The Picard–Lagrange Framework for Higher-Order Langevin Monte Carlo

Jaideep Mahajan*†, Kaihong Zhang*†, Feng Liang†, Jingbo Liu†

Abstract

Sampling from log-concave distributions is a central problem in statistics and machine learning. Prior work establishes theoretical guarantees for Langevin Monte Carlo algorithm based on overdamped and underdamped Langevin dynamics and, more recently, some third-order variants. In this paper, we introduce a new sampling algorithm built on a general Kth-order Langevin dynamics, extending beyond second- and third-order methods. To discretize the Kth-order dynamics, we approximate the drift induced by the potential via Lagrange interpolation and refine the node values at the interpolation points using Picard-iteration corrections, yielding a flexible scheme that fully utilizes the acceleration of higher-order Langevin dynamics. For targets with smooth, strongly log-concave densities, we prove dimension-dependent convergence in Wasserstein distance: the sampler achieves ε-accuracy within $\widetilde{O}(d^{\frac{K-1}{2K-3}}\varepsilon^{-\frac{2}{2K-3}})$ gradient evaluations for $K \geq 3$. To our best knowledge, this is the first sampling algorithm achieving such query complexity. The rate improves with the order K increases, yielding better rates than existing first to third-order approaches.

Contents

*These authors contributed equally to this work.

†Department of Statistics, University of Illinois Urbana-Champaign.

jaideep3@illinois.edu, kaihong5@illinois.edu, liangf@illinois.edu, jingbol@illinois.edu

1 Introduction

Sampling from high-dimensional log-concave distributions remains a central algorithmic primitive in statistics, machine learning, and scientific computing. Formally, the goal is to generate approximate

samples from a target density

$$p^*(x) \;\propto\; \exp(-U(x)), \qquad x \in \mathbb{R}^d, \tag{1}$$

where $U : \mathbb{R}^d \to \mathbb{R}$ is a smooth, strongly convex potential function. Computing the normalizing constant $Z = \int_{\mathbb{R}^d} e^{-U(x)} dx$ directly is infeasible, as numerical integration scales exponentially in d. This motivates the study of direct sampling algorithms that bypass explicit evaluation of Z.

Beyond the design of such algorithms, a key objective is to characterize their complexity, defined as the number of oracle queries required to obtain an approximate sample, and to understand its dependence on the dimension d and the error tolerance ϵ. Throughout, we assume only access to a black-box gradient oracle, which, given a point $x \in \mathbb{R}^d$, returns $\nabla U(x)$.

A widely used paradigm is to employ the Langevin diffusion, which is the solution to the stochastic differential equation (SDE)

$$\mathrm{d}X(t) = -\nabla U(X(t))\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t. \tag{2}$$

Under mild assumptions, the Langevin diffusion admits p^* as its unique stationary distribution and converges to it exponentially fast, without explicit dependence on dimension. However, in practice, the continuous-time dynamics must be discretized, e.g., via the Euler–Maruyama scheme, leading to discretization error that scales poorly with both the dimension and the conditioning of the problem. (Dalalyan, 2017; Dalalyan and Karagulyan, 2019; Durmus and Moulines, 2019; Vempala and Wibisono, 2019; Durmus et al., 2019; He et al., 2020). A recurring lesson is that while continuous-time convergence is rapid and dimension-free, discretization error becomes the main bottleneck.

One strategy to mitigate this issue is to augment the dynamics, creating smoother trajectories that are easier to discretize. For example, the second-order (or underdamped) Langevin algorithm *lifts* the dynamics from the original d-dimensional space to a $2d$-dimensional space, pairing each position vector X_1 with a momentum variable X_2:

$$\begin{cases} \mathrm{d}X_1(t) = X_2(t)\,\mathrm{d}t, \\ \mathrm{d}X_2(t) = -\gamma X_2(t)\,\mathrm{d}t - \nabla U(X_1(t))\,\mathrm{d}t + \sqrt{2\gamma}\,\mathrm{d}B_t, \end{cases}$$

where $(X_1(t), X_2(t)) \in \mathbb{R}^d \times \mathbb{R}^d$ and $\gamma > 0$ is a friction coefficient. The target distribution p^* is recovered as the marginal over X_1. The lift introduces additional smoothness in the trajectory of X_1, which has been exploited to design more accurate discretizations and improve query complexity (Cheng et al., 2018b,a; Dalalyan and Riou-Durand, 2020; Shen and Lee, 2019; Ma et al., 2021; Zhang et al., 2023).

Building on this idea, recent work has demonstrated that *further lifting* can yield algorithmic gains: a third-order Langevin system augments $(X_1(t), X_2(t))$ with an additional auxiliary variable, achieving sharper complexity bounds than its second-order counterpart under the same smoothness/log-

concavity assumptions (Mou et al., 2021). In parallel, the probability literature has analyzed higher-order systems in continuous time via hypocoercivity and Lyapunov techniques (Villani, 2009; Monmarché, 2023), suggesting a natural family of lifted SDEs. Taken together, these developments raise the following central question:

> *Can one develop a principled, scalable framework for K-th order Langevin dynamics whose discretizations systematically improve query complexity as K grows?*

Our starting point is a K-th order generalization of Langevin dynamics that couples the position variable with $K-1$ auxiliary variables in a structured way, ensuring exponential convergence to p^* in continuous time. The main challenge lies in discretization: the difficulty reduces to evaluating time integrals of the nonlinear force ∇U along the trajectory.

For third-order dynamics, Mou et al. (2021) proposed a splitting scheme in which the nonlinear force is approximated by a Lagrange polynomial, so that the required values at interpolation nodes can be read off along a linearized path. However, extending this splitting framework to general order K appears considerably more involved.

Without splitting, the use of polynomial approximation requires knowledge of the $\nabla U(.)$ at the interpolation nodes. To avoid this limitation, we employ a *Picard iteration* (see, e.g., Lee et al. (2018); Shen and Lee (2019); Anari et al. (2024); Yu and Dalalyan (2025)), which iteratively refines the node values and thereby yields consistent approximations at the interpolation points. This construction integrates polynomial approximation directly into the Picard framework, leveraging the smoothness of higher-order dynamics while retaining first-order gradient queries per update. The resulting discretization is systematic, stable under stochastic forcing, and scalable to arbitrary order K.

Our contributions are as follows:

1. We introduce the *Picard–Lagrange* discretization framework for higher-order Langevin dynamics, which leverages higher-order smoothness without requiring higher-order derivatives, and yields stable, tractable updates.

2. We show that the query complexity improves with K under higher-order smoothness assumptions, supported by non-asymptotic bounds on discretization error and mixing time in Wasserstein distance.

Related works. The most closely related work is the recent study of Dang et al. (2025), which employs a Taylor expansion combined with a splitting scheme under substantially stronger regularity assumptions. In particular, their method requires access to derivatives of ∇U up to arbitrary order K, enabling a high-order expansion. In contrast, our framework operates under the standard first-order oracle model—only evaluations of ∇U are assumed available—yet achieves comparable convergence and complexity guarantees.

Beyond these higher-order schemes, several alternative directions have been explored. Mirror and proximal sampling methods modify the geometry or incorporate proximal maps (Wibisono, 2019;

Chewi et al., 2020; Jiang, 2021; Salim and Richtarik, 2020; Ahn and Chewi, 2021; Fan et al., 2023), while the Metropolis-adjusted Langevin algorithm (MALA) augments the discretized chain with a Metropolis correction to ensure exact stationarity (Lee et al., 2018; Dwivedi et al., 2019; Wu et al., 2022; Chen et al., 2020; Altschuler and Chewi, 2024). In contrast, our contribution retains the classical Langevin diffusion as the underlying model and introduces new high-order discretization schemes.

While Langevin-based algorithms are well understood for log-concave targets, their convergence can deteriorate sharply on non-log-concave distributions, often scaling exponentially with the dimension or failing to converge altogether. To address such settings, diffusion-based Monte Carlo samplers have been proposed (Huang et al., 2023, 2024), which construct a forward noising process and learn approximate score functions to reverse it. Other work (Wu et al., 2024; Li et al., 2024, 2025) develops diffusion-based samplers that add Gaussian noise via a forward process and generate samples by simulating reverse-time dynamics. These methods estimate noisy scores from finite samples rather than assume oracle access to the density or its gradient, and thus fall outside the Langevin MC framework we study here.

Organization. The remainder of the paper is organized as follows. Section 2 introduces the notation and assumptions that will be used throughout. We then develop the higher-order Langevin dynamics in Section 3, and present the discretization scheme in Section 4. Our main convergence results are stated in Section 5, with a brief outline of the proof strategy given in Section 6. Finally, Section 7 concludes with a discussion, and technical lemmas are deferred to the appendix.

2 Preliminaries

We denote the standard Euclidean norm by $\|\cdot\|$ and the inner product by $\langle\cdot,\cdot\rangle$. To quantify higher-order derivatives, we next introduce the notion of a tensor spectral norm.

Definition 1 (Tensor spectral norm)**.** For a vector–valued k-th order tensor T, we define

$$\|T\|_{\mathrm{tsr}} := \sup_{v_1,\dots,v_{k-1}\in S^{d-1}} \big\|T\cdot[v_1,\dots,v_{k-1}]\big\|_2,$$

where S^{d-1} is the unit sphere in $\mathbb{R}^d$, and "$\cdot$" denotes the natural tensor contraction.

2.1 Assumptions on U

We make the following assumptions on the potential function U in (1).

Assumption 1 (Strong convexity and smoothness)**.** There exist positive constants $m \le L$ such that

$$\frac{m}{2}\|x'-x\|^2 \;\le\; U(x') - U(x) - \langle \nabla U(x), x'-x\rangle \;\le\; \frac{L}{2}\|x'-x\|^2, \qquad \forall\, x, x' \in \mathbb{R}^d.$$

Assumption 1 is equivalent to $mI_d \preceq \nabla^2 U(x) \preceq LI_d$.

Assumption 2 (Higher–order-smoothness)**.** There exist constants $L_1, \ldots, L_{K-1}$ such that

$$\left\| D^i \nabla U(x) \right\|_{\text{tsr}} \;\leq\; L_i, \qquad \forall\, x \in \mathbb{R}^d,\; i \leq K-1,$$

that is, the i-th order derivatives of ∇U are uniformly bounded in tensor spectral norm.

The effectiveness of higher-order polynomial approximations relies on higher-order smoothness of the underlying function. Assumption 2 ensures this by bounding the higher-order derivatives of U, which in turn controls the Lagrange-interpolation remainder when approximating the drift. Specifically, time derivatives of $\nabla U(X(t))$ up to order $K-1$ expand into higher-order derivatives of ∇U, each bounded by the constants L_i.

Assumption 3 (Centered potential)**.** Without loss of generality, assume $\nabla U(0) = 0$ and $U(0) = 0$, which can be enforced by shifting U.

2.2 Convergence Metrics and Notations

Consider an iterative algorithm that generates a random vector $\widehat{X}(nh)$ at the n-th step, corresponding to time $t = nh$ with step size h. Let $\hat{\pi}^{(n)}$ denote the law of $\hat{X}(nh)$. We study convergence $\hat{\pi}^{(n)} \to \pi$, where π is the target measure with density p^*. The distance between two measures is quantified using the *Wasserstein-2* distance:

$$\mathcal{W}_2^2(p, q) = \inf_{\gamma \in \Gamma(p,q)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 \, d\gamma(x, y),$$

where $\Gamma(p, q)$ is the set of couplings of p and q, with each coupling γ being a joint distribution whose marginals are p and q.

We use standard asymptotic notation to express convergence rates: for $f, g : \mathbb{R}^n \to [0, \infty)$, we write $f = O(g)$ if $\exists\, C > 0$ with $f(x) \leq Cg(x)$, $f = \Omega(g)$ if $\exists\, c > 0$ with $f(x) \geq cg(x)$, and $f = \tilde{O}(g)$ if $\exists\, C_1, C_2 > 0$ with $f(x) \leq C_1 g(x)(\log g(x))^{C_2}$, all for sufficiently large $\|x\|$.

3 Higher Order Langevin Dynamics

We begin with the family of SDEs of the form

$$\mathrm{d}X(t) \;=\; -(D + Q)\, \nabla H(X(t))\, \mathrm{d}t \;+\; \sqrt{2D}\, \mathrm{d}B_t, \tag{3}$$

where D is a constant positive semidefinite matrix and Q is a constant skew-symmetric matrix. It can be shown (Shi et al., 2012; Ma et al., 2015) that for any such choice of D and Q, the SDE (3) admits

$$p^*(x) \;\propto\; \exp\left(-H(x)\right)$$

as the invariant distribution. Notably, both the underdamped Langevin dynamics (Cheng et al., 2018b) and the more recent third-order Langevin dynamics (Mou et al., 2021) are special cases of this framework. This naturally raises the question of whether further acceleration can be achieved by extending to even higher-order dynamics.

For general $K \geq 3$, we expand the ambient space as $X = (X_1, X_2, \ldots, X_K)$, where $X_i \in \mathbb{R}^d$ and X_1 denotes the variable of interest, define $H(X) = U(X_1) + \frac{1}{2}\sum_{i=2}^{K} \|X_i\|^2$, and introduce the matrices

$$D = \big(\mathrm{diag}(0, \ldots, 0, \gamma)_{K\times K} \otimes I_d\big), \quad Q = \begin{bmatrix} 0 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -\gamma & \ddots & \vdots \\ 0 & \gamma & 0 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -\gamma \\ 0 & \cdots & 0 & \gamma & 0 \end{bmatrix}_{K\times K} \otimes I_d,$$

for some $\gamma > 0$. The resulting K-th order Langevin dynamics on $\mathbb{R}^{Kd}$ take the form

$$\mathrm{d}X(t) = -\big(D + Q\big) \begin{pmatrix} \nabla U(X_1(t)) \\ X_2(t) \\ \vdots \\ X_K(t) \end{pmatrix} dt + \sqrt{2D}\,\mathrm{d}B_t, \tag{4}$$

and the marginal law of X_1 under its invariant distribution is the target measure.

Thus, the framework in (4) provides a unified formulation of higher-order Langevin dynamics, generalizing both underdamped and third-order cases. In the next section, we turn to the problem of discretizing these dynamics and developing a practical algorithm.

4 Discretization

To implement the dynamics in practice, we must discretize (4). In this section, we provide an overview of the approach, with complete technical details postponed to the Appendix B.

To make the process (4) more tractable, we separate the drift into a linear part that can be solved exactly and a nonlinear part involving the potential.

$$\mathrm{d}X(t) \;=\; A\,X(t)\,\mathrm{d}t \;+\; g(X(t))\,\mathrm{d}t \;+\; \sqrt{2D}\,\mathrm{d}B_t. \tag{5}$$

where we define matrix A and g as follows:

$$g(X(t)) \;=\; \begin{pmatrix} \mathbf{0}_d \\ -\nabla U\big(X_1(t)\big) \\ \mathbf{0}_d \\ \vdots \\ \mathbf{0}_d \end{pmatrix} \in \mathbb{R}^{Kd}, \quad A \;=\; \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \gamma & \ddots & \vdots \\ 0 & -\gamma & 0 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \gamma \\ 0 & \cdots & 0 & -\gamma & -\gamma \end{bmatrix}_{K\times K} \otimes I_d.$$

To illustrate the discretization, consider one step from 0 to h. Applying the variation-of-constants formula to (5) for $0 < \tau \leq 1$ yields

$$X(\tau h) = e^{\tau h A} X(0) + h \int_0^\tau e^{(\tau - s)hA}\, g\big(X(sh)\big)\, \mathrm{d}s \;+\; W(\tau), \tag{6}$$

where $W(\tau) \sim \mathcal{N}(0, \Sigma(\tau))$. The Gaussian term arises from the Brownian increment, and its covariance $\Sigma(\tau)$ depends only on A and D, making it explicitly tractable. Detailed expressions for $\Sigma(\tau)$ are provided in Appendix B.

The only part of (6) that is not directly computable is the nonlinear integral

$$\int_0^\tau e^{(\tau - s)hA}\, g(X(sh))\, \mathrm{d}s.$$

The challenge here is that the drift term $g(X(sh))$ depends on the entire trajectory of X over $[0, \tau]$, which is not explicitly available. To address this, we approximate the time–dependence of the drift by replacing $g(X(sh))$ with a polynomial surrogate that can be integrated against the exponential kernel.

Specifically, we construct a degree-$(M-1)$ Lagrange interpolant from a set of collocation nodes in $[0, 1]$. For simplicity we take equispaced nodes

$$c_1 = 0, \quad c_j = \tfrac{j-1}{M-1}, \quad c_M = 1, \qquad j = 1, \ldots, M,$$

and let $\{\ell_j\}_{j=1}^M$ denote the associated Lagrange basis polynomials. The interpolant is then

$$P(s;\, X) \;:=\; \sum_{j=1}^{M} \ell_j(s)\, g\big(X(c_j h)\big), \quad s \in [0, 1].$$

Moreover, since $P(s; X)$ is polynomial, the kernel integral further simplifies to a weighted combination of drift evaluations at the collocation points:

$$\int_0^\tau e^{(\tau - s)hA}\, g(X(sh))\, \mathrm{d}s \;\approx\; \sum_{j=1}^{M} \alpha_j(\tau, h)\, g\big(X(c_j h)\big).$$

The weights $\alpha_j(\tau, h)$ depend only on A, h, τ, and the chosen nodes $\{c_j\}$; their explicit expressions are given in Appendix B. Using this approximation, one obtains the following update:

$$\widehat{X}(\tau h) = e^{\tau h A} X(0) + h \sum_{j=1}^{M} \alpha_j(\tau, h)\, g\big(\widehat{X}(c_j h)\big) \;+\; W(\tau), \tag{7}$$

This approximation makes the nonlinear integral explicit, but it also introduces a subtle difficulty. The expression for $\widehat{X}(\tau h)$ now involves the intermediate evaluations of ∇U at $\{\widehat{X}(c_j h)\}_{j=1}^{M}$ that appear inside the interpolant P. In other words, the update at time τh depends implicitly on future points along the same step.

This implicit dependence is the key obstacle: the update of X at time τh requires values at the collocation nodes that are themselves defined only through the update. To break this circularity, we adopt a fixed-point strategy based on Picard iterations. Starting from an initial guess for the nodal values, we repeatedly substitute them into (7) until the iterates converge. The resulting scheme is both stable and computationally tractable, and its flow is illustrated in Figure 1.

Formally, we initialize the nodal values by setting them equal to the current state $X(0)$:

$$\widehat{X}^{[0]}(c_j h) = X(0), \qquad j = 1, \ldots, M.$$

This simple initialization provides a consistent starting point for the iteration.

Using these provisional values, we build the interpolant $P(s;\, \widehat{X}^{[0]})$ and plug it into (7), giving the first update for $k = 1, \ldots, M$:

$$\widehat{X}^{[1]}(c_k h) = e^{c_k h A} X(0) + h \sum_{j=1}^{M} \alpha_j(c_k, h)\, g\big(\widehat{X}^{[0]}(c_j h)\big) \;+\; W(c_k).$$

This produces refined values $\{\widehat{X}^{[1]}(c_k h)\}_{k=1}^{M}$, which in turn yield a better interpolant $P(s;\, \widehat{X}^{[1]})$. Substituting this improved approximation back into the update leads to the next iterate, for example

$$\widehat{X}^{[2]}(c_k h) = e^{c_k h A} X(0) + h \sum_{j=1}^{M} \alpha_j(c_k, h)\, g\big(\widehat{X}^{[1]}(c_j h)\big) \;+\; W(c_k).$$

Iterating in this way steadily improves the approximation, and the procedure is guaranteed to converge under mild conditions. In practice, we stop after a fixed number ν^* of Picard iterations, which balances accuracy and cost. The complete routine is outlined in Algorithm 1. For clarity, explicit formulas for the weights $\alpha_j(c_k, h)$, the Gaussian terms $W_n(c_k)$, and the covariance Σ_C are provided in Appendix B.

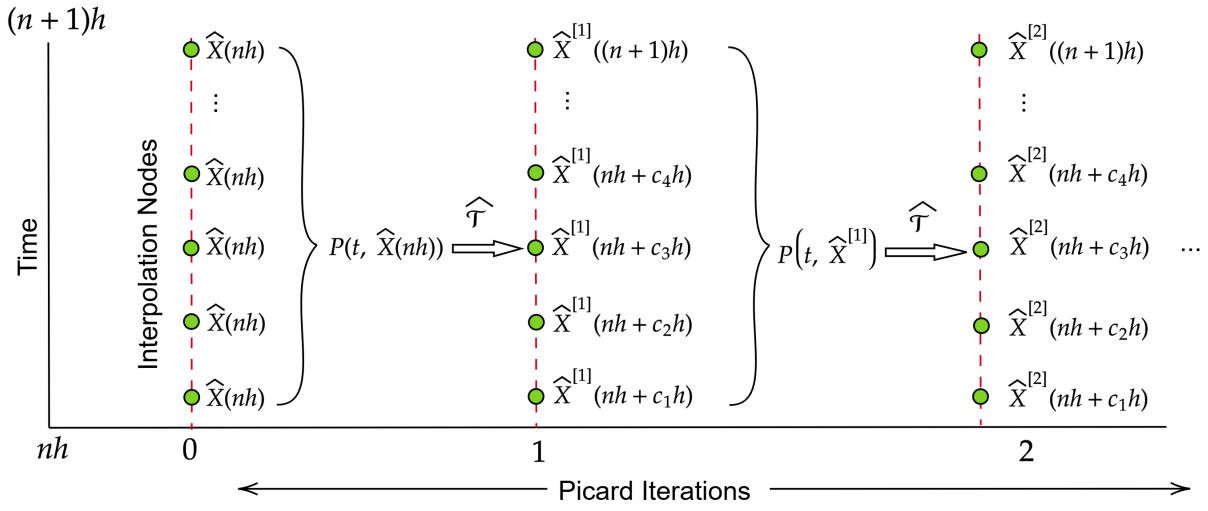

Figure 1: Schematic representation of Picard iterations. The procedure begins with constant values, which are used to form a Lagrange interpolant. Applying the update operator $\widehat{\mathcal{T}}$ at the interpolation nodes yields the next set of values, which then feed into the following Picard iteration.

5 Main Results

In this section, we present our main theorem of the convergence guarantee for our Higher-order Langevin Monte Carlo in Algorithm 1.

Theorem 1. *Under Assumption 1–3 and let $K \geq 3$ be the order of Langevin dynamics. Let $\widehat{\pi}^{(N)}$ be the law of $\widehat{X}_1(Nh) \in \mathbb{R}^d$ after N steps from Algorithm 1 with the number of Picard iterations $\nu_* \geq K-1$ and the number of nodes for Lagrange interpolation $M = K-1$. Then for any error tolerance $\varepsilon \in (0,1)$ it suffices to choose*

$$h = O\left(\left(\frac{\varepsilon^2}{d^{K-1}\log\frac{d}{\varepsilon^2}}\right)^{\frac{1}{2K-3}}\right), \quad N = \Omega\left(\frac{1}{h}\log\frac{d}{\varepsilon^2}\right),$$

to guarantee

$$W_2\left(\widehat{\pi}^{(N)}, \pi\right) \leq \varepsilon.$$

Equivalently, the total number of gradient evaluations needed to achieve $W_2 \leq \varepsilon$ is

$$\widetilde{O}\left(d^{\frac{K-1}{2K-3}}\,\varepsilon^{-\frac{2}{2K-3}}\right),$$

where $\widetilde{O}$ hides polylogarithmic factors in d and $1/\varepsilon$.

Remark 5.1. For underdamped Langevin dynamics ($K = 2$), the best known query complexity is $\widetilde{O}(d^{1/3}\,\varepsilon^{-2/3})$ (Shen and Lee, 2019). The third-order scheme of Mou et al. (2021), under α-th order smoothness assumptions, achieves $\widetilde{O}(d^{1/4}\,\varepsilon^{-1/2} + d^{1/2}\varepsilon^{-1/(\alpha-1)})$. Our general result shows

Algorithm 1 Higher-order Langevin Monte Carlo

Require: $\nabla U : \mathbb{R}^d \to \mathbb{R}^d$; Order K; Initial state $\widehat{X}(0) \in \mathbb{R}^{Kd}$; Total number of steps N; Step size h; Number of collocation points M with nodes $0 = c_1 < \cdots < c_{M-1} < c_M = 1$; Number of Picard iterations ν_*.

Precompute matrices $\{e^{c_k h A}\}$, $\{\alpha_j(c_k, h)\}$, and the noise covariance Σ_C.
for $n = 1, \dots, N$ **do**
  Sample the joint Gaussian noise vector $[W_n(c_1)^\top, \dots, W_n(c_M)^\top]^\top \sim \mathcal{N}(0, \Sigma_C)$.
  for $k = 1, \dots, M$ **do**
    $\widehat{X}^{[0]}(nh + c_k h) \leftarrow \widehat{X}(nh)$.
  end for
  for $\nu = 1, \dots, \nu_*$ **do**
    for $k = 1, \dots, M$ **do**
      $\widehat{X}^{[\nu+1]}(nh + c_k h) \leftarrow e^{c_k h A}\, \widehat{X}(nh)$
      $\quad + h \sum_{j=1}^{M} \alpha_j(c_k, h)\, g\big(\widehat{X}^{[\nu]}(nh + c_j h)\big) + W_n(c_k)$
    end for
  end for
  Update $\widehat{X}((n+1)h) \leftarrow \widehat{X}^{[\nu_*]}((n+1)h)$
end for
return $\widehat{X}(Nh)$.

that for every $K > 3$, the proposed sampler achieves ε-accuracy within $\widetilde{O}(d^{\frac{K-1}{2K-3}}\, \varepsilon^{-\frac{2}{2K-3}})$ gradient evaluations, thereby improving the dependence on the accuracy parameter ε compared to existing methods. The dimension exponent $\frac{K-1}{2K-3}$ converges to $1/2$ as $K \to \infty$, which is suboptimal relative to the $d^{1/3}$ scaling achieved by Shen and Lee (2019) for the underdamped Langevin dynamics.

