北京大学

博士研究生学位论文

题目：**复杂随机过程和数值方法的遍历性和长时间误差**

姓　　名：叶胥达

学　　号：2001110036

院　　系：数学科学学院

专　　业：计算数学

研究方向：科学计算、随机方法和应用分析

导师姓名：周珍楠 (西湖大学)

☑ 学术学位　　☐ 专业学位

二〇二五 年 六 月

版权声明

摘要

从高维分布中采样是计算数学与物理学中的核心问题，其应用涵盖分子动力学、贝叶斯推断以及机器学习等领域。高效且可扩展的采样算法对于探索复杂系统、计算热力学性质以及解决高维积分问题至关重要。本论文重点研究了两种重要采样方法的遍历性和长时间行为：路径积分分子动力学（Path Integral Molecular Dynamics, PIMD）和随机分组方法（Random Batch Method, RBM）。这些方法解决了高维采样中的关键挑战，包括保持计算效率、确保大系统中的算法收敛性以及在长时间模拟中实现高精度。

路径积分分子动力学是一种广泛用于计算量子热平均值的方法，这些热平均值对于理解量子系统的平衡性质至关重要。PIMD 通过将量子正则系综映射为由 N 个珠子组成的经典环聚合物系统，使得采样对应的 Boltzmann 分布成为可能。然而，确保随机动力学的遍历性在 N 上保持一致（即收敛速率不依赖于珠子的数量）是一个重大的理论挑战。本论文严格证明了欠阻尼 Langevin 动力学在 PIMD 中的遍历性对 N 保持一致，并以相对熵的形式进行了衡量。该结果不仅填补了 PIMD 理论理解中的重要空白，还利用了广义 Γ 演算，为高维随机过程提供了显式的收敛率。

随机分组方法是一种针对相互作用粒子系统的高效计算方法，后者的所有成对相互作用的计算成本通常随粒子数量平方增长。通过将粒子随机分组为批次，RBM 将计算复杂度降低为线性增长，同时近似保持系统的统计特性。本论文聚焦于随机分组相互作用粒子系统（Random Batch Interacting Particle System, RB–IPS），证明了其遍历性在 N 上保持一致，并将其与目标分布的 Wasserstein-1 误差界定为 $O(\sqrt{h})$，其中 h 为时间步长。此外，本论文分析了离散 RB–IPS 的长时间行为，证明其长时间误差以 $O(e^{-\lambda t} + \sqrt{h})$ 衰减，其中 λ 为独立于粒子数量的收敛速率。我们还将这些结果扩展到平均场极限，证明了 RB–IPS 在逼近 McKean–Vlasov 过程的平衡分布时的稳健性和可扩展性。

数值实验验证了理论结果，并展示了 PIMD 和 RBM 在解决高维和大规模采样问题中的实际效率。这些结果表明，这些方法能够在保持计算可行性的同时，实现高精度采样，即使在粒子数量庞大或配置维度高的系统中也是如此。这两种方法的遍历性对 N 保持一致的性质尤为显著，确保了其在实际应用中的可靠性和效率。

总之，本论文在理论和实践上对 PIMD 和 RBM 的理解做出了重要贡献，提供了严格的证明、定量的误差分析以及经过验证的数值性能。这些进展为进一步探索可扩展采样算法及其在计算物理、随机分析和机器学习中的应用奠定了基础。

关键词：路径积分分子动力学，随机分组方法，遍历性，采样，高维系统。

Ergodicity and Long-Time Error of Complicated Stochastic Processes and Numerical Methods

Xuda Ye (Computational Mathematics)
Directed by Prof. Zhennan Zhou

ABSTRACT

Sampling from high-dimensional distributions is a cornerstone of computational mathematics and physics, with applications spanning molecular dynamics, Bayesian inference, and machine learning. Efficient and scalable sampling algorithms are indispensable for exploring complex systems, computing thermodynamic properties, and solving high-dimensional integration problems. This thesis focuses on the ergodicity and long-time behavior of two prominent sampling methods: Path Integral Molecular Dynamics (PIMD) and the Random Batch Method (RBM). These methods address critical challenges in high-dimensional sampling, such as maintaining computational efficiency, ensuring convergence of algorithms in large systems, and achieving high accuracy over long-time simulations.

Path Integral Molecular Dynamics (PIMD) is a widely used approach for computing quantum thermal averages, which are crucial for understanding the equilibrium properties of quantum systems. By mapping the quantum canonical ensemble to a classical ring-polymer system with N beads, PIMD enables sampling of the corresponding Boltzmann distribution. However, ensuring that the stochastic dynamics exhibit uniform-in-N ergodicity—where the convergence rate is independent of the number of beads—is a significant theoretical challenge. In this thesis, we rigorously prove the uniform-in-N ergodicity of the underdamped Langevin dynamics in PIMD, measured in terms of relative entropy. This result not only fills an important gap in the theoretical understanding of PIMD but also leverages the generalized Γ calculus to provide explicit convergence rates for high-dimensional stochastic processes.

The Random Batch Method (RBM) offers a computationally efficient approach for simulating interacting particle systems, where the cost of computing all pairwise interactions typically grows quadratically with the number of particles. By randomly grouping particles into batches, RBM reduces this computational complexity to linear scaling while approximately preserving the system's statistical properties. This thesis focuses on the Random Batch Inter-

acting Particle System (RB–IPS), establishing its uniform-in-N ergodicity and bounding the Wasserstein-1 error from the target distribution by $O(\sqrt{h})$, where h is the time step size. Furthermore, we analyze the long-time behavior of the discrete RB–IPS, proving that its long-time error decays as $O(e^{-\lambda t} + \sqrt{h})$, where λ is a convergence rate independent of the number of particles. These results are extended to the mean-field limit, where the RB–IPS approximates the invariant distribution of the McKean–Vlasov process, further demonstrating the method's robustness and scalability.

Numerical experiments validate the theoretical findings and highlight the practical efficiency of PIMD and RBM in addressing high-dimensional and large-scale sampling challenges. The results demonstrate the ability of these methods to achieve accurate sampling while maintaining computational feasibility, even in systems with a large number of particles or high-dimensional configurations. The uniform-in-N properties of both methods are particularly noteworthy, as they ensure reliability and efficiency for real-world applications.

In summary, this thesis provides significant contributions to the theoretical and practical understanding of PIMD and RBM, offering rigorous proofs, quantitative error analysis, and validated numerical performance. These advancements pave the way for further exploration of scalable sampling algorithms and their applications in computational physics, stochastic analysis, and machine learning.

KEY WORDS: Path Integral Molecular Dynamics, Random Batch Method, Ergodicity, Sampling, High-dimensional Systems.

Contents

Chapter 1 Introduction

In a sampling problem, the goal is to draw samples from a target probability distribution $\pi(x)$, which is often expressed in the form $\pi(x) \propto \exp(-V(x))$, where $V(x)$ represents a potential function. A sampling algorithm aims to generate a sequence of samples $\{X_k\}_{k=0}^{\infty}$ such that the statistical properties of $\pi(x)$ can be inferred from this sequence.

Sampling is a fundamental problem in both computational physics and mathematics, with diverse applications in fields such as molecular dynamics, Bayesian inference, and machine learning. In computational physics, particularly in molecular dynamics, sampling methods are indispensable for exploring the configuration space of many-body systems and for calculating thermodynamic averages. The target distribution $\pi(x)$ often corresponds to the Boltzmann distribution, representing the equilibrium state of the system (Frenkel and Smit, 2023; Tuckerman, 2023; Leimkuhler and Matthews, 2015). Accurate sampling enables the estimation of properties such as free energy, entropy, and transport coefficients, which are critical for understanding the macroscopic behavior of complex systems.

In machine learning, sampling methods generate new data points that are similar to the training data. They are also used to train generative models, which can generate new data points that are indistinguishable from real data (Goodfellow et al., 2020). A prominent example of sampling methods in data science is the diffusion model, which is a generative model that learns to generate new data points by reversing a gradual noising process (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2020; Croitoru et al., 2023). Diffusion models have achieved state-of-the-art results in image generation, and they can also be used for other tasks such as text generation and audio generation. However, diffusion models can be computationally expensive to train, and they can also be prone to generating blurry or unrealistic samples.

Provided the strongly science motivated sampling problems, we require efficient sampling methods for exploring complex distributions. Some popular sampling methods include Markov chain Monte Carlo (MCMC) (Metropolis and Ulam, 1949; Hastings, 1970; Neal et al., 2011), importance sampling (Kong et al., 1994; Neal, 2001; Tokdar and Kass, 2010) and Stochastic Gradient Langevin Dynamics (SGLD) (Welling and Teh, 2011; Chen et al., 2014). MCMC methods work by constructing a Markov chain whose stationary distribution is the target distribution; importance sampling methods work by sampling from a proposal distribution that is similar to the target distribution; while SGLD is a variant of the Langevin dynamics (Leimkuh-

ler and Matthews, 2015; Haile, 1992) that uses stochastic gradients to sample from the target distribution. The choice of the sampling method depends on the specific problem being addressed, including the dimensionality of the space, the complexity of the distribution, and the available computational resources.

The importance of sampling methods becomes even more pronounced in high-dimensional problems, which are common in both molecular dynamics and data science. In molecular dynamics, the dimensionality grows rapidly with the number of particles, leading to a vast configuration space that must be efficiently explored to ensure accurate results. Similarly, in data science, high-dimensional sampling is crucial for tasks such as posterior inference, generative modeling, and high-dimensional optimization. The curse of dimensionality renders naive methods infeasible, necessitating the development of scalable and efficient algorithms. Advances in sampling not only enhance our ability to solve these challenging problems but also drive progress in related areas such as optimization, statistical physics, and machine learning.

Given the importance of high-dimensional sampling problems, understanding the accuracy of sampling algorithms is crucial. The sampling quality of an algorithm can be assessed in several ways. Let $\pi(x)$ be the target distribution in $\mathbb{R}^d$, and let $\{X_k\}_{k\geqslant 0}$ be the sequence of samples produced by the algorithm. If our goal is to compute the high-dimensional integral $\pi(f) := \int_{\mathbb{R}^d} f(x)\pi(x)\mathrm{d}x$ for a specific test function (or observable function) $f(x)$, then the sampling accuracy can be conveniently determined from the time average error:

$$\text{time average error at the } K\text{-th step} = \frac{1}{K}\sum_{k=0}^{K-1} f(X_k) - \pi(f), \tag{1.1}$$

where the integer K signifies the number of steps in the algorithm. Another approach to characterizing sampling accuracy is to compare the distance between distribution laws. Suppose we have a metric $d(\cdot,\cdot)$ on the probability space $\mathcal{P}(\mathbb{R}^d)$, and let $\mu_k = \mathrm{Law}(X_k)$ be the distribution law of X_k at the k-th step. Then, the distribution law error of the algorithm is given by

$$\text{distribution law error at the } K\text{-th step} = d(\mu_K, \pi). \tag{1.2}$$

In both (1.1) and (1.2), we expect the sampling accuracy to systematically improve as the number of steps K increases. However, verifying such convergence is not trivial, and an essential theoretical approach—ergodicity—is required to rigorously analyze the sampling accuracy.

In this paper, we study sampling algorithms based on the discretization of Stochastic Differential Equations (SDEs). Ergodicity characterizes the rate at which the distribution law of an SDE converges to the target distribution. On one hand, ergodicity captures the exponential

convergence rate of the distribution law, indicating the time required to obtain an effective sample point. On the other hand, it ensures the stability of the SDE, facilitating the analysis of the long-time error in sampling algorithms derived from time discretization.

Various methods exist for proving the ergodicity of a given SDE. In this work, we primarily employ reflection coupling (Eberle, 2011) and the generalized Γ-calculus (Monmarché, 2019). A brief review of popular techniques for establishing ergodicity is provided in Section 2.4, where we also introduce the triangle inequality framework. The triangle inequality framework offers a quantitative estimate of the long-time error in sampling algorithms by leveraging both the ergodicity and the finite-time error.

In the following, we aim to rigorously analyze the ergodicity and long-time behavior of two specific sampling algorithms: Path Integral Molecular Dynamics (PIMD) and Random Batch Method (RBM).

The PIMD is designed to compute the thermal average of a specific observable within a quantum canonical ensemble. Quantum thermal averages are vital in describing the thermal properties of complex quantum systems, including chemical reaction rates (Miller, 1993; Clary, 1998) and phase transitions (Voth, 1993; Sondhi et al., 1997). The PIMD approach is inspired by Feynman's path integral formulation (Feynman, 2018), which maps the original quantum canonical ensemble to a classical Boltzmann distribution of a ring-polymer system with N beads (Schweizer et al., 1981; Marx and Parrinello, 1996; Liu et al., 2016). Accurate computation of thermal averages requires $N \to \infty$, making it essential to investigate whether the stochastic process used to sample the target Boltzmann distribution exhibits *uniform-in-N* ergodicity. This property ensures that the convergence rate of the stochastic process is independent of N.

The RBM, introduced by Jin et al. (2020), is an efficient sampling algorithm for the interacting particle system (IPS). The IPS is a fundamental model in molecular dynamics, widely used to describe the collective behavior of particles interacting through pairwise forces. Such systems are crucial for studying a variety of physical and chemical phenomena, including molecular simulations (Frenkel and Smit, 2023), plasma dynamics (Nicholson, 1983), and chemical reaction modeling (Haile, 1992). In these systems, each particle experiences forces from all others, making the simulation computationally challenging as the number of particles N increases. Specifically, the cost of computing all pairwise interactions grows quadratically with N, i.e., $O(N^2)$, creating a bottleneck for large-scale simulations.

Beyond their role in molecular dynamics, the IPS has a significant mathematical prop-

erty: their behavior as N becomes very large can be approximated by the mean-field limit, which is described by the McKean–Vlasov process (MVP) (McKean, 1967; Sznitman, 1991; Chaintron and Diez, 2021, 2022). The MVP is a nonlinear stochastic model that captures the average effect of all particles in the system, making it a powerful tool in statistical physics, population dynamics, and emerging areas like machine learning (Carrillo et al., 2018). This mean-field approximation allows to study macroscopic properties of particle systems without the computational expense of simulating all individual interactions. However, the accuracy of this approach relies on the ability to efficiently sample the particle system, particularly for long-time simulations where statistical averages are critical.

The contributions of this thesis are as follows:

Path Integral Molecular Dynamics (PIMD): We establish a rigorous proof of uniform-in-N ergodicity for the underdamped Langevin dynamics associated with the N-bead ring polymer, quantified via the relative entropy (Theorem 3.2). This result addresses a critical theoretical gap by guaranteeing that the convergence rate toward the invariant distribution remains independent of N, even as the dimensionality grows. Such uniform convergence is pivotal for ensuring the scalability and reliability of PIMD in practical high-dimensional quantum simulations. Additionally, this finding demonstrates a novel application of the generalized Γ calculus, allowing explicit derivation of convergence rates for complex stochastic processes, further broadening its utility in computational mathematics and quantum physics. Our major proof technique is generalized Γ calculus, which is able to produce explicit convergence rate for the underdamped Langevin dynamics.

Random Batch Method (RBM): Our focus is on the Random Batch Interacting Particle System (RB–IPS), a key approximation for large-scale particle simulations. We prove that RB–IPS achieves uniform-in-N ergodicity (Theorem 4.1) and demonstrate that the Wasserstein-1 error from the target distribution is bounded by $O(\sqrt{h})$, where h denotes the step size (Theorem 4.3). Furthermore, we analyze the long-time behavior of time-discrete RB–IPS, establishing a bound on its long-time error of $O(e^{-\lambda t} + \sqrt{h})$, with λ as a convergence rate independent of N and h (Theorem 4.5). Leveraging the propagation of chaos, we also derive the long-time error for discrete RB–IPS in approximating the invariant distribution of the McKean–Vlasov process (Theorem 4.9). These contributions provide a robust theoretical foundation for the efficiency and accuracy of RB–IPS in large-scale simulations, reinforcing its potential to significantly reduce computational costs while maintaining high fidelity in the statistical properties of particle systems. Our proof technique is the reflection coupling, which is convenient to

validate the ergodicity of the IPS with nonconvex potentials.

Collectively, these results advance the theoretical understanding of ergodicity and long-time error analysis for high-dimensional stochastic processes, with direct implications for molecular dynamics, statistical physics, and computational algorithms in scientific computing.

The thesis is organized as follows. Chapter 1 introduces the background and motivation for this study, highlighting the importance of high-dimensional sampling algorithms and the specific challenges addressed in this work. Chapter 2 provides a comprehensive review of the theoretical and methodological foundations, including Path Integral Molecular Dynamics (PIMD), the Random Batch Method (RBM), and key analytical tools such as the generalized Γ calculus and coupling techniques. Chapter 3 focuses on the uniform-in-N ergodicity of PIMD, presenting rigorous proofs and illustrating their implications for sampling quantum thermal averages. Chapter 4 addresses the RB–IPS, detailing its uniform-in-N ergodicity, approximation error, and long-time behavior, with an extension to the mean-field limit.

Chapter 2 Background

The background knowledge necessary for this thesis is presented in this section, providing the theoretical foundation and methodological tools essential for the analysis in later chapters. We begin by introducing two key concepts—ergodicity and long-time error, which are fundamental to understanding the convergence properties of stochastic sampling algorithms. Next, we provide an overview of Path Integral Molecular Dynamics (PIMD) and the Random Batch Method (RBM), the two primary sampling methods studied in this work. Their mathematical formulations and practical relevance are discussed in detail. Additionally, we review theoretical approaches for analyzing ergodicity and long-time behavior, including the generalized Γ calculus and reflection coupling techniques, which serve as the main analytical tools in the subsequent chapters. These foundations will be directly applied in Chapters 3 and 4 to establish rigorous results on the ergodicity and error analysis of PIMD and RBM, respectively.

2.1 Basic notions: ergodicity and long-time error

Given the Markov process $(x_t)_{t\geqslant 0}$ in $\mathbb{R}^d$, define the Markov semigroup $(P_t)_{t\geqslant 0}$ as:

$$(P_t f)(x) = \mathbb{E}\big[f(x_t) : x_0 = x\big], \quad x \in \mathbb{R}^d, \tag{2.1}$$

where $f(\cdot)$ is a test function in $C^\infty(\mathbb{R}^d)$. The dual of $(P_t)_{t\geqslant 0}$ is the dual semigroup $(\mathcal{P}_t)_{t\geqslant 0}$, which acts on probability distributions on $\mathbb{R}^d$, which are denoted by the set $\mathcal{P}(\mathbb{R}^d)$①. Specifically, for any initial distribution ν with the initial state $x_0 \sim \nu$ of the Markov process, the notation $\nu\mathcal{P}_t \in \mathcal{P}(\mathbb{R}^d)$ represents the distribution law of x_t.

If the infinitesimal generator of $(x_t)_{t\geqslant 0}$ is $\mathcal{L}$, then the Markov semigroup $(P_t)_{t\geqslant 0}$ and the dual semigroup $(\mathcal{P}_t)_{t\geqslant 0}$ can be expressed as:

$$P_t = e^{t\mathcal{L}}, \quad \mathcal{P}_t = e^{t\mathcal{L}^*}, \quad t \geqslant 0,$$

where $\mathcal{L}^*$ is the adjoint operator of $\mathcal{L}$ in $L^2(\mathbb{R}^d)$. This operator is also known as the forward Kolmogorov operator or the Fokker–Planck operator.

Ergodicity characterizes the long-time convergence of stochastic processes. If the Markov

① The letter P in the Markov semigroup $(P_t)_{t\geqslant 0}$, the dual semigroup $(\mathcal{P}_t)_{t\geqslant 0}$ and the probability distribution set $\mathcal{P}(\mathbb{R}^d)$ is distinguished by the fonts.

process $(x_t)_{t\geqslant 0}$ has an invariant distribution $\pi \in \mathcal{P}(\mathbb{R}^d)$, meaning:

$$\pi \mathcal{P}_t = \pi, \quad \forall t \geqslant 0, \tag{2.2}$$

and there exist constants C, λ, and a metric $d(\cdot,\cdot)$ on probability distributions such that:

$$d(\nu \mathcal{P}_t, \pi) \leqslant C e^{-\lambda t} d(\nu, \pi), \quad \forall t \geqslant 0, \tag{2.3}$$

for any probability distribution $\nu \in \mathcal{P}(\mathbb{R}^d)$, then $(x_t)_{t\geqslant 0}$ exhibits ergodicity.

Remark 2.1. *The metric $d(\cdot,\cdot)$ does not need to be symmetric nor means to be a metric space. Common choices of $d(\cdot,\cdot)$ include the total variation, the relative entropy (KL divergence), and the Wasserstein distance.*

When $d(\cdot,\cdot)$ is the relative entropy $H(\cdot|\cdot)$, the ergodicity property (2.3) can be equivalently expressed as:

$$\mathrm{Ent}_\pi(\mathcal{P}_t f) \leqslant C e^{-\lambda t} \mathrm{Ent}_\pi(f), \quad \forall t \geqslant 0, \tag{2.4}$$

where $\mathrm{Ent}_\pi(f)$, the entropy of a positive test function f, is defined as:

$$\mathrm{Ent}_\pi(f) = \int_{\mathbb{R}^d} f \log f \mathrm{d}\pi - \int_{\mathbb{R}^d} f \mathrm{d}\pi \log \int_{\mathbb{R}^d} f \mathrm{d}\pi. \tag{2.5}$$

In particular, by setting $f = \frac{\mathrm{d}\nu}{\mathrm{d}\pi}$ in (2.4), we recover the inequality (2.3).

The stochastic process $(x_t)_{t\geqslant 0}$ must be discretized in the time t for implementation on a computer, and the discretization error depends on the step size h and the order of the numerical method. As a simple example, consider the stochastic differential equation (SDE):

$$\dot{x}_t = b(x_t) + \sigma \dot{B}_t, \tag{2.6}$$

where $b(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ is the drift force, $\sigma > 0$ is the diffusion coefficient, and $(B_t)_{t\geqslant 0}$ is the Brownian motion in $\mathbb{R}^d$. Moreover, the Euler–Maruyama integrator provides a straightforward discretization for the SDE (2.6) with strong order one:

$$\tilde{x}_{(n+1)h} = \tilde{x}_{nh} + b(\tilde{x}_{nh})h + \sigma\sqrt{h}\xi_n, \quad \xi_n \sim \mathcal{N}(0, I_d). \tag{2.7}$$

In general, let $(\tilde{x}_t)_{t\geqslant 0}$ be the numerical approximation of $(x_t)_{t\geqslant 0}$, and $(\tilde{\mathcal{P}}_t)_{t\geqslant 0}$ be its dual semigroup. If there exists an error function $\varepsilon(h)$ such that $\lim_{h\to 0} \varepsilon(h) = 0$ and

$$d(\nu \tilde{\mathcal{P}}_t, \pi) \leqslant C e^{-\lambda t} + \varepsilon(h), \tag{2.8}$$

then the long-time error of $(\tilde{x}_t)_{t\geqslant 0}$ is said to exhibit exponential decay. Specifically, $\tilde{x}_t$ provides reliable samples of the target distribution by choosing a sufficiently large evolution time t and

a sufficiently small step size h.

In conclusion, ergodicity describes the long-time behavior of a stochastic process, ensuring convergence to the invariant distribution and effective exploration of the state space. It is a fundamental property for understanding the stability and convergence of stochastic dynamics.

The long-time error, in contrast, quantifies the accuracy of numerical algorithms in sampling the invariant distribution, accounting for discretization and step size effects. Together, these concepts provide a framework for evaluating the theoretical and practical efficiency of stochastic simulation methods, crucial for applications in statistical physics, machine learning, and computational chemistry.

2.2 Path integral molecular dynamics

Path Integral Molecular Dynamics (PIMD) is a fundamental tool for computing quantum thermal averages. In this section, we demonstrate how the path integral formulation reformulates the quantum canonical ensemble into the problem of sampling a ring polymer system composed of N beads.

Consider the quantum system in $\mathbb{R}^d$ given by the Hamiltonian

$$\hat{H} = -\frac{\Delta}{2} + V(\hat{x}), \tag{2.9}$$

where Δ and $\hat{x}$ are the Laplacian and position operatoes in $\mathbb{R}^d$, and $V(\cdot)$ is a real-valued potential function in $\mathbb{R}^d$. When the quantum system is at a constant temperature $T = 1/\beta$, the state of the system is described by the canonical ensemble with the density operator $e^{-\beta\hat{H}}$, and thus the partition function is $\mathcal{Z} = \mathrm{Tr}[e^{-\beta\hat{H}}]$. We assume the observable operator $O(\hat{x})$ depends only on the position operator $\hat{x}$, where $O(\cdot)$ is a real-valued function in $\mathbb{R}^d$. Then the quantum thermal average of the system is defined as the canonical average of $O(\hat{x})$, namely

$$\langle O(\hat{x})\rangle_\beta = \frac{1}{\mathcal{Z}}\mathrm{Tr}[e^{-\beta\hat{H}}O(\hat{x})]. \tag{2.10}$$

Here, using the quantum bra-ket notation (Dirac notation), we can also write

$$\langle O(\hat{x})\rangle_\beta = \frac{1}{\mathcal{Z}}\int_{\mathbb{R}^d} O(x)\,\langle x|e^{-\beta\hat{H}}|x\rangle\,\mathrm{d}x. \tag{2.11}$$

To compute $\langle O(\hat{x})\rangle_\beta$, we only need to sample the probability density proportional to $\langle x|e^{-\beta\hat{H}}|x\rangle$. Let $N \in \mathbb{N}$ be an integer and $x_1 \in \mathbb{R}^d$ be a fixed position. Using the equality

$$\mathrm{Id} = \int_{\mathbb{R}^d} |x\rangle\,\langle x|\,\mathrm{d}x$$

and define $\beta_N = \beta/N$, we have the approximation

$$\begin{aligned}
\langle x_1|e^{-\beta\hat{H}}|x_1\rangle &= \int \langle x_1|e^{-\beta_N\hat{H}}|x_2\rangle\cdots\langle x_N|e^{-\beta_N\hat{H}}|x_1\rangle \,\mathrm{d}x_2\cdots x_N \qquad (2.12)\\
&= \int \prod_{j=1}^{N} \left\langle x_j\middle|e^{-\beta_N\hat{H}}\middle|x_{j+1}\right\rangle \mathrm{d}x_2\cdots\mathrm{d}x_N\\
&\approx \int \prod_{j=1}^{N} \left\langle x_j\middle|e^{-\frac{\beta_N}{2}V}e^{\frac{\beta_N}{2}\Delta}e^{-\frac{\beta_N}{2}V}\middle|x_{j+1}\right\rangle \mathrm{d}x_2\cdots\mathrm{d}x_N\\
&= \int e^{-\beta_N\sum_{j=1}^{N}V(x_j)}\left[\prod_{j=1}^{N}\left\langle x_j\middle|e^{\frac{\beta_N}{2}\Delta}\middle|x_{j+1}\right\rangle\right]\mathrm{d}x_2\cdots\mathrm{d}x_N\\
&= \int e^{-\beta_N\sum_{j=1}^{N}V(x_j)}\left[\prod_{j=1}^{N}\frac{1}{(2\pi\beta_N)^{\frac{d}{2}}}e^{-\frac{1}{2\beta_N}|x_j-x_{j+1}|^2}\right]\mathrm{d}x_2\cdots\mathrm{d}x_N\\
&= \frac{1}{(2\pi\beta_N)^{\frac{dN}{2}}}\int e^{-\mathcal{E}_N(x_1,\cdots,x_N)}\mathrm{d}x_2\cdots\mathrm{d}x_N, \qquad (2.13)
\end{aligned}$$

where the energy function $\mathcal{E}_N(x_1,\cdots,x_N)$ is defined by

$$\mathcal{E}_N(x_1,\cdots,x_N) = \frac{1}{2\beta_N}\sum_{j=1}^{N}|x_j - x_{j+1}|^2 + \beta_N\sum_{j=1}^{N}V(x_j). \qquad (2.14)$$

Here, we admit the periodic boundary condition for $\{x_j\}_{j=1}^{N}$, namely, $x_{N+1} = x_1$. Therefore, the quantum thermal average $\langle O(\hat{x})\rangle_\beta$ can be approximated as

$$\langle O(\hat{x})\rangle_\beta \approx \int_{\mathbb{R}^{dN}}\left(\frac{1}{N}\sum_{j=1}^{N}O(x_j)\right)\pi_N(x_1,\cdots,x_N)\mathrm{d}x_1\cdots\mathrm{d}x_N, \qquad (2.15)$$

where $\pi_N(x_1,\cdots,x_N) \propto e^{-\mathcal{E}_N(x_1,\cdots,x_N)}$ is a classical Boltzmann distribution in $\mathbb{R}^{dN}$.

Note that $\mathcal{E}_N(x_1,\cdots,x_N)$ can be viewed as the energy of a ring polymer system of N beads, where the adjacent beads (x_j, x_{j+1}) are connected by a spring potential, and each bead x_j feels the system potential $V(x_j)$. Figure 2.1 shows an example of the ring polymer system with 10 beads. Moreover, as the number of beads N tends to infinity, the energy function $\mathcal{E}_N(x_1,\cdots,x_N)$ has a continuum limit (Lu et al., 2020)

$$\mathcal{E}_\infty(x(\cdot)) = \frac{1}{2}\int_0^\beta |x'(\tau)|^2\mathrm{d}\tau + \int_0^\beta V(x(\tau))\mathrm{d}\tau, \qquad (2.16)$$

where $x : [0,\beta] \to \mathbb{R}^d$ is a parameterized continuous loop in the interval $[0,\beta]$.

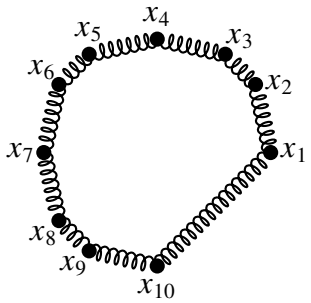


Figure 2.1 Example of a ring polymer with 10 beads.

Remark 2.2. *The approximation used in* (2.13) *is the symmetric operator splitting (Ames, 2014; Glowinski et al., 2017), and the local approximation error is* $O(\beta_N^3)$. *Therefore, the approximation error in the quantum thermal average in* (2.15) *is formally* $O(N\beta_N^3)$. *However, a rigorous justification of the approximation error in* N *is still unknown in literature.*

Since the accurate calculation of the quantum thermal average must require the number of beads N to be sufficiently large, it is essential to obtain a sampling algorithm for the target distribution $\pi_N(x_1, \cdots, x_N)$ with uniform-in-N ergodicity, namely, the convergence rate does not depend on the number of beads N.

2.3 Random batch method

Consider a system of N particles represented by a collection of position variables $\boldsymbol{x}_t = \{x_t^i\}_{i=1}^N$, where $x_t^i \in \mathbb{R}^d$ denotes the position of the i-th particle. The interacting particle system (IPS) $\boldsymbol{x}_t$ evolves according to the overdamped Langevin dynamics:

$$\dot{x}_t^i = b(x_t^i) + \frac{1}{N-1}\sum_{j\neq i} K(x_t^i - x_t^j) + \sigma \dot{B}_t^i, \quad i = 1, \ldots, N. \tag{2.17}$$

Here, $b(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ represents the drift force, $K(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ is the interaction force, $\sigma > 0$ is a scalar constant, and $\{B_t^i\}_{i=1}^N$ are N independent standard Brownian motions in $\mathbb{R}^d$.

As the number of particles $N \to \infty$, the IPS (2.17) formally converges to the nonlinear McKean–Vlasov process (MVP), defined as:

$$\dot{\bar{x}}_t = b(\bar{x}_t) + (K * \bar{\mu}_t)(\bar{x}_t) + \sigma \dot{B}_t, \tag{2.18}$$

where $\bar{\mu}_t$ represents the distribution law of $\bar{x}_t$ in $\mathbb{R}^d$, B_t is standard Brownian motion, and $*$ denotes the convolution operator, defined as:

$$(K * \mu)(x) = \int_{\mathbb{R}^d} K(x-y)\mu(dy), \quad x \in \mathbb{R}^d. \tag{2.19}$$

Under suitable assumptions on the drift force $b(x)$ and the interaction force $K(x)$, the IPS (2.17) admits a unique invariant distribution $\pi(\boldsymbol{x})$ in $\mathbb{R}^{dN}$. The primary objective is to sample $\pi(\boldsymbol{x})$ efficiently. For instance, if $b(x) = -\nabla U(x)$ and $K(x) = -\nabla V(x)$ for some potential functions $U(x)$ and $V(x)$, with $V(x)$ being even and $\sigma = \sqrt{2}$, then $\pi(\boldsymbol{x})$ is explicitly given by:

$$\pi(\boldsymbol{x}) \propto \exp\left(-\sum_{i=1}^{N} U(x^i) - \frac{1}{N-1}\sum_{1\leqslant i<j\leqslant N} V(x^i - x^j)\right), \quad \boldsymbol{x} \in \mathbb{R}^{dN}. \tag{2.20}$$

To simulate the IPS (2.17) numerically, time discretization is required at each time step. For an IPS with N particles, calculating all interaction forces $\{K(x_t^i - x_t^j)\}_{i\neq j}$ involves $O(N^2)$ complexity per time step, leading to inefficiency for large N. Therefore, it is desirable to employ approximate simulation methods that can reduce the computational cost while maintaining the reliability of the samples generated from the invariant distribution $\pi(\boldsymbol{x})$.

The Random Batch Method (RBM) proposed in Jin et al. (2020) is a simple random algorithm to reduce the computational cost per time step from $O(N^2)$ to $O(N)$. Nowadays, the RBM has been used to simulate complicated chemical systems (Jin et al., 2021; Li et al., 2020b; Liang et al., 2021, 2023) and accelerate the particle ensemble methods (Li et al., 2020a; Carrillo et al., 2021; Ha et al., 2021), and is also able to combine with the variance reduction techniques (Pareschi and Zanella, 2024; Xu et al., 2024). In these applications, the RBM is not only an efficient algorithm for the evolution of the system, it also preserves the invariant distribution $\pi(\boldsymbol{x})$ in an approximate sense, thus can be used to obtain statistical samples of the invariant distribution of the IPS (2.17).

The idea of the RBM is illustrated as follows. Let $h > 0$ be the time step for batch divisions. For each $n \geqslant 0$, let the index set $\{1, \cdots, N\}$ be randomly divided into q batches $\{C_1, \cdots, C_q\}$, where each batch C has size $p = N/q$. The IPS (2.17) within the time interval $t \in [nh, (n+1)h)$ is approximated as the particle system $\{y_t^i\}_{i=1}^N$ in $\mathbb{R}^{dN}$, given by the SDE

$$\dot{y}_t^i = b(y_t^i) + \frac{1}{p-1}\sum_{j\neq i, j\in C} K(y_t^i - y_t^j) + \sigma \dot{B}_t^i, \quad i \in C, \quad t \in [nh, (n+1)h), \tag{2.21}$$

where C is the unique batch that contains i. For the next time interval, the previous division is discarded and another random division is employed for the dynamics (2.21). In the following, the dynamical system (2.21) will be referred to as the random batch interacting particle system (RB–IPS), as a comparison to the IPS (2.17). Note that both (2.17) and (2.21) are exactly integrated in time, and thus there is no error due to time discretization.

Furthermore, if we employ the Euler–Maruyama integrator to discretize the RB–IPS

(2.21), then we obtain the following discrete RB–IPS:

$$Y^i_{n+1} = Y^i_n + \left(b(Y^i_n) + \frac{1}{p-1} \sum_{j \neq i, j \in C} K(Y^i_n - Y^j_n) \right) h + \sigma(B^i_{n(h+1)} - B^i_{nh}), \quad i \in C. \tag{2.22}$$

The discrete RB–IPS (2.22) requires only $O(Np)$ complexity to compute the interaction forces in each time step, which is a significant advance in simulation efficiency compared to the $O(N^2)$ complexity of the discrete IPS. Specifically, selecting a batch size of $p = 2$ results in a complexity of $O(N)$ per time step.

In this work, we rigorously establish the ergodicity of the RB–IPS (2.21) and quantitatively estimate the long-time error of the discrete RB–IPS (2.22). These results highlight the reliability of the RBM as an efficient sampling approach for IPS. By reducing computational complexity while preserving key statistical properties, the RBM provides a practical and robust method for simulating large-scale particle systems in the long time.

2.4 Theoretical approaches for ergodicity and long-time behavior

Ergodicity is a fundamental property of stochastic processes, signifying that the long-time behavior of the process reflects its statistical equilibrium. Specifically, a stochastic process is ergodic if time-averaged quantities converge to their ensemble averages, as defined by the invariant distribution of the process. This property ensures that the dynamics of the system explore the entire state space sufficiently over time, avoiding being trapped in specific regions.

The concept of ergodicity plays a pivotal role in stochastic analysis for both theoretical and practical reasons. Theoretically, it provides a rigorous foundation for the convergence of stochastic processes, allowing for the derivation of key results such as the existence and uniqueness of invariant measures. Practically, ergodicity underpins the validity of sampling methods, ensuring that long trajectories of a stochastic process yield reliable statistical estimates of the target distribution. Without ergodicity, these estimates may be biased or incomplete, particularly in high-dimensional systems where efficient exploration of the state space is crucial.

The theoretical approaches to studying ergodicity can be broadly classified into two categories: probabilistic approaches and PDE-based approaches.

Among the probabilistic methods, coupling is the most commonly used technique. For stochastic processes driven by Brownian motions, coupling involves constructing coupled processes with shared noise to demonstrate contractivity. Synchronous coupling is frequently employed for systems with convex potentials (Lindvall and Rogers, 1996; Chen and Li, 1989). More recently, reflection coupling has gained popularity in addressing non-convex sampling

problems (Eberle, 2011; Eberle et al., 2019). Reflection coupling is also instrumental in analyzing the contractivity of numerical integrators (Chak and Monmarché, 2023; Schuh and Whalley, 2024), providing direct insights into the long-time error of numerical methods. Another widely used result is Harris' ergodic theorem (Mattingly et al., 2002; Hairer and Mattingly, 2011; Sanz-Serna and Palencia, 1985), which is also proven using coupling techniques. Harris' theorem is particularly valuable as it characterizes the invariant distribution through Lyapunov conditions, offering a versatile framework for establishing ergodicity in a wide range of stochastic systems.

For non-degenerate diffusion processes, functional inequalities (Bakry et al., 2014; Wang, 2006) are a widely used PDE-based approach to establish ergodicity. For instance, the Poincaré inequality and the log-Sobolev inequality lead to exponential convergence in the sense of χ^2 divergence and relative entropy (KL divergence), respectively. For degenerate diffusion processes, Villani's hypocoercivity framework (Villani, 2009a) is a prominent method that employs a modified H^1 norm to derive ergodicity by capturing both diffusion and transport effects. Alternatively, the generalized Γ calculus (Monmarché, 2018, 2019) extends the hypocoercivity framework by leveraging functional inequalities, offering a versatile tool for analyzing the ergodicity of degenerate systems.

In this work, we primarily utilize the generalized Γ calculus to analyze PIMD and the reflection coupling technique for RBM. Additionally, we introduce the triangle inequality framework to investigate the long-time error of numerical methods.

