Two-scale coupling for preconditioned Hamiltonian Monte Carlo in infinite dimensions

Nawaf Bou-Rabee[1] · Andreas Eberle[2]

Abstract

We derive non-asymptotic quantitative bounds for convergence to equilibrium of the exact preconditioned Hamiltonian Monte Carlo algorithm (pHMC) on a Hilbert space. As a consequence, explicit and dimension-free bounds for pHMC applied to high-dimensional distributions arising in transition path sampling and path integral molecular dynamics are given. Global convexity of the underlying potential energies is not required. Our results are based on a two-scale coupling which is contractive in a carefully designed distance.

✉ Nawaf Bou-Rabee
   nawaf.bourabee@rutgers.edu

   Andreas Eberle
   eberle@uni-bonn.de

[1]   Department of Mathematical Sciences, Rutgers University Camden, 311 N 5th Street, Camden, NJ 08102, USA

[2]   Institut für Angewandte Mathematik, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany

1 Introduction

Hamiltonian or Hybrid Monte Carlo (HMC) methods are a class of Markov Chain Monte Carlo (MCMC) methods originating in statistical physics [20] which have become increasingly popular in various application areas [11,48,53,60,64]. Their success is in particular due to empirically observed convergence acceleration compared to more traditional, random-walk based methods. The basic idea in HMC is to define an MCMC method with the help of an artificial Hamiltonian dynamics whose *only* purpose is to accelerate convergence to equilibrium. This Hamiltonian dynamics is designed to leave invariant a product of the target measure and a fictitious Gaussian measure in an artifical velocity variable. First rigorous theoretical results supporting the empirical evidence have only been established recently. In particular, geometric ergodicity has been verified in [10,22,49], and quantitative convergence bounds have been derived in the strongly convex case in [52], and under more general assumptions in [9], both by applying coupling methods.

Since many applications are high dimensional, a key issue is to understand the dependence of the convergence bounds on the dimension. Here, we study the problem of dimension dependence for a special class of models that is relevant for several important applications including Path Integral Molecular Dynamics (PIMD) [14–17,31,32,43,44,50,57,58], Transition Path Sampling (TPS) [6,59,63,65], and Bayesian inverse problems [7,18,41,68]. For the class of models we consider, a corresponding HMC Markov chain relying on a preconditioned Hamiltonian dynamics can be defined directly on the infinite dimensional state space [3]. This suggests that one might hope for dimension-free convergence bounds for the corresponding Markov chains on finite-dimensional discretizations of the state space. Corresponding dimension-free convergence rates to equilibrium have been established for the preconditioned Crank-Nicholson (pCN) algorithm [36] and for the Metropolis-adjusted Langevin algorithm (MALA) [24], but a corresponding result for HMC is not known so far.

The goal of this paper is to fill this gap. To this end we extend the coupling approach developed for HMC in the finite dimensional case in [9], and combine it with a two-scale coupling approach for stochastic dynamics on infinite dimensional Hilbert spaces that originates in [33,34,54,55] and has been further developed in [71]. The splitting into "low modes" and "high modes" in the two-scale coupling can be traced back to contraction results for the stochastic Navier-Stokes equations [55], and analogous results in the deterministic setting [28]; see [54] for a detailed review.

Our object of study is the *exact preconditioned HMC algorithm* (pHMC) with fixed durations on a Hilbert space, i.e., the (preconditioned) Hamiltonian dynamics is exactly integrated (or, in practical terms, the integration is carried out with very small step sizes). Here, preconditioning corresponds to an appropriate choice of the kinetic energy which involves picking the mass operator equal to the stiffness operator (or inverse covariance) associated to the Gaussian reference measure of the target probability measure. This choice of kinetic energy ensures that the corresponding pHMC algorithm is more amenable to numerical approximation and Metropolis-adjustment than HMC without preconditioning [3,11].

We prove that the transition kernel of the Markov chain induced by the pHMC algorithm is contracting in a suitable Wasserstein/Kantorovich metric with a rate that

depends transparently on the duration of the Hamiltonian flow, the eigenvalues of the covariance operator of the Gaussian reference measure, and the regularity of the preconditioned Hamiltonian dynamics. The results are given in a more general setting that includes pHMC as a special case, and also covers other types of dynamics and preconditioning strategies. As a consequence of our general results, we derive dimension-free bounds for pHMC applied to finite-dimensional approximations arising in TPS and PIMD.

Before stating our results in detail, we conclude with a brief outlook. The results below apply only to pHMC with exact integration of the Hamiltonian dynamics. In practice, the Hamiltonian dynamics is numerically approximated, to obtain *numerical* versions of pHMC that are implementable on a computer. The time integrator of choice for pHMC is the symmetric splitting integrator introduced in [3]. Unlike other splittings for the Hamiltionian dynamics, this approximation has an acceptance rate that is uniform with respect to the spatial step size associated with the discretization of the Hilbert space [11, § 8]. Time discretization creates a bias in the invariant measure that can be avoided by a Metropolis adjustment [11,70]. We would expect that for unadjusted numerical HMC based on the integrator proposed in [3], similar contraction results as stated below hold if the time step size is chosen sufficiently small (but independently of the dimension). Under additional regularity assumptions, one could also hope for dimension free bounds for the Metropolis adjusted version. First steps in this direction are carried out in [9, § 2.5.4] in the finite dimensional case, and in [62, § 4] in a strongly convex infinite dimensional case, but a full study in the general case would be lengthy and go beyond the scope of the current work.

Alternatively to preconditioning, it is also possible (but more tricky) to implement non-preconditioned HMC, which corresponds to injecting white noise in the velocity variable. In this case, the corresponding Hamiltonian dynamics is highly oscillatory in high modes [61]. Therefore, convergence bounds for exact HMC without preconditioning on an infinite dimensional Hilbert space can be expected to hold only if the durations are randomized [60], and in numerical implementations, strongly stable integrators [43,44] have to be used in order to be able to choose the step size independently of the dimension. Furthermore, scaling limit results show that for Metropolis adjusted HMC applied to i.i.d. product measures on high dimensional state spaces, the step size has to be chosen of order $O(d^{-1/4})$ to avoid degeneracy of the acceptance probabilities [2,30,42].

We now state our main results in Sect. 2, and consider applications to TPS and PIMD in Sect. 3. The remaining sections contain the proofs of all results.

2 Main results

Let $\mathcal{H}$ be a separable and real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $|\cdot|$. Let $\mathcal{C} : \mathcal{H} \to \mathcal{H}$ be a positive compact symmetric linear operator. By the spectral theorem, the eigenfunctions $\{e_i\}_{i \in \mathbb{N}}$ of $\mathcal{C}$ form a complete orthonormal basis of $\mathcal{H}$ with corresponding eigenvalues $\{\lambda_i\}_{i \in \mathbb{N}}$ which we arrange in descending order, i.e., $\lambda_1 \geq \lambda_2 \geq \cdots$. The positivity condition means that $\lambda_j > 0$ for all $j \in \mathbb{N}$, and by compactness, if $\dim(\mathcal{H}) = \infty$ then $\lim_{j \to \infty} \lambda_j = 0$. Any function $x \in \mathcal{H}$ can be

represented in spectral coordinates by the expansion

$$x = \sum x_j e_j \quad \text{where } x_j := \langle x, e_j \rangle. \tag{2.1}$$

Moreover, for all $s \in \mathbb{R}$, the operator $\mathcal{C}^s$ is defined via the spectral decomposition of $\mathcal{C}$. We introduce the family of inner products and norms given by

$$\langle x, y \rangle_s := \langle x, \mathcal{C}^{-s} y \rangle = \langle \mathcal{C}^{-s/2} x, \mathcal{C}^{-s/2} y \rangle, \quad |x|_s := \langle x, x \rangle_s^{1/2} \tag{2.2}$$

for $x, y \in \mathcal{H}^s$. Here for $s \geq 0$, $\mathcal{H}^s$ denotes the Hilbert space consisting of all $x \in \mathcal{H}$ with $|x|_s < \infty$, whereas for $s < 0$, $\mathcal{H}^s$ is the completion of $\mathcal{H}$ w.r.t. $|x|_s$. Note that $\mathcal{H} = \mathcal{H}^0$, and for $s > 0$, $\mathcal{H}^s \subset \mathcal{H} \subset \mathcal{H}^{-s}$. Furthermore, the linear operator $\mathcal{C}$ restricts or extends (depending on whether $s > 0$ or $s < 0$) to a linear isometry from $\mathcal{H}^s$ to $\mathcal{H}^{s+2}$ which will again be denoted by $\mathcal{C}$. This setup is consistent with the framework for infinite-dimensional Bayesian inverse problems [3,18,68]. Here, typically, $s \in (0, 1)$.

We will now introduce the pHMC method for approximate sampling from a probability measure μ that has a density w.r.t. a Gaussian measure μ_0 on one of the Hilbert spaces $\mathcal{H}^s$. Afterwards, in Sect. 2.2, we will introduce a more general family of Markov chains on Hilbert spaces that includes the Markov chain associated to pHMC as a special case. In Sect. 2.3, we introduce a new coupling for these Markov chains that combines ideas from [9,71]. Then in Sects. 2.4 and 2.5, we state our main contraction result for these couplings, and derive quantitative error bounds.

2.1 Exact preconditioned Hamiltonian Monte Carlo

Let $\mu_0 = \mathcal{N}(0, \mathcal{C})$ denote the centered Gaussian measure whose covariance operator w.r.t. the inner product $\langle \cdot, \cdot \rangle$ is $\mathcal{C}$ [5]. If $\mathcal{C}$ is trace class then μ_0 is supported on $\mathcal{H}$. More generally, we fix $s \in (-\infty, 1)$ and assume that μ_0 is supported on the corresponding Hilbert space $\mathcal{H}^s$. This is ensured by the following assumption:

Assumption 2.1 The operator $\mathcal{C}^{1-s}$ is trace class, i.e.,

$$\text{trace}(\mathcal{C}^{1-s}) = \sum_{j=1}^{\infty} \lambda_j^{1-s} < \infty.$$

A realization ξ from μ_0 can be generated using the expansion

$$\xi = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \rho_j e_j, \quad \{\rho_i\} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1).$$

For $\xi \sim \mu_0$, Assumption 2.1 implies $\mathbb{E} |\xi|_s^2 = \text{trace}(\mathcal{C}^{1-s}) < \infty$, and thus, ξ is indeed a Gaussian random variable on $\mathcal{H}^s$.

Remark 2.2 To avoid confusion, we stress that the covariance operator of a Gaussian measure is a non-intrinsic object that depends on the choice of an inner product. In

particular, the covariance operator of μ_0 w.r.t. the $\mathcal{H}^s$ inner product is $\mathcal{C}^{1-s}$. Nonetheless, in what follows, we always define the covariance operator with respect to the $\mathcal{H}$ inner product, and in this sense, the measure μ_0 has covariance operator $\mathcal{C}$.

Exact preconditioned Hamiltonian Monte Carlo (pHMC) is an MCMC method for approximate sampling from probability distributions on a Hilbert space that have the general form

$$\mu(dx) \propto \exp(-U(x))\mu_0(dx), \quad \mu_0 = \mathcal{N}(0, \mathcal{C}), \tag{2.3}$$

where U is a function on a Hilbert space on which the Gaussian measure μ_0 is supported. The pHMC method generates a Markov chain on this Hilbert space with transition step

$$x \mapsto X'(x) \quad \text{where} \quad X'(x) = q_T(x, \xi). \tag{2.4}$$

Here $\xi \sim \mathcal{N}(0, \mathcal{C})$, and the duration $T : \Omega \to \mathbb{R}_+$ is in general an independent random variable with a given distribution ν (e.g. $\nu = \delta_r$ or $\nu = \text{Exp}(\lambda^{-1})$). We will only consider the case where $T \in (0, \infty)$ is a given deterministic constant. Moreover,

$$\phi_t(x, v) = (q_t(x, v), v_t(x, v)) \qquad (t \in [0, \infty))$$

is the exact flow of the Hilbert space valued ODE given by

$$\frac{d}{dt}q_t = v_t, \quad \frac{d}{dt}v_t = -q_t - \mathcal{C}\mathcal{D}U(q_t), \quad (q_0(x, v), v_0(x, v)) = (x, v). \tag{2.5}$$

Formally, (2.5) is a preconditioned Hamiltonian dynamics for the Hamiltonian

$$H(x, v) = U(x) + \langle x, \mathcal{C}^{-1}x \rangle/2 + \langle v, \mathcal{C}^{-1}v \rangle/2,$$

where the covariance operator $\mathcal{C}$ is used for preconditioning. A key property of (2.5) is that it leaves invariant the probability measure

$$\hat{\mu}(dx\,dv) \propto \exp(-U(x))\,\mathcal{N}(0, \mathcal{C})(dx)\,\mathcal{N}(0, \mathcal{C})(dv), \tag{2.6}$$

on phase space, and in turn, this implies that the transition kernel of pHMC defined by $\pi(x, B) = \mathbb{P}[X'(x) \in B]$ leaves μ in (2.3) invariant [3].

Below, our key assumption in this setup will be that U is a gradient Lipschitz function on the Hilbert space $\mathcal{H}^s$ where the reference measure μ_0 is supported:

Assumption 2.3 The target measure μ is a probability measure on $\mathcal{H}^s$ that is absolutely continuous with respect to μ_0. The relative density is proportional to $\exp(-U)$ where $U : \mathcal{H}^s \to [0, \infty)$ is a Fréchet differentiable function satisfying the gradient Lipschitz condition

$$|\partial_h U(x) - \partial_h U(y)| \le L_g\,|x - y|_s\,|h|_s \qquad \text{for all } x, y, h \in \mathcal{H}^s$$

for some finite positive constant L_g.

In Assumption 2.3, $\partial_h U$ denotes the directional derivative of U in direction h. We also use the notation DU to denote the differential of U, i.e., $(DU)(x)$ is the linear functional on $\mathcal{H}^s$ defined by $(DU)(x)[h] = (\partial_h U)(x)$. Identifying the dual space of $\mathcal{H}^s$ with $\mathcal{H}^{-s}$, Assumption 2.3 shows that we can view DU as a Lipschitz continuous function from $\mathcal{H}^s$ to $\mathcal{H}^{-s}$, i.e.,

$$|DU(x) - DU(y)|_{-s} \leq L_g \, |x - y|_s \qquad \text{for all } x, y \in \mathcal{H}^s. \tag{2.7}$$

Recalling that $\mathcal{C}$ is an isometry from $\mathcal{H}^{-s}$ to $\mathcal{H}^{2-s}$, and $\mathcal{H}^{2-s}$ is continuously embedded into $\mathcal{H}^s$ for $s < 1$, we see that Assumption 2.3 implies that the drift function

$$b(x) := -x - \mathcal{C} DU(x) \tag{2.8}$$

occurring in (2.5) is a Lipschitz continuous map from $\mathcal{H}^s$ to $\mathcal{H}^s$.

Remark 2.4 The global Lipschitz condition in Assumption 2.3 is principally the same as Condition 3.2 in [3] except here the domain $\mathcal{H}^s$ of the potential energy is defined in terms of the covariance operator itself rather than in terms of an auxiliary operator with related eigenfunctions and eigenvalues.

2.2 General setting

We now introduce a more general setup that includes the Markov chain induced by pHMC as a special case. We fix $s \in \mathbb{R}$, and we assume that $b : \mathcal{H}^s \to \mathcal{H}^s$ is a Lipschitz continuous function. Let $\phi_t(x, v) = (q_t(x, v), v_t(x, v))$ denote the exact flow of the Hilbert space valued ODE given by

$$\frac{d}{dt} q_t = v_t, \quad \frac{d}{dt} v_t = b(q_t), \quad (q_0(x, v), v_0(x, v)) = (x, v). \tag{2.9}$$

As above, we fix a constant duration $T \in (0, \infty)$ and consider the Markov chain on $\mathcal{H}^s$ with transition step

$$x \mapsto X'(x) := q_T(x, \xi), \qquad \xi \sim \mathcal{N}(0, \widetilde{\mathcal{C}}) \tag{2.10}$$

where $\widetilde{\mathcal{C}}$ is a linear operator on $\mathcal{H}$ with the same eigenfunctions as $\mathcal{C}$.

