

A Gentle Stochastic Thermostat for Molecular Dynamics

Ben Leimkuhler · Emad Noorizadeh · Florian Theil

Received: 16 May 2008 / Accepted: 24 March 2009 / Published online: 8 April 2009
© Springer Science+Business Media, LLC 2009

Abstract We discuss a dynamical technique for sampling the canonical measure in molecular dynamics. We present a method that generalizes a recently proposed scheme (Samoletov et al., J. Stat. Phys. 128:1321–1336, 2007), and which controls temperature by use of a device similar to that of Nosé dynamics, but adds random noise to improve ergodicity. In contrast to Langevin dynamics, where noise is added directly to each physical degree of freedom, the new scheme relies on an indirect coupling to a single Brownian particle. For a model with harmonic potentials, we show under a mild non-resonance assumption that we can recover the canonical distribution. In spite of its stochastic nature, experiments suggest that it introduces a relatively weak perturbative effect on the physical dynamics, as measured by perturbation of temporal autocorrelation functions. The kinetic energy is well controlled even in the early stages of a simulation.

Keywords Ergodicity · Hypocoellipticity · Temperature control · Molecular dynamics · Nosé-Hoover thermostat

1 Introduction

Molecular dynamics requires the use of auxiliary devices for control of the ensemble. In many computations it is desirable that these devices do not substantially corrupt dynamical processes, i.e. that they represent weak perturbations of dynamics.

Consider a physical system described by a Hamiltonian energy function $H(q, p)$, $q, p \in \mathbb{R}^n$. The (forward) trajectories of the corresponding Hamiltonian dynamics are defined for $t \geq 0$ by $q = q(t; q, p)$; $p = p(t; q, p)$, where $\dot{q} = \frac{\partial H}{\partial p}$, $\dot{p} = -\frac{\partial H}{\partial q}$, $q(0) = q$,

B. Leimkuhler (✉) · E. Noorizadeh
The Maxwell Institute and School of Mathematics, University of Edinburgh, Edinburgh EH9 3JZ, UK
e-mail: b.leimkuhler@ed.ac.uk

F. Theil
Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK

$p(0) = p$. With respect to the canonical measure, static observables are functions $O = O(q, p)$, computable by phase space averaging:

$$\langle O \rangle = \int_{\mathbb{R}^{2n}} O(q, p) d\rho_\beta(q, p), \tag{1}$$

where $d\rho_\beta(q, p) = \rho_\beta(q, p)dq dp$,

$$\rho_\beta(q, p) = \frac{1}{Z} e^{-\beta H(q,p)}, \quad Z = \int_{\mathbb{R}^{2n}} e^{-\beta H(q,p)} dq dp$$

is the Boltzmann-Gibbs distribution and β is inverse temperature. The time average can be defined with respect to a phase space curve (typically a non-Hamiltonian or stochastic-dynamics trajectory) $\Gamma = \{(\hat{q}(t), \hat{p}(t)) | t \geq 0\}$ is defined as:

$$\langle O \rangle_\Gamma = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau O(\hat{q}(t), \hat{p}(t)) dt. \tag{2}$$

In the typical case, the computation of long-time averages can be reduced to a sampling problem, i.e. identifying an appropriate means to generate curves Γ such that almost surely (i.e., for almost all initial conditions) $\langle O \rangle_\Gamma = \langle O \rangle$. In many cases the process used to generate sampling trajectories is a perturbation of Hamiltonian dynamics, but this is not essential.

Molecular dynamics is also used to compute dynamic observables, i.e. generalized autocorrelation functions. For example, the canonically weighted momentum autocorrelation function is given by

$$v(\tau) = \frac{1}{\alpha^2} \int_{\mathbb{R}^{2n}} p \cdot p(\tau; q, p) d\rho_\beta(q, p),$$

where, $\alpha^2 = \int_{\mathbb{R}^{2n}} p \cdot p d\rho_\beta(q, p)$, defined in terms of the Hamiltonian trajectories of the system. The most accurate procedure for generating autocorrelation functions at a given temperature is as follows: first, generate a set of well equilibrated initial conditions (at given temperature) through some strongly randomizing process. Next, from each initial point, run a microcanonical (Hamiltonian dynamics) simulation and compute the associated autocorrelation function. The canonically weighted autocorrelation function can then be obtained from a weighted sum of the results for all initial conditions. This straightforward method has the benefit of being trivially parallelizable, however it may introduce artificial nonequilibrium effects at the start of each trajectory, and therefore each trajectory will require an equilibration. As a practical device it is often desirable to compute both static and dynamic observables from a single dynamics. These considerations motivate the search for methods sampling from the canonical distribution that alter the original Hamiltonian evolution in a minimal way, so that autocorrelation functions can be directly computed in a single run. Our goal is therefore to find dynamics which generates trajectories which are close to microcanonical trajectories on finite time intervals sufficiently long to allow the recovery of autocorrelation functions.

Canonical sampling may be achieved using stochastic and deterministic methods. A stochastic method models a heat bath in contact with the system. In this approach the heat bath acts on the system by adding random forces to the system which will be appropriately balanced by a diffusion process (according to the fluctuation-dissipation theorem). The best known representative of this class of methods is the Langevin-thermostat [5]. It replaces

Newtonian dynamics by stochastic dynamics:

$$\frac{dq}{dt} = M^{-1}p, \quad (3)$$

$$dp = -\nabla V dt - \frac{1}{2}\beta\gamma^2 p dt + \gamma dW, \quad (4)$$

where we have assumed that $H(q, p) = \frac{1}{2}p^T M p + V(q)$, M is a mass matrix, γ represents the size of perturbation, and W is a vector of n independent Brownian motions. It is shown that the Boltzmann-Gibbs distribution is the unique invariant measure for the process generated by (3)–(4) [18, 19].

A common deterministic method is known as Nosé-Hoover dynamics [10, 24]. This method augments the physical system with one additional variable ξ called a thermostat variable. The thermostat models an artificial heat bath and is coupled to all the degrees of freedom of the physical system. Moreover the dynamics of ξ is governed by a heuristic auxiliary equation which forces the spontaneous kinetic energy of the system, per degree of freedom, to oscillate around a given target temperature. Nosé-Hoover dynamics takes the form

$$\dot{q} = M^{-1}p, \quad (5)$$

$$\dot{p} = -\nabla V - \xi p, \quad (6)$$

$$Q\dot{\xi} = p^T M^{-1}p - \frac{n}{\beta}, \quad (7)$$

where q and p are position and momentum vectors, respectively, n is the number of degrees of freedom, $\beta^{-1} = k_B T$, k_B is the Boltzmann constant, and T is temperature. The parameter Q is an (artificial) thermostat coefficient that influences the coupling of the heat bath to the system. It can be checked that the distribution with density function

$$\rho_\beta^{\text{aug}} \propto \exp\left(-\beta\left(H + \frac{Q}{2}\xi^2\right)\right),$$

where H is the Hamiltonian of the physical system, is invariant with respect to the flow of (5)–(7) [10, 24]. The Nosé-Hoover evolution is close to the Hamiltonian evolution for sufficiently large systems in the sense that autocorrelation functions computed using Nosé dynamics have been found to have an error inversely proportional to the system dimension (as demonstrated by numerical experiments in [7]), hence the computation of dynamic observables is not severely effected. On the other hand, there are examples that show that the Nosé-Hoover thermostat is not always a sampling method for the canonical distribution. Specifically, the crucial assumption of ergodicity, essential in the proof that Nosé-Hoover dynamics samples the canonical ensemble, may be violated.

