Theoretical and numerical comparison of some sampling methods in Molecular Dynamics.

Eric CANCES[1], Frédéric LEGOLL[2], Gabriel STOLTZ[1,3]

[1] CERMICS, ENPC (Marne-la-Vallée, France)

[2] LAMI, ENPC (Marne-la-Vallée, France)

[3] CEA/DAM (Bruyères-le-Châtel, France)

http://cermics.enpc.fr/∼stoltz/

References

- E. Cancès, F. Legoll, and G. Stoltz, Theoretical and numerical comparison of some sampling methods for molecular dynamics, IMA Preprint, **2040** (2005)

- Presentation and preprints available at the URL `http://cermics.enpc.fr/~stoltz/`

What we want to do

- Microscopic description of a system of M particles
  $(q, p) = (q_1, \ldots, q_M,\ p_1, \ldots, p_M)$
- Energy $H(q,p) = \sum_{i=1}^{M} \frac{p_i^2}{2m_i} + V(q_1, \ldots, q_M)$.
- Computation of (static) equilibrium properties of a system

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

where μ is the canonical probability measure (NVT)

$$d\mu(q,p) = Z^{-1} \exp(-\beta H(q,p))\, dqdp,$$

with $\beta = 1/k_{\mathrm{B}}T$.

- Manifold $\mathcal{M}$ ($= \mathbb{T}^{3N}$ if PBC for example).

Some observables

- Pressure

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

- (Kinetic) Temperature

$$A(q,p) = \frac{1}{3Mk_{\mathrm{B}}}\sum_{i=1}^{M}\frac{p_i^2}{m_i};$$

- Specific heat capacity

$$C_v = \frac{\mathcal{N}_a}{Mk_BT^2}(\langle H^2\rangle - \langle H\rangle^2).$$

Why is it difficult?

- Real issue = sampling the configurational part of the measure = high dimensional problem!

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

- Approximation of $< A >$ of the form

$$\frac{1}{N}\sum_{n=0}^{N-1} A(q^n, p^n) \simeq \int_{\mathcal{M}} A(q)\,d\mu(q)$$

- How can sequences $\{q^n\}$ be generated? Convergence and rate of convergence?

Different paradigmic methods

Type 1 $(q^n)_{n\in\mathbb{N}}$ are i.i.d. random variables, law of density $f(q) = Z_q^{-1}\mathrm{e}^{-\beta V(q)}$;

Type 2 $(q^n)_{n\in\mathbb{N}}$ realization of a continuous state-space Markov chain, leaving π invariant (id. μ);

Type 3 $(q^n)_{n\in\mathbb{N}}$ is an approximation $(q_{t_n})_{n\in\mathbb{N}}$ of the realization of a stochastic process $(q_t)_{t\geq 0}$ leaving π invariant (id. μ);

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 the trajectory of an extended deterministic system, which has an invariant measure $d\rho$, whose projection is $d\mu$.

Outline of the comparison

- (Quick) Presentation of different methods:
    1. Purely stochastic methods (Rejection, MIS)
    2. Mixed stochastic/molecular dynamics methods (HMC, Langevin)
    3. Deterministic methods (Nosé-Hoover and beyond)

- A numerical comparison on a benchmark system

- Perspectives

Purely stochastic methods

Rejection algorithm

- Sampling according to a density f can be done using a hat function g such that $f \leq cg$ if appropriately reweighting proposals generated according to g (i.i.d. random variables)

Algorithme 1 (Rejection) *For $n \geq 0$,*

1. *generate $\widetilde{q} \in \mathcal{M}$ according to g and compute $r = \dfrac{f(\widetilde{q})}{cg(\widetilde{q})}$;*
2. *draw $s \sim \mathcal{U}[0,1]$*
3. *if $s > r$ go back to (1) (reject the proposition $\widetilde{q}$), otherwise go to (4);*
4. *set $q^n = \widetilde{q}$ (accept proposition $\widetilde{q}$); replace n by $n+1$ and go back to (1).*

Rejection (convergence)

- Convergence of the empirical mean: LLN $S_N(A) = \sum_{n=0}^{N-1} A(q^n)$. Then

$$\pi(|A|) < +\infty \Rightarrow \lim_{N\to\infty} \frac{1}{N} S_N(A) = \int_{\mathcal{M}} A\,d\pi \quad \text{a.s.}$$

- Rate of convergence given by TCL :

$$\pi(|A|^2) < +\infty \Rightarrow S_N(\bar{A}) \underset{N\to\infty}{\rightarrow} \mathcal{N}(0,1),$$

with $\gamma_A > 0$, $\bar{A} = A - \int_{\mathcal{M}} A\,d\pi$.

