Separable Shadow Hybrid Monte Carlo: An efficient propagator for macromolecules.

Chris R. Sweet

Computer Science and Engineering

University of Notre Dame

In collaboration with Bob Skeel (Perdue), Scott Hampton and Jesus Izaguirre (Notre Dame)

University of Notre Dame
Computer Science and Engineering

Cambridge University. July 2007

The sampling problem

- Molecular dynamics (MD) is an important sampling method for biomolecules as it can easily generate long trajectories in phase-space.
- MD methods introduce a bias which depends on the integration time-step, even when using symplectic methods such as Verlet.
- Markov chain Monte Carlo methods are rigorous sampling methods but generating successive position sets is a challenge for large molecules.
- Hybrid Monte Carlo (HMC) combines these methods to give rigorous sampling and good exploration of phase-space.
- HMC scales poorly with system size.

Hybrid Monte Carlo (HMC) method

Given a separable probability density function, for example a Hamiltonian model sampling from the canonical ensemble

$$\begin{aligned}\rho(X,P) &\propto \exp(-\beta\mathcal{H}) = \exp(-\beta\mathcal{T})\exp(-\beta\mathcal{U}) \\ &\propto \rho_x(X)\rho_p(P)\end{aligned}$$

With mapping $(X',P') = \Psi(X,P)$, the HMC method is:

- MG step: Generate P from ρ_p
- MD step: (a) mapping $(X',P') = \Psi(X,P)$
  (b) Accept with probability
  $\min\{1, \rho(X',P')/\rho(X,P)\}$
  (c) If rejected chose X

The acceptance probability depends on $\sigma^2_{\mathcal{H}} \propto N$.

Shadow Hamiltonians

- Since the acceptance ratio of HMC depends on N it scales poorly with system size, limiting it use for Biomolecules.
- For symplectic methods it can be shown that the MD numerical results are the exact solution for a 'Shadow Hamiltonian'.
- The value of the Shadow Hamiltonian is constant for all phase-space variables and hence the probability of acceptance would be 1.
- The generation of P is now from a probability density which we have no expression for.
- Need to calculate the Shadow Hamiltonian easily.

Shadow momenta/Hamiltonian generation

- By a simple modification to the Hamiltonian the equation $\mathcal{H}_{[2k]}(x(t)) = \frac{1}{2}\dot{x}(t)^T \mathbf{J} x(t)$ can be used to calculate the Shadow Hamiltonian. The $\dot{x}(t)$ term can be calculated using an interpolative scheme to the desired order of accuracy.
- To generate a set of momenta from the probability density function from the Shadow phase-space the following scheme is used:
  Define the new Hamiltonian $\tilde{\mathcal{H}} = \min\{\mathcal{H}, \mathcal{H}_{[2k]} - c\}$
  (a) Generate P from ρ_P
  (b) Accept with probability

$$\min\{1, \exp(-\beta(\mathcal{H}_{[2k]} - c - \mathcal{H}))\}$$

  (c) Repeat until P accepted

Shadow Hybrid Monte Carlo (SHMC)

The Shadow hybrid Monte Carlo method can be defined as:

- MG step: Generate P from $\tilde{\rho}(X,P) \propto \exp(-\beta\tilde{\mathcal{H}})$
- MD step: (a) mapping $(X', P') = \Psi(X,P)$

  (b) Accept with probability

  $$\min\{1, \tilde{\rho}(X',P')/\tilde{\rho}(X,P)\}$$

  (c) If rejected chose X
- Since $\tilde{\mathcal{H}}$ is nearly conserved by the numerical method the acceptance at the MD stage will be very high.
- The acceptance rate of the MG step will be dependent on the constant c.
- Any observable A must be reweighted to obtain the correct canonical distribution so that

