Stoch PDE: Anal Comp
https://doi.org/10.1007/s40072-021-00211-z

Mixing rates for Hamiltonian Monte Carlo algorithms in finite and infinite dimensions

Nathan E. Glatt-Holtz[1] · **Cecilia F. Mondaini**[2]

Received: 15 July 2020 / Revised: 30 July 2021 / Accepted: 26 August 2021

Abstract

We establish the geometric ergodicity of the preconditioned Hamiltonian Monte Carlo (HMC) algorithm defined on an infinite-dimensional Hilbert space, as developed in Beskos et al. (Stoch Process Appl 121(10):2201–2230, 2011). This algorithm can be used as a basis to sample from certain classes of target measures which are absolutely continuous with respect to a Gaussian measure. Our work addresses an open question posed in Beskos et al. (2011), and provides an alternative to a recent proof based on exact coupling techniques given in Bou-Rabee and Eberle (Two-scale coupling for preconditioned Hamiltonian Monte Carlo in infinite dimensions , 2019). The approach here establishes convergence in a suitable Wasserstein distance by using the weak Harris theorem together with a generalized coupling argument. We also show that a law of large numbers and central limit theorem can be derived as a consequence of our main convergence result. Moreover, our approach yields a novel proof of mixing rates for the classical finite-dimensional HMC algorithm. As such, the methodology we develop provides a flexible framework to tackle the rigorous convergence of other Markov Chain Monte Carlo algorithms. Additionally, we show that the scope of our result includes certain measures that arise in the Bayesian approach to inverse PDE problems, cf. Stuart (Acta Numer 19:451–559, 2010). Particularly, we verify all of the required assumptions for a certain class of inverse problems involving the recovery of a divergence free vector field from a passive scalar, Borggaard et al. (SIAM/ASA J Uncertain Quant 8(3):1036–1060, 2020).

Keywords Hamiltonian Monte Carlo (HMC) · Infinite dimensional Hamiltonian systems · Markov Chain Monte Carlo (MCMC) · Statistical sampling · Bayesian inversion · Advection-diffusion equations · Passive scalar transport

Mathematics Subject Classification 62C10 · 11K45 · 37K99

Extended author information available on the last page of the article

Published online: 18 September 2021

1 Introduction

It has long been appreciated that Markov chains can be employed as an effective computational tool to sample from probability measures. Starting from a desired 'target' probability distribution μ on a space $\mathbb{H}$ one seeks a Markov transition kernel P for which μ is an invariant and which moreover maintains desirable mixing properties with respect to this μ. In particular in Bayesian statistics [12,28,51,63,64,78] and in computational chemistry [21,24–26,39,40,54,59,69,70] such Markov chain Monte Carlo methods (MCMC) play a critical role by efficiently resolving high-dimensional distributions possessing complex multimodal and correlation structures which typically arise. However, notwithstanding their broad use in a variety of applications, the theoretical and practical understanding of the mixing rates of these chains remains poorly understood.

The initial mathematical foundation of MCMC methods was set in the late 40's by Metropolis and Ulam in [67], and later improved with the development of the Metropolis-Hastings algorithm in [50,66]. Further notable developments in the late 80's and 90's derived MCMC algorithms based on suitable Hamiltonian [29,71] and Langevin dynamical systems [2,38]. See e.g. [8,57,76] for a further general overview of the field. In view of exciting applications for the Bayesian approach to PDE inverse problems and in transition path sampling [12,17,18,28,45–48,63,74,75,78], an important recent advance in the MCMC literature [6,7,23,81] concerns the development of algorithms which are well defined on infinite-dimensional spaces. These methods have the scope to partially beat the 'curse of dimensionality' since one expects that the number of samples required to effectively resolve the target distribution to be independent of the degree of numerical discretization. However validating such claims of efficacy concerning this recently discovered class of infinite dimensional algorithms both in theory and in practice is an exciting and rapidly developing direction in current research.

This work provides an analysis of mixing rates for one particular class of methods among the MCMC algorithms mentioned above, known as Hybrid or Hamiltonian Monte Carlo (HMC) sampling; cf. [6,29,57,72]. For HMC sampling the general idea consists in taking advantage of a Hamiltonian dynamic taylored to the structure of the target μ, a distribution which functions as the marginal onto position space of the Gibbs measure associated to the dynamics. As such this 'Hamiltonian approach' produces nonlocal and nonsymmetric moves on the state space, allowing for more effective sampling from distributions with complex correlation structures in comparison to more traditional random walk based methods. Indeed the efficacy of the HMC approach has led to its widespread adoption in the statistics community as exemplified for example by the success of the STAN software package [36,79]. However, notwithstanding notable recent work, the theoretical understanding of optimal mixing rates for HMC based methods remains rather incomplete both in terms of optimal tuning of algorithmic parameters and in terms of the allowable structure of the target measure admitted by the theory [3–5,13–16,18,32,58,60–62].

We are particularly focused here on a version of HMC introduced in [6] where the authors consider a preconditioned Hamiltonian dynamics in order to derive a sampler which is well defined in the infinite-dimensional Hilbert space setting. While recent work [3,12,18] has shown that this 'infinite-dimensional' algorithm can be quite effective in practice, the question of rigorous justification of mixing rates posed in [6] as an open problem has only very recently been addressed in the work [13] in the case of *exact* (i.e. non-temporally-discretized) and preconditioned HMC. In [13], the authors follow an approach based on an exact coupling method recently considered in [14,34]. Here we develop an alternative approach to establishing mixing rates for preconditioned HMC based on the so called weak Harris theorem [20,41,43, 44] combined with suitable 'nudging' in the velocity variable which plays an analogous role to that provided by the classical Foias-Prodi estimate in the ergodic theory of certain classes of nonlinear SPDEs; cf. [37,55,65]. As such we believe the alternative approach that we consider here to be more flexible in certain ways providing a basis for further future analysis of MCMC algorithms. Furthermore, our approach for the exact dynamics developed here can be modified to derive mixing rates in the more interesting and practical case for discretized HMC. This later challenge will be taken up in future work.

Our main results can be summarized as follows. We show exponential mixing rates for the exact preconditioned HMC with respect to an appropriate Wasserstein distance in the space of probability measures on $\mathbb{H}$. For suitable observables, we show that this mixing implies a strong law of large numbers and a central limit theorem. In addition, we use very similar arguments to obtain a novel proof of mixing rates for the finite-dimensional HMC. Finally, the second part of the paper is concerned with the application of the theoretical mixing result to the PDE inverse problem of determining a background flow from partial observations of a passive scalar that is advected by the flow. A careful analysis of this inverse problem within a Bayesian framework is carried out in [12], where the authors also provide numerical simulations showing the effectiveness of the infinite-dimensional HMC algorithm from [6] in approximating the target distribution in this case. Here our task is to show that this example, for suitable observations of the passive scalar, satisfies all the conditions needed for our theoretical mixing result to hold, thus complementing the numerical experiments in [12] with rigorous mixing rates. In the sequel we provide a more detailed summary of the results obtained in the bulk of this manuscript.

1.1 Overview of the main results

The preconditioned Hamiltonian Monte Carlo algorithm from [6] which we analyze here can be described as follows. Fix a separable Hilbert space $\mathbb{H}$ with norm $|\cdot|$ and inner product $\langle\cdot,\cdot\rangle$. Let $\mathcal{B}(\mathbb{H})$ denote the associated Borel σ-algebra and let $\Pr(\mathbb{H})$ denote the set of Borel probability measures on $\mathbb{H}$. Suppose we wish to consider a target measure $\mu \in \Pr(\mathbb{H})$ which is given in the Gibbsian form

$$\mu(d\mathbf{q}) \propto \exp(-U(\mathbf{q}))\mu_0(d\mathbf{q}), \tag{1.1}$$

where $U : \mathbb{H} \to \mathbb{R}$ is a potential function. Here μ_0 is a probability measure on $\mathbb{H}$ typically corresponding to the prior distribution when we consider a μ derived as a Bayesian posterior. Following a standard formulation in the infinite dimensional setting, we assume in what follows that μ_0 is a centered Gaussian distribution on $\mathbb{H}$, i.e. $\mu_0 = \mathcal{N}(0, \mathcal{C})$, with $\mathcal{C}$ being a symmetric, strictly positive-definite, trace-class linear operator on $\mathbb{H}$.

Consider the following preconditioned Hamiltonian dynamics

$$\frac{d\mathbf{q}_t}{dt} = \mathbf{v}_t, \quad \frac{d\mathbf{v}_t}{dt} = -\mathbf{q}_t - \mathcal{C}DU(\mathbf{q}_t), \quad \text{with initial condition } (\mathbf{q}_0, \mathbf{v}_0) \in \mathbb{H} \times \mathbb{H}, \tag{1.2}$$

where $\mathbf{v} \in \mathbb{H}$ denotes a 'velocity' variable, so that (1.2) describes the evolution of the 'position-velocity' pair $(\mathbf{q}, \mathbf{v})$ in the extended phase space $\mathbb{H} \times \mathbb{H}$. Here we adopt the notation $\mathbf{q}_t$ and $\mathbf{v}_t$ to denote the value at time t of the variables $\mathbf{q}$ and $\mathbf{v}$, respectively. The associated Hamiltonian function, a formal invariant of the flow in (1.2), is given by

$$H(\mathbf{q}, \mathbf{v}) = \langle \mathcal{C}^{-1}\mathbf{q}, \mathbf{q}\rangle + U(\mathbf{q}) + \langle \mathcal{C}^{-1}\mathbf{v}, \mathbf{v}\rangle \quad \text{for suitable } (\mathbf{q}, \mathbf{v}) \in \mathbb{H} \times \mathbb{H}.$$

The exact preconditioned HMC algorithm works as follows. Starting from any $\mathbf{q}_0 \in \mathbb{H}$, draw $\mathbf{v}_0 \sim \mathcal{N}(0, \mathcal{C})$ and run the Hamiltonian dynamics with initial condition $(\mathbf{q}_0, \mathbf{v}_0)$ for a chosen temporal duration $T > 0$. Thus a forward step is proposed as the projection on the $\mathbf{q}$-coordinate of the solution of (1.2) starting from $(\mathbf{q}_0, \mathbf{v}_0)$ at time T, i.e. $\mathbf{q}_T(\mathbf{q}_0, \mathbf{v}_0)$. The associated Markov transition kernel $P : \mathbb{H} \times \mathcal{B}(\mathbb{H}) \to [0, 1]$ is then given as

$$P(\mathbf{q}_0, A) = \mathbb{P}(\mathbf{q}_T(\mathbf{q}_0, \mathbf{v}_0) \in A) \quad \text{with } \mathbf{v}_0 \sim \mathcal{N}(0, \mathcal{C}), \tag{1.3}$$

for every $A \in \mathcal{B}(\mathbb{H})$. We adopt the notation P^n for n steps of the Markov kernel P and recall that P acts as

$$\nu P(\cdot) = \int P(\mathbf{q}, \cdot)\nu(d\mathbf{q}), \quad P\Phi(\cdot) = \int \Phi(\mathbf{q})P(\cdot, d\mathbf{q})$$

on measures $\nu \in \Pr(\mathbb{H})$ and observables $\Phi : \mathbb{H} \to \mathbb{R}$, respectively. This kernel P leaves invariant the desired target probability measure μ given in (1.1), namely $\mu P = \mu$, as was demonstrated in [6] and recalled in 13 below. Clearly, in practice, one is not able to integrate (1.2) exactly so that one must instead resort to suitable numerical discretizations. These numerical integration schemes are designed so as to ensure that fundamental properties of Hamiltonian dynamics are preserved, such as time reversibility and volume-preservation or 'symplectiness' –see e.g. [16] for a survey. In this work we only analyze the exact dynamics, as the discretized case requires additional techniques and will be the subject of future work.

Let us now sketch a simplified version of our main result, given in rigorous and complete detail in Theorem 26 below. Our mixing result for the Markov kernel P defined in (1.3) is given with respect to a suitably constructed Wasserstein distance on

$\mathrm{Pr}(\mathbb{H})$. Namely, starting from $\varepsilon > 0$ and $\eta > 0$, consider $\tilde{\rho} : \mathbb{H} \times \mathbb{H} \to \mathbb{R}^+$ defined as

$$\tilde{\rho}(\mathbf{q}, \tilde{\mathbf{q}}) := \sqrt{\left(\frac{|\mathbf{q} - \tilde{\mathbf{q}}|}{\varepsilon} \wedge 1\right)\left(1 + \exp(\eta|\mathbf{q}|^2) + \exp(\eta|\tilde{\mathbf{q}}|^2)\right)}. \tag{1.4}$$

Here ε corresponds to the small scales at which we can match small perturbations in the initial position $\mathbf{q}_0$ with a corresponding perturbation in the initial velocity $\mathbf{v}_0$ in (1.2). On the other hand, for sufficiently small $\eta > 0$, the function $V(\mathbf{q}) = \exp(\eta|\mathbf{q}|^2)$ is a *Foster–Lyapunov* (or, simply, *Lyapunov*) function for P in the sense of Definition 19 and Proposition 20 below.

The mapping $\tilde{\rho}$ is a *distance-like function* in $\mathbb{H}$, i.e. it is a symmetric and lower-semicontinuous non-negative function such that $\tilde{\rho}(\mathbf{q}, \tilde{\mathbf{q}}) = 0$ holds if and only if $\mathbf{q} = \tilde{\mathbf{q}}$. We denote by $\mathcal{W}_{\tilde{\rho}} : \mathrm{Pr}(\mathbb{H}) \times \mathrm{Pr}(\mathbb{H}) \to \mathbb{R}^+ \cup \{\infty\}$ the following extension of $\tilde{\rho}$ to $\mathrm{Pr}(\mathbb{H})$:

$$\mathcal{W}_{\tilde{\rho}}(\nu_1, \nu_2) = \inf_{\Gamma \in \mathfrak{C}(\nu_1, \nu_2)} \int_{\mathbb{V} \times \mathbb{V}} \tilde{\rho}(\mathbf{q}, \tilde{\mathbf{q}}) \Gamma(d\mathbf{q}, d\tilde{\mathbf{q}}), \tag{1.5}$$

where $\mathfrak{C}(\nu_1, \nu_2)$ denotes the set of all *couplings* of ν_1 and ν_2, i.e. the set of all measures $\Gamma \in \mathrm{Pr}(\mathbb{H} \times \mathbb{H})$ with marginals ν_1 and ν_2. We notice that, on the other hand, the mapping $\rho(\mathbf{q}, \tilde{\mathbf{q}}) = (|\mathbf{q} - \tilde{\mathbf{q}}|/\varepsilon) \wedge 1$ defines a standard metric in $\mathbb{H}$. As such, its associated extension $\mathcal{W}_\rho$ to $\mathrm{Pr}(\mathbb{H})$ coincides with the usual Wasserstein-1 distance, [83].

With the above notation, we have the following convergence result. For the complete, detailed and general formulation, see Theorem 26 below.

Theorem 1 *Suppose that $\mathcal{C}$ is a symmetric strictly positive-definite trace class operator and that $U \in C^2(\mathbb{H})$ satisfies the global bound*

$$L_1 := \sup_{\mathbf{q} \in \mathbb{H}} |D^2 U(\mathbf{q})| < \infty \tag{1.6}$$

and the following dissipativity condition

$$|\mathbf{q}|^2 + \langle \mathbf{q}, \mathcal{C} DU(\mathbf{q}) \rangle \geq L_2 |\mathbf{q}|^2 - L_3 \quad \text{for all } \mathbf{q} \in \mathbb{H}, \tag{1.7}$$

for some constants $L_2 > 0$ and $L_3 \geq 0$. Let λ_1 denote the largest eigenvalue of $\mathcal{C}$.

Then, there exists an integration time $T = T(\lambda_1, L_1, L_2)$ for which the associated Markov kernel P as defined in (1.3) satisfies, with respect to $\tilde{\rho}$ defined in (1.4),

$$\mathcal{W}_{\tilde{\rho}}(\nu_1 P^n, \nu_2 P^n) \leq c_1 e^{-c_2 n} \mathcal{W}_{\tilde{\rho}}(\nu_1, \nu_2) \quad \text{for any } \nu_1, \nu_2 \in \mathrm{Pr}(\mathbb{H}) \text{ and } n \in \mathbb{N}, \tag{1.8}$$

for some $\varepsilon > 0$ as in (1.4) and some positive constants c_1, c_2 which depend only on the integration time $T > 0$, the constants L_i, $i = 1, 2, 3$, associated to the potential function U, and the covariance operator $\mathcal{C}$. In particular, (1.8) implies that μ defined in (1.1) is the unique invariant measure for P. Moreover, taking $\nu_1 = \delta_{\mathbf{q}_0}$, the Dirac *delta concentrated at some* $\mathbf{q}_0 \in \mathbb{H}$*, and* $\nu_2 = \mu$*, it follows from* (1.8) *that* $P^n(\mathbf{q}_0, \cdot)$ *converges exponentially to* μ *with respect to* $\mathcal{W}_{\tilde{\rho}}$ *as* $n \to \infty$*. In addition, for any suitably regular observable* $\Phi : \mathbb{H} \to \mathbb{R}$*,*

$$\left| P^n \Phi(\mathbf{q}_0) - \int \Phi(\mathbf{q}')\mu(dq') \right| \leq L_\Phi c_1 e^{-nc_2} \int \sqrt{1 + \exp(\eta|\mathbf{q}_0|^2) + \exp(\eta|\mathbf{q}'|^2)}\mu(d\mathbf{q}'),$$

for all $n \in \mathbb{N}$*, for some* $\eta > 0$ *and* $L_\Phi > 0$*.*

Further, taking $\{Q_n(\mathbf{q}_0)\}_{n\in\mathbb{N}}$ *to be the process associated to* $\{P^n\}_{n\in\mathbb{N}}$ *starting from* $\mathbf{q}_0 \in \mathbb{H}$*, i.e.* $Q_n(\mathbf{q}_0) \sim P(Q_{n-1}(\mathbf{q}_0), \cdot)$ *we have, for any* $\mathbf{q}_0 \in \mathbb{H}$ *and any suitably regular observable* $\Phi : \mathbb{H} \to \mathbb{R}$*, that*

$$X_n := \frac{1}{n}\sum_{k=1}^{n} \Phi(Q_k(\mathbf{q}_0)) - \int \Phi(\mathbf{q})\mu(d\mathbf{q}) \to 0 \quad \text{as } n \to \infty \text{ almost surely}$$

and that

$$\mathbb{P}(a < \sqrt{n}X_n \leq b) \to \frac{1}{\sqrt{2\pi\sigma^2}}\int_a^b e^{-\frac{x^2}{2\sigma^2}}dx \quad \text{as } n \to \infty \text{ for any } a, b \in \mathbb{R} \text{ with } a < b,$$

where $\sigma = \sigma(\Phi)$*. In other words,* $\{Q_n(\mathbf{q}_0)\}_{n\geq 0}$ *satisfies a strong law of large numbers (SLLN) and a central limit theorem (CLT).*

With similar arguments as used in the proof of Theorem 1 (cf. Theorem 26), we can also provide a new proof of mixing rates for the classical finite-dimensional HMC algorithm, as specified by the dynamics (7.2). This is carried out in Theorem 28 below, and complemented by further comparisons with the assumptions in the main infinite-dimensional result in Remark 30.

Having formulated our mixing result for the exact HMC algorithm associated with (1.1) we would like to be able to demonstrate that the conditions (1.6)–(1.7) which we impose on the potential U can be verified in concrete examples specifically as would apply to the Bayesian approach to PDE inverse problems. Here, as an illustrative example, we consider the problem of recovering a divergence free fluid flow $\mathbf{q}$ from the sparse and noisy observation of a passive solute $\theta(\mathbf{q})$ as was recently studied in [11,12].

To be specific let

$$\partial_t\theta + \mathbf{q}\cdot\nabla\theta = \kappa\Delta\theta, \quad \theta(0) = \theta_0 \tag{1.9}$$

where the solution evolves on the periodic box $\mathbb{T}^2$, namely $\theta : [0, \infty) \times \mathbb{T}^2 \to \mathbb{R}$ and $\kappa > 0$ is a fixed diffusion parameter. Given a sufficiently regular initial condition $\theta_0 : \mathbb{T}^2 \to \mathbb{R}$, which we take to be known in advance, we specify the (linear) observation procedure, i.e.

$$\mathcal{O}(\theta) := \left\{ \int_0^\infty \int_{\mathbb{T}^2} \theta(t, x)K_j(t, x)dxdt \right\}_{j=1}^m \tag{1.10}$$

where $m \geq 1$ represents the number of separate observations of θ and K_j are the associated 'observation kernels'.

Here we notice that the general formulation (1.10) allows for a broad class of examples corresponding to specific functions K_j, as long as the integrals in (1.10) are well-defined, which of course depends on the regularity of the solution θ of (1.9), see Proposition 31 below. In particular, we may consider the case of pointwise in time observations by taking $K_j(t, x) = \delta_{t_j}(t) f_j(x)$, for any finite collections of times $\{t_j\}_{j=1}^m \subset [0, \infty)$ and functions $\{f_j\}_{j=1}^m \subset L^2(\mathbb{T}^2)$. Here the f_j's can be taken e.g. as basis functions of the Hilbert space $L^2(\mathbb{T}^2)$ to account for spectral in space observations, or as $f_j(x) = |A_j|^{-1} \mathbb{1}_{A_j}(x)$ for some bounded set $A_j \subset \mathbb{T}^2$ to represent spatial observations given as local averages. Further, we could also take $K_j(t, x) = \delta_{t_j}(t)\delta_{x_j}(x)$ for any finite set of spatial locations $\{x_j\}_{j=1}^m \subset \mathbb{T}^2$, thus representing the case of observations which are pointwise both in space and time. Clearly, other examples could be given by combining these different types of spatial observations with other kinds of temporal observations, such as local time averages, spectral etc. It is also notable that our theory below treats various linear observations of derivatives of θ and moreover is easily modifiable to include certain nonlinear observations of θ i.e. L^p norms of θ etc., see Sect. 8 below.

Positing an additive observation noise η, we have the following statistical model linking any suitably regular, divergence free, $\mathbf{q} : \mathbb{T}^2 \to \mathbb{R}^2$ with a resulting data set $\mathcal{Y}$ as

$$\mathcal{Y} = \mathcal{O}(\theta(\mathbf{q})) + \eta,$$

where $\theta(\mathbf{q})$ represents the solution of (1.10) corresponding to $\mathbf{q}$ so that $\theta(\mathbf{q})$ sits in an appropriate solution space which we specify in rigorous detail below in Proposition 31.

Following the Bayesian statistical inversion formalism [28,51], given a fixed observation $\mathcal{Y} \in \mathbb{R}^m$ and a prior distribution μ_0 on a suitable Hilbert space of divergence free, periodic vector fields and a probability density function $p_\eta : \mathbb{R}^m \to \mathbb{R}$ for the observation noise η, we obtain a posterior distribution

$$\mu^{\mathcal{Y}}(d\mathbf{q}) \propto \exp(-U(\mathbf{q}))\mu_0(d\mathbf{q}) \quad \text{where} \quad U(\mathbf{q}) = -\log(p_\eta(\mathcal{Y} - \mathcal{O}(\theta(\mathbf{q})))). \tag{1.11}$$

see e.g. [28], [12, Appendix C]. For simplicity of presentation, we focus here on the typical situation where $\eta \sim N(0, \Gamma)$, with Γ a symmetric, strictly positive definite covariance operator on $\mathbb{R}^m$. In this case U takes the form

$$U(\mathbf{q}) = |\Gamma^{-1/2}(\mathcal{Y} - \mathcal{O}(\theta(\mathbf{q})))|^2, \tag{1.12}$$

where $|\cdot|$ represents the usual Euclidean norm on $\mathbb{R}^m$.

Our main results here, Proposition 34 and Corollary 35, show that when the observations satisfy an inequality of the form

$$|\mathcal{O}(\theta)| \leq c_0 \sup_{t \leq t^*} \int_{\mathbb{T}^2} |\theta(t, x)|^2 dx, \tag{1.13}$$

for some $t^* \geq 0$, which in particular includes the examples of observations which are pointwise in time and spectral in space or local averages in space, then we can verify the conditions imposed on the potential function U (cf. (1.6) and more generally Assumption 8 below) and in particular establish suitable global bounds on D^2U. On the other hand, when the observations satisfy instead an inequality such as

$$|\mathcal{O}(\theta)| \leq c_0 \sup_{t \leq t^*, x \in \mathbb{T}^2} |\theta(t, x)| \tag{1.14}$$

for some $t^* \geq 0$, which includes in particular the example of space-time pointwise observations, or for observations involving gradients or other higher order derivatives of θ, we can only show local bounds on D^2U.

Overview of the proof

Our proof follows the approach of the weak Harris theorem developed in [43], which is an elegant generalization of the classical Harris mixing results, [42,49,68]. It establishes necessary conditions for two point contraction at small, intermediate and large scales in a fashion well adapted to the Wasserstein metric, a notion of distance which is crucially needed for many types of processes evolving on infinite dimensional spaces. We should emphasize the authors in [43] provide clarity and flexibility in their approach by developing a class of distance-like functions (cf. (1.4)) which allows one to establish global contractivity directly and thus avoiding the need for intricate pathwise coupling constructions considered elsewhere in the literature.

As such, the main difficulties here lie in showing that the necessary assumptions of the weak Harris theorem are valid in our context. These assumptions amount to showing, with respect to $\rho : \mathbb{H} \times \mathbb{H} \to [0, 1]$ defined as $\rho(\mathbf{q}, \tilde{\mathbf{q}}) = 1 \wedge (|\mathbf{q} - \tilde{\mathbf{q}}|/\varepsilon)$, with $\varepsilon > 0$ fixed, that the following is true: there exists $m \in \mathbb{N}$ sufficiently large such that

(i) P^m is ρ-contracting, i.e. there exists $0 < \delta_1 < 1$ such that

$$\mathcal{W}_\rho(P^m(\mathbf{q}_0, \cdot), P^m(\tilde{\mathbf{q}}_0, \cdot)) \leq \delta_1 \rho(\mathbf{q}, \tilde{\mathbf{q}}) \quad \text{for all } \mathbf{q}_0, \tilde{\mathbf{q}}_0 \in \mathbb{H} \text{ with } \rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0) < 1; \tag{1.15}$$

(ii) For level sets of the form $A_K := \{\mathbf{q} \in \mathbb{H} : |q| \leq K\}$, for $K > 0$, A_K is ρ-*small* for P^m, i.e. there exists $0 < \delta_2 < 1$ and $m \geq 1$ such that

$$\mathcal{W}_\rho(P^m(\mathbf{q}_0, \cdot), P^m(\tilde{\mathbf{q}}_0, \cdot)) \leq 1 - \delta_2 \quad \text{for all } \mathbf{q}_0, \tilde{\mathbf{q}}_0 \in A_K. \tag{1.16}$$

Finally we need a Lyapunov condition:

(iii) For a suitable $V : \mathbb{H} \to \mathbb{R}^+$ that

$$P^n V(\mathbf{q}) \leq C\kappa^n V(\mathbf{q}) + K, \tag{1.17}$$

for every $\mathbf{q} \in \mathbb{H}$ and $n \geq 1$, where $\kappa \in (0, 1)$ and $C, K > 0$ are independent of $\mathbf{q}$ and n.

Roughly speaking the conditions (i)–(iii) correspond to establishing a two-point contraction at small, intermediate and large scales respectively.

Following an approach developed in the stochastic PDE literature [20,37,41,43,55,65], the idea consists in establishing (i) and (ii) above without explicitly constructing a coupling between $P^m(\mathbf{q}_0, \cdot)$ and $P^m(\tilde{\mathbf{q}}_0, \cdot)$. Instead, we construct an 'approximate' coupling by defining a modified process $\widetilde{P}(\mathbf{q}_0, \tilde{\mathbf{q}}_0, \cdot)$ in a control-like approach. We define the process $\widetilde{P}$ by imposing a suitable 'shift' in the initial velocity $\mathbf{v}_0$ in (1.3) depending on the initial positions $\mathbf{q}_0, \tilde{\mathbf{q}}_0$. Namely, for a fixed integration time $T > 0$, we take

$$\widetilde{P}(\mathbf{q}_0, \tilde{\mathbf{q}}_0, A) := \mathbb{P}(\mathbf{q}_T(\tilde{\mathbf{q}}_0, \tilde{\mathbf{v}}_0) \in A) \quad \text{with } \tilde{\mathbf{v}}_0 = \mathbf{v}_0 + \mathcal{S}(\mathbf{q}_0, \tilde{\mathbf{q}}_0), \quad \mathbf{v}_0 \sim \mathcal{N}(0, \mathcal{C}), \tag{1.18}$$

for every $A \in \mathcal{B}(\mathbb{H})$. Here we consider a shift $\mathcal{S}(\mathbf{q}_0, \tilde{\mathbf{q}}_0)$ which is inspired by estimates developed in [14]; $\mathcal{S}$ is defined so as to ensure a suitable contraction between two solutions of (1.2) starting from $(\mathbf{q}_0, \mathbf{v}_0)$ and $(\tilde{\mathbf{q}}_0, \tilde{\mathbf{v}}_0)$ at the final time $T > 0$.

Since ρ is a metric in $\mathbb{H}$, the corresponding extension $\mathcal{W}_\rho$ is a metric in $\Pr(\mathbb{H})$ and in fact coincides with the Wasserstein-1 distance. Thus, by the triangle inequality,

$$\begin{aligned} &\mathcal{W}_\rho(P^m(\mathbf{q}_0, \cdot), P^m(\tilde{\mathbf{q}}_0, \cdot)) \\ &\quad \le \mathcal{W}_\rho(P^m(\mathbf{q}_0, \cdot), \widetilde{P}^m(\mathbf{q}_0, \tilde{\mathbf{q}}_0, \cdot)) + \mathcal{W}_\rho(\widetilde{P}^m(\mathbf{q}_0, \tilde{\mathbf{q}}_0, \cdot), P^m(\tilde{\mathbf{q}}_0, \cdot)), \end{aligned} \tag{1.19}$$

where $\widetilde{P}^m$ denotes the m-fold iteration of $\widetilde{P}$, corresponding to a sequence $(\mathbf{v}_0^{(1)}, \ldots, \mathbf{v}_0^{(m)})$ of initial velocities drawn from $\mathcal{N}(0, \mathcal{C})$ and shifted as in (1.18) with $\mathbf{q}_0, \tilde{\mathbf{q}}_0$ replaced with the starting positions from each iteration. In view of establishing (1.15) and (1.16), the first term on the right-hand side of (1.19) is estimated by first showing a contraction result between two solutions of (1.2) starting from $(\mathbf{q}_0, \mathbf{v}_0)$ and $(\tilde{\mathbf{q}}_0, \tilde{\mathbf{v}}_0)$ with respect to ρ in $\mathbb{H}$, which is then extended to $\mathcal{W}_\rho$ in $\Pr(\mathbb{H})$. Such contraction result follows solely from assumption (1.6) on the potential function U together with a smallness assumption on the integration time T; see Proposition 18 below. Moreover, assumption (1.6) implies that the only possible source of nonlinearity in the dynamics (1.2), i.e. DU, is Lipschitz, which in particular guarantees the well-posedness of (1.2) as we detail in Proposition 12.

The second term on the right-hand side (1.19) represents a 'cost of control' term and in fact the tuning parameter ε appearing in ρ specifies the scales at which this cost does not 'become too large'. We estimate this term with the help of Girsanov's theorem from which we obtain a bound in terms of the Radon-Nikodym derivative between the law σ_m of the velocity path $(\mathbf{v}_0^{(1)}, \ldots, \mathbf{v}_0^{(m)})$ and the law $\tilde{\sigma}_m$ of the associated shifted velocity path $(\tilde{\mathbf{v}}_0^{(1)}, \ldots, \tilde{\mathbf{v}}_0^{(m)})$, i.e. Girsanov provides us with $d\sigma_m/d\tilde{\sigma}_m$. Here we notice that, in order to guarantee that $d\sigma_m/d\tilde{\sigma}_m$ is well-defined, we define the shift $\mathcal{S}$ in (1.18) to be in a finite-dimensional subspace of $\mathbb{H}$ (cf. (5.7)). Indeed, looking at the case $m = 1$ for simplicity, notice that if $\mathbf{v}_0 \sim \mathcal{N}(0, \mathcal{C})$ then $\tilde{\mathbf{v}}_0 \sim \mathcal{N}(\mathcal{S}(\mathbf{q}_0, \tilde{\mathbf{q}}_0), \mathcal{C})$ and, by the Feldman-Hajek theorem (see, e.g., [27, Theorem 2.23]), $\mathcal{N}(0, \mathcal{C})$ and $\mathcal{N}(\mathcal{S}(\mathbf{q}_0, \tilde{\mathbf{q}}_0), \mathcal{C})$ are mutually singular unless $\mathcal{S}(\mathbf{q}_0, \tilde{\mathbf{q}}_0)$ belongs to the Cameron-Martin space of $\mathcal{N}(0, \mathcal{C})$. Notably, the Cameron-Martin space of $\mathcal{N}(0, \mathcal{C})$

has $\mathcal{N}(0, \mathcal{C})$-measure zero when $\mathbb{H}$ is infinite-dimensional. This illustrates the fact that two measures in an infinite-dimensional space are frequently mutually singular. However, by considering a velocity shift $\mathcal{S}$ that belongs to an N-dimensional subspace $\mathbb{H}_N \subset \mathbb{H}$, for some $N \in \mathbb{N}$, we can show that σ_m and $\tilde{\sigma}_m$ are mutually absolutely continuous, with an estimate of $d\sigma_m/d\tilde{\sigma}_m$, and thus of the second term in (1.19), that depends on the dimension N. Here N is chosen so as to obtain a suitable contraction between different trajectories of (1.2) and hence to provide a useful estimate of the first term in (1.19) (see Propositions 18 and 22). For this purpose, N must be chosen to be sufficiently large, but is nevertheless a fixed parameter depending only on the potential function U through the constant L_1 from (1.6) (see (3.20) below).

The third part of the proof consists in showing that such V is a Lyapunov function for P as given in Proposition 20 below. Here, in addition to quadratic exponential function $V(\mathbf{q}) = \exp(\eta|q|^2)$ as in (1.4) we in fact show that any function of the form $V(\mathbf{q}) = |\mathbf{q}|^i$, $i \in \mathbb{N}$, is also a Lyapunov function. The result of Proposition 20 follows from both assumptions (1.6) and (1.7) on the potential U together with a smallness assumption on the integration time T. Notably, assumption (1.7) on U is only imposed in order to obtain this Lyapunov structure. Indeed, condition (1.7) provides a coercivity-like property for DU in (1.2) which, when complemented with the smallness assumption on T, allows us to show the required exponential decay of such functions V modulo a constant, thus proving the Lyapunov property.

It remains to leverage the spectral gap now established, (1.8), to prove a Law of Large numbers (LLN) and Central Limit Theorem (CLT) type result for the implied Markov process. While this implication is extensively developed in the literature, and recently generalized to the situation where the spectral gap appears in the Wasserstein sense [53,56], it was not immediately clear that these results are easily applied as a black box to our situation. Instead, for clarity of presentation, we provide an independent proof of the LLN and CLT in an appendix which is carefully adapted to our situation where the $\tilde{\rho}$ in (1.8) is only distance-like. While we are in particular following the road map laid out in [53], we believe our proof may be of some independent interest.

Organization of the manuscript

The rest of the manuscript is organized as follows. In Sect. 2 we provide the complete details of our mathematical setting including the assumptions on the covariance operator $\mathcal{C}$ and the potential U in (1.2). Section 3 provides certain a priori bounds on (1.2) and concludes with the low-mode nudging bound that we use to synchronize the positions of two processes by suitably coupling their momenta. Lyapunov estimates on the exact Hamiltonian Monte Carlo dynamics are given in Sect. 4. In Sect. 5 we combine the bounds in the previous two sections to establish the pointwise contractivity of the Markovian dynamics, namely the so called ρ-small and ρ-contractivity conditions. The main result on geometric ergodicity is stated rigorously in Sect. 6 followed by the proof using the weak Harris theorem [43]. Section 7 details how our approach also provides a novel proof for the finite dimensional setting. Finally in Sect. 8 we establish that the conditions of the main theorem apply to the Bayesian statistical inversion problem of estimating a divergence free vector field $\mathbf{q}$ from the partial observation of a scalar quantity advected by the flow. Section 1 shows how the law of large numbers and the central limit theorem follow in our setting from our main result on spectral gaps.