- Ask for a good choice of g (*rejection control*): the acceptance rate usually decreases drastically with increasing space dimension!

Metropolized Independence Sampler

- "Markov chain" version of Rejection
- Metropolis-Hastings algorithm with i.i.d. proposals (density g)

Algorithme 2 (Metropolized independence sampling) *Set* $w = f/g$*. Given* q^0*, for* $n \geq 1$*,*

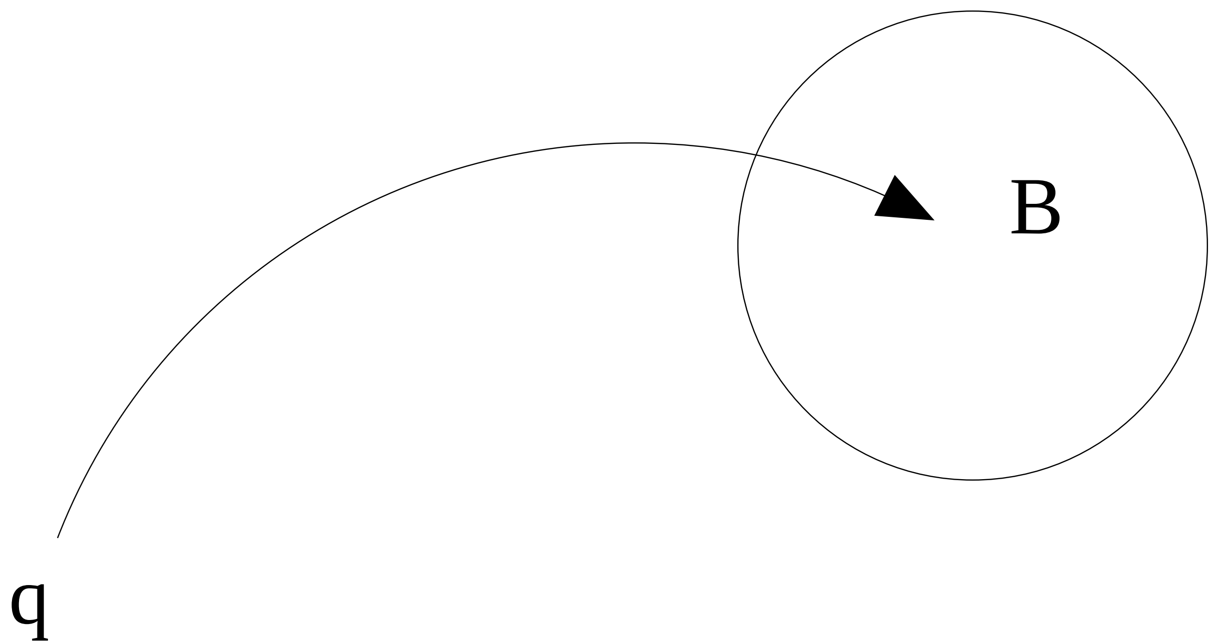
1. *generate* $\tilde{q} \in \mathcal{M}$ *according to* g
2. *draw* $s \sim \mathcal{U}[0, 1]$
3. *if* $s \leq \min\left\{1, \dfrac{w(\tilde{q})}{w(q^n)}\right\}$*, set* $q^{n+1} = \tilde{q}$*, otherwise set* $q^{n+1} = q^n$*;*
4. *replace* n *by* $n + 1$ *and go back to (1).*

- cf. Metropolis-Hastings scheme with proposal density $q(x, \cdot) \equiv g(\cdot)$

$$p(x, y) = \min\left(1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}\right)$$

Some elements on Markov chain convergence

- Reference: S.P. Meyn et R.L. Tweedie, *Markov Chains and Stochastic Stability*, Springer (1993)
- A Markov chain is characterized by its transition kernel P: if $q \in \mathcal{M}$ and $B \in \mathcal{B}(\mathcal{M})$ is a Borel set of M, then $P(q, B)$ is the probability to reach B starting from q in one iteration.



