RANDOMIZED HAMILTONIAN MONTE CARLO

BY NAWAF BOU-RABEE[1] AND JESÚS MARÍA SANZ-SERNA[2]

Rutgers University–Camden and Universidad Carlos III de Madrid

Tuning the durations of the Hamiltonian flow in Hamiltonian Monte Carlo (also called Hybrid Monte Carlo) (HMC) involves a tradeoff between computational cost and sampling quality, which is typically challenging to resolve in a satisfactory way. In this article, we present and analyze a randomized HMC method (RHMC), in which these durations are i.i.d. exponential random variables whose mean is a free parameter. We focus on the small time step size limit, where the algorithm is rejection-free and the computational cost is proportional to the mean duration. In this limit, we prove that RHMC is geometrically ergodic under the same conditions that imply geometric ergodicity of the solution to underdamped Langevin equations. Moreover, in the context of a multidimensional Gaussian distribution, we prove that the sampling efficiency of RHMC, unlike that of constant duration HMC, behaves in a regular way. This regularity is also verified numerically in non-Gaussian target distributions. Finally, we suggest variants of RHMC for which the time step size is not required to be small.

1. Introduction. In the present article, we suggest a randomized version of the Hamiltonian Monte Carlo (also called Hybrid Monte Carlo) algorithm that, under very general hypotheses, may be proved to be *geometrically ergodic*. The Hamiltonian Monte Carlo (HMC) algorithm is a general purpose Markov Chain Monte Carlo (MCMC) tool for sampling from a probability distribution Π [11, 22, 31, 33]. It offers the potential of generating proposed moves that are far away from the current location of the chain and yet may be accepted with high probability. The algorithm is based on integrating a Hamiltonian system and possesses two free parameters: the duration of the Hamiltonian flow and the time step size of the integrator. Unfortunately, the performance of HMC depends crucially on the values assigned by the user to those parameters; while for some parameter values HMC may be highly efficient, it is well known that, as discussed below, there are values for which the algorithm, in its simplest form, is not even ergodic. The Randomized Hybrid Monte Carlo (RHMC) addresses these shortcomings of HMC.

We target probability distributions of the form

$$(1) \qquad \Pi(dq) = C_0^{-1} \exp(-\Phi(q)) \, dq, \qquad C_0 = \int_{\mathbb{R}^D} \exp(-\Phi(q)) \, dq,$$

where the negative loglikelihood $\Phi : \mathbb{R}^D \to \mathbb{R}$ is seen as a potential energy function. We assume that Φ is at least differentiable and such that $C_0 < \infty$. As is the case with other MCMC methods, HMC does not require knowing the normalization factor C_0. HMC enlarges the state space from $\mathbb{R}^D$ to $\mathbb{R}^{2D}$ and considers the Boltzmann–Gibbs distribution in this space

$$(2) \qquad \Pi_{\mathrm{BG}}(dq, dp) = C_0^{-1} (2\pi)^{-D/2} \exp(-H(q, p)) \, dq \, dp,$$

where the artificial Hamiltonian function (or total energy) $H : \mathbb{R}^{2D} \to \mathbb{R}$ is taken to be

$$(3) \qquad H(q, p) = \frac{|p|^2}{2} + \Phi(q).$$

The vector p plays the role of a mechanical momentum and the term $|p|^2/2$ is the corresponding kinetic energy. The target Π is the marginal of Π_{BG} on q; the marginal on p is, of course, the standard D-dimensional Gaussian $\mathcal{N}(0, 1)^D$. More complicated kinetic energies of the form $(1/2) p^T M^{-1} p$, with M a symmetric positive definite mass matrix, may also be used, but, for notational simplicity, we restrict our study to the Hamiltonian (3).

The basic idea of HMC is encapsulated in the following procedure, where the duration $\lambda > 0$ is a (deterministic) parameter whose value is specified by the user.

ALGORITHM 1.1 (HMC).   Given the duration parameter $\lambda > 0$ and the current state of the chain $X_0 \in \mathbb{R}^D$, the method outputs a state $X_1 \in \mathbb{R}^D$ as follows:

Step 1. Generate a D-dimensional random vector $\xi_0 \sim \mathcal{N}(0, 1)^D$.
Step 2. Evolve over $[0, \lambda]$ Hamilton's equations associated to (3)

$$(4) \qquad \begin{aligned} \dot{q} &= p, \\ \dot{p} &= -\nabla \Phi(q), \end{aligned}$$

with initial condition $(q(0), p(0)) = (X_0, \xi_0)$.
Step 3. Output $X_1 = q(\lambda)$.

Since Step 2 conserves the Boltzmann–Gibbs distribution, it is clear that the mapping $X_0 \mapsto X_1$ preserves the target Π and, therefore, may be used to generate a Markov chain having Π as an invariant distribution (in fact the resulting chain is actually reversible with respect to Π). Note that Step 1 is easy to perform since it only involves generating a D-dimensional normal random vector. This step is the only source of randomness in determining X_1 conditional on X_0. The Hamiltonian flow in Step 2 is what, in principle, enables HMC to make large moves in

state space that reduce correlations in the Markov chain $\{X_i\}$. Roughly speaking, one may hope that, by increasing λ, X_1 moves away from X_0, thus reducing correlation. However, simple examples show that this outcome is far from assured. Indeed, for the univariate normal distribution with $\Phi(q) = q^2/2$, Hamilton's equations coincide with those of the harmonic oscillator, $(d/dt)q = -p$, $(d/dt)q = p$, and the flow is a rotation in the (q, p) plane with period 2π. It is easy to see (see Section 5 for a fuller discussion) that, if X_0 is taken from the target distribution, as λ increases from 0 to $\pi/2$, the correlation between X_1 and X_0 decreases and for $\lambda = \pi/2$, X_1 and X_0 are independent. However, increasing λ beyond $\pi/2$ will cause an increase of the correlation and for $\lambda = \pi$, $X_1 = -X_0$ and the chain is not ergodic. For general distributions, it is likely that a small λ will lead to a highly correlated chain, while choosing λ too large may cause the Hamiltonian trajectory to make a U-turn and fold back on itself, thus increasing correlation [17].

In practice, a formula for the exact solution used in Step 2 is unavailable and a numerical solution is used instead. Thus, in addition to the duration λ of the Hamiltonian flow in Step 2, another key parameter in the HMC method is the time step size Δt used to generate this numerical solution. To correct the bias introduced by time discretization error, a Metropolis–Hastings accept-reject step is also added [16, 28]. In order to keep the Metropolis–Hastings ratio simple, typically a volume-preserving and reversible method is used to numerically simulate Hamilton's equations in Step 2 [13]. The integrator of choice is the Verlet method, which is second-order accurate and, like Euler's rule, only requires one new evaluation of the gradient $\nabla\Phi(q)$ per step. Unfortunately, time discretization does not remove the complex dependence of correlation on the duration parameter λ. For instance, in the preceding example where $\Phi(q) = q^2/2$, it is easy to check that if λ is close to an integer multiple of π and $\Delta t > 0$ is suitably chosen, the Verlet numerical integration will result, for each X_0, in $X_1 = -X_0$ (a move that will be accepted by the Metropolis–Hasting step). To avoid such poor performance, HMC is typically operated with values of Δt that are randomized [23, 31]. Since, due to stability restrictions, explicit integrators cannot operate with arbitrarily large values of the time step, Δt is typically chosen from a uniform distribution in an (often narrow) interval $(\Delta t_{\min}, \Delta t_{\max})$. In any case, the fact remains that increasing the duration parameter will increase the computational cost and may impair the quality of the sampling.

In this paper, we randomize the duration of the Hamiltonian flow, cf. [8, 23, 31]. More precisely, in RHMC the lengths of the time intervals of integration of the Hamiltonian dynamics at the different steps of the Markov chain are identically distributed exponential random variables; these durations are mutually independent and independent of the state of the chain. In what follows, we are primarily interested in the case where the procedure uses the exact Hamiltonian flow (or, in practical terms, where the integration is carried out with such a small value of Δt which ensures that essentially all steps of the chain result in acceptance). This leaves the mean duration as the only free parameter. In this exact integration

scenario, we prove that, regardless of the choice of the mean duration, RHMC is geometrically ergodic. Furthermore, we show that the dependence of the performance of RHMC on this mean duration parameter is simpler than the dependence of the performance of HMC on its constant duration parameter. A full discussion of the situation where time-discretization errors are taken into account requires heavy use of numerical analysis techniques and will be the subject of a future publication. Nevertheless, in Section 6, we present two variants of RHMC, based on the ideas of Markov chain approximation methods (MCAM) [7, 20], that replace the Hamiltonian flow by a numerical approximation and may be treated with the infinitesimal tools we employ in the exact integration scenario.

Section 2 provides a description of the RHMC method and its infinitesimal generator L. In Section 3, we prove that the measure Π_{BG} is infinitesimally invariant for L. We then construct a Lyapunov function for RHMC that is of the same form as that used for the Langevin equations and requires similar assumptions on the potential energy function [27, 35, 37]. Here, it is important to point out that, while the Langevin dynamics explicitly includes friction, the dissipative behavior of RHMC comes from the randomization of the momentum. In particular, if the chain is at a location of high potential energy, the Hamiltonian dynamics will typically change a large part of the potential energy into kinetic energy; then, with high probability, the next momentum randomization will decrease the kinetic energy. With this Lyapunov function, we extend a local semimartingale representation of the process. It also follows that Dynkin's formula holds for functions that satisfy a mild growth condition. Using Dynkin's formula, we prove that the measure Π_{BG} is invariant for RHMC. Using a generating function for the Hamiltonian flow and a Duhamel formula for the Markov semigroup associated with the RHMC process, we prove that the transition probability distribution of RHMC satisfies a minorization condition. We then invoke Harris's theorem to conclude that RHMC is geometrically ergodic with respect to Π_{BG}, for any choice of the mean duration parameter λ.

Section 4 considers the model problem of a multidimensional Gaussian target distribution [23, 31]. For both RHMC and HMC, explicit formulas are derived for (i) the integrated autocorrelation time associated to the standard estimator of the mean of the target distribution; and (ii) the equilibrium mean-squared displacement. These formulas imply that the sampling efficiency of RHMC behaves in a simple way, while the sampling efficiency of HMC displays complicated dependence on the duration parameter. Section 5 presents numerical tests for a two-dimensional double well potential energy and a potential energy used in the chemistry literature for a pentane molecule. These tests support our theoretical findings. Two variants of RHMC that do not assume that integration errors are negligible are suggested in Section 6. The first of them is easily proved to have a Lyapunov function under suitable hypotheses but introduces a bias in the target distribution; the second removes the bias by allowing momentum flips at a rate dictated by a Metropolis–Hastings ratio. There is an Appendix devoted to Harris's theorem.

To summarize, the main theoretical contributions of this paper are the following:

- a proof of a Foster–Lyapunov drift condition for the infinitesimal generator of RHMC;
- a solution to a martingale problem for RHMC;
- a proof that infinitesimal invariance of the measure Π_{BG} and Dynkin's formula imply that the measure Π_{BG} is invariant for RHMC;
- a minorization condition for RHMC;
- a proof that Π_{BG} is the unique invariant measure of RHMC and that RHMC is geometrically ergodic (which combines all of the previous results); and
- introduced two practical implementations of RHMC, including one that is unbiased.

Let us finish this introduction by pointing out that the RHMC method is related to Anderson's impulsive thermostat common in molecular dynamics, which describes a molecular system interacting with a heat bath [2, 12, 21]. The molecular system is modeled using Hamiltonian dynamics, and its interaction with a heat bath is modeled by collisions that cause an instantaneous randomization of the momentum of a randomly chosen particle. The times between successive collisions are assumed to be i.i.d. exponential random variables. In [12], E and Li prove that the Anderson thermostat on a hyper-torus is geometrically ergodic. Since the state space is bounded, the main issue in that proof is the derivation of a minorization condition for the process. Our proof of geometric ergodicity of the RHMC method can be modified to extend their result to an unbounded space.

2. RHMC method. Here, we provide step-by-step instructions to produce an RHMC trajectory, and afterwards, introduce the infinitesimal generator L of RHMC.

The RHMC method generates a right-continuous with left limits (càdlàg) Markov process Z_t. While Algorithm 1.1 was formulated in $\mathbb{R}^D$, the process $Z_t = (Q_t, P_t)$ is defined in the enlarged space $\mathbb{R}^{2D}$ to include the possibility of partial randomization of the momentum, as in the generalized hybrid Monte Carlo of Horowitz [1, 18, 19]. This process Z_t can be simulated by iterating the following steps. The mean duration $\lambda > 0$ and the Horowitz angle $\phi \in (0, \pi/2]$ are deterministic parameters.

ALGORITHM 2.1 (RHMC). Given the current time t_0 and the current state $Z_{t_0} = (Q_{t_0}, P_{t_0})$, the method computes the next momentum randomization time $t_1 > t_0$ and the path of the process $Z_s = (Q_s, P_s)$ over $(t_0, t_1]$ as follows:

Step 1. Update time via $t_1 = t_0 + \delta t_0$ where $\delta t_0 \sim \text{Exp}(1/\lambda)$.
Step 2. Evolve over $[t_0, t_1]$ Hamilton's equations (4) associated with (3) with initial condition $(q(t_0), p(t_0)) = (Q_{t_0}, P_{t_0})$.
Step 3. Set

$$Z_s = (Q_s, P_s) = (q(s), p(s)) \qquad \text{for } t_0 \leq s < t_1.$$

Step 4. Randomize momentum by setting

$$Z_{t_1} = \big(q(t_1), \cos(\phi) p(t_1) + \sin(\phi)\xi\big),$$

where $\xi \sim \mathcal{N}(0, 1)^D$.