2 Preliminaries

This section collects various mathematical preliminaries and sets down the precise assumptions which we use below in the statements of the main results of the paper.

2.1 The Gaussian reference measure

Let $\mathbb{H}$ be a separable and real Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $|\cdot|$. We take $\mathcal{N}(0,\mathcal{C})$ to denote the centered normal distribution on $\mathbb{H}$ with covariance operator $\mathcal{C}$. See e.g. [10,27] for generalities concerning Gaussian measures on Hilbert space. In this paper we *always* assume that $\mathcal{C}$ satisfies the following conditions.

Assumption 2 $\mathcal{C} : \mathbb{H} \to \mathbb{H}$ is a trace class, symmetric and strictly positive definite linear operator. Thus, by the spectral theorem, we have a complete orthonormal basis $\{\mathbf{e}_i\}_{i\in\mathbb{N}}$ of $\mathbb{H}$ which are the eigenfunctions of $\mathcal{C}$. We write corresponding eigenvalues $\{\lambda_i\}_{i\in\mathbb{N}}$ in non-increasing order and note that the trace class condition amounts to

$$\mathrm{Tr}(\mathcal{C}) := \sum_i \lambda_i < \infty. \tag{2.1}$$

We will also make frequent use of fractional powers of $\mathcal{C}$ which we define as follows.

Definition 3 For any $\gamma \in \mathbb{R}$, we define fractional power $\mathcal{C}^\gamma$ of $\mathcal{C}$ by

$$\mathcal{C}^\gamma \mathbf{f} = \sum_i \lambda_i^\gamma \langle \mathbf{f}, \mathbf{e}_i\rangle \mathbf{e}_i ,$$

which makes sense for any $\mathbf{f} \in \mathbb{H}_\gamma$. Here $\mathbb{H}_\gamma$ is defined as

$$\mathbb{H}_\gamma = \{\mathbf{f} \in \mathbb{H} |\ |f|_\gamma < \infty\} \quad \text{where } |\mathbf{f}|_\gamma^2 := |\mathcal{C}^{-\gamma}\mathbf{f}|^2 = \sum_i \lambda_i^{-2\gamma} \langle \mathbf{f}, \mathbf{e}_i\rangle^2 \tag{2.2}$$

when $\gamma \geq 0$. For $\gamma < 0$, $\mathbb{H}_\gamma$ is defined as the dual of $\mathbb{H}_{-\gamma}$ relative to $\mathbb{H}$. In addition, for every $\gamma \in \mathbb{R}$, we define the inner product $\langle\cdot,\cdot\rangle_\gamma = \langle \mathcal{C}^{-\gamma}\cdot, \mathcal{C}^{-\gamma}\cdot\rangle$.

According to Definition 3, it follows that $\mathbb{H}_{-\tilde{\gamma}} \subseteq \mathbb{H}_{-\gamma}$ for every $\gamma, \tilde{\gamma} \in \mathbb{R}$ with $\gamma \geq \tilde{\gamma}$. Moreover, note that $\mathbb{H}_{1/2}$ is the Cameron-Martin space associated with $\mathcal{N}(0,\mathcal{C})$ with inner product $\langle\cdot,\cdot\rangle_{1/2} = \langle \mathcal{C}^{-1/2}\cdot, \mathcal{C}^{-1/2}\cdot\rangle$ and norm $|\cdot|_{1/2} = |\mathcal{C}^{-1/2}\cdot|$; see [27, Chapter 2].

In terms of these fractional spaces $\mathbb{H}_\gamma$ we have the following 'Poincaré' and 'reverse-Poincaré' inequalities. For this purpose and for later use we define, for $N \geq 1$,

$$\Pi_N \mathbf{f} = \sum_{j \leq N} \langle \mathbf{f}, \mathbf{e}_j \rangle \mathbf{e}_j, \quad \Pi^N \mathbf{f} = \sum_{j > N} \langle \mathbf{f}, \mathbf{e}_j \rangle \mathbf{e}_j, \tag{2.3}$$

namely the projection of $\mathbf{f} \in \mathbb{H}$ onto 'low' and 'high' modes.

Lemma 4 *Given any $\gamma, \tilde{\gamma} \in \mathbb{R}$ with $\gamma \geq \tilde{\gamma}$, the following hold:*

$$\left|\mathcal{C}^\gamma \mathbf{f}\right| \leq \lambda_1^{(\gamma-\tilde{\gamma})} \left|\mathcal{C}^{\tilde{\gamma}} \mathbf{f}\right|, \tag{2.4}$$

when $\mathbf{f} \in \mathbb{H}_{-\tilde{\gamma}}$. Moreover, for any $N \geq 1$,

$$\left|\mathcal{C}^\gamma \Pi^N \mathbf{f}\right| \leq \lambda_{N+1}^{(\gamma-\tilde{\gamma})} \left|\mathcal{C}^{\tilde{\gamma}} \Pi^N \mathbf{f}\right|, \tag{2.5}$$

for any $\mathbf{f} \in \mathbb{H}_{-\tilde{\gamma}}$.

In certain applications, one may wish to define the Markovian dynamics associated to (1.2) only on $\mathbb{H}_\gamma$ for some $\gamma \in (0, 1/2)$, which is a strict subset of $\mathbb{H}$. For this reason, in what follows we consider our underlying phase space to be more generally given by $\mathbb{H}_\gamma$, for some $\gamma \in [0, 1/2)$. This leads us to introduce the following additional assumption which will sometimes be imposed:

Assumption 5 For some $\gamma \in [0, 1/2)$, $\mathcal{C}^{1-2\gamma}$ is trace class. Namely,

$$\mathrm{Tr}(\mathcal{C}^{1-2\gamma}) := \sum_i \lambda_i^{1-2\gamma} < \infty. \tag{2.6}$$

Under Assumption 5 we have the following regularity property

Lemma 6 *Suppose that μ_0 is $\mathcal{N}(0, \mathcal{C})$ defined on $\mathbb{H}$ with $\mathcal{C}$ under Assumption 2, Assumption 5. Then μ_0 is also $\mathcal{N}(0, \mathcal{C}^{1-2\gamma})$ defined on $\mathbb{H}_\gamma$.*

Remark 7 We typically think of the covariance $\mathcal{C}$ as a 'smoothing operator'. A simple example of $\mathcal{C}$ satisfying the above assumptions is A^{-1} where $A = -\partial_{xx}$ is the second derivative on $[0, \pi]$ endowed with Dirichlet boundary conditions. Note that, with this choice of $\mathcal{C}$, the spaces $\mathbb{H}_\gamma$ correspond to the usual L^2-based Sobolev space $H^{\gamma/2}$ with the Cameron-Martin space given by H^1. A more involved variation on this theme will be considered below in Sect. 8 when we consider an application to a PDE inverse problem.

2.2 Conditions on the potential

In what follows we impose the following regularity conditions on the potential energy function U from (1.1). Note that in particular assumption $(B1)$ below is compatible with the setting imposed in [6]; see Remark 11 below.

Assumption 8 For a fixed value of $\gamma \in [0, 1/2)$ the potential in (1.2) $U : \mathbb{H}_\gamma \to \mathbb{R}$ is twice Fréchet differentiable and

(B1) There exists $L_1 > 0$ such that

$$|D^2U(\mathbf{f})|_{\mathcal{L}_2(\mathbb{H}_\gamma)} = |\mathcal{C}^\gamma D^2U(\mathbf{f})\mathcal{C}^\gamma|_{\mathcal{L}_2(\mathbb{H})} \le L_1 \tag{2.7}$$

for any $\mathbf{f} \in \mathbb{H}_\gamma$, where $|\cdot|_{\mathcal{L}_2(\mathbb{H}_\gamma)}$ and $|\cdot|_{\mathcal{L}_2(\mathbb{H})}$ denote the usual operator norms for real valued bilinear operators defined on $\mathbb{H}_\gamma \times \mathbb{H}_\gamma$ and on $\mathbb{H} \times \mathbb{H}$, respectively.

(B2) There exists $L_2 > 0$ and $L_3 \ge 0$ such that, for this value of $\gamma \in [0, 1/2)$

$$|\mathbf{f}|_\gamma^2 + \langle \mathbf{f}, \mathcal{C}DU(\mathbf{f})\rangle_\gamma \ge L_2\, |\mathbf{f}|_\gamma^2 - L_3 \tag{2.8}$$

for every $\mathbf{f} \in \mathbb{H}_\gamma$.

A number of remarks are in order regarding Assumption 8:

Remark 9 (i) Assumption 8 (B1) and the mean value theorem imply that

$$|DU(\mathbf{f}) - DU(\mathbf{g})|_{-\gamma} \le L_1\, |\mathbf{f} - \mathbf{g}|_\gamma \tag{2.9}$$

for any $\mathbf{f}, \mathbf{g} \in \mathbb{H}_\gamma$ and, in particular,

$$|DU(\mathbf{f})|_{-\gamma} \le L_1\, |\mathbf{f}|_\gamma + L_0 \tag{2.10}$$

for every $\mathbf{f} \in \mathbb{H}_\gamma$, where $L_0 = |DU(0)|_{-\gamma}$. Inequalities (2.9) and (2.10) will be used extensively in the analysis below.

(ii) If U satisfies, in addition, the following property:

(B3) There exists $L_4 \in [0, \lambda_1^{-1+2\gamma})$ and $L_5 \ge 0$ such that

$$|DU(\mathbf{f})|_{-\gamma} \le L_4\, |\mathbf{f}|_\gamma + L_5, \quad \text{for any } \mathbf{f} \in \mathbb{H}_\gamma, \tag{2.11}$$

then (B2) is automatically satisfied. Indeed, we have

$$\begin{aligned} |\mathbf{f}|_\gamma^2 + \langle \mathbf{f}, \mathcal{C}DU(\mathbf{f})\rangle_\gamma &\ge |\mathbf{f}|_\gamma^2 - |\langle \mathbf{f}, \mathcal{C}DU(\mathbf{f})\rangle_\gamma| \ge |\mathbf{f}|_\gamma^2 - |\mathbf{f}|_\gamma\, |\mathcal{C}^{1-\gamma}DU(\mathbf{f})| \\ &\ge |\mathbf{f}|_\gamma^2 - \lambda_1^{1-2\gamma}\, |\mathbf{f}|_\gamma\, |DU(\mathbf{f})|_{-\gamma}\,, \end{aligned} \tag{2.12}$$

where the last inequality follows from Lemma 4 and the fact that $\gamma \in [0, 1/2)$. Using (2.11) in (2.12) and Young's inequality, we obtain

$$\begin{aligned} |\mathbf{f}|_\gamma^2 + \langle \mathbf{f}, \mathcal{C}DU(\mathbf{f})\rangle_\gamma &\ge (1 - \lambda_1^{1-2\gamma}L_4)\, |\mathbf{f}|_\gamma^2 - \lambda_1^{1-2\gamma}L_5\, |\mathbf{f}|_\gamma \\ &\ge \frac{1 - \lambda_1^{1-2\gamma}L_4}{2}\, |\mathbf{f}|_\gamma^2 - C, \end{aligned}$$

where $C \in \mathbb{R}^+$ is a constant depending on $\lambda_1^{1-2\gamma}$, L_4, L_5. Notice that, in particular, if U satisfies (B1) with $L_1 \in [0, \lambda_1^{-1+2\gamma})$, then (B3) is verified with $L_4 = L_1$ and $L_5 = L_0$ (cf. (2.10)).

(iii) Assumptions (B1) and (B2) imply that the constants L_1 and L_2 satisfy the following relation:

$$L_2 \leq 1 + \lambda_1^{1-2\gamma} L_1. \tag{2.13}$$

Indeed, from (B2), Lemma 4 and (2.10), we obtain that

$$\begin{aligned}(L_2 - 1)\,|\mathbf{f}|_\gamma^2 - L_3 \leq \langle \mathbf{f}, \mathcal{C}DU(\mathbf{f})\rangle_\gamma &\leq \lambda_1^{1-2\gamma}\,|\mathbf{f}|_\gamma\,|DU(\mathbf{f})|_{-\gamma} \\ &\leq \lambda_1^{1-2\gamma} L_1\,|\mathbf{f}|_\gamma^2 + L_0\lambda_1^{1-2\gamma}\,|\mathbf{f}|_\gamma \\ &\leq (\delta + \lambda_1^{1-2\gamma} L_1)\,|\mathbf{f}|_\gamma^2 + \frac{(L_0\lambda_1^{1-2\gamma})^2}{4\delta},\end{aligned}$$

for any $\delta > 0$, so that

$$(L_2 - 1 - \lambda_1^{1-2\gamma} L_1 - \delta)\,|\mathbf{f}|_\gamma^2 \leq L_3 + \frac{(L_0\lambda_1^{1-2\gamma})^2}{4\delta}$$

holds for any $\mathbf{f} \in \mathbb{H}_\gamma$, and every $\delta > 0$, which implies (2.13).

This paper is concerned with sampling from probability distributions on $\mathbb{H}$ that have a density with respect to $\mathcal{N}(0, \mathcal{C})$ which are of the form (1.1). In order that this is indeed the case and furthermore to ensure the invariance of μ with respect to the Markovian dynamics defined with respect to (1.2), we assume the following condition.

Assumption 10 Taking $\gamma \in [0, 1/2)$ as in Assumption 8 we suppose that, for any $\varepsilon > 0$ there exists an $M = M(\varepsilon) \in \mathbb{R}$, such that

$$U(\mathbf{f}) \geq M - \varepsilon|\mathbf{f}|_\gamma^2 \quad \text{for any } \mathbf{f} \in \mathbb{H}_\gamma.$$

Remark 11 We notice that Assumption 8 $(B1)$ and 10 above are equivalent to conditions 3.2 and 3.3 imposed in [6]. Indeed such assumptions are applied there in order to show the well-posedness of the dynamics in (1.2) as well as to show that the measure μ defined in (1.1) is an invariant measure associated to (1.2). Such results are recalled in Propositions 12 and 13 below, respectively. However, as pointed out in the introduction, condition Assumption 8 $(B2)$ is further imposed in our setting in order to obtain the Lyapunov structure (1.17), which together with the contractivity and smallness properties (1.15)–(1.16) allows us to obtain our main convergence result, Theorem 26 below.

2.3 Well-posedness of the Hamiltonian dynamics

In the following proposition, we recall a well-posedness result of the Hamiltonian dynamics in (1.2), as shown in [6]. We consider the usual norm on the product space

$\mathbb{H}_\gamma \times \mathbb{H}_\gamma$ with the slight abuse of notation:

$$|(\mathbf{q}, \mathbf{v})|_\gamma := |\mathbf{q}|_\gamma + |\mathbf{v}|_\gamma \quad \text{for all } (\mathbf{q}, \mathbf{v}) \in \mathbb{H}_\gamma \times \mathbb{H}_\gamma. \tag{2.14}$$

Proposition 12 *Suppose $\mathcal{C}$ satisfies Assumption 2 and that U maintains Assumption 8, (B1). Let $\gamma \in [0, 1/2)$ be as in Assumption 8.*

(i) For any $(\mathbf{q}_0, \mathbf{v}_0) \in \mathbb{H}_\gamma \times \mathbb{H}_\gamma$, there exists a unique $(\mathbf{q}, \mathbf{v}) = (\mathbf{q}(\mathbf{q}_0, \mathbf{v}_0), \mathbf{v}(\mathbf{q}_0, \mathbf{v}_0))$ with

$$(\mathbf{q}, \mathbf{v}) \in C^1(\mathbb{R}; \mathbb{H}_\gamma \times \mathbb{H}_\gamma) \tag{2.15}$$

and obeying (1.2). The resulting solution operators $\{\Xi_t\}_{t\in\mathbb{R}}$ defined via

$$\Xi_t(\mathbf{q}_0, \mathbf{v}_0) = \mathbf{q}_t(\mathbf{q}_0, \mathbf{v}_0)$$

are all continuous maps from $\mathbb{H}_\gamma \times \mathbb{H}_\gamma$ to $\mathbb{H}_\gamma$.

(ii) Under the additional restriction on $\mathcal{C}$ of Assumption 5 and fixing an integration time $T > 0$ the random variable

$$Q_1(\mathbf{q}_0) = \mathbf{q}_T(\mathbf{q}_0, \mathbf{v}_0), \quad \mathbf{v}_0 \sim \mathcal{N}(0, \mathcal{C})$$

is well defined in $\mathbb{H}_\gamma$ for any $\mathbf{q}_0 \in \mathbb{H}_\gamma$. Moreover

$$P(\mathbf{q}_0, A) := \mathbb{P}(Q_1(\mathbf{q}_0) \in A) \tag{2.16}$$

defines a Feller Markov transition kernel on $\mathbb{H}_\gamma$.

Proof The first item follows from a standard Banach fixed point argument, i.e. it suffices to show that, given any $(\mathbf{q}_0, \mathbf{v}_0) \in \mathbb{H}_\gamma \times \mathbb{H}_\gamma$ and any $t_0 \in \mathbb{R}$, the mapping

$$G(\mathbf{p}, \mathbf{u})(t) := (\mathbf{q}_0, \mathbf{v}_0) + \int_{t_0}^{t} (\mathbf{u}(s), -\mathbf{p}(s) - \mathcal{C}DU(\mathbf{p}(s))ds,$$

is a contraction mapping on the space of continuous $(\mathbb{H}_\gamma \times \mathbb{H}_\gamma)$-valued functions defined on $I := [t_0 - \delta, t_0 + \delta] \subset \mathbb{R}$, that is on $C(I; \mathbb{H}_\gamma \times \mathbb{H}_\gamma)$, for some $\delta > 0$ sufficiently small independent of $(\mathbf{q}_0, \mathbf{v}_0)$ and t_0.

Observe that, with (2.9) and (2.4),

$$|\mathcal{C}^{1-\gamma}(DU(\mathbf{p}) - DU(\tilde{\mathbf{p}}))| \leq \lambda_1^{1-2\gamma} L_1 |\mathcal{C}^{-\gamma}(\mathbf{p} - \tilde{\mathbf{p}})| \quad \text{for all } \mathbf{p}, \tilde{\mathbf{p}} \in \mathbb{H}_\gamma. \tag{2.17}$$

Thus, for any $(\mathbf{p}, \mathbf{u}), (\tilde{\mathbf{p}}, \tilde{\mathbf{u}}) \in C(I; \mathbb{H}_\gamma \times \mathbb{H}_\gamma)$, using (2.14) and (2.17),

$$\sup_{t\in I} |G(\mathbf{p}, \mathbf{u})(t) - G(\tilde{\mathbf{p}}, \tilde{\mathbf{u}})(t)|_\gamma \leq \delta(1 + \lambda_1^{1-2\gamma} L_1) \sup_{t\in I} |(\mathbf{p}, \mathbf{u}) - (\tilde{\mathbf{p}}, \tilde{\mathbf{u}})|_\gamma .$$

Therefore, G is a contraction mapping on $C(I; \mathbb{H}_\gamma \times \mathbb{H}_\gamma)$ for $\delta < (1 + \lambda_1^{1-2\gamma} L_1)^{-1}$. Similar argumentation establishes the desired continuity of Ξ_t, thus completing the proof. □

2.4 Formulation of the preconditioned Hamiltonian Monte Carlo chain

Having fixed an integration time $T > 0$, we denote by $Q_n(\mathbf{q}_0)$ as a random variable arising as the n step dynamics of the exact Preconditioned Hamiltonian Monte Carlo (PHMC) chain (2.16) starting from $\mathbf{q}_0 \in \mathbb{H}$. Namely, we iteratively draw $Q_n(\mathbf{q}_0) \sim P(Q_{n-1}(\mathbf{q}_0), \cdot)$ for $n \geq 1$ starting from $Q_0(\mathbf{q}_0) = \mathbf{q}_0$. We can write $Q_n(\mathbf{q}_0)$ more explicitly as a transformation of the sequence of Gaussian draws for the velocity as follows: Let $\mathbb{H}^{\otimes n}$ denote the product of n copies of $\mathbb{H}$. Given a sequence $\{\mathbf{v}_0^{(j)}\}_{j \in \mathbb{N}}$ of i.i.d. draws from $\mathcal{N}(0, \mathcal{C})$, we denote by $\mathbf{V}_0^{(n)}$ the noise path

$$\mathbf{V}_0^{(n)} := (\mathbf{v}_0^{(1)}, \ldots, \mathbf{v}_0^{(n)}) \sim \mathcal{N}(0, \mathcal{C})^{\otimes n}, \tag{2.18}$$

where $\mathcal{N}(0, \mathcal{C})^{\otimes n}$ denotes the measure on $\mathbb{H}^{\otimes n}$ given as the product of n copies of $\mathcal{N}(0, \mathcal{C})$. Taking $\mathcal{B}(\mathbb{H})$ to be the Borel σ-algebra on $\mathbb{H}$, we define $Q_1(\mathbf{q}_0) : \mathbb{H} \to \mathbb{H}$ to be the Borel random variable defined as

$$Q_1(\mathbf{q}_0)(\mathbf{v}_0^{(1)}) = \mathbf{q}_t(\mathbf{q}_0, \mathbf{v}_0^{(1)}) \quad \text{where } \mathbf{v}_0^{(1)} \sim \mathcal{N}(0, \mathcal{C}).$$

Iteratively, we define for every $n \geq 2$ the Borel random variable $Q_n(\mathbf{q}_0) : \mathbb{H}^{\otimes n} \to \mathbb{H}$ given by

$$Q_n(\mathbf{q}_0)(\mathbf{V}_0^{(n)}) = \mathbf{q}_t(Q_{n-1}(q_0)(\mathbf{V}_0^{(n-1)}), \mathbf{v}_0^{(n)}) \quad \text{where } \mathbf{V}_0^{(n)} \sim \mathcal{N}(0, \mathcal{C})^{\otimes n}. \tag{2.19}$$

With these notations we can write the n-step iterated transition kernels as

$$P^n(\mathbf{q}_0, A) := \mathbb{P}(Q_n(\mathbf{q}_0) \in A) \tag{2.20}$$

for any $\mathbf{q}_0 \in \mathbb{H}_\gamma$ and $A \in \mathcal{B}(\mathbb{H}_\gamma)$. Or, equivalently, $P^n(\mathbf{q}_0, \cdot)$ is the push-forward of $\mathcal{N}(0, \mathcal{C})^{\otimes n}$ by the mapping $Q_n(\mathbf{q}_0)$, i.e.

$$P^n(\mathbf{q}_0, A) = Q_n(\mathbf{q}_0)^* \mathcal{N}(0, \mathcal{C})^{\otimes n}(A) = \mathcal{N}(0, \mathcal{C})^{\otimes n}(Q_n(\mathbf{q}_0)^{-1}(A)) \tag{2.21}$$

for every $\mathbf{q}_0 \in \mathbb{H}_\gamma$ and $A \in \mathcal{B}(\mathbb{H}_\gamma)$.

We recall an invariance result for (1.1) from [6] in our setting.

Proposition 13 *Under the conditions given in Proposition 12 and additionally imposing Assumption 10 we have that*

$$\mathfrak{M}(d\mathbf{q}, d\mathbf{v}) \propto e^{-U(q)} \mu_0(d\mathbf{q}) \times \mu_0(d\mathbf{v})$$

defines a probability measure on $\mathbb{H}_\gamma \times \mathbb{H}_\gamma$ *which is invariant under* $\{\Xi_t\}_{t\geq 0}$ *namely*

$$\int_{\mathbb{H}_\gamma \times \mathbb{H}_\gamma} f(\Xi_t(\mathbf{q}, \mathbf{v}))\mathfrak{M}(d\mathbf{q}, d\mathbf{v}) = \int_{\mathbb{H}_\gamma \times \mathbb{H}_\gamma} f(\mathbf{q}, \mathbf{v})\mathfrak{M}(d\mathbf{q}, d\mathbf{v})$$

holds for every $f \in C_b(\mathbb{H}_\gamma \times \mathbb{H}_\gamma)$ *and every* $t \geq 0$. *As a consequence,* μ *given in* (1.1) *is a Borel probability measure on* $\mathbb{H}_\gamma$ *which is invariant for* P *defined by* (2.16).

3 A priori bounds for the deterministic dynamics

This section provides various a priori bounds on the dynamics specified by (1.2). The proofs rely solely on the bound on D^2U given in (2.7). In fact, they are obtained by using inequalities (2.9) and (2.10), that follow as a consequence of (2.7).

Proposition 14 *Impose Assumptions 2 and 8, (B1) and fix any* $T \in \mathbb{R}^+$ *satisfying*

$$T \leq (1 + \lambda_1^{1-2\gamma} L_1)^{-1/2}, \tag{3.1}$$

where the constant L_1 *is given in* (2.9) *and* λ_1 *is the top eigenvalue of* $\mathcal{C}$. *Then the dynamics defined by* (1.2) *maintains the bounds*

$$\begin{aligned}\sup_{t\in[0,T]} &|\mathbf{q}_t(\mathbf{q}_0, \mathbf{v}_0) - (\mathbf{q}_0 + t\mathbf{v}_0)|_\gamma \\ &\leq (1 + \lambda_1^{1-2\gamma} L_1)T^2 \max\{|q_0|_\gamma\,, |\mathbf{q}_0 + T\mathbf{v}_0|_\gamma\} + \lambda_1^{1-2\gamma} L_0 T^2\end{aligned} \tag{3.2}$$

and

$$\begin{aligned}\sup_{t\in[0,T]} |\mathbf{v}(t) - \mathbf{v}_0|_\gamma &\leq (1+\lambda_1^{1-2\gamma} L_1)T[1 + (1 + \lambda_1^{1-2\gamma} L_1)T^2] \max\left\{|\mathbf{q}_0|_\gamma\,, |\mathbf{q}_0 + T\mathbf{v}_0|_\gamma\right\} \\ &+\lambda_1^{1-2\gamma} L_0 T[1 + (1 + \lambda_1^{1-2\gamma} L_1)T^2],\end{aligned} \tag{3.3}$$

for any $(\mathbf{q}_0, \mathbf{v}_0) \in \mathbb{H}_\gamma \times \mathbb{H}_\gamma$, *with* L_0 *as given in* (2.10).

Proof Integrating the first equation in (1.2) twice and then applying the operator $\mathcal{C}^{-\gamma}$, we obtain

$$\mathcal{C}^{-\gamma}\mathbf{q}_t = \mathcal{C}^{-\gamma}(\mathbf{q}_0 + t\mathbf{v}_0) - \int_0^t \int_0^s \left[\mathcal{C}^{-\gamma}\mathbf{q}_\tau + \mathcal{C}^{1-\gamma} DU(\mathbf{q}_\tau)\right] d\tau ds, \tag{3.4}$$

for each $t \in [0, T]$. From Lemma 4 and inequality (2.10), we obtain

$$\begin{aligned}|\mathbf{q}_t - (\mathbf{q}_0 + t\mathbf{v}_0)|_\gamma &\leq (1 + \lambda_1^{1-2\gamma} L_1) \int_0^t \int_0^s |\mathbf{q}_\tau|_\gamma \, d\tau ds + \lambda_1^{1-2\gamma} L_0 \frac{T^2}{2} \\ &\leq (1 + \lambda_1^{1-2\gamma} L_1) \int_0^t \int_0^s |\mathbf{q}_\tau - (\mathbf{q}_0 + \tau\mathbf{v}_0)|_\gamma \, d\tau ds\end{aligned}$$

$$
\begin{aligned}
&+ (1+\lambda_1^{1-2\gamma}L_1)\int_0^t\int_0^s |\mathbf{q}_0+\tau\mathbf{v}_0|_\gamma\, d\tau ds + \lambda_1^{1-2\gamma}L_0\frac{T^2}{2}\\
&\le (1+\lambda_1^{1-2\gamma}L_1)\frac{T^2}{2}\sup_{\tau\in[0,T]}|\mathbf{q}_\tau-(\mathbf{q}_0+\tau\mathbf{v}_0)|_\gamma\\
&+ (1+\lambda_1^{1-2\gamma}L_1)\frac{T^2}{2}\max\{|\mathbf{q}_0|_\gamma, |\mathbf{q}_0+T\mathbf{v}_0|_\gamma\} + \lambda_1^{1-2\gamma}L_0\frac{T^2}{2}.
\end{aligned} \tag{3.5}
$$

Here note that, using the convexity of the function $f(\tau)=|\mathbf{q}_0+\tau\mathbf{v}_0|_\gamma$, we have

$$
\sup_{\tau\in[0,T]}|\mathbf{q}_0+\tau\mathbf{v}_0|_\gamma \le \max\{|\mathbf{q}_0|_\gamma, |\mathbf{q}_0+T\mathbf{v}_0|_\gamma\} \tag{3.6}
$$

which we used in the final bound in (3.5). Thus, using assumption (3.1) and taking the supremum with respect to $t\in[0,T]$ in (3.5), we conclude the first bound (3.2).

Turn next to second bound (3.3), integrating the second equation in (1.2) once and using Lemma 4 and inequality (2.10) again, we have

$$
\begin{aligned}
|\mathbf{v}_t-\mathbf{v}_0|_\gamma &\le (1+\lambda_1^{1-2\gamma}L_1)\int_0^t|\mathbf{q}_s|_\gamma\, ds+\lambda_1^{1-2\gamma}L_0 t && (3.7)\\
&\le (1+\lambda_1^{1-2\gamma}L_1)T\sup_{s\in[0,T]}|\mathbf{q}_\tau|_\gamma+\lambda_1^{1-2\gamma}L_0T && (3.8)
\end{aligned}
$$

for every $t\in[0,T]$. From (3.2), it follows that

$$
\sup_{t\in[0,T]}|\mathbf{q}_s|_\gamma \le [1+(1+\lambda_1^{1-2\gamma}L_1)T^2]\max\{|\mathbf{q}_0|_\gamma, |\mathbf{q}_0+T\mathbf{v}_0|_\gamma\}+\lambda_1^{1-2\gamma}L_0T^2. \tag{3.9}
$$

Hence, we conclude (3.3) from (3.7) and (3.9), completing the proof. □

Proposition 15 *Impose Assumptions 2, 8, (B1) and consider any $T\in\mathbb{R}^+$ satisfying*

$$
T\le(1+\lambda_1^{1-2\gamma}L_1)^{-1/2}, \tag{3.10}
$$

where L_1 is as in (2.7) and λ_1 is the top eigenvalue of $\mathcal{C}$. Then, for any $(\mathbf{q}_0,\mathbf{v}_0)$, $(\tilde{\mathbf{q}}_0,\tilde{\mathbf{v}}_0)\in\mathbb{H}_\gamma\times\mathbb{H}_\gamma$,

$$
\begin{aligned}
&\sup_{t\in[0,T]}|\mathbf{q}_t(\mathbf{q}_0,\mathbf{v}_0)-\mathbf{q}_t(\tilde{\mathbf{q}}_0,\tilde{\mathbf{v}}_0)-(\mathbf{q}_0-\tilde{\mathbf{q}}_0)-t(\mathbf{v}_0-\tilde{\mathbf{v}}_0)|_\gamma\\
&\le(1+\lambda_1^{1-2\gamma}L_1)T^2\max\left\{|\mathbf{q}_0-\tilde{\mathbf{q}}_0|_\gamma, |(\mathbf{q}_0-\tilde{\mathbf{q}}_0)+T(\mathbf{v}_0-\tilde{\mathbf{v}}_0)|_\gamma\right\}.
\end{aligned} \tag{3.11}
$$

Remark 16 Observe that, given any $\mathbf{q}_0,\tilde{\mathbf{q}}_0,\mathbf{v}_0\in\mathbb{H}_\gamma$, by choosing

$$
\tilde{\mathbf{v}}_0:=\mathbf{v}_0+\frac{1}{T}(\mathbf{q}_0-\tilde{\mathbf{q}}_0), \tag{3.12}
$$

then under (3.11) we obtain

$$|\mathcal{C}^{-\gamma}\left[\mathbf{q}_T(\mathbf{q}_0, \mathbf{v}_0) - \mathbf{q}_T(\tilde{\mathbf{q}}_0, \tilde{\mathbf{v}}_0)\right]| \leq (1 + \lambda_1^{1-2\gamma} L_1) T^2 |\mathcal{C}^{-\gamma}(\mathbf{q}_0 - \tilde{\mathbf{q}}_0)|, \tag{3.13}$$

which thus yields a contraction when $T < (1 + \lambda_1^{1-2\gamma} L_1)^{-1/2}$. This observation for the initial conditions in (3.12) has previously been employed in [14] and, in the finite dimensional case where $\mathbb{H} = \mathbb{R}^k$ for some $k \in \mathbb{N}$, this bound can be used directly as a crucial step towards establishing the ρ-smallness and ρ-contraction conditions for the weak Harris theorem in [43], as we illustrate below in Sect. 7.

The idea behind definition (3.12) comes from the fact that for the simplified version of the dynamics in (1.2) where $d\mathbf{v}_t/dt = 0$, the positions of two associated trajectories starting from $(\mathbf{q}_0, \mathbf{v}_0)$ and $(\tilde{\mathbf{q}}_0, \tilde{\mathbf{v}}_0)$, with $\tilde{\mathbf{v}}_0$ as in (3.12), will coincide at time T. With a similar line of reasoning, one could consider a slighly better approximation of the dynamics in (1.2) by assuming instead $U = 0$, in which case the associated dynamics $d\mathbf{q}_t/dt = \mathbf{v}_t$, $d\mathbf{v}_t/dt = \mathbf{q}_t$ describes the motion of a simple pendulum. Here by defining $\tilde{\mathbf{v}}_0 = \mathbf{v}_0 + (\mathbf{q}_0 - \tilde{\mathbf{q}}_0)(\cos T / \sin T)$ one again concludes that the positions of two trajectories starting from $(\mathbf{q}_0, \mathbf{v}_0)$ and $(\tilde{\mathbf{q}}_0, \tilde{\mathbf{v}}_0)$ coincide after time T. While we could obtain similar results by using the latter approach, this would require the same type of assumptions we already impose in the first case, thus not showing a significant difference at least at the theoretical level. For simplicity, we then chose the first approach for our presentation. We remark however that the second approach, as being associated to a better approximation of (1.2), could lead to slightly less stringent constants on the conditions for the integration time T in comparison to (3.10).

More generally, we may view (3.12) as addressing a control problem. In fact, the methodology of the weak Harris theorem developed here could in principle allow the use of a wide variety of controls. More specifically, we are interested in any 'reasonable' mapping $\Psi : \mathbb{H}_\gamma \times \mathbb{H}_\gamma \times \mathbb{H}_\gamma \to \mathbb{H}_\gamma$ such that, for any $\mathbf{q}_0, \tilde{\mathbf{q}}_0, \mathbf{v}_0 \in \mathbb{H}_\gamma$ and any suitable value of $T > 0$, one would have

$$\mathbf{q}_T(\mathbf{q}_0, \mathbf{v}_0) \approx \mathbf{q}_T(\tilde{\mathbf{q}}_0, \Psi(\mathbf{q}_0, \tilde{\mathbf{q}}_0, \mathbf{v}_0)).$$

In this connection one might hope to make a more delicate use of the Hamiltonian dynamics, presumably tailored to the fine properties of a particular potential U of interest, to obtain refined results on convergence to equilibrium. In particular, we expect that the constraints imposed on T by Proposition 14 are overzealous, and could potentially be improved by a different type of control.

