Lecture Notes on Molecular Dynamics

Sebastian Reich*

Department of Mathematics, University of Potsdam, D-14415 Potsdam, Germany

(Dated: January 17, 2006)

*Electronic address: sreich@math.uni-potsdam.de

Contents

I. Why does molecular dynamics work? 2

A. Introduction 2

B. Modified equation analysis and conservation of energy 4

C. Shadowing 7

D. The Liouville picture 8

II. Oscillatory systems 9

A. Trigonometric approximation of Gaussian processes 9

B. Hamiltonian systems and heat baths 10

C. Multiple time stepping methods and numerical resonances 11

D. Mollified methods and regularized equations 12

III. Hybrid Monte Carlo methods 13

A. Generalized hybrid Monte Carlo (GHMC) algorithm 13

1. Molecular dynamics Monte Carlo (MDMC) 13

2. Partial momentum refreshment 14

3. Special cases of GHMC 14

4. Applications of hybrid Monte Carlo methods 15

B. Targeted shadow hybrid Monte Carlo (TSHMC) algorithm 15

1. Molecular dynamics step and modified energies 16

2. Alternative momentum updates 17

C. Sampling aspects 18

I. WHY DOES MOLECULAR DYNAMICS WORK?

A. Introduction

Classical molecular dynamics (MD) can be characterized in terms of Hamiltonian equations of motion

$$\dot{q} = M^{-1}p, \qquad \dot{p} = -\nabla V(q) \tag{1}$$

with atomistic positions $q \in \mathbb{R}^{3N}$ and momenta $p \in \mathbb{R}^{3N}$, N the number of atoms, $M \in \mathbb{R}^{3N\times 3N}$ the (diagonal) mass matrix, and $V : \mathbb{R}^{3N} \to \mathbb{R}$ the potential energy function.

The system of equations (1) can be written more compactly as

$$\dot{z} = J\nabla H(z) \tag{2}$$

with state variable $z = (q^T, p^T)^T$, Hamiltonian

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

and structure matrix

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}. \tag{4}$$

Alternatively, we may derive the equivalent Euler-Lagrange equation

$$M\ddot{q} + \nabla V(q) = 0 \tag{5}$$

by extremizing the functional (action)

$$\mathcal{L}[q] = \int_0^T \left\{ \frac{1}{2}\dot{q}^T M\dot{q} - V(q) \right\} dt \tag{6}$$

over all paths $q(t)$, $0 \le t \le T$, with fixed endpoints $q(0) = q_a$ and $q(T) = q_b$.

The flow map of a differential equation (DE)

$$\dot{z} = f(z) = J\nabla H(z) \tag{7}$$

is denoted by

$$z(t) = \phi_{t,f}(z_0) \tag{8}$$

or, in case of Hamiltonian DEs,

$$z(t) = \phi_{t,H}(z_0), \tag{9}$$

respectively, where $z_0 = z(0)$. We consider numerical approximations using one-step methods of order $k \ge 1$, i.e.

$$z^{n+1} = \Psi_{\Delta t}(z^n), \qquad \Psi_{\Delta t}(z) - \phi_{\Delta t,f}(z) = \mathcal{O}(\Delta t^{k+1}). \tag{10}$$

Example. Störmer-Verlet (SV) method:

$$p^{n+1/2} = p^n - \frac{\Delta t}{2}\nabla V(q^n), \quad q^{n+1} = q^n + \Delta t M^{-1} p^{n+1/2}, \quad p^{n+1} = p^{n+1/2} - \frac{\Delta t}{2}\nabla V(q^{n+1}), \tag{11}$$

which is equivalent to the position-only two-step leapfrog formulation

$$M\left(q^{n+1} - 2q^n + q^{n-1}\right) = -\Delta t^2 \nabla V(q^n). \tag{12}$$

□

The SV method is symplectic and time-reversible. A map $\Psi : \mathbb{R}^{6N} \to \mathbb{R}^{6N}$ is called symplectic if the Jacobian $D\Psi(z)$ satisfies

$$D\Psi(z)^T J^{-1} D\Psi(z) = J^{-1}. \tag{13}$$

A map Ψ is called time-reversible with respect to the involution $\hat{z} = Sz$, defined by $\hat{q} = q$, $\hat{p} = -p$, if

$$S\Psi(Sz) = \Psi^{-1}(z). \tag{14}$$

The leapfrog formulation of the SV method extremizes a discrete functional (action)

$$L[\{q^n\}] = \sum_{n=0}^{N-1} \left\{ \frac{1}{2}\left(\frac{q^{n+1} - q^n}{\Delta t}\right)^T M \left(\frac{q^{n+1} - q^n}{\Delta t}\right) - \frac{1}{2}\left(V(q^{n+1}) + V(q^n)\right) \right\} \Delta t. \tag{15}$$

Here again variations are taken with both end points $q_0 = q_a$ and $q_N = q_b$, $N = T/\Delta t$, held fixed. This puts the leapfrog method into the context of boundary value problems (BVP).

