Hybrid Monte Carlo in Path Space

Frank Pinski, University of Cincinnati

Joint work with
A. Beskos, J.M. Sanz-Serna, and A.M. Stuart

July 19, 2011

Lennard-Jones: 13 atoms

(Loading Video...)

Jellinek, Beck and Berry, J. Chem. Phys., 1987.

Outline

1. The General Problem
2. The Starting Point - Brownian Dynamics
   - Quadratic Variation
3. Onsager-Machlup Functional
4. The Continuum Limit
5. Sampling the Measure
   - Langevin Hairer, Stuart, Voss, Wiberg, Comm. in Math. Sciences, 2005
   - Preconditioning
   - Metropolis Adjusted Beskos, Roberts, Stuart, Voss, Stoch. & Dyn., 2008
   - Hybrid Monte Carlo Beskos, Pinski, Sanz-Serna, Stuart, Stoch. Proc. & App., accepted
6. Conclusion and Discussion

Starting Point - Brownian Dynamics

Sample Boltmann Distribution: $P_B \propto \exp(-V/T)$

$$dx = F\,du + \sqrt{2\,T}\,dW = T\,D \log P_B\,du + \sqrt{2\,T}\,dW$$

F is the force $F = -DV$
T is the temperature
u is the **time along the path**
dW is the standard Wiener Process

If a large energy barrier exists, the transition becomes a **rare event**.

The thrust of this work is to find an efficient way of sampling the transition paths themselves in a thermodynamically significant manner.

Brownian Paths

Finite Representation of a path

$$x_{i+1} - x_i = F_i \Delta u + \sqrt{2\,T\,\Delta u}\;\xi_i \quad \text{with} \quad P_G(\xi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2}\right)$$

Note: Path is continuous and almost nowhere differentiable.

Quadratic Variation

$$Q_v = \sum_i \left(x_{i+1} - x_i\right)^2 = 2\,T\,U$$

U is the length of the path

The grid space Δu must be sufficiently small to resolve the quadratic variation.

The high-frequency modes are dominated by noise.

Onsager-Machlup Functional

$$x_{i+1} - x_i = F_i \Delta u + \sqrt{2\,T\,\Delta u}\;\xi_i \quad \text{with} \quad P_G(\xi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2}\right)$$

The path probability can be constructed (OM functional)

$$\pi_{path} \propto \prod_i \exp\left(-\frac{\xi_i^2}{2}\right) = \exp\left(-\frac{\Delta u}{4T} \sum_i \left(\frac{\Delta x}{\Delta u} - F_i\right)^2\right)$$

Now consider a double ended path: fix both ends. These boundary conditions are chosen so that the starting configuration lies in one free-energy basin, and the other end lies in another basin.

The relative probability of such a path is expressed in terms of the path positions themselves. The noise history is implicitly included in the expression.

Continuum Limit

OM Functional: $\pi_{path} \propto \exp\left(-\frac{\Delta u}{4T}\sum_i \left(\frac{\Delta x}{\Delta u} - F_i\right)^2\right)$

Continuum Limit (The Measure – informally)

$\pi_{path} \propto \exp\left(-\frac{\mathbb{I}_0}{2T}\right)$ where $\mathbb{I}_0 = \frac{1}{2}\langle x,\, Lx\rangle + \langle 1,\, G(x)\rangle$

$\langle ...\rangle$ is the usual inner product,
$L = -d^2/du^2$ (non-negative) and
$G = \frac{1}{2}|DV|^2 - T\triangle V$

The function G is sometimes denoted as the path potential.
In the continuum limit, G contains the Laplacian of the particle potential. Note, that in equilibrium, $\mathbb{E}\left(|DV|^2\right) = T\,\mathbb{E}\left(\triangle V\right)$,

Need Girsanov's theorem and Ito's formula to justify these steps.