6 Proofs

In this section, we provide the proofs of Theorem 1. This section is organized as follows: Section 6.1 introduces the setting and notation required for the proof. In Section 6.2, we present a roadmap for the proof by decomposing the error into three components: convergence of the higher-order Langevin dynamics, the Lagrange interpolation error, and the Picard convergence error. Sections 6.3 deal with three errors separately.

6.1 Settings

Operator for Continuous and Discretized Process. Let $\mathcal{C}\big([0,h], \mathbb{R}^{Kd}\big)$ be the space of continuous paths on $[0,h]$ with values in $\mathbb{R}^{Kd}$. Recall that $\nabla H(x) = (\nabla U(x_1), x_2, \dots, x_K)$. For an initial value $y \in \mathbb{R}^{Kd}$, define the operator $\mathcal{T}_y : \mathcal{C}\big([0,h], \mathbb{R}^{Kd}\big) \to \mathcal{C}\big([0,h], \mathbb{R}^{Kd}\big)$ by

$$\big(\mathcal{T}_y[X]\big)(t) := y - \int_0^t (D+Q)\, \nabla H\big(X(s)\big)\, \mathrm{d}s + \int_0^t \sqrt{2D}\, \mathrm{d}B_s, \quad 0 \le t \le h.$$

With this notation, X solves the higher-order Langevin dynamic (4) on $[0,h]$ if and only if X satisfies the fixed point equation $X = \mathcal{T}_{X(0)}[X]$ on $[0,h]$.

In Algorithm 1, we approximate ∇U by Lagrange interpolation $P(s;X) := \sum_{j=1}^{M} \ell_j\left(\frac{s}{h}\right) \nabla U\big(X_1(c_j h)\big)$ (using normalized time s/h). Given $y \in \mathbb{R}^{Kd}$ and an input path $X \in \mathcal{C}([0,h],\mathbb{R}^{Kd})$, we first define $\tilde{X} = \tilde{X}[X;y] \in \mathcal{C}([0,h],\mathbb{R}^{Kd})$ to be the path on $[0,h]$ solving the following SDE

$$\tilde{X}(t) = y - \int_0^t (D+Q) \begin{bmatrix} P(s;X) \\ \tilde{X}_2(s) \\ \vdots \\ \tilde{X}_K(s) \end{bmatrix} \mathrm{d}s + \int_0^t \sqrt{2D}\, \mathrm{d}B_s, \quad 0 \le t \le h,$$

where $\tilde{X}_i(s)$ denotes the i-th block of $\tilde{X}(s)$. We then define the discretized operator $\widehat{\mathcal{T}}_y : \mathcal{C}\big([0,h],\mathbb{R}^{Kd}\big) \to \mathcal{C}\big([0,h],\mathbb{R}^{Kd}\big)$ by

$$\widehat{\mathcal{T}}_y[X] := \tilde{X}[X;y].$$

One can show that the operators $\mathcal{T}_y$ and $\widehat{\mathcal{T}}_y$ admit unique fixed points under the uniform norm $\|\cdot\|_\infty := \sup_{t\in[0,h]} \|\cdot(t)\|$ for any $y \in \mathbb{R}^{Kd}$ and sufficiently small h, see Appendix F.1 (Lemma F.1, Lemma F.2) for the proof of existence of fixed points. We denote the fixed point of $\mathcal{T}_y$ and $\widehat{\mathcal{T}}_y$ by X_y^* and $\widehat{X}_y^*$ correspondingly. Note that by definition $X_y^*(t)$ is the exact solution of the Langevin dynamics (4) on $[0,h]$ initialized at y.

Picard Iterates. Given y, define the Picard sequence on $[0,h]$ by

$$\widehat{X}_y^{[0]}(t) \equiv y, \quad \widehat{X}_y^{[\nu+1]} := \widehat{\mathcal{T}}_y\big[\widehat{X}_y^{[\nu]}\big], \ \nu \ge 0.$$

Starting from $y = \widehat{X}(nh)$, Algorithm 1 outputs the next state by evaluating the ν_*-th Picard iterate at the end of the step:

$$\widehat{X}((n+1)h) = \widehat{X}_y^{[\nu_*]}(h).$$

In Appendix F.2 (Lemma F.4), we provide more details to show that this representation is equivalent to the output of Algorithm 1.

6.2 Error Decomposition

Let $\|\cdot\|_S$ be the norm induced by a fixed $S \succ 0$, and thus $\|z\|_S^2 \le \|S\|_{\mathrm{op}} \|z\|^2$. Under synchronous coupling on each step and $X(0) \sim \pi$ (therefore $X(t) \sim \pi$ for all $t \ge 0$),

$$W_2^2\big(\pi, \hat{\pi}^{(N)}\big) \le \|S^{-1}\|_{\mathrm{op}}\, \mathbb{E}\big[\|X(Nh) - \widehat{X}(Nh)\|_S^2\big].$$

Let $E_n := \mathbb{E}\|X(nh) - \widehat{X}(nh)\|_S^2$. On the interval $[nh,(n+1)h]$, let X_y^* and $\widehat{X}_y^*$ be the fixed–point paths of $\mathcal{T}_y$ and $\widehat{\mathcal{T}}_y$, respectively, driven by the same Brownian process. Adding and subtracting

$X^*_{\widehat{X}(nh)}(h)$ and $\widehat{X}^*_{\widehat{X}(nh)}(h)$ yields

$$E_{n+1} = \mathbb{E}\big\|X((n+1)h) - \widehat{X}((n+1)h)\big\|_S^2$$
$$\leq 3\underbrace{\mathbb{E}\big\|X^*_{X(nh)}(h) - X^*_{\widehat{X}(nh)}(h)\big\|_S^2}_{\text{(I) Continuous time convergence}} + 3\underbrace{\mathbb{E}\big\|X^*_{\widehat{X}(nh)}(h) - \widehat{X}^*_{\widehat{X}(nh)}(h)\big\|_S^2}_{\text{(II) Interpolation Error}} + 3\underbrace{\mathbb{E}\big\|\widehat{X}^*_{\widehat{X}(nh)}(h) - \widehat{X}^{[\nu_*]}_{\widehat{X}(nh)}(h)\big\|_S^2}_{\text{(III) Picard Convergence Error}}. \tag{8}$$

- Term (I) is the error from running the exact continuous dynamics over one step starting from two different initial conditions, $X(nh)$ and $\widehat{X}(nh)$.
- Term (II) is the error from replacing ∇U by its Lagrange interpolant P.
- Term (III) is the finite–iteration error when approximating the fixed point of $\widehat{\mathcal{T}}_{\widehat{X}(nh)}$ by ν_* Picard steps.

See Figure 2 for an illustration of this one-step error decomposition. We will bound each of the three error components in Section 6.3. Combining these three terms, we obtain the following result.

Proposition 1 (One–step error bound)**.** *There exist constants $C_0, C_1, C_2, h_0 > 0$ (independent of h, d, n) such that, for all $h \in (0, h_0]$, $\nu_* \geq K - 1$, and integers $N \geq 0$,*

$$E_{N+1} \leq C_0 e^{-C_S(N+1)h} d + C_1 N h^{2K-2} d^{K-1} + C_2 h^{2\nu_*} d.$$

The proof of Proposition 1 is provided in Appendix D.2. Choosing h, N and ν_* so that each term on the right-hand side is at most $\varepsilon^2/3$ yields Theorem 1. We provide more details for the proof of Theorem 1 in Appendix D.1.

6.3 Bounds for the Three Error Components

Continuous Time Convergence. To bound term (I) in (8), we use the result of continuous-time convergence of the higher–order Langevin dynamics (4). It is well known that first–order Langevin dynamics converge exponentially fast to stationarity (see, e.g., Bakry et al. (2013)); a similar property holds for higher–order dynamics by constructing a suitable quadratic Lyapunov function. The following lemma shows exponential contraction in the S–norm for some suitable positive-definite matrix S.

Proposition 2. *Let the processes $\{X_t\}$ and $\{Y_t\}$ follow the Kth-order Langevin process in (4) with $\gamma \geq \gamma_0$ for some constant γ_0, starting from X_0 and $Y_0 \in \mathbb{R}^{Kd}$. Then there exists a matrix $S \succ 0$ and a constant $C_S > 0$ that only depends on (γ, K, m, L) such that there exists a coupling*

$$\bar{\zeta} \in \Gamma\big(p_t(\cdot \mid X_0),\, p_t(\cdot \mid Y_0)\big),$$

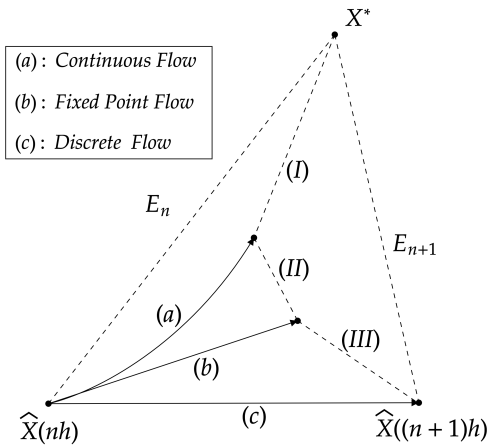

Figure 2: An illustration for one–step error decomposition: Starting from the current iterate $\widehat{X}(nh)$, **(a)** runs the continuous process (equivalently, the fixed–point flow of $\mathcal{T}$), yielding $\widehat{X}(nh) \to X^*_{\widehat{X}(nh)}$; **(b)** runs the fixed–point flow of the discretized operator $\widehat{\mathcal{T}}$, yielding $\widehat{X}(nh) \to \widehat{X}^*_{\widehat{X}(nh)}$; **(c)** is the algorithmic update, $\widehat{X}(nh) \to \widehat{X}((n+1)h)$. The three labeled distances match the terms in (8): **(I)** contraction of continuous process, **(II)** interpolation error $\|X^*_{\widehat{X}(nh)}(h) - \widehat{X}^*_{\widehat{X}(nh)}(h)\|_S^2$, **(III)** Picard convergence error $\|\widehat{X}^*_{\widehat{X}(nh)}(h) - \widehat{X}((n+1)h)\|_S^2$. Hence $E_{n+1} \le (\mathrm{I}) + (\mathrm{II}) + (\mathrm{III})$.

for all $(X_t, Y_t) \sim \bar{\zeta}$ *and all* $t \ge 0$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\,(X_t - Y_t)^\top S(X_t - Y_t) \;\le\; -2C_S\,(X_t - Y_t)^\top S(X_t - Y_t).$$

The proof of Proposition 2 is provided in Appendix C.1. Conditioning on $\mathcal{F}_{nh}$ and applying Proposition 2 with $X_0 = X(nh)$, $Y_0 = \widehat{X}(nh)$, and $t = h$ yields

$$\mathbb{E}\big\|X^*_{X(nh)}(h) - X^*_{\widehat{X}(nh)}(h)\big\|_S^2 \le e^{-2C_S\,h}\,\mathbb{E}\|X(nh) - \widehat{X}(nh)\|_S^2,$$

which controls term (I) by the contraction of the continuous process.

Interpolation Error. Henceforth, unless otherwise stated, we drop the dependence on the initialization $y = \widehat{X}(nh)$ and simply write X^* and $\widehat{X}^*$ for the fixed points of the operators $\mathcal{T}_y$ and $\widehat{\mathcal{T}}_y$, respectively. For the Lagrange interpolation error term (II), we have the following proposition.

Proposition 3 (Interpolation error within one step)**.** *Fix a step size* $h > 0$ *and an index* n*. Let* γ *be as in the higher–order Langevin process* (4) *and* C_S *be defined in Proposition 2. Then*

$$\big\|X^*(h) - \widehat{X}^*(h)\big\|_S^2 \;\le\; e^{-C_S h}\,\frac{9\gamma\,\|S\|_{\mathrm{op}}}{C_S}\,h\,I_U,$$

where

$$I_U \;:=\; \sup_{s\in[nh,(n+1)h]} \big\|\nabla U\big(\widehat{X}_1^*(s)\big) - P\big(s,\widehat{X}_1^*\big)\big\|^2. \tag{9}$$

Here the subscript 1 *denotes the first block (the* x_1*–component) of* $\widehat{X}^*$*.*

The proof of Proposition 3 can be found in Appendix D.3. Proposition 3 implies that

$$\mathbb{E}\,\|X^*(h) - \widehat{X}^*(h)\|_S^2 = O(h)\,\mathbb{E}[I_U],$$

where $\mathbb{E}[I_U]$ is the Lagrange interpolation error incurred when approximating ∇U along the fixed-point path of $\widehat{\mathcal{T}}_y$ by a Lagrange polynomial. If we take $K-1$ nodes for the interpolant, standard Lagrange interpolation analysis shows that

$$\mathbb{E}[I_U] = O\big(h^{2K-2}\big)\;\cdot \sup_{t\in[nh,(n+1)h]} \left\| \frac{d^{K-1}}{dt^{K-1}} \nabla U\big(\widehat{X}_1^*(t)\big)\right\|^2.$$

We then control the higher–order derivative term using the higher–order tensor bounds for ∇U in Assumption 2. Formally, we have the following lemma.

Lemma 4. *Fix a step size* $h>0$ *and an integer order* $K\ge 2$*. On each interval* $[nh,(n+1)h]$*, define the Lagrange interpolant of* $t\mapsto \nabla U\big(\widehat{X}_1^*(t)\big)$ *from the* $K-1$ *uniformly spaced nodes* $\{nh+c_jh\}_{j=1}^{K-1}$ *with* $c_j\in[0,1]$*. Let* I_U *be as in (9). Under Assumption 1-3 with smoothness parameters* $\{L_i\}_{i=1}^{K-1}$*, there exists a constant* $C_{\mathrm{IR}}>0$ *depending only on* $\big(K,\{L_i\}_{i\le K-1},\{c_j\}\big)$ *and the parameter* γ *from matrices* D *and* Q*, but not on* $h,d,n,$ *or* N*, such that, uniformly for all* $n=0,1,\dots,N-1$*,*

$$\mathbb{E}\,I_U \;\le\; C_{\mathrm{IR}}\,N\,h^{\,2K-2}\,d^{\,K-1}.$$

Remark 6.1. We take $K-1$ nodes because, by the definition of D in (4), the Gaussian noise is added only into the last block; consequently, the first coordinate X_1 admits time derivatives up to order $K-1$. The higher–order dynamics thus produce a sufficiently smooth path for X_1 to mitigate the nonsmoothness introduced by Brownian motion.

Remark 6.2. The dimension factor originates from controlling the $(K-1)$-st *time* derivative of $t\mapsto\nabla U\big(\widehat{X}_1^*(t)\big)$ that appears in the error term of the Lagrange polynomial approximation. By Faà di Bruno's formula, this time derivative of $U(\widehat{X}_1^*(t))$ can be expressed in terms of higher derivatives of U acting on time derivatives of $\widehat{X}_1^*$. Under Assumption 2, these contributions reduce to bounds involving moments of $\|\widehat{X}^*(t)\|$. In particular, controlling derivatives up to order $K-1$ requires bounding $\sup_{t\in[nh,(n+1)h]}\|\widehat{X}^*(t)\|^{K-1}$, which is of order $d^{\,K-1}$ with high probability. See the Proof of Lemma 4 in Appendix E for more details.

Picard Convergence Term (III) is the convergence of the finite Picard iteration to its fixed point. We can show that the operator $\widehat{\mathcal{T}}_y$ is a contraction whenever h is small enough, i.e.,

$$\left\|\widehat{\mathcal{T}}_y[X] - \widehat{\mathcal{T}}_y[Y]\right\|_\infty \le \rho \left\|X - Y\right\|_\infty$$

for some $\rho = O(h) < 1$. We refer the reader to Appendix F.1.2 (Lemma F.2) for more details.

We use the following proposition to control the error from finite Picard iterations.

Proposition 5 (Picard iteration error)**.** *Assume the contraction factor ρ of $\widehat{\mathcal{T}}$ satisfies $\rho = O(h) < 1$. Let E_n be the one-step error defined in (8). Then, for every $\nu \ge 0$,*

$$\mathbb{E} \sup_{t\in[0,h]} \left\|\widehat{X}^*(t) - \widehat{X}^{[\nu]}(t)\right\|^2 \le \frac{\rho^{2\nu}}{(1-\rho)^2} \left(O(h^2)\, E_n + O(dh)\right).$$

See Appendix D.4 for the proof of Proposition 5.

Remark 6.3. Let $\|Z\|_{\mathcal{H}} := \left(\mathbb{E}\sup_{t\in[0,h]} \|Z(t)\|^2\right)^{1/2}$. Using the identity $\widehat{X}^* - \widehat{X}^{[\nu]} = \sum_{k=\nu}^{\infty}(\widehat{X}^{[k+1]} - \widehat{X}^{[k]})$ and the contraction of $\widehat{\mathcal{T}}_y$ with factor $\rho = O(h) < 1$, we have $\|\widehat{X}^* - \widehat{X}^{[\nu]}\|_{\mathcal{H}} \le \sum_{k=\nu}^{\infty} \rho^k \|\widehat{X}^{[1]} - \widehat{X}^{[0]}\|_{\mathcal{H}} = \frac{\rho^\nu}{1-\rho} \|\widehat{X}^{[1]} - \widehat{X}^{[0]}\|_{\mathcal{H}}$. We can further bound the initial increment by $\|\widehat{X}^{[1]} - \widehat{X}^{[0]}\|_{\mathcal{H}}^2 \lesssim O(h^2)E_n + O(dh)$, which yields Proposition 5. See more details of the proof in Appendix D.4.

7 Discussion

We introduced a higher-order Langevin Monte Carlo built on a general K-th-order Langevin dynamics and discretized it via a Picard–Lagrange scheme. We established non-asymptotic convergence in Wasserstein distance and proved query complexity of $\widetilde{O}\left(d^{\frac{K-1}{2K-3}}\varepsilon^{-\frac{2}{2K-3}}\right)$ for $K \ge 3$; this improves the ε-dependence over the existing first-, second- (underdamped), and third-order Langevin Monte Carlo and continues to improve as K increases.

While the ε-dependence improves as the order K increases, our current bound carries a dimension factor $d^{1/2+\delta}$ (for a small $\delta > 0$) when K is large, which is weaker than the best-known $d^{1/3}$ for underdamped (second-order). It would be valuable to design an improved interpolation or discretization scheme whose dimension dependence also improves with K, so that increasing K improves the complexity in both ε and d.

Acknowledgments

This research was supported in part by NSF Grant DMS-2515510.

References

Ahn, K. and Chewi, S. (2021). Efficient constrained sampling via the mirror-langevin algorithm. *Advances in Neural Information Processing Systems*, 34:28405–28418.

Altschuler, J. M. and Chewi, S. (2024). Faster high-accuracy log-concave sampling via algorithmic warm starts. *Journal of the ACM*, 71(3):1–55.

Anari, N., Chewi, S., and Vuong, T.-D. (2024). Fast parallel sampling under isoperimetry. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 161–185. PMLR.

Arnold, A. and Erb, J. (2014). Sharp entropy decay for hypocoercive and non-symmetric Fokker-Planck equations with linear drift. Technical Report ASC Report 29/2014, Institute of Analysis and Scientific Computing, TU Wien.

Arnold, A., Jin, S., and Wöhrer, T. (2020). Sharp decay estimates in local sensitivity analysis for evolution equations with uncertainties: from odes to linear kinetic equations. *Journal of Differential Equations*, 268(3):1156–1204.

Bakry, D., Gentil, I., and Ledoux, M. (2013). *Analysis and geometry of Markov diffusion operators*, volume 348. Springer Science & Business Media.

Chen, Y., Dwivedi, R., Wainwright, M. J., and Yu, B. (2020). Fast mixing of metropolized hamiltonian monte carlo: Benefits of multi-step gradients. *Journal of Machine Learning Research*, 21(92):1–72.

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

Cheng, X., Chatterji, N. S., Bartlett, P. L., and Jordan, M. I. (2018b). Underdamped Langevin MCMC: A non-asymptotic analysis. In *Conference on learning theory*, pages 300–323. PMLR.

Chewi, S., Erdogdu, M. A., Li, M., Shen, R., and Zhang, M. S. (2024). Analysis of Langevin Monte Carlo from poincare to log-sobolev. *Foundations of Computational Mathematics*, pages 1–51.

Chewi, S., Le Gouic, T., Lu, C., Maunu, T., Rigollet, P., and Stromme, A. (2020). Exponential ergodicity of mirror-langevin diffusions. *Advances in Neural Information Processing Systems*, 33:19573–19585.

Constantine, G. and Savits, T. (1996). A multivariate Faa di Bruno formula with applications. *Transactions of the American Mathematical Society*, 348(2):503–520.

Dalalyan, A. S. (2017). 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.

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

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

Dang, T., Gurbuzbalaban, M., Islam, M. R., Yao, N., and Zhu, L. (2025). High-order Langevin Monte Carlo algorithms. *arXiv preprint arXiv:2508.17545*.

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

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

Dwivedi, R., Chen, Y., Wainwright, M. J., and Yu, B. (2019). Log-concave sampling: Metropolis-hastings algorithms are fast. *Journal of Machine Learning Research*, 20(183):1–42.

Fan, J., Yuan, B., and Chen, Y. (2023). Improved dimension dependence of a proximal algorithm for sampling. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 1473–1521. PMLR.

He, Y., Balasubramanian, K., and Erdogdu, M. A. (2020). On the ergodicity, bias and asymptotic normality of randomized midpoint sampling method. *Advances in Neural Information Processing Systems*, 33:7366–7376.

Huang, X., Dong, H., Hao, Y., Ma, Y.-A., and Zhang, T. (2023). Reverse diffusion monte carlo. *arXiv preprint arXiv:2307.02037*.

Huang, X., Zou, D., Dong, H., Ma, Y.-A., and Zhang, T. (2024). Faster sampling without isoperimetry via diffusion-based monte carlo. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 2438–2493. PMLR.

Jiang, Q. (2021). Mirror langevin monte carlo: the case under isoperimetry. *Advances in Neural Information Processing Systems*, 34:715–725.

Lee, Y. T., Song, Z., and Vempala, S. S. (2018). Algorithmic theory of odes and sampling from well-conditioned logconcave densities. *arXiv preprint arXiv:1812.06243*.

Li, G., Wei, Y., Chen, Y., and Chi, Y. (2024). Towards non-asymptotic convergence for diffusion-based generative models. The Twelfth International Conference on Learning Representations.

Li, G., Zhou, Y., Wei, Y., and Chen, Y. (2025). Faster diffusion models via higher-order approximation. *arXiv preprint arXiv:2506.24042*.

Ma, Y.-A., Chatterji, N. S., Cheng, X., Flammarion, N., Bartlett, P. L., and Jordan, M. I. (2021). Is there an analog of nesterov acceleration for gradient-based MCMC? *Bernoulli*, 27(3):1942–1992.

Ma, Y.-A., Chen, T., and Fox, E. (2015). A complete recipe for stochastic gradient MCMC. *Advances in neural information processing systems*, 28.

Mishkov, R. L. (2000). Generalization of the formula of Faa di Bruno for a composite function with a vector argument. *International Journal of Mathematics and Mathematical Sciences*, 24(7):481–491.

Monmarché, P. (2023). Almost sure contraction for diffusions on $\mathbb{R}^d$. application to generalized Langevin diffusions. *Stochastic Processes and their Applications*, 161:316–349.

Mou, W., Ma, Y.-A., Wainwright, M. J., Bartlett, P. L., and Jordan, M. I. (2021). High-order Langevin diffusion yields an accelerated MCMC algorithm. *Journal of Machine Learning Research*, 22(42):1–41.

Salim, A. and Richtarik, P. (2020). Primal dual interpretation of the proximal stochastic gradient langevin algorithm. *Advances in Neural Information Processing Systems*, 33:3786–3796.

Shen, R. and Lee, Y. T. (2019). The randomized midpoint method for log-concave sampling. *Advances in Neural Information Processing Systems*, 32.

Shi, J., Chen, T., Yuan, R., Yuan, B., and Ao, P. (2012). Relation of a new interpretation of stochastic differential equations to Ito process. *Journal of Statistical Physics*, 148(3):579–590.

Stoer, J., Bulirsch, R., Bartels, R., Gautschi, W., and Witzgall, C. (1980). *Introduction to Numerical Analysis*, volume 1993. Springer.

Vempala, S. and Wibisono, A. (2019). Rapid convergence of the unadjusted Langevin algorithm: Isoperimetry suffices. *Advances in neural information processing systems*, 32.

Villani, C. (2009). *Hypocoercivity*, volume 202. American Mathematical Society.

Wibisono, A. (2019). Proximal langevin algorithm: Rapid convergence under isoperimetry. *arXiv preprint arXiv:1911.01469*.

Wu, K., Schmidler, S., and Chen, Y. (2022). Minimax mixing time of the metropolis-adjusted langevin algorithm for log-concave sampling. *Journal of Machine Learning Research*, 23(270):1–63.

