Physique statistique numérique

Gabriel STOLTZ

gabriel.stoltz@enpc.fr

(CERMICS, Ecole des Ponts & MATHERIALS team, INRIA Paris)

"Décision dans l'incertain", mars 2023

General references (1)

- Computational Statistical Physics
  - D. Frenkel and B. Smit, *Understanding Molecular Simulation, From Algorithms to Applications* (Academic Press, 2002)
  - M. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation* (Oxford, 2010)
  - M. P. Allen and D. J. Tildesley, *Computer simulation of liquids* (Oxford University Press, 1987)
  - T. Schlick, *Molecular Modeling and Simulation* (Springer, 2002)
- Computational Statistics [my personal references... many more out there!]
  - J. Liu, *Monte Carlo strategies in scientific computing*, Springer, 2008
  - W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds), *Markov chain Monte Carlo in practice* (Chapman & Hall, 1996)
  - C. P. Robert and G. Casella, *Monte Carlo Statistical Methods* (Springer, 2004)
- Convergence of Markov chains
  - S. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability* (Cambridge University Press, 2009)
  - R. Douc, E. Moulines, P. Priouret and P. Soulier, *Markov chains* (Springer, 2018)

General references (2)

- Sampling the Boltzmann–Gibbs measure
  - L. Rey-Bellet, Ergodic properties of Markov processes, *Lecture Notes in Mathematics*, **1881** 1–39 (2006)
  - E. Cancès, F. Legoll and G. Stoltz, Theoretical and numerical comparison of some sampling methods, *Math. Model. Numer. Anal.* **41**(2) (2007) 351-390
  - T. Lelièvre, M. Rousset and G. Stoltz, *Free Energy Computations: A Mathematical Perspective* (Imperial College Press, 2010)
  - B. Leimkuhler and C. Matthews, *Molecular Dynamics: With Deterministic and Stochastic Numerical Methods* (Springer, 2015).
  - T. Lelièvre and G. Stoltz, Partial differential equations and stochastic methods in molecular dynamics, *Acta Numerica* **25**, 681-880 (2016)

- Machine learning and sampling
  - C. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006)
  - D. Barber, *Bayesian Reasoning and Machine Learning* (Cambridge University Press, 2012)
  - K. P. Murphy, *Probabilistic Machine Learning: An Introduction* (MIT Press, 2023)

Sampling high-dimensional probability measures

Statistical physics (1)

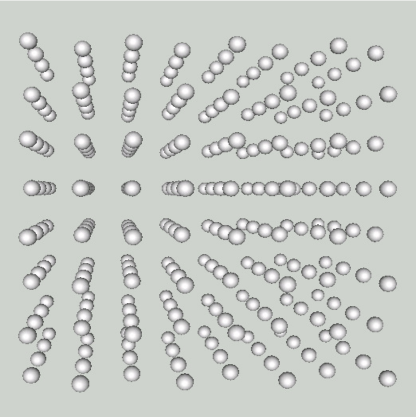
- **Aims of computational statistical physics**
  - numerical microscope
  - computation of average properties, static or dynamic

- **Orders of magnitude**
  - distances $\sim 1\ \mathring{A} = 10^{-10}$ m
  - energy per particle $\sim k_{\mathrm{B}}T \sim 4 \times 10^{-21}$ J at room temperature
  - atomic masses $\sim 10^{-26}$ kg
  - time $\sim 10^{-15}$ s
  - number of particles $\sim \mathcal{N}_A = 6.02 \times 10^{23}$

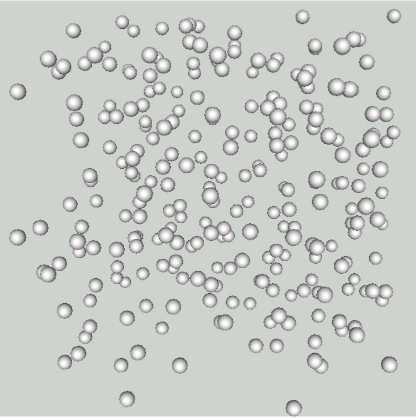
- **"Standard" simulations**
  - 10^6 particles ["world records": around 10^9 particles]
  - integration time: (fraction of) ns ["world records": (fraction of) μs]

Statistical physics (2)

What is the melting temperature of argon?