Recently, a scheme was presented by Samoletov, Chaplain and Dettman [30] which aims to combine the advantages of the Langevin-thermostat with those of the Nosé-Hoover method; although without proof of ergodicity. Our method generalizes this scheme; it can be viewed as a Nosé-Hoover method in which the thermostat variable is a Brownian particle. For the same case considered in [30], we consider the rigorous foundation of the method in the context of a harmonic potential energy, stating a simple condition under which the method samples the canonical ensemble. The harmonic case is important, as many physical models contain strong harmonic components that trap energy for long time, hence cause

difficulty for thermostats, e.g. crystalline solids [14] and biomolecular models [1]. In full generality, our scheme incorporates a device which may be used to increase ergodicity in those special cases where the non-resonance condition is violated.

We also consider via numerical experiments the issue of the perturbation of dynamics, showing for several examples that autocorrelation functions are relatively mildly perturbed by our method. Besides [30], some other recent articles have used a similar combination of stochastic and deterministic dynamics. Bussi et al. [6] developed a sampling method introducing a stochastic perturbation of velocities, while reducing the extent of random perturbation of the system compare to the Langevin dynamics. On the other hand their method relies on an auxiliary dynamics for kinetic energy and there is no clear case that it can improve the ergodicity. A method related to ours was also suggested by Quigley and Probert [27] for integration in the isothermal-isobaric ensemble. The primary distinction between our approach and others in the literature is that we provide not only a new method (which generalizes all the ones of which we are aware) but also an analysis of ergodicity, making use of the concept of *hypoellipticity* with respect to the operator defining the right hand side of the Fokker-Planck equations. The technique used here for analysis of ergodicity is motivated by recent approaches in [8, 18, 19].

2 Equations of Motion and Invariant Measure

Consider the following family of stochastic dynamics:

$$\frac{dq}{dt} = M^{-1} p, \tag{8}$$

$$\frac{dp}{dt} = -\nabla V(q) - A(\xi)p, \tag{9}$$

$$d\xi = \frac{1}{\mu} \left(p^T M^{-1} p - \frac{n}{\beta} \right) dt - \frac{1}{2} \mu \beta \sigma^2 \xi dt + \sigma dW, \tag{10}$$

where M is a positive definite symmetric matrix, $q, p \in \mathbb{R}^n$, $A(\xi) = \xi \text{Id} + MS(t, \xi)$, $S \in \mathbb{R}^{n \times n}$ is skew-symmetric (i.e., $S^T = -S$), $\xi \in \mathbb{R}$, W is the standard Brownian motion, $\mu > 0$, $\sigma \in \mathbb{R}$ and $\beta = \frac{1}{k_B T} > 0$ is the inverse temperature. For $\sigma = 0$ and $A = \text{Id}$, one obtains the classic Nosé-Hoover thermostat.

The Hamiltonian takes the usual form $H(q, p) = \frac{1}{2} p^T M^{-1} p + V(q)$ and we assume that the potential function is bounded below (i.e., $V : \mathbb{R}^{2n} \rightarrow (a, \infty)$), the augmented Boltzmann-Gibbs density is defined by

$$\rho_\beta^{\text{aug}}(q, p, \xi) := \frac{1}{Z} \exp\left(-\beta\left(H(q, p) + \frac{\mu}{2} \xi^2\right)\right), \tag{11}$$

where

$$Z = \int_{\mathbb{R}^N} dq dp d\xi \exp\left(-\beta\left(H(q, p) + \frac{\mu}{2} \xi^2\right)\right)$$

is the partition function and $N = 2n + 1$.

A function $\rho(q, p, \xi)$ is invariant if it satisfies the stationary Fokker-Planck equation $L^* \rho = 0$ where

$$L^* \rho = -\nabla_q \cdot [\rho M^{-1} p] + \nabla_p \cdot [\rho(\nabla V + Ap)] - \frac{\partial}{\partial \xi} \left[\left(\frac{p^T M^{-1} p - n\beta^{-1}}{\mu} - \frac{1}{2} \mu \beta \sigma^2 \xi \right) \rho \right] + \frac{1}{2} \frac{\partial^2}{\partial \xi^2} [\rho \sigma^2]. \tag{12}$$

It can be checked by inspection the ρ_β^{aug} is invariant.

Indeed, since M is symmetric and S is skew-symmetric one has that $\text{tr}(MS) = 0$, therefore $\nabla_p \cdot (Ap) = n\xi$. One obtains the following expressions for the individual terms in (12):

$$\begin{aligned} \nabla_q \cdot [\rho_\beta^{\text{aug}} p] &= \rho_\beta^{\text{aug}} [-\beta \nabla V \cdot M^{-1} p], \\ \nabla_p \cdot [\rho_\beta^{\text{aug}} (\nabla V + Ap)] &= \rho_\beta^{\text{aug}} [-\beta M^{-1} p \cdot \nabla V - \beta \xi p \cdot M^{-1} p + n\xi], \\ \frac{\partial}{\partial \xi} [\rho_\beta^{\text{aug}} ((p^T M^{-1} p - n\beta^{-1})/\mu - \mu\beta)] &= \rho_\beta^{\text{aug}} [-\beta \xi (p^T M^{-1} p - n\beta^{-1}) \\ &\quad + (\mu\beta)^2 \Sigma \xi^2 - \mu\beta \Sigma], \\ \frac{\partial^2}{\partial \xi^2} [\Sigma \rho_\beta^{\text{aug}}] &= \rho_\beta^{\text{aug}} [\Sigma (\mu\beta)^2 \xi^2 - \mu\beta \Sigma], \end{aligned}$$

where $\Sigma = \frac{1}{2} \sigma^2$. Thus ρ_β^{aug} is invariant if

$$-\beta \xi p^2 + \xi + \beta \xi p^2 - \xi - (\mu\beta)^2 \sigma \xi^2 + \xi - (\mu\beta)^2 \sigma \xi^2 - \mu\beta \Sigma = 0 \tag{13}$$

holds for all $\xi, p \in \mathbb{R}$. Since the matrix S drops out, this is clearly the case. Note that $\sigma = 0$ is admissible, hence we have recovered as a special case Hoover’s classic result [10, 11].