Wu, Y., Chen, Y., and Wei, Y. (2024). Stochastic runge-kutta methods: Provable acceleration of diffusion models. *arXiv preprint arXiv:2410.04760*.

Yu, L. and Dalalyan, A. (2025). Parallelized midpoint randomization for Langevin Monte Carlo. *Stochastic Processes and their Applications*, page 104764.

Zhang, S., Chewi, S., Li, M., Balasubramanian, K., and Erdogdu, M. A. (2023). Improved discretization analysis for underdamped Langevin Monte Carlo. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 36–71. PMLR.

Organization of the Appendix

Appendix A recalls the notation and settings from the paper. Appendix B provides additional details of Algorithm 1. Appendix C establishes continuous-time convergence to the invariant distribution via a Lyapunov argument. Appendix D contains proofs of the main results (Theorem 1 and Propositions 1, 3, and 5). Appendix E proves Lemma 1 by controlling the interpolation error I_U using high-probability bounds, Faà di Bruno's formula, and higher-order derivative upper bound. Appendix F collects technical details: existence and uniqueness of fixed points for $\mathcal{T}_y$ and $\widehat{\mathcal{T}}_y$, and an equivalent formulation of Algorithm 1. Appendix G gathers auxiliary lemmas.

A Settings

In this section, we recall some of the notations and settings from the paper that are used in the technical proofs presented in the appendix. Recall that in Section 6.1, we defined the following operators:

- **Continuous time operator** $\mathcal{T}_y : \mathcal{C}\big([0,h],\mathbb{R}^{Kd}\big) \to \mathcal{C}\big([0,h],\mathbb{R}^{Kd}\big)$:

$$\big(\mathcal{T}_y[X]\big)(t) := y - \int_0^t (D+Q)\,\nabla H\big(X(s)\big)\,\mathrm{d}s + \int_0^t \sqrt{2D}\,\mathrm{d}B_s, \quad 0 \le t \le h. \tag{10}$$

- **Discretized operator** $\widehat{\mathcal{T}}_y : \mathcal{C}\big([0,h],\mathbb{R}^{Kd}\big) \to \mathcal{C}\big([0,h],\mathbb{R}^{Kd}\big)$, where for given process X, $\widehat{\mathcal{T}}_y[X]$ is the solution $\tilde{X}$ to the following SDE:

$$\tilde{X}(t) = y - \int_0^t (D+Q) \begin{bmatrix} P(s;X) \\ \tilde{X}_2(s) \\ \vdots \\ \tilde{X}_K(s) \end{bmatrix} \mathrm{d}s + \int_0^t \sqrt{2D}\,\mathrm{d}B_s, \quad 0 \le t \le h, \tag{11}$$

  where the Lagrange interpolation is defined by

$$P(s;X) := \sum_{j=1}^{M} \ell_j\left(\tfrac{s}{h}\right)\,\nabla U\big(X_1(c_j h)\big). \tag{12}$$

  Equivalently, defining $J := \mathrm{diag}(0,1,\dots,1) \otimes I_d \in \mathbb{R}^{Kd\times Kd}$, $e_2 = (0,1,0,\dots,0)^\top \in \mathbb{R}^K$ and $A := -(D+Q)\,J$, we have the following decomposition:

$$\big(\widehat{\mathcal{T}}_y[X]\big)(t) = \tilde{X}(t) = y + \int_0^t A\,\tilde{X}(s)\,ds \;-\; \int_0^t (e_2 \otimes I_d)\,P(s;X)\,ds \;+\; \int_0^t \sqrt{2D}\,dB_s. \tag{13}$$

We denote the fixed point of $\mathcal{T}_y$ and $\widehat{\mathcal{T}}_y$ by X_y^* and $\widehat{X}_y^*$ correspondingly.

Picard Iterates. Given y, define the Picard sequence on $[0, h]$ by

$$\widehat{X}_y^{[0]}(t) \equiv y, \quad \widehat{X}_y^{[\nu+1]} := \widehat{\mathcal{T}}_y\big[\widehat{X}_y^{[\nu]}\big], \ \nu \geq 0.$$

Starting from $y = \widehat{X}(nh)$, Algorithm 1 outputs the next state by evaluating the ν_*-th Picard iterate at the end of the step:

$$\widehat{X}((n+1)h) = \widehat{X}_y^{[\nu_*]}(h).$$

Additional Notations. For the Lagrange basis $\{\ell_j\}_{j=1}^M$ defined on iterval $[0, 1]$, we define the standard Lebesgue constant

$$\Gamma_\phi \ := \ \sup_{\tau \in [0,1]} \sum_{j=1}^{M} \big|\ell_j(\tau)\big|.$$

Unless otherwise indicated, we write $a \lesssim b$ to mean that there exists a constant $C > 0$, independent of n, d, N, and h, such that $a \leq C\, b$. Here, n is the step index, N the total number of steps in Algorithm 1, h the step size, and d the dimension of the target distribution $\pi \propto e^{-U(x)}$. For a vector $x \in \mathbb{R}^d$, we use $\|x\|$ and $\|x\|_2$ to denote its Euclidean norm. For a matrix $A \in \mathbb{R}^{M \times N}$, we use $\|A\|$ to denote the operator norm induced by the Euclidean norm, i.e. $\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$. For any positive semidefinite matrix S and vector $x \in \mathbb{R}^d$, we define the S-norm as follows: $\|x\|_S := \sqrt{x^\top S x}$. For a matrix A, we denote by $\lambda_{\min}(A)$, $\lambda_{\max}(A)$, and $\lambda(A)$ its smallest eigenvalue, largest eigenvalue, and the set of all its eigenvalues, respectively. The symbols $\preceq$ and $\succeq$ denote the Loewner order between matrices, and $\prec$ and $\succ$ denote the corresponding strict inequalities. For two probability distributions p and q, we denote by $\Gamma(p, q)$ the set of all their couplings, i.e., the set of joint distributions on the product space whose marginals are p and q, respectively. For a complex number $x \in \mathbb{C}$, $\mathrm{Re}(x)$ denotes its real part.

B Additional Details of Algorithm 1

Recall that the K-th order Langevin dynamics on $\mathbb{R}^{Kd}$ is defined as followed:

$$dX_t = -\big(D + Q\big) \begin{pmatrix} \nabla U\big(X_{1,t}\big) \\ X_{2,t} \\ \vdots \\ X_{K,t} \end{pmatrix} dt \ + \ \sqrt{2\,D}\ dB_t, \tag{14}$$

where $X_t = \left(X_{1,t}^\top, X_{2,t}^\top, \ldots, X_{K,t}^\top\right)^\top \in \mathbb{R}^{Kd}$, $B_t \in \mathbb{R}^{Kd}$ is a standard Brownian motion, and

$$D = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \gamma \end{pmatrix}_{K\times K} \otimes I_d, \quad Q = \begin{pmatrix} 0 & -1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & -\gamma & \ddots & \vdots & \vdots \\ 0 & \gamma & 0 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & -\gamma & 0 \\ 0 & \vdots & 0 & \gamma & 0 & -\gamma \\ 0 & \vdots & 0 & \cdots & \gamma & 0 \end{pmatrix}_{K\times K} \otimes I_d. \tag{15}$$

Here D is symmetric positive semidefinite and Q is skew-symmetric. Define $J := \mathrm{diag}(0, 1, \ldots, 1) \otimes I_d \in \mathbb{R}^{Kd\times Kd}$ and $e_2 = (0, 1, 0, \ldots, 0)^\top \in \mathbb{R}^K$.With this notation, we can decompose the drift into linear and nonlinear parts:

$$A \;:=\; -\,(D+Q)\,J, \quad g(X) \;:=\; -(e_2 \otimes I_d)\,\nabla U(X_1). \tag{16}$$

So the higher-order Langevin dynamic in (14) can be written as

$$dX_t \;=\; A\,X_t\,dt \;+\; g(X_t)\,dt \;+\; \sqrt{2\,D}\,dB_t. \tag{17}$$

Fix a step size $h > 0$ and grid points $0 = t_0 < t_1 < \cdots < t_N = T$ with uniform mesh $t_i - t_{i-1} = h$. For $n = 0, \ldots, N-1$ and $\tau \in [0, 1]$, conditioning on $X_n := X(t_n)$, the variation-of-constants formula for (17) yields

$$X(nh + \tau h) = e^{\tau h A} X(nh) + h \int_0^\tau e^{(\tau - s)hA} g\big(X(nh + sh)\big)\, ds + W_n(\tau), \tag{18}$$

where $W_n(\tau) = \int_0^{\tau h} e^{(\tau h - s)A}\,\sqrt{2D}\,dB_{nh+s} \sim \mathcal{N}(0, \Sigma(\tau))$ with $\Sigma(\tau) := 2\int_0^{\tau h} e^{(\tau h - s)A}\sqrt{DD^\top}e^{(\tau h - s)A^\top}\,ds$. We now use the Lagrange polynomial to approximate the nonlinear term g. Choose equispaced collocation nodes $0 = c_1 < \cdots < c_M = 1$ (i.e. $c_j = \frac{j-1}{M-1}$) on $[0, 1]$, and let $\{\ell_j\}_{j=1}^M$ be their Lagrange polynomials basis on $[0, 1]$:

$$\ell_j(\sigma) \;=\; \prod_{\substack{k=1 \\ k\neq j}}^{M} \frac{\sigma - c_k}{c_j - c_k}, \qquad j = 1, \ldots, M. \tag{19}$$

For $n \geq 0$, let $\widehat{X}(nh) \in \mathbb{R}^{Kd}$ denote the numerical approximation to the exact continuous process $X(nh)$ at step n. For collocation nodes c_k, let $\widehat{X}_{n,k} \in \mathbb{R}^{Kd}$ denote the approximations to $X(nh + c_k h)$, $k = 1, \ldots, M$. Note that at $k = M$, $\widehat{X}_{n,M}$ is an approximation of $X((n+1)h)$.

For $\sigma \in [0, 1]$, suppose we have some approximation of the points $\widehat{X}_{n,j} \approx X(nh + c_j h)$ for all

$j = 1, \dots, M$, then we can approximate g by

$$g\big(X(nh + \sigma h)\big) \;\approx\; \sum_{j=1}^{M} \ell_j(\sigma)\, g\big(\widehat{X}_{n,j}\big). \tag{20}$$

Then, we can approximate the first integral in (18) by

$$h \int_0^\tau e^{(\tau-\sigma)hA}\, g\big(X(nh + \sigma h)\big)\, d\sigma \approx h \sum_{j=1}^{M} \int_0^\tau e^{(\tau-\sigma)hA}\, \ell_j(\sigma)\, d\sigma\, g(\widehat{X}_{n,j}) := h \sum_{j=1}^{M} \alpha_j(\tau, h) g(\widehat{X}_{n,j}), \tag{21}$$

where $\alpha_j(\tau, h) := \int_0^\tau e^{(\tau-\sigma)hA}\, \ell_j(\sigma)\, d\sigma$. Plug in the approximation (21) to the SDE in (18) and set $\tau = c_k$, we get the equation for unknown points $\widehat{X}_{n,k} \approx X(nh + c_k h)$:

$$\widehat{X}_{n,k} = e^{c_k hA} \widehat{X}(nh) \;+\; h \sum_{j=1}^{M} \alpha_j(c_k, h)\, g(\widehat{X}_{n,j}) \;+\; W_n(c_k). \tag{22}$$

Notice that the coefficient matrix $(\alpha_j(c_k, h))_{jk}$ can be precomputed given h and matrix A, then $(\widehat{X}_{n,1}, \dots, \widehat{X}_{n,M})$ forms nonlinear fixed-point equations, and we can solve $(\widehat{X}_{n,1}, \dots, \widehat{X}_{n,M})$ using the Picard iterations: we first initialize $\widehat{X}_{n,k}^{(0)} := \widehat{X}(nh)$ for $k = 1, \dots, M$, then iterate for $\nu = 1, \dots, \nu_*$:

$$\widehat{X}_{n,k}^{(\nu+1)} := e^{c_k hA} \widehat{X}_n + h \sum_{j=1}^{M} \alpha_j(c_k, h)\, g\big(\widehat{X}_{n,j}^{(\nu)}\big) + W_n(c_k).$$

Since we set $c_M = 1$, we have $\widehat{X}((n+1)h) = \widehat{X}_{n,M}^{(\nu_*)}$ being the update for next step. The joint Gaussian vector

$$\big[W_n(c_1)^\top, \dots, W_n(c_M)^\top\big]^\top \sim \mathcal{N}(0, \Sigma_C) \tag{23}$$

has block covariance

$$\mathrm{Cov}\big(W_n(c_i), W_n(c_j)\big) = 2 \int_0^{\min(c_i, c_j)h} e^{(c_i h - s)A} \sqrt{DD^\top}\, e^{(c_j h - s)A^\top}\, ds, \tag{24}$$

and $(W_n(\cdot))_n$ are independent across n.

C Continuous Time Convergence

It can be shown (Shi et al., 2012; Ma et al., 2015) that the higher-order Langevin process X_t as defined in (4) admits an invariant distribution with density

$$p^*(x) \propto \exp\left(-H(x)\right), \quad H(x) = U(x_1) + \frac{1}{2}\sum_{i=2}^{K} \|x_i\|^2.$$

In this section, we prove that the law of $X(t)$ converges to p^* at an exponential rate in the Wasserstein-2 distance. To establish convergence, we employ a Lyapunov function. Let two processes follow the dynamics (4), started from the initial distributions p_0 and p^*, respectively. For a symmetric positive definite matrix $S \succ 0$, define

$$\mathcal{L}_t := \inf_{\zeta_t \in \Gamma(p_t, p^*)} \mathbb{E}_{(X(t), X^*) \sim \zeta_t} \left[(X(t) - X^*)^\top S (X(t) - X^*) \right].$$

With the Lyapunov function in place, we now formalize the continuous-time behavior. The following theorem establishes that the continuous-time dynamics converge exponentially fast under a suitable Lyapunov matrix S. Although the discretization analysis does not use this main result directly, its guarantee of exponential convergence to equilibrium serves as the main motivation for constructing a discrete-time scheme that preserves the same contraction behavior.

Theorem C.1. *Let $X(0) \sim p_0$ and $X^* \sim p^*$. Let $\{X(t)\}_{t \geq 0}$ follow the Kth-order Langevin process defined in (4) with a sufficiently large damping parameter $\gamma > \gamma_0$, for some $\gamma_0 > 0$ that ensures contractivity , and started from p_0 and p^*, respectively. Then there exist a symmetric positive definite matrix $S \succ 0$ and a constant $C_S > 0$, depending only on m, L, γ, and K, such that for all $t \geq 0$,*

$$\begin{aligned} &\inf_{\zeta_t \in \Gamma(p_t, p^*)} \mathbb{E}_{(X(t), X^*) \sim \zeta_t} \left[(X(t) - X^*)^\top S (X(t) - X^*) \right] \\ \leq & e^{-2C_S t} \inf_{\zeta_0 \in \Gamma(p_0, p^*)} \mathbb{E}_{(X(0), X^*) \sim \zeta_0} \left[(X(0) - X^*)^\top S (X(0) - X^*) \right]. \end{aligned}$$

Moreover, letting $\lambda_{\min}(S)$ and $\lambda_{\max}(S)$ denote the minimal and maximal eigenvalues of S, inequality above implies exponential convergence in the Wasserstein-2 distance:

$$\begin{aligned} W_2^2(p_t, p^*) \; &\leq \frac{1}{\lambda_{\min}(S)} \inf_{\zeta_t \in \Gamma(p_t, p^*)} \mathbb{E}_{(X(t), X^*) \sim \zeta_t} \Big[(X(t) - X^*)^\top S (X(t) - X^*) \Big] \\ &\leq \frac{\lambda_{\max}(S)}{\lambda_{\min}(S)} \, e^{-2C_S t} \, W_2^2(p_0, p^*). \end{aligned}$$

Proof. Let $\tilde{\zeta}$ denote the coupling constructed in Proposition C.4. Define

$$\hat{\zeta}_t(X(t), X^*(t)) = \mathbb{E}_{(X(0), X^*(0)) \sim \zeta_0^*} \big[\tilde{\zeta}(X(t), X^*(t) \mid X(0), X^*(0)) \big].$$

Since $\hat{\zeta}_t$ is a valid coupling between p_t and p_t^*, Proposition C.4 together with Gronwall's inequality

yields

$$\begin{aligned}
&\inf_{\zeta_t \in \Gamma(p_t, p_t^*)} \mathbb{E}_{(X(t),X^*(t))\sim\zeta_t}\left[(X(t)-X^*(t))^\top S(X(t)-X^*(t))\right] \\
&\qquad \le \mathbb{E}_{(X(t),X^*(t))\sim\hat{\zeta}_t}\left[(X(t)-X^*(t))^\top S(X(t)-X^*(t))\right] \\
&\qquad = \mathbb{E}_{(X(0),X^*(0))\sim\zeta_0^*}\left[\mathbb{E}_{(X(t),X^*(t))\sim\tilde{\zeta}(\,\cdot\,|X(0),X^*(0))}\left[(X(t)-X^*(t))^\top S(X(t)-X^*(t))\right]\right] \\
&\qquad \le e^{-2C_S t}\,\mathbb{E}_{(X(0),X^*(0))\sim\zeta_0^*}\left[(X(0)-X^*(0))^\top S(X(0)-X^*(0))\right] \\
&\qquad = e^{-2C_S t} \inf_{\zeta_0 \in \Gamma(p_0, p^*)} \mathbb{E}_{(X(0),X^*)\sim\zeta_0}\left[(X(0)-X^*)^\top S(X(0)-X^*)\right].
\end{aligned}$$

This proves Theorem C.1.

□

C.1 Proof of Proposition 2

To prove this proposition, we follow the approach developed in Monmarché (2023); Arnold and Erb (2014); Arnold et al. (2020). Our first objective is to identify a matrix S that yields the desired contraction. For this purpose only, we first introduce a convenient reparametrization of the dynamics.

Write the state as

$$X(t) = \big(X_1(t),\, Y(t)\big) \in \mathbb{R}^d \times \mathbb{R}^{(K-1)d}.$$

Then the continuous-time process (4) can be rewritten as

$$\mathrm{d}X(t) = P\,Y(t)\,\mathrm{d}t, \qquad \mathrm{d}Y(t) = -P^\top \nabla U(X(t))\,\mathrm{d}t - \gamma\,\tilde{Q}\,Y(t)\,\mathrm{d}t + \sqrt{2\gamma\,\tilde{D}}\,\mathrm{d}B_t, \tag{25}$$

where B_t is a standard $(K-1)d$-dimensional Brownian motion and the matrices $P \in \mathbb{R}^{d\times(K-1)d}$, $\tilde{Q}, \tilde{D} \in \mathbb{R}^{(K-1)d\times(K-1)d}$ are defined by

$$P = \begin{pmatrix} I_d & 0 & \cdots & 0 \end{pmatrix}, \quad \tilde{Q} = \begin{pmatrix} 0 & -I_d & 0 & \cdots & 0 \\ I_d & 0 & -I_d & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & I_d & 0 & -I_d \\ 0 & \cdots & 0 & I_d & I_d \end{pmatrix}, \quad \tilde{D} = \begin{pmatrix} 0_d & & & & 0 \\ & \ddots & & & \vdots \\ & & 0_d & & 0 \\ & & & 0_d & 0 \\ 0 & \cdots & 0 & 0 & I_d \end{pmatrix}.$$

Let $b(x_1, y)$ denote the drift of (25), then for $x_1 \in \mathbb{R}^d$, $y \in \mathbb{R}^{(K-1)d}$, (and correspondingly for $x = (x_1, y) \in \mathbb{R}^{Kd}$),

$$b(x_1, y) = b(x) = -(D+Q)\,\nabla H(x). \tag{26}$$

We write J_b for the Jacobian of b; explicitly,

$$J_b(x) = -(D+Q)\begin{pmatrix} \nabla^2 U(x_1) & & & & 0 \\ & \ddots & & & \vdots \\ & & I_d & & 0 \\ & & & I_d & 0 \\ 0 & \cdots & 0 & 0 & I_d \end{pmatrix}. \tag{27}$$

To this end, we state the following set of conditions, adapted from prior work (see Monmarché (2023)), and later show that our dynamics satisfy them.

Condition 1. There exist constants $m, L > 0$ such that

$$\forall\, x \in \mathbb{R}^d, \qquad m\, I_d \;\preceq\; \nabla^2 U(x) \;\preceq\; L\, I_d.$$

Condition 2. There exist $\kappa > 0$ and a symmetric positive–definite matrix $\tilde{N} \in \mathbb{R}^{(K-1)d \times (K-1)d}$ such that

$$\tilde{N}\tilde{Q} + \tilde{Q}^\top \tilde{N} \;\succeq\; 2\kappa\, \tilde{N}.$$

Condition 3. Decompose $\tilde{Q}$ as

$$\tilde{Q} \;=\; \begin{pmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} \\ \tilde{Q}_{21} & \tilde{Q}_{22} \end{pmatrix},$$

where the blocks have sizes

$$\tilde{Q}_{11} \in \mathbb{R}^{d\times d}, \quad \tilde{Q}_{12} \in \mathbb{R}^{d\times (K-2)d}, \quad \tilde{Q}_{21} \in \mathbb{R}^{(K-2)d\times d}, \quad \tilde{Q}_{22} \in \mathbb{R}^{(K-2)d\times (K-2)d}.$$

Require that $\tilde{Q}_{22}$ be invertible and that

$$E \;:=\; \tilde{Q}_{11} - \tilde{Q}_{12}\tilde{Q}_{22}^{-1}\tilde{Q}_{21}$$

is symmetric positive definite. Moreover, set $H := \tilde{Q}_{12}\tilde{Q}_{22}^{-1}$.

The following lemma ensures that the higher-order Langevin dynamics (25) satisfies all the above conditions.

Lemma C.2. *The dynamics* (25) *satisfy Conditions 1–3.*

Proof. Condition 1 follows directly from the smoothness and strong convexity assumptions on U (Assumption 1).

For Condition 2, write $\widetilde{Q} = \widetilde{Q}_{\text{can}} \otimes I_d$ with the $(K-1) \times (K-1)$ "canonical" backbone matrix

$$\widetilde{Q}_{\text{can}} = \begin{pmatrix} 0 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 1 & 0 & -1 \\ 0 & \cdots & 0 & 1 & 1 \end{pmatrix}.$$

Hence $\lambda(\widetilde{Q}) = \lambda(\widetilde{Q}_{\text{can}})$ (with $\lambda(\cdot)$ denoting the spectrum, with multiplicities scaled by d), and eigenvectors of $\widetilde{Q}$ are $w = v \otimes e_j$, where v is an eigenvector of $\widetilde{Q}_{\text{can}}$ and $\{e_j\}_{j=1}^d$ is the canonical basis of $\mathbb{R}^d$.

Given that $\widetilde{Q}$ and $\widetilde{Q}_{\text{can}}$ share the same spectrum, we now show that

$$\min\{\text{Re}(\lambda) : \lambda \in \lambda(\widetilde{Q}_{\text{can}})\} > 0.$$

Decompose

$$\widetilde{Q}_{\text{can}} = \frac{\widetilde{Q}_{\text{can}} + \widetilde{Q}_{\text{can}}^\top}{2} + \frac{\widetilde{Q}_{\text{can}} - \widetilde{Q}_{\text{can}}^\top}{2} =: L + M,$$

where L is Hermitian and M is skew–Hermitian. If $\widetilde{Q}_{\text{can}} x = \lambda x$ for some $x = (x_1, \ldots, x_{K-1}) \in \mathbb{C}^{K-1} \setminus \{0\}$, then by the standard Rayleigh-quotient identity for the symmetric part,

$$\text{Re}(\lambda) = \frac{x^* L x}{x^* x} = \frac{|x_{K-1}|^2}{|x|^2}.$$

We claim $x_{K-1} \neq 0$. Indeed, if $x_{K-1} = 0$, then the last row of $\widetilde{Q}_{\text{can}} x = \lambda x$ gives $x_{K-2} = 0$; iterating backwards yields $x_{K-3} = 0, \ldots, x_1 = 0$, a contradiction to $x \neq 0$. Hence $x_{K-1} \neq 0$ and thus $\text{Re}(\lambda) > 0$ for every eigenvalue λ of $\widetilde{Q}_{\text{can}}$. The same conclusion then holds for $\widetilde{Q} = \widetilde{Q}_{\text{can}} \otimes I_d$.

Given the above result that $\min\{\text{Re}(\lambda) : \lambda \in \lambda(\widetilde{Q})\} > 0$, the works Arnold and Erb (2014, Lemma 4.3), Arnold et al. (2020, Section 2.1), and Dang et al. (2025, Appendix A) provide an explicit construction of a symmetric positive–definite matrix $\tilde{N}$ such that

$$\tilde{N}\widetilde{Q} + \widetilde{Q}^\top \tilde{N} \succeq 2\kappa \tilde{N}.$$

We refer the reader to these references for the complete construction and reproduce the key details in Appendix C.2 for completeness.