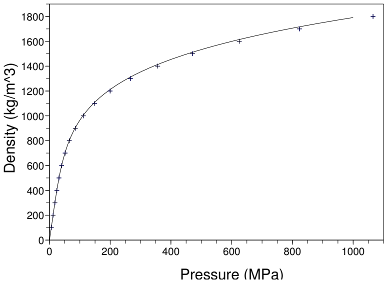
(a) Solid argon (low temperature)

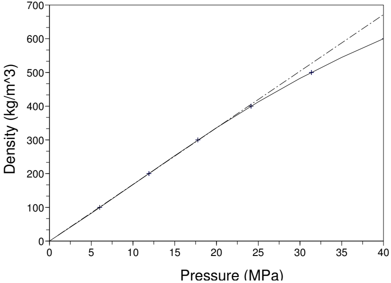


(b) Liquid argon (high temperature)

Statistical physics (3)

"Given the structure and the laws of interaction of the particles, what are the macroscopic properties of the matter composed of these particles?"

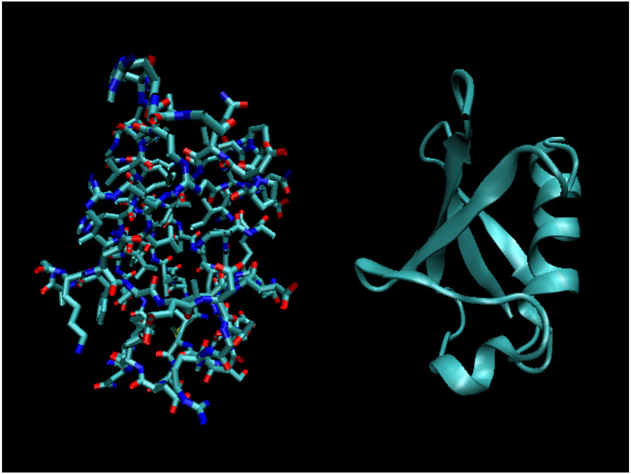




Equation of state (pressure/density diagram) for argon at $T = 300$ K

Statistical physics (4)

What is the structure of the protein? What are its typical conformations, and what are the transition pathways from one conformation to another?



Statistical physics (5)

• Microstate of a classical system of N particles:

$$(q,p) = (q_1,\dots,q_N,\ p_1,\dots,p_N) \in \mathcal{E}$$

Positions q (configuration), momenta p (to be thought of as $M\dot{q}$)

• In the simplest cases, $\mathcal{E} = \mathcal{D} \times \mathbb{R}^{3N}$ with $\mathcal{D} = \mathbb{R}^{3N}$ or $\mathbb{T}^{3N}$

• More complicated situations can be considered: molecular constraints defining submanifolds of the phase space

• Hamiltonian $H(q,p) = E_{\mathrm{kin}}(p) + V(q)$, where the kinetic energy is

$$E_{\mathrm{kin}}(p) = \frac{1}{2}\, p^T M^{-1} p, \qquad M = \begin{pmatrix} m_1 \,\mathrm{Id}_3 & & 0 \\ & \ddots & \\ 0 & & m_N \,\mathrm{Id}_3 \end{pmatrix}.$$

Statistical physics (6)

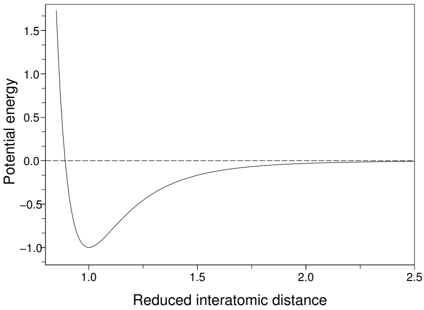
- All the physics is contained in V
  - ideally derived from quantum mechanical computations
  - in practice, empirical potentials for large scale calculations

