

Hybrid Monte Carlo

Dieter W. Heermann

Monte Carlo Methods

2015

- In conventional Monte-Carlo (MC) calculations of condensed matter systems, such as an N -particle system with a Hamiltonian $\mathcal{H} = \mathcal{U}$, only local moves (displacement of a single particle) are made.
- Updating more than one particle typically results in a prohibitively low average acceptance probability $\langle P_A \rangle$.
- This implies large relaxation times and high autocorrelations especially for macromolecular systems.
- In a Molecular Dynamics (MD) simulation, with $\mathcal{H} = \mathcal{T} + \mathcal{U}$, on the other hand, global moves are made.
- The MD scheme, however, is prone to errors and instabilities due to the finite step size in time.
- In order to introduce temperature in the microcanonical context, isokinetic MD schemes are often used.
- However, they do not yield the canonical probability distribution, unlike Monte-Carlo calculations.

- The Hybrid Monte-Carlo (HMC) method combines the advantages of Molecular Dynamics and Monte-Carlo methods:
 - it allows for global moves (which essentially consist in integrating the system through *phase* space);
 - HMC is an exact method, i.e., the ensemble averages do not depend on the step size chosen;
 - algorithms derived from the method do not suffer from numerical instabilities due to finite step size as MD algorithms do;
 - and temperature is incorporated in the correct statistical mechanical sense.

- In the HMC scheme global moves can be made while keeping the average acceptance probability $\langle P_A \rangle$ high.
- One global move in *configuration* space consists in integrating the system through *phase* space for a fixed time t using some discretization scheme (δt denotes the step size)

$$\begin{aligned}
 g^{\delta t} : \mathbf{R}^{6N} &\longrightarrow \mathbf{R}^{6N} \\
 (x, p) &\longrightarrow g^{\delta t}(x, p) =: (x', p')
 \end{aligned}$$

of Hamilton's equations

$$\begin{aligned}
 \frac{dx}{dt} &= \frac{\partial \mathcal{H}}{\partial p} \\
 \frac{dp}{dt} &= -\frac{\partial \mathcal{H}}{\partial x}.
 \end{aligned} \tag{1}$$

- Since the system is moved deterministically through phase space, the conditional probability of suggesting configuration x' starting at x is given by

$$p_C(x \rightarrow x') dx' = p_C(p) dp. \quad (2)$$

- The initial momenta are drawn from a Gaussian distribution at inverse temperature β :

$$p_C(p) \propto e^{-\beta \sum_{j=1}^N \frac{p_j^2}{2m}}. \quad (3)$$

- Thus

$$P_A((x, p) \rightarrow g^{\delta t}(x, p)) = \min\{1, e^{-\beta\delta\mathcal{H}}\}, \quad (4)$$

where

$$\delta\mathcal{H} = \mathcal{H}(g^{\delta t}(x, p)) - \mathcal{H}(x, p)$$

is the discretization error associated with $g^{\delta t}$. Using the algebraic identity

$$e^{-\mathcal{H}(x,p)} \min\{1, e^{-\delta\mathcal{H}}\} = e^{-\mathcal{H}(g^{\delta t}(x,p))} \min\{e^{\delta\mathcal{H}}, 1\} \quad (5)$$

- it can be shown that for a discretization scheme which is *time-reversible*

$$g^{-\delta t} \circ g^{\delta t} = id \quad (6)$$

- and *area-preserving*

$$\det \frac{\partial g^{\delta t}(x, p)}{\partial (x, p)} = 1, \quad (7)$$

detailed balance is satisfied:

$$\begin{aligned} p(x)p_M(x \rightarrow x') dx dp &= p(x)p_C(p)P_A((x, p) \rightarrow g^{\delta t}(x, p)) dx dp \\ &= p(x')p_C(p')P_A(g^{\delta t}(x, p) \rightarrow (x, p)) dx dp \\ &= p(x')p_C(p')P_A((x', p') \rightarrow g^{-\delta t}(x', p')) dx dp \\ &= p(x')p_C(p')P_A((x', p') \rightarrow g^{-\delta t}(x', p')) dx' dp' \\ &= p(x')p_M(x' \rightarrow x) dx' dp'. \end{aligned}$$

- Thus, provided the discretization scheme used is *time-reversible* and *area-preserving*, the HMC algorithm generates a Markov chain with the stationary probability distribution $p(x)$.