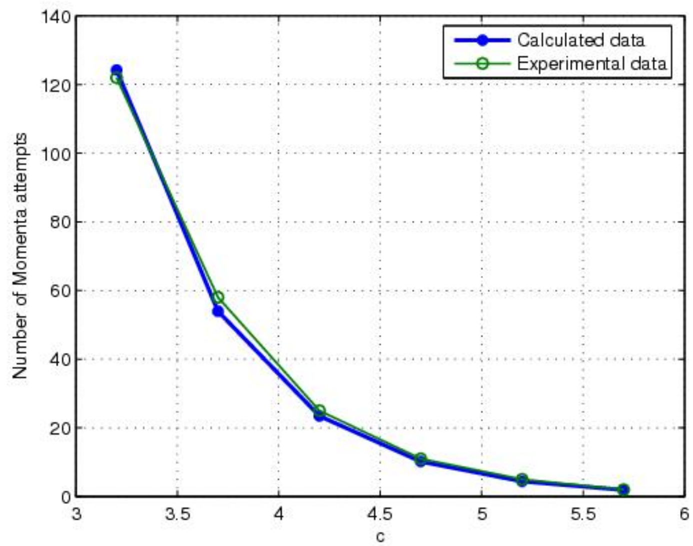
  $$\langle A \rangle = \sum_i \omega_i A_i / \sum_i \omega_i, \;\; \omega_i = \exp(-\beta(\mathcal{H} - \tilde{\mathcal{H}}))$$

Selection of 'c' parameter for MG stage

The acceptance ratio R_{MG} at the MG stage will increase for large *c* and we can show

$$R_{MG} \approx \exp\left(\frac{1}{2}\beta(2c + \beta\sigma^2_{\Delta\mathcal{H}} - 2\mu_{\Delta\mathcal{H}})\right)$$

This will be a significant factor as it takes several MD steps to calculate the Shadow Hamiltonian.



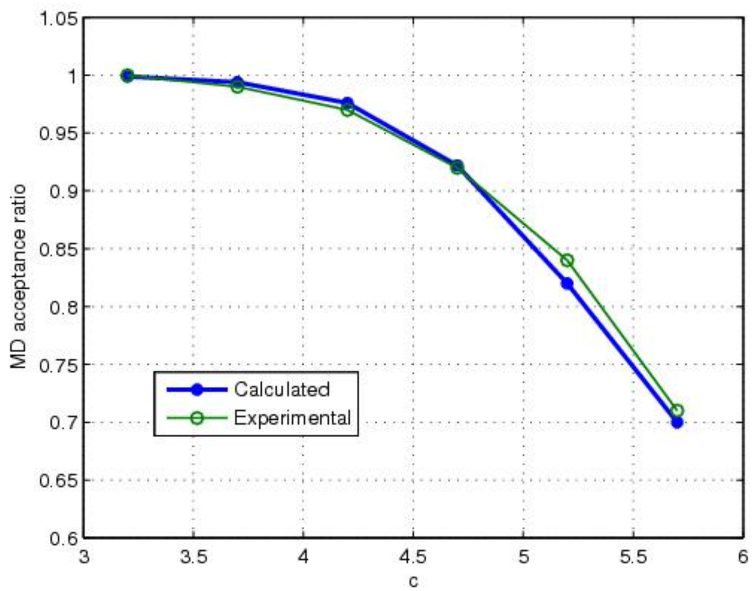
Selection of 'c' parameter for MD stage

The acceptance ratio R_{MD} at the MG stage will degrease for large *c* and we can show

$$R_{MD} = (1 - S_H) + 0.5 S_H^2$$

Where

$$S_H = \frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(\frac{\mu_{\Delta\mathcal{G}} - c}{\sqrt{2}\sigma_{\Delta\mathcal{G}}}\right)$$