B. Modified equation analysis and conservation of energy

Conservation of a modified energy can be concluded quite easily for the leapfrog formulation

$$M\frac{q^{n+1} - 2q^n + q^{n-1}}{\Delta t^2} = -\nabla V(q^n). \tag{16}$$

We note that

$$\frac{q^{n+1} - 2q^n + q^{n-1}}{\Delta t^2} = \ddot{q}(t_n) + \frac{\Delta t^2}{12} q^{(iv)}(t_n) + \mathcal{O}(\Delta t^4) \tag{17}$$

and, hence, obtain a modified equation

$$M\ddot{q} + \frac{\Delta t^2}{12} M q^{(iv)} = -\nabla V(q) \tag{18}$$

up to terms of order Δt^4. Furthermore, multiplication by $\dot{q}^T$ yields

$$\frac{d}{dt}\left[\frac{1}{2}\dot{q}^T M\dot{q} + V(q)\right] = -\frac{\Delta t^2}{12}\dot{q}^T M q^{(iv)} \tag{19}$$

$$= -\frac{\Delta t^2}{12}\frac{d}{dt}\left[\dot{q}^T M q^{(iii)} - \frac{1}{2}\ddot{q}^T M\ddot{q}\right] \tag{20}$$

and, finally, the modified energy

$$E_{\Delta t} = \frac{1}{2}\dot{q}^T M\dot{q} + V(q) + \frac{\Delta t^2}{12}\left[\dot{q}^T M q^{(iii)} - \frac{1}{2}\ddot{q}^T M\ddot{q}\right]. \tag{21}$$

This formula can easily be generalized to higher-order approximations and can be verified along numerical trajectories by replacing time derivatives by high-order finite-difference approximations. We will come back to this observation in the third lecture on hybrid Monte Carlo methods.

A more sophisticated approach is to define a modified DE

$$\dot{z} = \tilde{f}_{\Delta t}(z) \tag{22}$$

such that

$$\Psi_{\Delta t}(z) \approx \phi_{\Delta t, \tilde{f}_{\Delta t}}(z). \tag{23}$$

In case of a symplectic method, such as the SV method, the modified DE is Hamiltonian, i.e.

$$\dot{z} = \tilde{f}_{\Delta t}(z) = J\nabla\tilde{H}_{\Delta t}(z). \tag{24}$$

In case of analytic functions, Cauchy's estimate and a bit of sweating leads to an estimate of type

$$\|\Psi_{\Delta t}(z) - \phi_{\Delta t, \tilde{f}_{\Delta t}}(z)\| \leq c_1 e^{-c_2/\Delta t}. \tag{25}$$

Since

$$\tilde{H}_{\Delta t}(z) - H(z) = \mathcal{O}(\Delta t^k), \tag{26}$$

near conservation of $H(z^n)$ over exponentially long simulation intervals follows.

We now take a step back and consider a non-autonomous modified DE defined by

$$\dot{z} = \frac{d}{dt}\Psi_t(z_0) \tag{27}$$

$$= \left[\frac{\partial}{\partial t}\Psi_t\right] \circ \Psi_t^{-1}(z) \tag{28}$$

$$= f(z) + R_{\Delta t}(z,t) =: \tilde{f}_{\Delta t}(z,t) \tag{29}$$

for $0 \le t \le \Delta t$. We clearly have by construction that

$$\phi_{\Delta t, \tilde{f}_{\Delta t}} = \Psi_{\Delta t} \tag{30}$$

(exact embedding into a flow map). By periodic extension of $\tilde{f}_{\Delta t}(x,t)$ for all $t \ge 0$, we obtain

$$\phi_{n\Delta t, \tilde{f}_{\Delta t}} = \Psi_{\Delta t}^n. \tag{31}$$

Unfortunately, $\tilde{f}_{\Delta t}(z,t)$ is not continuous in t. A more sophisticated embedding works with

$$\hat{\Psi}_t := \phi_{t-\Delta t\xi(t/\Delta t), f} \circ \Psi_{\Delta t \xi(t/\Delta t)} \tag{32}$$

and

$$\tilde{f}_{\Delta t} := \left[\frac{\partial}{\partial t}\hat{\Psi}_t\right] \circ \hat{\Psi}_t^{-1}, \tag{33}$$

where $\xi(s)$ satisfies $\xi(0) = 0$, $\xi(1) = 1$, $\xi^{(m)}(0) = \xi^{(m)}(1) = 0$ for $m \ge 1$. Note that $\hat{\Psi}_{\Delta t} = \Psi_{\Delta t}$. More work is needed to make $\tilde{f}_{\Delta t}$ analytic in t (see P.C. Moan).