Some elements on Markov chain convergence (2)

- Convergence of the empirical mean along one trajectory provided
  1. the canonical measure is invariant:

$$\forall B \in \mathcal{B}(\mathcal{M}), \quad \pi(B) = \int_{\mathcal{M}} P(q, B)\pi(dq);$$

  2. P satisfies an accessibility condition

$$\forall q \in \mathcal{M}, \forall B \in \mathcal{B}(\mathcal{M}), \exists n \in \mathbb{N} \quad \lambda^{\mathrm{Leb}}(B) > 0 \Rightarrow P^n(q, B) > 0$$

- A Rate of convergence can be precised in some cases (TCL for Markov chains)
- For example, the MIS is ergodic whenever $\mathrm{supp}(f) \subset \mathrm{supp}(g)$

Stochastically perturbed Molecular dynamics

Molecular dynamics requires perturbations!

- Molecular dynamics = Hamiltonian dynamics on the manifold

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

  where $E_0 = H(q_0, p_0)$ is the initial energy of the system.

- Requires perturbations to sample all submanifolds $T^*\mathcal{M}(E_0)$ (for $E_0 \geq 0$)
- Types of pertubations :
  1. strong, at discrete times (HMC)
  2. smooth but permanent (Langevin)

Hybrid Monte Carlo (HMC)

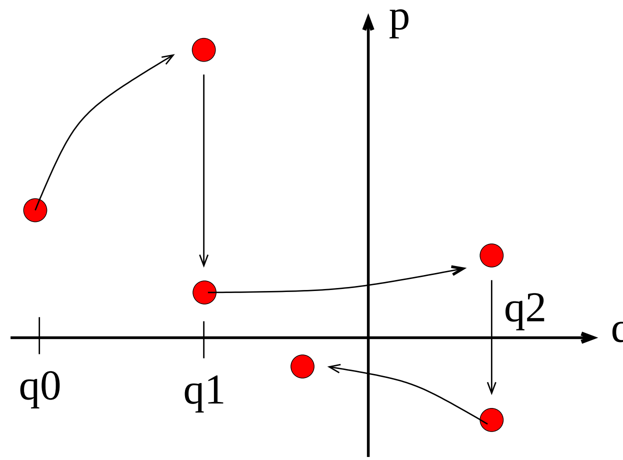
Markov chain in the configuration space (Duane et al. 1987, Schuette et al. 1999). Starting from q^n:

- generate momenta p^n according to $\mathcal{P}(p)dp = 1/Z_p \; e^{-\beta p^2/2m} dp$
- compute (an approximation of) the flow $\Phi_\tau(q^n, p^n) = (\tilde{q}^{n+1}, \tilde{p}^{n+1})$ of Newton's equations, i.e. integrate

$$\dot{q}_i = \frac{p_i}{m_i}, \quad \dot{p}_i = -\nabla_{q_i} V(q) \tag{1}$$

  on a time τ starting from (q^n, p^n).

- accept $\tilde{q}^{n+1}$ and set $q^{n+1} = \tilde{q}^{n+1}$ with a probability $\min\left(1, \exp -\beta(\tilde{E} - E_n)\right)$; otherwise set $q^{n+1} = q^n$.



Two parameters : τ and Δt.

Méthodes mixtes : HMC (2)

- Ergodicity requires assumptions on V (cf. harmonic oscillator)
- Convergence of the empirical mean under rather intricate assumptions : C. Schütte, Habilitation thesis (1998).
- A simplified result :

  Theorem 1 (Cancès, Legoll, Stoltz (2005)) *Assume that $V \in C^1(\mathcal{M})$ is bounded from above and ∇V is globally Lipschitz. Consider a sequence of points (q_n) generated by the HMC algorithm. Then, for almost all starting points $q^0 \in \mathcal{M}$,*

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

- Ingredients of the proof: least-action principle, continuity of the flow, Hausdorff measure

Langevin dynamics

- Type: discretization of a Markov process
- Hypo-elliptic SDE

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

where W_t is a $3M$-dimensional Wiener process

- Fluctuation/dissipation relation

$$\sigma = (2\xi/\beta)^{1/2}.$$

- Ergodicity can be proven under minimal assumptions on V

Langevin dynamics: discretization

- Discretization (BBK algorithm)

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

- numerical fluctuation/dissipation relation

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

so that the kinetic temperature is correct (theoretical analysis when $\nabla V = 0$, numerical experiments otherwise)