On the other hand, in the infinite dimensional Hilbert space setting which we are primarily focused on here, even (3.12) is insufficient for the aim of establishing contractivity in the Markovian dynamics, as the law of this choice of $\tilde{\mathbf{v}}_0$ is not generically absolutely continuous with respect to the law of $\mathbf{v}_0$; cf. Propositions 22 and 24 below. We proceed instead by using the refinement (3.19) which is shown to produce a contraction in Proposition 18. Here we are making use of some of the intuition and approach to ergodicity in the stochastic fluids literature, cf. [37,55,65]. In these works one modifies the noise path on low modes with the expectation that if one induces a contraction on the large scale dynamics for sufficiently many low frequency modes then the high frequencies (or small scales) will also contract, being enslaved to the behavior of the system at large scales. This effect, sometimes referred as a Foias-Prodi bound [35], is widely observed in the fluids and infinite dimensional dynamical systems literature.

Proof *(Proof of Proposition 15)* Let $\mathbf{z}_t = \mathbf{q}_t(\mathbf{q}_0, \mathbf{v}_0) - \mathbf{q}_t(\widetilde{\mathbf{q}}_0, \widetilde{\mathbf{v}}_0)$ and $\mathbf{w}_t = d\mathbf{z}_t/dt$. Then, for any $t > 0$, $\mathbf{z}_t$ satisfies

$$\frac{d^2\mathbf{z}_t}{dt^2} = -\mathbf{z}_t - \mathcal{C}g(t) \tag{3.14}$$

where

$$g(t) := DU(\mathbf{q}_t(\mathbf{q}_0, \mathbf{v}_0)) - DU(\mathbf{q}_t(\widetilde{\mathbf{q}}_0, \widetilde{\mathbf{v}}_0)). \tag{3.15}$$

Therefore, for every $t \geq 0$,

$$\mathcal{C}^{-\gamma}\mathbf{z}_t = \mathcal{C}^{-\gamma}(\mathbf{z}_0 + t\mathbf{w}_0) - \int_0^t \int_0^s [\mathcal{C}^{-\gamma}\mathbf{z}_\tau + \mathcal{C}^{1-\gamma}g(\tau)]d\tau ds.$$

By using Lemma 4 and inequality (2.9), we obtain

$$\begin{aligned} |\mathbf{z}_t - (\mathbf{z}_0 + t\mathbf{w}_0)|_\gamma &\leq \int_0^t \int_0^s \left[|\mathbf{z}_\tau|_\gamma + \lambda_1^{1-2\gamma}\,|g(\tau)|_{-\gamma}\right] d\tau ds \\ &\leq (1 + \lambda_1^{1-2\gamma} L_1) \int_0^t \int_0^s |\mathbf{z}_\tau|_\gamma \, d\tau ds. \end{aligned}$$

The remaining portion of the proof follows analogously as in the proof of (3.2). □

In view of Remark 16 the bounds in Proposition 15 are not sufficient for our application to prove the ρ-contractivity and ρ-smallness conditions for the weak Harris theorem below in Sect. 5. For this purpose we consider a modified version of (3.12) where the shift only involves a low-modes finite-dimensional approximation of $\mathbf{q}_0 - \tilde{\mathbf{q}}_0$.

Before proceeding let us introduce some notation. Split $\mathbb{H}$ into a space $\mathbb{H}_N := \text{span}\{\mathbf{e}_1, \cdots, \mathbf{e}_N\}$ and its orthogonal complement $\mathbb{H}^N$; so that $\mathbb{H} = \mathbb{H}_N \oplus \mathbb{H}^N$ where N satisfies the second condition in (3.20), below. Recall, as in (2.3), that, given $\mathbf{f} \in \mathbb{H}$, we denote by $\Pi_N \mathbf{f}$ and $\Pi^N \mathbf{f}$ the orthogonal projections onto $\mathbb{H}_N$ and $\mathbb{H}^N$, respectively. This splitting is defined such that the Lipschitz constant of the projection of $-\mathcal{C}DU(\mathbf{f})$ onto $\mathbb{H}^N$ is at most $1/4$.

For any $\gamma \in [0, 1/2)$ and $\alpha \in \mathbb{R}^+$, we consider the following auxiliary norm:

$$|\mathbf{f}|_{\gamma,\alpha} := |\Pi_N \mathbf{f}|_\gamma + \alpha|\Pi^N \mathbf{f}|_\gamma, \text{ for any } \mathbf{f} \in \mathbb{H}_\gamma. \tag{3.16}$$

Remark 17 Notice that $|\cdot|_{\gamma,\alpha}$ is equivalent to $|\cdot|_\gamma$ and

$$\min\{1, \alpha\}\,|\mathbf{f}|_\gamma \leq |\mathbf{f}|_{\gamma,\alpha} \leq \sqrt{2}\max\{1, \alpha\}\,|\mathbf{f}|_\gamma\,, \quad \text{for all } \mathbf{f} \in \mathbb{H}_\gamma. \tag{3.17}$$

In particular, for α defined as in (3.21) below, we have

$$|\mathbf{f}|_\gamma \leq |\mathbf{f}|_{\gamma,\alpha} \leq \sqrt{2}\alpha \, |\mathbf{f}|_\gamma \,, \quad \text{for all } \mathbf{f} \in \mathbb{H}_\gamma. \tag{3.18}$$

Proposition 18 *Impose Assumptions 2, 8, (B1). Let $(\mathbf{q}_0, \mathbf{v}_0)$, $(\tilde{\mathbf{q}}_0, \tilde{\mathbf{v}}_0) \in \mathbb{H}_\gamma \times \mathbb{H}_\gamma$ such that*

$$\Pi^N \tilde{\mathbf{v}}_0 = \Pi^N \mathbf{v}_0 \quad \textit{and} \quad \Pi_N \tilde{\mathbf{v}}_0 = \Pi_N \mathbf{v}_0 + T^{-1}(\Pi_N \mathbf{q}_0 - \Pi_N \tilde{\mathbf{q}}_0). \tag{3.19}$$

Assume that $T \in \mathbb{R}^+$ and $N \in \mathbb{N}$ satisfy

$$T \leq \frac{1}{[2(1 + \lambda_1^{1-2\gamma} L_1)]^{1/2}} \quad \textit{and} \quad \lambda_{N+1}^{1-2\gamma} \leq \frac{1}{4L_1}, \tag{3.20}$$

and let

$$\alpha = 4(1 + \lambda_1^{1-2\gamma} L_1). \tag{3.21}$$

Here γ is specified in Assumption 8, L_1 is as in (2.7) and λ_j represent the eigenvalues of $\mathcal{C}$ in descending order as in Assumption 2. Then,

$$|\mathbf{q}_T(\mathbf{q}_0, \mathbf{v}_0) - \mathbf{q}_T(\tilde{\mathbf{q}}_0, \tilde{\mathbf{v}}_0)|_{\gamma,\alpha} \leq \kappa_1 \, |\mathbf{q}_0 - \tilde{\mathbf{q}}_0|_{\gamma,\alpha} \,, \tag{3.22}$$

where $|\cdot|_{\gamma,\alpha}$ is the norm defined in (3.16) and

$$\kappa_1 = 1 - \frac{T^2}{12}.$$

Proof As in the proof of Proposition 15, let us denote $\mathbf{z}_t := \mathbf{q}_t(\mathbf{q}_0, \mathbf{v}_0) - \mathbf{q}_t(\tilde{\mathbf{q}}_0, \tilde{\mathbf{v}}_0)$ and $\mathbf{w}_t = d\mathbf{z}_t/dt$, for all $t \geq 0$. Notice that

$$\Pi_N \mathbf{z}_0 + T \Pi_N \mathbf{w}_0 = 0 \quad \text{and} \quad \Pi^N \mathbf{w}_0 = 0. \tag{3.23}$$

Applying $\mathcal{C}^{-\gamma}$ to (3.14), projecting onto $\mathbb{H}_N$ and integrating, yields

$$\mathcal{C}^{-\gamma} \Pi_N \mathbf{z}_T = -\int_0^T \int_0^s \Big[\mathcal{C}^{-\gamma} \Pi_N \mathbf{z}_\tau + \mathcal{C}^{1-\gamma} \Pi_N g(\tau) \Big] d\tau ds,$$

with $g(\cdot)$ defined as in (3.15). Thus, using (2.4) in Lemma 4 and (2.7) of Assumption 8, we estimate

$$\begin{aligned} |\Pi_N \mathbf{z}_T|_\gamma \leq & \int_0^T \int_0^s \Big[|\mathbf{z}_\tau|_\gamma + \lambda_1^{1-2\gamma} \, |g(\tau)|_{-\gamma} \Big] d\tau ds \leq (1 + \lambda_1^{1-2\gamma} L_1) \frac{T^2}{2} \sup_{s \in [0,T]} |\mathbf{z}_s|_\gamma \\ = & \frac{\alpha T^2}{8} \sup_{s \in [0,T]} |\mathbf{z}_s|_\gamma \,. \end{aligned} \tag{3.24}$$

On the other hand, by Duhamel's formula, we have

$$\mathbf{z}_T = \mathbf{z}_0 \cos(T) + \mathbf{w}_0 \sin(T) - \int_0^T \sin(T-s)\, \mathcal{C} g(s) ds,$$

and hence, with (3.23),

$$\mathcal{C}^{-\gamma} \Pi^N \mathbf{z}_T = \mathcal{C}^{-\gamma} \Pi^N \mathbf{z}_0 \cos(T) - \int_0^T \sin(T-s)\, \mathcal{C}^{1-\gamma} \Pi^N g(s) ds$$

Now, using (ii) of Lemma 4 and $(B1)$ of Assumption 8, we estimate

$$\begin{aligned}\left|\Pi^N \mathbf{z}_T\right|_\gamma &\leq \left|\Pi^N \mathbf{z}_0\right|_\gamma \cos(T) + \lambda_{N+1}^{1-2\gamma} L_1 \int_0^T \sin(T-s)\, |\mathbf{z}_s|_\gamma \, ds \\ &\leq \left|\Pi^N \mathbf{z}_0\right|_\gamma \cos(T) + \frac{1-\cos(T)}{4} \sup_{s\in[0,T]} |\mathbf{z}_s|_\gamma \,.\end{aligned}$$

where for the final inequality we used the second condition in (3.20). Therefore, using that $\cos(s) \leq 1 - s^2/2 + s^4/24$ and $1-\cos(s) \leq s^2/2$ for every $s \in \mathbb{R}$, yields

$$\left|\Pi^N \mathbf{z}_T\right|_\gamma \leq \left(1 - \frac{T^2}{2} + \frac{T^4}{24}\right) \left|\Pi^N \mathbf{z}_0\right|_\gamma + \frac{T^2}{8} \sup_{s\in[0,T]} |\mathbf{z}_s|_\gamma \,. \tag{3.25}$$

From Proposition 15 and a bound as in (3.6) it follows that

$$\sup_{s\in[0,T]} |\mathbf{z}_s|_\gamma \leq [1 + (1+\lambda_1^{1-2\gamma} L_1) T^2] \max\left\{ |\mathbf{z}_0|_\gamma \,, |\mathbf{z}_0 + T\mathbf{w}_0|_\gamma \right\}.$$

However from (3.23) we have $\mathbf{z}_0 + T\mathbf{w}_0 = \Pi^N \mathbf{z}_0$, so that $\max\{|\mathbf{z}_0|_\gamma \,, |\mathbf{z}_0 + T\mathbf{w}_0|_\gamma\} = |\mathbf{z}_0|_\gamma$. With this and the first condition in (3.20), we therefore obtain

$$\sup_{s\in[0,T]} |\mathbf{z}_s|_\gamma \leq [1 + (1+\lambda_1^{1-2\gamma} L_1) T^2]\, |\mathbf{z}_0|_\gamma \leq \frac{3}{2} |\mathbf{z}_0|_\gamma \,. \tag{3.26}$$

Using (3.26) in (3.24) and in (3.25), we obtain

$$|\Pi_N \mathbf{z}_T|_\gamma \leq \frac{3\alpha T^2}{16} |\mathbf{z}_0|_\gamma$$

and

$$\left|\Pi^N \mathbf{z}_T\right|_\gamma \leq \left(1 - \frac{T^2}{2} + \frac{T^4}{24}\right) \left|\Pi^N \mathbf{z}_0\right|_\gamma + \frac{3T^2}{16} |\mathbf{z}_0|_\gamma \,,$$

so that finally

$$
\begin{aligned}
|\mathbf{z}_T|_{\gamma,\alpha} &= |\Pi_N \mathbf{z}_T|_\gamma + \alpha \left|\Pi^N \mathbf{z}_T\right|_\gamma \leq \frac{3\alpha T^2}{8} |\mathbf{z}_0|_\gamma + \alpha \left(1 - \frac{T^2}{2} + \frac{T^4}{24}\right) \left|\Pi^N \mathbf{z}_0\right|_\gamma \\
&\leq \frac{3\alpha T^2}{8} |\Pi_N \mathbf{z}_0|_\gamma + \alpha \left(1 - \frac{T^2}{8} + \frac{T^4}{24}\right) \left|\Pi^N \mathbf{z}_0\right|_\gamma . \qquad (3.27)
\end{aligned}
$$

From the first condition in (3.20) and the definition of α in (3.21), it follows in particular that $\alpha T^2 \leq 2$ and also $T \leq 1$, so that $T^4 \leq T^2$. Therefore, from (3.27), we have

$$
\begin{aligned}
|\mathbf{z}_T|_{\gamma,\alpha} &\leq \frac{3}{4} |\Pi_N \mathbf{z}_0|_\gamma + \alpha \left(1 - \frac{T^2}{12}\right) \left|\Pi^N \mathbf{z}_0\right|_\gamma \\
&\leq \max\left\{1 - \frac{T^2}{12}, \frac{3}{4}\right\} |\mathbf{z}_0|_{\gamma,\alpha} = \left(1 - \frac{T^2}{12}\right) |\mathbf{z}_0|_{\gamma,\alpha} ,
\end{aligned}
$$

where the equality above follows again from the fact that $T \leq 1$, by the first condition in (3.20). This completes the proof. $\square$

4 Foster–Lyapunov structure

This section provides the details of the Foster–Lyapunov structure for the Markov kernel P defined by (2.16) under Assumption 5 and 8. First, we recall the underlying definition:

Definition 19 We say that $V : \mathbb{H}_\gamma \to \mathbb{R}^+$ is a *Foster–Lyapunov* (or, simply, a *Lyapunov*) function for the Markov kernel P if V is integrable with respect to $P^n(\mathbf{q}, \cdot)$ for every $\mathbf{q} \in \mathbb{H}$ and $n \in \mathbb{N}$, and satisfies the following inequality

$$
P^n V(\mathbf{q}) \leq C\kappa^n V(\mathbf{q}) + K \quad \text{for all } \mathbf{q} \in \mathbb{H} \text{ and } n \in \mathbb{N}, \qquad (4.1)
$$

for some constants $\kappa \in (0, 1)$ and $C, K > 0$.

With this definition in hand the main result of this section is as follows:

Proposition 20 *Impose Assumption 2, 5 and 8 and suppose that $T \in \mathbb{R}^+$ satisfies*

$$
T \leq \min\left\{\frac{1}{[2(1 + \lambda_1^{1-2\gamma} L_1)]^{1/2}}, \frac{L_2^{1/2}}{2\sqrt{6}(1 + \lambda_1^{1-2\gamma} L_1)}\right\}, \qquad (4.2)
$$

where L_1 and L_2 are defined as in (2.7), (2.8), respectively and λ_1 is the largest eigenvalue of $\mathcal{C}$. Then, the functions

$$
V_{1,i}(\mathbf{q}) = |\mathbf{q}|_\gamma^i , \quad i \in \mathbb{N}, \qquad (4.3)
$$

and

$$V_{2,\eta}(\mathbf{q}) = \exp(\eta\, |\mathbf{q}|_\gamma^2), \tag{4.4}$$

with $\eta \in \mathbb{R}^+$ *satisfying*

$$\eta < \left[c\, \mathrm{Tr}(\mathcal{C}^{1-2\gamma})\left(L_2^{-1} + T^2\right)\right]^{-1}, \tag{4.5}$$

for a suitable absolute constant $c \in \mathbb{R}^+$*, are Lyapunov functions for the Markov kernel* P *defined in* (2.16)*.*

Remark 21 Before delving into the proof, some heuristic remarks are in order here concerning why we might expect a dissipative structure à la (4.1) for the HMC chain (2.20) not withstanding non-dissipative nature of Hamiltonian systems in general. Starting from a current position $\mathbf{q}_0$, we draw an initial velocity $\mathbf{v}_0 \sim \mu_0 = N(0, \mathcal{C})$. If $\mathbf{q}_0$ is sufficiently far from the origin and the core of the distribution μ_0 then, on average, the Hamiltonian system (1.2) starts at $(\mathbf{q}_0, \mathbf{v}_0)$ with a potential energy which is large with respect to its kinetic energy. We may therefore expect that the dynamics (1.2) converts some of this potential energy into kinetic energy. Thus, while total energy is conserved along the Hamiltonian path, we may expect this energy to change its form and to be converted from potential to kinetic energy. This transfer of energy is then lost on average when we reset the velocity component as we start the next step of the chain.

The forthcoming bounds reflect that such an energy conversion can be made explicit and quantitative at the level of the simple case of the pendulum $d\bar{\mathbf{q}}_t/dt = \bar{\mathbf{v}}_t, d\bar{\mathbf{v}}/dt = -\bar{\mathbf{q}}_t$. Our estimates then show that the presence of the 'nonlinear term' $\mathcal{C}DU(\mathbf{q}_t)$ does not change this picture at least for a small time and so long as the tail condition (2.8) holds. We may expect that other relevant mechanisms for energy transfer from potential to kinetic energy may be exploited to a similar effect in future studies.

Proof We start by showing that $V_{1,2}(\mathbf{q}) = |\mathbf{q}|_\gamma^2$ is a Lyapunov function for P. First, notice $\frac{d}{dt}\, |\mathbf{q}_t|_\gamma^2 = 2\langle \mathbf{q}_t, \mathbf{v}_t\rangle_\gamma$ so that

$$|\mathbf{q}_T|_\gamma^2 = |\mathbf{q}_0|_\gamma^2 + 2\int_0^T \langle \mathbf{q}_s, \mathbf{v}_s\rangle_\gamma ds. \tag{4.6}$$

Moreover, from (1.2)

$$\frac{d}{ds}\langle \mathbf{q}_s, \mathbf{v}_s\rangle_\gamma = |\mathbf{v}_s|_\gamma^2 - |\mathbf{q}_s|_\gamma^2 - \langle \mathbf{q}_s, \mathcal{C}\nabla U(\mathbf{q}_s)\rangle_\gamma. \tag{4.7}$$

Hence, using Assumption 8, (B2),

$$
\begin{aligned}
\langle \mathbf{q}_s, \mathbf{v}_s\rangle_\gamma &= \langle \mathbf{q}_0, \mathbf{v}_0\rangle_\gamma + \int_0^s \Big[|\mathbf{v}_\tau|_\gamma^2 - |\mathbf{q}_\tau|_\gamma^2 - \langle \mathbf{q}_\tau, \mathcal{C}\nabla U(\mathbf{q}_\tau)\rangle_\gamma \Big] d\tau \\
&\le \langle \mathbf{q}_0, \mathbf{v}_0\rangle_\gamma + \int_0^s \Big[|\mathbf{v}_\tau|_\gamma^2 - L_2\, |\mathbf{q}_\tau|_\gamma^2 + L_3 \Big] d\tau,
\end{aligned} \tag{4.8}
$$

for any $s \ge 0$. Using (4.8) in (4.6), we obtain

$$
|\mathbf{q}_T|_\gamma^2 \le |\mathbf{q}_0|_\gamma^2 + 2T\langle \mathbf{q}_0, \mathbf{v}_0\rangle_\gamma + 2\int_0^T \int_0^s \Big[|\mathbf{v}_\tau|_\gamma^2 - L_2\, |\mathbf{q}_\tau|_\gamma^2 + L_3 \Big] d\tau ds. \tag{4.9}
$$

From Proposition 14, (3.3) and hypothesis (4.2), it follows that

$$
|\mathbf{v}_\tau|_\gamma \le \frac{7}{4}\, |\mathbf{v}_0|_\gamma + \frac{3}{2}(1+\lambda_1^{1-2\gamma} L_1)\tau\, |\mathbf{q}_0|_\gamma + \frac{3}{2}\lambda_1^{1-2\gamma} L_0 \tau,
$$

so that

$$
|\mathbf{v}_\tau|_\gamma^2 \le \frac{49}{8}\, |\mathbf{v}_0|_\gamma^2 + 9(1+\lambda_1^{1-2\gamma} L_1)^2\tau^2\, |\mathbf{q}_0|_\gamma^2 + 9(\lambda_1^{1-2\gamma} L_0)^2\tau^2, \tag{4.10}
$$

which holds for any $\tau \ge 0$. Moreover, from (3.2) and using hypothesis (4.2) again, we obtain that

$$
|\mathbf{q}_\tau - (\mathbf{q}_0 + \tau \mathbf{v}_0)|_\gamma \le \frac{|\mathbf{q}_0|_\gamma}{2} + \frac{\tau}{2}\, |\mathbf{v}_0|_\gamma + \lambda_1^{1-2\gamma} L_0 \tau^2,
$$

so that

$$
|\mathbf{q}_\tau|_\gamma \ge \frac{|\mathbf{q}_0|_\gamma}{2} - \frac{3}{2}\tau\, |\mathbf{v}_0|_\gamma - \lambda_1^{1-2\gamma} L_0 \tau^2
$$

and, consequently,

$$
2\, |\mathbf{q}_\tau|_\gamma^2 \ge \frac{|\mathbf{q}_0|_\gamma^2}{4} - 9\tau^2\, |\mathbf{v}_0|_\gamma^2 - 4(\lambda_1^{1-2\gamma} L_0)^2\tau^4.
$$

Thus, from (2.13) and (4.2), it follows that

$$
\begin{aligned}
-2L_2\, |\mathbf{q}_\tau|_\gamma^2 &\le -\frac{L_2}{4}\, |\mathbf{q}_0|_\gamma^2 + 9L_2\tau^2\, |\mathbf{v}_0|_\gamma^2 + 4L_2(\lambda_1^{1-2\gamma} L_0)^2\tau^4 \\
&\le -\frac{L_2}{4}\, |\mathbf{q}_0|_\gamma^2 + 9(1+\lambda_1^{1-2\gamma} L_1)\tau^2\, |\mathbf{v}_0|_\gamma^2 + 4(1+\lambda_1^{1-2\gamma} L_1)(\lambda_1^{1-2\gamma} L_0)^2\tau^4 \\
&\le -\frac{L_2}{4}\, |\mathbf{q}_0|_\gamma^2 + \frac{9}{2}\, |\mathbf{v}_0|_\gamma^2 + 2(\lambda_1^{1-2\gamma} L_0)^2\tau^2,
\end{aligned} \tag{4.11}
$$

for any $\tau \ge 0$. Using (4.10) and (4.11) in (4.9), yields

$$
\begin{aligned}
|\mathbf{q}_T|_\gamma^2 \leq & \left(1 + \frac{3}{2}(1+\lambda_1^{1-2\gamma} L_1)^2 T^4 - \frac{L_2}{8} T^2\right) |\mathbf{q}_0|_\gamma^2 \\
& + 2T \langle \mathbf{q}_0, \mathbf{v}_0 \rangle_\gamma + \frac{67}{8} T^2 \, |\mathbf{v}_0|_\gamma^2 + \frac{5}{3}(\lambda_1^{1-2\gamma} L_0)^2 T^4 + L_3 T^2. \qquad (4.12)
\end{aligned}
$$

By hypothesis (4.2), we have that $3(1+\lambda_1^{1-2\gamma} L_1)^2 T^4/2 \leq L_2 T^2/16$. Thus,

$$
1 + \frac{3}{2}(1+\lambda_1^{1-2\gamma} L_1)^2 T^4 - \frac{L_2}{8} T^2 \leq 1 - \frac{L_2}{16} T^2 \leq e^{-\frac{L_2 T^2}{16}}, \qquad (4.13)
$$

where we used the fact that $1 - x \leq e^{-x}$, for every $x \geq 0$. Using (4.13) in (4.12) and taking expected values on both sides of the resulting inequality, and noting that, by symmetry $\mathbb{E}\langle \mathbf{q}_0, \mathbf{v}_0 \rangle_\gamma = 0$ we obtain

$$
P V_{1,2}(\mathbf{q}_0) = \mathbb{E}\, |\mathbf{q}_T|_\gamma^2 \leq e^{-\frac{L_2 T^2}{16}} \, |\mathbf{q}_0|_\gamma^2 + \left(\frac{67}{8} \operatorname{Tr}(\mathcal{C}^{1-2\gamma}) + \frac{5}{3}(\lambda_1^{1-2\gamma} L_0)^2 T^2 + L_3 \right) T^2. \qquad (4.14)
$$

Hence, after iterating on the result in (4.14) n times, we have

$$
\begin{aligned}
P^n V_{1,2}(\mathbf{q}_0) &= \mathbb{E}\, |Q_n(\mathbf{q}_0)|_\gamma^2 \\
&\leq e^{-\frac{n L_2 T^2}{16}} \, |\mathbf{q}_0|_\gamma^2 + \left(\frac{67}{8} \operatorname{Tr}(\mathcal{C}^{1-2\gamma}) + \frac{5}{3}(\lambda_1^{1-2\gamma} L_0)^2 T^2 + L_3 \right) T^2 \sum_{j=0}^{n-1} e^{-\frac{j L_2 t^2}{16}}.
\end{aligned} \qquad (4.15)
$$

Notice that

$$
T^2 \sum_{j=0}^{n-1} e^{-\frac{j L_2 T^2}{16}} \leq \frac{T^2}{1 - e^{-\frac{L_2 t^2}{16}}} \leq \frac{48}{L_2},
$$

where in the last inequality we used that $x/(1-e^{-x}) \leq e \leq 3$, for every $0 \leq x \leq 1$. Thus,

$$
P^n V_{1,2}(\mathbf{q}_0) \leq e^{-\frac{n L_2 T^2}{16}} \, |\mathbf{q}_0|_\gamma^2 + \left(\frac{67}{8} \operatorname{Tr}(\mathcal{C}^{1-2\gamma}) + \frac{5}{3}(\lambda_1^{1-2\gamma} L_0)^2 T^2 + L_3 \right) \frac{48}{L_2},
$$

which shows 19 for $V_{1,2}$.

We turn now to establish 19 in the general case of $V_{1,i}$, for any $i \in \mathbb{N}$. Here, invoking Young's inequality to estimate the term $2T \langle \mathbf{q}_0, \mathbf{v}_0 \rangle_\gamma$ in (4.12) as

$$
2T \langle \mathbf{q}_0, \mathbf{v}_0 \rangle_\gamma \leq \frac{L_2 T^2}{32} \, |\mathbf{q}_0|_\gamma^2 + \frac{32}{L_2} \, |\mathbf{v}_0|_\gamma^2 \, ,
$$

and using again that $3(1+\lambda_1^{1-2\gamma}L_1)^2T^4/2 \le L_2T^2/16$, it follows from (4.12) that

$$|\mathbf{q}_T|_\gamma^2 \le \left(1-\frac{L_2T^2}{32}\right)|\mathbf{q}_0|_\gamma^2 + \left(\frac{67}{8}T^2+\frac{32}{L_2}\right)|\mathbf{v}_0|_\gamma^2 + \frac{5}{3}(\lambda_1^{1-2\gamma}L_0)^2T^4 + L_3T^2. \tag{4.16}$$

Invoking the basic inequalities $1-x\le e^{-x}$ and $(x+y)^{1/2}\le x^{1/2}+y^{1/2}$, valid for every $x, y\ge 0$, we obtain, for any $i\ge 1$,

$$\begin{aligned}|\mathbf{q}_T|_\gamma^i &\le e^{-\frac{L_2T^2 i}{64}}|\mathbf{q}_0|_\gamma^i + C\sum_{j=1}^{i}\left(e^{-\frac{L_2T^2}{64}}|\mathbf{q}_0|_\gamma\right)^j\left(|\mathbf{v}_0|_\gamma^{i-j}+1\right)\\ &\le e^{-\frac{L_2T^2 i}{65}}|\mathbf{q}_0|_\gamma^i + \tilde{C}\left(|\mathbf{v}_0|_\gamma^i+1\right),\end{aligned} \tag{4.17}$$

where in the second inequality we invoked Young's inequality to estimate each term inside the sum, and with C and $\tilde{C}$ being positive constants depending on $i, \lambda_1, \gamma, T, L_0, L_2$ and L_3. Since $\mathbf{v}_0 \sim \mathcal{N}(0,\mathcal{C})$, by Fernique's theorem (see, e.g., [27, Theorem 2.7]) we have that $\mathbb{E}\,|\mathbf{v}_0|_\gamma^i < \infty$ for every $i\in\mathbb{N}$. Therefore, we conclude the result for $V_{1,i}$ after taking expected values in (4.17) and iterating n times on the resulting inequality.

Finally, let us show 19 for $V_{2,\eta}$ as in (4.4). Multiplying by η, taking the exponential and expected value on both sides of (4.16), it follows that

$$\begin{aligned}PV_2(\mathbf{q}_0) &= \mathbb{E}\exp\left(\eta\,|\mathbf{q}_T|_\gamma^2\right)\\ &\le \exp\left(\eta\left(1-\frac{L_2T^2}{32}\right)|\mathbf{q}_0|_\gamma^2\right)\exp\left(\frac{5}{3}\eta(\lambda_1^{1-2\gamma}L_0)^2T^4+\eta L_3T^2\right)\\ &\quad\cdot\mathbb{E}\exp\left[\eta\left(\frac{32}{L_2}+\frac{67}{8}T^2\right)|\mathbf{v}_0|_\gamma^2\right].\end{aligned} \tag{4.18}$$

Recalling $\mathbf{v}_0\sim\mathcal{N}(0,\mathcal{C})$ and the assumption $\eta < \left[2\,\mathrm{Tr}(\mathcal{C}^{1-2\gamma})\left(\frac{32}{L_2}+\frac{67}{8}T^2\right)\right]^{-1}$, we have, again by Fernique's theorem [27, Proposition 2.17], and Lemma 4 that

$$\mathbb{E}\exp\left[\eta\left(\frac{32}{L_2}+\frac{67}{8}T^2\right)|\mathbf{v}_0|_\gamma^2\right] \le \left[1-2\eta\left(\frac{32}{L_2}+\frac{67}{8}T^2\right)\mathrm{Tr}(\mathcal{C}^{1-2\gamma})\right]^{-1/2}. \tag{4.19}$$

Thus, denoting $\tilde{\kappa}_2 = 1-L_2T^2/32$ and

$$R = \exp\left(\frac{5}{3}\eta(\lambda_1^{1-2\gamma}L_0)^2T^4+\eta L_3T^2\right)\left[1-2\eta\left(\frac{32}{L_2}+\frac{67}{8}T^2\right)\mathrm{Tr}(\mathcal{C}^{1-2\gamma})\right]^{-1/2},$$

we obtain from (4.18) and (4.19) that

$$PV_{2,\eta}(\mathbf{q}_0) \le R\exp\left(\eta\tilde{\kappa}_2\,|\mathbf{q}_0|_\gamma^2\right) = R\exp\left(\eta\,|\mathbf{q}_0|_\gamma^2\right)^{\tilde{\kappa}_2}$$

$$
\begin{aligned}
&\le \tilde{\kappa}_2 V_2(\mathbf{q}_0) + R^{\frac{1}{1-\tilde{\kappa}_2}}(1-\tilde{\kappa}_2) = \tilde{\kappa}_2 V_2(\mathbf{q}_0) + R^{\frac{32}{L_2T^2}}\frac{L_2T^2}{32} \\
&\le e^{-\frac{L_2T^2}{32}} V_2(\mathbf{q}_0) + R^{\frac{32}{L_2T^2}}\frac{L_2T^2}{32} \qquad (4.20)
\end{aligned}
$$

where the second estimate follows by Young's inequality. We conclude 19 for $V_{2,\eta}$ after using (4.20) n times iteratively. The proof is now complete. □

5 Pointwise contractivity bounds for the Markovian dynamics

This section details two pointwise contractivity bounds for the Markovian dynamics of the PHMC chain (2.16) in a suitably tuned Wasserstein-Kantorovich metric. These bounds provide crucial ingredients needed for the weak Harris theorem, namely the so called 'ρ-contractivity' and 'ρ-smallness' conditions, which, together with the Lyapunov structure identified in Proposition 20, form the core of the proof of Theorem 26.

Our contraction results are given with respect to an underlying metric $\rho : \mathbb{H}_\gamma \times \mathbb{H}_\gamma \to [0, 1]$ defined as

$$
\rho(\mathbf{q}, \tilde{\mathbf{q}}) := \frac{|\mathbf{q} - \tilde{\mathbf{q}}|_\gamma}{\varepsilon} \wedge 1, \qquad (5.1)
$$

where γ is given in Assumption 8. On the other hand, $\varepsilon > 0$ is a tuning parameter which specifies the small scales in our problem and is determined by (5.3) in such a fashion as to produce a contraction in (5.2). Recall that the Wasserstein distance on the space of probability measures on $\mathbb{H}_\gamma$ induced by ρ is given as in (1.5) with $\tilde{\rho}$ replaced by ρ, and denoted by $\mathcal{W}_\rho$.

The first result yielding 'ρ-contractivity' (cf. [43, Definition 4.6]) is given as follows:

Proposition 22 *Suppose Assumptions 2, 5 and 8 are satisfied and choose an integration time $T > 0$ and $N \in \mathbb{N}$ maintaining the condition (3.20). Fix any $\varepsilon > 0$ defining the associated metric ρ as in (5.1). Then, for every $n \in \mathbb{N}$ and for every $\mathbf{q}_0, \tilde{\mathbf{q}}_0 \in \mathbb{H}_\gamma$ such that $\rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0) < 1$, we have*

$$
\mathcal{W}_\rho(P^n(\mathbf{q}_0, \cdot), P^n(\tilde{\mathbf{q}}_0, \cdot)) \le \kappa_3 \rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0) \qquad (5.2)
$$

where recall that P^n is n steps of the PHMC kernel (2.20) and $\mathcal{W}_\rho$ is the Wasserstein distance, as in (1.8), associated with ρ. Here

$$
\kappa_3 = \kappa_3(n) := \kappa_2(n) + \frac{2\sqrt{2}\lambda_N^{-\frac{1}{2}+\gamma}(1+\lambda_1^{1-2\gamma}L_1)\varepsilon}{T(1-\kappa_1^2)^{1/2}} = \kappa_2(n) + \frac{\sqrt{2}\lambda_N^{-\frac{1}{2}+\gamma}\alpha\varepsilon}{2T(1-\kappa_1^2)^{1/2}}, \qquad (5.3)
$$

where

$$\kappa_2(n) := 4\sqrt{2}(1 + \lambda_1^{1-2\gamma} L_1)\kappa_1^n = \sqrt{2}\alpha\kappa_1^n, \quad \kappa_1 := 1 - \frac{T^2}{12}, \tag{5.4}$$

$T > 0$ is the integration time in (2.16), L_1 is the Lipschitz constant of DU as in (2.7) and λ_1 is the largest eigenvalue of $\mathcal{C}$ and, in regards to α, recall (3.21).

Remark 23 If $N \in \mathbb{N}$ is the smallest natural number for which the corresponding condition in (3.20) holds, i.e.

$$N = \min\left\{ n \in \mathbb{N} \, : \, \lambda_{n+1}^{1-2\gamma} \leq \frac{1}{4L_1} \right\},$$

then κ_3 from (5.3) above can be given in the more explicit form

$$\kappa_3 = \kappa_3(n) := \kappa_2(n) + \frac{4\sqrt{2}L_1^{1/2}(1 + \lambda_1^{1-2\gamma} L_1)\varepsilon}{T(1-\kappa_1^2)^{1/2}} = \kappa_2(n) + \frac{\sqrt{2}L_1^{1/2}\alpha\varepsilon}{T(1-\kappa_1^2)^{1/2}},$$

with κ_2 defined exactly as in (5.4) above.

