

THEORETICAL AND NUMERICAL COMPARISON OF SOME SAMPLING METHODS FOR MOLECULAR DYNAMICS

ERIC CANCÈS^{1,2}, FRÉDÉRIC LEGOLL^{1,3} AND GABRIEL STOLTZ^{1,4}

Abstract. The purpose of the present article is to compare different phase-space sampling methods, such as purely stochastic methods (Rejection method, Metropolized independence sampler, Importance Sampling), stochastically perturbed Molecular Dynamics methods (Hybrid Monte Carlo, Langevin Dynamics, Biased Random Walk), and purely deterministic methods (Nosé-Hoover chains, Nosé-Poincaré and Recursive Multiple Thermostats (RMT) methods). After recalling some theoretical convergence properties for the various methods, we provide some new convergence results for the Hybrid Monte Carlo scheme, requiring weaker (and easier to check) conditions than previously known conditions. We then turn to the numerical efficiency of the sampling schemes for a benchmark model of linear alkane molecules. In particular, the numerical distributions that are generated are compared in a systematic way, on the basis of some quantitative convergence indicators.

Mathematics Subject Classification. 82B80, 37M25, 65C05, 65C40.

Received July 30, 2005.

Phase-space integrals are widely used in Statistical Physics to relate the macroscopic properties of a system to the elementary phenomena at the microscopic scale [17]. In constant temperature (NVT) molecular simulations, these integrals take the form

$$\langle A \rangle = \int_{T^*\mathcal{M}} A(q, p) d\mu(q, p). \quad (1)$$

In the above expression, \mathcal{M} denotes the position space (also called the *configuration space*), and $T^*\mathcal{M}$ its cotangent space. Typically, $\mathcal{M} = \mathbb{T}^{3N}$ (a torus of dimension $3N$) for simulations with periodic boundary conditions (PBC) and N atoms in the simulation cell. In this case, $T^*\mathcal{M} = \mathbb{T}^{3N} \times \mathbb{R}^{3N}$. Let us note that, for biological systems currently studied, N is typically more than 100 000. A generic element of the position space \mathcal{M} will be denoted by $q = (q_1, \dots, q_N)$ and a generic element of the momentum space \mathbb{R}^{3N} by $p = (p_1, \dots, p_N)$. The so-called canonical probability measure μ appearing in (1) is given by

$$d\mu(q, p) = Z^{-1} \exp(-\beta H(q, p)) dq dp, \quad (2)$$

Keywords and phrases. Sampling methods, canonical ensemble, Molecular Dynamics.

¹ CERMICS, École Nationale des Ponts et Chaussées, 6-8 avenue Blaise Pascal, 77455 Marne-la-Vallée Cedex 2, France. cances@cermics.enpc.fr; legoll@cermics.enpc.fr; stoltz@cermics.enpc.fr

² INRIA, Domaine de Voluceau-Rocquencourt, B.P. 105, 78153 Le Chesnay Cedex, France.

³ Institute for Mathematics and its Applications, University of Minnesota, 400 Lind Hall, 207 Church Street SE, Minneapolis MN 55455, USA.

⁴ CEA/DAM Ile-de-France, BP 12, 91680 Bruyères-le-Châtel, France.

where $\beta = 1/k_B T$ (T denotes the temperature and k_B the Boltzmann constant) and where H denotes the Hamiltonian of the molecular system:

$$H(q, p) = \frac{1}{2} p^T M^{-1} p + V(q). \quad (3)$$

In the above expression, $M = \text{Diag}(m_1, \dots, m_N)$ where m_i is the mass of the i -th particle, and V is the potential energy experienced by the N particles. In simulations with PBC, the potential energy V takes into account all the particles in the simulation cell, as well as their periodic images.

The constant Z in (2) is the normalization constant defined as

$$Z = \int_{T^* \mathcal{M}} \exp(-\beta H(q, p)) \, dq \, dp,$$

and is also called the partition function in Statistical Physics. Since the Hamiltonian H is separable, the canonical measure is of the form

$$d\mu(q, p) = d\pi(q) d\kappa(p),$$

where

$$d\kappa(p) = \mathcal{P}(p) \, dp = Z_p^{-1} \exp\left(-\frac{\beta}{2} p^T M^{-1} p\right) \, dp, \quad (4)$$

and

$$d\pi(q) = f(q) \, dq = Z_q^{-1} e^{-\beta V(q)} \, dq. \quad (5)$$

The positive numbers Z_q and $Z_p = (2\pi/\beta)^{3N/2} \prod_{i=1}^N m_i^{3/2}$ are normalization constants. Notice that we implicitly assume that the measures μ and π are probability measures, which is the case when $f \in L^1(\mathcal{M})$. In the case of an isolated system, the potential energy V is invariant under global translation and rotation of the system, and therefore $f \notin L^1(\mathcal{M})$. It is nevertheless possible to give a sense to (1) by quotienting out rigid body translations.

The function $A(q, p)$ in (1) is an observable, that is, a function on the phase-space related to some physical property of the macroscopic system under consideration. This property can be static, in which case the function A does not depend on time, or dynamical, in which case the function A depends on the time evolution of the system.

Let us present a few examples. For a Lennard-Jones fluid, a classical static quantity is the constant volume specific heat C_V . It is often defined in terms of energy fluctuations, and reads, for a system of N identical particles [54],

$$C_V = \frac{\mathcal{N}_a}{N k_B T^2} (\langle H^2 \rangle - \langle H \rangle^2),$$

where \mathcal{N}_a is the Avogadro number. Another example is the bulk pressure P in a liquid, given by $P = \langle A \rangle$ with

$$A(q, p) = \frac{1}{3|\mathcal{M}|} \sum_{i=1}^N \left(\frac{p_i^2}{m_i} - q_i \cdot \frac{\partial V}{\partial q_i}(q) \right),$$

where $|\mathcal{M}|$ is the volume occupied by the system. Many other static thermodynamical quantities can in fact be computed in the same way. Structural properties such as packing fractions in granular media or radial distribution functions in fluids are also static quantities that are classically considered.

Transport coefficients are examples of dynamical properties. For instance, the self diffusion coefficient in a system of N identical particles of mass m can be computed by the Einstein relation [54]:

$$D = \lim_{t \rightarrow +\infty} \frac{1}{6Nt} \left\langle \sum_{i=1}^N |q_i(t) - q_i(0)|^2 \right\rangle,$$

where $q_i(t)$ is the position of the i -th particle at time t . An alternative expression is the Green-Kubo formula based on the integrated velocity autocorrelation function [54]:

$$D = \frac{1}{3Nm^2} \int_0^{+\infty} \left\langle \sum_{i=1}^N p_i(t) \cdot p_i(0) \right\rangle dt,$$

where $p_i(t)$ is the momentum of the i -th particle at time t . Other classical examples are the shear viscosity of a fluid or its thermal diffusivity [54].

Let us now make precise the way the trajectories $(q(t), p(t))_{t \geq 0}$ are computed in practice. Usually, $(q(t), p(t))$ is taken to be equal to $\Phi_t(q(0), p(0))$ where

$$\begin{aligned} \Phi_t : T^*\mathcal{M} &\rightarrow T^*\mathcal{M} \\ (q, p) &\mapsto \Phi_t(q, p) \end{aligned} \tag{6}$$

is the flow associated with the Hamiltonian (3). In other words, $\Phi_t(q^0, p^0)$ is the solution at time t of the Hamiltonian equations¹

$$\begin{cases} \frac{dq(t)}{dt} = \frac{\partial H}{\partial p}(q(t), p(t)) = M^{-1}p(t) \\ \frac{dp(t)}{dt} = -\frac{\partial H}{\partial q}(q(t), p(t)) = -\nabla V(q(t)) \end{cases} \tag{7}$$

with initial conditions $(q(0), p(0)) = (q^0, p^0)$. Choosing NVE (constant energy) trajectories is perfectly justified for computing NVT (constant temperature) dynamical properties for some physical systems such as rarefied gases or molecular beams. The situation is less clear in the condensed phase but we can nevertheless remark that this choice is consistent in the sense that the canonical (NVT) measure (2) is invariant under the microcanonical (NVE) dynamics (7). We therefore consider here the following class of time-dependent thermodynamic integrals:

$$\langle B \rangle(t) = \int_{T^*\mathcal{M}} B(\Phi_t(q, p), (q, p)) d\mu, \tag{8}$$

where $B : T^*\mathcal{M} \times T^*\mathcal{M} \rightarrow \mathbb{R}$ is a regular function, Φ_t is defined by (6) and the measure $d\mu$ is given by (2). Transport coefficients and (auto)correlation functions can be expressed as such averages.

For some systems, Newton equations do not model properly the dynamics. This is for instance the case of biological molecules in implicit solvent, whose evolution is usually described by a Langevin equation. For such a system, the friction parameter has a physical interpretation, and can thus be fitted to an all-atom system (with an explicit solvent) simulation. Hence it is sometimes of interest to compute (8) where Φ_t is not the Hamiltonian flow, but the flow of some other dynamics (*e.g.* the Langevin dynamics). We will not consider such cases in this article.

An accurate numerical computation of time-dependent thermodynamical integrals asks first for a good sampling of the starting points, distributed according to the canonical distribution. These points should not be too numerous – one must be able to run short Hamiltonian trajectories (one or several) starting from each point with reasonable computer resources. Notice that techniques such as importance sampling do not seem appropriate here since they generate numerous points to sample the canonical measure $d\mu$, many of them with very small weights. As a consequence, when (8) is computed, a lot of time is wasted in computing trajectories with very low weights. Such methods are relevant for the computation of static integrals only. A simple computation confirms these heuristic arguments: the cost of computing a single trajectory over a given physical time interval $[0, T]$ scales as $(\Delta t)^{-1}$. The total cost is of order $O(N(\Delta t)^{-1})$, where N is the number of starting

¹We will assume in the sequel of this article that these equations, supplemented with some initial conditions, have a unique solution. This is for instance the case when ∇V is a globally Lipschitz function.

points. Therefore, for a fixed computational cost, there is a trade-off to be made between the accuracy of the sampling of $d\mu$ (scaled by N) and the accuracy of the numerical integration of (7) (given by Δt).

The purpose of this article is twofold: on the one hand, we provide new convergence results on the Hybrid Monte-Carlo sampling scheme (see Sect. 2.2); on the other hand, we review various numerical methods which are used in molecular simulation to compute integrals such as (1) or (8), and we compare their efficiency on simple alkane molecules. More precisely, we consider the issue of sampling from the canonical measure (2). Since it is straightforward to sample from the momentum distribution (4) (it is a product of independent Gaussian densities), the actual issue is to sample efficiently from the (position space) measure π given by (5).

All the methods considered in this article consist in generating a sequence of points $(q^n)_{n \in \mathbb{N}}$ in the position space. Depending on the method,

Type 1: $(q^n)_{n \in \mathbb{N}}$ is a sequence of independent realizations of a given random variable of density $f(q) = \frac{1}{Z_q} e^{-\beta V(q)}$; this is the case for the standard Rejection and for the Rejection control methods;

Type 2: $(q^n)_{n \in \mathbb{N}}$ is a realization of a continuous state-space Markov chain, for which π is an invariant measure; this is the case for the Metropolized independence sampler and for the Hybrid Monte Carlo method;

Type 3: $(q^n)_{n \in \mathbb{N}}$ is an approximation of $(q_{t_n})_{n \in \mathbb{N}}$ where $(q_t)_{t \geq 0}$ (resp. $(q_t, p_t)_{t \geq 0}$) is a sample path of a stochastic process on \mathcal{M} (resp. on $T^*\mathcal{M}$), for which π (resp. μ) is an invariant measure; this is the case for the biased Random-Walk (resp. for the Langevin dynamics);

Type 4: $(q^n)_{n \in \mathbb{N}}$ is an approximation of $(q(t_n))_{n \in \mathbb{N}}$ where $(q(t), p(t), x(t))_{t \geq 0}$ is a trajectory of a deterministic extended dynamical system (q and p are the physical variables, while x represents some additional variables; see Sect. 3 for more details); this extended dynamical system is such that it preserves a measure $d\rho$ whose projection on the physical variables q, p is the measure $d\mu$ given by (2); this is the case for Nosé-Hoover, Nosé-Poincaré and Recursive Multiple Thermostat methods.

The first two questions under examination, which are relevant for all the methods mentioned above, are the following:

- **Question 1.** An observable $A(q)$ on \mathcal{M} being given, does the empirical mean $\frac{1}{N} \sum_{n=0}^{N-1} A(q^n)$ converge

to the space average $\int_{\mathcal{M}} A(q) d\pi(q)$?

- **Question 2.** If so, can the speed of convergence be estimated?

For methods of Type 1, the answers to Questions 1 and 2 are obviously positive and are direct consequences of the Law of Large Number (LLN) and of the Central Limit Theorem (CLT) for independent identically distributed (i.i.d.) random variables. For the methods of Type 2, Questions 1 and 2 can be positively answered, at least for compact position spaces \mathcal{M} and under some assumptions on the potential energy V . For Question 1, the point is to check (see Thm. 1 below) that

$$\pi \text{ is an invariant probability measure of the Markov chain,} \tag{9}$$

and that the probability transition kernel $P(q, \cdot)$ of the Markov chain² satisfies the accessibility condition

$$\forall q \in \mathcal{M}, \quad \forall B \in \mathcal{B}(\mathcal{M}), \quad \mu^{\text{Leb}}(B) > 0 \Rightarrow P(q, B) > 0, \tag{10}$$

where $\mathcal{B}(\mathcal{M})$ is the Borel σ -algebra of \mathcal{M} and μ^{Leb} is the Lebesgue measure on \mathcal{M} . Turning to Question 2, a convergence rate of $N^{-1/2}$ can be obtained when the transition kernel P has some regularity properties, and provided some Lyapunov condition holds true (see Thm. 2 and condition (13) below).

²If $q \in \mathcal{M}$ and B is a Borel set of \mathcal{M} , $P(q, B)$ is the probability for the Markov chain to be in B when starting from q .

For the methods of Type 3, analogous results can be stated at the continuous level (for the underlying Markov processes). In computations, discrete-time approximations are used, and one recovers the case of a Markov chain, and the same kind of results as for methods of Type 2 hold true. For methods of Type 4, no general convergence result is known.

In the case when the sequence $(q^n)_{n \in \mathbb{N}}$ originates from a Markov chain on \mathcal{M} or from a discretized stochastic process on \mathcal{M} or on $T^*\mathcal{M}$ (methods of Types 2 and 3), additional questions arise. Indeed, instead of considering *one* realization starting from a given initial data, it is also possible to generate samples with the same computational cost by considering *several* shorter realizations starting either all from the same point or from different points (which constitute a pre-existing initial distribution). In this case, typical convergence results involve weighted total variation norms for the probability measures that are generated. In the sequel, we will often refer to this kind of convergence as the “convergence of densities” since, when the n -step probability transition kernel³ $P^n(q, \cdot)$ of the Markov chain and the invariant probability measure both admit densities with respect to the Lebesgue measure, the convergence in total variation norm implies the L^1 convergence of the densities. We can thus formulate the following two questions:

- **Question 3.** Does $\|P^n(q, \cdot) - \pi\|$ converge to zero when n goes to infinity for some (weighted) total variation norm?
- **Question 4.** If so, can the speed of convergence be estimated?

Again, if π is an invariant probability measure and if the accessibility condition (10) holds true, the answer to Question 3 is positive (see Thms. 3 and 4 below). A geometric convergence rate in ρ^n for some $\rho \in (0, 1)$ in some weighted total variation norm can also be obtained when the transition kernel P has some weak regularity properties and provided some Lyapunov condition holds true (namely condition (23) below, see Thm. 8). Let us point out that the Lyapunov condition (23) providing geometric convergence of the densities is not of the same nature as the condition (13) providing a convergence rate of the average along one sample path.

Let us mention that, in some applications, integrals such as (1) are sometimes computed using Blue Moon sampling techniques [9, 11, 15]. In this case, integrals over submanifolds (generally hypersurfaces) of \mathcal{M} have to be estimated. For such computations, the theoretical analysis is the same as the one we present in this article. From the numerical viewpoint, algorithms adapted to the constraint of sampling a hypersurface (and not the whole space) have to be used, namely projected algorithms for stochastic dynamics (see *e.g.* [12]) and SHAKE or RATTLE algorithms for deterministic evolutions (see [19], Chap. VII.1.4). The numerical analysis of the sampling properties of these specific algorithms is out of the scope of this article.

The article is organized as follows. We first describe and compare from a theoretical point of view the most popular methods to sample from the canonical distribution. In Section 1, we consider purely stochastic methods; stochastically perturbed Molecular Dynamics methods and deterministic thermostating methods are presented in Sections 2 and 3 respectively. In particular, in Section 2.2, we present some new convergence results for the Hybrid Monte Carlo scheme (see Thms. 7, 9 and 10). A summary of the main known results is presented in Table 1. We refer to the corresponding sections for notations and further explanations.

We then turn to a practical application of those methods in the case of linear alkane molecules in Section 4. The fact that some methods may work better than others, and that this depends on the situation at hand, is commonly accepted. However, these beliefs are usually only based on some qualitative comparisons, or on comparison with experimental data. In the latter case, discrepancies between numerical results and experimental results can come both from numerical and modelling approximations, so it is not easy to draw conclusions specifically on the numerical methods. Comparing the methods in a *quantitative* way is one of the main purposes of this article.

³For $q \in \mathcal{M}$ and B a Borel set of \mathcal{M} , $P^n(q, B)$ is the probability for the Markov chain to be in B when starting from q after exactly n steps. It is inductively defined from P by $P^0(q, B) = \mathbf{1}_B(q)$ and the induction rule

$$P^n(q, B) = \int_{\mathcal{M}} P(q, dq') P^{n-1}(q', B).$$

TABLE 1. Summary of the different sampling methods and their properties. The following shortenings have been used: MH (Metropolis-Hastings scheme), MD (Molecular Dynamics), i.i.d. r.v. (independently and identically distributed random variables), LLN (usual Law of Large Numbers, *i.e.* for i.i.d. variables), MC LLN (LLN for Markov chains), MP LLN (LLN for Markov processes).