With $H := -(I_d, \ldots, I_d) \in \mathbb{R}^{d \times (K-2)d}$ we verify directly that $H\,\widetilde{Q}_{22} = \widetilde{Q}_{12}$. It follows that $E := \widetilde{Q}_{11} - \widetilde{Q}_{12}\widetilde{Q}_{22}^{-1}\widetilde{Q}_{21} = \widetilde{Q}_{11} - H\,\widetilde{Q}_{21} = I_d$, which is symmetric positive definite. this proves Condition 3. □

Since $E = I_d \succ 0$ and $\tilde{N} \succ 0$, we can fix constants $h_i > 0$ $(i = 1, \dots, 5)$ such that

$$\tilde{N} P^\top P \tilde{N} \preceq h_1 \tilde{N}, \qquad \frac{1}{h_2} I_d \preceq E \preceq h_3 I_d, \tag{28}$$

and

$$\begin{pmatrix} I_d & -H \\ 0 & 0 \end{pmatrix} \preceq h_4 \tilde{N}, \qquad \begin{pmatrix} I_d & -H \\ -H^\top & 0 \end{pmatrix} \preceq h_5 \tilde{N}. \tag{29}$$

Note that, here we can take $h_2 = h_3 = 1$.

The following theorem is taken from (Monmarché, 2023, Theorem 9); see that reference for the proof.

Theorem C.3. *Define*

$$\gamma_0 := 2\sqrt{\frac{h_1 L}{\kappa}} \max\left(\sqrt{h_2 h_5}, \sqrt{\frac{h_4}{\kappa}} \right). \tag{30}$$

If $\gamma \geq \gamma_0$, then under Conditions 1–3 there exist a positive definite matrix S such that the drift of the process (25) satisfies

$$\forall\, x \in \mathbb{R}^{Kd}, \qquad S\, J_b(x) + J_b^\top(x) S \preceq -2\, C_s\, S,$$

with

$$C_s = \min\left(\frac{m}{3 h_3 \gamma}, \frac{\gamma \kappa}{6} \right).$$

Moreover, S is a positive definite matrix such that

$$\frac{1}{2} \begin{pmatrix} I_d & 0 \\ 0 & \frac{\kappa}{L h_1} \tilde{N} \end{pmatrix} \preceq S \preceq \frac{3}{2} \begin{pmatrix} I_d & 0 \\ 0 & \frac{\kappa}{L h_1} \tilde{N} \end{pmatrix}.$$

Remark C.1. Based on the construction of $\tilde{N}$ provided in Appendix C.2, both $\|\tilde{N}\|_{\mathrm{op}}$ and $\|\tilde{N}^{-1}\|_{\mathrm{op}}$ are bounded by constants that are independent of the ambient dimension d. As a result of Lemma C.5, all constants appearing in inequality (28), as well as the eigenvalues of the associated matrix S, are dimension-free. In particular, the parameters h_i and κ, and consequently the constants C_S, $\lambda_{\min}(S)$, and $\lambda_{\max}(S)$, remain independent of d.

With all these tools in hand, we restate Proposition 2 below for completeness before presenting its proof.

Proposition C.4. *Let the processes $\{X(t)\}$ and $\{X^*(t)\}$ follow the Kth-order Langevin process in (4) with $\gamma \geq \gamma_0$ for some γ_0 as defined in (30). Let the initial conditions be $X(0)$ and $X^*(0) \in \mathbb{R}^{Kd}$. Then there exists a coupling*

$$\bar{\zeta} \in \Gamma\big(p_t(X(t) \mid X(0)),\, p_t^*(X^*(t) \mid X^*(0))\big)$$

of the laws of $X(t)$ and $X^*(t)$, a symmetric positive definite matrix $S \succ 0$, and a constant $C_S > 0$, depending only on m, L, and K, such that

$$\frac{\mathrm{d}}{\mathrm{d}t}\,(X(t)-X^*(t))^\top S(X(t)-X^*(t)) \;\le\; -2C_S\,(X(t)-X^*(t))^\top S(X(t)-X^*(t)), \quad \textit{for all } (X(t),X^*(t)) \sim \bar{\zeta}. \tag{31}$$

Proof. Consider the processes $X(t)$ and $X^*(t)$ in the statement. Under synchronous coupling, applying (27) yields

$$\begin{aligned} d(X(t)-X^*(t)) &= -(D+Q)\big(\nabla H(X(t)) - \nabla H(X^*(t))\big)\,dt \\ &= \big(b(X(t)) - b(X^*(t))\big)\,dt \\ &= \Bigg[\int_0^1 \underbrace{J_b\big(X^*(t)+\lambda(X(t)-X^*(t))\big)}_{=:\widetilde{J}_b(t,\lambda)}\,d\lambda\Bigg](X(t)-X^*(t))\,dt, \end{aligned}$$

where $b(\cdot)$ is the drift and J_b its Jacobian (27). Since U is m-strongly convex and L-smooth on $\mathbb{R}^d$, the (vector-valued) mean-value theorem (fundamental theorem of calculus) on open convex sets yields the last equality.

Consequently, using Theorem C.3 we obtain

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\,(X(t)-X^*(t))^\top S(X(t)-X^*(t)) &= (X(t)-X^*(t))^\top\left(S\int_0^1 \widetilde{J}_b(t,\lambda)\,d\lambda + \int_0^1 \widetilde{J}_b(t,\lambda)^\top\,d\lambda\,S\right)(X(t)-X^*(t)) \\ &= \int_0^1 (X(t)-X^*(t))^\top\big(S\widetilde{J}_b(t,\lambda) + \widetilde{J}_b(t,\lambda)^\top S\big)(X(t)-X^*(t))\,d\lambda \\ &\le \int_0^1 -2C_S\,(X(t)-X^*(t))^\top S(X(t)-X^*(t))\,d\lambda \\ &= -2C_S\,(X(t)-X^*(t))^\top S(X(t)-X^*(t)). \end{aligned}$$

This finishes the proof of Proposition C.4. □

C.2 Construction of Matrix $\tilde{N}$

We briefly outline here the construction of $\widetilde{N}$, which satisfies Condition 2. The full details can be found in Arnold and Erb (2014, Lemma 4.3), Arnold et al. (2020, Section 2.1), and Dang et al. (2025, Appendix A).

We begin by observing that $\widetilde{Q}$ can be written in the Kronecker product form

$$\widetilde{Q} \;=\; \widetilde{Q}_{\text{can}} \otimes I_d, \qquad \widetilde{Q}_{\text{can}} \;:=\; \begin{pmatrix} 0 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 1 & 0 & -1 \\ 0 & \cdots & 0 & 1 & 1 \end{pmatrix}.$$

This representation implies that $\widetilde{Q}$ and the $(K-1)\times(K-1)$ matrix $\widetilde{Q}_{\text{can}}$ share the same spectrum. Let $\mu = \min\{\mathrm{Re}(\lambda) : \lambda \in \lambda(\widetilde{Q})\} > 0$. Furthermore, if v_i $(1 \le i \le K-1)$ are the eigenvectors (or generalized eigenvectors) of $\widetilde{Q}_{\text{can}}$ and e_j $(1 \le j \le d)$ are the standard basis vectors in $\mathbb{R}^d$, then

$$w_{ij} \;=\; v_i \otimes e_j, \qquad 1 \le i \le K-1,\ 1 \le j \le d,$$

form the eigenvectors (respectively generalized eigenvectors) of $\widetilde{Q}$.

Case 1: $\widetilde{Q}$ is non-defective [1] (diagonalizable).

We first consider the simpler case where $\widetilde{Q}$ is non-defective (diagonalizable), in which case we can define

$$\tilde{N} \;=\; \sum_k w_k w_k^\dagger, \qquad k \in \{(i,j) : 1 \le i \le K-1,\ 1 \le j \le d\},$$

where † denotes the conjugate transpose. Since $\{w_k\}$ forms a basis of $\mathbb{C}^{d(K-1)}$, the matrix $\tilde{N}$ is symmetric and positive definite. If any w_k is complex, its conjugate $\overline{w}_k$ is also an eigenvector corresponding to the conjugate eigenvalue $\overline{\lambda}_k$. Noting that each λ_k is an eigenvalue of $\widetilde{Q}$, we obtain

$$\widetilde{Q}\tilde{N} + \tilde{N}\widetilde{Q}^\top \;=\; \sum_k (\lambda_k + \overline{\lambda}_k)\, w_k w_k^\dagger \;\succeq\; 2\,\mu\,\tilde{N}.$$

Hence, the desired inequality (Condition 2) holds in the non-defective case.

Case 2: $\widetilde{Q}$ is defective, with a trivial Jordan block corresponding to the eigenvalue of minimal real part.

Let us now move on to the defective case. If at least one eigenvalue of $\widetilde{Q}$ is defective, we include the generalized eigenvectors and proceed as follows. Let $A^{-1}\widetilde{Q}A = J = \mathrm{diag}(J_1, \ldots, J_N)$ be the Jordan normal form, where each J_n is of length ℓ_n with eigenvalue λ_n. For now, let us assume that all Jordan blocks corresponding to eigenvalues with $\mathrm{Re}(\lambda_n) = \mu$ are trivial, i.e., $\ell_n = 1$.

Define a block-diagonal positive matrix

$$B := \mathrm{diag}(B_1, \ldots, B_N), \qquad B_n := \mathrm{diag}(b_n^{\ell_n}, \ldots, b_n^1),$$

[1] An eigenvalue is *defective* if its geometric multiplicity is strictly less than its algebraic multiplicity. Equivalently, a matrix is *non-defective* if it is diagonalizable over $\mathbb{C}$.

whose entries are given by

$$b_n^1 = 1, \qquad b_n^j = c_j\,(\tau_n)^{2(1-j)}, \quad j = 2, \ldots, \ell_n, \tag{32}$$

where $c_1 = 1$, $c_j = 1 + (c_{j-1})^2$, and $\tau_n := 2(\mathrm{Re}(\lambda_n) - \mu) \geq 0$. For $\ell_n = 1$, $B_n = 1$ and

$$J_n B_n + B_n J_n^\dagger = (\lambda_n + \overline{\lambda}_n) B_n \succeq 2\mu B_n.$$

When $\ell_n > 1$, one checks via the principal-minor test and recursion that

$$J_n B_n + B_n J_n^\dagger - 2\mu B_n \succeq 0,$$

so that $J_n B_n + B_n J_n^\dagger \succeq 2\mu B_n$ for all n. Therefore, in total we have,

$$JB + BJ^\dagger \succeq 2\mu B.$$

Multiplying by A and $A^\dagger$ on both sides yields

$$A^{-1}\widetilde{Q}A\,B + B\,A^\dagger \widetilde{Q}^T (A^{-1})^\dagger \succeq 2\mu B,$$

which implies

$$\widetilde{Q}ABA^\dagger + ABA^\dagger \widetilde{Q}^T \succeq 2\mu ABA^\dagger.$$

Finally, setting $\widetilde{N} := ABA^\dagger$ gives

$$\widetilde{Q}\widetilde{N} + \widetilde{N}\widetilde{Q}^\top \succeq 2\mu \widetilde{N}.$$

Case 3: $\widetilde{Q}$ is defective, with a non-trivial Jordan block corresponding to the eigenvalue of minimal real part.

In the case where there exists a nontrivial Jordan block $J_{\tilde{n}}$ corresponding to an eigenvalue with $\mathrm{Re}(\lambda_{\tilde{n}}) = \mu$, we modify the construction (32) by choosing (instead of τ_n),

$$\tau_{\tilde{n}} := \mu > 0.$$

Hence,

$$J_{\tilde{n}} B_{\tilde{n}} + B_{\tilde{n}} J_{\tilde{n}}^\dagger \;\succeq\; \mu B_{\tilde{n}}, \qquad \text{(a slightly weaker inequality)}$$

and Condition 2 follows.

Lemma C.5. *All constants appearing in inequality (28), as well as the eigenvalues associated with the matrix S, are independent of the ambient dimension d.*

Proof. (The case where $\widetilde{Q}$ is non-diagonalizable is treated analogously.) From the Kronecker

representation

$$\widetilde{N} \;=\; \left(\sum_{i=1}^{K-1} v_i v_i^{\dagger}\right) \otimes I_d,$$

it follows that both $\|\widetilde{N}\|_{\mathrm{op}}$ and $\|\widetilde{N}^{-1}\|_{\mathrm{op}}$ are independent of the dimension d.

Next, note that $E = I_d$ and $P = (I_d\ 0\ \cdots\ 0)$, yielding $h_1 = \|\widetilde{N}P^{\top}P\|_{\mathrm{op}}$; hence h_1, h_2, and h_3 are dimension-free. To control h_4 and h_5, consider $H = -(I_d, \ldots, I_d)$. Then one can easily show that

$$(1 + (K-1)/2)\, I_{(K-1)d} - \begin{pmatrix} I_d & -H \\ 0 & 0 \end{pmatrix} \succeq 0.$$

Which implies that choosing $h_4 = (1 + (K-1)/2)\|\widetilde{N}^{-1}\|_{\mathrm{op}}$ is sufficient. A similar argument, based on the decomposition

$$\begin{pmatrix} I_d & -H \\ -H^{\top} & 0 \end{pmatrix} = \begin{pmatrix} I_d & -H \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -H^{\top} & 0 \end{pmatrix},$$

yields $h_5 = (1+K)\|\widetilde{N}^{-1}\|_{\mathrm{op}}$.

Finally, by construction of S,

$$\frac{1}{2}\begin{pmatrix} I_d & 0 \\ 0 & \dfrac{\kappa}{Lh_1}\widetilde{N} \end{pmatrix} \;\preceq\; S \;\preceq\; \frac{3}{2}\begin{pmatrix} I_d & 0 \\ 0 & \dfrac{\kappa}{Lh_1}\widetilde{N} \end{pmatrix}.$$

Since the spectrum of $\widetilde{N}$ does not depend on d, the minimum and maximum eigenvalues of S are likewise dimension-independent. □

D Proof of Main Results

D.1 Proof of Theorem 1

Recall the synchronous coupling bound

$$W_2^2\big(\pi, \hat{\pi}^{(N)}\big) \;\le\; \|S^{-1}\|_{\mathrm{op}}\, \mathbb{E}\big[\|X(Nh) - \widehat{X}(Nh)\|_S^2\big], \tag{33}$$

where π is the stationary law of the continuous dynamics as defined in (4), $\hat{\pi}^{(N)}$ is the law of the output of Algorithm 1 after N steps of size h, and both processes are driven by the same Brownian path on each step. Throughout we assume $X(0) \sim \pi$ and $\widehat{X}(0) = 0$, so the initial continuous process is stationary. We provide the formal version of Theorem 1 below.

Theorem D.1. *Assume that U satisfies Assumption 1-3, fix $K \ge 3$, and run Algorithm 1 with Picard iterations $\nu_* \ge K-1$ at each step and $K-1$ equispaced collocation nodes $\{c_j\}_{j=1}^{K-1}$ (i.e., $M = K-1$ in Algorithm 1). Let $C_S > 0$ be the S-contractivity constant from Proposition C.4, and let C_0^*, C_1^*, C_2^* be the constants from Proposition D.2. There exists a constant $h_0 > 0$ (depending*

only on γ, m, L and K, and the collocation nodes $\{c_j\}$, but independent of d, N, and ε) such that if, for a target accuracy $\varepsilon \in (0,1)$,

$$h \;\le\; \min\left\{ h_0,\; \left[\frac{C_S\,\varepsilon^2}{3\,C_1^*\,d^{\,K-1}\,\log\!\left(\frac{3\,C_0^*\,d}{\varepsilon^2}\right)}\right]^{\frac{1}{2K-3}},\; \left[\frac{\varepsilon^2}{3\,C_2^*\,d}\right]^{\frac{1}{2K-2}} \right\}, \tag{34}$$

and

$$N \;\ge\; \left\lceil \frac{1}{C_S\,h}\,\log\!\left(\frac{3\,C_0^*\,d}{\varepsilon^2}\right)\right\rceil, \tag{35}$$

then the Wasserstein-2 distance after N steps satisfies

$$W_2^2\big(\pi,\hat\pi^{(N)}\big) \;\le\; \|S^{-1}\|_{\mathrm{op}}\;\varepsilon^2. \tag{36}$$

Proof. From Proposition D.2, for $h>0$ small enough so that $\rho := 2Lh\Gamma_\phi \le \frac{1}{2}$, there exist constants $C_0^*, C_1^*, C_2^* > 0$ (independent of h, d, N) such that

$$E_{N+1} := \mathbb{E}\,\|X((N+1)h) - \widehat{X}((N+1)h)\|_S^2 \;\le\; C_0^*\,e^{-C_S(N+1)h}\,d \;+\; C_1^*\,N\,h^{\,2K-2}\,d^{\,K-1} \;+\; C_2^*\,h^{\,2\nu_*}\,d. \tag{37}$$

Using (33) and (37),

$$W_2^2\big(\pi,\hat\pi^{(N+1)}\big) \;\le\; \|S^{-1}\|_{\mathrm{op}}\Big(C_0^*e^{-C_S(N+1)h}\,d + C_1^*Nh^{2K-2}\,d^{K-1} + C_2^*h^{2\nu_*}\,d\Big).$$

We now bound each contribution under the choices (34)–(35).

By (35), $e^{-C_S(N+1)h} \le \exp\!\Big(-\log\frac{3C_0^*d}{\varepsilon^2}\Big) = \frac{\varepsilon^2}{3C_0^*d}$, hence

$$C_0^*e^{-C_S(N+1)h}\,d \;\le\; \frac{\varepsilon^2}{3}.$$

Using (35) we have $N \le \frac{1}{C_Sh}\log\!\big(\frac{3C_0^*d}{\varepsilon^2}\big)$, so

$$C_1^*Nh^{2K-2}d^{K-1} \;\le\; \frac{C_1^*}{C_S}\,\log\!\Big(\frac{3C_0^*d}{\varepsilon^2}\Big)\,h^{2K-3}d^{K-1}.$$

By the second term in the minimum of (34), the right-hand side is $\le \varepsilon^2/3$. By the third term in the minimum of (34), and using $\nu_* \ge K-1$ we have $C_2^*h^{2\nu_*}d \le \varepsilon^2/3$. □

D.2 Proof of Proposition 1

In this section, we focus on the one–step update and establish the corresponding discretization bound. The formal statement of Proposition 1 is given below.

Proposition D.2. *Let $E_n := \mathbb{E}\,\|X(nh) - \widehat{X}(nh)\|_S^2$ and assume the step size $h>0$ is small enough*

so that $h \le \bar{h}$ *with*

$$\bar{h} \;:=\; \min\Big\{(3\gamma(L\vee 1))^{-1},\ \tfrac{\log 2}{3\gamma},\ (4L\Gamma_\phi)^{-1},\ h_1\Big\},$$

where h_1 *is the stepsize upper bound (independent of* d, n*) from Lemma E.9 and* Γ_ϕ *is defined in Lemma F.2. Then there exist constants* $C_a, C_1, C_2 > 0$ *(independent of* h, d, n*) such that, for all* $n \ge 0$*,*

$$E_{n+1} \;\le\; a_h\, E_n \;+\; c_h, \qquad a_h := e^{-2C_S h} \;+\; C_a\, h^{2\nu_*+2}, \qquad c_h := C_1\, N h^{2K-1} d^{K-1} \;+\; C_2\, h^{2\nu_*+1}\, d. \tag{38}$$

Moreover, there exists $h_0 := \min\Big\{\bar{h},\ \frac{1}{4C_S},\ \Big(\frac{C_S}{2C_a}\Big)^{\frac{1}{2\nu_*+1}}\Big\}$ *such that for all* $h \in (0, h_0]$ *and any* $\nu_* \in \mathbb{N}$*,*

$$a_h \;\le\; 1 - C_S\, h,$$

and hence, for all $N \ge 0$*,*

$$E_{N+1} \;\le\; C_0^*\, e^{-C_S(N+1)h}\, d \;+\; C_1^\star\, N\, h^{2K-2}\, d^{K-1} \;+\; C_2^\star\, h^{2\nu_*}\, d, \tag{39}$$

for some constants $C_0^*, C_1^\star, C_2^\star > 0$ *that do not depend on* h, d, N*.*

Proof. **One–step recursion.** From the error decomposition in Section 6.2,

$$\begin{aligned} E_{n+1} &= \mathbb{E}\big\|X((n+1)h) - \widehat{X}((n+1)h)\big\|_S^2 \\ &\le 3\,\underbrace{\mathbb{E}\big\|X^*_{X(nh)}(h) - X^*_{\widehat{X}(nh)}(h)\big\|_S^2}_{\text{(I) Continuous time convergence}} + 3\,\underbrace{\mathbb{E}\big\|X^*_{\widehat{X}(nh)}(h) - \widehat{X}^*_{\widehat{X}(nh)}(h)\big\|_S^2}_{\text{(II) Interpolation Error}} + 3\,\underbrace{\mathbb{E}\big\|\widehat{X}^*_{\widehat{X}(nh)}(h) - \widehat{X}^{[\nu_*]}_{\widehat{X}(nh)}(h)\big\|_S^2}_{\text{(III) Picard Convergence Error}}. \end{aligned}$$

For (I), by Proposition 2 (or see Proposition C.4 for a formal version),

$$\mathbb{E}\big\|X^*_{X(nh)}(h) - X^*_{\widehat{X}(nh)}(h)\big\|_S^2 \le e^{-2C_S\, h}\, E_n.$$

For (II), by Proposition 3 and Lemma 4 (or see Proposition D.3, Lemma E.2) gives

$$\text{(II)} \le e^{-C_S h}\, \frac{9\|S\|_{\mathrm{op}}}{C_S}\, h\, \mathbb{E}\, I_{U,n} \;\le\; C_1\, N h^{2K-1} d^{K-1}.$$

For (III), Proposition 5 (or Proposition D.5) with $\nu = \nu_*$ yields

$$\text{(III)} \le \frac{\rho^{2\nu_*}}{(1-\rho)^2}\, C_\Delta(h), \qquad C_\Delta(h) = 54h^2(L^2+1)e^{3h^2\|A\|^2}\, E_n + \Delta(h),$$

with $\Delta(h) = O(dh)$ from Lemma D.4. Since $\rho = \Theta(h)$, we have $\rho^{2\nu_*} = \Theta(h^{2\nu_*})$; multiplying by $\|S\|_{\mathrm{op}}$ gives

$$\|S\|_{\mathrm{op}}\text{(III)} \;\le\; C_a\, h^{2\nu_*+2}\, E_n \;+\; C_2\, h^{2\nu_*+1}\, d,$$

for suitable $C_a, C_2 > 0$. Combining the three terms proves (38). All bounds above are valid for h

smaller enough such that $h \le \bar{h}$, which ensures the contraction of operators $\mathcal{T}_y$ and $\widehat{\mathcal{T}}_y$ (Lemmas F.1, F.2) and the applicability of Lemma E.2.

Global bound. Use $e^{-2C_S h} \le 1 - 2C_S h + 2C_S^2 h^2$ to get $a_h \le 1 - 2C_S h + 2C_S^2 h^2 + C_a h^{2\nu_*+2}$. For any fixed $\nu_* \ge 1$, with $h_0 := \min\left\{\bar{h}, \frac{1}{4C_S}, \left(\frac{C_S}{2C_a}\right)^{\frac{1}{2\nu_*+1}}\right\}$ we have $2C_S^2 h^2 + C_a h^{2\nu_*+2} \le C_S h$ for $h \le h_0$; hence $a_h \le 1 - C_S h$. Solving the recursion $E_{n+1} \le a_h E_n + c_h$ yields $E_{N+1} \le a_h^{N+1} E_0 + c_h(1 - a_h^{N+1})/(1 - a_h)$. Since $1 - a_h \ge C_S h$ and $a_h^{N+1} \le e^{-C_S(N+1)h}$,

$$E_{N+1} \le e^{-C_S(N+1)h} E_0 + (C_S h)^{-1} c_h.$$

Since we initialize the algorithm with $\hat{X}(0) = 0$ and by Lemma G.4, $\mathbb{E}\|X\|^2 = O(d)$ for $X \sim \rho(dx) \propto e^{-U(x_1) - \frac{1}{2}\sum_{k=2}^K \|x_k\|^2} dx$, we have $E_0 = O(d)$. Finally, insert $c_h = C_1 h^{2K-1} d^{K-1} + C_2 h^{2\nu_*+1} d$ and absorb constants to obtain (39). □

D.3 Proof of Proposition 3

We provide the formal version of Proposition 3 below.

Proposition D.3 (Interpolation error within one step)**.** *Fix a step size $h > 0$ and index n, and set $y := \widehat{X}(nh)$. Let X^* and $\widehat{X}^*$ be the fixed points of $\mathcal{T}_y$ and $\widehat{\mathcal{T}}_y$ on $[0, h]$ (as defined in (10) and (11)), driven by the same Brownian motion (synchronous coupling). Then*

$$\left\|X^*(h) - \widehat{X}^*(h)\right\|_S^2 \le e^{-C_S h} \frac{9\gamma \|S\|_{\mathrm{op}}}{C_S} h\, I_U, \tag{40}$$

where

$$I_U := \sup_{s \in [nh, (n+1)h]} \left\|\nabla U\big(\widehat{X}_1^*(s)\big) - P\big(s, \widehat{X}_1^*(s)\big)\right\|^2.$$

Here the subscript 1 denotes the first block (the x_1–component) of $\widehat{X}^*$.