2.4.1 Generalized Γ calculus

The generalized Γ calculus developed in Monmarché (2018, 2019) is a functional approach to study the ergodicity of stochastic processes with degenerate diffusions. In particular, the generalized Γ calculus can be applied on the underdamped Langevin dynamics. The generalized Γ calculus is based on the functional inequalities (Bakry et al., 2014; Wang, 2006), and can be viewed as a variant of Villani's hypocoercivty theory (Villani, 2009a).

The generalized Γ calculus is able to provide explicit convergence rates for complicated stochastic processes. Except for the PIMD considered in this thesis, we also note the applications of the generalized Γ calculus in the mean-field interacting particle system (Guillin and Monmarché, 2021) and the Langevin dynamics with singular potentials (Baudoin et al., 2021).

Standard results in functional inequalities We begin with reviewing the standard results in the functional inequalities. Let $(x_t)_{t\geqslant 0}$ be a diffusion process (in the sense of Chapter 1.10.1

of Bakry et al. (2014)) in $\mathbb{R}^d$, and $(P_t)_{t\geqslant 0}$ be the correspondingly Markov semigroup. As is stated in (2.1), this means for any positive test function $f(x)$ in $C^\infty(\mathbb{R}^d)$,

$$(P_t f)(x) = \mathbb{E}\big[f(x_t) : x_0 = x\big], \quad x \in \mathbb{R}^d.$$

Let $\mathcal{L}$ be the infinitesimal generator of $(x_t)_{t\geqslant 0}$, and π be the invariant distribution. Then $\mathcal{L}$ is a self-adjoint elliptic operator in the Hilbert space $L^2(\pi)$. To characterize the convergence of the diffusion process, we employ the relative entropy defined in (2.5):

$$\mathrm{Ent}_\pi(f) = \int_{\mathbb{R}^d} f \log f \mathrm{d}\pi - \int_{\mathbb{R}^d} f \mathrm{d}\pi \log \int_{\mathbb{R}^d} f \mathrm{d}\pi.$$

Next, we introduce a crucial notion in the functional inequalities, the Γ operator (also known as the carré du champ operator, see Chapter 1.4.2 of Bakry et al. (2014)).

Definition 2.1. *For a diffusion process with generator $\mathcal{L}$, the Γ_1, Γ_2 operators are defined by*

$$\Gamma_1(f,g) = \frac{1}{2}(\mathcal{L}(fg) - g\mathcal{L}f - f\mathcal{L}g), \tag{2.23}$$

$$\Gamma_2(f,g) = \frac{1}{2}(\mathcal{L}\Gamma_1(f,g) - \Gamma_1(f,\mathcal{L}g) - \Gamma_1(g,\mathcal{L}f)). \tag{2.24}$$

When the test functions $f = g$, we simply write $\Gamma_1(f) = \Gamma_1(f,f)$. Using the property of the diffusion operator $\mathcal{L}$, we have the following properties of the Γ_1 operator.

Lemma 2.1. *For any test function f, we have $\Gamma_1(f) \geqslant 0$, and thus the Cauchy inequality*

$$\Gamma_1(f,g)^2 \leqslant \Gamma_1(f)\Gamma_1(g). \tag{2.25}$$

If a is a smooth function in $\mathbb{R}^d$, then for any test functions f, g,

$$\mathcal{L}(a(f)) = a'(f)\mathcal{L}f + a''(f)\Gamma_1(f), \tag{2.26}$$

$$\Gamma_1(a(f),g) = a'(f)\Gamma_1(f,g). \tag{2.27}$$

The related discussions can be found in Chapter 1.4.2 of Bakry et al. (2014) and Lemma 6 of Monmarché (2019).

We note that the generator $\mathcal{L}$ can be conversely determined by the Γ_1 operator and the invariant distribution π. In fact, for any test functions f, g we have the equality

$$\int_{\mathbb{R}^d} g\mathcal{L}f \mathrm{d}\pi = -\int_{\mathbb{R}^d} \Gamma_1(f,g)\mathrm{d}\pi. \tag{2.28}$$

In this case, we say the generator $\mathcal{L}$ is determined by the pair (π, Γ_1). See the related discussions in Chapter 3.1 of Bakry et al. (2014).

The inequality relation of Γ_1 and Γ_2 provides a convenient condition to verify the log-Sobolev inequality (see Chapter 5 of Bakry et al. (2014) for reference).

Theorem 2.1. *For a diffusion process with generator $\mathcal{L}$, if there is a constant $\rho > 0$ such that*

$$\Gamma_2(f) \geqslant \rho\Gamma_1(f), \quad \text{for any test function } f, \tag{2.29}$$

then we have the log-Sobolev inequality (denoted as $LS(\rho)$)

$$\mathrm{Ent}_\pi(f) \leqslant \frac{1}{2\rho}\int_{\mathbb{R}^d}\frac{\Gamma_1(f)}{f}\mathrm{d}\pi, \quad \text{for any test function } f, \tag{2.30}$$

and thus the exponential decay of the relative entropy

$$\mathrm{Ent}_\pi(P_t f) \leqslant e^{-2\rho t}\mathrm{Ent}_\pi(f), \quad \forall t \geqslant 0. \tag{2.31}$$

The proof can be found in Proposition 5.7.1 and Theorem 5.2.1 of Bakry et al. (2014).

A useful property is the bounded perturbation, which provides a convenient criterion to prove the log-Sobolev inequality beyond directly verifying the inequality (2.29).

Theorem 2.2. *Suppose a diffusion process determined by (π, Γ_1) satisfies $LS(\rho)$ in $\mathbb{R}^d$. Let π' be a probability distribution in $\mathbb{R}^d$ such that*

$$\frac{1}{M} \leqslant \frac{\mathrm{d}\pi'}{\mathrm{d}\pi}(x) \leqslant M, \quad \forall x \in \mathbb{R}^d, \tag{2.32}$$

then the diffusion process determined by (π', Γ_1) satisfies $LS(M^{-2}\rho)$.

The proof can be found in Proposition 5.1.6 of Bakry et al. (2014). This result shows that the new diffusion process determined by (π', Γ_1) still satisfies the log-Sobolev inequality, however the convergence rate reduces from ρ to $M^{-2}\rho$.

Another useful property is the tensorization of the log-Sobolev inequalities, allowing to establish the log-Sobolev inequality in higher dimensions.

Theorem 2.3. *Suppose two diffusion processes determined by (π, Γ_1) and (π', Γ_1') satisfy $LS(\rho)$ and $LS(\rho')$ in $\mathbb{R}^d$ respectively, then the product diffusion process determined by $(\pi \otimes \pi', \Gamma_1 \oplus \Gamma_1')$ satisfies $LS(\min\{\rho, \rho'\})$ in $\mathbb{R}^{2d}$.*

The proof can be found in Proposition 5.2.7 of Bakry et al. (2014). This result implies the convergence rate of the product diffusion process is determined by the smaller one in ρ and ρ'.

Generalized Γ operator The generalized Γ operator is the central tool in the generalized Γ calculus, and can be viewed as the generalization of the Γ_1 and Γ_2 operators.

Definition 2.2. *Let $\Phi(f)$ is a local operator on the test functions, namely, $\Phi(f)$ only depends on the value and derivatives of f. For a diffusion process with generator $\mathcal{L}$, define the generalized Γ operator by*

$$\Gamma_\Phi(f) = \frac{1}{2}(\mathcal{L}\Phi(f) - \mathrm{d}\Phi(f)\cdot\mathcal{L}f), \tag{2.33}$$

where $\mathrm{d}\Phi(f)\cdot g$ for two test functions f, g is defined as

$$\mathrm{d}\Phi(f)\cdot g = \lim_{s\to 0}\frac{\Phi(f+sg)-\Phi(f)}{s}. \tag{2.34}$$

When $\Phi(f)$ is a quadratic form, it is convenient to obtain the expression of $\Gamma_\Phi(f)$.

Lemma 2.2. *Suppose C_1, C_2 are two linear operators and $\Phi(f) = C_1 f\cdot C_2 f$, then*

$$\Gamma_\Phi(f) = \Gamma_1(C_1 f, C_2 f) + \frac{1}{2}C_1 f\cdot[\mathcal{L}, C_2]f + \frac{1}{2}[\mathcal{L}, C_1]f\cdot C_2 f, \tag{2.35}$$

where $[\cdot,\cdot]$ is the operator commutator.

The following proof is adapted from Lemma 5 of Monmarché (2019).

Proof of Lemma 2.2. Use $\mathrm{d}\Phi(f)\cdot g = C_1 f\cdot C_2 g + C_1 g\cdot C_2 f$ from Definition 2.2. ∎

Using Lemmas 2.1 & 2.2, it is convenient to derive some useful properties of $\Gamma_\Phi(f)$.

Lemma 2.3. *If $\Phi(f) = f\log f$, then*

$$\Gamma_\Phi(f) = \Gamma_{\cdot\log(\cdot)} = \frac{\Gamma_1(f)}{2f}. \tag{2.36}$$

If $\Phi(f) = |Cf|^2$ for some linear operator C, then

$$\Gamma_\Phi(f) = \Gamma_{|C\cdot|^2}(f) \geqslant Cf\cdot[\mathcal{L}, C]f. \tag{2.37}$$

If $\Phi(f) = |Cf|^2/f$ for some linear operator C, then

$$\Gamma_\Phi(f) = \Gamma_{|C\cdot|^2/\cdot}(f) \geqslant \frac{Cf\cdot[\mathcal{L}, C]f}{f}. \tag{2.38}$$

The following proof is partially adapted from Lemma 7 of Monmarché (2019).

Proof of Lemma 2.3. To prove (2.36), we use Definition 2.2 and derive

$$\Gamma_{\cdot\log(\cdot)}(f) = \frac{1}{2}\big(\mathcal{L}(f\log f) - \mathrm{d}(f\log f)\cdot\mathcal{L}f\big). \tag{2.39}$$

Using the definition of the Γ_1 operator, we have

$$\mathcal{L}(f\log f) = \mathcal{L}f\cdot\log f + f\mathcal{L}(\log f) + 2\Gamma_1(f, \log f)$$

$$= \mathcal{L}f \cdot \log f + f\left(\frac{1}{f}\mathcal{L}f - \frac{1}{f^2}\Gamma_1(f)\right) + \frac{2}{f}\Gamma_1(f,f)$$
$$= (1+\log f)\mathcal{L}f + \frac{1}{f}\Gamma_1(f).$$

On the other hand,

$$\mathrm{d}(f\log f)\cdot\mathcal{L}f = (1+\log f)\mathcal{L}f.$$

Hence from (2.39) we obtain the desired equality (2.36).

To prove (2.37), we only need to choose $C_1 = C_2 = C$ in Lemma 2.2, then

$$\Gamma_{|C\cdot|^2}(f) = \Gamma_1(Cf) + Cf\cdot[\mathcal{L},C]f \geqslant Cf\cdot[\mathcal{L},C]f, \tag{2.40}$$

which completes the proof.

The proof of (2.38) is a bit more complicated. For the operator $|C\cdot|^2/\cdot$, we have

$$\begin{aligned}\mathcal{L}\left(\frac{|Cf|^2}{f}\right) &= \frac{1}{f}\mathcal{L}(|Cf|^2) + |Cf|^2\mathcal{L}\left(\frac{1}{f}\right) + 2\Gamma_1\left(|Cf|^2,\frac{1}{f}\right)\\ &= \frac{1}{f}\mathcal{L}(|Cf|^2) + |Cf|^2\left(-\frac{\mathcal{L}f}{f^2} + \frac{2}{f^3}\Gamma_1(f)\right) + \frac{4Cf\cdot\Gamma_1(Cf,f)}{f^2},\end{aligned} \tag{2.41}$$

and

$$\mathrm{d}\left(\frac{|Cf|^2}{f}\right)\cdot\mathcal{L}f = \frac{\mathrm{d}(|Cf|^2)\cdot\mathcal{L}f}{f^2} - |Cf|^2\frac{Lf}{f^2}. \tag{2.42}$$

Hence from (2.41), (2.42) and Definition 2.2, we obtain

$$\begin{aligned}\Gamma_{|C\cdot|^2/\cdot}(f) &= \frac{1}{2}\left[\mathcal{L}\left(\frac{|Cf|^2}{f}\right) - \mathrm{d}\left(\frac{|Cf|^2}{f}\right)\cdot Lf\right]\\ &= \frac{\mathcal{L}(|Cf|^2) - \mathrm{d}(|Cf|^2)\cdot\mathcal{L}f}{2f} + \frac{|Cf|^2\Gamma_1(f)}{f^3} + \frac{2Cf\cdot\Gamma_1(Cf,f)}{f^2}\\ &= \frac{\Gamma_{|C\cdot|^2}(f)}{f} + \frac{|Cf|^2\Gamma_1(f)}{f^3} + \frac{2Cf\cdot\Gamma_1(Cf,f)}{f^2}.\end{aligned}$$

Then using the expression of $\Gamma_{|C\cdot|^2}(f)$ in (2.40) and Cauchy inequality in (2.25),

$$\begin{aligned}\Gamma_{|C\cdot|^2/\cdot}(f) &\geqslant \frac{\Gamma_1(Cf) + Cf\cdot[\mathcal{L},C]f}{f} + \frac{|Cf|^2\Gamma_1(f,f)}{f^3}\\ &\quad - 2\sqrt{\frac{|Cf|^2\Gamma_1(f,f)}{f^3}}\sqrt{\frac{\Gamma_1(Cf)}{f}} \geqslant \frac{Cf\cdot[\mathcal{L},C]f}{f}.\end{aligned}$$

Hence we obtain the desired inequality (2.38). ■

Next we show that the generalized Γ operator $\Gamma_\Phi(f)$ can be related to the time derivative of the local operator Φ.

Lemma 2.4. *Given the constant $t > 0$, for any $s \in [0, t]$, we have the equality*

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[P_s\Phi(P_{t-s}f)(x)\right] = 2P_s\Gamma_\Phi(P_{t-s}f)(x).$$

As a consequence, for any $t \geqslant 0$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d}\Phi(P_t f)\mathrm{d}\pi = -2\int_{\mathbb{R}^d}\Gamma_\Phi(P_t f)\mathrm{d}\pi.$$

Proof of Lemma 2.4. Note that the Markov semigroup $P_t = e^{t\mathcal{L}}$. Using the chain rule, we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s}\left[P_s\Phi(P_{t-s}f)\right] &= \mathcal{L}P_s\Phi(P_{t-s}f) + P_s\frac{\mathrm{d}}{\mathrm{d}s}\left[\Phi(P_{t-s}f)\right] \\
&= \mathcal{L}P_s\Phi(P_{t-s}f) + P_s\lim_{r\to 0}\frac{\Phi(P_{t-s-r}f) - \Phi(P_{t-s}f)}{r} \\
&= \mathcal{L}P_s\Phi(P_{t-s}f) - P_s\mathrm{d}\Phi(P_{t-s}f)\cdot\mathcal{L}P_{t-s}f \\
&= P_s\left(\mathcal{L}\Phi_{t-s}f - \mathrm{d}\Phi(P_{t-s}f)\cdot\mathcal{L}P_{t-s}f\right) \\
&= 2P_s\Gamma_\Phi(P_{t-s}f).
\end{aligned}$$

Integrating the equality over the distribution π, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}s}\int_{\mathbb{R}^d}P_s\Phi(P_{t-s}f)\mathrm{d}\pi = 2\int_{\mathbb{R}^d}\Gamma_\Phi(P_{t-s}f)\mathrm{d}\pi. \tag{2.43}$$

Since π is the invariant distribution, replacing $t - s$ by s in (2.43), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}s}\int_{\mathbb{R}^d}\Phi(P_t f)\mathrm{d}\pi = -2\int_{\mathbb{R}^d}\Gamma_\Phi(P_t f)\mathrm{d}\pi,$$

which completes the proof. ∎

Convergence in entropy-like functional Now we derive the main theorem in the generalized Γ calculus, which states that the entropy-like functional $W_\pi(f)$ exhibits exponential decay under specific functional inequalities. This result connects the geometric properties of the underlying stochastic process with the rate of convergence to equilibrium.

Theorem 2.4. *Let $(x_t)_{t\geqslant 0}$ be a diffusion process with the invariant distribution π. If for two local operators $\Phi_1(f)$ and $\Phi_2(f)$, there hold the functional inequalities*

$$0 \leqslant \int_{\mathbb{R}^d}\Phi_1(f)\mathrm{d}\pi - \Phi_1\left(\int_{\mathbb{R}^d}f\mathrm{d}\pi\right) \leqslant c\int_{\mathbb{R}^d}\Phi_2(f)\mathrm{d}\pi, \tag{2.44}$$

$$\Gamma_{\Phi_2}(f) \geqslant \rho\Phi_2(f) - m\Gamma_{\Phi_1}(f), \tag{2.45}$$

for some constants $c, \rho, m > 0$, then by defining the entropy-like functional

$$W_\pi(f) = m\left(\int_{\mathbb{R}^d}\Phi_1(f)\mathrm{d}\pi - \Phi_1\left(\int_{\mathbb{R}^d}f\mathrm{d}\pi\right)\right) + \int_{\mathbb{R}^d}\Phi_2(f)\mathrm{d}\pi, \tag{2.46}$$

we have the exponential decay

$$W_\pi(P_t f) \leqslant \exp\left(-\frac{2\rho t}{1+mc}\right) W_\pi(f), \quad \forall t \geqslant 0.$$

The following proof is adapted from Lemma 3 of Monmarché (2019).

Proof of Theorem 2.4. Using Lemma 2.4 and (2.45), we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} W_\pi(P_t f) &= \frac{\mathrm{d}}{\mathrm{d}t}\left[m\int_{\mathbb{R}^d} \Phi_1(P_t f)\mathrm{d}\pi + \int_{\mathbb{R}^d} \Phi_2(P_t f)\mathrm{d}\pi\right] \\ &= -2\int_{\mathbb{R}^d} \left(m\Gamma_{\Phi_1} + \Gamma_{\Phi_2}\right)(P_t f)\mathrm{d}\pi \leqslant -2\rho \int_{\mathbb{R}^d} \Phi_2(P_t f)\mathrm{d}\pi. \end{aligned} \tag{2.47}$$

Using (2.44) and the definition of $W_\pi(f)$, we have

$$W_\pi(f) = m\left(\int_{\mathbb{R}^d} \Phi_1(f)\mathrm{d}\pi - \Phi_1\left(\int_{\mathbb{R}^d} f\mathrm{d}\pi\right)\right) + \int_{\mathbb{R}^d} \Phi_2(f)\mathrm{d}\pi \leqslant (1+mc)\int_{\mathbb{R}^d} \Phi_2(f)\mathrm{d}\pi.$$

Hence (2.47) implies

$$\frac{\mathrm{d}}{\mathrm{d}t} W_\pi(P_t f) \leqslant -\frac{2\rho}{1+mc} W_\pi(P_t f), \quad \forall t \geqslant 0,$$

yielding the desired result. ∎

Example: ergodicity of Langevin dynamics We employ the functional inequalities and the generalized Γ calculus to study the ergodicity of two specific examples, the overdamped and underdamped Langevin dynamics. Let $V(x)$ be the potential function in $\mathbb{R}^d$ and $\pi(x) \propto e^{-V(x)}$ be the target distribution. The assumptions on $V(x)$ are provided as follows.

Assumption 2.1. *The potential function $V(x) \in C^2(\mathbb{R}^d)$ satisfies*

(i) $V(x)$ can be decomposed as $V^c(x) + V^b(x)$, and for some constants $a, M_1 > 0$,

$$\nabla^2 V^c(x) \succcurlyeq aI_d, \quad |V^b(x)| \leqslant M_1, \quad \forall x \in \mathbb{R}^d.$$

(ii) For some constant $M_2 > 0$,

$$-M_2 I_d \preccurlyeq \nabla^2 V(x) \preccurlyeq M_2 I_d, \quad \forall x \in \mathbb{R}^d.$$

Here, I_d donotes the identity matrix of size d, and $\preccurlyeq, \succcurlyeq$ define the Loewner order in symmetric matrices. Assumption (i) shows that $V(x)$ is the sum of a strongly convex potential and a globally bounded potential, and $V(x)$ itself does not require to be globally convex.

The overdamped Langevin dynamics for sampling the distribution $\pi(x)$ reads

$$\dot{x} = -\nabla V(x) + \sqrt{2}\dot{B}, \tag{2.48}$$

where B is the standard Brownian motion in $\mathbb{R}^d$. The following result states the exponential convergence of the overdamped Langevin dynamics (2.48) in the sense of the relative entropy. The proof is based on the functional inequalities.

Theorem 2.5. *Assume (i). Let $(P_t)_{t\geqslant 0}$ be the Markov semigroup of the overdamped Langevin dynamics (2.48), and $\pi \in \mathcal{P}(\mathbb{R}^d)$ be the invariant distribution. Then for any test function f in $\mathbb{R}^d$,*

$$\mathrm{Ent}_\pi(P_t f) \leqslant e^{-2\lambda_1 t}\mathrm{Ent}_\pi(f), \quad \forall t \geqslant 0, \tag{2.49}$$

where the convergence rate $\lambda_1 = e^{-2M_1}a > 0$.

Proof of Theorem 2.5. We first prove the log-Sobolev inequality for the dynamics

$$\dot{x} = -\nabla V^c(x) + \sqrt{2}\dot{B}, \tag{2.50}$$

where the potential function $V^c(x)$ is convex. The invariant distribution of (2.50) is clearly $\pi^c(x) \propto e^{-V^c(x)}$. By direct calculation, the Γ operators for (2.50) are given by

$$\begin{aligned}\Gamma_1(f,g) &= \nabla f \cdot \nabla g,\\ \Gamma_2(f,g) &= \nabla^2 f : \nabla^2 g + \nabla f \cdot \nabla V^c(x) \cdot \nabla g,\end{aligned}$$

where $A : B = \sum A_{ij}B_{ij}$ denotes the double inner product. Since $V^c(x)$ is convex, we have

$$\Gamma_2(f) \geqslant a\Gamma_1(f), \quad \text{for any test function } f, \tag{2.51}$$

which corresponds to the condition in (2.29). Therefore, applying Theorem 2.1, we obtain the log-Sobolev inequality $LS(a)$ for the overdamped Langevin dynamics (2.50) determined by (π^c, Γ_1). Next, note that the target distribution π satisfies the inequality

$$\frac{\pi(x)}{\pi^c(x)} = e^{-V^b(x)} \in \left[e^{-M_1}, e^{M_1}\right], \quad \forall x \in \mathbb{R}^d,$$

hence, by using the bounded perturbation property (Theorem 2.2), we conclude that $LS(e^{-2M_1}a)$ holds for the overdamped Langevin dynamics (2.48) determined by (π, Γ_1). ∎

The underdamped Langevin dynamics for sampling the distribution $\pi(x)$ reads

$$\begin{cases}\dot{x} = v,\\ \dot{v} = -\nabla V(x) - v + \sqrt{2}\dot{B},\end{cases} \tag{2.52}$$

and the invariant distribution is given by

$$\mu(x,v) \propto \exp\left(-\frac{|v|^2}{2} - V(x)\right).$$

The following result states the exponential convergence of the underdamped Langevin dynamics (2.52) in the sense of the relative entropy. The proof is based on the generalized Γ calculus.

Theorem 2.6. *Assume (i)(ii). Let $(P_t)_{t\geqslant 0}$ be the Markov semigroup of the underdamped Langevin dynamics (2.52), and $\mu \in \mathcal{P}(\mathbb{R}^{2d})$ be the invariant distribution. Define the entropy-like functional*

$$W_\mu(f) = 2(M_2+1)^2 \mathrm{Ent}_\mu(f) + \int_{\mathbb{R}^{2d}} \frac{|\nabla_v f|^2 + |\nabla_v f - \nabla_x f|^2}{f} \mathrm{d}\mu, \tag{2.53}$$

Then for any test function f in $\mathbb{R}^{2d}$,

$$W_\mu(P_t f) \leqslant e^{-2\lambda_2 t} W_\mu(f), \quad \forall t \geqslant 0, \tag{2.54}$$

where the convergence rate $\lambda_2 = \frac{1}{14}(M_2+1)^{-2}\min\{e^{-2M_1}a, 1\}$.

Proof of Theorem 2.6. Let $\sigma(v) \propto e^{-\frac{|v|^2}{2}}$ be the target distribution in the v variable. Define the Γ_1 operators in the v and x variables as

$$\Gamma_1^v(f,g) = \nabla_v f \cdot \nabla_v g, \quad \Gamma_1^x(f,g) = \nabla_x f \cdot \nabla_x g.$$

Since the potential function $\frac{1}{2}|v|^2$ is strongly convex in $\mathbb{R}^d$, the overdamped Langevin dynamics determined by (σ, Γ_1^v) satisfies $LS(1)$ in $\mathbb{R}^d$. According to Theorem 2.5, the overdamped Langevin dynamics determined by (π, Γ_1^x) satisfies $LS(e^{-2M_1}a)$ in $\mathbb{R}^d$. Using $\mu(x,v) = \pi(x)\otimes\sigma(v)$ and the tensorization of the log-Sobolev inequalities (Theorem 2.3), we conclude that the overdamped Langevin dynamics determined by (π, Γ_1) satisfies $LS(\min\{e^{-2M_1}a, 1\})$ in $\mathbb{R}^{2d}$, where the Γ_1 operator is defined by

$$\Gamma_1(f,g) = \Gamma_1^x(f,g) + \Gamma_1^v(f,g) = \nabla_x f \cdot \nabla_x g + \nabla_v f \cdot \nabla_v g.$$

The log-Sobolev inequality $LS(\min\{e^{-2M_1}a, 1\})$ explicitly reads

$$\min\{e^{-2M_1}a, 1\}\mathrm{Ent}_\mu(f) \leqslant \frac{1}{2}\int_{\mathbb{R}^d} \frac{|\nabla_x f|^2 + |\nabla_v f|^2}{f} \mathrm{d}\mu, \quad \text{for any test function } f. \tag{2.55}$$

The inequality (2.55) alone cannot establish the convergence of the underdamped Langevin dynamics (2.52) because it pertains to an overdamped Langevin dynamics.

To address this, we introduce the local operators

$$\Phi_1(f) = f\log f, \quad \Phi_2(f) = \frac{|\nabla_v f|^2 + |\nabla_v f - \nabla_x f|^2}{f} \tag{2.56}$$

for the underdamped Langevin dynamics (2.52). The construction of $\Phi_1(f)$ and $\Phi_2(f)$ is

directly inspired by Example 3 of Monmarché (2019). Using the inequality

$$\frac{|\nabla_x f|^2 + |\nabla_v f|^2}{3f} \leqslant \frac{|\nabla_v f|^2 + |\nabla_v f - \nabla_x f|^2}{f},$$

we derive from the log-Sobolev inequality (2.55) that

$$\frac{1}{3}\min\{e^{-2M_1}a, 1\}\left[\int_{\mathbb{R}^{2d}} \Phi_1(f)\mathrm{d}\mu - \Phi_1\left(\int_{\mathbb{R}^{2d}} f\mathrm{d}\mu\right)\right] \leqslant \int_{\mathbb{R}^{2d}} \Phi_2(f)\mathrm{d}\mu, \tag{2.57}$$

hence the inequality (2.44) holds with $c = 3\max\{e^{2M_1}a^{-1}, 1\}$.

Using Lemma 2.3, we deduce

$$\Gamma_{\Phi_1}(f) = \frac{|\nabla_v f|^2}{2f}, \quad \Gamma_{\Phi_2}(f) \geqslant \frac{\nabla_v f \cdot [\mathcal{L}, \nabla_v] f + (\nabla_v f - \nabla_x f) \cdot [\mathcal{L}, \nabla_v - \nabla_x] f}{f}. \tag{2.58}$$

Here, $\mathcal{L}$ is the infinitesimal generator of the underdamped Langevin dynamics (2.52),

$$\mathcal{L} = v \cdot \nabla_x - (\nabla V(x) + v) \cdot \nabla_v + \Delta_v.$$

The operator commutators are given by

$$[\mathcal{L}, \nabla_v] = \nabla_v - \nabla_x, \quad [\mathcal{L}, \nabla_x] = \nabla^2 V(x) \cdot \nabla_v.$$

Using the estimate of $\Gamma_{\Phi_2}(f)$ in (2.58), we obtain

$$\Gamma_{\Phi_2}(f) \geqslant \frac{(2\nabla_v f - \nabla_x f) \cdot (\nabla_v f - \nabla_x f) - (\nabla_v f - \nabla_x f) \cdot \nabla^2 V \cdot \nabla_v f}{f}. \tag{2.59}$$

For convenience, let

$$\mathcal{P} = \nabla_v f - \nabla_x f \in \mathbb{R}^d, \quad \mathcal{Q} = \nabla_v f \in \mathbb{R}^d,$$

so that

$$\Gamma_{\Phi_1}(f) = \frac{|\mathcal{Q}|^2}{2f}, \quad \Phi_2(f) = \frac{|\mathcal{P}|^2 + |\mathcal{Q}|^2}{f}, \quad \Gamma_{\Phi_2}(f) \geqslant \frac{(\mathcal{P} + \mathcal{Q}) \cdot \mathcal{P} - \mathcal{P} \cdot \nabla^2 V \cdot \mathcal{Q}}{f}.$$

Using the boundedness of $\nabla^2 V$ in Assumption (ii), we derive from (2.59) that

$$\Gamma_{\Phi_2}(f) \geqslant \frac{|\mathcal{P}|^2 - (M_2 + 1)|\mathcal{P}||\mathcal{Q}|}{f}. \tag{2.60}$$

Combining (2.58) and (2.60), we obtain the functional inequality

$$\Gamma_{\Phi_2}(f) - \frac{1}{2}\Phi_2 + 2(M_2+1)^2\Gamma_{\Phi_1}(f) \geqslant \frac{\frac{1}{2}|\mathcal{P}|^2 - (M_2 + 1)|\mathcal{P}||\mathcal{Q}| + \frac{1}{2}(M_2 + 1)^2|\mathcal{Q}|^2}{f} \geqslant 0, \tag{2.61}$$

which implies we can take $\rho = \frac{1}{2}$ and $m = 2(M_2 + 1)^2$ in (2.45). According to Theorem 2.4,

the convergence rate in the entropy-like functional $W_\mu(f)$ satisfies

$$\lambda_2 = \frac{\rho}{1+mc} \geqslant \frac{1}{14}(M_2+1)^{-2}\min\{e^{-2M_1}a, 1\},$$

which completes the proof. ■

Remark 2.3. *It can be observed that the convergence rate obtained in the underdamped Langevin dynamics (2.52) is smaller than the overdamped Langevin dynamics (2.48). Technically, this is because the generalized Γ calculus still relies on the functional inequalities on the overdamped Langevin dynamics. In practice, however, it is believed that underdamped Langevin dynamics has a faster convergence rate Ma et al. (2021), which has been rigorously justified in Lu and Wang (2020).*

2.4.2 Reflection coupling technique

Overview of reflection coupling Reflection coupling is a probabilistic technique used to establish the ergodicity of stochastic processes and the contractivity of numerical integrators, particularly when the potential function is non-convex. While synchronous coupling is typically applied to strongly convex functions, reflection coupling focuses on coupling the Brownian motions of two copies of a stochastic process and proving contractivity with respect to an appropriate metric.

Initially, reflection coupling was employed to demonstrate the ergodicity of the overdamped Langevin dynamics (Eberle, 2011, 2016). With refined coupling schemes, it has also been used to establish the ergodicity of Andersen dynamics (Bou-Rabee and Eberle, 2022), Hamiltonian Monte Carlo (Bou-Rabee et al., 2020; Bou-Rabee and Eberle, 2021, 2023), and underdamped Langevin dynamics (Eberle et al., 2019). More recently, reflection coupling has been applied to verify the contractivity of numerical integrators, such as generalized Hamiltonian Monte Carlo (Gouraud et al., 2022; Chak and Monmarché, 2023) and the UBU integrator (Schuh and Whalley, 2024). Stochastic gradient integrators have also been analyzed using similar techniques in Li et al. (2023); Leimkuhler et al. (2024).

Construction of distance function To illustrate the principle of reflection coupling, consider the following overdamped Langevin dynamics in $\mathbb{R}^d$:

$$\dot{x}_t = b(x_t) + \sigma\dot{B}_t, \tag{2.62}$$

where $b(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ represents the drift force, $\sigma > 0$ is a scalar constant, and $(B_t)_{t\geq 0}$ denotes standard Brownian motion in $\mathbb{R}^d$. The contraction property of the drift force $b(\cdot)$ is

described by the following function:

$$\kappa(r) = \inf\left\{ -\frac{2}{\sigma^2}\frac{(x-y)\cdot(b(x)-b(y))}{|x-y|^2} : x, y \in \mathbb{R}^d, |x-y| = r \right\}. \tag{2.63}$$

If $b(x) = -\nabla U(x)$ is in gradient form, $\kappa(r)$ characterizes the convexity of the potential function $U(x)$. Specifically, if $\nabla^2 U(x)$ is strongly convex outside a finite ball, $\kappa(r)$ becomes positive for sufficiently large r. This leads to the following reasonable assumption about $\kappa(r)$.

Assumption 2.2. *For the drift function $b(x)$, the function $\kappa(r)$ defined in (2.63) satisfies*

- *$\kappa(r)$ is continuous for $r \in (0, +\infty)$;*
- *$\kappa(r)$ has a lower bound for $r \in (0, +\infty)$;*
- *$\varliminf_{r\to\infty} \kappa(r) > 0$.*

Assumption 2.2 allows to construct a special distance function $f(r)$ satisfying a differential inequality related to Itô calculus.

Lemma 2.5. *Under Assumption 2.2, there exists a function $f(r)$ in $r \in [0, +\infty)$ satisfying*

- *$f(0) = 0$, and $f(r)$ is concave and strictly increasing in $[0, +\infty)$;*
- *$f(r) \in C^2[0, +\infty)$ and there exists a constant $c_0 > 0$ such that*

$$f''(r) - \frac{1}{4} r\kappa(r) f'(r) \leqslant -\frac{c_0}{2} f(r), \quad \forall r \geqslant 0. \tag{2.64}$$

- *There exists a constant $\varphi_0 > 0$ such that*

$$\varphi_0 r \leqslant f(r) \leqslant r, \quad \forall r \geqslant 0. \tag{2.65}$$

The constants c_0, φ_0 only depend on the function $\kappa(r)$.

Proof of Lemma 2.5. Utilizing the positivity of $\kappa(r)$, define the constants $R_0, R_1 \geqslant 0$ by

$$R_0 := \inf\{R \geqslant 0 : \kappa(r) \geqslant 0, \forall r \geqslant R\},$$
$$R_1 := \inf\{R \geqslant R_0 : \kappa(r)R(R - R_0) \geqslant 16, \forall r \geqslant R\}.$$

Then we then have $\kappa(r) \geqslant 0$ for $r \geqslant R_0$ and $\kappa(r)R_1(R_1 - R_0) \geqslant 16$ for $r \geqslant R_1$. Given the function $\kappa(r)$, define the auxiliary functions $\varphi(r), \Phi(r), g(r)$ by

$$\varphi(r) = \exp\left(-\frac{1}{4}\int_0^r s\kappa(s)^- \mathrm{d}s \right), \quad \Phi(r) = \int_0^r \varphi(s)\mathrm{d}s,$$
$$g(r) = \begin{cases} 1 - \frac{1}{2}\int_0^r \frac{\Phi(s)}{\varphi(s)}\mathrm{d}s \Big/ \int_0^{R_1} \frac{\Phi(s)}{\varphi(s)}\mathrm{d}s, & r \leqslant R_1, \\ \frac{1}{2} - \frac{\eta(r - R_1)}{1 + 4\eta(r - R_1)}, & r > R_1, \end{cases}$$

where $x^- = -\min\{x, 0\}$ is the negative part of $x \in \mathbb{R}$ and the constant $\eta > 0$ is defined by

$$\eta = -g'(R_1) = \frac{1}{2}\frac{\Phi(R_1)}{\varphi(R_1)} \Big/ \int_0^{R_1} \frac{\Phi(s)}{\varphi(s)} \mathrm{d}s.$$

Finally, the distance function $f(r)$ is defined as

$$f(r) = \int_0^r \varphi(s) g(s) \mathrm{d}s. \tag{2.66}$$

The only difference of the construction of above from the original one in Eberle (2011) is the definition of $g(r)$ for $r > R_1$. In our choice, $g(r)$ is differentiable at $r = R_1$ so that $f(r) \in C^2$, while in the original proof $f(r) \in C^1$ and $f'(r)$ is absolutely continuous.

It is easy to verify the following properties of the functions $f(r), \varphi(r), \Phi(r)$ and $g(r)$:

1. $0 < \varphi(r) \leqslant 1$, $\frac{1}{4} \leqslant g(r) \leqslant 1$. $\varphi(0) = g(0) = 1$. $\Phi(0) = 0$.
2. The derivatives of φ and g are given by

$$\varphi'(r) = -\frac{1}{4} r \kappa(r)^- \varphi(r), \quad g'(r) = -\frac{1}{2}\frac{\Phi(r)}{\varphi(r)} \Big/ \int_0^{R_1} \frac{\Phi(s)}{\varphi(s)} \mathrm{d}s, \quad 0 \leqslant r \leqslant R_1.$$

   Hence $\varphi'(0) = g'(0) = 0$ and $\varphi'(r) \leqslant 0, g'(r) \leqslant 0$ for all $r \geqslant 0$.

3. The second derivative of $f(r)$ is given by

$$f''(r) = \varphi(r) g'(r) + \varphi'(r) g(r) \leqslant 0, \tag{2.67}$$

   which implies $f(r)$ is concave for all $r \geqslant 0$.

4. When $r > R_0$, $\varphi(r)$ equals to a constant φ_1 given by

$$\varphi(r) \equiv \varphi_1 = \exp\left(-\frac{1}{4} \int_0^{R_0} s \kappa(s)^- \mathrm{d}s \right),$$

   Since $\varphi(r) \geqslant \varphi_1$ and $g(r) \geqslant \frac{1}{4}$ for all $r \geqslant 0$, we obtain

$$f'(r) = \varphi(r) g(r) \geqslant \frac{\varphi_1}{4} \Longrightarrow f(r) \geqslant \frac{\varphi_1}{4} r. \tag{2.68}$$

   Denote the constant $\varphi_0 = \frac{\varphi_1}{4}$, then we have $f(r) \geqslant \varphi_0 r$ for any $r \geqslant 0$.

5. Since $g(r) \leqslant 1$, $\Phi(r)$ provides an upper bound of $f(r)$:

$$\Phi(r) = \int_0^r \varphi(s) \mathrm{d}s \geqslant \int_0^r \varphi(s) g(s) \mathrm{d}s = f(r). \tag{2.69}$$

   From $\Phi''(r) = \varphi'(r) \leqslant 0$, we conclude $\Phi(r)$ is also concave for $r \in [0, +\infty)$.