Name	Rejection and Rejection control	Metropolized independence sampler (MIS)	Hybrid Monte-Carlo (HMC)	Biased Random-Walk	Langevin dynamics	Deterministic dynamics
Method	Sampling from the true density	MH with independent proposals	MH with MD proposals	Elliptic diffusion	Hypoelliptic diffusion	Extended MD system
Type	i.i.d variables	Markov chain	Markov chain	Markov process	Markov process	ODE
Questions 1, 2	LLN	MC LLN (conditions on the proposal function)	MC LLN (conditions on the potential energy)	MP LLN (Lyapunov condition)	MP LLN (Lyapunov condition)	Open question
	Any textbook	Section 1.3 and [43]	Section 2.2.2	Section 2.3.1	Section 2.4.1	
Questions 3, 4	-	Uniform ergodicity when a bounding function exists	Ergodicity	Geometric ergodicity (Lyapunov condition)	Geometric ergodicity (Lyapunov condition)	Open question
		Section 1.3	Section 2.2.2	Section 2.3.2	Section 2.4.1	
Numerical discretization	-	-	MH with velocity-Verlet	Euler-Maruyama or MALA	BBK algorithm or higher order schemes	Operator splitting
			Section 2.2.5	Section 2.3.3	Section 2.4.2	Section 3
Type	-	-	Markov chain	Markov chain	Markov chain	ODE discretization
Convergence	-	-	Same techniques and results as for the continuous scheme	Classical MC techniques	No result for usual schemes / results for specific schemes	Open question
			Section 2.2.5	Section 2.3.3 and [57]	Section 2.4.2	
Free parameters	Sampling function g	Proposal function g	Time step Δt , Integration time τ	Time step Δt	Time step Δt , Friction coefficient ξ	Number/values of thermostat masses, time step Δt
Rule	g "close to" f	g "close to" f	"Not too much rejection"	Acceptance rate $\simeq 0.5$	$\xi \Delta t$ "small" (0.01)	

1. PURELY STOCHASTIC METHODS

Purely stochastic methods consist in generating points in the position space according to the measure $d\pi(q) = f(q) dq$ given by (5), without referring to any physical dynamics of the system.

We briefly recall here four methods, the Rejection, Rejection control, Importance sampling, and Metropolized sampling methods. They all make use of a reference positive probability distribution $g(q)$, such that (i) it is easy to generate samples from g , and (ii) g is a “good” approximation of f , in a sense that will be made precise below.

1.1. Rejection method

The Rejection method [35] requires the knowledge of a probability density g which bounds f from above up to a multiplicative factor $c > 0$:

$$f \leq cg, \quad (11)$$

and from which it is easy to generate samples. For instance, when $\mathcal{M} = \mathbb{T}^{3N}$ (molecular system with periodic boundary conditions) and the potential energy V is bounded from below, a uniform density g may be used (but its efficiency is likely to be very poor). A less simple example is dealt with in [70]. The idea of the method is to draw proposals according to the density g and to accept them with probability $f/(cg)$.

Actually, a bound on the (non-normalized) distribution $\tilde{f}(q) = Z_q f(q) = e^{-\beta V(q)}$ is sufficient to run the algorithm. Such a bound reads $\tilde{f} \leq \tilde{c}g$, and is much easier to establish in practice since the normalization constant Z_q is unknown and very difficult to estimate. The proposals are then accepted with probability $\tilde{f}/(\tilde{c}g)$.

Finding a function g such that the constant c appearing in (11) is small is very important. It is indeed well-known [35] that, on average, generating one sample point requires c draws, that is c evaluations of the potential energy V , which is by far the most computationally expensive part of the calculation. This constant c is therefore of paramount importance. When the system dimension is small, it is usually possible to find g such that c is not too large, and therefore the method is very efficient. But when c is very large, the method is totally inefficient. In molecular simulation, it is usually very difficult to construct efficient sampling functions g for systems involving more than a few atoms. This can however still be done for some specific systems, such as crystals at low temperature, using Taylor expansions around the equilibrium position, and controlling the relevance of the expansion by Rejection control techniques (see Sect. 1.2 below). We refer to [70] for some examples of this approach.

Since the points generated by the Rejection algorithm are independent realizations of some random variable, usual convergence results such as the Law of Large Numbers and the Central Limit Theorem apply [18]. Let A be some observable over the position space, $(q^n)_{0 \leq n \leq N-1}$ be the sample generated by the method, and let us set

$$S_N(A) = \sum_{n=0}^{N-1} A(q^n). \quad (12)$$

If $\pi(|A|) < +\infty$, then the Law of Large Numbers holds true:

$$\lim_{N \rightarrow \infty} \frac{1}{N} S_N(A) = \int_{\mathcal{M}} A(q) f(q) dq = \int_{\mathcal{M}} A d\pi \quad \text{a.s.}$$

If $\pi(|A|^2) < +\infty$, then the Central Limit Theorem holds true: there exists $\gamma_A > 0$ such the following convergence in law holds:

$$(N\gamma_A)^{-1/2} S_N(\bar{A}) \xrightarrow{N \rightarrow \infty} \mathcal{N}(0, 1),$$

where $\bar{A} = A - \int_{\mathcal{M}} A d\pi$ and $\mathcal{N}(0, 1)$ is the standard Gaussian random variable.

1.2. Rejection control

It is often tricky to find a function g such that (11) is satisfied everywhere in \mathcal{M} . However, it is sometimes possible to find a sampling function g for which (11) is satisfied for most proposals \tilde{q} generated from g . In this case, the Rejection method presented in the previous section can be somewhat modified so that the non-global character of the bound is taken into account.

The Rejection control scheme [10, 35] allows one to handle proposals that violate the inequality (11) by an appropriate *a posteriori* reweighting. Let us just note here that this scheme can be recast [10] as an Importance sampling scheme, a method we will recall in Section 1.4.

1.3. Metropolized independence sampler

When c is large, the Rejection method may require many evaluations of the potential energy V . As c is unknown in practice, it is difficult to estimate *a priori* the computational efficiency of the method. Therefore, a stochastic method with a fixed computational cost could provide an interesting alternative.

The Metropolized independence sampler (MIS), presented *e.g.* in [35], Section 5.4.2, is one such method. Basically, it is a Metropolis-Hastings algorithm [22, 44] with i.i.d. proposals. Therefore, the generated sequence of points forms a Markov chain (see [46] for some definitions and properties of continuous state-space Markov chains).

We assume that the potential energy V is continuous. Considering an everywhere positive probability density g , let us set $w(q) = \frac{f(q)}{g(q)}$. The algorithm is as follows:

Algorithm 1.1 (metropolized independence sampling). *Consider an initial point q^0 . For $n \geq 1$,*

- (1) *generate a point \tilde{q} in \mathcal{M} from the density g ;*
- (2) *generate a random number $s \sim \mathcal{U}[0, 1]$;*
- (3) *if $s \leq \min \left\{ 1, \frac{w(\tilde{q})}{w(q^n)} \right\}$, set $q^{n+1} = \tilde{q}$, otherwise, set $q^{n+1} = q^n$;*
- (4) *replace n by $n + 1$ and go back to step (1).*

Let us now recall some convergence results for this method. We denote by P the transition kernel of this Markov chain.

1.3.1. Convergence of the average along one sample path

Let us denote by A some observable on the position space and by $(q^n)_{n \in \mathbb{N}}$ one realization of the MIS Markov chain starting from a given q^0 . The question under examination is that of the convergence of the empirical mean $\frac{1}{N} S_N(A)$ toward $\int_{\mathcal{M}} A(q) d\pi(q)$ where π is the canonical measure defined by (5) and $S_N(A)$ is defined by (12).

By construction, π is an invariant measure [35]. Therefore, condition (9) is satisfied. Condition (10) is also trivially satisfied whenever the support of f is a subset of the support of g . This is the case here since we have chosen a function g whose support is the whole position space \mathcal{M} .

Since conditions (9) and (10) are satisfied, a Law of Large Numbers (LLN) holds for almost all starting points, and Question 1 can therefore be answered positively. Indeed, recall the following theorem:

Theorem 1 ([46], Thm. 17.1.7). *Suppose conditions (9) and (10) are satisfied. Then, for any measurable function $A \in L^1(\pi)$,*

$$\lim_{N \rightarrow \infty} \frac{1}{N} S_N(A) = \int_{\mathcal{M}} A d\pi \quad a.s.$$

for almost all starting points $q^0 \in \mathcal{M}$, where $S_N(A)$ is defined by (12).

To obtain a convergence rate on $S_N(A)$, an additional condition is needed, such as:

$$\begin{aligned} &\text{There exist two measurable functions } L \geq \min\{1, A\} \text{ and } W \geq 0, \text{ a real number } b \\ &\text{and a petite set } C \text{ such that} \\ &\Delta W(q) \leq -L(q) + b\mathbf{1}_C(q), \quad \pi(W^2) < +\infty, \end{aligned} \tag{13}$$

where A is the observable under consideration and $\Delta W(q)$ is defined by

$$\forall q \in \mathcal{M}, \Delta W(q) = (PW)(q) - W(q) = \int_{\mathcal{M}} P(q, dy)W(y) - W(q). \tag{14}$$

The definition of petite sets can be found in [46]. Let us make the following remark, which will be very useful:

Remark 1. Under some regularity conditions that will always be met here (including the fact that the chain is weak Feller [46], Chap. 6), all compact subsets of \mathcal{M} are petite sets and the Markov chain is Doeblin [14]. As a consequence, when the state space \mathcal{M} is compact, the condition (13) holds true (choose $C = \mathcal{M}$, W and L arbitrary smooth functions and take b large enough).

Condition (13) allows one to obtain a Central Limit Theorem (CLT). For a given measurable function A such that $\pi(|A|) < +\infty$, let us formally define the function \hat{A} by the following Poisson equation:

$$-\Delta \hat{A} = A - \pi(A), \tag{15}$$

where Δ is defined as in (14). It is not clear in general whether \hat{A} is well-defined. This turns out to be the case when condition (13) is satisfied, and allows to state a CLT:

Theorem 2 ([46], Thm. 17.5.3). *Assume conditions (9), (10) and (13) hold true, and let A be a function such that $|A| \leq L$. Let $S_N(A)$ be defined by (12). There exists a function \hat{A} which satisfies (15), and the constant $\gamma_A^2 := \pi(\hat{A}^2 - (P\hat{A})^2)$ is well-defined, non-negative and finite. If $\gamma_A^2 > 0$, then, defining $\bar{A} = A - \pi(A)$,*

$$(N\gamma_A^2)^{-1/2}S_N(\bar{A}) \xrightarrow{N \rightarrow \infty} \mathcal{N}(0, 1),$$

this convergence being in law.

Since conditions (9), (10) and (13) are satisfied for the MIS chain, Question 2 can be answered positively for almost all starting points q^0 .

1.3.2. Convergence of the densities

To handle convergence of densities, it is necessary to introduce the total variation norm for a signed Borel measure ν , defined as

$$\|\nu\| = \sup_{h \text{ measurable}, |h| \leq 1} |\nu(h)| = \sup_{A \in \mathcal{B}(\mathcal{M})} \nu(A) - \inf_{A \in \mathcal{B}(\mathcal{M})} \nu(A). \tag{16}$$

Notice that convergence in total variation implies weak convergence.

Definition 1. A chain on \mathcal{M} is *ergodic* when

$$\forall q \in \mathcal{M}, \quad \lim_{n \rightarrow \infty} \|P^n(q, \cdot) - \pi\| = 0$$

where π is the invariant measure and P^n is the n -step probability transition kernel.

Recall the following theorem:

Theorem 3 ([46], Thm. 13.3.4). *If conditions (9) and (10) hold true, then*

$$\|P^n(q, \cdot) - \pi\| \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

for π -almost all starting points q .

The convergence in total variation norm implies convergence of the expectations only for bounded observables A . It is therefore not sufficient in practice. Fortunately, the ergodicity results can be strengthened in a straightforward way. For a given measurable non-negative function $W \geq 1$, let us define the W -total variation norm for a signed Borel measure μ as

$$\|\mu\|_W = \sup_{h \text{ measurable, } |h| \leq W} |\mu(h)|. \tag{17}$$

Then Theorem 3 can be readily extended to π -integrable functions A .

Theorem 4 ([46], Thm. 14.0.1). *Suppose that $A \geq 1$ is measurable and $\pi(|A|) < +\infty$. If conditions (9) and (10) hold true, then for π -almost all $q \in \mathcal{M}$,*

$$\|P^n(q, \cdot) - \pi\|_{|A|} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Since conditions (9) and (10) are satisfied, the MIS Markov chain is ergodic and Theorems 3 and 4 hold true. This answers Question 3.

Under an assumption which is reminiscent of the Rejection method setting, a simple uniform convergence rate (independent of the starting point q^0) can be obtained:

Theorem 5 ([43], Thm. 2.1). *If the probability density g used in the metropolized independence sampling scheme is such that*

$$\exists c, \forall q \in \mathcal{M}, \quad f(q) \leq cg(q),$$

then the scheme is geometrically ergodic with a uniform bound. In this case, for all $q^0 \in \mathcal{M}$,

$$\|P^n(q^0, \cdot) - \pi\| \leq (1 - c^{-1})^n.$$

This theorem gives an answer to Question 4. Note that in the particular case when $c = 1$ (that is when $f = g$ since both functions are densities of probability), the convergence is already achieved for $n = 1$. This is actually clear since in this case the MIS scheme samples from the true density!

1.4. Importance sampling

Importance sampling is a well-known general stochastic integration method. The underlying idea is to recast the integral $\mathbb{E}_\pi(A) = \int_{\mathcal{M}} A(q) f(q) dq$ as

$$\mathbb{E}_\pi(A) = \int_{\mathcal{M}} \left(A(q) \frac{f(q)}{g(q)} \right) g(q) dq$$

and to approximate the latter integral through a random sample $(q^n)_{0 \leq n \leq N-1}$ drawn according to the density g (see e.g. [35], Sect. 2).

The choice of the trial function g is crucial for the overall efficiency of the method. It should be a good approximation of f or, better, of $f(q)A(q)$. Since f is typically of exponential or Gaussian form, and A is most often bounded by a polynomial, f is usually the most important term in the product $f(q)A(q)$ as far as sampling issues are concerned. Besides, in applications, it is often the case that several integrals have to be computed, with different functions A . So g is often looked for as a good approximation of f .

Let us note that, for the computation of static quantities, the importance sampling method based on a density g outperforms the Rejection method based on the same density g [10].

2. STOCHASTICALLY PERTURBED MOLECULAR DYNAMICS METHODS

We first present in Section 2.1 the general framework of deterministic microcanonical (NVE) MD. In Section 2.2, we describe the Hybrid Monte Carlo (HMC) method, from both the theoretical and the numerical viewpoints, and give some new convergence results (see Thms. 7, 9 and 10). We then present the biased Random-Walk (BRW) in Section 2.3, and the Langevin dynamics in Section 2.4.

We assume in the sequel that $T^*\mathcal{M}$ is globally diffeomorphic to $\mathcal{M} \times \mathbb{R}^{3N}$, and actually identify the two sets for simplicity. We also assume that \mathcal{M} is globally diffeomorphic to \mathbb{R}^{3N} in Sections 2.3 and 2.4, and identify the two sets as well. Straightforward modifications allow to handle the other cases (such as systems with periodic boundary conditions or isolated systems parametrized by rigid-body motions and internal coordinates).

2.1. General framework for NVE molecular dynamics

The equations of motion (7) associated with the Hamiltonian (3) can be numerically integrated *e.g.* by the celebrated velocity-Verlet algorithm [76]

$$\begin{cases} p^{n+1/2} = p^n - \frac{\Delta t}{2} \nabla V(q^n) \\ q^{n+1} = q^n + \Delta t M^{-1} p^{n+1/2} \\ p^{n+1} = p^{n+1/2} - \frac{\Delta t}{2} \nabla V(q^{n+1}) \end{cases} \quad (18)$$

where Δt is the time step. The velocity-Verlet scheme is an *explicit* integrator: recall that in Statistical Physics one often considers systems with a large number of particles, making implicit algorithms untractable. The numerical flow associated to the velocity-Verlet algorithm shares two qualitative properties with the exact flow of (7): it is *time reversible* and *symplectic*, which are very important properties as far as the long time numerical integration of Hamiltonian dynamics is concerned (see [19], Chaps. VIII and IX, and [33]). This algorithm also asks for a unique evaluation of the forces $F = -\nabla V$ per time step. For all these reasons, it is the most commonly used algorithm in molecular dynamics.

The dynamics (7) cannot be used to generate points according to the canonical measure, because the energy (3) is preserved by the flow. Hence, the trajectory of the system remains on the submanifold of constant energy

$$T^*\mathcal{M}(E_0) = \{(q, p) \in T^*\mathcal{M}; H(q, p) = E_0\}$$

where $E_0 = H(q_0, p_0)$ is the energy of the initial data. Under some assumptions, the dynamics (7) can be used to compute microcanonical (NVE) ensemble averages, that is, averages over $T^*\mathcal{M}(E_0)$. The numerical analysis of this method (in the very simple case of completely integrable systems) can be read in [7, 8, 31]. To generate points according to the canonical measure, there is a need for stochastic perturbations to ensure that different energy levels will be explored, and eventually all of them. These considerations straightforwardly extend to the numerical case since symplectic methods such as (18) almost preserve the energy over extremely long times [19], Chapter IX.

2.2. Hybrid Monte Carlo

2.2.1. Presentation of the method

The Hybrid Monte Carlo method allows one to generate points in the position space distributed according to the canonical measure (5). It aims at combining the advantages of molecular dynamics (that approximates the physical dynamics of the system) and of Monte Carlo methods (that explore the position space more globally). It is in fact a Metropolis-Hastings algorithm, in which proposals are constructed using the NVE Hamiltonian flow of the system. This method has been first introduced by Duane *et al.* in [13] and partially analyzed from a mathematical viewpoint by Schütte in [63]. It has been used in [64, 65] to identify the metastable conformations of some biological systems. This method can be seen as a generalization of the Andersen thermostat method [3].