An alternative embedding has been suggested for the SV method (and general splitting methods) by J. Wisdom. He uses

$$M\ddot{q} = -2\pi \sum_{m=-\infty}^{m=+\infty} \delta\left(\frac{2\pi t}{\Delta t} - 2\pi m\right) \nabla V(q) = -\sum_{m=-\infty}^{m=+\infty} \cos\left(\frac{2\pi m}{\Delta t} t\right) \nabla V(q) \tag{34}$$

with modified time-dependent Hamiltonian

$$\tilde{H}_{\Delta t} = \frac{1}{2} p^T M^{-1} p + \sum_{m=-\infty}^{m=+\infty} \cos\left(\frac{2\pi m}{\Delta t} t\right) V(q). \tag{35}$$

For a symplectic time-stepping method the non-autonomous DE is Hamiltonian, i.e.

$$\dot{z} = J\nabla \tilde{H}_{\Delta t}(z,t) = J\nabla \left(H(z) + R_{\Delta t}(z,t)\right). \tag{36}$$

The drift in energy is given by

$$\tilde{H}_{\Delta t}(z(t),t) = \tilde{H}_{\Delta t}(z(0),0) + \int_0^t \frac{\partial}{\partial s} R_{\Delta t}(z(s),s)\, ds. \tag{37}$$

Suppose now that $z(t)$ is quasi-periodic, i.e.

$$z(t) = \mathrm{Re}\left\{\sum_{k\in\mathbb{Z}^d} z_m\, e^{i(\omega\cdot k)t}\right\}, \tag{38}$$

with frequency vector $\omega \in \mathbb{R}^d$. Then

$$\int_0^t \frac{\partial}{\partial s} R_{\Delta t}(z(s), s)\, ds = \sum_{m \neq 0, k \in \mathbb{Z}^d} r_{m,k} \int_0^t e^{i(k\cdot\omega + 2\pi m/\Delta t)\, s}\, ds \tag{39}$$

$$= \sum_{m \neq 0, k \in \mathbb{Z}^d} r_{m,k} \frac{e^{i(k\cdot\omega + 2\pi m/\Delta t)\, s}}{i(k \cdot \omega + 2\pi m/\Delta t)}. \tag{40}$$

Numerically induced drift in energy can be expected for

$$k \cdot \omega + \frac{2\pi m}{\Delta t} \approx 0 \tag{41}$$

(numerical resonances). Not relevant for $\Delta t \to 0$ as the Fourier coefficients $r_{m,k}$ decay exponentially fast for analytic functions and standard backward error analysis can be recovered by averaging over time. However, non-autonomous modified equation analysis is relevant for large (multiple) time stepping methods.

C. Shadowing

Given a numerical trajectory $\{z^n\}_{n=0}^L$, can we find an exact solution $z(t)$ of the given DE such that, roughly speaking,

$$\|z(t_n) - z^n\| \leq \delta \tag{42}$$

for all $0 \leq n \leq L$? This shadowing property holds for hyperbolic systems and for Hamiltonian systems that are hyperbolic on hyper-surfaces of constant energy.

For Hamiltonian systems, shadowing can be rephrased as follows: Given a numerical trajectory $z^n = ((q^n)^T, (p^n)^T)^T$ computed with the SV method, can we find a solution of the Lagrangian variational problem

$$\mathcal{L}[q] = \int_0^T \left[\frac{1}{2}\dot{q}^T M \dot{q} - V(q)\right] dt \tag{43}$$

subject to the boundary conditions $q(0) = q^0$ and $q(T) = q^N$, $T = \Delta t\, N$? A necessary condition is that any point in coordinate space can be connected to another point by a solution of the corresponding Euler-Lagrange equation. Furthermore, convexity would imply that the boundary value problem is well-posed. Under these conditions shadowing would be possible. Positive results are available over short time intervals (see work by Müller & Ortiz, 2004).

Shadowing combined with backward error analysis can be used to show that a symplectic time-stepping method will 'shadow' trajectories of the modified equations over exponentially long time periods implying, for example, that time averages (expectation values) can be computed reliably for mixing systems (Reich, SINUM, 1999).

D. The Liouville picture

Let us now consider the noise-driven DE

$$\dot{z} = f(z) + \gamma(z)\dot{W}, \tag{44}$$

where $W(t)$ denotes standard multivariate Brownian motion.

If we consider initial data distributed according to a measure with density $\rho_0(z)$, then the density of solutions $z(t)$ satisfies the Fokker-Plank equation

$$\rho_t = \mathcal{L}^*\rho. \tag{45}$$

The generator $\mathcal{L}$ is defined by

$$\mathcal{L}\phi = f \cdot \nabla\phi + \frac{1}{2}\Gamma : \nabla(\nabla\phi) \tag{46}$$

and its dual by $\langle \mathcal{L}\phi, \rho\rangle = \langle \phi, \mathcal{L}^*\rho\rangle$. Here $\langle ., .\rangle$ is the standard L_2 inner product, $A : B = \text{trace}\,(A^T B)$ denotes the inner product on the space of $d \times d$ matrices, $\Gamma(z) = \gamma(z)\gamma(z)^T$, and $(\nabla v)_{ij} = \partial v_i/\partial z_j$.

The spectrum of $\mathcal{L}^*$ encodes all relevant dynamic information such as rate of mixing, meta-stable states, invariant density. Most theoretical results (see survey by Weinan E & Eric Vanden-Eijnden) are available for

$$\dot{q} = -\nabla V(q) + \gamma\dot{W} \tag{47}$$

while very little is known for deterministic Hamiltonian systems $\gamma(z) \equiv 0$ except for hyperbolic systems (see work by Liverani).