Now we prove the inequality (2.64) with the constant c_0 defined by

$$\frac{1}{c_0} = \int_0^{R_1} \frac{\Phi(s)}{\varphi(s)} \mathrm{d}s. \tag{2.70}$$

1. When $r \leqslant R_1$, using the inequality $f(r) \leqslant \Phi(r)$,

$$\begin{aligned} f''(r) &= \varphi'(r)g(r) + \varphi(r)g'(r) \\ &= -\frac{1}{4}r\kappa(r)^-\varphi(r)g(r) - \frac{1}{2}\Phi(r)\bigg/\int_0^{R_1}\frac{\Phi(s)}{\varphi(s)}\mathrm{d}s \\ &\leqslant \frac{1}{4}r\kappa(r)f'(r) - \frac{1}{2}f(r)\bigg/\int_0^{R_1}\frac{\Phi(s)}{\varphi(s)}\mathrm{d}s, \end{aligned}$$

   hence (2.64) holds with c_0 defined in (2.70).

2. When $r > R_1$, we have $f'(r) \geqslant \frac{\varphi_1}{4}$ and $f''(r) \leqslant 0$. Hence by the definition of R_1 and the concavity of $\Phi(r)$ with $\Phi(0) = 0$, we obtain the inequality

$$f''(r) - \frac{1}{4}r\kappa(r)f'(r) \leqslant -\frac{1}{16}r\kappa(r)\varphi_0 \leqslant -\frac{\varphi_1}{R_1 - R_0}\frac{r}{R_1} \leqslant -\frac{\varphi_1}{R_1 - R_0}\frac{\Phi(r)}{\Phi(R_1)}. \tag{2.71}$$

   Since $\varphi(r) \equiv \varphi_1$ for $r \geqslant R_0$, the primitive function $\Phi(r)$ is linear in r, i.e.,

$$\Phi(r) = \Phi(R_0) + (r - R_0)\varphi_1, \quad r \geqslant R_0.$$

   In particular, $\Phi(R_1) = \Phi(R_0) + (R_1 - R_0)\varphi_1$, hence

$$\int_{R_0}^{R_1}\frac{\Phi(s)}{\varphi(s)}\mathrm{d}s = \frac{\Phi(R_0)}{\varphi_1}(R_1 - R_0) + \frac{1}{2}(R_1 - R_0)^2 \geqslant \frac{1}{2}(R_1 - R_0)\frac{\Phi(R_1)}{\varphi_1}. \tag{2.72}$$

   Combining the inequalities (2.71) and (2.72), we obtain

$$f''(r) - \frac{1}{4}r\kappa(r)f'(r) \leqslant -\frac{1}{2}\Phi(r)\bigg/\int_{R_0}^{R_1}\frac{\Phi(s)}{\varphi(s)}\mathrm{d}s \leqslant -\frac{1}{2}f(r)\bigg/\int_0^{R_1}\frac{\Phi(s)}{\varphi(s)}\mathrm{d}s, \tag{2.73}$$

   hence (2.64) holds with the constant c_0 defined in (2.70).

Finally, it is easy to see $\varphi_0 r \leqslant f(r) \leqslant r$ for any $r \geqslant 0$. ∎

Coupling to ergodicity To study the ergodicity of the overdamped Langevin dynamics (2.62), we define two copies $(x_t)_{t\geqslant 0}$ and $(\bar{x}_t)_{t\geqslant 0}$ of (2.62) in the following Brownian noise reflection coupling scheme:

$$\begin{cases} \dot{x}_t = b(x_t) + \sigma\dot{B}_t, \\ \dot{\bar{x}}_t = b(\bar{x}_t) + \sigma(I_d - 2e_t e_t^\top)\dot{B}_t, \end{cases} \quad \text{when } t < T, \tag{2.74}$$

where $T := \inf\{t \geqslant 0 : x_t = \bar{x}_t\}$ denotes the collision time of the two copies. For $t \geqslant T$, the two copies are defined to be identical, namely, $x_t = \bar{x}_t$. Moreover, e_t is the unit vector

$$e_t = \frac{x_t - \bar{x}_t}{|x_t - \bar{x}_t|} \in \mathbb{R}^d, \tag{2.75}$$

so that $I_d - 2e_t e_t^\top$ is the Householder transform Householder (1958) of x_t and $\bar{x}_t$, as shown in Figure 2.2.

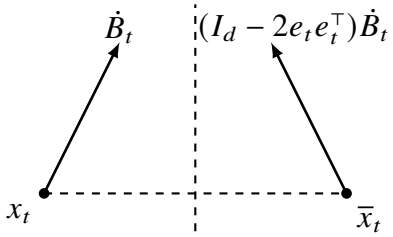


Figure 2.2 Graphic illustration of the reflection coupling.

Let $z_t = x_t - \bar{x}_t$ and $r_t = |z_t|$. Then z_t satisfies the SDE

$$\dot{z}_t = b(x_t) - b(\bar{x}_t) + 2\sigma\dot{W}_t, \tag{2.76}$$

where $(W_t)_{t\geqslant 0}$ is a one-dimensional Brownian motion defined by $\dot{W}_t = e_t^\top \dot{B}_t$.

Recall the distance function $f(r)$ defined in Lemma 2.5. Utilizing Itô calculus, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} f(r_t) = \frac{z_t}{r_t} \cdot (b(x_t) - b(\bar{x}_t)) f'(r_t) + 2\sigma^2 f''(r_t) + 2\sigma f'(r_t)\dot{W}_t. \tag{2.77}$$

Taking the expectation in both sides, we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[f(r_t)] &= \mathbb{E}\left[\frac{z_t}{r_t} \cdot (b(x_t) - b(\bar{x}_t)) f'(r_t) + 2\sigma^2 f''(r_t)\right] \\ &\leqslant \mathbb{E}\left[-\frac{\sigma^2}{2} r_t^2 f'(r_t) + 2\sigma^2 f''(r_t)\right] \leqslant -c_0\sigma^2\mathbb{E}[f(r_t)],\end{aligned}$$

and thus we have the exponential decay of $\mathbb{E}[f(r_t)]$:

$$\mathbb{E}[f(r_t)] \leqslant e^{-\beta t}\mathbb{E}[f(r_0)], \quad \forall t \geqslant 0, \tag{2.78}$$

where the convergence rate $\beta = c_0\sigma^2$.

Remark 2.4. *The positivity of σ is crucial for the validity of the reflection coupling (2.74), ensuring that the convergence rate β is also positive.*

The inequality (2.78) is sufficient to show the convergence in the Wasserstein-1 distance. Before presenting the main theorem, we introduce the Wasserstein-f distance corresponding to the distance function $f(r)$:

Definition 2.3. *For the distance function $f(r)$, define the Wasserstein-f distance as*

$$\mathcal{W}_f(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} \int f(|x-y|)\gamma(\mathrm{d}x\mathrm{d}y), \tag{2.79}$$

where $\Pi(\mu, \nu)$ denotes the set of joint distributions in $\mathbb{R}^d \times \mathbb{R}^d$ whose marginal distributions in the x, y variables are exactly μ, ν.

In particular, if $f(r) \equiv r$, the corresponding Wasserstein-f distance becomes the usual Wasserstein-1 distance $\mathcal{W}_1(\cdot, \cdot)$.

Since the distance function $f(r)$ defined in Lemma 2.5 satisfies $\varphi_0 r \leqslant f(r) \leqslant r$, the corresponding Wasserstein-f distance satisfies the inequality

$$\varphi_0 \mathcal{W}_1(\mu, \nu) \leqslant \mathcal{W}_f(\mu, \nu) \leqslant \mathcal{W}_1(\mu, \nu). \tag{2.80}$$

Remark 2.5. *In the field of optimal transport (Villani, 2009b), the joint distribution $\gamma \in \Pi(\mu, \nu)$ is commonly referred to as the transport plan between μ and ν.*

Remark 2.6. *Since the distance function $f(r)$ is concave, $\mathcal{W}_f(\cdot, \cdot)$ does not necessarily satisfy the triangle inequality and, therefore, does not define a metric space.*

Given the initial distributions $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, if the initial random variables $x_0 \sim \mu$ and $\bar{x}_0 \sim \nu$ are chosen such that

$$\mathbb{E}\big[f(|x_0 - \bar{x}_0|)\big] = \mathcal{W}_f(\mu, \nu),$$

then we immediately obtain the Wasserstein-f contractivity from the inequality (2.78).

Theorem 2.7. *Under Assumption 2.2, let $(\mathcal{P}_t)_{t \geqslant 0}$ be the dual semigroup of the overdamped Langevin dynamics (2.62), and $f(r), c_0, \varphi_0$ be defined as in Lemma 2.5. Then for any distributions $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, we have*

$$\mathcal{W}_f(\mu \mathcal{P}_t, \nu \mathcal{P}_t) \leqslant e^{-\beta t} \mathcal{W}_f(\mu, \nu), \quad \forall t \geqslant 0, \tag{2.81}$$

where the convergence rate $\beta = c_0 \sigma^2$. As a consequence,

$$\mathcal{W}_1(\mu \mathcal{P}_t, \nu \mathcal{P}_t) \leqslant \frac{1}{\varphi_0} e^{-\beta t} \mathcal{W}_1(\mu, \nu), \quad \forall t \geqslant 0. \tag{2.82}$$

Example: non-convex potential We consider a simple example of a non-convex potential function and calculate the convergence rate β using Lemma 2.5. Let the potential function $U(x)$ be defined as

$$U(x) = \frac{1}{2} x^2 + 2 \sin x, \quad x \in \mathbb{R}^d, \tag{2.83}$$

with the corresponding drift force given by $b(x) = -\nabla U(x) = -x - 2 \cos x$. Let the diffusion coefficient $\sigma = \sqrt{2}$, and the underlying overdamped Langevin dynamics (2.62) reads

$$\dot{x}_t = -x_t - 2 \cos x_t + \sqrt{2} \dot{B}_t. \tag{2.84}$$

For real numbers $x, y \in \mathbb{R}^d$ with $x - y = r \geqslant 0$, we have

$$\frac{b(x) - b(y)}{x - y} = 1 - \frac{4 \sin \frac{r}{2}}{r} \sin \frac{x + y}{2} \geqslant 1 - \frac{4 \sin \frac{r}{2}}{r},$$

hence the distance function $\kappa(r)$ in (2.63) is explicitly given by $\kappa(r) = 1 - \frac{4 \sin \frac{r}{2}}{r}$.

According to the proof of Lemma 2.5, the constants $R_0 \approx 3.790989$ and $R_1 \approx 6.307840$. Furthermore, we plot the functions $\kappa(r), \varphi(r)$ and $\Phi(r)$ in Figure 2.3.

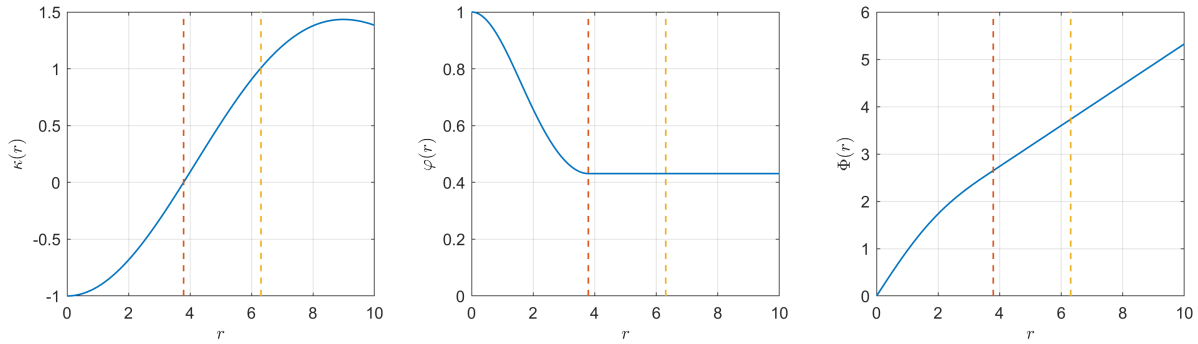


Figure 2.3 The functions $\kappa(r), \varphi(r)$ and $\Phi(r)$ in Lemma 2.5. The red and yellow dashed lines denote the constants R_0 and R_1.

Finally, the convergence rate β in Theorem 2.7 is calculated as

$$\beta = c_0 \sigma^2 = 2 \left(\int_0^{R_1} \frac{\Phi(s)}{\varphi(s)} \mathrm{d}s \right)^{-1} \approx 0.068906. \tag{2.85}$$

2.4.3 Triangle inequality framework

In general, the long-time error analysis for numerical methods is more challenging than the finite-time error analysis, and the latter being relatively standard and well-documented in textbooks, such as Chapter 7.5 of E et al. (2021). However, Shardlow and Stuart (2000); Mattingly et al. (2002, 2010) introduced a specialized approach—referred to in this paper as the triangle inequality framework—to address the difficulties of long-time error analysis. This framework has also been recently reviewed in Schuh and Souttar (2024) for the applications in the multiscale methods.

The core idea of the triangle inequality framework is straightforward: in addition to leveraging the results from finite-time error analysis, one only requires the geometric ergodicity of the stochastic dynamics to extend the analysis to the long-time regime. The basic idea can be summarized in the following lemma.

Lemma 2.6. *Let $(x_t)_{t \geqslant 0}$ and $(\tilde{x}_t)_{t \geqslant 0}$ be stochastic processes in $\mathbb{R}^d$ with dual semigroups $(\mathcal{P}_t)_{t \geqslant 0}$ and $(\tilde{\mathcal{P}}_t)_{t \geqslant 0}$ respectively. Given the metric $d(\cdot, \cdot)$ on $\mathcal{P}(\mathbb{R}^d)$, assume $(\mathcal{P}_t)_{t \geqslant 0}$ has an invariant*

distribution $\pi \in \mathcal{P}(\mathbb{R}^d)$ *and there exist constants* $C, \beta > 0$ *such that*

$$d(\nu\mathcal{P}_t, \pi) \leqslant Ce^{-\beta t} d(\nu, \pi), \quad \forall \nu \in \mathcal{P}(\mathbb{R}^d); \tag{2.86}$$

and for any $T > 0$*, there exists a constant* $\varepsilon(T)$ *such that*

$$\sup_{0 \leqslant t \leqslant T} d(\nu\mathcal{P}_t, \nu\tilde{\mathcal{P}}_t) \leqslant \varepsilon(T), \quad \forall \nu \in \mathcal{P}(\mathbb{R}^d). \tag{2.87}$$

Then there exist constants $T_0, \lambda > 0$ *such that*

$$d(\nu\tilde{\mathcal{P}}_t, \pi) \leqslant 2\varepsilon(T_0) + 2M_0 e^{-\lambda t}, \quad \forall t \geqslant 0, \tag{2.88}$$

where $M_0 := \sup_{s \in [0, T_0]} d(\nu\tilde{\mathcal{P}}_s, \pi)$.

In Lemma 2.6, $(\tilde{x}_t)_{t \geqslant 0}$ is interpreted as an approximation of $(x_t)_{t \geqslant 0}$. For instance, $(x_t)_{t \geqslant 0}$ could represent the solution to a SDE, while $(\tilde{x}_t)_{t \geqslant 0}$ corresponds to its time discretization. The inequality (2.86) characterizes the ergodicity of $(x_t)_{t \geqslant 0}$, with $\beta > 0$ denoting the convergence rate. Similarly, the inequality (2.87) measures the distance between the dual semigroups $(\mathcal{P}_t)_{t \geqslant 0}$ and $(\tilde{\mathcal{P}}_t)_{t \geqslant 0}$ over a finite time period. Typically, this finite-time error is determined by strong error estimates, where the error term $\varepsilon(T)$ usually vanishes as the step size of the discretization approaches zero.

Finally, the long-time error of $(\tilde{x}_t)_{t \geqslant 0}$ is characterized by the difference between the distributions $\nu\tilde{\mathcal{P}}_t$ and π. For sufficiently large t, the choice of T_0 remains constant. Thus, Lemma 2.6 demonstrates that only the ergodicity and finite-time error estimates are needed to deduce the long-time error.

Proof of Lemma 2.6. For any $T > 0$ and $t \geqslant T$, we have the inequality

$$\begin{aligned} d(\nu\tilde{\mathcal{P}}_t, \pi) &\leqslant d(\nu\tilde{\mathcal{P}}_{t-T}\tilde{\mathcal{P}}_T, \nu\tilde{\mathcal{P}}_{t-T}\mathcal{P}_T) + d(\nu\tilde{\mathcal{P}}_{t-T}\mathcal{P}_T, \pi\mathcal{P}_T) \\ &\leqslant \varepsilon(T) + Ce^{-\beta T} d(\nu\tilde{\mathcal{P}}_{t-T}, \pi). \end{aligned}$$

By choosing $T = T_0$ such that $Ce^{-\beta T_0} = \frac{1}{2}$, we have

$$d(\nu\tilde{\mathcal{P}}_t, \pi) \leqslant \varepsilon(T_0) + \frac{1}{2} d(\nu\tilde{\mathcal{P}}_{t-T_0}, \pi), \quad \forall t \geqslant T_0.$$

By induction on the integer $n \geqslant 0$, we obtain

$$d(\nu\tilde{\mathcal{P}}_t, \pi) \leqslant 2\left(1 - \frac{1}{2^n}\right)\varepsilon(T_0) + \frac{1}{2^n} d(\nu\tilde{\mathcal{P}}_{t-nT_0}, \pi), \quad \forall t \geqslant nT_0.$$

For any $t \in [0, +\infty)$, there exists a unique integer $n \geqslant 0$ such that $t \in [nT_0, (n+1)T_0)$. Then

$$d(\nu\tilde{\mathcal{P}}_t, \pi) \leqslant 2\varepsilon(T_0) + 2^{1-t/T_0} \sup_{s\in[0,T_0]} d(\nu\tilde{\mathcal{P}}_s, \pi),$$

which implies the long-time error estimate (2.88) with $\lambda = \ln 2/T_0$. ■

Remark 2.7. *In this work, the distance $d(\cdot,\cdot)$ is chosen as the Wasserstein-1 distance, and estimate of the finite-time error $\varepsilon(T)$ comes from the strong error estimate. As a consequence, we can only obtain half-order convergence in the step size h. In practice, $d(\cdot,\cdot)$ can also be chosen as the total variation, see Durmus and Moulines (2017) for example.*

Finally, we remark that the triangle inequality framework is remotely reminiscent of the well-known Lax equivalence theorem (Sanz-Serna and Palencia, 1985) in numerical analysis. Here, the ergodicity (2.86) serves as the stability and it helps translate the finite-time error estimate (2.87) to the long-time error estimate (2.88) without sacrificing the accuracy order.

Chapter 3 Dimension-free ergodicity of PIMD

Within the path integral formulation, the calculation of the quantum thermal average $\langle O(\hat{x})\rangle_\beta$ is reduced to a sampling problem for a classical Boltzmann distribution

$$\pi_N(x_1,\cdots,x_N) \propto \exp\left(-\frac{1}{2\beta_N}\sum_{j=1}^{N}|x_j-x_{j+1}|^2-\beta_N\sum_{j=1}^{N}V(x_j)\right), \tag{3.1}$$

where $\beta_N = \beta/N$ and we adimit the periodic boundary condition $x_{N+1} = x_1$. In this section, we first introduce the normal mode coordinates, which diagonalizes the spring potential $\sum_{j=1}^{N}|x_j - x_{j+1}|$ of the ring polymer. Second, we derive the preconditioned underdamped Langevin dynamics for sampling $\pi_N(x_1,\cdots,x_N)$. Finally, we prove the underlying Langevin dynamics has uniform-in-N ergodicity in the sense of the relative entropy.

3.1 Normal mode coordinates

Since the energy function $\mathcal{E}_N(x_1,\cdots,x_N)$ has a proper inifinite bead limit, it is natural to ask whether we can sample the corresponding Boltzmann distribution $\pi_N(x_1,\cdots,x_N)$ with a uniform-in-N convergence rate. However, the direct simulation of the ring polymer system is notorious for the stiffness in the ring polymer energy (Marx and Parrinello, 1996; Ceriotti et al., 2010)—the spring potential $\sum_{j=1}^{N}|x_j-x_{j+1}|^2$ has high mode frequencies when the number of beads N is large, making it impossible to stably discretize the dynamics with an $O(1)$ step size. In practice, there are three major approaches to resolve the stiffness issue: (i) precondition the dynamics to slow the time scale of high-frequencies modes (Durlak et al., 2009; Lu et al., 2020; Bou-Rabee and Eberle, 2021); (ii) apply staging coordinates to decouple the dynamics (Cao and Martyna, 1996; Liu et al., 2016); (ii) use strongly stable numerical integrators (Korol et al., 2019, 2020). In the thesis, we employ the preconditioning approach, which requires the normal mode coordinates to diagonalize the spring potential into different Fourier modes.

In order to diagonalize the spring potential, we note that the eigenvalues and eigenvectors

of the periodic Laplacian matrix

$$L = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 & -1 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ -1 & 0 & 0 & \cdots & -1 & 2 \end{bmatrix} \in \mathbb{R}^{N\times N}$$

are explicitly given by

$$\lambda_k = 4\sin^2\left(\frac{\pi k}{N}\right) \geqslant 0, \quad v_k = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 \\ e^{i\frac{2\pi k}{N}} \\ e^{i\frac{4\pi k}{N}} \\ \vdots \\ e^{i\frac{2(N-1)\pi k}{N}} \end{bmatrix} \in \mathbb{R}^{N\times 1}, \quad k = 0, 1, \cdots, N-1.$$

As a consequence, we introduce the following definition of the normal mode coordinates.

Definition 3.1. *Given the bead positions $\{x_j\}_{j=1}^N$ in $\mathbb{R}^d$, define the normal mode coordinates $\{\xi_k\}_{k=0}^{N-1}$ in $\mathbb{R}^d$ by*

$$\xi_k = \beta_N \sum_{j=1}^{N} x_j c_{j,k}, \tag{3.2}$$

where $\beta_N = \beta/N$, and $c_{j,k}$ is the discrete Fourier coefficient defined by the following rule:

- *If N is odd, then for each $j = 1, \cdots, N$,*

$$c_{j,0} = \frac{1}{\sqrt{\beta}},$$

$$c_{j,2k-1} = \sqrt{\frac{2}{\beta}}\sin\left(\frac{2\pi k j}{N}\right), \quad c_{j,2k} = \sqrt{\frac{2}{\beta}}\cos\left(\frac{2\pi k j}{N}\right), \quad k = 1, \cdots, \frac{N-1}{2}.$$

- *If N is even, then for each $j = 1, \cdots, N$,*

$$c_{j,0} = \frac{1}{\sqrt{\beta}}, \quad c_{j,N-1} = \frac{(-1)^j}{\sqrt{\beta}},$$

$$c_{j,2k-1} = \sqrt{\frac{2}{\beta}}\sin\left(\frac{2\pi k j}{N}\right), \quad c_{j,2k} = \sqrt{\frac{2}{\beta}}\cos\left(\frac{2\pi k j}{N}\right), \quad k = 1, \cdots, \frac{N}{2} - 1.$$

It is easy to verify $c_{j,k}$ satisfies the orthogonal condition:

$$\beta_N \sum_{j=1}^{N} c_{j,k} c_{j,k'} = \delta(k-k') = \begin{cases} 1, & \text{if } k = k', \\ 0, & \text{if } k \neq k', \end{cases} \tag{3.3}$$

then we obtain the formula for the inverse transform

$$x_j = \sum_{k=0}^{N-1} \xi_k c_{j,k}, \quad j = 1, \cdots, N, \tag{3.4}$$

and the spring potential can be equivalently written as

$$\frac{1}{2\beta_N} \sum_{j=1}^{N} |x_j - x_{j+1}|^2 = \frac{1}{2} \sum_{k=0}^{N-1} \omega_k^2 |\xi_k|^2. \tag{3.5}$$

Therefore, in the normal mode coordinates $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_N) \in \mathbb{R}^{dN}$, we can equivalently write the ring polymer energy function $\mathcal{E}_N(x_1, \cdots, x_N)$ as

$$\mathcal{E}_N(\boldsymbol{\xi}) = \frac{1}{2} \sum_{k=0}^{N-1} \omega_k^2 |\xi_k|^2 + \beta_N \sum_{j=1}^{N} V\left(\sum_{k=0}^{N-1} \xi_k c_{j,k} \right), \tag{3.6}$$

where the normal mode frequencies $\{\omega_k\}_{k=0}^{N-1}$ are given by

$$\omega_0 = 0, \quad \omega_{2k-1} = \omega_{2k} = \frac{2}{\beta_N} \sin\left(\frac{k\pi}{N}\right), \quad k = 1, \cdots, \left\lfloor \frac{N}{2} \right\rfloor. \tag{3.7}$$

The target Boltzmann distribution $\pi_N(x_1, \cdots, x_N)$ is now expressed as $\pi_N(\boldsymbol{\xi}) \propto e^{-\mathcal{E}_N(\xi)}$, defined in the normal mode coordinates $\{\xi_k\}_{k=0}^{N-1}$. In Figure 3.1, we plot the Fourier coefficients $c_{k,j}$ for the first five modes, along with the growth curve of the normal mode frequency ω_k.

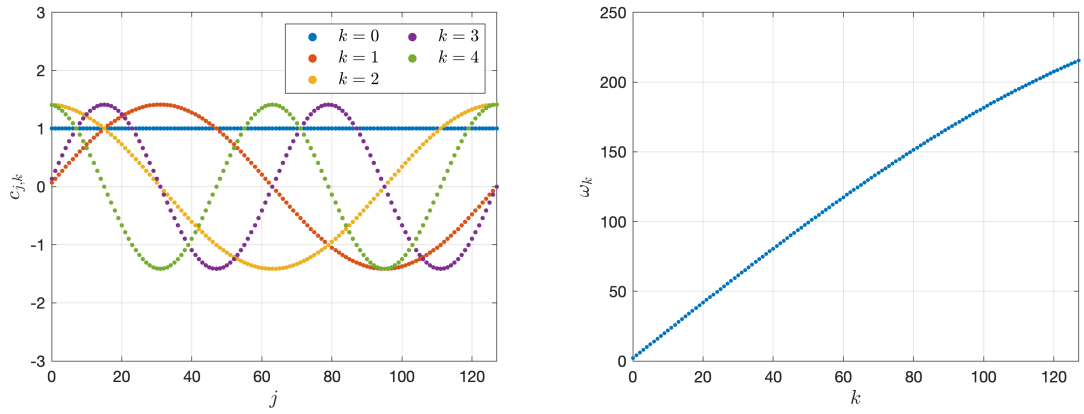


Figure 3.1 Left: Fourier coefficients $c_{j,k}$ in the first five modes. Right: Growth curve of the normal mode frequency ω_k. The number of beads $N = 128$, and the inverse temperature $\beta = 1$.

Remark 3.1. *The transform from the positions $\{x_j\}_{j=1}^{N}$ to the normal mode coordinates $\{\xi_k\}_{k=0}^{N-1}$ is orthogonal, hence the Jacobian matrix has a constant determinant. This means that to com-*

pute the quantum thermal average, sampling $\pi_N(x_1, \cdots, x_N)$ *in the positions and sampling* $\pi_N(\xi_0, \cdots, \xi_{N-1})$ *in the normal mode coordinates are equivalent.*

3.2 Preconditioned underdamped Langevin dynamics in PIMD

We derive the underdamped Langevin dynamics for sampling the target distribution $\pi_N(\boldsymbol{\xi})$ in the normal mode coordinates $\boldsymbol{\xi} = \{\xi_k\}_{k=0}^{N-1}$. The internal frequency of the k-th mode is $\omega_k = \frac{2}{\beta_N} \sin \frac{k\pi}{N} \sim \frac{2k\pi}{\beta}$; in particular, the frequency of the first mode is $\omega_0 = 0$. Inspired from Lu et al. (2020); Bou-Rabee and Eberle (2021), we introduce a constant $a > 0$ and define $V^a(x) = V(x) - \frac{a}{2}|x|^2$, then the energy function $\mathcal{E}_N(\boldsymbol{\xi})$ can be written as

$$\mathcal{E}_N(\boldsymbol{\xi}) = \frac{1}{2} \sum_{k=0}^{N-1} (\omega_k^2 + a)|\xi_k|^2 + \beta_N \sum_{j=1}^{N} V^a\left(\sum_{k=0}^{N-1} \xi_k c_{j,k} \right), \tag{3.8}$$

and the frequency of the first mode becomes $a > 0$.

To construct the underdamped Langevin dynamics sampling $\pi_N(\boldsymbol{\xi}) \propto e^{-\mathcal{E}_N(\boldsymbol{\xi})}$, we introduce the auxiliary momentum variables $\boldsymbol{\eta} = \{\eta_k\}_{k=0}^{N-1}$ and write

$$\begin{cases} \dot{\xi}_k = \eta_k, \\ \dot{\eta}_k = -(\omega_k^2 + a)\xi_k - \beta_N \sum_{j=1}^{N} \nabla V^a(x_j(\boldsymbol{\xi}))c_{j,k} - \eta_k + \sqrt{2}\dot{B}_k, \end{cases} \tag{3.9}$$

where $\{B_k\}_{k=1}^N$ are independent Brownian motions in $\mathbb{R}^d$. Although the underdamped Langevin dynamics (3.9) preserves $\pi_N(\boldsymbol{\xi})$ as the invariant distribution, it appears that (3.9) embeds strong stiffness due to frequency $\omega_k \sim \frac{2k\pi}{\beta}$ grows linearly with the mode index k. The preconditioning technique in Lu et al. (2020) modifies (3.9) by adding a scaling coefficient $(\omega_k^2 + a)^{-1}$ to the drift force, namely,

$$\begin{cases} \dot{\xi}_k = \eta_k, \\ \dot{\eta}_k = -\xi_k - \dfrac{\beta_N}{\omega_k^2 + a} \sum_{j=1}^{N} \nabla V^a(x_j(\boldsymbol{\xi}))c_{j,k} - \eta_k + \sqrt{\dfrac{2}{\omega_k^2 + a}}\dot{B}_k, \end{cases} \quad (k = 0, 1, \cdots, N-1) \tag{3.10}$$

where $\{x_j\}_{j=1}^N$ are determined by the normal mode transform (3.4). As k grows large, the dynamics of (ξ_k, η_k) in the preconditioned underdamped Langevin dynamics (3.10) is close to

$$\begin{cases} \dot{\xi}_k = \eta_k, \\ \dot{\eta}_k = -\xi_k - \eta_k + \sqrt{\dfrac{2}{\omega_k^2 + a}}\dot{B}_k, \end{cases} \quad (k = 0, 1, \cdots, N-1) \tag{3.11}$$

which is a linear stochastic process can be integrated with $O(1)$ step size. In this way, the stiffness embedded in the vanilla underdamped Langevin dynamics (3.9) is resolved.

Remark 3.2. *An alternative way to handle the stiffness is to use a strongly stable numerical integrator, for example the BCOCB integrator introduced in Korol et al. (2019). The BCOCB integrator directly discretizes the vanilla underdamped Langevin dynamics (3.9), while the discussion on the numerical stability is subtler compared to the preconditioning technique.*

Remark 3.3. *The preconditioning technique can also be applied without introducing the artificial parameter $a > 0$. For example, one may use the scaling coefficient ω_k^{-2} for each mode index $k \geqslant 1$, and the resulting preconditioned underdamped Langevin dynamics reads*

$$\begin{cases} \dot{\xi}_0 = \eta_0, \\ \dot{\eta}_0 = -\dfrac{\sqrt{\beta}}{N}\displaystyle\sum_{j=1}^{N} \nabla V(x_j(\boldsymbol{\xi})) - \eta_0 + \sqrt{2}\dot{B}_0, \\ \dot{\xi}_k = \eta_k, \\ \dot{\eta}_k = -\dfrac{\beta_N}{\omega_k^2}\displaystyle\sum_{j=1}^{N} \nabla V(x_j(\boldsymbol{\xi}))c_{j,k} - \eta_k + \dfrac{\sqrt{2}}{\omega_k}\dot{B}_k. \end{cases} \qquad (k = 1, \cdots, N-1) \tag{3.12}$$

The alternate Langevin dynamics (3.12) can be used when the potential $V(x)$ is defined in a periodic region. However, in this paper, we focus on the Langevin dynamics (3.10), as it is more convenient for studying ergodicity.

Remark 3.4. *In addition to Langevin thermostats, several other options are available for sampling the target Boltzmann distribution. Commonly used alternatives include the Hamiltonian Monte Carlo (HMC) method Betancourt (2017), the Andersen thermostat E and Li (2008), and the Nosé–Hoover thermostat Evans and Holian (1985). By comparison, the convergence analysis of Langevin dynamics is more extensively studied and understood.*

3.3 Assumptions and results

In this section, we present the assumptions and results in the ergodicity of the preconditioned underdamped Langevin dynamics (3.10). The invariant distribution of (3.10) is

$$\mu_N(\boldsymbol{\xi}, \boldsymbol{\eta}) \propto \exp\left\{-\frac{1}{2}\sum_{k=0}^{N-1}(\omega_k^2 + a)(|\xi_k|^2 + |\eta_k|^2) - \beta_N \sum_{j=1}^{N} V^a\left(\sum_{k=0}^{N-1}\xi_k c_{j,k}\right)\right\}, \tag{3.13}$$

whose marginal distribution in $\boldsymbol{\xi}$ is exactly the target distribution

$$\pi_N(\boldsymbol{\xi}) \propto \exp\left\{-\frac{1}{2}\sum_{k=0}^{N-1}(\omega_k^2+a)|\xi_k|^2 - \beta_N\sum_{j=1}^{N}V^a\left(\sum_{k=0}^{N-1}\xi_k c_{j,k}\right)\right\}, \tag{3.14}$$

For the convenience of the proof, we also introduce the overdamped form of (3.10):

$$\dot{\xi}_k = -\xi_k - \frac{\beta_N}{\omega_k^2+a}\sum_{j=1}^{N}\nabla V^a(x_j(\boldsymbol{\xi}))c_{j,k} + \sqrt{\frac{2}{\omega_k^2+a}}\dot{B}_k, \quad (k=0,1,\cdots,N-1) \tag{3.15}$$

where $\{B_k\}_{k=0}^{N-1}$ are independent Brownian motions in $\mathbb{R}^d$. It is clear that the invariant distribution of (3.15) is $\pi_N(\boldsymbol{\xi})$.

Assumption 3.1. *Given the constant $a>0$, the potential function*

$$V^a(x) = V(x) - \frac{a}{2}|x|^2, \quad x\in\mathbb{R}^d,$$

is twice differentiable in $\mathbb{R}^d$, and for some constants $M_1, M_2 \geqslant 0$,

(i) $V^a(x)$ can be decomposed as $V^c(x)+V^b(x)$, where $\nabla^2 V^c(x) \succcurlyeq O_d$ and $|V^b(x)| \leqslant M_1$ for any $x\in\mathbb{R}^d$;

(ii) $-M_2 I_d \preccurlyeq \nabla^2 V^a(x) \preccurlyeq M_2 I_d$ for any $x\in\mathbb{R}^d$.

Assumption (i) can be interpreted as: $V^a(x)$ is the sum of a globally convex potential $V^c(x)$ and a globally bounded potential $V^b(x)$.

The ergodicity results of the overdamped and underdamped Langevin dynamics are summarized in Table 3.1.

Dynamics	overdamped (3.15)	underdamped (3.10)
Assumption	(i) $\Rightarrow$ Theorem 3.1	(i)(ii) $\Rightarrow$ Theorem 3.2
Distribution	$\pi_N(\boldsymbol{\xi})$ in (3.14)	$\mu_N(\boldsymbol{\xi},\boldsymbol{\eta})$ in (3.13)
Ergodicity	$\mathrm{Ent}_{\pi_N}(P_t f) \leqslant e^{-2\lambda_1 t}\mathrm{Ent}_{\pi_N}(f)$	$W_{\mu_N}(P_t f) \leqslant e^{-2\lambda_2 t}W_{\mu_N}(f)$

Table 3.1 The uniform-in-N ergodicity of the Langevin dynamics in PIMD.

Here, the convergence rates λ_1 and λ_2 are explicitly given by

$$\lambda_1 = \exp(-4\beta M_1), \quad \lambda_2 = \frac{a^2}{3M_2^2+5a^2}\exp(-4\beta M_1).$$

We note that $\lambda_2 < \lambda_1$ is because the generalized Γ calculus—the proof technique for the underdamped Langevin dynamics (3.10)—is based on the log-Sobolev inequality for the overdamped Langevin dynamics (3.15). See the related discussions in Remark 2.3.

Remark 3.5. *In principle, the overdamped Langevin dynamics* (3.15) *can also be used to sample the target distribution* $\pi_N(\boldsymbol{\xi})$. *Nevertheless, the underdamped form* (3.10) *is preferred in the physics community because* (3.10) *is a direct generalization of the Hamiltonian dynamics. From the mathematical perspective, justifying the ergodicity of the underdamped Langevin dynamics is much more difficult than the overdamped case, because the infinitesimal generator is a hypoelliptic operator rather than a strongly elliptic one (Villani, 2009a).*

Remark 3.6. *The ergodicity of PIMD has also been established in its preconditioned Hamiltonian Monte Carlo (pHMC) formulation, as shown in Theorem 3.5 of Bou-Rabee and Eberle (2021). However, the time duration restriction imposed on pHMC, as specified in Equation (3.22), is quite stringent and challenging to verify in practical applications. In contrast, the underdamped Langevin dynamics* (3.10) *offers a more straightforward approach, as it can be directly discretized for computing the quantum thermal average.*

3.4 Uniform-in-N ergodicity of Langevin dynamics in PIMD

In this section we prove the uniform-in-N ergodicity of the overdamped Langevin dynamics (3.15) and the underdamped Langevin dynamics (3.10). A introduction of the proof techniques—the functional inequalities and the generalized Γ calculus—can be found in Chapter 2.4.1. In particular, the proofs of Theorem 2.5 and 2.6 provide the blueprints for the proofs of Theorem 3.1 and 3.2 in this section.

Theorem 3.1 (overdamped)**.** *Assume (i). Let* $(P_t)_{t\geqslant 0}$ *be the Markov semigroup of the overdamped Langevin dynamics* (3.15)*, then for any test function* $f(\boldsymbol{\xi})$ *in* $\mathbb{R}^{dN}$*,*

$$\mathrm{Ent}_{\pi_N}(P_t f) \leqslant e^{-2\lambda_1 t}\mathrm{Ent}_{\pi_N}(f), \quad \forall t \geqslant 0,$$

where the convergence rate $\lambda_1 = \exp(-4\beta M_1)$.

The proof is based on the log-Sobolev inequality for the distribution $\pi_N(\boldsymbol{\xi})$.