Our second main result corresponding to 'ρ-smallness' (cf. [43, Definition 4.4]) is given as:

Proposition 24 *Assume the same hypotheses from Proposition 22. Let $M \geq 0$ and take*

$$A = \left\{ \mathbf{q} \in \mathbb{H}_\gamma \, : \, |\mathbf{q}|_\gamma \leq M \right\}.$$

Then, for every $n \in \mathbb{N}$ and every $\varepsilon > 0$ we have for the corresponding ρ defined by (5.1) that

$$\mathcal{W}_\rho(P^n(\mathbf{q}_0, \cdot), P^n(\tilde{\mathbf{q}}_0, \cdot)) \leq 1 - \kappa_4 \tag{5.5}$$

for every $\mathbf{q}_0, \tilde{\mathbf{q}}_0 \in A$, where

$$\begin{aligned}\kappa_4 = \kappa_4(n) &:= \frac{1}{2}\exp\left(-\frac{256 L_1(1 + \lambda_1^{1-2\gamma} L_1)^2 M^2}{T^2(1-\kappa_1^2)} \right) - \frac{2M\kappa_2(n)}{\varepsilon} \\ &= \frac{1}{2}\exp\left(-\frac{16 L_1 \alpha^2 M^2}{T^2(1-\kappa_1^2)} \right) - \frac{2M\kappa_2(n)}{\varepsilon},\end{aligned}$$

with κ_1 and κ_2 as defined in (5.4), and α as defined in (3.21).

Before proceeding with the proofs of Propositions 22 and 24, we introduce some further preliminary terminology and general background. Set an integration time $T > 0$ in the definition of the transition kernel P of the PHMC chain, (2.16). For each $n \in \mathbb{N}$, let $\mathbb{H}^{\otimes n}$ denote the space given as the product of n copies of $\mathbb{H}$. Moreover, given a sequence $\{\mathbf{v}_0^{(j)}\}_{j\in\mathbb{N}}$ of i.i.d. draws from $\mathcal{N}(0,\mathcal{C})$, we denote by $\mathbf{V}_0^{(n)} = (\mathbf{v}_0^{(1)},\ldots,\mathbf{v}_0^{(n)})$ the noise path for the first $n \geq 1$ steps, as in (2.18). We then have $\mathbf{V}_0^{(n)} \sim \mathcal{N}(0,\mathcal{C})^{\otimes n}$, with $\mathcal{N}(0,\mathcal{C})^{\otimes n}$ denoting the product of n independent copies of $\mathcal{N}(0,\mathcal{C})$.

For simplicity of notation, we set from now on

$$\sigma := \mathcal{N}(0,\mathcal{C}), \quad \sigma_n := \mathcal{N}(0,\mathcal{C})^{\otimes n}.$$

For every $\mathbf{q}_0, \tilde{\mathbf{q}}_0 \in \mathbb{H}_\gamma$, with γ as in (2.6), (2.7), and $N \in \mathbb{N}$ as in Proposition 18, we consider $\widetilde{Q}_1(\tilde{\mathbf{q}}_0,\mathbf{q}_0) : \mathbb{H} \to \mathbb{H}$ to be the random variable defined as

$$\widetilde{Q}_1(\mathbf{q}_0,\tilde{\mathbf{q}}_0)(\mathbf{v}_0^{(1)}) = \mathbf{q}_T(\tilde{\mathbf{q}}_0, \mathbf{v}_0^{(1)} + T^{-1}\Pi_N(\mathbf{q}_0 - \tilde{\mathbf{q}}_0))$$

where $\mathbf{v}_0^{(1)} \sim \sigma$. Iteratively we define, for $n \geq 2$, the random variables $\widetilde{Q}_n(\mathbf{q}_0,\tilde{\mathbf{q}}_0) : \mathbb{H}^{\otimes n} \to \mathbb{H}$ as

$$\widetilde{Q}_n(\mathbf{q}_0,\tilde{\mathbf{q}}_0)(\mathbf{V}_0^{(n)}) := q_T(\widetilde{Q}_{n-1}(\mathbf{q}_0,\tilde{\mathbf{q}}_0)(\mathbf{V}_0^{(n-1)}), \mathbf{v}_0^{(n)} + \mathcal{S}_n(\mathbf{V}_0^{(n-1)})), \tag{5.6}$$

where $\mathbf{V}_0^{(n)} \sim \sigma_n$, and

$$\mathcal{S}_n(\mathbf{V}_0^{(n-1)}) := T^{-1}\Pi_N[Q_{n-1}(\mathbf{q}_0)(\mathbf{V}_0^{(n-1)}) - \widetilde{Q}_{n-1}(\mathbf{q}_0,\tilde{\mathbf{q}}_0)(\mathbf{V}_0^{(n-1)})]. \tag{5.7}$$

We therefore obtain the shifted noise path

$$\tilde{\mathbf{V}}_0^{(n)} = (\mathbf{v}_0^1 + \mathcal{S}_1, \mathbf{v}_0^{(2)} + \mathcal{S}_2(\mathbf{V}_0^{(1)}),\ldots,\mathbf{v}_0^{(n)} + \mathcal{S}_n(\mathbf{V}_0^{(n-1)})), \tag{5.8}$$

where $\mathcal{S}_1 = T^{-1}\Pi_N(\mathbf{q}_0 - \tilde{\mathbf{q}}_0)$.

Let $\tilde{\sigma}_n := \mathrm{Law}(\tilde{\mathbf{V}}_0^{(n)})$. In order to simplify notation, let us denote

$$\boldsymbol{\mathcal{S}}_n(\mathbf{V}_0^{(n)}) = (\mathcal{S}_1, \mathcal{S}_2(\mathbf{V}_0^{(1)}),\ldots,\mathcal{S}_n(\mathbf{V}_0^{(n-1)})) \tag{5.9}$$

and

$$\boldsymbol{\mathcal{R}}_n(\mathbf{V}_0^{(n)}) = \mathbf{V}_0^{(n)} + \boldsymbol{\mathcal{S}}_n(\mathbf{V}_0^{(n)}), \tag{5.10}$$

so that $\tilde{\mathbf{V}}_0^{(n)} = \boldsymbol{\mathcal{R}}_n(\mathbf{V}_0^{(n)})$. Thus, $\tilde{\sigma}_n$ is the push-forward of σ_n by the mapping $\boldsymbol{\mathcal{R}}_n : \mathbb{H}^{\otimes n} \to \mathbb{H}^{\otimes n}$, i.e. $\tilde{\sigma}_n = \boldsymbol{\mathcal{R}}_n^*\sigma_n$. Now put, for every $n \in \mathbb{N}$ and $A \in \mathcal{B}(\mathbb{H})$,

$$\widetilde{P}^n(\mathbf{q}_0,\tilde{\mathbf{q}}_0,A) = \widetilde{Q}_n(\mathbf{q}_0,\tilde{\mathbf{q}}_0)^*\sigma_n(A) = \sigma_n(\widetilde{Q}_n(\mathbf{q}_0,\tilde{\mathbf{q}}_0)^{-1}(A)). \tag{5.11}$$

Notice that $\widetilde{P}^n(\mathbf{q}_0,\tilde{\mathbf{q}}_0,\cdot)$ can be equivalently written as

$$\widetilde{P}^n(\mathbf{q}_0,\tilde{\mathbf{q}}_0,A) = Q_n(\tilde{\mathbf{q}}_0)^*(\boldsymbol{\mathcal{R}}_n^*\sigma_n)(A) = Q_n(\tilde{\mathbf{q}}_0)^*\tilde{\sigma}_n(A). \tag{5.12}$$

With these notations in place we have the following estimate which we will use several times below in establishing Propositions 22 and 24. The proof follows immediately from Proposition 18 and Remark 17.

Lemma 25 *We are maintaining the same hypotheses as in Proposition 22. Then, starting from any* $\mathbf{q}_0, \tilde{\mathbf{q}}_0 \in \mathbb{H}_\gamma$ *we have that for all* $n \geq 1$,

$$\left| Q_n(\mathbf{q}_0)(\mathbf{V}_0^{(n)}) - \widetilde{Q}_n(\mathbf{q}_0, \tilde{\mathbf{q}}_0)(\mathbf{V}_0^{(n)}) \right|_\gamma \leq \kappa_2 \left| \mathbf{q}_0 - \tilde{\mathbf{q}}_0 \right|_\gamma \quad \textit{for every } \mathbf{V}_0^{(n)} \in \mathbb{H}^{\otimes n},$$

where Q_n *and* $\widetilde{Q}_n$ *are defined as in* (2.19) *and* (5.6)*, respectively, and* κ_2 *is as in* (5.4)*. Therefore,*

$$\mathbb{E} \left| Q_n(\mathbf{q}_0) - \tilde{Q}_n(\mathbf{q}_0, \tilde{\mathbf{q}}_0) \right|_\gamma \leq \kappa_2 \left| \mathbf{q}_0 - \tilde{\mathbf{q}}_0 \right|_\gamma . \tag{5.13}$$

We also recall additional notions of distances in the space of Borel probability measures on a given complete metric space (X, d), denoted $\Pr(X)$, with the associated Borel σ-algebra denoted as $\mathcal{B}(X)$. Namely, the *total variation* distance is defined as

$$\|\nu - \tilde{\nu}\|_{\mathrm{TV}} := \sup_{A \in \mathcal{B}(X)} |\nu(A) - \tilde{\nu}(A)| \tag{5.14}$$

for any $\nu, \tilde{\nu} \in \Pr(X)$. On the other hand when $\tilde{\nu} \ll \nu$, i.e. when $\tilde{\nu}$ is absolutely continuous with respect to ν, the *Kullback-Leibler Divergence* is defined as

$$D_{\mathrm{KL}}(\tilde{\nu}|\nu) := \int_X \log\left(\frac{d\tilde{\nu}}{d\nu}(\mathbf{V})\right) d\tilde{\nu}(d\mathbf{V}). \tag{5.15}$$

Recall that for the trivial metric

$$\rho_0(\mathbf{q}, \tilde{\mathbf{q}}) := \begin{cases} 1 & \text{if } \mathbf{q} \neq \tilde{\mathbf{q}} \\ 0 & \text{if } \mathbf{q} = \tilde{\mathbf{q}}, \end{cases}$$

the associated Wasserstein distance $\mathcal{W}_{\rho_0}$ coincides with the total variation distance. On the other hand, Pinsker's inequality (see e.g. [82]) states that

$$\|\nu - \tilde{\nu}\|_{\mathrm{TV}} \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(\tilde{\nu}|\nu)}, \tag{5.16}$$

for any $\nu, \tilde{\nu} \in \Pr(X)$, $\tilde{\nu} \ll \nu$. Moreover, as showed e.g. in [20, Appendix],

$$\|\nu - \tilde{\nu}\|_{\mathrm{TV}} \leq 1 - \frac{1}{2} \exp\left(-D_{\mathrm{KL}}(\tilde{\nu}|\nu)\right) \tag{5.17}$$

for all $\nu, \tilde{\nu} \in \Pr(X)$, $\tilde{\nu} \ll \nu$.

Proof *(Proof of Proposition 22)* Fix any $\mathbf{q}_0, \tilde{\mathbf{q}}_0 \in \mathbb{H}_\gamma$ such that $\rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0) < 1$. Then, recalling the notation (5.11) and using that ρ is a metric on $\mathbb{H}$ we have

$$\mathcal{W}_\rho(P^n(\mathbf{q}_0, \cdot), P^n(\tilde{\mathbf{q}}_0, \cdot)) \leq \mathcal{W}_\rho(P^n(\mathbf{q}_0, \cdot), \widetilde{P}^n(\mathbf{q}_0, \tilde{\mathbf{q}}_0, \cdot)) + \mathcal{W}_\rho(\widetilde{P}^n(\mathbf{q}_0, \tilde{\mathbf{q}}_0, \cdot), P^n(\tilde{\mathbf{q}}_0, \cdot)). \tag{5.18}$$

Notice that

$$\begin{aligned}\mathcal{W}_\rho(P^n(\mathbf{q}_0, \cdot), \widetilde{P}^n(\mathbf{q}_0, \tilde{\mathbf{q}}_0, \cdot)) \leq& \mathbb{E}\rho(Q_n(\mathbf{q}_0), \widetilde{Q}_n(\mathbf{q}_0, \tilde{\mathbf{q}}_0)) \leq \frac{1}{\varepsilon}\mathbb{E}\left|Q_n(\mathbf{q}_0) - \widetilde{Q}_n(\mathbf{q}_0, \tilde{\mathbf{q}}_0)\right|_\gamma \\ \leq& \frac{\kappa_2}{\varepsilon}\left|\mathbf{q}_0 - \tilde{\mathbf{q}}_0\right|_\gamma = \kappa_2\rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0),\end{aligned} \tag{5.19}$$

where the last inequality follows from Lemma 25.

For the second term in (5.18), it follows from the coupling lemma (see e.g. [55, Lemma 1.2.24]) and the fact that $\rho \leq 1$ that

$$\mathcal{W}_\rho(\widetilde{P}^n(\mathbf{q}_0, \tilde{\mathbf{q}}_0, \cdot), P^n(\tilde{\mathbf{q}}_0, \cdot)) \leq \left\| \widetilde{P}^n(\mathbf{q}_0, \tilde{\mathbf{q}}_0, \cdot) - P^n(\tilde{\mathbf{q}}_0, \cdot)\right\|_{\mathrm{TV}}. \tag{5.20}$$

From (2.21) and (5.12), we have

$$\left\| \widetilde{P}^n(\mathbf{q}_0, \tilde{\mathbf{q}}_0, \cdot) - P^n(\tilde{\mathbf{q}}_0, \cdot)\right\|_{\mathrm{TV}} = \left\| Q_n(\tilde{\mathbf{q}}_0)^*\tilde{\sigma}_n - Q_n(\tilde{\mathbf{q}}_0)^*\sigma_n\right\|_{\mathrm{TV}}.$$

Moreover, from the definition of the total variation distance in (5.14) and inequality (5.16), we infer

$$\left\| Q_n(\tilde{\mathbf{q}}_0)^*\tilde{\sigma}_n - Q_n(\tilde{\mathbf{q}}_0)^*\sigma_n\right\|_{\mathrm{TV}} \leq \|\tilde{\sigma}_n - \sigma_n\|_{\mathrm{TV}} \leq \sqrt{\frac{1}{2}D_{\mathrm{KL}}(\tilde{\sigma}_n|\sigma_n)}. \tag{5.21}$$

As a consequence of Girsanov's Theorem, we obtain

$$\frac{d\sigma_n}{d\tilde{\sigma}_n}(\boldsymbol{\mathcal{R}}_n(\mathbf{V})) = \exp\left(\frac{1}{2}|\mathcal{C}^{-1/2}\mathbf{V}|^2 - \frac{1}{2}|\mathcal{C}^{-1/2}\boldsymbol{\mathcal{R}}_n(\mathbf{V})|^2\right) \quad \text{for any } \mathbf{V} \in \mathbb{H}_{1/2}^{\otimes n}, \tag{5.22}$$

with $\boldsymbol{\mathcal{R}}_n$ as defined in (5.10). Thus,

$$\begin{aligned}D_{\mathrm{KL}}(\tilde{\sigma}_n|\sigma_n) &= \int \log\left(\frac{d\tilde{\sigma}_n}{d\sigma_n}(\mathbf{V})\right)\tilde{\sigma}_n(d\mathbf{V}) = -\int \log\left(\frac{d\sigma_n}{d\tilde{\sigma}_n}(\mathbf{V})\right)\tilde{\sigma}_n(d\mathbf{V}) \\ &= -\int \log\left(\frac{d\sigma_n}{d\tilde{\sigma}_n}(\boldsymbol{\mathcal{R}}_n(\mathbf{V}))\right)\sigma_n(d\mathbf{V}) = \int\left(-\frac{1}{2}|\mathbf{V}|_{1/2}^2 + \frac{1}{2}|\boldsymbol{\mathcal{R}}_n(\mathbf{V})|_{1/2}^2\right)\sigma_n(d\mathbf{V}) \\ &= \int\left(\langle\boldsymbol{\mathcal{S}}_n(\mathbf{V}), \mathbf{V}\rangle_{1/2} + \frac{1}{2}|\boldsymbol{\mathcal{S}}_n(\mathbf{V})|_{1/2}^2\right)\sigma_n(d\mathbf{V}) = \frac{1}{2}\int |\boldsymbol{\mathcal{S}}_n(\mathbf{V})|_{1/2}^2\sigma_n(d\mathbf{V}) \\ &= \frac{1}{2}\sum_{j=1}^n \mathbb{E}|\mathcal{S}_j(\cdot)|_{1/2}^2.\end{aligned} \tag{5.23}$$

Here note that, taking $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ and $\mathbf{V}^j = (\mathbf{v}_1, \dots, \mathbf{v}_j)$ for $j \leq n$ we have

$$\begin{aligned}\int \langle \boldsymbol{\mathcal{S}}_n(\mathbf{V}), \mathbf{V}\rangle_{1/2}\sigma_n(d\mathbf{V}) &= \sum_{j=1}^{n} \int \langle \mathcal{S}_j(\mathbf{V}^{j-1}), \mathbf{v}_j\rangle_{1/2}\sigma_n(d\mathbf{V})\\ &= \sum_{j=1}^{n} \int \int \langle \mathcal{S}_j(\mathbf{V}^{j-1}), \mathbf{v}_j\rangle_{1/2}\sigma(d\mathbf{v}_j)\sigma_{j-1}(d\mathbf{V}^{j-1}) = 0,\end{aligned}$$

which justifies dropping this term in (5.23). Now, from the definition of $\mathcal{S}_j$ in (5.7), (2.5) in Lemma 4 and (3.18) it follows that

$$\begin{aligned}|\mathcal{S}_j(\mathbf{V}_0^{j-1})|_{1/2}^2 \leq \lambda_N^{-1+2\gamma} \left|\mathcal{S}_j(\mathbf{V}_0^{j-1})\right|_{\gamma}^2 &\leq \lambda_N^{-1+2\gamma} \left|\mathcal{S}_j(\mathbf{V}_0^{j-1})\right|_{\gamma,\alpha}^2\\ &\leq T^{-2}\lambda_N^{-1+2\gamma}\kappa_1^{2(j-1)} |\mathbf{q}_0 - \tilde{\mathbf{q}}_0|_{\gamma,\alpha}^2\\ &\leq T^{-2}\lambda_N^{-1+2\gamma}\kappa_1^{2(j-1)}2\alpha^2 |\mathbf{q}_0 - \tilde{\mathbf{q}}_0|_{\gamma}^2,\end{aligned}$$

for each $j \geq 1$, with α as defined in (3.21). Therefore,

$$D_{\mathrm{KL}}(\tilde{\sigma}_n|\sigma_n) \leq \frac{\lambda_N^{-1+2\gamma}\alpha^2}{T^2} |\mathbf{q}_0 - \tilde{\mathbf{q}}_0|_{\gamma}^2 \sum_{j=1}^{n} \kappa_1^{2(j-1)} \leq \frac{\lambda_N^{-1+2\gamma}\alpha^2}{T^2(1-\kappa_1^2)} |\mathbf{q}_0 - \tilde{\mathbf{q}}_0|_{\gamma}^2, \tag{5.24}$$

so that, combining this observation with (5.20)–(5.21), and our standing assumption that $\rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0) < 1$,

$$\mathcal{W}_\rho(\widetilde{P}^n(\mathbf{q}_0, \tilde{\mathbf{q}}_0, \cdot), P^n(\tilde{\mathbf{q}}_0, \cdot)) \leq \frac{\lambda_N^{-\frac{1}{2}+\gamma}\alpha}{\sqrt{2}T(1-\kappa_1^2)^{1/2}} |\mathbf{q}_0 - \tilde{\mathbf{q}}_0|_{\gamma} = \frac{\lambda_N^{-\frac{1}{2}+\gamma}\alpha\varepsilon}{\sqrt{2}T(1-\kappa_1^2)^{1/2}} \rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0). \tag{5.25}$$

We therefore conclude (5.2) from (5.18), (5.19) and (5.25), completing the proof of Proposition 22. □

Proof *(Proof of Proposition 24)* We proceed similarly as in the proof of Proposition 22 starting with the splitting (5.18). Fix any $\mathbf{q}_0, \tilde{\mathbf{q}}_0 \in A$. The first term after inequality (5.18) is estimated exactly as in (5.19), so that

$$\mathcal{W}_\rho(P^n(\mathbf{q}_0, \cdot), \widetilde{P}^n(\mathbf{q}_0, \tilde{\mathbf{q}}_0, \cdot)) \leq \frac{\kappa_2}{\varepsilon} |\mathbf{q}_0 - \tilde{\mathbf{q}}_0|_{\gamma} \leq \frac{2M\kappa_2}{\varepsilon}.$$

The second term in (5.18) is estimated by using (5.17) and (5.24) as

$$\mathcal{W}_\rho(\widetilde{P}^n(\mathbf{q}_0, \tilde{\mathbf{q}}_0, \cdot), P^n(\tilde{\mathbf{q}}_0, \cdot)) \leq \|\tilde{\sigma}_n - \sigma_n\|_{\mathrm{TV}} \leq 1 - \frac{1}{2}\exp\left(-D_{\mathrm{KL}}(\tilde{\sigma}_n|\sigma_n)\right)$$

$$\leq 1 - \frac{1}{2} \exp\left(- \frac{\lambda_N^{-1+2\gamma} \alpha^2}{T^2(1-\kappa_1^2)} |\mathbf{q}_0 - \tilde{\mathbf{q}}_0|_\gamma^2 \right),$$

with α as defined in (3.21). Hence, together with (5.18) and using that $\mathbf{q}_0, \tilde{\mathbf{q}}_0 \in A$, we conclude (5.5). □

6 Main result

Having obtained in the previous sections a Foster–Lyapunov structure (4.1) together with the smallness and contractivity properties (5.2)–(5.5) for the Markov kernel P in (1.3), we are now ready to proceed with the proof of our main result. As pointed out in the introduction, the spectral gap (6.2) below follows as a consequence of the weak Harris theorem given the aforementioned properties.

We provide a self-contained presentation of the weak Harris approach in this section both for completeness and in order to make some of the constants in the proof more explicit. We start by noticing that it is enough to show (6.2) for ν_1, ν_2 being Dirac measures, say concentrated at points $\mathbf{q}_0, \tilde{\mathbf{q}}_0 \in \mathbb{H}_\gamma$. The proof is then split into three possible cases for such points: $\rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0) < 1$ ('close to each other'); $\rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0) = 1$ with $V(\mathbf{q}_0) + V(\tilde{\mathbf{q}}_0) > 4K_V$ ('far from the origin'); and $\rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0) = 1$ with $V(\mathbf{q}_0) + V(\tilde{\mathbf{q}}_0) \leq 4K_V$ ('close to the origin'). The first case follows from the contraction result in Proposition 22 together with the Lyapunov structure from Proposition 20. The second case follows entirely from the Lyapunov property. Lastly, the third case follows by invoking the smallness result in Proposition 24 as well as the Lyapunov structure. Finally, the second part of our main result, namely (6.4)–(6.6), follows essentially from the spectral gap (6.2) by invoking Propositions 40, 43 and 46, which are all proved in detail in Appendix 1.

Theorem 26 *Fix $\gamma \in [0, 1/2)$. Suppose Assumptions 2, 5, 8 and 10 are satisfied and choose an integration time $T > 0$ such that*

$$T \leq \min \left\{ \frac{1}{[2(1+\lambda_1^{1-2\gamma} L_1)]^{1/2}}, \frac{L_2^{1/2}}{2\sqrt{6}(1+\lambda_1^{1-2\gamma} L_1)} \right\}. \tag{6.1}$$

Here the constants L_1, L_2 are as in (2.7) and (2.8) and λ_1 is the largest eigenvalue of the covariance operator $\mathcal{C}$ defined as in Assumption 2. Let $V : \mathbb{H}_\gamma \to \mathbb{R}^+$ be a Lyapunov function for the Markov kernel P defined in (2.16) of the form (4.3) or (4.4). Then, there exists $\varepsilon > 0$, $C_1 > 0$ and $C_2 > 0$ such that, for every $\nu_1, \nu_2 \in \Pr(\mathbb{H})$ with support included in $\mathbb{H}_\gamma$,

$$\mathcal{W}_{\tilde{\rho}}(\nu_1 P^n, \nu_2 P^n) \leq C_1 e^{-C_2 n} \mathcal{W}_{\tilde{\rho}}(\nu_1, \nu_2) \quad \textit{for all } n \in \mathbb{N}, \tag{6.2}$$

where $\tilde{\rho} : \mathbb{H}_\gamma \times \mathbb{H}_\gamma \to \mathbb{R}^+$ is the distance-like function given by

$$\tilde{\rho}(\mathbf{q}, \tilde{\mathbf{q}}) = \sqrt{\rho(\mathbf{q}, \tilde{\mathbf{q}})(1 + V(\mathbf{q}) + V(\tilde{\mathbf{q}}))} \quad \textit{for all } \mathbf{q}, \tilde{\mathbf{q}} \in \mathbb{H}_\gamma,$$

with ρ as defined in (5.1).

Moreover, with respect to μ defined in (1.1), i.e. the invariant measure for P (cf. Proposition 13), the following results hold: for any observable $\Phi : \mathbb{H}_\gamma \to \mathbb{R}$ such that

$$L_\Phi := \sup_{q\in\mathbb{H}_\gamma} \frac{\max\{2|\Phi(\mathbf{q})|, \sqrt{\varepsilon}|D\Phi(\mathbf{q})|_{\mathcal{L}(\mathbb{H}_\gamma)}\}}{\sqrt{1+V(\mathbf{q})}} < \infty, \tag{6.3}$$

with $|\cdot|_{\mathcal{L}(\mathbb{H}_\gamma)}$ denoting the standard operator norm of a linear functional on $\mathbb{H}_\gamma$, we have

$$\left| P^n\Phi(\mathbf{q}) - \int \Phi(\mathbf{q}')\mu(dq') \right| \le L_\Phi C_1 e^{-nC_2} \int \sqrt{1+V(\mathbf{q})+V(\mathbf{q}')}\mu(d\mathbf{q}'), \tag{6.4}$$

for every $n \in \mathbb{N}$ and $\mathbf{q} \in \mathbb{H}_\gamma$. On the other hand, taking $\{Q_k(\mathbf{q}_0)\}_{k\ge 0}$ to be any process associated to $\{P^k(\mathbf{q}_0,\cdot)\}_{k\ge0}$ as in (2.21), we have, for any measurable observable maintaining (6.3), that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \Phi(Q_k(\mathbf{q})) = \int \Phi(\mathbf{q}')\mu(d\mathbf{q}'), \quad \textit{almost surely}, \tag{6.5}$$

for all $\mathbf{q} \in \mathbb{H}_\gamma$. Furthermore,

$$\sqrt{n}\left[\frac{1}{n}\sum_{k=1}^{n}\Phi(Q_k(\mathbf{q})) - \int \Phi(\mathbf{q}')\mu(d\mathbf{q}'))\right] \Rightarrow \mathcal{N}(0,\sigma^2(\Phi)) \quad \textit{as } n\to\infty, \tag{6.6}$$

for all $\mathbf{q} \in \mathbb{H}_\gamma$, i.e. the expression in the left-hand side of (6.6) converges weakly to a real-valued gaussian random variable with mean zero and covariance $\sigma^2(\Phi)$, where $\sigma^2(\Phi)$ is specified explicitly as (A.36) below, with μ^* replaced by μ.

Proof We claim it suffices to show that there exists $\varepsilon > 0$, $C_1 > 0$ and $C_2 > 0$ such that

$$\mathcal{W}_{\tilde{\rho}}(P^n(\mathbf{q}_0,\cdot), P^n(\tilde{\mathbf{q}}_0,\cdot)) \le C_1 e^{-C_2 n}\tilde{\rho}(\mathbf{q}_0,\tilde{\mathbf{q}}_0) \quad \text{for all } \mathbf{q}_0, \tilde{\mathbf{q}}_0 \in \mathbb{H}_\gamma \text{ and } n \in \mathbb{N}. \tag{6.7}$$

Indeed, since $\tilde{\rho}$ is lower-semicontinuous and non-negative, it follows from [83, Theorem 4.8] that

$$\mathcal{W}_{\tilde{\rho}}(\nu_1 P^n, \nu_2 P^n) \le \int \mathcal{W}_{\tilde{\rho}}(P^n(\mathbf{q}_0,\cdot), P^n(\tilde{\mathbf{q}}_0,\cdot))\Gamma(d\mathbf{q}_0, d\tilde{\mathbf{q}}_0)$$

for all $\Gamma \in \mathfrak{C}(\nu_1,\nu_2)$ and $n \in \mathbb{N}$.

Clearly, if ν_1 and ν_2 have supports included in $\mathbb{H}_\gamma$, then $\Gamma \in \mathfrak{C}(\nu_1, \nu_2)$ has support included in $\mathbb{H}_\gamma \times \mathbb{H}_\gamma$. Hence, if (6.7) holds then

$$\mathcal{W}_{\tilde{\rho}}(\nu_1 P^n, \nu_2 P^n) \leq C_1 e^{-C_2 n} \int \tilde{\rho}(\mathbf{q}_0, \tilde{\mathbf{q}}_0) \Gamma(d\mathbf{q}_0, d\tilde{\mathbf{q}}_0) \tag{6.8}$$

for all $\Gamma \in \mathfrak{C}(\nu_1, \nu_2)$ and $n \in \mathbb{N}$, which implies (6.2).

In order to show (6.7), we consider an auxiliary metric defined as

$$\tilde{\rho}_\beta(q, \tilde{q}) = \sqrt{\rho(\mathbf{q}, \tilde{\mathbf{q}})(1 + \beta V(\mathbf{q}) + \beta V(\tilde{\mathbf{q}}))}, \quad \text{for all } \mathbf{q}, \tilde{\mathbf{q}} \in \mathbb{H}_\gamma,$$

with the additional parameter $\beta > 0$ to be appropriately chosen below; cf. (6.18). Notice that $\tilde{\rho}$ and $\tilde{\rho}_\beta$ are equivalent. Indeed,

$$(\min\{1, \beta\})^{1/2}\, \tilde{\rho}(\mathbf{q}, \tilde{\mathbf{q}}) \leq \tilde{\rho}_\beta(\mathbf{q}, \tilde{\mathbf{q}}) \leq (\max\{1, \beta\})^{1/2}\, \tilde{\rho}(\mathbf{q}, \tilde{\mathbf{q}}), \quad \text{for all } \mathbf{q}, \tilde{\mathbf{q}} \in \mathbb{H}_\gamma. \tag{6.9}$$

We now show that

$$\mathcal{W}_{\tilde{\rho}_\beta}(P^n(\mathbf{q}_0, \cdot), P^n(\tilde{\mathbf{q}}_0, \cdot)) \leq \kappa_5(n) \tilde{\rho}_\beta(\mathbf{q}_0, \tilde{\mathbf{q}}_0) \quad \text{for all } n \geq 1 \text{ and } \mathbf{q}_0, \tilde{\mathbf{q}}_0 \in \mathbb{H}_\gamma, \tag{6.10}$$

such that, for suitably chosen $\varepsilon > 0$, $\beta > 0$, and for $n_0 \in \mathbb{N}$ sufficiently large we have $\kappa_5(n) < 1$ for every $n \geq n_0$. We then subsequently use this bound to establish (6.7) as in (6.27) below.

The analysis leading to (6.10) is split into three cases:

<u>*Case1 :*</u> *Suppose that* $\rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0) < 1$, *so that* $\rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0) = |\mathbf{q}_0 - \tilde{\mathbf{q}}_0|_\gamma\, \varepsilon^{-1}$.

By Hölder's inequality, we obtain

$$\begin{aligned}
&\mathcal{W}_{\tilde{\rho}_\beta}(P^n(\mathbf{q}_0, \cdot), P^n(\tilde{\mathbf{q}}_0, \cdot))^2 \\
&\leq \inf_{\Gamma \in \mathfrak{C}(\delta_{\mathbf{q}_0} P^n, \delta_{\tilde{\mathbf{q}}_0} P^n)} \left\{ \left(\int \rho(\mathbf{q}, \tilde{\mathbf{q}}) \Gamma(d\mathbf{q}, d\tilde{\mathbf{q}}) \right) \left(\int (1 + \beta V(\mathbf{q}) + \beta V(\tilde{\mathbf{q}})) \Gamma(d\mathbf{q}, d\tilde{\mathbf{q}}) \right) \right\} \\
&= \big(1 + \beta P^n V(\mathbf{q}_0) + \beta P^n V(\tilde{\mathbf{q}}_0)\big)\, \mathcal{W}_\rho(P^n(\mathbf{q}_0, \cdot), P^n(\tilde{\mathbf{q}}_0, \cdot)).
\end{aligned} \tag{6.11}$$

From Propositions 20 and 22, it follows that

$$\begin{aligned}
\mathcal{W}_{\tilde{\rho}_\beta}(P^n(\mathbf{q}_0, \cdot), P^n(\tilde{\mathbf{q}}_0, \cdot))^2 &\leq \big(1 + \beta \kappa_V^n V(\mathbf{q}_0) + \beta \kappa_V^n V(\tilde{\mathbf{q}}_0) + 2\beta K_V\big) \kappa_3 \rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0) \\
&\leq (1 + \beta V(\mathbf{q}_0) + \beta V(\tilde{\mathbf{q}}_0) + 2\beta K_V) \kappa_3 \rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0) \\
&\leq \kappa_3 (1 + 2\beta K_V)\, (1 + \beta V(\mathbf{q}_0) + \beta V(\tilde{\mathbf{q}}_0))\, \rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0) \\
&= \kappa_3 (1 + 2\beta K_V) \left(\tilde{\rho}_\beta(\mathbf{q}_0, \tilde{\mathbf{q}}_0)\right)^2.
\end{aligned} \tag{6.12}$$

<u>*Case2 :*</u> *Suppose that* $\rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0) = 1$ *and* $V(\mathbf{q}_0) + V(\tilde{\mathbf{q}}_0) > 4K_V$.

Since $\rho(\cdot,\cdot) \leq 1$ and again invoking Proposition 20 we obtain

$$
\begin{aligned}
\mathcal{W}_{\tilde{\rho}_\beta}(P^n(\mathbf{q}_0,\cdot), P^n(\tilde{\mathbf{q}}_0,\cdot))^2 & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (6.13)\\
&\leq 1 + \beta P^n V(\mathbf{q}_0) + \beta P^n V(\tilde{\mathbf{q}}_0)\\
&\leq 1 + \beta \kappa_V^n V(\mathbf{q}_0) + \beta \kappa_V^n V(\tilde{\mathbf{q}}_0) + 2\beta K_V\\
&= \frac{1+2\beta K_V}{1+3\beta K_V}(1+3\beta K_V) + \kappa_V^n \beta (V(\mathbf{q}_0) + V(\tilde{\mathbf{q}}_0))\\
&\leq \max\left\{\frac{1+2\beta K_V}{1+3\beta K_V}, 4\kappa_V^n\right\}\left(1+3\beta K_V + \frac{\beta}{4}(V(\mathbf{q}_0)+V(\tilde{\mathbf{q}}_0))\right)\\
&< \max\left\{\frac{1+2\beta K_V}{1+3\beta K_V}, 4\kappa_V^n\right\}(1+\beta V(\mathbf{q}_0) + \beta V(\tilde{\mathbf{q}}_0))\\
&= \max\left\{\frac{1+2\beta K_V}{1+3\beta K_V}, 4\kappa_V^n\right\}\left(\tilde{\rho}_\beta(\mathbf{q}_0,\tilde{\mathbf{q}}_0)\right)^2. \qquad (6.14)
\end{aligned}
$$

<u>$Case 3:$</u> *Suppose that* $\rho(\mathbf{q}_0,\tilde{\mathbf{q}}_0) = 1$ *and* $V(\mathbf{q}_0) + V(\tilde{\mathbf{q}}_0) \leq 4K_V$.