If the process generated by (8)–(10) is ergodic, then a generalization of Birkhoff’s ergodic theorem [2, 15, 25, 26] implies that long trajectories can be used to sample any static observable with respect to the measure ρ_β^{aug} , i.e. there exists a set $U \subset \mathbb{R}^N$ with full measure such that

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau O(q(t), p(t)) dt = \int_{\mathbb{R}^N} O(q, p) d\rho_\beta^{\text{aug}}(q, p, \xi) = \int_{\mathbb{R}^{2n}} O(q, p) d\rho_\beta(q, p),$$

almost surely for all initial values $(q(0), p(0), \xi(0)) \in U$.

For our purposes U needs to be invariant under the flow, i.e.

$$(M^{-1} p, -\nabla V(q)) + \text{span}\{(0, Ap)\} \in TU(q, p), \tag{14}$$

where $TU(q, p)$ is the tangent space of U at (q, p) . We state a basic result which relates ergodicity to regularity of solutions of the Fokker-Planck equation; for similar results see [8, 18, 19], as well as earlier works on ergodicity of Markov chains [21, 28, 29].

Theorem 1 *Let $U \subset \mathbb{R}^N$ be open, connected and invariant under the flow in the sense of (14). If all solutions ρ of $L^* \rho = 0$ are continuous on U , then ρ_β^{aug} is the unique invariant measure on U .*

Proof We show first that the set of ergodic invariant measures is countable. Let ν be an ergodic invariant measure. Since ν is invariant, the Fokker-Planck equation $L^* \nu = 0$ is satisfied. Therefore the measure ν has a continuous density $f \in C(U)$. Define

$$\mathcal{D} = \{f \in C(U) \mid f \text{ is the density of an invariant ergodic measure } \nu\}.$$

Birkhoff’s ergodic theorem implies for each pair $f, g \in \mathcal{D}$ that either $\text{int}(\text{supp}(f)) \cap \text{int}(\text{supp}(g)) = \emptyset$ or $f = g$. Let $K_i \subset U$ be a countable family of bounded open sets such that $\bigcup_{i=1}^\infty K_i = U$. For each i we define the set

$$\mathcal{D}_i = \{f \in \mathcal{D} \mid \text{int}(\text{supp}(f)) \cap K_i \neq \emptyset\}.$$

Since the interiors of the supports of the densities in \mathcal{D} are disjoint

$$\sum_{f \in \mathcal{D}_i} \text{meas}(\text{supp}(f) \cap K_i) \leq \text{meas}(K_i) < \infty.$$

A convergent sum can have only countably many nonzero terms, thus \mathcal{D}_i is countable for each i . This shows that \mathcal{D} is countable.

By the decomposition theorem for invariant measures there exists a countable index set \mathcal{J} , weights $\lambda_j \in [0, 1]$, $j \in \mathcal{J}$ and densities $f_j \in \mathcal{D}$ such that $f_j \neq f_l$ if $j \neq l$ and $\rho_\beta^{\text{aug}} = \sum_{j \in \mathcal{J}} \lambda_j f_j$. Let j be such that $\lambda_j > 0$ and assume that there exists $z \in \partial\text{supp}(f_j)$. Then, by continuity of f_j for every $\varepsilon > 0$ there exists $\delta > 0$ such that $f_j(z') < \varepsilon$ for all $|z - z'| < \delta$. But this is impossible since $\inf_{|z-z'|<1} \rho_\beta^{\text{aug}}(z') > 0$ and $\rho_\beta^{\text{aug}}(z') = f_j(z')$ for all $z' \in \text{int}(\text{supp}(f_j))$.

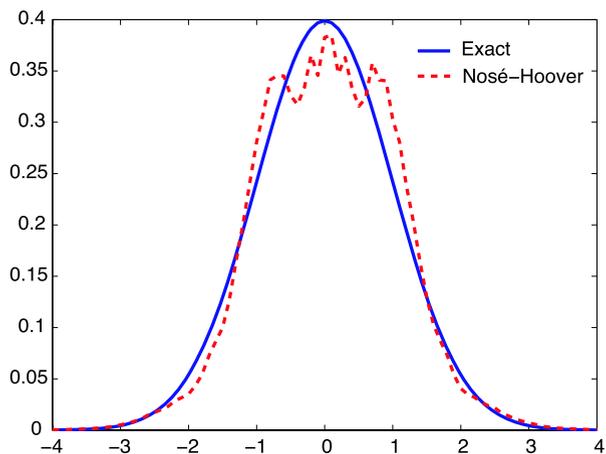
Therefore, $\partial\text{supp}(f_j) = \emptyset$ for every j , and by connectedness of U either $\text{supp}(f_j) = U$ or $\text{supp}(f_j) = \emptyset$. This implies that there exists precisely one $j \in \mathcal{J}$ such that $\lambda_j > 0$. Thus we have shown that $\rho_\beta^{\text{aug}} = f_j$. Since the support of ρ_β^{aug} is U there can be no further ergodic invariant measure and thus $\mathcal{D} = \{\rho_\beta^{\text{aug}}\}$. □

3 Ergodicity and Hypocoellipticity of L^*

It is well known that in general the Nosé-Hoover thermostat ($\sigma = 0$) is not ergodic, see [10, 16, 17]. The most notorious case is given by the harmonic oscillator, where $n = M = \beta = \mu = 1$, $A = \xi$ (see Fig. 1).

When $\sigma \neq 0$, the Fokker-Planck equation $L^* \rho = 0$ changes type. This is captured by the concept of *hypoellipticity*.

Fig. 1 Non-ergodicity of Nosé-Hoover: the graph compares the approximated density of momentum with the exact density



Definition 1 Let $U \subset \mathbb{R}^N$ be open. A second order differential operator L with C^∞ coefficients is hypoelliptic on U if all distributional solutions ρ of the differential equation $L\rho = 0$ are C^∞ .

If U is connected an obvious necessary condition for hypoellipticity of L^* on U is that U cannot be written as a union of several invariant sets. In the case of the Hoover-Langevin thermostat, if q is an equilibrium (i.e. $\nabla V(q) = 0$), then $(q, 0, \xi) \notin U$ for every ξ if L^* is U -hypoelliptic. Moreover, if V is quadratic, then the span of any collection of eigenspaces is invariant.

A sufficient criterion for hypoellipticity is provided by Hörmander’s condition.

Definition 2 Let $U \subset \mathbb{R}^N$ be open, the vector fields $X_0, \dots, X_r : U \rightarrow \mathbb{R}^N$ satisfy Hörmander’s condition at $z \in U$ if the vector space generated by the iterated brackets

$$X_0(z), \dots, X_r(z), [X_i, X_j](z), [X_i, [X_j, X_k]](z) \dots$$

is \mathbb{R}^N .

Typically we will choose $r = 1$ and

$$X_0 = \left(M^{-1}p, -\nabla V - Ap, \frac{1}{\mu} \left(p^T M^{-1}p - \frac{n}{\beta} \right) - \frac{1}{2} \mu \beta \sigma^2 \xi \right), \quad X_1 = (0, 0, \sigma). \tag{15}$$

The main application of the Hörmander’s condition is Hörmander’s theorem [9, 12, 23].