Step 5. Output $X_1 = q(t_1)$.

On test functions $f \in C^1(\mathbb{R}^{2D})$, the infinitesimal generator of Z is given by

(5)
$$Lf(q, p) = \overbrace{\lambda^{-1}\mathbb{E}\{f\big(\Gamma(q, p)\big) - f(q, p)\}}^{\text{momentum randomization operator}}$$
$$+ \underbrace{p^T \nabla_q f(q, p) - \nabla_q \Phi(q)^T \nabla_p f(q, p)}_{\text{Liouville operator}}.$$

The expectation in the momentum randomization operator is over the random variable $\Gamma(q, p)$ defined as

(6)
$$\Gamma(q, p) = \big(q, \cos(\phi)p + \sin(\phi)\xi\big),$$

where $\xi \sim \mathcal{N}(0, 1)^D$ and $\phi \in (0, \pi/2]$. A sample path of this process is given by the Hamiltonian flow associated with (3) with intermittent and instantaneous jumps in momentum. The random times between successive jumps are independent and exponentially distributed with mean λ. The Horowitz angle ϕ is a deterministic parameter that governs how much the momentum immediately after a jump depends on the value of the momentum immediately prior to a jump. The case $\phi = 0$ in (6) has to be excluded because it leads to $\Gamma(q, p) = (q, p)$ and then the generator L reduces to the Liouville operator associated with the Hamiltonian H, which is in general not ergodic with respect to Π_{BG}. Note that, if $\phi = \pi/2$, the random vector $\Gamma(q, p)$ does not depend on p (complete momentum randomization).

3. Geometric ergodicity of RHMC.

3.1. *Overview.* In this section, we prove that Π_{BG} is the unique invariant probability measure of RHMC and that RHMC is geometrically ergodic. Our main tool to prove geometric ergodicity is Harris's theorem. In the Appendix, we recall this theorem in the present context of a continuous-time Markov process with an uncountable state space.

A main ingredient in Harris's theorem is Hypothesis A.1 on the existence of a Lyapunov function, which we refer to as a *Foster–Lyapunov drift condition*. We formulate this condition in terms of an abstract infinitesimal generator whose precise meaning is given in Definition A.1. Note that this definition of an infinitesimal generator accommodates the Lyapunov function $V : \mathbb{R}^{2D} \to \mathbb{R}$ introduced below because for any $t > 0$ the process

$$V(Z_t) - V(z) - \int_0^t LV(Z_s)\,ds, \qquad Z_0 = z \in \mathbb{R}^{2D},$$

is always a local martingale, and hence, according to Definition A.1 the function V belongs to the domain of L. Note that here we assume that this Lyapunov function is continuously differentiable, since the operator L involves partial derivatives. After showing that V is a Lyapunov function for L, we apply this Lyapunov function to solve the martingale problem for the operator L on functions that are C^1 with globally Lipschitz continuous derivative. This solution is used to show that Π_{BG} is an invariant measure for RHMC. We then prove that the transition probabilities of RHMC satisfy a minorization condition given in Hypothesis A.2. With these pieces in place, we invoke Harris's theorem to conclude that Π_{BG} is the unique invariant measure for RHMC, and that RHMC is geometrically ergodic.

To establish a Foster–Lyapunov drift condition, we use a Lyapunov function V originally introduced to prove stochastic stability of a Hamiltonian system with dissipation and random impulses; see Section 2.2 and equation (14) of [35]. See also Section 3 of [27] and Section 2 of [37] for an application of this Lyapunov function to prove geometric ergodicity of the solution to underdamped Langevin equations with nonglobally Lipschitz coefficients. This Lyapunov function $V : \mathbb{R}^{2D} \to \mathbb{R}$ is of the form

$$(7) \qquad V(z) = H(z) + c_1 \langle q, p \rangle + c_2 \frac{|q|^2}{2}, \qquad z = (q, p) \in \mathbb{R}^{2D},$$

where H is the Hamiltonian given earlier in (3), and c_1 and c_2 are constants given by

$$(8) \qquad c_1 = \frac{\lambda^{-1}}{4} \sin^2(\phi) \quad \text{and} \quad c_2 = \lambda^{-1} c_1 (1 - \cos(\phi)).$$

Since $\lambda > 0$ and $\phi \in (0, \pi/2]$, note from (8) that both c_1 and c_2 are positive.

Throughout this section, we will use the following conditions on the potential energy $\Phi(q)$. Note that not all of these assumptions will be required for every statement, but we find it notationally convenient to have a single set of assumptions to refer to.

HYPOTHESIS 3.1. *The potential energy* $\Phi \in C^2(\mathbb{R}^D)$ *satisfies the following conditions*:

H1. *One has* $\Phi(q) \geq 0$ *and* $\int_{\mathbb{R}^D} |q|^2 \exp(-\Phi(q)) \, dq < \infty$.
H2. *Let* c_1 *and* c_2 *be the constants appearing in* (7). *Then there exist* $a > 0$ *and* $b \in (0, 1)$ *such that*

$$\frac{1}{2} \langle \nabla_q \Phi(q), q \rangle \geq b \Phi(q) + \frac{(c_1 b)^2 + c_2 b(1 - b)}{2(1 - b)} |q|^2 - a$$

for all $q \in \mathbb{R}^D$.

We stress that these assumptions on the potential energy function are typically made to prove geometric ergodicity of the solution to underdamped Langevin equations. For instance, see equation (13) of [35], Hypothesis 1.1 in [37] and Condition 3.1 in [27]. Hypothesis **H1** and (7) imply that for any constants $c_1, c_2 \in \mathbb{R}$, we have that

$$\int_{\mathbb{R}^{2D}} V(z) \Pi_{\mathrm{BG}}(dz) < \infty.$$

In other words, this Lyapunov function is integrable with respect to Π_{BG}.

The hypothesis that the potential is bounded from below by itself guarantees that the kinetic energy and, therefore, the momenta are bounded as time increases. It follows that the configuration variables grow at most linearly with time and, therefore, solutions of Hamilton's equations are well defined for all real time.

3.2. *The measure Π_{BG} is an infinitesimally invariant measure.* As expected, the Boltzmann–Gibbs distribution in (2) is an infinitesimally invariant probability measure for the process Z. By implication the target Π is infinitesimally invariant for the process Q. To state this result, let $C_c^\infty(\mathbb{R}^{2D})$ denote the space of compactly supported smooth functions from $\mathbb{R}^{2D}$ to $\mathbb{R}$.

PROPOSITION 3.1. *Suppose Hypothesis 3.1 **H1** holds. Then for any $f \in C_c^\infty(\mathbb{R}^{2D})$ we have that*

$$\int_{\mathbb{R}^{2D}} Lf(z) \Pi_{\mathrm{BG}}(dz) = 0.$$

PROOF. Hypothesis 3.1 **H1** guarantees that $\int_{\mathbb{R}^{2D}} e^{-H(z)} \, dz < \infty$, and hence, the measure Π_{BG} is a probability measure. Note that Π_{BG} is an invariant probability measure for the momentum randomization operator in (5), and hence,

$$\int_{\mathbb{R}^{2D}} \mathbb{E}\{f(\Gamma(q, p)) - f(q, p)\} \Pi_{\mathrm{BG}}(dq, dp) = 0.$$

Moreover, integration by parts shows that the Liouville operator leaves Π_{BG} infinitesimally invariant. In particular, the boundary terms resulting from the integration by parts vanish because f is compactly supported. $\quad\square$

Later in this section, we strengthen this result to Π_{BG} is the unique invariant probability measure for RHMC.

3.3. *Foster–Lyapunov drift condition.* The following lemma is remarkable because it states that the infinitesimal generator L possesses a Lyapunov function, even though RHMC does not incorporate explicit dissipation.

LEMMA 3.2. *Suppose Hypothesis 3.1 holds, and $V(z)$ is given by (7). Then there exist real positive constants c_1, c_2, γ and K such that*

$$(9) \qquad LV(z) \leq -\gamma V(z) + K$$

for all $z \in \mathbb{R}^{2D}$. Moreover, V is nonnegative and

$$\lim_{|z| \to \infty} V(z) = \infty.$$

This lemma implies that Hypothesis A.1 of the Harris's theorem holds. The proof below shows that the momentum randomizations in RHMC are the source of dissipation in RHMC.

PROOF. Let $z = (q, p) \in \mathbb{R}^{2D}$. From (7), note that

$$\nabla_q V = \nabla_q \Phi + c_1 p + c_2 q,$$

$$\nabla_p V = p + c_1 q$$

and if we set $u = \cos(\phi)p + \sin(\phi)\xi$, note from (6) that

$$\mathbb{E}|u|^2 = D\sin^2(\phi) + \cos^2(\phi)|p|^2,$$

$$\mathbb{E}\langle q, u \rangle = \cos(\phi)\langle q, p \rangle,$$

where the expected value is taken over the D-dimensional standard normal vector $\xi \sim \mathcal{N}(0, 1)^D$. We recall that we excluded the case $\phi = 0$ in the definition of u in (6). A direct calculation shows that

$$(10) \qquad \begin{aligned} LV &= \lambda^{-1} D \frac{1}{2} \sin^2(\phi) - c_1 \langle \nabla_q \Phi, q \rangle \\ &\quad + \left(2c_1 - \lambda^{-1} \sin^2(\phi)\right) \frac{|p|^2}{2} + \left(\lambda^{-1} c_1 (\cos(\phi) - 1) + c_2\right) \langle q, p \rangle. \end{aligned}$$

We now choose c_1 and c_2 such that

$$\begin{cases} 2c_1 - \lambda^{-1} \sin^2(\phi) = -2c_1, \\ \lambda^{-1} c_1 (\cos(\phi) - 1) + c_2 = 0. \end{cases}$$

In other words, we pick c_1 and c_2 to eliminate the $\langle q, p \rangle$ cross term in LV, and to rescale the $(|p|^2)/2$ term so that it matches the coefficient of the $\langle \nabla_q \Phi, q \rangle$ term. Solving these equations yields (8). With this choice of c_1 and c_2, (10) simplifies to

$$(11) \qquad LV = -2c_1 \left(\frac{|p|^2}{2} + \frac{1}{2} \langle \nabla_q \Phi, q \rangle \right) + D\lambda^{-1} \frac{1}{2} \sin^2(\phi).$$

Let b be the constant appearing in Hypothesis 3.1 **H2**. Applying Cauchy's inequality with $\delta > 0$ to the $\langle q, p \rangle$ cross term in (7) yields

$$bV(q, p) \leq \frac{b}{2}|p|^2 + b\Phi(q) + c_1 b\left(\frac{\delta}{2}|p|^2 + \frac{1}{2\delta}|q|^2\right) + \frac{c_2 b}{2}|q|^2.$$

Choosing $\delta = (1 - b)/(c_1 b)$ and invoking Hypothesis 3.1 **H2**, we obtain

(12)
$$bV(q, p) \leq \frac{|p|^2}{2} + b\Phi(q) + \frac{(c_1 b)^2 + c_2 b(1 - b)}{2(1 - b)} |q|^2$$

$$\leq \frac{|p|^2}{2} + \frac{1}{2}\langle\nabla\Phi(q), q\rangle + a.$$

Together (11) and (12) imply that the desired Foster–Lyapunov drift condition holds with $\gamma = 2c_1 b$ and for some $K > 0$.

To complete the proof, recall that $\Phi \geq 0$ by Hypothesis 3.1 **H1**, and thus, it suffices to check that the quadratic form

$$\frac{|p|^2}{2} + c_1\langle q, p\rangle + c_2\frac{|q|^2}{2}$$

appearing in $V(q, p)$ is positive definite. This condition is met when

$$c_2 > c_1^2 > 0 \quad \Longrightarrow \quad (1 - \cos(\phi)) > \frac{1}{4}\sin^2(\phi) > 0$$

which holds for all $\phi \in (0, \pi/2]$. $\quad\square$

REMARK 3.3. The proof of Lemma 3.2 shows that $\gamma \propto \lambda^{-1}\sin^2(\phi)$. Thus, we see that if λ is smaller or ϕ is closer to $\pi/2$, then γ becomes larger. This result is expected since momentum randomizations are the source of dissipation in RHMC, and smaller λ implies more randomizations of momentum, and larger ϕ implies the momentum is randomized more completely.

REMARK 3.4. It follows from Lemma 3.2 that

$$\mathbb{E}_z V(Z_t) \leq e^{-\gamma t}V(z) + \frac{K}{\gamma}(1 - e^{-\gamma t}).$$

See, for example, the proof of Theorem 6.1 in [29].

3.4. *Martingale problem.* Here, we use Lemma 3.2 to solve the martingale problem for the operator L on functions that are C^1 with globally Lipschitz continuous derivative.