- An example: Lennard-Jones pair interactions to describe noble gases

$$V(q_1,\ldots,q_N) = \sum_{1\leqslant i<j\leqslant N} v(|q_j - q_i|)$$

$$v(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

$$\text{Argon:} \quad \begin{cases} \sigma = 3.405 \times 10^{-10}\ \text{m} \\ \varepsilon/k_{\mathrm{B}} = 119.8\ \ \text{K} \end{cases}$$



Statistical physics (7)

- Macrostate of the system described by a probability measure

Equilibrium thermodynamic properties (pressure,...)

$$\mathbb{E}_\mu(\varphi) = \int_{\mathcal{E}} \varphi(q,p)\, \mu(dq\, dp)$$

- Choice of thermodynamic ensemble
  - least biased measure compatible with the observed macroscopic data
  - Volume, energy, number of particles, ... fixed exactly or in average
  - Equivalence of ensembles (as $N \to +\infty$)

- Canonical ensemble $=$ measure on (q,p), average energy fixed H

$$\mu_{\mathrm{NVT}}(dq\, dp) = Z_{\mathrm{NVT}}^{-1}\, \mathrm{e}^{-\beta H(q,p)}\, dq\, dp$$

with $\beta = \dfrac{1}{k_{\mathrm{B}} T}$ the Lagrange multiplier of the constraint $\displaystyle\int_{\mathcal{E}} H\, \rho\, dq\, dp = E_0$

Standard techniques to sample probability measures (1)

- The basis is the generation of numbers uniformly distributed in $[0,1]$

- Deterministic sequences which look like they are random...
  - Early methods: linear congruential generators ("chaotic" sequences)

$$x_{n+1} = a x_n + b \mod c, \qquad u_n = \frac{x_n}{c-1}$$

  - Known defects: short periods, point alignments, etc, which can be (partially) patched by cleverly combining several generators

- More recent algorithms: shift registers, such as Mersenne-Twister
$\rightarrow$ defaut choice in *e.g.* Python, available in the GNU Scientific Library

- Randomness tests: various flavors

Standard techniques to sample probability measures (2)

- Classical distributions are obtained from the uniform distribution by...
  - inversion of the cumulative function $F(x) = \int_{-\infty}^{x} f(y)\,dy$ (which is an increasing function from $\mathbb{R}$ to $[0,1]$)

$$X = F^{-1}(U) \sim f(x)\,dx$$

Proof: $\mathbb{P}\{a < X \leqslant b\} = \mathbb{P}\{a < F^{-1}(X) \leqslant b\} = \mathbb{P}\{F(a) < U \leqslant F(b)\} = F(b) - F(a) = \int_a^b f(x)\,dx$

Example: exponential law of density $\lambda \mathrm{e}^{-\lambda x} \mathbf{1}_{\{x \geqslant 0\}}$, $F(x) = \mathbf{1}_{\{x \geqslant 0\}}(1 - \mathrm{e}^{-\lambda x})$, so that $X = -\frac{1}{\lambda} \ln U$

  - change of variables: standard Gaussian $G = \sqrt{-2 \ln U_1} \cos(2\pi U_2)$

Proof: $\mathbb{E}(f(X,Y)) = \frac{1}{2\pi} \int_{\mathbb{R}^2} f(x,y)\, \mathrm{e}^{-(x^2+y^2)/2}\,dx\,dy = \int_0^{+\infty} f\left(\sqrt{r}\cos\theta, \sqrt{r}\sin\theta\right) \frac{1}{2} \mathrm{e}^{-r/2}\,dr\,\frac{d\theta}{2\pi}$