Proof of Theorem 3.1. Introduce the potential function of the ring polymer energy

$$\mathcal{V}_N^a(\boldsymbol{\xi}) = \beta_N \sum_{j=1}^{N} V^a(x_j(\boldsymbol{\xi})) = \beta_N \sum_{j=1}^{N} V^a\left(\sum_{k=0}^{N-1} \xi_k c_{j,k}\right), \tag{3.16}$$

then we can simplify the overdamped Langevin dynamics (3.15) as

$$\dot{\xi}_k = -\xi_k - \frac{1}{\omega_k^2 + a}\nabla_{\xi_k}\mathcal{V}_N^a(\boldsymbol{\xi}) + \sqrt{\frac{2}{\omega_k^2 + a}}\dot{B}_k. \tag{3.17}$$

Similar to the proof of Theorem 2.5, we prove the log-Sobolev inequality for (3.17) by studying its convex counterpart. The potential function corresponding to $V^c(x)$ is given by

$$\mathcal{V}_N^c(\boldsymbol{\xi}) = \beta_N \sum_{j=1}^{N} V^c(x_j(\boldsymbol{\xi})) = \beta_N \sum_{j=1}^{N} V^c\bigg(\sum_{k=0}^{N-1} \xi_k c_{j,k}\bigg).$$

Since $V^c(x)$ is globally convex in $\mathbb{R}^d$, we can deduce $\mathcal{V}_N^c(\boldsymbol{\xi})$ is globally convex in $\mathbb{R}^{dN}$. In fact, for any $\boldsymbol{\theta} = (\theta_0, \theta_1, \cdots, \theta_{N-1}) \in \mathbb{R}^{dN}$, we calculate

$$\begin{aligned}\boldsymbol{\theta} \cdot \nabla^2 \mathcal{V}_N^c(\boldsymbol{\xi}) \cdot \boldsymbol{\theta} &= \sum_{k,l=0}^{N-1} \theta_k \cdot \bigg(\beta_N \sum_{j=1}^{N} \nabla^2 V^c(x_j(\boldsymbol{\xi})) c_{j,k} c_{j,l}\bigg) \cdot \theta_l \\ &= \beta_N \sum_{j=1}^{N} \bigg(\sum_{k=0}^{N-1} \theta_k c_{j,k}\bigg) \cdot \nabla^2 V(x_j(\boldsymbol{\xi})) \cdot \bigg(\sum_{l=0}^{N-1} \theta_l c_{j,l}\bigg) \geqslant 0.\end{aligned}$$

The overdamped Langevin dynamics driven by the potential $\mathcal{V}_N^c(\boldsymbol{\xi})$ reads

$$\dot{\xi}_k = -\xi_k - \frac{1}{\omega_k^2 + a} \nabla_{\xi_k} \mathcal{V}_N^c(\boldsymbol{\xi}) + \sqrt{\frac{2}{\omega_k^2 + a}} \dot{B}_k, \tag{3.18}$$

whose infinitesimal generator is given by

$$\mathcal{L}^c = \sum_{k=0}^{N-1} \bigg(\xi_k + \frac{1}{\omega_k^2 + a} \nabla_{\xi_k} \mathcal{V}_N^c(\boldsymbol{\xi})\bigg) \cdot \nabla_{\xi_k} + \sum_{k=0}^{N-1} \frac{1}{\omega_k^2 + a} \Delta_{\xi_k}, \tag{3.19}$$

and the invariant distribution is

$$\pi_N^c(\xi) \propto \exp\bigg\{ -\frac{1}{2} \sum_{k=0}^{N-1} (\omega_k^2 + a) |\xi_k|^2 - \beta_N \sum_{j=1}^{N} V^c\bigg(\sum_{k=0}^{N-1} \xi_k c_{j,k})\bigg)\bigg\}. \tag{3.20}$$

By direct calculation, the Γ operators (see Definition 2.2) corresponding to $\mathcal{L}^c$ are given by

$$\begin{aligned}\Gamma_1(f, g) &= \sum_{k=0}^{N-1} \frac{\nabla_{\xi_k} f \cdot \nabla_{\xi_k} g}{\omega_k^2 + a}, \\ \Gamma_2^c(f, g) &= \sum_{k,l=0}^{N-1} \frac{\nabla^2_{\xi_k \xi_l} f : \nabla^2_{\xi_k \xi_l} g}{(\omega_k^2 + a)(\omega_l^2 + a)} + \sum_{k=0}^{N-1} \frac{\nabla_{\xi_k} f \cdot \nabla_{\xi_k} g}{\omega_k^2 + a} + \sum_{k,l=0}^{N-1} \frac{\nabla_{\xi_k} f \cdot \nabla^2_{\xi_k \xi_l} \mathcal{V}_N^c(\boldsymbol{\xi}) \cdot \nabla_{\xi_l} g}{(\omega_k^2 + a)(\omega_l^2 + a)}.\end{aligned}$$

Utilizing the convexity of the potential function $\mathcal{V}_N^c(\boldsymbol{\xi})$, we obtain

$$\Gamma_2^c(f) \geqslant \sum_{k=0}^{N-1} \frac{|\nabla_{\xi_k} f|^2}{\omega_k^2 + a} + \sum_{k,j=0}^{N-1} \frac{\nabla_{\xi_k} f \cdot \nabla^2_{\xi_k \xi_j} \mathcal{V}_N^c(\boldsymbol{\xi}) \cdot \nabla_{\xi_j} f}{(\omega_k^2 + a)(\omega_j^2 + a)} \geqslant \sum_{k=0}^{N-1} \frac{|\nabla_{\xi_k} f|^2}{\omega_k^2 + a} = \Gamma_1(f). \tag{3.21}$$

Hence from Theorem 2.1 we derive the log-Sobolev inequality

$$\mathrm{Ent}_{\pi_N^c}(f) \leqslant \frac{1}{2} \int_{\mathbb{R}^{dN}} \frac{\Gamma_1(f)}{f} \mathrm{d}\pi_N^c = \frac{1}{2} \sum_{k=0}^{N-1} \frac{1}{\omega_k^2 + a} \int_{\mathbb{R}^{dN}} \frac{|\nabla_{\xi_k} f|^2}{f} \mathrm{d}\pi_N^c. \tag{3.22}$$

In the final step, we apply the bounded perturbation property (Theorem 2.2) to transform (3.22) to the log-Sobolev inequality for $\pi_N(\boldsymbol{\xi})$. Let $\mathcal{Z}_N$ and $\mathcal{Z}_N^c$ be the normalization constants of the distributions $\pi_N(\xi)$ and $\pi_N^c(\xi)$, namely,

$$\mathcal{Z}_N = \int_{\mathbb{R}^{dN}} \exp\bigg(-\frac{1}{2}\sum_{k=0}^{N-1}(\omega_k^2+a)|\xi_k|^2 - \beta_N \sum_{j=1}^{N} V^a(x_j(\boldsymbol{\xi}))\bigg)\mathrm{d}\xi,$$
$$\mathcal{Z}_N^c = \int_{\mathbb{R}^{dN}} \exp\bigg(-\frac{1}{2}\sum_{k=0}^{N-1}(\omega_k^2+a)|\xi_k|^2 - \beta_N \sum_{j=1}^{N} V^c(x_j(\boldsymbol{\xi}))\bigg)\mathrm{d}\xi.$$

Using the inequality $|V^a(x) - V^c(x)| = |V^b(x)| \leqslant M_1$, we have

$$|\mathcal{V}_N^a(\xi) - \mathcal{V}_N^c(\xi)| \leqslant \beta_N \sum_{j=1}^{N} |V^b(x_j(\boldsymbol{\xi}))| \leqslant \beta M_1,$$

and thus the normalization constants $\mathcal{Z}_N$ and $\mathcal{Z}_N^c$ satisfy

$$\frac{\mathcal{Z}_N}{\mathcal{Z}_N^c} \in \big[\exp(-\beta M_1), \exp(\beta M_1)\big].$$

As a result, the density functions $\pi_N(\xi)$ and $\pi_N^c(\xi)$ satisfy

$$\frac{\pi_N^c(\xi)}{\pi_N(\xi)} = \frac{\mathcal{Z}_N}{\mathcal{Z}_N^c}\exp\bigg(\beta_N\sum_{j=1}^{N} V^b(x_j(\boldsymbol{\xi}))\bigg) \in \big[\exp(-2\beta M_1), \exp(2\beta M_1)\big]. \tag{3.23}$$

Using the bounded perturbation (Theorem 2.2), we obtain from (3.22) that

$$\exp(-4\beta M_1)\mathrm{Ent}_{\pi_N}(f) \leqslant \frac{1}{2}\sum_{k=0}^{N-1}\frac{1}{\omega_k^2+a}\int_{\mathbb{R}^{dN}}\frac{|\nabla_{\xi_k} f|^2}{f}\mathrm{d}\pi_N. \tag{3.24}$$

Hence for the rate $\lambda_1 = \exp(-4\beta M_1)$, the relative entropy has exponential decay,

$$\mathrm{Ent}_{\pi_N}(P_t f) \leqslant e^{-2\lambda_1 t}\mathrm{Ent}_{\pi_N}(f), \quad \forall t \geqslant 0,$$

which completes the proof. ∎

Theorem 3.2 (underdamped)**.** *Assume (i)(ii). Let $(P_t)_{t\geqslant 0}$ be the Markov semigroup of the underdamped Langevin dynamics (3.10), and define the entropy-like functional*

$$W_{\mu_N}(f) = \bigg(\frac{M_2^2}{a^2}+1\bigg)\mathrm{Ent}_{\mu_N}(f) + \sum_{k=0}^{N-1}\frac{1}{\omega_k^2+a}\int_{\mathbb{R}^{2dN}}\frac{|\nabla_{\eta_k} f - \nabla_{\xi_k} f|^2 + |\nabla_{\eta_k} f|^2}{f}\mathrm{d}\mu_N. \tag{3.25}$$

Then for any test function $f(\boldsymbol{\xi},\boldsymbol{\eta})$ in $\mathbb{R}^{2dN}$,

$$W_{\mu_N}(P_t f) \leqslant e^{-2\lambda_2 t} W_{\mu_N}(f), \quad \forall t \geqslant 0,$$

where the convergence rate $\lambda_2 = \frac{a^2}{3M_2^2+5a^2}\exp(-4\beta M_1)$.

The proof is based on the generalized Γ calculus.

Proof of Theorem 3.2. First, we establish the uniform-in-N log-Sobolev inequality for the distribution $\mu_N(\xi,\eta)$. In Theorem 3.1, we have proved the log-Sobolev inequality $\pi_N(\xi)$:

$$\lambda_1 \mathrm{Ent}_{\pi_N}(f) \leqslant \frac{1}{2}\sum_{k=0}^{N-1}\frac{1}{\omega_k^2+a}\int_{\mathbb{R}^{dN}}\frac{|\nabla_{\xi_k} f|^2}{f}\mathrm{d}\pi_N, \tag{3.26}$$

where the convergence rate $\lambda_1 = \exp(-4\beta M_1)$. On the other hand, for the Gaussian distribution

$$\sigma_N(\boldsymbol{\eta}) \propto \exp\left\{-\frac{1}{2}\sum_{k=0}^{N-1}(\omega_k^2+a)|\eta_k|^2\right\} \tag{3.27}$$

we apply Theorem 3.1 and obtain the log-Sobolev inequality

$$\mathrm{Ent}_{\sigma_N}(f) \leqslant \frac{1}{2}\sum_{k=0}^{N-1}\frac{1}{\omega_k^2+a}\int_{\mathbb{R}^{dN}}\frac{|\nabla_{\eta_k} f|^2}{f}\mathrm{d}\sigma_N. \tag{3.28}$$

Since the invariant distribution $\mu_N(\boldsymbol{\xi},\boldsymbol{\eta}) = \pi_N(\boldsymbol{\xi})\otimes\sigma_N(\boldsymbol{\eta})$, Theorem 2.3 implies that the tensorization of (3.26) and (3.28) yields a new log-Sobolev inequality

$$\lambda_1 \mathrm{Ent}_{\mu_N}(f) \leqslant \frac{1}{2}\sum_{k=0}^{N-1}\frac{1}{\omega_k^2+a}\int_{\mathbb{R}^{2dN}}\frac{|\nabla_{\xi_k} f|^2+|\nabla_{\eta_k} f|^2}{f}\mathrm{d}\mu_N, \tag{3.29}$$

where the convergence rate is determined by the smaller one of the rates $\lambda_1 = \exp(-4\beta M_1)$ and 1, which is λ_1 itself.

We continue using the notation in the proof of Theorem 3.1,

$$\mathcal{V}_N^a(\boldsymbol{\xi}) = \beta_N\sum_{j=1}^{N}V^a\left(\sum_{k=0}^{N-1}\xi_k c_{j,k}\right), \quad \boldsymbol{\xi}\in\mathbb{R}^{dN},$$

which denotes the external potential of the ring polymer. Note that the generator of the underdamped Langevin dynamics (3.10) is given by

$$\mathcal{L} = \sum_{k=0}^{N-1}\eta_k\cdot\nabla_{\xi_k} - \sum_{k=0}^{N-1}\left(\xi_k+\eta_k+\frac{1}{\omega_k^2+a}\nabla_{\xi_k}\mathcal{V}_N^a(\boldsymbol{\xi})\right)\cdot\nabla_{\eta_k} + \sum_{k=0}^{N-1}\frac{\Delta_{\eta_k}}{\omega_k^2+a}, \tag{3.30}$$

then it is easy to derive the expressions of the commutators:

$$[\mathcal{L},\nabla_{\xi_k}] = \nabla_{\eta_k} + \sum_{l=0}^{N-1}\frac{1}{\omega_l^2+a}\nabla^2_{\xi_k\xi_l}\mathcal{V}_N^a(\xi)\cdot\nabla_{\eta_l}, \quad [\mathcal{L},\nabla_{\eta_k}] = \nabla_{\eta_k} - \nabla_{\xi_k}. \tag{3.31}$$

Inspired from Example 3 of Monmarché (2019), introduce the local operators

$$\Phi_1(f) = f\log f, \quad \Phi_2(f) = \sum_{k=0}^{N-1}\frac{1}{\omega_k^2+a}\frac{|\nabla_{\eta_k} f-\nabla_{\xi_k} f|^2+|\nabla_{\eta_k} f|^2}{f},$$

then from Lemma 2.3 we obtain the generalized Γ operator

$$\Gamma_{\Phi_1}(f) = \frac{1}{2}\sum_{k=0}^{N-1}\frac{1}{\omega_k^2+a}\frac{|\nabla_{\eta_k} f|^2}{f}.$$

In order to compute $\Gamma_{\Phi_2}(f)$, we write $\Phi_2(f) = \sum_{k=0}^{N-1}\frac{1}{\omega_k^2+a}\Phi_{2,k}(f)$, where

$$\Phi_{2,k}(f) = \frac{|\nabla_{\eta_k} f - \nabla_{\xi_k} f|^2 + |\nabla_{\eta_k} f|^2}{f}, \quad k = 0, 1, \cdots, N-1.$$

Utilizing Lemma 2.3 again, we have the following estimate of $\Gamma_{\Phi_{2,k}}(f)$:

$$\begin{aligned} f\cdot\Gamma_{\Phi_{2,k}}(f) &\geqslant (\nabla_{\eta_k} f - \nabla_{\xi_k} f)\cdot[\mathcal{L}, \nabla_{\eta_k} - \nabla_{\xi_k}]f + \nabla_{\eta_k} f\cdot[\mathcal{L}, \nabla_{\eta_k}]f \\ &= |\nabla_{\eta_k} f - \nabla_{\xi_k} f|^2 - (\nabla_{\eta_k} f - \nabla_{\xi_k} f)\cdot\sum_{l=0}^{N-1}\frac{1}{\omega_l^2+a}\nabla^2_{\xi_k\xi_l}\mathcal{V}_N^a(\boldsymbol{\xi})\cdot\nabla_{\eta_l} f. \end{aligned}$$

Taking the summation over $k = 0, 1, \cdots, N-1$, we obtain

$$f\cdot\Gamma_{\Phi_2}(f) \geqslant \sum_{k=0}^{N-1}\frac{|\nabla_{\eta_k} f - \nabla_{\xi_k} f|^2}{\omega_k^2+a} - \sum_{k,l=0}^{N-1}\frac{\nabla_{\eta_k} f - \nabla_{\xi_k} f}{\omega_k^2+a}\cdot\nabla^2_{\xi_k\xi_l}\mathcal{V}_N^a(\boldsymbol{\xi})\cdot\frac{\nabla_{\eta_l} f}{\omega_l^2+a}. \tag{3.32}$$

To further simplify the expression of $\Gamma_{\Phi_2}(f)$, define the vectors $\mathcal{P}, \mathcal{Q}\in\mathbb{R}^{dN}$ by

$$\mathcal{P} = \left\{\frac{\nabla_{\eta_k} f - \nabla_{\xi_k} f}{\sqrt{\omega_k^2+a}}\right\}_{k=0}^{N-1}\in\mathbb{R}^{dN}, \quad \mathcal{Q} = \left\{\frac{\nabla_{\eta_k} f}{\sqrt{\omega_k^2+a}}\right\}_{k=0}^{N-1}\in\mathbb{R}^{dN}, \tag{3.33}$$

then the inequality (3.32) can be equivalently written as

$$\Gamma_{\Phi_2}(f) \geqslant \frac{|\mathcal{P}|^2 - \mathcal{P}^\top\boldsymbol{\Sigma}\mathcal{Q}}{f}, \tag{3.34}$$

where the symmetric matrix $\boldsymbol{\Sigma}\in\mathbb{R}^{dN\times dN}$ is given by

$$\Sigma_{kl} = \frac{1}{\sqrt{(\omega_k^2+a)(\omega_l^2+a)}}\nabla^2_{\xi_k\xi_l}\mathcal{V}_N^a(\boldsymbol{\xi})\in\mathbb{R}^{d\times d}, \quad k, l = 0, 1, \cdots, N-1.$$

For any $\boldsymbol{\theta} = (\theta_0, \theta_1, \cdots, \theta_{N-1})\in\mathbb{R}^{dN}$, we have

$$\begin{aligned} \boldsymbol{\theta}^\top\boldsymbol{\Sigma}\boldsymbol{\theta} &= \sum_{k,l=0}^{N-1}\frac{\theta_k}{\sqrt{\omega_k^2+a}}\cdot\nabla^2_{\xi_k\xi_l}\mathcal{V}_N^a(\boldsymbol{\xi})\cdot\frac{\theta_l}{\sqrt{\omega_l^2+a}} \\ &= \beta_N\sum_{j=1}^{N}\left(\sum_{k,l=0}^{N-1}\frac{\theta_k c_{j,k}}{\sqrt{\omega_k^2+a}}\cdot\nabla^2_{\xi_k\xi_l}V^a(x_j(\boldsymbol{\xi}))\cdot\frac{\theta_l c_{j,l}}{\sqrt{\omega_l^2+a}}\right), \end{aligned}$$

which implies

$$|\boldsymbol{\theta}^\top \boldsymbol{\Sigma} \boldsymbol{\theta}| \leqslant \beta_N M_2 \sum_{j=1}^{N} \left| \sum_{k=0}^{N-1} \frac{\theta_k c_{j,k}}{\sqrt{\omega_k^2 + a}} \right|^2 \leqslant \frac{\beta_N M_2}{a} \sum_{k=0}^{N-1} |\theta_k|^2 \sum_{k=0}^{N-1} |c_{j,k}|^2 = \frac{M_2}{a} \sum_{k=0}^{N-1} |\theta_k|^2.$$

Hence we arrive at the inequality

$$-\frac{M_2}{a} I_{dN} \preccurlyeq \Sigma \preccurlyeq \frac{M_2}{a} I_{dN}. \tag{3.35}$$

In conclusion, the local operators $\Phi_1(f)$, $\Phi_2(f)$ and their generalized Γ operators satisfy

$$\Phi_1(f) = f \log f, \quad \Phi_2(f) = \frac{|\mathcal{P}|^2 + |\mathcal{Q}|^2}{f}, \tag{3.36}$$

$$\Gamma_{\Phi_1}(f) = \frac{|\mathcal{Q}|^2}{2f}, \quad \Gamma_{\Phi_2}(f) \geqslant \frac{|\mathcal{P}|^2 - \frac{M_2}{a}|\mathcal{P}||\mathcal{Q}|}{f}. \tag{3.37}$$

Finally, we derive the functional inequalities satisfied by $\Phi_1(f)$ and $\Phi_2(f)$. The log-Sobolev inequality (3.29) implies

$$\lambda_1 \mathrm{Ent}_{\mu_N}(f) \leqslant \frac{1}{2} \int_{\mathbb{R}^{2dN}} \frac{|\mathcal{P} + \mathcal{Q}|^2 + |\mathcal{Q}|^2}{f} \mathrm{d}\mu_N \leqslant \frac{3}{2} \int_{\mathbb{R}^{2dN}} \frac{|\mathcal{P}|^2 + |\mathcal{Q}|^2}{f} \mathrm{d}\mu_N,$$

hence with the expressions of $\Phi_1(f)$ and $\Phi_2(f)$, we can equivalently write

$$\frac{2\lambda_1}{3} \left(\int_{\mathbb{R}^{2dN}} \Phi_1(f) \mathrm{d}\mu_N - \Phi_1 \left(\int_{\mathbb{R}^{2dN}} f \mathrm{d}\mu_N \right) \right) \leqslant \int_{\mathbb{R}^{2dN}} \Phi_2(f) \mathrm{d}\mu_N. \tag{3.38}$$

On the other hand, (3.37) implies

$$\Gamma_{\Phi_2}(f) - \frac{1}{2}\Phi_2(f) + \left(\frac{M_2^2}{a^2} + 1 \right) \Gamma_{\Phi_1}(f) \geqslant \frac{(|\mathcal{P}| - \frac{M_2}{a}|\mathcal{Q}|)^2}{2f} \geqslant 0. \tag{3.39}$$

Collecting the functional inequalities (3.38) and (3.39), we can directly apply Theorem 2.4 with constants

$$c = \frac{3}{2\lambda_1}, \quad \rho = \frac{1}{2}, \quad m = \frac{M_2^2}{a^2} + 1$$

and obtain the convergence rate

$$\lambda_2 = \frac{\rho}{2(1 + mc)} = \frac{\lambda_1}{2\lambda_1 + 3(\frac{M_2^2}{a^2} + 1)} \leqslant \frac{a^2}{3M_2^2 + 5a^2} \exp(-4\beta M_1),$$

which completes the proof. ∎

In Theorem 3.2, we established the explicit convergence rate λ_2 for the entropy-like quantity $W_{\mu_N}(P_t f)$. Next, we demonstrate that for a carefully chosen initial distribution $\nu_N \in$

$\mathcal{P}(\mathbb{R}^{2dN})$, the test function

$$f_N(\boldsymbol{\xi},\boldsymbol{\eta}) = \frac{\nu_N(\boldsymbol{\xi},\boldsymbol{\eta})}{\mu_N(\boldsymbol{\xi},\boldsymbol{\eta})}$$

ensures $W_{\mu_N}(f_N)$ remains uniform-in-N. This uniformity is crucial, as it guarantees that $H(\nu_N\mathcal{P}_t|\mu_N)$ does not diverge as the number of beads $N \to \infty$.

We select the initial distribution $\nu_N(\boldsymbol{\xi},\boldsymbol{\eta}) = \sigma_N(\boldsymbol{\xi})\otimes\sigma_N(\boldsymbol{\eta})$, where $\sigma_N(\cdot)$ is the Gaussian distribution defined in (3.27). Since the target distribution is $\mu_N(\boldsymbol{\xi},\boldsymbol{\eta}) = \pi_N(\boldsymbol{\xi})\otimes\sigma_N(\boldsymbol{\eta})$, the test function f_N depends only on the $\boldsymbol{\xi}$ variable and is expressed as

$$f_N(\boldsymbol{\xi}) = \frac{\sigma_N(\boldsymbol{\xi})}{\pi_N(\boldsymbol{\xi})} \propto \exp\left\{\beta_N \sum_{j=1}^{N} V^a\left(\sum_{k=0}^{N-1}\xi_k c_{j,k}\right)\right\}. \tag{3.40}$$

Consequently, the entropy-like quantity $W_{\mu_N}(f_N)$ simplifies to

$$\begin{aligned} W_{\mu_N}(f_N) &= \left(\frac{M_2^2}{a^2}+1\right)\mathrm{Ent}_{\pi_N}(f_N) + \sum_{k=0}^{N-1}\frac{1}{\omega_k^2+a}\int_{\mathbb{R}^{dN}}\frac{|\nabla_{\xi_k}f_N|^2}{f_N}\mathrm{d}\pi_N \\ &\leqslant e^{4\beta M_1}\left(\frac{M_2^2}{a^2}+2\right)\sum_{k=0}^{N-1}\frac{1}{\omega_k^2+a}\int_{\mathbb{R}^{dN}}\frac{|\nabla_{\xi_k}f_N|^2}{f_N}\mathrm{d}\pi_N, \end{aligned} \tag{3.41}$$

where the log-Sobolev inequality in (3.24) is applied. Therefore, the task reduces to estimating the RHS of (3.41). In the following, we assume $\nabla V(0) = 0$. If this condition is not met, the global minimum of the potential function $V(x)$ can be shifted to the origin without loss of generality.

Lemma 3.1. *Assume (i)(ii) and $\nabla V(0) = 0$. Let $f_N(\boldsymbol{\xi})$ be the test function in (3.40), then*

$$\sum_{k=0}^{N-1}\frac{1}{\omega_k^2+a}\int_{\mathbb{R}^{dN}}\frac{|\nabla_{\xi_k}f_N|^2}{f_N}\mathrm{d}\pi_N \leqslant M_2^2\left(\frac{1}{a}+3\beta^2\right)^2. \tag{3.42}$$

As a consequence, the entropy-like quantity $W_{\mu_N}(f_N)$ is bounded by

$$W_{\mu_N}(f_N) \leqslant e^{4\beta M_1}M_2^2\left(\frac{M_2^2}{a^2}+2\right)\left(\frac{1}{a}+3\beta^2\right)^2. \tag{3.43}$$

Proof of Lemma 3.1. We begin with the following inequality:

$$\sum_{k=0}^{N-1}\frac{1}{\omega_k^2+a} \leqslant \frac{1}{a}+3\beta^2. \tag{3.44}$$

Using the explicit expression of the frequency ω_k in (3.7), we obtain

$$\sum_{k=0}^{N-1}\frac{1}{\omega_k^2+a} \leqslant \sum_{|k|\leqslant\frac{N}{2}}\frac{1}{a+\frac{4}{\beta_N^2}\sin^2\frac{k\pi}{N}} \leqslant \sum_{|k|\leqslant\frac{N}{2}}\frac{1}{a+\frac{4}{\beta_N^2}\left(\frac{2}{\pi}\frac{k\pi}{N}\right)^2}$$

$$\leqslant \sum_{k\in\mathbb{Z}} \frac{1}{a+\frac{16k^2}{\pi^2\beta^2}} \leqslant \frac{1}{a} + \frac{\pi^2\beta^2}{8}\sum_{k=1}^{\infty}\frac{1}{k^2} = \frac{1}{a} + \frac{\pi^4}{48}\beta^2.$$

Since $\frac{\pi^4}{48} < 3$, we obtain the desired inequality (3.44).

Next we prove (3.42). Using $\nabla V(0) = 0$ we have $\nabla V^a(0) = 0$ and thus the inequality

$$|\nabla V^a(x)| \leqslant M_2|x|, \quad \forall x \in \mathbb{R}^d.$$

Utilizing the expression $f_N(\boldsymbol{\xi}) = \frac{\sigma_N(\boldsymbol{\xi})}{\pi_N(\boldsymbol{\xi})}$, we have

$$\int_{\mathbb{R}^{dN}} \frac{|\nabla_{\xi_k} f_N|^2}{f_N} \mathrm{d}\pi_N = \int_{\mathbb{R}^{dN}} \frac{|\nabla_{\xi_k} f_N|^2}{f_N^2} \mathrm{d}\sigma_N. \tag{3.45}$$

From (3.40), $\frac{\nabla_{\xi_k} f_N}{f_N}$ is explicitly given by

$$\frac{\nabla_{\xi_k} f_N}{f_N} = \beta_N \sum_{j=1}^{N} \nabla V^a(x_j(\boldsymbol{\xi}))c_{j,k},$$

and hence by Cauchy's inequality

$$\begin{aligned}\frac{|\nabla_{\xi_k} f_N|^2}{f_N^2} &\leqslant \beta_N^2 \sum_{j=1}^{N} \left|\nabla V^a(x_j(\boldsymbol{\xi}))\right|^2 \sum_{j=1}^{N} |c_{j,k}|^2 \\ &\leqslant \beta_N M_2^2 \sum_{j=1}^{N} |x_j(\boldsymbol{\xi})|^2 = M_2^2 \sum_{k=0}^{N-1} |\xi_k|^2.\end{aligned}$$

As a consequence, we obtain

$$\begin{aligned}\int_{\mathbb{R}^{dN}} \frac{|\nabla_{\xi_k} f_N|^2}{f_N^2} \mathrm{d}\sigma_N &\leqslant M_2^2 \int_{\mathbb{R}^{dN}} \left(\sum_{k=0}^{N-1} |\xi_k^2|\right) \mathrm{d}\sigma_N = M_2^2 \sum_{k=0}^{N-1} \mathbb{E}\left[|\xi_k|^2 : \boldsymbol{\xi} \sim \sigma_N\right] \\ &= M_2^2 \sum_{k=0}^{N-1} \frac{1}{\omega_k^2 + a} \leqslant M_2^2\left(\frac{1}{a} + 3\beta^2\right),\end{aligned}$$

for each $k = 0, 1, \cdots, N-1$. Hence we finally obtain

$$\sum_{k=0}^{N-1} \frac{1}{\omega_k^2 + a} \int_{\mathbb{R}^{dN}} \frac{|\nabla_{\xi_k} f_N|^2}{f_N^2} \mathrm{d}\sigma_N \leqslant M_2^2\left(\frac{1}{a} + 3\beta^2\right) \sum_{k=0}^{N-1} \frac{1}{\omega_k^2 + a} \leqslant M_2^2\left(\frac{1}{a} + 3\beta^2\right)^2,$$

which completes the proof. ■

Utilizing Lemma 3.1, we establish the exponential decay of the relative entropy $H(\nu_N \mathcal{P}_t|\pi_N)$, where both the coefficient and the convergence rate remain independent of the number of beads N.

Theorem 3.3. *Assume (i)(ii) and* $\nabla V(0) = 0$. *Let* $(\mathcal{P}_t)_{t\geqslant 0}$ *be the dual semigroup of the under-*

damped Langevin dynamics (3.10)*, and choose the initial distribution* $\nu_N \in \mathcal{P}(\mathbb{R}^{dN})$ *as*

$$\nu_N(\boldsymbol{\xi}, \boldsymbol{\eta}) \propto \exp\left\{-\frac{1}{2}\sum_{k=0}^{N-1}(\omega_k^2 + a)(|\xi_k|^2 + |\eta_k|^2)\right\},$$

then we have the inequality

$$H(\nu_N \mathcal{P}_t | \pi_N) \leqslant 2M_2^2\left(\frac{1}{a} + 3\beta^2\right)^2 e^{4\beta M_1 - \lambda_2 t}, \quad \forall t \geqslant 0, \tag{3.46}$$

where the convergence rate $\lambda_2 = \frac{a^2}{3M_2^2+5a^2}\exp(-4\beta M_1)$.

3.5 Relation to Matsubara mode PIMD

We propose an alternative derivation of the Path Integral Molecular Dynamics (PIMD), which is based on taking the continuum limit of the ring polymer energy in (2.16), rather than directly diagonalizing the ring polymer energy in (2.14). This approach leads to a different set of mode frequencies, denoted as $\{\bar{\omega}_k\}_{k=0}^{N-1}$, which are referred to as the Matsubara frequencies (Chandler and Wolynes, 1981; Willatt, 2017; Althorpe, 2024). Consequently, the resulting method is also known as the Matsubara mode PIMD.

Recall that the continuum limit of the ring polymer energy is given by

$$\mathcal{E}_\infty(x(\cdot)) = \frac{1}{2}\int_0^\beta |x'(\tau)|^2\, d\tau + \int_0^\beta V(x(\tau))\, d\tau,$$

where $x(\cdot)$ is a continuous loop defined over the interval $[0, \beta]$. To diagonalize the kinetic energy part of $\mathcal{E}_\infty(x(\cdot))$, we consider the following eigenvalue problem with periodic boundary conditions:

$$-\ddot{c}_k(\tau) = \bar{\omega}_k^2 c_k(\tau), \quad \tau \in [0, \beta]. \tag{3.47}$$

The eigenvalues and eigenfunctions are explicitly given by

$$\begin{aligned}
&\bar{\omega}_0 = 0, && c_0(\tau) = \sqrt{\frac{1}{\beta}};\\
&\bar{\omega}_{2k-1} = \frac{2k\pi}{\beta}, && c_{2k-1}(\tau) = \sqrt{\frac{2}{\beta}}\sin\left(\frac{2k\pi\tau}{\beta}\right), \quad k = 1, 2, \cdots;\\
&\bar{\omega}_{2k} = \frac{2k\pi}{\beta}, && c_{2k}(\tau) = \sqrt{\frac{2}{\beta}}\cos\left(\frac{2k\pi\tau}{\beta}\right), \quad k = 1, 2, \cdots.
\end{aligned}$$

The eigenvalues $\{\bar{\omega}_k\}$ and eigenfunctions $\{c_k(\cdot)\}$ are closely related to the normal mode frequencies $\{\omega_k\}$ and coefficients $\{c_{j,k}\}$ as defined in Definition 3.1. Specifically, we have the

following relationship:

$$\lim_{N\to\infty}\omega_k = \lim_{N\to\infty}\frac{2}{\beta_N}\sin\left(\frac{\lceil\frac{k}{2}\rceil\pi}{N}\right) = \frac{2\lceil\frac{k}{2}\rceil\pi}{\beta} = \bar{\omega}_k, \quad k = 0, 1, \ldots, \tag{3.48}$$

and when N is odd, $c_k(j\beta_N) = c_{j,k}$ for $j = 1, \ldots, N$. Since $\bar{\omega}_k$ represents the infinite-bead limit of ω_k, $\{\bar{\omega}_k\}_{k=0}^{\infty}$ are referred to as the Matsubara frequencies, and the corresponding eigenfunctions $\{c_k(\cdot)\}_{k=0}^{\infty}$ are known as the Matsubara modes in chemical literature.

Using the Matsubara modes, any continuous loop $x(\cdot)$ in $[0, \beta]$ can be represented as

$$x(\tau) = \sum_{k=0}^{\infty}\xi_k c_k(\tau), \quad \tau \in [0, \beta], \tag{3.49}$$

where $\{\xi_k\}_{k=0}^{N-1}$ are referred to as the Matsubara coordinates. Using the Matsubara coordinates, we can express the energy function $\mathcal{E}_\infty(x(\cdot))$ as

$$\mathcal{E}_\infty(\boldsymbol{\xi}) = \frac{1}{2}\sum_{k=0}^{\infty}\bar{\omega}_k^2|\xi_k|^2 + \int_0^\beta V\left(\sum_{k=0}^{N-1}\xi_k c_k(\tau)\right)d\tau. \tag{3.50}$$

Formally, the preconditioned underdamped Langevin dynamics that samples $e^{-\mathcal{E}_\infty(\boldsymbol{\xi})}$ is

$$\begin{cases}\dot{\xi}_k = \eta_k, \\ \dot{\eta}_k = -\xi_k - \dfrac{1}{\bar{\omega}_k^2 + a}\displaystyle\int_0^\beta \nabla V^a(x(\tau))c_k(\tau)\,d\tau - \eta_k + \sqrt{\frac{2}{\bar{\omega}_k^2 + a}}\dot{B}_k,\end{cases} \tag{3.51}$$

where $\{B_k\}_{k=0}^{\infty}$ are independent Brownian motions in $\mathbb{R}^d$. The infinite-dimensional Langevin dynamics in (3.51) corresponds exactly to the stochastic partial differential equation (SPDE) introduced in Equation (21) of Lu et al. (2020).

Although (3.51) does not involve any approximation error, it is infinite-dimensional. Therefore, we must truncate the number of modes to a finite integer N, and the continuous integral must be approximated numerically. A natural truncation scheme is given by

$$\begin{cases}\dot{\xi}_k = \eta_k, \\ \dot{\eta}_k = -\xi_k - \dfrac{\beta_N}{\bar{\omega}_k^2 + a}\displaystyle\sum_{j=1}^{N}\nabla V^a(x_j(\boldsymbol{\xi}))c_{j,k} - \eta_k + \sqrt{\frac{2}{\bar{\omega}_k^2 + a}}\dot{B}_k,\end{cases} \tag{3.52}$$

where the values of $\{x_j(\boldsymbol{\xi})\}_{j=1}^{N}$ are given by

$$x_j(\boldsymbol{\xi}) = \sum_{k=0}^{N-1}\xi_k c_{j,k}.$$

The truncated Langevin dynamics in (3.52) is referred to as the Matsubara mode PIMD, which

differs from the Langevin dynamics in (3.10) in the mode frequencies.

Denote the invariant distribution of the Matsubara mode PIMD (3.52) as

$$\bar{\mu}_N(\boldsymbol{\xi},\boldsymbol{\eta}) \propto \exp\left\{-\frac{1}{2}\sum_{k=0}^{N-1}(\bar{\omega}_k^2+a)(|\xi_k|^2+|\eta_k|^2)-\beta_N\sum_{j=1}^{N}V^a\left(\sum_{k=0}^{N-1}\xi_k c_{j,k}\right)\right\}. \tag{3.53}$$

Employing the same approach with Theorem 3.2, we can prove the uniform-in-N ergodicity of (3.52), and the convergence remains the same.

Theorem 3.4 (Matsubara)**.** *Assume (i)(ii). Let $(P_t)_{t\geqslant 0}$ be the Markov semigroup of the Matsubara mode PIMD (3.52), then for any test function $f(\boldsymbol{\xi},\boldsymbol{\eta})$ in $\mathbb{R}^{2dN}$,*

$$W_{\bar{\mu}_N}(P_t f) \leqslant e^{-2\lambda_2 t} W_{\bar{\mu}_N}(f), \quad \forall t \geqslant 0,$$

where the convergence rate $\lambda_2 = \frac{a^2}{3M_2^2+5a^2}\exp(-4\beta M_1)$.

Our results demonstrate that both the standard PIMD (with normal mode frequencies) and the Matsubara mode PIMD (with Matsubara frequencies) exhibit uniform-in-N ergodicity. However, this does not imply that their approximation errors with respect to N are similar. As shown in the numerical tests, the convergence rates of the standard PIMD and the Matsubara mode PIMD are $O(1/N^2)$ and $O(1/N)$, respectively. This indicates that the standard PIMD (3.10) is more suitable for accurately calculating the quantum thermal average.

Remark 3.7. *In chemical literature, Matsubara frequencies are primarily used in Matsubara dynamics Willatt (2017); Althorpe (2024), which is a computational tool for evaluating quantum correlation functions (QCT). In this context, errors mainly arise from approximating the real-time quantum dynamics, and the Matsubara frequencies result from discretizing the imaginary-time path integral to reflect the periodic boundary conditions in quantum systems. However, the Matsubara mode PIMD is not a standard technique for calculating quantum thermal averages. Unlike the standard PIMD, which uses normal mode frequencies, the Matsubara mode PIMD introduces a different formulation and leads to distinct approximation errors.*

3.6 Numerical tests

In the numerical tests, we compute the quantum thermal average using both the standard PIMD (3.10) and the Matsubara mode PIMD (3.52). The primary objectives are to evaluate the approximation accuracy with respect to the number of beads N and to verify the uniform-in-N ergodicity of the Langevin dynamics. To achieve this, we fix the step size h as a small constant

while varying the inverse temperature β and the number of beads N. The numerical examples include a one-dimensional potential and a three-dimensional spherical potential.