LEMMA 3.5. *Suppose Hypothesis* 3.1 *holds. For all globally Lipschitz and continuously differentiable functions* $f : \mathbb{R}^{2D} \to \mathbb{R}$, *for any initial condition* $Z_0 = z \in \mathbb{R}^{2D}$, *and for any* $t > 0$, *the local martingale*

$$M_t^f = f(Z_t) - f(z) - \int_0^t Lf(Z_s)\,ds$$

is a martingale.

A key ingredient in the proof given below is the Lyapunov function for L from Lemma 3.2. In particular, since globally Lipschitz functions grow at most linearly, the class of functions that appear in Lemma 3.5 are bounded by this Lyapunov function. Moreover, Dynkin's formula holds for this class of functions:

$$(13) \qquad \mathbb{E}_z f(Z_t) = f(z) + \int_0^t \mathbb{E}_z Lf(Z_s)\,ds, \qquad Z_0 = z, t \geq 0.$$

See, for example, Chapter 1 of [9] for more discussion on Dynkin's formula.

PROOF. In this proof, we use the well-known fact (see, e.g., Corollary 3, Chapter II in [32]) that a local martingale with integrable quadratic variation is a martingale. Let $[M^f](t)$ denote the quadratic variation of M_t^f on the interval $[0, t]$. The global Lipschitz assumption on f implies that there exists a constant $L_f > 0$ such that

$$\left| f(z_1) - f(z_2) \right| \leq L_f |z_1 - z_2|$$

for all $z_1, z_2 \in \mathbb{R}^{2D}$. Moreover, since the process Z_t satisfies Hamilton's differential equations in between consecutive momentum randomizations, and since the process $f(Z_t) - M_t^f$ is continuous and of finite variation, the quadratic variation of M^f is equal to the sum of the squares of the jumps in $f(Z_t)$. Thus,

$$[M^f](t) = [f](t) = \sum_{0 < s \leq t} \left(f(Z_s) - f(Z_{s-}) \right)^2 \leq \sum_{0 < s \leq t} L_f^2 |Z_s - Z_{s-}|^2.$$

In other words, the global Lipschitz property of f enables bounding the quadratic variation of the scalar-valued process $f(Z_t)$ by the quadratic variation of the components of the process Z_t.

Let $\{t_i\}$ be the sequence of random times at which the momentum randomizations occur. This sequence can be produced by iterating the recurrence relation: $t_{i+1} = t_i + \delta t_i$ with initial condition $t_0 = 0$ and where $\{\delta t_i\}$ are i.i.d. exponential random variables with mean λ. Let N_t be the (random) number of momentum randomizations that have occurred up to time t. Note that N_t is a Poisson process with intensity λ^{-1}, and hence, a.s. bounded. This permits us to interchange expectation and summation in the following inequalities:

$$\mathbb{E}_z[M^f](t) \leq L_f^2 \mathbb{E}_z \sum_{1 \leq i \leq N_t} |Z_{t_i} - Z_{t_i-}|^2$$

$$\leq L_f^2 \lambda^{-1} \sum_{i>0} \mathbb{E}_z \left\{ |P_{t_i} - P_{t_i-}|^2 (t \wedge t_{i+1} - t \wedge t_i) \right\}$$

$$\leq 2L_f^2 \lambda^{-1} \sum_{i>0} \mathbb{E}_z \left\{ (|P_{t_i}|^2 + |P_{t_i-}|^2)(t \wedge t_{i+1} - t \wedge t_i) \right\}$$

$$\leq 4L_f^2 C_2 \lambda^{-1} \sum_{i>0} \mathbb{E}_z \left\{ (V(Z_{t_i}) + V(Z_{t_i-}))(t \wedge t_{i+1} - t \wedge t_i) \right\}$$

$$\leq 8L_f^2 C_2 \lambda^{-1} t \left(V(z) + \frac{K}{\gamma} \right),$$

where in the last two steps we used the Lyapunov function given in Lemma 3.2, and in addition, we introduced the positive constant $C_2 = 2c_2/(c_2 - c_1^2)$ with c_1 and c_2 being the constants defined in (8). Thus, we may conclude that M_t^f is a martingale for any $t > 0$.

Alternatively, we could have used the compensator of $[M^f](t)$:

$$\langle M^f \rangle(t) = \int_0^t \left(Lf^2(Z_s) - 2f(Z_s)Lf(Z_s) \right) ds$$

$$= \lambda^{-1} \int_0^t \mathbb{E}\big(f\big(\Gamma(Z_s)\big) - f(Z_s) \big)^2 ds$$

which would give a similar bound on $\mathbb{E}_z[M^f](t)$. $\quad\square$

3.5. *The measure Π_{BG} is an invariant measure.* In this part, we combine Proposition 3.1 and Lemma 3.5 to prove that Π_{BG} is an invariant probability measure for RHMC.

LEMMA 3.6. *Suppose Hypothesis 3.1 holds. For any $f \in C_c^\infty(\mathbb{R}^{2D})$ and for any $t > 0$,*

$$\int_{\mathbb{R}^{2D}} \mathbb{E}_z f(Z_t) \Pi_{\mathrm{BG}}(dz) = \int_{\mathbb{R}^{2D}} f(z) \Pi_{\mathrm{BG}}(dz).$$

To prove this lemma, we use Dynkin's formula and condition on a fixed sequence of jump times to exploit the fact that the Hamiltonian flow and the momentum randomization individually leave Π_{BG} invariant.

PROOF. Let $f \in C_c^\infty(\mathbb{R}^{2D})$ and $z = (q, p) \in \mathbb{R}^{2D}$. Referring to (5), since $f \in C_c^\infty(\mathbb{R}^{2D})$ the function

$$\mathbb{E} f\big(\Gamma(z)\big) = \mathbb{E} f\big(q, \cos(\phi)p + \sin(\phi)\xi\big)$$

is smooth, compactly supported in the q component, and bounded in the p component, and hence, $Lf \in C_b^\infty(\mathbb{R}^{2D})$. Moreover, since any smooth function with compact support is globally Lipschitz continuous, we can invoke Lemma 3.5 to conclude that for any $f \in C_c^\infty(\mathbb{R}^{2D})$ and for any $t > 0$, the process

$$f(Z_t) - f(z) - \int_0^t Lf(Z_s)\, ds, \qquad Z_0 = z,$$

is a martingale. Thus, Dynkin's formula holds:

$$\mathbb{E}_z f(Z_t) = f(z) + \int_0^t \mathbb{E}_z Lf(Z_s)\, ds,$$

and in particular,

$$
\int_{\mathbb{R}^{2D}} \mathbb{E}_z f(Z_t) \Pi_{\mathrm{BG}}(dz) = \int_{\mathbb{R}^{2D}} f(z) \Pi_{\mathrm{BG}}(dz)
$$

(14)

$$
+ \int_0^t \left(\int_{\mathbb{R}^{2D}} \mathbb{E}_z L f(Z_s) \Pi_{\mathrm{BG}}(dz) \right) ds,
$$

where we used Fubini's theorem to interchange time and space integrals. This interchange (and subsequent ones) are valid since the function Lf is bounded by the Lyapunov function, which is an integrable function under the measure Π_{BG}. We next argue that the second term on the right-hand side of (14) vanishes due to Lemma 3.1 and some basic properties of Hamiltonian flows (volume and Hamiltonian preserving) and the momentum randomization map (Boltzmann–Gibbs preserving).

For this purpose, and as in Lemma 3.2, let $\{t_i\}$ denote a realization of the sequence of random times at which the momentum randomizations occur. For any $t > 0$, let $\vartheta_t : \mathbb{R}^{2D} \to \mathbb{R}^{2D}$ be the Hamiltonian flow map associated to H. We recall that the jump times and momentum randomizations are mutually independent, and that the number of jumps in any finite time interval is a.s. finite. By change of variables, and using the volume and Hamiltonian preserving properties of the Hamiltonian flow ϑ_s, note that

$$
\int_{\mathbb{R}^{2D}} Lf(\vartheta_s(z)) \Pi_{\mathrm{BG}}(dz) = \int_{\mathbb{R}^{2D}} Lf(z) \Pi_{\mathrm{BG}}(dz)
\tag{15}
$$

holds for any $s \in [0, t]$. In addition, since Π_{BG} is an invariant measure for the momentum randomization map and $Lf \in C_b^\infty(\mathbb{R}^{2D})$, we have the identity:

$$
\int_{\mathbb{R}^{2D}} \mathbb{E} Lf(\Gamma(z)) \Pi_{\mathrm{BG}}(dz) = \int_{\mathbb{R}^{2D}} Lf(z) \Pi_{\mathrm{BG}}(dz).
\tag{16}
$$

These facts motivate us to decompose the process Z_s into its Hamiltonian and momentum randomization pieces. To do this, and with a slight abuse of notation, we regard the process $Z_s : \mathbb{R}^{2D} \to \mathbb{R}^{2D}$ as an evolution operator and decompose it via

$$
Z_s(z) = \begin{cases} \vartheta_{s-t_k}(Z_{t_k}), & \text{if } t_k \le s < t_{k+1}, \\ \Gamma(\vartheta_{t_{k+1}-t_k}(Z_{t_k})), & \text{if } s = t_{k+1} \end{cases}
$$

for all $s \ge 0$. To leverage this decomposition, we use $\{t_i\}$ to split the time integral appearing in the second term of the right-hand side of (14) into time intervals between consecutive momentum randomizations:

$$
\mathbb{E} \int_0^t Lf(Z_s(z)) ds = \mathbb{E} \left\{ \sum_{k \ge 0} \int_{t_k \wedge t}^{t_{k+1} \wedge t} Lf(\theta_s \circ \Gamma \circ \theta_{\delta t_{k-1}} \circ \cdots \circ \theta_{\delta t_0}(z)) ds \right\}.
$$

In this form, we can take advantage of the independence between momentum randomizations and jump times, in order to simplify this term. In particular, we condition on the jump times and then average over individual momentum randomizations to obtain

$$\mathbb{E}\int_0^t \int_{\mathbb{R}^{2D}} Lf\big(Z_s(z)\big)\Pi_{\mathrm{BG}}(dz)\,ds$$

$$= \mathbb{E}\left\{\sum_{k\geq 0}\int_{t_k\wedge t}^{t_{k+1}\wedge t}\int_{\mathbb{R}^{2D}} Lf\big(\theta_s\circ\Gamma\circ\theta_{\delta t_{k-1}}\circ\cdots\circ\theta_{\delta t_0}(z)\big)\Pi_{\mathrm{BG}}(dz)\,ds\right\}$$

$$= \mathbb{E}\left\{\sum_{k\geq 0}\int_{t_k\wedge t}^{t_{k+1}\wedge t}\int_{\mathbb{R}^{2D}} Lf\big(\theta_s\circ\Gamma\circ\theta_{\delta t_{k-1}}\circ\cdots\circ\theta_{\delta t_1}\circ\Gamma(z)\big)\Pi_{\mathrm{BG}}(dz)\,ds\right\}$$

$$= \mathbb{E}\left\{\sum_{k\geq 0}\int_{t_k\wedge t}^{t_{k+1}\wedge t}\int_{\mathbb{R}^{2D}} Lf\big(\theta_s\circ\Gamma\circ\theta_{\delta t_{k-1}}\circ\cdots\circ\theta_{\delta t_1}(z)\big)\Pi_{\mathrm{BG}}(dz)\,ds\right\}$$

$$= \cdots = \cdots = \int_0^t\int_{\mathbb{R}^{2D}} Lf(z)\Pi_{\mathrm{BG}}(dz)\,ds,$$

where we sequentially used (15) and (16) from initial time 0 up to final time s. Note that to use (16) one has to average over the Gaussian random vector associated to the ith momentum randomization for $1\leq i\leq k$ in the inner-most expectation. Substituting this result back into (14), we obtain

$$\int_{\mathbb{R}^{2D}}\mathbb{E}_z f(Z_t)\Pi_{\mathrm{BG}}(dz) = \int_{\mathbb{R}^{2D}} f(z)\Pi_{\mathrm{BG}}(dz) + t\int_{\mathbb{R}^{2D}} Lf(z)\Pi_{\mathrm{BG}}(dz).$$

To finish, we invoke Lemma 3.1, which implies that Π_{BG} is infinitesimally invariant for L, and hence, the second term on the right-hand side of this equation vanishes, as required. □

3.6. *Minorization condition.* For $t\geq 0$, let P_t denote the transition semigroup of the Markov process Z_t

$$P_t f(z) = \mathbb{E}f\big(Z_t(z)\big), \qquad Z_0(z) = z$$

and let $\Pi_{t,z}$ denote the associated transition probability distribution

$$P_t f(z) = \int_{\mathbb{R}^{2D}} f(w)\Pi_{t,z}(dw).$$

Recall that the process Z_t only moves by either the Hamiltonian flow for a random duration or momentum randomizations that are instantaneous. Thus, we expect the semigroup to not have the strong Feller property [10], since it lacks a sufficient regularizing effect. Nevertheless, we can prove a minorization condition for this process by using the weaker regularizing effect of the momentum randomizations.

PROPOSITION 3.7. *Suppose Hypothesis* 3.1 *holds. For every compact set* $\Omega \subset \mathbb{R}^{2D}$, *there exist a probability measure* η *over* Ω, $\varepsilon > 0$ *and* $t > 0$ *such that*

$$\Pi_{t,z}(\cdot) \geq \varepsilon \eta(\cdot)$$

holds for all $z \in \Omega$.