Assumption 2.5 $\widetilde{\mathcal{C}} : \mathcal{H} \to \mathcal{H}$ is a symmetric linear operator with eigenfunctions $\{e_i\}_{i \in \mathbb{N}}$ and corresponding eigenvalues $\{\widetilde{\lambda}_i\}_{i \in \mathbb{N}}$. Moreover, the operator $\widetilde{\mathcal{C}} \mathcal{C}^{-s}$ is trace class, i.e.,

$$\text{trace}(\widetilde{\mathcal{C}} \mathcal{C}^{-s}) = \sum_{j=1}^{\infty} \widetilde{\lambda}_j \lambda_j^{-s} < \infty.$$

For $\xi \sim \mathcal{N}(0, \widetilde{C})$, this implies $\mathbb{E}\,|\xi|_s^2 = \mathrm{trace}(\widetilde{C}C^{-s}) < \infty$, and thus, ξ in (2.10) is a Gaussian random variable on $\mathcal{H}^s$. Let $\pi(x, B) = \mathbb{P}[X'(x) \in B]$ denote the corresponding transition kernel. In particular, in the case where b is given by (2.8) and $\widetilde{C} = C$, we recover the Markov chain associated to pHMC. When $\widetilde{C} \neq C$, the choice $b(x) = -\widetilde{C}C^{-1}x - \widetilde{C}DU(x)$ ensures that the corresponding partially preconditioned dynamics in (2.9) leaves invariant the probability measure $\mu(dx)\mathcal{N}(0, \widetilde{C})(dv)$.

Our main result rests on the assumption that the Hilbert space $\mathcal{H}^s$ can be split into a finite dimensional subspace $\mathcal{H}^{s,\ell}$ ("the low modes") and its orthogonal complement $\mathcal{H}^{s,h}$ ("the high modes") such that $b(x)$ is close to a linear map on $\mathcal{H}^{s,h}$. More precisely, fix $n \in \mathbb{N}$. Let $\mathcal{H}^{s,\ell} := \mathrm{span}\{e_1, \dots, e_n\}$, and let $\mathcal{H}^{s,h}$ denote its orthogonal complement, i.e., $\mathcal{H}^{s,h}$ is the closure in $\mathcal{H}^s$ of $\mathrm{span}\{e_{n+1}, e_{n+2}, \dots\}$. Thus $\mathcal{H}^s = \mathcal{H}^{s,\ell} \oplus \mathcal{H}^{s,h}$. For any $x \in \mathcal{H}^s$, we denote by x^ℓ and x^h the orthogonal projections onto $\mathcal{H}^{s,\ell}$ and $\mathcal{H}^{s,h}$, respectively.

Assumption 2.6 b is a function from $\mathcal{H}^s$ to $\mathcal{H}^s$ such that $b(0) = 0$. Moreover it satisfies the following conditions:

(B1) There exists $L \in [1, \infty)$ such that

$$|b(x) - b(y)|_s \leq L\,|x - y|_s \quad \text{for all } x, y \in \mathcal{H}^s. \qquad (2.11)$$

(B2) There exists $n \in \mathbb{N}$ such that

$$\left| b^h(x) - b^h(y) + x^h - y^h \right|_s \leq \frac{1}{3}\,|x - y|_s \quad \text{for all } x, y \in \mathcal{H}^s. \qquad (2.12)$$

(B3) There exist $K > 0$ and $A \geq 0$ such that

$$\langle x, b(x) \rangle_s \leq -K\,|x|_s^2 + A \quad \text{for any } x \in \mathcal{H}^s. \qquad (2.13)$$

Condition (B1) is a global Lipschitz condition. Since $b(0) = 0$, it implies the linear growth condition $|b(x)|_s \leq L\,|x|_s$, and this condition and (B3) imply that $K \leq L$. Condition (B2) says that in the high modes, $b(x)$ behaves essentially as a linear drift. Finally, Condition (B3) is a standard drift condition which implies that the Markov chain has a Foster–Lyapunov function. It is similar to other conditions in the literature that consider Markov processes on unbounded spaces based on second-order dynamical systems including Hypothesis (H2) in [10], Equation (13) of [66], Hypothesis 1.1 in [69], Condition 3.1 in [56], and Assumption 1.2 in [25].

Lemma 2.7 (Foster–Lyapunov function) *Suppose that Assumptions* 2.5 *and* 2.6 *hold. Then for any $T > 0$ satisfying $LT^2 \leq \frac{1}{48}\frac{K}{L}$ we have*

$$\mathbb{E}\left[\left|X'(x)\right|_s^2\right] \leq \left(1 - \frac{KT^2}{2}\right)|x|_s^2 + 5(A + \mathrm{trace}(\widetilde{C}C^{-s}))T^2 \quad \text{for all } x \in \mathcal{H}^s.$$

The proof of this lemma is given in Sect. 5.

Example 2.1 *(pHMC)* Suppose that b is given by (2.8) and U satisfies the global Lipschitz condition in Assumption 2.3 with Lipschitz constant L_g. Then condition (B1) holds with Lipschitz constant $L = 1 + \lambda_1^{1-s} L_g$ and Condition (B2) holds with $n = \inf\{k \in \mathbb{N} : \lambda_{k+1}^{1-s} < 1/(3L_g)\}$. Indeed, by (2.7),

$$|b(x) - b(y)|_s \leq |x - y|_s + \left| \mathcal{C}^{1-s}(DU(x) - DU(y)) \right|_{-s} \leq (1 + \lambda_1^{1-s} L_g) |x - y|_s \,,$$

and $\left| b^h(x) - b^h(y) + x^h - y^h \right|_s \leq \lambda_{n+1}^{1-s} L_g |x - y|_s \leq (1/3) |x - y|_s$ as required. Moreover, the drift condition (B3) can be verified in examples, see Sect. 3.

2.3 Two-scale coupling

We now introduce a coupling for the transition steps of two copies of the Markov chain starting at different initial conditions x and y. We use a synchronous coupling of the high modes in $\mathcal{H}^{s,h}$ and a different coupling for the low modes in $\mathcal{H}^{s,\ell}$ that together enable us to derive a weak form of contractivity. Note that the covariance operator $\widetilde{C}$ has a bounded inverse on the finite dimensional subspace $\mathcal{H}^\ell$. Therefore, for $\mathsf{h} \in \mathcal{H}^\ell$, the Gaussian measure $\mathcal{N}(\mathsf{h}, \widetilde{C})$ is absolutely continuous w.r.t. $\mathcal{N}(0, \widetilde{C})$ with relative density

$$\rho_{\mathsf{h}}(x) = \exp\left(\langle \widetilde{C}^{-1}\mathsf{h}, x \rangle - \langle \widetilde{C}^{-1}\mathsf{h}, \mathsf{h} \rangle/2 \right). \tag{2.14}$$

Let $\gamma > 0$ be a positive constant. The precise value of the parameter γ will be chosen in an appropriate way below. The coupling transition step is given by $(x, y) \mapsto (X'(x, y), Y'(x, y))$ where

$$X'(x, y) = q_T(x, \xi), \quad \text{and} \quad Y'(x, y) = q_T(y, \eta) \tag{2.15}$$

with $\xi \sim \mathcal{N}(0, \widetilde{C})$ and η defined in high/low components as $\eta^h := \xi^h$ and

$$\eta^\ell := \begin{cases} \xi^\ell + \gamma z^\ell & \text{if } \mathcal{U} \leq \rho_{-\gamma z^\ell}(\xi^\ell), \\ \mathcal{R}\xi^\ell & \text{otherwise.} \end{cases} \tag{2.16}$$

Here $\mathcal{U} \sim \mathrm{Unif}(0, 1)$ is independent of ξ, $z := x - y$, and the reflection operator $\mathcal{R}$ is defined by

$$\mathcal{R} := \widetilde{C}^{1/2}(I - 2e^\ell \langle e^\ell, \cdot \rangle)\widetilde{C}^{-1/2}, \quad \text{where} \quad e^\ell := \widetilde{C}^{-1/2}z^\ell / \left| \widetilde{C}^{-1/2}z^\ell \right|. \tag{2.17}$$

Due to Assumption 2.6 (B2), the component in $\mathcal{H}^{s,h}$ of the resulting coupled dynamics is contracting in a finite time interval as a result of the linear part of the drift in (2.9). Moreover, the coupling of the components of the initial velocities in $\mathcal{H}^{s,\ell}$ is similar to the coupling in [9] which is inspired by a related coupling for second order Langevin diffusions [25]. It is defined in such a way that $\xi^\ell - \eta^\ell = -\gamma z^\ell$ occurs with the maximal possible probability. As illustrated in Fig. 1, and proven later in
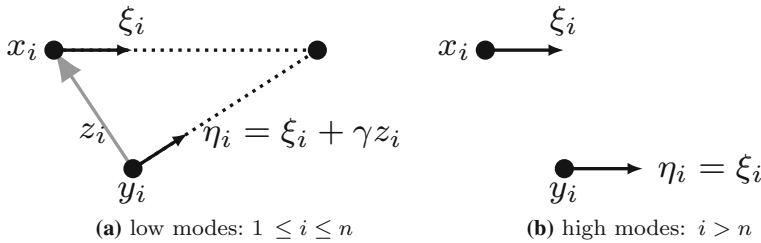
(a) low modes: $1 \leq i \leq n$ **(b)** high modes: $i > n$

Fig. 1 Two-Scale Coupling. A diagram illustrating the two-scale coupling in the case $\gamma = T^{-1}$. (a) In the low modes ($1 \leq i \leq n$), the initial velocities are coupled such that the final positions $q_T^\ell(x, \xi) = q_T^\ell(y, \eta)$ are equal when $b \equiv 0$. (b) In the high modes ($i > n$), the initial velocities are synchronously coupled

Lemma 4.3, the reason for this choice is that the projection of the difference process on $\mathcal{H}^{s,\ell}$, i.e., $q_t^\ell(x, \xi) - q_t^\ell(y, \eta)$, is contracting in a finite time interval if the difference $\xi^\ell - \eta^\ell$ of the initial velocities is negatively proportional to the difference of the initial positions $x^\ell - y^\ell$. Note that if $b(x) = 0$ or $b(x) = -x$ then the optimal choices of γ would be $\gamma = T^{-1}$ and $\gamma = \cot(T)$, respectively, because for these choices, $X'(x, y) = Y'(x, y)$ if $\mathcal{U} \leq \rho_{-\gamma z^\ell}(\xi^\ell)$. In the case where $\xi^\ell - \eta^\ell \neq -\gamma z^\ell$, a reflection coupling is applied. The corresponding reflection $\mathcal{R}$ is an orthogonal transformation w.r.t. the inner product $\langle x, y \rangle_{\widetilde{\mathcal{C}}} = \langle \widetilde{\mathcal{C}}^{-1/2} x, \widetilde{\mathcal{C}}^{-1/2} y \rangle$ induced by the covariance operator $\widetilde{\mathcal{C}}$ on $\mathcal{H}^\ell$.

In order to verify that $(X'(x, y), Y'(x, y))$ is indeed a coupling of the transition probabilities $\pi(x, \cdot)$ and $\pi(y, \cdot)$, we remark that the distribution of η is $\mathcal{N}(0, \widetilde{\mathcal{C}})$ since, by definition of η^ℓ in (2.16) and a change of variables,

$$
\begin{aligned}
&P[\eta^\ell \in B] \\
&= E\left[I_B(\xi^\ell + \gamma z^\ell) \rho_{-\gamma z^\ell}(\xi^\ell) \wedge 1 \right] + E\left[I_B(\mathcal{R}\xi^\ell)\left(1 - \rho_{-\gamma z^\ell}(\xi^\ell)\right)^+ \right] \\
&= E\left[\rho_{\gamma z^\ell}(\xi^\ell) I_B(\xi^\ell) \rho_{-\gamma z^\ell}(\xi^\ell - \gamma z^\ell) \wedge 1 \right] + E\left[I_B(\xi^\ell)\left(1 - \rho_{-\gamma z^\ell}(\mathcal{R}\xi^\ell)\right)^+ \right] \\
&= E\left[I_B(\xi^\ell) 1 \wedge \rho_{\gamma z^\ell}(\xi^\ell) \right] + E\left[I_B(\xi^\ell)\left(1 - \rho_{\gamma z^\ell}(\xi^\ell)\right)^+ \right] = P[\xi^\ell \in B]
\end{aligned}
$$

for any measurable set B. Here $a \wedge b$ denotes the minimum of real numbers a and b, $I_B(\cdot)$ denotes the indicator function for the set B, and we have used that $\mathcal{N}(0, \widetilde{\mathcal{C}})$ is invariant under the reflection $\mathcal{R}$, $\mathcal{R} z^\ell = -z^\ell$, and by (2.14), $\rho_{-h}(x - h)\rho_h(x) = 1$. A similar calculation shows that

$$
P[\eta^\ell \neq \xi^\ell + \gamma z^\ell] = E\left[\left(1 - \rho_{-\gamma z^\ell}(\xi^\ell)\right)^+\right] = d_{\mathrm{TV}}(\mathcal{N}(0, \widetilde{\mathcal{C}}), \mathcal{N}(\gamma z^\ell, \widetilde{\mathcal{C}})) \quad (2.18)
$$

where d_{TV} is the total variation distance. Hence, by the coupling characterization of the total variation distance, $\eta^\ell = \xi^\ell + \gamma z^\ell$ does indeed hold with the maximal possible probability. Note that if z is not in the reproducing kernel Hilbert space of the covariance operator $\widetilde{\mathcal{C}}$ then the probability of the event $\eta^\ell \neq \xi^\ell + \gamma z^\ell$ in (2.18)

tends to one as the number of low modes increases. This explains why it is necessary to split the Hilbert space and apply a two-scale coupling.

2.4 Contractivity

We now state our main contraction bound for the coupling introduced above. We first define a norm $\|\!|\cdot|\!\|_\alpha$ on $\mathcal{H}^s$ where the high modes are weighted by $\alpha > 0$:

$$\|\!|x|\!\|_\alpha \;=\; \left|\widetilde{\mathcal{C}}^{-\frac{1}{2}}x^\ell\right| + \alpha\left|x^h\right|_s \;=\; \left|\widetilde{\mathcal{C}}^{-\frac{1}{2}}\mathcal{C}^{\frac{s}{2}}x^\ell\right|_s + \alpha\left|x^h\right|_s. \tag{2.19}$$

Let $\sigma_{min} = \min_{1\le i\le n}\{\widetilde{\lambda}_i^{-1/2}\lambda_i^{s/2}\}$ and $\sigma_{max} = \max_{1\le i\le n}\{\widetilde{\lambda}_i^{-1/2}\lambda_i^{s/2}\}$. Note that

$$\sigma_{min}\left|x^\ell\right|_s \;\le\; \left|\widetilde{\mathcal{C}}^{-1/2}\mathcal{C}^{s/2}x^\ell\right|_s \;\le\; \sigma_{max}\left|x^\ell\right|_s \quad \text{for all } x \in \mathcal{H}^s. \tag{2.20}$$

Thus $\|\!|\cdot|\!\|_\alpha$ and $|\cdot|_s$ are equivalent norms with

$$\min(\sigma_{min},\alpha)\,|x|_s \;\le\; \|\!|x|\!\|_\alpha \;\le\; \sqrt{2}\,\max(\sigma_{max},\alpha)\,|x|_s \quad \text{for all } x \in \mathcal{H}^s. \tag{2.21}$$

Remark 2.8 If the dimension is infinite then the operator $\widetilde{\mathcal{C}}^{-1}\mathcal{C}^s$ is unbounded on $\mathcal{H}^s$, because its inverse is trace class. Nonetheless, $\|\!|x|\!\|_\alpha$ is a well-defined norm for any $x \in \mathcal{H}^s$ because the operator $\widetilde{\mathcal{C}}^{-\frac{1}{2}}\mathcal{C}^{\frac{s}{2}}$ appearing in $\|\!|x|\!\|_\alpha$ only acts on the projection x^ℓ of x onto the finite dimensional space $\mathcal{H}^{s,\ell}$.

As we will see below, even when U is non-convex, we can still obtain contractivity with respect to a semimetric $\rho : \mathcal{H}^s \times \mathcal{H}^s \to [0, \infty)$ of the form

$$\rho(x, y) \;=\; \sqrt{f(\|\!|x - y|\!\|_\alpha)(1 + \epsilon\,|x|_s^2 + \epsilon\,|y|_s^2)}, \qquad x, y \in \mathcal{H}^s, \tag{2.22}$$

where $f : [0, \infty) \to [0, \infty)$ is a concave function given by

$$f(r) = \int_0^r e^{-at}\,I_{\{t\le R\}}\,dt \;=\; \frac{1}{a}\left(1 - e^{-a\,r\wedge R}\right), \tag{2.23}$$

and where $R > 0$, $a > 0$, and $\epsilon > 0$ are parameters to be specified below. The semimetric ρ is similar to the one introduced in [9] in order to prove contractivity of the HMC transition step in the finite dimensional case. In general, ρ is not a metric, since the triangle inequality might be violated. Note that f is non-decreasing, and constant when $r \ge R$.