In the standard HMC setting, the sequence of generated positions forms a Markov chain of order one defined as follows:

Algorithm 2.1 (Hybrid Monte Carlo). *Consider an initial position $q^0 \in \mathcal{M}$ and $\tau > 0$. For $n \geq 0$,*

- (1) *generate momenta p^n according to the canonical distribution (4) and compute the energy $E^n = H(q^n, p^n)$ of the configuration (q^n, p^n) ;*
- (2) *compute $\Phi_\tau(q^n, p^n) = (q^{n,\tau}, p^{n,\tau})$, that is, integrate the NVE equations of motion (7) on the time interval $[0, \tau]$ starting from the initial data (q^n, p^n) ;*
- (3) *compute the energy $E^{n,\tau} = H(q^{n,\tau}, p^{n,\tau})$ of the new phase-space configuration. Accept the proposal $q^{n,\tau}$ with probability $p = \min(1, e^{-\beta(E^{n,\tau} - E^n)})$; more precisely, generate a random number $\alpha \in [0, 1]$ according to the uniform law, and set $q^{n+1} = q^{n,\tau}$ if $\alpha \leq p$ and $q^{n+1} = q^n$ otherwise;*
- (4) *replace n by $n + 1$ and go back to step (1).*

Let us emphasize that the proposal $q^{n,\tau}$ would always be accepted at step (3) if the NVE equations of motion, that are energy conserving, were integrated exactly. In practice, the time-step Δt used in the numerical integrator (18) can be chosen larger than in standard applications of MD since the dynamics of the system used to generate proposals is not constrained to accurately reproduce the physical dynamics of the system. On the other hand, it should not be too large; otherwise, the rejection rate would be large and the efficiency of the method would be low.

Let us notice that in the standard HMC method, only the end points of the MD trajectories are part of the sample. It is not completely clear whether taking into account the intermediate points of the generated MD trajectories in the sample would bias the sampling, *e.g.* if the final point is rejected, should these intermediate points be kept? See [49] for some work in this direction.

Let us also mention that there exist several refinements of the standard HMC scheme. In order to improve the acceptance rate, one could use a criterion based on a shadow Hamiltonian to accept or reject the new point [20, 27]. The idea is that this shadow Hamiltonian is preserved more accurately than the Hamiltonian (3) by the numerical trajectory. The bias introduced by this modification is corrected by a convenient reweighting, in the spirit of importance sampling. Another improvement consists in generating, after each NVE trajectory of length τ , some new momenta which are correlated with the previous ones [26, 28]. Of course, both approaches can be combined [1].

2.2.2. Convergence of the average along one realization

As above, let us denote by A some observable on the position space and by $(q^n)_{n \in \mathbb{N}}$ one realization of the HMC Markov chain starting from a given q^0 . Let Π_1 be the first coordinate field of the phase-space: $\Pi_1(q, p) = q$.

Convergence results for the HMC scheme have been published by Schütte in [63]. In this proof, the NVE Hamiltonian flow is assumed to satisfy two conditions:

- (1) a *mixing* condition, which reads as follows (see [63], Assumption 4.27): for every pair of open subsets $B, C \subset \mathcal{M}$, there exists $n_0 \in \mathbb{N}$ such that

$$\forall n \geq n_0, \int_B T^n \mathbf{1}_C(q) f(q) dq > 0,$$

where f is given by (5) and the function Tu is defined for any function $u : \mathcal{M} \rightarrow \mathbb{R}$ by

$$Tu(q) = \int_{\mathbb{R}^{3N}} u(\Pi_1 \Phi_{-\tau}(q, p)) \mathcal{P}(p) dp, \quad (19)$$

where Φ_τ is defined by (6). This condition amounts to a certain accessibility of the whole position space when starting from any point.

- (2) a so-called *momentum invertibility of the flow* condition (see [63], Def. 4.1). The flow Φ_τ is called momentum-invertible if the two following conditions hold true:

- For almost every $q \in \mathcal{M}$, there is an open set $M(q) \subset \mathbb{R}^{3N}$ such that the function $y_q : p \mapsto \Pi_1 \Phi_{-\tau}(q, p)$ is locally invertible in $M(q)$, that is, $\det \nabla_p y_q \neq 0$ for $p \in M(q)$.
- There is an $\eta > 0$ such that

$$\operatorname{ess\,inf}_{q \in \mathcal{M}} \int_{M(q)} \mathcal{P}(p) \, dp = \eta.$$

This condition states that the transition probabilities are bounded from below in some sense.

The following convergence result is given in [63]:

Theorem 6 ([63], Lem. 4.31 and Thm. A.24). *Under the assumptions (1) and (2) recalled above, for any measurable function $A \in L^1(\pi)$, it follows*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} A(q^n) = \int_{\mathcal{M}} A \, d\pi \quad a.s. \tag{20}$$

for almost all starting points $q^0 \in \mathcal{M}$, where $(q^n)_{n \in \mathbb{N}}$ is the sequence of points generated by the HMC Algorithm 2.1 where, at step (2), the NVE equations of motion (7) are exactly integrated.

Note that ergodicity results have also been proved [63], Corollary 4.33, as well as convergence results on the numerical flow [63], page 96 (in this latter case, $(q^n)_{n \in \mathbb{N}}$ in (20) is the sequence of points generated by the HMC Algorithm 2.1 where the NVE equations of motion (7) are now numerically integrated).

The conditions (1) and (2) recalled above are difficult to check in practice, and furthermore, it is not clear whether they are necessary. We present here a new convergence result, that does not require these assumptions.

Let us first consider the case when the NVE equations of motion are integrated exactly. The transition kernel P of the HMC Markov chain is defined by

$$\forall (q, B) \in \mathcal{M} \times \mathcal{B}(\mathcal{M}), \quad P(q, B) = \int_{\mathbb{R}^{3N}} \mathbf{1}_{\{\Pi_1 \Phi_\tau(q, p) \in B\}} \mathcal{P}(p) \, dp, \tag{21}$$

where the density \mathcal{P} is the canonical distribution on the momentum space given by (4).

As the phase-space canonical measure $\mu = \pi \otimes \kappa$ is an invariant measure for Φ_τ , it is clear that the position-space canonical measure π is an invariant measure for the HMC Markov chain (see *e.g.* [35], Sect. 9.3, for details). Therefore, condition (9) holds true.

We now consider the accessibility condition (10). This condition is not satisfied in general, for any potential energy. Consider for example a one-dimensional particle ($\mathcal{M} = \mathbb{R}$) of mass $m = 1$ subjected to the potential energy $V(q) = \frac{1}{2} q^2$. Then the trajectory $q(t)$ starting from q^0 with momentum p^0 is given by

$$q(t) = q^0 \cos(t) + p^0 \sin(t).$$

As already noticed by Mackenze in [36], taking $\tau = 2\pi$ leads to $q(\tau) = q^0$ whatever the choice of p^0 . The condition (10) is therefore clearly not satisfied, and the Markov chain is not ergodic. Of course this spurious effect only arises for special choices of τ . It is also linked to the fact that the period of the trajectory of the harmonic oscillator does not depend on the initial momentum.

To prove the accessibility condition (10), a first way is to make the additional assumption that the potential energy is bounded from above. We acknowledge that this assumption is often not satisfied in practice. Nevertheless, for some potential energies that do not satisfy this assumption, it is still possible to prove an accessibility condition by some explicit constructions, specific to the system at hand (see [70] for the case of a singular central potential energy). We will also consider in Section 2.2.6 another possibility, based on random integration times τ , that can be used for a larger class of potentials.

We now turn to proving the accessibility condition (10) under the assumption that V is bounded from above. This is the result of the following lemmas.

Lemma 1 (HMC accessibility – exact flow). *Let $\tau > 0$. Assume that V is in $C^1(\mathcal{M})$ and is bounded from above. Then for any $q, q' \in \mathcal{M}$ and any neighborhood \mathcal{V}' of q' , there holds*

$$P(q, \mathcal{V}') > 0.$$

Proof. The proof is based on the least action principle (LAP). Let us denote by

$$S(\phi) = \int_0^\tau \left(\frac{1}{2} \dot{\phi}^T(t) M \dot{\phi}(t) - V(\phi(t)) \right) dt$$

the action associated with the path $\phi \in \mathcal{H} = \{\phi \in H^1([0, \tau], \mathcal{M}) \mid \phi(0) = q, \phi(\tau) = q'\}$. Since V is bounded from above, there exists E_0 such that $V(q) \leq E_0$ for all $q \in \mathcal{M}$. Thus, S is bounded from below:

$$S(\phi) \geq - \int_0^\tau V(\phi(t)) dt \geq -E_0\tau.$$

Therefore, there exists a minimizing sequence $(\phi_n)_{n \in \mathbb{N}} \in \mathcal{H}$ such that $S(\phi_n) \rightarrow \inf_{\phi \in \mathcal{H}} S(\phi) = s > -\infty$. Without restriction, it can be assumed that $s \leq S(\phi_n) \leq s + 1$ for all $n \in \mathbb{N}$. Thus,

$$\int_0^\tau \dot{\phi}_n^T(t) M \dot{\phi}_n(t) dt = 2S(\phi_n) + 2 \int_0^\tau V(\phi_n(t)) dt \leq 2S(\phi_n) + 2\tau E_0 \leq 2(s + 1) + 2\tau E_0.$$

Therefore, $(\dot{\phi}_n)_{n \in \mathbb{N}}$ is bounded in $L^2([0, \tau], \mathcal{M})$. The sequence $(\phi_n)_{n \in \mathbb{N}}$ is then bounded in the space $H^1([0, \tau], \mathcal{M})$. Let $\phi \in H^1([0, \tau], \mathcal{M})$ such that (up to extraction) $\phi_n \rightharpoonup \phi$ in $H^1([0, \tau], \mathcal{M})$ -weak and $\phi_n \rightarrow \phi$ almost everywhere. Since \mathcal{H} is convex and closed in $H^1([0, \tau], \mathcal{M})$, the limit ϕ is actually in \mathcal{H} . Besides, it is easy to check that $\liminf_{n \rightarrow \infty} S(\phi_n) \geq S(\phi)$ (by lower semi-continuity on the kinetic energy and Fatou lemma on the potential energy), and this gives immediately

$$\inf_{\psi \in \mathcal{H}} S(\psi) = \min_{\psi \in \mathcal{H}} S(\psi) = S(\phi).$$

Thus ϕ minimizes S on \mathcal{H} . Therefore, the equation

$$M \ddot{\phi} = -\nabla V(\phi) \tag{22}$$

holds true on $(0, \tau)$ in the distributions sense. By standard regularity results, $\phi \in C^2([0, \tau], \mathcal{M})$ and (22) holds true in the sense of continuous functions. Hence the function ϕ is simply the solution of the Hamiltonian dynamics with $\phi(0) = q, \phi(\tau) = q'$ and initial velocity $\dot{\phi}(0)$.

Consider eventually a neighborhood \mathcal{V}' of q' . Then $P(q, \mathcal{V}') > 0$ is a straightforward consequence of the continuity of the solutions of (22) with respect to the initial velocity $\dot{\phi}(0)$. \square

Lemma 1 gives accessibility from any point to any open set. It is therefore not enough for condition (10) to hold true since it requires accessibility from one point to any arbitrary Borel set of positive Lebesgue measure. This asks for some regularity of the transition kernel, and in fact, some regularity of the dynamics, inferred from stronger assumptions on the potential energy V . More precisely, we have the following lemma:

Lemma 2 (HMC irreducibility – exact flow). *Assume that $V \in C^1(\mathcal{M})$ is bounded from above and ∇V is a globally Lipschitz function. Then the transition kernel of the HMC Markov chain satisfies*

$$\forall q \in \mathcal{M}, \forall B \in \mathcal{B}(\mathcal{M}), \mu^{\text{Leb}}(B) > 0 \Rightarrow P(q, B) > 0.$$

Proof. Consider $B \in \mathcal{B}(\mathcal{M})$ such that $\mu^{\text{Leb}}(B) > 0$, and $q \in \mathcal{M}$. We want to show that $P(q, B) > 0$ for P defined by (21). For the sake of simplicity, we assume here that all particle masses are equal to 1.

The proof is based on volume conservation in the phase space: any Borel set of final positions of positive measure can be reached from q and a set of momenta of positive measure. Denote $I_B(q) = \{p \in \mathbb{R}^{3N} \mid \Pi_1 \Phi_\tau(q, p) \in B\}$, and consider the function $\theta : I_B(q) \mapsto B$ such that $\theta(p) = \Pi_1 \Phi_\tau(q, p)$. This function is surjective according to the proof of the accessibility Lemma 1, so that $\theta(I_B(q)) = B$. Moreover, $P(q, B) = \int_{I_B(q)} \mathcal{P}(p) dp$. Therefore,

since \mathcal{P} is positive and continuous, it is enough to show that $\mu^{\text{Leb}}(I_B(q)) > 0$ in order to get $P(q, B) > 0$.

We proceed by contradiction. Suppose $\mu^{\text{Leb}}(I_B(q)) = 0$. We first note that θ is Lipschitz (of constant $\text{Lip}(\theta)$) since ∇V is continuous and globally Lipschitz by assumption, and $\tau > 0$ is fixed. Indeed, denote C the Lipschitz constant of ∇V and note that a solution of the equations of motion can be written as

$$q(\tau) = q + p\tau - \int_0^\tau (\tau - s)\nabla V(q(s)) ds.$$

For two different initial momenta p_1 and p_2 , we have

$$|q_1(t) - q_2(t)| \leq |p_1 - p_2|t + C \int_0^t (t - s)|q_1(s) - q_2(s)|ds.$$

By Gronwall lemma, there exists $c_\tau < +\infty$ such that

$$|q_1(\tau) - q_2(\tau)| \leq c_\tau |p_1 - p_2|,$$

hence θ is Lipschitz.

Since the Lebesgue measure and the Hausdorff measure \mathcal{H}^{3N} agree on \mathbb{R}^{3N} (see [16], Sect. 2.2, Thm. 2), and since the behavior of the Hausdorff measure under Lipschitz mappings is known [16], Section 2.4, Theorem 1, we obtain

$$\mu^{\text{Leb}}(B) = \mu^{\text{Leb}}(\theta(I_B(q))) = \mathcal{H}^{3N}(\theta(I_B(q))) \leq \text{Lip}(\theta)^{3N} \mathcal{H}^{3N}(I_B(q)) = \text{Lip}(\theta)^{3N} \mu^{\text{Leb}}(I_B(q)) = 0.$$

This gives $\mu^{\text{Leb}}(B) = 0$, in contradiction with the assumption $\mu^{\text{Leb}}(B) > 0$. □

Since conditions (9) and (10) are satisfied, a Law of Large Numbers (LLN) holds true for almost all starting points (see Thm. 1). We can therefore answer positively to Question 1:

Theorem 7. *Assume that $V \in C^1(\mathcal{M})$ is bounded from above and ∇V is a globally Lipschitz function. Let $(q^n)_{n \in \mathbb{N}}$ be the sequence of points generated by the HMC Algorithm 2.1 where, at step (2), the NVE equations of motion (7) are exactly integrated. Then*

$$\frac{1}{N} \sum_{n=0}^{N-1} A(q^n) \rightarrow \int_{\mathcal{M}} A(q) d\pi \quad a.s.$$

for almost all starting points $q^0 \in \mathcal{M}$.

2.2.3. Convergence of the densities

Since condition (9) is satisfied, and condition (10) holds true under the above assumptions on the potential energy (V is C^1 , bounded from above and ∇V is globally Lipschitz), the HMC Markov chain is ergodic (see Thm. 3). In particular,

$$\|P^n(q^0, \cdot) - \pi\| \rightarrow 0$$

for almost all starting points $q^0 \in \mathcal{M}$, where $\|\cdot\|$ denotes the total variation norm (16). We also get convergence in the $|A|$ -total variation norm (17) provided $\pi(|A|) < +\infty$ and $|A| \geq 1$ (see Thm. 4). This answers Question 3.

2.2.4. *Convergence rates*

We have not been able to state more sophisticated convergence results (Central Limit Theorem, geometric ergodicity) in the general HMC framework since they require stronger results on the Markov chain such as a drift condition (13) or a Lyapunov condition such as

$$\begin{aligned} &\text{There exist a measurable function } W \geq 1, \text{ real numbers } c > 0 \text{ and } b, \\ &\text{and a petite set } C \text{ such that} \\ &\forall q \in \mathcal{M}, \Delta W(q) \leq -cW(q) + b\mathbf{1}_C, \end{aligned} \tag{23}$$

where $\Delta W(q)$ is defined by (14). Let us make the following remark:

Remark 2. Under some regularity conditions that will always be met here (including the fact that the chain is weak Feller [46], Chap. 6), and when \mathcal{M} is compact, condition (23) is straightforwardly satisfied with the choice $C = \mathcal{M}$ (in view of Rem. 1, \mathcal{M} is a petite set and the Markov chain is Doeblin [14]) for any arbitrary smooth function W (taking b large enough).

When the state space is compact, conditions (13) and (23) hold true (in view of Rems. 1 and 2). We thus obtain a positive answer to Question 2 (see Thm. 2). We also obtain a positive answer to Question 4, in view of the following theorem:

Theorem 8 ([46], Thm. 15.0.1). *Assume conditions (9), (10) and (23) hold true. Then there exist $\rho < 1$ and $R < +\infty$ such that, for all q satisfying $W(q) < +\infty$,*

$$\|P^n(q, \cdot) - \pi\|_W \leq RW(q) \rho^n,$$

where P^n is the n -step probability transition kernel and $\|\cdot\|_W$ is the norm defined by (17).

2.2.5. *Numerical implementation: Method and convergence results*

It is standard to use the velocity-Verlet scheme (18) to integrate numerically the trajectories over times $\tau = k\Delta t$ for some integer k . Let us point out that the acceptance/rejection step (3) in Algorithm 2.1 ensures that the HMC Markov chain correctly samples the canonical measure π , so that no bias is introduced by the numerical discretization. The situation will be different for the Biased Random-Walk and the Langevin equation (see Sects. 2.3.3 and 2.4.2). We denote by $P_{\Delta t}$ the transition kernel of the Markov chain using the velocity-Verlet integrator (18) with time-step Δt .

The theoretical proof of convergence for the numerical version of HMC follows the same lines as the proof of convergence for the exact version using the Hamiltonian flow. The only difference lies in the additional acceptance/rejection step which does not modify the structure of the chain (for it does not change the accessibility properties of the chain). We only precise here the changes that have to be considered for the accessibility Lemma.