We proceed as in (6.11), but now use Proposition 24 to estimate the term $\mathcal{W}_\rho(P^n(\mathbf{q}_0,\cdot), P^n(\tilde{\mathbf{q}}_0,\cdot))$. First, let $M_V > 0$ be such that

$$
\left\{\mathbf{q} \in \mathbb{H}_\gamma : V(\mathbf{q}) \leq 4K_V\right\} = \left\{\mathbf{q} \in \mathbb{H}_\gamma : |\mathbf{q}|_\gamma \leq M_V\right\}.
$$

Notice that the specific definition of M_V depends on the choice of Lyapunov function V (which defines the constant K_V, cf. (4.3)–(4.4)). Thus, for any $\mathbf{q}_0, \tilde{\mathbf{q}}_0 \in \left\{\mathbf{q} \in \mathbb{H}_\gamma : V(\mathbf{q}) \leq 4K_V\right\}$ from Proposition 24, it follows that

$$
\mathcal{W}_\rho(P^n(\mathbf{q}_0,\cdot), P^n(\tilde{\mathbf{q}}_0,\cdot)) \leq 1 - \kappa_4,
$$

where

$$
\kappa_4 = \kappa_4(n) := \frac{1}{2}\exp\left(-\frac{16 L_1 \alpha^2 M_V^2}{T^2(1-\kappa_1^2)}\right) - \frac{2 M_V \kappa_2(n)}{\varepsilon}, \qquad (6.15)
$$

with κ_1 and κ_2 as defined in (5.4) and $\alpha = 4(1 + \lambda_1^{1-2\gamma} L_1)$ (cf. (3.21)). Hence,

$$
\begin{aligned}
\mathcal{W}_{\tilde{\rho}_\beta}(P^n(\mathbf{q}_0,\cdot), P^n(\tilde{\mathbf{q}}_0,\cdot))^2 &\leq (1-\kappa_4)\left(1 + \beta\kappa_V^n\left(V(\mathbf{q}_0) + V(\tilde{\mathbf{q}}_0)\right) + 2\beta K_V\right)\\
&\leq (1-\kappa_4)(1 + 2(1+2\kappa_V^n)\beta K_V)\\
&\leq (1-\kappa_4)(1 + 2(1+2\kappa_V^n)\beta K_V)\left(\tilde{\rho}_\beta(\mathbf{q}_0,\tilde{\mathbf{q}}_0)\right)^2. \qquad (6.16)
\end{aligned}
$$

From (6.12), (6.13) and (6.16), we now obtain the bound (6.10) with $\kappa_5 = \kappa_5(n)$ defined as

$$\kappa_5(n) = \left(\max\left\{ (1+2\beta K_V)\kappa_3(n), \max\left\{ \frac{1+2\beta K_V}{1+3\beta K_V}, 4\kappa_V^n \right\}, \right.\right.$$
$$\left.\left. (1-\kappa_4(n))(1+2(1+2\kappa_V^n)\beta K_V) \right\} \right)^{1/2}. \tag{6.17}$$

We claim that if we now choose $\varepsilon > 0$, $\beta > 0$ satisfying

$$\varepsilon \leq \frac{T(1-\kappa_1^2)^{1/2}}{8\sqrt{2}\alpha L_1^{1/2}} \quad \text{and} \quad \beta \leq \frac{1}{12 K_V} \exp\left(-\frac{16 L_1 \alpha^2 M_V^2}{T^2(1-\kappa_1^2)} \right), \tag{6.18}$$

and $n_0 \in \mathbb{N}$ satisfying

$$\kappa_1^{n_0} \leq \min\left\{ \frac{1}{4\sqrt{2}\alpha}, \frac{\varepsilon}{8\sqrt{2}\alpha M_V} \exp\left(-\frac{16 L_1 \alpha^2 M_V^2}{T^2(1-\kappa_1^2)} \right) \right\} \quad \text{and} \quad \kappa_V^{n_0} \leq \frac{1}{8}, \tag{6.19}$$

then indeed we have

$$\kappa_5(n) \leq \kappa_5(n_0) \leq \left(\max\left\{ \frac{1+2\beta K_V}{1+3\beta K_V}, 1 - \frac{1}{16} \exp\left(-\frac{32 L_1 \alpha^2 M_V^2}{T^2(1-\kappa_1^2)} \right) \right\} \right)^{1/2} < 1 \tag{6.20}$$

for all $n \geq n_0$, as we desired in the estimate (6.10).

To see this bound in (6.20) observe that since $\kappa_1^{n_0} \leq (4\sqrt{2}\alpha)^{-1}$ and ε satisfies the first inequality in (6.18), then it follows from the definitions of κ_2 and κ_3 in (5.4) and (5.3), respectively, that

$$\kappa_2(n) \leq \frac{1}{4} \quad \text{and} \quad \kappa_3(n) \leq \frac{3}{8} \quad \text{for all } n \geq n_0. \tag{6.21}$$

From (6.18), we have in particular that $\beta \leq (12 K_V)^{-1}$. Together with (6.21), this yields

$$(1+2\beta K_V)\kappa_3(n) \leq \frac{1}{2} \quad \text{for all } n \geq n_0. \tag{6.22}$$

Moreover, since $\kappa_V^{n_0} \leq 1/8$, then

$$\max\left\{ \frac{1+2\beta K_V}{1+3\beta K_V}, 4\kappa_V^n \right\} \leq \max\left\{ \frac{1+2\beta K_V}{1+3\beta K_V}, \frac{1}{2} \right\} = \frac{1+2\beta K_V}{1+3\beta K_V}. \tag{6.23}$$

Also, from the definition of κ_2 in (5.4) and the first condition in (6.19), it follows that κ_4, defined in (6.15), satisfies

$$\kappa_4(n) \geq \frac{1}{4} \exp\left(-\frac{16 L_1 \alpha^2 M_V^2}{T^2(1-\kappa_1^2)} \right) \quad \text{for all } n \geq n_0. \tag{6.24}$$

Thus, with condition (6.18) on β, we obtain

$$(1-\kappa_4(n))(1+3\beta K_V) \leq 1 - \frac{1}{16}\exp\left(-\frac{32L_1\alpha^2 M_V^2}{T^2(1-\kappa_1^2)}\right) \quad \text{for all } n \geq n_0. \quad (6.25)$$

Combining now (6.17), (6.22), (6.23) and (6.25) we now conclude (6.20).

We turn now to show that (6.10) implies (6.7) and, consequently, (6.2). First note that, by the same arguments as in (6.7)–(6.8) we have that (6.10) implies $\mathcal{W}_{\tilde{\rho}_\beta}(\nu_1 P^n, \nu_2 P^n) \leq \kappa_5 \mathcal{W}_{\tilde{\rho}_\beta}(\nu_1, \nu_2)$ for all $n \geq n_0$ and $\nu_1, \nu_2 \in \Pr(\mathbb{H}_\gamma)$ with support included in $\mathbb{H}_\gamma$. Now, for any $n \in \mathbb{N}$, we can write $n = mn_0 + k$, for some $m, k \in \mathbb{N}$ with $k \leq n_0 - 1$. Thus,

$$\begin{aligned}
\mathcal{W}_{\tilde{\rho}_\beta}(P^n(\mathbf{q}_0, \cdot), P^n(\tilde{\mathbf{q}}_0, \cdot)) &= \mathcal{W}_{\tilde{\rho}_\beta}(P^{mn_0+k}(\mathbf{q}_0, \cdot), P^{mn_0+k}(\tilde{\mathbf{q}}_0, \cdot)) \\
&\leq \kappa_5(n_0)^m \mathcal{W}_{\tilde{\rho}_\beta}(P^k(\mathbf{q}_0, \cdot), P^k(\tilde{\mathbf{q}}_0, \cdot)) \\
&\leq \kappa_5(n_0)^m \kappa_5(k) \tilde{\rho}_\beta(\mathbf{q}_0, \tilde{\mathbf{q}}_0) \\
&\leq \kappa_5(n_0)^{\frac{n}{n_0}-1} \kappa_5(n_0-1) \tilde{\rho}_\beta(\mathbf{q}_0, \tilde{\mathbf{q}}_0),
\end{aligned}$$

where in the last inequality we used that κ_5 is a non-increasing function of n. Moreover, from the equivalence between $\tilde{\rho}$ and $\tilde{\rho}_\beta$ in (6.9), we obtain

$$\begin{aligned}
\mathcal{W}_{\tilde{\rho}}(P^n(\mathbf{q}_0, \cdot), P^n(\tilde{\mathbf{q}}_0, \cdot)) &\leq \left(\frac{\max\{1,\beta\}}{\min\{1,\beta\}}\right)^{1/2} \kappa_5(n_0)^{\frac{n}{n_0}-1} \kappa_5(n_0-1) \tilde{\rho}(\mathbf{q}_0, \tilde{\mathbf{q}}_0) \\
&\leq \left(\frac{\max\{1,\beta\}}{\min\{1,\beta\}}\right)^{1/2} \frac{\kappa_5(n_0-1)}{\kappa_5(n_0)} \exp\left(n \log\left(\kappa_5(n_0)^{\frac{1}{n_0}}\right)\right) \tilde{\rho}(\mathbf{q}_0, \tilde{\mathbf{q}}_0) \quad \text{for all } n \in \mathbb{N}.
\end{aligned} \quad (6.26)$$

Therefore, with the constants

$$C_1 := \left(\frac{\max\{1,\beta\}}{\min\{1,\beta\}}\right)^{1/2} \frac{\kappa_5(n_0-1)}{\kappa_5(n_0)} \quad \text{and} \quad C_2 := -\log\left(\kappa_5(n_0)^{\frac{1}{n_0}}\right), \quad (6.27)$$

(6.7) and consequently (6.2) are now established.

Finally, the second part of the proof, namely (6.4)–(6.6) under assumption (6.3), follow as a direct consequence of Proposition 40 and 43 combined with Proposition 46.

□

7 Implications for the finite dimensional setting

The approach given above can be modified in a straightforward fashion to provide a novel proof of the ergodicity of the exact HMC algorithm in finite dimensions. We detail this connection in this section. We abuse notation and use the same terminology for the analogous constants and operators from the infinite-dimensional case introduced in the previous sections.

We take our phase space to be $\mathbb{H} = \mathbb{R}^k$, $k \in \mathbb{N}$, endowed with the Euclidean inner product and norm, which are denoted by $\langle \cdot, \cdot \rangle$ and $|\cdot|$, respectively. Similarly to (1.1) above we fix a target probability measure of the form

$$\mu(d\mathbf{q}) \propto \exp(-U(\mathbf{q}))\mu_0(d\mathbf{q}) \quad \text{with } \mu_0 = \mathcal{N}(0, \mathcal{C}), \tag{7.1}$$

where $\mathcal{C}$ is a symmetric strictly positive-definite covariance matrix. Here we aim to sample from μ using the dynamics

$$\frac{d\mathbf{q}}{dt} = \mathcal{M}^{-1}\mathbf{p} \quad \frac{d\mathbf{p}}{dt} = -\mathcal{C}^{-1}\mathbf{q} - DU(\mathbf{q}) \tag{7.2}$$

corresponding to the Hamiltonian

$$H(\mathbf{q}, \mathbf{p}) = \frac{1}{2}\langle \mathcal{C}^{-1}\mathbf{q}, \mathbf{q}\rangle + U(\mathbf{q}) + \frac{1}{2}\langle \mathcal{M}^{-1}\mathbf{p}, \mathbf{p}\rangle, \tag{7.3}$$

where $\mathcal{M}$ is a user-specified ‘mass matrix’ which we suppose to be symmetric and strictly positive definite; and $U : \mathbb{R}^k \to \mathbb{R}$ is a C^2 potential function. Let us denote by $\lambda_{\mathcal{M}}$ and $\Lambda_{\mathcal{M}}$ the smallest and largest eigenvalues of $\mathcal{M}$. Analogously, let $\lambda_{\mathcal{C}}$ and $\Lambda_{\mathcal{C}}$ be the smallest and largest eigenvalues of $\mathcal{C}$.

We impose the following conditions on the potential function U (cf. Assumption 8 above):

Assumption 27

(F1) There exists a constant $L_1 \geq 0$ such that

$$|D^2U(\mathbf{f})| \leq L_1 \quad \text{for any } \mathbf{f} \in \mathbb{R}^k. \tag{7.4}$$

(F2) There exist constants $L_2 > 0$ and $L_3 \geq 0$ such that

$$|\mathcal{M}^{-1/2}\mathcal{C}^{-1/2}\mathbf{f}|^2 + \langle \mathbf{f}, \mathcal{M}^{-1}DU(\mathbf{f})\rangle \geq L_2|\mathcal{M}^{-1/2}\mathcal{C}^{-1/2}\mathbf{f}|^2 - L_3 \quad \text{for any } \mathbf{f} \in \mathbb{R}^k. \tag{7.5}$$

Note that under (7.4), U is globally Lipschitz so that (7.2) yields a well defined dynamical system on $C^1(\mathbb{R}, \mathbb{R}^k)$ as above in Proposition 12. Furthermore, similarly as in Remark 9, we have:

(i) From (7.4), it follows that

$$|DU(\mathbf{f})| \leq L_1|\mathbf{f}| + L_0 \quad \text{for every } \mathbf{f} \in \mathbb{R}^k. \tag{7.6}$$

where $L_0 = |DU(0)|$.

(ii) If $|DU(\mathbf{f})| \leq L_4|\mathbf{f}| + L_5$ for some $L_4 \in [0, \lambda_{\mathcal{M}}(\Lambda_{\mathcal{M}}\Lambda_{\mathcal{C}})^{-1})$ and $L_5 \geq 0$, then (7.5) follows.

(iii) Assumptions $(F1)$ and $(F2)$ imply that

$$L_2 \leq 1 + \Lambda_{\mathcal{M}} \Lambda_{\mathcal{C}} \lambda_{\mathcal{M}}^{-1} L_1. \tag{7.7}$$

Fixing an integration time $T > 0$, and under the given conditions on $\mathcal{C}$, $\mathcal{M}$ and U in (7.2) we have a well-defined Feller Markov transition kernel defined as

$$P(\mathbf{q}_0, A) = \mathbb{P}(q_T(\mathbf{q}_0, \mathbf{p}_0) \in A) \tag{7.8}$$

for any $\mathbf{q}_0 \in \mathbb{R}^k$ and any Borel set $A \subset \mathbb{R}^k$, where

$$\mathbf{p}_0 \sim N(0, \mathcal{M}). \tag{7.9}$$

Here, following previous notation, $q_T(\mathbf{q}_0, \mathbf{p}_0)$ is the solution of (7.2) at time T starting from the initial position $\mathbf{q}_0 \in \mathbb{R}^k$ and momentum $\mathbf{p}_0 \in \mathbb{R}^k$. The n-fold iteration of the kernel P is denoted as P^n.

As in Theorem 26, we measure the convergence of P^n using a suitable Wasserstein distance. In this case, we take

$$\tilde{\rho}(\mathbf{q}, \tilde{\mathbf{q}}) = \sqrt{\rho(\mathbf{q}, \tilde{\mathbf{q}})(1 + V(\mathbf{q}) + V(\tilde{\mathbf{q}}))} \quad \text{where} \quad \rho(\mathbf{q}, \tilde{\mathbf{q}}) = \frac{|\mathbf{q} - \tilde{\mathbf{q}}|}{\varepsilon} \wedge 1 \tag{7.10}$$

and V is a Foster–Lyapunov function defined as either $V(\mathbf{q}) = V_{1,i}(\mathbf{q}) = |\mathbf{q}|^i$, $i \in \mathbb{N}$, or as $V(\mathbf{q}) = V_{2,\eta}(\mathbf{q}) = \exp(\eta |\mathbf{q}|^2)$, with $\eta > 0$ satisfying

$$\eta < \left[2 \operatorname{Tr}(\mathcal{M}) \left(\frac{67}{8} T^2 + \frac{32}{L_2 (\Lambda_{\mathcal{M}} \Lambda_{\mathcal{C}})^{-1}} \right) \lambda_{\mathcal{M}}^{-2} \right]^{-1}. \tag{7.11}$$

We then consider the corresponding Wasserstein distance $\mathcal{W}_{\tilde{\rho}}$ and prove the theorem below concerning the exact HMC kernel P.

Theorem 28 *Consider the Markov kernel P defined as (7.8), (7.9) from the dynamics (7.2). We suppose that $\mathcal{M}$ and $\mathcal{C}$ in (7.2) are both symmetric and strictly positive definite and we assume that the potential function U satisfies Assumption 27. In addition, we impose the following condition on the integration time $T > 0$:*

$$T \leq \min \left\{ \frac{1}{\left[2 \lambda_{\mathcal{M}}^{-1} (\lambda_{\mathcal{C}}^{-1} + L_1) \right]^{1/2}}, \frac{L_2^{1/2} (\Lambda_{\mathcal{M}} \Lambda_{\mathcal{C}})^{-1/2}}{2\sqrt{6} \lambda_{\mathcal{M}}^{-1} (\lambda_{\mathcal{C}}^{-1} + L_1)} \right\}, \tag{7.12}$$

where $\lambda_{\mathcal{M}}$ and $\Lambda_{\mathcal{M}}$ denote the smallest and largest eigenvalues of $\mathcal{M}$, while $\lambda_{\mathcal{C}}$ and $\Lambda_{\mathcal{C}}$ denote the smallest and largest eigenvalues of $\mathcal{C}$, respectively.

Then P has a unique ergodic invariant measure given by μ in (7.1). Moreover, P satisfies the following spectral gap condition with respect to the Wasserstein distance

$\mathcal{W}_{\tilde\rho}$ associated to $\tilde\rho$ defined in (7.10): For all ν_1, ν_2 Borel probability measures on $\mathbb{R}^k$,

$$\mathcal{W}_{\tilde\rho}(\nu_1 P^n, \nu_2 P^m) \le C_1 e^{-C_2 n} \mathcal{W}_{\tilde\rho}(\nu_1, \nu_2) \quad \text{for all } n \in \mathbb{N}, \tag{7.13}$$

where the constants $C_1, C_2, \varepsilon > 0$ are independent of ν_1, ν_2 and k, and can be given explicitly as depending exclusively on L_1, L_2, L_3, T, $\mathcal{M}$ and $\mathcal{C}$.

Remark 29 Similarly as in Theorem 26, we can also show that (7.13) implies a convergence result with respect to suitable observables as in (6.4), as well as a strong law of large numbers and a central limit theorem analogous to (6.5)–(6.6).

Proof The proof follows very similar steps to the results from Sects. 3, 4, 5 and 6, so we only point out the main differences.

From (7.2), it follows that

$$\frac{d^2\mathbf{q}}{dt^2} = -\mathcal{M}^{-1}\mathcal{C}^{-1}\mathbf{q} - \mathcal{M}^{-1}DU(\mathbf{q}),$$

so that, after integrating with respect to $t \in [0, T]$ twice, we have

$$\mathbf{q}_t - (\mathbf{q}_0 + t\mathcal{M}^{-1}\mathbf{p}_0) = -\int_0^t \int_0^s \left(\mathcal{M}^{-1}\mathcal{C}^{-1}\mathbf{q}_\tau + \mathcal{M}^{-1}DU(\mathbf{q}_\tau)\right) d\tau ds \tag{7.14}$$

Using that

$$|\mathcal{M}^{-1}\mathbf{f}| \le \lambda_{\mathcal{M}}^{-1}|\mathbf{f}| \quad \text{and} \quad |\mathcal{C}^{-1}\mathbf{f}| \le \lambda_{\mathcal{C}}^{-1}|\mathbf{f}| \quad \text{for every } \mathbf{f} \in \mathbb{R}^k,$$

together with (7.6) and the condition $T \le [\lambda_{\mathcal{M}}^{-1}(\lambda_{\mathcal{C}}^{-1} + L_1)]^{-1/2}$, one obtains, analogously to (3.2) and (3.3),

$$\sup_{t\in[0,T]} |\mathbf{q}_t - (\mathbf{q}_0 + t\mathcal{M}^{-1}\mathbf{p}_0)| \le \lambda_{\mathcal{M}}^{-1}(\lambda_{\mathcal{C}}^{-1} + L_1)T^2 \max\left\{|\mathbf{q}_0|, |\mathbf{q}_0 + T\mathcal{M}^{-1}\mathbf{p}_0|\right\} + \lambda_{\mathcal{M}}^{-1}L_0T^2 \tag{7.15}$$

and

$$\begin{aligned}\sup_{t\in[0,T]} |\mathbf{p}_t - \mathbf{p}_0| \le &(\lambda_{\mathcal{C}}^{-1} + L_1)t\left[1 + \lambda_{\mathcal{M}}^{-1}(\lambda_{\mathcal{C}}^{-1} + L_1)t^2\right] \max\left\{|\mathbf{q}_0|, |\mathbf{q}_0 + T\mathcal{M}^{-1}\mathbf{p}_0|\right\} \\ &+ L_0 t\left[1 + \lambda_{\mathcal{M}}^{-1}(\lambda_{\mathcal{C}}^{-1} + L_1)t^2\right].\end{aligned} \tag{7.16}$$

Moreover, analogously to (3.11), we obtain that for every $(\mathbf{q}_0, \mathbf{p}_0), (\tilde{\mathbf{q}}_0, \tilde{\mathbf{p}}_0) \in \mathbb{R}^k \times \mathbb{R}^k$,

$$\begin{aligned}&\sup_{t\in[0,T]} |\mathbf{q}_t(\mathbf{q}_0, \mathbf{p}_0) - \mathbf{q}_t(\tilde{\mathbf{q}}_0, \tilde{\mathbf{p}}_0) - [(\mathbf{q}_0 - \tilde{\mathbf{q}}_0) + t\mathcal{M}^{-1}(\mathbf{p}_0 - \tilde{\mathbf{p}}_0)]| \\ &\quad\le \lambda_{\mathcal{M}}^{-1}(\lambda_{\mathcal{C}}^{-1} + L_1)T^2 \max\left\{|\mathbf{q}_0 - \tilde{\mathbf{q}}_0|, |\mathbf{q}_0 - \tilde{\mathbf{q}}_0 + t\mathcal{M}^{-1}(\mathbf{p}_0 - \tilde{\mathbf{p}}_0)|\right\}.\end{aligned}$$

In particular, if $\tilde{\mathbf{p}}_0 = \mathbf{p}_0 + \mathcal{M}(\mathbf{q}_0 - \tilde{\mathbf{q}}_0)T^{-1}$ then

$$\sup_{t\in[0,T]} |\mathbf{q}_t(\mathbf{q}_0, \mathbf{p}_0) - \mathbf{q}_t(\tilde{\mathbf{q}}_0, \tilde{\mathbf{p}}_0)| \le \lambda_{\mathcal{M}}^{-1}(\lambda_{\mathcal{C}}^{-1} + L_1)T^2|\mathbf{q}_0 - \tilde{\mathbf{q}}_0| \le \frac{1}{2}|\mathbf{q}_0 - \tilde{\mathbf{q}}_0|. \tag{7.17}$$

We also show that $V(\mathbf{q}) = |\mathbf{q}|^i$ with $i \ge 1$ or $V(\mathbf{q}) = \exp(\eta|\mathbf{q}|^2)$, with $\eta > 0$ satisfying (7.11), all verify a Foster–Lyapunov structure as in Definition 19. The proof follows as in Proposition 20, with the difference starting from (4.7), which is now written as

$$\frac{d}{ds}\langle \mathbf{q}_s, \mathcal{M}^{-1}\mathbf{p}_s\rangle = |\mathcal{M}^{-1}\mathbf{p}_s|^2 - |\mathcal{M}^{-1/2}\mathcal{C}^{-1/2}\mathbf{q}_s|^2 - \langle \mathbf{q}_s, \mathcal{M}^{-1}DU(\mathbf{q}_s)\rangle. \tag{7.18}$$

Using now $(F2)$ from Assumption 27 and the inequalities

$$|\mathcal{M}^{-1/2}\mathbf{f}| \ge \Lambda_{\mathcal{M}}^{-1/2}|\mathbf{f}| \quad \text{and} \quad |\mathcal{C}^{-1/2}\mathbf{f}| \ge \Lambda_{\mathcal{C}}^{-1/2}|\mathbf{f}| \quad \text{for all } \mathbf{f} \in \mathbb{R}^k,$$

we obtain from (7.18) that

$$|\mathbf{q}_T|^2 \le |\mathbf{q}_0|^2 + 2T\langle \mathbf{q}_0, \mathcal{M}^{-1}\mathbf{p}_0\rangle + 2\int_0^T \int_0^s \left[\lambda_{\mathcal{M}}^{-2}|\mathbf{p}_\tau|^2 - L_2(\Lambda_{\mathcal{M}}\Lambda_{\mathcal{C}})^{-1}|\mathbf{q}_\tau|^2 + L_3\right] d\tau\, ds. \tag{7.19}$$

Then, with (7.7), the a priori bounds (7.15)–(7.16) and the fact that $2\lambda_{\mathcal{M}}^{-1}(\lambda_{\mathcal{C}}^{-1} + L_1)T^2 \le 1$ from hypothesis (7.12), we arrive at

$$\begin{aligned} |\mathbf{q}_T|^2 \le{}& \left(1 + \frac{3}{2}\lambda_{\mathcal{M}}^{-2}(\lambda_{\mathcal{C}}^{-1} + L_1)^2T^4 - \frac{L_2}{8}(\Lambda_{\mathcal{M}}\Lambda_{\mathcal{C}})^{-1}T^2\right)|\mathbf{q}_0|^2 + 2T\langle \mathbf{q}_0, \mathcal{M}^{-1}\mathbf{p}_0\rangle \\ &+ \frac{67}{8}\lambda_{\mathcal{M}}^{-2}T^2|\mathbf{p}_0|^2 + \frac{3}{2}L_0^2\lambda_{\mathcal{M}}^{-2}T^4 + \frac{L_0^2}{6}\lambda_{\mathcal{M}}^{-2}T^4 + L_3T^2. \end{aligned} \tag{7.20}$$

From the second condition in hypothesis (7.12) it follows that $(3/2)\lambda_{\mathcal{M}}^{-2}(\lambda_{\mathcal{C}}^{-1} + L_1)^2T^4 \le (L_2/16)(\Lambda_{\mathcal{M}}\Lambda_{\mathcal{C}})^{-1}T^2$, so that after taking expected values in (7.20) we obtain

$$\mathbb{E}|\mathbf{q}_T|^2 \le \exp\left(-\frac{L_2}{16}(\Lambda_{\mathcal{M}}\Lambda_{\mathcal{C}})^{-1}T^2\right)|\mathbf{q}_0|^2 + \left(\frac{67}{8}\lambda_{\mathcal{M}}^{-2}\operatorname{Tr}(\mathcal{M}) + \frac{5}{3}\lambda_{\mathcal{M}}^{-2}L_0^2T^2 + L_3\right)T^2.$$

Now proceeding analogously as in (4.15)–(4.20), we obtain that for $V : \mathbb{R}^k \to \mathbb{R}$ given either as $V(\mathbf{q}) = |\mathbf{q}|^i$, $i \in \mathbb{N}$, or $V(\mathbf{q}) = \exp(\eta|\mathbf{q}|^2)$, with $\eta > 0$ satisfying (7.11), there exist constants $\kappa_V \in [0, 1)$ and $K_V > 0$ such that

$$P^nV(\mathbf{q}_0) \le \kappa_V^n V(\mathbf{q}_0) + K_V \quad \text{for all } \mathbf{q}_0 \in \mathbb{R}^k, \text{ for all } n \in \mathbb{N}, \tag{7.21}$$

i.e. these are Lyapunov functions for P.

Let $(\mathbb{R}^k)^n$ denote the product of n copies of $\mathbb{R}^k$ and let $\mathcal{N}(0, \mathcal{M})^{\otimes n}$ denote the product of n copies of $\mathcal{N}(0, \mathcal{M})$. Analogously to Sect. 5, given $\mathbf{q}_0 \in \mathbb{R}^k$ and a sequence $\{\mathbf{p}_0^{(j)}\}_{j\in\mathbb{N}}$ of i.i.d. draws from $\mathcal{N}(0,\mathcal{M})$, we denote $\mathbf{P}_0^{(n)}=(\mathbf{p}_0^{(1)},\ldots,\mathbf{p}_0^{(n)})$, for all $n\in\mathbb{N}$, and take $Q_n(\mathbf{q}_0,\cdot):(\mathbb{R}^k)^n\to\mathbb{R}^k$, according to

$$Q_1(\mathbf{q}_0,\mathbf{p}_0^{(1)})=\mathbf{q}_T(\mathbf{q}_0,\mathbf{p}_0^{(1)}),\quad Q_n(\mathbf{q}_0,\mathbf{P}_0^{(n)})=\mathbf{q}_T(Q_{n-1}(\mathbf{q}_0,\mathbf{P}_0^{(n-1)}),\mathbf{p}_0^{(n)}).$$

for all $n\geq 2$.

Similarly, given any $\mathbf{q}_0,\tilde{\mathbf{q}}_0\in\mathbb{R}^k$ we take $\widetilde{Q}_n(\mathbf{q}_0,\tilde{\mathbf{q}}_0,\cdot):(\mathbb{R}^k)^n\to\mathbb{R}^k$ to be the random variables starting from

$$\widetilde{Q}_1(\mathbf{q}_0,\tilde{\mathbf{q}}_0,\mathbf{p}_0^{(1)})=\mathbf{q}_T(\tilde{\mathbf{q}}_0,\mathbf{p}_0^{(1)}+T^{-1}\mathcal{M}(\mathbf{q}_0-\tilde{\mathbf{q}}_0)),$$

then defined for each integer $n\geq 2$ as

$$\widetilde{Q}_n(\mathbf{q}_0,\tilde{\mathbf{q}}_0,\mathbf{P}_0^{(n)})=\mathbf{q}_T(\widetilde{Q}_{n-1}(\mathbf{q}_0,\tilde{\mathbf{q}}_0,\mathbf{P}_0^{(n-1)}),\mathbf{p}_0^{(n)}+\mathcal{S}_n(\mathbf{P}_0^{(n-1)}))$$

with

$$\mathcal{S}_n(\mathbf{P}_0^{(n-1)})=T^{-1}\mathcal{M}\left[Q_{n-1}(\mathbf{q}_0,\mathbf{P}_0^{(n-1)})-\widetilde{Q}_{n-1}(\mathbf{q}_0,\tilde{\mathbf{q}}_0,\mathbf{P}_0^{(n-1)})\right]\quad\text{for all } n\geq 2. \tag{7.22}$$

We also denote

$$\boldsymbol{\mathcal{S}}_n(\mathbf{P}_0^{(n)})=(\mathcal{S}_1,\mathcal{S}_2(\mathbf{P}_0^{(1)}),\ldots,\mathcal{S}_n(\mathbf{P}_0^{(n-1)})),\quad\text{with } \mathcal{S}_1=T^{-1}(\mathbf{q}_0-\tilde{\mathbf{q}}_0),$$

and $\boldsymbol{\Psi}_n(\mathbf{P}_0^{(n)})=\mathbf{P}_0^{(n)}+\boldsymbol{\mathcal{S}}_n(\mathbf{P}_0^{(n)})$. Thus, by using inequality (7.17) n times iteratively, we obtain that

$$|Q_n(\mathbf{q}_0,\mathbf{P}_0^{(n)})-\widetilde{Q}_n(\mathbf{q}_0,\tilde{\mathbf{q}}_0,\mathbf{P}_0^{(n)})|\leq\frac{1}{2^n}|\mathbf{q}_0-\tilde{\mathbf{q}}_0|, \tag{7.23}$$

for all $\mathbf{P}_0^{(n)}\in(\mathbb{R}^k)^n$.

Let $\sigma_n=\mathrm{Law}(\mathbf{P}_0^{(n)})=\mathcal{N}(0,\mathcal{M})^{\otimes n}$ and $\tilde{\sigma}_n=\mathrm{Law}(\boldsymbol{\Psi}_n(\mathbf{P}_0^{(n)}))=\boldsymbol{\Psi}_n^*\nu_n$. Analogously as in Propositions 22 and 24, we obtain that the distance-like function ρ defined in (7.10) satisfies contractivity and smallness properties with respect to the Markov operator P^n for n sufficiently large. Here, the main difference lies in the estimate of Kullback-Leibler Divergence $D_{\mathrm{KL}}(\tilde{\sigma}_n|\sigma_n)$, (5.15). Proceeding similarly as in (5.22)–(5.23), we arrive at

$$D_{\mathrm{KL}}(\tilde{\sigma}_n|\sigma_n)\leq\frac{1}{2}\sum_{j=1}^{n}\mathbb{E}|\mathcal{M}^{-1/2}\mathcal{S}_j(\cdot)|^2.$$

Using (7.23), it follows that for every $j\in\{1,\ldots,n\}$ and $\mathbf{P}_0^{(j-1)}\in(\mathbb{R}^k)^{(j-1)}$

$$|\mathcal{M}^{-1/2}\mathcal{S}_j(\mathbf{P}_0^{(j-1)})|^2\leq\lambda_{\mathcal{M}}^{-1}|\mathcal{S}_j(\mathbf{P}_0^{(j-1)})|^2$$

$$\leq \lambda_{\mathcal{M}}^{-1}\Lambda_{\mathcal{M}}^2 T^{-2}|Q_{j-1}(\mathbf{q}_0)(\mathbf{P}_0^{(j-1)}) - \widetilde{Q}_{j-1}(\mathbf{q}_0, \tilde{\mathbf{q}}_0)(\mathbf{P}_0^{(j-1)})|^2$$
$$\leq \frac{\lambda_{\mathcal{M}}^{-1}\Lambda_{\mathcal{M}}^2 T^{-2}}{2^{(j-1)2}}|\mathbf{q}_0 - \tilde{\mathbf{q}}_0|^2,$$

where in the second inequality we used that $|\mathcal{M}\cdot|^2 \leq \Lambda_{\mathcal{M}}^2|\cdot|^2$. Hence,

$$D_{\mathrm{KL}}(\tilde{\sigma}_n|\sigma_n) \leq \frac{\lambda_{\mathcal{M}}^{-1}\Lambda_{\mathcal{M}}^2 T^{-2}}{2}|\mathbf{q}_0 - \tilde{\mathbf{q}}_0|^2 \sum_{j=1}^{n} \frac{1}{2^{(j-1)2}} \leq \frac{4\Lambda_{\mathcal{M}}^2}{\lambda_{\mathcal{M}} T^2}|\mathbf{q}_0 - \tilde{\mathbf{q}}_0|^2. \quad (7.24)$$

By using (7.24), one obtains analogously as in Proposition 22 that for every $n \in \mathbb{N}$ and for every $\mathbf{q}_0, \tilde{\mathbf{q}}_0 \in \mathbb{R}^k$ such that $\rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0) < 1$, we have

$$\mathcal{W}_\rho(P^n(\mathbf{q}_0,\cdot), P^n(\tilde{\mathbf{q}}_0,\cdot)) \leq \left(\frac{1}{2^n} + \frac{\sqrt{2}\Lambda_{\mathcal{M}}\varepsilon}{\lambda_{\mathcal{M}}^{1/2} T}\right)\rho(\mathbf{q}_0, \tilde{\mathbf{q}}_0). \quad (7.25)$$

Moreover, analogously as in Proposition 24, we obtain that, given $M \geq 0$, for every $\mathbf{q}_0, \tilde{\mathbf{q}}_0 \in A := \{\mathbf{q} \in \mathbb{R}^k \,:\, |\mathbf{q}| \leq M\}$, it holds:

$$\mathcal{W}_\rho(P^n(\mathbf{q}_0,\cdot), P^n(\tilde{\mathbf{q}}_0,\cdot)) \leq 1 - \frac{1}{2}\exp\left(-\frac{16\Lambda_{\mathcal{M}}^2 M^2}{\lambda_{\mathcal{M}} T^2}\right) + \frac{M}{2^{n-1}\varepsilon}. \quad (7.26)$$

The remaining portion of the proof now follows as for Theorem 26, by combining (7.21), (7.25) and (7.26). □

Remark 30 From condition (7.12) on the integration time T, we see how the upper bound could potentially degenerate to zero in case the eigenvalues of $\mathcal{C}$ and/or the eigenvalues of $\mathcal{M}$ decay to zero as the dimension of $\mathbb{R}^k$ increases. Moreover, if the eigenvalues of $\mathcal{M}$ decrease to zero (i.e. $\lambda_{\mathcal{M}} \to 0$) or increase to infinity (i.e. $\Lambda_{\mathcal{M}} \to \infty$) with respect to k, then, for fixed n, ε and T, the upper bound in (7.25) increases to infinity, and the first two terms in the upper bound in (7.26) increase to 1. This would imply that the convergence rate in (7.13), which is directly proportional to the upper bounds in (7.25)–(7.26) and inversely proportional to T, would become ‘slower’ as k increases. In other words, the number n of iterations necessary for the distance between $\nu_1 P^n$ and $\nu_2 P^n$ to decay within a given $\delta > 0$ would increase with the dimension k. This type of behavior is commonly known as the ‘curse of dimensionality’.

