. Towards a complete analysis of Langevin Monte Carlo: Beyond Poincaré inequality. In *The Thirty Sixth Annual Conference on Learning Theory*, pp. 1–35. PMLR, 2023.

Ariel Neufeld and Ying Zhang. Non-asymptotic estimates for accelerated high order Langevin Monte Carlo algorithms. *arXiv preprint arXiv:2405.05679*, 2024.

Gilles Pagès and Fabien Panloup. Unadjusted Langevin algorithm with multiplicative noise: Total variation and Wasserstein bounds. *The Annals of Applied Probability*, 33(1):726–779, 2023.

Chenxu Pang, Xiaojie Wang, and Yue Wu. Projected Langevin Monte Carlo algorithms in non-convex and super-linear setting. *Journal of Computational Physics*, 526:113754, 2025.

Grigorios A Pavliotis. *Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations*, volume 60. Springer, 2014.

Sotirios Sabanis and Ying Zhang. Higher order Langevin Monte Carlo algorithm. *Electronic Journal of Statistics*, 13:3805–3850, 2019.

Feng-Yu Wang. Exponential contraction in Wasserstein distances for diffusion semigroups with negative curvature. *Potential Anal.*, 53(3):1123–1144, 2020.

Xiaojie Wang and Bin Yang. When Langevin Monte Carlo meets randomization: New sampling algorithms with non-asymptotic error bounds beyond log-concavity and gradient Lipschitzness. *arXiv preprint arXiv:2509.25630*, 2026.

Bin Yang and Xiaojie Wang. Non-asymptotic error bounds in $\mathcal{W}_2$-distance with sqrt (d) dimension dependence and first order convergence for Langevin Monte Carlo beyond log-concavity. In *Proceedings of the 42nd International Conference on Machine Learning*, pp. 71358–71382, 2025.

Bin Yang and Xiaojie Wang. Accelerating Langevin Monte Carlo via Efficient Stochastic Runge–Kutta Methods beyond Log-Concavity. *International Conference on Machine Learning*, 2026.

Lu Yu, Avetik Karagulyan, and Arnak Dalalyan. Langevin Monte Carlo for strongly log-concave distributions: Randomized midpoint revisited. In *ICLR International Conference on Learning Representations*, 2024.

A PROOFS OF FINITE-TIME ERROR BOUNDS

Proof of Proposition 3.3. The proof is divided into two parts. Before proceeding with the proof, we note that C denotes a positive constant depending only on $\mu, \mu', L_1, L_1', L_2, T$, whose value may change from line to line.

• **Part I: Analysis of finite-time error bounds for RT-LMC**.

Next we analyze the error bounds of the proposed scheme. Denoting

$$e_n := X(t_n) - Y_n,$$

one can easily derive

$$\begin{aligned} e_{n+1} &= X(t_{n+1}) - Y_{n+1} \\ &= e_n + \int_{t_n}^{t_{n+1}} f(X(s))\,\mathrm{d}s - hf(\bar{Y}_n^\tau) - \sqrt{2}f'(Y_n)\int_{t_n}^{t_{n+1}} \big[W(s) - W(t_n^\tau)\big]\mathrm{d}s. \end{aligned}$$

Before proceeding further, we introduce an auxiliary process

$$\widehat{X}(t_n^\tau) := X(t_n) + \tau_n h f(X(t_n)) + \sqrt{2}\big(W(t_n^\tau) - W(t_n)\big).$$

Then

$$\begin{aligned} e_{n+1} = e_n &+ h\Big[f\big(\widehat{X}(t_n^\tau)\big) - f(\bar{Y}_n^\tau)\Big] \\ &+ \sqrt{2}\Big[f'(X(t_n)) - f'(Y_n)\Big]\int_{t_n}^{t_{n+1}} \big[W(s) - W(t_n^\tau)\big]\mathrm{d}s + \widehat{\widehat{R}}_{n+1}, \end{aligned}$$

where

$$\widehat{\widehat{R}}_{n+1} := \int_{t_n}^{t_{n+1}} f(X(s))\,\mathrm{d}s - hf\big(\widehat{X}(t_n^\tau)\big) - \sqrt{2}f'(X(t_n))\int_{t_n}^{t_{n+1}}\big[W(s) - W(t_n^\tau)\big]\mathrm{d}s.$$

By iteration, we derive

$$\begin{aligned} e_{n+1} = h\sum_{i=0}^{n}\Big[f\big(\widehat{X}(t_i^\tau)\big) - f(\bar{Y}_i^\tau)\Big] &\\ + \sqrt{2}\sum_{i=0}^{n}\Big[f'(X(t_i)) - f'(Y_i)\Big]\int_{t_i}^{t_{i+1}}\big[W(s) - W(t_i^\tau)\big]\mathrm{d}s &\\ + \sum_{i=0}^{n}\widehat{\widehat{R}}_{i+1}. & \end{aligned}$$