Lemma 3 (HMC accessibility – numerical flow). *Let $\tau > 0$. Assume that V is in $C^1(\mathcal{M})$ and is bounded from above on \mathcal{M} , and consider the numerical discretization scheme (18). Then for any $q, q' \in \mathcal{M}$ and any neighborhood \mathcal{V}' of q' , there holds*

$$P_{\Delta t}(q, \mathcal{V}') > 0.$$

Proof. The proof of Lemma 1 is based on the minimization of the action S over some space \mathcal{H} . Here, we extend this proof to the discretized case using a convenient approximation of this variational problem. There are several ways to discretize the variational problem, leading to different numerical schemes. In particular, the velocity-Verlet algorithm can be derived by minimizing the discretized action [38]

$$S_{\Delta t}(\Phi) = \Delta t \sum_{i=0}^{k-1} \left[\frac{1}{2} \left(\frac{q^{i+1} - q^i}{\Delta t} \right)^2 - \frac{V(q^{i+1}) + V(q^i)}{2} \right], \tag{24}$$

where $\tau = k\Delta t$ (we again assumed here that all particle masses are equal to 1).

The minimization is performed on the sequences $\Phi = \{q^0, q^1, \dots, q^k\}$ with the constraints $q^0 = q$ and $q^k = q'$. The quantity $S_{\Delta t}$ is still bounded from below for a potential energy bounded from above. Hence, there exists a minimizing sequence $(\Phi_n)_{n \in \mathbb{N}} = (\{q^{0,n}, q^{1,n}, \dots, q^{k,n}\})_{n \in \mathbb{N}}$. Each difference $q^{i+1,n} - q^{i,n}$ is easily seen to be bounded, thus each component $q^{i,n}$ is in fact bounded. We can consider $\bar{\Phi} = (\bar{q}^0, \dots, \bar{q}^k)$ such that, upon extraction, we have $q^{i,n} \rightarrow \bar{q}^i$ when $n \rightarrow \infty$ for each i . Moreover, $S(\bar{\Phi}) = \min_{\Phi} S(\Phi)$. The optimality conditions then read

$$\bar{q}^{i+1} = 2\bar{q}^i - \bar{q}^{i-1} - \Delta t^2 \nabla V(\bar{q}^i)$$

for $1 \leq i \leq k - 1$. We recognize the Verlet scheme. As in addition $\bar{q}^0 = q$ and $\bar{q}^k = q'$, this shows that given two points q, q' , there is a path connecting them using a numerical velocity-Verlet trajectory with initial velocity $\bar{p}^0 = \frac{\bar{q}^1 - \bar{q}^0}{\Delta t} + \frac{\Delta t}{2} \nabla V(\bar{q}^0)$. By continuity, for initial velocities close to \bar{p}^0 , the endpoint of the resulting trajectory remains in a neighborhood of q' . The proof is complete. \square

We can now state a Law of Large Number theorem (see Thm. 1):

Theorem 9. *Assume that $V \in C^1(\mathcal{M})$ is bounded from above and ∇V is globally Lipschitz. Let $(q^n)_{n \in \mathbb{N}}$ be the sequence of points generated by the HMC Algorithm 2.1 where, at step (2), the NVE equations of motion (7) are numerically integrated by (18). Then*

$$\frac{1}{N} \sum_{n=0}^{N-1} A(q^n) \rightarrow \int_{\mathcal{M}} A(q) d\pi \quad a.s.$$

for almost all starting points $q^0 \in \mathcal{M}$.

2.2.6. Random Time Hybrid Monte Carlo

In order to prove convergence of the classical HMC scheme, we have assumed in the previous section that the potential energy is bounded from above. As explained in the discussion just above Lemma 1, another possibility is to modify the HMC scheme as in [36]. The modification consists in transforming the fixed parameter τ into a random variable, distributed with a density $\mathcal{T}(\tau)$. This ensures that resonance effects are avoided. We call this scheme ‘‘Random Time Hybrid Monte Carlo’’ (RTHMC).

The only property required on \mathcal{T} is that \mathcal{T} is continuous and positive on \mathbb{R}_+ . The corresponding Markov transition kernel reads, for $q \in \mathcal{M}$ and $B \in \mathcal{B}(\mathcal{M})$,

$$P(q, B) = \int_{\mathbb{R}^{3N} \times \mathbb{R}_+} \mathbf{1}_{\{\Pi_1 \Phi_{\tau}(q,p) \in B\}} \mathcal{P}(p) \mathcal{T}(\tau) dp d\tau. \tag{25}$$

Notice that π is still an invariant probability measure for this Markov chain, so condition (9) holds true. Therefore, to get convergence results, we only need to show condition (10). This is done in two steps, as for the classical HMC scheme.

The first lemma states that there is a positive probability to go from one state q to any neighborhood of any state q' in one RTHMC iteration.

Lemma 4 (RTHMC accessibility). *Assume that $V \in C^1(\mathcal{M})$ and $D^2V \in L^\infty(\mathbb{R}^3)$. Then for any $q_0, q_1 \in \mathcal{M}$, there exists $\tau^* > 0$ such that, for all $0 < \tau \leq \tau^*$, there exists $p \in \mathbb{R}^{3N}$ with $\Pi_1 \Phi_{\tau}(q_0, p) = q_1$.*

Proof. A similar idea is used in [63] in a slightly different context. If V is identically equal to zero, then going from q_0 to q_1 is possible through the choice of (say) the initial momenta $p^* = M(q_1 - q_0)/\tau$ for some evolution time $\tau > 0$. We then consider the rescaled equation

$$M\ddot{q}_\epsilon(t) = -\epsilon \nabla V(q_\epsilon(t)) \tag{26}$$

and the associated flow ϕ_ϵ . Setting

$$F(\epsilon, p) = \phi_\epsilon(\tau, q_0, p) - q_1,$$

the function F is $C^1(\mathbb{R} \times \mathbb{R}^{3N})$ (we use here the assumption $D^2V \in L^\infty(\mathbb{R}^3)$), $F(0, p^*) = 0$ and $\partial_p F(0, p^*) = \tau M^{-1}$ is invertible. In view of the implicit function theorem, there exists $\epsilon^* > 0$ such that for all $0 \leq \epsilon \leq \epsilon^*$, there exists p_ϵ such that $F(\epsilon, p_\epsilon) = 0$.

This shows (by the change of variables $t \rightarrow \epsilon t$ in (26) for $0 < \epsilon \leq \epsilon^*$) that $\Pi_1 \Phi_{\epsilon\tau}(q_0, p_\epsilon/\epsilon) = q_1$. □

Condition (10) can then be obtained in the same way as for the classical HMC scheme, the proof following the same lines as for Lemma 2.

Lemma 5 (RTHMC irreducibility). *Provided that $V \in C^1(\mathcal{M})$ and $D^2V \in L^\infty(\mathcal{M})$, the transition kernel (25) of the RTHMC Markov chain satisfies condition (10).*

Proof. Consider $B \in \mathcal{B}(\mathcal{M})$ such that $\mu^{\text{Leb}}(B) > 0$, and $q \in \mathcal{M}$. We want to show that $P(q, B) > 0$ for P defined by (25). For the sake of simplicity, we assume here that all particle masses are equal to 1.

The proof relies on the fact that, for a given q and for $\tau > 0$ small enough, the mapping $p \mapsto \Pi_1 \Phi_\tau(q, p)$ is invertible. Denote $J_B(q, \tau) = \{p \in \mathbb{R}^{3N} \mid \Pi_1 \Phi_\tau(q, p) \in B\}$, and consider $\psi_\tau : J_B(q, \tau) \rightarrow B$ such that $\psi_\tau(p) = \Pi_1 \Phi_\tau(q, p)$.

We first show that ψ_τ is an injective function for $\tau > 0$ small enough. From the equations of motion,

$$\psi_\tau(p) = q + p\tau - \int_0^\tau (\tau - s)\nabla V(\psi_s(p)) \, ds.$$

Hence

$$\nabla_p \psi_\tau(p) = \tau \text{Id} - \int_0^\tau (\tau - s)D^2V(\psi_s(p)) \cdot \nabla_p \psi_s(p) \, ds. \tag{27}$$

Set $\alpha_R(s) = \sup_{|p| \leq R} \|\nabla_p \psi_s(p) - s \text{Id}\|_\infty$. Since ∇V is a globally Lipschitz function, we have

$$\alpha_R(\tau) \leq C \left(\int_0^\tau (\tau - s)\alpha_R(s) \, ds + \frac{\tau^3}{6} \right) \tag{28}$$

with $C = \|D^2V\|_{L^\infty(\mathcal{M})}$. We now consider $\tau_c^R = \sup\{\tau'; \alpha_R(\tau) \leq \tau/2 \text{ for all } \tau \in [0, \tau']\}$. From (28), we obtain that $\tau_c^R \geq \sqrt{2/C}$. Hence, we have

$$\forall \tau \in [0, \sqrt{2/C}], \alpha_R(\tau) \leq \frac{\tau}{2}.$$

Inserting this inequality in (28), we also obtain that

$$\forall \tau \in [0, \sqrt{2/C}], \alpha_R(\tau) \leq C \frac{\tau^3}{4}.$$

It follows that

$$\alpha(s) = \sup_{p \in \mathbb{R}^{3N}} \|\nabla_p \psi_s(p) - s \text{Id}\|_\infty \leq C \frac{\tau^3}{4}. \tag{29}$$

Now,

$$\begin{aligned} (\psi_\tau(p_1) - \psi_\tau(p_2)) \cdot (p_1 - p_2) &= \int_0^1 (p_1 - p_2) \cdot \nabla_p \psi_\tau(p_2 + s(p_1 - p_2)) \cdot (p_1 - p_2) \, ds \\ &= \int_0^1 (p_1 - p_2) \cdot (\nabla_p \psi_\tau(p_2 + s(p_1 - p_2)) - \tau \text{Id}) \cdot (p_1 - p_2) \, ds \\ &\quad + \tau |p_1 - p_2|^2 \end{aligned}$$

Let us suppose that $\psi_\tau(p_1) = \psi_\tau(p_2)$. Then

$$\tau|p_1 - p_2|^2 \leq \alpha(\tau) |p_1 - p_2|^2 \leq \frac{\tau}{2} |p_1 - p_2|^2$$

and we obtain $p_1 = p_2$. Hence, the mapping $J_B(q, \tau) \ni p \mapsto \psi_\tau(p) \in B$ is an injective function for $\tau \leq \sqrt{2/C}$.

We now show that this mapping is onto. We consider, for $q' \in B$, the C^1 function

$$G(\tau, p, q') = \psi_\tau(p) - q'.$$

Let us fix $q^* \in B$ such that, for all $\epsilon > 0$, $\mu^{\text{Leb}}(B \cap B_\epsilon(q^*)) > 0$. Lemma 4 shows that there exists $\tau^* > 0$ such that

$$\forall \tau, 0 < \tau < \min(\tau^*, \sqrt{2/C}), \quad \exists p \in \mathbb{R}^{3N} \text{ s.t. } G(\tau, p, q^*) = 0.$$

Since $\partial_p G = \partial_p \psi_\tau$ is invertible (using (27) and the bound (29)), we obtain from the implicit function theorem that there exists a neighborhood $\mathcal{V}_\tau(p)$ of p and a neighborhood $\mathcal{V}_\tau(q^*)$ of q^* such that, for any $q' \in \mathcal{V}_\tau(q^*)$, there exists $p' \in \mathcal{V}_\tau(p)$ with $G(\tau, p', q') = 0$. This gives the desired result.

Thus, for $0 < \tau < \min(\tau^*, \sqrt{2/C})$, the mapping ψ_τ is one-to-one from $\mathcal{V}_\tau(p)$ onto $\mathcal{V}_\tau(q^*)$. Using (29), we also have $\text{Det}(\nabla_p \psi_\tau(p)) = \tau^{3N}(1 + o(1))$ uniformly in p . Hence, the mapping ψ_τ is invertible and $\text{Det}(\nabla_p \psi_\tau^{-1}(q)) = \tau^{-3N}(1 + o(1))$.

We are now in position to show that $P(q, B) > 0$. By contradiction, assume $P(q, B) = 0$. Then

$$\int_{\mathbb{R}^{3N}} \mathbf{1}_{\{\Pi_1 \Phi_\tau(x, p) \in B\}} \mathcal{P}(p) dp = 0 \text{ for almost all } \tau. \text{ Therefore, for almost all } 0 < \tau < \min(\tau^*, \sqrt{2/C}), \text{ we have}$$

$$\int_{\mathbb{R}^{3N}} \mathbf{1}_{\{\Pi_1 \Phi_\tau(x, p) \in B \cap \mathcal{V}_\tau(q^*)\}} \mathcal{P}(p) dp = 0. \text{ Thus, a change of variable shows that}$$

$$\int_{B \cap \mathcal{V}_\tau(q^*)} \mathcal{P}(\psi_\tau^{-1}(q)) |\text{Jac}(\psi_\tau^{-1}(q))| dq = 0$$

for almost all $0 < \tau < \min(\tau^*, \sqrt{2/C})$. This is however not possible since \mathcal{P} is continuous and positive, $\mu^{\text{Leb}}(B \cap \mathcal{V}_\tau(q^*)) > 0$, and $|\text{Jac}(\psi_\tau^{-1}(q))| \sim \tau^{-3N}$ when $\tau \rightarrow 0$ so that $|\text{Jac}(\psi_\tau^{-1}(q))| > 0$ for τ small enough. \square

We then get convergence of the average along a sample path (see Thm. 1):

Theorem 10. *Assume that $V \in C^2(\mathcal{M})$ and $D^2V \in L^\infty(\mathcal{M})$. Let $(q^n)_{n \in \mathbb{N}}$ be the sequence of points generated by the RTHMC algorithm where the NVE equations of motion (7) are exactly integrated. Then*

$$\frac{1}{N} \sum_{n=0}^{N-1} A(q^n) \rightarrow \int_{\mathcal{M}} A(q) d\pi \text{ a.s.}$$

for almost all starting points $q^0 \in \mathcal{M}$.

We also obtain ergodicity and convergence of the densities as for the classical HMC scheme under the assumptions of Lemma 5 (see Thm. 3).

For the numerical discretization, we have to consider times $\tau_n = n\Delta t$, and a probability \mathcal{T} on \mathbb{N} such that $\mathcal{T}(n) > 0$ for all n (a Poisson law for instance). The time-step Δt has to be chosen small enough such that no resonance effect can appear.

2.3. Biased Random-Walk

The so-called biased Random-Walk, also known as the Brownian dynamics, or the overdamped Langevin dynamics, is defined by the fictitious dynamics

$$dq_t = -\nabla V(q_t)dt + \sigma dW_t, \tag{30}$$

where $(W_t)_{t \geq 0}$ is a $3N$ -dimensional standard Wiener process and $\sigma = (2/\beta)^{1/2}$. The term ‘‘biased’’ refers to the fact that the brownian trajectories are affected by the drift term $-\nabla V$ which tends to draw them toward the local minima of V . The infinitesimal generator \mathcal{A} associated with the biased Random-Walk (30) is defined by

$$\mathcal{A}g = -\nabla V \cdot \nabla g + \frac{\sigma^2}{2} \Delta g, \tag{31}$$

for $g \in C^2(\mathbb{R}^{3N})$. We denote by P^t the Markov semigroup associated with (30). Trajectorial existence and uniqueness for (30) is classical for globally Lipschitz force-fields [21,37], namely for potential energies V satisfying for some positive constant L

$$\forall(x, y) \in \mathbb{R}^{3N} \times \mathbb{R}^{3N}, \quad |\nabla V(x) - \nabla V(y)| \leq L|x - y|. \tag{32}$$

When this condition is not satisfied, it is possible to conclude to trajectorial existence and uniqueness for locally Lipschitz force-fields under the following hypothesis [21,37]: there exist a function $W(q) \in C^2(\mathbb{R}^{3N})$ that goes to infinity at infinity and a positive constant c such that

$$\mathcal{A}W \leq cW. \tag{33}$$

Besides, under assumption (32) or (33), one can prove that the Markov process (30) is Feller [45].

From the Fokker-Planck equation associated with (30), it is easy to check that

$$\pi \text{ is an invariant probability measure of (30),} \tag{34}$$

where π is the canonical position space distribution (5).

2.3.1. Convergence of the time average along one sample path

Let us consider the time average

$$S_T(A) = \frac{1}{T} \int_0^T A(q_t^x) dt, \tag{35}$$

where q_t^x is a sample path of (30) with the deterministic initial condition $q_0 = x$. Convergence results analogous to the results obtained for Markov chains can be extended to Markov processes, with an average (35) still taken only over one realization of the process (see [72] for a seminal contribution (that also considers discretization issues), [73, 74] for improvements and refinements, and [53] for a recent review).

To obtain an almost sure convergence of $S_T(A)$ to the position space average (and thus a positive answer to Question 1), the following theorem can be used:

Theorem 11 ([45], Thm. 8.1). *Assume that the process q_t defined by (30) is Feller, that condition (34) holds true as well as the following condition:*

$$\text{For all } t, \text{ for all } q \in \mathbb{R}^{3N} \text{ and all open sets } \mathcal{O} \subset \mathbb{R}^{3N}, \quad P^t(q, \mathcal{O}) > 0. \tag{36}$$

Then, for π -almost every $q \in \mathbb{R}^{3N}$ and for any $A \in L^1(\pi)$,

$$\lim_{T \rightarrow \infty} S_T(A) = \int_{\mathbb{R}^{3N}} A(q) d\pi \quad a.s.$$

If ∇V is globally Lipschitz, then (36) holds true by standard results [59]. In other cases, a simple way to check condition (36) is to use a controllability argument inspired from [42], Lemma 3.4. Central Limit Theorems (which would provide a convergence rate of $S_T(A)$ towards its limit and thus provide an answer to Question 2) can also be stated. We refer for example to [25].

2.3.2. *Convergence of the densities*

Ergodicity holds true whenever conditions (34) and (36) are satisfied (see [45], Thm. 6.1). Question 3 can therefore be answered positively. To get an exponential convergence rate (in the W -total variation norm (17)), that is, to answer Question 4, one needs to show the stronger condition

$$\mathcal{A}W(q) \leq -cW(q) + b\mathbf{1}_C(q), \tag{37}$$

where $W \geq 1$ is a measurable function going to infinity at infinity, $c > 0$, $b \in \mathbb{R}$ and C is a compact set (compare this condition with condition (23) for Markov chains). We do not address this question in the present article (see [42, 73, 74] for examples of such studies).

2.3.3. *Numerical implementation*

The Euler-Maruyama numerical scheme associated to (30) reads, when taking integration steps $h = \Delta t^2/2$:

$$q^{n+1} = q^n - \frac{\Delta t^2}{2} \nabla V(q^n) + \beta^{-1/2} \Delta t R^n, \tag{38}$$

where $(R^n)_{n \in \mathbb{N}}$ is a sequence of i.i.d. $3N$ -dimensional standard Gaussian random vectors.