To prove this proposition, it is convenient to introduce the following operator:

$$\tag{17} \mathcal{A}f(z) = \mathbb{E}f(\Gamma(z)) = \int_{\mathbb{R}^D} f(q, \eta) g_\phi(p, \eta) \, d\eta,$$

where $g_\phi : \mathbb{R}^{2D} \to \mathbb{R}^+$ is defined as

$$\tag{18} g_\phi(p, \eta) = (2\pi \sin^2 \phi)^{-D/2} \exp\left(-\frac{|\eta - p\cos\phi|^2}{2\sin^2\phi}\right).$$

Note that the operator $\mathcal{A}$ appears in the infinitesimal generator in (5). For any $t \geq 0$, let $\theta_t : \mathbb{R}^{2D} \to \mathbb{R}^{2D}$ denote the Hamiltonian flow associated with (3) and $\mathcal{C}_{\theta_t}$ denote the composition operator for θ_t defined as

$$\mathcal{C}_{\theta_t} f(z) = f(\theta_t(z)).$$

This Hamiltonian flow can be characterized by a generating function $S_t(q_0, q_1)$ [24, 25, 34]. Specifically, if $|q_1 - q_0|$ and $t > 0$ are sufficiently small, then $(q_1, p_1) = \theta_t(q_0, p_0)$ satisfy the following system of equations:

$$\tag{19} \begin{aligned} p_0 &= -D_1 S_t(q_0, q_1), \\ p_1 &= D_2 S_t(q_0, q_1). \end{aligned}$$

Here, D_i denotes the derivative with respect to the ith component of S_t. Moreover, the generating function can be written as

$$\tag{20} S_t(q_0, q_1) = \int_0^t \left(\frac{1}{2}|\dot{q}(s)|^2 - \Phi(q(s))\right) ds,$$

where $q : [0, t] \to \mathbb{R}^D$ solves the Euler–Lagrangian equations $\ddot{q} = -\nabla\Phi(q)$ with boundary conditions $q(0) = q_0$ and $q(t) = q_1$. In discrete mechanics, the equations (19) and the generating function S_t are known as discrete Hamilton's equations and the (exact) discrete Lagrangian, respectively [25].

PROOF. We adapt to our setting some ideas from the proof of Theorem 2.3 in [12]. To establish the desired result, we use the (weak) regularizing effect of the operator $\mathcal{A}$ on a function f in the momentum degrees of freedom. Since the Hamiltonian flow is regular [24], this regularizing effect can be transferred to the position degrees of freedom of f. Similar results appear in Lemma 2.2 of [12] and the proof of Theorem 2.2 of [6].

Specifically, a change of variables shows that

$$(21) \qquad (\mathcal{A}\mathcal{C}_{\theta_t}\mathcal{A}f)(q,p) = \int_{\mathbb{R}^D} \int_{\mathbb{R}^D} f(q_1, p_1) g_{t,\phi}((q,p),(q_1,p_1)) \, dq_1 \, dp_1,$$

where we have introduced the transition density of the operator $\mathcal{A}\mathcal{C}_{\theta_t}\mathcal{A}$:

$$g_{t,\phi}((q,p),(q_1,p_1))$$
$$= |\det D_{12} S_t(q,q_1)| g_\phi(p, -D_1 S_t(q,q_1)) g_\phi(D_2 S_t(q,q_1), p_1)$$

in terms of g_ϕ in (18) and S_t in (20).

To take advantage of (21), we consider the following Duhamel formula:

$$e^{\lambda^{-1}(s-t)} \mathcal{C}_{\theta_{t-s}} P_s f(z) \big|_{s=0}^{s=t} = P_t f(z) - e^{-\lambda^{-1}t} \mathcal{C}_{\theta_t} f(z)$$
$$= \lambda^{-1} \int_0^t e^{\lambda^{-1}(s-t)} \mathcal{C}_{\theta_{t-s}} \mathcal{A} P_s f(z) \, ds.$$

A second application of this Duhamel formula implies that

$$P_t f(z) = e^{-\lambda^{-1}t} \mathcal{C}_{\theta_t} f(z) + \lambda^{-1} \int_0^t e^{-\lambda^{-1}t} \mathcal{C}_{\theta_{t-t_1}} \mathcal{A} \mathcal{C}_{\theta_{t_1}} f(z) \, dt_1$$
$$+ \lambda^{-2} \int_0^t \int_0^{t_1} e^{\lambda^{-1}(t_1-t)} e^{\lambda^{-1}(t_2-t_1)} \mathcal{C}_{\theta_{t-t_1}} \mathcal{A} \mathcal{C}_{\theta_{t_1-t_2}} \mathcal{A} P_{t_2} f(z) \, dt_2 \, dt_1.$$

A third application of this formula yields

$$(22) \qquad P_t f(z) \geq \lambda^{-2} \int_0^t \int_0^{t_1} e^{\lambda^{-1}(t_2-t)} \mathcal{C}_{\theta_{t-t_1}} \mathcal{A} \mathcal{C}_{\theta_{t_1-t_2}} \mathcal{A} \mathcal{C}_{\theta_{t_2}} f(z) \, dt_2 \, dt_1$$
$$\geq \lambda^{-2} \int_{\varepsilon t}^t \int_0^{t_1-\varepsilon t_1} e^{\lambda^{-1}(t_2-t)} \mathcal{C}_{\theta_{t-t_1}} \underbrace{\mathcal{A} \mathcal{C}_{\theta_{t_1-t_2}} \mathcal{A}} \mathcal{C}_{\theta_{t_2}} f(z) \, dt_2 \, dt_1$$

for ε sufficiently small. Since $\varepsilon t \leq t_1 \leq t$ and $0 \leq t_2 \leq t_1 - \varepsilon t_1$, we have that $\varepsilon^2 t \leq t_1 - t_2 \leq t$. Combining this result with (21) applied to the bracketed term in (22), yields the desired result. $\square$

Next, we show that Hypothesis A.2 of Harris's theorem holds for the transition probabilities of the RHMC method. To state this proposition, we will use the total variation (TV) distance between measures. Recall that

$$(23) \qquad \|\mu - \nu\|_{\mathrm{TV}} = 2 \sup_A |\mu(A) - \nu(A)|,$$

where the supremum runs over all measurable sets. In particular, the total variation distance between two probability measures is two if and only if they are mutually singular.

LEMMA 3.8. *Suppose Hypothesis* 3.1 *holds and let* $V(z)$ *be the Lyapunov function from Lemma* 3.2. *For every* $E > 0$, *there exist* $t > 0$ *and* $\varepsilon > 0$ *such that*

$$\left\| \Pi_t(z_1, \cdot) - \Pi_t(z_2, \cdot) \right\|_{\mathrm{TV}} \le 2(1 - \varepsilon)$$

holds for all $z_1, z_2 \in \mathbb{R}^{2D}$ *satisfying* $V(z_1) \vee V(z_2) < E$.

PROOF. Proposition 3.7 implies the following transition probability $\tilde{\Pi}_{t,z}$ is well defined:

$$\tilde{\Pi}_{t,z}(\cdot) = \frac{1}{1 - \varepsilon} \Pi_{t,z}(\cdot) - \frac{\varepsilon}{1 - \varepsilon} \eta(\cdot)$$

for any z satisfying $V(z) < E$. Therefore,

$$\left\| \Pi_{t,z_1}(\cdot) - \Pi_{t,z_2}(\cdot) \right\|_{\mathrm{TV}} = (1 - \varepsilon) \left\| \tilde{\Pi}_{t,z_1}(\cdot) - \tilde{\Pi}_{t,z_2}(\cdot) \right\|_{\mathrm{TV}}$$

for all z_1, z_2 satisfying $V(z_1) \vee V(z_2) < E$. Since the TV norm is bounded by 2, one obtains the desired result. $\square$

3.7. *Main result*: *Geometric ergodicity.* With a Lyapunov function and minorization condition in hand, we are now in position to state a main result of this paper.

THEOREM 3.9. *Suppose Hypothesis* 3.1 *holds and let* $V(z)$ *be the Lyapunov function from Lemma* 3.2. *Then the Markov process induced by* L *has a unique invariant probability measure given by* Π_{BG}. *Furthermore, there exist positive constants* r *and* C *such that*

$$(24) \qquad \left\| \Pi_{t,z} - \Pi_{\mathrm{BG}} \right\|_{\mathrm{TV}} \le C V(z) e^{-rt}$$

for all $t \ge 0$ *and all* $z \in \mathbb{R}^{2D}$.

PROOF. According to Lemma 3.2, the generator L satisfies a Foster–Lyapunov drift condition. Moreover, its associated transition probabilities satisfy Lemma 3.8. Hence, Theorem A.1 implies that (i) the process possesses a unique invariant distribution, and (ii) the transition probability of the process converges to this invariant distribution geometrically fast in the TV metric. Since, by Proposition 3.1, Π_{BG} is an infinitesimally invariant measure for the process, and that compactly supported functions are in the domain of L, it follows that Π_{BG} is the unique invariant measure for the process. $\square$

4. Model problem. For simplicity, we assume in this section and the next that the Horowitz angle is chosen as $\phi = \pi/2$, that is, the momentum randomizations are complete rather than partial.

We quantify the sampling capabilities of the RHMC method given in Algorithm 2.1 on a model problem in which the target is a multivariate Gaussian distribution with uncorrelated components, some of them with small variances. We also compare against the standard HMC method given in Algorithm 1.1, where the duration is a fixed parameter. This model problem is discussed in [31] and analyzed further in [5]. This distribution can be interpreted as a truncation of an infinite-dimensional normal distribution on a Hilbert space [4]. The potential energy function is given by

$$(25) \qquad \Phi(q_1, \ldots, q_D) = \sum_{i=1}^{D} \frac{1}{2\sigma_i^2} q_i^2,$$

where σ_i^2 is the variance in the ith component.

4.1. *The process.* For (25), a sample trajectory of the RHMC method satisfies

$$(26) \qquad q_i(t_{n+1}) = \cos\left(\frac{\delta t_n}{\sigma_i}\right) q_i(t_n) + \sigma_i \sin\left(\frac{\delta t_n}{\sigma_i}\right) \xi_{i,n}$$

with $q_i(0)$ given and where $\{t_n\}$ are random jump times related by

$$t_n = t_{n-1} + \delta t_{n-1},$$

with $t_0 = 0$. Here, we have introduced the following sequences of random variables:

$$\{\xi_{i,n}\} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1), \qquad \{\delta t_n\} \overset{\text{i.i.d.}}{\sim} \text{Exp}\left(\frac{1}{\lambda}\right),$$

where i (resp., n) runs over the components of the target distribution (resp., jump times).

The solution of the stochastic difference equation (26) is given by

$$(27) \qquad q_i(t_n) = \prod_{j=0}^{n-1} \cos\left(\frac{\delta t_j}{\sigma_i}\right) q_i(0) + \sum_{j=0}^{n-1} \sigma_i \sin\left(\frac{\delta t_j}{\sigma_i}\right) \prod_{k=j+1}^{n-1} \cos\left(\frac{\delta t_k}{\sigma_i}\right) \xi_{i,j}.$$

(We adhere to the standard convention that $\prod_{k=m}^{n} \cdot$ takes the value 1 if $m > n$.) At steady-state the ith component of this process is normally distributed with mean zero and variance σ_i^2, that is,

$$q_i(t) \overset{\text{d}}{\to} \mathcal{N}(0, \sigma_i^2) \qquad \text{as } t \to \infty.$$

The corresponding solution for HMC is given by formula (27) with constant, as opposed to random, durations δt_n. Note that, for constant duration, formula (26)

makes apparent that, at stationarity, the correlation between $q_i(t_{n+1})$ and $q_i(t_n)$ is $\cos(\delta t_n / \sigma_i)$. When λ is an even integer multiple of $\pi \sigma_i$, $q_i(t_{n+1}) = q_i(t_n)$; for odd multiples, $q_i(t_{n+1}) = -q_i(t_n)$. For those value of λ, the chain is not ergodic; for values of λ close to an integer multiple of $\pi \sigma_i$, the performance of the chain may be expected to be poor.

4.2. *Integrated Autocorrelation Time* (*IAC*). The first measure of the quality of the samples provided by the algorithm, we consider is the integrated autocorrelation time (IAC) associated with estimating the mean of the target distribution. The natural estimator for this purpose is

$$\hat{f}_{i,N} = \frac{1}{N+1} \sum_{j=0}^{N} q_i(t_j).$$

As shown in, for example, Chapter IV of [3], asymptotically as $n \to \infty$, the variance of the estimator behaves as $\hat{\sigma}_i^2 / n$, where the constant $\hat{\sigma}_i^2$ is called the asymptotic variance. This may be computed by means of the formula

$$(28) \qquad \hat{\sigma}_i^2 = \mathrm{Var}(q_i(0)) + 2 \sum_{j=1}^{\infty} \mathrm{cov}(q_i(0), q_i(t_j)).$$

This holds for any random variables $q_i(t_j)$ for which both $\mathbb{E}(q_i(t_j))$ and $\mathbb{E}(q_i(t_j)q_i(t_{j+k}))$ are independent of j [it is not necessary, in particular, that the $q_i(t_j)$ originate from a Markov chain], provided that the series converges absolutely. If successive samples are independent, then the series vanishes and the asymptotic variance equals the variance σ_i^2 of $q_i(0)$. The integrated autocorrelation time (IAC) of the estimator $\hat{f}_{i,N}$ is defined as the ratio $\hat{\sigma}_i^2 / \sigma_i^2$. It follows from (27) that, at stationarity of the chain,

$$\mathrm{cov}(q_i(0), q_i(t_j)) = \mathbb{E}(q_i(0)q_i(t_j))$$

$$= \sigma_i^2 \mathbb{E} \prod_{k=0}^{j-1} \cos\left(\frac{\delta t_k}{\sigma_i}\right)$$

$$= \sigma_i^2 \prod_{k=0}^{j-1} \mathbb{E} \cos\left(\frac{\delta t_k}{\sigma_i}\right) = \sigma_i^2 \left(\frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}\right)^j$$

and hence,

$$(29) \qquad \mathrm{IAC}_i^{\mathrm{RHMC}} = 1 + 2\frac{\sigma_i^2}{\lambda^2}.$$

Thus, the IAC in each component monotonically decreases with increasing λ and then plateaus at unity.