Note that

$$\widehat{X}(t_i^\tau) = X(t_i) + \tau_i h f(X(t_i)) + \sqrt{2}\big(W(t_i^\tau) - W(t_i)\big),$$

and thus

$$\widehat{X}(t_i^\tau) - \bar{Y}_i^\tau = e_i + \tau_i h\big[f(X(t_i)) - f(Y_i)\big].$$

The remainder term $\widehat{\widehat{R}}_{i+1}$ can be decomposed as follows:

$$\begin{aligned} \widehat{\widehat{R}}_{i+1}^{(1)} := \int_{t_i}^{t_{i+1}}\Big\{f(X(s)) - \sqrt{2}f'(X(t_i))\big[W(s) - W(t_i)\big]\Big\}\mathrm{d}s &\\ - \Big\{hf(X(t_i^\tau)) - \sqrt{2}hf'(X(t_i))\big[W(t_i^\tau) - W(t_i)\big]\Big\}, & \end{aligned}$$

$$\widehat{\widehat{R}}_{i+1}^{(2)} := hf(X(t_i^\tau)) - hf\big(\widehat{X}(t_i^\tau)\big),$$

so that

$$\widehat{\widehat{R}}_{i+1} = \widehat{\widehat{R}}_{i+1}^{(1)} + \widehat{\widehat{R}}_{i+1}^{(2)}.$$

Under the gradient Lipschitz condition (21), one can easily treat $\widehat{\widehat{R}}_{i+1}^{(2)}$ as follows:

$$\begin{aligned} \big|\widehat{\widehat{R}}_{i+1}^{(2)}\big| &\le hL_1\big|X(t_i^\tau) - \widehat{X}(t_i^\tau)\big| \\ &\le hL_1\left|\int_{t_i}^{t_i^\tau}\big[f(X(s)) - f(X(t_i))\big]\mathrm{d}s\right| \\ &\le hL_1\int_{t_i}^{t_{i+1}}\big|f(X(s)) - f(X(t_i))\big|\mathrm{d}s \\ &\le hL_1^2\int_{t_i}^{t_{i+1}}\big|X(s) - X(t_i)\big|\mathrm{d}s. \end{aligned}$$

Thanks to the Hölder continuity in time for the Langevin dynamics (see Lemma A.1 of Wang & Yang (2026)),

$$\sup_{t\ge 0}\big\|X(t+\theta) - X(t)\big\|_{L^p(\Omega;\mathbb{R}^d)} \le C_p\Big(\sqrt{d} + \|X_0\|_{L^p(\Omega;\mathbb{R}^d)}\Big)\theta^{1/2}, \qquad p \ge 1,$$

for $\theta \in (0, h]$, one can obtain

$$\big\|\widehat{\widehat{R}}_{i+1}^{(2)}\big\|_{L^2(\Omega;\mathbb{R}^d)} \le L_1^2 C_p\Big(\sqrt{d} + \|X_0\|_{L^2(\Omega;\mathbb{R}^d)}\Big)h^{5/2}.$$

Before coming to $\widehat{\widehat{R}}_{i+1}^{(1)}$, we introduce a notation

$$\phi(s) := f(X(s)) - \sqrt{2} f'(X(t_i))\big[W(s) - W(t_i)\big], \qquad s \in [t_i, t_{i+1}],$$

so that $\widehat{\widehat{R}}_{i+1}^{(1)}$ can be rewritten as

$$\widehat{\widehat{R}}_{i+1}^{(1)} = \int_{t_i}^{t_{i+1}} \phi(s)\, \mathrm{d}s - h\phi(t_i^\tau).$$

It is easy to see

$$\mathbb{E}_\tau\left[\widehat{\widehat{R}}_{i+1}^{(1)}\right] = 0,$$

and

$$\mathbb{E}_\tau\left[\left\langle \widehat{\widehat{R}}_{i+1}^{(1)}, \widehat{\widehat{R}}_{j+1}^{(1)} \right\rangle\right] = 0, \qquad for \qquad i \neq j,$$

where we used the fact that $\{\tau_n\}_{n\geq 0}$ is a sequence of i.i.d. random variables. As a consequence,

$$\mathbb{E}_\tau\left[\left|\sum_{i=0}^{n} \widehat{\widehat{R}}_{i+1}^{(1)}\right|^2\right] = \sum_{i=0}^{n} \mathbb{E}_\tau\left[\left|\widehat{\widehat{R}}_{i+1}^{(1)}\right|^2\right].$$