Leapfrog Method Shadow Hamiltonian

The ‘truncated’ shadow Hamiltonian for the Leapfrog method is

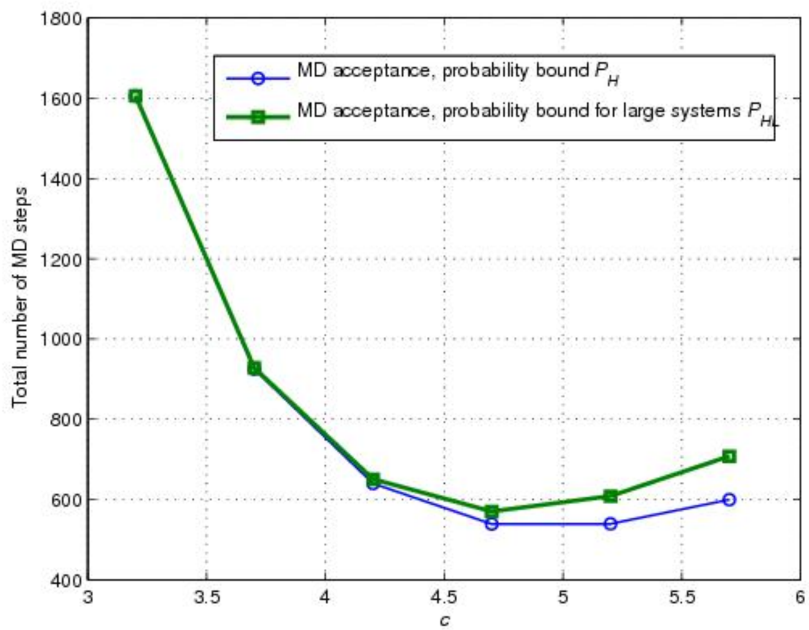
TexPoint Display

Conclusions

- SHMC can be used for efficient sampling of biomolecules without introducing the bias associated with MD.
- SHMC can be used for systems which are orders of magnitude larger than HMC can achieve, typically over 20,000 atoms.
- The *c* parameter introduced by SHMC can be calculated explicitly with data from some additional MG stages.
- The increase in variance observed when re-weighting observables from SHMC can be controlled easily by selecting the step-size, a fourth order relationship for Verlet.

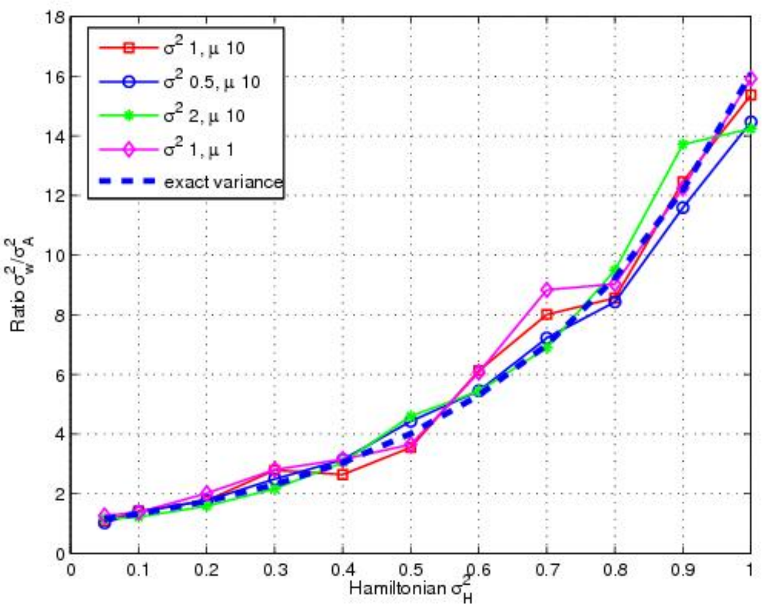
Optimum 'c' parameter

Since the choice of *c* is a compromise we can find an optimum depending on the number of MD steps n_{MD} and MG steps n_{MG} and the variance of the Hamiltonian $\sigma^2_{\Delta\mathcal{H}}$ to give the total steps $n_s = \frac{n_{MG}}{R_{MG}} + \frac{n_{MD}}{R_{MD}}$



Variance of observables for SHMC

The reweighting stage of the SHMC method can be shown to increase the variance of the observable σ_ω^2 according to $\sigma_\omega^2 = \exp(\beta^2\sigma_{\mathcal{H}}^2)\sigma_A^2$ where σ_A^2 is the original variance and $\sigma_{\mathcal{H}}^2$ the variance of the Hamiltonian as seen in the figure below.



We can control $\sigma_{\mathcal{H}}^2$ by changing the step-size according to

$$\Delta t_T = \Delta t_H \sqrt{\frac{\sigma_T}{\sigma_{\mathcal{H}}}}$$

SHMC compared to HMC for large N

We can now compare the efficiency of the SHMC and HMC methods for different *N*, assuming $\sigma^2_{\Delta\mathcal{H}} \propto N$, compared to the Hamiltonian variance for a box of 216 water molecules.