Maps such as $\Psi_{\Delta t}$ and $\phi_{\Delta t,H}$ also generate linear operators (called Frobenius-Perron operators), which characterize the propagtion of densities of solutions. The spectrum of such FP operators could be compared to assess the properties of exact dynamics versus numerical approximations. This is probably one of the big challenges in numerical ODE theory.

Another, entirely different, approach to molecular dynamics is to discretized phase space into 'boxes' and to computed transition probabilities between 'boxes'. This amounts effectively to discretizing the operator $\mathcal{L}^*$ and to compute the eigenvalues of the resulting truncation. This approach is not feasible, in general, due to the inherent computational complexity. Attempts have been made to find a reduced phase space over which a Markovian assumption is still valid and, hence, to reduce the computational complexity (Grubmüller, Schütte et al, Stuart, Swope et al, Pande et al, Kevrekidis).

II. OSCILLATORY SYSTEMS

'Stochastic' (exponential decay of correlation) behavior for deterministic dynamics can be shown for hyperbolic systems. This lecture will introduce another way to obtain stochastic behavior in deterministic systems. This approach is based on introducing a limit of infinitely many oscillatory degrees of freedom and a subsequent reduction to a single distinguished degrees of freedom. Most of the presentation is based on a survey article by Givon, Kupfermann & Stuart in Nonlinearity, 2004.

We consider deterministic Hamiltonian systems and introduce a splitting of the state variable z into $(x, y) \in X \times Y$ with $\dim Y \to \infty$. We are interested in situations where the solution $\rho(x, y, t)$ of the Liouville equations (Fokker-Plank with $\Gamma \equiv 0$)

$$\rho_t = \mathcal{L}^* \rho \tag{48}$$

can be approximated by

$$\rho(x, y, t) \approx \rho(x, y)\, \bar{\rho}(x, t) \tag{49}$$

for finite times $0 \leq t \leq T$, where T is large enough to study meta-stable behavior and where $\bar{\rho}$ satisfies a Fokker-Plank equations with $\Gamma \neq 0$.

A. Trigonometric approximation of Gaussian processes

Mean zero Gaussian processes $\Xi(t)$ are completely characterized by their auto-covariance function

$$R(\tau) = \mathbb{E}\left[\Xi(t)\Xi(t+\tau)\right]. \tag{50}$$

We consider the approximation of Gaussian processes by finite series of the form

$$\Xi_K(t) = \frac{1}{K^b} \sum_{j=1}^{K} F(\omega_j) \left[\xi_j \cos \omega_j t + \eta_j \sin \omega_j t\right] \tag{51}$$

where ξ_j, η_j, $j = 1, \ldots, K$, are mutually independent i.i.d. sequences with $\xi_j, \eta_j \sim \mathcal{N}(0,1)$. One finds that

$$\begin{aligned} R_K(\tau) &= \mathbb{E}\left[\Xi_K(t)\, \Xi_K(t+\tau)\right] && (52)\\ &= \frac{1}{K^{2b}} \sum_{j=1}^{K} F^2(\omega_j)\, \cos \omega_j \tau. && (53) \end{aligned}$$

Example. If $a \in (0,1)$, $2b = 1 - a$, $\omega_j = N^\alpha \zeta_j$, $\{\zeta_j\}$ is an i.i.d. sequence with $\zeta_j \sim \mathcal{U}[0,1]$, $\Delta\omega = N^a/N$, then

$$R_K(\tau) = \sum_{j=1}^{K} F^2(\omega_j)\, \cos(\omega_j \tau)\, \Delta\omega, \tag{54}$$

which, as $K \to \infty$, is a Monte Carlo approximation to the Fourier-cosine transform of $F^2(\omega)$:

$$R(\tau) = \int_0^\infty F^2(\omega)\, \cos(\omega\tau)\, d\omega. \tag{55}$$

Under suitable decay conditions on $F^2(\omega)$, $R_K(\tau)$ converges to $R(\tau)$ point-wise and in $L_1[0,T]$, $T > 0$ arbitrary. As a specific example consider

$$F^2(\omega) = \frac{2\alpha/\pi}{\alpha^2 + \omega^2}, \qquad \alpha > 0, \tag{56}$$

then

$$R(\tau) = e^{-\alpha|\tau|} \tag{57}$$

and $\Xi(t)$ becomes an OU process, i.e.

$$\dot{\Xi} = -\alpha \Xi + (2\alpha)^{1/2} \dot{W}. \tag{58}$$

□

B. Hamiltonian systems and heat baths

Consider the model problem

$$H(Q,P,q,p) = \frac{1}{2}P^2 + V(Q) + \frac{1}{2}\sum_{j=1}^{K} p_j^2/m_j + \frac{1}{2}\sum_{j=1}^{K} k_j \left(q_j - Q\right)^2, \tag{59}$$

where $Q, P, q_j, p_j \in \mathbb{R}$. The equations of motion are

$$\ddot{Q} + V'(Q) = \sum_{j=1}^{K} (q_j - Q), \tag{60}$$

$$\ddot{q}_j + \omega_j^2 (q_j - Q) = 0, \tag{61}$$

where $\omega_j^2 = k_j/m_j$. Here (Q_0, P_0) are given while the initial conditions (q_j^0, p_j^0) are drawn from a distribution with density proportional to $\exp(-\beta H)$, i.e.