A very similar calculation shows that the corresponding formula for the HMC method is given by

$$(30) \qquad \mathrm{IAC}_i^{\mathrm{HMC}} = \frac{1 + \cos(\frac{\lambda}{\sigma_i})}{1 - \cos(\frac{\lambda}{\sigma_i})}.$$

Unlike (29), the IAC for the HMC method is an oscillatory function of λ: as discussed in the Introduction it is possible that increasing λ (and, therefore, increasing computational costs) results in higher correlation.

It is useful to discuss the following choices of λ:

- λ is an even multiple of $\pi \sigma_i$ leading to $q_i(t_{n+1}) = q_i(t_n)$. The chain is not ergodic and formula (28) is not valid, due to the divergence of the series. Note however that the estimator yields the value $q_i(0)$; when $q(0)$ is taken from the stationary distribution, the variance of the estimator is therefore σ_i^2, regardless of the number of samples. Thus, the asymptotic variance is not finite, which agrees with (30), even though this formula was derived from (28), invalid in this case.

- λ is an even multiple of $\pi \sigma_i$, with $q_i(t_{n+1}) = -q_i(t_n)$. The chain is not ergodic and formula (28) is not valid, due to the divergence of the series. However, for the particular statistics being considered, the estimation is exceptionally good. If the number of samples is even, the estimator gives the exact value 0, and, for an odd number, the estimator yields the very accurate value $q_i(0)/(N+1)$ [with an $\mathcal{O}(N^{-1})$, rather than $\mathcal{O}(N^{-1/2})$ error]. Therefore, the asymptotic variance vanishes, in agreement with (30).

4.3. *Mean Squared Displacement* (*MSD*).   As another metric for the quality of sampling, we consider the single-step, equilibrium mean-squared displacement of the RHMC and HMC methods. This statistic quantifies how far apart successive samples are. A direct calculation using (26) shows that

$$\begin{aligned}
\mathrm{MSD}^{\mathrm{RHMC}} &= \sum_{i=1}^{D} \mathbb{E}|q_i(t_1) - q_i(0)|^2 \\
&= \sum_{i=1}^{D} \mathbb{E}\left\{ \left| \left(\cos\left(\frac{t_1}{\sigma_i}\right) - 1 \right) q_i(0) + \sigma_i \sin\left(\frac{t_1}{\sigma_i}\right) \right|^2 \right\} \\
&= \sum_{i=1}^{D} \mathbb{E}\left\{ \left(\cos\left(\frac{t_1}{\sigma_i}\right) - 1 \right)^2 + \sin^2\left(\frac{t_1}{\sigma_i}\right) \right\} \sigma_i^2 \\
&= \sum_{i=1}^{D} \mathbb{E}\left\{ 2 - 2\cos\left(\frac{t_1}{\sigma_i}\right) \right\} \sigma_i^2,
\end{aligned}$$

which implies that

$$(31) \qquad \text{MSD}^{\text{RHMC}} = \sum_{i=1}^{D} \frac{2\lambda^2 \sigma_i^2}{\sigma_i^2 + \lambda^2}.$$

Note that MSD^{RHMC} monotonically increases with increasing λ and then plateaus at $\sum_{i=1}^{D} 2\sigma_i^2$. Since in the present scenario of small time steps, the computational cost of the algorithm is proportional to λ, the function $\text{MSD}^{\text{RHMC}}/\lambda$ may be taken as a measure of the efficiency of the algorithm. For λ small, this function increases with λ; it reaches a maximum at

$$(32) \qquad \lambda_{\max} = \left(\frac{\sum_{i=1}^{D} \sigma_i^4}{\sum_{i=1}^{D} \sigma_i^2} \right)^{1/2}$$

and then decreases. The quantity (32) is approximately $\max_{1 \leq i \leq D} \sigma_i$ and the conclusion is that, from this point of view, the best value of λ coincides with the standard deviation of the least constrained variate. Taking λ above the optimal value will not decrease the mean square displacement (as distinct from the situation for HMC we discuss next), but will waste computational resources.

A similar calculation shows that the single-step, equilibrium mean-squared displacement of the HMC is given by

$$(33) \qquad \text{MSD}^{\text{HMC}} = \sum_{i=1}^{D} 2 \left(1 - \cos\left(\frac{\lambda}{\sigma_i} \right) \right) \sigma_i^2$$

which is an oscillatory function of λ.

5. Numerical testing.

5.1. *Standard normal distribution.* We first numerically illustrate the analysis in Section 4 assuming $D = 1$ and unit variance.

Figure 1 corresponds to IAC. In the left panel (HMC), for λ close to π or 3π there is a clear discrepancy between the function in (30) and the empirical estimate based on a single trajectory: due to the divergence of the series (28) the software package used to measure IAC does not work satisfactorily. The right panel corresponds to RHMC. Note the horizontal asymptote at unity, which is consistent with (29).

Figure 2 shows results for MSD. On the left (HMC), the lack of ergodicity at the values π, 3π entails, in spite of the large number (10^6) of samples, discrepancies between the value at stationarity in (33) and the empirical value along the trajectory considered. We observe the monotonic behavior on the right (RHMC) with a horizontal asymptote at 2.
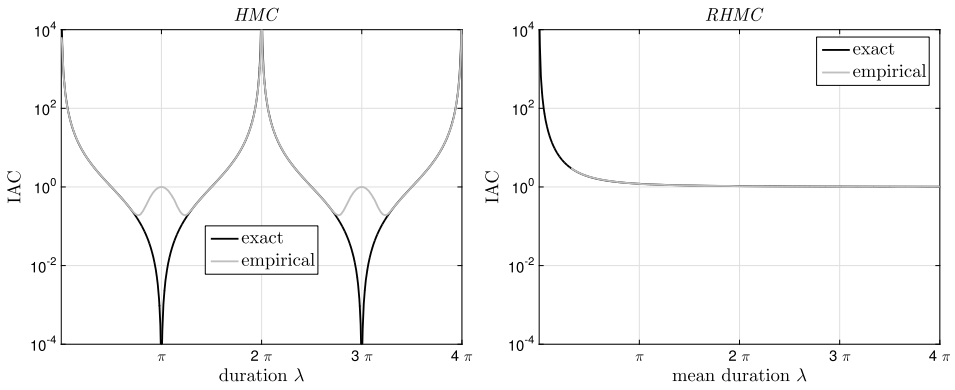
FIG. 1. *IAC for Normal Distribution. The left (resp., right) panel of this figure plots* IAC^HMC *(resp.,* IAC^RHMC*) vs. duration (resp., mean duration) λ for the HMC (resp., RHMC) method applied to a standard normal distribution. The black lines in the left (resp., right) panel plot* IAC^HMC *(resp.,* IAC^RHMC*) as given in* (30) [*resp.,* (29)]. *The grey lines show an empirical estimate obtained by using an output trajectory with* 10^6 *samples and the ACOR software package* [14, 36].

5.2. *Ten-dimensional normal distribution.*   We next consider the case $D = 10$ and

$$\sigma_i = \frac{i}{D}, \qquad 1 \le i \le D.$$

Figure 3 compares the IAC of HMC and RHMC. Figure 4 refers to MSD. The pathologies of HMC in the one-dimensional case discussed above are also manifest here, but, with 10 different frequencies in the Hamiltonin system, there are more
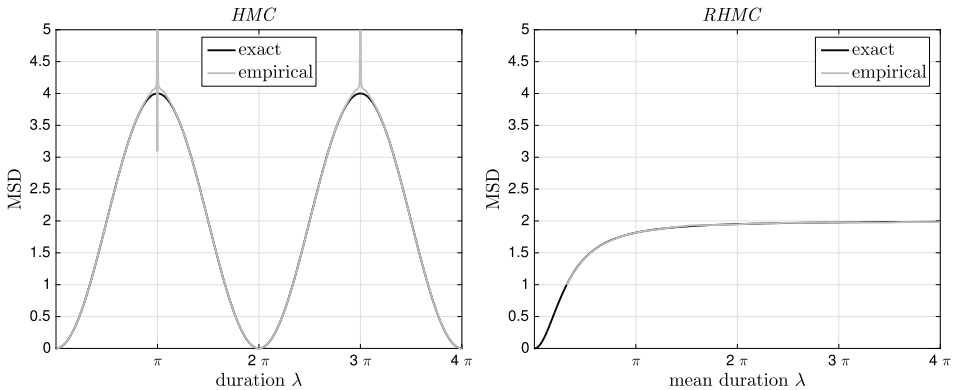

FIG. 2. *MSD for Normal Distribution. The left (resp., right) panel of this figure plots* MSD^HMC *(resp.,* MSD^RHMC*) vs. duration (resp. mean duration) λ for the HMC (resp., RHMC) method applied to a standard normal distribution. The black lines in the left (resp. right) panel plot* MSD^HMC *(resp.,* MSD^RHMC*) as given in* (33) [*resp.,* (31)]. *The grey lines show an empirical estimate obtained by using an output trajectory with* 10^6 *samples.*
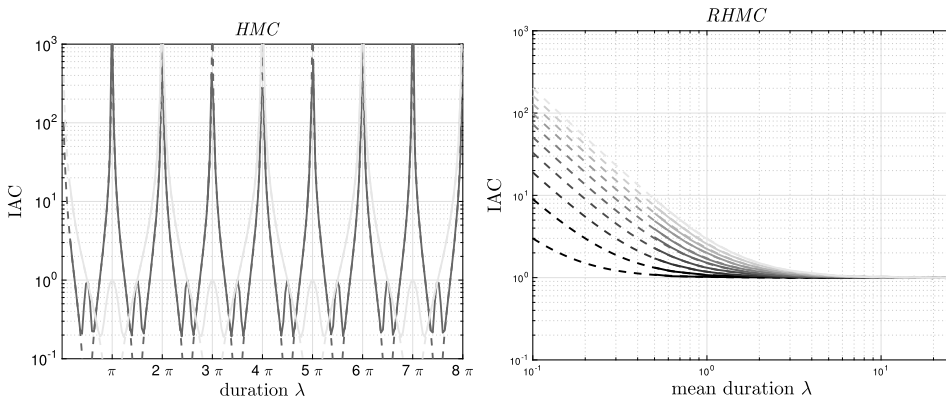
FIG. 3. *IAC for Ten-Dimensional Normal Distribution. The left panel plots* IAC_i^{HMC} *given in* (30) *as a function of the duration* λ *with* $i = 10$ (*dashed light grey*) *and* $i = 5$ (*dashed dark grey*); *and an approximation of* IAC_i^{HMC} *using a time series:* $\{q_i(t_k)\}_{1 \le k \le 10^6}$ *and the ACOR software package with* $i = 10$ (*solid light grey*) *and* $i = 5$ (*solid dark grey*) [14, 36]. *The dashed lines in the right panel plot* IAC_i^{RHMC} *given in* (29) *as a function of the mean duration* λ *for* $1 \le i \le 10$. *The solid lines plot an approximation of* IAC_i^{RHMC} *using a time series:* $\{q_i(t_k)\}_{1 \le k \le 10^6}$ *and the ACOR software package* [14, 36]. *Different shades of grey in the right panel indicate different components of the target distribution, with darker shades corresponding to lower variance.*

resonant choices of λ leading to lack of ergodicity. In the right panel of Figure 4, we see that MSD^{RHMC} has a horizontal asymptote at

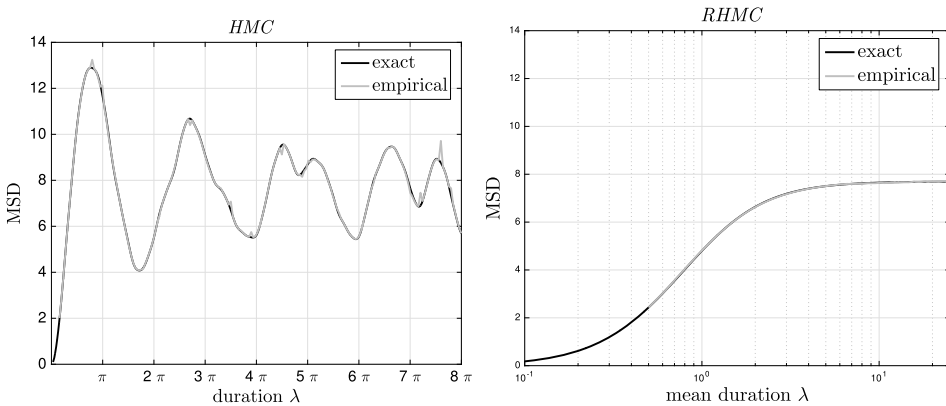$$\sum_{i=1}^{10} 2\sigma_i^2 \approx 7.7.$$

FIG. 4. *MSD for Ten-Dimensional Normal Distribution. The solid black lines in the left* (*resp., right*) *panel plot the formulas given in* (33) [*resp.,* (31)] *as a function of the duration* (*resp., mean duration*) λ. *The solid grey lines in the left* (*resp., right*) *panel plot an approximation of* MSD^{HMC} (*resp.,* MSD^{RHMC}) *associated to the time series:* $\{q_i(t_k)\}_{1 \le k \le 10^6}$ *versus* λ *for* $1 \le i \le D$.
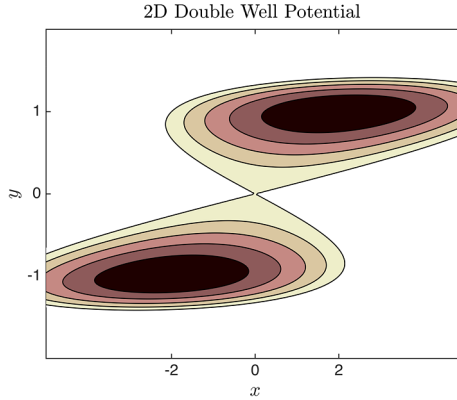
FIG. 5.   *Contours of* $\Phi(x)$ *in* (34). *This potential energy function has local minima located at* $x^{\pm} = (\pm 2, \pm 1)$ *with* $\Phi(x^+) = \Phi(x^-)$, *and a saddle point at the origin.*

To summarize, the behavior of MSD of the HMC method as a function of the duration is complex, whereas for the RHMC method the dependence is monotonic and plateaus.