Time average error and autocorrelation function To evaluate the accuracy and efficiency of the simulation, we introduce two key metrics: the time average error, which quantifies the error in computing the quantum thermal average, and the autocorrelation function, which measures the convergence rate. While these concepts are primarily explained using the standard PIMD (3.10), their definitions for the Matsubara mode PIMD (3.52) are analogous.

The target distribution for the standard PIMD is $\pi_N(\boldsymbol{\xi})$, as defined in (3.14). Accordingly, the quantum thermal average $\langle O(\hat{q})\rangle_\beta$ is approximated by

$$\langle O(\hat{x})\rangle_\beta \approx \langle O(\hat{x})\rangle_{\beta,N} := \int_{\mathbb{R}^{dN}} \left[\frac{1}{N}\sum_{j=1}^{N} O\left(\sum_{k=0}^{N-1}\xi_k c_{j,k}\right)\right]\pi_N(\boldsymbol{\xi})\,d\boldsymbol{\xi}. \tag{3.54}$$

Furthermore, let $\xi_k(t)$ and $\eta_k(t)$ be solutions of the underdamped Langevin dynamics (3.10). The accuracy of the quantum thermal average $\langle O(\hat{q})\rangle_\beta$ can be characterized by the time average error,

$$e(\beta,N,T) := \frac{1}{T}\int_0^T \left[\frac{1}{N}\sum_{j=0}^{N-1} O\left(\sum_{k=0}^{N-1}\xi_k(t)c_{j,k}\right)\right]dt - \langle O(\hat{x})\rangle_\beta, \tag{3.55}$$

which depends on the number of beads N and the simulation time T.

The autocorrelation function of the normal mode coordinates $\{\xi_k\}_{k=0}^{N-1}$ is defined by

$$C_k(\beta,N,\Delta T) := \frac{\langle(\xi_k(t)-\langle\xi_k\rangle)(\xi_k(t+\Delta T)\rangle - \langle\xi_k\rangle)}{\langle\xi_k(t)-\langle\xi_k\rangle\rangle^2}, \quad k=0,1,\ldots,N-1, \tag{3.56}$$

where $\langle f(t)\rangle := \lim_{T\to\infty} T^{-1}\int_0^T f(t)\,dt$ denotes the time average of a function f, and ΔT is the time interval between successive measurements of ξ_k. The exponential decay of $C_k(\beta,N,\Delta T)$ characterizes the convergence behavior of the k-th mode.

Time discretization: BAOAB integrator The time discretization of the underdamped Langevin dynamics is performed using the BAOAB integrator (Leimkuhler and Matthews, 2015; Liu et al., 2016), a widely used numerical scheme in molecular dynamics based on operator splitting. In the standard PIMD, (3.10) can be expressed as:

$$\begin{cases} \dot{\xi}_k = \eta_k, \\ \dot{\eta}_k = -\xi_k - \dfrac{\beta_N}{\omega_k^2+a}\nabla_{\xi_k}\mathcal{V}_N^a(\boldsymbol{\xi}) - \eta_k + \sqrt{\dfrac{2}{\omega_k^2+a}}\dot{B}_k, \end{cases} \tag{3.57}$$

where $\mathcal{V}_N^a(\boldsymbol{\xi})$ is the potential function defined in (2.14). To construct the integrator, (3.57) is split into three parts:

$$\mathcal{A}: \begin{cases} \dot{\xi}_k = \eta_k, \\ \dot{\eta}_k = 0, \end{cases} \quad \mathcal{B}: \begin{cases} \dot{\xi}_k = 0, \\ \dot{\eta}_k = -\xi_k - \dfrac{\beta_N}{\omega_k^2 + a} \nabla_{\xi_k} \mathcal{V}_N^a(\boldsymbol{\xi}), \end{cases} \quad O: \begin{cases} \dot{\xi}_k = 0, \\ \dot{\eta}_k = -\eta_k + \sqrt{\dfrac{2}{\omega_k^2 + a}} \dot{B}_k, \end{cases}$$

where $\mathcal{A}, \mathcal{B}$, and O have explicit solutions. The solutions for $\mathcal{A}$ and $\mathcal{B}$ are:

$$\mathcal{A}_t : \xi_k(t) = \xi_k(0) + \eta_k(0)t, \tag{3.58}$$

$$\mathcal{B}_t : \eta_k(t) = -\left(\xi_k(0) + \frac{\beta_N}{\omega_k^2 + a} \nabla_{\xi_k} \mathcal{V}_N^a(\boldsymbol{\xi}(0))\right) t. \tag{3.59}$$

For O, which represents a linear Ornstein–Uhlenbeck process, the solution is:

$$O_t : \eta_k(t) = e^{-t}\eta_k(0) + \sqrt{\frac{1 - e^{-2t}}{\omega_k^2 + a}} \theta_k, \tag{3.60}$$

where $\{\theta_k\}_{k=0}^{N-1}$ are independent Gaussian random variables sampled from $\mathcal{N}(0, I_d)$.

Finally, the BAOAB integrator with step size h is constructed as:

$$(\boldsymbol{\xi}^{m+1}, \boldsymbol{\eta}^{m+1}) = \left(\mathcal{B}_{h/2} \circ \mathcal{A}_{h/2} \circ O_h \circ \mathcal{A}_{h/2} \circ \mathcal{B}_{h/2}\right)(\boldsymbol{\xi}^m, \boldsymbol{\eta}^m), \quad m = 0, 1, \ldots. \tag{3.61}$$

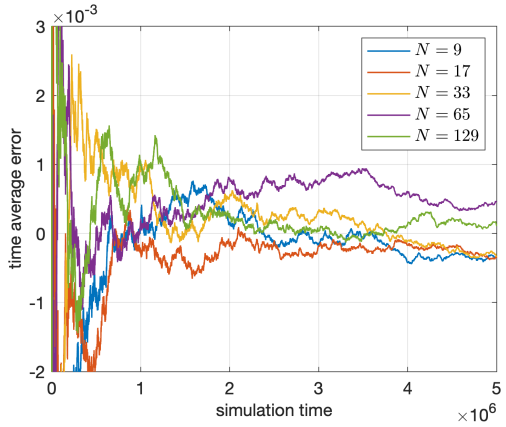
Since the BAOAB integrator is based on symmetric operator splitting, the time discretization error is $O(h^2)$ in the weak sense, ensuring both accuracy and efficiency.

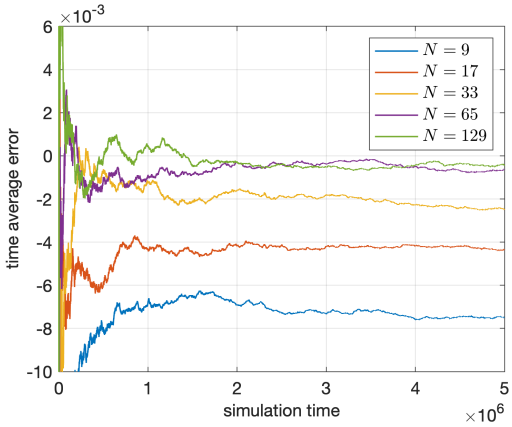
Example: 1D potential Let the potential $V(x)$ and the observable $O(x)$ be defined as:

$$V(x) = \frac{1}{2}x^2 + x\cos x, \quad O(x) = \sin\left(\frac{\pi}{2}x\right), \quad x \in \mathbb{R}. \tag{3.62}$$

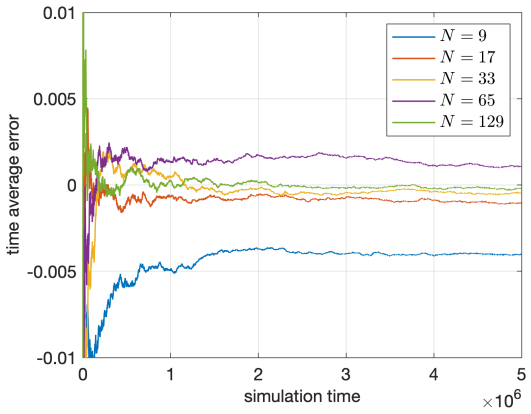
The exact quantum thermal average $\langle O(\hat{x})\rangle_\beta$ is computed using the spectral method with Gauss–Hermite quadrature. For the simulations, the step size is fixed at $h = \frac{1}{16}$, and the simulation time is set to $T = 5 \times 10^6$.

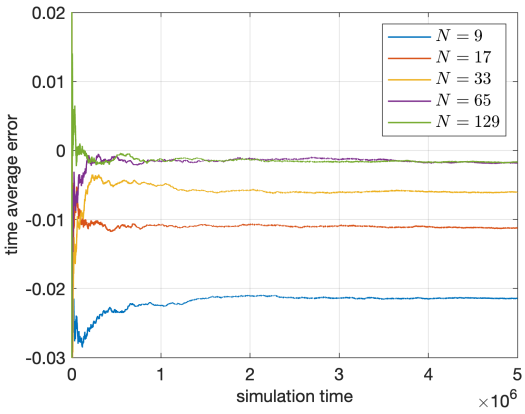
The time average error $e(\beta, N, T)$, as defined in (3.55), is shown in Figure 3.2. The tests are conducted for inverse temperatures $\beta = 1, 2, 4, 8$, and the number of modes $N = 9, 17, 33, 65, 129$. The left and right columns of Figure 3.2 illustrate the results for the standard PIMD and the Matsubara mode PIMD, respectively.



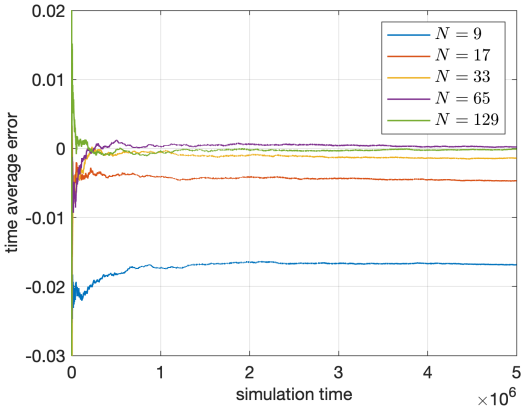


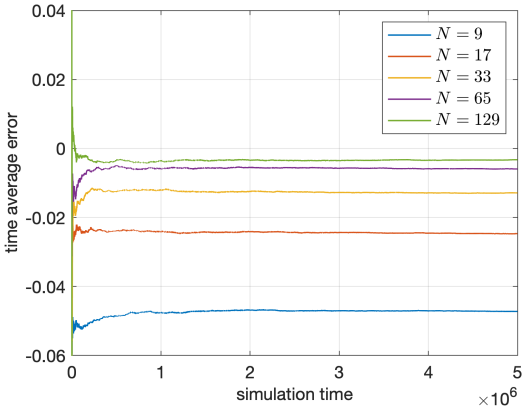
($\beta = 1$)



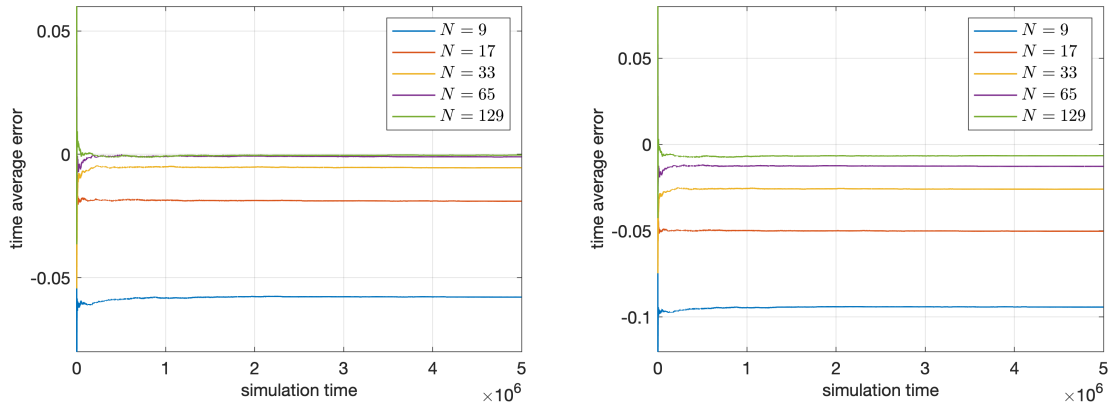


($\beta = 2$)





($\beta = 4$)

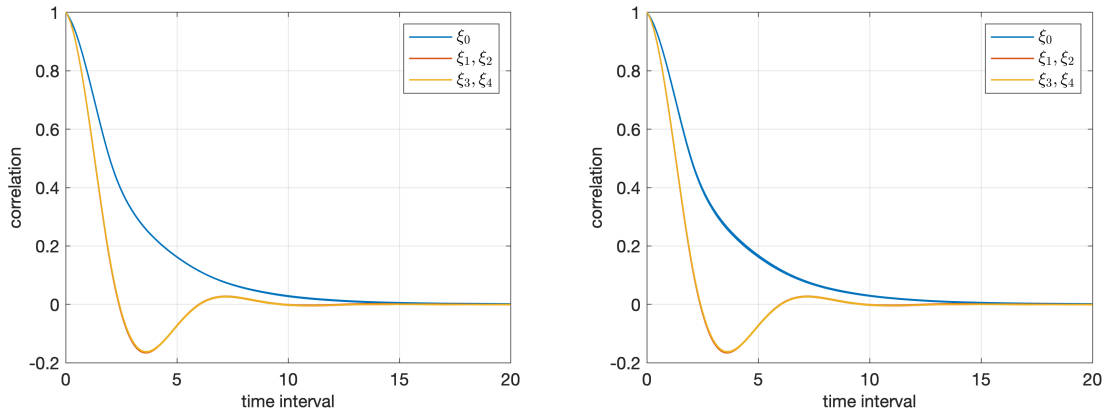


($\beta = 8$)

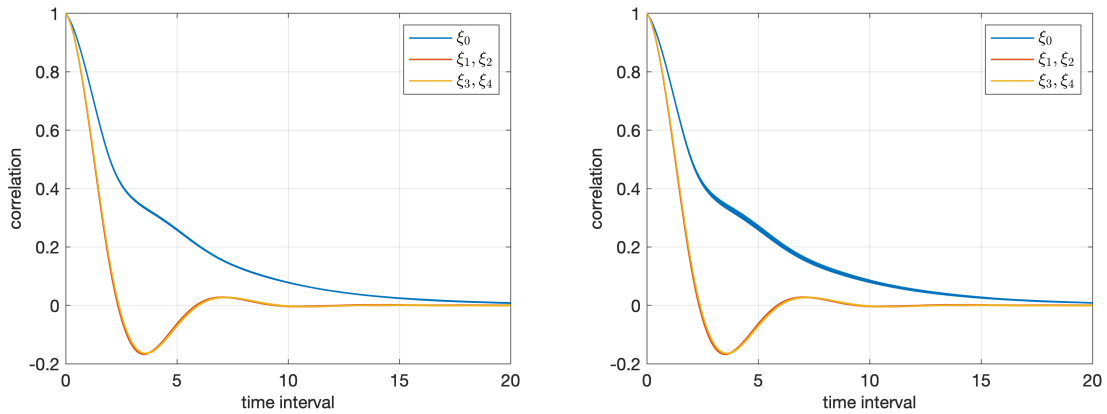
Figure 3.2 Time average error in computing the quantum thermal average for the 1D potential (3.62). Left: standard PIMD. Right: Matsubara mode PIMD. Top to bottom: inverse temperatures $\beta = 1, 2, 4, 8$.

Figure 3.2 demonstrates that the standard PIMD achieves better accuracy than the Matsubara mode PIMD across all temperatures and requires fewer modes N for convergence. The numerical results suggest that the standard PIMD converges with an order of $O(1/N^2)$, whereas the Matsubara mode PIMD converges with an order of $O(1/N)$.

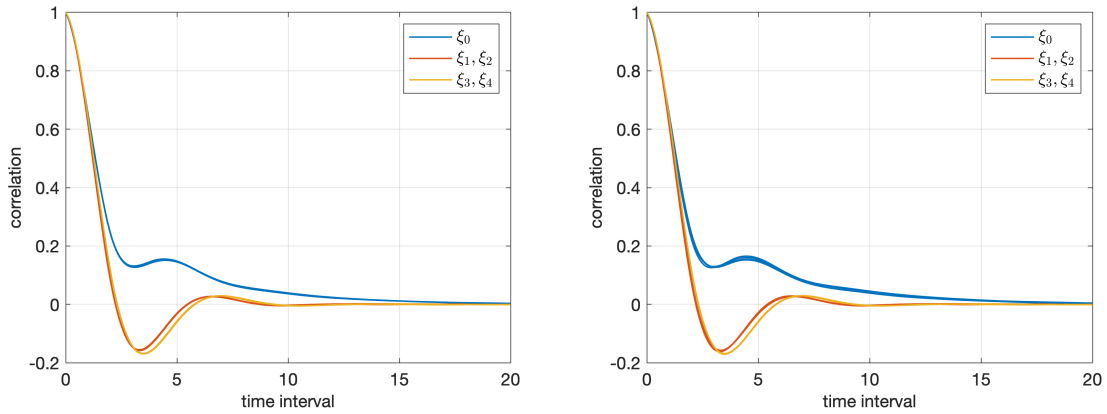
Next, we compute the autocorrelation functions for the first five mode coordinates $\{\xi_k\}_{k=0}^4$ at inverse temperatures $\beta = 1, 2, 4, 8$ and mode numbers $N = 9, 17, 33, 65, 129$. The autocorrelation functions $C_k(\beta, N, \Delta T)$, as defined in (3.56), are plotted against ΔT in Figure 3.3.



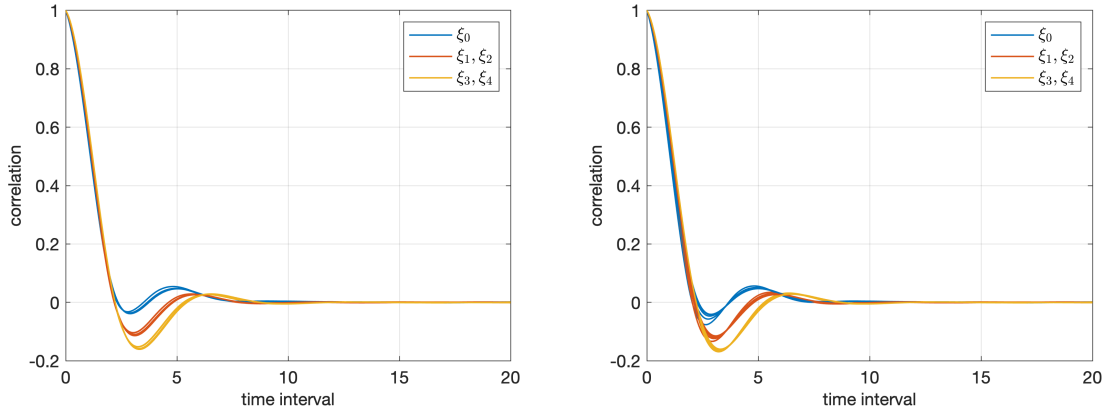
($\beta = 1$)



($\beta = 2$)



($\beta = 4$)



($\beta = 8$)

Figure 3.3 Autocorrelation functions for the 1D potential (3.62). Left: standard PIMD. Right: Matsubara mode PIMD. Top to bottom: $\beta = 1, 2, 4, 8$. The first mode ξ_0 is shown in blue, ξ_1 and ξ_2 in red, and ξ_3 and ξ_4 in yellow.

Figure 3.3 shows that the correlation functions for various N coincide, confirming that both the standard PIMD and the Matsubara mode PIMD exhibit uniform-in-N ergodicity. Furthermore, the separation of correlation functions for different k-modes highlights that convergence rates

vary among modes. At lower temperatures (larger β), high-frequency modes tend to exhibit longer correlation times.

Example: 3D spherical potential Consider the 3D spherical potential

$$V(x) = \frac{1}{2}|x|^2 + \frac{1}{\sqrt{|x|^2 + 0.2^2}}, \quad |x| = \sqrt{x_1^2 + x_2^2 + x_3^2}, \tag{3.63}$$

where we aim to capture the probability distribution of $|x|$, the Euclidean distance from the origin in $\mathbb{R}^3$. Using the density operator $e^{-\beta\hat{H}}$, the distribution of $|x|$ can be expressed through the density function:

$$\rho(r) = \frac{1}{\mathcal{Z}} \int_{\mathbb{R}^3} \langle x|e^{-\beta\hat{H}}|x\rangle \, \delta(|x| - r)\mathrm{d}x, \quad r \geqslant 0, \quad \mathcal{Z} = \mathrm{Tr}\left[e^{-\beta\hat{H}}\right]. \tag{3.64}$$

This radial distribution $\rho(r)$ characterizes observable functions dependent on $|x|$.

For the simulation, we set the inverse temperature $\beta = 4$, the step size $h = \frac{1}{32}$, and the simulation time $T = 5 \times 10^6$. In Figure 3.4, the density function $\rho(r)$ is plotted for varying numbers of modes $N = 3, 5, 9, 17, 33$.

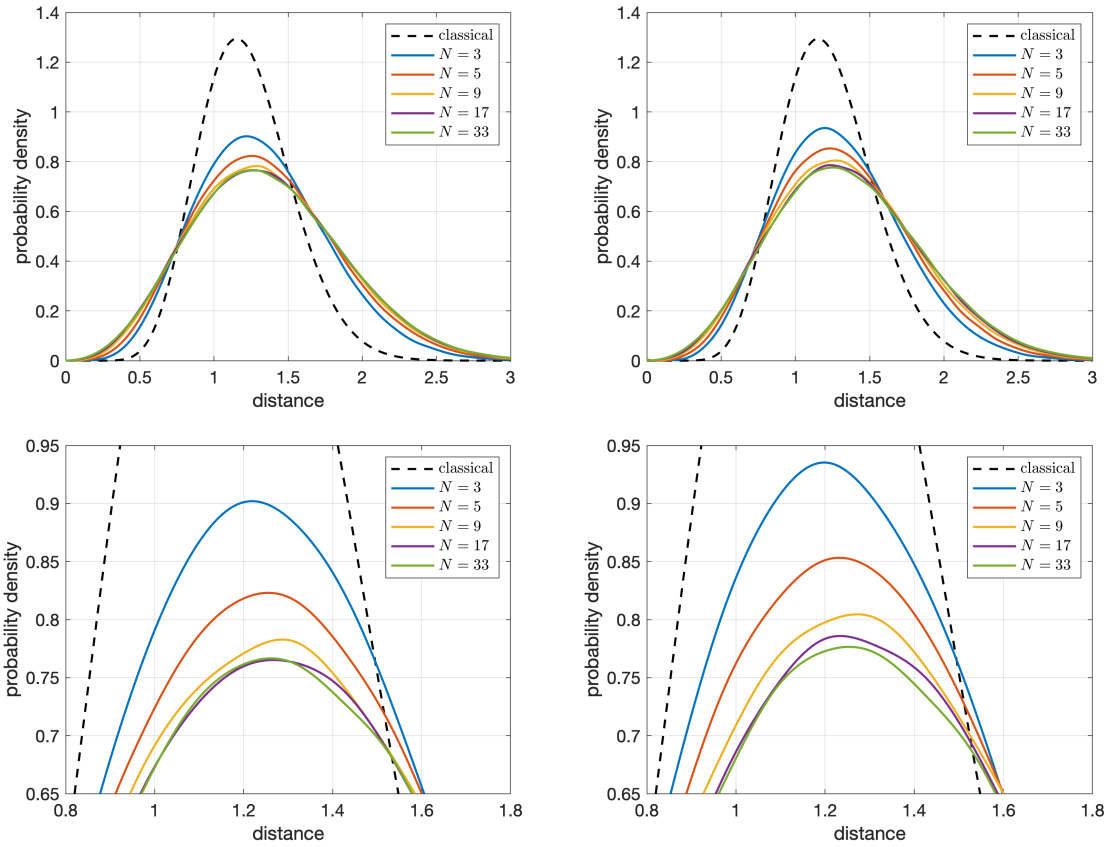


Figure 3.4 Probability density of $|x|$ in the simulation of the PIMD. Left: Matsubara mode PIMD. Right: standard PIMD. Top and bottom graphs use different scales.

Figure 3.4 shows that as the number of modes N increases, the density function transitions from the classical distribution (black dashed curve) to the quantum limit. Both the standard PIMD and Matsubara mode PIMD accurately compute the density function $\rho(r)$; however, the standard PIMD exhibits better accuracy than the Matsubara mode PIMD when the number of modes N is small.

3.7 Brief summary

In this work, we have conducted a comprehensive theoretical analysis of Path Integral Molecular Dynamics (PIMD). A key achievement is our proof of the uniform-in-N ergodicity of the underdamped Langevin dynamics within PIMD (Theorem 3.2), where N represents the number of beads. This result marks a significant advancement in the theoretical understanding of PIMD and demonstrates a novel application of the generalized Γ calculus.

A crucial open question remains: how can the approximation error in terms of N be rigorously quantified? Resolving this would further enhance the theoretical framework and inform practical simulations. This work lays the groundwork for addressing such challenges, bridging theoretical insights and practical applications in PIMD.

Chapter 4 Ergodicity and long-time error of RBM

In this chapter, our primary objective is to establish the ergodicity of the RB–IPS model (2.21) and to quantify the long-time error of the discrete RB–IPS scheme (2.22). To achieve this, we utilize the reflection coupling technique presented in Section 2.4.2 for proving ergodicity and adopt the triangle inequality framework outlined in Section 2.4.3 for estimating the long-time error. Notably, we aim to demonstrate that the convergence rate and error coefficients are independent of both the number of particles N and the step size h.

For clarity in the analysis, let $(\mathcal{P}_t)_{t\geqslant 0}$ denote the continuous dual semigroup of the IPS (2.17), $(\bar{\mathcal{P}}_t)_{t\geqslant 0}$ the continuous dual semigroup of the MVP (2.17), $(\mathcal{Q}_n^h)_{n\geqslant 0}$ the discrete dual semigroup of the RB–IPS (2.21), and $(\tilde{\mathcal{Q}}_n^h)_{n\geqslant 0}$ the discrete dual semigroup of the discrete RB–IPS (2.22). Specifically, for any initial distribution $\nu \in \mathcal{P}(\mathbb{R}^{dN})$, $\nu\mathcal{P}_t$ represents the distribution law of the IPS at time t, while $\nu\mathcal{Q}_n^h$ and $\nu\tilde{\mathcal{Q}}_n^h$ represent the distribution laws of the RB–IPS and discrete RB–IPS at the n-th step, respectively. We also list the related notations and the main results in Table 4.1.

Dynamics	Symbol	Semigroup	Invariant	Ergodicity	Long-time error
IPS (2.17)	$(\boldsymbol{x}_t)_{t\geqslant 0}$	$(\mathcal{P}_t)_{t\geqslant 0}$	π	Theorem 4.2	–
MVP (2.18)	$(\bar{x}_t)_{t\geqslant 0}$	$(\bar{\mathcal{P}}_t)_{t\geqslant 0}$	$\bar{\pi}$	Theorem 4.7	Theorem 4.8
RB–IPS (2.21)	$(\boldsymbol{y}_t)_{t\geqslant 0}$	$(\mathcal{Q}_n^h)_{n\geqslant 0}$	π^h	Theorem 4.1	Theorem 4.3
discrete RB–IPS (2.22)	$(\boldsymbol{Y}_n)_{n\geqslant 0}$	$(\tilde{\mathcal{Q}}_n^h)_{n\geqslant 0}$	–	–	Theorem 4.5

Table 4.1 The stochastic processes studied within the RBM, and the corresponding results on the ergodicity and the long-time error.

4.1 Uniform-in-N ergodicity of RB–IPS

Basic assumptions Compared to the overdamped Langevin dynamics (2.62) for a single particle in $\mathbb{R}^d$, the RB–IPS model involving N particles $\{y_t^i\}_{i=1}^N$, defined as

$$\dot{y}_t^i = b(y_t^i) + \frac{1}{p-1}\sum_{j\neq i, j\in C} K(y_t^i - y_t^j) + \sigma\dot{B}_t^i, \quad i \in C, \quad t \in [nh, (n+1)h),$$

includes additional pairwise interaction terms $K(y_t^i - y_t^j)$. Consequently, in addition to the contraction condition on the drift force $b(x)$ stated in Assumption 2.2, we also impose a bound-

edness condition on the interaction force $K(\cdot)$.

Assumption 4.1. *For the drift force $b(\cdot):\mathbb{R}^d\to\mathbb{R}^d$, there exists a constant L_0 such that*

$$|b(x)|\leqslant L_0(|x|+1),\quad |\nabla b(x)|\leqslant L_0,\quad \forall x\in\mathbb{R}^d. \tag{4.1}$$

For the interaction force $K(\cdot):\mathbb{R}^d\to\mathbb{R}^d$, there exists a constant L_1 such that

$$\max\left\{K(x),\nabla K(x),\nabla^2 K(x)\right\}\leqslant L_1,\quad \forall x\in\mathbb{R}^d. \tag{4.2}$$

For notation convenience, denote the interaction force on the i-th particle as

$$\gamma^i(\boldsymbol{x})=\frac{1}{p-1}\sum_{j\neq i,j\in C}K(x^i-x^j),\quad \forall \boldsymbol{x}=(x_1,\cdots,x_N)\in\mathbb{R}^{dN}, \tag{4.3}$$

where C is the batch which contains i. Then the RB–IPS (2.21) can be shortly written as

$$\dot{y}_t^i=b(y_t^i)+\gamma^i(\boldsymbol{y}_t)+\sigma\dot{B}_t^i,\quad i=1,\cdots,N. \tag{4.4}$$

Due to the Lipschitz condition on $K(\cdot)$ in Assumption 4.1, we have the inequality

$$\begin{aligned}\sum_{i=1}^N|\gamma^i(\boldsymbol{x})-\gamma^i(\bar{\boldsymbol{x}})|&\leqslant\frac{1}{p-1}\sum_{i=1}^N\left[\sum_{j\neq i,j\in C}\left|K(x^i-x^j)-K(\bar{x}^i-\bar{x}^j)\right|\right]\\&\leqslant\frac{1}{p-1}\sum_{i=1}^N\left[\sum_{j\neq i,j\in C}L_1\left(|x^i-\bar{x}^i|+|x_j-\bar{x}_j|\right)\right]\\&=2L_1\sum_{i=1}^N|x^i-\bar{x}^i|.\end{aligned}$$

Hence we obtain the inequality

$$\sum_{i=1}^N|\gamma^i(\boldsymbol{x})-\gamma^i(\bar{\boldsymbol{x}})|\leqslant 2L_1\sum_{i=1}^N|x^i-\bar{x}^i|,\quad \forall\boldsymbol{x},\bar{\boldsymbol{x}}\in\mathbb{R}^{dN}. \tag{4.5}$$

Construction of reflection coupling Let $(\boldsymbol{y}_t)_{t\geqslant 0}$ and $(\bar{\boldsymbol{y}}_t)_{t\geqslant 0}$ be two copies of the RB–IPS (2.21) in $\mathbb{R}^{dN}$, and we define the reflection coupling scheme as

$$\begin{cases}\dot{y}_t^i=b(y_t^i)+\gamma^i(\boldsymbol{y}_t)+\sigma\left(\mathrm{rc}(r_t^i)\dot{B}_t^i+\mathrm{sc}(r_t^i)\dot{\bar{B}}_t^i\right),\\ \dot{\bar{y}}_t^i=b(\bar{y}_t^i)+\gamma^i(\bar{\boldsymbol{y}}_t)+\sigma\left(\mathrm{rc}(r_t^i)(I-2e_t^i(e_t^i)^\top)\dot{B}_t^i+\mathrm{sc}(r_t^i)\dot{\bar{B}}_t^i\right),\end{cases} \tag{4.6}$$

where we employ the notations

$$z_t^i=y_t^i-\bar{y}_t^i\in\mathbb{R}^d,\quad r_t^i=|z_t^i|,\quad e_t^i=\frac{z_t^i}{r_t^i}=\frac{y_t^i-\bar{y}_t^i}{|y_t^i-\bar{y}_t^i|},\quad i=1,\cdots,N, \tag{4.7}$$

and $\{B_t^i\}_{i=1}^N$ and $\{\bar{B}_t^i\}_{i=1}^N$ are two sets of independent Brownian motions in $\mathbb{R}^d$. In addition, $\mathrm{rc}(r)$ and $\mathrm{sc}(r)$ are two smooth functions in $r \in [0, +\infty)$ satisfying

- $\mathrm{rc}^2(r) + \mathrm{sc}^2(r) = 1$ for any $r \geqslant 0$;
- $\mathrm{rc}(r) = 0$ for $r \leqslant \frac{\delta}{2}$ and $\mathrm{rc}(r) = 1$ for $r \geqslant \delta$.

Here, $\delta > 0$ is a flexible parameter in the construction of the coupling. Figure 4.1 provides a simplified illustration of the coupling scheme described in (4.6).

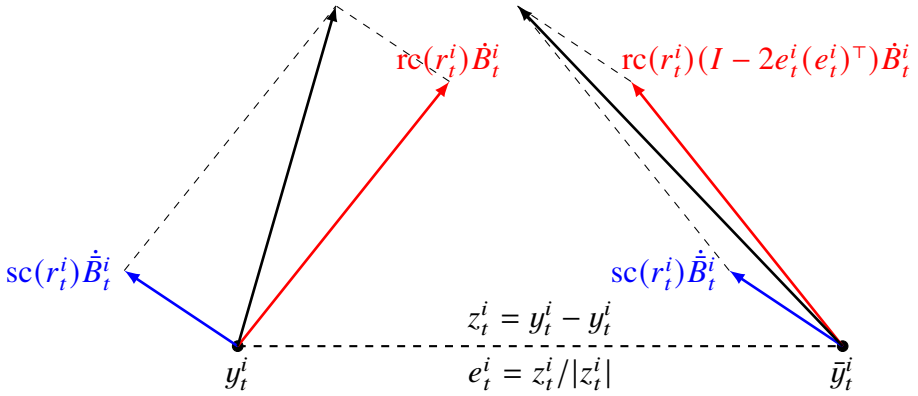


Figure 4.1 The coupling scheme between the two copies $\boldsymbol{y}_t$ and $\bar{\boldsymbol{y}}_t$: blue arrows represent the synchronous coupling, while red arrows represent the reflection coupling.

Remark 4.1. *The key distinction between the coupling scheme* (4.6) *for the RB–IPS and the coupling scheme* (2.74) *for the overdamped Langevin dynamics is the inclusion of the auxiliary functions* $\mathrm{rc}(r)$ *and* $\mathrm{sc}(r)$*. These functions facilitate a smooth transition between synchronous coupling (*$\mathrm{sc}(r) \equiv 1$*) and reflection coupling (*$\mathrm{rc}(r) \equiv 1$*), preventing particles from becoming stuck together. The importance of this smooth transition lies in the need to define the coupling dynamics consistently for all N particles, as detailed in Section 6 of Eberle et al. (2019).*

Remark 4.2. *By Lévy's characterization (Lévy, 1940), the normalizing condition* $\mathrm{rc}^2(r) + \mathrm{sc}^2(r) = 1$ *ensures that both copies* $(\boldsymbol{y}_t)_{t\geqslant 0}$ *and* $(\bar{\boldsymbol{y}}_t)_{t\geqslant 0}$ *are driven by standard Brownian motions in the coupled dynamics* (4.6).

Remark 4.3. *The random divisions at each time step lead to a different definition of* $\gamma^i(\cdot)$ *at each step. However, the batch divisions for* $(\boldsymbol{y}_t)_{t\geqslant 0}$ *and* $(\bar{\boldsymbol{y}}_t)_{t\geqslant 0}$ *remain consistent and identical at every time step.*

From the coupled dynamics (4.6), the position displacement $z_t^i = y_t^i - \bar{y}_t^i$ satisfies the SDE

$$\dot{z}_t^i = b(y_t^i) - b(\bar{y}_t^i) + \gamma^i(\boldsymbol{y}_t) - \gamma^i(\bar{\boldsymbol{y}}_t) + 2\sigma\,\mathrm{rc}(r_t^i)\frac{z_t^i}{r_t^i}\dot{W}_t^i, \tag{4.8}$$

where W_t^i is the one-dimensional Brownian motion defined by $\dot{W}_t^i = (e_t^i)^\top \dot{B}_t^i$. Note that the synchronous coupling part vanishes in (4.8), and the diffusion coefficient $\sigma \mathrm{rc}(Z_t^i)$ solely comes from the reflection coupling. Furthermore, we can verify that $r_t^i = |z_t^i|$ satisfies the SDE

$$\dot{r}_t^i = \frac{z_t^i}{r_t^i} \cdot \big(b(y_t^i) - b(\bar{y}_t^i) + \gamma^i(\boldsymbol{y}_t) - \gamma^i(\bar{\boldsymbol{y}}_t)\big) + 2\sigma \mathrm{rc}(r_t^i)\dot{W}_t^i. \tag{4.9}$$

With the distance function $f(r)$ introduced in Lemma 2.5, we apply Itô calculus and obtain

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} f(r_t^i) = \frac{z_t^i}{r_t^i} \cdot \big(b(y_t^i) - b(\bar{y}_t^i) + \gamma^i(\boldsymbol{y}_t) - \gamma^i(\bar{\boldsymbol{y}}_t)\big) f'(r_t^i) \\ + 2\sigma^2 \mathrm{rc}^2(r_t^i) f''(r_t^i) + 2\sigma \mathrm{rc}(r_t^i) f'(r_t^i) \dot{B}_t^i, \end{aligned} \tag{4.10}$$

which is an analogue to (2.77) in the proof for the overdamped Langevin dynamics (2.62). Taking the expectation in both sides of (4.10), we obtain the changing rate of $\mathbb{E}[f(r_t^i)]$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[f(r_t^i)] = \mathbb{E}\left[\frac{z_t^i}{r_t^i} \cdot \big(b(y_t^i) - b(\bar{y}_t^i) + \gamma^i(\boldsymbol{y}_t) - \gamma^i(\bar{\boldsymbol{y}}_t)\big) f'(r_t^i) + 2\sigma^2 \mathrm{rc}^2(r_t^i) f''(r_t^i)\right]. \tag{4.11}$$

We note that the interaction force $\gamma^i(\cdot)$ in (4.10) is a random function, as it depends on the choice of the batch division. Consequently, the expectation in (4.11) involves the random selection of the batch division.

Estimate of changing rates We need to estimate the RHS of (4.11) to prove the ergodicity. The following inequality is essential in estimating the changing rate of $\mathbb{E}[f(r_t)]$.