$$q_j^0 = Q_0 + (1/\beta k_j)^{1/2} \xi_j, \quad p_j^0 = (m_j/\beta)^{1/2} \eta_j. \tag{62}$$

Equation (61) can be integrated out and we obtain an integro-differential equation for Q:

$$\ddot{Q} + V'(Q) + \int_0^t R_K(t-s)\, \dot{Q}(s)\, ds = \beta^{-1/2} \Xi_K(t), \tag{63}$$

where

$$R_K(\tau) = \sum_{j=1}^{K} k_j \cos \omega_j \tau \tag{64}$$

and

$$\Xi_K(t) = \sum_{j=1}^{K} k_j^{1/2} \left[\xi \cos \omega_j t + \eta_j \sin \omega_j t \right]. \tag{65}$$

Under appropriate conditions (63) converges weakly on any bounded time interval to the stochastic integro-differential equation

$$\ddot{Q} + V'(Q) + \int_0^t R(t-s)\, \dot{Q}(s)\, ds = \beta^{-1/2} \Xi(t). \tag{66}$$

If $R(\tau) = e^{-\alpha|\tau|}$, then

$$\ddot{Q} + V'(Q) = s, \tag{67}$$

$$\dot{s} + \alpha\, s = -P + (2\alpha/\beta)^{1/2} \dot{W}. \tag{68}$$

The analysis becomes much nicer if a single particle is coupled to a Hamiltonian wave equation. See work by Eckmann et al. Cotter has applied this theory to particle methods coupled to Hamiltonian wave equations to derive sub-grid (turbulence) models.

C. Multiple time stepping methods and numerical resonances

Heat bath models and indeed molecular dynamics introduce nearly harmonic oscillations of very high frequency compared to the time-scales of interest. How can one design efficient numerical time-stepping methods for such systems?

The basic idea of multiple-time stepping is to integrate the fast degrees of motion (e.g., equation (61) with fixed Q) with a small inner time step $\Delta\tau$, while the remaining equations are solved with a larger outer time-step $\Delta t = K \cdot \Delta\tau$. In principle, one can take $K \to \infty$ and $\Delta\tau \to 0$, i.e., keep Δt fixed. Then a time-stepping method $\Psi_{\Delta t}$ is obtained for which $\omega_j \Delta t$ is not necessarily small. Autonomous backward error analysis breaks down in this context and one has to use an embedding into a non-autonomous Hamiltonian system (provided the method $\Psi_{\Delta t}$ is symplectic. However, following (37), we now have to struggle with numerically induced resonances for any pair $k \in \mathbb{Z}^d$ and $m \in \mathbb{Z}$, $m \neq 0$ for which

$$k \cdot (\omega \Delta t) + 2\pi m \approx 0. \tag{69}$$

unless the corresponding Fourier coefficient $r_{m,k}$ simultaneously vanishes!

D. Mollified methods and regularized equations

Skeel and co-workers have developed two methods called MOLLY and EQUILIBRIUM that eliminates the troublesome numerical drift introduced by near resonances of the form (69). The idea is to average the slow forces along fast oscillations, i.e., roughly speaking, a slow potential energy contribution $V_{\rm slow}(q)$ is replaced by $V_{\rm slow}(\mathcal{A}(q))$, where $\mathcal{A}$ is a mollifier that maps the instantaneous value of $q(t)$ to its locally time-averaged value $\langle q \rangle(t)$.

The mollifier $\mathcal{A}(q)$ may be viewed as a form of regularization. Similar types of regularization appear in computational fluid dynamics (CFD) and numerical weather prediction (NWP). However, large time step methods for NWP most often use the implicit midpoint rule to approximate fast waves/oscillations. This effectively results in a slowing down of oscillations with frequency ω to a numerical value $\Omega = (\Delta t)^{-1} \tan^{-1}(\omega \Delta t)$ and the most severe numerical resonances are avoided since $\Omega \Delta t \leq \pi$. See Wood, Staniforth & Reich for more results.

For CFD and NWP these methods are a huge success story. The speed-up in the context of MD is much less spectacular (perhaps a factor of 10). Implicit time-stepping methods are not in use.

See also the work by Hairer, Lubich & Hochbruck on exponential integrators and filtering.

III. HYBRID MONTE CARLO METHODS

A. Generalized hybrid Monte Carlo (GHMC) algorithm

We describe the generalized hybrid Monte Carlo (GHMC) algorithm of Kennedy & Pendleton for a molecular Hamiltonian (energy function)

$$H(q,p) = \frac{1}{2}p^T M^{-1} p + V(q). \tag{70}$$

We begin by recalling that a Markov process will converge to some distribution of configurations if it is constructed out of updates each of which has the desired distribution as a fixed point, and which taken together are ergodic. The GHMC algorithm for the generation of the canonical density function

$$\rho_{\rm can}(q,p) = \frac{1}{Z}\exp(-\beta H(q,p)), \quad \text{with} \quad \beta = 1/k_B T,\ Z = \int \exp(-\beta H(q,p))\, dqdp, \tag{71}$$

is constructed of two such steps.