From the definition of ϕ, we can infer

$$\begin{aligned}
\phi(s) &= f(X(t_i)) + \int_0^1 f'\Big(X(t_i) + \theta\big(X(s) - X(t_i)\big)\Big)\big(X(s) - X(t_i)\big)\, d\theta \\
&\quad - \sqrt{2} f'(X(t_i))\big[W(s) - W(t_i)\big] \\
&= f(X(t_i)) + \int_0^1 f'\Big(X(t_i) + \theta\big(X(s) - X(t_i)\big)\Big)\, d\theta \\
&\quad \cdot \Big(\big[X(s) - \sqrt{2}\, W(s)\big] - \big[X(t_i) - \sqrt{2}\, W(t_i)\big]\Big) \\
&\quad + \sqrt{2} \int_0^1 \Big[f'\Big(X(t_i) + \theta\big(X(s) - X(t_i)\big)\Big) - f'(X(t_i))\Big]\, d\theta\, \big[W(s) - W(t_i)\big].
\end{aligned}$$

Under the gradient Lipschitz and the Hessian Lipschitz condition (23), one can show

$$
\begin{aligned}
&\big\|\phi(s)-\phi(t_i^\tau)\big\|_{L^2(\Omega;\mathbb{R}^d)}\\
\le&\bigg\|\int_0^1 f'\Big(X(t_i)+\theta\big(X(s)-X(t_i)\big)\Big)\,d\theta\\
&\qquad\cdot\Big(\big[X(s)-\sqrt{2}\,W(s)\big]-\big[X(t_i)-\sqrt{2}\,W(t_i)\big]\Big)\bigg\|_{L^2(\Omega;\mathbb{R}^d)}\\
&+\sqrt{2}\bigg\|\int_0^1\Big[f'\Big(X(t_i)+\theta\big(X(s)-X(t_i)\big)\Big)-f'(X(t_i))\Big]\,d\theta\\
&\qquad\cdot\big[W(s)-W(t_i)\big]\bigg\|_{L^2(\Omega;\mathbb{R}^d)}\\
&+\bigg\|\int_0^1 f'\Big(X(t_i)+\theta\big(X(t_i^\tau)-X(t_i)\big)\Big)\,d\theta\\
&\qquad\cdot\Big(\big[X(t_i^\tau)-\sqrt{2}\,W(t_i^\tau)\big]-\big[X(t_i)-\sqrt{2}\,W(t_i)\big]\Big)\bigg\|_{L^2(\Omega;\mathbb{R}^d)}\\
&+\sqrt{2}\bigg\|\int_0^1\Big[f'\Big(X(t_i)+\theta\big(X(t_i^\tau)-X(t_i)\big)\Big)-f'(X(t_i))\Big]\,d\theta\\
&\qquad\cdot\big[W(t_i^\tau)-W(t_i)\big]\bigg\|_{L^2(\Omega;\mathbb{R}^d)}.\\
\le&\,L_1\Big\|\big[X(s)-\sqrt{2}\,W(s)\big]-\big[X(t_i)-\sqrt{2}\,W(t_i)\big]\Big\|_{L^2(\Omega;\mathbb{R}^d)}\\
&+\sqrt{2}\,L_2\big\|X(s)-X(t_i)\big\|_{L^4(\Omega;\mathbb{R}^d)}\big\|W(s)-W(t_i)\big\|_{L^4(\Omega;\mathbb{R}^d)}\\
&+L_1\Big\|\big[X(t_i^\tau)-\sqrt{2}\,W(t_i^\tau)\big]-\big[X(t_i)-\sqrt{2}\,W(t_i)\big]\Big\|_{L^2(\Omega;\mathbb{R}^d)}\\
&+\sqrt{2}\,L_2\big\|X(t_i^\tau)-X(t_i)\big\|_{L^2(\Omega;\mathbb{R}^d)}\big\|W(t_i^\tau)-W(t_i)\big\|_{L^2(\Omega;\mathbb{R}^d)}\\
=:&\,J_1+J_2+J_3+J_4.
\end{aligned}
$$