Lemma 4.1. *Under Assumptions 2.2 & 4.1, let $f(r), c_0, \varphi_0$ be defined as in Lemma 2.5. Given $\delta > 0$, let $\mathrm{rc}(z)$ be a smooth continuous function with $|\mathrm{rc}(z)| \leqslant 1$ and $\mathrm{rc}(z) = 1$ for $|z| \geqslant \delta$. If the constant L_1 in Assumption 4.1 satisfies*

$$L_1 \leqslant \frac{1}{4} c_0 \varphi_0 \sigma^2,$$

then the following inequality holds with $\beta = \frac{1}{2} c_0 \sigma^2$:

$$\begin{aligned} \sum_{i=1}^N \Bigg(\frac{z^i}{r^i} \cdot \big(b(y^i) - b(\bar{y}^i) + \gamma^i(\boldsymbol{y}) - \gamma^i(\bar{\boldsymbol{y}})\big) f'(r^i) \\ + 2\sigma^2 \mathrm{rc}^2(r^i) f''(r^i) \Bigg) \leqslant N m(\delta) - \beta \sum_{i=1}^N f(r^i), \end{aligned} \tag{4.12}$$

where $\boldsymbol{y}, \bar{\boldsymbol{y}} \in \mathbb{R}^{dN}$, $\boldsymbol{z} = \boldsymbol{y} - \bar{\boldsymbol{y}}$, $r^i = |z^i|$ and $m(\delta)$ is defined by

$$m(\delta) = \frac{\sigma^2}{2} \sup_{r < \delta} \left(r\kappa(r)^-\right) + c_0 \sigma^2 \delta. \tag{4.13}$$

Here $x^- = -\min\{x, 0\}$ denotes the negative part of $x \in \mathbb{R}$.

Proof of Lemma 4.1. The LHS of (4.12) can be written as $I = I_1 + I_2 + I_3$,

$$I_1 = \sum_{i=1}^{N} \frac{z^i}{r^i} \cdot (b(y^i) - b(\bar{y}^i)) f'(r^i),$$
$$I_2 = \sum_{i=1}^{N} \frac{z^i}{r^i} \cdot (\gamma^i(\boldsymbol{y}) - \gamma^i(\bar{\boldsymbol{y}})) f'(r^i),$$
$$I_3 = 2\sigma^2 \sum_{i=1}^{N} \mathrm{rc}^2(r^i) f''(r^i).$$

Next we estimate I_1, I_2, I_3 respectively. By the definition of $\kappa(r)$ in (2.63), we obtain

$$I_1 \leqslant -\frac{\sigma^2}{2} \sum_{i=1}^{N} r^i \kappa(r^i) f'(r^i). \tag{4.14}$$

Using the Lipschitz condition in 4.5 and $f(r) \geqslant \varphi_0 r$,

$$I_2 \leqslant \sum_{i=1}^{N} |\gamma^i(\boldsymbol{y}) - \gamma^i(\bar{\boldsymbol{y}})| \leqslant 2L_1 \sum_{i=1}^{N} r^i \leqslant \frac{2L_1}{\varphi_0} \sum_{i=1}^{N} f(r^i). \tag{4.15}$$

Applying the estimate of $f''(r)$ in Lemma 2.5, we have

$$\begin{aligned} I_3 &\leqslant \frac{\sigma^2}{2} \sum_{i=1}^{N} r^i \kappa(r^i) \mathrm{rc}^2(r^i) f'(r^i) - c_0 \sigma^2 \sum_{i=1}^{N} \mathrm{rc}^2(r^i) f(r^i) \\ &= \frac{\sigma^2}{2} \sum_{i=1}^{N} r^i \kappa(r^i) f'(r^i) - c_0 \sigma^2 \sum_{i=1}^{N} f(r^i) \\ &\quad \underbrace{-\frac{\sigma^2}{2} \sum_{i=1}^{N} r^i \kappa(r^i)(1 - \mathrm{rc}^2(r^i)) f'(r^i)}_{I_{31}} + \underbrace{c_0 \sigma^2 \sum_{i=1}^{N} (1 - \mathrm{rc}^2(r^i)) f(r^i)}_{I_{32}}. \end{aligned} \tag{4.16}$$

Then we estimate I_{31} and I_{32} in (4.16). Note that $1 - \mathrm{rc}^2(r^i) = 0$ if $r_i \geqslant \delta$, we have

$$\begin{aligned} I_{31} &= -\frac{\sigma^2}{2} \sum_{i=1}^{N} r^i \kappa(r^i)(1 - \mathrm{rc}^2(r^i)) f'(r^i) \\ &\leqslant \frac{\sigma^2}{2} \sum_{i: r^i < \delta} r^i \kappa(r^i)^- f'(r^i) \leqslant \frac{\sigma^2}{2} \sum_{i: r^i < \delta} r^i \kappa(r^i)^- \leqslant \frac{N\sigma^2}{2} \sup_{r < \delta} \left(r\kappa(r)^- \right). \end{aligned}$$

In a similar way, using $f(r) \leqslant r$ we obtain

$$I_{32} = c_0 \sigma^2 \sum_{i=1}^{N} (1 - \mathrm{rc}^2(r^i)) f(r^i) \leqslant c_0 \sigma^2 \sum_{i: r^i < \delta} f(r^i) \leqslant c_0 N \sigma^2 \delta.$$

From the definition of $m(\delta)$ in (4.13), we obtain the estimate of I_3:

$$I_3 \leqslant \frac{\sigma^2}{2}\sum_{i=1}^{N} r^i \kappa(r^i) f'(r^i) - c_0\sigma^2 \sum_{i=1}^{N} f(r^i) + Nm(\delta). \tag{4.17}$$

Summation over the inequalities (4.14), (4.15) and (4.17) for I_1, I_2, I_3 gives

$$I \leqslant -\left(c_0\sigma^2 - \frac{2L_1}{\varphi_0}\right)\sum_{i=1}^{N} f(r^i) + Nm(\delta).$$

When the Lipschitz constant $L_1 \leqslant \frac{1}{4}c_0\varphi_0\sigma^2$, we obtain

$$I \leqslant -\frac{1}{2}c_0\sigma^2 \sum_{i=1}^{N} f(r^i) + Nm(\delta) = Nm(\delta) - \beta\sum_{i=1}^{N} f(r^i),$$

which is exactly the result we need. ∎

Note that the distance function $f(r)$, the upper bound of L_1 and the contraction rate β are all independent of parameter δ, thus we may pass δ to the limit 0 without changing the value of β.

Using Lemma 4.1, it is convenient to obtain the contractivity of the coupled dynamics (4.6). Denote the distance between the two sets of particles $\boldsymbol{y}, \bar{\boldsymbol{y}} \in \mathbb{R}^{dN}$ by

$$\rho(\boldsymbol{y}, \bar{\boldsymbol{y}}) = \frac{1}{N}\sum_{i=1}^{N} f(|y^i - \bar{y}^i|), \tag{4.18}$$

then we have the following result on the contractivity of the RB–IPS (2.21).

Lemma 4.2. *Under Assumptions 2.2 & 4.1, let $f(r), c_0, \varphi_0$ be defined as in Lemma 2.5. If the constant L_1 in Assumption 4.1 satisfies*

$$L_1 \leqslant \frac{1}{4}c_0\varphi_0\sigma^2,$$

then for $(\boldsymbol{y}_t)_{t\geqslant 0}$ and $(\bar{\boldsymbol{y}}_t)_{t\geqslant 0}$ evolved by the coupled RB–IPS (4.6), we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\big[\rho(\boldsymbol{y}_t, \bar{\boldsymbol{y}}_t)\big] \leqslant m(\delta) - \beta\cdot\mathbb{E}\big[\rho(\boldsymbol{y}_t, \bar{\boldsymbol{y}}_t)\big], \tag{4.19}$$

where $\beta = \frac{1}{2}c_0\varphi_0\sigma^2$ and $m(\delta)$ is defined in (4.13).

Proof of Lemma 4.2. Using $\rho(\boldsymbol{y}_t, \bar{\boldsymbol{y}}_t) = \frac{1}{N}\sum_{i=1}^{N} f(r_t^i)$, according to the equality (4.11) we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\big[\rho(\boldsymbol{y}_t, \bar{\boldsymbol{y}}_t)\big] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\frac{z_t^i}{r_t^i}\cdot\big(b(y_t^i) - b(\bar{y}_t^i) + \gamma^i(\boldsymbol{y}_t) - \gamma^i(\bar{\boldsymbol{y}}_t)\big)f'(r_t^i) + 2\sigma^2\mathrm{rc}^2(r_t^i)f''(r_t^i)\right].$$

Applying the estimate in Lemma 4.1, we immediately obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\big[\rho(\boldsymbol{y}_t, \bar{\boldsymbol{y}}_t)\big] \leqslant m(\delta) - \beta\cdot\mathbb{E}\big[\rho(\boldsymbol{y}_t, \bar{\boldsymbol{y}}_t)\big],$$

which completes the proof. ∎

Main theorem: ergodicity of RB–IPS Similar to Definition 2.3 for the overdamped Langevin dynamics, we introduce the normalized Wasserstein distances for distributions in $\mathbb{R}^{dN}$.

Definition 4.1. *For the distance function $f(r)$, define the normalized Wasserstein-f distance as*

$$\mathcal{W}_f(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbb{R}^{dN}\times\mathbb{R}^{dN}} \left(\frac{1}{N}\sum_{i=1}^{N} f(|x^i-y^i|)\right)\gamma(\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}), \tag{4.20}$$

where $\Pi(\mu,\nu)$ denotes the set of joint distributions in $\mathbb{R}^d\times\mathbb{R}^d$ whose marginal distributions in the $\boldsymbol{x},\boldsymbol{y}$ variables are exactly μ,ν.

In particular, if $f(r)\equiv r$, the corresponding normalized Wasserstein-f distance becomes the normalized Wasserstein-1 distance

$$\mathcal{W}_1(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbb{R}^{dN}\times\mathbb{R}^{dN}} \left(\frac{1}{N}\sum_{i=1}^{N} |x^i-y^i|\right)\gamma(\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}). \tag{4.21}$$

The normalization factor N^{-1} in Definition 4.1 ensures that the distance between two distributions in $\mathbb{R}^{dN}$ is still $O(1)$ for sufficiently large N. According to Lemma 2.5, we have

$$\varphi_0\mathcal{W}_1(\mu,\nu) \leqslant \mathcal{W}_f(\mu,\nu) \leqslant \mathcal{W}_1(\mu,\nu) \tag{4.22}$$

for any distributions μ,ν in $\mathbb{R}^{dN}$.

We have the following theorem on the ergodicity of the RB–IPS (2.21).

Theorem 4.1 (ergodicity of RB–IPS)**.** *Under Assumptions 2.2 & 4.1, let $f(r), c_0, \varphi_0$ be defined as in Lemma 2.5. Let $(\mathcal{Q}_n^h)_{n\geqslant 0}$ be the dual semigroup of the RB–IPS (2.21). If the constant L_1 in Assumption 4.1 satisfies*

$$L_1 \leqslant \frac{1}{4}c_0\varphi_0\sigma^2,$$

then for any probability distributions μ,ν in $\mathbb{R}^{dN}$, we have

$$\mathcal{W}_f(\mu\mathcal{Q}_n^h,\nu\mathcal{Q}_n^h) \leqslant e^{-\beta nh}\mathcal{W}_f(\mu,\nu), \quad \forall n\geqslant 0, \tag{4.23}$$

where $\beta=\frac{1}{2}c_0\varphi_0\sigma^2$. As a consequence, in the normalized Wasserstein-1 distance we have

$$\mathcal{W}_1(\mu\mathcal{Q}_n^h,\nu\mathcal{Q}_n^h) \leqslant \frac{1}{\varphi_0}e^{-\beta nh}\mathcal{W}_1(\mu,\nu), \quad \forall n\geqslant 0. \tag{4.24}$$

Proof of Theorem 4.1. For given distributions μ,ν in $\mathbb{R}^{dN}$, let $\gamma\in\Pi(\mu,\nu)$ satisfies

$$\int_{\mathbb{R}^{dN}} \left(\frac{1}{N}\sum_{i=1}^{N} f(|x^i-y^i|)\right)\gamma(\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}) \leqslant \mathcal{W}_f(\mu,\nu)+\varepsilon, \tag{4.25}$$

where $\varepsilon > 0$ is an arbitrary small constant. Let $(\boldsymbol{y}_t)_{t\geqslant 0}$ and $(\bar{\boldsymbol{y}}_t)_{t\geqslant 0}$ be evolved by the coupled RB–IPS (4.6) with the initial value $(\boldsymbol{y}_0, \bar{\boldsymbol{y}}_0) \sim \gamma$, then we have $\boldsymbol{y}_{nh} \sim \mu\mathcal{Q}_n^h$ and $\bar{\boldsymbol{y}}_{nh} \sim \nu\mathcal{Q}_n^h$.

Recall in Lemma 4.2 the following inequality holds for $t \in [nh, (n+1)h)$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\big[\rho(\boldsymbol{y}_t, \bar{\boldsymbol{y}}_t)\big] \leqslant m(\delta) - \beta \cdot \mathbb{E}\big[\rho(\boldsymbol{y}_t, \bar{\boldsymbol{y}}_t)\big], \tag{4.26}$$

Integrating the inequality (4.26) for $t \in [nh, (n+1)h)$ yields

$$\mathbb{E}\big[\rho(\boldsymbol{y}_{(n+1)h}, \bar{\boldsymbol{y}}_{(n+1)h})\big] \leqslant e^{-\beta h}\mathbb{E}\big[\rho(\boldsymbol{y}_{nh}, \bar{\boldsymbol{y}}_{nh})\big] + \frac{m(\delta)(1-e^{-\beta h})}{\beta}. \tag{4.27}$$

Repeating the inequality (4.27), we obtain for any $n \geqslant 0$,

$$\begin{aligned}\mathbb{E}\big[\rho(\boldsymbol{y}_{nh}, \bar{\boldsymbol{y}}_{nh})\big] &\leqslant e^{-\beta nh}\mathbb{E}\big[\rho(\boldsymbol{y}_0, \bar{\boldsymbol{y}}_0)\big] + \frac{m(\delta)(1-e^{-\beta nh})}{\beta} \\ &\leqslant e^{-\beta nh}\mathcal{W}_f(\boldsymbol{y}_0, \bar{\boldsymbol{y}}_0) + \frac{m(\delta)(1-e^{-\beta nh})}{\beta} + \varepsilon.\end{aligned}$$

Using the definition of the normalized Wasserstein-f distance in Definition 4.1, we have

$$\mathcal{W}_f(\mu\mathcal{Q}_n^h, \nu\mathcal{Q}_n^h) \leqslant e^{-\beta nh}\mathcal{W}_f(\mu, \nu) + \frac{m(\delta)(1-e^{-\beta nh})}{\beta} + \varepsilon. \tag{4.28}$$

Note that the evolution of μQ_n^h and νQ_n^h does not depend on the coupling scheme, we can directly pass δ and ε to 0 in (4.28) and obtain

$$\mathcal{W}_f(\mu\mathcal{Q}_n^h, \nu\mathcal{Q}_n^h) \leqslant e^{-\beta nh}\mathcal{W}_f(\mu, \nu),$$

which completes the proof. ∎

Since the IPS (2.17) is a special case of the RB–IPS (2.21) with batch size $p = N$, the ergodicity of the IPS follows directly from Theorem 4.1, thereby recovering Corollary 9 of Eberle et al. (2019).

Theorem 4.2 (ergodicity of IPS)**.** *Under Assumptions 2.2 & 4.1, let $f(r), c_0, \varphi_0$ be defined as in Lemma 2.5. Let $(\mathcal{P}_t)_{t\geqslant 0}$ be the dual semigroup of the IPS (2.21). If the constant L_1 in Assumption 4.1 satisfies*

$$L_1 \leqslant \frac{1}{4}c_0\varphi_0\sigma^2,$$

then for any probability distributions μ, ν in $\mathbb{R}^{dN}$, we have

$$\mathcal{W}_f(\mu\mathcal{P}_t, \nu\mathcal{P}_t) \leqslant e^{-\beta t}\mathcal{W}_f(\mu, \nu), \quad \forall n \geqslant 0, \tag{4.29}$$

where $\beta = \frac{1}{2}c_0\varphi_0\sigma^2$. *As a consequence, in the normalized Wasserstein-1 distance we have*

$$\mathcal{W}_1(\mu\mathcal{P}_t, \nu\mathcal{P}_t) \leqslant \frac{1}{\varphi_0} e^{-\beta t}\mathcal{W}_1(\mu, \nu), \quad \forall n \geqslant 0. \tag{4.30}$$

It is crucial that all constants in Theorem 4.1, including c_0, φ_0, β, are independent of both the number of particles N and the step size h. This dimension-free property provides a solid theoretical foundation for studying the mean-field behavior of the RB–IPS (2.21). It is also worth noting that the condition requiring L_1 to be sufficiently small is essential for ensuring dimension-free ergodicity. Without this condition, when the mean-field McKean–Vlasov process admits multiple invariant distributions, the convergence rate can degenerate to zero as $N \to \infty$. See, for example, Durmus et al. (2020).

4.2 Approximation error of RB–IPS

In Chapter 4.1, we established the uniform-in-N ergodicity of the RB–IPS (2.21), demonstrating its effectiveness as an approximate sampling method for the target distribution $\pi(\boldsymbol{x})$ in $\mathbb{R}^{dN}$. However, to fully understand the sampling error of the RB–IPS, it is essential to evaluate the bias of its invariant distribution. Specifically, we need to quantitatively estimate $\mathcal{W}_1(\pi, \pi^h)$, where π^h denotes the invariant distribution of the RB–IPS.

Existence and uniqueness of invariant distribution Leveraging the ergodicity results established in Theorems 4.1 and 4.2, we can readily confirm the existence and uniqueness of the invariant distribution using the Banach fixed point theorem, even when the drift force $b(\cdot)$ and the interaction force $K(\cdot)$ are not in gradient form.

Lemma 4.3. *Under Assumptions 2.2 & 4.1, let* $f(r), c_0, \varphi_0$ *be defined as in Lemma 2.5. If the constant* L_1 *in Assumption 4.1 satisfies*

$$L_1 \leqslant \frac{1}{4}c_0\varphi_0\sigma^2,$$

then we have the following:

1. *The IPS* $(\boldsymbol{x}_t)_{t\geqslant 0}$ *defined in (2.17) as a (continuous) Markov process has a unique invariant distribution* $\pi \in \mathcal{P}(\mathbb{R}^{dN})$*;*
2. *The RB–IPS* $(\boldsymbol{y}_{nh})_{n\geqslant 0}$ *defined in (2.21) as a (discrete) Markov chain has a unique invariant distribution* $\pi^h \in \mathcal{P}(\mathbb{R}^{dN})$.

The following proof is inspired from Corollary 3 of Ref. Eberle et al. (2019).

Proof of Lemma 4.3. Let $\mathcal{P}_1(\mathbb{R}^{dN})$ be the set of the probability distributions in $\mathbb{R}^{dN}$ with finite first moments, namely,

$$\mathcal{P}_1(\mathbb{R}^{dN}) := \left\{ \mu \in \mathcal{P}(\mathbb{R}^{dN}) : \int_{\mathbb{R}^{dN}} \left(\frac{1}{N} \sum_{i=1}^{N} |x^i| \right) \mu(\mathrm{d}\boldsymbol{x}) < +\infty \right\}, \tag{4.31}$$

then $(\mathcal{P}_1(\mathbb{R}^{dN}), \mathcal{W}_1)$ forms a complete metric space, where $\mathcal{W}_1(\cdot,\cdot)$ is the normalized Wasserstein-1 distance in Definition 4.1.

In Theorem 4.2, we have proved for any distributions μ, ν in $\mathbb{R}^{dN}$,

$$\mathcal{W}_1(\mu\mathcal{P}_t, \nu\mathcal{P}_t) \leqslant \frac{1}{\varphi_0} e^{-\beta t} \mathcal{W}_1(\mu, \nu), \quad \forall t \geqslant 0.$$

By picking a constant $T > 0$ such that $\frac{1}{\varphi_0} e^{-\beta T} = \frac{1}{2}$, we obtain the inequality

$$\mathcal{W}_1(\mu\mathcal{P}_T, \nu\mathcal{P}_T) \leqslant \frac{1}{2} \mathcal{W}_1(\mu, \nu). \tag{4.32}$$

Hence the mapping $\nu \mapsto \nu P_T$ is contractive in $(\mathcal{P}_1(\mathbb{R}^{dN}), \mathcal{W}_1)$. From the Banach fixed point theorem (see Chapter 5.1 of Kreyszig (1991) for reference), this mapping has a fixed point $\pi_0 \in \mathcal{P}_1(\mathbb{R}^{dN})$, i.e.,

$$\pi_0 = \pi_0 \mathcal{P}_T.$$

Define the probability measure in $\mathbb{R}^{dN}$ by

$$\pi = \frac{1}{T} \int_0^T \pi_0 \mathcal{P}_s \mathrm{d}s,$$

then from the semigroup property of $(\mathcal{P}_t)_{t \geqslant 0}$, for any $t \geqslant 0$ we have

$$\pi \mathcal{P}_t = \frac{1}{T} \int_0^T (\pi_0 \mathcal{P}_s) \mathcal{P}_t \mathrm{d}s = \frac{1}{T} \int_0^T \pi_0 \mathcal{P}_{s+t} \mathrm{d}s. \tag{4.33}$$

Since the family of distributions $\{\pi_0 \mathcal{P}_t\}_{t \geqslant 0}$ has the period T, from (4.33) we have

$$\pi \mathcal{P}_t = \frac{1}{T} \int_0^T \pi_0 \mathcal{P}_{s+t} \mathrm{d}s = \frac{1}{T} \int_0^T \pi_0 \mathcal{P}_s \mathrm{d}s = \pi. \tag{4.34}$$

Therefore, π is the invariant distribution of the IPS $(\boldsymbol{x}_t)_{t \geqslant 0}$. The uniqueness of π follows from the ergodicity of the IPS proved in Theorem 4.2.

For the RB–IPS, the proof is similar. Given $h > 0$, from Theorem 4.1 we have

$$\mathcal{W}_1(\mu\mathcal{Q}_n^h, \nu\mathcal{Q}_n^h) \leqslant \frac{1}{\varphi_0} e^{-\beta n h} \mathcal{W}_1(\mu, \nu).$$

Then we choose an integer $M \in \mathbb{N}$ such that $\frac{1}{\varphi_0} e^{-\beta M h} \leqslant \frac{1}{2}$, and thus

$$\mathcal{W}_1(\mu\mathcal{Q}_M^h, \nu\mathcal{Q}_M^h) \leqslant \frac{1}{2} \mathcal{W}_1(\mu, \nu), \tag{4.35}$$

hence the mapping $\nu \mapsto \nu Q_M^h$ is contractive. From the Banach fixed point theorem, this mapping has a fixed point $\pi_0^h \in \mathcal{P}_1(\mathbb{R}^{dN})$, i.e.,

$$\pi_0^h = \pi_0^h Q_M^h. \tag{4.36}$$

Define the probability distribution in $\mathbb{R}^{dN}$ by

$$\pi^h = \frac{1}{M} \sum_{k=0}^{M-1} \pi_0^h Q_k^h, \tag{4.37}$$

then from the semigroup property of $(Q_n^h)_{n\geqslant 0}$, for any $n \geqslant 0$ we have

$$\pi^h Q_n^h = \frac{1}{M} \sum_{k=0}^{M-1} (\pi_0 Q_k^h) Q_n^h = \frac{1}{M} \sum_{k=0}^{M-1} \pi_0^h Q_k^h = \pi^h. \tag{4.38}$$

Therefore, π^h is the invariant distribution of the RB–IPS $(\boldsymbol{y}_{nh})_{n\geqslant 0}$. The uniqueness of π^h follows from the contractivity in Theorem 4.1. ∎

Strong error of RB–IPS in finite time We establish the uniform-in-time moments for both the IPS (2.17) and the RB–IPS (2.21). In particular, the fourth moment estimate is crucial, as it is needed for the strong error estimate.

Lemma 4.4. *Under Assumptions 2.2 & 4.1, if the initial distribution $\nu \in \mathcal{P}(\mathbb{R}^{dN})$ satisfies*

$$\max_{1\leqslant i\leqslant N} \int_{\mathbb{R}^{dN}} |x^i|^4 \nu(\mathrm{d}\boldsymbol{x}) \leqslant M,$$

for some constant M, then there exists a constant $C = C(d, \sigma, \kappa(\cdot), L_0, L_1, M)$ such that the fourth moments of the IPS (2.17) and the RB–IPS (2.21) are bounded by

$$\sup_{t\geqslant 0} \mathbb{E}\big[|x_t^i|^4\big] \leqslant C, \quad \sup_{t\geqslant 0} \mathbb{E}\big[|y_t^i|^4\big] \leqslant C, \quad i = 1, \cdots, N.$$

The following proof is inspired from Lemma 3.3 of Jin et al. (2020).

Proof of Lemma 4.4. We only prove the moment bound for the RB–IPS, because the IPS can be viewed as a special case of the RB–IPS when the batch size $p = N$. We aim to prove the following inequality:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}|y_t^i|^4 \leqslant -\beta \cdot \mathbb{E}|y_t^i|^4 + C, \quad i = 1, \cdots, N \tag{4.39}$$

for some positive constants $\beta, C > 0$. By Itô calculus, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}|y_t^i|^4 = 4\mathbb{E}\Big\{|y_t^i|^2\Big(y_t^i \cdot b(y_t^i) + y_t^i \cdot \gamma^i(\boldsymbol{y}_t)\Big)\Big\} + 2(d+2)\sigma^2\mathbb{E}|y_t^i|^2, \tag{4.40}$$

where the interaction force $\gamma^i(x)$ is given by (4.3).

On the one hand, the definition of $\kappa(r)$ yields the inequality

$$-x \cdot (b(x) - b(0)) \geqslant \frac{\sigma^2}{2}\kappa(|x|)|x|^2, \quad \forall x \in \mathbb{R}^d.$$

Hence the drift force part in (4.40) is bounded by

$$|y_t^i|^2 y_t^i \cdot b(y_t^i) \leqslant C|y_t^i|^3 - \frac{\sigma^2}{2}\kappa(|y_t^i|)|y_t^i|^2. \tag{4.41}$$

On the other hand, $|\gamma^i(\boldsymbol{x})| \leqslant L_1$ implies the interaction force part in (4.40) is bounded by

$$|y_t^i|^2 y_t^i \cdot \gamma^i(\boldsymbol{y}_t) \leqslant C|y_t^i|^3. \tag{4.42}$$

Combining the inequalities (4.41) and (4.42), from (4.40) we deduce

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}|y_t^i|^4 \leqslant -2\sigma^2\mathbb{E}\big(\kappa(|y_t^i|)|y_t^i|^2\big) + C\big(\mathbb{E}|y_t^i|^3 + \mathbb{E}|y_t^i|^2\big). \tag{4.43}$$

According to Assumption 2.2, there exist constants $\delta, R > 0$ such that $\kappa(r) \geqslant \delta$ for $r \geqslant R$. Also $\kappa(r)$ has a uniform lower bound for $r > 0$. Hence we have the inequality

$$-\kappa(r)r^4 = (\delta - \kappa(r))r^4 - \delta r^4 \leqslant C - \delta r^4, \quad \forall r \geqslant 0,$$

and thus

$$-\mathbb{E}\big(\kappa(|y_t^i|)|y_t^i|^4\big) \leqslant C - \delta \cdot \mathbb{E}|y_t^i|^4. \tag{4.44}$$

Hence from (4.43) and (4.44) we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}|y_t^i|^4 \leqslant -2\sigma^2\delta \cdot \mathbb{E}|y_t^i|^4 + C(\mathbb{E}|y_t^i|^3 + \mathbb{E}|y_t^i|^2 + 1). \tag{4.45}$$

Using the interpolation inequality, $\mathbb{E}|y_t^i|^3$ and $\mathbb{E}|y_t^i|^2$ can be bounded by $\mathbb{E}|y_t^i|^4$ plus constant. Therefore, (4.45) implies

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}|y_t^i|^4 \leqslant -\sigma^2\delta \cdot \mathbb{E}|y_t^i|^4 + C \tag{4.46}$$

for some constant C, which is exactly the inequality (4.39). Finally, it is easy to deduce

$$\sup_{t\geqslant 0}\mathbb{E}\big[|y_t^i|^4\big] \leqslant C,$$

which completes the proof. ■

We now present the strong error analysis between the IPS (2.17) and the RB–IPS (2.21). As is typical in strong error analysis, we assume that the Brownian motions $\{B_t^i\}_{i=1}^N$ in both the IPS and RB–IPS are identical. Therefore, the difference between the IPS and the RB–IPS arises solely from the interaction forces.

Lemma 4.5. *Under Assumptions 2.2 & 4.1, if the initial distribution $\nu \in \mathcal{P}(\mathbb{R}^{dN})$ satisfies*

$$\max_{1\leqslant i\leqslant N}\int_{\mathbb{R}^{dN}}|x^i|^4\nu(\mathrm{d}\boldsymbol{x})\leqslant M,$$

for some constant M, then for any $T>0$, there exists a constant $C=C(d,\sigma,\kappa(\cdot),L_0,L_1,M,T)$ such that the IPS (2.17) and the RB–IPS (2.21) satisfy

$$\sup_{0\leqslant t\leqslant T}\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}|x_t^i-y_t^i|^2\right)\leqslant C\left(\frac{h}{p-1}+h^2\right). \tag{4.47}$$

As a consequence, we have the inequality

$$\sup_{0\leqslant n\leqslant T/h}\mathcal{W}_1(\nu\mathcal{P}_{nh},\nu\mathcal{Q}_n^h)\leqslant C\sqrt{\frac{h}{p-1}+h^2}. \tag{4.48}$$

The proof can be found in Theorem 3.1 of Jin et al. (2020). When the batch size p is small, $\sqrt{h/(p-1)}$ dominates the Wasserstein error $\mathcal{W}_1(\nu\mathcal{P}_{nh},\nu\mathcal{Q}_n^h)$. In this sense, the Wasserstein error $\mathcal{W}_1(\nu\mathcal{P}_{nh},\nu\mathcal{Q}_n^h)$ has at least half-order convergence with respect to the time step h.

Main theorem: estimate of $\mathcal{W}_1(\pi,\pi^h)$ We are now prepared to estimate the approximation error of the RB–IPS (2.21), specifically the difference between the invariant distributions $\mathcal{W}_1(\pi,\pi^h)$. In particular, we utilize the triangle inequality framework to transition from $\mathcal{W}_1(\nu\mathcal{P}_{nh},\nu\mathcal{Q}_n^h)$ over finite time to $\mathcal{W}_1(\pi,\pi^h)$ in the long-time limit.

Theorem 4.3. *Under Assumptions 2.2 & 4.1, let c_0,φ_0 be defined as in Lemma 2.5, and Let π and π^h be the invariant distributions of the IPS (2.17) and the RB–IPS (2.21) respectively. If the constant L_1 in Assumption 4.1 satisfies*

$$L_1\leqslant\frac{1}{4}c_0\varphi_0\sigma^2,$$

then there exists a constant $C=C(d,\sigma,\kappa(\cdot),L_0)$ such that

$$\mathcal{W}_1(\pi,\pi^h)\leqslant C\sqrt{\frac{h}{p-1}+h^2}. \tag{4.49}$$

Proof of Theorem 4.3. Let ν_0 be the distribution in $\mathbb{R}^{dN}$ with all the N particles frozen at origin, then using Lemma 4.4 we have the uniform-in-time moment estimates

$$\sup_{t\geqslant 0}\left\{\max_{1\leqslant i\leqslant N}\int_{\mathbb{R}^{dN}}|x^i|^4(\nu_0\mathcal{P}_t)(\mathrm{d}\boldsymbol{x})\right\}\leqslant C_1,\quad \sup_{n\geqslant 0}\left\{\max_{1\leqslant i\leqslant N}\int_{\mathbb{R}^{dN}}|x^i|^4(\nu_0\mathcal{Q}_n^h)(\mathrm{d}\boldsymbol{x})\right\}\leqslant C_1, \tag{4.50}$$

where the constant $C_1=C_1(d,\sigma,\kappa(\cdot),L_0,L_1)$. Since π and π^h are the long-time limits of $\nu_0\mathcal{P}_t$ and $\nu_0\mathcal{Q}_n^h$ in the metric space $(\mathcal{P}_1(\mathbb{R}^{dN}),\mathcal{W}_1)$ respectively, from the moment estimate (4.50)

we have

$$\int_{\mathbb{R}^{dN}} \left(\frac{1}{N}\sum_{i=1}^{N}|x^i|\right)\pi(\mathrm{d}\boldsymbol{x}) \leqslant C, \quad \int_{\mathbb{R}^{dN}} \left(\frac{1}{N}\sum_{i=1}^{N}|x^i|\right)\pi^h(\mathrm{d}\boldsymbol{x}) \leqslant C_1.$$

Hence the Wasserstein error $\mathcal{W}_1(\pi,\pi^h)$ is bounded by

$$\mathcal{W}_1(\pi,\pi^h) \leqslant \int_{\mathbb{R}^{dN}} \left(\frac{1}{N}\sum_{i=1}^{N}|x^i|\right)\pi(\mathrm{d}\boldsymbol{x}) + \int_{\mathbb{R}^{dN}} \left(\frac{1}{N}\sum_{i=1}^{N}|x^i|\right)\pi^h(\mathrm{d}\boldsymbol{x}) \leqslant C_1.$$

In the following, we may assume the step size $h \leqslant C_1$. Instead of directly measuring the Wasserstein error $\mathcal{W}_1(\pi,\pi^h)$, we fix a constant $T>0$ and study the quantity $\mathcal{W}_1(\nu_0\mathcal{P}_T,\pi^h)$. By Theorem 4.1, for any integer $n \geqslant 0$ we have the inequality

$$\begin{aligned}\mathcal{W}_1(\nu_0\mathcal{P}_T,\pi^h) &= \mathcal{W}_1(\nu_0\mathcal{P}_T,\pi^h\mathcal{Q}_n^h)\\ &\leqslant \mathcal{W}_1(\nu_0\mathcal{P}_T\mathcal{Q}_n^h,\pi^h\mathcal{Q}_n^h) + \mathcal{W}_1(\nu_0\mathcal{P}_T,\nu_0\mathcal{P}_T\mathcal{Q}_n^h)\\ &\leqslant \frac{1}{\varphi_0}e^{-\beta nh}\mathcal{W}_1(\nu_0\mathcal{P}_T,\pi^h) + \mathcal{W}_1(\nu_0\mathcal{P}_T,\nu_0\mathcal{P}_T\mathcal{Q}_n^h).\end{aligned}$$

Since the step size $h \leqslant C_1$, we can choose the integer n as

$$n = \left\lceil \frac{\log(2/\varphi_0)}{\beta h} \right\rceil,$$

then we have the inequalities $\frac{1}{\varphi_0}e^{-\beta nh} \leqslant \frac{1}{2}$ and

$$nh \leqslant \left(\frac{\log(2/\varphi_0)}{\beta h}+1\right)h \leqslant \frac{\log(2/\varphi_0)}{\beta} + C_1,$$

which implies nh has an upper bound uniform in N and h. For this chosen n we have

$$\begin{aligned}\mathcal{W}_1(\nu_0\mathcal{P}_T,\pi^h) &\leqslant 2\cdot\mathcal{W}_1(\nu_0\mathcal{P}_T,\nu_0\mathcal{P}_T\mathcal{Q}_n^h)\\ &\leqslant 2\cdot\mathcal{W}_1(\nu_0\mathcal{P}_T,\nu_0\mathcal{P}_T\mathcal{P}_{nh}) + 2\cdot\mathcal{W}_1(\nu_0\mathcal{P}_T\mathcal{P}_{nh},\nu_0\mathcal{P}_T\mathcal{Q}_n^h)\\ &\leqslant \frac{1}{\varphi_0}e^{-\beta T}\mathcal{W}_1(\nu_0,\nu_0\mathcal{P}_{nh}) + 2\cdot\mathcal{W}_1(\nu_0\mathcal{P}_T\mathcal{P}_{nh},\nu_0\mathcal{P}_T\mathcal{Q}_n^h).\end{aligned}$$

Fixing the integer $n \geqslant 0$, we pass to the limit $T\to\infty$ and obtain

$$\mathcal{W}_1(\pi,\pi^h) \leqslant 2\varlimsup_{T\to\infty}\mathcal{W}_1(\nu_0\mathcal{P}_T\mathcal{P}_{nh},\nu_0\mathcal{P}_T\mathcal{Q}_n^h). \tag{4.51}$$

Since $\nu_0\mathcal{P}_T$ always has finite fourth moments as in (4.50), using Lemma 4.5 we have

$$\mathcal{W}_1(\nu_0\mathcal{P}_T\mathcal{P}_{nh},\nu_0\mathcal{P}_T\mathcal{Q}_n^h) \leqslant C\sqrt{\frac{h}{p-1}+h^2}, \quad \forall T>0, \tag{4.52}$$

where the constant $C = C(d,\sigma,\kappa(\cdot),L_0,L_1)$. Combining (4.51) and (4.52) we finally obtain

the estimate of the Wasserstein-1 error $\mathcal{W}_1(\pi, \pi^h)$:

$$\mathcal{W}_1(\pi, \pi^h) \leqslant C\sqrt{\frac{h}{p-1} + h^2},$$

which completes the proof. ■

We stress that the constant C in Theorem 4.3 does not depend on the number of particles N, the batch size p and the step size h.

Remark 4.4. *When the batch size $p = N$, there is RB–IPS (2.21) coincides with the IPS (2.17). However, this consistency property is not reflected in the expression for $\mathcal{W}_1(\pi, \pi^h)$ because the strong error analysis was used to derive $\mathcal{W}_1(\pi, \pi^h)$. Additionally, the strong error analysis only achieves half-order convergence with respect to the step size h. For a more refined result, we refer to the recent work (Huang et al., 2024), which establishes first-order convergence using the entropy method.*

4.3 Long-time error of discrete RB–IPS

In this section, we analyze the long-time error of the discrete RB–IPS (2.22). Given the established ergodicity of both the IPS (2.17) and the RB–IPS (2.21), our focus shifts to estimating the strong error of the discrete RB–IPS. By applying the triangle inequality framework from Section 2.4.3, we can effectively study its long-time behavior.