1. *Molecular dynamics Monte Carlo (MDMC)*

This in turn consists of three parts:

(i) *Molecular dynamics* (MD): an approximate integration of Hamilton's equations of motion

$$\dot{q} = M^{-1}p, \qquad \dot{p} = F(q), \qquad F(q) := -\nabla_q V(q), \tag{72}$$

with the leapfrog/Störmer-Verlet method over L steps and step-size Δt. The resulting map $U_\tau : (q,p) \to (q',p')$, $\tau = L\Delta t$, is exactly area preserving and time-reversible.

(ii) A *momentum flip* $\mathcal{F} : (q,p) \to (q,-p)$.

(iii) *Monte Carlo* (MC): a Metropolis accept/reject test

$$(q',p') = \begin{cases} \mathcal{F}\cdot U_\tau(q,p) & \text{with probability } \min(1,\exp(-\beta\,\delta H)) \\ (q,p) & \text{otherwise} \end{cases}, \tag{73}$$

with

$$\delta H := H(U_\tau(q,p)) - H(q,p) = H(\mathcal{F}\cdot U_\tau(q,p)) - H(q,p). \tag{74}$$

Molecular dynamics Monte Carlo (MDMC) satisfies detailed balance since $(\mathcal{F}\cdot U_\tau)^2 = \mathrm{id}$ and U_τ is volume conserving.

2. Partial momentum refreshment

We first apply an extra momentum flip $\mathcal{F}$ so that the trajectory is reversed upon an MC rejection instead of on an acceptance. For high rejection rates this momentum flip leads to an undesirable *Zitterbewegung* (going forward and backward) in the molecular trajectories.

The momenta p are now mixed with a normal (Gaussian) i.i.d. distributed noise vector $\Xi \in \mathbb{R}^{3N}$ and the complete partial momentum refreshment step is given by

$$\begin{pmatrix} p' \\ \Xi' \end{pmatrix} = \begin{pmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{pmatrix} \cdot \mathcal{F} \begin{pmatrix} p \\ \Xi \end{pmatrix} \tag{75}$$

where

$$\Xi = \beta^{-1/2} M^{1/2} \xi, \qquad \xi_i \sim \mathrm{N}(0,1),\ i = 1, \ldots, 3N \tag{76}$$

and $0 \leq \phi \leq \pi/2$. Here $\mathrm{N}(0,1)$ denotes the normal distribution with zero mean and unit variance.

If p and Ξ are both distributed according to the same normal (Gaussian) distribution, then so are p' and Ξ'. This special property of Gaussian random variables under an orthogonal transformation (75) makes it possible to conduct the partial momentum refreshment step without a Metropolis accept/reject test.

3. Special cases of GHMC

Several well-known algorithms are special cases of GHMC:

- The usual hybrid Monte Carlo (HMC) algorithm is the special case where $\phi = \pi/2$. The momentum flips may be ignored in this case since $p' = \Xi$ in (75) and the previous value of p is entirely discarded.

- The choice $\phi = 0$ corresponds to constant energy molecular dynamics under the assumption that the propagator U_τ conserves energy exactly.

- Langevin Monte Carlo algorithms correspond to $L = 1$; i.e., a single MD time-step with $\tau = \Delta t$ and an arbitrary $0 < \phi \leq \pi/2$. The single step ($L = 1$) may be replaced by a small number of MD steps ($L = 10, \ldots, 100$). Langevin Monte Carlo recovers stochastic Langevin molecular dynamics

$$\dot{q} = M^{-1} p, \qquad \dot{p} = F(q) - \gamma p + \sigma \dot{W}, \tag{77}$$

provided $\phi = (2\gamma\Delta t)^{1/2}$, $\gamma > 0$ a constant and σ is determined by the standard fluctuation-dissipation relation. In this regime, we find that (75) without the momentum flip $\mathcal{F}$ reduces to

$$p' \approx (1 - \gamma\Delta t)\, p + (2\gamma\Delta t)^{1/2}\Xi \tag{78}$$

and one may view the GHMC algorithm as a mean to simulate *stochastic molecular dynamics* (instead of using GHMC as a pure *sampling* device).

4. Applications of hybrid Monte Carlo methods

A main application of HMC algorithms is provided by free energy calculations. It can be used to implement blue-moon sampling as well as metadynamics (see survey by E & Vanden-Eijnden).

B. Targeted shadow hybrid Monte Carlo (TSHMC) algorithm

A high acceptance rate is a desirable property of any Monte Carlo scheme. In fact, one of the reasons for the introduction of the HMC method was its vastly superior acceptance rate over standard Monte Carlo methods. However the acceptance rate of HMC degrades with the size of the simulated molecular system. Furthermore, in light of the modified Hamiltonian, it would appear that essentially no rejections are necessary at all for a symplectic integration method such as Störmer-Verlet. In fact, that is indeed the case up to a small rejection rate caused by the truncation of a modified Hamiltonian $\tilde{H}_{\Delta t}$ after a finite number of terms. A practical HMC algorithm based on modified Hamiltonians was first proposed by Izaguirre & Hampton. Akhmatskaya & Reich proposed a variant of their SHMC method with two important modifications:

(i) a simplified evaluation of the modified energy (Hamiltonian),

(ii) a modified and more flexible momentum update.