Proof. Let $\delta X(t) := X^*(t) - \widehat{X}^*(t)$ for $t \in [0, h]$. Under the fixed-point representations and synchronous coupling, δX solves

$$\frac{d}{dt}\delta X(t) = -(D+Q)\begin{bmatrix} \nabla U\big(X_1^*(t)\big) - \nabla U\big(\widehat{X}_1^*(t)\big) \\ \delta X_2(t) \\ \vdots \\ \delta X_K(t) \end{bmatrix} - (D+Q)\begin{bmatrix} r(t) \\ 0_d \\ \vdots \\ 0_d \end{bmatrix}, \tag{41}$$

where

$$r(t) := \nabla U\big(\widehat{X}_1^*(t)\big) - P\big(t, \widehat{X}_1^*\big). \tag{42}$$

By the mean-value theorem, $\nabla U(X_1^*) - \nabla U(\widehat{X}_1^*) = H_t\,(X_1^* - \widehat{X}_1^*)$ with $mI_d \preceq H_t \preceq LI_d$. Define

$J_b := -(D+Q)\,\mathrm{diag}(H_t, I_d, \ldots, I_d)$ and $R(t) = (r(t), 0_d, \ldots, 0_d)^T$. Then

$$\frac{d}{dt}\delta X(t) = J_b\,\delta X(t) \;-\; (D+Q)\,R(t). \tag{43}$$

Define $E(t) := \|\delta X(t)\|_S^2$. Differentiating and using the contraction $SJ_b + J_b^T S \preceq -2C_S S$ in Theorem C.3,

$$\begin{aligned}
\frac{d}{dt}E(t) &= \Big\langle \delta X(t), S\,\frac{d}{dt}\delta X(t)\Big\rangle + \Big\langle \frac{d}{dt}\delta X(t), S\,\delta X(t)\Big\rangle \\
&= \Big\langle \delta X(t),\, S\Big(J_b\,\delta X(t) - (D+Q)R(t)\Big)\Big\rangle + \Big\langle J_b\,\delta X(t) - (D+Q)R(t),\, S\,\delta X(t)\Big\rangle \\
&= \Big\langle \delta X(t),\, \big(SJ_b + J_b^\top S\big)\,\delta X(t)\Big\rangle \;-\; 2\Big\langle \delta X(t),\, S(D+Q)R(t)\Big\rangle \\
&\le -2C_S\,\|\delta X(t)\|_S^2 \;+\; 2\,\|\delta X(t)\|_S\,\Big\|S^{1/2}(D+Q)R(t)\Big\|.
\end{aligned}$$

Apply Young's inequality with any $\eta > 0$:

$$\|\delta X(t)\|_S\,\Big\|S^{1/2}(D+Q)R(t)\Big\| \;\le\; \frac{\eta}{2}\|\delta X(t)\|_S^2 + \frac{1}{2\eta}\|S^{1/2}(D+Q)R(t)\|^2.$$

Since $\|S^{1/2}(D+Q)R(t)\|^2 \le \|S\|_{\mathrm{op}}\|D+Q\|^2\,\|r(t)\|^2$, we get

$$\frac{d}{dt}E(t) \;\le\; (\eta - 2C_S)\,E(t) \;+\; \frac{\|S\|_{\mathrm{op}}\|D+Q\|^2}{\eta}\,\|r(t)\|^2.$$

Lemma G.1 implies that $\|D+Q\| \le 3\gamma$. With $E(0) = 0$, Grönwall's inequality implies, for any $\theta \in [0,1]$,

$$E(\theta h) \;\le\; e^{(\eta-2C_S)\theta h}\,\frac{9\gamma\|S\|_{\mathrm{op}}}{\eta}\int_0^{\theta h}\|r(t)\|^2 dt \;\le\; e^{(\eta-2C_S)\theta h}\,\frac{9\gamma\|S\|_{\mathrm{op}}}{\eta}\,\theta h\,\sup_{t\in[0,h]}\|r(t)\|^2,$$

which is (40) with $I_U = \sup_{s\in[nh,(n+1)h]}\|\nabla U(\widehat{X}_1^*(s)) - P(s,\widehat{X}_1^*)\|^2$ and $\eta = C_S$. □

D.4 Proof of Proposition 5

Now we consider the Picard iteration convergence when the number of iterations $\nu \to \infty$. Fix $h > 0$ and a step index n. Let $y := \widehat{X}(nh) \in \mathbb{R}^{Kd}$ and consider the operator $\widehat{\mathcal{T}}_y$ on $[0,h]$. Initialize the Picard sequence with the constant path

$$\widehat{X}_y^{[0]}(t) \equiv y, \quad t \in [0,h],$$

and for $\nu \ge 1$ set $\widehat{X}_y^{[\nu]} := \widehat{\mathcal{T}}_y\big[\widehat{X}_y^{[\nu-1]}\big]$. Recall the contraction ratio from Lemma F.2,

$$\rho \;:=\; 2\,L\,h\,\Gamma_\phi \quad \text{with} \quad \rho \le \tfrac{1}{2}.$$

Before starting the proof of Proposition 5, we introduce the following lemma.

Lemma D.4 (Picard convergence)**.** *Under Assumption 1 and the condition for step-size h in Lemma F.2 (so that $\rho \le \frac{1}{2}$), with $y = \widehat{X}(nh)$, for all $\nu \ge 2$,*

$$\mathbb{E} \sup_{t\in[0,h]} \big\|\widehat{X}_y^{[\nu]}(t) - \widehat{X}_y^{[\nu-1]}(t)\big\|^2 \;\le\; \rho^{2\nu-2}\Big[54\,\gamma\,h^2(L^2+1)\,e^{3h^2\|A\|^2}\,\mathbb{E}\,\|\widehat{X}(nh) - X(nh)\|^2 \;+\; \Delta(h)\Big],$$

where

$$\Delta(h) := e^{3h^2\|A\|^2}\Big\{54\,\gamma\,h^2\Big(\mathbb{E}\,\|\nabla U\big(X_1(nh)\big)\|^2 + \mathbb{E}\,\|X(nh)\|^2\Big) \;+\; 24\,d\,h\Big\}.$$

Moreover, using the moment bound for $\mathbb{E}\|\nabla U(X_1)\|^2$ and $\mathbb{E}\|X\|^2$ in Lemma G.4, and let $x_\star \in \arg\min U$, then we have

$$\Delta(h) \;\le\; e^{3h^2\|A\|^2}\Big\{54\,\gamma\,h^2\Big(\tfrac{L^2}{m}\,d + 2\|x_\star\|^2 + \tfrac{2}{m}d + (K-1)d\Big) + 24\,d\,h\Big\} = O(dh). \tag{44}$$

Proof. By the contraction property of $\widehat{\mathcal{T}}_y$ in Lemma F.2,

$$\sup_{t\in[0,h]} \big\|\widehat{X}_y^{[\nu]}(t) - \widehat{X}_y^{[\nu-1]}(t)\big\| \le \rho \sup_{t\in[0,h]} \big\|\widehat{X}_y^{[\nu-1]}(t) - \widehat{X}_y^{[\nu-2]}(t)\big\| \le \rho^{\nu-1} \sup_{t\in[0,h]} \big\|\widehat{X}_y^{[1]}(t) - \widehat{X}_y^{[0]}(t)\big\|.$$

Taking squares and expectation yields

$$\mathbb{E} \sup_{t\in[0,h]} \big\|\widehat{X}_y^{[\nu]}(t) - \widehat{X}_y^{[\nu-1]}(t)\big\|^2 \le \rho^{2\nu-2}\,\mathbb{E} \sup_{t\in[0,h]} \big\|\widehat{X}_y^{[1]}(t) - \widehat{X}_y^{[0]}(t)\big\|^2. \tag{45}$$

It remains to bound the initial increment $\widehat{X}_y^{[1]}(t) - \widehat{X}_y^{[0]}(t)$. With $\widehat{X}_y^{[0]}(t) \equiv y = \widehat{X}(nh)$, the interpolation satisfies $P(s; y) \equiv \nabla U\big(\widehat{X}_1(nh)\big)$. By the decomposition of $\widehat{\mathcal{T}}_y$ in (13),

$$\widehat{X}_y^{[1]}(t) - \widehat{X}_y^{[0]}(t) = \int_0^t A\big(\widehat{X}_y^{[1]}(s) - \widehat{X}_y^{[0]}(s)\big)\,ds - \int_0^t (D+Q)\begin{bmatrix}\nabla U\big(\widehat{X}_1(nh)\big)\\ \widehat{X}_2(nh)\\ \vdots\\ \widehat{X}_K(nh)\end{bmatrix} ds + \int_0^t \sqrt{2D}\,dB_s.$$

By the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}\mathbb{E} \sup_{t\in[0,h]} \|\widehat{X}_y^{[1]}(t) - \widehat{X}_y^{[0]}(t)\|^2 \le 3h\,\|A\|^2 \int_0^h \mathbb{E} \sup_{u\in[0,s]} \|\widehat{X}_y^{[1]}(u) - \widehat{X}_y^{[0]}(u)\|^2 ds \\ + 3h^2\|D+Q\|^2\Big(\mathbb{E}\,\|\nabla U(\widehat{X}_1(nh))\|^2 + \mathbb{E}\,\|\widehat{X}(nh)\|^2\Big) \\ + 3\,\mathbb{E} \sup_{t\in[0,h]} \Big\|\int_0^t \sqrt{2D}\,dB_s\Big\|^2.\end{aligned} \tag{46}$$

For the last term we use Doob's L^2 inequality and the Itô isometry:

$$\mathbb{E}\sup_{t\in[0,h]}\Big\|\int_0^t\sqrt{2D}\,dB_s\Big\|^2\le 4\,\mathbb{E}\Big\|\int_0^h\sqrt{2D}\,dB_s\Big\|^2=4\,\mathrm{tr}(2D)\,h=8\,\mathrm{tr}(D)\,h=8dh.$$

Next, control the second $\widehat{X}$-terms in (46) by adding and subtracting the continuous process X at time $t=nh$:

$$\mathbb{E}\,\|\nabla U(\widehat{X}_1(nh))\|^2\le 2L^2\,\mathbb{E}\,\|\widehat{X}(nh)-X(nh)\|^2+2\,\mathbb{E}\,\|\nabla U(X_1(nh))\|^2,$$
$$\mathbb{E}\,\|\widehat{X}(nh)\|^2\le 2\,\mathbb{E}\,\|\widehat{X}(nh)-X(nh)\|^2+2\,\mathbb{E}\,\|X(nh)\|^2.$$

Plugging these into the (46) and applying Grönwall's inequality to $s\mapsto\mathbb{E}\sup_{u\in[0,s]}\|\widehat{X}_y^{[1]}(u)-\widehat{X}_y^{[0]}(u)\|^2$ gives

$$\mathbb{E}\sup_{t\in[0,h]}\|\widehat{X}_y^{[1]}(t)-\widehat{X}_y^{[0]}(t)\|^2\le e^{3h^2\|A\|^2}\Big\{6h^2\|D+Q\|^2\Big(L^2\mathbb{E}\|\widehat{X}(nh)-X(nh)\|^2+\mathbb{E}\|\nabla U(X_1(nh))\|^2\Big)$$
$$+6h^2\|D+Q\|^2\Big(\mathbb{E}\|\widehat{X}(nh)-X(nh)\|^2+\mathbb{E}\|X(nh)\|^2\Big)+24dh\Big\}.\quad(47)$$

Collecting the $\mathbb{E}\|\widehat{X}-X\|^2$ terms yields the stated $6h^2(L^2+1)\|D+Q\|^2e^{3h^2\|A\|^2}\mathbb{E}\|\widehat{X}-X\|^2$, and the remainder is $\Delta(h)$ as defined. Lemma G.1 shows that $\|D+Q\|\le 3\gamma$. Combining with the (45) proves the lemma. □

Now the following is the formal statement of Proposition 5.

Proposition D.5 (Picard iteration error)**.** *Fix $h>0$ and the step n. Let $y:=\widehat{X}(nh)$ and let $X_\phi^*[y]$ be the fixed point of $\widehat{\mathcal{T}}_y$ on $[0,h]$. Define the Picard sequence $\widehat{X}_y^{[0]}(t)\equiv y$ and $\widehat{X}_y^{[\nu+1]}:=\widehat{\mathcal{T}}_y\big[\widehat{X}_y^{[\nu]}\big]$. Assume the step-size condition of Lemma F.2 holds so that $\rho:=2Lh\Gamma_\phi\le\frac12$. Let*

$$C_\Delta(h)\;:=\;54h^2(L^2+1)\,e^{3h^2\|A\|^2}\,\mathbb{E}\big\|\widehat{X}(nh)-X(nh)\big\|^2\;+\;\Delta(h),$$

with $\Delta(h)$ as in Lemma D.4. Then, for every $\nu\ge0$,

$$\mathbb{E}\sup_{t\in[0,h]}\big\|\widehat{X}_y^*(t)-\widehat{X}_y^{[\nu]}(t)\big\|^2\;\le\;\frac{\rho^{2\nu}}{(1-\rho)^2}\,C_\Delta(h).\quad(48)$$

Proof. Work in the Banach space $\big(\mathcal{H},\|\cdot\|_{\mathcal{H}}\big)$ with $\|Z\|_{\mathcal{H}}:=\Big(\mathbb{E}\sup_{t\in[0,h]}\|Z(t)\|^2\Big)^{1/2}$. Since,

$$\widehat{X}_y^*-\widehat{X}_y^{[\nu]}\;=\;\sum_{k=\nu}^{\infty}\big(\widehat{X}_y^{[k+1]}-\widehat{X}_y^{[k]}\big),$$

by the triangle inequality in $\|\cdot\|_{\mathcal{H}}$,

$$\|\widehat{X}_y^* - \widehat{X}_y^{[\nu]}\|_{\mathcal{H}} \;\le\; \sum_{k=\nu}^{\infty} \big\|\widehat{X}_y^{[k+1]} - \widehat{X}_y^{[k]}\big\|_{\mathcal{H}}.$$

Lemma D.4 gives, for each $k \ge 0$,

$$\big\|\widehat{X}_y^{[k+1]} - \widehat{X}_y^{[k]}\big\|_{\mathcal{H}} \;=\; \Big(\mathbb{E}\sup_{t\in[0,h]} \|\widehat{X}_y^{[k+1]}(t) - \widehat{X}_y^{[k]}(t)\|^2\Big)^{1/2} \;\le\; \rho^k\,\sqrt{C_\Delta(h)}.$$

Therefore

$$\|\widehat{X}_y^* - \widehat{X}_y^{[\nu]}\|_{\mathcal{H}} \;\le\; \sqrt{C_\Delta(h)}\sum_{k=\nu}^{\infty}\rho^k \;=\; \frac{\rho^\nu}{1-\rho}\,\sqrt{C_\Delta(h)}.$$

Squaring both sides yields (48). □

E Proof of Lemma 4

First we recall the definition of the Lagrange interpolation operator used in our scheme.

Definition 2. Let $X : [0,h] \to \mathbb{R}^d$ denote a given flow. For a fixed integer $M \ge 2$, define the equispaced interpolation nodes

$$s_i = \frac{i-1}{M-1}\,h, \qquad i = 1,2,\dots,M.$$

The *The Lagrange interpolation operator* P is defined for $t \in [0,h]$ by

$$P(t;X) = \sum_{i=1}^{M} \nabla U\big(X(s_i)\big) \prod_{\substack{j=1\\ j\neq i}}^{M} \frac{t-s_j}{s_i-s_j}. \tag{49}$$

Remark E.1. This definition is equivalent to its reparameterized form on $[0,1]$: with $s = t/h$ and $u_i = s_i/h$, the basis polynomials satisfy $\ell_i(t) = \tilde{\ell}_i(t/h)$ and hence $P(t;X) = \tilde{P}(t/h;X)$.

The following lemma provides a standard error estimate for the Lagrange interpolation with equispaced nodes:

Lemma E.1 (Stoer et al. (1980))**.** *For a curve $(x_t : 0 \le t \le \ell)$ in $\mathbb{R}^d$, let $(\Phi(t;x) \mid 0 \le t \le \ell)$ be the $(\alpha-1)$-order Lagrange polynomial defined at the α nodes. Then the interpolation error is bounded as*

$$\sup_{0\le t\le \ell} \|x_t - \Phi(t;x)\|_2 \;\le\; \frac{\ell^\alpha}{\alpha!}\sup_{0\le t\le \ell}\left\|\frac{\mathrm{d}^\alpha}{\mathrm{d}t^\alpha}x_t\right\|. \tag{50}$$

With this definition and error bound in place, we now turn to Lemma 1, which we restate below.

Lemma E.2. *Let h'' be defined in Lemma E.9. Fix a step size $0 < h \le h''$ and an integer order*

$K \geq 2$. *On each interval* $[nh, (n+1)h]$, *define the Lagrange interpolant of* $t \mapsto \nabla U\big(\widehat{X}_1^*(t)\big)$ *from the* $K-1$ *uniformly spaced nodes* $\{nh + c_j h\}_{j=1}^{K-1}$ *with* $c_j \in [0,1]$. *Let*

$$I_U \;:=\; \sup_{s\in[nh,(n+1)h]} \big\|\nabla U\big(\widehat{X}_1^*(s)\big) - P\big(s, \widehat{X}_1^*\big)\big\|^2. \tag{51}$$

Under Assumption 1-3 with smoothness parameters $\{L_i\}_{i=1}^K$, *there exists a constant* $C_{\mathrm{IR}} > 0$ *depending only on* $\big(K, \{L_i\}_{i\leq K}, \{c_j\}\big)$ *and the parameter* γ *from matrices* D *and* Q, *but not on* h, d, n, *or* N, *such that, uniformly for all* $n = 0, 1, \ldots, N-1$,

$$\mathbb{E}\, I_U \;\leq\; C_{\mathrm{IR}}\, N\, h^{\,2K-2}\, d^{\,K-1}. \tag{52}$$

Proof. Suppose we use $(K-1)$ nodes to construct the polynomial interpolant. By Lemma E.1, the approximation error I_U satisfies

$$\sqrt{I_U} = \sup_{t\in[nh,(n+1)h]} \big\|\nabla U\big(\widehat{X}_1^*(s)\big) - P\big(t, \widehat{X}_1^*(s)\big)\big\| \leq \frac{h^{K-1}}{(K-1)!} \sup_{t\in[nh,(n+1)h]} \left\| \frac{d^{K-1}}{dt^{K-1}} \nabla U\big(\widehat{X}_1^*(t)\big) \right\|.$$

We use two auxiliary Lemmas whose proofs are deferred to Part I (Subsection E.1) and Part II (Subsection E.2) below. Part I (Lemma E.4) provides an inductive bound on higher–order time derivatives of ∇U along the flow; Part II (Lemma E.9) provides high order moment bound for $\sup_{t\in[0,1]} \|\widehat{X}^*(nh + th)\|$.

Using the bound for higher-order derivatives of ∇U from Part I (Lemma E.4) and squaring both sides, we obtain

$$I_U \leq \frac{h^{2K-2}}{(K-1)!^2}\, C_{K-1}^2 \left[2 + 2 \sup_{t\in I} \|\widehat{X}^*\|^{2K-2}\, \mathbb{1}_{\{\sup_{t\in I}\|\widehat{X}^*\|\geq 1\}}\right].$$

Taking expectations and using the high–moment bound from Part II (Lemma E.9),we obtain

$$\begin{aligned}
\mathbb{E}\, I_U &\leq \frac{h^{2K-2}}{(K-1)!^2}\, C_{K-1}^2 \big[2 + 6N\, (C_f d)^{K-1}\, (K-1)!\big] \\
&\leq h^{2K-2} d^{K-1} N\, \frac{C_{K-1}^2 \big[2 + 6\, C_f^{K-1} (K-1)!\big]}{(K-1)!^2}.
\end{aligned}$$

This establishes the desired bound and completes the proof. □

E.1 Bounding I_U: Part I

Our first step uses Assumption 2 to control the higher–order derivatives of U. Here we reinstate Assumption 2.

Assumption 2. Let T be an m-th order tensor. Its tensor norm is defined as

$$\|T\|_{\text{tsr}} \;=\; \sup_{v_i \in S^{d-1}} \big\|T \cdot [v_1, v_2, \ldots, v_{m-1}]\big\|,$$

where S^{d-1} denotes the unit sphere in $\mathbb{R}^d$. We assume that for all $2 \le i \le K$,

$$\|\nabla^{(i)} U\|_{\text{tsr}} \;\le\; L_i.$$

The following lemma is a direct consequence.

Lemma E.3. *For any vectors $\Delta_1, \ldots, \Delta_k \in \mathbb{R}^d$,*

$$\|D^{(k+1)} U[\Delta_1, \ldots, \Delta_k]\|_2 \;\le\; L_{max} \prod_{i=1}^{k} \|\Delta_i\|_2, \tag{53}$$

where $L_{max} = \max_{1 \le i \le K} L_i$.

With this multilinear control in hand, we now derive bounds on the time derivatives of ∇U along the trajectory $\widehat{X}_1^*(t)$. For brevity, we write $\widehat{X}_1^* := \widehat{X}_1^*(t)$, suppressing the dependence on t. We further use $\widehat{X}^*$ to denote the entire Kd-dimensional trajectory vector.

Lemma E.4. *Let $C_1 = L$. Then, for each $n \le K-1$, the n-th time derivative satisfies*

$$\sup_{t \in I} \left\| \frac{d^n}{dt^n} \nabla U(\widehat{X}_1^*) \right\| \;\le\; C_n \left[\mathbb{1}_{\{\sup_{t \in I} \|\widehat{X}^*\| < 1\}} + \sup_{t \in I} \|\widehat{X}^*\|^n \, \mathbb{1}_{\{\sup_{t \in I} \|\widehat{X}^*\| \ge 1\}} \right], \tag{54}$$

where I denote the interpolation interval $[nh, (n+1)h]$. The constant C_n depends solely on the preceding constants $\{C_i\}_{i<n}$ and on $L_{\max}$.

Proof. We use induction to prove the statement.

For $n = 1$,

$$\begin{aligned}
\sup_{t \in I} \left\| \frac{d}{dt} \nabla U(\widehat{X}_1^*) \right\| &\le \sup_{t \in I} \left\| D^{(2)}[U(\widehat{X}_1^*)] \right\|_{tsr} \left\| \frac{d\widehat{X}_1^*}{dt} \right\| \\
&\le L_{max} \sup_{t \in I} \|\widehat{X}_2^*\| \\
&\le L_{max} \sup_{t \in I} \|\widehat{X}^*\|.
\end{aligned}$$

Here $C_1 = L_{max}$. Assume that for $n \le K-2$, the bound (54) holds. Let us try to bound the $(n+1)$-th derivative. Using Faà di Bruno's formula (Lemma E.11) together with the bound on

multilinear operations (Lemma E.3), we obtain

$$\begin{aligned}
\sup_{t\in I}\left\|\frac{d^{n+1}\nabla U(\widehat{X}_1^*)}{dt^{n+1}}\right\| &\le \sup_{t\in I}\sum_{k=0}^{n+1}\sum_{j_i}\frac{(n+1)!}{j_1!1!^{j_1}j_2!\dots j_{n+2-k}!(n+2-k)!^{j_{n+2-k}}}\|D^{(k+1)}(U)\|_{tsr}\prod_{l=1}^{n+2-k}\left\|\frac{d^l\widehat{X}_1^*}{dt^l}\right\|^{j_l}\\
&\le L_{max}\sum_{k=0}^{n+1}\sum_{j_i}\frac{(n+1)!}{j_1!1!^{j_1}j_2!\dots j_{n+2-k}!(n+2-k)!^{j_{n+2-k}}}\sup_{t\in I}\prod_{l=1}^{n+2-k}\left\|\frac{d^l\widehat{X}_1^*}{dt^l}\right\|^{j_l}\\
&\le L_{max}\sum_{k=0}^{n+1}\sum_{j_i}\frac{(n+1)!}{j_1!1!^{j_1}j_2!\dots j_{n+2-k}!(n+2-k)!^{j_{n+2-k}}}\\
&\qquad\prod_{l=1}^{n+2-k}\left(2^{l-2}\gamma^{l-1}\sup_{t\in I}\|\widehat{X}^*\|+\gamma^{l-1}\sum_{i=0}^{l-2}C_{p,i}\sup_{t\in I}\left\|\frac{d^i}{dt^i}\nabla U(\widehat{X}_1^*)\right\|\right)^{j_l}.
\end{aligned}$$

In the last inequality, we used Lemma E.16 to control the higher-order derivatives of the fixed point.