Langevin Sampling

Let the path evolve as a function of algorithmic time t:

$$\frac{\partial x}{\partial t} = 2\,T\,D \log \pi_{path} + \sqrt{4\,T}\,\frac{\partial w}{\partial t} = \frac{\partial^2 x}{\partial u^2} + DG + \sqrt{4\,T}\,\frac{\partial w}{\partial t}$$

This is subject to the imposed Boundary Conditions.

Stiff: different modes have different relaxation times.

Use "preconditioning," to arrive at the SPDE:

$$\frac{\partial x}{\partial t} = -x - L^{-1}\,DG + \sqrt{4\,T}\,\frac{dB}{dt}$$

with the last term being the unit Brownian Bridge, and again $L = -d^2/du^2$

All modes evolve at the same rate (democratic).

Metropolis Adjusted - Smart Monte Carlo

We want to sample paths from the measure (almost surely infinite), that we can write informally as

$\pi_{path} \propto \exp\left(-\frac{\mathbb{I}_0}{2\,T}\right)$ with $\mathbb{I}_0 = \frac{1}{2}\langle x,\, Lx\rangle + \langle 1,\, G(x)\,\rangle$

Use an approximation to $\frac{\partial x}{\partial t} = -x - L^{-1}\,DG + \sqrt{4\,T}\;B(t)$
as a (biased) way of generating a proposed path evolution.
Use a Metropolis-Hasting criterion to accept or reject the proposal.

Set Δt in the approximation of the SPDE to allow efficient movement through path space.

Proposed moves using small values of Δt tend to be accepted but a large number of steps are needed to move appreciably far away from current path.
Proposed moves using large values of Δt tend to be rejected as the integration errors become prohibitively large.

Hybrid Monte Carlo (HMC)

We want to sample paths from the measure

$$\pi_{path} \propto \exp\left(-\frac{\mathbb{I}_0}{2T}\right) \qquad \text{and} \qquad \mathbb{I}_0 = \frac{1}{2}\langle x,\, Lx\rangle + \langle 1,\, G(x)\,\rangle$$

Augment $\mathbb{I}_0$ to include "Kinetic Energy" thereby forming $\mathbb{I}$:

$$\mathbb{I} = \frac{1}{2}\langle p,\, M^{-1}p\rangle + \frac{1}{2}\langle x,\, Lx\rangle + \langle 1,\, G(x)\,\rangle$$

where M is the mass matrix.

The path p is composed of auxiliary variables, corresponding to momenta: they are conjugate to x, they do not alter the stationary distribution of paths and their (Gaussian) distribution is known.

HMC has four (4) ingredients.

1. Choice of mass Matrix M.
2. Hamiltonian Flow
3. Integrator that is reversible and volume conserving
4. Accept/reject criteria (based on "energy drift")

HMC - Mass Matrix

The effective Hamiltonian can be informally written as

$$\mathbb{I} = \tfrac{1}{2}\langle p,\, M^{-1}p\rangle + \tfrac{1}{2}\langle x,\, Lx\rangle + \langle 1,\, G(x)\,\rangle$$

We choose $M^{-1} = L$.

By inspecting the above equation for $\mathbb{I}$, we see that we keep the desired feature that all the modes evolve at the same rate.

The paths x and p are conditioned bridges, both having the same quadratic variation.

HMC - Hamiltonian Flow

The effective Hamiltonian can be informally written as

$$\mathbb{I} = \tfrac{1}{2}\langle p,\ Lp\rangle + \tfrac{1}{2}\langle x,\ Lx\rangle + \langle 1,\ G(x)\,\rangle$$

Use Hamilton's equations:

$$\frac{\partial x}{\partial t} = \frac{\partial\, \mathbb{I}}{\partial p} = L\, p$$

$$\frac{\partial p}{\partial t} = -\frac{\partial\, \mathbb{I}}{\partial x} = -L\, x - DG$$

We can combine these and get

$$\frac{\partial^2 x}{\partial t^2} = -x - L^{-1} DG$$

Note: we have seen something similar to the above equation before.