The resulting TSHMC method avoids, in particular, the introduction of a factor c in the modified energy used for SHCM, which is difficult to tune.

1. Molecular dynamics step and modified energies

The probability of a change in energy $\beta\,\delta H$ having the value ξ when averaged over the equilibrium distribution of starting points on phase space is

$$P_{\delta H}(\xi) = \frac{1}{Z}\int \exp(-\beta\, H)\,\delta(\xi - \beta\,\delta H)\,dqdp \tag{79}$$

$$\approx \frac{1}{4\pi\beta\langle\delta H\rangle}\exp\left[-\frac{(\xi - \beta\langle\delta H\rangle)^2}{4\beta\langle\delta H\rangle}\right] \tag{80}$$

as a expected from the central limit theorem. The average acceptance rate for the MDMC part of the GHMC method is given by

$$P_{\text{acc}} = \langle\min(1, \exp(-\beta\,\delta H))\rangle \tag{81}$$

$$= \int_{-\infty}^{+\infty} P_{\delta H}(\xi)\ \min(1, \exp(-\xi)\,d\xi \tag{82}$$

$$= \text{erfc}\left(\frac{1}{2}\sqrt{\beta\langle\delta H\rangle}\right) \tag{83}$$

where $\langle\delta H\rangle$ denotes the expectation value of energy fluctuations in the Hamiltonian H (in case of the standard hybrid Monte Carlo method) or the modified Hamiltonian $\mathcal{H}_{\Delta t}$ in case of SHMC/TSHMC. It should be kept in mind that δH denotes fluctuations in the energy after L MD steps with step-size Δt. Theoretical results for a chain of harmonic oscillators and numerical evidence suggest that

$$\langle\delta H\rangle = \mathcal{O}(V\Delta t^4) \tag{84}$$

for the Störmer-Verlet method and

$$\langle\delta H\rangle = \mathcal{O}(V\Delta t^{2k}) \tag{85}$$

for a k-th order modified Hamiltonian $\tilde{H}_{\Delta t}$, where V is the volume of the simulation domain.

It follows that the efficiency gain due to a higher acceptance rate in the molecular dynamics Monte Carlo step with fixed molecular dynamics trajectory length $\tau = L\Delta t = const.$ is proportional to $V^{1/4}$ for modified Hamiltonians $\mathcal{H}_{\Delta t}$ of sufficiently high order.

Results by Takaishi furthermore imply that the optimal acceptance rate for a k-th order method is given by

$$\langle P_{\text{acc}}\rangle_{\text{opt}} = \exp(-1/k). \tag{86}$$

The following fourth-order modified energy (Hamiltonian) for the leapfrog/Störmer-Verlet method has been given by Skeel & Hardy:

$$E_{\Delta t} = H(q^n, p^n) + \frac{1}{4}\delta^2 U^n + \frac{\Delta t}{6}(p^n)^T M^{-1}\mu\delta F^n + \frac{5\Delta t^2}{24}(F^n)^T M^{-1} F^n + \frac{\Delta t^2}{12}(F^n)^T M^{-1}\delta^2 F^n. \tag{87}$$

Here a superscript n denotes evaluation at q^n, the centered difference operator is defined by $\delta w^n = w^{n+1/2} - w^{n-1/2}$, the averaging operator is defined by $\mu w^n = \frac{1}{2}w^{n+1/2} + \frac{1}{2}w^{n-1/2}$, and values $q^{n\pm 1}$ are defined in terms of (q^n, p^n) by the leapfrog/Störmer-Verlet method.

Another option is to numerical evaluate the modified energy (21). For the SV method this is probably the most straightforward way to obtain modified energies. But the approach does not in an obvious manner extend to other symplectic methods.

The standard Metropolis acceptance/rejection criterion for the MDMC part of the GHMC algorithm is now replaced by

$$\min\left(1, \exp\left[-\beta\{E_{\Delta t}(q', p') - E_{\Delta t}(q, p)\}\right]\right). \tag{88}$$

Note that time averages need to include the factor

$$w^m = e^{\beta(\hat{E}^m_{\Delta t} - E^m)}, \tag{89}$$

where E^m is the value of the given energy after completion of the m^{th} SHMC/TSHMC step and $E^m_{\Delta t}$ is the modified energy, respectively, i.e., averages of an observable Ω are computed according to the formula:

$$\langle\Omega\rangle = \frac{\frac{1}{M}\sum_{m=1}^{M}\Omega(q^m, p^m)\, w^m}{\frac{1}{M}\sum_{m=1}^{M} w^m}. \tag{90}$$

This is a standard re-weighting procedure for simulations in modified ensembles.

To avoid the *Zitterbewegung* caused by the momentum reversal necessary for rejected MD steps, it is desirable to pick a step-size Δt such that the rejection rate after L MD steps is kept sufficiently small (e.g. below 10%).