If $\sup_{t\in I}\|\widehat{X}^*\|<1$ then

$$\begin{aligned}
\sup_{t\in I}\left\|\frac{d^{n+1}\nabla U(\widehat{X}_1^*)}{dt^{n+1}}\right\| &\le L_{max}\sum_{k=0}^{n+1}\sum_{j_i}\frac{(n+1)!}{j_1!1!^{j_1}j_2!\dots j_{n+2-k}!(n+2-k)!^{j_{n+2-k}}}\prod_{l=1}^{m-k+1}\left(2^{l-2}\gamma^{l-1}+\sum_{i=0}^{(l-2)}C_{p,i}C_l\right)^{j_l}\\
&= C_{n+1}.
\end{aligned}$$

If $\sup_{t\in I}\|\widehat{X}^*\|\ge 1$, then

$$\begin{aligned}
\sup_{t\in I}\left\|\frac{d^{n+1}\nabla U(\widehat{X}_1^*)}{dt^{n+1}}\right\| &\le L_{max}\sum_{k=0}^{n+1}\sum_{j_i}\frac{(n+1)!}{j_1!1!^{j_1}j_2!\dots j_{n+2-k}!(n+2-k)!^{j_{n+2-k}}}\\
&\qquad\prod_{l=1}^{m-k+1}\left(2^{l-2}\gamma^{l-1}+\sum_{i=0}^{(l-2)}C_{p,i}C_l\right)^{j_l}\sup_{t\in I}\|\widehat{X}^*\|^{l\cdot j_l}\\
&\le L_{max}\sup_{t\in I}\|\widehat{X}^*\|^{n+1}\sum_{k=0}^{n+1}\sum_{j_i}\frac{(n+1)!}{j_1!1!^{j_1}j_2!\dots j_{n+2-k}!(n+2-k)!^{j_{n+2-k}}}\\
&\qquad\prod_{l=1}^{m-k+1}\left(2^{l-2}\gamma^{l-1}+\sum_{i=0}^{(l-2)}C_{p,i}C_l\right)^{j_l}\\
&\le C_{n+1}\sup_{t\in I}\|\widehat{X}^*\|^{n+1}.
\end{aligned}$$

In both these cases we define

$$C_{n+1}=L_{max}\sum_{k=0}^{n+1}\sum_{j_i}\frac{(n+1)!}{j_1!1!^{j_1}j_2!\dots j_{n+2-k}!(n+2-k)!^{j_{n+2-k}}}\prod_{l=1}^{m-k+1}\left(2^{l-2}\gamma^{l-1}+\sum_{i=0}^{(l-2)}C_{p,i}C_l\right)^{j_l},$$

which is only dependent on C_i, $C_{p,i}$ for $i \leq n$ and L_{max}. Note that this quantity is independent of d. Combining these two results, we have

$$\left\| \frac{d^{n+1}}{dt^{n+1}} \nabla U(\widehat{X}_1^*) \right\| \leq C_{n+1} \left[\mathbb{1}_{\{\sup_{t\in I} \|\widehat{X}^*\|<1\}} + \sup_{t\in I} \|\widehat{X}^*\|^{n+1} \mathbb{1}_{\{\sup_{t\in I} \|\widehat{X}^*\|\geq 1\}} \right].$$

□

E.2 Bounding I_U: Part II

In this subsection, we derive moment bounds for $\sup_{t\in[0,1]} \|\widehat{X}^*(nh+th)\|$. Before proceeding, we first introduce the following uniform bound for Brownian motion.

Lemma E.5. *Let $(B_t)_{t\geq 0}$ be a d-dimensional standard Brownian motion, $h > 0$ denote the stepsize and $N \geq 1$ be an integer denoting the maximum number of iterations. For $C_b \geq 0$ define the events*

$$\mathcal{G}_n(h, C_b) \;:=\; \Big\{ \sup_{0\leq t\leq h} \|B_{nh+t} - B_{nh}\| \;\leq\; C_b \Big\}, \qquad n = 0, 1, \ldots, N-1.$$

Then

$$\mathbb{P}\left(\bigcap_{n=0}^{N-1} \mathcal{G}_n(h, C_b) \right) \;\geq\; 1 - 3N \exp\left(-\frac{C_b^2}{6dh} \right).$$

Equivalently,

$$\mathbb{P}\left(\max_{0\leq n\leq N-1} \sup_{0\leq t\leq h} \|B_{nh+t} - B_{nh}\| \;\geq\; C_b \right) \;\leq\; 3N \exp\left(-\frac{C_b^2}{6dh} \right).$$

In particular, for any $\delta \in (0,1)$, choosing

$$C_b \;=\; \sqrt{6dh \, \log\frac{3N}{\delta}}$$

ensures $\mathbb{P}\Big(\bigcap_{n=0}^{N-1} \mathcal{G}_n(h, C_b\Big) \geq 1 - \delta$.

Proof. By stationary increments, for each n, $\sup_{0\leq t\leq h} \|B_{nh+t} - B_{nh}\| \overset{d}{=} \sup_{0\leq t\leq h} \|B_t\|$. The one-interval tail bound $\mathbb{P}\big(\sup_{0\leq t\leq h} \|B_t\| \geq C_b\big) \leq 3\exp\big(-C_b^2/(6dh)\big)$ (See, i.e., Chewi et al. (2024, lemma 34)) then implies, by a union bound over $n = 0, \ldots, N-1$,

$$\mathbb{P}\left(\bigcup_{n=0}^{N-1} \mathcal{G}_n(h, C_b)^c \right) \leq \sum_{n=0}^{N-1} 3\exp\left(-\frac{C_b^2}{6dh} \right) = 3N \exp\left(-\frac{C_b^2}{6dh} \right),$$

which yields the claim. □

Given a starting point y, we analyze the first Picard iterate on $[0, h]$. On the event where the Brownian increment is suitably bounded (the "nice" event), the following lemma provides an explicit

upper bound on its deviation from y.

Lemma E.6. *Let $y \in \mathbb{R}^{Kd}$ and y_1 be the first K element of y. Define the Picard iterates on $[0,h]$ by $\widehat{X}_y^{[0]}(t) \equiv y$ and $\widehat{X}_y^{[1]} := \widehat{\mathcal{T}}_y\big[\widehat{X}_y^{[0]}\big]$. Then, on the event $\bigcap_{n=0}^{N-1} \mathcal{G}_n(h, C_b)$, the following bound holds:*

$$\sup_{0\le t\le h} \big\|\widehat{X}_y^{[1]}(t) - y\big\| \;\le\; e^{\|A\|h}\Big[h\,\|D{+}Q\|\big(L\|y_1\| + \|y\|\big) + \sqrt{2\gamma}\,C_b\Big],$$

where A is the matrix defined in (13). In particular, if $h \le (\log 2)/\|A\|$ (so $e^{\|A\|h} \le 2$), then

$$\sup_{0\le t\le h} \big\|\widehat{X}_y^{[1]}(t) - y\big\| \;\le\; 2\Big[h\,\|D{+}Q\|\big(L\|y_1\| + \|y\|\big) + \sqrt{2\gamma}\,C_b\Big].$$

Proof. For $t \in [0,h]$,

$$\begin{aligned}\sup_{0\le u\le t} \big\|\widehat{X}_y^{[1]}(u) - y\big\| \le \|A\| \int_0^t \sup_{0\le r\le s} \big\|\widehat{X}_y^{[1]}(r) - y\big\|\,\mathrm{d}s \\ + h\,\|D{+}Q\|\Big(L\|y_1\| + \|y\|\Big) + \sqrt{2\gamma} \sup_{0\le u\le h} \big\|B_u\big\|.\end{aligned}$$

Applying Grönwall's inequality at $t = h$ gives

$$\sup_{0\le u\le h} \big\|\widehat{X}_y^{[1]}(u) - y\big\| \le e^{\|A\|h}\left[h\,\|D{+}Q\|\Big(L\|y_1\| + \|y\|\Big) + \sqrt{2\gamma} \sup_{0\le u\le h} \big\|B_u\big\|\right].$$

□

We now establish a high-probability bound that controls the discrete evolution of the algorithm across steps.

Lemma E.7. *For $h \le h'$, where $h' := \min\left\{\frac{C_S}{18\gamma^2 L^2\,\|S\|\|S^{-1}\|},\ \left(\frac{C_S}{5184\gamma^4\|S\|\|S^{-1}\|(1+L)^2(1+2L\Gamma_\phi)^2}\right)^{1/3}\right\}$, define the event $\bigcap_{n=0}^{N-1} \mathcal{G}_n(h, C_b)$. Then the following bound holds:*

$$\|\widehat{X}((n+1)h)\|_S^2 \;\le\; \frac{1152h\gamma^3 C_b^2\,\|S\|(1+2L\Gamma_\phi)^2}{C_S} + \frac{4\gamma C_b^2\|S\|}{h\,C_S}, \tag{55}$$

Proof. At each algorithmic step, we focus on the final Picard iteration ν^*. Note that at the start of each algorithmic step we initialize the Picard iterations at the current state, i.e., set $y := \widehat{X}(nh)$

(For notational convenience, we omit y in the proof.) Then the following relation holds:

$$\begin{aligned}\widehat{X}((n+1)h) = \widehat{X}^{[\nu^*]}(h) &= \widehat{X}(nh) - \int_0^h (D+Q) \begin{bmatrix} P(s;\widehat{X}^{[\nu^*-1]}) \\ \widehat{X}_2^{[\nu^*]}(s) \\ \vdots \\ \widehat{X}_K^{[\nu^*]}(s) \end{bmatrix} \mathrm{d}s + \sqrt{2}\int_0^h D\,\mathrm{d}B_s \\ &= \underbrace{\left(\widehat{X}(nh) - \int_0^h (D+Q) \begin{bmatrix} \nabla U(\widehat{X}_1(nh)) \\ \widehat{X}_2(nh) \\ \vdots \\ \widehat{X}_K(nh) \end{bmatrix} \mathrm{d}s\right)}_{\text{(I)}} \\ &\quad + \underbrace{\left(-\int_0^h (D+Q) \begin{bmatrix} P(s;\widehat{X}^{[\nu^*-1]}) - \nabla U(\widehat{X}_1(nh)) \\ \widehat{X}_2^{[\nu^*]}(s) - \widehat{X}_2(nh) \\ \vdots \\ \widehat{X}_K^{[\nu^*]}(s) - \widehat{X}_K(nh) \end{bmatrix} \mathrm{d}s\right)}_{\text{(II)}} + \sqrt{2}\int_0^h \sqrt{D}\,\mathrm{d}B_s.\end{aligned} \tag{56}$$

Consider these two terms separately (first term in the S-norm):

$$\begin{aligned}\text{(I)} &= \left\| \widehat{X}(nh) - \int_0^h (D+Q) \begin{bmatrix} \nabla U(\widehat{X}_1(nh)) \\ \widehat{X}_2(nh) \\ \vdots \\ \widehat{X}_K(nh) \end{bmatrix} \mathrm{d}s \right\|_S^2 \\ &= \left\| \widehat{X}(nh) - \int_0^h \int_0^1 J_b(\lambda \widehat{X}(nh))\mathrm{d}\lambda \begin{bmatrix} \widehat{X}_1(nh) \\ \widehat{X}_2(nh) \\ \vdots \\ \widehat{X}_K(nh) \end{bmatrix} \mathrm{d}s \right\|_S^2 \\ &= \|\widehat{X}(nh)\|_S^2 + 2h\int_0^1 \widehat{X}(nh)^\top S J_b(\lambda \widehat{X}(nh))\widehat{X}(nh)\mathrm{d}\lambda + h^2\|S\|\|J_b\|^2\|\widehat{X}(nh)\|^2 \\ &\le \|\widehat{X}(nh)\|_S^2 - 2h\, C_S\|\widehat{X}(nh)\|_S^2 + h^2\|S\|\|S^{-1}\|\|J_b\|^2\|\widehat{X}(nh)\|_S^2.\end{aligned}$$

Here we apply the same mean-value theorem argument as in Proposition C.4, using the two endpoints $\widehat{X}(nh)$ and $\mathbf{0}$, to obtain the final inequality.

Let us deal with the second term now (in l_2 norm):

$$\text{(II)} = \left\| \int_0^h (D+Q) \begin{bmatrix} P\big(s;\, \widehat{X}^{[\nu^*-1]}\big) - \nabla U\big(\widehat{X}_1(nh)\big) \\ \widehat{X}_2^{[\nu^*]}(s) - \widehat{X}_2(nh) \\ \vdots \\ \widehat{X}_K^{[\nu^*]}(s) - \widehat{X}_K(nh) \end{bmatrix} \mathrm{d}s \right\|_2$$

$$\leq Lh\,\Gamma_\phi \,\|D+Q\| \sup_{[0,h]} \big\|\widehat{X}^{[\nu^*-1]}(s) - \widehat{X}(nh)\big\| \;+\; h\,\|D+Q\| \sup_{[0,h]} \big\|\widehat{X}^{[\nu^*]}(s) - \widehat{X}(nh)\big\| \tag{57}$$

$$\leq \rho\,\|D+Q\| \sup_{[0,h]} \big\|\widehat{X}^{[\nu^*-1]}(s) - \widehat{X}(nh)\big\| \;+\; h\,\|D+Q\| \sup_{[0,h]} \big\|\widehat{X}^{[\nu^*]}(s) - \widehat{X}(nh)\big\| \tag{58}$$

$$\leq \rho\,\|D+Q\| \frac{1-\rho^{\nu^*-1}}{1-\rho} \sup_{[0,h]} \big\|\widehat{X}^{[1]}(s) - \widehat{X}(nh)\big\| \;+\; h\,\|D+Q\| \frac{1-\rho^{\nu^*}}{1-\rho} \sup_{[0,h]} \big\|\widehat{X}^{[1]}(s) - \widehat{X}(nh)\big\| \tag{59}$$

$$\leq 2\left[\rho\,\|D+Q\| \frac{1-\rho^{\nu^*-1}}{1-\rho} + h\,\|D+Q\| \frac{1-\rho^{\nu^*}}{1-\rho}\right] \left[h\,\|D+Q\|\big(L\|\widehat{X}_1(nh)\| + \|\widehat{X}(nh)\|\big) + \sqrt{2\gamma}\,C_b\right]. \tag{60}$$

We now explain each inequality in detail:

- (57) follows from the Lipschitz continuity of ∇U and the definition of the interpolant $P(s; \widehat{X}^{[\nu^*-1]})$.
- (58) applies the definition of $\rho = 2Lh\,\Gamma_\phi$.
- (59) uses the Picard contraction property established in Lemma F.2 with rate $\rho = 2Lh\,\Gamma_\phi$, and then applies a geometric-sum argument to accumulate the iterates.
- (60) follows from Lemma E.6, which bounds the first Picard increment.

Next, we express (II) in terms of the current state $\widehat{X}(nh)$, noting that with $\rho := 2Lh\,\Gamma_\phi \leq \frac{1}{2}$ we have

$$\begin{aligned}
\text{(II)} &\leq 2h\,\|D+Q\|^2(1+L)\left[\rho\frac{1-\rho^{\nu^*-1}}{1-\rho} + h\frac{1-\rho^{\nu^*}}{1-\rho}\right]\|\widehat{X}(nh)\| \\
&\quad + 2\sqrt{2\gamma}\,C_b\|D+Q\|\left[\rho\,\frac{1-\rho^{\nu^*-1}}{1-\rho} + h\,\frac{1-\rho^{\nu^*}}{1-\rho}\right] \\
&\leq 2h\,\|D+Q\|^2(1+L)(2\rho+2h)\|\widehat{X}(nh)\| \;+\; 2\sqrt{2\gamma}\,C_b\,\|D+Q\|(2\rho+2h) \qquad (\because \rho \leq 1/2) \\
&\leq 4h^2\,\|D+Q\|^2(1+L)(1+2L\Gamma_\phi)\|\widehat{X}(nh)\| \;+\; 4\sqrt{2\gamma}\,C_b\,\|D+Q\|(1+2L\Gamma_\phi)h. \qquad (\because \rho = 2Lh\,\Gamma_\phi)
\end{aligned}$$

We now proceed to analyze the recursion in the S-norm. Combining (56) with the bounds for terms

(I) and (II) (Converting into S-norm), we obtain

$$\begin{aligned}\|\widehat{X}((n+1)h)\|_S^2 &\le \|\widehat{X}(nh)\|_S^2 - 2h\,C_S\,\|\widehat{X}(nh)\|_S^2 + h^2\|S\|\|S^{-1}\|\|J_b\|^2\|\widehat{X}(nh)\|_S^2 \\ &\quad + 32\|S\|\|S^{-1}\|\,h^4\|D{+}Q\|^4(1+L)^2(1+2L\Gamma_\phi)^2\|\widehat{X}(nh)\|_S^2 \\ &\quad + 64\|S\|\gamma C_b^2\,h^2\|D{+}Q\|^2(1+2L\Gamma_\phi)^2 + 2\gamma C_b^2\|S\| \\ &\le \Big(1 - 2h\,C_S + h^2\|S\|\|S^{-1}\|\|J_b\|^2 + 32h^4\|S\|\|S^{-1}\|\|D{+}Q\|^4(1+L)^2(1+2L\Gamma_\phi)^2\Big)\|\widehat{X}(nh)\|_S^2 \\ &\quad + 64\|S\|\gamma C_b^2\,h^2\|D{+}Q\|^2(1+2L\Gamma_\phi)^2 + 2\gamma C_b^2\|S\| \\ &\le \Big(1 - \tfrac{C_S h}{2}\Big)\|\widehat{X}(nh)\|_S^2 + 576h^2\gamma^3 C_b^2\,\|S\|(1+2L\Gamma_\phi)^2 + 2\gamma C_b^2\|S\|.\end{aligned}$$

Here the last inequality follows from the bounds on $\|D{+}Q\|$ and $\|J_b\|$ (Lemmas G.1 and G.2), and by enforcing the bound

$$\Big(1 - 2h\,C_S + 9h^2\gamma^2L^2\|S\|\|S^{-1}\| + 2592h^4\gamma^4\|S\|\|S^{-1}\|(1+L)^2(1+2L\Gamma_\phi)^2\Big) \;\le\; 1 - \tfrac{C_S}{2}h,$$

by choosing a sufficiently small stepsize h such that

$$h \;\le\; \min\left\{\frac{C_S}{18\gamma^2L^2\,\|S\|\|S^{-1}\|},\ \left(\frac{C_S}{5184\gamma^4\|S\|\|S^{-1}\|(1+L)^2(1+2L\Gamma_\phi)^2}\right)^{1/3}\right\}.$$

Starting from the initialization $\widehat{X}(0) = 0$, the recursion yields for all n,

$$\|\widehat{X}((n+1)h)\|_S^2 \le \frac{1152h\gamma^3C_b^2\,\|S\|(1+2L\Gamma_\phi)^2}{C_S} + \frac{4\gamma C_b^2\|S\|}{h\,C_S}.$$

□

Fix $n \in \{0,\dots,N{-}1\}$. On the high–probability event $\bigcap_{n=0}^{N-1}\mathcal{G}_n(h,C_b)$ (which occurs with probability at least $1-\delta$), next we bound the Picard fixed point initialized at the current state $\widehat{X}(nh)$.

Lemma E.8. *For any $k \in \mathbb{N}$ and any stepsize*

$$h \;\le\; \min\left\{h',\ \frac{1}{768\,\gamma},\ \left(\frac{C_S}{20736\,\gamma^3\,\|S\|\,\|S^{-1}\|\,(1+2L\Gamma_\phi)^2}\right)^{1/2}\right\},$$

where h' is from Lemma E.7, the following bound holds with probability at least $1-\delta$:

$$\sup_{t\in[0,1]}\|\widehat{X}^*_y(nh+th)\| \lesssim \left[d\,\log\frac{3N}{\delta}\right]^{1/2}, \tag{61}$$

where $y = \widehat{X}(nh)$.

Proof. For notational convenience, we suppress the subscript y, though we emphasize that the

Picard fixed point is initialized at $y = \widehat{X}(nh)$. Under the event $\bigcap_{n=0}^{N-1} \mathcal{G}_n(h, C_b)$, we have

$$\begin{aligned}
\sup_{t\in[0,1]} \|\widehat{X}^*(nh+th)\|^2 &\le 2 \sup_{t\in[0,1]} \|\widehat{X}^*(nh+th) - \widehat{X}(nh)\|^2 + 2\|\widehat{X}(nh)\|^2 \\
&\le \frac{2}{(1-\rho)^2} \sup_{t\in[0,1]} \|\widehat{X}^{[1]}(nh+th) - \widehat{X}(nh)\|^2 + 2\|\widehat{X}(nh)\|^2 \qquad \text{(consequence of Lemma F.2)} \\
&\le \frac{2}{(1-\rho)^2} \left[8h^2\|D+Q\|^2(L+1)^2\|\widehat{X}(nh)\|^2 + 16\gamma C_b^2\right] + 2\|\widehat{X}(nh)\|^2 \qquad \text{(by Lemma E.6)} \\
&\le \left[\frac{16h^2\|D+Q\|^2(L+1)^2}{(1-\rho)^2} + 2\right] \|S^{-1}\|\|\widehat{X}(nh)\|_S^2 + \frac{32\gamma C_b^2}{(1-\rho)^2},
\end{aligned}$$

where in the second inequality, we applied the Picard contraction at rate $\rho \le 1/2$ (consequence of Lemma F.2). Noting that $\rho \le \frac{1}{2}$ and assuming $576h^2\gamma^2(L+1)^2 < 1$ (absorbing $\|D+Q\|$ using the bound from Lemma G.1), we deduce from Lemma E.7 that

$$\begin{aligned}
\sup_{t\in[0,1]} \|\widehat{X}^*(nh+th)\|^2 &\le 3\|S^{-1}\|\|\widehat{X}(nh)\|_S^2 + 128\gamma C_b^2 \\
&\le \frac{3456h\gamma^3 C_b^2 \|S\|\|S^{-1}\|(1+2L\Gamma_\phi)^2}{C_S} + \frac{12\gamma C_b^2\|S\|\|S^{-1}\|}{h\, C_S} + 128\gamma C_b^2. \qquad (62)
\end{aligned}$$

Recall that $C_b^2 = 6dh\log\left(\frac{3N}{\delta}\right)$. Define

$$K_1 \;:=\; \frac{\gamma^3\|S\|\,\|S^{-1}\|\,(1+2L\Gamma_\phi)^2}{C_S}, \qquad K_2 \;:=\; \frac{\|S\|\,\|S^{-1}\|\,\gamma}{C_S}.$$

Then the right-hand side of (62) rewrites as

$$\begin{aligned}
&\frac{3456h\,\gamma^3 C_b^2\,\|S\|\,\|S^{-1}\|\,(1+2L\Gamma_\phi)^2}{C_S} + \frac{12\,\gamma C_b^2\,\|S\|\,\|S^{-1}\|}{h\,C_S} + 128\,\gamma C_b^2 \\
&= d\,\log\left(\tfrac{3N}{\delta}\right)\Big(20736\,K_1\,h^2 \;+\; 72\,K_2 \;+\; 768\,\gamma\,h\Big).
\end{aligned}$$

Choose the stepsize

$$h \;\le\; \min\left\{\frac{1}{768\,\gamma},\; \frac{1}{\sqrt{20736\,K_1}}\right\}.$$

Then

$$d\,\log\left(\tfrac{3N}{\delta}\right)\Big(288\,K_1h^2 + 72\,K_2 + 768\,h\Big) \;\le\; d\,\log\left(\tfrac{3N}{\delta}\right)(72\,K_2+2).$$

Consequently,

$$\sup_{t\in[0,1]} \|\widehat{X}^*(nh+th)\|^2 \;\le\; \frac{3456h\,\gamma^3 C_b^2\,\|S\|\,\|S^{-1}\|\,(1+2L\Gamma_\phi)^2}{C_S} + \frac{12\,\gamma C_b^2\,\|S\|\,\|S^{-1}\|}{h\,C_S} + 128\,\gamma C_b^2 \;\le\; C_f\,d\,\log\frac{3N}{\delta},$$

for a constant $C_f := 72\,K_2 + 2$ that is independent of h and d. Since $\mathbb{P}\big(\bigcap_{n=0}^{N-1}\mathcal{G}_n(h,C_b)\big) \ge 1-\delta$, it

follows that

$$\mathbb{P}\left(\sup_{t\in[0,1]} \|\widehat{X}^*(nh+th)\| \le \left[C_f\, d\, \log\frac{3N}{\delta} \right]^{1/2} \right) \;\ge\; 1-\delta.$$

□

Lemma E.9. *Assume $h \le h''$, where $h'' := \min\{H_1, H_2, H_3, H_4\}$ and*

$$H_1 := \frac{C_S}{2\,\|S\|\,\|S^{-1}\|\,\|J_b\|}, \qquad H_2 := \left(\frac{C_S}{5184\gamma^4\,\|S\|\,\|S^{-1}\|(1+L)^2(1+2L\Gamma_\phi)^2} \right)^{1/3},$$
$$H_3 := \frac{1}{768\,\gamma}, \qquad H_4 := \left(\frac{C_S}{20736\,\gamma^3\,\|S\|\,\|S^{-1}\|\,(1+2L\Gamma_\phi)^2} \right)^{1/2}.$$

Let

$$Z \;:=\; \sup_{t\in[0,1]} \big\|\widehat{X}^*_y(nh+th)\big\|,$$

for $y = \widehat{X}(nh)$. Then for every integer $k \ge 1$,

$$\mathbb{E}\big[Z^{2k}\big] \;\le\; 3N\,(C_f d)^k\,\Gamma(k+1) \;=\; 3N\,(C_f d)^k\,k!\,.$$

In particular, $\mathbb{E}\big[Z^{2k}\big] \lesssim N\,d^k$, with a constant independent of d, h, and N.

Proof. For $k \ge 1$,

$$\mathbb{E}\big[Z^{2k}\big] = \int_0^\infty \mathbb{P}\big(Z^{2k} > t\big)\,dt = \int_0^\infty \mathbb{P}\big(Z > u\big)\,2k\,u^{2k-1}\,du.$$

By the tail estimate of Lemma E.8, we obtain

$$\mathbb{E}\big[Z^{2k}\big] \le 2k\cdot 3N \int_0^\infty u^{2k-1}\exp\Big(-\frac{u^2}{C_f d}\Big)\,du.$$

Let $a := 1/(C_f d)$. The standard integral $\int_0^\infty u^{2k-1}e^{-au^2}\,du = \frac{1}{2}a^{-k}\Gamma(k)$ yields

$$\mathbb{E}\big[Z^{2k}\big] \le 2k\cdot 3N\cdot\frac{1}{2}\,a^{-k}\,\Gamma(k) = 3N\,(C_f d)^k\,k\,\Gamma(k) = 3N\,(C_f d)^k\,\Gamma(k+1),$$

as claimed. □

E.3 Supporting Lemma for Interpolation Error

E.3.1 Faà di Bruno's Formula

In this section, we introduce Faà di Bruno's formula for composite functions. Our primary interest lies in the setting where

$$f : \mathbb{R}^d \to \mathbb{R}^d, \qquad g : \mathbb{R} \to \mathbb{R}^d,$$

so that we study higher-order derivatives of the composition $f \circ g : \mathbb{R} \to \mathbb{R}^d$. General multivariate formulations of Faà di Bruno's formula can be found in the existing literature (Constantine and Savits (1996); Mishkov (2000) etc). For completeness, we provide here a short proof by induction tailored to this special case.