By denoting

$$Z(s):=X(s)-\sqrt{2}\,W(s),$$

one easily gets

$$Z(s)-Z(t_i)=\int_{t_i}^{s} f(X(r))\,dr,$$

and similarly

$$Z(t_i^\tau)-Z(t_i)=\int_{t_i}^{t_i^\tau} f(X(r))\,dr.$$

Therefore

$$
\begin{aligned}
J_1+J_3&\le L_1\big\|Z(s)-Z(t_i)\big\|_{L^2(\Omega;\mathbb{R}^d)}+L_1\big\|Z(t_i^\tau)-Z(t_i)\big\|_{L^2(\Omega;\mathbb{R}^d)}\\
&\le 2L_1\int_{t_i}^{t_{i+1}}\big\|f(X(r))\big\|_{L^2(\Omega;\mathbb{R}^d)}\,dr\\
&\le 2L_1\Big(\widehat{C}\sqrt{d}+\|X_0\|_{L^2(\Omega;\mathbb{R}^d)}\Big)h.
\end{aligned}
$$

Using the Hölder continuity in time for the Langevin dynamics and the Hölder inequality gives

$$J_2+J_4\le 2\sqrt{2}\,L_2\Big(\widehat{C}d+\sqrt{d}\,\|X_0\|_{L^2(\Omega;\mathbb{R}^d)}\Big)h\le 2\sqrt{2}\,L_2\Big((\widehat{C}+1)d+\|X_0\|^2_{L^4(\Omega;\mathbb{R}^d)}\Big)h.$$

As a consequence,

$$\big\|\phi(s)-\phi(t_i^\tau)\big\|_{L^2(\Omega;\mathbb{R}^d)} \le \Big(2L_1\widehat{C}+2\sqrt{2}\,L_2(\widehat{C}+1)\Big)\Big(d+\|X_0\|^2_{L^4(\Omega;\mathbb{R}^d)}\Big)h.$$

Recall that

$$\begin{aligned}\mathbb{E}_\tau\left[\left|\sum_{i=0}^n \widehat{\widehat{R}}_{i+1}^{(1)}\right|^2\right] &= \sum_{i=0}^n \mathbb{E}_\tau\left[\left|\widehat{\widehat{R}}_{i+1}^{(1)}\right|^2\right]\\ &= \sum_{i=0}^n \mathbb{E}_\tau\left[\left|\int_{t_i}^{t_{i+1}}\big[\phi(s)-\phi(t_i^\tau)\big]\,\mathrm{d}s\right|^2\right]\\ &\le h\sum_{i=0}^n\int_{t_i}^{t_{i+1}}\mathbb{E}_\tau\Big[|\phi(s)-\phi(t_i^\tau)|^2\Big]\,\mathrm{d}s.\end{aligned}$$

and therefore

$$\begin{aligned}\mathbb{E}\left[\left|\sum_{i=0}^n \widehat{\widehat{R}}_{i+1}^{(1)}\right|^2\right] &\le h\sum_{i=0}^n\int_{t_i}^{t_{i+1}}\big\|\phi(s)-\phi(t_i^\tau)\big\|^2_{L^2(\Omega;\mathbb{R}^d)}\,\mathrm{d}s\\ &\le C(t_n,L_1,L_2,\widehat{C})\Big(d+\|X_0\|^2_{L^4(\Omega;\mathbb{R}^d)}\Big)^2h^3,\end{aligned}$$

where $C(t_n,L_1,L_2,\widehat{C})$ depends on $t_n,L_1,L_2,\widehat{C}$.

From the error equation, we deduce

$$\begin{aligned}\|e_{n+1}\|_{L^2(\Omega;\mathbb{R}^d)} &\le hL_1\sum_{i=0}^n\big\|\widehat{X}(t_i^\tau)-\bar{Y}_i^\tau\big\|_{L^2(\Omega;\mathbb{R}^d)}\\ &\quad+\sqrt{2}\,L_2\sum_{i=0}^n\|e_i\|_{L^2(\Omega;\mathbb{R}^d)}\,h^{3/2}\\ &\quad+\left\|\sum_{i=0}^n\widehat{\widehat{R}}_{i+1}\right\|_{L^2(\Omega;\mathbb{R}^d)}.\end{aligned}$$

where we used the gradient Lipschitz condition and the Hessian Lipschitz condition. Noting that

$$\begin{aligned}\big\|\widehat{X}(t_i^\tau)-\bar{Y}_i^\tau\big\|_{L^2(\Omega;\mathbb{R}^d)} &\le \|e_i\|_{L^2(\Omega;\mathbb{R}^d)}+h\big\|f(X(t_i))-f(Y_i)\big\|_{L^2(\Omega;\mathbb{R}^d)}\\ &\le (1+L_1h)\|e_i\|_{L^2(\Omega;\mathbb{R}^d)},\end{aligned}$$

we further derive

$$\begin{aligned}\|e_{n+1}\|_{L^2(\Omega;\mathbb{R}^d)} &\le h\Big(L_1+L_1^2h+\sqrt{2}\,L_2\sqrt{h}\Big)\sum_{i=0}^n\|e_i\|_{L^2(\Omega;\mathbb{R}^d)}\\ &\quad+C(t_{n+1},L_1,L_2,\widehat{C},C_p)\Big(d+\|X_0\|^2_{L^4(\Omega;\mathbb{R}^d)}\Big)h^{3/2}.\end{aligned}$$