Remark 2.9 The semimetric (2.22) incorporates, in a multiplicative or weighted way, the quadratic Foster–Lyapunov function for pHMC from Lemma 2.7 with weight ϵ. The idea to use semimetrics of this general form to study contraction properties of Markov processes goes back to [12,35]; see also [26].

Lemma 2.7 implies that the coupling transition $(x, y) \mapsto (X'(x, y), Y'(x, y))$ also has a quadratic Foster–Lyapunov function: If $LT^2 \leq \frac{1}{48} \frac{K}{L}$ then

$$\mathbb{E}\left[|X'(x, y)|_s^2 + |Y'(x, y)|_s^2\right] \leq \left(1 - \frac{KT^2}{2}\right)(|x|_s^2 + |y|_s^2) + 10(A + \text{trace}(\widetilde{\mathcal{C}}\mathcal{C}^{-s}))T^2.$$

We fix a finite, positive constant R satisfying

$$R \geq 8\sqrt{40}(A + \text{trace}(\widetilde{\mathcal{C}}\mathcal{C}^{-s}))^{1/2}\sigma_{max}LK^{-1/2}. \tag{2.24}$$

In our main result below, we choose $\alpha := 4\sigma_{max}L$. In this case, the choice of R in (2.24) guarantees that a strict drift condition

$$\mathbb{E}\left[|X'(x, y)|_s^2 + |Y'(x, y)|_s^2\right] \leq \left(1 - KT^2/4\right)(|x|_s^2 + |y|_s^2) \tag{2.25}$$

holds for all (x, y) satisfying $\|x - y\|_\alpha \geq R$, because by (2.21) and since $L \geq 1$,

$$\|x - y\|_\alpha \leq 4\sqrt{2}\,\sigma_{max}L\,|x - y|_s \leq 8\sigma_{max}L\sqrt{|x|_s^2 + |y|_s^2}. \tag{2.26}$$

The asymptotic strict drift condition in (2.25) allows us to split the proof of contractivity into two parts: (i) $\|x - y\|_\alpha \geq R$ where any coupling is contracting in ρ due to (2.25), and (ii) $\|x - y\|_\alpha < R$, where ρ is contracting due to the specially designed two-scale coupling.

Theorem 2.10 *Suppose that Assumption 2.6 holds. Let $T > 0$ satisfy*

$$\frac{\sigma_{max}}{\sigma_{min}}LT^2 \leq \min\left(\frac{1}{48}\frac{K}{L}, \frac{1}{256LR^2}\frac{\sigma_{min}}{\sigma_{max}}\right) \tag{2.27}$$

Let α, γ, a, and ϵ be given by

$$\alpha := 4\sigma_{max}L, \tag{2.28}$$

$$\gamma := \min\left(T^{-1}, R^{-1}/4\right), \tag{2.29}$$

$$a := T^{-1}, \tag{2.30}$$

$$\epsilon := (1/160)(A + \text{trace}(\widetilde{\mathcal{C}}\mathcal{C}^{-s}))^{-1}e^{-R/T}. \tag{2.31}$$

Then for any $x, y \in \mathcal{H}^s$, we have

$$E[\rho(X'(x, y), Y'(x, y))] \leq e^{-c}\rho(x, y), \qquad where \tag{2.32}$$

$$c = \min\left(\frac{1}{16}KT^2, \frac{1}{128}T\max(R, T)\,e^{-\max(R,T)/T}\right). \tag{2.33}$$

The proof of this theorem is given in Sect. 6.

Remark 2.11 The rate in (2.33) is similar to the rate in the finite-dimensional case found in Theorem 2.3 of Ref. [9]. The main difference is that the condition on LT^2 in (2.27) now reflects the effect of preconditioning.

2.5 Quantitative bounds for distance to the invariant measure

Theorem 2.10 establishes global contractivity of the transition kernel $\pi(x, dy)$ w.r.t. the Kantorovich distance based on the underlying semimetric ρ, which for probability measures ν, η on $\mathcal{H}^s$ is defined as

$$\mathcal{W}_\rho(\nu, \eta) = \inf_{\gamma \in C(\nu, \eta)} \int \rho(x, y) \, \gamma(dx \, dy) = \inf_{X' \sim \nu, Y' \sim \eta} E\left[\rho(X', Y')\right]$$

where the infimum is over all couplings γ of ν and η. Moreover, it implies quantitative bounds for the standard L^1 Wasserstein distance

$$\mathcal{W}^{s,1}(\nu, \mu) = \inf_{\gamma \in C(\nu, \mu)} \int |x - y|_s \, \gamma(dx \, dy) = \inf_{X' \sim \nu, Y' \sim \mu} E\left[|X' - Y'|_s\right]$$

with respect to the invariant measure μ on $\mathcal{H}^s$. Let $M_1(\nu) := \int |x|_s \, \nu(dx)$.

Corollary 2.12 *Suppose that Assumption 2.6 holds. Let $T \in (0, R)$ satisfy (2.27). Then for any $k \in \mathbb{N}$ and for any probability measures ν, η on $\mathcal{H}^s$,*

$$\mathcal{W}_\rho(\nu\pi^k, \eta\pi^k) \le e^{-ck} \mathcal{W}_\rho(\nu, \eta), \qquad and \tag{2.34}$$

$$\mathcal{W}^{s,1}(\nu\pi^k, \mu) \le C \left(1 + \sqrt{\epsilon} M_1(\nu) + (1/4)K^{-1/2} e^{-R/(2T)}\right) e^{-ck} \tag{2.35}$$

where the rate c and the constant ϵ are given explicitly by (2.33) and (2.31), and

$$C = \max\left(2T\sigma_{min}^{-1}, \, 23 \, (A + \mathrm{trace}(\widetilde{\mathcal{C}}\mathcal{C}^{-s}))^{1/2} e^{R/(2T)}\right). \tag{2.36}$$

In particular, for a given constant $\delta \in (0, \infty)$, the L^1 Wasserstein distance $\Delta(k) = \mathcal{W}^{s,1}(\nu\pi^k, \mu)$ w.r.t. μ after k steps of the chain with initial distribution ν satisfies $\Delta(k) \le \delta$ provided

$$k \ge \frac{1}{c} \log \frac{C \left(1 + \sqrt{\epsilon} M_1(\nu) + (1/4)K^{-1/2} e^{-R/(2T)}\right)}{\delta}. \tag{2.37}$$

The corollary is a rather direct consequence of Theorem 2.10. A short proof is included in Sect. 6.

Remark 2.13 *(Quantitative bounds for ergodic averages)* MCMC methods are often applied to approximate expectation values w.r.t. the target distribution by ergodic averages of the Markov chain. Our results (e.g. (2.34)) directly imply completely explicit bounds for bias and variances, as well as explicit concentration inequalities for these ergodic averages in the case of pHMC. Indeed, the general results by Joulin and

Ollivier [40] show that such bounds follow directly from an L^1 Wasserstein contraction w.r.t. an arbitrary metric ρ, which is precisely the statement shown above.

3 Applications

3.1 Transition path sampling

Here we discuss the use of pHMC in transition path sampling (TPS). As an application of Theorem 2.10, we obtain dimension-free contraction rates for exact preconditioned HMC in this context. Fix a time horizon $\tau > 0$ (not to be confused with the duration parameter in preconditioned HMC which we denote by T). The aim of TPS [4,37, 38,65] is to sample from a *diffusion bridge* or *conditioned diffusion*, i.e., from the conditional law $\nu_{a,b}$ of the solution $X : [0, \tau] \to \mathbb{R}^d$ to a d-dimensional stochastic differential equation of the form

$$dX(t) = -\nabla\Psi(X(t))\,ds + dW(t) \tag{3.1}$$

given both initial and final conditions

$$X(0) = a \quad \text{and} \quad X(\tau) = b.$$

Here $\Psi : \mathbb{R}^d \to \mathbb{R}$ is a given potential energy function and W is a d-dimensional standard Brownian motion. TPS is particularly relevant to molecular dynamics where the states a and b represent different configurations of a molecular system [6,59,63].

We first recenter: Let $\mu = \nu \circ \theta_M^{-1}$ denote the law of the recentered bridge where $\theta_M(x) = x - M$ is the translation on path space by the mean $M(t) = a + (t/\tau)(b-a)$ of the Brownian bridge from a to b. Then by Girsanov's theorem, the measure μ is absolutely continuous with respect to the law μ_0 of the Brownian bridge from 0 to 0 [4,39]. Moreover, the measure μ_0 is the centered Gaussian measure on the Hilbert space $\mathcal{H} = L^2([0, \tau], \mathbb{R}^d)$ with covariance operator $\mathcal{C} = -\Delta_D^{-1}$ where Δ_D is the Dirichlet Laplacian, and the relative density of μ with respect to μ_0 is proportional to $\exp(-U(x))$ where the function $U(x)$ is defined in terms of the so-called path potential energy function $G : \mathbb{R}^d \to \mathbb{R}$ as follows

$$U(x) = \int_0^\tau G(x(t) + M(t))dt \quad \text{where} \quad G(\cdot) = \frac{1}{2}|\nabla\Psi(\cdot)|^2 - \frac{1}{2}\Delta\Psi(\cdot). \tag{3.2}$$

In the main convergence result given below, we make the following regularity assumption on G.

Assumption 3.1 The function $G : \mathbb{R}^d \to \mathbb{R}$ is continuously differentiable. Moreover, $\nabla G(0) = 0$, and ∇G is uniformly bounded and globally Lipschitz continuous, i.e., there exist finite constants M_G, L_G such that for all $x, y \in \mathbb{R}^d$,

$$|\nabla G(x)| \le M_G, \quad \text{and} \quad |\nabla G(x) - \nabla G(y)| \le L_G|x - y|.$$

This regularity assumption frequently holds in molecular dynamics applications, since the configuration space of molecular systems is usually taken to be a fixed cubic box with periodic boundary conditions [1,8,13,23,29,46,67]. In this case, we can lift the TPS problem to the covering space $\mathbb{R}^d$ by extending the path potential to a periodic function on this space. Thus after recentering the coordinate system, Assumption 3.1 is satisfied whenever G is C^2.

To implement TPS on a computer, we use the finite difference method to approximate the infinite-dimensional distribution $\mu(dx) \propto \exp(-U(x))\mathcal{N}(0, \mathcal{C})(dx)$ by a finite-dimensional probability measure μ_m. Other approximations, e.g., Galerkin or finite-element, are also possible and should yield similar results. We focus on the finite difference method because it is widely used in practice. Discretize the interval $[0, \tau]$ into $m + 2$ evenly spaced grid points

$$\mathsf{t}_j = \tau j/(m+1), \quad j = 0, \ldots, m+1. \tag{3.3}$$

The space of paths on $\mathbb{R}^d$ is then approximated by the finite-dimensional space $\mathbb{R}^{md}$. Specifically, we write $x \in \mathbb{R}^{md}$ as

$$x = (x_{1:d}, \ldots, x_{m+1:m+d})$$

where the j-th component $x_{j+1:j+d} := (x_{j+1}, \ldots, x_{j+d})$ is a d-dimensional vector that can be viewed as an approximation of $x(\mathsf{t}_j)$ for $j = 1, \ldots, m$. The Dirichlet Laplacian Δ_D is approximated by the $md \times md$ Dirichlet Laplacian matrix $\mathbf{\Delta}_{D,m}$ with (i, j)-th entry

$$(\mathbf{\Delta}_{D,m})_{i,j} = \begin{cases} -2\,(\tau/(m+1))^{-2} & \text{if } |i-j| = 0, \\ (\tau/(m+1))^{-2} & \text{if } |i-j| = d, \\ 0 & \text{otherwise.} \end{cases}$$

The covariance operator $\mathcal{C}$ is approximated by the $md \times md$ matrix $\mathcal{C} = -\mathbf{\Delta}_{D,m}^{-1}$, and the Hilbert space $\mathcal{H}$ is represented by $\mathbb{R}^{md}$ with inner product given by the weighted dot product $\langle x, y \rangle = \frac{\tau}{m+1} x \bullet y$. The functional (3.2) is discretized as

$$U_m(x) = \frac{\tau}{m+1} G_m(x), \quad \text{where} \quad G_m(x) = \sum_{j=1}^{m} G(x_{j+1:j+d} + \mathsf{M}(\mathsf{t}_j)).$$

Note that if the vector x contains the grid values of a smooth function x, then $U_m(x) \to U(x)$ as $m \to \infty$. In summary, the infinite-dimensional path distribution $\mu(dx)$ is approximated by the finite-dimensional probability measure $\mu_m(dx)$ with non-normalized density $\exp\left(-U_m(x) - \frac{1}{2}\langle x, \mathcal{C}^{-1} x \rangle\right)$.

To approximately sample from μ_m, we use pHMC with transition step in (2.10). This corresponds to a Markov chain on $\mathbb{R}^{md}$ with transition step

$$x \mapsto X'(x) := q_T(x, \xi), \quad \xi \sim \mathcal{N}\left(0, \frac{m+1}{\tau}\mathcal{C}\right) \tag{3.4}$$

where $\boldsymbol{q}_t$ solves

$$\frac{d}{dt}\boldsymbol{q}_t = \boldsymbol{v}_t, \quad \frac{d}{dt}\boldsymbol{v}_t = \boldsymbol{b}(\boldsymbol{q}_t), \quad (\boldsymbol{q}_0(\boldsymbol{x},\boldsymbol{v}), \boldsymbol{v}_0(\boldsymbol{x},\boldsymbol{v})) = (\boldsymbol{x},\boldsymbol{v}) \in \mathbb{R}^{2md}, \qquad (3.5)$$

with $\boldsymbol{b}(\boldsymbol{x}) = -\boldsymbol{x} - \mathcal{C}\nabla G_m(\boldsymbol{x})$. Let π_m denote the transition kernal of (3.4).

Theorem 3.2 (Transition Path Sampling) *Suppose that Assumption* 3.1 *holds. Let* $\kappa := 2(\tau^2/\pi^2)L_G$, $m_\ell = \lfloor\sqrt{3\kappa}\rfloor$, $n = m_\ell d$, *and* $m^\star = \lceil(m_\ell+1)\pi/2\rceil$. *Let* R, c, C *and* ϵ *be defined as*

$$R := 16\sqrt{20}\pi\kappa^{1/2}(1+\kappa)((\tau/\pi)^3 M_G^2 + d)^{1/2}, \qquad (3.6)$$

$$c := \min((1/32)T^2, (1/128)T^2\max(R,T)e^{-\max(R,T)/T}), \qquad (3.7)$$

$$C := \max(T\tau, 23((\tau^5/\pi^4)M_G^2 + d\tau^2/3)^{1/2}e^{R/(2T)}), \qquad (3.8)$$

$$\epsilon := (\tau^5 M_G^2)^{-1}e^{-R/T}. \qquad (3.9)$$

Suppose that the duration parameter $T \in (0, R)$ *satisfies*

$$2\sqrt{3\kappa}(1+\kappa)T^2 \leq \min\left(\frac{1}{96(1+\kappa)}, \frac{1}{512\sqrt{3\kappa}(1+\kappa)R^2}\right). \qquad (3.10)$$

Then for any $m > m^\star$, $k \in \mathbb{N}$, *and probability measure* ν_m *on* $\mathbb{R}^{md}$,

$$\mathcal{W}^{0,1}(\nu_m\pi_m^k, \mu_m) \leq Ce^{-ck}\left(1 + \sqrt{\epsilon}M_1(\nu_m) + (1/8)e^{-R/(2T)}\right). \qquad (3.11)$$

Remark 3.3 Note that the upper bound in (3.11) depends on dimension only through the initial distribution. The dimension independence in the other terms of this bound reflects that the finite-dimensional pHMC algorithm in (3.4) converges to an infinite-dimensional pHMC algorithm whose transition kernel satisfies an infinite-dimensional analog of this quantitative bound.

A proof of this result is given in Sect. 7.1.

3.2 Path integral molecular dynamics

Here we discuss the use of pHMC for path-integral molecular dynamics (PIMD), and as an application of Theorem 2.10, obtain dimension-free contraction rates for preconditioned HMC in this context. PIMD is used to compute exact Boltzmann properties and approximate dynamical properties of quantum mechanical systems [14]. The technique is based on Feynman's path-integral formulation of quantum statistical mechanics [27], and the observation that the quantum Boltzmann statistical mechanics of a quantum system can be reproduced by the *classical* Boltzmann statistical mechanics of a ring-polymer system [14].