- Parameters : Δt, ξ (rule: $\xi \Delta t / 2m_i$ "not too large")

Purely deterministic methods

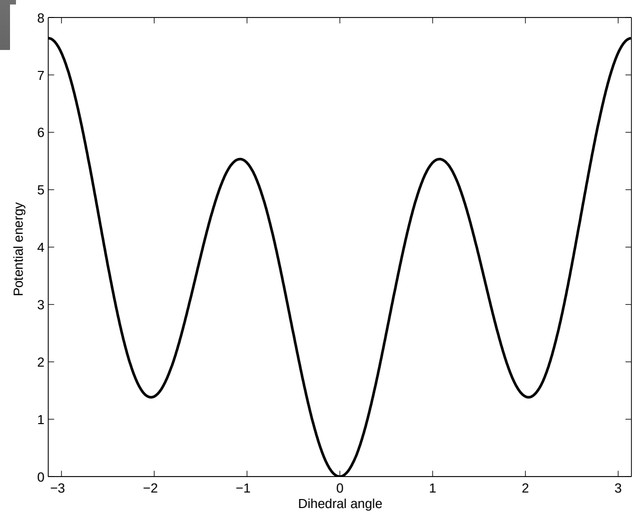
Nosé-Hoover methods and beyond

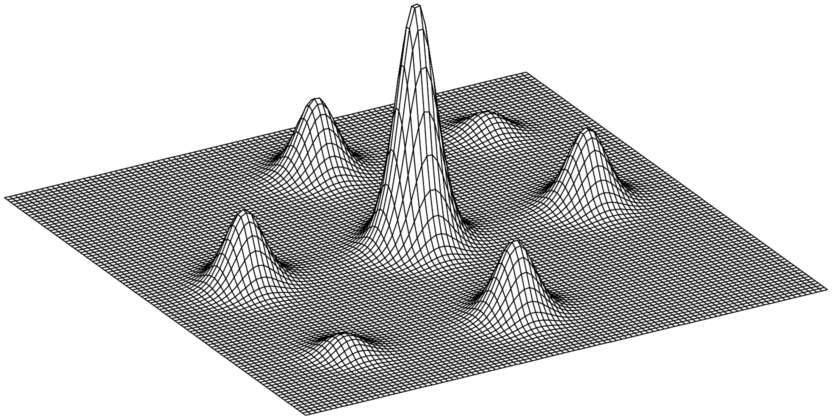
- Extended system (q, p, η, ξ) + ODE system
- Projection of the dynamics onto the variables (q, p): canonical measure invariant
- No theoretical proof of convergence (Non-convergence in some cases)
- Generalizations to account for these failures: Nosé-Hoover chain, Recursive Multiple Thermostats
- Discretization: reversible-in-time algorithms, symplectic or measure-preserving

Numerical comparison

Numerical benchmark

Linear alkane model in the united-atom setting CH_3-$(CH_2)_n$-CH_3 (Ryckaert/Bellemans)





Left : dihedral angle potential.

Right : Empirical probability distribution projected onto the plane (ϕ_1, ϕ_2) for a pentane molecule at T = 300 K (Importance sampling, $M = 10^9$ configurations).

Testing the numerical convergence

- Convergence indicator ?
- Usually, tests on the kinetic part (mean, first moments)
- Or convergence of given physical observables
- Here, direct comparison of projections of empirical and analytical distributions onto the dihedral angle plane in the simple case when there are no Lennard-Jones interactions

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

Results

- Purely stochastic methods are efficient in low-dimensional systems
- Déterministic become competitive when the dimension of space increases
- Stochastically perturbed MD have a robust behavior w.r.t. space dimension
- Improvements: performing several short trajectories (parallelization), evolving an initial distribution, undersampling

Perspectives

- Blue Moon sampling for free-energy differences = dynamics on a submanifold
- At low temperatures, the methods as such may fail to be ergodic: simulated-annealing type strategies (M. Rousset and G. Stoltz, *An interacting particle system approach for molecular dynamics*), also allowing computation of free-energy differences
- Computation of dynamical properties

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

Which dynamics should be used? Perturbation of the NVE dynamics through stochastic forcing at the boundary (E. Cancès and G. Stoltz, *Thermal boundary conditions for the computation of dynamical properties with molecular dynamics*)