5.3. *Two-dimensional double well.*   Consider a Brownian particle with the following two-dimensional double-well potential energy function:

$$(34) \qquad \Phi(x_1, x_2) = 5(x_2^2 - 1)^2 + 1.25\left(x_2 - \frac{x_1}{2}\right)^2,$$

whose contours are displayed in Figure 5. The left (resp., right) panel of Figure 6 plots the IAC τ versus duration λ for estimating the mean of the time series $\{f(Q_1(t_{i-1}), Q_2(t_{i-1}))\}_{1 \le i \le 10^6}$ where $f(x, y) = 2x + y$ produced by a HMC
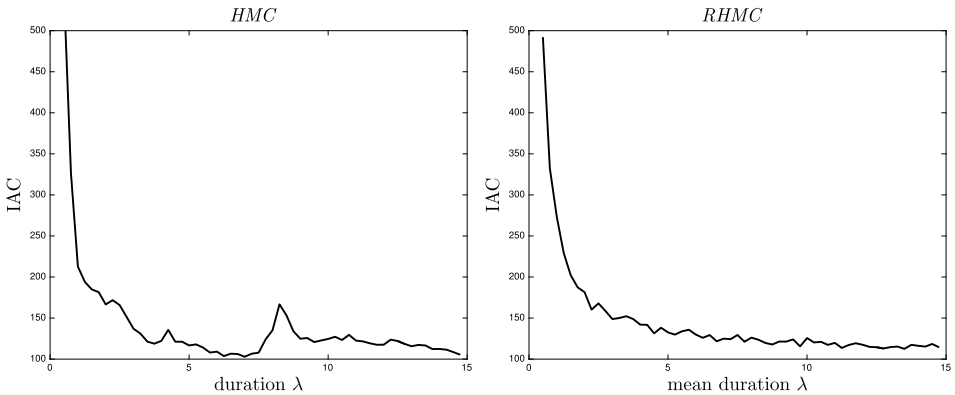

FIG. 6.   *IAC for* 2D *Double Well. The left* (*resp., right*) *panel plot an estimate of the IAC associated to the time series* $\{2Q_1(t_{i-1}) + Q_2(t_{i-1})\}_{1 \le i \le 10^6}$ *versus duration* (*resp., mean duration*) λ *for a HMC* (*resp., RHMC*) *method applied to the potential energy function graphed in Figure* 5.
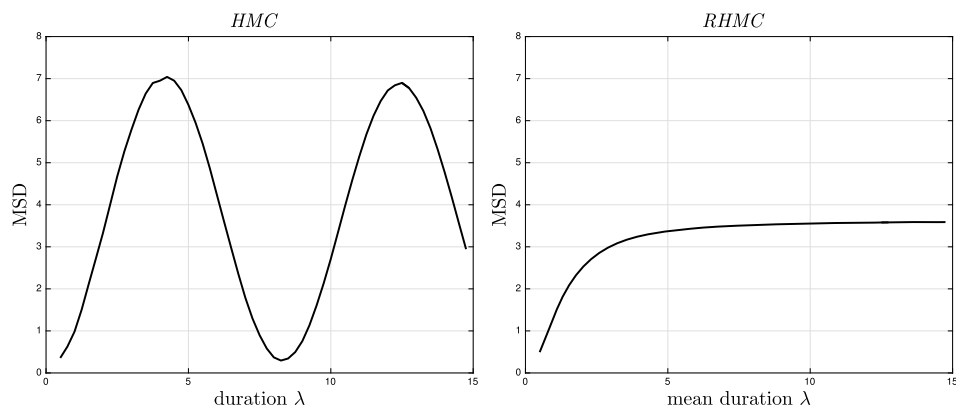
FIG. 7. *MSD for 2D Double Well. The left (resp., right) panel plot an estimate of the equilibrium MSD vs. duration (resp., mean duration) λ for a HMC (resp., RHMC) method applied to the potential energy function graphed in Figure 5.*

(resp., RHMC) scheme. This test function is the dot product between the configuration vector $(Q_1(t_{i-1}), Q_2(t_{i-1}))$ and the line connecting the two wells or span$((2, 1))$. For the same output trajectory, the left (resp., right) panel of Figure 7 plots an estimate of the equilibrium mean-squared displacement produced by a HMC (resp., RHMC) scheme.

5.4. *Fifteen-dimensional pentane molecule.* Consider sampling from the equilibrium distribution of a chain of five particles with a potential energy that is a function of bond lengths, bond angles, dihedral angles and inter-particles distances. This potential energy is described in greater detail in [8], which is based on the model described in [26]. Figure 8 illustrates and describes the three types of minima of the potential energy for the parameter values selected. Due to symmetries, each type corresponds to several modes of the distribution. For instance, there is stable configuration where particles 1 and 5 are above the plane of particles 2, 3,
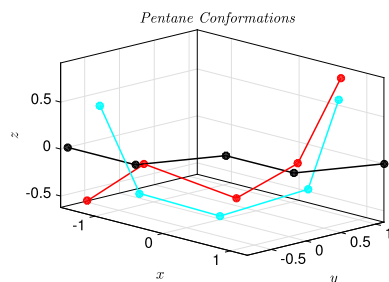

FIG. 8. *15D Pentane Molecule. This figure shows the three types of potentially stable conformations of the pentane molecule.*
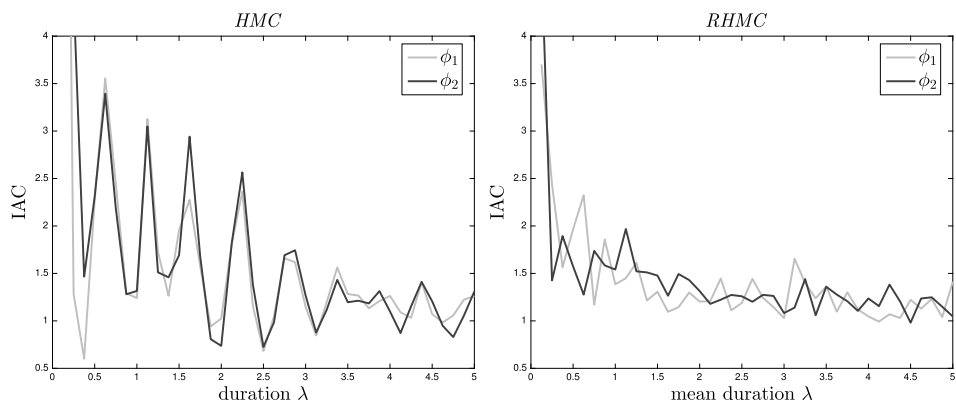
FIG. 9.   *IAC for* 15*D Pentane. The left* (*resp., right*) *panel plot an estimate of the IAC associated to the dihedral angles* (*labelled* ϕ_1 *and* ϕ_2 *in the figure legends*) *vs. duration* (*resp., mean duration*) λ *for a HMC* (*resp., RHMC*) *method applied to the potential energy function of the pentane molecule described in the text. The time series each consist of* 10^4 *samples.*

4 and another of the same type where they are below. Figure 9 (resp., 10) compares the IAC (resp., MSD) of HMC and RHMC. For this moderate dimensional distribution, we again observe a complex dependence of the performance of HMC on λ.

6. Outlook. All of the preceding developments have taken place in the exact integration scenario. The most obvious modification of Algorithm 2.1 that takes into account integration errors approximates the Hamiltonian flow by a volume-
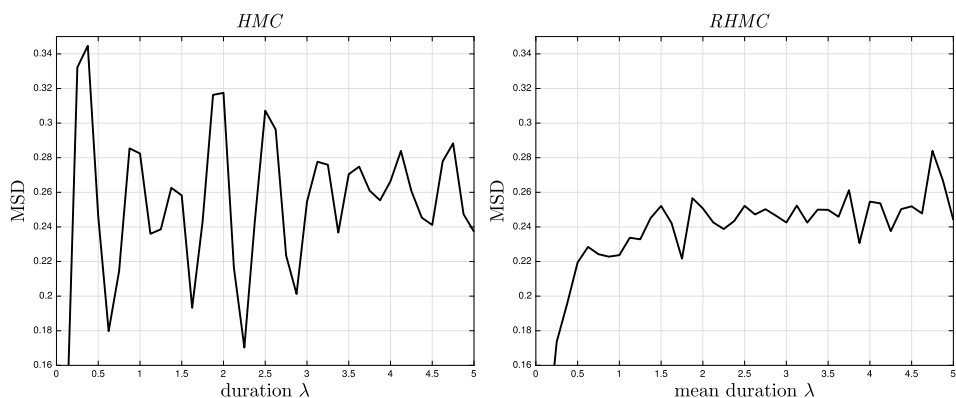

FIG. 10.   *MSD for* 15*D Pentane. The left* (*resp., right*) *panel plot an estimate of the equilibrium MSD vs. duration* (*resp., mean duration*) λ *for a HMC* (*resp., RHMC*) *method applied to the potential energy function of the pentane molecule described in the text. The time series each consist of* 10^4 *samples.*

preserving, reversible integrator (such as Verlet) with time step Δt, updates time via $t_1 = t_0 + M \Delta t$, where M is a random number of time-steps that is geometrically distributed with mean $\lambda/\Delta t$, and uses an accept-reject mechanism to remove the bias due to the energy errors introduced by this integrator. The study of that algorithm involves many numerical analysis technicalities, as it is necessary to derive global error bounds for the numerical solution that are valid over integration legs of arbitrarily long length $M \Delta t$. Such developments do not have much relation with the mathematical techniques used here so far and are out of the scope of the present article.

In this section, we suggest two variants of Algorithm 2.1 that do not use the exact solution of the Hamiltonian dynamics. These variants are based on approximating the gradients in (5) with the help of a numerical integrator and are therefore similar to the recently introduced generalized Markov chain Approximation Methods (MCAM) [7, 20]. We shall not be concerned with comparing the efficiency of the modifications of Algorithm 2.1 introduced here with that of standard HMC and related techniques.

In what follows, for a given numerical integrator, we denote by $\theta_{\Delta t}$ the map that advances the solution of Hamilton's equations over a single time-step of length Δt; a typical example is provided by the (velocity) Verlet integrator $(q_1, p_1) = \theta_{\Delta t}(q, p)$:

$$\binom{q_1}{p_1} = \begin{pmatrix} q + \Delta t p - \dfrac{(\Delta t)^2}{2} \nabla \Phi(q) \\ p - \dfrac{\Delta t}{2} \big(\nabla \Phi(q) + \nabla \Phi(q_1)\big) \end{pmatrix}.$$

6.1. *Variant #1.* This variant is defined by the following approximation to the generator (5) of RHMC:

$$L_h^{(1)} f(q, p) = \lambda^{-1} \mathbb{E}\{ f(\Gamma(q, p)) - f(q, p) \} + h^{-1} \big(f(\theta_h(q, p)) - f(q, p) \big),$$

where $h > 0$ is a parameter; since the Liouville operator acting on a test function f represents the time-derivative of f along the solutions of Hamilton's equations, it is clear that, for consistent integrators and smooth test functions,

$$L_h^{(1)} f(q, p) = Lf(q, p) + O(h).$$

The random jump times and the embedded chain of the Markov process generated by $L_h^{(1)}$ can be produced by iterating the following algorithm.

ALGORITHM 6.1. Given a duration parameter $\lambda > 0$, a step length parameter $h > 0$, the current time t_0, and the current state $(\tilde{Q}(t_0), \tilde{P}(t_0)) \in \mathbb{R}^{2D}$, output an updated state $(\tilde{Q}(t_1), \tilde{P}(t_1))$ at the random time t_1 using two steps:

Step 1. Generate an exponential random variable δt with mean $h\lambda/(h + \lambda)$, and update time via $t_1 = t_0 + \delta t$.