Strong error of discrete RB–IPS Since the strong error between the IPS and the RB–IPS is already established in Lemma 4.5, it remains to estimate the strong error between the RB–IPS and the discrete RB–IPS,

$$\sup_{0\leqslant n\leqslant T/h} \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N} |y^i_{nh} - Y^i_n|^2\right),$$

which accounts solely for the discretization error. We first prove the following result:

Lemma 4.6. *Under Assumption 4.1, if the initial distribution $\nu \in \mathcal{P}(\mathbb{R}^{dN})$ satisfies*

$$\max_{1\leqslant i\leqslant N} \int_{\mathbb{R}^{dN}} |x^i|^2 \nu(\mathrm{d}\boldsymbol{x}) \leqslant M,$$

then for any $T > 0$, there exists a constant $C = C(d, \sigma, L_0, L_1, M, T)$ such that the RB–IPS (2.21) satisfies

$$\sup_{0\leqslant t\leqslant T} \mathbb{E}|y^i_t|^2 \leqslant C, \qquad \sup_{t\in[nh,(n+1)h\wedge T)} \mathbb{E}|y^i_t - y^i_{nh}|^2 \leqslant Ch. \tag{4.53}$$

Proof of Lemma 4.6. By Itô calculus, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}|y_t^i|^2 = 2y_t^i \cdot (b(y_t^i) + \gamma^i(\boldsymbol{y}_t) + \sigma \dot{B}_t^i) + d\sigma^2,$$

where $\gamma^i(\cdot)$ is the interaction force defined in (4.3). Taking the expectation, we have

$$\mathbb{E}|y_t^i|^2 = \mathbb{E}|y_0^i|^2 + 2\int_0^t y_s^i \cdot \big(b(y_s^i) + \gamma^i(\boldsymbol{y}_s)\big)\mathrm{d}s + d\sigma^2 t. \tag{4.54}$$

On the one hand, $\gamma^i(\cdot)$ is uniformly bounded by L_1, hence

$$2\int_0^t y_s^i \cdot \gamma^i(\boldsymbol{y}_s)\mathrm{d}s \leqslant 2L_1 \int_0^t |y_s^i|\mathrm{d}s \leqslant L_1 \int_0^t |y_s^i|^2\mathrm{d}s + L_1 t. \tag{4.55}$$

On the other hand, using $|b(x)| \leqslant L_0(|x| + 1)$ we have

$$\begin{aligned} 2\int_0^t y_s^i \cdot b(y_s^i)\mathrm{d}s &\leqslant 2L_0 \int_0^t (|y_s^i|^2 + |y_s^i|)\mathrm{d}s \\ &\leqslant L_0 \int_0^t (3|y_s^i|^2 + 1)\mathrm{d}s \leqslant 3L_0 \int_0^t |y_s^i|^2 + L_0 t. \end{aligned} \tag{4.56}$$

Using the inequalities (4.54), (4.55) and (4.56), we obtain

$$2\int_0^t y_s^i \cdot \big(b(y_s^i) + \gamma^i(\boldsymbol{y}_s)\big)\mathrm{d}s \leqslant (3L_0 + L_1)\int_0^t |y_s^i|^2\mathrm{d}s + (L_0 + L_1)t. \tag{4.57}$$

Let $L := 3L_0 + L_1 + d\sigma^2$, then for any $t \in [0, T]$, we have

$$\mathbb{E}|y_t^i|^2 \leqslant \mathbb{E}|y_0^i|^2 + L\int_0^t |y_s^i|^2\mathrm{d}s + Lt \leqslant M + LT + L\int_0^t |y_s^i|^2\mathrm{d}s. \tag{4.58}$$

Using the Gronwall inequality (Lemma 7.21 of E et al. (2021)), we obtain from (4.58)

$$\mathbb{E}|y_t^i|^2 \leqslant (M + LT)\exp(LT), \quad t \in [0, T],$$

which yields the first inequality of Lemma 4.6. For the second inequality, we note that

$$y_t^i - y_{nh}^i = \int_{nh}^t \big(b(y_s^i) + \gamma^i(\boldsymbol{y}_s)\big)\mathrm{d}s + \sigma(W_t^i - W_{nh}^i). \tag{4.59}$$

Hence using Cauchy inequality, we obtain

$$\mathbb{E}|y_t^i - y_{nh}^i|^2 \leqslant 2h\int_{nh}^t \mathbb{E}\big|b(y_s^i) + \gamma^i(\boldsymbol{y}_s)\big|^2\mathrm{d}s + 2d\sigma^2 h. \tag{4.60}$$

Using the inequality $|b(x^i) + \gamma^i(\boldsymbol{x})| \leqslant L(|x^i| + 1)$, we obtain

$$\mathbb{E}|y_t^i - y_{t_n}^i|^2 \leqslant 4L^2 h\int_{nh}^t \big(\mathbb{E}|y_s^i|^2 + 1\big)\mathrm{d}s + 2d\sigma^2 h \leqslant 4L^2 TCh + 2d\sigma^2 h = Ch, \tag{4.61}$$

which produces the desired result. ∎

Subsequently, we can derive the strong error between the RB–IPS and the discrete RB–IPS.

Lemma 4.7. *Under Assumption 4.1, if the initial distribution $\nu \in \mathcal{P}(\mathbb{R}^{dN})$ satisfies*

$$\max_{1\leqslant i\leqslant N}\int_{\mathbb{R}^{dN}}|x^i|^2\nu(\mathrm{d}\boldsymbol{x})\leqslant M,$$

then for any $T>0$, there exists a constant $C=C(d,\sigma,L_0,L_1,M,T)$ such that the RB–IPS (2.21) and the discrete RB–IPS (2.22) satisfy

$$\sup_{0\leqslant n\leqslant T/h}\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}|y^i_{nh}-Y^i_n|^2\right)\leqslant Ch. \tag{4.62}$$

Proof of Lemma 4.7. Let the step size $h\leqslant T$. Define the trajectory difference $e^i_n=y^i_{nh}-Y^i_n$, and let $F^i(\boldsymbol{x})$ be th total force on the i-th particle, namely,

$$F^i(\boldsymbol{x})=b(x^i)+\gamma^i(\boldsymbol{x}),\quad \boldsymbol{x}\in\mathbb{R}^{dN}.$$

Then we can write the RB–IPS (2.21) and the discrete RB–IPS (2.22) as

$$y^i_{(n+1)h}=y^i_{nh}+\int_{nh}^{(n+1)h}F^i(\boldsymbol{y}_t)\mathrm{d}t+\sigma\Delta B^i_n, \tag{4.63}$$

$$Y^i_{n+1}=Y^i_n+\int_{nh}^{(n+1)h}F^i(\boldsymbol{Y}_n)\mathrm{d}t+\sigma\Delta B^i_n, \tag{4.64}$$

where $\Delta B^i_n:=W^i_{(n+1)h}-W^i_{nh}\sim\mathcal{N}(0,h)$. Then e^i_n satisfies the recurrence relation

$$e^i_{n+1}=e^i_n+\int_{nh}^{(n+1)h}(F^i(\boldsymbol{y}_t)-F^i(\boldsymbol{Y}_n))\mathrm{d}t. \tag{4.65}$$

Squaring both sides of (4.65) and applying Cauchy inequality, we obtain

$$\begin{aligned}
|e^i_{n+1}|^2&\leqslant(1+h)|e^i_n|^2+\left(1+\frac{1}{h}\right)\left(\int_{nh}^{(n+1)h}(F^i(\boldsymbol{y}_t)-F^i(\boldsymbol{Y}_n))\mathrm{d}t\right)^2\\
&\leqslant(1+h)|e^i_n|^2+(1+h)\int_{nh}^{(n+1)h}|F^i(\boldsymbol{y}_t)-F^i(\boldsymbol{Y}_n)|^2\mathrm{d}t\\
&\leqslant(1+h)|e^i_n|^2+2(1+h)\int_{nh}^{(n+1)h}|b(y^i_t)-b(Y^i_n)|^2\mathrm{d}t\\
&\qquad+2(1+h)\int_{nh}^{(n+1)h}|\gamma^i(\boldsymbol{y}_t)-\gamma^i(\boldsymbol{Y}_n)|^2\mathrm{d}t.
\end{aligned} \tag{4.66}$$

On the one hand, the Lipschitz condition on the drift force $b(\cdot)$ implies

$$\int_{nh}^{(n+1)h}|b(y^i_t)-b(Y^i_n)|^2\leqslant L_0^2\int_{nh}^{(n+1)h}|y^i_t-Y^i_n|^2\mathrm{d}t. \tag{4.67}$$

On the other hand, the Lipschitz condition on the interaction force γ^i implies

$$|\gamma^i(\boldsymbol{y}_t)-\gamma^i(\boldsymbol{Y}_n)|\leqslant L_1|y^i_t-Y^i_n|+\frac{L_1}{p-1}\sum_{j\neq i,j\in C}|y^j_t-Y^j_n|,$$

where C is the batch which contains i, and p is the batch size. By Cauchy inequality, we have

$$\begin{aligned}|\gamma^i(\boldsymbol{y}_t)-\gamma^i(\boldsymbol{Y}_n)|^2 &\leqslant 2L_1^2|y_t^i-Y_n^i|^2+2L_1^2\bigg(\frac{1}{p-1}\sum_{j\neq i,j\in C}|y_t^j-Y_n^j|\bigg)^2\\ &\leqslant 2L_1^2|y_t^i-Y_n^i|^2+\frac{2L_1^2}{p-1}\sum_{j\neq i,j\in C}|y_t^j-Y_n^j|^2.\end{aligned}$$

Hence we obtain the inequality

$$\int_{nh}^{(n+1)h}|\gamma^i(\boldsymbol{y}_t)-\gamma^i(\boldsymbol{Y}_n)|^2\mathrm{d}t\leqslant 2L_1^2\int_{nh}^{(n+1)h}|y_t^i-Y_n^i|^2\mathrm{d}t+\frac{2L_1^2}{p-1}\sum_{j\neq i,j\in C}\int_{nh}^{(n+1)h}|y_t^j-Y_n^j|^2\mathrm{d}t. \tag{4.68}$$

Combining the inequalities (4.66), (4.67) and (4.68), e_{n+1}^i can be bounded by

$$\begin{aligned}|e_{n+1}^i|^2 \leqslant{}& (1+h)|e_n^i|^2+(1+h)(2L_0^2+4L_1^2)\int_{nh}^{(n+1)h}|y_t^i-Y_n^i|^2\mathrm{d}t\\ &+(1+h)\frac{4L_1^2}{p-1}\sum_{j\neq i,j\in C}\int_{nh}^{(n+1)h}|y_t^j-Y_n^j|^2\mathrm{d}t.\end{aligned}$$

Summation over i gives

$$\sum_{i=1}^N|e_{n+1}^i|^2\leqslant(1+h)\sum_{i=1}^N|e_n^i|^2+(1+h)(2L_0^2+8L_1^2)\sum_{i=1}^N\int_{nh}^{(n+1)h}|y_t^i-Y_n^i|^2\mathrm{d}t. \tag{4.69}$$

Note that from Lemma 4.6

$$\mathbb{E}|y_t^i-Y_n^i|^2\leqslant 2\mathbb{E}|y_t^i-y_{nh}^i|^2+2\mathbb{E}|y_{nh}^i-Y_n^i|^2\leqslant Ch+2\mathbb{E}|e_n^i|^2,$$

hence integrating in the time interval $[nh,(n+1)h)$ gives

$$\int_{nh}^{(n+1)h}\mathbb{E}|y_t^i-Y_n^i|^2\mathrm{d}t\leqslant Ch^2+2h\mathbb{E}|e_n^i|^2.$$

Taking the expectation in (4.69), we obtain

$$\begin{aligned}\sum_{i=1}^N\mathbb{E}|e_{n+1}^i|^2 &\leqslant (1+h)\sum_{i=1}^N\mathbb{E}|e_n^i|^2+C(1+h)\bigg(Nh^2+h\sum_{i=1}^N\mathbb{E}|e_n^i|^2\bigg)\\ &\leqslant (1+Ch)\sum_{i=1}^N\mathbb{E}|e_n^i|^2+CNh^2.\end{aligned} \tag{4.70}$$

Note that $e_0^i\equiv 0$, the discrete Gronwall inequality applying on (4.70) gives

$$\frac{1}{N}\sum_{i=1}^N\mathbb{E}|e_n^i|^2\leqslant h\Big((1+Ch)^n-1\Big)\leqslant e^{Cnh}h\leqslant Ch,$$

which implies the strong error is bounded by Ch for $0\leqslant n\leqslant T/h$. ∎

Combining the results in Lemma 4.5 and Lemma 4.7, we obtain the strong error of the discrete RB–IPS (2.22) in a finite time.

Lemma 4.8. *Under Assumption 4.1, if the initial distribution $\nu \in \mathcal{P}(\mathbb{R}^{dN})$ satisfies*

$$\max_{1\leqslant i\leqslant N}\int_{\mathbb{R}^{dN}}|x^i|^4\nu(\mathrm{d}\boldsymbol{x})\leqslant M,$$

then for any $T>0$, there exists a constant $C=C(d,\sigma,L_0,L_1,M,T)$ such that the IPS (2.17) and the discrete RB–IPS (2.22) satisfy

$$\sup_{0\leqslant n\leqslant T/h}\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}|x^i_{nh}-Y^i_n|^2\right)\leqslant Ch. \tag{4.71}$$

As a consequence, we have the inequality

$$\sup_{0\leqslant n\leqslant T/h}\mathcal{W}_1(\nu\mathcal{P}_{nh},\nu\tilde{\mathcal{Q}}^h_n)\leqslant C\sqrt{h}. \tag{4.72}$$

Note that the requirement on the moment of the initial distribution ν increases to fourth-order, as stipulated in Lemma 4.5.

Uniform-in-time moment estimate To analyze the long-time behavior of the discrete RB–IPS (2.22), it is necessary to establish that the discrete RB–IPS maintains uniform-in-time fourth moments. The proof differs slightly from Lemma 4.4 due to the involvement of time discretization in the discrete RB–IPS.

Lemma 4.9. *Under Assumptions 2.2, there exists a constant $\alpha_0=\alpha_0(\sigma,\kappa(\cdot))>0$ such that*

$$\sup_{x\in\mathbb{R}^d}\left(x\cdot b(x)+\alpha_0|x|^2\right)<+\infty. \tag{4.73}$$

Proof of Lemma 4.9. Using the definition of the function $\kappa(r)$ in (2.63), we have

$$-\frac{x\cdot(b(x)-b(0))}{|x|^2}\geqslant\frac{\sigma^2}{2}\kappa(|x|),\quad\forall x\in\mathbb{R}^d,$$

which can be written as

$$-\frac{x\cdot b(x)}{|x|^2}\geqslant\frac{\sigma^2}{2}\kappa(|x|)-\frac{x\cdot b(0)}{|x|^2},\quad\forall x\in\mathbb{R}^d. \tag{4.74}$$

Since $\kappa(r)$ has asymptotic positivity, we can choose the constant α_0 as

$$\alpha_0=\frac{\sigma^2}{4}\varliminf_{r\to\infty}\kappa(r)>0.$$

Taking the limit inferior in (4.74) as $|x|\to\infty$, we obtain

$$\varliminf_{|x|\to\infty}\left(-\frac{x\cdot b(x)}{|x|^2}\right)\geqslant 2\alpha_0>0.$$

Then there exists a constant R such that when $|x| \geqslant R$, there holds the inequality

$$-x \cdot b(x) \geqslant \alpha_0 |x|^2, \quad \forall |x| \geqslant R.$$

Hence $x \cdot b(x) + \alpha_0 |x|^2 \leqslant 0$ when $|x| \geqslant R$. Since the drift function $b(x)$ is continuous, we conclude that $x \cdot b(x) + \alpha_0 |x|^2$ is globally bounded in $\mathbb{R}^d$. ∎

Next, we require the following general result to bound the fourth moments.

Lemma 4.10. *Under Assumptions 2.2 & 4.1, let α_0 be defined as in Lemma 4.9 and define*

$$h_0(\sigma, \kappa(\cdot), L_0) = \frac{1}{2} \min\left\{\frac{\alpha_0}{L_0^2}, \frac{1}{\alpha_0}\right\}. \tag{4.75}$$

Then there exists a constant $C = C(d, \sigma, \kappa(\cdot), L_0, L_1)$ such that for any step size $0 \leqslant h \leqslant h_0$ and $\gamma \in \mathbb{R}^d$ with $|\gamma| \leqslant L_1$, there holds the inequality

$$|x + b(x)h + \gamma h|^4 \leqslant \left(1 - \frac{\alpha h}{2}\right)|x|^4 + Ch. \tag{4.76}$$

Proof of Lemma 4.10. Using Lemma 4.9, there exists $C_0 = C_0(d, \sigma, \kappa(\cdot), L_0)$ such that

$$-x \cdot b(x) \geqslant \alpha_0 |x|^2 - C_0, \quad \forall x \in \mathbb{R}^d.$$

Hence we have the following estimate of $|x + b(x)h|^2$:

$$\begin{aligned} |x + b(x)h|^2 &= |x|^2 + 2x \cdot b(x)h + |b(x)|^2 h^2 \\ &\leqslant |x|^2 + 2(C_0 - \alpha_0 |x|^2)h + 2L_0^2(|x|^2 + 1)h^2 \\ &= (1 + 2L_0^2 h^2 - 2\alpha_0 h)|x|^2 + Ch. \end{aligned} \tag{4.77}$$

Since the step size $h \leqslant \frac{\alpha_0}{2L_0^2}$, the inequality (4.77) implies

$$|x + b(x)h|^2 \leqslant (1 - \alpha_0 h)|x|^2 + Ch, \tag{4.78}$$

Square both sides of (4.78) and utilize $h \leqslant \frac{1}{2\alpha}$, then

$$\begin{aligned} |x + b(x)h|^4 &\leqslant (1 - \alpha h)^2 |x|^4 + C|x|^2 h + Ch^2 \\ &\leqslant (1 - \alpha h)^2 |x|^4 + \frac{\alpha}{2}|x|^4 h + Ch \\ &= (1 - \alpha h)|x|^4 + Ch. \end{aligned}$$

Using the interpolation inequality, for some constant $C = C(d, \sigma, \kappa(\cdot), L_0, L_1)$ we have

$$|x + b(x)h + \gamma h|^4 \leqslant \left(1 + \frac{1}{3}\alpha h\right)|x + b(x)h|^4 + \frac{C}{h^3}|\gamma h|^4$$

$$\leqslant \left(1 - \frac{1}{3}\alpha h\right)(1 + \alpha h)|x|^4 + Ch$$
$$\leqslant \left(1 - \frac{1}{2}\alpha h\right)|x|^4 + Ch,$$

which completes the proof. ∎

Now we can obtain the uniform-in-time fourth moments of the discrete RB–IPS (2.22).

Theorem 4.4. *Under Assumptions 2.2 & 4.1, let h_0 be defined as in Lemma 4.10. If the initial distribution $\nu \in \mathcal{P}(\mathbb{R}^{dN})$ satisfies*

$$\max_{1\leqslant i\leqslant N} \int_{\mathbb{R}^{dN}} |x^i|^4 \nu(\mathrm{d}\boldsymbol{x}) \leqslant M,$$

and the step size $h \leqslant h_0$, then there exists a constant $C = C(d, \sigma, \kappa(\cdot), L_0, L_1, M)$ such that

$$\sup_{n\geqslant 0} \mathbb{E}|Y_n^i|^4 \leqslant C, \quad i = 1, \cdots, N. \tag{4.79}$$

Proof of Theorem 4.4. The update scheme of the discrete RB–IPS (2.22) is given by

$$Y_{n+1}^i = Y_n^i + b(Y_n^i)h + \gamma^i(\boldsymbol{Y}_n)h + \sigma\Delta B_n^i, \tag{4.80}$$

where $\Delta B_n^i \sim \mathcal{N}(0, h)$, and $\gamma^i(\cdot)$ is defined in (4.3). Since the random variable ΔB_n^i is independent of the numerical solution $\boldsymbol{Y}_n$, we have

$$\begin{aligned}\mathbb{E}|Y_{n+1}^i|^4 &= \mathbb{E}|Y_n^i + b(Y_n^i)h + \gamma^i(\boldsymbol{Y}_n)h|^4 + 6\,\mathbb{E}|Y_n^i + b(Y_n^i)h + \gamma^i(\boldsymbol{Y}_n)h|^2\,\mathbb{E}|\sigma\Delta B_n^i|^2 + \mathbb{E}|\sigma\Delta B_n^i|^4 \\ &\leqslant \mathbb{E}|Y_n^i + b(Y_n^i)h + \gamma^i(\boldsymbol{Y}_n)h|^4 + Ch\,\mathbb{E}|Y_n^i + b(Y_n^i)h + \gamma^i(\boldsymbol{Y}_n)h|^2 + Ch^2 \\ &= \left(1 + \frac{1}{8}\alpha h\right)\mathbb{E}|Y_n^i + b(Y_n^i)h + \gamma^i(\boldsymbol{Y}_n)h|^4 + Ch. \end{aligned} \tag{4.81}$$

Applying Lemma 4.10 in (4.81), we obatin

$$\mathbb{E}|Y_{n+1}^i|^4 \leqslant \left(1 + \frac{1}{8}\alpha h\right)\left(\left(1 - \frac{1}{2}\alpha h\right)\mathbb{E}|Y_n^i|^4 + Ch\right) + Ch \leqslant \left(1 - \frac{1}{4}\alpha h\right)\mathbb{E}|Y_n^i|^4 + Ch.$$

yielding the desired moment bound. ∎

Main theorem: estimate of $\mathcal{W}_1(\nu\tilde{Q}_n^h, \pi)$ Using the uniform-in-time moments, we can derive the long-time error of the RB–IPS (2.22). As the RB–IPS is a discrete-time Markov chain, a discrete version of the triangle inequality framework (Lemma 2.6) is required.

Lemma 4.11. *Given $m \in \mathbb{N}$, $\varepsilon > 0$ and $q \in (0, 1)$. If a nonnegative sequence $\{a_n\}_{n\geqslant 0}$ satisfies*

$$a_n \leqslant \varepsilon + qa_{n-m}, \quad \forall n \geqslant m, \tag{4.82}$$

then

$$a_n \leqslant \frac{\varepsilon}{1-q} + Mq^{\frac{n}{m}-1}, \quad \forall n \geqslant 0, \tag{4.83}$$

where $M = \max_{0\leqslant k\leqslant m-1} a_k$.

Proof of Lemma 4.11. By induction on the integer $s \geqslant 1$, it is easy to verify if $n \geqslant sm$, then

$$a_n \leqslant \varepsilon\frac{1-q^s}{1-q} + q^s a_{n-sm}. \tag{4.84}$$

For any integer $n \geqslant 0$, let $n = sm + r$ for some integer $s \geqslant 0$ and $r \in \{0, 1, \cdots, m-1\}$. Then

$$a_n \leqslant \frac{\varepsilon}{1-q} + Mq^s \leqslant \frac{\varepsilon}{1-q} + Mq^{\frac{n}{m}-1}, \tag{4.85}$$

yielding the inequality (4.83). ∎

Combining the finite time strong error (Lemma 4.8) and the geometric ergodicity of the IPS (Theorem 4.2), we employ the triangle inequality in Lemma 4.11 to estimate the long-time error of the discrete RB–IPS.

Theorem 4.5. *Under Assumptions 2.2 & 4.1, let c_0, φ_0 be defined as in Lemma 2.5, and h_0 be defined as in Lemma 4.10. If the initial distribution $\nu \in \mathcal{P}(\mathbb{R}^{dN})$ satisfies*

$$\max_{1\leqslant i\leqslant N} \int_{\mathbb{R}^{dN}} |x^i|^4 \nu(\mathrm{d}\boldsymbol{x}) \leqslant M,$$

and the constant L_1 in Assumption 4.1 and the step size h satisfy

$$L_1 \leqslant \frac{c_0\varphi_0\sigma^2}{4}, \quad h \leqslant h_0,$$

then there exist positive constants $\lambda = \lambda(d, \sigma, \kappa(\cdot), L_0)$ and $C = C(d, \sigma, \kappa(\cdot), L_0, M)$ such that the discrete RB–IPS (2.22) satisfies

$$\mathcal{W}_1(\nu\tilde{\mathcal{Q}}_n^h, \pi) \leqslant C\sqrt{h} + Ce^{-\lambda nh}, \quad \forall n \geqslant 0. \tag{4.86}$$

Proof of Theorem 4.5. For any given integers $n \geqslant m$, we have the following triangle inequality

$$\mathcal{W}_1(\nu\tilde{\mathcal{Q}}_n^h, \pi) \leqslant \mathcal{W}_1(\nu\tilde{\mathcal{Q}}_{n-m}^h\tilde{\mathcal{Q}}_m^h, \nu\tilde{\mathcal{Q}}_{n-m}^h\mathcal{P}_{mh}) + \mathcal{W}_1(\nu\tilde{\mathcal{Q}}_{n-m}^h\mathcal{P}_{mh}, \pi\mathcal{P}_{mh}). \tag{4.87}$$

By Theorem 4.4, $\nu\tilde{\mathcal{Q}}_{n-m}^h$ has uniform-in-time fourth order moment, i.e., there exists a constant $M' = M'(d, \sigma, \kappa(\cdot), L_0, M)$ such that

$$\max_{1\leqslant i\leqslant N} \left\{ \sup_{n\geqslant m} \int_{\mathbb{R}^{dN}} |x^i|^4 \big(\nu\tilde{\mathcal{Q}}_{n-m}^h\big)(\mathrm{d}\boldsymbol{x}) \right\} \leqslant M'. \tag{4.88}$$

Hence by Lemma 4.8, there exists a constant $C_1 = C_1(d, \sigma, \kappa(\cdot), L_0, M, mh)$ such that

$$\mathcal{W}_1(\nu\tilde{\mathcal{Q}}_{n-m}^h\tilde{\mathcal{Q}}_m^h, \nu\tilde{\mathcal{Q}}_{n-m}^h\mathcal{P}_{mh}) \leqslant C_1\sqrt{h}, \quad \forall n \geqslant m. \tag{4.89}$$

Note that the constant C_1 depends on the upper bound of mh, which signifies the evolution time of the discrete IPS (2.22). By Theorem 4.2, there exist constants $C_0, \beta > 0$ depending only on σ and $\kappa(\cdot)$ such that

$$\mathcal{W}_1(\nu\tilde{\mathcal{Q}}^h_{n-m}\mathcal{P}_{mh}, \pi\mathcal{P}_{mh}) \leqslant C_0 e^{-\beta mh}\mathcal{W}_1(\nu\tilde{\mathcal{Q}}^h_{n-m}, \pi), \quad \forall n \geqslant m. \tag{4.90}$$

From (4.87), (4.89) and (4.90) we employ the triangle inequality and obtain

$$\mathcal{W}_1(\nu\tilde{\mathcal{Q}}^h_n, \pi) \leqslant C_1\sqrt{h} + C_0 e^{-\beta mh}\mathcal{W}_1(\nu_0\tilde{\mathcal{Q}}^h_{n-m}, \pi), \quad \forall n \geqslant m. \tag{4.91}$$

For given time step $h > 0$, we wish to choose m to satisfy $C_0 e^{-\beta mh} = \frac{1}{e}$, so that Lemma 4.11 can be applied. However, m is restricted to be an integer, thus our choice is

$$m = \left\lceil \frac{\log C_0 + 1}{\beta h} \right\rceil.$$

It is easy to check mh has an upper bound independent of h,

$$mh \leqslant \left(\frac{\log C_0 + 1}{\beta h} + 1\right)h \leqslant \frac{\log C_0 + 1}{\beta} + \frac{1}{2\alpha}, \tag{4.92}$$

hence the constant C_1 in (4.89) can be made independent of h, i.e., $C_1 = C_1(d, \sigma, \kappa(\cdot), L_0, M)$. Note that for this choice of m we have $C_0 e^{-\beta mh} \leqslant \frac{1}{e}$, and (4.91) implies

$$\mathcal{W}_1(\nu\tilde{\mathcal{Q}}^h_n, \pi) \leqslant C_1\sqrt{h} + \frac{1}{e}\mathcal{W}_1(\nu_0\tilde{\mathcal{Q}}^h_{n-m}, \pi), \quad \forall n \geqslant m.$$

Applying Lemma 4.11 with $a_n := \mathcal{W}_1(\nu_0\tilde{\mathcal{Q}}^h_n, \pi)$, we have

$$\mathcal{W}_1(\nu\tilde{\mathcal{Q}}^h_n, \pi) \leqslant C_1\sqrt{h} + M_0 e^{1-\frac{n}{m}}, \quad \forall n \geqslant 0, \tag{4.93}$$

where the constant

$$M_0 := \sup_{0\leqslant k\leqslant m-1} \mathcal{W}_1(\nu\tilde{\mathcal{Q}}^h_k, \pi) \leqslant \sup_{k\geqslant 0} \mathcal{W}_1(\nu\tilde{\mathcal{Q}}^h_k, \pi). \tag{4.94}$$

Introduce the normalized first moment for $\nu \in \mathcal{P}(\mathbb{R}^{dN})$ by

$$\mathcal{M}_1(\nu) = \int_{\mathbb{R}^{dN}} \left(\frac{1}{N}\sum_{i=1}^N |x^i|\right)\nu(\mathrm{d}\boldsymbol{x}),$$

then the normalized Wasserstein-1 distance in (4.94) is bounded by

$$\mathcal{W}_1(\nu\tilde{\mathcal{Q}}^h_k, \pi) \leqslant \mathcal{M}_1(\nu\tilde{\mathcal{Q}}^h_k) + \mathcal{M}_1(\pi). \tag{4.95}$$

On the one hand, $\nu\tilde{\mathcal{Q}}^h_k$ has uniform-in-time fourth moments according to Theorem 4.4, hence

there exists a constant $C_2 = C_2(d, \sigma, \kappa(\cdot), L_0, M)$ such that

$$\sup_{k\geqslant 0} \mathcal{M}_1(\nu\tilde{\mathcal{Q}}_k^h) \leqslant C_2. \tag{4.96}$$

On the other hand, by Lemma 4.4, for the invariant distribution π of the IPS (2.17), there exists a constant $C_2 = C_2(d, \sigma, \kappa(\cdot), L_0, M)$ such that

$$\mathcal{M}_1(\pi) \leqslant C_2. \tag{4.97}$$

Combining the inequalities (4.95), (4.96) and (4.97), we obtain

$$\mathcal{W}_1(\nu\tilde{\mathcal{Q}}_n^h, \pi) \leqslant C_1\sqrt{h} + C_2 e^{-\frac{n}{m}}, \quad \forall n \geqslant 0, \tag{4.98}$$

where both constants C_1, C_2 only depend on $(d, \sigma, \kappa(\cdot), L_0, M)$. Note that by the choice of m in (4.92) ensures that

$$\frac{n}{m} \geqslant \frac{n}{\frac{\log C_0+1}{\beta h}+1} \geqslant \frac{\beta n h}{\log C_0 + \beta/(2\alpha) + 1},$$

hence by defining the constant

$$\lambda := \frac{\beta}{\log C_0 + \beta/(2\alpha) + 1}, \tag{4.99}$$

there holds $e^{-n/m} \leqslant e^{-\lambda n h}$. Hence (4.98) implies

$$\mathcal{W}_1(\nu\tilde{\mathcal{Q}}_n^h, \pi) \leqslant C\sqrt{h} + Ce^{-\lambda n h}, \quad \forall n \geqslant 0,$$

which is exactly the long-time sampling error of the discrete RB–IPS (2.22). ■

Note that the sampling error $\mathcal{W}_1(\nu\tilde{\mathcal{Q}}_n^h, \pi)$ approaches 0 as the number of steps $n \to 0$ and the step size $h \to \infty$. Hence we obtain the consistency of the discrete RB–IPS (2.21) in sampling the invariant distribution π of the IPS.

Remark 4.5. *As shown in* (4.99)*, the decay rate λ in the discrete RB–IPS* (2.22) *is smaller than the convergence rate β of the IPS* (2.17) *and the RB–IPS* (2.21)*. However, λ is guaranteed to be independent of the number of particles N, the batch size p, and the step size h.*

Remark 4.6. *The batch size p does not explicitly appear in the long-time error* (4.86) *because the discretization error, which is of half-order in the step size h, dominates the random batch error described in Theorem 4.3.*

4.4 Long-time error in approximating MVP

The discrete RB–IPS (2.22) can also function as a sampling algorithm for the MVP (2.18). Since the MVP is the mean-field limit of the IPS (2.17), the sampling is accurate only when the number of particles $N \to \infty$. To estimate the sampling error of the discrete RB–IPS, it is necessary to evaluate the difference between the invariant distributions $\pi \in \mathcal{P}(\mathbb{R}^{dN})$ of the IPS and $\bar{\pi} \in \mathcal{P}(\mathbb{R}^d)$ of the MVP. This analysis falls under the classical topic of propagation of chaos (McKean, 1967; Sznitman, 1991; Guillin and Monmarché, 2021).

Propagation of chaos While the IPS (2.17) is defined for N particles, the MVP (2.18) characterizes the evolution of the distribution law of a single particle. To compare the difference between the IPS and the MVP, we introduce N copies $\{\bar{x}_t^i\}_{i=1}^N$ of the MVP:

$$\dot{\bar{x}}_t^i = b(\bar{x}_t^i) + (K * \bar{\mu}_t)(\bar{x}_t^i) + \sigma \dot{B}_t^i, \tag{4.100}$$

where the initial values $\{\bar{x}_0^i\}_{i=1}^N$ are drawn independently from the same distribution $\nu \in \mathcal{P}(\mathbb{R}^d)$, $\{B_t^i\}_{i=1}^N$ are N independent Brownian motions in $\mathbb{R}^d$ (identical to those of the IPS), and $\bar{\mu}_t \in \mathcal{P}(\mathbb{R}^d)$ is the distribution law of each $\bar{x}_t^i$. Since the distribution laws of x_t^i for different i are identical, we omit the superscript i in $\bar{\mu}_t$.

The key difference between the IPS (2.17) and the MVP system (4.100) is that the particles in $\{x_t^i\}_{i=1}^N$ interact with each other, whereas the particles in $\{\bar{x}_t^i\}_{i=1}^N$ are fully decoupled, meaning the evolution of the N particles in $\{\bar{x}_t^i\}_{i=1}^N$ is independent. In the synchronous coupling of the IPS and the MVP system, we assume that for some distribution $\nu \in \mathcal{P}(\mathbb{R}^d)$, the initial values

$$x_0^i = \bar{x}_0^i \sim \nu, \quad i = 1, \cdots, N,$$

are independently sampled from ν, and the Brownian motions $\{B_t^i\}_{i=1}^N$ are identical. The strong error between the IPS and the MVP system is a well-established result in the propagation of chaos (see Theorem 3.1 of Chaintron and Diez (2021) for reference).

Theorem 4.6. *Under Assumption 4.1, for any $T > 0$, there exists a constant $C = C(d, \sigma, L_0, L_1, T)$ such that the IPS (2.17) and the MVP system (4.100) satisfy*

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E}\left[\sup_{0 \leqslant t \leqslant T} |x_t^i - \bar{x}_t^i|^2 \right] \leqslant \frac{C}{N}. \tag{4.101}$$

We note that the IPS (2.17) in this paper differs slightly from the setting in Chaintron and Diez

(2021), where the interaction force $\gamma^i(\boldsymbol{x})$ is defined as

$$\gamma^i(\boldsymbol{x}) = \frac{1}{N}\sum_{j=1}^{N} K(x^i - x^j) \tag{4.102}$$

instead of

$$\gamma^i(\boldsymbol{x}) = \frac{1}{N-1}\sum_{j\neq i} K(x^i - x^j). \tag{4.103}$$

This minor difference in the definition of γ^i does not affect the final result of the propagation of chaos. The proof of Theorem 4.6 under the setting (4.103) can be found in Proposition 4.2 of Jin and Li (2022).

We can also express the result of Theorem 4.6 in terms of semigroups. For a given $\nu \in \mathcal{P}(\mathbb{R}^d)$, let $\nu^{\otimes N} \in \mathcal{P}(\mathbb{R}^{dN})$ denote the product distribution of N independent copies of ν. Then, $\nu^{\otimes N}\mathcal{P}_t \in \mathcal{P}(\mathbb{R}^{dN})$ represents the distribution law of the IPS (2.17), while $\nu^{\otimes N}\bar{\mathcal{P}}_t^{\otimes N} \in \mathcal{P}(\mathbb{R}^{dN})$ represents the distribution law of the MVP system (4.100). Theorem 4.6 implies

$$\sup_{0\leqslant t\leqslant T} \mathcal{W}_1(\nu^{\otimes N}\bar{\mathcal{P}}_t^{\otimes N}, \nu^{\otimes N}\mathcal{P}_t) \leqslant \frac{C}{\sqrt{N}}. \tag{4.104}$$

Using the strong error of the discrete RB–IPS (2.22) in Lemma 4.8, we obtain the following finite-time error estimate.

Lemma 4.12. *Under Assumption 4.1, if the initial distribution $\nu \in \mathcal{P}(\mathbb{R}^d)$ satisfies*

$$\int_{\mathbb{R}^d} |x|^4 \nu(\mathrm{d}x) \leqslant M,$$

then for any $T > 0$, there exist constants $C_1 = C_1(d, \sigma, L_0, L_1, M, T)$ and $C_2 = C_2(d, \sigma, L_0, L_1, T)$ such that the MVP system (4.100) and the discrete RB–IPS (2.22) satisfy

$$\sup_{0\leqslant n\leqslant T/h} \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N} |\bar{x}_{nh}^i - Y_n^i|^2\right) \leqslant Ch + \frac{C}{N}. \tag{4.105}$$

As a consequence, we have the inequality

$$\sup_{0\leqslant n\leqslant T/h} \mathcal{W}_1(\nu^{\otimes N}\bar{\mathcal{P}}_{nh}^{\otimes N}, \nu^{\otimes N}\tilde{\mathcal{Q}}_n^h) \leqslant C\sqrt{h} + \frac{C}{\sqrt{N}}. \tag{4.106}$$

Estimate of $\mathcal{W}_1(\bar{\pi}^{\otimes N}, \pi)$ To analyze the long-time behavior of the MVP (2.18), it is necessary to quantify the difference $\mathcal{W}_1(\bar{\pi}^{\otimes N}, \pi)$. This can be obtained using the ergodicity of the MVP (2.18), as stated below.

Theorem 4.7 (ergodicity of MVP)**.** *Under Assumptions 2.2 & 4.1, let $f(r), c_0, \varphi_0$ be defined as in Lemma 2.5, and $(\bar{\mathcal{P}}_t)_{t\geqslant 0}$ be the dual semigroup of the MVP (2.21). If the constant L_1 in*

Assumption 4.1 satisfies

$$L_1 \leqslant \frac{1}{4} c_0 \varphi_0 \sigma^2,$$

then for any probability distributions μ, ν in $\mathbb{R}^d$, we have

$$\mathcal{W}_f(\mu \bar{\mathcal{P}}_t, \nu \bar{\mathcal{P}}_t) \leqslant e^{-\beta t} \mathcal{W}_f(\mu, \nu), \quad n = 0, 1, \cdots, \tag{4.107}$$

where $\beta = \frac{1}{2} c_0 \varphi_0 \sigma^2$. As a consequence, in the normalized Wasserstein-1 distance we have

$$\mathcal{W}_1(\mu \bar{P}_t, \nu \bar{P}_t) \leqslant \frac{1}{\varphi_0} e^{-\beta t} \mathcal{W}_1(\mu, \nu), \quad n = 0, 1, \cdots. \tag{4.108}$$

The proof can be conveniently obtained from the ergodicity of the IPS (2.17) in Theorem 4.2, whose mean-field limit is exactly the MVP (2.18).