HMC - Integrator

Consider the second order equation and convert it to

$$v = \frac{\partial x}{\partial t} \qquad \textit{and} \qquad \frac{\partial v}{\partial t} = -x - L^{-1} DG$$

Splitting of the Verlet integrator:

1. Half step $\qquad w_i = v_i - \frac{h}{2}\, L^{-1} DG_i$
2. Full step – Rotation

$$\begin{pmatrix} x_{i+1} \\ w_{i+1} \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_i \\ w_i \end{pmatrix}$$

3. Half step $\qquad v_{i+1} = w_{i+1} - \frac{h}{2}\, L^{-1} DG_{i+1}$

$$\cos\theta = \cos h \quad \textit{or} \quad \frac{4 - h^2}{4 + h^2} \qquad \sin\theta = \sin h \quad \textit{or} \quad \frac{4\,h}{4 + h^2}$$

Integration scheme is Reversible and Volume Conserving.
For finite representations, this Verlet splitting preserves the Quadratic Variation of the evolving path.

HMC: Metropolis-Hastings Criterion

The value of $\mathbb{I}$ is almost surely infinite in the continuum limit.

Must devise a method to calculate differences in $\mathbb{I}$ as the path evolves without subtracting large (possibly infinite) numbers.

At the end of every MD step, $\Delta\mathbb{I}$ can be tracked.

$$\begin{aligned}\Delta\mathbb{I} \;=\; & \langle 1,\, G_{i+1}\rangle - \langle 1,\, G_i\rangle \\ & + \frac{h^2}{8}\Big(\langle DG_{i+1},\, L^{-1}DG_{i+1}\rangle - \langle DG_i,\, L^{-1}DG_i\rangle\Big) \\ & - \frac{h}{2\,\sin\theta}\Big(\langle DG_{i+1},\, x_{i+1}-x_i\rangle - \langle DG_i,\, x_i - x_{i+1}\rangle\Big)\end{aligned}$$

Accumulate the changes as one performs MD integration. If step size, h, is small, drift in $\mathbb{I}$ is minimal, the evolved path will be accepted. For large step sizes, the integration error will be substantial, and the entire sequence of paths will be rejected.