2. Alternative momentum updates

We already discussed that it is not necessary to completely re-sample the momenta as done in the standard HMC scheme. Instead one may take the set of given momenta p and modifies it by a vector Ξ to obtain a new set given by

$$p' = p + \sigma\,\Xi. \tag{91}$$

Here $\sigma > 0$ is a free parameter and Ξ is sampled from a Boltzmann distribution, i.e., Ξ is a vector of independent Gaussian random variables with mean zero and variance $k_B T$. Smaller values of σ lead to smaller perturbations in the momenta.

Following work by Cotter & Reich on dissipative particle dynamics (DPD), the following more general update has been suggested: (91):

$$p' = p + \sigma \sum_{k=1}^{K} \nabla h_k(q)\, \Xi_k, \tag{92}$$

where σ and $\Xi = (\Xi_1, \ldots, \Xi_K)^T$ are defined as before, and the functions $h_k(q)$, $k = 1, \ldots, K$, can be chosen quite arbitrarily. The particular choice

$$h_k(q) = \phi(r_{ij}), \qquad r_{ij} = \|q_i - q_j\|, \tag{93}$$

$k = 1, \ldots, (N-1)N/2$, ϕ a given function of inter-particle distances r_{ij}, transforms (92) into an update very similar to what is used in dissipative particle dynamics (DPD). An attractive feature of such an update is its conservation of linear and angular momenta:

$$\sum_{i=1}^{N} p_i = \sum_{i=1}^{N} p'_i, \qquad \sum_{i=1}^{N} q_i \times p_i = \sum_{i=1}^{N} q'_i \times p'_i. \tag{94}$$

Given a new set of momenta p', we need to evaluate the corresponding modified energy $E_{\Delta t}(q, p')$. This step requires time-stepping the equations of motion two steps forward and backward in time and, hence, two additional force field evaluations are needed. We then apply (88) in its slightly modified form:

$$\min\left(1, \exp\left[-\beta\{E_{\Delta t}(q, p') - E_{\Delta t}(q, p)\}\right]\right). \tag{95}$$

It is again easily verified that the momentum update (92) combined with the Metropolis criterion (95) satisfies detailed balance and preserves the canonical density corresponding to $\hat{E}_{\Delta t}$. Hence we may conclude that the TSHMC method (without re-weighting) constitutes a Markov chain Monte Carlo method which samples from the canonical density $\rho_{\text{can}} \sim \exp(-\beta \hat{E}_{\Delta t})$.

C. Sampling aspects

In this section, we look at means to assess the *sampling* efficiency of Monte Carlo methods.

Let $(q^1, q^1, q^2, \ldots, q^M)$ be a sequence of molecular configurations generated by an equilibrated Monte Carlo simulation, and let $\langle \Omega(q) \rangle$ denote the expectation value of some 'observable' Ω for q distributed according to the canonical ensemble. For simplicity we assume that $\langle \Omega(q) \rangle = 0$. We may define an unbiased estimator $\bar{\Omega}$ over the finite sequence of configurations by

$$\bar{\Omega} := \frac{1}{M} \sum_{m=1}^{M} \Omega(q^m), \tag{96}$$

so $\langle \bar{\Omega} \rangle = \langle \Omega \rangle = 0$.

Let

$$C_\Omega(l) := \frac{\langle \Omega(q^{1+l})\, \Omega(q^1) \rangle}{\langle (\Omega(q^1)^2 \rangle}$$

denote the autocorrelation function for Ω. If the Markov process is ergodic, then for large l,

$$|C_\Omega(l)| \leq \lambda_{\max}^l := e^{l/M_{\exp}},$$

where $\lambda_{\max}$ is the second-largest eigenvalue of the Markov matrix and $M_{\exp}$ is the exponential autocorrelation time. If $M \gg M_{\exp}$ then

$$\langle (\bar{\Omega})^2 \rangle = \{1 + 2A_\Omega\} \frac{\langle (\Omega)^2 \rangle}{M} \left[1 + \mathcal{O}(M_{\exp}/M)\right], \tag{97}$$

where

$$A_\Omega := \sum_{l=1}^{\infty} C_\Omega(l) \tag{98}$$

is the integrated autocorrelation function of the operator Ω.

This result tells us that on average $1+2A_\Omega$ correlated measurements are needed to reduce the variance by the same amount as a single truly independent measurement. The quantity A_Ω provides a good measurement for the efficiency of a Monte Carlo algorithm (assuming nearly identical work to generate the sequence $\{q^m\}_{m=1}^M$ of molecular conformations).

Work is underway to improve the quality of the momentum refreshment step for TSHMC.

Method	numerical $\langle V_{\text{tors}} \rangle$	theoretical $\langle V_{\text{tors}} \rangle$	$\langle A_\Omega \rangle$	AR
HMC	2.6322 ± 0.0575	2.6313	67.7950 ± 11.1430	78%
SHMC	2.6370 ± 0.0587	2.6313	56.3774 ± 10.6648	98%
TSHMC	2.6336 ± 0.0156	2.6313	46.0230 ± 7.4815	99%

TABLE I: Simulation results for a single butane molecule with $\Delta t = 6$ fs and $\tau = 600$ fs. The modified Hamiltonian is fourth order.