  - using the rejection method

Find a probability density g and a constant $c \geqslant 1$ such that $0 \leqslant f(x) \leqslant c g(x)$. Generate i.i.d. variables $(X^n, U^n) \sim g(x)\,dx \otimes \mathcal{U}[0,1]$, compute $r^n = \frac{f(X^n)}{c g(X^n)}$, and accept X^n if $r^n \geqslant U^n$

Standard techniques to sample probability measures (3)

- The previous methods work only
  - for low-dimensional probability measures
  - when the normalization constants of the probability density are known (or at least bounds, as for rejection sampling)

- In more complex cases, one needs to resort to trajectory averages

Ergodic methods

$$\frac{1}{N_{\mathrm{iter}}} \sum_{n=1}^{N_{\mathrm{iter}}} \varphi(x^n) \xrightarrow[N_{\mathrm{iter}} \to +\infty]{} \int \varphi \, d\mu$$

- **Find methods for which**
  - the convergence is guaranteed? (and in which sense?)
  - error estimates are available? (typically with Central Limit Theorem)

Standard techniques to sample probability measures (4)

- Assume that $x^n \sim \pi$ are idependently and identically distributed (i.i.d.)

Law of Large Numbers for $\varphi \in L^1(\pi)$

$$S_{N_{\rm iter}} = \frac{1}{N_{\rm iter}} \sum_{n=1}^{N_{\rm iter}} \varphi(x^n) \xrightarrow[N_{\rm iter}\to+\infty]{} \mathbb{E}_\pi(\varphi) = \int_{\mathcal{X}} \varphi \, d\pi \quad \text{almost surely}$$

Central Limit Theorem for $\varphi \in L^2(\pi)$

$$\sqrt{N_{\rm iter}} \left(S_{N_{\rm iter}} - \int \varphi \, d\pi \right) \xrightarrow[N_{\rm iter}\to+\infty]{\text{law}} \mathcal{N}(0,\sigma_\varphi^2), \ \ \sigma_\varphi^2 = \int_{\mathcal{X}} \left[\varphi - \mathbb{E}_\pi(\varphi)\right]^2 \, d\pi$$

- This should be thought of in practice as $S_{N_{\rm iter}} \simeq \mathbb{E}_\pi(\varphi) + \dfrac{\sigma_\varphi}{\sqrt{N_{\rm iter}}}\mathcal{G}$

Metropolis–Hastings algorithms

Metropolis-Hastings algorithm (1)

- Markov chain method[1,2], on position space

  - Given q^n, propose $\tilde{q}^{n+1}$ according to transition probability $T(q^n, \tilde{q})$
  - Accept the proposition with probability $\min\left(1, r(q^n, \widetilde{q}^{n+1}\right)$ where

$$r(q,q') = \frac{T(q',q)\,\nu(q')}{T(q,q')\,\nu(q)}, \qquad \nu(dq) \propto \mathrm{e}^{-\beta V(q)}.$$

  If acception, set $q^{n+1} = \tilde{q}^{n+1}$; otherwise, set $q^{n+1} = q^n$.

- Example of proposals
  - Gaussian displacement $\tilde{q}^{n+1} = q^n + \sigma\, G^n$ with $G^n \sim \mathcal{N}(0, \mathrm{Id})$
  - Biased random walk[3,4] $\tilde{q}^{n+1} = q^n - \sigma^2 \nabla V(q^n) + \sqrt{\frac{2\sigma^2}{\beta}}\, G^n$

---

[1] Metropolis, Rosenbluth ($\times 2$), Teller ($\times 2$), *J. Chem. Phys.* (1953)

[2] W. K. Hastings, *Biometrika* (1970)

[3] G. Roberts and R.L. Tweedie, *Bernoulli* (1996)

Metropolis-Hastings algorithm (2)