Consider N interacting quantum particles in 3D with potential energy operator given by

$$\hat{V} = V(\hat{q}_1, \ldots, \hat{q}_N) \tag{3.12}$$

where $\hat{q}_i$ is the three-dimensional position operator of particle i and $V : \mathbb{R}^d \to \mathbb{R}$ is a potential energy function where $d = 3N$ [32]. The thermal equilibrium properties of this system are described by the quantum mechanical Boltzmann partition function,

$$Q = \mathrm{Tr}[e^{-\beta \hat{V}}] \tag{3.13}$$

where β is an inverse temperature parameter. For some $\mathsf{a} > 0$, suppose that the potential energy function can be written as

$$V(\cdot) = \frac{1}{2}\mathsf{a}|\cdot|^2 + G(\cdot)$$

where $G : \mathbb{R}^d \to \mathbb{R}$. Then the partition function Q can be written as the expected value of a Gaussian random variable on loop space as follows

$$Q = \mathbb{E}[e^{-U(\xi)}], \quad \text{where} \quad \xi \sim \mu_0 = \mathcal{N}(0, \mathcal{C}_{\mathsf{a}}), \tag{3.14}$$

and the covariance operator $\mathcal{C}_{\mathsf{a}}$ of the Gaussian reference measure is defined in terms of the Laplacian with periodic boundary conditions Δ_P on $L^2([0, \beta], \mathbb{R}^d)$ as follows

$$\mathcal{C}_{\mathsf{a}} = (-\Delta_P + \mathsf{a}I)^{-1},$$

where I is the identity operator and the potential energy $U(x)$ is given by

$$U(x) = \int_0^\beta G(x(\mathsf{t}))\mathsf{dt}. \tag{3.15}$$

The probability measures μ_0 and $\mu(dx) \propto \exp(-U(x)) \mathcal{N}(0, \mathcal{C}_{\mathsf{a}})(dx)$ are supported on the loop space consisting of all periodic continuous paths $x : [0, \beta] \to \mathbb{R}^d$. They are similar to the corresponding measures considered for Transition Path Sampling, but there is an additional, artificially introduced parameter a appearing in $\mathcal{C}_{\mathsf{a}}$. This parameter is essential because Δ_P is not invertible since it has a zero (leading) eigenvalue corresponding to the 'centroid mode' [50].

To implement PIMD on a computer, we use finite-differences to truncate the infinite-dimensional path distribution μ to a finite-dimensional one μ_m by discretizing the interval $[0, \beta]$ into $m + 1$ grid points

$$\mathsf{t}_j = \beta j / m, \quad j = 0, \ldots, m. \tag{3.16}$$

The space of loops on $\mathbb{R}^d$ is approximated by the finite-dimensional space $\mathbb{R}^{md}$. Specifically, we write $\boldsymbol{x} \in \mathbb{R}^{md}$ as

$$\boldsymbol{x} = (\boldsymbol{x}_{1:d}, \ldots, \boldsymbol{x}_{m+1:m+d})$$

where $x_{j+1:j+d} := (x_{j+1}, \ldots, x_{j+d})$ is a d-dimensional vector that can be viewed as an approximation of $x(\mathsf{t}_j)$ for $j = 1, \ldots, m$.

Remark 3.4 Comparing (3.3)–(3.16), note that the number of grid points in TPS, resp. PIMD, is $m + 2$, resp. $m + 1$. Nonetheless, in both cases path and loop space are approximated by $\mathbb{R}^{md}$. The difference in the number of grid points is due to the boundary conditions: in TPS the Dirichlet boundary conditions eliminate two unknown d-dimensional vectors, whereas in PIMD the periodic boundary conditions eliminate only one unknown d-dimensional vector. Thus, the total number of unknowns in both cases is md.

The periodic Laplacian Δ_P is approximated by the $md \times md$ discrete periodic Laplacian matrix $\Delta_{P,m}$ with (i, j)-th entry

$$(\Delta_{P,m})_{i,j} = \begin{cases} -2\,(\beta/m)^{-2} & \text{if } |i - j| = 0, \\ (\beta/m)^{-2} & \text{if } (i - j) \bmod (md) = d \text{ or } (j - i) \bmod (md) = d, \\ 0 & \text{otherwise.} \end{cases}$$

Naturally, the covariance operator $\mathcal{C}_{\mathsf{a}}$ is approximated by the $md \times md$ matrix $\mathcal{C}_{\mathsf{a}} = (-\Delta_{P,m} + \mathsf{a}I_{md \times md})^{-1}$ where $I_{md \times md}$ is the $md \times md$ identity matrix, and the infinite-dimensional Hilbert space $\mathcal{H}$ is represented by $\mathbb{R}^{md}$ with inner product given by the weighted scalar product $\langle x, y \rangle = \frac{\beta}{m} x \bullet y$. The functional in (3.2) is discretized as

$$U_m(x) = \frac{\beta}{m} G_m(x), \quad \text{where} \quad G_m(x) = \sum_{j=1}^{m} G(x_{j+1:j+d}).$$

In summary, the infinite-dimensional path distribution $\mu(dx)$ is approximated by the finite-dimensional distribution $\mu_m(dx) \propto \exp(-U_m(x))\mathcal{N}(0, \frac{m}{\beta}\mathcal{C}_{\mathsf{a}})(dx)$.

In this context, pHMC generates a Markov chain on $\mathbb{R}^{md}$ with invariant measure μ_m and with transition step given by

$$x \mapsto X'(x) := q_T(x, \xi), \quad \xi \sim \mathcal{N}(0, \frac{m}{\beta}\mathcal{C}_{\mathsf{a}}) \tag{3.17}$$

where q_t solves (3.5) with $b(x) = -x - \mathcal{C}_{\mathsf{a}}\nabla G_m(x)$.

Theorem 3.5 Path Integral Molecular Dynamics *Suppose that Assumption* 3.1 *holds. Let* $\kappa := 6a^{-1}L_G$, $m_\ell = \lceil \sqrt{3L_G/2}(\beta/\pi) \rceil$, $n = 2m_\ell d - d$, *and* $m^\star = \lceil 2\pi m_\ell \rceil$. *Let* R, c, C *and* ϵ *be defined as*

$$R := 16\sqrt{20}(1 + \kappa)^{3/2}((1/2)\beta a^{-1}M_G^2 + 2d(\beta^2 a + 1))^{1/2}, \tag{3.18}$$

$$c := \min((1/32)T^2, (1/128)T^2 \max(R, T)e^{-\max(R,T)/T}), \tag{3.19}$$

$$C := \max(2Ta^{1/2}, 23((1/2)\beta a^{-2}M_G^2 + 2d(a^{-1} + \beta^2))^{1/2}e^{R/(2T)}), \tag{3.20}$$

$$\epsilon := (1/80)a^2(\beta M_G^2)^{-1}e^{-R/T}. \tag{3.21}$$

Suppose that the duration parameter $T \in (0, R)$ satisfies

$$(1 + \kappa)^{3/2} T^2 \leq \min\left(\frac{1}{96(1 + \kappa)}, \frac{1}{256(1 + \kappa)^{3/2} R^2}\right). \qquad (3.22)$$

Then for any $m > m^\star$, $k \in \mathbb{N}$, and probability measure ν_m on $\mathbb{R}^{md}$, (3.11) holds for the transition kernel of (3.17).

A proof of this result is given in Sect. 7.2.

3.3 Numerical illustration of couplings

Before turning to the proofs of our main results, we test the two-scale coupling defined by (2.15) numerically on the following distributions.

- A TPS distribution with the three-well path potential energy function illustrated in Fig. 2a. The initial conditions of the components of the coupling are taken to be paths that pass through the two channels that connect the bottom two wells located at $x^\pm \approx (\pm 1.048, -0.042)$.
- A PIMD distribution where the underlying potential energy is the negative logarithm of the normal mixture density illustrated in Fig. 2b. The mixture components are twenty two-dimensional Gaussian distributions with covariance matrix given by the 2×2 identity matrix and with mean vectors given by 20 independent samples from the uniform distribution over the rectangle $[0, 10] \times [0, 10]$. The energy barriers are not large. The potential energy in this example is adapted from [45,47]. The initial paths are taken to be two unit circles one centered at $(1, 1)$ and the other centered at $(9, 9)$. The parameter a is selected to be 0.1.
- A PIMD distribution where the underlying potential energy is the negative logarithm of the Laplace mixture density illustrated in Fig. 2c. The mixture components are twenty two-dimensional (regularized) Laplace distributions using the same covariance matrix and mean vectors as in the preceding example. However, unlike the preceding example, in this example the underlying potential is only asymptotically convex. The initial paths are taken to be two unit circles one centered at $(1, 1)$ and the other centered at $(9, 9)$. The parameter a is selected to be 0.1.
- A PIMD distribution where the underlying potential energy is the banana-shaped potential energy illustrated in Fig. 2d. This function is highly non-convex and unimodal with a global minimum at the point $(1, 1)$. This minimum lies in a long, narrow, banana shaped valley. The initial paths are taken to be small circles centered at $(\pm 4, 16)$. The parameter a is selected to be 1.0.

For the TPS and PIMD distributions we use the finite-dimensional approximations described in Sects. 3.1 and 3.2, respectively. The resulting semi-discrete evolution equations are discretized in time using a strongly stable symmetric splitting integrator [3,43,44]. We describe this integrator in the specific context of TPS, since a very similar method is used for PIMD with $\mathcal{C}_a$ replacing $\mathcal{C}$ in the dynamics, and the covariance matrix $(m/\beta)\mathcal{C}_a$ replacing $((m + 1)/\tau)\mathcal{C}$ in the velocity randomization step. First, split (3.5) with $\boldsymbol{b}(\boldsymbol{x}) = \boldsymbol{x} - \mathcal{C}\nabla G_m(\boldsymbol{x})$ into

$$(A): \quad \dot{\boldsymbol{q}}_t = \boldsymbol{v}_t, \quad \dot{\boldsymbol{v}}_t = -\boldsymbol{q}_t \tag{3.23}$$

$$(B): \quad \dot{\boldsymbol{q}}_t = 0, \quad \dot{\boldsymbol{v}}_t = -\mathcal{C}\nabla G_m(\boldsymbol{q}_t) \tag{3.24}$$

with corresponding flows explicitly given by

$$\varphi_t^{(A)}(\boldsymbol{q}_0, \boldsymbol{v}_0) = (\cos(t)\boldsymbol{q}_0 + \sin(t)\boldsymbol{v}_0, -\sin(t)\boldsymbol{q}_0 + \cos(t)\boldsymbol{v}_0)$$

$$\varphi_t^{(B)}(\boldsymbol{q}_0, \boldsymbol{v}_0) = (\boldsymbol{q}_0, \boldsymbol{v}_0 - t\mathcal{C}\nabla G_m(\boldsymbol{q}_0)).$$

Given a time step size $\Delta t > 0$, and using these exact solutions, a Δt step of the symmetric splitting integrator we use is given by

$$\psi_{\Delta t} = \varphi_{\Delta t/2}^{(B)} \circ \varphi_{\Delta t}^{(A)} \circ \varphi_{\Delta t/2}^{(B)}. \tag{3.25}$$

In order to mitigate the effect of periodicities or near periodicities in the underlying dynamics, we choose the number of integration steps to be geometrically distributed with mean $T/\Delta t$. The idea of duration randomization has a long history [10,11,13,19, 21,51]. The initial velocity is taken to be an md-dimensional standard normal vector with covariance matrix $((m+1)/\tau)\mathcal{C}$ and a Metropolis accept/reject step is added to ensure the algorithm leaves invariant μ_m [11,70]. In summary, we use the following transition step in the simulations.

Algorithm 3.1 *(Numerical Randomized pHMC)* Denote by $T > 0$ the duration parameter and let $\psi_{\Delta t}$ be the time integrator described in (3.25). Given the current state of the chain $\boldsymbol{x} \in \mathbb{R}^{md}$, the algorithm outputs the next state of the chain $X \in \mathbb{R}^{md}$ as follows.

Step 1 Generate a d-dimensional random vector $\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{0}, ((m+1)/\tau)\mathcal{C})$.
Step 2 Generate a geometric random variable k supported on the set $\{1, 2, 3, ...\}$ with mean $T/\Delta t$.
Step 3 Output $X = \gamma\widetilde{\boldsymbol{q}}_k + (1-\gamma)\boldsymbol{x}$ where $(\widetilde{\boldsymbol{q}}_k, \widetilde{\boldsymbol{v}}_k) = \psi_{\Delta t}^k(\boldsymbol{x}, \boldsymbol{\xi})$, and given $\boldsymbol{\xi}$ and k, γ is a Bernoulli random variable with parameter α defined as

$$\alpha = \min\{1, \exp\left(-[\mathcal{E}(\widetilde{\boldsymbol{q}}_k, \widetilde{\boldsymbol{v}}_k) - \mathcal{E}(\boldsymbol{x}, \boldsymbol{\xi})]\right)\}$$

where $\mathcal{E}(\boldsymbol{x}, \boldsymbol{v}) = (1/2)\langle \boldsymbol{v}, \mathcal{C}^{-1}\boldsymbol{v}\rangle + U_m(\boldsymbol{x}) + (1/2)\langle \boldsymbol{x}, \mathcal{C}^{-1}\boldsymbol{x}\rangle$.

We stress that $\boldsymbol{\xi}$ and k from (Step 1) and (Step 2) are mutually independent and independent of the state of the Markov chain associated to pHMC. We pick the time step size Δt of the integrator sufficiently small to ensure that 99% of proposal moves are accepted on average in (Step 3).

Realizations of the coupling process are shown in Fig. 2. We chose parameters only for visualization purposes. The different components of the coupling are shown as different color dots. The insets of the figures show the distance between the components of the coupling as a function of the number of steps.

Figure 3 shows the average time after which the distance between the components of the coupling is for the first time within 10^{-12}. To produce this figure, we generated
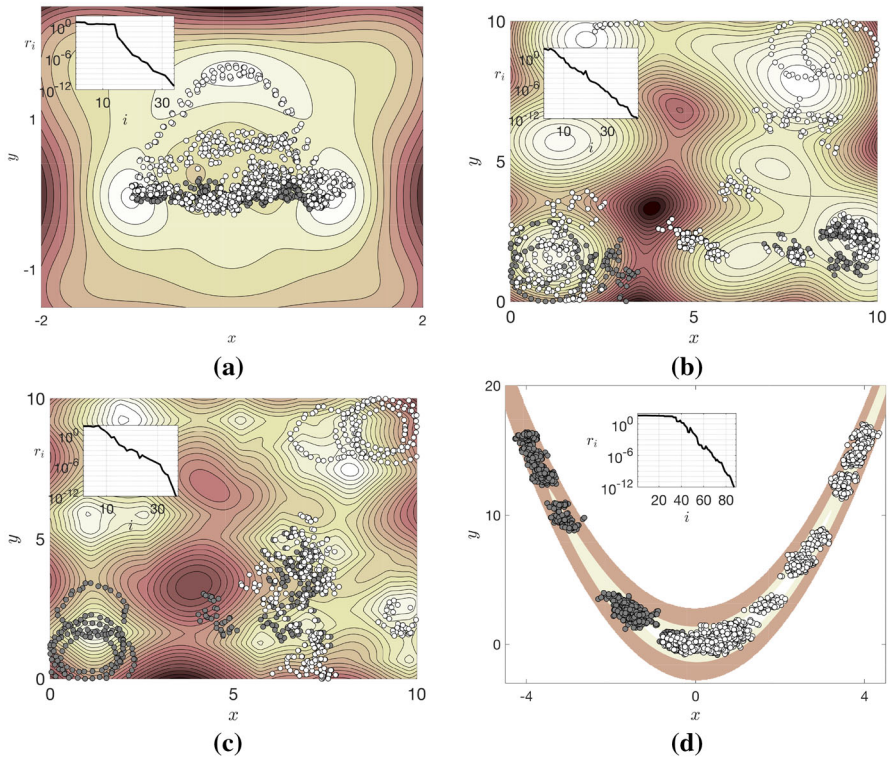
Fig. 2 Realizations of Two-Scale Coupling. This figure illustrates realizations of the coupling with $T = \gamma^{-1}$. The different components of the coupling are shown as different color dots. A contour plot of the underlying potential energy function is shown in the background. The inset plots the distance r_i between the components of the coupling as a function of the step index i. The simulation is terminated when this distance first reaches 10^{-12}. In (a), (b), (c), and (d), this occurs in 34, 44, 38, and 88 steps, respectively

one hundred samples of the coupled process for one hundred different values of the duration parameter T. As indicated in the figure legends, the coupling parameter γ is set equal to either:

- $\gamma = 0$ which corresponds to a synchronous coupling of the initial velocities;
- $\gamma = T^{-1}$ which corresponds to the optimal coupling of the initial velocities when $b(x) = 0$; and,
- $\gamma = \cot(T)$ which corresponds to the optimal coupling of the initial velocities when $b(x) = -x$.