Lemma E.10 (Faà di Bruno formula for the composition $f \circ g$)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}^d$ be C^m, and let $g : \mathbb{R} \to \mathbb{R}^d$ be C^m as well. Then for each $m \geq 1$, the m-th derivative of the composition $t \mapsto f(g(t))$ is given by*

$$\frac{\mathrm{d}^m f}{\mathrm{d}t^m} = \sum D^{(k)}(f) \cdot \left[\underbrace{\frac{\mathrm{d}g}{\mathrm{d}t}, \ldots, \frac{\mathrm{d}g}{\mathrm{d}t}}_{b_1 \text{ times}}, \underbrace{\frac{\mathrm{d}^2 g}{\mathrm{d}t^2}, \ldots, \frac{\mathrm{d}^2 g}{\mathrm{d}t^2}}_{b_2 \text{ times}}, \ldots, \underbrace{\frac{\mathrm{d}^m g}{\mathrm{d}t^m}, \ldots, \frac{\mathrm{d}^m g}{\mathrm{d}t^m}}_{b_m \text{ times}} \right], \tag{63}$$

where $D^{(k)} f(g(t))$ denotes the k-th derivative tensor of f at $g(t)$, i.e. the symmetric k-linear map

$$D^{(k)} f(g(t)) : (\mathbb{R}^d)^k \to \mathbb{R}^d,$$

which acts on vectors $v_1, \ldots, v_k \in \mathbb{R}^d$ as $D^{(k)} f(g(t)) \cdot [v_1, \ldots, v_k]$. Here the dot "$\cdot$" represents the natural tensor contraction, and the summation runs over all partitions of $\{1, 2, \ldots, m\}$. Each partition is represented by integers $b_1, \ldots, b_m$, where b_i is the number of blocks of size i. These satisfy

$$\sum_{i=1}^{m} i\, b_i = m, \qquad \sum_{i=1}^{m} b_i = k,$$

with k denoting the total number of blocks in the partition.

Proof. We proceed by induction on m.

For the base case $m = 1$, we have

$$\frac{d}{dt} f(g(t)) = D^{(1)} f(g(t)) \cdot \big[g'(t)\big],$$

which agrees with the claimed formula.

Now assume that the statement holds for some $m \geq 1$. Consider the case $m + 1$. Every partition of $\{1, 2, \ldots, m+1\}$ can be obtained uniquely from a partition of $\{1, 2, \ldots, m\}$ by adjoining the element $m + 1$.

Case 1: $m + 1$ forms a new singleton block. In this case, the number of blocks of size 1 increases by one, and the total number of blocks increases by one. On the analytic side, this corresponds to differentiating the factor $D^{(k)} f(g(t))$, yielding $D^{(k+1)} f(g(t))$, and adding one more factor of $g'(t)$ to the multilinear map:

$$D^{(k+1)} f(g(t)) \cdot \big[\underbrace{g'(t), \ldots, g'(t)}_{b_1+1 \text{ times}}, \underbrace{g''(t), \ldots, g''(t)}_{b_2 \text{ times}}, \ldots, \underbrace{g^{(m)}(t), \ldots, g^{(m)}(t)}_{b_m \text{ times}} \big].$$

Case 2: $m+1$ is added to an existing block of size i. In this case, the number of blocks of size i decreases by one, while the number of blocks of size $i+1$ increases by one, and the total number of blocks remains unchanged. If we started with b_i such blocks, there are b_i possible ways to add $m+1$. Analytically, this corresponds to differentiating one of the $g^{(i)}(t)$ factors to produce $g^{(i+1)}(t)$:

$$D^{(k+1)}f(g(t)) \cdot \big[\dots, \underbrace{g^{(i)}(t), \dots, g^{(i)}(t)}_{b_i-1 \text{ times}}, \underbrace{g^{(i+1)}(t), \dots, g^{(i+1)}(t)}_{b_{i+1}+1 \text{ times}}, \dots\big].$$

In both cases, each partition of $\{1, \dots, m\}$ corresponds uniquely to a partition of $\{1, \dots, m+1\}$, and the combinatorial coefficients are preserved. Therefore the summation in (63) holds for $m+1$, completing the induction. □

For convenience, Faà di Bruno's formula can also be expressed in terms of Bell multi-tuple polynomials, as follows.

Lemma E.11 (Bell Multi-Tuple Representation of Faà di Bruno)**.** *For integers $m, k \geq 0$, define*

$$B_{m,k}(\Delta_1, \dots, \Delta_{m-k+1}) := \sum \frac{m!}{j_1!\, 1!^{j_1}\, j_2!\, 2!^{j_2} \cdots j_{m-k+1}!\, (m-k+1)!^{j_{m-k+1}}} \times \big[\underbrace{\Delta_1, \dots, \Delta_1}_{j_1 \text{ times}}, \dots, \underbrace{\Delta_{m-k+1}, \dots, \Delta_{m-k+1}}_{j_{m-k+1} \text{ times}}\big], \tag{64}$$

where each $\Delta_i \in \mathbb{R}^d$, and the sum runs over all tuples $(j_1, \dots, j_{m-k+1})$ of nonnegative integers satisfying

$$j_1 + j_2 + \cdots + j_{m-k+1} = k, \qquad j_1 + 2j_2 + \cdots + (m-k+1)j_{m-k+1} = m.$$

Then Faà di Bruno's formula for the composition $f \circ g$ can be written equivalently as

$$\frac{d^m}{dt^m} f(g(t)) = \sum_{k=0}^{m} D^{(k)}f(g(t)) \cdot B_{m,k}\Big(g^{(1)}(t),\, g^{(2)}(t),\, \dots,\, g^{(m-k+1)}(t)\Big). \tag{65}$$

E.3.2 Derivatives of the Lagrange Polynomial

We now provide conservative bounds on the n-th derivative of a Lagrange polynomial. Let $f = f(t)$ be the function of interest, and let P_t denote its Lagrange interpolating polynomial, which serves as an approximation of f. We consider $k+1$ interpolation nodes

$$t_0,\, t_1,\, \dots,\, t_k, \qquad t_j = t_0 + jh, \quad j = 0, \dots, k,$$

with constant step size $h > 0$.

In the Newton form (Stoer et al., 1980), the interpolating polynomial P_t can be expressed as

$$P_t \;=\; a_0 \;+\; \sum_{i=1}^{K} a_i \prod_{j=0}^{i-1}(t - t_j), \tag{66}$$

where the coefficients $a_i = [f(t_0), f(t_1), \ldots, f(t_i)]$ are the finite divided differences of f. These coefficients satisfy the recursive identity

$$[f(t_0), f(t_1), \ldots, f(t_i)] \;=\; \frac{[f(t_1), f(t_2), \ldots, f(t_i)] \;-\; [f(t_0), f(t_1), \ldots, f(t_{i-1})]}{t_i - t_0}. \tag{67}$$

Having expressed the interpolant in the Newton form, we now establish a few useful lemmas that will be instrumental in the proof. In particular, we next derive a more general recursive relation for finite divided differences.

Lemma E.12 (Order-reduction of divided differences)**.** *Let $t_r = t_0 + rh$ for $r = 0, 1, \ldots, k$ be equally spaced nodes with step size $h > 0$ and let $D_r^{(k)} = f[t_r, t_{r+1}, \ldots, t_{r+k}]$, be the k-th order divided difference starting at point $f(t_r)$. For $i \le k$, let $g_r = D_r^i = f[x_r, \ldots, x_{r+i}]$, be i-th order finite divided difference starting at point $f(x_r)$. Then for $m = k - i$,*

$$D_0^{(k)} = \frac{i!}{k! h^m} \sum_{j=0}^{m} (-1)^{m+j} \binom{m}{j} g_j. \tag{68}$$

This lemma expresses k-th order finite divided difference in terms of i-th order finite divided differences.

Proof. We will prove this by induction over m. For the base case $m = 1$, i.e. $i = k - 1$, we have

$$D_0^{(k)} = [f(t_0), f(t_1), \ldots, f(t_k)] = \frac{1}{kh} \left(-f[t_0, \ldots, t_{k-1}] + f[t_1, \ldots, t_k] \right). \tag{69}$$

This is true by the recursive relation (67).

Let us assume that the relation is true for $m = l$, that is, $i = k - l$. Then

$$\begin{aligned}
D_0^{(k)} &= \frac{(k-l)!}{k!h^l}\sum_{j=0}^{l}(-1)^{l+j}\binom{m}{j}D_j^{k-l} \\
&= \frac{(k-l)!}{k!h^l}\sum_{j=0}^{l}(-1)^{l+j}\binom{l}{j}\left(\frac{D_{j+1}^{k-l-1}-D_j^{k-l-1}}{(k-l)h}\right) \\
&= \frac{(k-l-1)!}{k!h^{l+1}}\sum_{j=0}^{l}(-1)^{l+j}\binom{l}{j}\left(D_{j+1}^{k-l-1}-D_j^{k-l-1}\right) \\
&= \frac{(k-l-1)!}{k!h^{l+1}}\left(\sum_{j=0}^{l}(-1)^{l+j}\binom{l}{j}D_{j+1}^{k-l-1}+\sum_{j=0}^{l}(-1)^{l+j+1}\binom{l}{j}D_j^{k-l-1}\right) \\
&= \frac{(k-l-1)!}{k!h^{l+1}}\left((-1)^{l+1}D_0^{k-l-1}+\sum_{j=1}^{l}(-1)^{l+j+1}D_j^{k-l-1}\left[\binom{l}{j+1}+\binom{l}{j}\right]+(-1)^{2l}D_{l+1}^{k-l-1}\right) \\
&= \frac{(k-l-1)!}{k!h^{l+1}}\left((-1)^{l+1}D_0^{k-l-1}+\sum_{j=1}^{l}(-1)^{l+j+1}\binom{l+1}{j}D_j^{k-l-1}+(-1)^{2l}D_{l+1}^{k-l-1}\right) \\
&= \frac{(k-l-1)!}{k!h^{l+1}}\sum_{j=0}^{l+1}(-1)^{l+1+j}\binom{l+1}{j}D_j^{k-l-1}.
\end{aligned}$$

Thus, the expression also holds for $m = l + 1$, which completes the proof. □

To analyze the derivatives of the Newton form, we first record following useful identity.

Lemma E.13.

$$\frac{d^n}{dt^n}\prod_{j=0}^{i-1}(t-t_j) = \begin{cases} 0, & \text{if } n > i, \\ n!, & \text{if } n = i, \\ n!\displaystyle\sum_{0\le j_1<\cdots<j_n\le i-1}\ \prod_{\substack{j=0 \\ j\notin\{j_1,\ldots,j_n\}}}^{i-1}(t-t_j), & \text{if } n < i. \end{cases} \tag{70}$$

Consequently, for $(k+1)$ equally spaced points $t_0, \ldots, t_k$, with difference h, if $t \in [t_0, t_k]$, we have

$$\left|\frac{d^n}{dt^n}\prod_{j=0}^{i-1}(t-t_j)\right| \le \begin{cases} 0, & \text{if } n > i, \\ n!, & \text{if } n = i, \\ n!k^{i-n}h^{i-n}\binom{i}{n}, & \text{if } n < i. \end{cases} \tag{71}$$

Proof. We apply the general Leibniz rule for the n-th derivative of a product:

$$\frac{d^n}{dt^n}\left(\prod_{j=0}^{i-1}(t-t_j)\right) = \sum_{\substack{k_0+\cdots+k_{i-1}=n \\ k_j\geq 0}} \binom{n}{k_0,\ldots,k_{i-1}} \prod_{j=0}^{i-1} \frac{d^{k_j}}{dt^{k_j}}(t-t_j).$$

Note:

$$\frac{d^{k_j}}{dt^{k_j}}(t-t_j) = \begin{cases} (t-t_j), & \text{if } k_j = 0, \\ 1, & \text{if } k_j = 1, \\ 0, & \text{if } k_j \geq 2. \end{cases}$$

Only terms with $k_j \in \{0,1\}$ contribute, and the sum is over all subsets of size n from $\{0,\ldots,i-1\}$, with those positions differentiated once and the rest undifferentiated. Therefore:

$$\frac{d^n}{dt^n}\prod_{j=0}^{i-1}(t-t_j) = n! \sum_{0\leq j_1<\cdots<j_n\leq i-1} \prod_{\substack{j=0 \\ j\notin\{j_1,\ldots,j_n\}}}^{i-1} (t-t_j).$$

This completes the proof.

□

Lemma E.14 (Mean Value Theorem: Divided Differences)**.** *For any $i+1$ pairwise distinct points $t_0,\ldots,t_i$ in the domain of an i-times differentiable function f, there exists an interior point*

$$\xi \in (\min\{t_0,\ldots,t_i\}, \max\{t_0,\ldots,t_i\})$$

such that:

$$f[t_0,\ldots,t_i] = \frac{f^{(i)}(\xi)}{i!}. \tag{72}$$

Finally, we derive bounds for the derivatives of the Lagrange polynomial:

Lemma E.15. *The n-th derivative of the Lagrange interpolating polynomial, constructed on $(k+1)$ equally spaced nodes $t_0,\ldots,t_k$ with spacing h, admits the following bound in terms of the n-th derivative of the underlying function f:*

$$\|P_t^{(n)}\|_2 \leq C_{p,n} \sup_{t\in[t_0,t_k]} \|f^{(n)}(t)\|_2 \tag{73}$$

where $C_{p,n}$ is a constant only depending on k,n.

Proof. Taking the n-th derivative of the Newton form (66) gives

$$P_t^{(n)} = \sum_{i=0}^{k-n} a_{n+i} \frac{d^n}{dt^n} \prod_{j=0}^{n+i-1} (t-t_j). \tag{74}$$

Here a_{n+i} is the $(n+i)$-th order divided difference of f. To express this in terms of n-th order divided differences, we apply Lemma (E.12), yielding

$$\|P_t^{(n)}\|_2 \le \sum_{i=0}^{k-n} \left(\frac{n!}{(n+i)!\, h^i} \sum_{j=0}^{i} \binom{i}{j} \|D_j^{(n)}\|_2 \right) \left| \frac{d^n}{dt^n} \prod_{j=0}^{n+i-1} (t - t_j) \right|. \tag{75}$$

By the mean value theorem for divided differences (Lemma (E.14)), each term $\|D_j^{(n)}\|$ is bounded by

$$M_n := \sup_{t \in [t_0, t_k]} \|f^{(n)}(t)\|.$$

Substituting this bound and invoking Lemma (E.13), we obtain

$$\begin{aligned} \|P_t^{(n)}\| &\le \sum_{i=0}^{k-n} \left(\frac{n!}{(n+i)!\, h^i} \sum_{j=0}^{i} \binom{i}{j} M_n \right) n!\, k^i h^i \\ &\le M_n \left((n!)^2 \sum_{i=0}^{k-n} \frac{(2k)^i}{(n+i)!} \right). \end{aligned}$$

This completes the proof. □

E.3.3 k-th Derivative of $\widehat{X}_1^*$

We next establish a bound on the derivatives of the fixed-point trajectory $\widehat{X}_1^*$ within a single interpolation interval.

Lemma E.16. *Let $I \subset \mathbb{R}$ be an interval on which the Lagrange interpolant $P(t; \widehat{X}_1^*)$ of $\widehat{X}_1^*$ is constructed (using any fixed set of nodes in I). Fix an integer $2 \le k \le K-1$. Then*

$$\left\| \frac{d^k \widehat{X}_1^*}{dt^k} \right\| \;\le\; 2^{\,k-2} \gamma^{k-1} \|\widehat{X}^*(t)\| + \gamma^{k-1} \sum_{i=0}^{k-2} C_{p,i} \sup_{t \in I} \left\| \frac{d^i}{dt^i} \nabla U\big(\widehat{X}_1^*\big) \right\|, \tag{76}$$

where the constants $C_{p,i} > 0$ depend only on the number of interpolation nodes and the derivative order i.

Proof. We proceed in four steps.

Step 1: Parity decomposition. Write $P_t := P(t; \widehat{X}_1^*)$ for the Lagrange interpolant built from $\widehat{X}_1^*$. Direct differentiation of the chain $\{\widehat{X}_i^*\}_{i \ge 1}$ (See Lemma F.3) yields the initial identities

$$\left\| \frac{d\widehat{X}_1^*}{dt} \right\| = \left\| \widehat{X}_2^* \right\|, \qquad \left\| \frac{d^2 \widehat{X}_1^*}{dt^2} \right\| \le \gamma \left(\|P_t\| + \left\| \widehat{X}_3^* \right\| \right),$$

$$\left\|\frac{d^3\widehat{X}_1^*}{dt^3}\right\| \le \gamma^2\left(\left\|\frac{d}{dt}P_t\right\| + \left\|\widehat{X}_2^*\right\| + \left\|\widehat{X}_4^*\right\|\right), \qquad \left\|\frac{d^4\widehat{X}_1^*}{dt^4}\right\| \le \gamma^3\left(\left\|\frac{d^2}{dt^2}P_t\right\| + \|P_t\| + 2\left\|\widehat{X}_3^*\right\| + \left\|\widehat{X}_5^*\right\|\right),$$

$$\left\|\frac{d^5\widehat{X}_1^*}{dt^5}\right\| \le \gamma^4\left(\left\|\frac{d^3}{dt^3}P_t\right\| + \left\|\frac{d}{dt}P_t\right\| + 2\left\|\widehat{X}_2^*\right\| + 3\left\|\widehat{X}_4^*\right\| + \left\|\widehat{X}_6^*\right\|\right),$$

$$\left\|\frac{d^6\widehat{X}_1^*}{dt^6}\right\| \le \gamma^5\left(\left\|\frac{d^4}{dt^4}P_t\right\| + \left\|\frac{d^2}{dt^2}P_t\right\| + \|P_t\| + 5\left\|\widehat{X}_3^*\right\| + 4\left\|\widehat{X}_5^*\right\| + \left\|\widehat{X}_7^*\right\|\right), \dots$$

Ignoring the factor of γ for the moment, one can verify by induction on k that the expansions naturally split according to parity: for even indices $k = 2r$,

$$\left\|\frac{d^{2r}}{dt^{2r}}\widehat{X}_1^*(t)\right\| \lesssim \sum_{s=1}^{r} a_{2r,2s+1}\left\|\widehat{X}_{2s+1}^*(t)\right\| + \sum_{i=0}^{r-1}\left\|\frac{d^{2i}}{dt^{2i}}P_t\right\|, \tag{77}$$

and for $k = 2r+1$,

$$\left\|\frac{d^{2r+1}}{dt^{2r+1}}\widehat{X}_1^*(t)\right\| \lesssim \sum_{s=1}^{r+1} a_{2r+1,2s}\left\|\widehat{X}_{2s}^*(t)\right\| + \sum_{i=0}^{r-1}\left\|\frac{d^{2i+1}}{dt^{2i+1}}P_t\right\|, \tag{78}$$

with integer coefficients $a_{k,j}, b_{k,i}$ depending only on k.

Step 2: Coefficient relations and boundary terms. From the recursion of the chain (and the antisymmetry of the matrix Q) we have, for $r \ge 1$,

$$\begin{aligned} &a_{2r+1,2s} = a_{2r,2s-1} + a_{2r,2s+1}, \qquad a_{2r+1,2} = a_{2r,3}, \qquad a_{2r+1,2r+2} = 1, \\ &a_{2r,2s+1} = a_{2r-1,2s} + a_{2r-1,2s+2}, \qquad a_{2r,2r+1} = 1, \end{aligned} \tag{79}$$

and $a_{1,2} = a_{2,3} = 1$, with $a_{k,j} = 0$ outside their index ranges. Define the parity–sums

$$E_r := \sum_{s=1}^{r} a_{2r,2s+1}, \qquad O_r := \sum_{s=1}^{r+1} a_{2r+1,2s}.$$

Step 3: Summation identities $O_r = 2E_r$ and $E_r \le 2O_{r-1}$. Using (79),

$$\begin{aligned} O_r &= a_{2r+1,2} + \sum_{s=2}^{r} a_{2r+1,2s} + a_{2r+1,2r+2} \\ &= a_{2r,3} + \sum_{s=2}^{r}\left(a_{2r,2s-1} + a_{2r,2s+1}\right) + a_{2r,2r+1} \\ &= \sum_{u=1}^{r} a_{2r,2u+1} + \sum_{u=1}^{r} a_{2r,2u+1} = 2E_r, \end{aligned}$$

so $O_r = 2E_r$. For E_r,

$$\begin{aligned} E_r &= \sum_{s=1}^{r-1} a_{2r,2s+1} + a_{2r,2r+1} \\ &= \sum_{s=1}^{r-1} \left(a_{2r-1,\,2s} + a_{2r-1,\,2s+2}\right) + a_{2r-1,2r} \\ &= \sum_{s=1}^{r-1} a_{2r-1,\,2s} + \sum_{s=1}^{r-1} a_{2r-1,\,2s+2} + a_{2r-1,2r} \\ &= O_{r-1} + O_{r-1} - a_{2r-1,2} \\ &\le 2O_{r-1}. \end{aligned}$$

Finally, $O_0 = a_{1,2} = 1$ and $E_1 = a_{2,3} = 1$.

Step 4: Growth bounds and the derivative estimate. From $O_r = 2E_r$ and $E_r \le 2O_{r-1}$ we obtain

$$O_r \le 4\,O_{r-1}, \qquad E_r \le 4\,E_{r-1} \qquad (r \ge 1),$$

whence by iteration

$$E_r \le 4^{\,r-1} = 2^{2r-2}, \qquad O_r \le 2 \cdot 4^{\,r-1} = 2^{2r-1} \qquad (r \ge 1).$$

Let $A_k := \sum_{j=2}^{k+1} a_{k,j}$, so $A_{2r} = E_r$ and $A_{2r+1} = O_r$. Then, for all $k \ge 2$,

$$A_k \le 2^{\,k-2}.$$

Then the triangle inequality applied to (77)–(78) yields, for every $2 \le k \le K-1$,

$$\left\| \tfrac{d^k}{dt^k} \widehat{X}_1^*(t) \right\| \;\le\; \gamma^{k-1} \left(A_k \max_{2 \le j \le k+1} \|\widehat{X}_j^*(t)\| \;+\; \sum_{i=0}^{k-2} \left\| \tfrac{d^{\,i}}{dt^{\,i}} P_t \right\| \right) \;\le\; \gamma^{k-1} \left(2^{\,k-2}\, \|\widehat{X}^*(t)\| \;+\; \sum_{i=0}^{k-2} \left\| \tfrac{d^{\,i}}{dt^{\,i}} P_t \right\| \right).$$

Finally, for $t \in I$, Lemma (E.15) implies

$$\left\| \frac{d^k \widehat{X}_1^*}{dt^k} \right\| \;\le\; 2^{\,k-2} \gamma^{k-1}\, \|\widehat{X}^*(t)\| + \gamma^{k-1} \sum_{i=0}^{k-2} C_{p,i} \sup_{t \in I} \left\| \frac{d^i}{dt^i} \nabla U\big(\widehat{X}_1^*\big) \right\|. \tag{80}$$

□

F Technical Details

F.1 Existence of Fixed Point

In this subsection, we provide the proofs of the existence of fixed point for operators $\mathcal{T}_y$ and $\widehat{\mathcal{T}}_y$ as defined in Section 6.1 (or see (10), (11)).