Theorem 2 Let $U \subset \mathbb{R}^N$ be open. If $X_0, X_1 : U \rightarrow \mathbb{R}^N$ are two vector fields that satisfy Hörmander’s condition at every $z \in U$, then the operator L^* which is defined by

$$L^* \rho(z) := - \sum_{i=1}^N \frac{\partial}{\partial z_i} (\rho(z) X_{0,i}(z)) + \frac{1}{2} \sum_{i,j=1}^N \frac{\partial^2}{\partial z_i \partial z_j} (\rho(z) X_{1,i}(z) X_{1,j}(z))$$

is hypoelliptic.

Hypoellipticity clearly provides smoothness of solutions required for application of Theorem 1, hence the flow induced by (8)–(10) is ergodic if we can find an open, connected set U with full measure such that the vector fields X_0 and X_1 satisfy Hörmander’s condition at every $z \in U$.

A simple case where this can be done is given by quadratic Hamiltonians, where

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

Only a mild assumption on the spectrum of B is needed in the case where $A(\xi) = \xi$. For general forces we conjecture that L^* remains hypoelliptic, but it is difficult to verify this analytically due to the long calculation of iterated Lie brackets and the need to show that they are linearly independent. However for Langevin dynamics (3)–(4) it is possible to verify hypoellipticity for bounded Lipschitz forces since one does not need to iterate brackets [18, 19].

Theorem 3 Let $M, B \in \mathbb{R}^{n \times n}$ be two symmetric and positive definite matrices such that

$$\omega_k \neq \omega_l \quad \text{for all } k \neq l, \tag{16}$$

where $\omega_k = \varphi_k^T M^{-1} B \varphi_k$ are the eigenvalues and $\varphi_1, \dots, \varphi_n \in \mathbb{R}^n$ are the normalized eigenvectors of $M^{-1} B$. If $H(q, p) = \frac{1}{2} p^T M^{-1} p + \frac{1}{2} q^T B q$ and

$$U = \left\{ (q, p) \mid \prod_{k=1}^n ((q \cdot \varphi_k)^2 + (p \cdot \varphi_k)^2) \neq 0 \right\} \times \mathbb{R}, \tag{17}$$

then the vector fields $X_0 = (p, -Bq - \xi p, \frac{1}{\mu}(p^T M^{-1} p - n\beta^{-1}) - \frac{1}{2}\mu\beta\sigma^2\xi)$ and $X_1 = (0, 0, \sigma)$ satisfy the Hörmander condition at each $(q, p, \xi) \in U$. In particular the process generated by (8)–(10) is ergodic on U .

We conjecture that if the matrix A in (9) is random, then Theorem 3 holds almost surely without the non-resonance assumption (16).

Conjecture 1 Let M, B be symmetric, positive definite matrices. If $H(q, p) = \frac{1}{2}(p^T M^{-1} p + q^T B q)$, $A = \xi \text{Id} + SM$ where $S = G - G^T$ and $G \in \mathbb{R}^{n \times n}$ is a random matrix with iid Gaussian entries, then for almost every realization of S the flow generated by (8)–(10) is ergodic on U (defined by (17)).

The theorem is sharp in the sense that if one of the assumption (16), (17) is violated, then the dynamics generated by equations (8)–(10) is not ergodic. Indeed, assume that B is a diagonal matrix and $q_i(t = 0) = p_i(t = 0) = 0$ for some i . Clearly $q_i(t), p_i(t) = 0$ for all t and thus the evolution is not ergodic.

Assume next that $n = 3$ and $M = B = \text{Id}$ (the identity matrix). Define the subspace

$$S = \text{span}\{(q_0, 0), (p_0, 0), (0, q_0), (0, p_0)\} \subset \mathbb{R}^6,$$

where q_0 and p_0 are the initial values of q and p . Again, it can be seen easily that S is invariant. Since S is 4-dimensional the evolution is not ergodic.

A nontrivial quadratic Hamiltonian that satisfies (16) is a harmonic chain with clamped end-particles where $V(q) = \frac{1}{2} \sum_{i=0}^n (q_{i+1} - q_i)^2$ and $q_0 = q_{n+1} = 0$. Then $\partial V(q)/\partial q_i = -q_{i-1} + 2q_i - q_{i+1}$ if $i \in \{1, \dots, n\}$. Without the clamping assumption the Hamiltonian H is translation invariant and $Z = \int_{\mathbb{R}^{2(n+2)}} dq dp \exp(-\beta H)$ does not exist. Define the discrete sine-transform as follows: $(\mathcal{F}q)_k = \hat{q}_k = \frac{2}{n+1} \sum_{i=1}^n \sin(\pi i k / (n + 1)) q_i$, such that $q_i = \sum_{k=1}^n \hat{q}_k \sin(\pi i k / (n + 1))$. One obtains that $|\hat{q}| = |q|$ and

$$\mathcal{F}(-q_{i-1} + 2q_i - q_{i+1})(k) = 2(1 - \cos(\pi k / (n + 1))) \hat{q} = \omega_k \hat{q}(k).$$

Since the dispersion relation ω is strictly increasing with k , inequality (16) is satisfied.

Proof of Theorem 3 We can assume without loss of generality that $\sigma = \mu = 1 = \beta = 1$ and $M = \text{Id}$. Furthermore, we assume that B is diagonal, hence $H(q, p) = \frac{1}{2} \sum_{k=1}^n (\omega_k q_k^2 + p_k^2)$. This assumption does not involve any loss of generality since it amounts to choosing the coordinate system which is created by the eigenvectors $\varphi_1 \dots \varphi_n$.

After these simplifications the vector fields X_0 and X_1 assume the form

$$X_0 = \left(p, -Bq - \xi p, p^2 - n - \frac{1}{2}\xi \right), \quad X_1 = (0, 0, 1).$$

Next, we define recursively the following sequence of vector fields:

$$Z_k = \frac{1}{2}[Y_k, X_3], \quad Y_{k+1} = -\frac{1}{2}[Z_k, X_3],$$

where

$$\begin{aligned} X_2 &= \left[X_1, \left(p^2 - n - \frac{1}{2}\xi \right) X_1 - X_0 \right] = (0, p, 0), \\ X_3 &= X_0 - \left(p^2 - n - \frac{1}{2}\xi \right) X_1 + \xi X_2 = (p, -Bq, 0), \\ Y_1 &= [X_2, X_3] = (p, Bq, 0). \end{aligned}$$

Induction yields that

$$Y_k = (B^{k-1}p, B^kq, 0), \quad Z_k = (B^kq, -B^k p, 0), \quad k = 1, 2, \dots$$