4 A priori bounds

In this section we gather several bounds for the dynamics and for the coupling that will be crucial in the proof of our main result.
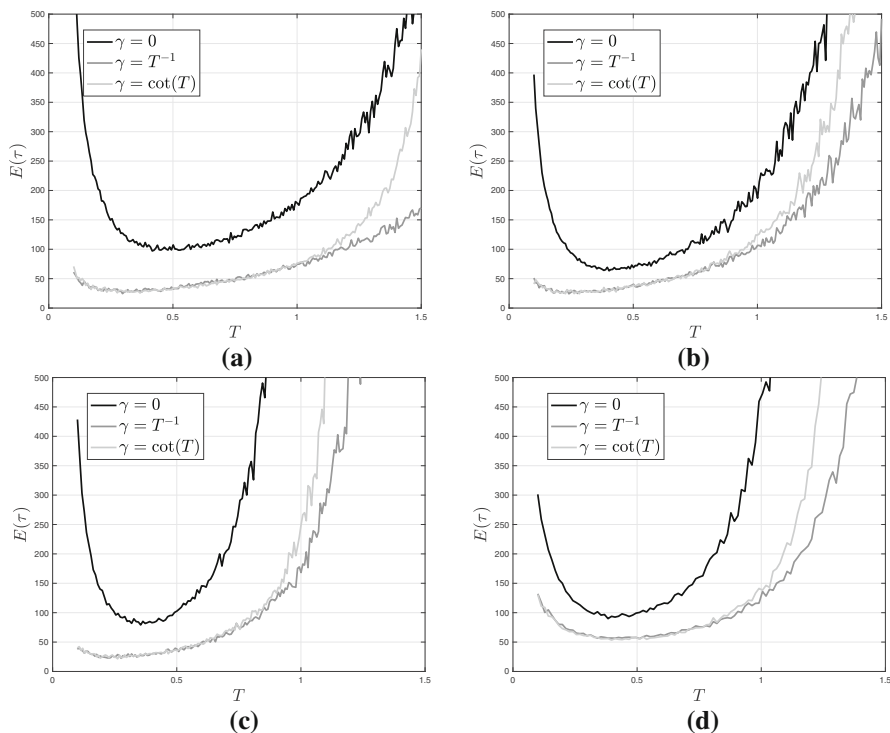
Fig. 3 Mean Coupling Times. This figure illustrates the average of the random time τ after which the distance between the components of the coupling is for the first time within 10^{-8}. The estimated average is plotted as a function of the duration T of the Hamiltonian dynamics for $\gamma = 0$ (black), $\gamma = T^{-1}$ (gray), and $\gamma = \cot(T)$ (light gray). In all cases, note that the minimum of the function is smaller and occurs at a smaller value of T when $\gamma = T^{-1}$ or $\gamma = \cot(T)$, and the difference between the minima for $\gamma = T^{-1}$ and $\gamma = \cot(T)$ is slight, because these minima occur at $T \leq 1$ where $\cot(T) \approx T^{-1}$

4.1 Bounds for the dynamics

In the following, we assume throughout that Assumption 2.6 is satisfied, and

$$Lt^2 \leq 1. \tag{4.1}$$

Recall that $\phi_t = (q_t, v_t)$ denotes the flow of (2.9). With the exception of using a different norm, the proofs of Lemmas 4.1 and 4.2 below are identical to the proofs of Lemmas 3.1 and 3.2 in Ref. [9] and therefore not repeated.

Lemma 4.1 *For any $x, v \in \mathcal{H}^s$,*

$$\sup_{r \leq t} |q_r(x, v) - (x + rv)|_s \leq Lt^2 \max(|x|_s, |x + tv|_s), \quad and \tag{4.2}$$

$$\sup_{r \leq t} |v_r(x, v) - v|_s \leq Lt \sup_{r \leq t} |q_r(x, v)|_s$$

$$\leq Lt(1 + Lt^2) \max(|x|_s, |x + tv|_s). \tag{4.3}$$

In particular,

$$\sup_{r \le t} |q_r(x, v)|_s \le 2 \max(|x|_s, |x + tv|_s), \quad \text{and} \tag{4.4}$$

$$\sup_{r \le t} |v_r(x, v)|_s \le |v|_s + 2Lt \max(|x|_s, |x + tv|_s). \tag{4.5}$$

Lemma 4.2 *For any $x, y, u, v \in \mathcal{H}^s$,*

$$\sup_{r \le t} |q_r(x, u) - q_r(y, v) - (x - y) - r(u - v)|_s$$

$$\le Lt^2 \max \left(|x - y|_s, |(x - y) + t(u - v)|_s \right), \quad \text{and} \tag{4.6}$$

$$\sup_{r \le t} |v_r(x, u) - v_r(y, v) - (u - v)|_s \le Lt \sup_{r \le t} |q_r(x, u) - q_r(y, v)|_s$$

$$\le Lt(1 + Lt^2) \max \left(|x - y|_s, |(x - y) + t(u - v)|_s \right). \tag{4.7}$$

In particular,

$$\sup_{r \le t} |q_r(x, u) - q_r(y, v)|_s \le (1 + Lt^2) \max(|x - y|_s, |(x - y) + t(u - v)|_s). \tag{4.8}$$

Lemma 4.1 is used in the proof of the Foster–Lyapunov drift condition in Lemma 2.7. Lemma 4.2 is used in the proof of Lemma 4.3 below.

4.2 Bounds related to two-scale coupling

The following lemma is used in the proof of Theorem 2.10 to obtain a contraction for the two-scale coupling when the distance between the components of the coupling is sufficiently small, i.e., $\|x - y\|_\alpha < R$.

Lemma 4.3 *Suppose that $\gamma > 0$ and $t > 0$ satisfy $\gamma t \le 1$ and $Lt^2 \le 1/4$. Then for any $x, y, u, v \in \mathcal{H}^s$ such that $v^h = u^h$ and $v^\ell = u^\ell + \gamma(x^\ell - y^\ell)$, we have*

$$(i) \left| \tilde{\mathcal{C}}^{-\frac{1}{2}} \mathcal{C}^{\frac{s}{2}} (q_t^\ell(x, u) - q_t^\ell(y, v)) \right|_s \le \left(1 - \gamma t + \frac{5}{8} \frac{\sigma_{max}}{\sigma_{min}} Lt^2 \right) \left| \tilde{\mathcal{C}}^{-\frac{1}{2}} \mathcal{C}^{\frac{s}{2}} (x^\ell - y^\ell) \right|_s$$

$$+ \frac{5}{8} \sigma_{max} Lt^2 \left| x^h - y^h \right|_s$$

$$(ii) \left| q_t^h(x, u) - q_t^h(y, v) \right|_s \le \left(1 - \frac{1}{4} t^2 \right) \left| x^h - y^h \right|_s + \frac{1}{4} \sigma_{min}^{-1} t^2 \left| \tilde{\mathcal{C}}^{-\frac{1}{2}} \mathcal{C}^{\frac{s}{2}} (x^\ell - y^\ell) \right|_s$$

Proof Let $\mathcal{G}(x) = b(x) + x$, $z_t = q_t(x, u) - q_t(y, v)$, and $w_t = \frac{dz_t}{dt}$. By (2.9),

$$\frac{d}{dt} z_t^\ell = w_t^\ell, \quad \frac{d}{dt} w_t^\ell = b^\ell(q_t(x, u)) - b^\ell(q_t(y, v)),$$

$$\frac{d}{dt} z_t^h = w_t^h, \quad \frac{d}{dt} w_t^h = -z_t^h + \mathcal{G}^h(q_t(x, u)) - \mathcal{G}^h(q_t(y, v)),$$

with $w_0 = -\gamma z_0^\ell$. These are second order linear ordinary differential equations, perturbed by a nonlinearity. A variation of constants ansatz shows that they are equivalent to the equations

$$z_t^\ell = (1 - \gamma t)z_0^\ell + \int_0^t (t - r)\left(b^\ell(q_r(x, u)) - b^\ell(q_r(y, v))\right) dr,$$

$$z_t^h = \cos(t)z_0^h + \int_0^t \sin(t - r)\left(\mathcal{G}^h(q_r(x, u)) - \mathcal{G}^h(q_r(y, v))\right) dr.$$

Since $t^2 \leq Lt^2 \leq 1/4$, $\gamma t \leq 1$, and by Assumption 2.6 (B1) and (B2) and (2.20),

$$\left|\widetilde{\mathcal{C}}^{-\frac{1}{2}}\mathcal{C}^{\frac{s}{2}}z_t^\ell\right|_s \leq (1 - \gamma t)\left|\widetilde{\mathcal{C}}^{-\frac{1}{2}}\mathcal{C}^{\frac{s}{2}}z_0^\ell\right|_s + \sigma_{max}L\int_0^t (t - r)\,|z_r|_s\,dr,$$

$$\leq (1 - \gamma t)\left|\widetilde{\mathcal{C}}^{-\frac{1}{2}}\mathcal{C}^{\frac{s}{2}}z_0^\ell\right|_s + \sigma_{max}\frac{Lt^2}{2}\sup_{r \leq t}|z_r|_s, \tag{4.9}$$

$$\left|z_t^h\right|_s \leq |\cos(t)|\left|z_0^h\right|_s + \frac{1}{3}\int_0^t |\sin(t - r)|\,|z_r|_s\,dr,$$

$$\leq \left(1 - \frac{t^2}{2} + \frac{t^4}{24}\right)\left|z_0^h\right|_s + \frac{t^2}{6}\sup_{r \leq t}|z_r|_s, \tag{4.10}$$

Here in (4.10) we used that $t \leq 1/2$, and hence $|\cos(t)| \leq 1 - (1/2)t^2 + (1/24)t^4$, and $|\sin(t - r)| \leq t$. Since $w_0 = -\gamma z_0^\ell$ and $\gamma t \leq 1$, Lemma 4.2 and (2.20) imply

$$\sup_{r \leq t}|z_r|_s \leq \frac{5}{4}\left(\left|z_0^\ell\right|_s + \left|z_0^h\right|_s\right) \leq \frac{5}{4}\left(\sigma_{min}^{-1}\left|\widetilde{\mathcal{C}}^{-\frac{1}{2}}\mathcal{C}^{\frac{s}{2}}z_0^\ell\right|_s + \left|z_0^h\right|_s\right).$$

Inserting this estimate into (4.9) and (4.10), and again using $t^2 \leq 1/4$ yields,

$$\left|\widetilde{\mathcal{C}}^{-\frac{1}{2}}\mathcal{C}^{\frac{s}{2}}z_t^\ell\right|_s \leq \left(1 - \gamma t + \frac{5}{8}\frac{\sigma_{max}}{\sigma_{min}}Lt^2\right)\left|\widetilde{\mathcal{C}}^{-\frac{1}{2}}\mathcal{C}^{\frac{s}{2}}z_0^\ell\right|_s + \frac{5}{8}\sigma_{max}Lt^2\left|z_0^h\right|_s,$$

$$\left|z_t^h\right|_s \leq \left(1 - \frac{1}{4}t^2\right)\left|z_0^h\right|_s + \frac{1}{4}\sigma_{min}^{-1}t^2\left|\widetilde{\mathcal{C}}^{-\frac{1}{2}}\mathcal{C}^{\frac{s}{2}}z_0^\ell\right|_s \qquad \text{as required.}$$

$\square$

Recall that the two-scale coupling that we consider ensures that $\xi^\ell - \eta^\ell = -\gamma z^\ell$ with the maximal possible probability, where $z = x - y$. The following lemma enables us to control the probability that $\xi^\ell - \eta^\ell \neq -\gamma z^\ell$ for small distances $\|z\|_\alpha < R$.

Lemma 4.4 *For any choice of γ, $P[\xi^\ell - \eta^\ell \neq -\gamma z^\ell] \leq \left|\gamma\widetilde{\mathcal{C}}^{-\frac{1}{2}}\mathcal{C}^{\frac{s}{2}}z^\ell\right|_s / \sqrt{2\pi}$.*

Since $\widetilde{\mathcal{C}}\mathcal{C}^{-s}$ is a trace class operator on $\mathcal{H}^s$, note that $\widetilde{\mathcal{C}}^{-\frac{1}{2}}\mathcal{C}^{\frac{s}{2}}$ is not a bounded operator. Nonetheless, the bound appearing in Lemma 4.4 is finite because the operator $\widetilde{\mathcal{C}}^{-\frac{1}{2}}\mathcal{C}^{\frac{s}{2}}$ appearing in that bound only acts on z^ℓ, i.e., the n-dimensional projection of z onto the finite dimensional space $\mathcal{H}^{s,\ell}$.
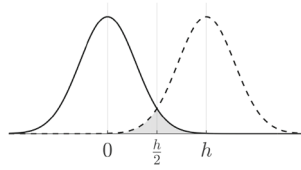
Fig. 4 The total variation distance between the one-dimensional normal distributions $\mathcal{N}(0, 1)$ and $\mathcal{N}(h, 1)$ equals one minus the area of the shaded region. Therefore, $d_{\mathrm{TV}}(\mathcal{N}(0, 1), \mathcal{N}(h, 1)) = 2\mathcal{N}(0, 1)[(0, h/2)]$

Proof Recall from (2.18) that $P[\xi^\ell - \eta^\ell \neq -\gamma z^\ell] = d_{\mathrm{TV}}(\mathcal{N}(0, \widetilde{\mathcal{C}}), \mathcal{N}(\gamma z^\ell, \widetilde{\mathcal{C}}))$. Let $\widetilde{\mathcal{C}}^\ell$ denote the restriction of $\widetilde{\mathcal{C}}$ to $\mathcal{H}^\ell$. Then since $z^\ell \in \mathcal{H}^\ell$, and by scale invariance of the total variation distance,

$$
\begin{aligned}
P[\xi^\ell - \eta^\ell \neq -\gamma z^\ell] &= d_{\mathrm{TV}}(\mathcal{N}(0, \widetilde{\mathcal{C}}^\ell), \mathcal{N}(\gamma z^\ell, \widetilde{\mathcal{C}}^\ell)) \\
&= d_{\mathrm{TV}}(\mathcal{N}(0, I_\ell), \mathcal{N}(\gamma \widetilde{\mathcal{C}}^{-1/2} z^\ell, I_\ell)) = d_{\mathrm{TV}}(\mathcal{N}(0, 1), \mathcal{N}(|\gamma \widetilde{\mathcal{C}}^{-1/2} z^\ell|, 1)) \\
&= 2\mathcal{N}(0, 1)\left[(0, |\gamma \widetilde{\mathcal{C}}^{-1/2} z^\ell/2|)\right] \leq |\gamma \widetilde{\mathcal{C}}^{-1/2} z^\ell|/\sqrt{2\pi},
\end{aligned}
$$

see Fig. 4 for the last equation. $\qquad\square$

Next we gather some elementary inequalities on the function f in (2.23) needed in the proof of Theorem 2.10. To state these results, let f'_- denote the left derivative of f which satisfies

$$
f'_-(r) = \begin{cases} e^{-ar} & \text{for } r \leq R, \\ 0 & \text{for } r > R. \end{cases}
$$

Lemma 4.5 *For all $r, \widetilde{r} > 0$, the function f in (2.23) satisfies:*

   (i) $f(\widetilde{r}) - f(r) \leq f'_-(r)(\widetilde{r} - r)$.
   (ii) $f(\widetilde{r}) - f(r) \leq a^{-1} f'_-(r)$.
   (iii) If $r \leq R$ then $\max(1, aR)e^{-\max(1,aR)} \leq r f'_-(r)/f(r) \leq 1$.