- The probability distribution is entirely determined by the detailed balance condition.
- Therefore neither $p(x)$ nor any ensemble averages depend on the step size δt chosen.
- However, the average acceptance probability $\langle P_A \rangle$, because of (4), depends on the average discretization error $\langle \delta \mathcal{H} \rangle$ and hence does depend on δt .
- It can be shown that for $(\varrho, T) \neq (\varrho_c, T_c)$

$$\langle P_A \rangle = \operatorname{erfc}\left(\frac{1}{2}\sqrt{\beta\langle\delta\mathcal{H}\rangle}\right)$$

is a good approximation for sufficiently large systems ($N \rightarrow \infty$) and small step sizes ($\delta t \rightarrow 0$).

- From normalization and the area-preserving property one has

$$\langle e^{-\beta\delta\mathcal{H}} \rangle = 1. \quad (8)$$

- Equation (8) can be expanded into cumulants

$$\langle \delta\mathcal{H} \rangle = \frac{\beta}{2} \langle (\delta\mathcal{H} - \langle \delta\mathcal{H} \rangle)^2 \rangle + \dots$$

- In order to obtain a nonzero average acceptance probability $\langle P_A \rangle$ in the limit $N \rightarrow \infty$ one has to let $\delta t \rightarrow 0$, keeping $\langle (\delta\mathcal{H} - \langle \delta\mathcal{H} \rangle)^2 \rangle$ fixed.

- In this limit higher-order cumulants will vanish. The resulting distribution of the discretization error will thus be gaussian with mean and width related through

$$\langle \delta \mathcal{H} \rangle = \frac{\beta}{2} \langle (\delta \mathcal{H} - \langle \delta \mathcal{H} \rangle)^2 \rangle. \quad (9)$$

- From (4) and (9) one has in this case

$$\begin{aligned} \langle P_A \rangle &= \frac{1}{\sqrt{2\pi \langle (\delta \mathcal{H} - \langle \delta \mathcal{H} \rangle)^2 \rangle}} \int_{-\infty}^{\infty} dt \min\{1, e^{-\beta t}\} e^{-\frac{(t - \langle \delta \mathcal{H} \rangle)^2}{2 \langle (\delta \mathcal{H} - \langle \delta \mathcal{H} \rangle)^2 \rangle}} \\ &= \operatorname{erfc}\left(\frac{1}{2} \sqrt{\beta \langle \delta \mathcal{H} \rangle}\right). \end{aligned} \quad (10)$$

- The square root in (10) is always well defined since (8) implies

$$\langle \delta \mathcal{H} \rangle \geq 0.$$

- Equality holds in the limit $\delta t \rightarrow 0$, where energy is conserved exactly and $\langle P_A \rangle = 1$.
- Increasing the step size will result in a lower average acceptance probability $\langle P_A \rangle$. Varying δt , the average acceptance probability $\langle P_A \rangle$ can thus be adjusted to minimize autocorrelations.
- The momenta do not necessarily have to be drawn from the Gaussian distribution.

- A particularly simple and computationally efficient alternative to would be a uniform momentum distribution.
- This choice, however, did not prove successful, since a cut-off has to be introduced for computational reasons. This cut-off must be taken into account in P_A , leading to a very low average acceptance probability $\langle P_A \rangle$.
- It is clear that instead of choosing a discretization scheme of Hamilton's equations (1) any time-reversible and area-preserving discrete mapping can be used to propagate the system through phase space.