After these preparations we can show that the vectors $X_1, Y_1, Z_1, \dots, Y_{n-1}, Z_{n-1}, Y_n, Z_n$ span \mathbb{R}^{2n+1} . Clearly, it suffices to demonstrate that for each $\eta, \mu \in \mathbb{R}^n$ there exist coefficients $a_1, b_1, \dots, a_n, b_n \in \mathbb{R}$ such that

$$\sum_{k=1}^n (a_k Y_k + b_k Z_k) = \sum_{k=1}^n (a_k B^{k-1}p + b_k B^kq, a_k B^kq - b_k B^k p) = (\eta, \mu). \tag{18}$$

Since the matrix B is diagonal (18) is equivalent to

$$\begin{pmatrix} \text{diag}(B^{-1}p) & \text{diag}(q) \\ \text{diag}(q) & -\text{diag}(p) \end{pmatrix} \begin{pmatrix} \mathcal{V}a \\ \mathcal{V}b \end{pmatrix} = \begin{pmatrix} \eta \\ \mu \end{pmatrix},$$

where $\mathcal{V}_{kl} = \omega_k^l, k, l = 1, \dots, n$ is a Vandermonde matrix with determinant

$$\det(\mathcal{V}) = \prod_k \omega_k \prod_{k>l} (\omega_k - \omega_l).$$

Set now

$$\tilde{a} = \mathcal{V}a, \quad \tilde{b} = \mathcal{V}b, \tag{19}$$

then the k -th components of \tilde{a}, \tilde{b} solve of the linear system

$$\begin{pmatrix} \omega_k^{-1} p_k & q_k \\ q_k & -p_k \end{pmatrix} \begin{pmatrix} \tilde{a}_k \\ \tilde{b}_k \end{pmatrix} = \begin{pmatrix} \eta_k \\ \mu_k \end{pmatrix},$$

i.e. $\begin{pmatrix} \tilde{a}_k \\ \tilde{b}_k \end{pmatrix} = \frac{1}{\omega_k^{-1} p_k^2 + q_k^2} \begin{pmatrix} p_k & q_k \\ q_k & -p_k \end{pmatrix} \begin{pmatrix} \eta_k \\ \mu_k \end{pmatrix}.$

The coefficient vectors a and b are obtained by inverting the relation (19) which is possible since we have assumed that the eigenvalues ω_i are pairwise different from each other and bigger than zero, thus the determinant of \mathcal{V} is nonzero. □

It is important to note that, as with all dynamics-based and stochastic dynamics-based schemes, numerical truncation errors will likely distort the distribution. (These effects are discussed in detail in [3] in the setting of Nosé dynamics.)

4 Numerical Results

In this section we run a series of tests on the system (8)–(10) to investigate the validity of the invariant measure ρ_β^{aug} and its applications. We used the following discretization for the system (8)–(10)

$$\begin{aligned} Q &:= q^n + \frac{\Delta t}{2} p^n, \\ P &:= p^n - \frac{\Delta t}{2} \nabla V(q^{n+1/2}) - \frac{\Delta t}{2} \bar{\xi} P, \\ \bar{\xi} &:= \xi^n + \frac{\Delta t}{2} \mu^{-1} \left(\sum \frac{P_i^2}{m_i} - \frac{n}{\beta} \right) - \frac{\Delta t}{4} \sigma^2 \mu \beta \bar{\xi} + \frac{\sigma}{2} \sqrt{\Delta t} W, \\ p^{n+1} &:= 2P - p^n, \\ \xi^{n+1} &:= 2\bar{\xi} - \xi^n, \\ q^{n+1} &:= Q + \frac{\Delta t}{2} p^{n+1}. \end{aligned}$$

Here W is a standard normal random variable. This method is semi-implicit (requiring an iteration to solve at each step), but it is important to note that only one force evaluation is required at a timestep. In practice the method therefore has the cost of an explicit integrator such as comparable methods [5, 22, 32] for Langevin dynamics. After preparing the numerical experiments, we obtained the following explicit discretization:

$$\begin{aligned} P &:= p^n - \frac{\Delta t}{2} \nabla V(q^n), \\ Q &:= q^n + \frac{\Delta t}{2} P, \\ P &:= \exp(-\delta t \xi^n / 2) P, \\ \xi^{n+1} &:= \xi^n + \frac{\Delta t}{\mu} \left(\sum \frac{P_i^2}{m_i} - \frac{n}{\beta} \right) + \sigma \sqrt{\Delta t} W - \frac{\Delta t \sigma^2}{4\mu} (\xi^n + \xi^{n+1}), \\ P &:= \exp(-\Delta t \xi^{n+1} / 2) P, \\ q^{n+1} &:= Q + \frac{\Delta t}{2} P, \\ p^{n+1} &:= P - \frac{\Delta t}{2} \nabla V(q^{n+1}). \end{aligned}$$

The implicit equation at the 4th step of this algorithm is linear, so can be easily solved exactly. Tests confirmed that the performance of the two methods was very similar.

Alternative discretizations may be obtained by following the procedures described in [4, 5, 20, 22, 31, 32].

4.1 Harmonic Oscillator

First we investigate the dynamics of (8)–(10) for the case where the energy of the system is given by a Hamiltonian of the form

$$H(q, p) = \frac{p^2}{2m} + \omega^2 \frac{q^2}{2}.$$

In our experiment we chose $\omega = m = 1, \beta = 1.0, \mu = 0.5, \sigma = 5.0$ and $\Delta t = 0.01$. The parameter μ influences the control on temperature and σ influences the coupling between system and the heat bath. To verify that our dynamics generates the Boltzmann-Gibbs distribution, the distribution of momentum is compared to $\sqrt{\frac{\beta}{2\pi m}} e^{-\beta \frac{p^2}{2m}}$. This is demonstrated in Fig. 2.

In order to quantify the error in the distribution generated by (8)–(10), we define the following norm. For a given interval (a, b) , define

$$D_n(x) = \left(\frac{1}{M} \sum_{i=1}^M \left(\phi_{K_i}(x) - \int_{K_i} d\rho_\beta \right)^2 \right)^{\frac{1}{2}}, \tag{20}$$

where x is a set of size n samples generated by the dynamics, (K_1, \dots, K_M) are M partitions of (a, b) and $\phi_{K_i}(x)$ is the observed density of samples in x which belong to the partition K_i . We postulate that the convergence of $D_n(x)$ toward zero implies the law of large numbers