- The normalization constant in the canonical measure needs not be known

- Transition kernel: accepted moves + rejection

$$P(q, dq') = \min\left(1, r(q,q')\right) T(q,q')\, dq' + \left(1 - \alpha(q)\right) \delta_q(dq'),$$

where $\alpha(q) \in [0,1]$ is the probability to accept a move starting from q:

$$\alpha(q) = \int_{\mathcal{D}} \min\left(1, r(q,q')\right) T(q,q')\, dq'.$$

- Rejection rate $1 - \alpha(q) \sim \sigma$ for RWMH, and σ^3 for MALA

- The canonical measure is reversible with respect to ν

$$P(q, dq')\nu(dq) = P(q', dq)\nu(dq')$$

This implies invariance: $\displaystyle \int_{\mathcal{D}} \int_{\mathcal{D}} \varphi(q') P(q, dq')\, \nu(dq) = \int_{\mathcal{D}} \varphi(q)\, \nu(dq)$

Metropolis-Hastings algorithm (3)

- Proof: Detailed balance on the absolutely continuous parts

$$\begin{aligned}\min\left(1, r(q,q')\right) T(q,dq')\nu(dq) &= \min\left(1, r(q',q)\right) r(q,q') T(q,dq')\nu(dq) \\ &= \min\left(1, r(q',q)\right) T(q',dq)\nu(dq')\end{aligned}$$

using successively $\min(1,r) = r\min\left(1, \frac{1}{r}\right)$ and $r(q,q') = \dfrac{1}{r(q',q)}$

- Equality on the singular parts $(1-\alpha(q))\,\delta_q(dq')\nu(dq) = (1-\alpha(q'))\delta_{q'}(dq)\nu(dq')$

$$\begin{aligned}\int_{\mathcal{D}}\int_{\mathcal{D}} \phi(q,q')\,(1-\alpha(q))\,\delta_q(dq')\nu(dq) &= \int_{\mathcal{D}} \phi(q,q)(1-\alpha(q))\nu(dq) \\ &= \int_{\mathcal{D}}\int_{\mathcal{D}} \phi(q,q')(1-\alpha(q'))\delta_{q'}(dq)\nu(dq')\end{aligned}$$

- Note: other acceptance ratios $R(r)$ possible as long as $R(r) = rR(1/r)$, but the Metropolis ratio $R(r) = \min(1,r)$ is optimal in terms of asymptotic variance[5]

[5] P. Peskun, *Biometrika* (1973)

Metropolis-Hastings algorithm (4)

- Irreducibility: for almost all q_0 and any set $\mathcal{S}$ of positive measure, there exists n such that

$$P^n(q_0,\mathcal{S}) = \int_{x\in\mathcal{D}} P(q_0,dx)\, P^{n-1}(x,\mathcal{S}) > 0$$

- Assume also aperiodicity (comes from rejections)

- Pathwise ergodicity[6] $\displaystyle \lim_{N_{\mathrm{iter}}\to+\infty} \frac{1}{N_{\mathrm{iter}}} \sum_{n=1}^{N_{\mathrm{iter}}} \varphi(q^n) = \int_{\mathcal{D}} \varphi(q)\,\nu(dq)$

- Central limit theorem for Markov chains under additional assumptions:

$$\sqrt{N_{\mathrm{iter}}} \left| \frac{1}{N_{\mathrm{iter}}} \sum_{n=1}^{N_{\mathrm{iter}}} \varphi(q^n) - \int_{\mathcal{D}} \varphi(q)\,\nu(dq) \right| \xrightarrow[N_{\mathrm{iter}}\to+\infty]{\mathrm{law}} \mathcal{N}(0,\sigma_\varphi^2)$$

[6] S. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability* (1993)

Metropolis-Hastings algorithm (5)

- The asymptotic variance σ_φ^2 takes into account the correlations:

$$\sigma_\varphi^2 = \mathrm{Var}_\nu(\varphi) + 2\sum_{n=1}^{+\infty} \mathbb{E}_\nu\Big[\big(\varphi(q^0) - \mathbb{E}_\nu(\varphi)\big)\big(\varphi(q^n) - \mathbb{E}_\nu(\varphi)\big)\Big]$$