Proof Property (i) follows from the fact that f is concave. Since f is non-decreasing and constant for $r \geq R$, (ii) is trivially true in the cases $\widetilde{r} \leq r$ and $r \geq R$. In the case $r < \min(\widetilde{r}, R)$,

$$
f(\widetilde{r}) - f(r) = \frac{1}{a}\left(e^{-ar} - e^{-a\min(\widetilde{r}, R)}\right) \leq \frac{1}{a} f'_-(r).
$$

Combining these cases gives (ii). Let

$$
g(x) := \frac{x}{e^x - 1} \quad \text{so that} \quad \frac{r f'_-(r)}{f(r)} = g(ar).
$$

Property (iii) then follows because g decreases with x, $\lim_{x \to 0} g(x) = 1$ and $g(x) \geq \max(1, x)e^{-\max(1, x)}$. $\qquad\square$

5 Proof of Foster–Lyapunov drift condition

Before giving the proof of Lemma 2.7, here are some preparatory results. Using the shorthand notation $\varrho(t) = |q_t(x, \xi)|_s^2$ and $\varphi(t) = |v_t(x, \xi)|_s^2$, (2.9) implies

$$\varrho'(t) = 2\langle q_t(x, \xi), v_t(x, \xi)\rangle_s,$$
$$\varrho''(t) = 2\left(\varphi(t) + \langle q_t(x, \xi), b(q_t(x, \xi))\rangle_s\right),$$
$$\varphi'(t) = 2\langle v_t(x, \xi), b(q_t(x, \xi))\rangle_s.$$

Hence, by Assumption 2.6 (B1) and (B3), we have the differential inequalities

$$\varrho''(t) \le 2\left(\varphi(t) - K\varrho(t) + A\right), \qquad -\varrho''(t) \le 2L\varrho(t). \tag{5.1}$$

The following formula comes from two applications of integration by parts and is valid for any $k \in \mathbb{N}$ and for any twice differentiable function $g : \mathbb{R} \to \mathbb{R}$,

$$\int_0^T g(r)(T-r)^k dr = \frac{g(0)T^{k+1}}{k+1} + \frac{g'(0)T^{k+2}}{(k+1)(k+2)} + \int_0^T \frac{g''(r)(T-r)^{k+2}}{(k+1)(k+2)} dr. \tag{5.2}$$

We also require the following inequalities

$$\sup_{r \le T} \varrho(r) \le 4(|x|_s + T |\xi|_s)^2 \le 8(\varrho(0) + T^2\varphi(0)), \tag{5.3}$$

$$\sup_{r \le T} \varphi(r) \le 8L^2T^2\varrho(0) + 2(1 + 2LT^2)^2\varphi(0) \le 8L^2T^2\varrho(0) + \frac{9}{2}\varphi(0), \tag{5.4}$$

which follow from Lemma 4.1 and the assumption $LT^2 \le 1/48$.

Proof of Lemma 2.7 Apply in turn (5.1), then (5.2) with $g(r) = \varrho(r)$, and then (5.1) again, to obtain

$$E\left[|X'(x)|_s^2\right] = E\left[\rho(T)\right] = \rho(0) + T\,E[\rho'(0)] + \int_0^T (T-r)E\left[\varrho''(r)\right] dr$$

$$\le |x|_s^2 + 2\int_0^T (T-r)\,E\left[\varphi(r) - K\varrho(r) + A\right] dr$$

$$\le (1 - KT^2)\,|x|_s^2 + AT^2 + 2\int_0^T (T-r)E\left[\varphi(r)\right] dr$$

$$-2K \int_0^T \frac{(T-r)^3}{3!} E\left[\varrho''(r)\right] dr$$

$$\le (1 - KT^2)\,|x|_s^2 + AT^2 + T^2 E\left[\sup_{r \le T} \varphi(r)\right] + \frac{1}{6}KLT^4\,E\left[\sup_{r \le T} \varrho(r)\right]$$

$$\le \left(1 - KT^2 + 8L^2T^4 + \frac{4}{3}KLT^4\right)|x|_s^2 + AT^2 + \left(\frac{9}{2} + \frac{4}{3}KLT^4\right)T^2 E\,|\xi|_s^2$$

where in the last step we applied (5.3) and (5.4). Since we assume $LT^2 \leq (1/48)(K/L)$, note that $8L^2T^4 \leq (1/6)KT^2$, and since also $KT^2 \leq LT^2 \leq 1/4$,

$$E\left[\left|X'(x)\right|_s^2\right] \leq \left(1 - KT^2/2\right)|x|_s^2 + 5(\text{trace}(\widetilde{\mathcal{C}}\mathcal{C}^{-s}) + A)T^2$$

as required. □

6 Proofs of main results

Proof of Theorem 2.10 The parameters γ, a and ϵ have been chosen in (2.29), (2.30), and (2.31) such that the following conditions are satisfied:

$$\gamma T \leq 1, \tag{6.1}$$
$$\gamma R \leq 1/4, \tag{6.2}$$
$$aT = 1, \tag{6.3}$$
$$(\sigma_{max}/\sigma_{min})LT \leq \gamma/4, \tag{6.4}$$
$$\epsilon(A + \text{trace}(\widetilde{\mathcal{C}}\mathcal{C}^{-s})) = \max(1, R/T)\, e^{-\max(1, R/T)}/160. \tag{6.5}$$

Indeed, (6.1) and (6.2) hold by selection of γ in (2.29); (6.3) holds by selection of a in (2.30); (6.4) holds because (2.27) implies that

$$(\sigma_{max}/\sigma_{min})T \leq \min(T^{-1}/4, R^{-1}/16) = \gamma/4 \tag{6.6}$$

by selection of γ in (2.29); and (6.5) holds by selection of ϵ in (2.31).

Let $z = x - y$, $W = \xi - \eta$, $r = \|z\|_\alpha$, $R' = \|X'(x, y) - Y'(x, y)\|_\alpha$, $G = 1 + \epsilon(|x|_s^2 + |y|_s^2)$, $G' = 1 + \epsilon(|X'|_s^2 + |Y'|_s^2)$, $F = f(r)$ and $F' = f(R')$. We consider separately the cases where $r < R$ and $r \geq R$.

(i) Contractivity for $r < R$. Expand

$$E[F' - F] = I + II, \quad \text{where} \quad \begin{cases} I = E[F' - F; \ W^{\cdot} = -z^{\cdot}], \\ II = E[F' - F; \ W^{\cdot} \neq -z^{\cdot}]. \end{cases} \tag{6.7}$$

Let $Z_T = q_T(x, \xi) - q_T(y, \eta)$. By Lemma 4.5 (i), Lemma 4.3, (6.4) and (2.28),

$$\begin{aligned} I &\leq f'(r)E[R' - r; \ W^\ell = -\gamma z^\ell] \\ &= f'(r)E\left[\left|\widetilde{\mathcal{C}}^{-\frac{1}{2}}\mathcal{C}^{\frac{s}{2}} Z_T^\ell\right|_s - \left|\widetilde{\mathcal{C}}^{-\frac{1}{2}}\mathcal{C}^{\frac{s}{2}} z^\ell\right|_s + \alpha\left(\left|Z_T^h\right|_s - \left|z^h\right|_s\right); \ W^\ell = -\gamma z^\ell\right] \\ &\leq f'(r)\left((-\gamma T + \frac{5}{8}\frac{\sigma_{max}}{\sigma_{min}}LT^2 + \frac{1}{4}\sigma_{min}^{-1}\alpha T^2)\left|\widetilde{\mathcal{C}}^{-\frac{1}{2}}\mathcal{C}^{\frac{s}{2}} z^\ell\right|_s\right. \end{aligned}$$

$$+ \left(-\frac{1}{4} \alpha T^2 + \frac{5}{8} \sigma_{max} L T^2 \right) \left| z^h \right|_s \right) (1 - P[W^\ell \neq -\gamma z^\ell])$$

$$\leq -f'(r) \left(\frac{19}{32} \gamma T \left| \widetilde{\mathcal{C}}^{-\frac{1}{2}} \mathcal{C}^{\frac{s}{2}} z^\ell \right|_s + \frac{3}{32} \alpha T^2 \left| z^h \right|_s \right) (1 - P[W^\ell \neq -\gamma z^\ell]) \quad (6.8)$$

Moreover, by Lemmas 4.4 and 4.5 (ii),

$$P[W^\ell \neq -\gamma z^\ell] \leq \frac{\gamma R}{\sqrt{2\pi}} \leq \frac{1}{4} \frac{1}{\sqrt{2\pi}} < \frac{1}{10}, \quad \text{and} \quad (6.9)$$

$$\text{II} \leq a^{-1} f'(r) P[w^\ell \neq -\gamma z^\ell] \leq f'(r) \frac{2}{5} \gamma T \left| \widetilde{\mathcal{C}}^{-\frac{1}{2}} \mathcal{C}^{\frac{s}{2}} z^\ell \right|_s, \quad (6.10)$$

where in (6.9) we used (6.2) and in (6.10) we used (6.3) and $\sqrt{2\pi} > 5/2$.
Inserting (6.8), (6.9), and (6.10) into (6.7), and using Lemma 4.5 (iii), gives

$$\begin{aligned}
E[F' - F] &\leq f'(r) \left(-\tfrac{1}{8} \gamma T \left| \widetilde{\mathcal{C}}^{-\frac{1}{2}} \mathcal{C}^{\frac{s}{2}} z^\ell \right|_s - \tfrac{1}{12} \alpha T^2 \left| z^h \right|_s \right) \\
&\leq -\tfrac{1}{12} \min(\gamma T, T^2) \, r f'(r) \\
&\leq -\tfrac{1}{12} \min(1, T/(4R), T^2) \max(1, R/T) \, e^{-\max(1, R/T)} \, f(r) \\
&\leq -c_1 F. \quad (6.11)
\end{aligned}$$

Here we have introduced $c_1 := (1/12) T^2 \max(1, R/T) e^{-\max(1, R/T)}$, and we have used (2.29), (6.3), and the fact that $T^2 \leq \min(1, T/(4R))$ by (6.6).
Furthermore, by Lemma 2.7,

$$E[G' - G] \leq 10\epsilon (A + \text{trace}(\widetilde{\mathcal{C}} \mathcal{C}^{-s})) T^2 \leq (3/4) c_1 G, \quad (6.12)$$

where in the last step we eliminated ϵ using (6.5).
The Cauchy-Schwarz inequality, (6.11) and (6.12) now imply

$$\begin{aligned}
E[\rho(X', Y')] = E[\sqrt{F'G'}] &\leq \sqrt{E[F']} \sqrt{E[G']} \quad (6.13) \\
&\leq \sqrt{(1 - c_1) F} \sqrt{(1 + 3c_1/4) G} \leq \sqrt{1 - c_1/4} \sqrt{FG} \\
&\leq \exp(-c_1/8) \, \rho(x, y), \quad (6.14)
\end{aligned}$$

where in the last step we used $1 - \mathsf{x} \leq e^{-\mathsf{x}}$ with $\mathsf{x} = c_1/4$.

(ii) *Contractivity for* $r \geq R$. In this case, by (2.24) and (2.26), we have $|x|_s^2 + |y|_s^2 \geq 40(A + \text{trace}(\widetilde{\mathcal{C}} \mathcal{C}^{-s}))/K$, and we can apply the Foster–Lyapunov drift condition in (2.25) to (6.13) to obtain

$$E[\rho(X', Y')] \leq \sqrt{F} \sqrt{E[G']} \leq \sqrt{F} \left(1 + \epsilon \left(1 - \frac{KT^2}{4} \right) (|x|_s^2 + |y|_s^2) \right)^{1/2}$$

$$\leq \sqrt{F} \left(1 - \frac{5}{2} \epsilon (A + \text{trace}(\widetilde{\mathcal{C}} \mathcal{C}^{-s})) T^2 + \epsilon \left(1 - \frac{KT^2}{8} \right) (|x|_s^2 + |y|_s^2) \right)^{1/2}$$

$$\leq (1 - c_2)^{1/2} \, \rho(x, y) \, \leq \, \exp\left(-c_2/2\right) \rho(x, y) \tag{6.15}$$

where $c_2 := \min\left(KT^2/8, \, T^2 \max(1, R/T)e^{-\max(1,R/T)}/64\right)$.

(iii) Global Contraction. Let $c := \min(c_1/8, c_2/2) = c_2/2$. By combining the bounds in (6.14) and (6.15), we see that for any $x, y \in \mathcal{H}^s$,

$$E[\rho(X', Y')] \leq e^{-c} \rho(x, y).$$

$\square$

Proof of Corollary 2.12 The Wasserstein contraction in (2.34) follows directly from Theorem 2.10, see e.g. [9, Corollary 2.8] for a similar result. The bound in (2.35) then follows from (2.34) by comparing ρ to the metric on $\mathcal{H}^s$. Indeed, recall that by (2.23) and (2.30), $f(r) = (1 - e^{-\min(r,R)/T})T$. Let $x, y \in \mathcal{H}^s$, and let $r = \|\|x - y\|\|_\alpha$. Suppose first that $r \leq T$. Then $f(r) \geq r/e$, and thus

$$\rho(x, y) \, \geq \, \sqrt{f(r)} \, \geq \, \sqrt{r/e} \, \geq \, r/\sqrt{eT} \, \geq \, |x - y|_s \min(\sigma_{min}, \alpha)/\sqrt{eT}$$

where in the last step we used (2.21). Now suppose that $r > T$. Then since also $R \geq T$ by assumption, $f(r) \geq (1 - e^{-1})T$, and thus we obtain

$$\rho(x, y) \, \geq \, \sqrt{(1 - e^{-1})T\epsilon}\sqrt{|x|_s^2 + |y|_s^2} \, \geq \, \sqrt{(1 - e^{-1})T\epsilon/2} \, |x - y|_s \, .$$

Combining both cases and noting that by (2.28), $\alpha \geq \sigma_{min}$, we see that

$$|x - y|_s \, \leq \, \max\left(\sqrt{eT}/\sigma_{min}, \sqrt{2}/\sqrt{(1 - e^{-1})T\epsilon}\right) \rho(x, y) \quad \text{for all } x, y,$$

which implies an analogue bound for the corresponding Wasserstein distances $\mathcal{W}^{s,1}$ and $\mathcal{W}_\rho$. Conversely, since $f(r) \leq T$ for all r,

$$\rho(x, y) \, \leq \, \sqrt{T}\sqrt{1 + \epsilon|x|_s^2 + \epsilon|y|_s^2} \, \leq \, \sqrt{T}(1 + \sqrt{\epsilon}|x|_s + \sqrt{\epsilon}|y|_s) \quad \text{for all } x, y.$$

Therefore, with C defined by (2.36), we obtain

$$\mathcal{W}^{s,1}(\nu\pi^k, \mu) \, = \, \mathcal{W}^{s,1}(\nu\pi^k, \mu\pi^k) \, \leq \, CT^{-1/2} \, \mathcal{W}_\rho(\nu\pi^k, \mu\pi^k)$$
$$\leq \, CT^{-1/2}e^{-ck} \, \mathcal{W}_\rho(\nu, \mu) \, \leq \, C\left(1 + \sqrt{\epsilon}M_1(\nu) + \sqrt{\epsilon}M_1(\mu)\right)e^{-ck}$$

for all $k \in \mathbb{N}$ and all probability measures ν on $\mathcal{H}^s$. Finally, by Lemma 2.7 and (2.31), we have $\sqrt{\epsilon}M_1(\mu) \leq (1/4)K^{-1/2}e^{-R/(2T)}$. $\square$

7 Proofs of results from Sect. 3 (applications)

7.1 Proofs of results for TPS

To prove Theorem 3.2, we compare the eigenvalues of $\mathcal{C}$ to $\mathcal{C}$. Note that these eigenvalues each have multiplicity d, and to account for this, define the index function $\varphi(k, j) = d(k-1) + j$. Then the eigenvalues of $\mathcal{C}$ are

$$\lambda_{\varphi(k,j)} = \left(\frac{\tau}{k\pi}\right)^2, \quad k \in \mathbb{N}, \quad 1 \le j \le d, \tag{7.1}$$

and the eigenvalues of $\mathcal{C}$ are

$$\mathbf{\Lambda}_{\varphi(k,j)} = \lambda_{\varphi(k,j)} \left(\frac{\theta_k}{\sin(\theta_k)}\right)^2, \quad \theta_k := \frac{k\pi}{2(m+1)}, \quad 1 \le k \le m, \ 1 \le j \le d. \tag{7.2}$$

The following lemma helps estimate the error of the eigenvalues of the approximation $\mathcal{C}$ relative to those of $\mathcal{C}$.