By the Gronwall inequality, we can arrive at

$$\sup_{0\le n\le N}\|e_n\|_{L^2(\Omega;\mathbb{R}^d)} \le C(T)\Big(d+\|X_0\|^2_{L^4(\Omega;\mathbb{R}^d)}\Big)\,h^{3/2},$$

where $C(T):=C(L_1,L_2,\widehat{C},C_p)e^{T(L_1+L_1^2+\sqrt{2}L_2)}$ is nondecreasing in T.

• **Part II: Analysis of finite-time error bounds for RRK-LMC**.

Similarly as before, we define a remainder term

$$\begin{aligned}\check{R}_{n+1} :=& \int_{t_n}^{t_{n+1}} f(X(s))\,\mathrm{d}s - hf\big(\widehat{X}(t_n^\tau)\big)\\ &-\sqrt{2}\,f\left(X(t_n)+\int_{t_n}^{t_{n+1}}\big[W(s)-W(t_n^\tau)\big]\,\mathrm{d}s\right)\\ &+\sqrt{2}\,f(X(t_n)).\end{aligned}$$

Then the true solution of OLD can be rewritten as

$$X(t_{n+1}) = X(t_n) + hf\big(\widehat{X}(t_n^\tau)\big) + \sqrt{2}\, f\bigg(X(t_n) + \int_{t_n}^{t_{n+1}} \big[W(s) - W(t_n^\tau)\big]\mathrm{d}s\bigg) - \sqrt{2}\, f(X(t_n)) + \check{R}_{n+1}.$$

Likewise,

$$e_n^z := X(t_n) - Z_n$$

satisfies

$$\begin{aligned} e_{n+1}^z = e_n^z &+ h\Big[f\big(\widehat{X}(t_n^\tau)\big) - f(\bar{Z}_n^\tau)\Big] \\ &+ \sqrt{2}\int_0^1 f'\bigg(X(t_n) + \theta\int_{t_n}^{t_{n+1}} \big[W(s) - W(t_n^\tau)\big]\mathrm{d}s\bigg)\, d\theta \int_{t_n}^{t_{n+1}} \big[W(s) - W(t_n^\tau)\big]\mathrm{d}s \\ &- \sqrt{2}\int_0^1 f'\bigg(Z_n + \theta\int_{t_n}^{t_{n+1}} \big[W(s) - W(t_n^\tau)\big]\mathrm{d}s\bigg)\, d\theta \int_{t_n}^{t_{n+1}} \big[W(s) - W(t_n^\tau)\big]\mathrm{d}s \\ &+ \check{R}_{n+1}. \end{aligned}$$

and noting $e_0^z = 0$.

By iteration, we further derive

$$\begin{aligned} e_n^z = h\sum_{i=0}^{n-1}&\Big[f\big(\widehat{X}(t_i^\tau)\big) - f(\bar{Z}_i^\tau)\Big] \\ &+ \sqrt{2}\sum_{i=0}^{n-1}\int_0^1 \bigg\{ f'\bigg(X(t_i) + \theta\int_{t_i}^{t_{i+1}} \big[W(s) - W(t_i^\tau)\big]\mathrm{d}s\bigg) \\ &\qquad - f'\bigg(Z_i + \theta\int_{t_i}^{t_{i+1}} \big[W(s) - W(t_i^\tau)\big]\mathrm{d}s\bigg)\bigg\} d\theta \\ &\cdot \int_{t_i}^{t_{i+1}} \big[W(s) - W(t_i^\tau)\big]\mathrm{d}s + \sum_{i=0}^{n-1} \check{R}_{i+1}. \end{aligned}$$

Owing to the gradient Lipschitz and Hessian Lipschitz conditions, we arrive at

$$\begin{aligned} \|e_n^z\|_{L^2(\Omega;\mathbb{R}^d)} \le hL_1 &\sum_{i=0}^{n-1} \big\|\widehat{X}(t_i^\tau) - \bar{Z}_i^\tau\big\|_{L^2(\Omega;\mathbb{R}^d)} \\ &+ \sqrt{2}\, L_2 \sum_{i=0}^{n-1} \|e_i^z\|_{L^2(\Omega;\mathbb{R}^d)}\, h^{3/2} \\ &+ \bigg\|\sum_{i=0}^{n-1} \check{R}_{i+1}\bigg\|_{L^2(\Omega;\mathbb{R}^d)}. \end{aligned}$$