For globally Lipschitz force-fields, the Euler-Maruyama scheme (38) converges: if the process q_t defined by (30) is ergodic, then the numerical Markov chain is ergodic and its invariant measure is close to the invariant measure of the original process (for Δt small enough) ([42], Thm. 7.3).

However, for non-globally Lipschitz force-fields, it is not sufficient to consider the discretization (38) of the diffusion process alone. Indeed, examples of non-globally Lipschitz force-fields are known for which the Euler-Maruyama scheme fails [42, 57]. There are two ways out of this situation. First, convenient discretizations of (38) using some implicit integration can be used. Under some assumptions on the potential energy V , the corresponding numerical scheme converges: first, there exists an invariant probability measure for the Markov chain formalizing the algorithm; second, empirical averages of observables (with at most polynomial growth) converge to position space averages up to $O(\Delta t)$ terms (see [74]). However, implicit methods become untractable for large systems. Another approach may then be considered, the so-called ‘‘Metropolis-adjusted Langevin⁴ algorithm’’ (MALA), proposed by Roberts and Tweedie in [57], which corrects the Euler-Maruyama discretization (38) by an additional acceptance/rejection step in a Metropolis-Hastings fashion. Therefore, there is no bias in the measure sampled. The algorithm consists in generating proposal steps using (38), and accepting or rejecting them according to a Metropolis-Hastings rule. Note that choosing the time step h of the MALA algorithm such that $h = \Delta t^2/2$ makes the comparison between the MALA algorithm and the Hybrid Monte Carlo methods easier (see [70] for further details).

In the case of the MALA algorithm, using a potential energy $V \in C^1(\mathbb{R}^{3N})$ is enough to satisfy condition (10). Since π is by construction an invariant probability measure (and therefore condition (9) holds true), the Markov chain formalizing the algorithm is ergodic for almost all starting points, and the convergence results stated in Theorems 1 and 3 apply. On the other hand, conditions ensuring the Central Limit Theorem and geometric ergodicity (conditions (13) and (23), see Thms. 2 and 8) are not easy to check. We refer to [57, 58] for such studies.

The only adjustable parameter of the algorithm is the time-step Δt . The rejection rate is a good indicator of efficiency. It is indeed well-known that a good sampling is a trade-off between decorrelation (to this end, larger time-steps are required) and acceptance rate (the larger the time-step, the larger the rejection rate). We refer for example to [56] where it is shown that, for tensorized distributions, the asymptotical optimal acceptance rate, when the dimension of the position space \mathcal{M} goes to infinity, is 0.574. This theoretical result does not extend to more complicated situations. However, numerical experiments show that an acceptance/rejection rate about 50% leads to a rather efficient method.

⁴The term ‘‘Langevin’’ does not refer here to the Langevin dynamics as known in the Physics literature (see Sect. 2.4). In the Probability and Statistics fields, it is, for some authors, the name for the biased Random-Walk.

In Section 4, we present numerical results obtained both with the Euler-Maruyama scheme and with the MALA scheme.

2.4. Langevin dynamics

The paradigm of Langevin dynamics is to introduce in the Newton equations of motion (7) some fictitious brownian forces modelling fluctuations, balanced by viscous damping forces modelling dissipation. More precisely, the equations of motion read here

$$\begin{cases} dq_t = M^{-1}p_t dt \\ dp_t = -\nabla V(q_t) dt - \xi M^{-1}p_t dt + \sigma dW_t \end{cases} \quad (39)$$

where $(W_t)_{t \geq 0}$ is a $3N$ -dimensional Wiener process. The parameters ξ and σ represent the magnitude of the fluctuations and of the dissipation respectively, and are linked by the fluctuation-dissipation relation:

$$\sigma = (2\xi/\beta)^{1/2}, \quad (40)$$

where $\beta = 1/k_B T$. Therefore, there remains one adjustable parameter in the model. Let us remark that the biased Random-Walk (30) is obtained from the Langevin dynamics (39) by letting the mass matrix M go to zero and by setting $\xi = 1$, which amounts here to rescaling the time.

The proof of trajectorial existence and uniqueness follows the same lines as for the biased Random-Walk case, with the same kind of assumptions (globally Lipschitz force fields ∇V or a Lyapunov condition analogous to (33)). It is straightforward to show that the canonical probability measure (2) is a steady state of the Fokker-Planck equation associated with (39).

2.4.1. Convergence results

The same results hold true for the Langevin process as the ones stated in Sections 2.3.1 and 2.3.2 for the biased Random-Walk, the proofs following the same lines. We refer to [42] for further details concerning condition (36) (where \mathbb{R}^{3N} is to be replaced by $\mathbb{R}^{3N} \times \mathbb{R}^{3N}$ and P^t is now the Markov semigroup associated with the Langevin dynamics). We also refer to [23] for a remarkable work allowing, under some assumptions of local regularity and growth at infinity on the potential energy V , to obtain geometrical convergence of the density $P^t(q, \cdot)$ toward the invariant measure, in some weighted Sobolev norms. In particular, estimates of the convergence rate involving M , ξ , β and V , can be explicitly derived.

Questions 1 and 3 can therefore be answered positively. Question 4 can also be answered positively when a convenient drift condition can be stated (condition (37) where \mathcal{A} is now the infinitesimal generator associated to (39)).

2.4.2. Numerical implementation

There are several ways to compute numerically an invariant distribution using a Langevin dynamics: with a Metropolized scheme as for the biased Random-Walk case (see [61] for an application to Variational Monte-Carlo); with convenient discretizations and a step-size Δt sufficiently small ensuring the sampling from an invariant measure close to the canonical measure (2); or by extending usual NVE schemes used in deterministic MD simulations to the case of the Langevin dynamics. The latter way is the most convenient in many applications, and allows usually to take larger time steps than for pure NVE simulations since the scheme is intrinsically more stable in view of its dissipative properties. Unfortunately, to our knowledge, there is no theoretical proof of convergence for the resulting scheme. Let us now detail successively the last two approaches.

General results of error analysis hold true for the numerical discretization of the Langevin equation for globally Lipschitz force fields [42]. In this case, the resulting numerical Markov chain is ergodic for usual discretization schemes (including the Euler-Maruyama discretization) and their invariant measures are close to the invariant measure of the original process (for Δt small enough).

The results are not the same for only locally Lipschitz force fields. Some classes of discretized schemes however behave properly under additional assumptions on the potential energy. This is the case for the so-called split-step Backward Euler-method proposed in [42]. Applied to the Langevin equation (39), this algorithm reads

$$\begin{cases} q^{n+1} = q^n + \Delta t M^{-1} p^* \\ p^* = p^n - \xi \Delta t M^{-1} p^* - \Delta t \nabla V(q^{n+1}) \\ p^{n+1} = p^* + \sigma \sqrt{\Delta t} R^n \end{cases} \quad (41)$$

where $(R^n)_{n \in \mathbb{N}}$ is a sequence of $3N$ -dimensional i.i.d. Gaussian random vectors. Unfortunately, this method is implicit (see the first two equations, to be solved for (q^{n+1}, p^*)), therefore not convenient for MD simulations of large systems. The following *explicit* scheme is therefore preferred

$$\begin{cases} p^* = p^n - \xi \Delta t M^{-1} p^* - \Delta t \nabla V(q^n) \\ q^{n+1} = q^n + \Delta t M^{-1} p^* \\ p^{n+1} = p^* + \sigma \sqrt{\Delta t} R^n \end{cases} \quad (42)$$

where $(R^n)_{n \in \mathbb{N}}$ is a sequence of $3N$ -dimensional i.i.d. Gaussian random vectors. The above algorithm is an extension of the symplectic Euler scheme to the stochastic setting.

We now turn to the numerical analysis of (42). Let us denote by \mathcal{F}_n the σ -algebra of events up to and including the n -th iteration. We need to prove condition (9) and condition (10) to state a Law of Large Number theorem (see Thm. 1). The accessibility condition (10) is easily seen to be satisfied (by arguments similar to those of Sect. 2.3.1 in this time discrete case). We now prove condition (9), that is, the existence of an invariant probability measure. For this purpose, we need to make some assumptions on the potential energy V , similar to those of [42], to state a Lyapunov inequality for the discretized process. Indeed, we want to make use of the following theorem:

Theorem 12 ([42], Thm. 2.5). *Denote by P the transition kernel associated to the Markov chain formalizing (42), assumed to be Feller. Assume that (10) is satisfied and that there exist a function $W_{\Delta t}(q, p) \geq 1$, going to infinity at infinity, and two real numbers $b \in (0, 1)$ and $c > 0$ such that*

$$\mathbb{E}(W_{\Delta t}(q^{n+1}, p^{n+1}) | \mathcal{F}_n) \leq b \mathbb{E}(W_{\Delta t}(q^n, p^n)) + c, \quad (43)$$

where (q^n, p^n) is the discrete trajectory given by (42). Then there exists an invariant probability measure $\mu_{\Delta t}$, and condition (9) holds true.

The numerical scheme then converges (with respect to the measure $d\mu_{\Delta t}$) in the sense of Questions 1 to 4. The question of estimating the distance between $\mu_{\Delta t}$ and the canonical measure μ has been addressed in *e.g.* [42, 74].

Let us now find $W_{\Delta t}$, b and c satisfying (43). We assume that the potential energy V is in $C^2(\mathbb{R}^{3N})$ and satisfies a one-sided Lipschitz condition: there exists $C > 0$ such that

$$\forall a, b \in \mathbb{R}^{3N}, (\nabla V(a) - \nabla V(b)) \cdot (a - b) \leq C|a - b|^2. \quad (44)$$

We also assume that there exist $A, B > 0$ such that

$$\forall q \in \mathbb{R}^{3N}, -\nabla V(q) \cdot M^{-1}q \leq A - B \left(V(q) + \frac{\xi^2}{4} q^T M^{-1}q \right). \quad (45)$$

These conditions are satisfied for example for potential energies growing quadratically at infinity. The following result, strongly inspired from [42], can then be stated:

Lemma 6. *Let (q^n, p^n) be the discrete trajectory given by (42). Let us assume that V is bounded from below and let us set $m = \max \{m_1, \dots, m_N\}$,*

$$W(q, p) = 1 + \frac{1}{2} p^T M^{-1} p + \frac{\xi^2}{4} q^T M^{-1} q + V(q) - \inf V + \frac{\xi}{2} p^T M^{-1} q \tag{46}$$

and $W_{\Delta t}(q, p) = W(q, p) + \frac{\xi}{4m^2} \Delta t |p|^2$. Let us assume that (44) and (45) are satisfied, and that

$$0 \leq \Delta t \leq \frac{\xi}{\xi^2/m + 4C}. \tag{47}$$

Then $W_{\Delta t}$ satisfies (43) for some $c > 0, 0 < b < 1$.

We just give here the outline of the proof, and refer to [70] for further details.

Proof. Consider the numerical scheme (42). Some computations give

$$\begin{aligned} W(q^{n+1}, p^*) - W(q^n, p^n) &\leq -\frac{\xi \Delta t}{2m^2} \left(1 - \frac{\xi \Delta t}{2m}\right) |p^*|^2 - \frac{\xi \Delta t}{2} \nabla V(q^n) \cdot M^{-1} q^n \\ &\quad + V(q^n + \Delta t M^{-1} p^*) - V(q^n) - \Delta t \nabla V(q^n) \cdot M^{-1} p^*. \end{aligned}$$

The one-sided Lipschitz condition (44) allows to handle the term $V(q^n + \Delta t M^{-1} p^*) - V(q^n) - \Delta t \nabla V(q^n) \cdot M^{-1} p^*$. The condition (45) allows to handle the term $-\frac{\xi \Delta t}{2} \nabla V(q^n) \cdot M^{-1} q^n$. When (47) is satisfied, it then follows

$$W(q^{n+1}, p^*) - W(q^n, p^n) \leq A \frac{\xi \Delta t}{2} - B \frac{\xi \Delta t}{2} \left(V(q^n) + \frac{\xi^2}{4} q^n \cdot M^{-1} q^n \right) - \frac{\xi}{4m^2} \Delta t |p^*|^2. \tag{48}$$

Recalling $W_{\Delta t}(q, p) = W(q, p) + \frac{\xi}{4m^2} \Delta t |p|^2$, we obtain

$$\begin{aligned} W_{\Delta t}(q^{n+1}, p^*) - W_{\Delta t}(q^n, p^n) &\leq A \frac{\xi \Delta t}{2} - B \frac{\xi \Delta t}{2} \left(V(q^n) + \frac{\xi^2}{4} q^n \cdot M^{-1} q^n \right) - \frac{\xi}{4m^2} \Delta t |p^n|^2 \\ &\leq A \frac{\xi \Delta t}{2} - B' W_{\Delta t}(q^n, p^n) \end{aligned}$$

for some $B' > 0$. The final step $p^{n+1} = p^* + \sigma \sqrt{\Delta t} R^n$ leads to

$$\mathbb{E}(W_{\Delta t}(q^{n+1}, p^{n+1}) \mid \mathcal{F}_n) = \mathbb{E}(W_{\Delta t}(q^{n+1}, p^*)) + \mathbb{E}|\sigma \sqrt{\Delta t} R^n|^2,$$

so that

$$\mathbb{E}(W_{\Delta t}(q^{n+1}, p^{n+1}) \mid \mathcal{F}_n) \leq b \mathbb{E}(W_{\Delta t}(q^n, p^n)) + c \tag{49}$$

for some $c > 0, 0 < b < 1$. □

Let us now turn to the last approach, and describe the algorithms that are the most used in practice, but for which there are no convergence results at this date (only consistency results are known). Some classical algorithms are the ones described in [2]. The idea is to exactly integrate the dynamics when forces $F(q(t))$ vary linearly with respect to time. In practice, forces are interpolated in time between two successive time steps. Another approach is to generalize the Verlet algorithm. One such algorithm is the BBK algorithm, proposed by Brunger *et al.* [6]. Another example is the quasi-symplectic algorithm of [47]. A third approach, more recent, is

to design algorithms based on an operator splitting method. The Langevin Impulse algorithm, proposed in [67], is such an algorithm. It is not completely understood which integration scheme is the most efficient [48, 68, 79], especially because the comparison benchmarks vary from one field to another.

We focus on the sequel on the BBK algorithm, which is well-suited only for small values of ξ [48, 62] (otherwise, algorithms from [2] or the Langevin impulse scheme [67] should be used). It is a modification of the usual velocity-Verlet scheme obtained by adding a term $-\xi \frac{p_i}{m_i} + \frac{\sigma_i}{\sqrt{\Delta t}} R_i^n$ to the force f_i exerted on particle i (the relation between ξ and σ_i will be made precise below). This may explain its popularity since it only asks for slight modifications of standard MD codes. The random forcing terms R_i^n ($i \in \{1, \dots, N\}$ is the label of the particles, n is the iteration index) are standard i.i.d. Gaussian random variables. The scheme reads:

$$\begin{cases} p_i^{n+1/2} = p_i^n + \frac{\Delta t}{2} \left(-\nabla_{q_i} V(q^n) - \xi \frac{p_i^n}{m_i} + \frac{\sigma_i}{\sqrt{\Delta t}} R_i^n \right), \\ q_i^{n+1} = q_i^n + \Delta t \frac{p_i^{n+1/2}}{m_i}, \\ p_i^{n+1} = \frac{1}{1 + \frac{\xi \Delta t}{2m_i}} \left(p_i^{n+1/2} - \frac{\Delta t}{2} \nabla_{q_i} V(q^{n+1}) + \sigma_i \frac{\sqrt{\Delta t}}{2} R_i^{n+1} \right). \end{cases} \tag{50}$$

We now make precise the relation between ξ and σ_i by considering the case when there are no forces. When $\nabla V = 0$, the BBK algorithm reads

$$\left(1 + \frac{\xi}{2m_i} \Delta t \right) p_i^{n+1} = \left(1 - \frac{\xi}{2m_i} \Delta t \right) p_i^n + \sigma_i \frac{\sqrt{\Delta t}}{2} (R_i^n + R_i^{n+1}). \tag{51}$$

We see that, if $\mathbb{E}(p_i^n) = 0$, then $\mathbb{E}(p_i^{n+1}) = 0$. Choosing p_i^0 such that $\mathbb{E}(p_i^0) = 0$, we have $\mathbb{E}(p_i^n) = 0$ for all n . Let us now denote by $K_i^n = \mathbb{E}((p_i^n)^2)$ the variance of p_i^n . Setting $\gamma_i = \frac{\xi \Delta t}{2m_i}$, one has

$$K_i^{n+1} = \left(\frac{1 - \gamma_i}{1 + \gamma_i} \right)^2 K_i^n + \frac{3\sigma_i^2 \Delta t}{(1 + \gamma_i)^3}.$$

The above recursion is of the general form $x_{n+1} = ax_n + b$, and has a fixed point provided $a < 1$, which is always the case here since $\gamma_i > 0$. This fixed point K_i^∞ is such that

$$\frac{1}{m_i} K_i^\infty = \frac{3\sigma_i^2}{2\xi(1 + \gamma_i)}. \tag{52}$$

Setting σ_i to the value

$$\sigma_i^{\Delta t} = \sqrt{\frac{2\xi(1 + \gamma_i)}{\beta}} = \sqrt{\frac{2\xi}{\beta} \left(1 + \frac{\xi \Delta t}{2m_i} \right)}, \tag{53}$$

we see that $K_i^\infty = \frac{3m_i}{\beta}$, which is indeed the expected value (the kinetic temperature is correct). Note that (53) gives the magnitude of the random forcing that should be used in numerical simulations if one wants the kinetic temperature to be correct. Otherwise, if σ is chosen according to (40), the time-averaged kinetic temperature is lower than the target temperature T (see [70] for some numerical illustrations), and the error is of order Δt , as can be seen from (52). This is consistent with the results obtained in [79] from a modified equation approach. Note that using (53) instead of (40) does not improve the configurational sampling accuracy (the error on the configurational sampling is of order Δt with both choices (40) and (53)).

Another modification of the BBK algorithm has been proposed in [66]. It amounts to using the same Gaussian random variables in the first and the third lines of (50). In this case, there is no bias on the kinetic temperature with the choice (40).

3. DETERMINISTIC MOLECULAR DYNAMICS SAMPLING

We now turn in this section to purely deterministic methods. These methods rely on the following idea: a system in the canonical ensemble can be considered as a system interacting with an external heat bath, the interaction being such that, at equilibrium, the physical system variables are distributed according to the canonical measure (2). Thus, the idea is to consider an extended system composed of the physical variables and some additional variables modelling the bath. Various dynamics have been proposed in this vein.