Lemma 7.1 *For any $m \in \mathbb{N}$, for all $1 \le k \le m$, and for $1 \le j \le d$,*

(E1) $|\mathbf{\Lambda}_{\varphi(k,j)} - \lambda_{\varphi(k,j)}| = \mathbf{\Lambda}_{\varphi(k,j)} - \lambda_{\varphi(k,j)} \le \lambda_{\varphi(k,j)} \dfrac{k^2\pi^2}{6(m+1)^2} = \dfrac{\tau^2}{6(m+1)^2},$

(E2) $\left(\dfrac{\mathbf{\Lambda}_1}{\mathbf{\Lambda}_{\varphi(k,j)}}\right)^{1/2} \le \left(\dfrac{\lambda_1}{\lambda_{\varphi(k,j)}}\right)^{1/2} \left(1 + \dfrac{\pi^2}{16(m+1)^2}\right).$

Proof This lemma is an easy consequence of the elementary inequalities

$$\frac{1}{2} < 1 - \frac{\theta^2}{6} < \frac{\sin(\theta)}{\theta} < 1 \quad \text{and} \quad 1 < \frac{\theta}{\sin(\theta)} < 1 + \frac{\theta^2}{4} < \frac{5}{3} \tag{7.3}$$

which are valid for $0 < \theta < \pi/2$. Indeed, (7.1), (7.2) and (7.3) imply

$$\begin{aligned}
|\mathbf{\Lambda}_{\varphi(k,j)} - \lambda_{\varphi(k,j)}| &= \mathbf{\Lambda}_{\varphi(k,j)} - \lambda_{\varphi(k,j)} \\
&= \lambda_{\varphi(k,j)} \left(\left(\frac{\theta_k}{\sin(\theta_k)}\right)^2 - 1\right) \\
&= \lambda_{\varphi(k,j)} \left(\frac{\theta_k}{\sin(\theta_k)} + 1\right)\left(\frac{\theta_k}{\sin(\theta_k)} - 1\right) \\
&\le \lambda_{\varphi(k,j)} \frac{2}{3}\theta_k^2 = \lambda_{\varphi(k,j)} \frac{k^2\pi^2}{6(m+1)^2} = \frac{\tau^2}{6(m+1)^2}
\end{aligned}$$

as required for (E1). For (E2), we use (7.2) to write

$$\left(\frac{\mathbf{\Lambda}_1}{\mathbf{\Lambda}_{\varphi(k,j)}}\right)^{1/2} = \left(\frac{\lambda_1}{\lambda_{\varphi(k,j)}}\right)^{1/2} \frac{\sin(\theta_k)}{\theta_k} \frac{\theta_1}{\sin(\theta_1)}.$$

Hence, by (7.3),

$$\left(\frac{\mathbf{\Lambda}_1}{\mathbf{\Lambda}_{\varphi(k,j)}}\right)^{1/2} \le \left(\frac{\lambda_1}{\lambda_{\varphi(k,j)}}\right)^{1/2} \left(1 + \frac{\theta_1^2}{4}\right) = \left(\frac{\lambda_1}{\lambda_{\varphi(k,j)}}\right)^{1/2} \left(1 + \frac{\pi^2}{16(m+1)^2}\right)$$

as required for (E2). □

Proof of Theorem 3.2 This result is an application of Corollary 2.12. Since $\mathcal{C}$ is a finite-dimensional matrix, Assumption 2.1 holds for $\mathcal{C}$ with $s = 0$, and since we choose $\widetilde{\mathcal{C}} = \mathcal{C}$, Assumption 2.5 also holds with $s = 0$. Therefore, to apply Corollary 2.12, it suffices to check that: (i) Assumption 2.6 holds with dimension-free constants L, n, K, and A, (ii) the dimension-free R defined in (3.6) satisfies condition (2.24), and (iii) the dimension-free condition (3.10) on the duration T implies (2.27) holds. We then invoke Corollary 2.12 to conclude convergence in the standard L^1 Wasserstein distance.

Verify Assumption 2.6 (B1)-(B3). For (B1), note that

$$\begin{aligned}|\boldsymbol{b}(\boldsymbol{x}) - \boldsymbol{b}(\boldsymbol{y})| &\le (1 + \mathbf{\Lambda}_1 L_G)|\boldsymbol{x} - \boldsymbol{y}| \\ &\le (1 + 2\lambda_1 L_G)|\boldsymbol{x} - \boldsymbol{y}| = (1 + 2(\tau^2/\pi^2)L_G)|\boldsymbol{x} - \boldsymbol{y}|\end{aligned}$$

where in the last step we used Lemma 7.1 (E1) which implies that $\mathbf{\Lambda}_1 - \lambda_1 \le \lambda_1\pi^2/(6(m+1)^2) \le \lambda_1$ since $m \ge 1$. Thus, (B1) holds with $L = 1 + \kappa$ since $\kappa = 2(\tau^2/\pi^2)L_G$. For (B2), since $n = m_\ell d = \varphi(m_\ell, d)$, $n + 1 = \varphi(m_\ell + 1, 1)$ and

$$\begin{aligned}|\boldsymbol{b}^h(\boldsymbol{x}) + \boldsymbol{x}^h - \boldsymbol{b}^h(\boldsymbol{y}) - \boldsymbol{y}^h| &\le \mathbf{\Lambda}_{\varphi(m_\ell+1,1)} L_G|\boldsymbol{x} - \boldsymbol{y}| \\ &\le \lambda_{\varphi(m_\ell+1,1)} L_G|\boldsymbol{x} - \boldsymbol{y}| + (\mathbf{\Lambda}_{\varphi(m_\ell+1,1)} - \lambda_{\varphi(m_\ell+1,1)})L_G|\boldsymbol{x} - \boldsymbol{y}| \\ &\le 2\lambda_{\varphi(m_\ell+1,1)} L_G|\boldsymbol{x} - \boldsymbol{y}| \le (1/3)|\boldsymbol{x} - \boldsymbol{y}|\end{aligned}$$

where in the second to last step we used

$$\mathbf{\Lambda}_{\varphi(m_\ell+1,1)} - \lambda_{\varphi(m_\ell+1,1)} \le \lambda_{\varphi(m_\ell+1,1)}(m_\ell + 1)^2\pi^2/(6(m+1)^2) \le \lambda_{\varphi(m_\ell+1,1)}$$

which follows from Lemma 7.1 (E1) since $m \ge (m_\ell + 1)\pi/2$, and in the last step, we used that $m_\ell + 1 \ge \sqrt{6L_G}\tau/\pi$. Hence, (B2) holds with $n = m_\ell d$. For (B3),

$$\begin{aligned}\langle \boldsymbol{x}, \boldsymbol{b}(\boldsymbol{x})\rangle &\le -|\boldsymbol{x}|^2 + |\boldsymbol{x}||\mathcal{C}\nabla G_m(\boldsymbol{x})| \le -(1/2)|\boldsymbol{x}|^2 + (1/2)|\mathcal{C}\nabla G_m(\boldsymbol{x})|^2 \\ &\le -(1/2)|\boldsymbol{x}|^2 + (1/2)\mathbf{\Lambda}_1^2 M_G^2\tau \le -(1/2)|\boldsymbol{x}|^2 + \lambda_1^2\tau M_G^2\end{aligned}$$

where in the last step we used

$$\mathbf{\Lambda}_1^2 - \lambda_1^2 = 2\lambda_1(\mathbf{\Lambda}_1 - \lambda_1) + (\mathbf{\Lambda}_1 - \lambda_1)^2 \le \lambda_1^2$$

which follows from (E1) since $m \ge 1$. Thus, (B3) holds with $K = 1/2$ and $A = \lambda_1^2\tau M_G^2 = (\tau^5/\pi^4)M_G^2$. To summarize, Assumption 2.6 holds with dimension-

independent constants $L = 1 + \kappa$, $n = m_\ell d$ where $m_\ell = \lfloor \sqrt{3\kappa} \rfloor$, $K = 1/2$, and $A = (\tau^5/\pi^4)M_G^2$.

Verify Conditions (2.24) & (2.27). To show that R defined in (3.6) satisfies condition (2.24) and that condition (3.10) on the duration parameter implies condition (2.27), in this paragraph we gather some additional bounds. Since $m_\ell \leq \sqrt{3\kappa}$, we have

$$\sigma_{max} = \mathbf{\Lambda}_{\varphi(m_\ell, 1)}^{-1/2} = \lambda_{\varphi(m_\ell, 1)}^{-1/2} (\sin(\theta_{m_\ell})/\theta_{m_\ell}) \leq m_\ell \pi/\tau \leq \sqrt{6 L_G}, \qquad (7.4)$$

$$\sigma_{min}^{-1} = \mathbf{\Lambda}_1^{1/2} = \lambda_1^{1/2} + \lambda_1^{1/2} (\theta_1/\sin(\theta_1) - 1) \leq 2\lambda_1^{1/2} = 2\tau/\pi, \qquad (7.5)$$

where in (7.4) and (7.5) we used (7.3). Moreover, by Lemma 7.1 (E2),

$$\frac{\sigma_{max}}{\sigma_{min}} = \left(\frac{\mathbf{\Lambda}_1}{\mathbf{\Lambda}_{\varphi(m_\ell, 1)}}\right)^{1/2} \leq \left(\frac{\lambda_1}{\lambda_{\varphi(m_\ell, 1)}}\right)^{1/2} \left(1 + \frac{\pi^2}{16(m+1)^2}\right) \leq 2m_\ell \qquad (7.6)$$

since $m \geq 1$ and $\sqrt{\lambda_1/\lambda_{\varphi(m_\ell, 1)}} = m_\ell$, and by Lemma 7.1 (E1),

$$\text{trace}(\mathcal{C}) \leq \text{trace}(\mathcal{C}) + \sum_{i=1}^{md} (\mathbf{\Lambda}_i - \lambda_i) \leq \text{trace}(\mathcal{C}) + (d\tau^2/6)\, m/(m+1)^2$$

$$\leq 2\,\text{trace}(\mathcal{C}) = d\tau^2/3, \quad \text{since trace}(\mathcal{C}) = d\tau^2/6. \qquad (7.7)$$

Let $R_m = 8\sqrt{40}(A + \text{trace}(\mathcal{C}))^{1/2}\sigma_{max} L K^{-1/2}$ denote the RHS of (2.24). Then using (7.4), (7.7), $L = 1 + \kappa$, $K = 1/2$, and $A = M_G^2 \tau^5/\pi^4$, we have

$$R_m^2 \leq 128 \times 40 \times 6 L_G (1 + \kappa)^2 ((\tau^5/\pi^4) M_G^2 + \text{trace}(\mathcal{C}))$$

$$\leq 256 \times 20 \times \kappa (1 + \kappa)^2 \left(3(\tau^3/\pi^2) M_G^2 + d\pi^2\right)$$

$$\leq 256 \times 20\pi^2 \kappa (1 + \kappa)^2 ((\tau/\pi)^3 M_G^2 + d) = R^2$$

which implies that R defined in (3.6) satisfies (2.24). Moreover, by (7.6),

$$\frac{\sigma_{max}}{\sigma_{min}} L \leq 2m_\ell (1 + \kappa) \leq 2\sqrt{3\kappa}(1 + \kappa). \qquad (7.8)$$

Inserting (7.8) into the LHS and RHS of (2.27) gives (3.10). Thus, whenever T satisfies condition (3.10) then condition (2.27) holds.

Invoke Corollary 2.12. By Corollary 2.12 and using $K = 1/2$, as long as T satisfies (3.10),

$$\mathcal{W}^{0,1}(\nu_m \pi_m^k, \mu_m) \leq C_m e^{-ck}\left(1 + \sqrt{\epsilon_m} M_1(\nu_m) + (1/8)e^{-R/(2T)}\right) \qquad (7.9)$$

holds with the dimension-free rate in (3.7) and the constants:

$$C_m = \max(2T\sigma_{min}^{-1}, 23(A + \text{trace}(\mathcal{C}))^{1/2} e^{R/(2T)}),$$
$$\epsilon_m = (1/160)(A + \text{trace}(\mathcal{C}))^{-1} e^{-R/T}. \qquad (7.10)$$

These dimension-dependent constants can be upper bounded by dimension-free constants C and ϵ given in (3.8) and (3.9), by using $A = (\tau^5/\pi^4)M_G^2$, (7.5) and (7.7). Thus, (3.11) holds. □

7.2 Proofs of results for PIMD

To prove Theorem 3.5, we compare the eigenvalues of $\mathcal{C}_a$ to C_a. The leading eigenvalue of C_a has multiplicity d, while all of the other eigenvalues have multiplicity $2d$. If m is odd, then the leading eigenvalue of both $\mathcal{C}_a$ has multiplicity d, while all of the other eigenvalues of $\mathcal{C}_a$ have multiplicity $2d$. However, if m is even, then the trailing and leading eigenvalues of $\mathcal{C}_a$ have multiplicity d, while all of the other eigenvalues of of $\mathcal{C}_a$ have multiplicity $2d$. To account for these multiplicities, it is helpful to define the index function

$$\varphi(k, j) = \begin{cases} j & \text{if } k = 1, \ 1 \le j \le d, \\ 2d(k-2) + j + d & \text{if } k > 1, \ 1 \le j \le 2d. \end{cases}$$

For any $k \in \mathbb{N}$, the eigenvalues of C_a are

$$\lambda_{\varphi(k,j)} = \begin{cases} a^{-1} & \text{if } k = 1, \ 1 \le j \le d, \\ \dfrac{1}{a + \omega_k^2} & \text{if } k > 1, \ 1 \le j \le 2d. \end{cases} \tag{7.11}$$

For all $m \in \mathbb{N}$ and $k \in \{1, \ldots, \lceil (m+1)/2 \rceil\}$, the eigenvalues of $\mathcal{C}_a$ are

$$\Lambda_{\varphi(k,j)} = \begin{cases} a^{-1} & \text{if } k = 1, \ 1 \le j \le d, \\ \dfrac{1}{a + \omega_k^2 \sin^2(\theta_k)/\theta_k^2} & \begin{cases} \text{if } k > 1 \,\&\, k \ne \frac{m}{2} + 1, & 1 \le j \le 2d, \\ \text{if } k = \frac{m}{2} + 1, & 1 \le j \le d, \end{cases} \end{cases} \tag{7.12}$$

Here we have introduced

$$\theta_k := \frac{(k-1)\pi}{m} \quad \text{and} \quad \omega_k^2 := \frac{4(k-1)^2\pi^2}{\beta^2}. \tag{7.13}$$

Note that the definition in (7.12) includes odd or even values of m. The following lemma estimates the error of the eigenvalues of $\mathcal{C}_a$ relative to those of C_a.

Lemma 7.2 *For any $m \in \mathbb{N}$ and $k \in \{1, \ldots, \lceil (m+1)/2 \rceil\}$,*

(E1) $|\Lambda_{\varphi(k,1)} - \lambda_{\varphi(k,1)}| = \Lambda_{\varphi(k,1)} - \lambda_{\varphi(k,1)} \le \lambda_{\varphi(k,1)} 2\theta_k^2.$

(E2) $\left(\dfrac{\Lambda_1}{\Lambda_{\varphi(k,1)}} \right)^{1/2} \le \left(\dfrac{\lambda_1}{\lambda_{\varphi(k,1)}} \right)^{1/2}.$

Proof This lemma is an easy consequence of the inequalities in (7.3). For $k = 1$, (E1) and (E2) trivially hold since $\Lambda_{\varphi(k,1)} = \lambda_{\varphi(k,1)} = a^{-1}$. For $k > 1$, (7.12), (7.11), and

([7.3](#)) imply

$$
\begin{aligned}
|\mathbf{\Lambda}_{\varphi(k,j)} - \lambda_{\varphi(k,j)}| &= \mathbf{\Lambda}_{\varphi(k,j)} - \lambda_{\varphi(k,j)} \\
&= \left([\mathsf{a} + \omega_k^2 \sin^2(\theta_k)/\theta_k^2]^{-1} - \lambda_{\varphi(k,j)}\right) \\
&\leq \left([\mathsf{a} + \omega_k^2 - \omega_k^2\theta_k^2/3]^{-1} - \lambda_{\varphi(k,j)}\right) \quad \text{since } \sin^2(\theta_k)/\theta_k^2 \geq 1 - \theta_k^2/3 \\
&\leq \lambda_{\varphi(k,j)} \left([1 - \omega_k^2\theta_k^2/(3(\mathsf{a} + \omega_k^2))]^{-1} - 1\right) \\
&\leq \lambda_{\varphi(k,j)}\theta_k^2/3(1 - \theta_k^2/3)^{-1} \\
&\leq \lambda_{\varphi(k,j)}2\theta_k^2 \quad \text{as required for (E1).}
\end{aligned}
$$

For (E2) with $k > 1$, by ([7.12](#)), ([7.11](#)), and ([7.3](#)),

$$
\frac{\mathbf{\Lambda}_1}{\mathbf{\Lambda}_{\varphi(k,j)}} = 1 + \frac{\omega_k^2}{\mathsf{a}}\frac{\sin^2(\theta_k)}{\theta_k^2} \leq 1 + \frac{\omega_k^2}{\mathsf{a}} = \frac{\lambda_1}{\lambda_{\varphi(k,j)}}.
$$

Taking square roots of both sides then gives (E2). $\qquad\square$

Proof of Theorem 3.5 This proof is very similar to the Proof of Theorem [3.2](#) with some differences which are highlighted below.

Verify Assumption [2.6](#) *(B1)-(B3).* Since both $\mathcal{C}_\mathsf{a}$ and $\mathcal{C}_\mathsf{a}$ have leading eigenvalue a^{-1}, (B1) holds with $L = 1 + 6\mathsf{a}^{-1}L_G$. Similarly, (B3) holds with $K = 1/2$ and $A = (1/2)\beta\lambda_1^2 M_G^2 = (1/2)\beta\mathsf{a}^{-2}M_G^2$. Moreover, (B2) holds with $n = 2m_\ell d - d = \varphi(m_\ell, d)$, since $n + 1 = \varphi(m_\ell + 1, 1)$, $m_\ell \geq \sqrt{3L_G/2}(\beta/\pi)$, and $m \geq 2\pi m_\ell$.