Step 2. Generate a uniform random variable $u \sim U(0, 1)$ and set

$$\left(\tilde{Q}(t_1), \tilde{P}(t_1)\right) = \begin{cases} \Gamma(\tilde{Q}(t_0), \tilde{P}(t_0)), & u \le \dfrac{h}{h + \lambda}, \\ \theta_h(\tilde{Q}(t_0), \tilde{P}(t_0)), & \text{otherwise}, \end{cases}$$

where Γ is the momentum randomization map given in (6).

The random variables δt, u, and the random vector ξ in the momentum randomization map Γ are independent. In terms of the sequence of random jump states $\{(\tilde{Q}(t_i), \tilde{P}(t_i))\}$ and random jump times $\{t_i\}$, the trajectory of the process is

$$\left(Q(t), P(t)\right) = \left(\tilde{Q}(t_i), \tilde{P}(t_i)\right) \qquad \text{for } t \in [t_i, t_{i+1}).$$

For any $t > 0$, the time-average of a function $f : \mathbb{R}^{2D} \to \mathbb{R}$ along this trajectory is given by

$$\int_0^t f\left(Q(s), P(s)\right) ds = \sum_{0 \le i \le \infty} f\left(Q(t_i), P(t_i)\right)(t \wedge t_{i+1} - t \wedge t_i).$$

The mean holding time of this Markov jump process is constant and given by

$$(35) \qquad \mathbb{E}\delta t = \frac{h\lambda}{h + \lambda}.$$

If λ is large and h is small, this process mainly jumps from (q, p) to $\theta_h(q, p)$, with occasional randomizations of momentum; in other words, the jump states come from integration legs of the Hamiltonian dynamics interspersed with occasional momentum randomizations. Note that, while the holding times δt are random, the step-length parameter in the numerical integrator remains constant. Figure 11 illustrates the use of the algorithm in the case a one-dimensional standard normal target distribution.

We now show that the generator $L_h^{(1)}$ inherits the stochastic Lyapunov function of L. The hypotheses on the potential energy that we use imply that the tails of the target $\exp(-\Phi)$ are no lighter or heavier than those of a Gaussian distribution; it is likely that those hypotheses may be relaxed. The hypotheses on θ_h are satisfied by any reasonable integrator such as the Euler rule or the Verlet method.

PROPOSITION 6.1. *Suppose that*, *in addition to Hypothesis* 3.1, *the potential energy is twice-differentiable and satisfies the following conditions*:

1. *There exists $C > 0$ such that for all $q \in \mathbb{R}^D$:*

$$\left\| D^2\Phi(q) \right\| \le C.$$

2. *There exists $C > 0$ such that for all $q \in \mathbb{R}^D$:*
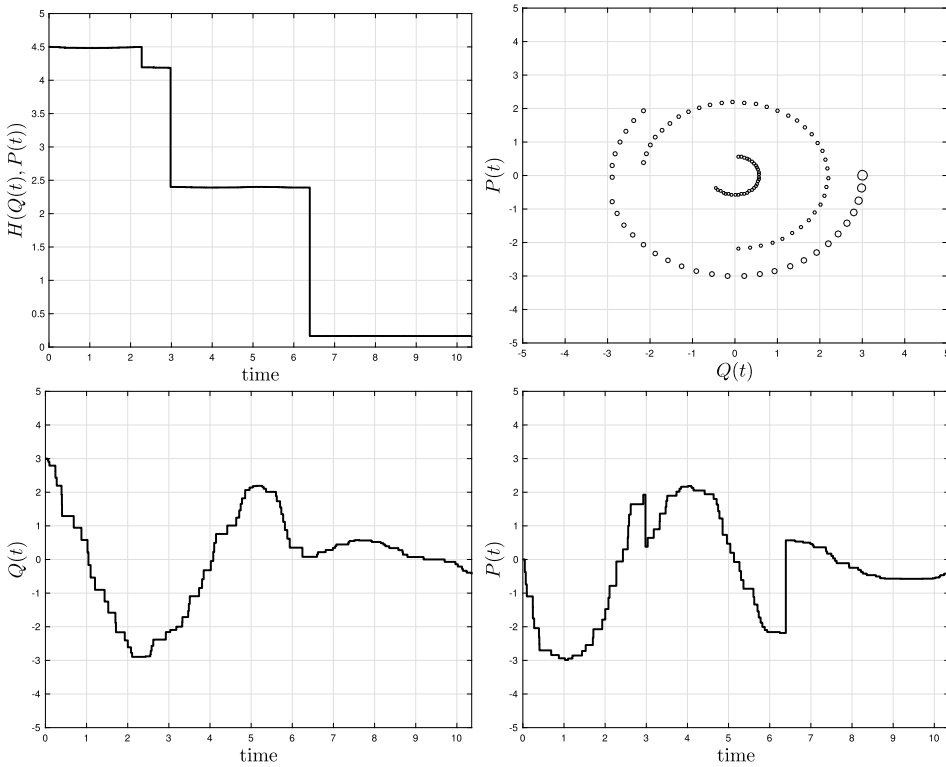
$$\Phi(q) \ge C\left(1 + |q|^2\right).$$

FIG. 11. *Sample Paths of Variant #1. These figures show a realization produced by iterating Algorithm 6.1 in the case of a one-dimensional normal target distribution. The step length is $h = 0.125$, and the duration $\lambda = \pi$ (these values are chosen for visualization purposes only). The top left panel shows the evolution of the total energy as a function of time. Note that there have been three momentum randomizations and that, between them, the total energy is essentially constant. The top right panel shows the evolution of the discrete trajectory in phase space. The size of the markers is related to t: points along the trajectory corresponding to larger values of t have smaller markers. The bottom panels show the position and momentum as functions of time. The holding time in each state is random, and $Q(t)$ and $P(t)$ are piecewise constant in time.*

Furthermore, assume that the integrator is such that, for any $f \in C^2(\mathbb{R}^{2D}, \mathbb{R})$ with the property that there exists $C_1 > 0$ such that

$$\|D^2 f(q, p)\| \le C_1 \qquad \forall (q, p) \in \mathbb{R}^{2D},$$

there exists a constant $C_2(f) > 0$ such that

$$f(\theta_h(q, p)) = f(q, p) + h\mathcal{L}f(q, p) + R_f(q, p, h),$$

where

$$|R_f(q, p, h)| \le C_2(f)(1 + H(q, p))h^2 \qquad \forall (q, p) \in \mathbb{R}^{2D}.$$

Then, for h sufficiently small, there exist $\gamma_h > 0$ and $K_h \geq 0$ such that

$$L_h^{(1)} V(q, p) \leq -\gamma_h V(q, p) + K_h \qquad \forall (q, p) \in \mathbb{R}^{2D}.$$

PROOF. By hypothesis, there exists a constant $C_V > 0$ such that

$$h^{-1}\big(V\big(\theta_h(q, p)\big) - V(q, p)\big) \leq \mathcal{L}V(q, p) + h\big(1 + V(q, p)\big)C_V.$$

Since V is a Lyapunov function for L, we have that

$$\begin{aligned} L_h^{(1)} V(q, p) &\leq LV(q, p) + h\big(1 + V(q, p)\big)C_V \\ &\leq -\gamma V(q, p) + K + h\big(1 + V(q, p)\big)C_V \\ &\leq -(\gamma - hC_V)V(q, p) + (K + hC_V) \end{aligned}$$

which gives the desired result with $\gamma_h = \gamma - hC_V$ and $K_h = K + hC_V$. $\square$

Unfortunately, due to the discretization error, the integrator θ_h does not preserve the Hamiltonian function exactly, and thus, the invariant measure of $L_h^{(1)}$ is not the invariant measure of L.

6.2. *Variant #2.* In order to correct the bias in $L_h^{(1)}$, we add to Algorithm 6.1 the possibility of additional jumps (momentum flips) from (q, p) to $\phi(q, p) = (q, -p)$, as follows.

ALGORITHM 6.2. Let $\nu(q, p) = \exp(-H(q, p))$ and define the Metropolis ratio

$$\alpha_h(q, p) = 1 \wedge \frac{\nu(\theta_h(q, p))}{\nu(q, p)}$$

for all $(q, p) \in \mathbb{R}^{2D}$. Given a duration parameter $\lambda > 0$, a step length parameter $h > 0$, the current time t_0, and the current state $(\tilde{Q}(t_0), \tilde{P}(t_0)) \in \mathbb{R}^{2D}$, the method outputs an updated state $(\tilde{Q}(t_1), \tilde{P}(t_1))$ at the random time t_1 using two steps:

Step 1. Generate an exponential random variable δt with mean $h\lambda/(h + \lambda)$, and update time via $t_1 = t_0 + \delta t$.
Step 2. Generate a uniform random variable $u \sim U(0, 1)$ and set

$$\begin{aligned} &\big(\tilde{Q}(t_1), \tilde{P}(t_1)\big) \\ &= \begin{cases} \Gamma\big(\tilde{Q}(t_0), \tilde{P}(t_0)\big), & u \leq \dfrac{h}{h + \lambda}, \\ \theta_h\big(\tilde{Q}(t_0), \tilde{P}(t_0)\big), & \dfrac{h}{h + \lambda} < u \leq \dfrac{h + \alpha_h(\tilde{Q}(t_0), \tilde{P}(t_0))\lambda}{h + \lambda}, \\ \phi\big(\tilde{Q}(t_0), \tilde{P}(t_0)\big), & \text{otherwise}, \end{cases} \end{aligned}$$

where Γ is the momentum randomization map given in (6).

The infinitesimal generator associated to Algorithm 6.2 is given by

$$
\begin{aligned}
(36) \quad L_h^{(2)} f(q, p) = {} & \lambda^{-1} \mathbb{E}\{ f(\Gamma(q, p)) - f(q, p) \} \\
& + h^{-1} \alpha_h(q, p)(f(\theta_h(q, p)) - f(q, p)) \\
& + h^{-1}(1 - \alpha_h(q, p))(f(\phi(q, p)) - f(q, p)).
\end{aligned}
$$

Like $L_h^{(1)}$, the generator $L_h^{(2)}$ induces a Markov jump process with a constant mean holding time given by (35). If λ is large and h is small, this process mainly jumps from (q, p) to $\theta_h(q, p)$, with occasional randomizations of momentum and momentum flips. As we show next, the weights α_h and $(1 - \alpha_h)$ in (36) have been chosen to ensure that, for suitable integrators, $L_h^{(2)}$ has the same infinitesimally invariant measure of L.

PROPOSITION 6.2. *Suppose that the integrator θ_h is reversible, that is,* $\theta_{\Delta t} \circ \phi \circ \theta_{\Delta t} = \phi$ *and* volume-preserving, *that is,* $\det(D\theta_h) = 1$. *Then, for all* $f \in C_c^\infty(\mathbb{R}^{2D})$, *we have that*

$$
\int_{\mathbb{R}^{2D}} L_h^{(2)} f(q, p) \nu(q, p) \, dq \, dp = 0.
$$

PROOF. Note that

$$
\int_{\mathbb{R}^{2D}} L_h^{(2)} f(q, p) \nu(q, p) \, dq \, dp
$$

$$
= \int_{\mathbb{R}^{2D}} \lambda^{-1} \mathbb{E}\{ f(\Gamma(q, p)) - f(q, p) \} \nu(q, p) \, dq \, dp
$$

$$
+ \frac{1}{h} \int_{\mathbb{R}^{2D}} (\nu(q, p) \wedge \nu(\theta_h(q, p)))(f(\theta_h(q, p)) - f(q, p)) \, dq \, dp
$$

$$
+ \frac{1}{h} \int_{\mathbb{R}^{2D}} (\nu(q, p) - \nu(q, p) \wedge \nu(\theta_h(q, p)))(f(\phi(q, p))
$$

$$
- f(q, p)) \, dq \, dp.
$$

As discussed in Proposition 3.1, the first integral on the right-hand side vanishes. For the next two integrals, since θ_h is volume-preserving and reversible by hypothesis, a change of variables implies that

$$
\int_{\mathbb{R}^{2D}} L_h^{(2)} f(q, p) \nu(q, p) \, dq \, dp
$$

$$
= \frac{1}{h} \int_{\mathbb{R}^{2D}} [\nu(\theta_h^{-1}(q, p)) \wedge \nu(q, p)] f(q, p) \, dq \, dp
$$

$$
- \frac{1}{h} \int_{\mathbb{R}^{2D}} [\nu(q, p) \wedge \nu(\theta_h(q, p))] f(q, p) \, dq \, dp
$$

$$-\frac{1}{h}\int_{\mathbb{R}^{2D}}\big[\nu\big(\phi(q,p)\big)\wedge\nu\big(\theta_h\circ\phi(q,p)\big)\big]f(q,p)\,dq\,dp$$

$$+\frac{1}{h}\int_{\mathbb{R}^{2D}}\big[\nu(q,p)\wedge\nu\big(\theta_h(q,p)\big)\big]f(q,p)\,dq\,dp=0,$$

where in the last step we used the fact that $\nu\circ\theta_h=\nu\circ\phi\circ\theta_h=\nu\circ\theta_h^{-1}$. □

We conjecture that Algorithm 6.2 is geometrically ergodic under suitable assumptions on the potential energy function and suitable choices of the reversible, volume preserving integrator. However, the proof of this appears to be involved because one needs to estimate carefully the behavior of the jump rate from (q,p) to $\theta_h(q,p)$ in regions where H is large, and consequently, the integrator has large errors. This analysis will be presented in a future publication.