Similarly, one knows

$$\begin{aligned} \big\|\widehat{X}(t_i^\tau) - \bar{Z}_i^\tau\big\|_{L^2(\Omega;\mathbb{R}^d)} &\le \|e_i^z\|_{L^2(\Omega;\mathbb{R}^d)} + h\big\|f(X(t_i)) - f(Z_i)\big\|_{L^2(\Omega;\mathbb{R}^d)} \\ &\le (1 + L_1 h)\|e_i^z\|_{L^2(\Omega;\mathbb{R}^d)}. \end{aligned}$$

and thus

$$\|e_n^z\|_{L^2(\Omega;\mathbb{R}^d)} \le h\Big(L_1 + L_1^2 h + \sqrt{2}\, L_2\sqrt{h}\Big)\sum_{i=0}^{n-1} \|e_i^z\|_{L^2(\Omega;\mathbb{R}^d)} + \bigg\|\sum_{i=0}^{n-1} \check{R}_{i+1}\bigg\|_{L^2(\Omega;\mathbb{R}^d)}.$$

Now it remains to bound the remainder terms. We first note

$$\begin{aligned}\check{R}_{i+1} &= \hat{\hat{R}}_{i+1} + \sqrt{2}\, f'(X(t_i)) \int_{t_i}^{t_{i+1}} \big[W(s) - W(t_i^\tau)\big]\mathrm{d}s \\ &\quad - \sqrt{2}\, f\bigg(X(t_i) + \int_{t_i}^{t_{i+1}} \big[W(s) - W(t_i^\tau)\big]\mathrm{d}s\bigg) + \sqrt{2}\, f(X(t_i)) \\ &= \hat{\hat{R}}_{i+1} + \int_0^1 \bigg[f'(X(t_i)) - f'\bigg(X(t_i) + \theta \int_{t_i}^{t_{i+1}} \big[W(s) - W(t_i^\tau)\big]\mathrm{d}s\bigg)\bigg] d\theta \\ &\qquad \cdot \int_{t_i}^{t_{i+1}} \big[W(s) - W(t_i^\tau)\big]\mathrm{d}s.\end{aligned}$$

As a consequence of the Hessian Lipschitz condition, we get

$$\begin{aligned}\bigg\|\sum_{i=0}^{n-1} \check{R}_{i+1}\bigg\|_{L^2(\Omega;\mathbb{R}^d)} &\le \bigg\|\sum_{i=0}^{n-1} \hat{\hat{R}}_{i+1}\bigg\|_{L^2(\Omega;\mathbb{R}^d)} + \sum_{i=0}^{n-1} \bigg\| L_2 \bigg| \int_{t_i}^{t_{i+1}} \big[W(s) - W(t_i^\tau)\big]\mathrm{d}s\bigg|^2\bigg\|_{L^2(\Omega;\mathbb{R}^d)} \\ &\le C(t_{n+1}, L_1, L_2, \widehat{C}, C_p)\Big(d + \|X_0\|^2_{L^4(\Omega;\mathbb{R}^d)}\Big)h^{3/2} + C(t_{n+1}, L_2) d\, h^2 \\ &\le C(t_{n+1}, L_1, L_2, \widehat{C}, C_p)\Big(d + \|X_0\|^2_{L^4(\Omega;\mathbb{R}^d)}\Big)h^{3/2}.\end{aligned}$$

Therefore, a Gronwall inequality gives

$$\sup_{0\le n\le N} \|e_n^z\|_{L^2(\Omega;\mathbb{R}^d)} \le C(T, L_1, L_2, \widehat{C}, C_p)\Big(d + \|X_0\|^2_{L^4(\Omega;\mathbb{R}^d)}\Big)h^{3/2}.$$

□

B PROOF OF UNIFORM-IN-TIME MOMENT BOUNDS

Proof of Proposition 3.6. For a uniform stepsize $h > 0$, let $t_n := nh$, $n \in \mathbb{N}$, and let $\mathcal{F}^\tau_{t_n}$ denote the σ-field generated by $\tau_0, \tau_1, \ldots, \tau_{n-1}$. Then we define

$$\mathcal{F}_{t_n} := \mathcal{F}^W_{t_n} \otimes \mathcal{F}^\tau_{t_n},$$

where $\{\mathcal{F}^W_t\}_{t\ge 0}$ is the filtration previously introduced for the Brownian motion. Throughout the proof, C denotes a positive constant depending only on μ, μ', L_1, L_1', whose value may change from line to line.