F.1.1 Operator $\mathcal{T}_y$

Lemma F.1 (Existence and uniqueness of the fixed point)**.** *Under Assumption 1, let* $L_H := \max\{L, 1\}$ *be the Lipschitz constant of* ∇H*. There exists*

$$h_* \;:=\; \frac{1}{3\,\gamma\,L_H} \;>\; 0,$$

such that for any $0 < h < h_*$ *and any* $y \in \mathbb{R}^{Kd}$*, the operator* $\mathcal{T}_y$ *admits a unique fixed point in* $\mathcal{C}\big([0,h], \mathbb{R}^{Kd}\big)$*.*

Proof. Work on the complete metric space $\big(\mathcal{C}([0,h], \mathbb{R}^{Kd}), \|\cdot\|_\infty\big)$ with $\|X\|_\infty := \sup_{t\in[0,h]} \|X(t)\|$. Let $X, Y \in \mathcal{C}([0,h], \mathbb{R}^{Kd})$ and set $\Delta(t) := (\mathcal{T}_y[X])(t) - (\mathcal{T}_y[Y])(t)$. The additive noise cancels, so

$$\Delta(t) = -\int_0^t (D+Q)\Big(\nabla H(X(s)) - \nabla H(Y(s))\Big)\, ds.$$

Taking norms and using the Lipschitzness of ∇H,

$$\|\Delta(t)\| \le \int_0^t \|D+Q\|\,\|\nabla H(X(s)) - \nabla H(Y(s))\|\, ds \le \|D+Q\|\, L_H \int_0^t \|X(s) - Y(s)\|\, ds.$$

Thus, for $t \in [0,h]$,

$$\|\Delta(t)\| \;\le\; \|D+Q\|\, L_H\, t\, \|X - Y\|_\infty \;\le\; 3\,\gamma\, L_H\, h\, \|X - Y\|_\infty,$$

where in the ast inequality we use $\|D+Q\| \le 3\gamma$ in Lemma G.1. Taking the supremum over $t \in [0,h]$ yields the contraction

$$\|\mathcal{T}_y[X] - \mathcal{T}_y[Y]\|_\infty \;\le\; \rho\, \|X - Y\|_\infty, \qquad \rho := 3\,\gamma\, L_H\, h.$$

If $h < h_* = (3\,\gamma L_H)^{-1}$, then $\rho < 1$, so $\mathcal{T}_y$ is a contraction. By Banach's fixed point theorem, $\mathcal{T}_y$ has a unique fixed point in $\mathcal{C}([0,h], \mathbb{R}^{Kd})$. □

F.1.2 Operator $\widehat{\mathcal{T}}_y$

For the Lagrange basis, we use the standard Lebesgue constant

$$\Gamma_\phi \;:=\; \sup_{\tau\in[0,1]} \sum_{j=1}^{M} \big|\ell_j(\tau)\big|.$$

Lemma G.3 shows that $\Gamma_\phi \le \frac{2^{M-1}\,(M-1)^{M-1}}{(M-1)!}$ which only depends on the number of collocation nodes M. Then we have the following lemma:

Lemma F.2. *Under Assumption 1 and let $\|\cdot\|_\infty := \sup_{t\in[0,h]} \|\cdot(t)\|$ be the uniform norm. Then, for*

$$h \;\le\; \min\!\left\{\frac{\log 2}{3\gamma},\; \frac{1}{4\,L\,\Gamma_\phi}\right\},$$

the operator $\widehat{\mathcal{T}}_y$ is a contraction on $\mathcal{C}\big([0,h],\mathbb{R}^{Kd}\big)$ with respect to $\|\cdot\|_\infty$.

Proof. Let $\tilde{X} = \widehat{\mathcal{T}}_y[X]$ and $\tilde{Y} = \widehat{\mathcal{T}}_y[Y]$. Let $A := -(D+Q)J$ with $J := \mathrm{diag}(0,1,\dots,1)\otimes I_d$ and $e_2 = (0,1,0,\dots,0)^\top \in \mathbb{R}^K$, then from (13),

$$\big(\widehat{\mathcal{T}}_y[X]\big)(t) = \tilde{X}(t) = y + \int_0^t A\,\tilde{X}(s)\,ds \;-\; \int_0^t (e_2\otimes I_d)\,P(s;X)\,ds \;+\; \int_0^t \sqrt{2D}\,dB_s. \tag{81}$$

Since $\|e_2\otimes I_d\| = 1$,

$$\|\tilde{X}(t)-\tilde{Y}(t)\| \le \int_0^t \|A\|\,\|\tilde{X}(s)-\tilde{Y}(s)\|\,ds \;+\; \int_0^t \|P(s;X)-P(s;Y)\|\,ds.$$

By Lipschitzness of ∇U and the definition of the Lagrange iterpolation P in (12),

$$\|P(s;X)-P(s;Y)\| \;\le\; L\sum_{j=1}^{M}\big|\ell_j(s)\big|\,\|X_1-Y_1\|_\infty \;\le\; L\,\Gamma_\phi\,\|X-Y\|_\infty,$$

and hence

$$\|\tilde{X}(t)-\tilde{Y}(t)\| \;\le\; \int_0^t \|A\|\,\|\tilde{X}(s)-\tilde{Y}(s)\|\,ds \;+\; L\,\Gamma_\phi\,t\,\|X-Y\|_\infty.$$

Applying Grönwall's inequality, taking $t\le h$ and applying Lemma G.1,

$$\|\tilde{X}(t)-\tilde{Y}(t)\| \;\le\; e^{\|A\|t}\,L\,\Gamma_\phi\,t\,\|X-Y\|_\infty \;\le\; e^{3\gamma h}\,L\,\Gamma_\phi\,h\,\|X-Y\|_\infty.$$

Choose $h\le\frac{\log 2}{3\gamma}$ so that $e^{3\gamma h}\le 2$, and $h\le(4L\Gamma_\phi)^{-1}$ so that the product is at most $1/2$. Taking the supremum over $t\in[0,h]$ yields

$$\|\widehat{\mathcal{T}}_y[X]-\widehat{\mathcal{T}}_y[Y]\|_\infty \;\le\; \tfrac{1}{2}\,\|X-Y\|_\infty,$$

which proves the contraction claim. □

The following lemma identifies the unique fixed point of $\widehat{\mathcal{T}}_y$ (proof of existence in Lemma F.2) with the unique solution of the interpolated SDE on $[0,h]$. Moreover, the Picard iterates generated by $\widehat{\mathcal{T}}_y$ converge uniformly on $[0,h]$ to this fixed point.

Lemma F.3. *Assume the stepsize condition of Lemma F.2 so that $\widehat{\mathcal{T}}_y : \mathcal{C}([0,h];\mathbb{R}^{Kd}) \to \mathcal{C}([0,h];\mathbb{R}^{Kd})$ is a contraction in $\|\cdot\|_\infty$. Let $\widehat{X}_y^*$ denote its unique fixed point. Then $\widehat{X}_y^*$ is the unique solution on $[0,h]$ of*

$$d\widehat{X}(t) = A\,\widehat{X}(t)\,dt - (e_2 \otimes I_d)\,P\left(t;\widehat{X}\right)dt + \sqrt{2D}\,dB_t, \qquad \widehat{X}(0) = y, \tag{82}$$

equivalently, for all $t \in [0,h]$,

$$\widehat{X}_y^*(t) = y + \int_0^t A\,\widehat{X}_y^*(s)\,ds - \int_0^t (e_2 \otimes I_d)\,P\left(s;\widehat{X}_y^*\right)ds + \int_0^t \sqrt{2D}\,dB_s. \tag{83}$$

Proof. Define the Picard sequence

$$\widehat{X}^{[0]}(t) \equiv y, \qquad \widehat{X}^{[\nu+1]} := \widehat{\mathcal{T}}_y[\widehat{X}^{[\nu]}], \quad \nu \geq 0.$$

By the definition of $\widehat{\mathcal{T}}_y$, each $\widehat{X}^{[\nu+1]}$ satisfies, for all $t \in [0,h]$,

$$\widehat{X}^{[\nu+1]}(t) = y + \int_0^t A\,\widehat{X}^{[\nu+1]}(s)\,ds - \int_0^t (e_2 \otimes I_d)\,P\Big(s;\widehat{X}^{[\nu]}\Big)\,ds + \int_0^t \sqrt{2D}\,dB_s. \tag{84}$$

By Lemma F.2 there is $\rho < 1/2$ such that

$$\|\widehat{X}^{[\nu+1]} - \widehat{X}^{[\nu]}\|_\infty = \|\widehat{\mathcal{T}}_y[\widehat{X}^{[\nu]}] - \widehat{\mathcal{T}}_y[\widehat{X}^{[\nu-1]}]\|_\infty \leq \rho\,\|\widehat{X}^{[\nu]} - \widehat{X}^{[\nu-1]}\|_\infty.$$

Thus $\{\widehat{X}^{[\nu]}\}_{\nu\geq 0}$ is Cauchy sequence in $\mathcal{C}([0,h];\mathbb{R}^{Kd})$ and converges uniformly to some $\widehat{X}_y^*$. By completeness and the Banach fixed point theorem, $\widehat{X}_y^*$ is the unique fixed point of $\widehat{\mathcal{T}}_y$.

To identify the limiting equation, fix $t \in [0,h]$ and pass to the limit $\nu \to \infty$ in (84). For the linear drift we use uniform convergence and linearity:

$$\sup_{u\leq t}\left\| \int_0^u A\big(\widehat{X}^{[\nu]}(s) - \widehat{X}_y^*(s)\big)\,ds\right\| \leq t\,\|A\|\,\|\widehat{X}^{[\nu]} - \widehat{X}_y^*\|_\infty \xrightarrow[\nu\to\infty]{} 0.$$

For the interpolated term we use the Lipschitz property of P from Lemma F.2: there exists $L_P < \infty$ such that $\|P(s;X) - P(s;Y)\| \leq L_P\|X - Y\|_\infty$ for all s and X, Y, hence

$$\sup_{u\leq t}\left\| \int_0^u (e_2 \otimes I_d)\Big(P(s;\widehat{X}^{[\nu]}) - P(s;\widehat{X}_y^*)\Big)ds\right\| \leq t\,L_P\,\|\widehat{X}^{[\nu]} - \widehat{X}_y^*\|_\infty \xrightarrow[\nu\to\infty]{} 0.$$

The stochastic integral $\int_0^t \sqrt{2D}\,dB_s$ is independent of ν and remains unchanged. Taking limits in (84) yields (83). Hence $\widehat{X}_y^*$ is a solution on $[0,h]$.

For uniqueness, let Z be any strong solution of (82). Then Z satisfies $Z = \widehat{\mathcal{T}}_y[Z]$, so Z is a fixed point of the contraction $\widehat{\mathcal{T}}_y$. By uniqueness of fixed points, $Z = \widehat{X}_y^*$ a.s. □

F.2 Alternative Representation of Algorithm 1

Algorithm 1 can be interpreted as performing Picard iterations of the discretized operator $\widehat{\mathcal{T}}_y$ on the interval $[0,h]$, initialized at $y=\widehat{X}(nh)$. The next lemma formalizes this equivalence by showing that the outputs of Algorithm 1 at Picard index ν coincide with the ν-th Picard iterates of $\widehat{\mathcal{T}}_y$ evaluated at the nodes $\{c_k h\}_{k=1}^M$.

Lemma F.4 (Equivalence of Algorithm 1 and Picard iterates)**.** *Fix $h>0$, and nodes $0<c_1<\cdots<c_M=1$. For a fixed time step n, let $\widehat{X}^{[\nu]}(nh+c_kh)$ denote the stage iterates produced by Algorithm 1 at Picard index ν and stage $k\in\{1,\dots,M\}$, initialized by $\widehat{X}^{[0]}(nh+c_kh)=\widehat{X}(nh)$ for all k. Let an initializer $y=\widehat{X}(nh)\in\mathbb{R}^{Kd}$ and let $\widehat{\mathcal{T}}_y$ be the discretized operator defined in Section 6.1 (or see (11)), and define the Picard sequence on $[0,h]$ by*

$$\widehat{Y}_y^{[0]}(t)\equiv y,\quad \widehat{Y}_y^{[\nu+1]}:=\widehat{\mathcal{T}}_y\big[\widehat{X}_y^{[\nu]}\big],\quad \nu\ge 0.$$

Then, for every $\nu\ge 0$ and every $k=1,\dots,M$,

$$\widehat{Y}_y^{[\nu]}(c_kh)\;=\;\widehat{X}^{[\nu]}(nh+c_kh). \tag{85}$$

In particular, with $c_M=1$,

$$\widehat{Y}_y^{[\nu_*]}(h)\;=\;\widehat{X}^{[\nu_*]}(nh+h)\;=\;\widehat{X}\big((n+1)h\big).$$

Proof. For any path $X\in\mathcal{C}([0,h],\mathbb{R}^{Kd})$, define the interpolant $P(s;X)=\sum_{j=1}^M \ell_j(s/h)\,\nabla U\big(X_1(c_jh)\big)$ as in (12). The decomposition in (13) shows that $\tilde{X}:=\mathcal{T}_{\phi,y}[X]$ solves, on $[0,h]$,

$$d\tilde{X}(t)=A\,\tilde{X}(t)\,dt\;-\;(e_2\otimes I_d)\,P(t;X)\,dt\;+\;\sqrt{2D}\,dB_{nh+t},\qquad \tilde{X}(0)=y.$$

By variation of constants, it satisfies

$$\tilde{X}(t)=e^{tA}y\;-\;\int_0^t e^{(t-s)A}\,(e_2\otimes I_d)\,P(s;X)\,ds\;+\;W_n(t/h),\qquad t\in[0,h]. \tag{86}$$

Set $t=c_kh$. Using $P(s;X)=\sum_j \ell_j(s/h)\,\nabla U(X_1(c_jh))$ and the definition

$$\alpha_j(\tau,h)\;:=\;\int_0^\tau e^{(\tau-\sigma)hA}\,\ell_j(\sigma)\,d\sigma,\qquad \theta\in[0,1],$$

we obtain from (86) that

$$\big(\widehat{\mathcal{T}}_y[X]\big)(c_kh)=e^{c_khA}y\;-\;h\sum_{j=1}^M \alpha_j(c_k,h)\,(e_2\otimes I_d)\,\nabla U\big(X_1(c_jh)\big)\;+\;W_n(c_k). \tag{87}$$

Recall the decomposition of the dynamics $dX=AX\,dt+g(X)\,dt+\sqrt{2D}\,dB_t$ with $g(X)=-(e_2\otimes$

$I_d)\,\nabla U(X_1)$. Then (87) is equivalently

$$\big(\widehat{\mathcal{T}}_y[X]\big)(c_k h) = e^{c_k h A} y \;+\; h \sum_{j=1}^{M} \alpha_j(c_k, h)\, g\big(X(c_j h)\big) \;+\; W_n(c_k). \tag{88}$$

Next we prove (85) by induction. Base case $\nu = 0$: by definition, $\widehat{Y}_y^{[0]}(c_k h) \equiv \widehat{X}^{[0]}(nh + c_k h) \equiv y$. Assume $\widehat{Y}_y^{[\nu]}(c_j h) = \widehat{X}^{[\nu]}(nh + c_j h)$ for all j. Apply (88) with $X = \widehat{Y}_y^{[\nu]}$:

$$\widehat{Y}_y^{[\nu+1]}(c_k h) = e^{c_k h A} y \;+\; h \sum_{j=1}^{M} \alpha_j(c_k, h)\, g\big(\widehat{Y}_y^{[\nu]}(c_j h)\big) \;+\; W_n(c_k).$$

By the inductive hypothesis, $g(\widehat{Y}_y^{[\nu]}(c_j h)) = g(\widehat{X}^{[\nu]}(nh + c_j h))$, hence

$$\widehat{Y}_y^{[\nu+1]}(c_k h) = e^{c_k h A} y \;+\; h \sum_{j=1}^{M} \alpha_j(c_k, h)\, g\big(\widehat{X}^{[\nu]}(nh + c_j h)\big) \;+\; W_n(c_k).$$

This is exactly the update in Algorithm 1 (Line 9), so $\widehat{Y}_y^{[\nu+1]}(c_k h) = \widehat{X}^{[\nu+1]}(nh + c_j h)$. This completes the induction. Taking $k = M$ with $c_M = 1$ gives $\widehat{Y}_y^{[\nu_*]}(h) = \widehat{X}^{[\nu_*]}(nh + h) = \widehat{X}((n+1)h)$. □

G Auxiliary Lemmas

Lemma G.1. *Let D, Q be as defined in (15) with $\gamma > 0$, and let $K \geq 2$. Let $\|\cdot\|$ denote the operator norm (induced by Euclidean norm) for a matrix. Then*

$$\sqrt{1+\gamma^2} \;\leq\; \|D+Q\| \;\leq\; \|D\| + \|Q\| \;\leq\; \gamma + \max\{\,1+\gamma,\; 2\gamma\,\} \;=\; \max\{\,1+2\gamma,\; 3\gamma\,\},$$

and, moreover,

$$\|(D+Q)\,\mathrm{diag}(0, 1, \ldots, 1) \otimes I_d\| \;\geq\; \sqrt{1+\gamma^2} \quad (\textit{with equality when } K = 2).$$

Proof. By construction $D = \mathrm{diag}(0, \ldots, 0, \gamma) \otimes I_d$, hence $\|D\| = \gamma$. Also $Q = T_\gamma \otimes I_d$, where $T_\gamma \in \mathbb{R}^{K \times K}$ is the (skew–symmetric) tridiagonal matrix

$$T_\gamma = \begin{pmatrix} 0 & -1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & -\gamma & \ddots & \vdots & \vdots \\ 0 & \gamma & 0 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & -\gamma & 0 \\ 0 & \vdots & 0 & \gamma & 0 & -\gamma \\ 0 & \vdots & 0 & \cdots & \gamma & 0 \end{pmatrix}.$$

Kronecker structure gives $\|Q\| = \|T_\gamma\|$. For any matrix A, $\|A\| \le \sqrt{\|A\|_1 \|A\|_\infty}$; we now compute $\|T_\gamma\|_1$ and $\|T_\gamma\|_\infty$. Calculating the row sums (in absolute value) of T_γ yields: $\|T_\gamma\|_\infty = \max\{1+\gamma,\, 2\gamma\}$. By the same pattern for column sums, $\|T_\gamma\|_1 = \max\{1+\gamma,\, 2\gamma\}$. Therefore

$$\|Q\| = \|T_\gamma\| \;\le\; \sqrt{\|T_\gamma\|_1 \|T_\gamma\|_\infty} \;=\; \max\{1+\gamma,\, 2\gamma\}.$$

Finally, by the triangle inequality,

$$\|D+Q\| \;\le\; \|D\| + \|Q\| \;\le\; \gamma + \max\{1+\gamma,\, 2\gamma\} \;=\; \max\{1+2\gamma,\, 3\gamma\}.$$

For the lower bounds, note first that $\|B \otimes I_d\| = \|B\|$. Write $D+Q = (\tilde{D}+\tilde{Q}) \otimes I_d$ with $\tilde{D} = \mathrm{diag}(0, \ldots, 0, \gamma)$ and $\tilde{Q} = T_\gamma$. For any matrix M, $\|M\| \ge \max_j \|Me_j\|_2$. Reading off the second column of $\tilde{D}+\tilde{Q}$ gives

$$(\tilde{D}+\tilde{Q})e_2 = \begin{cases} -e_1 + \gamma e_2, & K = 2, \\ -e_1 + \gamma e_3, & K \ge 3, \end{cases}$$

hence $\|(\tilde{D}+\tilde{Q})e_2\|_2 = \sqrt{1+\gamma^2}$ for all $K \ge 2$, and therefore $\|D+Q\| \ge \sqrt{1+\gamma^2}$.

Let $P := \mathrm{diag}(0, 1, \ldots, 1)$. Since $\|(D+Q)P \otimes I_d\| = \|(\tilde{D}+\tilde{Q})P\|$ and P leaves column 2 unchanged, the same column calculation yields

$$\|(D+Q)P \otimes I_d\| = \|(\tilde{D}+\tilde{Q})P\| \;\ge\; \|(\tilde{D}+\tilde{Q})Pe_2\|_2 = \sqrt{1+\gamma^2}.$$

□

Lemma G.2. *Let $b(x) = -(D+Q)\,\nabla H(x)$. Assume*

$$mI \;\preceq\; \nabla^2 U(x) \;\preceq\; LI \quad \textit{for all } x \in \mathbb{R}^d,$$

for $0 < m \le L < \infty$. Then, for all x,

$$\|J_b(x)\| \;\le\; L \max\{\,1+2\gamma,\; 3\gamma\,\}, \qquad \textit{where}\;\; J_b(x) := \nabla b(x) = -(D+Q)\,\nabla^2 H(x).$$

Proof. Taking operator norms and using Lemma G.1 yields

$$\|J_b(x)\| \;=\; \|(D+Q)\,\nabla^2 H(x)\| \;\le\; \|D+Q\|\,\|\nabla^2 H(x)\| \;\le\; L \max\{\,1+2\gamma,\; 3\gamma\,\},$$

since $\|\nabla^2 U(x)\|_{\mathrm{op}} \le L$ by assumption. This proves the upper bound. □

Lemma G.3. *Let $M \ge 2$ and take uniform nodes $c_j = (j-1)/(M-1)$ for $j = 1, \ldots, M$ on $[0,1]$. Let $\{\ell_j\}_{j=1}^M$ be the associated Lagrange basis and $\Gamma_\phi := \sup_{\tau \in [0,1]} \sum_{j=1}^M |\ell_j(\tau)|$ the Lebesgue constant. Then*

$$\Gamma_\phi \;\le\; \frac{2^{M-1}\,(M-1)^{M-1}}{(M-1)!}.$$

Proof. Fix $\tau \in [0,1]$ and $j \in \{1,\dots,M\}$. Using the product form,

$$|\ell_j(\tau)| = \prod_{\substack{k=1\\k\neq j}}^{M} \frac{|\tau - c_k|}{|c_j - c_k|} \le \prod_{\substack{k=1\\k\neq j}}^{M} \frac{1}{|c_j - c_k|}.$$

For the uniform grid, $|c_j - c_k| = |j-k|/(M-1)$, so

$$\prod_{k\neq j} |c_j - c_k| = \Big(\frac{1}{(M-1)}\Big)^{M-1} \prod_{k\neq j} |j-k| = \Big(\frac{1}{(M-1)}\Big)^{M-1} (j-1)!\,(M-j)!.$$

Hence

$$|\ell_j(\tau)| \le \frac{(M-1)^{M-1}}{(j-1)!\,(M-j)!}.$$

Summing over j and using $\sum_{j=1}^{M} \binom{M-1}{j-1} = 2^{M-1}$,

$$\sum_{j=1}^{M} |\ell_j(\tau)| \le (M-1)^{M-1} \sum_{j=1}^{M} \frac{1}{(j-1)!\,(M-j)!} = \frac{(M-1)^{M-1}}{(M-1)!} \sum_{j=1}^{M} \binom{M-1}{j-1} = \frac{2^{M-1}\,(M-1)^{M-1}}{(M-1)!}.$$

Taking the supremum over $\tau \in [0,1]$ gives the claim. □

Lemma G.4 (Stationary moment bounds)**.** *Let $H(x) = U(x_1) + \frac{1}{2}\sum_{k=2}^{K} \|x_k\|^2$ and $\rho(dx) \propto e^{-H(x)}dx$ be the stationary law of the dynamics, with U satisfying Assumption 1. Let $x_\star \in \arg\min U$ and $X = (X_1,\dots,X_K) \in \mathbb{R}^{Kd}$. Then:*

$$\mathbb{E}_\rho\big[\|X_1 - x_\star\|^2\big] \le \frac{d}{m}, \tag{89}$$

$$\mathbb{E}_\rho\big[\|\nabla U(X_1)\|^2\big] \le L^2\,\mathbb{E}_\rho\big[\|X_1 - x_\star\|^2\big] \;\le\; \frac{L^2}{m}\,d, \tag{90}$$

$$\mathbb{E}_\rho\big[\|X\|^2\big] = \mathbb{E}_\rho\big[\|X_1\|^2\big] + \sum_{k=2}^{K} \mathbb{E}_\rho\big[\|X_k\|^2\big] \;\le\; \big(2\|x_\star\|^2 + \tfrac{2d}{m}\big) \;+\; (K-1)\,d. \tag{91}$$

Proof. **Bound for $\mathbb{E}\|X_1 - x_\star\|^2$.** Write $\pi_U(dx_1) \propto e^{-U(x_1)}dx_1$ for the marginal stationary law of X_1; under ρ we have the factorization $\rho(dx) = \pi_U(dx_1) \otimes \big(\otimes_{k=2}^{K} \mathcal{N}(0, I_d)\big)$. Because U is m-strongly convex, $U(x_1) \ge U(x_\star) + \frac{m}{2}\|x_1 - x_\star\|^2$, so e^{-U} has Gaussian tails. Thus, for $g(x_1) = x_1 - x_\star$ (whose divergence is $\nabla\cdot g \equiv d$), integration by parts yields

$$\int_{\mathbb{R}^d} \langle \nabla U(x_1),\, x_1 - x_\star\rangle\, e^{-U(x_1)}\,dx_1 = \int_{\mathbb{R}^d} \nabla\cdot g(x_1)\, e^{-U(x_1)}\,dx_1 = d\int_{\mathbb{R}^d} e^{-U(x_1)}\,dx_1,$$

where the boundary term vanishes since e^{-U} has tails. Normalizing gives

$$\mathbb{E}_{\pi_U}\big[\langle \nabla U(X_1),\, X_1 - x_\star\rangle\big] = d.$$

By strong convexity and $\nabla U(x_\star) = 0$, $\langle \nabla U(x_1), x_1 - x_\star \rangle \geq m\|x_1 - x_\star\|^2$. Taking expectation, $d \geq m\,\mathbb{E}_{\pi_U}\|X_1 - x_\star\|^2$. Since the X_1-marginal of ρ is π_U, the same bound holds under ρ, proving (89).

Bound for $\mathbb{E}\|\nabla U(X_1)\|^2$. First, by L-smoothness and $\nabla U(x_\star) = 0$, $\|\nabla U(x_1)\| \leq L\|x_1 - x_\star\|$. Squaring and taking expectations then using (89) gives $\mathbb{E}\|\nabla U(X_1)\|^2 \leq (L^2/m)\, d$.

Bound for $\mathbb{E}\|X\|^2$. Under ρ, $X_k \sim \mathcal{N}(0, I_d)$ for $k \geq 2$, hence $\mathbb{E}\|X_k\|^2 = d$. For X_1, $\|X_1\|^2 = \|X_1 - x_\star + x_\star\|^2 \leq 2\|X_1 - x_\star\|^2 + 2\|x_\star\|^2$. Taking expectations and applying (89) gives $\mathbb{E}\|X_1\|^2 \leq 2(d/m) + 2\|x_\star\|^2$. □