Proof of Theorem 4.7. Given the distributions $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, by Theorem 4.2 we have

$$\mathcal{W}_1(\mu^{\otimes N} \mathcal{P}_t, \nu^{\otimes N} \mathcal{P}_t) \leqslant \frac{1}{\varphi_0} e^{-\beta t} \mathcal{W}_1(\mu^{\otimes N}, \nu^{\otimes N}) = \frac{1}{\varphi_0} e^{-\beta t} \mathcal{W}_1(\mu, \nu), \quad \forall t \geqslant 0.$$

Here, $(P_t)_{t \geqslant 0}$ is the semigroup of the IPS in $\mathbb{R}^{dN}$, and $\mathcal{W}_1(\mu, \nu)$ denotes the usual Wasserstein-1 distance between μ and ν. Using the triangle inequality, we have

$$\begin{aligned}
&\mathcal{W}_1(\mu \bar{\mathcal{P}}_t, \nu \bar{\mathcal{P}}_t) \\
&= \mathcal{W}_1(\mu^{\otimes N} \bar{\mathcal{P}}_t^{\otimes N}, \nu^{\otimes N} \bar{\mathcal{P}}_t^{\otimes N}) \\
&\leqslant \mathcal{W}_1(\mu^{\otimes N} \bar{\mathcal{P}}_t^{\otimes N}, \mu^{\otimes N} \mathcal{P}_t) + \mathcal{W}_1(\nu^{\otimes N} \bar{\mathcal{P}}_t^{\otimes N}, \nu^{\otimes N} \mathcal{P}_t) + \mathcal{W}_1(\mu^{\otimes N} \mathcal{P}_t, \nu^{\otimes N} \mathcal{P}_t) \\
&\leqslant \mathcal{W}_1(\mu^{\otimes N} \bar{\mathcal{P}}_t^{\otimes N}, \mu^{\otimes N} \mathcal{P}_t) + \mathcal{W}_1(\nu^{\otimes N} \bar{\mathcal{P}}_t^{\otimes N}, \nu^{\otimes N} \mathcal{P}_t) + \frac{1}{\varphi_0} e^{-\beta t} \mathcal{W}_1(\mu, \nu).
\end{aligned} \tag{4.109}$$

By Theorem 4.6, for given $t > 0$ there exists a constant $C_0 = C_0(d, \sigma, L_0, L_1, t)$ such that

$$\mathcal{W}_1(\mu^{\otimes N} \bar{\mathcal{P}}_t^{\otimes N}, \mu^{\otimes N} \mathcal{P}_t) \leqslant \frac{C_0}{N}. \tag{4.110}$$

Using the inequalities (4.109) and (4.110) we obtain

$$\mathcal{W}_1(\mu \bar{\mathcal{P}}_t, \nu \bar{\mathcal{P}}_t) \leqslant \frac{2C_0}{\sqrt{N}} + \frac{1}{\varphi_0} e^{-\beta t} \mathcal{W}_1(\mu, \nu).$$

Fix $t > 0$ and let $N \to \infty$, we obtain

$$\mathcal{W}_1(\mu \bar{\mathcal{P}}_t, \nu \bar{\mathcal{P}}_t) \leqslant C e^{-\beta t} \mathcal{W}_1(\mu, \nu),$$

yielding the desired result. ∎

Next it is convenient to apply the triangle inequality framework to estimate $\mathcal{W}_1(\bar{\pi}^{\otimes N}, \pi)$.

Theorem 4.8. *Under Assumptions 2.2 & 4.1, let $f(r), c_0, \varphi_0$ be defined as in Lemma 2.5, and $(\bar{\mathcal{P}}_t)_{t\geqslant 0}$ be the dual semigroup of the MVP (2.21). If the constant L_1 in Assumption 4.1 satisfies*

$$L_1 \leqslant \frac{1}{4} c_0 \varphi_0 \sigma^2,$$

then the MVP has a unique invariant distribution $\bar{\pi} \in \mathcal{P}(\mathbb{R}^d)$, and

$$\mathcal{W}_1(\nu \bar{\mathcal{P}}_t, \bar{\pi}) \leqslant \frac{1}{\varphi_0} e^{-\beta t} \mathcal{W}_1(\nu, \bar{\pi}), \quad \forall t \geqslant 0, \tag{4.111}$$

where $\beta = \frac{1}{2} c_0 \varphi_0 \sigma^2$. Furthermore, there exists a constant $C = C(d, \sigma, \kappa(\cdot), L_0)$ such that

$$\mathcal{W}_1(\bar{\pi}^{\otimes N}, \pi) \leqslant \frac{C}{\sqrt{N}}. \tag{4.112}$$

Since $(\bar{\mathcal{P}}_t)_{t\geqslant 0}$ is a nonlinear dual semigroup, we cannot use the same technique as in the linear case. Our proof below is partially inspired from Theorem 5.1 of Cañizo and Mischler (2023).

Proof of Theorem 4.8. First we prove the existence of the invariant distribution $\bar{\pi} \in \mathcal{P}(\mathbb{R}^d)$ of the MVP (2.18). Choose the constant T which satisfies $\frac{1}{\varphi_0} e^{-\beta T} = \frac{1}{2}$, then we have

$$\mathcal{W}_1(\mu \bar{\mathcal{P}}_T, \nu \bar{\mathcal{P}}_T) \leqslant \frac{1}{2} \mathcal{W}_1(\mu, \nu)$$

for any probability distributions $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$. Hence the mapping $\mu \mapsto \mu \bar{\mathcal{P}}_T$ is contractive in the complete metric space $(\mathcal{P}_1(\mathbb{R}^d), \mathcal{W}_1)$. Using the Banach fixed point theorem, there exists a unique fixed point $\bar{\pi} \in \mathcal{P}_1(\mathbb{R}^d)$ such that

$$\bar{\pi} \bar{\mathcal{P}}_T = \bar{\pi}.$$

Since $(\bar{\mathcal{P}}_t)_{t\geqslant 0}$ forms a semigroup, for any $t \geqslant 0$ we have

$$\left(\bar{\pi} \bar{\mathcal{P}}_t\right) \bar{\mathcal{P}}_T = \bar{\pi} \bar{\mathcal{P}}_t,$$

which implies $\bar{\pi} \bar{\mathcal{P}}_t \in \mathcal{P}_1(\mathbb{R}^d)$ is the invariant distribution of the operator $\bar{\mathcal{P}}_T$. Due to the uniqueness of the invariant distribution $\bar{\pi}$ for the operator $\bar{\mathcal{P}}_T$, we obtain

$$\bar{\pi} \bar{\mathcal{P}}_t = \bar{\pi}, \quad \forall t \geqslant 0,$$

hence $\bar{\pi} \in \mathcal{P}_1(\mathbb{R}^d)$ is the invariant distribution of the semigroup $(\bar{\mathcal{P}}_t)_{t\geqslant 0}$.

Next we estimate the difference between the invariant distributions $\pi \in \mathcal{P}_1(\mathbb{R}^{dN})$ and $\bar{\pi} \in \mathcal{P}_1(\mathbb{R}^d)$. We choose the constant T according to $\frac{1}{\varphi_0} e^{-\beta T} = \frac{1}{2}$. Using the propagation of chaos in Theorem 4.6, there exists a constant $C = C(d, \sigma, \kappa(\cdot), L_0)$ such that

$$\mathcal{W}_1(\bar{\pi}^{\otimes N}, \pi) = \mathcal{W}_1(\bar{\pi}^{\otimes N} \bar{\mathcal{P}}_T^{\otimes N}, \pi \mathcal{P}_T)$$

$$
\begin{aligned}
&\leqslant \mathcal{W}_1(\bar{\pi}^{\otimes N}\bar{\mathcal{P}}_T^{\otimes N}, \bar{\pi}^{\otimes N}\mathcal{P}_T) + \mathcal{W}_1(\bar{\pi}^{\otimes N}\mathcal{P}_T, \pi\mathcal{P}_T) \\
&\leqslant \frac{C}{\sqrt{N}} + \frac{1}{\varphi_0} e^{-\beta T}\mathcal{W}_1(\bar{\pi}^{\otimes N}, \pi) \\
&= \frac{C}{\sqrt{N}} + \frac{1}{2}\mathcal{W}_1(\bar{\pi}^{\otimes N}, \pi),
\end{aligned}
$$

which implies the inequality $\mathcal{W}_1(\bar{\pi}^{\otimes N}, \pi) \leqslant \frac{C}{\sqrt{N}}$. ∎

Main theorem: long-time error in approximating MVP Combining Theorem 4.5 and Theorem 4.8, we immediately obtain the following result of long-time sampling error of the discrete RB–IPS (2.22) in approximating the invariant distribution $\bar{\pi} \in \mathcal{P}(\mathbb{R}^d)$.

Theorem 4.9. *Under Assumptions 2.2 & 4.1, let c_0, φ_0 be defined as in Lemma 2.5, and h_0 be defined as in Lemma 4.10. If the initial distribution $\nu \in \mathcal{P}(\mathbb{R}^{dN})$ satisfies*

$$
\max_{1\leqslant i\leqslant N} \int_{\mathbb{R}^{dN}} |x^i|^4 \nu(\mathrm{d}\boldsymbol{x}) \leqslant M,
$$

and the constant L_1 in Assumption 4.1 and the step size h satisfy

$$
L_1 \leqslant \frac{c_0\varphi_0\sigma^2}{4}, \quad h \leqslant h_0,
$$

then there exist positive constants $\lambda = \lambda(d, \sigma, \kappa(\cdot), L_0)$, $C_1 = C_1(d, \sigma, \kappa(\cdot), L_0, M)$ and $C_2 = C_2(d, \sigma, \kappa(\cdot), L_0)$ such that the discrete RB–IPS (2.22) satisfies

$$
\mathcal{W}_1(\nu\tilde{\mathcal{Q}}_n^h, \bar{\pi}^{\otimes N}) \leqslant C_1\sqrt{h} + C_1 e^{-\lambda nh} + \frac{C_2}{\sqrt{N}}, \quad \forall n \geqslant 0. \tag{4.113}
$$

If we define the empirical distribution of the N-particle system $\{Y_t^i\}_{i=1}^N$ as

$$
\tilde{\mu}_n^h(x) = \frac{1}{N}\sum_{i=1}^N \delta(x - Y_t^i) \in \mathcal{P}(\mathbb{R}^d), \quad x \in \mathbb{R}^d, \tag{4.114}
$$

then $\tilde{\mu}_n^h(x)$ is a random probability distribution in $\mathbb{R}^d$. According to Proposition 2.14 of Hauray and Mischler (2014), we can also write (4.113) as

$$
\mathbb{E}\big[\mathcal{W}_1(\tilde{\mu}_n^h, \bar{\pi})\big] \leqslant C_1\sqrt{h} + C_1 e^{-\lambda nh} + \frac{C_2}{\sqrt{N}}, \quad \forall n \geqslant 0. \tag{4.115}
$$

Theorem 4.9 characterizes the long-time sampling error of the discrete RB–IPS (2.22) in approximating the invariant distribution $\bar{\pi} \in \mathcal{P}(\mathbb{R}^d)$. The error terms in (4.113) consist of three components: (1) $C_1\sqrt{h}$, error due to time discretization and random batch divisions; (2) $C_1 e^{-\lambda nh}$, error from the exponential convergence of the discrete RB–IPS; (3) $C_2/\sqrt{N}$, error from the uniform-in-time propagation of chaos. In particular, the constants C_1, C_2 and λ do

not depend on the number of particles N, the batch size p and the step size h, showing that the reliability of the discrete RB–IPS to sample the distribution $\bar{\pi}$.

4.5 Numerical tests

We perform numerical tests to evaluate the convergence and accuracy of the discrete RB–IPS (2.22) in sampling the IPS. Assume the drift force $b(x) = -\nabla U(x)$ and the interaction force $K(x) = -\nabla V(x)$ with $\sigma = \sqrt{2}$, so that the invariant distribution $\pi(\boldsymbol{x})$ is given by (2.20):

$$\pi(\boldsymbol{x}) \propto \exp\left(-\sum_{i=1}^{N} U(x^i) - \frac{1}{N-1}\sum_{1\leqslant i<j\leqslant N} V(x^i - x^j)\right), \quad \boldsymbol{x} \in \mathbb{R}^{dN}.$$

Example: non-convex potential with aggregation interaction Let the external potential $U(x)$ and the interaction potential $V(x)$ be given by

$$U(x) = \frac{1}{2}x^2 + 3.5e^{-(x-0.3)^2}, \quad V(x) = -e^{-x^2}, \tag{4.116}$$

and it is evident from Figure 4.2 that $U(x)$ is a non-convex double-well potential, while $V(x)$ induces an aggregation effect on the interacting particles.

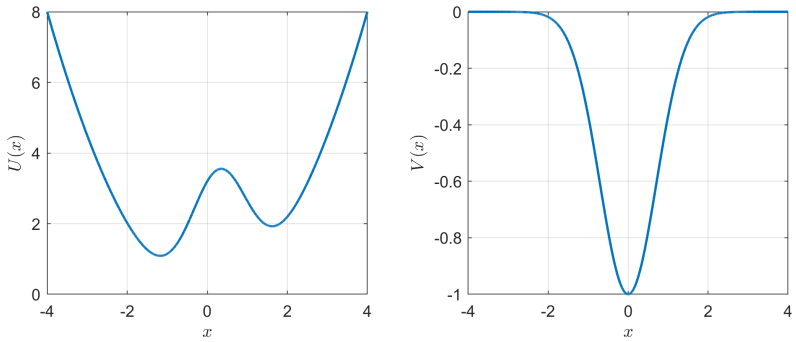


Figure 4.2 Graphs of the potential functions $U(x)$ and $V(x)$ in (4.116).

Next, we fix the number of particles to $N = 64$ and simulate the IPS (2.17) (without random batch approximations) with a step size of $h = \frac{1}{32}$ and a simulation time of $T = 10^7$. The sampling trajectory of the first particle, x_t^1, is shown in Figure 4.3, and the marginal distribution of the target distribution $\pi(\boldsymbol{x})$ is shown in Figure 4.4.

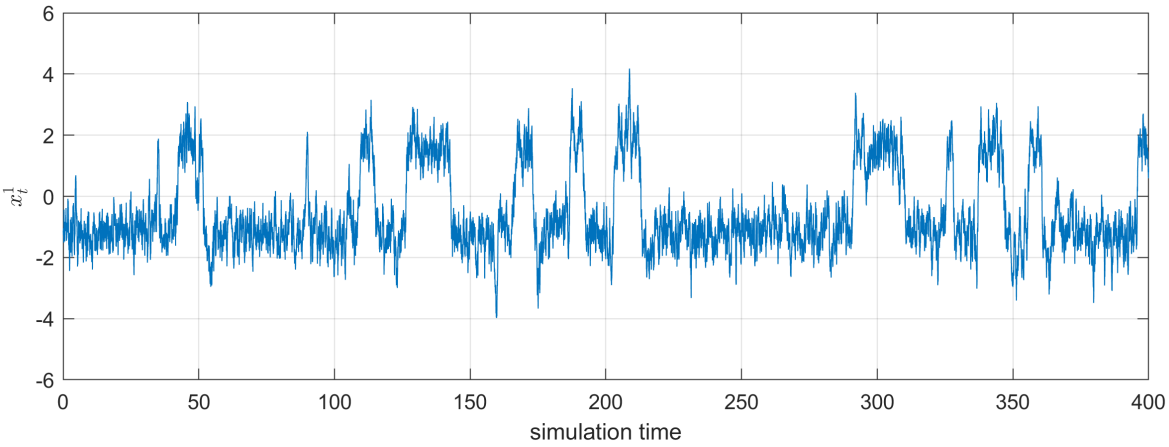


Figure 4.3 Sampling trajectory of x_t^1 from the IPS (2.17) over a finite time.

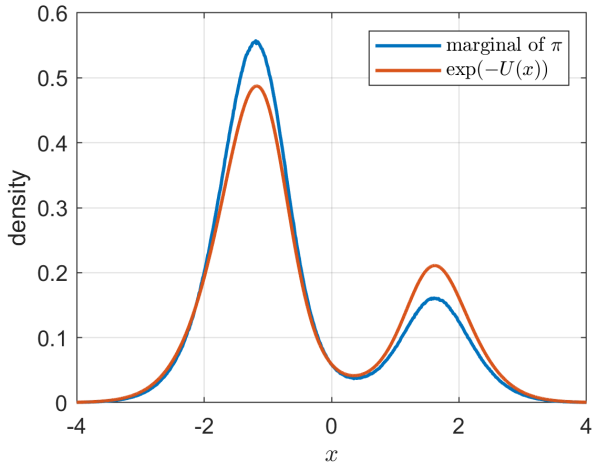


Figure 4.4 Marginal of $\pi(\boldsymbol{x})$ in blue, and the distribution $\exp(-U(x))$ in red.

From Figure 4.3, we observe that the IPS (2.17) is ergodic, with the particle x_t^1 alternately switching between the two basins of the potential function $U(x)$. In Figure 4.4, we see that the marginal distribution of $\pi(\boldsymbol{x})$ slightly differs from the distribution $\exp(-U(x))$. The presence of the interaction potential causes the target distribution to be more concentrated around the global minimum of $U(x)$.

In the following, we fix the simulation time at $T = 10^5$ and test the convergence of the discrete RB–IPS (2.22) with different choices of step size h. We also test the convergence of the discrete IPS, which can be viewed as a special case of the discrete RB–IPS with batch size $p = N$. The sampling errors of the algorithms are computed from the quadratic mean of the errors in calculating the averages of the following five test functions:

$$\left\{e^{-2x^2}, e^{-2(x-0.4)^2}, e^{-2(x+0.4)^2}, e^{-2(x-0.8)^2}, e^{-2(x+0.8)^2}\right\}.$$

We compute the sampling errors for these algorithms with step sizes $h = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}$ and batch sizes $p = 2, 4$. The results are shown in Figure 4.5 and Table 4.2.

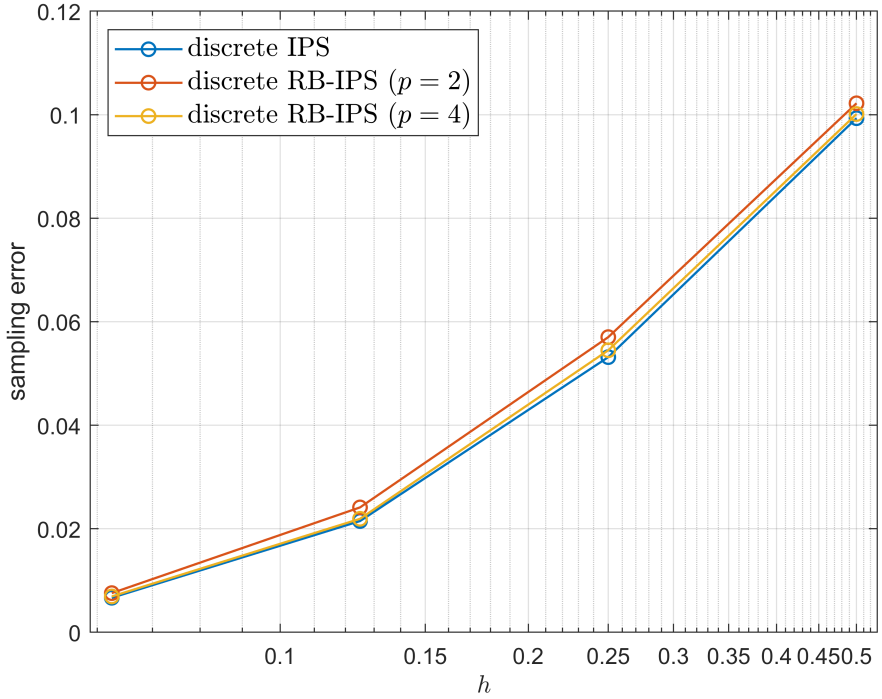


Figure 4.5 Sampling error of the discrete IPS and the discrete RB–IPS for various step sizes $h = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}$. The batch sizes p are set to 2 and 4.

step size h	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$
discrete IPS	0.0993	0.0532	0.0215	0.0067
discrete RB–IPS ($p = 2$)	0.1022	0.0570	0.0241	0.0076
discrete RB–IPS ($p = 4$)	0.1001	0.0545	0.0219	0.0069

Table 4.2 Sampling error of the discrete IPS and the discrete RB–IPS for various step sizes $h = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}$. The batch sizes p are set to 2 and 4.

From Figure 4.5 and Table 4.2, we observe that the discrete RB–IPS (2.22) exhibits a convergence order of 1 with respect to the step size h, which exceeds the theoretical result in Theorem 4.5. Additionally, since the interaction force is small and the discretization error dominates the error bound, the sampling results remain highly accurate even when $p = 2$.

4.6 Brief summary

The RB–IPS (2.21) serves as an efficient sampling method for interacting particle systems, with the discrete RB–IPS (2.22) being its time discretization. In this work, we rigorously

established the long-time behavior of the RB–IPS (Theorem 4.1) and analyzed the long-time error of the discrete RB–IPS (Theorem 4.5). We demonstrated that the convergence order with respect to the step size is $O(\sqrt{h})$ in general. Moreover, when the interaction force is moderately large, the convergence rate is independent of the number of particles N, the batch size p, and the step size h. By combining error analysis with the propagation of chaos, we further showed that the discrete RB–IPS can reliably sample the invariant distribution of the MVP (2.18) (Theorem 4.9).

Chapter 5 Conclusion

This thesis has explored fundamental aspects of ergodicity and long-time behavior in high-dimensional stochastic processes, focusing on two critical sampling methodologies: Path Integral Molecular Dynamics (PIMD) and the Random Batch Method (RBM). By addressing theoretical challenges and providing practical insights, the work contributes significantly to the field of computational mathematics and physics.

The main contributions of this thesis include a rigorous establishment of the uniform-in-N ergodicity of the underdamped Langevin dynamics in Path Integral Molecular Dynamics (PIMD), ensuring that the convergence rate remains independent of the number of beads. This advancement significantly enhances the theoretical applicability of PIMD. Additionally, the thesis provides a systematic study of the ergodicity properties of the Random Batch Interacting Particle System (RB–IPS), along with an in-depth analysis of the long-time error of the discrete RB–IPS. These results offer a solid theoretical foundation for related sampling methods.

Building upon these findings, several potential directions for future research are proposed. A rigorous error analysis of stochastic gradient-based sampling methods is an ongoing effort, aiming to further refine the theoretical framework. Moreover, leveraging machine learning techniques and extending generative models could lead to the development of new and efficient sampling algorithms for complex distribution sampling. Another promising direction involves designing specialized sampling algorithms tailored to practical applications, such as high-dimensional physical and data science problems, to enhance their performance in real-world scenarios.

In conclusion, this thesis has provided both theoretical insights and practical tools for high-dimensional sampling. The uniform-in-N properties and rigorous error analyses presented here not only deepen the understanding of PIMD and RBM but also establish a foundation for future innovations in computational mathematics and physics.

References

Althorpe, S. C. (2024). Path Integral Simulations of Condensed-Phase Vibrational Spectroscopy. *Annual Review of Physical Chemistry*, 75(1):397–420.

Ames, W. F. (2014). *Numerical Methods for Partial Differential Equations*. Academic Press.

Bakry, D., Gentil, I., Ledoux, M., et al. (2014). *Analysis and Geometry of Markov Diffusion Operators*, volume 103. Springer.

Baudoin, F., Gordina, M., and Herzog, D. P. (2021). Gamma Calculus Beyond Villani and Explicit Convergence Estimates for Langevin Dynamics with Singular Potentials. *Archive for Rational Mechanics and Analysis*, 241(2):765–804.

Betancourt, M. (2017). A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434*.

Bou-Rabee, N. and Eberle, A. (2021). Two-Scale Coupling for Preconditioned Hamiltonian Monte Carlo in Infinite Dimensions. *Stochastics and Partial Differential Equations: Analysis and Computations*, 9:207–242.

Bou-Rabee, N. and Eberle, A. (2022). Couplings for Andersen Dynamics. *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, 58(2):916–944.

Bou-Rabee, N. and Eberle, A. (2023). Mixing Time Guarantees for Unadjusted Hamiltonian Monte Carlo. *Bernoulli*, 29(1):75–104.

Bou-Rabee, N., Eberle, A., and Zimmer, R. (2020). Coupling and Convergence for Hamiltonian Monte Carlo. *The Annals of Applied Probability*, 30(3):1209–1250.

Cao, J. and Martyna, G. J. (1996). Adiabatic Path Integral Molecular Dynamics Methods. II. Algorithms. *The Journal of Chemical Physics*, 104(5):2028–2035.

Carrillo, J. A., Choi, Y.-P., Hauray, M., and Salem, S. (2018). Mean-Field Limit for Collective Behavior Models with Sharp Sensitivity Regions. *Journal of the European Mathematical Society*, 21(1):121–161.

Carrillo, J. A., Jin, S., and Tang, Y. (2021). Random Batch Particle Methods for the Homogeneous Landau Equation. *arXiv preprint arXiv:2110.06430*.

Cañizo, J. A. and Mischler, S. (2023). Harris-Type Results on Geometric and Subgeometric Convergence to Equilibrium for Stochastic Semigroups. *Journal of Functional Analysis*, 284(7):109830.

Ceriotti, M., Parrinello, M., Markland, T. E., and Manolopoulos, D. E. (2010). Efficient Stochastic Thermostatting of Path Integral Molecular Dynamics. *The Journal of Chemical Physics*, 133(12).

Chaintron, L.-P. and Diez, A. (2021). Propagation of Chaos: A Review of Models, Methods and Applications. II. Applications. *arXiv preprint arXiv:2106.14812*.

Chaintron, L.-P. and Diez, A. (2022). Propagation of Chaos: A Review of Models, Methods and Applications. I. Models and Methods. *arXiv preprint arXiv:2203.00446*.

Chak, M. and Monmarché, P. (2023). Reflection Coupling for Unadjusted Generalized Hamiltonian Monte Carlo in the Nonconvex Stochastic Gradient Case. *arXiv preprint arXiv:2310.18774*.

Chandler, D. and Wolynes, P. G. (1981). Exploiting the Isomorphism Between Quantum Theory and Classical Statistical Mechanics of Polyatomic Fluids. *The Journal of Chemical Physics*, 74(7):4078–4095.

Chen, M.-F. and Li, S.-F. (1989). Coupling Methods for Multidimensional Diffusion Processes. *The Annals of Probability*, pages 151–177.

Chen, T., Fox, E., and Guestrin, C. (2014). Stochastic Gradient Hamiltonian Monte Carlo. In *International conference on machine learning*, pages 1683–1691. PMLR.

Clary, D. C. (1998). Quantum Theory of Chemical Reaction Dynamics. *Science*, 279(5358):1879–1882.

Croitoru, F.-A., Hondru, V., Ionescu, R. T., and Shah, M. (2023). Diffusion Models in Vision: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10850–10869.

Durlak, P., Latajka, Z., and Berski, S. (2009). A Car–Parrinello and Path Integral Molecular Dynamics Study of the Intramolecular Lithium Bond in the Lithium 2-Pyridyl-N-Oxide Acetate. *The Journal of Chemical Physics*, 131(2).

Durmus, A., Eberle, A., Guillin, A., and Zimmer, R. (2020). An Elementary Approach to Uniform in Time Propagation of Chaos. *Proceedings of the American Mathematical Society*, 148(12):5387–5398.

Durmus, A. and Moulines, E. (2017). Nonasymptotic Convergence Analysis for the Unadjusted Langevin Algorithm. *The Annals of Applied Probability*, 27(3):1551–1587.

E, W. and Li, D. (2008). The Andersen Thermostat in Molecular Dynamics. *Communications on Pure and Applied Mathematics*, 61(1):96–136.

E, W., Li, T., and Vanden-Eijnden, E. (2021). *Applied Stochastic Analysis*, volume 199. American Mathematical Society.

Eberle, A. (2011). Reflection Coupling and Wasserstein Contractivity Without Convexity. *Comptes Rendus Mathématique*, 349(19-20):1101–1104.

Eberle, A. (2016). Reflection Couplings and Contraction Rates for Diffusions. *Probability Theory and Related Fields*, 166:851–886.

Eberle, A., Guillin, A., and Zimmer, R. (2019). Couplings and Quantitative Contraction Rates for Langevin Dynamics. *The Annals of Probability*, 47(3):1982–2010.

Evans, D. J. and Holian, B. L. (1985). The Nosé–Hoover Thermostat. *The Journal of Chemical Physics*, 83(8):4069–4074.

Feynman, R. P. (2018). *Statistical Mechanics: A Set of Lectures*. CRC Press.

Frenkel, D. and Smit, B. (2023). *Understanding Molecular Simulation: From Algorithms to Applications*. Elsevier.

Glowinski, R., Osher, S. J., and Yin, W. (2017). *Splitting Methods in Communication, Imaging, Science, and Engineering*. Springer.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative Adversarial Networks. *Communications of the ACM*, 63(11):139–144.

Gouraud, N., Le Bris, P., Majka, A., and Monmarché, P. (2022). HMC and Underdamped Langevin United in the Unadjusted Convex Smooth Case. *arXiv preprint arXiv:2202.00977*.

Guillin, A. and Monmarché, P. (2021). Uniform Long-Time and Propagation of Chaos Estimates for Mean Field Kinetic Particles in Non-Convex Landscapes. *Journal of Statistical Physics*, 185:1–20.

Ha, S.-Y., Jin, S., Kim, D., and Ko, D. (2021). Convergence Toward Equilibrium of the First-Order Consensus Model with Random Batch Interactions. *Journal of Differential Equations*, 302:585–616.

Haile, J. M. (1992). *Molecular Dynamics Simulation: Elementary Methods*. John Wiley & Sons.

Hairer, M. and Mattingly, J. C. (2011). Yet Another Look at Harris' Ergodic Theorem for Markov Chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI: Centro Stefano Franscini, Ascona, May 2008*, pages 109–117. Springer.

Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1):97–109.

Hauray, M. and Mischler, S. (2014). On Kac's Chaos and Related Problems. *Journal of Functional Analysis*, 266(10):6055–6157.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. *Advances in neural information processing systems*, 33:6840–6851.

Householder, A. S. (1958). Unitary Triangularization of a Nonsymmetric Matrix. *Journal of the ACM (JACM)*, 5(4):339–342.

Huang, Z., Jin, S., and Li, L. (2024). Mean Field Error Estimate of the Random Batch Method for Large Interacting Particle Systems. *arXiv preprint arXiv:2403.08336*.

Jin, S. and Li, L. (2022). On the Mean Field Limit of the Random Batch Method for Interacting Particle Systems. *Science China Mathematics*, pages 1–34.

Jin, S., Li, L., and Liu, J.-G. (2020). Random Batch Methods (RBM) for Interacting Particle Systems. *Journal of Computational Physics*, 400:108877.

Jin, S., Li, L., Xu, Z., and Zhao, Y. (2021). A Random Batch Ewald Method for Particle Systems with Coulomb Interactions. *SIAM Journal on Scientific Computing*, 43(4):B937–B960.

Kong, A., Liu, J. S., and Wong, W. H. (1994). Sequential Imputations and Bayesian Missing Data Problems. *Journal of the American statistical association*, 89(425):278–288.

Korol, R., Bou-Rabee, N., and Miller, T. F. (2019). Cayley Modification for Strongly Stable Path-Integral and Ring-Polymer Molecular Dynamics. *The Journal of Chemical Physics*, 151(12).

Korol, R., Rosa-Raíces, J. L., Bou-Rabee, N., and Miller, T. F. (2020). Dimension-Free Path-Integral Molecular Dynamics Without Preconditioning. *The Journal of Chemical Physics*, 152(10).

Kreyszig, E. (1991). *Introductory Functional Analysis with Applications*, volume 17. John Wiley & Sons.

Leimkuhler, B. and Matthews, C. (2015). Molecular Dynamics. *Interdisciplinary Applied Mathematics*, 39(1).

Leimkuhler, B., Paulin, D., and Whalley, P. A. (2024). Contraction Rate Estimates of Stochastic Gradient Kinetic Langevin Integrators. *ESAIM: Mathematical Modelling and Numerical Analysis*.

Li, L., Liu, J.-G., and Tang, Y. (2020a). Some Random Batch Particle Methods for the Poisson-Nernst-Planck and Poisson-Boltzmann Equations. *arXiv preprint arXiv:2004.05614*.

Li, L., Liu, J.-G., and Wang, Y. (2023). Geometric Ergodicity of SGLD via Reflection Coupling. *arXiv preprint arXiv:2301.06769*.

Li, L., Xu, Z., and Zhao, Y. (2020b). A Random-Batch Monte Carlo Method for Many-Body Systems with Singular Kernels. *SIAM Journal on Scientific Computing*, 42(3):A1486–A1509.

Liang, J., Xu, Z., and Zhao, Y. (2021). Random-Batch List Algorithm for Short-Range Molecular Dynamics Simulations. *The Journal of Chemical Physics*, 155(4).

Liang, J., Xu, Z., and Zhou, Q. (2023). Random Batch Sum-of-Gaussians Method for Molecular Dynamics Simulations of Particle Systems. *SIAM Journal on Scientific Computing*, 45(5):B591–B617.

Lindvall, T. and Rogers, L. C. G. (1996). On Coupling of Random Walks and Renewal Processes. *Journal of Applied Probability*, 33(1):122–126.

Liu, J., Li, D., and Liu, X. (2016). A Simple and Accurate Algorithm for Path Integral Molecular Dynamics with the Langevin Thermostat. *The Journal of Chemical Physics*, 145(2).

Lu, J., Lu, Y., and Zhou, Z. (2020). Continuum Limit and Preconditioned Langevin Sampling of the Path Integral Molecular Dynamics. *Journal of Computational Physics*, 423:109788.

Lu, J. and Wang, L. (2020). On Explicit L^2-Convergence Rate Estimate for Piecewise Deterministic Markov Processes. *arXiv preprint arXiv:2007.14927*.

Lévy, P. (1940). Sur Certains Processus Stochastiques Homogènes. *Compositio Mathematica*, 7:283–339.

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.

Marx, D. and Parrinello, M. (1996). Ab Initio Path Integral Molecular Dynamics: Basic Ideas. *The Journal of Chemical Physics*, 104(11):4077–4082.

Mattingly, J. C., Stuart, A. M., and Higham, D. J. (2002). Ergodicity for SDEs and Approximations: Locally Lipschitz Vector Fields and Degenerate Noise. *Stochastic Processes and Their Applications*, 101(2):185–232.

Mattingly, J. C., Stuart, A. M., and Tretyakov, M. V. (2010). Convergence of Numerical Time-Averaging and Stationary Measures via Poisson Equations. *SIAM Journal on Numerical Analysis*, 48(2):552–577.

McKean, H. P. (1967). Propagation of Chaos for a Class of Non-Linear Parabolic Equations. *Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic University, 1967)*, pages 41–57.

Metropolis, N. and Ulam, S. (1949). The Monte Carlo Method. *Journal of the American statistical association*, 44(247):335–341.

Miller, W. H. (1993). Beyond Transition-State Theory: A Rigorous Quantum Theory of Chemical Reaction Rates. *Accounts of Chemical Research*, 26(4):174–181.

Monmarché, P. (2018). Hypocoercivity in Metastable Settings and Kinetic Simulated Annealing. *Probability Theory and Related Fields*, 172(3):1215–1248.

Monmarché, P. (2019). Generalized Γ Calculus and Application to Interacting Particles on a Graph. *Potential Analysis*, 50(3):439–466.

Neal, R. M. (2001). Annealed Importance Sampling. *Statistics and computing*, 11:125–139.

Neal, R. M. et al. (2011). MCMC Using Hamiltonian Dynamics. *Handbook of Markov chain Monte Carlo*, 2(11):2.

Nicholson, D. R. (1983). *Introduction to Plasma Theory*, volume 1. Wiley.

Pareschi, L. and Zanella, M. (2024). Reduced Variance Random Batch Methods for Nonlocal PDEs. *Acta Applicandae Mathematicae*, 191(1):1–23.

Sanz-Serna, J. M. and Palencia, C. (1985). A General Equivalence Theorem in the Theory of Discretization Methods. *Mathematics of Computation*, 45(171):143–152.

Schuh, K. and Souttar, I. (2024). Conditions for Uniform in Time Convergence: Applications to Averaging, Numerical Discretisations and Mean-Field Systems. *arXiv preprint arXiv:2412.05239*.

Schuh, K. and Whalley, P. A. (2024). Convergence of Kinetic Langevin Samplers for Non-Convex Potentials. *arXiv preprint arXiv:2405.09992*.

Schweizer, K. S., Stratt, R. M., Chandler, D., and Wolynes, P. G. (1981). Convenient and Accurate Discretized Path Integral Methods for Equilibrium Quantum Mechanical Calculations. *The Journal of Chemical Physics*, 75(3):1347–1364.

Shardlow, T. and Stuart, A. M. (2000). A Perturbation Theory for Ergodic Markov Chains and Application to Numerical Approximations. *SIAM Journal on Numerical Analysis*, 37(4):1120–1137.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and Ganguli, S. (2015). Deep Unsupervised Learning Using Nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. PMLR.

Sondhi, S. L., Girvin, S. M., Carini, J. P., and Shahar, D. (1997). Continuous Quantum Phase Transitions. *Reviews of Modern Physics*, 69(1):315.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. (2020). Score-Based Generative Modeling Through Stochastic Differential Equations. *arXiv preprint arXiv:2011.13456*.

Sznitman, A.-S. (1991). Topics in Propagation of Chaos. *Ecole d'Eté de Probabilités de Saint-Flour XIX—1989*, 1464:165–251.

Tokdar, S. T. and Kass, R. E. (2010). Importance Sampling: A Review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1):54–60.

Tuckerman, M. E. (2023). *Statistical Mechanics: Theory and Molecular Simulation*. Oxford University Press.

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

Villani, C. (2009b). *Optimal Transport: Old and New*, volume 338. Springer.

Voth, G. A. (1993). Feynman Path Integral Formulation of Quantum Mechanical Transition-State Theory. *The Journal of Physical Chemistry*, 97(32):8365–8377.

Wang, F. (2006). *Functional Inequalities, Markov Semigroups, and Spectral Theory*. Elsevier.

Welling, M. and Teh, Y. W. (2011). Bayesian Learning via Stochastic Gradient Langevin Dynamics. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 681–688. Citeseer.

Willatt, M. J. (2017). *Matsubara Dynamics and Its Practical Implementation*. PhD thesis, University of Cambridge.

Xu, Z., Zhao, Y., and Zhou, Q. (2024). Variance-Reduced Random Batch Langevin Dynamics. *arXiv preprint arXiv:2411.01762*.

Publications

Ye, X., and Zhou, Z. (2021). Efficient sampling of thermal averages of interacting quantum particle systems with random batches. *The Journal of Chemical Physics*, 154(20).

Ye, X., and Zhou, Z. (2023). Dimension-free ergodicity of path integral molecular dynamics. *arXiv preprint arXiv:2307.06510*.

Jin, S., Li, L., Ye, X., and Zhou, Z. (2023). Ergodicity and long-time behavior of the Random Batch Method for interacting particle systems. *Mathematical Models and Methods in Applied Sciences*, 33(01), 67-102.

Ye, X., and Zhou, Z. (2024). Error analysis of time-discrete random batch method for interacting particle systems and associated mean-field limits. *IMA Journal of Numerical Analysis*, 44(3), 1660-1698.