HMC: importance of Verlet splitting

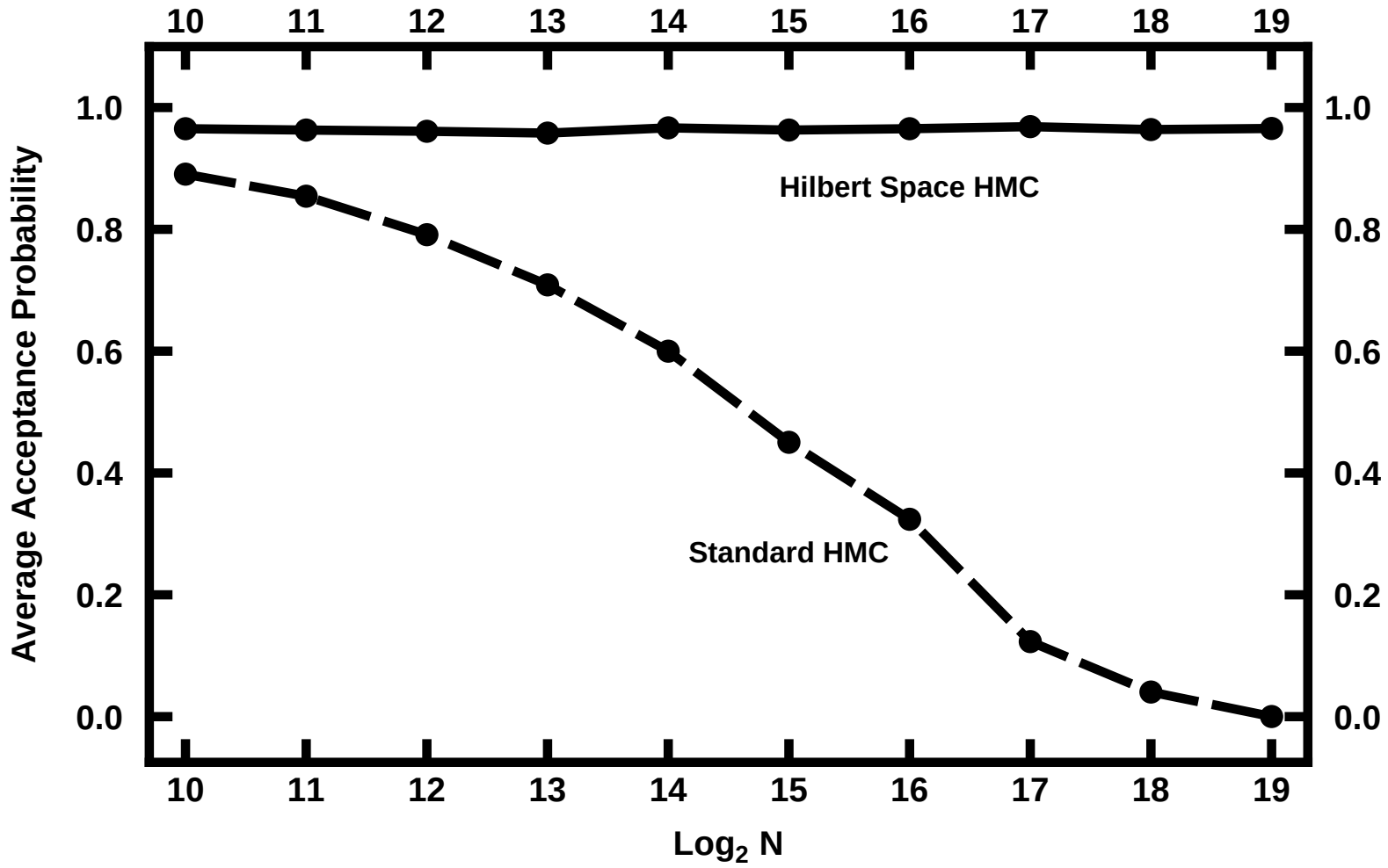

Figure: 1-d. Acceptance rate plotted as a function of N, the number of points used in the representation of the path.

Langevin vs HMC

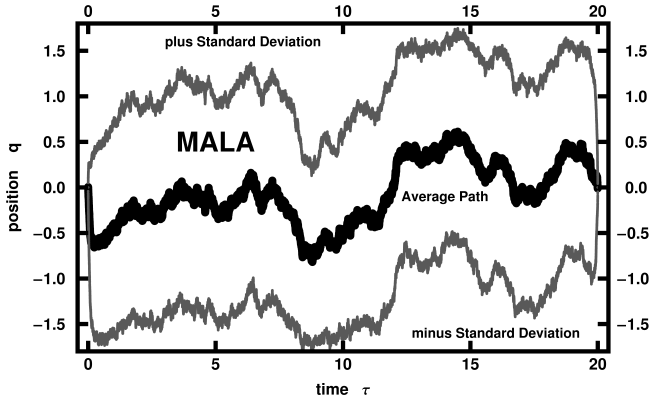

Figure: 1-d. Averages using Langevin with Metropolis-Hastings.

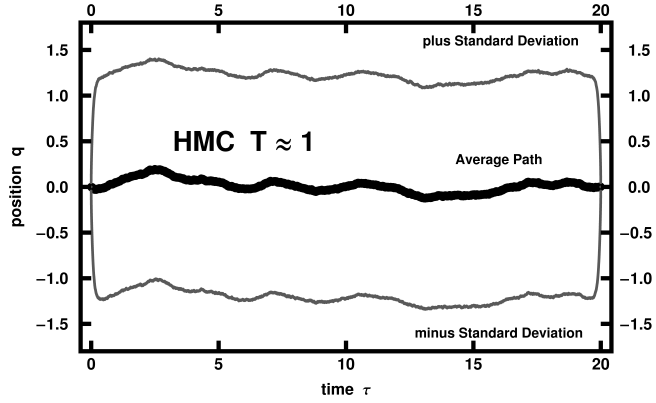

Figure: 1-d. Averages using the Hybrid Monte Carlo scheme.