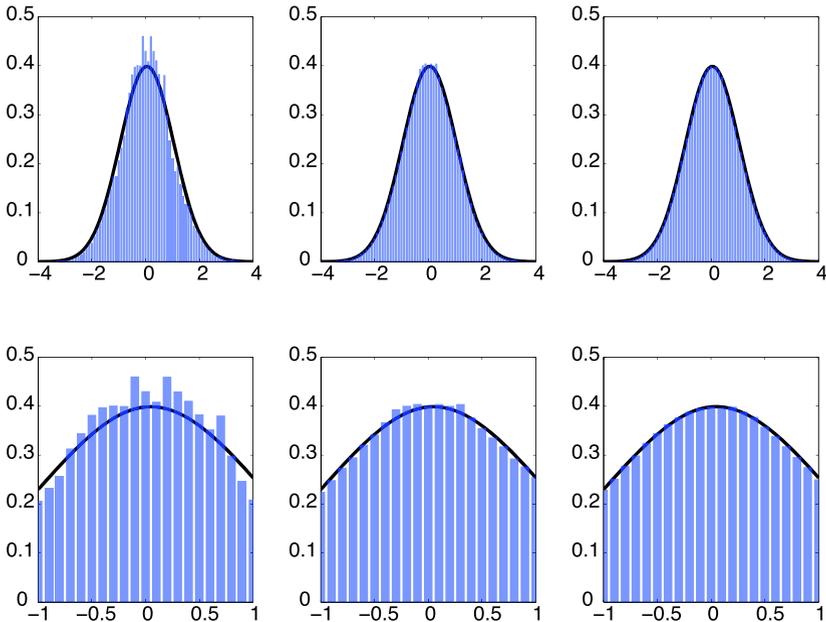


Fig. 2 Convergence of momentum distribution is verified for the harmonic oscillator. The solid line is the exact density and the approximated density is in bar style. The (left) column 10^5 steps, the (middle) column 10^6 of steps and the (right) column 10^7 steps, each of step of size $\Delta t = 0.01$

Fig. 3 The graph shows the error $D_n(x)$ in the approximated density of momentum against the number of samples n . The rate of convergence of the distribution of Hoover-Langevin is similar to Nosé-Hoover chain (NHC) and Langevin dynamics for the case of harmonic oscillator

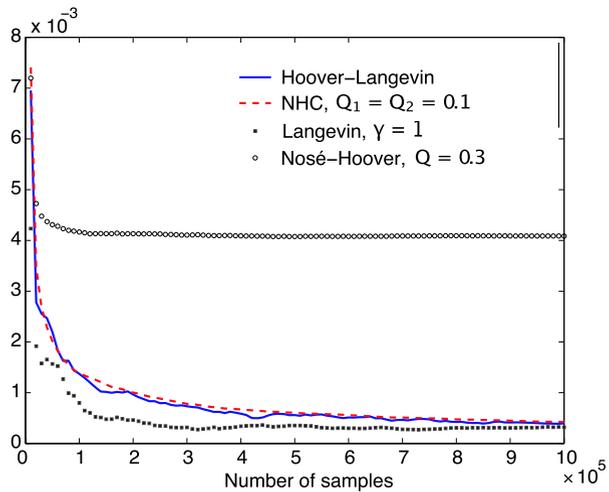


Table 1 Error (20) in distribution for p , p^2 and p^4 using Hoover-Langevin

	Error for 10^5 evolutions	Error for 10^6 evolutions	Error for 10^7 evolutions
p	0.201035×10^{-2}	0.454371×10^{-3}	0.167924×10^{-3}
p^2	0.912343×10^{-3}	0.207135×10^{-3}	0.444854×10^{-6}
p^4	0.130941×10^{-2}	0.251866×10^{-3}	0.487444×10^{-6}

and the rate of convergence of $D_n(x)$ is related to the rates of convergence of average of observable.

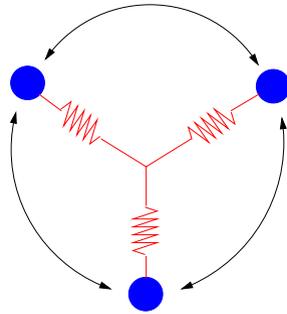
In Fig. 3, we compare the error norm $D_n(x)$ for the new dynamics (Hoover-Langevin) with other widely used sampling methods namely Nosé-Hoover chains (NHC) [17] (an extension of Nosé-Hoover where a chain of thermostats ξ_i with thermostat coefficient Q_i are attached to the system) and Langevin dynamics [5] to investigate the rate of convergence. We chose $\gamma = 1$ for Langevin and $Q_1 = Q_2 = 0.1$ for NHC which we observed to be optimal parameters for these methods. In order to reduce the inconsistency in the results due to the random noise, for each method, 100 different simulations with different initial conditions have been performed and the result illustrated in Fig. 3 is the mean of the 100 different results.

We also computed the errors (20) in distribution for p^2 and p^4 , which are presented in Table 1.

4.2 Discrepancy in the Dynamics

One important aspect of molecular dynamics (MD) is to capture macroscopic information from the dynamics of atoms or small constituent parts that form a material. Therefore it is essential to take care that the algorithm used in MD is not changing the dynamics of the physical system significantly. The new dynamics is designed to generate the canonical distribution by introducing a minimal perturbation to the system so that the dynamics of the thermostated system is close to the unperturbed system.

Fig. 4 Three particles of mass m are connected by springs to the origin and interacting with each other through Lennard-Jones (LJ) potential



Consider a two dimensional system consisting of three particles which are connected by springs with rest length to a fixed point at the origin (Fig. 4). The interaction between particles is modelled by Lennard-Jones potential,

$$U_{LJ}(r) = 4\epsilon \left[\left(\frac{\alpha}{r} \right)^{12} - \left(\frac{\alpha}{r} \right)^6 \right].$$

The Hamiltonian of the system is

$$H(q, p) = \sum_{i=1}^3 \frac{1}{2m_i} p_i^2 + \sum_{i=1}^3 \frac{1}{2} k (L - \|q_i\|)^2 + \sum_{i=1}^2 \sum_{j=i+1}^3 U_{LJ}(r_{ij}), \tag{21}$$

where L is the spring rest length, k is the spring constant, $r_{ij} = \|q_j - q_i\|$ and U_{LJ} is the Lennard-Jones potential. This is a challenging problem in terms of equilibration due to the locking of energy in springs.

In our simulation we took $\alpha = \epsilon = 1$, $k = 10$, $L = 1$, $m_i = 1$ for $i = 1, 2, 3$ and set the target temperature $T = 1$, $k_B = 1$. In order to measure the changes in the dynamics we look at the velocity autocorrelation function of the radial component of velocity,

$$v_{r_i}(t) = \frac{\dot{q}_i \cdot q_i}{\|q_i\|}. \tag{22}$$

To calculate the canonically weighted VAF function we first construct a set of 1000 random initial conditions $\{z_i\}$ from a canonical distribution at the target temperature. From each z_i we run a microcanonical simulation and calculate its VAF, the correct VAF is then obtained as a weighted average of VAFs from different initial conditions:

$$\bar{c}(\tau) = \frac{\sum_i c(\tau; z_i) \rho_\beta(z_i)}{\sum_i \rho_\beta(z_i)}, \tag{23}$$

where

$$c(\tau; z) = \lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T \frac{v_{r_1}(t; z) v_{r_1}(t + \tau; z)}{v_{r_1}(t; z) v_{r_1}(t; z)} dt, \tag{24}$$

with v_{r_1} representing, in this case, the radial velocity of the first particle of the system.