In this section, we first consider the Nosé-Hoover dynamics and its generalization to the Nosé-Hoover chains [24,40,52,75]. Then, we consider the Nosé-Poincaré method [5] and the Recursive Multiple Thermostats method, which has been recently proposed in [34].

3.1. The Nosé-Hoover and Nosé-Hoover chains methods

The Nosé-Hoover (NH) method, proposed by Hoover, consists in describing the heat bath by two scalar variables, its “position” η and its “momentum” ξ , and to *postulate* the following dynamics for the extended set of variables [24, 52]:

$$\begin{cases} \frac{dq_i}{dt} = \frac{p_i}{m_i} \\ \frac{dp_i}{dt} = -\nabla_{q_i} V - \frac{p_i \xi}{Q} \\ \frac{d\eta}{dt} = \frac{\xi}{Q} \\ \frac{d\xi}{dt} = \sum_{i=1}^N \frac{p_i^2}{m_i} - g k_B T \end{cases} \quad (54)$$

where V is the potential energy of the system, g is a parameter we will fix later and T is the target temperature. The parameter Q represents the mass of the thermostat; it is a free parameter that the user has to choose. The quantity

$$\tilde{H}_{\text{NH}} = \sum_{i=1}^N \frac{p_i^2}{2m_i} + V(q) + \frac{\xi^2}{2Q} + g k_B T \eta \quad (55)$$

is an invariant of the dynamics (54), which also preserves the measure

$$d\mu_{\text{NH}} = \exp(3N\eta) dq dp d\eta d\xi. \quad (56)$$

We refer to [17] for details on the origin of this dynamics. Let us just note here that (54) is not a Hamiltonian dynamics⁵. Since the dynamics preserves (55), it cannot be ergodic with respect to $d\mu_{\text{NH}}$. Let us introduce the manifold $\mathcal{M}_{\text{NH}}(E_0) = \left\{ (q, p, \eta, \xi) \in \mathbb{R}^{6N+2} \mid \tilde{H}_{\text{NH}}(q, p, \eta, \xi) = E_0 \right\}$ and the measure

$$d\rho_{\text{NH}} = \frac{d\sigma_{\text{NH}}}{\|\nabla \tilde{H}_{\text{NH}}\|_2}, \quad (57)$$

where $d\sigma_{\text{NH}}$ is the area measure induced on $\mathcal{M}_{\text{NH}}(E_0)$ by the measure (56), $\nabla \tilde{H}_{\text{NH}}$ is the gradient of (55) with respect to all variables and $\|\cdot\|_2$ is the Euclidian norm. Then $d\rho_{\text{NH}}$ is an invariant measure for the Nosé-Hoover dynamics (54).

⁵The Nosé-Hoover dynamics can be recast, after changing variables and time, as a Hamiltonian dynamics, the so-called Nosé dynamics [51]. However, the time of this dynamics does not correspond anymore to the physical time.

Suppose now that the dynamics is ergodic with respect to $d\rho_{\text{NH}}$ (note that this implies that \tilde{H}_{NH} is the unique invariant of (54)). Let us set $g = 3N$, where N is the number of particles. An easy computation (see [32, 75]) shows that the dynamics $(q(t), p(t))$ is ergodic with respect to the canonical measure (2), and thus provides a sampling of the phase space according to the canonical measure (at least before numerical discretization).

We emphasize the fact that, to the best of the authors knowledge, there is no rigorous proof in the literature showing that (54) is ergodic with respect to $d\rho_{\text{NH}}$. Furthermore, it has been numerically observed that, for some systems, the dynamics $(q(t), p(t))$ does not seem to sample the phase space according to the canonical measure. For instance, this is the case with the one-dimensional harmonic oscillator, for which it is actually observed that the trajectory stays in a ring, namely that there exist $c, C > 0$ such that $c \leq q^2(t) + p^2(t) \leq C$ for all t (see [40, 75]). Some mathematical analysis of this fact can be read in [32].

To circumvent this difficulty, a generalization of the Nosé-Hoover dynamics (54) has been proposed by Martyna *et al.* in [40]. The idea consists in coupling the physical variables with a first thermostat as in (54), and to couple this thermostat with a second one, which can be coupled to a third one, and so on. The variables now include $2M$ additional scalar variables η_j and ξ_j , $j = 1, \dots, M$, where the number M of thermostats is arbitrary. The corresponding dynamics is the so-called Nosé-Hoover chain dynamics (NHC) [40], in which there are M free parameters, Q_1, \dots, Q_M , representing the masses of the M thermostats. The dynamics preserves an invariant \tilde{H}_{NHC} and a measure $d\mu_{\text{NHC}}$ (which are the generalization of (55) and (56)).

As for the Nosé-Hoover dynamics, if the NHC dynamics is ergodic with respect to a measure $d\rho_{\text{NHC}}$ built in the same way as $d\rho_{\text{NH}}$, then the dynamics $(q(t), p(t))$ is ergodic with respect to the canonical measure. Provided that the number M of thermostats is large enough ($M \geq 3$ or 4 in practice), numerical simulations seem to show that this dynamics samples the phase space according to the canonical measure, even for systems such as the harmonic oscillator. Again, there is no rigorous proof showing that the NHC dynamics is *actually* ergodic with respect to $d\rho_{\text{NHC}}$.

Regarding numerical integration, it seems interesting to work with algorithms that preserve the qualitative structure of the dynamics, that is *time reversibility* and *measure preservation*. Reversible-in-time and measure-preserving algorithms have been proposed in [41] (let us just mention here that they are based on a splitting of the dynamics). Simulation results discussed in Section 4 have been obtained with these algorithms.

3.2. The Nosé-Poincaré and the Recursive Multiple Thermostat methods

Both the Nosé-Hoover and the Nosé-Hoover chain dynamics suffer from not being Hamiltonian dynamics. As a consequence, the quasi-conservation by the numerical flow of the invariants \tilde{H}_{NH} (see (55)) and \tilde{H}_{NHC} is not guaranteed. On the contrary, when working with a Hamiltonian dynamics, it is known that the energy can be preserved by the numerical flow over very long times, provided symplectic algorithms are used (see [19], Chap. IX, and [55]). Another problem with Nosé-Hoover chains is the choice of the number of thermostats as well as their masses Q_j , which seem to have an influence on the results.

The Recursive Multiple Thermostat method (RMT) has been recently proposed by Leimkuhler and Sweet [34] to solve the difficulties that have just been highlighted. It is a Hamiltonian dynamics which, like the Nosé-Hoover or Nosé-Hoover chains dynamics, couples the physical variables with a heat bath. This dynamics is a generalization of the Nosé-Poincaré (NP) method [5], which is also a Hamiltonian method. The Nosé-Poincaré method consists in adding a single thermostat, whereas the RMT method consists in adding an arbitrary number M of thermostats, which are all coupled together and to the physical particles. This is not the case in the Nosé-Hoover chain dynamics, where only the first thermostat is coupled to the physical particles (and not the other thermostats).

The Nosé-Poincaré method is based on the following Hamiltonian:

$$H_{\text{NP}}(q, p, \eta, \xi) = \eta \left(H \left(q, \frac{p}{\eta} \right) + \frac{\xi^2}{2Q} + gk_{\text{B}} T \ln \eta - H_0 \right), \tag{58}$$

where H is given by (3), H_0 is chosen such that $H_{\text{NP}} = 0$ for the initial conditions, and where Q is some free parameter. Sampling properties and numerical algorithms are discussed in [5]. Let us just mention here that, as for the Nosé-Hoover dynamics, one has to set $g = 3N$ if the only invariant of the dynamics is H_{NP} .

The motivation for introducing the RMT method is the observation that, at least for some systems, numerical results seem to depend much less on the thermostat masses (which are user-chosen parameters) than with the Nosé-Poincaré method (see [34, 71]).

The numerical results that are presented in Section 4 have been obtained with the algorithms proposed in [5] and [34]. Let us note that different algorithms may have different numerical stabilities, and so different abilities to adequately sample the phase space with a trajectory of a given number of time steps. A new algorithm for the RMT dynamics has been proposed very recently in [4].

4. NUMERICAL ILLUSTRATIONS

The different methods presented above can be used to compute numerical approximations of phase space integrals. In some cases, theoretical convergence rates can be obtained. Typically, when a CLT holds true, the error is bounded by $Cn^{-1/2}$ (where n is the number of evaluations of the potential energy and/or of the forces; see the Central Limit Theorem Thm. 2) for some unknown prefactor C , depending on both the system and the observable A . An important issue is the value of the prefactor in numerical computations, which can greatly vary from one method to another one.

However, since this prefactor depends on A , it is not easy to compare the different methods in a general way. After a brief description of the alkane model in Section 4.1, we present in Section 4.2 an abstract criterion defined without any explicit dependence on an observable A . The criterion measures the deviation between the empirical distributions and the canonical distribution. This comparison can be performed for a fixed sample size [70] (bearing in mind the computation of autocorrelation functions with a fixed computational cost for example), or, more fairly, at a fixed computational cost. Some improvements can also be achieved when combining different sampling techniques, or when resorting to strategies different from the computation of a single long trajectory. This is made precise in Section 4.3. In Section 4.4, we consider a specific case of observable A , which corresponds to a correlation function. The numerical results that are obtained with this physical choice illustrate the conclusions drawn from the abstract criterion in Section 4.2.

4.1. Description of the linear alkane molecule

Linear alkanes are chemical compounds of the form $\text{CH}_3\text{-(CH}_2\text{)}_n\text{-CH}_3$. In this study, the so-called united-atom model [60] is used, in which the conformation of the molecule is completely characterized by the positions of the Carbon atoms. The presence of the Hydrogen atom is implicitly taken into account in the definition of the interaction potential energy the Carbon atoms are subjected to. The Carbon atoms of the linear alkane molecule are indexed from 1 to N , and their positions are described by the vector $q = (q_1, \dots, q_N) \in (\mathbb{R}^3)^N$. We set $r_{i,j} = q_j - q_i$ and we denote by $d_{i,j} = |r_{i,j}|$ the distance between the Carbon atoms i and j .

In the model presented here, the interatomic potential energy involves two-, three-, and four-body interactions:

- two Carbon atoms connected by a covalent bond interact *via* a harmonic potential energy

$$V_2(d) = \frac{1}{2} k_0 (d - d_0)^2; \quad (59)$$

- two Carbon atoms that are separated by three covalent bonds or more interact *via* a Lennard-Jones potential energy

$$V_{\text{LJ}}(d) = 4\epsilon \left(\left(\frac{\sigma}{d} \right)^{12} - \left(\frac{\sigma}{d} \right)^6 \right).$$

The parameters ϵ and σ depend on the atoms that interact, and can have three values: $\epsilon_{\text{CH}_3-\text{CH}_3}$ and $\sigma_{\text{CH}_3-\text{CH}_3}$ when two CH_3 groups interact (the end groups), $\epsilon_{\text{CH}_3-\text{CH}_2}$ and $\sigma_{\text{CH}_3-\text{CH}_2}$ when an interior group interacts with an end group, and $\epsilon_{\text{CH}_2-\text{CH}_2}$ and $\sigma_{\text{CH}_2-\text{CH}_2}$ when two CH_2 groups interact;

- three consecutive Carbon atoms $C_i-C_{i+1}-C_{i+2}$ interact *via* the three-body interaction potential energy

$$V_3(\theta_i) = \frac{1}{2} k_\theta (\theta_i - \theta_0)^2, \quad (60)$$

where

$$\theta_i = \arccos \left(\frac{r_{i,i+1} \cdot r_{i+1,i+2}}{|r_{i,i+1}| \cdot |r_{i+1,i+2}|} \right) \quad (61)$$

is the bending angle of the $C_i-C_{i+1}-C_{i+2}$ chain;

- lastly, four consecutive Carbon atoms $C_i-C_{i+1}-C_{i+2}-C_{i+3}$ experience the four-body interaction potential energy

$$V_4(\phi_i) = u_{\text{tors}}(\cos \phi_i), \quad (62)$$

where ϕ_i is the dihedral angle defined by

$$\cos \phi_i = - \frac{(r_{i,i+1} \times r_{i+1,i+2}) \cdot (r_{i+1,i+2} \times r_{i+2,i+3})}{|(r_{i,i+1} \times r_{i+1,i+2})| \cdot |(r_{i+1,i+2} \times r_{i+2,i+3})|} \quad (63)$$

and where the function u_{tors} is given by

$$u_{\text{tors}}(x) = c_1(1 - x) + 2c_2(1 - x^2) + c_3(1 + 3x - 4x^3).$$

The potential energy of the linear alkane molecule eventually reads

$$V(q) = \sum_{i=1}^{N-1} V_2(d_{i+1,i}) + \sum_{i=1}^{N-2} V_3(\theta_i) + \sum_{i=1}^{N-3} V_4(\phi_i) + \sum_{i=1}^{N-4} \sum_{j=i+3}^N V_{\text{LJ}}(d_{i,j}), \quad (64)$$

where the term V_{LJ} depends on the type of interaction considered.

The values of the parameters d_0 , ϵ , σ , k_θ , θ_0 , c_1 , c_2 and c_3 are taken from [39]. In the system of units where the length unit is $l_0 = 1.53 \cdot 10^{-10}$ m and the energy unit is such that $k_B T = 1$ at $T = 300$ K, the time unit is $\bar{t} = 364$ fs, and the numerical values of the parameters are $d_0 = 1$, $\epsilon_{\text{CH}_3-\text{CH}_3} = 0.294$, $\epsilon_{\text{CH}_3-\text{CH}_2} = 0.241$, $\epsilon_{\text{CH}_2-\text{CH}_2} = 0.198$, $\sigma_{\text{CH}_3-\text{CH}_3} = \sigma_{\text{CH}_3-\text{CH}_2} = \sigma_{\text{CH}_2-\text{CH}_2} = 2.55$, $k_\theta = 208 \text{ rad}^{-2}$, $\theta_0 = 1.187 \text{ rad}$, $c_1 = 1.18$, $c_2 = -0.23$ and $c_3 = 2.64$. Notice that for these values of the parameters c_i , the function u_{tors} has a unique global minimum (at $\phi = 0$) and two local non-global minima. As far as the parameter k_0 is concerned, we set $k_0 = 1000$ (another possibility[39] is to constrain the C-C covalent bond length to be equal to d_0). We set the unit of mass such that the mass of each particle is equal to 1.

We note that $\sum_{i=1}^N \nabla_{q_i} V = 0$, and that $\sum_{i=1}^N q_i \times \nabla_{q_i} V = 0$. As a consequence, the Newton equations (7) not only preserve the energy, but also preserve the linear momentum $\sum_{i=1}^N p_i$ and the angular momentum $\sum_{i=1}^N q_i \times p_i$. Similarly, the Nosé-Hoover dynamics (54) also has additional invariants: besides (55), it preserves $e^\eta \sum_{i=1}^N p_i$ and $e^\eta \sum_{i=1}^N q_i \times p_i$. As a consequence, it cannot be ergodic with respect to (57). One can nevertheless recover correct sampling properties in the q variables by

- starting from an initial condition that satisfies $\sum_{i=1}^N p_i(0) = 0$ and $\sum_{i=1}^N q_i(0) \times p_i(0) = 0$, so that the linear and angular momenta are always equal to 0;
- setting $g = 3N - N_c$, where N_c is the number of conservation laws (besides the energy (55)).

In the case under study here, $N_c = 6$. The same kind of remarks also hold true for the Nosé-Hoover chain dynamics, the Nosé-Poincaré dynamics and the RMT method. The simulation results that we present below have been obtained with these choices. Note that there is no need for any modification for the stochastically perturbed MD methods.



FIGURE 1. Empirical probability distribution of the dihedral angles (ϕ_1, ϕ_2) of the pentane molecule generated with Importance sampling, for $\beta = 1$ (left) and $\beta = 2$ (right), with sample size $N = 10^9$ and $\epsilon_{\text{CH}_3-\text{CH}_3} = 0.29$, $\epsilon_{\text{CH}_3-\text{CH}_2} = 0$.

The linear pentane $\text{CH}_3-(\text{CH}_2)_3-\text{CH}_3$ is the shortest linear alkane for which a two-body Lennard-Jones interaction (coupling the variables $d_{i,i+1}$, θ_i and ϕ_i all together) has to be taken into account. In addition, it involves only two dihedral angles and these two angles essentially determine the conformation of the molecule. Indeed, the covalent stretching and bending potential energies (namely, V_2 and V_3) are stiff and consequently the bond lengths and bending angles are statistically close to their equilibrium values at room temperature. Therefore, the linear pentane molecule is a good test case for it allows a simple reduced representation of the conformation while being a non-trivial model in which the internal degrees of freedom are coupled all together. For completeness, tests on longer molecules are performed in order to investigate the robustness of the numerical methods with respect to increasing configurational space dimensions.

Some reference empirical densities for the dihedral angles obtained through Importance sampling techniques are presented in Figure 1. They correspond to pentane, with $N = 10^9$ sample points.

4.2. Discrepancy of sample points

In order to quantitatively assess the quality of the samples generated by the various methods described above, we use a discrepancy criterion. Recall that the discrepancy D_n of a sequence $x = \{x_m\}_{0 \leq m \leq n-1}$ with values in $[0, 1]^d$ is defined as (see [30])

$$D_n(x) = \sup_{y \in [0, 1]^d} \left| \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{1}_{\{x_m \in [0, y]\}} - \text{Volume}([0, y]) \right|, \quad (65)$$

where, for d -dimensional vectors y, z , we write $y \leq z$ when $y_i \leq z_i$ for all $1 \leq i \leq d$, and note $[0, y] = \{z \in [0, 1]^d, z \leq y\}$. The fact that $D_n(x) \rightarrow 0$ when $n \rightarrow \infty$ is equivalent (see [30], p.15) to the fact that, for any Riemann integrable function A defined on $[0, 1]^d$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=0}^{n-1} A(x_m) = \int_{[0, 1]^d} A(x) dx.$$

In addition, for functions A which have bounded variations $V_{\text{HK}}(A)$ in the sense of Hardy and Krause [50], the following error estimate holds true:

$$\left| \frac{1}{n} \sum_{m=0}^{n-1} A(x_m) - \int_{[0, 1]^d} A(x) dx \right| \leq V_{\text{HK}}(A) D_n(x). \quad (66)$$

If $A \in C^d([0, 1]^d)$, then its variation $V_{\text{HK}}(A)$ has a simple expression (see [50], p. 19). If $d = 2$, which is the case we will be interested in below, then

$$V_{\text{HK}}(A) = \int_{[0,1]^2} \left| \frac{\partial^2 A}{\partial x_1 \partial x_2} \right| dx + \int_0^1 \left| \frac{\partial A}{\partial x_1}(x_1, 1) \right| dx_1 + \int_0^1 \left| \frac{\partial A}{\partial x_2}(1, x_2) \right| dx_2.$$

As a consequence of (66), the convergence of $D_n(x)$ toward 0 implies the Law of Large Numbers, and the rate of convergence of $D_n(x)$ gives information about the convergence rate of the observable average.