A natural choice for the mass matrix $\mathcal{M}$ to avoid such unwanted behavior is given by $\mathcal{M} = \mathcal{C}^{-1}$ – this is the idea behind preconditioning in [6] which leads us to consider (1.2) in the infinite dimensional formulation. In this preconditioned case, one could use that $\lambda_{\mathcal{M}} = \Lambda_{\mathcal{C}}^{-1}$ and $\Lambda_{\mathcal{M}} = \lambda_{\mathcal{C}}^{-1}$ directly in (7.12) to obtain

$$T \leq \min\left\{\frac{1}{[2\Lambda_{\mathcal{C}}(\lambda_{\mathcal{C}}^{-1} + L_1)]^{1/2}}, \frac{L_2^{1/2}(\lambda_{\mathcal{C}}^{-1}\Lambda_{\mathcal{C}})^{-1/2}}{2\sqrt{6}\Lambda_{\mathcal{C}}(\lambda_{\mathcal{C}}^{-1} + L_1)}\right\}, \quad (7.27)$$

where the upper bound actually still degenerates to zero in case $\lambda_{\mathcal{C}} \to 0$ as $k \to \infty$ (corresponding to the trace-class assumption on $\mathcal{C}$ in the infinite-dimensional case). However, the inequalities that lead to the condition on T as in (7.27) would in fact be a rough overestimate in this case. Indeed, for $\mathcal{M} = \mathcal{C}^{-1}$, the term $\mathcal{M}^{-1}\mathcal{C}^{-1}\mathbf{q}_\tau$ in (7.14) is simply equal to $\mathbf{q}_\tau$ and thus we no longer estimate from above by $\lambda_{\mathcal{M}}^{-1}\lambda_{\mathcal{C}}^{-1}|\mathbf{q}_\tau|$ as in (7.15). Similarly, the term $|\mathcal{M}^{-1/2}\mathcal{C}^{-1/2}\mathbf{q}_s|^2$ in (7.18) is simply $|\mathbf{q}_s|^2$ and thus no longer estimated from below by $(\Lambda_{\mathcal{M}}\Lambda_{\mathcal{C}})^{-1}|\mathbf{q}_\tau|^2$ as in (7.19). With these changes, T is required to satisfy instead

$$T \leq \min\left\{\frac{1}{[2(1+\Lambda_{\mathcal{C}}L_1)]^{1/2}}, \frac{L_2^{1/2}}{2\sqrt{6}(1+\Lambda_{\mathcal{C}}L_1)}\right\},$$

which is consistent with condition (6.1) for $\Lambda_{\mathcal{C}} = \lambda_1$ (when $\gamma = 0$), and thus independent of k when $\Lambda_{\mathcal{C}}$ is uniformly bounded with respect to k.

On the other hand, replacing $\Lambda_{\mathcal{M}}$ with $\lambda_{\mathcal{C}}^{-1}$ in (7.25) and (7.26), we see that the same unwanted behavior is not removed here when $\lambda_{\mathcal{C}} \to 0$ as $k \to \infty$; i.e. the convergence rate would still degenerate with the dimension k. This emphasizes the need for considering 'shifts' in the momentum (or velocity) paths for the modified process $\widetilde{Q}_n(\mathbf{q}_0, \tilde{\mathbf{q}}_0, \cdot)$, $\mathbf{q}_0, \tilde{\mathbf{q}}_0 \in \mathbb{R}^k$, that are restricted to a fixed number of directions in $\mathbb{R}^k$, for every k, as done in (5.7) through the projection operator Π_N, with N sufficiently large but fixed (cf. (7.22)).

8 Application for the Bayesian estimation of divergence free flows from a passive scalar

In this section we establish some results concerning the degree of applicability of Theorem 26 to the PDE inverse problem of estimating a divergence free flow from a passive scalar as we described above in the introduction, cf. (1.9), (1.10), (1.11).

For this purpose, according to the conditions required in Assumption 8, we wish to establish suitable bounds on U, DU and D^2U. Of course such bounds are expected to depend crucially on the form of the observation operator $\mathcal{O}$. Here, adopting the notations $U^\xi = \langle DU, \xi\rangle$ and $U^{\xi,\tilde{\xi}} = \langle D^2U\xi, \tilde{\xi}\rangle$ for directional derivatives of U with respect to vectors $\xi, \tilde{\xi}$ in the phase space, we have that

$$U^\xi(\mathbf{q}) = -2\langle \Gamma^{-1/2}(\mathcal{Y} - \mathcal{O}(\theta(\mathbf{q}))), \Gamma^{-1/2}\mathcal{O}(\psi^\xi(\mathbf{q}))\rangle \tag{8.1}$$

and

$$\begin{aligned} &U^{\xi,\tilde{\xi}}(\mathbf{q}) \\ &= 2\langle \Gamma^{-1/2}\mathcal{O}(\psi^{\tilde{\xi}}(\mathbf{q})), \Gamma^{-1/2}\mathcal{O}(\psi^\xi(\mathbf{q}))\rangle - 2\langle \Gamma^{-1/2}(\mathcal{Y} - \mathcal{O}(\theta(\mathbf{q}))), \Gamma^{-1/2}\mathcal{O}(\psi^{\xi,\tilde{\xi}}(\mathbf{q}))\rangle \end{aligned} \tag{8.2}$$

where $\psi^\xi(\mathbf{q}) = \psi^\xi(t; \mathbf{q})$ obeys

$$\partial_t\psi^\xi + \mathbf{q}\cdot\nabla\psi^\xi = \kappa\Delta\psi^\xi - \xi\cdot\nabla\theta(\mathbf{q}), \quad \psi^\xi(0; \mathbf{q}) = 0 \tag{8.3}$$

and $\psi^{\xi,\tilde{\xi}}(\mathbf{q}) = \psi^{\xi,\tilde{\xi}}(t;\mathbf{q})$ satisfies

$$\partial_t \psi^{\xi,\tilde{\xi}} + \mathbf{q} \cdot \nabla \psi^{\xi,\tilde{\xi}} = \kappa \Delta \psi^{\xi,\tilde{\xi}} - \tilde{\xi} \cdot \nabla \psi^{\xi} - \xi \cdot \nabla \psi^{\tilde{\xi}}, \quad \psi^{\xi,\tilde{\xi}}(0;\mathbf{q}) = 0, \tag{8.4}$$

for any suitable $\xi, \tilde{\xi}$.

8.1 Mathematical setting of the advection diffusion equation, associated bounds

In order to place (1.11) in a rigorous functional setting we adapt some results from [11,12]. In view of (8.3), (8.4) we consider a slightly more general version of (1.9) where we include an external forcing term $f : [0, T] \times \mathbb{T}^2 \to \mathbb{R}$, namely,

$$\partial_t \phi + \mathbf{q} \cdot \nabla \phi = \kappa \Delta \phi + f, \quad \phi(0) = \phi_0. \tag{8.5}$$

Specially, we need to estimate terms appearing in the gradient and Hessian of U involving solutions of (8.5) with certain forcing terms; cf. (8.3), (8.4) below.

We adopt the notation $H^s(\mathbb{T}^2)$ for the Sobolev space of periodic functions with $s \geq 0$ derivatives in L^2. Here we denote $\Lambda^s = (-\Delta)^{s/2}$. Thus, the associated $H^s(\mathbb{T}^2)$ norms are given by $\| \cdot \|_s = \| \Lambda^s \cdot \|_0$ where $\| \cdot \|_0$ is the usual $L^2(\mathbb{T}^2)$ norm. We also make use of the negative Sobolev spaces $H^{-s}(\mathbb{T}^2)$ for $s \geq 0$ defined via duality relative to $L^2(\mathbb{T}^2)$ with the norms reading as

$$\| f \|_{-s} = \sup_{\|\xi\|_s = 1} \langle f, \xi \rangle \tag{8.6}$$

where $\langle \cdot, \cdot \rangle$ is the usual duality pairing so that $\langle f, \xi \rangle = \int_{\mathbb{T}^2} f \xi dx$ when $f \in L^2(\mathbb{T}^2)$. All other norms are denoted as $\| \cdot \|_X$ where X is the associated space i.e. L^∞. We abuse notation and use the same naming convention $H^s(\mathbb{T}^2)$ and associated norm $\| \cdot \|_s$ for periodic, divergence free vector fields with s derivatives in $L^2(\mathbb{T}^2)$.

We have the following proposition adapted from [11]:

Proposition 31 (Well-Posedness and Continuity of the solution map for (8.5))

(i) Fix any $s \geq 0$ and suppose that $\mathbf{q} \in H^s(\mathbb{T}^2)$, $\phi_0 \in H^s(\mathbb{T}^2) \cap L^\infty(\mathbb{T}^2)$ and $f \in L^2_{loc}([0, \infty); H^{s-1}(\mathbb{T}^2))$. Then there exists a unique $\phi = \phi(\mathbf{q}, \phi_0, f)$ such that

$$\begin{aligned} &\phi \in L^2_{loc}([0, \infty); H^{s+1}(\mathbb{T}^2)) \cap L^\infty([0, \infty); H^s(\mathbb{T}^2)), \\ &\frac{\partial \phi}{\partial t} \in L^2_{loc}([0, \infty); H^{s-1}(\mathbb{T}^2)) \end{aligned} \tag{8.7}$$

so that in particular[1]

$$\phi \in C([0, \infty); H^s(\mathbb{T}^2))$$

[1] See e.g. [80, Lemma 3.1.2].

and where ϕ solves (8.5) at least weakly. Additionally ϕ maintains the bounds

$$\frac{d}{dt}\|\phi\|_0^2 + 2\kappa\|\phi\|_1^2 = 2\int f\phi dx, \tag{8.8}$$

$$\sup_{t\in[0,t^*]}\|\phi(t)\|_{L^\infty} \leq \|\phi_0\|_{L^\infty} + \int_0^{t^*}\|f\|_{L^\infty}dt, \text{ for any } t^* > 0. \tag{8.9}$$

When $s > 0$ we have

$$\frac{d}{dt}\|\phi\|_s^2 + \kappa\|\phi\|_{s+1}^2 \leq c\|\phi\|_s^2\|\mathbf{q}\|_s^a + 2\int \Lambda^s f \Lambda^s \phi\, dx \tag{8.10}$$

where the constant $c = c(\kappa, s)$, $a = a(\kappa, s)$ are independent of $\mathbf{q}$.

(ii) Let $\phi^{(j)} = \phi(\mathbf{q}_j, \phi_{0,j}, f_j)$ for $j = 1, 2$ be two solutions of (8.5) corresponding to data $\mathbf{q}_j, \phi_{0,j}, f_j$ satisfying the conditions in part (i). Then, taking $\psi = \phi^{(1)} - \phi^{(2)}$, $\mathbf{p} = \mathbf{q}_1 - \mathbf{q}_2$, we have

$$\frac{d}{dt}\|\psi\|_0^2 + \kappa\|\psi\|_1^2 \leq c\|\mathbf{p}\|_0^2\|\phi^{(1)}\|_{L^\infty}^2 + c\|f_1 - f_2\|_{-1}^2 \tag{8.11}$$

with $c = c(\kappa)$ independent of $\mathbf{q}_1, \mathbf{q}_2$. Furthermore, in the case when $s > 0$ we have

$$\frac{d}{dt}\|\psi\|_s^2 + \kappa\|\psi\|_{s+1}^2 \leq c\|\psi\|_s^2\|\mathbf{q}_1\|_s^a + c\|\mathbf{p}\|_s^2\|\phi^{(2)}\|_{s+1}^2 + c\|f_1 - f_2\|_{s-1}^2, \tag{8.12}$$

where the constants $c = c(\kappa, s)$, $a = a(\kappa, s)$ are again independent of $\mathbf{q}_1, \mathbf{q}_2$.

Proposition 31 immediately yields quantitate bounds on derivatives of $\theta(\mathbf{q})$ in its advecting flow $\mathbf{q}$ which solve (8.3), (8.4). In turn these bounds provide the quantitative foundation for the estimates on DU and D^2U below in Proposition 34 and Corollary 35.

Proposition 32 *Fix any $s > 0$ and $\theta_0 \in L^\infty(\mathbb{T}^2) \cap H^s(\mathbb{T}^2)$. Then the map from $H^s(\mathbb{T}^2)$ to $C([0, \infty); H^s(\mathbb{T}^2))$ that associates to each $\mathbf{q} \in H^s(\mathbb{T}^2)$ the corresponding solution $\theta(\mathbf{q}) := \theta(\cdot; \mathbf{q}, \theta_0)$ of (1.9) is a C^2 function. Denote $\psi^\xi(\mathbf{q})$ and $\psi^{\xi,\tilde{\xi}}(\mathbf{q})$ as the directional derivatives of θ in the directions $\xi, \tilde{\xi} \in H^s(\mathbb{T}^2)$. Then $\psi^\xi(\mathbf{q})$ and $\psi^{\xi,\tilde{\xi}}(\mathbf{q})$ obey (8.3) and (8.4), respectively, with regularity (8.7) in the sense of Proposition 31. Furthermore,*

(i) For any $\mathbf{q}, \xi \in H^s(\mathbb{T}^2)$, $t^* > 0$ we have

$$\sup_{t\leq t^*}\|\psi^\xi(t; \mathbf{q})\|_0^2 + \int_0^{t^*}\|\psi^\xi(t; \mathbf{q})\|_1^2 dt \leq ct^*\|\xi\|_0^2 \tag{8.13}$$

and

$$\sup_{t\le t^*}\|\psi^{\xi}(t;\mathbf{q})\|_0^2+\int_0^{t^*}\|\psi^{\xi}(t;\mathbf{q})\|_1^2dt\le c\|\xi\|_s^2 \tag{8.14}$$

where $c=c(\|\theta_0\|_{L^\infty},\kappa)$ *is independent of* $\mathbf{q}$, ξ *and* t^*. *Furthermore,*

$$\sup_{t\le t^*}\|\psi^{\xi}(t;\mathbf{q})\|_s^2+\int_0^{t^*}\|\psi^{\xi}(t;\mathbf{q})\|_{s+1}^2dt\le c\|\xi\|_s^2\exp(ct^*\|\mathbf{q}\|_s^a) \tag{8.15}$$

where the constant $c=c(s,\|\theta_0\|_s,\kappa)$ *is independent of* $\mathbf{q}$, ξ *and* $t^*>0$*; and* a *is precisely the constant from* (8.10).

(ii) On the other hand, given any $\mathbf{q},\xi,\tilde{\xi}\in H^s(\mathbb{T}^2)$, $t^*>0$

$$\sup_{t\le t^*}\|\psi^{\xi,\tilde{\xi}}(t;\mathbf{q})\|_0^2+\int_0^{t^*}\|\psi^{\xi,\tilde{\xi}}(t;\mathbf{q})\|_1^2dt\le c(\|\xi\|_s^4+\|\tilde{\xi}\|_s^4) \tag{8.16}$$

where $c=c(s,\|\theta_0\|_{L^\infty},\|\theta_0\|_s,\kappa)$ *is independent of* $\mathbf{q},\xi,\tilde{\xi}$ *and* t^*. *Moreover,*

$$\sup_{t\le t^*}\|\psi^{\xi,\tilde{\xi}}(t;\mathbf{q})\|_s^2\le c(\|\xi\|_s^4+\|\tilde{\xi}\|_s^4)\exp(t^*c\|\mathbf{q}\|_s^a) \tag{8.17}$$

for a constant $c=c(s,\|\theta_0\|_{L^\infty},\|\theta_0\|_s,\kappa)$ *independent of* $\mathbf{q},\xi,\tilde{\xi}$ *and* $t^*>0$.

Remark 33 With suitable technical adjustments, Proposition 32 can be extended to the case of Dirichlet boundary condition following the main steps in the proof presented below.

Before turning to the details of the proof let us recall some useful inequalities. Firstly the Sobolev embedding theorem in dimension $d=2$ is given as

$$\|g\|_{L^p}\le c\|g\|_{H^r}\quad \text{for any } r\ge 1-\frac{2}{p},\ \text{with } 2\le p<\infty, \tag{8.18}$$

for any $g:\mathbb{T}^2\to\mathbb{R}$ in $H^r(\mathbb{T}^2)$, where the universal constant c depends only on p and r. We also make use of the Leibniz-Kato-Ponce inequality which takes the general form

$$\|\Lambda^r(fg)\|_{L^m}\le C(\|\Lambda^r f\|_{L^{p_1}}\|g\|_{L^{q_1}}+\|f\|_{L^{p_2}}\|\Lambda^r g\|_{L^{q_2}}) \tag{8.19}$$

valid for any $r\ge 0$, $1<m<\infty$ and $1<p_i,q_i\le\infty$ with $m^{-1}=p_j^{-1}+q_j^{-1}$ for $j=1,2$ and where C is a positive constant depending only on r,m,p_1,q_1,p_2,q_2.

Proof The claimed regularity for $\psi^{\xi},\psi^{\xi,\tilde{\xi}}$ follows from Proposition 31 and the forthcoming formal estimates leading to (8.13)–(8.17) which can be justified in the context of an appropriate regularization scheme. We begin by showing (8.13). From (8.8), namely multiplying (8.3) by ψ^ξ and integrating we have

$$\frac{1}{2}\frac{d}{dt}\|\psi^\xi\|_0^2 + \kappa\|\nabla\psi^\xi\|_0^2 = -\int_{\mathbb{T}^2} \xi\cdot\nabla\theta(\mathbf{q})\psi^\xi dx. \tag{8.20}$$

Integrating by parts and using that ξ is divergence free

$$\left|\int_{\mathbb{T}^2} \xi\cdot\nabla\theta(\mathbf{q})\psi^\xi dx\right| = \left|\int_{\mathbb{T}^2} \xi\cdot\nabla\psi^\xi\theta(\mathbf{q}) dx\right| \le \|\theta(\mathbf{q})\|_{L^\infty}\|\nabla\psi^\xi\|_0\|\xi\|_0 \tag{8.21}$$

Invoking the Maximum principle as in (8.9) we obtain that

$$\|\theta(t;\mathbf{q})\|_{L^\infty} \le \|\theta_0\|_{L^\infty} \quad \text{for any } t\ge 0, \tag{8.22}$$

and hence

$$\frac{d}{dt}\|\psi^\xi\|_0^2 + \kappa\|\nabla\psi^\xi\|_0^2 \le c\|\xi\|_0^2.$$

This immediately implies the first estimate (8.13). For showing (8.14), we estimate (8.21) differently, namely

$$\left|\int_{\mathbb{T}^2} \xi\cdot\nabla\theta(\mathbf{q})\psi^\xi dx\right| \le \|\xi\|_p\|\nabla\theta(\mathbf{q})\|_0\|\psi^\xi\|_q$$

with $1 < p, q < \infty$ such that $\frac{1}{p}+\frac{1}{q}=\frac{1}{2}$. With the Sobolev inequality (8.18) and noting that $q\to 2$ when $p\to\infty$ we can find p and q in this range such that

$$\begin{aligned}\left|\int_{\mathbb{T}^2} \xi\cdot\nabla\theta(\mathbf{q})\psi^\xi dx\right| &\le \|\xi\|_s\|\nabla\theta(\mathbf{q})\|_0\|\nabla\psi^\xi\|_0\\ &\le \frac{\kappa}{2}\|\nabla\psi^\xi\|_0^2 + c\|\xi\|_s^2\|\nabla\theta(\mathbf{q})\|_0^2,\end{aligned}$$

which in combination with (8.20) yields

$$\frac{d}{dt}\|\psi^\xi\|_0^2 + \kappa\|\nabla\psi^\xi\|_0^2 \le c\|\xi\|_s^2\|\nabla\theta(\mathbf{q})\|_0^2. \tag{8.23}$$

Integrating (8.8) for $f=0$ with respect to time, we have

$$\sup_{s\le t^*}\|\theta(\mathbf{q})\|_0^2 + \kappa\int_0^{t^*}\|\nabla\theta(\mathbf{q})\|_0^2 dt \le \|\theta_0\|_0^2 \tag{8.24}$$

Hence from (8.23) and (8.24), it follows that

$$\sup_{t\le t^*}\|\psi^\xi\|_0^2 + \kappa\int_0^{t^*}\|\nabla\psi^\xi\|_0^2 dt \le c\|\xi\|_s^2\int_0^{t^*}\|\nabla\theta(\mathbf{q})\|_0^2 dt \le c\|\theta_0\|_0^2\|\xi\|_s^2,$$

finishing the proof of (8.14).

Turning to $H^s(\mathbb{T}^2)$ estimates we refer to (8.10) which translates to

$$\frac{d}{dt}\|\psi^\xi\|_s^2 + \kappa\|\nabla\psi^\xi\|_s^2 \le c\|\psi^\xi\|_s^2\|\mathbf{q}\|_s^a - 2\int \Lambda^s(\xi\cdot\nabla\theta(\mathbf{q}))\Lambda^s\psi^\xi\,dx. \tag{8.25}$$

Invoking Hölder's inequality and the Leibniz bound (8.19) we estimate

$$\left|\int \Lambda^s(\xi\cdot\nabla\theta(\mathbf{q}))\Lambda^s\psi^\xi\,dx\right| \le c\|\Lambda^s\psi^\xi\|_{L^p}(\|\Lambda^s\xi\|_0\|\Lambda^1\theta(\mathbf{q})\|_{L^q} + \|\xi\|_{L^q}\|\Lambda^{s+1}\theta(\mathbf{q})\|_0) \tag{8.26}$$

valid whenever $1 < p, q < \infty$ and maintains $1 - \frac{1}{p} = \frac{1}{2} + \frac{1}{q}$ i.e. $q = 2p/(p-2)$. Again with the Sobolev inequality (8.18) and noting that $q \to 2$ when $p \to \infty$ we can find p and q in this range such that

$$\begin{aligned}\left|\int \Lambda^s(\xi\cdot\nabla\theta(\mathbf{q}))\Lambda^s\psi^\xi\,dx\right| &\le c\|\Lambda^{s+1}\psi^\xi\|_0\|\Lambda^s\xi\|_0\|\Lambda^{s+1}\theta(\mathbf{q})\|_0\\ &\le \frac{\kappa}{4}\|\Lambda^{s+1}\psi^\xi\|_0^2 + c\|\Lambda^s\xi\|_0^2\|\Lambda^{s+1}\theta(\mathbf{q})\|_0^2.\end{aligned} \tag{8.27}$$

Combining this bound with (8.25) yields the inequality

$$\frac{d}{dt}\|\psi^\xi\|_s^2 + \frac{\kappa}{2}\|\nabla\psi^\xi\|_s^2 \le c\|\psi^\xi\|_s^2\|\mathbf{q}\|_s^a + c\|\xi\|_s^2\|\theta(\mathbf{q})\|_{s+1}^2 \tag{8.28}$$

so that with the Gronwall inequality we obtain

$$\sup_{r\le t^*}\|\psi^\xi\|_s^2 \le \|\xi\|_s^2\exp(ct^*\|\mathbf{q}\|_s^a)\int_0^{t^*}\|\theta(\mathbf{q})\|_{s+1}^2 dt$$

A second application of (8.10), this time with $f = 0$, yields

$$\begin{aligned}\kappa\int_0^{t^*}\|\theta(\mathbf{q})\|_{s+1}^2 dt &\le ct^*\|\mathbf{q}\|_s^a\sup_{t\le t^*}\|\theta(\mathbf{q})\|_s^2 \le ct^*\|\mathbf{q}\|_s^a\exp(ct^*\|\mathbf{q}\|_s^a)\|\theta_0\|_s^2\\ &\le c\exp(ct^*\|\mathbf{q}\|_s^a)\|\theta_0\|_s^2.\end{aligned} \tag{8.29}$$

Combining the previous two bounds we find, for any $t^* \ge 0$,

$$\sup_{t\le t^*}\|\psi^\xi\|_s^2 \le c\exp(ct^*\|\mathbf{q}\|_s^a)\|\xi\|_s^2\|\theta_0\|_s^2. \tag{8.30}$$

Integrating (8.28) in time and invoking (8.29), (8.30)

$$
\begin{aligned}
&\kappa \int_0^{t^*} \|\nabla\psi^\xi\|_s^2 dt \\
&\quad \le ct^* \sup_{t\le t^*} \|\psi^\xi\|_s^2 \|\mathbf{q}\|_s^a + c\|\xi\|_s^2 \int_0^{t^*} \|\theta(\mathbf{q})\|_{s+1}^2 dt \le c\|\theta_0\|_s^2 \|\xi\|_s^2 \exp(ct^*\|\mathbf{q}\|_s^a)
\end{aligned}
$$

and hence we now obtain (8.15).

We next provide estimates for $\psi^{\xi,\tilde{\xi}}$. As before we begin by addressing the L^2 case, namely (8.16). We take the inner product in L^2 of (8.4) with $\psi^{\xi,\tilde{\xi}}$ and integrate to obtain, as in (8.8),

$$
\frac{1}{2}\frac{d}{dt}\|\psi^{\xi,\tilde{\xi}}\|_0^2 + \kappa\|\nabla\psi^{\xi,\tilde{\xi}}\|_0^2 = -\int \tilde{\xi}\cdot\nabla\psi^\xi\psi^{\xi,\tilde{\xi}} - \int \xi\cdot\nabla\psi^{\tilde{\xi}}\psi^{\xi,\tilde{\xi}} := I. \quad (8.31)
$$

Integrating by parts and using Hölder's inequality the right hand side is estimated as

$$
|I| \le (\|\xi\|_{L^p} + \|\tilde{\xi}\|_{L^p})(\|\psi^\xi\|_{L^q} + \|\psi^{\tilde{\xi}}\|_{L^q})\|\nabla\psi^{\xi,\tilde{\xi}}\|_0
$$

for $p^{-1} + q^{-1} = 2^{-1}$. Choosing p, q appropriately and then applying the Sobolev embedding, (8.18), we find

$$
\begin{aligned}
|I| &\le (\|\xi\|_s + \|\tilde{\xi}\|_s)(\|\psi^\xi\|_1 + \|\psi^{\tilde{\xi}}\|_1)\|\nabla\psi^{\xi,\tilde{\xi}}\|_0 \\
&\le c(\|\xi\|_s^2 + \|\tilde{\xi}\|_s^2)(\|\psi^\xi\|_1^2 + \|\psi^{\tilde{\xi}}\|_1^2) + \frac{\kappa}{2}\|\nabla\psi^{\xi,\tilde{\xi}}\|_0^2.
\end{aligned}
$$

Hence, using this bound with (8.31) and then applying (8.13) we infer (8.16).

We turn finally to the $H^s(\mathbb{T}^2)$ estimates for $\psi^{\xi,\tilde{\xi}}$. Here (8.10) becomes

$$
\frac{d}{dt}\|\psi^{\xi,\tilde{\xi}}\|_s^2 + \kappa\|\nabla\psi^{\xi,\tilde{\xi}}\|_s^2 \le c\|\psi^{\xi,\tilde{\xi}}\|_s^2\|\mathbf{q}\|_s^a - 2\int \Lambda^s(\tilde{\xi}\cdot\nabla\psi^\xi + \xi\cdot\nabla\psi^{\tilde{\xi}})\Lambda^s\psi^{\xi,\tilde{\xi}}\,dx. \quad (8.32)
$$

Estimating the last term above in a similar fashion in (8.27) above leads to

$$
\begin{aligned}
&\left|\int \Lambda^s(\tilde{\xi}\cdot\nabla\psi^\xi + \xi\cdot\nabla\psi^{\tilde{\xi}})\Lambda^s\psi^{\xi,\tilde{\xi}}\,dx\right| \\
&\quad \le \frac{\kappa}{2}\|\psi^{\xi,\tilde{\xi}}\|_{s+1}^2 + c(\|\xi\|_s^2 + \|\tilde{\xi}\|_s^2)(\|\psi^\xi\|_{s+1}^2 + \|\psi^{\tilde{\xi}}\|_{s+1}^2). \quad (8.33)
\end{aligned}
$$

Combining the previous two bounds (8.32), (8.33) and then making use of Gronwall inequality and (8.15) we obtain

$$
\begin{aligned}
\sup_{r\le t^*}\|\psi^{\xi,\tilde{\xi}}\|^2 &\le c(\|\xi\|_s^2 + \|\tilde{\xi}\|_s^2)\exp(ct^*\|\mathbf{q}\|_s^a)\int_0^{t^*}(\|\psi^\xi\|_{s+1}^2 + \|\psi^{\tilde{\xi}}\|_{s+1}^2)dt \\
&\le c(\|\xi\|_s^4 + \|\tilde{\xi}\|_s^4)\exp(ct^*\|\mathbf{q}\|_s^a),
\end{aligned}
$$

which establishes the final bound, (8.17), completing the proof. □

8.2 Bounds on the potential U and its derivatives

With these preliminary bounds on (8.5) and hence Proposition 32 in hand we turn to provide estimates for U defined as in (1.11). Recall that we seek to determine the extent to which Assumptions 8 and 10 applies for the certain classes of potential U which arise in this example, namely (1.12) subject to conditions on the observation operator (1.10). Of course, since U is positive, Assumption 10 holds regardless of our assumptions on $\mathcal{O}$.

Regarding the assumptions on $\mathcal{O}$ we consider the following three situations. Fix an observation time window $t^* > 0$. Firstly, we suppose that $\mathcal{O}$ satisfies an inequality of the form

$$|\mathcal{O}(\phi)| \leq c_0 \sup_{t \leq t^*} \|\phi(t)\|_0 \tag{8.34}$$

for $\phi \in C([0, t^*]; L^2(\mathbb{T}^2))$, which is verified in particular for the examples with pointwise in time and spectral in space observations or that of spatial (volumetric) averages. On the other hand, addressing in particular the example of pointwise in both space and time observations, we consider the case when $\mathcal{O}$ satisfies an inequality of the form

$$|\mathcal{O}(\phi)| \leq c_0 \sup_{t \leq t^*} \|\phi(t)\|_{L^\infty} \tag{8.35}$$

for $\phi \in C([0, t^*] \times \mathbb{T}^2)$. Finally, for estimates involving gradients or other derivatives of ϕ we assume that, for some $s > 0$,

$$|\mathcal{O}(\phi)| \leq c_0 \sup_{t \leq t^*} \|\phi(t)\|_{H^s} \tag{8.36}$$

valid for $\phi \in C([0, t^*]; H^s(\mathbb{T}^2))$.

Let us begin with estimates on DU and D^2U in negative Sobolev space which in turn yield the conditions in Assumption 8 on the $\mathbb{H}_\gamma$ spaces, (2.2), defined relative to a covariance operator $\mathcal{C}$ of the Gaussian prior μ_0 in (1.11).

Proposition 34 *Let U be defined as in (1.12) for a fixed $\mathcal{Y} \in \mathbb{R}^m$ and Γ a symmetric strictly positive definite matrix.*

(i) When $\mathcal{O}$ satisfies (8.34), U is twice Fréchet differentiable in $H^{s'}(\mathbb{T}^2)$ for any $s' > 0$. In this case for any $s' \geq 0$

$$\|DU(\mathbf{q})\|_{-s'} \leq M_1 < \infty \tag{8.37}$$

for a constant $M_1 = M_1(s', \kappa, t^*, \theta_0, c_0, \mathcal{Y}, \Gamma)$ *which is independent of* $\mathbf{q}$.[2] *Furthermore, assuming now that* $s' > 0$ *we have*

$$\|D^2U(\mathbf{q})\|_{\mathcal{L}_2(H^{s'}(\mathbb{T}^2))} \leq M_2 < \infty, \tag{8.38}$$

where $\|\cdot\|_{\mathcal{L}_2(H^{s'}(\mathbb{T}^2))}$ *denotes the standard operator norm of a real-valued bilinear operator on* $H^{s'}(\mathbb{T}^2) \times H^{s'}(\mathbb{T}^2)$ *(see* (8.42)*), and* $M_2 = M_2(s', \kappa, \theta_0, c_0, \mathcal{Y}, \Gamma)$ *is a constant independent of* $\mathbf{q}$.

(ii) In the case (8.36) *for* $\mathcal{O}$ *we have once again that* U *is twice Fréchet differentiable in* $H^s(\mathbb{T}^2)$ *for the given value of* $s > 0$ *in* (8.36)*. Here, for any* $s' \geq s$,

$$\|DU(\mathbf{q})\|_{-s'} \leq M \exp(c\|\mathbf{q}\|_{s'}^a) \tag{8.39}$$

and

$$\|D^2U(\mathbf{q})\|_{\mathcal{L}_2(H^{s'}(\mathbb{T}^2))} \leq M \exp(c\|\mathbf{q}\|_{s'}^a) \tag{8.40}$$

where $c = c(s', \kappa, t^*, \theta_0, c_0, \mathcal{Y}, \Gamma)$, $M = M(s', \kappa, t^*, \theta_0, c_0, \mathcal{Y}, \Gamma)$ *are independent of* $\mathbf{q}$ *and* $a > 0$ *is precisely the constant appearing in* (8.10).

(iii) Finally under the assumption that $\mathcal{O}$ *obeys* (8.35)*,* U *is twice Fréchet differentiable in* $H^{s'}(\mathbb{T}^2))$ *for any* $s' > 1$*. In this case, when* $s' > 1$*, we again have the bounds* (8.39)*,* (8.40).

Proof We start with the proof of (8.37). Notice that, referring back to (8.1) and using the condition (8.34), we have

$$|U^\xi(\mathbf{q})| \leq c(1 + \sup_{t \leq t^*} \|\theta(\mathbf{q})\|_0) \cdot \sup_{t \leq t^*} \|\psi^\xi(\mathbf{q})\|_0,$$

for any $\mathbf{q} \in L^2(\mathbb{T}^2)$, $\xi \in H^{s'}(\mathbb{T}^2)$ and $c = c(\Gamma^{-1/2}, \mathcal{Y}, c_0)$. Observe that for any $s' \geq 0$ we have

$$\|DU(\mathbf{q})\|_{-s'} = \sup_{\|\xi\|_{s'}=1} |U^\xi(\mathbf{q})|. \tag{8.41}$$

Thus, invoking the bounds (8.24), (8.13) when $s' = 0$ or (8.14) for the case $s' > 0$, we obtain (8.37).

We turn next to the proof of (8.38). In this case, working from (8.2) and again making use of the condition (8.34),

$$|U^{\xi,\tilde{\xi}}(\mathbf{q})| \leq c \sup_{t \leq t^*} \|\psi^{\tilde{\xi}}(\mathbf{q})\|_0 \cdot \sup_{t \leq t^*} \|\psi^\xi(\mathbf{q})\|_0 + c(1 + \sup_{t \leq t^*} \|\theta(\mathbf{q})\|_0) \cdot \sup_{t \leq t^*} \|\psi^{\xi,\tilde{\xi}}(\mathbf{q})\|_0$$

[2] Note furthermore that M_1 is independent of t^* in the case when $s' > 0$, cf. (8.13), (8.14).

for any $\xi, \tilde{\xi} \in H^{s'}(\mathbb{T}^2)$, where $c = c(\Gamma^{-1/2}, \mathcal{Y}, c_0)$. Here using

$$\|D^2 U(\mathbf{q})\|_{\mathcal{L}_2(H^{s'}(\mathbb{T}^2))} = \sup_{\|\xi\|_{s'}=\|\tilde{\xi}\|_{s'}=1} |U^{\xi,\tilde{\xi}}(\mathbf{q})| \tag{8.42}$$

and the bounds (8.24), (8.14), (8.16), the desired estimate (8.38) now follows.