Proof: Consider $\widetilde{\varphi} = \varphi - \mathbb{E}_\nu(\varphi)$ and the average $\widetilde{\Phi}_{N_{\mathrm{iter}}} = \dfrac{1}{N_{\mathrm{iter}}}\displaystyle\sum_{n=1}^{N_{\mathrm{iter}}} \widetilde{\varphi}(q^n)$

Compute $N_{\mathrm{iter}}\mathbb{E}_\nu\left(\widetilde{\Phi}_{N_{\mathrm{iter}}}^2\right) = \dfrac{1}{N_{\mathrm{iter}}}\displaystyle\sum_{n,m=0}^{N_{\mathrm{iter}}} \mathbb{E}_\nu\Big(\widetilde{\varphi}(q^n)\widetilde{\varphi}(q^m)\Big)$

Stationarity $\mathbb{E}_\nu\Big(\widetilde{\varphi}(q^n)\widetilde{\varphi}(q^m)\Big) = \mathbb{E}_\nu\Big(\widetilde{\varphi}(q^{n-m})\widetilde{\varphi}(q^0)\Big)$ for $n \geqslant m$, implies

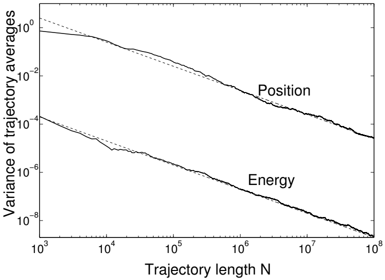
$$N_{\mathrm{iter}}\mathbb{E}_\nu\left(\widetilde{\Phi}_{N_{\mathrm{iter}}}^2\right) = \mathbb{E}_\nu\Big(\widetilde{\varphi}\left(q^0\right)^2\Big) + 2\sum_{n=1}^{N_{\mathrm{iter}}}\left(1 - \frac{n}{N_{\mathrm{iter}}}\right)\mathbb{E}_\nu\Big(\widetilde{\varphi}(q^n)\widetilde{\varphi}(q^0)\Big)$$

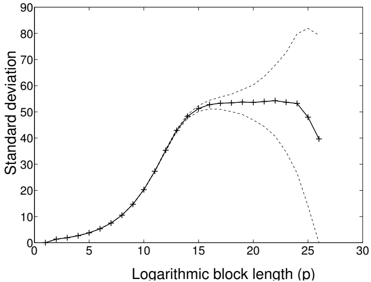
Metropolis-Hastings algorithm (6)

- Estimation of σ_φ^2 by block averaging (batch means)

$$\sigma_\varphi^2 = \lim_{N,M\to+\infty} \frac{N}{M} \sum_{k=1}^{M} \left(\Phi_N^k - \Phi_{NM}^1\right)^2, \qquad \Phi_N^k = \frac{1}{N} \sum_{n=(k-1)N+1}^{kN} \varphi(q^n)$$

Expected $\Phi_N^k \sim \displaystyle\int_{\mathcal{X}} \varphi \, d\nu + \frac{\sigma_\varphi}{\sqrt{N}} \mathscr{G}^k$, with $\mathscr{G}^k$ i.i.d.





Metropolis-Hastings algorithm (7)

- Useful rewriting: number of correlated steps $\sigma_\varphi^2 = N_{\mathrm{corr}} \mathrm{Var}_\nu(\varphi)$

- Numerical efficiency: trade-off between acceptance and sufficiently large moves in space to reduce autocorrelation (rejection rate around 0.5)[7]

- Refined Monte Carlo moves such as
  - "non physical" moves
  - parallel tempering
  - replica exchanges
  - Hybrid Monte-Carlo

- A way to stabilize discretization schemes for SDEs

[7] Roberts/Gelman/Gilks (1997), ..., Jourdain/Lelièvre/Miasojedow (2012)

Hybrid Monte–Carlo

The Hamiltonian dynamics (1)