Figure 5 compares the radial VAF for Hoover-Langevin with those obtained by other methods. The parameters are chosen with the criteria to achieve the correct distribution:

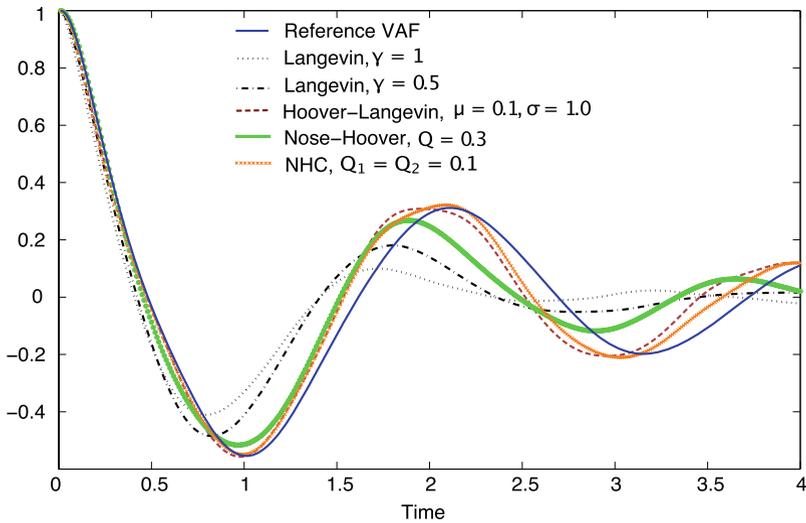


Fig. 5 Autocorrelation function $c_1(\tau)$, computed using Hoover-Langevin, Langevin, Nose-Hoover and NHC, and compared to the velocity autocorrelation of canonically averaged microcanonical ($\bar{c}_1(\tau)$) dynamics

$\mu = 0.1, \sigma = 1$ for Hoover-Langevin, $\gamma = 1$ and $\gamma = 0.5$ for Langevin, $Q = 0.3$ for Nosé-Hoover and $Q_1 = Q_2 = 0.1$ for NHC. We used these values of the Langevin parameter so that the error in its distribution is of the same size of the error in the distribution for Hoover-Langevin. Moreover, we observed that for $\gamma < 0.5$ the temperature fails to reach its target value within the simulation time, we elaborate more on temperature in the next subsection. As can be seen from Fig. 5, Hoover-Langevin follows the VAF of microcanonical (unperturbed dynamics) very closely, whereas the Langevin dynamics for $\gamma = 1.0$ and $\gamma = 0.5$ profoundly changes the VAF, since it perturbs every degree of freedom by adding random noise. Using smaller values of γ would improve the result for the VAF for Langevin dynamics, albeit at the expense of further perturbing the distribution obtained on a fixed time interval. Hence we compare the VAF of Langevin and Hoover-Langevin for the same level of perturbation needed for each method to approximate the Gibbs measure with the same accuracy, with a given amount of computational effort, Table 2 shows that Langevin dynamics and Hoover-Langevin approximate the Gibbs measure with very close accuracy for the chosen values of parameters. This illustrates that the dynamics of Hoover-Langevin has the characteristic of deterministic thermostats of being close to the original dynamics despite the fact that it is a stochastic method.

The error in VAF and the error in distribution for Hoover-Langevin, Nose-Hoover, NHC and Langevin method are shown in Table 2. It worth noting that Langevin fails to produce the correct qualitative approximation of VAF as is visible in Fig. 5.

4.3 Temperature Control

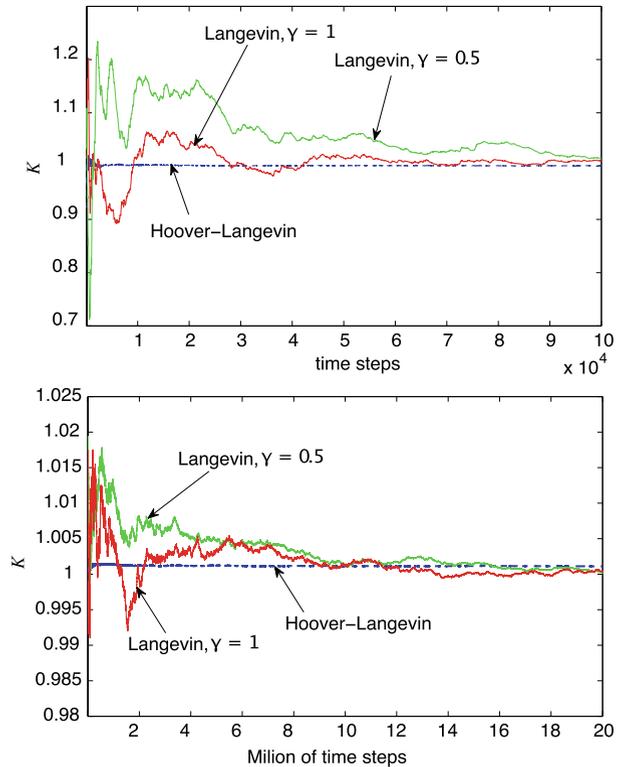
One important feature of the new dynamics is the control feedback loop in the dynamics which stabilizes the cumulative average kinetic energy of the system near the target temperature. Cumulative average kinetic energy is defined by

$$K(t) = \frac{1}{t} \int_0^t n^{-1} p^T(s) M^{-1} p(s) ds.$$

Table 2 Comparison of root mean square of error on $[0, 4]$ of VAF and the error in distribution using (20) for 10^6 of $\Delta t = 0.01$ evaluations

Method	Parameters	Error in distribution	Error on $[0, 4]$ of VAF
Hoover-Langevin	$\mu = 0.1, \sigma = 1$	0.270198×10^{-3}	0.0675
Hoover-Langevin	$\mu = 0.1, \sigma = 10$	0.232064×10^{-3}	0.0578
Langevin	$\gamma = 0.5$	0.252864×10^{-3}	0.1018
Langevin	$\gamma = 1$	0.228635×10^{-3}	0.1383
NHC	$Q_1 = Q_2 = 0.1$	0.275997×10^{-3}	0.0603
Nosé-Hoover	$Q = 0.3$	0.165209×10^{-2}	0.0807

Fig. 6 The (top) panel shows cumulative kinetic energy during 10^5 of time steps ($\Delta t = 0.01$) simulation. $K(t)$ computed by Hoover-Langevin dynamics reaches 1 (the target temperature) and stays close to 1, whereas it takes longer for Langevin dynamics to reach the target temperature and the deviation is greater. The (lower) panel shows the slow convergence of temperature over twenty million time steps



In Fig. 6 we compare the $K(t)$ of the Hoover-Langevin with the Langevin dynamics for the system (21). We used $\mu = 0.1, \sigma = 1$ for Hoover-Langevin and $\gamma = 1$ and $\gamma = 0.5$ for Langevin, both methods produce correct Gibbs measure in the long term, but the convergence of $K(t)$ is much slower for Langevin dynamics. Note that this experiment does not demonstrate convergence to equilibrium; for general convergence to equilibrium one needs to compare the spectral gaps of the generator of the process [13, 33], which is difficult to do for Hoover-Langevin because it is highly degenerate.