We next address (8.39), (8.40). Here (8.1) and (8.36) result in

$$|U^{\xi}(\mathbf{q})| \le c(1 + \sup_{t\le t^*} \|\theta(\mathbf{q})\|_s) \cdot \sup_{t\le t^*} \|\psi^{\xi}(\mathbf{q})\|_s \tag{8.43}$$

and similarly, with (8.2),

$$|U^{\xi,\tilde{\xi}}(\mathbf{q})| \le c \sup_{t\le t^*} \|\psi^{\tilde{\xi}}(\mathbf{q})\|_s \cdot \sup_{t\le t^*} \|\psi^{\xi}(\mathbf{q})\|_s + c(1 + \sup_{t\le t^*} \|\theta(\mathbf{q})\|_s) \cdot \sup_{t\le t^*} \|\psi^{\xi,\tilde{\xi}}(\mathbf{q})\|_s \tag{8.44}$$

for any $\xi, \tilde{\xi} \in H^{s'}(\mathbb{T}^2)$, $s' \ge s$. Thus, invoking (8.10) (with $f \equiv 0$), (8.15), (8.17) with (8.41)–(8.44), we obtain (8.39), (8.40) establishing the second item.

Regarding the final item (iii) observe that (8.1), (8.2) and the Sobolev embedding of $H^s(\mathbb{T}^2) \subset L^\infty(\mathbb{T}^2)$ when $s > 1$ we obtain bounds as in (8.43), (8.44) under (8.35) for any $s > 1$. We therefore conclude this final item arguing as in the previous case. The proof is now complete. $\square$

Drawing upon Proposition 34 we now draw certain conclusions on the scope of applicability of Assumption 8 to (1.11). For this purpose suppose $\mathcal{C}$ is a symmetric, positive, trace class operator on $L^2(\mathbb{T}^2)$. Following the notations introduced above in (2.2) we consider the fractional powers of $\mathcal{C}$ and associated spaces $\mathbb{H}_\gamma$ with norm $|\mathbf{q}|_\gamma = \|\mathcal{C}^{-\gamma}\mathbf{q}\|_0$ for $\gamma \ge 0$, so that in particular we have the notation $|\mathbf{q}| = \|\mathbf{q}\|_0$. We have the following corollary:

Corollary 35 *Let $\mathcal{C}$ be a symmetric, positive, trace class operator on $L^2(\mathbb{T}^2)$. Assume that for some $s > 0$, and some $\gamma \in (0, 1/2)$ there is a constant c_1 such that*

$$\|\mathbf{q}\|_s \le c_1 |\mathbf{q}|_\gamma = c_1 \|\mathcal{C}^{-\gamma}\mathbf{q}\|_0 \quad \textit{for all } \mathbf{q} \in \mathbb{H}_\gamma, \tag{8.45}$$

so that $\mathbb{H}_\gamma \subset H^s(\mathbb{T}^2)$.

(i) Under the spectral observation assumption, (8.34), Assumption 8 and 10 hold for U and the given $\mathcal{C}$. Additionally, if for this value of γ, $\mathcal{C}^{1-2\gamma}$ is trace class in the sense of (2.6), so that Assumption 5 holds, then Theorem 26 applies to (1.11).

(ii) Under (8.36), assuming that (8.45) holds for the value of $s > 0$ in (8.36) we have that

$$|DU(\mathbf{q})|_{-\gamma} \le M \exp(c|\mathbf{q}|_\gamma^a) \tag{8.46}$$

and that

$$\|\mathcal{C}^{\gamma} D^2 U(\mathbf{q}) \mathcal{C}^{\gamma}\|_{\mathcal{L}_2(\mathbb{H}_0)} \leq M \exp(c|\mathbf{q}|_{\gamma}^{a}) \tag{8.47}$$

where $\|\cdot\|_{\mathcal{L}(\mathbb{H}_0)}$ *here denotes the standard operator norm of a real-valued bilinear operator on* $\mathbb{H}_0 \times \mathbb{H}_0$*, and again the constants* $c = c(s', \kappa, t^*, \theta_0, c_0, c_1 \mathcal{Y}, \Gamma)$*,* $M = M(s', \kappa, t^*, \theta_0, c_0, c_1, \mathcal{Y}, \Gamma)$ *are independent of* $\mathbf{q}$ *and* $a > 0$ *is as in* (8.10)*.*

(iii) In the case (8.35)*, if* (8.45) *holds for some* $s > 1$ *then we again have the bounds* (8.46)*,* (8.47) *for the corresponding values of* γ*.*

Proof Regarding the first item we proceed to establish the conditions (2.7) and (2.8). Observe that under (8.45)

$$c_1^2 \|D^2 U(\mathbf{q})\|_{\mathcal{L}_2(H^s(\mathbb{T}^2))} \geq \|\mathcal{C}^{\gamma} D^2 U(\mathbf{q}) \mathcal{C}^{\gamma}\|_{\mathcal{L}_2(\mathbb{H}_0)} \tag{8.48}$$

so that with (8.38) we infer (2.7). For (2.8) we demonstrate the stronger condition (2.11). Again, due to (8.45) we have

$$c_1 \|DU(\mathbf{q})\|_{-s} \geq |DU(\mathbf{q})|_{-\gamma} \tag{8.49}$$

so that (2.11) follows from (8.37).

Regarding the second and third items we simply apply (8.48), (8.49) now in combination with (8.39) and (8.40). The proof is complete. □

Remark 36 Let A be the Stokes operator in dimension 2 with periodic boundary conditions. Of course for any given $s > 0$ the condition (8.45) is fulfilled when $\mathcal{C} = (A)^{-\kappa/2}$ for any κ such that $\kappa \geq s/\gamma$. Here note, in regards to Assumption 5, $\mathcal{C} = (A)^{-\kappa/2}$ has the eigenvalues $\lambda_j \approx |j|^{\kappa/2}$. Thus (2.6) entails the additional requirement $\kappa > 2/(1-2\gamma)$.

Note however that the examples considered in [12] involved a covariance $\mathcal{C}$ with exponentially decaying spectrum so that (8.45) applies for any $s \geq 0$ and (2.6) for any $0 \leq \gamma < 1/2$.

Remark 37 [Improved bounds in the time independent case]

We expect that improved, $\mathbf{q}$-independent bounds on (8.3) and (8.4) can be achieved through more sophisticated parabolic regularity techniques. In turn this could improve bounds obtainable for DU and D^2U in the case of point observations (8.35). Whatever the mechanism, we note that the numerical results in [12] suggest good mixing occurs for the Hamiltonian Monte Carlo algorithm in this case of point observations notwithstanding the fact that our current results do not cover this situation.

In this connection it is notable that a global bound on DU and D^2U and hence the conditions for 26 can be achieved for point observations in the time-stationary analogue of (8.5) thanks to [1]. Let

$$\mathbf{q} \cdot \nabla\theta = \kappa \Delta\theta + f \tag{8.50}$$

on $\mathbb{T}^2$ for a given fixed $f : \mathbb{T}^2 \to \mathbb{R}$, $\kappa > 0$. We can consider, similarly to above, the statistical inversion problem of recovering a divergence free $\mathbf{q}$ from the sparse observation of the resulting solution $\theta : \mathbb{T}^2 \to \mathbb{R}$. In this case, following the Bayesian approach we again obtain a posterior measure of the form (1.11) with U given analogously to (1.12) in the case of Gaussian observation noise.

As previously the task of estimating DU and D^2U entails suitable estimates for

$$\mathbf{q} \cdot \nabla \psi^{\xi} = \kappa \Delta \psi^{\xi} - \xi \cdot \nabla \theta(\mathbf{q}),$$

and

$$\mathbf{q} \cdot \nabla \psi^{\xi,\tilde{\xi}} = \kappa \Delta \psi^{\xi,\tilde{\xi}} - \tilde{\xi} \cdot \nabla \psi^{\xi} - \xi \cdot \nabla \psi^{\tilde{\xi}}.$$

over suitable directions $\xi, \tilde{\xi}$.

Suppose that ϕ obeys

$$\mathbf{q} \cdot \nabla \phi = \kappa \Delta \phi + g \tag{8.51}$$

for some $\mathbf{q} : \mathbb{T}^2 \to \mathbb{R}^2$, divergence free and $g : \mathbb{T}^2 \to \mathbb{R}$. According to [1, Lemma 1.3] we have that[3]

$$\|\phi\|_{L^\infty} \le c\|g\|_{L^p} \tag{8.52}$$

for any $p > 1$ where crucially the constant $c = c(p, \kappa)$ is independent of $\mathbf{q}$. Applying (8.52) and carrying out other standard manipulations we have that

$$\|\theta(\mathbf{q})\|_{L^\infty}^2 + \|\nabla \theta(\mathbf{q})\|_0^2 \le c\|f\|_0^2 \tag{8.53}$$

for $c = c(\kappa)$ independent of $\mathbf{q}$. As such a second application of (8.52), Sobolev embedding, (8.18), and (8.53) yields

$$\|\psi^{\xi}\|_{L^\infty} \le c\|\xi\|_s\|f\|_0 \tag{8.54}$$

for any $s > 0$ where the constant $c = c(s, \kappa)$ is again independent of $\mathbf{q}$. Moreover, using that $\mathbf{q}$ is divergence free and (8.53)

$$\|\nabla \psi^{\xi}\|_0 \le c\|\xi\|_0\|f\|_0 \tag{8.55}$$

with $c = c(s, \kappa)$ independent of $\mathbf{q}$. Finally (8.52) followed by

$$\|\psi^{\xi,\tilde{\xi}}\|_{L^\infty} \le c(\|\xi\|_s^2 + \|\tilde{\xi}\|_s^2). \tag{8.56}$$

[3] The result [1] is stated for (8.50) supplemented with Dirichlet boundary conditions but pursuing the proof it is clear that this bound also applies in the spatially periodic case.

for any $s > 0$ where $c = c(s, \kappa)$ does not depend on $\mathbf{q}$. Thus, arguing as in Proposition 32 but making use of (8.54), (8.56) we can therefore conclude that whenever

$$|\mathcal{O}(\phi)| \leq c_0 \|\phi\|_{L^\infty},$$

bounds as in (8.37), (8.38) must hold.

9 Outlook

This work provides an illustration of the power and efficacy of the weak Harris theorem as a tool for the analysis of mixing in infinite-dimensional MCMC methods. Specifically our work addresses a Hilbert space version from [6] of the Hamiltonian Monte Carlo method. Notwithstanding recent progress in this setting of infinite dimensional MCMC algorithms, the understanding of mixing rates and the relatedly optimal choice of algorithmic parameters remains in its infancy. Let us therefore point out a number of interesting questions remaining to be studied which we plan to address in future work.

One immediate avenue concerns the analysis of numerically discretized versions of the HMC algorithm (2.16) which must be used in practice. Here the Metropolization step, which is used to correct for the bias introduced by the discretization of (1.2), must be accounted for. In a similar vein it would be useful to have error bounds between the adjusted and unadjusted versions of the algorithm.

It is also worth noting that there are a number of variations on the infinite dimensional HMC algorithm from [6] now available in the literature whose mixing properties are poorly understood, particularly as we regard these different algorithms in comparative perspective. For example we note the Second-Order Langevin Hamiltonian (SOLHMC) methods in [73] and the Riemannian (geometric) HMC approach developed in [3,9].

Although the above analysis is a nontrivial first step towards a better understanding of (1.3) one may nevertheless view the time step condition (6.1) as restricting the scope of our analysis to a perturbation of the linear Gaussian case; cf. Remark 16. It is notable that similar small time step condition also appears in all the other recent studies of the HMC algorithm that we are aware of [13,14,32,58]. We conjecture that for many problems of interest this restriction on T may be far from optimal from the point of view of mixing rates. Indeed this bound on T (6.1) turns on our treatment of the Lyapunov structure in Proposition 20 and on the nudging scheme in Proposition 18 which could presumably be improved with a more delicate treatment of the Hamiltonian dynamics (1.2). As a starting point it would be of great interest to find some simple settings in finite dimensions where this could be carried out.

As already noted above in the introduction, a primary motivation for considering infinite dimensional MCMC methods concerns the Bayesian approach to PDE inverse problems. While several large scale numerical studies have been carried out for some specific problems a more systematic gallery of examples on which the performance of algorithms have been experimentally tested would be desirable. Here our results presented in Sect. 8 show that analysis of conditions on the potential U in (1.2) as arising from the Bayesian approach to PDE inverse problems can be quite involved. Indeed, in the case of the advection-diffusion problem we consider here, it is not clear that we can obtain a global Hessian bound on U for interesting classes of observations, such as space-time point observations. Thus it would be useful to develop an analysis that only requires that U is locally Lipschitz. More broadly, further examples of PDE inverse problems as found in e.g. [78] should be analytically studied in this context to obtain a broader sense of the variety of relevant conditions on U.

Acknowledgements This work would not have been able to be completed in its current form without the extensive feedback and input of Nawaf Bou-Rabee during two visits to Tulane university in the spring and summer of 2019. We would also like to acknowledge D. Albritton, G. Didier, J. Foldes, E. T. Holtzchetti, J. Krometis, S. McKinley, J. Mattingly, S. Punshon-Smith, G. Richards, V. Sverak for further fruitful discussions surrounding on this work. Our efforts here were partially supported under the grants DMS-1313272 (NEGH), DMS-1816551 (NEGH), the Simons Foundation travel grant under 515990 (NEGH), and DMS-2009859 (CFM).

Data Availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Appendix: Consequences for convergence of observables

Let P be a Markov kernel on a Polish space $\mathbb{V}$ and take $\{Q_n(\mathbf{q}_0)\}_{n\geq 1, \mathbf{q}_0 \in \mathbb{V}}$ to be the Markov process associated with P starting from $\mathbf{q}_0 \in \mathbb{V}$. Suppose that μ_* is an invariant measure for P. In addition to quantifying various abstract notions of distance, i.e. the Wasserstein metric, between the measures μP^n and μ_*, we are typically interested in estimating

$$\left| P^n \Phi(\mathbf{q}_0) - \int \Phi(\mathbf{q}')\mu_*(d\mathbf{q}') \right| \tag{A.1}$$

and also

$$\left| \frac{1}{n} \sum_{k=1}^{n} \Phi(Q_k(\mathbf{q}_0)) - \int \Phi(\mathbf{q}')\mu_*(d\mathbf{q}') \right| \tag{A.2}$$

for concrete observables $\Phi : \mathbb{V} \to \mathbb{R}$ and starting from any initial $\mathbf{q}_0 \in \mathbb{V}$.

Typically, contraction bounds as in (6.2) and (7.13) which we demonstrated above can be used to establish estimates for quantities like (A.1), (A.2). Indeed, if the $\tilde{\rho}$ appearing in the bounds (6.2) and (7.13) was actually a metric then the Kantorovich-Wasserstein duality would immediately imply bounds for (A.1). Moreover, a number of results in the literature, e.g. [41,52,53,55,56,77], yield a law of large numbers, central limit theorems type convergence results from Wasserstein contraction bounds as desired in (A.2). This appendix proceeds to show that useful bounds for (A.1), (A.2) can still be achieved in our setting without presuming that the underlying distance $\tilde{\rho}$ is a metric. Notwithstanding the significant literature on such convergence results we expect our approach here to be of novel interest even when the underlying distance is a metric.

In order to proceed, let us recall a few basic definitions:

Definition 38 We say that $\ell : \mathbb{V}\times\mathbb{V} \to \mathbb{R}^+$ is a *distance-like function* if ℓ is symmetric, lower-semicontinuous and it holds that $\ell(\mathbf{q}, \tilde{\mathbf{q}}) = 0$ if and only if $\mathbf{q} = \tilde{\mathbf{q}}$. We define $\mathcal{W}_\ell : \mathrm{Pr}(\mathbb{V}) \times \mathrm{Pr}(\mathbb{V}) \to \mathbb{R}^+ \cup \{+\infty\}$ to be the following Wasserstein-like extension of ℓ to $\mathrm{Pr}(\mathbb{V}) \times \mathrm{Pr}(\mathbb{V})$:

$$\mathcal{W}_\ell(\nu_1, \nu_2) = \inf_{\Gamma\in\mathfrak{C}(\nu_1,\nu_2)} \int_{\mathbb{V}\times\mathbb{V}} \ell(\mathbf{q}, \tilde{\mathbf{q}})\Gamma(d\mathbf{q}, d\tilde{\mathbf{q}}),$$

where $\mathfrak{C}(\nu_1, \nu_2)$ is the set of all couplings of $\nu_1, \nu_2 \in Pr(\mathbb{V})$.[4]

Relative to a given distance-like function ℓ we define ℓ-Lipschitz in the obvious way as:

Definition 39 Given a distance-like function $\ell : \mathbb{V}\times\mathbb{V} \to \mathbb{R}^+$, we say that $\Phi : \mathbb{V} \to \mathbb{R}$ is ℓ-Lipschitz with Lipschitz constant $L_\Phi > 0$ if

$$|\Phi(\mathbf{q}) - \Phi(\mathbf{q}')| \le L_\Phi \ell(\mathbf{q}, \mathbf{q}')$$

for any $\mathbf{q}, \mathbf{q}' \in \mathbb{V}$. We denote the set of ℓ-Lipschitz functions as Lip_ℓ.

In order to verify that an observable Φ is ℓ-Lipschitz for the class of distance like functions employed above, see Proposition 46 below.

Results for (A.1) can be drawn by using the following proposition.

Proposition 40 *Let $\ell : \mathbb{V} \times \mathbb{V} \to \mathbb{R}^+$ be a distance-like function as in Definition 38. Then, for every $\nu_1, \nu_2 \in Pr(\mathbb{V})$ and every ℓ-Lipschitz function $\Phi : \mathbb{V} \to \mathbb{R}$,*

$$\mathcal{W}_\ell(\nu_1, \nu_2) \ge \frac{1}{L_\Phi} \left| \int \Phi(\mathbf{q})\nu_1(d\mathbf{q}) - \int \Phi(\mathbf{q}')\nu_2(d\mathbf{q}') \right|, \tag{A.3}$$

where L_Φ is the Lipschitz constant associated with Φ. In particular, for any Markov kernel P,

$$\left| P^n\Phi(\mathbf{q}_0) - \int \Phi(\mathbf{q})\nu(d\mathbf{q}) \right| \le L_\Phi \mathcal{W}_\ell(P^n(\mathbf{q}_0, \cdot), \nu), \tag{A.4}$$

valid for any measure $\nu \in Pr(\mathbb{V})$, $\mathbf{q}_0 \in \mathbb{V}$ and ℓ-Lipschitz function Φ.

Proof Fix $\nu_1, \nu_2 \in \mathrm{Pr}(\mathbb{V})$ and let $\Gamma \in \mathfrak{C}(\nu_1, \nu_2)$. Note that

$$\begin{aligned} \left| \int \Phi(\mathbf{q})\nu_1(d\mathbf{q}) - \int \Phi(\mathbf{q}')\nu_2(d\mathbf{q}') \right| &\le \int \left|\Phi(\mathbf{q}) - \Phi(\mathbf{q}')\right| \Gamma(d\mathbf{q}, d\mathbf{q}') \\ &\le L_\Phi \int l(\mathbf{q}, \mathbf{q}')\Gamma(d\mathbf{q}, d\mathbf{q}'). \end{aligned} \tag{A.5}$$

[4] The mapping $\mathcal{W}_\ell$ is also called the ‘optimal transport cost functional’ in the optimal transport literature; see, e.g., [83].

Inequality (A.3) then follows by taking the infimum in (A.5) over all $\Gamma \in \mathfrak{C}(\nu_1, \nu_2)$. □

We next present a first version of the strong law of large numbers (SLLN) relevant for certain classes of mixing Markov processes. Note that this first result does not require a spectral gap condition but see Proposition 43 below where we additionally establish criteria for a central limit theorem under the stronger assumption of a spectral gap.

Proposition 41 *Suppose that P is a Markov kernel with a unique invariant measure μ_*. We denote the associated Markov process as $\{Q_k(\mathbf{q}_0)\}_{k\geq 0, q_0 \in \mathbb{V}}$. Let ℓ be a distance-like function and introduce the notation*

$$G(\mathbf{q}_0) := \sum_{k=0}^{\infty} \mathcal{W}_\ell(P^k(\mathbf{q}_0, \cdot), \mu_*). \tag{A.6}$$

Then, for any $\mathbf{q}_0 \in \mathbb{V}$ such that

$$G(\mathbf{q}_0) + \sup_{n\geq 1} \mathbb{E}[G(Q_n(\mathbf{q}_0))^2] < \infty \tag{A.7}$$

and such that, for some $\bar{\mathbf{q}} \in \mathbb{V}$,

$$\sup_{n\geq 1} \mathbb{E}[\ell(Q_n(\mathbf{q}_0), \bar{\mathbf{q}})^2] < \infty, \tag{A.8}$$

we have that, for each $\Phi \in Lip_\ell$,

$$\lim_{n\to\infty} \left| \frac{1}{n} \sum_{k=1}^{n} \Phi(Q^k(\mathbf{q}_0)) - \int \Phi(\mathbf{q}')\mu_*(d\mathbf{q}') \right| = 0, \tag{A.9}$$

almost surely.[5]

Remark 42 The scope of applicability of Propositions 40 and 41 reaches beyond Proposition 43 below which is more specialized to our setting. See, for example, the sub-geometric rates of convergence in the Wasserstein distance given in [19,30,31].

Proof Take $\{\mathcal{F}_n\}_{n\geq 1}$ to be the filtration associated with the Markov process $\{Q_k(\mathbf{q}_0)\}_{k\geq 0, q_0\in\mathbb{V}}$. Given any $\Phi \in \mathrm{Lip}_\ell$, we define

$$M_n^\Phi := \sum_{k=0}^{\infty} \big(\mathbb{E}(\bar{\Phi}(Q_k(\mathbf{q}_0)|\mathcal{F}_n) - \mathbb{E}(\bar{\Phi}(Q_k(\mathbf{q}_0))\big) \tag{A.10}$$

[5] Note that under (A.7) every $\mathrm{Lip}_\ell \subset L^1(\mu_*)$ so that $\int \Phi(\mathbf{q}')\mu_*(d\mathbf{q}')$ is a well defined, finite quantity.

where

$$\bar{\Phi}(\mathbf{q}_0) := \Phi(\mathbf{q}_0) - \int \Phi(\tilde{\mathbf{q}})\mu_*(d\tilde{\mathbf{q}}) \tag{A.11}$$

Invoking the Markov property,

$$M_n^\Phi = \sum_{k=0}^{n} \bar{\Phi}(Q_k(\mathbf{q}_0)) + \sum_{k=0}^{\infty} \left(P^{k+1}\bar{\Phi}(Q_n(\mathbf{q}_0)) - P^k\bar{\Phi}(\mathbf{q}_0) \right), \tag{A.12}$$

so that, rearranging, we have

$$\begin{aligned} \frac{1}{n}\sum_{k=0}^{n} \Phi(Q_k(\mathbf{q}_0)) - \int \Phi(\tilde{\mathbf{q}})\mu_*(d\tilde{\mathbf{q}}) &= \frac{1}{n}\sum_{k=0}^{\infty} \left(P^k\bar{\Phi}(\mathbf{q}_0) - P^{k+1}\bar{\Phi}(Q_n(\mathbf{q}_0)) \right) + \frac{M_n^\Phi}{n} \\ &:= T_1^{(n)} + T_2^{(n)}. \end{aligned} \tag{A.13}$$

Let us show that, for each of the terms $T_j^{(n)}$, $\lim_{n\to\infty} T_j^{(n)} = 0$ a.s. in order to infer the desired conclusion.

Start with $T_1^{(n)}$. Here note that, with (A.4),

$$|T_1^{(n)}| \leq L_\Phi \frac{G(\mathbf{q}_0) + G(Q_n(\mathbf{q}_0))}{n}, \tag{A.14}$$

where L_Φ is the Lipschitz constant associated with Φ. Form the sets $A_n := \{|T_1^{(n)}| \geq n^{-1/4}\}$. With (A.14) and the Markov inequality we find

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}(A_n) \leq L_\Phi \sum_{n=1}^{\infty} \frac{\mathbb{E}\,(G(\mathbf{q}_0) + G(Q_n(\mathbf{q}_0))^2}{n^{3/2}} &\leq 2L_\Phi(G(\mathbf{q}_0)^2 \\ + \sup_{n\geq 1} \mathbb{E}G(Q_n(\mathbf{q}_0))^2) \sum_{n=1}^{\infty} \frac{1}{n^{3/2}}. \end{aligned}$$

Hence, invoking the Borel-Cantelli lemma and the condition (A.7), we infer that $\mathbb{P}(A_n \text{ infinitely often}) = 0$ which amounts to the desired convergence for $T_1^{(n)}$.

Regarding the second term $T_2^{(n)}$, we claim that $\{M_n^\Phi\}_{n\geq 0}$ is a mean zero, square integrable martingale. From the definition of $\{M_n^\Phi\}_{n\in\mathbb{N}}$ in (A.13) it follows immediately that $M_0 = 0$. Now in view of (A.12), notice that for any $n \geq 1$ the increments $M_n^\Phi - M_{n-1}^\Phi$ have the form

$$M_n^\Phi - M_{n-1}^\Phi = \bar{\Phi}(Q_n(\mathbf{q}_0)) + \sum_{k=0}^{\infty} \left(P^{k+1}\bar{\Phi}(Q_n(\mathbf{q}_0)) - P^{k+1}\bar{\Phi}(Q_{n-1}(\mathbf{q}_0)) \right). \tag{A.15}$$

Thus, for any $n \geq 1$, using that $\bar{\Phi} \in \mathrm{Lip}_\ell$ and recalling the definition of G we have

$$\mathbb{E}(M_n^\Phi - M_{n-1}^\Phi)^2 \leq 4\bar{\Phi}(\bar{\mathbf{q}})^2 + 4L_\Phi^2 \left[\mathbb{E}\ell(\bar{\mathbf{q}}, Q_n(\mathbf{q}_0))^2 + \mathbb{E}G(Q_{n-1}(\mathbf{q}_0))^2 + \mathbb{E}G(Q_n(\mathbf{q}_0))^2\right] \tag{A.16}$$

where $\bar{\mathbf{q}} \in \mathbb{V}$ is selected as in (A.8). With (A.7), (A.8) and noticing that

$$\mathbb{E}M_n^2 = \mathbb{E}\left(\sum_{k=1}^n (M_k - M_{k-1})\right)^2 \leq c(n) \sum_{k=1}^n \mathbb{E}\left(M_k - M_{k-1}\right)^2,$$

we conclude that $\{M_n^\Phi\}_{n\in\mathbb{N}}$ is square integrable. To show that $\{M_n^\Phi\}_{n\in\mathbb{N}}$ is a martingale observe that for any $n \geq 0$, using standard properties of conditional expectations,

$$\mathbb{E}(M_{n+1}^\Phi | \mathcal{F}_n) = \sum_{k=0}^\infty \left(\mathbb{E}(\mathbb{E}(\bar{\Phi}(Q_k(\mathbf{q}_0)|\mathcal{F}_{n+1})|\mathcal{F}_n) - \mathbb{E}(\bar{\Phi}(Q_k(\mathbf{q}_0))\right) = M_n^\Phi. \tag{A.17}$$

With this in hand we recall a martingale convergence theorem from [22] (see also [55, Appendix A.12]) which can be stated as follows: Let $\{M_n\}_{n\in\mathbb{N}}$ be a square integrable, mean zero martingale. If

$$\sum_{k=1}^\infty \frac{\mathbb{E}(M_k - M_{k-1})^2}{k^2} < \infty \tag{A.18}$$

then

$$\lim_{n\to\infty} \frac{M_n}{n} = 0 \quad \text{almost surely.}$$

In view of the bound (A.16) and again invoking the standing conditions (A.7), (A.8) we find that the condition (A.18) is satisfied for $\{M_n^\Phi\}_{n\in\mathbb{N}}$ and hence we infer that $\lim_{n\to\infty} T_2^{(n)} = 0$ almost surely. The proof is now complete. □

In order to obtain rates of convergence for (A.2) we can furthermore establish a central limit theorem (CLT) result by now directly imposing a 'spectral gap' condition. For this stronger convergence result we again rely on the decomposition (A.10), (A.13) now in conjunction with a Martingale central limit result from [53] which we recall as Theorem 45 below.

Proposition 43 *Let P be a Markov kernel on a complete metric space $(\mathbb{V}, \rho)$. Take $\{Q_n(\mathbf{q}_0)\}_{n\geq 0, q_0\in\mathbb{V}}$ to be the associated Markov process. Let $V : \mathbb{V} \to \mathbb{R}^+$ be a function satisfying the following Lyapunov type assumption:*

$$\mathbb{E}[V(Q_n(\mathbf{q}_0))^2] \leq \kappa^n V(\mathbf{q}_0)^2 + K \tag{A.19}$$

for some constants $\kappa \in (0, 1)$, $K > 0$ *independent of* $n \geq 0$. *Consider the distance-like functions*

$$\ell_p(\mathbf{q}, \tilde{\mathbf{q}}) = \sqrt{[1 \wedge \rho(\mathbf{q}, \tilde{\mathbf{q}})](1 + V(\mathbf{q})^p + V(\tilde{\mathbf{q}})^p)} \tag{A.20}$$

for $p \geq 1$. *We assume that for* $p = 1, 2$ *the contraction condition*

$$\mathcal{W}_{\ell_p}(\nu_1 P^n, \nu_2 P^n) \leq c_1 e^{-c_2 n} \mathcal{W}_{\ell_p}(\nu_1, \nu_2) \quad \textit{for any } \nu_1, \nu_2 \in Pr(\mathbb{V}), \tag{A.21}$$

is maintained, where c_1, c_2 *are constants independent of* n *but which may depend on* p.

For $\Phi \in Lip_{\ell_1}$, *let*

$$X_n(\Phi) := \frac{1}{n} \sum_{k=1}^{n} \Phi(Q_k(\mathbf{q}_0)) - \int \Phi(\mathbf{q}') \mu_*(d\mathbf{q}'),$$

where μ_* *is the unique invariant measure for* P; *cf. Remark 44. Then, under these circumstances, for any* $\Phi \in Lip_{\ell_1}$,

$$X_n(\Phi) \to 0 \quad \textit{as } n \to \infty \tag{A.22}$$

almost surely and moreover

$$\sqrt{n} X_n(\Phi) \Rightarrow N(0, \sigma^2(\Phi)) \quad \textit{as } n \to \infty, \tag{A.23}$$

i.e. $\sqrt{n} X_n(\Phi)$ *converges weakly to a real-valued gaussian random variable with mean zero and covariance* $\sigma^2(\Phi)$, *where* $\sigma^2(\Phi)$ *is specified explicitly as* (A.36) *below.*

Remark 44 The condition (A.21) ensures the existence and uniqueness of the invariant measure μ_* as observed in [43]. Moreover, (A.19) implies the following moment bound for μ_*

$$\int V(\mathbf{q}')^2 \mu_*(d\mathbf{q}') \leq K < \infty. \tag{A.24}$$

As such, using that $\Phi \in \mathrm{Lip}_{\ell_1}$ and (A.20), we have

$$\int |\Phi(\mathbf{q}')| \mu_*(d\mathbf{q}') \leq |\Phi(\bar{\mathbf{q}})| + L_\Phi \left(1 + \sqrt{V(\bar{\mathbf{q}})} + \int \sqrt{V(\mathbf{q}')} \mu_*(d\mathbf{q}')\right)$$

for any $\bar{\mathbf{q}} \in \mathbb{V}$ so that with (A.24) we are guaranteed that $\int |\Phi(\mathbf{q}')| \mu_*(d\mathbf{q}') < \infty$.

Our proof relies on the following abstract result from [53, Theorem 5.1] which we reformulate here for clarity and the convenience of the reader.

Theorem 45 *Let* $\{M_n\}_{n \geq 0}$ *be a square integrable, mean zero martingale, relative to a filtration* $\{\mathcal{F}_n\}_{n \geq 0}$. *Assume that:*

(i) we have the uniform bound

$$\sup_{n\geq 0}\mathbb{E}(M_{n+1}-M_n)^2<\infty. \tag{A.25}$$

(ii) For every $\epsilon>0$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n-1}\mathbb{E}[(M_{m+1}-M_m)^2\mathbb{1}_{|M_{m+1}-M_m|\geq\epsilon\sqrt{n}}]=0. \tag{A.26}$$

(iii) For every $\epsilon>0$,

$$\lim_{k\to\infty}\limsup_{n\to\infty}\frac{1}{nk}\sum_{m=1}^{n}\sum_{j=(m-1)k}^{mk-1}\mathbb{E}\left[(1+(M_{j+1}-M_j)^2)\mathbb{1}_{|M_j-M_{(m-1)k}|\geq\epsilon\sqrt{nk}}\right]=0. \tag{A.27}$$

(iv) There exists a constant $\sigma^2\geq 0$ such that

$$\lim_{k\to\infty}\limsup_{n\to\infty}\frac{1}{n}\sum_{m=1}^{n}\mathbb{E}\left|\frac{1}{k}\sum_{j=(m-1)k}^{mk-1}\mathbb{E}((M_{j+1}-M_j)^2|\mathcal{F}_{(m-1)k})-\sigma^2\right|=0. \tag{A.28}$$

Then, under these four conditions,

$$\frac{M_n}{\sqrt{n}}\Rightarrow N(0,\sigma^2)\quad \textit{as } n\to\infty,$$

(that is in distribution) where σ^2 is the constant appearing in (A.28).

With this result in hand we turn to the proof of Proposition 43.

Proof *(Proof of Proposition 43)* To prove (A.22) we simply show that (A.21), (A.19) imply (A.7), (A.8), with $\ell=\ell_1$, so that we can directly apply Proposition 41. Observe that, for any $\bar{\mathbf{q}}\in\mathbb{V}$ we have

$$\sum_{k=0}^{\infty}\mathcal{W}_{\ell_1}(P^k(\bar{\mathbf{q}},\cdot),\mu_*)\leq\mathcal{W}_{\ell_1}(\delta_{\bar{\mathbf{q}}},\mu_*)\sum_{k=0}^{\infty}c_1e^{-c_2k}\leq c\left(1+\sqrt{V(\bar{\mathbf{q}})}+\int\sqrt{V(\mathbf{q}')}\mu_*(d\mathbf{q}')\right).$$

Noting that, with (A.24), we have $\int\sqrt{V(\mathbf{q}')}\mu_*(d\mathbf{q}')<\infty$ and with (A.19) we infer $\sup_{k\geq 0}\mathbb{E}V(Q_k(\mathbf{q}_0))<\infty$ so that (A.7) holds. Regarding (A.8) we have, for any $\mathbf{q}_0,\bar{\mathbf{q}}\in\mathbb{V}$

$$\sup_{n\geq 1}\mathbb{E}\ell_1(Q_n(\mathbf{q}_0),\bar{\mathbf{q}}) \leq c\left(1+\sup_{n\geq 1}\mathbb{E}\sqrt{V(Q_n(\mathbf{q}_0))}+\sqrt{V(\bar{\mathbf{q}})}\right) \leq c\left(1+\sqrt{V(\mathbf{q}_0)}+\sqrt{V(\bar{\mathbf{q}})}\right),$$

where the last inequality again follows from (A.19).