In this framework, we intend for example to characterize the repartition of sample points in the subset $[-\pi, \pi]^2$ of the (ϕ_i, ϕ_j) -plane for two of the dihedral angles ϕ_i, ϕ_j . This can be achieved by considering the marginal ν_{ij} of the canonical density π with respect to the other degrees of freedom. Unfortunately, there is no simple exact expression of this marginal. We therefore consider the situation when all $\epsilon = 0$ (that is when the Lennard-Jones interactions are all turned off), in which case the marginal has the simple expression

$$d\nu_{ij}(\phi_i, \phi_j) = Z_\phi^{-2} e^{-\beta V_4(\phi_i)} e^{-\beta V_4(\phi_j)} d\phi_i d\phi_j, \tag{67}$$

with V_4 given by (62).

We then introduce the discrepancy criterion

$$D_n(\{q^m\}) = \sup_{(\phi_i, \phi_j) \in [-\pi, \pi]^2} \left| \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{1}_{\{\phi_i^m \leq \phi_i, \phi_j^m \leq \phi_j\}} - \int_{\{\psi_i \leq \phi_i, \psi_j \leq \phi_j\}} d\nu_{ij}(\psi_i, \psi_j) \right|, \tag{68}$$

which provides a bound on the L^∞ distance between the empirical distribution functions and the exact ones. Notice that the second integral factorizes as

$$\int_{\{\psi_i \leq \phi_i, \psi_j \leq \phi_j\}} d\nu_{ij}(\psi_i, \psi_j) = Z_\phi^{-2} \int_{\psi_i \leq \phi_i} e^{-\beta V_4(\psi_i)} d\psi_i \int_{\psi_j \leq \phi_j} e^{-\beta V_4(\psi_j)} d\psi_j,$$

and can therefore easily be computed using standard numerical techniques.

Numerically, we compute an approximate value of D_n as follows. Suppose that we have partitioned the (ϕ_i, ϕ_j) -plane into K^2 boxes $B_{kl} = [\Phi_k, \Phi_{k+1}[\times [\Phi_l, \Phi_{l+1}[$ with $\Phi_k = -\pi + \frac{2k\pi}{K}$ for $0 \leq k \leq K - 1$. The supremum in (68) is now taken over a finite set of elements:

$$D_n^K(q) = \sup_{1 \leq k, l \leq K} \left| \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{1}_{\{\phi_i^m \leq \Phi_k, \phi_j^m \leq \Phi_l\}} - \int_{\{\psi_i \leq \Phi_k, \psi_j \leq \Phi_l\}} d\nu_{ij}(\psi_i, \psi_j) \right|. \tag{69}$$

We then compute the discrepancies for the sample points obtained by different methods with a fixed computational cost. The computational cost measures here the number of force or energy evaluations.

Implementation details, optimal choice of parameters and algorithmic costs are discussed more precisely in [70]. For each method⁶, parameters have been optimized in a certain range of values, in order to get the lowest discrepancy at a fixed computational cost. For HMC, we have chosen Δt in order to have an acceptance rate close to 90% (this yields $\Delta t = 0.025$), and we have optimized the integration time τ . For the BRW, we have considered time steps Δt ranging from 0.01 to 0.05 (the corresponding rejection rates with the MALA algorithm range from 2 to 98%). For the Langevin equation, we have made use of the BBK scheme with $\Delta t = 0.02$ and we have considered friction coefficients in the range $\xi = 0.1$ to $\xi = 100$. We have noticed that, for the HMC, BRW and Langevin equation methods, there exists a broad range of parameter values around the values that we have chosen that give similar results. Hence the precise values of the parameters do not

⁶The Nosé-Hoover method (resp. the Nosé-Poincaré method) has been considered as a special case of the Nosé-Hoover chain method (resp. of the RMT method).

seem to have a significant influence on the final results. For the deterministic methods, the time step Δt has been chosen so that invariants are numerically preserved with a relative accuracy of a few percents⁷. For the Nosé-Hoover method, we have looked for an optimized Q_1 in the range $0.1 \leq Q_1 \leq 10$. For the Nosé-Hoover chain method, we have restricted ourselves to a chain of $M = 2$ thermostats. We have worked with Q_2 from 0.01 to 10, with the choice $Q_1 = Q_2$ or $Q_1 = 15Q_2$ (see [40], App. B, for some explanation for this choice). The optimal thermostat mass for the Nosé-Poincaré has been looked for in the range $0.1 \leq Q_1 \leq 10$. For the RMT method, we have also restricted ourselves to the case of $M = 2$ thermostats. We have considered all the choices (Q_1, Q_2) considered for the Nosé-Hoover chain method, and we have also considered the choice $Q_2 = 2Q_1$, with $0.05 \leq Q_1 \leq 10$, following the strategy explained in [4]. Concerning the purely stochastic methods, we have worked with $g(q) = \tilde{Z}_q^{-1} \exp(-\beta\tilde{V}(q))$, where

$$\tilde{V}(q) = \sum_{i=1}^{N-1} V_2(d_{i+1,i}) + \sum_{i=1}^{N-2} V_3(\theta_i)$$

and \tilde{Z}_q is a normalization constant. The functions V_2 and V_3 are quadratic (see (59) and (60)), which makes it possible to actually sample from $g(q) dq$ (see [70] for more details). We also note that $\tilde{V}(q) \leq V(q)$ (recall that the Lennard-Jones interactions are all turned off), so that g satisfies the constraint (11).

The results are presented in Tables 2 to 4. For each method, 10 different simulations have been performed, and we give in the Tables the mean and the square-root of the variance (in brackets) of the 10 different results.

One can see that purely stochastic methods are very efficient for small alkane chains, but rapidly lose their efficiency when the length of the chain increases. Thus, the Langevin dynamics and the HMC method seem to be the most efficient methods, although other non purely stochastic methods also give good results. The Langevin, the HMC and the BRW (with Euler-Maruyama algorithm) methods keep the same efficiency whatever the length of the chain. This seems also to be the case for the NHC method. The efficiency of the BRW (with the MALA algorithm) decreases when the chain length increases. There seems to be a problem with the RMT method applied to the pentane molecule. A careful analysis of the results show that the numerical dihedral angle distribution corresponds to (67) but with a temperature significantly different from the target temperature. If longer chains are considered, this problem disappears and the RMT method results are of the same order of magnitude as the results from other methods (see Tabs. 3 and 4).

We can also see that, for short chains, the biased Random-Walk (MALA) is more efficient than the NHC method. However, for chains of 9 and 12 particles, the NHC method is more efficient. The biased Random-Walk with the Euler-Maruyama algorithm always seems to be a little more efficient than the biased Random-Walk with the MALA algorithm.

4.3. Improvement of the convergence rates

4.3.1. Convergence rate improvements using several shorter realizations

We already mentioned that, instead of running a single long trajectory, it might be more efficient, for a given computational cost, to run several shorter trajectories. This can be done for methods of type 2 to 4. For methods of types 2 and 3, this strategy relies on the following numerical approximation. Assuming that the methods are ergodic, it follows

$$\mathbb{E}_x(A(q^{N_1})) \rightarrow \int_{\mathcal{M}} A(q) d\pi \quad (70)$$

when $N_1 \rightarrow +\infty$. In some cases, this convergence is exponentially fast. The term $\mathbb{E}_x(A(q^{N_1}))$ is the expectation of the realizations of the chain conditioned at starting from $x \in \mathcal{M}$. It can be approximated by N_2 independent realizations of the Markov chain. Each realization is labelled by an index $k \in \{1, \dots, N_2\}$, and the associated

⁷Note that, for the NP method, the Hamiltonian (58) remains equal to 0 along the exact trajectory. We have considered two relative errors, first H_{NP}/H_0 , and second $H_{\text{NP}}^{\text{int}}/H_0$, where $H_{\text{NP}}^{\text{int}}$ is such that $H_{\text{NP}} = \eta(H_{\text{NP}}^{\text{int}} - H_0)$.

TABLE 2. Numerical results for the discrepancy (69) for the pentane (ϕ_1, ϕ_2) distribution in the case $\beta = 1$ and $K = 100$.

Method	Parameters	Discrepancy for 10^6 evaluations	Discrepancy for 10^7 evaluations
Importance sampling	-	0.00428 (0.00114)	0.00115 (1.60×10^{-4})
Rejection	-	0.00856 (0.00204)	0.00256 (4.98×10^{-4})
MIS	-	0.0228 (0.00416)	0.0225 (7.75×10^{-4})
HMC	$\tau = 10\Delta t, \Delta t = 0.025$	0.0389 (0.0183)	0.0119 (4.87×10^{-4})
BRW (Euler-Maruyama)	$\Delta t = 0.028$	0.0791 (0.0265)	0.0231 (0.00619)
BRW (MALA)	$\Delta t = 0.028$	0.104 (0.0446)	0.0343 (0.0139)
Langevin	$\Delta t = 0.02, \xi = 1$	0.0339 (0.0142)	0.0157 (0.00393)
NHC	$Q_1 = Q_2 = 0.05, \Delta t = 0.0025$	0.103 (0.036)	0.0456 (0.0117)
RMT	$Q_1 = 5, Q_2 = 7.5, \Delta t = 0.0025$	0.196 (0.142)	0.178 (0.177)

TABLE 3. Numerical results for the discrepancy (69) for the (ϕ_1, ϕ_3) distribution for C_9H_{20} in the case $\beta = 1$ and $K = 100$. The computational cost is fixed to 10^7 force or energy evaluations.

Method	Parameters	Discrepancy
Importance sampling	-	0.0205 (0.00544)
Rejection	-	0.192 (0.0379)
MIS	-	0.521 (0.0151)
HMC	$\tau = 10\Delta t, \Delta t = 0.02$	0.0261 (0.00846)
BRW (Euler-Maruyama)	$\Delta t = 0.025$	0.0402 (0.0229)
BRW (MALA)	$\Delta t = 0.025$	0.0477 (0.0129)
Langevin	$\Delta t = 0.025, \xi = 1$	0.0144 (0.00544)
NHC	$Q_1 = 0.15, Q_2 = 0.01, \Delta t = 0.0025$	0.0292 (0.0102)
NP	$Q = 5, \Delta t = 0.0025$	0.0386 (0.0095)

TABLE 4. Numerical results for the discrepancy (69) for the (ϕ_1, ϕ_3) distribution for $C_{12}H_{26}$ in the case $\beta = 1$ and $K = 100$. The computational cost is fixed to 10^7 force or energy evaluations.

Method	Parameters	Discrepancy
Importance sampling	-	0.102 (0.0436)
Rejection	-	1.0 (0.0)
MIS	-	0.493 (0.222)
HMC	$\tau = 10\Delta t, \Delta t = 0.02$	0.0207 (0.00730)
BRW (Euler-Maruyama)	$\Delta t = 0.023$	0.0312 (0.0102)
BRW (MALA)	$\Delta t = 0.023$	0.0610 (0.0201)
Langevin	$\Delta t = 0.025, \xi = 1$	0.0173 (0.00726)
NHC	$Q_1 = 0.15, Q_2 = 0.01, \Delta t = 0.0025$	0.0350 (0.00865)
RMT	$Q_1 = 5, Q_2 = 7.5, \Delta t = 0.0025$	0.0428 (0.0194)

sample path is $(q^{0,k}, \dots, q^{N-1,k})$. Notice that, for all samples, $q^{0,k} = x$. An approximation of $\mathbb{E}_x(A(q^{N_1}))$ is then obtained as

$$\mathbb{E}_x(A(q^{N_1})) \simeq I_{N_2}^{N_1}(x) = \frac{1}{N_2} \sum_{k=1}^{N_2} A(q^{N_1,k}). \tag{71}$$

TABLE 5. Numerical results for the discrepancy (69) for the pentane (ϕ_1, ϕ_2) distribution in the case $\beta = 1$ and $K = 100$, using a Langevin method with $\xi = 1$ and $\Delta t = 0.02$. The discrepancy has been computed with all points appearing in (72) (that is all points of the N_2 trajectories of length N_1), with a computational cost fixed to 10^7 force or energy evaluations.

Number N_2 of realizations	Discrepancy
1	0.0157 (0.00393)
5	0.0117 (0.00388)
10	0.0132 (0.00210)
20	0.0149 (0.00701)
50	0.0120 (0.00330)
100	0.0112 (0.00263)
200	0.0130 (0.00419)
500	0.0308 (0.00834)
1000	0.0528 (0.00740)

Notice that we expect the error between $I_{N_2}^{N_1}(x)$ and the space average $\int_{\mathcal{M}} A(q) d\pi$ to be of the form $C(x)\rho^{N_1} + C(x, N_1)N_2^{-1/2}$ for some $0 < \rho < 1$.

When a short trajectory of length N_1 is computed for N_2 realizations starting from a given initial point x , we can also consider the following approximation of the position space average

$$\int_{\mathcal{M}} A(q) d\pi \simeq \frac{1}{N_1} \sum_{m=0}^{N_1-1} I_{N_2}^m(x), \tag{72}$$

where the right hand side is the Cesaro average of (71).

The results are presented in Table 5 in the case of a Langevin sampling for the pentane molecule at $\beta = 1$. As can be seen, there is a slight improvement when generating several shorter trajectories, provided these trajectories remain long enough. Note however that such an improvement is not always observed. But we emphasize that there is no degradation of the results either. This is an interesting point since it allows a straightforward parallelization of the method.

4.3.2. *Convergence rate improvements at fixed computational cost, using an appropriate initial distribution*

Another improvement is as follows. Instead of considering a fixed initial point, we can make a first approximation of the canonical distribution. Let us denote by π^{N_3} the following approximation of π :

$$\pi^{N_3} = \frac{1}{N_3} \sum_{i=1}^{N_3} \delta_{x^i}.$$

For each initial point x^i ($1 \leq i \leq N_3$), an approximation (72) can be computed, for N_2 realizations of the Markov chain with trajectories of length N_1 . The total number of points generated in this way is therefore $N_1N_2N_3$. The important issue is then to optimize the choices of N_1 , N_2 and N_3 in order to have the best accuracy for a given total cost.

For the method to be efficient, the empirical measure π^{N_3} has to be a good approximation of π . To this end, the points x^i are chosen as follows. We first generate N^{tot} points $(y^1, \dots, y^{N^{\text{tot}}})$ with weights $(w_1, \dots, w_{N^{\text{tot}}})$, using (say) an Importance sampling method. We then generate N_3 points from this list with replacement with probabilities $(\frac{w_1}{W}, \dots, \frac{w_{N^{\text{tot}}}}{W})$ where $W = \sum_{i=1}^{N^{\text{tot}}} w_i$, and run one or several trajectories for each starting point. This can improve the rate of convergence of some methods. An example is the biased Random Walk at $\beta = 1$

TABLE 6. Numerical results for the discrepancy (69) for the pentane (ϕ_1, ϕ_2) distribution in the case $\beta = 1$ and $K = 100$, using a Langevin method with $\xi = 1$ and $\Delta t = 0.02$. The computational cost is fixed to 10^6 force or energy evaluations.

Undersampling rate	Discrepancy
1	0.0339 (0.0142)
5	0.0369 (0.0121)
10	0.0350 (0.00996)
50	0.0391 (0.0194)
100	0.0385 (0.0169)
500	0.0343 (0.0102)
1000	0.0539 (0.0173)

with $\Delta t = 0.028$ for 10^6 operations. We consider $N^{\text{tot}} = 10^4$, $N_3 = 99$, $N_1 = 10^4$ and $N_2 = 1$. The discrepancy is lowered from 0.104 (0.0446) (with $N_1 = 10^6$, $N_2 = 1$ and $N_3 = 1$, see Tab. 2) to 0.0430 (0.0144). In general, it is observed that convergence occurs faster when starting from an approximate distribution.

4.3.3. Effect of undersampling

As a final improvement, we can test the influence of a systematic undersampling, which consists in picking only some of the points generated instead of considering all of them. Indeed, some techniques generate points (q^0, \dots, q^{N-1}) that may be very much correlated, and it can happen that the sequence $(q^0, q^r, \dots, q^{sr})$, the undersampling rate r being such that $N - 1 = rs$, is better distributed than the original sequence.

The results are presented in Table 6 in the case of a Langevin sampling for pentane at $\beta = 1$. As can be seen, the efficiency of the method remains stable when undersampling the data. This is particularly interesting when computing autocorrelation functions or time-dependent integrals of the form (8) since a NVE trajectory has to be computed for each starting point generated from the canonical distribution.

Of course, it is still possible to try to improve the quality of a single realization by filtering out the corresponding sequence of configurations, as is done for NVE simulations in [7,8], but we will not detail this strategy any further.

4.4. Computation of correlation functions

We present, as a final application, the computation of some correlation function, namely the transition rate from the set $\mathcal{A} = \{q \in \mathcal{M} ; |\phi_1| \geq 1, |\phi_2| \geq 1\}$ (both dihedral angles are not in their ground states) to the set $\mathcal{B} = \{q \in \mathcal{M} ; |\phi_1| \leq 1, |\phi_2| \leq 1\}$ (both dihedral angles are in their ground states). This transition rate is expressed as

$$C(t) = \frac{\langle \mathbf{1}_{\mathcal{A}}(q^0) \mathbf{1}_{\mathcal{B}}(\Pi_1 \Phi_t(q, p)) \rangle}{\langle \mathbf{1}_{\mathcal{A}}(q^0) \rangle}. \quad (73)$$

We proceed as follows. We first sample $M = 10^4$ initial conditions according to the canonical measure $d\mu$ (at $\beta = 1$) using 10^6 force evaluations and the parameters given in Table 2 (*i.e.* in all cases except for the HMC algorithm, we undersample at rate 100 a single trajectory that always starts from the same equilibrium position; the HMC trajectory is undersampled at rate 10 only since $\tau = 10\Delta t$). We then integrate the Newton equations of motion from each initial condition using the velocity Verlet scheme (18), for a time $t = 100$ (with $\Delta t = 0.005$). This procedure is repeated 100 times. The results are presented in Figure 2, and are compared to a reference result obtained starting from 10^6 initial conditions sampled with a rejection method.