5 Conclusion

We have presented a new thermostat for generating the canonical distribution in molecular dynamics simulations. This thermostat is derived by combining Nosé-Hoover and Langevin dynamics together with the aim to achieve a provably correct distribution and at the same time minimizing the effect on the dynamics. The new method should be of interest in cases where one is concerned with computing the average of local observables which depend on small number of degrees of freedom, for instance, for calculating free energy of activated processes where the process occurs along a reaction coordinate which can be described as a function of the degrees of freedom of the system. This new thermostat is likely to be preferable for some non-equilibrium molecular dynamics simulations compared to the Langevin method, since it is close to the dynamics of the unperturbed system, and therefore interacts weakly with a non-equilibrium force acting on the system.

The new dynamics has an invariant probability measure ρ_β^{aug} which is proportional to the Boltzmann-Gibbs distribution and we have proved analytically that under a non-resonance assumption, an open, connected set U with full measure can be constructed such that ρ_β^{aug} is ergodic on U . Thus, when the new thermostat is applied to quadratic Hamiltonians without resonances the dynamics is ergodic. This has been verified in several simple examples.

Acknowledgements The authors thank Martin Hairer, Oliver Penrose, Gabriel Stolz and the anonymous referees for very helpful comments and suggestions. The second author was supported by funds from the UK Engineering and Physical Sciences Research Council, and the Maxwell Institute. We gratefully acknowledge the financial support of the EPSRC network “Mathematical Challenges of Molecular Dynamics: A Chemo-Mathematical Forum” (EP/F03685X/1).

References

1. Barth, E., Leimkuhler, B., Sweet, C.: Approach to thermal equilibrium in biomolecular simulation. *Lect. Notes Comput. Sci. Eng.* **49**, 125–140 (2005)
2. Birkhoff, G.D.: Proof of the ergodic theorem. *Proc. Natl. Acad. Sci. U.S.A.* **17**(12), 656 (1931)
3. Bond, S.D., Leimkuhler, B.J.: Molecular dynamics and the accuracy of numerically computed averages. *Acta Numer.* **16**, 1–65 (2007)
4. Bou-Rabee, N., Owjadi, H.: Boltzmann-Gibbs preserving stochastic variational integrator (2007). <http://arxiv.org/abs/0712.4123>
5. Brünger, A., Brooks, C.B., Karplus, M.: Stochastic boundary conditions for molecular dynamics simulations of st2 water. *J. Chem. Phys. Lett.* **105**(5), 495–500 (1984)
6. Bussi, G., Donadio, D., Parrinello, M.: Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014,101 (2007)
7. Evans, D., Holian, B.: The Nosé-Hoover thermostat. *J. Chem. Phys.* **83**, 4069–4074 (1985)
8. Hairer, M., Mattingly, J.: Ergodicity of the 2d Navier-Stokes equations with degenerate stochastic forcing. *Ann. Math.* **164**(3) (2006)
9. Helffer, B., Nier, F.: *Hypoelliptic Estimates and Spectral Theory for Fokker-Planck Operators and Witten Laplacians*. Springer, New York (2005)
10. Hoover, W.: Canonical dynamics: equilibrium phase space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985)
11. Hoover, W.G.: *Molecular Dynamics*. Springer, New York (1986)
12. Hörmander, L.: *The Analysis of Linear Partial Differential Operators*. Springer, New York (1985)
13. Ingrassia, S.: On the rate of convergence of the metropolis algorithm and Gibbs sampler by geometric bounds. *Ann. Appl. Probab.* **4**(2), 347–389 (1994)
14. Kaczmarek, M., Rurali, R., Hernández, E.: Reversible scaling simulations of the melting transition in silicon. *Phys. Rev. B* **69**, 214,105 (2004)
15. Khinchin, A.I.: *Mathematical Foundations of Statistical Physics*. Dover, New York (1949)
16. Legoll, F., Luskin, M., Moeckel, R.: Non-ergodicity of the Nosé-Hoover thermostatted harmonic oscillator. *Arch. Ration. Mech. Anal.* **184**, 449–463 (2007)

17. Martyna, G.J., Klein, M.L., Tuckerman, M.: Nosé-Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **97**(4), 2635–2643 (1992)
18. Mattingly, J.C., Stuart, A.M.: Geometric ergodicity of some hypo-elliptic diffusions for particle motions. *Markov Processes Relat. Fields* **8**(2), 199–214 (2002)
19. Mattingly, J.C., Stuart, A.M., Higham, D.J.: Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stoch. Process. Appl.* **101**(2), 185–232 (2002)
20. Melchionna, S.: Design of quasisymplectic propagators for Langevin dynamics. *J. Chem. Phys.* **127**, 044,108 (2007)
21. Meyn, S.P., Tweedie, R.: *Markov Chains and Stochastic Stability*. Springer, London (1993)
22. Milstein, G., Tretyakov, N.: Quasi-symplectic methods for Langevin-type equations. *IMA J. Numer. Anal.* **23**(3), 593–626 (2003)
23. Norris, J.: Simplified Malliavin calculus. In: *Séminaire de Probabilités XX 1984/85. Lecture Notes in Mathematics*, pp. 101–130. Springer, Berlin (1986)
24. Nosé, S.: A unified formulation of the constant temperature molecular dynamics method. *J. Chem. Phys.* **81**, 511–519 (1984)
25. Penrose, O.: *Foundations of Statistical Mechanics: a Deductive Treatment*. Pergamon, Elmsford (1970)
26. Petersen, K.: *Ergodic Theory*. Cambridge Studies in Advanced Mathematics, vol. 2. Cambridge University Press, Cambridge (1989)
27. Quigley, D., Probert, M.: Langevin dynamics in constant pressure extended systems. *J. Chem. Phys.* **120**, 11432 (2004)
28. Roberts, G.O., Tweedie, R.L.: Exponential convergence of Langevin diffusions and their discrete approximations. *Bernoulli* **2**(4), 341–363 (1995)
29. Rosenthal, J.: Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Am. Stat. Assoc.* **90**(430), 558–566 (1995)
30. Samoletov, A., Chaplain, M.A.J., Dettmann, C.P.: Thermostats for “slow” configurational modes. *J. Stat. Phys.* **128**, 1321–1336 (2007)
31. Skeel, R.D., Izaguirre, J.A.: An impulse integrator for Langevin dynamics. *Mol. Phys.* **100**, 3885 (2002)
32. Vanden-Eijnden, E., Ciccotti, G.: Second-order integrators for Langevin equations with holonomic constraints. *Chem. Phys. Lett.* **429**, 310–316 (2006)
33. Villani, C.: Topics in optimal transportation. *Am. Math. Soc.* **58** (2003)