We first prove the first assertion. By (21) and $U \in C^2(\mathbb{R}^d)$,

$$\|\nabla^2 U(x)\| \le L_1, \qquad x \in \mathbb{R}^d. \tag{32}$$

By (21),

$$\begin{aligned}&\mathbb{E}\Big[\big|h\big(\nabla U(\bar{Y}_n^\tau) - \nabla U(Y_n)\big)\big|^4 \,\Big|\, \mathcal{F}_{t_n}\Big] \\ &\quad \le L_1^4 h^4 \mathbb{E}\Big[\Big|-\tau_n h \nabla U(Y_n) + \sqrt{2}\big(W(t_n^\tau) - W(t_n)\big)\Big|^4 \,\Big|\, \mathcal{F}_{t_n}\Big] \\ &\quad \le 8L_1^4 h^8 |\nabla U(Y_n)|^4 + 32 L_1^4 h^4 \mathbb{E}\big[|W(t_n^\tau) - W(t_n)|^4 \,\big|\, \mathcal{F}_{t_n}\big].\end{aligned} \tag{33}$$

Since

$$\mathbb{E}\big[|W(t_n^\tau) - W(t_n)|^4 \,\big|\, \mathcal{F}_{t_n}\big] \le 3d^2h^2,$$

and equation 22 gives

$$|\nabla U(Y_n)|^4 \le 8L_1^4 |Y_n|^4 + 8(L_1')^4 d^2,$$

we obtain, for $h \le 1$,

$$\mathbb{E}\Big[\big|h\big(\nabla U(\bar{Y}_n^\tau) - \nabla U(Y_n)\big)\big|^4 \,\Big|\, \mathcal{F}_{t_n}\Big] \le 64 L_1^8 h^8 |Y_n|^4 + C d^2 h^6. \tag{34}$$

Recall that

$$\Delta\Xi_{n+1} = \int_{t_n}^{t_{n+1}} \big(W(s) - W(t_n)\big)\,\mathrm{d}s.$$

Conditionally on $\mathcal{F}_{t_n}$ and τ_n, the vector

$$\sqrt{2}\,\Delta W_{n+1} - \sqrt{2}\,\nabla^2 U(Y_n)\left[\Delta\Xi_{n+1} - h\big(W(t_n^\tau) - W(t_n)\big)\right]$$

is centered Gaussian. In particular, its conditional first and third moments vanish. For each Brownian coordinate, the conditional variance of the expression inside the square brackets is

$$h^3\left(\tfrac{1}{3} - \tau_n + \tau_n^2\right) \le \tfrac{h^3}{3}.$$

Consequently, using (32) and $L_1 h \le 1$, the covariance matrix of the centered Gaussian vector has trace at most

$$4dh + \tfrac{4}{3}L_1^2 dh^3 \le 6dh.$$

The Gaussian fourth-moment formula therefore yields

$$\mathbb{E}\left[\left|\sqrt{2}\,\Delta W_{n+1} - \sqrt{2}\,\nabla^2 U(Y_n)\left[\Delta\Xi_{n+1} - h\big(W(t_n^\tau) - W(t_n)\big)\right]\right|^2 \,\middle|\, \mathcal{F}_{t_n}\right] \le 6dh, \tag{35}$$

$$\mathbb{E}\left[\left|\sqrt{2}\,\Delta W_{n+1} - \sqrt{2}\,\nabla^2 U(Y_n)\left[\Delta\Xi_{n+1} - h\big(W(t_n^\tau) - W(t_n)\big)\right]\right|^4 \,\middle|\, \mathcal{F}_{t_n}\right] \le 108d^2h^2. \tag{36}$$

By (25) and (22),

$$\begin{aligned}
|Y_n - h\nabla U(Y_n)|^2 &= |Y_n|^2 - 2h\langle Y_n, \nabla U(Y_n)\rangle + h^2|\nabla U(Y_n)|^2 \\
&\le (1 - 2\mu h + 2L_1^2h^2)|Y_n|^2 + \big(2\mu' h + 2(L_1')^2h^2\big)d \\
&\le \big(1 - \tfrac{3}{2}\mu h\big)\,|Y_n|^2 + Cdh.
\end{aligned} \tag{37}$$