As can be seen from the results, the methods yielding large discrepancies (such as Nosé-Hoover and BRW) predict a correlation $C(t)$ quite different from the reference result. On the other hand, the HMC and Langevin methods give much better results, especially HMC.

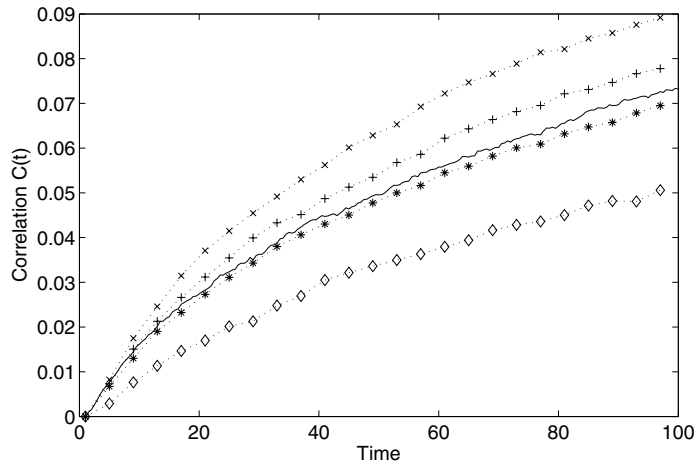


FIGURE 2. Plot of the correlation function $C(t)$ starting from initial conditions generated with the rejection method (solid line), BRW/EM (x), Langevin/BBK (+), HMC (*) and Nosé-Hoover chain (\diamond).

5. CONCLUSION AND PROSPECTS

In this article, we have reviewed usual sampling methods for Molecular Dynamics, and both considered theoretical convergence results and numerical efficiency. From a theoretical viewpoint, purely stochastic methods converge under minimal assumptions, whereas stochastically perturbed Molecular Dynamics methods require some assumptions on the potential energy V . On the other hand, nothing is known about the convergence of deterministic thermostating methods. In practice, the methods involve different parameters, and some are really easier to tune. For example, finding efficient importance sampling functions for purely stochastic methods is very difficult in general. The choice of the right parameters for Nosé-like dynamics also requires some careful exploration of the dynamics, and must be changed whenever the size of the system changes (though there are some heuristic rules to obtain an order of magnitude of the new parameters). On the other hand, the parameters for the methods relying on stochastically perturbed molecular dynamics are found to be valid for a wide range of systems (for the Langevin dynamics in reduced units for example, $\xi/m = 1$ is usually a good choice).

From a numerical viewpoint, it is observed in the test case of linear alkanes that purely stochastic methods have increasing difficulties to converge when the length of the chain increases, whereas stochastically perturbed Molecular Dynamics remain efficient whatever the length of the chain. Deterministic thermostating methods seem to be more efficient when the chain is longer. On the systems we have studied, the Langevin and the Hybrid Monte-Carlo methods seem to be the best methods (and have similar accuracies). Note that a discrepancy twice smaller means that a trajectory 4 times shorter can be used, which provides a significant computation time gain. This conclusion is confirmed by a model correlation function computation.

There remain, however, many interesting problems. First, new sampling methods, especially deterministic sampling methods, are still investigated, from both practical and theoretical viewpoints. Second, and more importantly, the algorithms presented in this article lose their efficiencies when the system presents several metastable states, some of them with life times of the order of magnitude of (or even larger than) the simulation time interval. Promising tracks have been explored to address these situations (see *e.g.* [9, 11, 15, 29, 65, 69, 77,

78, 80]) but robust, general purpose methods able to sample complex potential energy surfaces (such as those of large biological systems) are still lacking.

Acknowledgements. We thank Gilles Stoltz for useful references on statistical tests, Tony Lelièvre and Benjamin Jourdain for background references on SDE, Wilhelm Huisinga and Gilles Zérah for stimulating discussions during CEMRACS'04, Andrew Stuart for discussions and references and Claude Le Bris for his valuable comments and suggestions. This work was supported by the Action Concertée Incitative Nouvelles Interfaces des Mathématiques, "Simulation moléculaire", Ministère de la Recherche, France.

REFERENCES

- [1] E. Akhmatskaya and S. Reich, The targetted shadowing hybrid Monte Carlo (TSHMC) method, in *New Algorithms for Macromolecular Simulation, Lecture Notes in Computational Science and Engineering* **49**, B. Leimkuhler, C. Chipot, R. Elber, A. Laaksonen, A. Mark, T. Schlick, C. Schuette and R. Skeel Eds., Springer Verlag, Berlin and New York (2006) 145–158.
- [2] M.P. Allen and D.J. Tildesley, *Computer simulation of liquids*. Oxford Science Publications (1987).
- [3] H.C. Andersen, Molecular dynamics simulations at constant pressure and/or temperature *J. Chem. Phys.* **72** (1980) 2384–2393.
- [4] E. Barth, B.J. Leimkuhler, and C.R. Sweet, Approach to thermal equilibrium in biomolecular simulation. *Proceedings of AM3-2004 conference*, available at the URL <http://adrg.maths.ed.ac.uk/ADRG/FILES/Archive/BaLeSw2005.pdf>
- [5] S.D. Bond, B.J. Leimkuhler, and B.B. Laird, The Nosé-Poincaré method for constant temperature molecular dynamics. *J. Comput. Phys.* **151** (1999) 114–134.
- [6] A. Brünger, C.B. Brooks, and M. Karplus, Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem. Phys. Lett.* **105** (1983) 495–500.
- [7] E. Cancès, F. Castella, P. Chartier, E. Faou, C. Le Bris, F. Legoll and G. Turinici, High-order averaging schemes with error bounds for thermodynamical properties calculations by molecular dynamics simulations. *J. Chem. Phys.* **121** (2004) 10346–10355.
- [8] E. Cancès, F. Castella, P. Chartier, E. Faou, C. Le Bris, F. Legoll and G. Turinici, Long-time averaging for integrable Hamiltonian dynamics. *Numer. Math.* **100** (2005) 211–232.
- [9] E.A. Carter, G. Ciccotti, J.T. Hynes and R. Kapral, Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.* **156** (1989) 472–477.
- [10] Y. Chen, Another look at Rejection sampling through Importance sampling. *Discussion papers* **04-30**, Institute of Statistics and Decision Science, Duke University (2004).
- [11] G. Ciccotti, R. Kapral and E. Vanden-Eijnden, Blue Moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics. *Chem. Phys. Chem.* **6** (2005) 1809–1814.
- [12] G. Ciccotti, T. Lelièvre and E. Vanden-Eijnden, Projection of diffusions on submanifolds: Application to mean force computation. *CERMICS preprint* **309** (2006).
- [13] S. Duane, A.D. Kennedy, B. Pendleton and D. Roweth, Hybrid Monte Carlo. *Phys. Letters B.* **195** (1987) 216–222.
- [14] M. Duflo, *Random iterative models*. Springer, Berlin, New York (1997).
- [15] W. E, W. Ren and E. Vanden-Eijnden, Finite temperature string method for the study of rare events. *J. Phys. Chem. B* **109** (2005) 6688–6693.
- [16] L.C. Evans and R.F. Gariepy, *Measure Theory and Fine Properties of Functions, Studies in advanced mathematics*. CRC Press, Chapman and Hall (1991).
- [17] D. Frenkel and B. Smit, *Understanding Molecular Simulation, From Algorithms to Applications*, 2nd edn. Academic Press (2002).
- [18] G. Grimmett and D. Stirzaker, *Probability and Random Processes*. Oxford University Press (2001).
- [19] E. Hairer, C. Lubich and G. Wanner, *Geometric Numerical Integration, Structure-Preserving Algorithms For Ordinary Differential Equations, Springer Series in Computational Mathematics* **31**, 2nd edn. Springer-Verlag, Berlin (2006).
- [20] S. Hampton, P. Brenner, A. Wenger, S. Chatterjee and J.A. Izaguirre, Biomolecular Sampling: Algorithms, Test Molecules, and Metrics, in *New Algorithms for Macromolecular Simulation, Lecture Notes in Computational Science and Engineering* **49**, B. Leimkuhler, C. Chipot, R. Elber, A. Laaksonen, A. Mark, T. Schlick, C. Schuette and R. Skeel Eds., Springer Verlag, Berlin and New York (2006) 103–123.
- [21] R.Z. Has'minskii, *Stochastic Stability of Differential Equations*. Sijthoff and Noordhoff (1980).
- [22] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** (1970) 97–109.
- [23] F. Hérau and F. Nier, Isotropic hypoellipticity and trend to equilibrium for the Fokker-Planck equation with a high-degree potential. *Arch. Rational Mech. Anal.* **171** (2004) 151–218.
- [24] W.G. Hoover, Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31** (1985) 1695–1697.

- [25] F.C. Hoppensteadt, M. Rahman and B.D. Welfert, \sqrt{n} -Central limit theorems for Markov processes with applications to circular processes, preprint version (2003). Available at the URL <http://math.asu.edu/~bdw/PAPERS/CLT.pdf>
- [26] A.M. Horowitz, A generalized guided Monte Carlo algorithms. *Phys. Lett. B* **268** (1991) 247–252.
- [27] J.A. Izaguirre and S.S. Hampton, Shadow Hybrid Monte Carlo: an efficient propagator in phase space of macromolecules. *J. Comput. Phys.* **200** (2004) 581–604.
- [28] A.D. Kennedy and B. Pendleton, Cost of the generalised hybrid Monte Carlo algorithm for free field theory. *Nucl. Phys. B* **607** (2001) 456–510.
- [29] A. Laio and M. Parrinello, Escaping free energy minima. *Proc. Natl. Acad. Sci. USA* **99** (2002) 12562–12566.
- [30] B. Lapeyre, E. Pardoux and R. Sentis, *Méthodes de Monte Carlo pour les équations de transport et de diffusion, Mathématiques et applications* **29**, Springer (1998); B. Lapeyre, E. Pardoux and R. Sentis, translated by A. Craig and F. Craig, *Introduction to Monte-Carlo methods for transport and diffusion equations*. Oxford University Press (2003).
- [31] F. Legoll, *Molecular and Multiscale Methods for the Numerical Simulation of Materials*. Ph.D. thesis, University of Paris VI, France (2004).
- [32] F. Legoll, M. Luskin and R. Moeckel, Non-ergodicity of the Nosé-Hoover thermostatted harmonic oscillator. *Arch. Rat. Mech. Anal.* **184** (2007) 449–463.
- [33] B.J. Leimkuhler and S. Reich, *Simulating Hamiltonian dynamics, Cambridge monographs on applied and computational mathematics* **14**. Cambridge University Press (2005).
- [34] B.J. Leimkuhler and C.R. Sweet, A Hamiltonian formulation for recursive multiple thermostats in a common timescale. *SIAM J. Appl. Dyn. Syst.* **4** (2005) 187–216.
- [35] J.S. Liu, *Monte Carlo strategies in Scientific Computing*. Springer Series in Statistics (2001).
- [36] P.B. Mackenze, An improved hybrid Monte Carlo. *Phys. Lett. B.* **226** (1989) 369–371.
- [37] X. Mao, *Stochastic differential equations and applications*. Horwood, Chichester (1997).
- [38] J.E. Marsden and M. West, Discrete mechanics and variational integrators. *Acta Numer.* **10** (2001) 357–514.
- [39] M.G. Martin and J.I. Siepmann, Transferable potentials for phase equilibria. I. United-atom description of *n*-alkanes. *J. Phys. Chem.* **102** (1998) 2569–2577.
- [40] G.J. Martyna, M.L. Klein and M.E. Tuckerman, Nosé-Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **97** (1992) 2635–2643.
- [41] G.J. Martyna, M.E. Tuckerman, D.J. Tobias and M.L. Klein, Explicit reversible integrators for extended systems dynamics. *Mol. Phys.* **87** (1996) 1117–1157.
- [42] J.C. Mattingly, A.M. Stuart and D.J. Higham, Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise. *Stoch. Proc. Appl.* **101** (2002) 185–232.
- [43] K.L. Mengersen and R.L. Tweedie, Rates of convergence in the Hastings-Metropolis algorithm. *Ann. Statist.* **24** (1996) 101–121.
- [44] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21** (1953) 1087–1091.
- [45] S.P. Meyn and R.L. Tweedie, Stability of Markovian processes. II. Continuous-time processes and sampled chains. *Adv. Appl. Probab.* **24** (1993) 487–517.
- [46] S.P. Meyn and R.L. Tweedie, *Markov Chains and Stochastic Stability*. Springer (1993).
- [47] G.N. Milstein and M.V. Tretyakov, Quasi-symplectic methods for Langevin-type equations. *IMA J. Numer. Anal.* **23** (2003) 593–626.
- [48] B. Mishra and T. Schlick, The notion of error in Langevin dynamics: I. Linear analysis. *J. Chem. Phys.* **105** (1996) 299–318.
- [49] R.M. Neal, An improved acceptance procedure for the hybrid Monte-Carlo algorithm. *J. Comput. Phys.* **111** (1994) 194–203.
- [50] N. Niederreiter, *Random Number Generation and Quasi Monte-Carlo Methods*. Society for Industrial and Applied Mathematics (1992).
- [51] S. Nosé, A Molecular Dynamics method for simulations in the canonical ensemble, *Mol. Phys.* **52** (1984) 255–268.
- [52] S. Nosé, A unified formulation of the constant temperature Molecular Dynamics method, *J. Chem. Phys.* **81** (1985) 511–519.
- [53] G. Pagès, Sur quelques algorithmes récursifs pour les probabilités numériques. *ESAIM: PS* **5** (2001) 141–170.
- [54] D.C. Rapaport, *The Art of Molecular Dynamics Simulations*. Cambridge University Press (1995).
- [55] S. Reich, Backward error analysis for numerical integrators. *SIAM J. Numer. Anal.* **36** (1999) 1549–1570.
- [56] G.O. Roberts and J.S. Rosenthal, Optimal scaling of discrete approximations to Langevin diffusions. *J. Roy. Stat. Soc. B* **60** (1998) 255–268.
- [57] G.O. Roberts and R.L. Tweedie, Exponential convergence of Langevin diffusions and their discrete approximations. *Bernoulli* **2** (1996) 341–364.
- [58] G.O. Roberts and R.L. Tweedie, Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** (1996) 95–110.
- [59] L.C.G. Rogers, Smooth transition densities for one-dimensional probabilities. *Bull. London Math. Soc* **17** (1985) 157–161.
- [60] J.P. Ryckaert and A. Bellemans, Molecular dynamics of liquid alkanes. *Faraday Discuss.* **66** (1978) 95–106.
- [61] A. Scemama, T. Lelièvre, G. Stoltz, E. Cancès and M. Caffarel, An efficient sampling algorithm for Variational Monte Carlo. *J. Chem. Phys.* **125** (2006) 114105.
- [62] T. Schlick, *Molecular Modeling and Simulation*. Springer (2002).

- [63] C. Schütte, *Conformational dynamics: Modelling, Theory, Algorithm, and Application to Biomolecules*. Habilitation Thesis, Free University Berlin (1999).
- [64] C. Schütte and W. Huisinga, Biomolecular conformations can be identified as metastable sets of molecular dynamics, in *Handbook of Numerical Analysis (Special volume on computational chemistry)*, Vol. X, P.G. Ciarlet and C. Le Bris Eds., Elsevier (2003) 699–744.
- [65] C. Schütte, A. Fischer, W. Huisinga and P. Deuffhard, A direct approach to conformational dynamics based on Hybrid Monte-Carlo. *J. Comp. Phys.* **151** (1999) 146–168.
- [66] T. Shardlow, Splitting for dissipative particle dynamics. *SIAM J. Sci. Comput.* **24** (2003) 1267–1282.
- [67] R.D. Skeel, in *The graduate student's guide to numerical analysis, Springer Series in Computational Mathematics*, M. Ainsworth, J. Levesley and M. Marletta Eds., Springer-Verlag, Berlin (1999) 119–176.
- [68] R.D. Skeel and J.A. Izaguirre, An impulse integrator for Langevin dynamics. *Mol. Phys.* **100** (2002) 3885–3891.
- [69] M.R. Sorensen and A.F. Voter, Temperature accelerated dynamics for simulation of infrequent events. *J. Chem. Phys.* **112** (2000) 9599–9606.
- [70] G. Stoltz, *Quelques méthodes mathématiques pour la simulation moléculaire et multiéchelle*. Ph.D. Thesis (in preparation).
- [71] C.R. Sweet, *Hamiltonian Thermostatting Techniques for Molecular Dynamics Simulation*. Ph.D. Thesis, University of Leicester (2004).
- [72] D. Talay, Second-order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stoch. Stoch. Rep.* **29** (1990) 13–36.
- [73] D. Talay, Approximation of invariant measures of nonlinear Hamiltonian and dissipative stochastic differential equations, in *Progress in Stochastic Structural Dynamics*, R. Bouc and C. Soize Eds., Publication du L.M.A.-C.N.R.S. **152** (1999) 139–169.
- [74] D. Talay, Stochastic Hamiltonian dissipative systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Proc. Rel. Fields* **8** (2002) 163–198.
- [75] M.E. Tuckerman and G.J. Martyna, Understanding modern molecular dynamics: Techniques and applications. *J. Phys. Chem. B* **104** (2000) 159–178.
- [76] L. Verlet, Computer “experiments” on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **159** (1967) 98–103.
- [77] A.F. Voter, A method for accelerating the molecular dynamics simulation of infrequent events. *J. Chem. Phys.* **106** (1997) 4665–4677.
- [78] A.F. Voter, Parallel replica method for dynamics of infrequent events. *Phys. Rev. B* **57** (1998) 13985–13988.
- [79] W. Wang and R.D. Skeel, Analysis of a few numerical integration methods for the Langevin equation. *Mol. Phys.* **101** (2003) 2149–2156.
- [80] Z. Zhu, M.E. Tuckerman, S.O. Samuelson and G.J. Martyna, Using novel variable transformations to enhance conformational sampling in molecular dynamics. *Phys. Rev. Lett.* **88** (2002) 100201.