Hamiltonian dynamics

$$\left\{ \begin{aligned} \frac{dq(t)}{dt} &= \nabla_p H(q(t),p(t)) &&= M^{-1}p(t) \\ \frac{dp(t)}{dt} &= -\nabla_q H(q(t),p(t)) &&= -\nabla V(q(t)) \end{aligned} \right.$$

Assumed to be well-posed (*e.g.* when the energy is a Lyapunov function)

- Flow: $\phi_t(q_0,p_0)$ solution at time t starting from initial condition (q_0,p_0)

- Why Hamiltonian formalism? (instead of working with velocities?)
  - Note that the vector field is divergence-free

$$\mathrm{div}_q\Big(\nabla_p H(q(t),p(t))\Big) + \mathrm{div}_p\Big(-\nabla_q H(q(t),p(t))\Big) = 0$$

  - Volume preservation $\displaystyle\int_{\phi_t(B)} dq\,dp = \int_B dq\,dp$

The Hamiltonian dynamics (2)

- Other properties
  - Preservation of energy $H \circ \phi_t = H$

$$\frac{d}{dt}\Big[H\big(q(t),p(t)\big)\Big] = \nabla_q H(q(t),p(t))\cdot\frac{dq(t)}{dt} + \nabla_p H(q(t),p(t))\cdot\frac{dp(t)}{dt} = 0$$

  - Time-reversibility $\phi_{-t} = S \circ \phi_t \circ S$ where $S(q,p) = (q,-p)$

    Proof: use $S^2 = \mathrm{Id}$ and note that

$$S \circ \phi_{-t}(q_0,p_0) = \big(q(-t), -p(-t)\big)$$

    is a solution of the Hamiltonian dynamics starting from $(q_0,-p_0)$, as is $\phi_t \circ S(q_0,p_0)$. Conclude by uniqueness of solution.

  - Symmetry $\phi_{-t} = \phi_t^{-1}$ (in general, $\phi_{t+s} = \phi_t \circ \phi_s$)

The Hamiltonian dynamics (3)

- Numerical integration: usually Verlet scheme[8] (Strang splitting)

Störmer-Verlet scheme

$$\begin{cases} p^{n+1/2} = p^n - \dfrac{\Delta t}{2}\nabla V(q^n) \\ q^{n+1} = q^n + \Delta t\ M^{-1}p^{n+1/2} \\ p^{n+1} = p^{n+1/2} - \dfrac{\Delta t}{2}\nabla V(q^{n+1}) \end{cases}$$

- Properties:
  - Symplectic, symmetric, time-reversible
  - One force evaluation per time-step, linear stability condition $\omega\Delta t < 2$
  - In fact, $M\dfrac{q^{n+1} - 2q^n + q^{n-1}}{\Delta t^2} = -\nabla V(q^n)$

[8] L. Verlet, *Phys. Rev.* **159**(1) (1967) 98-105

Hybrid Monte Carlo (1)

- Measure $\mu(dq\,dp) = \mathrm{e}^{-\beta H(q,p)}\,dq\,dp$ with marginal $\nu(dq) = \mathrm{e}^{-\beta V(q)}\,dq$

- Markov chain in the configuration space[9,10]: parameters τ and Δt
  - generate momenta p^n according to $Z_p^{-1}\,\mathrm{e}^{-\beta p^T M^{-1} p/2}\,dp$
  - compute an approximation of the flow $\Phi_\tau(q^n,p^n) = (\tilde{q}^{n+1},\tilde{p}^{n+1})$ of the Hamiltonian dynamics (i.e. Verlet scheme with $\tau/\Delta t$ timesteps)
  - set $q^{n+1} = \tilde{q}^{n+1}$ with probability $\min\left(1, \mathrm{e}^{-\beta(H(\tilde{q}^{n+1},\tilde{p}^{n+1})-H(q^n,p^n))}\right)$; otherwise set $q^{n+1} = q^n$.