7. Conclusion. In this article, we have primarily studied HMC in the exact integration scenario. The RHMC method introduced here is a version of HMC where the durations of the Hamiltonian flows are independent, exponentially distributed random variables. This method has an infinitesimal generator which is a linear combination of a momentum randomization operator and a (differential) Liouville operator.

The analysis in Section 3 related the nonasymptotic properties of RHMC to those of the underdamped Langevin dynamics. We showed that, under standard hypotheses on the potential energy function, RHMC possesses a stochastic Lyapunov function of the same form as that of the underdamped Langevin dynamics. However, unlike underdamped Langevin dynamics, the trajectories produced by RHMC are not continuous functions of time due to the instantaneous momentum randomizations. This difference in regularity made it tricky to use standard approximate controllability arguments to establish a minorization condition for RHMC and an alternative approach was required. Our analysis showed that the mechanism for dissipation (and hence, exponential stability) in RHMC comes from the momentum randomizations; therefore, momentum randomizations play here the role played by the heat bath appearing in underdamped Langevin dynamics. For sampling, the main qualitative advantage of RHMC compared with underdamped Langevin dynamics is that the paths of RHMC have a stronger tendency to move consistently away from the current state of the chain.

In a model test problem, we carried out a quantitative analysis of the sampling performance of RHMC using integrated autocorrelation time and mean squared displacement as metrics. Our analysis showed that randomizing the durations of the Hamiltonian flows mitigates some artifacts associated to using Hamiltonian dynamics. In particular, we saw that these sampling metrics depend monotonically on the mean duration parameter. Numerical examples showed that this monotonicity persists for more general target distributions. In contrast, these sampling metrics

for classical HMC are a more complicated function of its (deterministic) duration parameter.

As an outlook to future developments of RHMC, we considered two approximations of RHMC based on spatially discretizing the infinitesimal generator of RHMC. This outlook introduced a new viewpoint to developing Hamiltonian-based MCMC methods. In particular, it transforms the problem of approximating the Hamiltonian flow from one of time discretizing Hamilton's equations into one of spatially discretizing its associated Liouville operator in high dimension in a way that generates a Markov process [7]. We showed how to construct such approximations so that they preserve the Boltzmann–Gibbs distribution. A complete analysis of these approximations to RHMC will be the subject of future work.

APPENDIX: HARRIS'S THEOREM

Harris's theorem states that if a Markov process admits a Lyapunov function such that its sublevel sets are "small," then it is geometrically ergodic. In this part, we recall this theorem for the convenience of the reader. Since RHMC has an infinitesimal generator, it is convenient to formulate the Foster–Lyapunov condition in Harris's theorem in terms of an infinitesimal generator. To make the domain of this generator sufficiently inclusive, we define the infinitesimal generator of a Markov process $X(t)$ on a Polish state space Ω equipped with probability measure $\mathbb{P}$ in the following way.

DEFINITION A.1. Let $D(\mathsf{L})$ be the set of all measurable functions $F : \Omega \to \mathbb{R}$ such that there exists a measurable function $G : \Omega \to \mathbb{R}$ with the property that for any $x \in \Omega$ and for all $t > 0$ the process

$$F\big(X(t)\big) - F(x) - \int_0^t G\big(X(s)\big) \, ds, \qquad X(0) = x,$$

is a local martingale adapted to the natural filtration of X under the probability measure $\mathbb{P}$. Then we define $\mathsf{L}F = G$ and call L the infinitesimal generator of the process $X(t)$ with domain $D(\mathsf{L})$.

This definition seems to be due to [9], Definition (14.15) in Chapter 1. In addition to the infinitesimal generator, denote by P_t the Markov semigroup of $X(t)$, and denote the transition probabilities of $X(t)$ by

$$\Pi_{t,x}(A) = \Pr\big(X(t) \in A \mid X(0) = x\big) \qquad \forall t \geq 0, \forall x \in \Omega.$$

Sufficient conditions for Harris's theorem to hold are Hypotheses A.1 and A.2 given below.

HYPOTHESIS A.1 (Foster–Lyapunov Drift Condition). *There exist a function* $\Psi : \Omega \to \mathbb{R}^+$ *and strictly positive constants* w *and* k *such that*

(37) $$\mathsf{L}\Psi(x) \leq -\mathsf{w}\Psi(x) + \mathsf{k},$$

for all $x \in \Omega$.

REMARK A.1. Hypothesis A.1 implies that

$$P_t \Psi(x) \le e^{-\mathsf{w}t} \Psi(x) + \frac{\mathsf{k}}{\mathsf{w}}(1 - e^{-\mathsf{w}t})$$

for every $t \ge 0$ and for every $x \in \Omega$.

HYPOTHESIS A.2. *There exist* $\mathsf{a} > 0$ *and* $t > 0$ *such that the sublevel set* $\{z \in \Omega \mid \Psi(z) \le 2\mathsf{k}/\mathsf{w}\}$ *is "small," that is,*

$$(38) \qquad \|\Pi_{t,x} - \Pi_{t,y}\|_{\mathrm{TV}} \le 2(1 - \mathsf{a})$$

for every pair x, y *satisfying* $\Psi(x) \vee \Psi(y) \le 2\mathsf{k}/\mathsf{w}$, *where the constants* k *and* w *are taken from Hypothesis* A.1.

THEOREM A.1 (Harris's theorem). *Consider a Markov process* $\mathsf{X}(t)$ *on* Ω *with generator* L *and transition probabilities* $\Pi_{t,x}$, *which satisfies Hypotheses* A.1 *and* A.2. *Then* $\mathsf{X}(t)$ *possesses a unique invariant probability measure* Π, *and there exist positive constants* C *and* r (*both depending only on the constants* w, k *and* a *appearing in the assumptions*) *such that*

$$\|\Pi_{t,x} - \Pi\|_{\mathrm{TV}} \le \mathsf{C} \exp(-\mathsf{r}t)\Psi(x),$$

for all $x \in \Omega$ *and for any* $t \ge 0$.

REMARK A.2. It follows from (37) that

$$\Pi(\Psi) \le \frac{\mathsf{k}}{\mathsf{w}}.$$

For further exposition and a proof of Harris's theorem in a general context, see the monograph [30] (or [15] for an alternative proof), and for a treatment in the specific context of stochastic differential equations, see [27].

REFERENCES

[1] AKHMATSKAYA, E. and REICH, S. (2008). GSHMC: An efficient method for molecular simulation. *J. Comput. Phys.* **227** 4934–4954. MR2414842

[2] ANDERSEN, H. C. (1980). Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **72** 2384.

[3] ASMUSSEN, S. and GLYNN, P. W. (2007). *Stochastic Simulation: Algorithms and Analysis. Stochastic Modelling and Applied Probability* **57**. Springer, New York. MR2331321

[4] BESKOS, A., PINSKI, F. J., SANZ-SERNA, J. M. and STUART, A. M. (2011). Hybrid Monte-Carlo on Hilbert spaces. *Stochastic Process. Appl.* **121** 2201–2230. MR2822774

[5] BLANES, S., CASAS, F. and SANZ-SERNA, J. M. (2014). Numerical integrators for the hybrid Monte Carlo method. *SIAM J. Sci. Comput.* **36** A1556–A1580. MR3233942

[6] BOU-RABEE, N. and OWHADI, H. (2010). Long-run accuracy of variational integrators in the stochastic context. *SIAM J. Numer. Anal.* **48** 278–297. MR2608370

[7] BOU-RABEE, N. and VANDEN-EIJNDEN, E. (2016). Continuous-time random walks for the numerical solution of stochastic differential equations. *Memoirs of the AMS*. To appear.

[8] CANCÈS, E., LEGOLL, F. and STOLTZ, G. (2007). Theoretical and numerical comparison of some sampling methods for molecular dynamics. *ESAIM Math. Model. Numer. Anal.* **41** 351–389. MR2339633

[9] DAVIS, M. H. A. (1993). *Markov Models and Optimization. Monographs on Statistics and Applied Probability* **49**. Chapman & Hall, London. MR1283589

[10] DA PRATO, G. and ZABCZYK, J. (1996). *Ergodicity for Infinite-Dimensional Systems. London Mathematical Society Lecture Note Series* **229**. Cambridge Univ. Press, Cambridge. MR1417491

[11] DUANE, S., KENNEDY, A. D., PENDLETON, B. J. and ROWETH, D. (1987). Hybrid Monte-Carlo. *Phys. Lett. B* **195** 216–222.

[12] E, W. and LI, D. (2008). The Andersen thermostat in molecular dynamics. *Comm. Pure Appl. Math.* **61** 96–136. MR2361305

[13] FANG, Y., SANZ-SERNA, J. M. and SKEEL, R. D. (2014). Compressible generalized hybrid Monte Carlo. *J. Chem. Phys.* **140** 174108.

[14] GOODMAN, J. Acor, statistical analysis of a time series. Available at http://www.math.nyu.edu/faculty/goodman/software/acor/. Accessed 06–18–06–2015.

[15] HAIRER, M. and MATTINGLY, J. C. (2011). Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis*, *Random Fields and Applications VI* 109–117. Springer, Basel.

[16] HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. MR3363437

[17] HOMAN, M. D. and GELMAN, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15** 1593–1623.

[18] HOROWITZ, A. M. (1991). A generalized guided Monte-Carlo algorithm. *Phys. Lett. A* **268** 247–252.

[19] KENNEDY, A. D. and PENDLETON, B. (2001). Cost of the generalised hybrid Monte Carlo algorithm for free field theory. *Nuclear Phys. B* **607** 456–510. MR1850796

[20] KUSHNER, H. J. and DUPUIS, P. (2001). *Numerical Methods for Stochastic Control Problems in Continuous Time*, 2nd ed. *Applications of Mathematics* (*New York*) **24**. Springer, New York. MR1800098

[21] LI, D. (2007). On the rate of convergence to equilibrium of the Andersen thermostat in molecular dynamics. *J. Stat. Phys.* **129** 265–287. MR2358805

[22] LIU, J. S. (2008). *Monte Carlo Strategies in Scientific Computing*, 2nd ed. Springer, New York. MR2401592

[23] MACKENZIE, P. B. (1989). An improved hybrid Monte Carlo method. *Phys. Lett. B* **226** 369–371.

[24] MARSDEN, J. E. and RATIU, T. S. (1999). *Introduction to Mechanics and Symmetry*, 2nd ed. *Texts in Applied Mathematics* **17**. Springer, New York. MR1723696

[25] MARSDEN, J. E. and WEST, M. (2001). Discrete mechanics and variational integrators. *Acta Numer.* **10** 357–514. MR2009697

[26] MARTIN, M. G. and SIEPMANN, J. I. (1998). Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. *J. Phys. Chem.*, *B* **102** 2569–2577.

[27] MATTINGLY, J. C., STUART, A. M. and HIGHAM, D. J. (2002). Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.* **101** 185–232. MR1931266

[28] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.

[29] MEYN, S. and TWEEDIE, R. (1993). Stability of Markovian processes III: Foster–Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.* **25** 518–548. MR1234295

[30] MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2509253

[31] NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 113–162. CRC Press, Boca Raton, FL. MR2858447

[32] PROTTER, P. E. (2004). *Stochastic Integration and Differential Equations*, 2nd ed. *Applications of Mathematics* (*New York*) **21**. Springer, Berlin. MR2020294

[33] SANZ-SERNA, J. M. (2014). Markov Chain Monte Carlo and numerical differential equations. In *Current Challenges in Stability Issues for Numerical Differential Equations* (L. Dieci and N. Guglielmi, eds.) **2082** 39–88. Springer, Berlin.

[34] SANZ-SERNA, J. M. and CALVO, M. P. (1994). *Numerical Hamiltonian Problems. Applied Mathematics and Mathematical Computation* **7**. Chapman & Hall, London. MR1270017

[35] SANZ-SERNA, J. M. and STUART, A. M. (1999). Ergodicity of dissipative differential equations subject to random impulses. *J. Differential Equations* **155** 262–284. MR1698555

[36] SOKAL, A. (1997). Monte Carlo methods in statistical mechanics: Foundations and new algorithms. In *Functional Integration* (*Cargèse*, 1996). *NATO Adv. Sci. Inst. Ser. B Phys.* **361** 131–192. Plenum, New York. MR1477456

[37] TALAY, D. (2002). Stochastic Hamiltonian systems: Exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Processes Related Fields* **8** 1–36. MR1924934

DEPARTMENT OF MATHEMATICAL SCIENCES
RUTGERS UNIVERSITY–CAMDEN
311 N 5TH STREET
CAMDEN, NEW JERSEY 08102
USA
E-MAIL: nawaf.bourabee@rutgers.edu

DEPARTAMENTO DE MATEMÁTICAS
UNIVERSIDAD CARLOS III DE MADRID
AVENIDA DE LA UNIVERSIDAD 30
E-28911 LEGANÉS (MADRID)
SPAIN
E-MAIL: jmsanzserna@gmail.com