Since $\mu h \le 1/4$,

$$\big(1 - \tfrac{3}{2}\mu h\big)^2 \le 1 - 2\mu h.$$

Expanding the fourth power and using the vanishing conditional odd moments, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\left|Y_n - h\nabla U(Y_n) + \sqrt{2}\,\Delta W_{n+1} - \sqrt{2}\,\nabla^2 U(Y_n)\left[\Delta\Xi_{n+1} - h\big(W(t_n^\tau) - W(t_n)\big)\right]\right|^4 \,\middle|\, \mathcal{F}_{t_n}\right] \\
&\quad\le |Y_n - h\nabla U(Y_n)|^4 + 36dh|Y_n - h\nabla U(Y_n)|^2 + 108d^2h^2 \\
&\quad\le (1 - 2\mu h)|Y_n|^4 + Cdh|Y_n|^2 + Cd^2h^2 \\
&\quad\le \big(1 - \tfrac{7}{4}\mu h\big)\,|Y_n|^4 + Cd^2h.
\end{aligned} \tag{38}$$

The last inequality follows from Young's inequality, which allows the mixed term to be bounded by

$$Cd|Y_n|^2 \le \tfrac{\mu}{4}|Y_n|^4 + Cd^2.$$

Rewrite the numerical update as

$$\begin{aligned}
Y_{n+1} = Y_n &- h\nabla U(Y_n) + \sqrt{2}\,\Delta W_{n+1} \\
&- \sqrt{2}\,\nabla^2 U(Y_n)\left[\Delta\Xi_{n+1} - h\big(W(t_n^\tau) - W(t_n)\big)\right] \\
&- h\big(\nabla U(\bar{Y}_n^\tau) - \nabla U(Y_n)\big).
\end{aligned}$$

For any $\varepsilon > 0$, convexity gives

$$|a+b|^4 \le (1+\varepsilon)^3|a|^4 + (1+\varepsilon^{-1})^3|b|^4.$$

Taking $\varepsilon = \mu h/16$ and using $\mu h \le 1/4$, we have

$$(1+\varepsilon)^3 \le 1 + \tfrac{\mu h}{4}, \qquad (1+\varepsilon^{-1})^3 \le \tfrac{2^{15}}{\mu^3 h^3}.$$

Applying this inequality to the preceding decomposition, and using (34) and (38), yields

$$\mathbb{E}\left[|Y_{n+1}|^4 \,\middle|\, \mathcal{F}_{t_n}\right] \le \left[\left(1 + \tfrac{\mu h}{4}\right)\left(1 - \tfrac{7}{4}\mu h\right) + \tfrac{2^{21}L_1^8 h^5}{\mu^3}\right]|Y_n|^4 + Cd^2h. \tag{39}$$

Here we used $h \le 1$ to absorb the additive Cd^2h^3 term into Cd^2h. The restriction $h \le \mu/(64L_1^2)$ implies

$$\frac{2^{21}L_1^8h^5}{\mu^3} \le \frac{1}{8}\mu h.$$

Therefore,

$$\left(1+\tfrac{\mu h}{4}\right)\left(1-\tfrac{7}{4}\mu h\right) + \frac{2^{21}L_1^8h^5}{\mu^3} \le 1 - \tfrac{11}{8}\mu h \le 1 - \tfrac{\mu h}{4}.$$

We conclude that

$$\mathbb{E}\left[|Y_{n+1}|^4 \,\middle|\, \mathcal{F}_{t_n}\right] \le \left(1-\tfrac{\mu h}{4}\right)|Y_n|^4 + Cd^2h. \tag{40}$$

Taking expectations in equation 40 and iterating, we obtain

$$\begin{aligned}\mathbb{E}|Y_n|^4 &\le \left(1-\tfrac{\mu h}{4}\right)^n \mathbb{E}|X_0|^4 + Cd^2h\sum_{j=0}^{n-1}\left(1-\tfrac{\mu h}{4}\right)^j \\ &\le e^{-\mu nh/4}\mathbb{E}|X_0|^4 + \tfrac{4C}{\mu}d^2.\end{aligned}$$

Taking fourth roots gives

$$\begin{aligned}\|Y_n\|_{L^4(\Omega;\mathbb{R}^d)} &\le e^{-\mu nh/16}\|X_0\|_{L^4(\Omega;\mathbb{R}^d)} + \left(\tfrac{4C}{\mu}\right)^{1/4}\sqrt{d} \\ &\le \bar{C}\sqrt{d} + \|X_0\|_{L^4(\Omega;\mathbb{R}^d)},\end{aligned}$$

where $\bar{C} = (4C/\mu)^{1/4}$. This proves the first assertion.

The uniform fourth-moment bound for the randomized Runge–Kutta scheme follows by an argument analogous to the previous proof. The gradient-difference correction is controlled using (21), yielding

$$\|Z_n\|_{L^4(\Omega;\mathbb{R}^d)} \le \bar{C}\sqrt{d} + \|X_0\|_{L^4(\Omega;\mathbb{R}^d)}, \qquad n \ge 0,$$

where $\bar{C}$ is independent of n, d, and h. We omit the details. □