Verify Conditions ([2.24](#)) & ([2.27](#)). By ([7.12](#)) and ([7.13](#)),

$$
\sigma_{min} = \mathsf{a}^{1/2}, \quad \sigma_{max} \leq \beta^{-1}\left(\beta^2\mathsf{a} + 4(m_\ell - 1)^2\pi^2\right)^{1/2} \leq \sqrt{\mathsf{a} + 6L_G}, \quad (7.14)
$$

since $m_\ell - 1 < \sqrt{3L_G/2}(\beta/\pi)$. Moreover, by Lemma [7.2](#) (E2),

$$
\frac{\sigma_{max}}{\sigma_{min}} \leq \left(1 + \frac{\omega_{m_\ell}^2}{\mathsf{a}}\right)^{1/2} = \left(1 + \frac{4(m_\ell - 1)^2\pi^2}{\beta^2\mathsf{a}}\right)^{1/2} \leq \left(1 + \frac{6L_G}{\mathsf{a}}\right)^{1/2}. \quad (7.15)
$$

Furthermore, by Lemma [7.2](#) (E1),

$$
\begin{aligned}
\text{trace}(\mathcal{C}_\mathsf{a}) &\leq \text{trace}(\mathcal{C}_\mathsf{a}) + 2d\sum_{k=1}^{\lceil(m+1)/2\rceil}(\mathbf{\Lambda}_{\varphi(k,1)} - \lambda_{\varphi(k,1)}) \\
&\leq \text{trace}(\mathcal{C}_\mathsf{a}) + d\beta^2 = \frac{d}{2\mathsf{a}} + \frac{d\beta}{4\sqrt{\mathsf{a}}}\left(1 + \frac{2}{e^{\sqrt{\mathsf{a}}\beta} - 1}\right) + d\beta^2 \\
&\leq 2d(\mathsf{a}^{-1} + \beta^2)
\end{aligned} \quad (7.16)
$$

where in the last step we used $1 + 2/(e^{2\mathsf{x}} - 1) < \mathsf{x} + \mathsf{x}^{-1}$ valid for all $\mathsf{x} > 0$.

Let $R_m = 8\sqrt{40}(A + \mathrm{trace}(\mathcal{C}))^{1/2}\sigma_{max}LK^{-1/2}$ denote the RHS of (2.24). Using $L = 1 + 6\mathsf{a}^{-1}L_G$, $K = 1/2$, $A = (1/2)(\beta/\mathsf{a}^2)M_G^2$, (7.14), and (7.16),

$$
\begin{aligned}
R_m^2 &\le 128 \times 40\mathsf{a}(1 + 6L_G\mathsf{a}^{-1})^3((1/2)\beta\mathsf{a}^{-2}M_G^2 + 2d(\beta^2 + \mathsf{a}^{-1})) \\
&\le 256 \times 20(1 + 6L_G\mathsf{a}^{-1})^3\left((1/2)\beta\mathsf{a}^{-1}M_G^2 + 2d(\beta^2\mathsf{a} + 1)\right) \\
&\le 256 \times 20(1 + \kappa)^3\left((1/2)\beta\mathsf{a}^{-1}M_G^2 + 2d(\beta^2\mathsf{a} + 1)\right) = R^2,
\end{aligned}
$$

which implies that R defined in (3.18) satisfies (2.24). Moreover, by (7.15),

$$
\left(\frac{\sigma_{max}}{\sigma_{min}}\right)L \le (1 + \kappa)^{3/2}. \tag{7.17}
$$

Inserting (7.17) into (2.27) gives (3.22).

Invoke Corollary 2.12. By Corollary 2.12, as long as T satisfies (3.22), (7.9) holds with the dimension-free rate in (3.19) and the constants in (7.10) with $\mathcal{C}$ replaced with $\mathcal{C}_\mathsf{a}$. Moreover, the dimension-dependent constants in (7.10) can be upper bounded by dimension-free constants C and ϵ given in (3.20) and (3.21), by using $A = (1/2)\beta M_G^2\mathsf{a}^{-2}$, (7.14) and (7.16). Thus, (3.11) holds for the transition kernel of (3.17). $\qquad\square$

References

1. Allen, M.P., Tildesley, D.J.: Computer Simulation of Liquids. Clarendon Press, Oxford (1987)
2. Beskos, A., Pillai, N.S., Roberts, G.O., Sanz-Serna, J.M., Stuart, A.M.: Optimal tuning of hybrid Monte-Carlo algorithm. Bernoulli **19**, 1501–1534 (2013)
3. Beskos, A., Pinski, F.J., Sanz-Serna, J.M., Stuart, A.M.: Hybrid Monte-Carlo on Hilbert spaces. Stoch. Proc. Appl. **121**(10), 2201–2230 (2011)
4. Beskos, A., Roberts, G., Stuart, A., Voss, J.: MCMC methods for diffusion bridges. Stoch. Dyn. **8**(03), 319–350 (2008)
5. Bogachev, V.I.: Gaussian Measures, vol. 62. American Mathematical Society, Providence (1998)
6. Bolhuis, P.G., Chandler, D., Dellago, C., Geissler, P.L.: Transition path sampling: throwing ropes over rough mountain passes, in the dark. Ann. Rev. Phys. Chem. **53**(1), 291–318 (2002)
7. Borggaard, J., Glatt-Holtz, N., Krometis, J.: A Bayesian Approach to Estimating Background Flows from a Passive Scalar, arXiv preprint arXiv:1808.01084 (2018)
8. Bou-Rabee, N.: Time integrators for molecular dynamics. Entropy **16**(1), 138–162 (2014)
9. Bou-Rabee, N., Eberle, A., Zimmer, R.: Coupling and Convergence for Hamiltonian Monte Carlo. Ann. Appl. Probab. (to appear) arXiv:1805.00452
10. Bou-Rabee, N., Sanz-Serna, J.M.: Randomized Hamiltonian Monte Carlo. Ann. Appl. Probab. **27**(4), 2159–2194 (2017)
11. Bou-Rabee, N., Sanz-Serna, J.M.: Geometric integrators and the Hamiltonian Monte Carlo method. Acta Numerica **27**, 113–206 (2018)
12. Butkovsky, O.: Subgeometric rates of convergence of Markov processes in the Wasserstein metric. Ann. Appl. Probab. **24**(2), 526–552 (2014)
13. Cancés, E., Legoll, F., Stoltz, G.: Theoretical and numerical comparison of some sampling methods for molecular dynamics. Math. Model. Numer. Anal. **41**, 351–389 (2007)
14. Chandler, D., Wolynes, P.G.: Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids. J. Chem. Phys. **74**(7), 4078–4095 (1981)
15. Craig, I.R., Manolopoulos, D.E.: Quantum statistics and classical mechanics: real time correlation functions from ring polymer molecular dynamics. J. Chem. Phys. **121**(8), 3368–3373 (2004)

16. Craig, I.R., Manolopoulos, D.E.: A refined ring polymer molecular dynamics theory of chemical reaction rates. J. Chem. Phys. **123**(3), 034102 (2005)
17. Craig, I.R., Manolopoulos, D.E.: Chemical reaction rates from ring polymer molecular dynamics. J. Chem. Phys. **122**(8), 084106 (2005)
18. Dashti, M., Stuart, A.M.: The Bayesian approach to inverse problems. In: Handbook of Uncertainty Quantification, pp. 311–428 (2017)
19. Duane, S.: Stochastic quantization versus the microcanonical ensemble: getting the best of both worlds. Nuclear Phys. B **257**, 652–662 (1985)
20. Duane, S., Kennedy, A.D., Pendleton, B.J., Roweth, D.: Hybrid Monte-Carlo. Phys. Lett. B **195**, 216–222 (1987)
21. Duane, S., Kogut, J.B.: The theory of hybrid stochastic algorithms. Nucl. Phys. B **275**(3), 398–420 (1986)
22. Durmus, A., Moulines, E., Saksman, E.: On the convergence of Hamiltonian Monte Carlo (April 2017). arXiv:1705.00166 [stat.CO]
23. Weinan, E., Li, D.: The Andersen thermostat in molecular dynamics. CPAM **61**, 96–136 (2008)
24. Eberle, A.: Error bounds for Metropolis-Hastings algorithms applied to perturbations of Gaussian measures in high dimensions. Ann. Appl. Probab. **24**(1), 337–377 (2014)
25. Eberle, A., Guillin, A., Zimmer, R.: Couplings and quantitative contraction rates for Langevin dynamics. Ann. Probab. **47**(4), 1982–2010 (2019)
26. Eberle, A., Guillin, A., Zimmer, R.: Quantitative Harris-type theorems for diffusions and Mckean–Vlasov processes. Trans. Am. Math. Soc. **371**(10), 7135–7173 (2019)
27. Feynman, R.P., Hibbs, A.R.: Quantum Mechanics and Path Integrals. McGraw-Hill, Cambridge (1965)
28. Foias, C., Prodi, G.: Sur le comportement global des solutions non-stationnaires des équations de Navier–Stokes en dimension 2. Rendiconti del Seminario Matematico della Università di Padova **39**, 1–34 (1967)
29. Frenkel, D., Smit, B.: Understanding Molecular Simulation: From Algorithms to Applications. Academic Press, Cambridge (2002)
30. Gupta, R., Kilcup, G.W., Sharpe, S.R.: Tuning the hybrid Monte Carlo algorithm. Phys. Rev. D **38**(4), 1278 (1988)
31. Habershon, S., Fanourgakis, G.S., Manolopoulos, D.E.: Comparison of path integral molecular dynamics methods for the infrared absorption spectrum of liquid water. J. Chem. Phys. **129**(7), 074501 (2008)
32. Habershon, S., Manolopoulos, D.E., Markland, T.E., Miller, T.F.: Ring-polymer molecular dynamics: quantum effects in chemical dynamics from classical trajectories in an extended phase space. Ann. Rev. Phys. Chem. **64**(1), 387–413 (2013)
33. Hairer, M.: Exponential mixing properties of stochastic PDEs through asymptotic coupling. Probab. Theory Relat. Fields **124**(3), 345–380 (2002)
34. Hairer, M., Mattingly, J.C.: Spectral gaps in Wasserstein distances and the 2D stochastic Navier–Stokes equations. Ann. Probab. **36**(6), 2050–2091 (2008)
35. Hairer, M., Mattingly, J.C., Scheutzow, M.: Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. Probab. Theory Relat. Fields **149**(1–2), 223–259 (2011)
36. Hairer, M., Stuart, A.M., Vollmer, S.J.: Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. Ann. Appl. Probab. **24**(6), 2455–2490 (2014)
37. Hairer, M., Stuart, A.M., Voss, J.: Analysis of SPDEs arising in path sampling part II: the nonlinear case. Ann. Appl. Probab. **17**(5/6), 1657–1706 (2007)
38. Hairer, M., Stuart, A.M., Voss, J.: Sampling conditioned diffusions. Trends Stoch. Anal. **353**, 159–186 (2009)
39. Hairer, M., Stuart, A.M., Voss, J., Wiberg, P.: Analysis of SPDEs arising in path sampling. Part I: the Gaussian case. Commun. Math. Sci. **3**(4), 587–603 (2005)
40. Joulin, A., Ollivier, Y.: Curvature, concentration and error estimates for Markov chain Monte Carlo. Ann. Probab. **38**(6), 2418–2442 (2010)
41. Kaipio, J., Somersalo, E.: Statistical and Computational Inverse Problems, Applied Mathematical Sciences, vol. 160. Springer, Berlin (2005)
42. Kennedy, A.D., Pendleton, B.: Acceptances and autocorrelations in hybrid Monte Carlo. Nucl. Phys. B Proc. Suppl. **20**, 118–121 (1991)
43. Korol, R., Bou-Rabee, N., Miller III, T.F.: Cayley modification for strongly stable path-integral and ring-polymer molecular dynamics. J. Chem. Phys. **151**(12), 124103 (2019)

44. Korol, R., Rosa-Races, J.L., Bou-Rabee, N., Miller III, T.F.: Dimension-free path-integral molecular dynamics without preconditioning. J. Chem. Phys. **152**(10), 104102 (2020)
45. Kou, S.C., Zhou, Q., Wong, W.H.: Discussion paper equi-energy sampler with applications in statistical inference and statistical mechanics. Ann. Stat. **34**, 1581–1619 (2006)
46. Leimkuhler, B., Matthews, C.: Molecular Dynamics. Springer, Berlin (2015)
47. Liang, F., Wong, W.H.: Real-parameter evolutionary Monte Carlo with applications to Bayesian mixture models. J. Am. Stat. Assoc. **96**(454), 653–666 (2001)
48. Liu, J.S.: Monte Carlo Strategies in Scientific Computing, 2nd edn. Springer, Berlin (2008)
49. Livingstone, S., Betancourt, M., Byrne, S., Girolami, M.: On the geometric ergodicity of Hamiltonian Monte Carlo. Bernoulli **25**(4A), 3109–3138 (2019)
50. Lu, J., Zhou, Z.: Continuum limit and preconditioned Langevin sampling of the path integral molecular dynamics (2018). arXiv preprint arXiv:1811.10995
51. Mackenze, P.B.: An improved hybrid Monte Carlo method. Phys. Lett. B **226**(3), 369–371 (1989)
52. Mangoubi, O., Smith, A.: Rapid mixing of Hamiltonian Monte Carlo on strongly log-concave distributions (2017). arXiv preprint arXiv:1708.07114
53. Mathias, R., Stoltz, G., Lelievre, T.: Free Energy Computations: A Mathematical Perspective. World Scientific, Singapore (2010)
54. Mattingly, J.: On recent progress for the stochastic Navier–Stokes equations, Journées Equations aux Dérivées Partielles, pp. 1–52. Université de Nantes, Nantes (2003)
55. Mattingly, J.C.: The Stochastic Navier–Stokes Equation: Energy Estimates and Phase Space Contraction. Doctoral thesis, Princeton University (1998)
56. Mattingly, J.C., Stuart, A.M., Higham, D.J.: Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. Stoch. Proc. Appl. **101**(2), 185–232 (2002)
57. Miller, T.F., Manolopoulos, D.E.: Quantum diffusion in liquid para-hydrogen from ring-polymer molecular dynamics. J. Chem. Phys. **122**(18), 184503 (2005)
58. Miller, T.F., Manolopoulos, D.E.: Quantum diffusion in liquid water from ring polymer molecular dynamics. J. Chem. Phys. **123**(15), 154504 (2005)
59. Miller III, T.F., Predescu, C.: Sampling diffusive transition paths. J. Chem. Phys. **126**(14), 144102 (2007)
60. Neal, R.M.: MCMC using Hamiltonian dynamics. Handb. Markov Chain Monte Carlo **2**, 113–162 (2011)
61. Petzold, L.R., Jay, L.O., Yen, J.: Numerical solution of highly oscillatory ordinary differential equations. Acta Numer. **6**, 437–483 (1997)
62. Pidstrigach, J.: Coupling and convergence for Hamiltonian Monte Carlo on Hilbert spaces, Master's Thesis Universität Bonn (2019)
63. Pinski, F.J., Stuart, A.M.: Transition paths in molecules at finite temperature. J. Chem. Phys. **132**(18), 184104 (2010)
64. Prokhorenko, S., Kalke, K., Nahas, Y., Bellaiche, L.: Large scale hybrid Monte Carlo simulations for structure and property prediction. Comput. Mater. **4**(1), 80 (2018)
65. Reznikoff, M.G., Vanden-Eijnden, E.: Invariant measures of stochastic partial differential equations and conditioned diffusions. Comptes Rendus Mathematique **340**(4), 305–308 (2005)
66. Sanz-Serna, J.M., Stuart, A.M.: Ergodicity of dissipative differential equations subject to random impulses. J. Differ. Equ. **155**(2), 262–284 (1999)
67. Stoltz, G.: Some mathematical methods for molecular and multiscale simulation, Ph.D. thesis, Ecole Nationale des Ponts et Chaussées (2007)
68. Stuart, A.M.: Inverse problems: a Bayesian perspective. Acta Numerica **19**, 451–559 (2010)
69. Talay, D.: Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. Markov Process. Relat. Fields **8**, 1–36 (2002)
70. Tierney, L.: A note on Metropolis–Hastings kernels for general state spaces. Ann. Appl. Probab. **8**(1), 1–9 (1998)
71. Zimmer, R.: Explicit contraction rates for a class of degenerate and infinite-dimensional diffusions. Stoch. Partial Differ. Equ. Anal. Comput. **5**(3), 368–399 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.