- Rejection rate of order Δt^2 when $\tau = \mathrm{O}(1)$, and Δt^3 for $\tau = \Delta t$

- **Various extensions**, including correlated momenta, random times τ, constraints, ...

- Ergodicity is an issue (quadratic potential with $\tau =$ period)

[9] S. Duane, A. Kennedy, B. Pendleton and D. Roweth, *Phys. Lett. B* (1987)
[10] Ch. Schütte, *Habilitation Thesis* (1999)

(Generalized) Hybrid Monte Carlo (1)

• Transformation $S = S^{-1}$ leaving $\mu(dx)$ invariant, *e.g.* $S(q,p) = (q,-p)$

• Assume that $r(x,x') = \dfrac{T(S(x'),S(dx))\,\pi(dx')}{T(x,dx')\,\pi(dx)}$ is defined and positive

Generalized Hybrid Monte Carlo

- given x^n, propose a new state $\tilde{x}^{n+1}$ from x^n according to $T(x^n,\cdot)$;
- accept the move with probability $\min\left(1, r(x^n,\tilde{x}^{n+1})\right)$, and set in this case $x^{n+1} = \tilde{x}^{n+1}$; otherwise, set $x^{n+1} = S(x^n)$.

• Reversibility up to S, *i.e.* $P(x,dx')\,\mu(dx) = P(S(x'),S(dx))\,\mu(dx')$

• Standard HMC: $T(q,dq') = \delta_{\Phi_\tau(q)}(dq')$, momentum reversal upon rejection (not important since momenta are resampled, but is important when momenta are partially resampled)

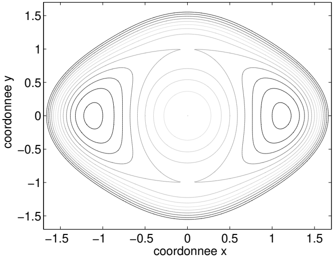
(Generalized) Hybrid Monte Carlo (2)

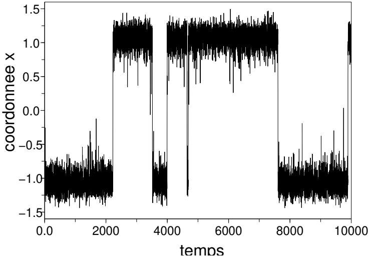
Complete algorithm ($M = \mathrm{Id}$, $\beta = 1$): starting from (q^n, p^n),

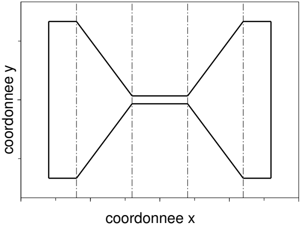
- Partially resample momenta as $p^{n+1/2} = \alpha p^n + \sqrt{1-\alpha^2}\, G^n$
- Perform one Verlet step as $(\widetilde{q}^{n+1}, \widetilde{p}^{n+1}) = \Phi_{\Delta t}(q^n, p^n)$
- Compute the acceptance probability $a^n = \mathrm{e}^{H(q^n,p^n) - H(\widetilde{q}^{n+1}, \widetilde{p}^{n+1})}$
- Sample $U^n \sim \mathcal{U}[0,1]$
- If $U^n \leqslant a^n$, set $(q^{n+1}, p^{n+1}) = (\widetilde{q}^{n+1}, \widetilde{p}^{n+1})$
  otherwise set $(q^{n+1}, p^{n+1}) = (q^n, -p^{n+1/2})$

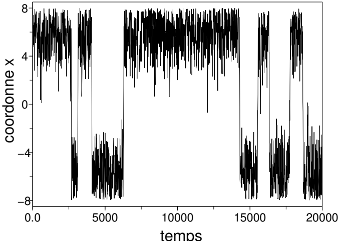
• Ergodicity no longer is an issue (irreducibility much easier to prove than for standard HMC)

Metastability: large variances...









Need for variance reduction! A lot remains to be done for actual systems...