Let us next turn to establish the convergence to normality, (A.23). Fix $\Phi \in \mathrm{Lip}_{\ell_1}$. Here, working from the identity (A.13), we have

$$\sqrt{n}X_n(\Phi) = \frac{1}{\sqrt{n}}\sum_{k=0}^{\infty}\left(P^k\bar{\Phi}(\mathbf{q}_0) - P^{k+1}\bar{\Phi}(Q_n(\mathbf{q}_0))\right) + \frac{M_n^{\Phi}}{\sqrt{n}} := \bar{T}_1^{(n)} + \bar{T}_2^{(n)}, \tag{A.29}$$

where M_n^{Φ} is the martingale defined as in (A.12). We would like to show that $\lim_{n\to\infty}\bar{T}_1^{(n)} = 0$ in probability and that $\bar{T}_2^{(n)}$ converges in distribution to a normal random variable in order to conclude (A.23) from the 'converging together lemma'; cf. [33].

Regarding the first term $\bar{T}_1^{(n)}$, with (A.4) and (A.21), it follows

$$\begin{aligned}|\bar{T}_1^{(n)}| &\leq \frac{L_{\Phi}}{\sqrt{n}}\sum_{k=0}^{\infty}(\mathcal{W}_{\ell_1}(P^k(\mathbf{q}_0,\cdot),\mu^*) + \mathcal{W}_{\ell_1}(P^{k+1}(Q_n(\mathbf{q}_0),\cdot),\mu^*))\\ &\leq \frac{c}{\sqrt{n}}(\mathcal{W}_{\ell_1}(\delta_{\mathbf{q}_0},\mu^*) + \mathcal{W}_{\ell_1}(\delta_{Q_n(\mathbf{q}_0)},\mu^*)) \leq \frac{c\left(1+\sqrt{V(\mathbf{q}_0)}+\sqrt{V(Q_n(\mathbf{q}_0))}\right)}{\sqrt{n}}\end{aligned}$$

where we used that ℓ_1 has the form (A.20) for the final bound. With this estimate and our assumption (A.19) we find that $\lim_{n\to\infty}\mathbb{E}|T_1^{(n)}| = 0$ so that $T_1^{(n)}$ decays to zero in probability as desired.

We address the second term $\bar{T}_n^{(2)}$ by verifying the conditions of Theorem 45. As in (A.16), (A.17), it is clear that $\{M_n^{\Phi}\}_{n\geq 0}$ is a mean zero square integrable martingale. We therefore proceed to establish each of the bounds (A.25)–(A.28) for $\{M_n^{\Phi}\}_{n\geq 0}$ in turn.

Start with (A.25). Working from the identity (A.15), we observe that, for any $m \geq 0$,

$$\begin{aligned}(M_{m+1}^{\Phi} - M_m^{\Phi})^4 &\leq c\bar{\Phi}(Q_{m+1}(\mathbf{q}_0))^4 + c\left(\sum_{k=0}^{\infty}P^{k+1}\bar{\Phi}(Q_{m+1}(\mathbf{q}_0)) - P^{k+1}\bar{\Phi}(Q_m(\mathbf{q}_0))\right)^4\\ &\leq c(\ell_1(Q_{m+1}(\mathbf{q}_0),0)^4 + V(Q_{m+1}(\mathbf{q}_0))^2 + V(Q_m(\mathbf{q}_0))^2 + 1)\\ &\leq c(V(Q_{m+1}(\mathbf{q}_0))^2 + V(Q_m(\mathbf{q}_0))^2 + 1)\end{aligned}$$

where we have used (A.4) and (A.21). Therefore, invoking (A.19), we have now shown

$$\sup_{m\geq 0}\mathbb{E}(M_{m+1}^{\Phi} - M_m^{\Phi})^4 < \infty \tag{A.30}$$

so that, in particular, (A.25) holds. Furthermore since, for any $\epsilon > 0$ and any $0 \leq m \leq n$

$$\begin{aligned}\mathbb{E}[(M^{\Phi}_{m+1} - M^{\Phi}_{m})^2 \mathbb{1}_{|M^{\Phi}_{m+1} - M^{\Phi}_{m}| \geq \epsilon\sqrt{n}}] &\leq \left(\mathbb{E}(M^{\Phi}_{m+1} - M^{\Phi}_{m})^4\right)^{1/2} \mathbb{P}(|M^{\Phi}_{m+1} - M^{\Phi}_{m}| \geq \epsilon\sqrt{n})^{1/2}\\ &\leq \frac{1}{\epsilon^2 n}\mathbb{E}(M^{\Phi}_{m+1} - M^{\Phi}_{m})^4\end{aligned}$$

we infer (A.26).

Regarding (A.27) we proceed in a similar fashion. For $(m-1)k \leq j \leq mk - 1$ and any $m, n, k \geq 1$ we have

$$\begin{aligned}&\mathbb{E}[(1+(M^{\Phi}_{j+1} - M^{\Phi}_{j})^2) \mathbb{1}_{|M^{\Phi}_{j} - M^{\Phi}_{(m-1)k}| \geq \epsilon\sqrt{nk}}]\\ &\quad\leq \frac{c}{\epsilon^{1/2}(nk)^{1/4}} \left(\mathbb{E}(1 + (M^{\Phi}_{j+1} - M^{\Phi}_{j})^4)\right)^{1/2} \left(\mathbb{E}|M^{\Phi}_{j} - M^{\Phi}_{(m-1)k}|\right)^{1/2}\end{aligned} \tag{A.31}$$

We estimate the last term between parentheses in (A.31) as

$$\mathbb{E}|M^{\Phi}_{j} - M^{\Phi}_{(m-1)k}| \leq \sum_{l=(m-1)k}^{j-1} \mathbb{E}|M^{\Phi}_{l+1} - M^{\Phi}_{l}| \leq c(j-(m-1)k) \leq ck, \tag{A.32}$$

where in the second inequality we used (A.25). Combining (A.31) and (A.32) now yields (A.27), where we notice carefully that having the lim sup as $n \to \infty$ applied first is crucial.

Let us turn to the final bound (A.28). Take

$$\Psi(\mathbf{q}, \tilde{\mathbf{q}}) := \left[\bar{\Phi}(\mathbf{q}) + \sum_{k=0}^{\infty}(P^{k+1}\bar{\Phi}(\mathbf{q}) - P^{k+1}\bar{\Phi}(\tilde{\mathbf{q}}))\right]^2 \tag{A.33}$$

Now for any $j \geq (m-1)k$ and with $m, k \geq 1$ we have

$$\begin{aligned}&\mathbb{E}((M^{\Phi}_{j+1} - M^{\Phi}_{j})^2 | \mathcal{F}_{(m-1)k})\\ &\quad= \mathbb{E}\Psi(Q_{(j+1-(m-1)k)+(m-1)k}(\mathbf{q}_0), Q_{(j-(m-1)k)+(m-1)k}(\mathbf{q}_0)) | \mathcal{F}_{(m-1)k})\\ &\quad= H_{j-(m-1)k}(Q_{(m-1)k}(\mathbf{q}_0))\end{aligned}$$

where we have used the Markov property at the last step. Here for any $l \geq 0$

$$H_l(\mathbf{q}_0) := \mathbb{E}\Psi(Q_{l+1}(\mathbf{q}_0), Q_l(\mathbf{q}_0)) = P^l \Gamma(\mathbf{q}_0)$$

with

$$\Gamma(\mathbf{q}_0) = \mathbb{E}\Psi(Q_1(\mathbf{q}_0), \mathbf{q}_0). \tag{A.34}$$

Working from these identities we find, again for any $j \geq (m-1)k$ and with $m, k \geq 1$

$$\frac{1}{k}\sum_{j=(m-1)k}^{mk-1} \mathbb{E}((M^{\Phi}_{j+1} - M^{\Phi}_{j})^2 | \mathcal{F}_{(m-1)k}) = \frac{1}{k}\sum_{j=(m-1)k}^{mk-1} H_{j-(m-1)k}(Q_{(m-1)k}(\mathbf{q}_0))$$
$$= \frac{1}{k}\sum_{j=0}^{k-1} H_j(Q_{(m-1)k}(\mathbf{q}_0)) = \frac{1}{k}\sum_{j=0}^{k-1} P^j \Gamma(Q_{(m-1)k}(\mathbf{q}_0)).$$

As such,

$$\frac{1}{n}\sum_{m=1}^{n} \mathbb{E}\left|\frac{1}{k}\sum_{j=(m-1)k}^{mk-1} \mathbb{E}((M^{\Phi}_{j+1} - M^{\Phi}_{j})^2 | \mathcal{F}_{(m-1)k}) - \sigma^2\right|$$
$$\leq \frac{1}{n}\sum_{m=1}^{n} P^{(m-1)k}\left(\frac{1}{k}\sum_{j=0}^{k-1} |P^l \Gamma(\mathbf{q}_0) - \sigma^2|\right), \tag{A.35}$$

which is valid for any $0 \leq \sigma^2 < \infty$.

With the aim of once again combining (A.4) with (A.21) we now take

$$\sigma^2 = \sigma^2(\Phi) := \int \Gamma(\mathbf{q})\mu_*(d\mathbf{q}). \tag{A.36}$$

with Γ as in (A.34). We will show presently that whenever Φ is ℓ_1-Lipshitz then Γ is ℓ_2-Lipshitz, namely (A.45) below. This being so, as in (A.24), it is clear that $\sigma^2(\Phi) < \infty$ for any ℓ_1-Lipshitz Φ. Moreover, invoking once again (A.4) and (A.21) we obtain that

$$\frac{1}{k}\sum_{j=0}^{k-1} |P^j \Gamma(\mathbf{q}_0) - \sigma^2(\Phi)| \leq \frac{L_\Gamma}{k}\sum_{j=0}^{k-1} \mathcal{W}_{\ell_2}(P^j(\mathbf{q}_0, \cdot), \mu^*) \leq \frac{c(1 + \sqrt{V(\mathbf{q}_0)})}{k}. \tag{A.37}$$

Combining (A.35), (A.37) with (A.19) we find

$$\frac{1}{n}\sum_{m=1}^{n} \mathbb{E}\left|\frac{1}{k}\sum_{j=(m-1)k}^{mk-1} \mathbb{E}((M^{\Phi}_{j+1} - M^{\Phi}_{j})^2 | \mathcal{F}_{(m-1)k}) - \sigma^2(\Phi)\right|$$
$$\leq \frac{c}{nk}\sum_{m=1}^{n} P^{(m-1)k}(1 + \sqrt{V(\mathbf{q}_0)}) \leq \frac{c}{nk}\sum_{m=1}^{n} (1 + \alpha^{(m-1)k}\sqrt{V(\mathbf{q}_0)}) \leq \frac{c(1 + \sqrt{V(\mathbf{q}_0)})}{k}$$

which yields the final item (A.28).

We therefore conclude the proof by showing that $\Gamma \in \mathrm{Lip}_{\ell_2}$ whenever $\Phi \in \mathrm{Lip}_{\ell_1}$. Observe that from (A.34) we have

$$\Gamma(\mathbf{q}) - \Gamma(\tilde{\mathbf{q}}) = \mathbb{E}\Big[(\sqrt{\Psi(Q_1(\mathbf{q}), \mathbf{q}} - \sqrt{\Psi(Q_1(\tilde{\mathbf{q}}), \tilde{\mathbf{q}}})(\sqrt{\Psi(Q_1(\mathbf{q}), \mathbf{q}} + \sqrt{\Psi(Q_1(\tilde{\mathbf{q}}), \tilde{\mathbf{q}}})\Big] \tag{A.38}$$

From (A.33) and invoking (A.4), (A.21) we have that

$$\begin{aligned}
&|\sqrt{\Psi(Q_1(\mathbf{q}), \mathbf{q}} - \sqrt{\Psi(Q_1(\tilde{\mathbf{q}}), \tilde{\mathbf{q}}})| \\
&\qquad \le |\bar{\Phi}(Q_1(\mathbf{q})) - \bar{\Phi}(Q_1(\tilde{\mathbf{q}}))| + |\sum_{k=0}^{\infty}(P^{k+1}\bar{\Phi}(Q_1(\mathbf{q})) - P^{k+1}\bar{\Phi}(Q_1(\tilde{\mathbf{q}})))| \\
&\qquad\quad + |\sum_{k=0}^{\infty}(P^{k+1}\bar{\Phi}(\mathbf{q}) - P^{k+1}\bar{\Phi}(\tilde{\mathbf{q}}))| &\text{(A.39)}\\
&\qquad \le c(\ell_1(Q_1(\mathbf{q}), Q_1(\tilde{\mathbf{q}})) + \ell_1(\mathbf{q}, \tilde{\mathbf{q}})). &\text{(A.40)}
\end{aligned}$$

On the other hand, again with (A.4), (A.11) and (A.21) we also obtain the bound

$$\begin{aligned}
&|\sqrt{\Psi(Q_1(\mathbf{q}), \mathbf{q}} + \sqrt{\Psi(Q_1(\tilde{\mathbf{q}}), \tilde{\mathbf{q}}})| \\
&\quad \le c\left(|\bar{\Phi}(Q_1(\mathbf{q}))| + |\bar{\Phi}(Q_1(\tilde{\mathbf{q}}))| + \mathcal{W}_{\ell_1}(\delta_{Q_1(\mathbf{q})}, \mu_*) + \mathcal{W}_{\ell_1}(\delta_{Q_1(\tilde{\mathbf{q}})}, \mu_*)\right. \\
&\quad\quad \left. + Wass_{\ell_1}(\delta_{\mathbf{q}}, \mu_*) + \mathcal{W}_{\ell_1}(\delta_{\tilde{\mathbf{q}}}, \mu_*)\right) \\
&\quad \le c\left(1 + \sqrt{V(Q_1(\mathbf{q}))} + \sqrt{V(Q_1(\tilde{\mathbf{q}}))} + \sqrt{V(\mathbf{q})} + \sqrt{V(\tilde{\mathbf{q}})}\right) \\
&\quad \le c\left(\sqrt{1 + V(Q_1(\mathbf{q})) + V(Q_1(\tilde{\mathbf{q}}))} + \sqrt{1 + V(\mathbf{q}) + V(\tilde{\mathbf{q}})}\right) &\text{(A.41)}
\end{aligned}$$

Now observe that, for any $\mathbf{q}, \tilde{\mathbf{q}} \in \mathbb{V}$

$$\ell_1(\mathbf{q}, \tilde{\mathbf{q}})\sqrt{1 + V(\mathbf{q}) + V(\tilde{\mathbf{q}})} \le 2\ell_2(\mathbf{q}, \tilde{\mathbf{q}}), \tag{A.42}$$

so that combining this simple observation with (A.38)–(A.41) we find

$$\begin{aligned}
&|\Gamma(\mathbf{q}) - \Gamma(\tilde{\mathbf{q}})| &\text{(A.43)}\\
&\quad \le c\mathbb{E}[(\ell_1(Q_1(\mathbf{q}), Q_1(\tilde{\mathbf{q}})) + \ell_1(\mathbf{q}, \tilde{\mathbf{q}}))(\sqrt{1 + V(Q_1(\mathbf{q})) + V(Q_1(\tilde{\mathbf{q}}))} \\
&\quad\quad + \sqrt{1 + V(\mathbf{q}) + V(\tilde{\mathbf{q}})})] \\
&\quad \le c\mathbb{E}\ell_2(Q_1(\mathbf{q}), Q_1(\tilde{\mathbf{q}})) + c\ell_1(\mathbf{q}, \tilde{\mathbf{q}})\mathbb{E}\left(\sqrt{1 + V(Q_1(\mathbf{q})) + V(Q_1(\tilde{\mathbf{q}}))}\right) \\
&\quad\quad + c\sqrt{1 + V(\mathbf{q}) + V(\tilde{\mathbf{q}})}\mathbb{E}\ell_1(Q_1(\mathbf{q}), Q_1(\tilde{\mathbf{q}})) + c\ell_2(\mathbf{q}, \tilde{\mathbf{q}}). &\text{(A.44)}
\end{aligned}$$

Now notice that, under (A.19) we have

$$\mathbb{E}\left(\sqrt{1 + V(Q_1(\mathbf{q})) + V(Q_1(\tilde{\mathbf{q}}))}\right) \le c\sqrt{1 + V(\mathbf{q}) + V(\tilde{\mathbf{q}})}.$$

On the other hand, notice that we may take $Q_1(\mathbf{q})$ and $Q_1(\tilde{\mathbf{q}})$ to be any coupling of $P(\mathbf{q}, \cdot)$ and $P(\tilde{\mathbf{q}}, \cdot)$ in (A.44). As such, with (A.44) and these two observations

$$\begin{aligned}|\Gamma(\mathbf{q}) - \Gamma(\tilde{\mathbf{q}})| &\leq \mathcal{W}_{\ell_2}(P(\mathbf{q}, \cdot), P(\tilde{\mathbf{q}}, \cdot)) + c\ell_1(\mathbf{q}, \tilde{\mathbf{q}})\sqrt{1 + V(\mathbf{q}) + V(\tilde{\mathbf{q}})}\\ &\quad + c\mathcal{W}_{\ell_1}(P(\mathbf{q}, \cdot), P(\tilde{\mathbf{q}}, \cdot))\sqrt{1 + V(\mathbf{q}) + V(\tilde{\mathbf{q}})} + c\ell_2(\mathbf{q}, \tilde{\mathbf{q}}),\end{aligned}$$

so that with (A.42) and a final invocation of (A.21), we have

$$|\Gamma(\mathbf{q}) - \Gamma(\tilde{\mathbf{q}})| \leq c\ell_2(\mathbf{q}, \tilde{\mathbf{q}}). \tag{A.45}$$

The proof is now complete. □

We conclude this section with the following proposition which gives a sufficient condition for a function to be ℓ-Lipschitz for a class of distance-like functions including those appearing in the main results of this work.

Proposition 46 *Let $(\mathbb{V}, \|\cdot\|)$ be a Banach space and consider distance-like functions of the form*

$$\ell(\mathbf{q}, \tilde{\mathbf{q}}) = \sqrt{\left(\frac{\|\mathbf{q} - \tilde{\mathbf{q}}\|}{\varepsilon} \wedge 1\right)(1 + V(\mathbf{q}) + V(\tilde{\mathbf{q}}))} \tag{A.46}$$

where we suppose that $\varepsilon > 0$ and $V : \mathbb{V} \to [0, \infty)$ is convex. Given any continuously differentiable function $\Phi : \mathbb{V} \to \mathbb{R}$, define

$$L_\Phi := \sup_{\mathbf{q} \in \mathbb{V}} \frac{\max\{2|\Phi(\mathbf{q})|, \sqrt{\varepsilon}\|D\Phi(\mathbf{q})\|\}}{\sqrt{1 + V(\mathbf{q})}}. \tag{A.47}$$

If $L_\Phi < \infty$ then Φ is ℓ-Lipschitz, with L_Φ providing a suitable Lipschitz constant.

Proof Fix any $\mathbf{q}, \tilde{\mathbf{q}} \in \mathbb{V}$. We consider separately the cases when $\|\mathbf{q} - \tilde{\mathbf{q}}\| > \varepsilon$ and when $\|\mathbf{q} - \tilde{\mathbf{q}}\| \leq \varepsilon$. In the first situation when $\|\mathbf{q} - \tilde{\mathbf{q}}\| > \varepsilon$ we estimate

$$|\Phi(\mathbf{q}) - \Phi(\tilde{\mathbf{q}})| \leq \sqrt{1 + V(\mathbf{q}) + V(\tilde{\mathbf{q}})}\left(\frac{|\Phi(\mathbf{q})|}{\sqrt{1 + V(\mathbf{q})}} + \frac{|\Phi(\tilde{\mathbf{q}})|}{\sqrt{1 + V(\tilde{\mathbf{q}})}}\right) \leq L_\Phi \ell(\mathbf{q}, \tilde{\mathbf{q}}).$$

Now consider the case when $\|\mathbf{q} - \tilde{\mathbf{q}}\| \leq \varepsilon$. Let $\mathbf{q}_s = \mathbf{q} + s(\tilde{\mathbf{q}} - \mathbf{q})$, for $s \in [0, 1]$ and observe that

$$\begin{aligned}|\Phi(\mathbf{q}) - \Phi(\tilde{\mathbf{q}})| &\leq \|\mathbf{q} - \tilde{\mathbf{q}}\| \int_0^1 \|D\Phi(\mathbf{q}_s)\| ds\\ &\leq \int_0^1 \sqrt{\left(\frac{\|\mathbf{q} - \tilde{\mathbf{q}}\|}{\varepsilon}\right)(1 + V(\mathbf{q}_s))} \cdot \frac{\sqrt{\varepsilon}\|D\Phi(\mathbf{q}_s)\|}{\sqrt{1 + V(\mathbf{q}_s)}} ds\\ &\leq L_\Phi \int_0^1 \sqrt{\left(\frac{\|\mathbf{q} - \tilde{\mathbf{q}}\|}{\varepsilon}\right)(1 + sV(\mathbf{q}) + (1 - s)V(\tilde{\mathbf{q}}))} ds \leq L_\Phi \ell(\mathbf{q}, \tilde{\mathbf{q}})\end{aligned}$$

where we have used the convexity of V for the penultimate bound. The proof is complete. □

References

1. Berestycki, H., Kiselev, A., Novikov, A., Ryzhik, L.: The explosion problem in a flow. J. Anal. Math. **110**(1), 31–65 (2010)
2. Besag, J.: Comments on "Representations of knowledge in complex systems" by U. Grenander and MI Miller. J. R. Stat. Soc. Ser. B **56**, 591–592 (1994)
3. Beskos, A., Girolami, M., Lan, S., Farrell, P., Stuart, A.: Geometric MCMC for infinite-dimensional inverse problems. J. Comput. Phys. **335**, 327–351 (2017)
4. Beskos, A., Kalogeropoulos, K., Pazos, E.: Advanced MCMC methods for sampling on diffusion pathspace. Stoch. Process. Appl. **123**(4), 1415–1453 (2013)
5. Beskos, A., Pillai, N., Roberts, G., Sanz-Serna, J., Stuart, A.: Optimal tuning of the hybrid Monte Carlo algorithm. Bernoulli **19**(5A), 1501–1534 (2013)
6. Beskos, A., Pinski, F.J., Sanz-Serna, J., Stuart, A.: Hybrid Monte-Carlo on Hilbert spaces. Stoch. Process. Appl. **121**(10), 2201–2230 (2011)
7. Beskos, A., Roberts, G., Stuart, A., Voss, J.: MCMC methods for diffusion bridges. Stoch. Dyn. **8**(03), 319–350 (2008)
8. Betancourt, M.: The convergence of Markov chain Monte Carlo methods: from the Metropolis method to Hamiltonian Monte Carlo. Ann. Phys. **531**(3), 1700214 (2019)
9. Betancourt, M., Byrne, S., Livingstone, S., Girolami, M.: The geometric foundations of Hamiltonian Monte Carlo. Bernoulli **23**(4A), 2257–2298 (2017)
10. Bogachev, V.: Gaussian Measures, 62. American Mathematical Society, Providence (1998)
11. Borggaard, J., Glatt-Holtz, N., Krometis, J.: On Bayesian consistency for flows observed through a passive scalar. Ann. Appl. Probab. **30**(4), 1762–1783 (2020)
12. Borggaard, J., Glatt-Holtz, N., Krometis, J.: A Bayesian approach to estimating background flows from a passive scalar. SIAM/ASA J. Uncertain. Quant. **8**(3), 1036–1060 (2020)
13. Bou-Rabee, N., Eberle, A.: Two-scale coupling for preconditioned Hamiltonian Monte Carlo in infinite dimensions (2019). ArXiv preprint arXiv:1909.07962 [math.PR]
14. Bou-Rabee, N., Eberle, A., Zimmer, R.: Coupling and convergence for Hamiltonian Monte Carlo Ann. Appl. Probab. (to appear) (2018). arXiv preprint arXiv:1805.00452
15. Bou-Rabee, N., Sanz-Serna, J.: Randomized Hamiltonian Monte Carlo. Ann. Appl. Probab. **27**(4), 2159–2194 (2017)
16. Bou-Rabee, N., Sanz-Serna, J.: Geometric integrators and the Hamiltonian Monte Carlo method. Acta Numer. **27**, 113–206 (2018)
17. Bui-Thanh, T., Ghattas, O.: An analysis of infinite dimensional Bayesian inverse shape acoustic scattering and its numerical approximation. SIAM/ASA J. Uncertain. Quant. **2**(1), 203–222 (2014)
18. Bui-Thanh, T., Nguyen, Q.: FEM-based discretization-invariant MCMC methods for PDE-constrained Bayesian inverse problems. Inverse Probl. Imaging **10**(4), 943–975 (2016)
19. Butkovsky, O.: Subgeometric rates of convergence of Markov processes in the Wasserstein metric. Ann. Appl. Probab. **24**(2), 526–552 (2014)
20. Butkovsky, O., Kulik, A., Scheutzow, M.: Generalized couplings and ergodic rates for SPDEs and other Markov models. arXiv preprint arXiv:1806.00395 (2018)
21. Chandler, D., Wolynes, P.: Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids. J. Chem. Phys. **74**(7), 4078–4095 (1981). https://doi.org/10.1063/1.441588
22. Chow, Y.: On a strong law of large numbers for martingales. Ann. Math. Stat. **38**(2), 610 (1967)
23. Cotter, S., Roberts, G., Stuart, A., White, D.: MCMC methods for functions: modifying old algorithms to make them faster. Stat. Sci. **28**(3), 424–446 (2013)
24. Craig, I., Manolopoulos, D.: Quantum statistics and classical mechanics: real time correlation functions from ring polymer molecular dynamics. J. Chem. Phys. **121**(8), 3368–3373 (2004). https://doi.org/10.1063/1.1777575
25. Craig, I., Manolopoulos, D.: A refined ring polymer molecular dynamics theory of chemical reaction rates. J. Chem. Phys. **123**(3), 034102 (2005). https://doi.org/10.1063/1.1954769

26. Craig, I., Manolopoulos, D.: Chemical reaction rates from ring polymer molecular dynamics. J. Chem. Phys. **122**(8), 084106 (2005). https://doi.org/10.1063/1.1850093
27. Da Prato, G., Zabczyk, J.: Stochastic Equations in Infinite Dimensions. Cambridge University Press, Cambridge (2014)
28. Dashti, M., Stuart, A.: The Bayesian approach to inverse problems. In: Ghanem, R., Higdon, D., Owhadi, H. (eds.) Handbook of Uncertainty Quantification, pp. 311–428. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-12385-1_7
29. Duane, S., Kennedy, A., Pendleton, B., Roweth, D.: Hybrid Monte-Carlo. Phys. Lett. B **195**, 216–222 (1987)
30. Durmus, A., Fort, G., Moulines, E.: Subgeometric rates of convergence in Wasserstein distance for Markov chains. In: Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, vol. 52, pp. 1799–1822. Institut Henri Poincaré (2016)
31. Durmus, A., Moulines, E.: Quantitative bounds of convergence for geometrically ergodic Markov chain in the Wasserstein distance with application to the Metropolis adjusted Langevin algorithm. Stat. Comput. **25**(1), 5–19 (2015)
32. Durmus, A., Moulines, E., Saksman, E.: On the convergence of Hamiltonian Monte Carlo. arXiv preprint arXiv:1705.00166 (2017)
33. Durrett, R.: Probability: Theory and Examples, vol. 49. Cambridge University Press, Cambridge (2019)
34. Eberle, A., Guillin, A., Zimmer, R.: Couplings and quantitative contraction rates for Langevin dynamics (2017). ArXiv preprint arXiv:1703.01617
35. 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)
36. Gelman, A., Lee, D., Guo, J.: Stan: a probabilistic programming language for Bayesian inference and optimization. J. Educ. Behav. Stat. **40**(5), 530–543 (2015)
37. Glatt-Holtz, N., Mattingly, J., Richards, G.: On unique ergodicity in nonlinear stochastic partial differential equations. J. Stat. Phys. (2016). https://doi.org/10.1007/s10955-016-1605-x
38. Grenander, U., Miller, M.: Representations of knowledge in complex systems. J. R. Stat. Soc. Ser. B (Methodol.) **56**(4), 549–581 (1994)
39. Habershon, S., Fanourgakis, G., 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). https://doi.org/10.1063/1.2968555
40. Habershon, S., Manolopoulos, D., Markland, T., Miller, T.: Ring-polymer molecular dynamics: quantum effects in chemical dynamics from classical trajectories in an extended phase space. Annu. Rev. Phys. Chem. **64**(1), 387–413 (2013). https://doi.org/10.1146/annurev-physchem-040412-110122
41. Hairer, M., Mattingly, J.: Spectral gaps in Wasserstein distances and the 2D stochastic Navier–Stokes equations. Ann. Probab. **36**(6), 2050–2091 (2008)
42. Hairer, M., Mattingly, J.: Yet another look at Harris' ergodic theorem for Markov chains. In: Seminar on Stochastic Analysis, Random Fields and Applications VI, pp. 109–117. Springer (2011)
43. Hairer, M., Mattingly, J., 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)
44. Hairer, M., Stuart, A., Vollmer, S.: Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. Ann. Appl. Probab. **24**(6), 2455–2490 (2014)
45. Hairer, M., Stuart, A., Voss, J.: Analysis of SPDEs arising in path sampling part II: the nonlinear case. Ann. Appl. Probab. **17**(5/6), 1657–1706 (2007)
46. Hairer, M., Stuart, A., Voss, J.: Sampling conditioned diffusions. Trends Stoch. Anal. **353**, 159–186 (2009)
47. Hairer, M., Stuart, A., Voss, J.: Signal processing problems on function space: Bayesian formulation, stochastic PDEs and effective MCMC methods. In: The Oxford Handbook of Nonlinear Filtering, pp. 833–873, Oxford Univ. Press, Oxford, (2011)
48. Hairer, M., Stuart, A., Voss, J., Wiberg, P.: Analysis of SPDEs arising in path sampling. Part I: the Gaussian case. Commun. Math. Sci. **3**(4), 587–603 (2005)
49. Harris, T.: The existence of stationary measures for certain Markov processes. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, vol. 2, pp. 113–124 (1956)
50. Hastings, W.: Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**(1), 97–109 (1970)

51. Kaipio, J., Somersalo, E.: Statistical and Computational Inverse Problems, Applied Mathematical Sciences, vol. 160. Springer, Berlin (2005)
52. Komorowski, T., Landim, C., Olla, S.: Fluctuations in Markov Processes: Time Symmetry and Martingale Approximation, vol. 345. Springer, Berlin (2012)
53. Komorowski, T., Walczuk, A.: Central limit theorem for Markov processes with spectral gap in the Wasserstein metric. Stoch. Process. Appl. **122**(5), 2155–2184 (2012)
54. Korol, R., Bou-Rabee, N., Miller III, T.: Cayley modification for strongly stable path-integral and ring-polymer molecular dynamics. J. Chem. Phys. (to appear) (2019). arXiv preprint arXiv:1907.07941
55. Kuksin, S., Shirikyan, A.: Mathematics of Two-Dimensional Turbulence, vol. 194. Cambridge University Press, Cambridge (2012)
56. Kulik, A.: Ergodic Behavior of Markov Processes: With Applications to Limit Theorems, vol. 67. Walter de Gruyter GmbH & Co KG, Berlin (2017)
57. Liu, J.: Monte Carlo Strategies in Scientific Computing, 2nd edn. Springer, Berlin (2008)
58. Livingstone, S., Betancourt, M., Byrne, S., Girolami, M.: On the geometric ergodicity of Hamiltonian Monte Carlo. Bernoulli **25**(4A), 3109–3138 (2019)
59. Lu, J., Zhou, Z.: Continuum limit and preconditioned Langevin sampling of the path integral molecular dynamics. Preprint (2018). arXiv:1811.10995
60. Mangoubi, O., Pillai, N., Smith, A.: Does Hamiltonian Monte Carlo mix faster than a random walk on multimodal densities? arXiv preprint arXiv:1808.03230 (2018)
61. Mangoubi, O., Smith, A.: Rapid mixing of Hamiltonian Monte Carlo on strongly log-concave distributions. arXiv preprint arXiv:1708.07114 (2017)
62. Mangoubi, O., Smith, A.: Mixing of Hamiltonian Monte Carlo on strongly log-concave distributions 2: Numerical integrators. In: The 22nd International Conference on Artificial Intelligence and Statistics, pp. 586–595 (2019)
63. Martin, J., Wilcox, L., Burstedde, C., Ghattas, O.: A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. SIAM J. Sci. Comput. **34**(3), A1460–A1487 (2012)
64. Marzouk, Y., Najm, H., Rahn, L.: Stochastic spectral methods for efficient Bayesian solution of inverse problems. J. Comput. Phys. **224**(2), 560–586 (2007)
65. Mattingly, J.: Exponential convergence for the stochastically forced Navier–Stokes equations and other partially dissipative dynamics. Commun. Math. Phys. **230**(3), 421–462 (2002)
66. Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E.: Equation of state calculations by fast computing machines. J. Chem. Phys. **21**(6), 1087–1092 (1953)
67. Metropolis, N., Ulam, S.: The Monte Carlo method. J. Am. Stat. Assoc. **44**(247), 335–341 (1949)
68. Meyn, S., Tweedie, R.: Markov Chains and Stochastic Stability. Springer, Berlin (2012)
69. Miller, T., Manolopoulos, D.: Quantum diffusion in liquid para-hydrogen from ring-polymer molecular dynamics. J. Chem. Phys. **122**(18), 184503 (2005). https://doi.org/10.1063/1.1893956
70. Miller, T., Manolopoulos, D.: Quantum diffusion in liquid water from ring polymer molecular dynamics. J. Chem. Phys. **123**(15), 154504 (2005). https://doi.org/10.1063/1.2074967
71. Neal, R.: Probabilistic inference using Markov chain Monte Carlo methods. Department of Computer Science, University of Toronto Toronto, ON, Canada (1993)
72. Neal, R.: MCMC using Hamiltonian dynamics. In: Handbook of Markov Chain Monte Carlo, vol. 2, pp. 113–162 (2011)
73. Ottobre, M., Pillai, N., Pinski, F., Stuart, A.: A function space HMC algorithm with second order Langevin diffusion limit. Bernoulli **22**(1), 60–106 (2016)
74. Petra, N., Martin, J., Stadler, G., Ghattas, O.: A computational framework for infinite-dimensional Bayesian inverse problems, part II: stochastic Newton MCMC with application to ice sheet flow inverse problems. SIAM J. Sci. Comput. **36**(4), A1525–A1555 (2014)
75. Reznikoff, M., Vanden-Eijnden, E.: Invariant measures of stochastic partial differential equations and conditioned diffusions. C. R. Math. **340**(4), 305–308 (2005)
76. Robert, C., Casella, G.: Monte Carlo Statistical Methods. Springer, Berlin (2013)
77. Rudolf, D.: Explicit error bounds for Markov chain Monte Carlo. arXiv preprint arXiv:1108.3201 (2011)
78. Stuart, A.: Inverse problems: a Bayesian perspective. Acta Numer. **19**, 451–559 (2010)
79. Team, S.D. : Stan modeling language users guide and reference manual. Technical report (2016)
80. Temam, R.: Navier–Stokes Equations: Theory and Numerical Analysis. AMS Chelsea Publishing, Providence, RI (2001).. ((**Reprint of the 1984 edition**))

81. Tierney, L.: A note on Metropolis–Hastings kernels for general state spaces. Ann. Appl. Probab. **8**(1), 1–9 (1998)
82. Tsybakov, A.: Introduction to Nonparametric Estimation. Springer Series in Statistics, Springer, New York, (2009)
83. Villani, C.: Optimal Transport: Old and New, vol. 338. Springer, Berlin (2008)

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

Authors and Affiliations

Nathan E. Glatt-Holtz[1] · Cecilia F. Mondaini[2]

✉ Cecilia F. Mondaini
cf823@drexel.edu
https://drexel.edu/coas/faculty-research/faculty-directory/mondaini-cecilia/

Nathan E. Glatt-Holtz
negh@tulane.edu
http://www.math.tulane.edu/~negh/

1 Department of Mathematics, Tulane University, New Orleans, USA

2 Department of Mathematics, Drexel University, Philadelphia, USA