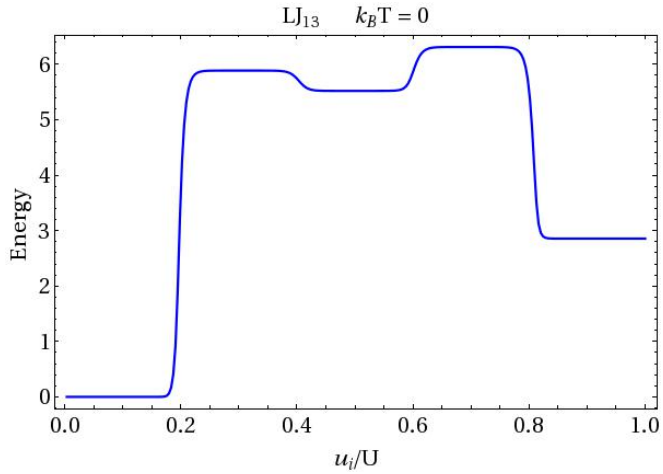

Figure: LJ_{13}: Energy along the path: beginning path.

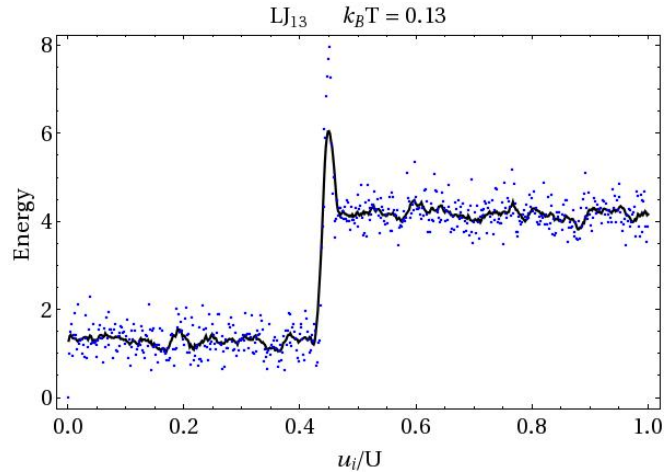

Figure: LJ_{13}: Energy along the path: annealed path.

Hybrid Monte Carlo Summary

We discussed the four (4) ingredients.

1. Choice of mass Matrix M.
2. Hamiltonian Flow
3. Integrator that is reversible and volume conserving
4. Accept/reject criteria (based on "energy drift")

1-dimension. Beskos, et al.
HMC is better than the Langevin (preconditioned) sampling.
Lennard-Jones systems.
With Langevin, the path evolves stochastically. The time step must be small since the hard-core nature of the potential makes the energy penalty prohibitively large for large step sizes.
In the HMC method, the systematic push of the MD integration allows the path to be shoved to new regions of path space using an energetically favorable route. But to do so, it needs thousands of integration steps.

Conclusion - Message to take home

- Algorithms on function space are robust to mesh refinement.
- The Langevin and Hybrid Monte Carlo (HMC) methods were described.
  - The methods are similar in many ways.
  - A series of short random bursts nudge the path along in the Langevin method.
  - A long, steady push is supplied to the path in the HMC algorithm.
- Sampling path space: HMC is more efficient.
  - The cost of the long deterministic integration is more than offset by the larger distances transversed during the sustained push that it provides.
  - The stochastic sampling, inherent in the Langevin approach, results in a more diffusive evolution of the path.