Hamiltonian Monte Carlo with Graphical Applications

Ho Chung Leon Law and Nathan Cunningham

OxWaSP 2015
Department Of Statistics
University Of Oxford

October 22, 2015

Hamiltonian Dynamics

Hamiltonian dynamics is the system described by a pair of differential equations with coordinates $(\mathbf{q}, thbfp) \in \mathbb{R}^{2d}$.
For $i = 1 \ldots d$,

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} \tag{1}$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} \tag{2}$$

where H is the Hamiltonian and is a function of $(\mathbf{q}, \mathbf{p})$.

Target Density and Energy functions

From statistical mechanics, given some energy function $E(x)$ in some system, we can express its canonical distribution as:

$$Pr(x) = \frac{1}{N} \exp(-E(x)) \tag{3}$$

where N is a normalising constant. We can rewrite our target density as:

$$U(\mathbf{q}) = -\log[\pi(\mathbf{q}) L(\mathbf{q}|D)] \tag{4}$$

where $\mathbf{q} \in \mathbb{R}^d$

$$P(\mathbf{q}) = \pi(\mathbf{q}) L(\mathbf{q}|D) \qquad q \in \mathbb{R}^d \tag{5}$$

is the target density.

Construction: Introduction of Auxillary Variable

Introduce an auxiliary variable $\mathbf{p} \in \mathbb{R}^d$ with energy function $K(\mathbf{p})$ and density function $Q(\mathbf{p})$.
Define the joint density of $(\mathbf{q}, \mathbf{p})$:

$$\begin{aligned} P_{\text{joint}}(\mathbf{q}, \mathbf{p}) &= \frac{1}{Z} \exp(-U(\mathbf{q})/T) \exp(-K(\mathbf{p})/T) \\ &= \frac{1}{Z} \exp(-H(\mathbf{q}, \mathbf{p})/T) \end{aligned} \tag{6}$$

where $H(\mathbf{q}, \mathbf{p}) = U(\mathbf{q}) + K(\mathbf{p})$.
Note $H(\mathbf{q}, \mathbf{p})$ the energy function for the joint state $(\mathbf{q}, \mathbf{p})$ distribution.

Properties of Hamiltonian dynamics

- Hamiltonian remains constant.
- Define $T_s : (\mathbf{q}(t), \mathbf{p}(t)) \rightarrow (\mathbf{q}(t+s), \mathbf{p}(t+s))$, the arrow represents the evolution of the dynamics. The mapping T_s is reversible.
- Hamiltonian dynamics preserves volume in the $(\mathbf{q}, \mathbf{p})$ space, i.e. the image of T_s from some region R has the same volume as region R.

Hamiltonian Dynamics and Hamiltonian Monte Carlo

Idea: Construct a Hamiltonian and Markov chain that make use of these properties.

Illustration: $U(q) = \frac{q^2}{2}$

Transfer of 'Energy' between $U(q)$ and $P(q)$ after giving particle some random momentum, a normal is commonly used.

The 'Ideal' Algorithm

1. initial $\mathbf{q}$
2. $\mathbf{p} \sim \mathcal{MN}(\mathbf{0}, \mathbf{M})$
3. Given $(\mathbf{q}, \mathbf{p})$, simulate Hamiltonian dynamics for some time and obtain $(\mathbf{q}^*, \mathbf{p}^*)$.
4. Negate $\mathbf{p}^*$ to ensure reversibility.
5. Accept $(\mathbf{q}^*, \mathbf{p}^*)$ as the next step in the Markov chain with probability 1.

World is not ideal: LeapFrog

Can not solve analytically. S: Discretize and pick L and ϵ.

$$p_i(t+\epsilon/2) = p_i(t) - (\epsilon/2)\frac{\partial U}{\partial q_i}(\mathbf{q}(t)) \tag{7}$$

$$q_i(t+\epsilon) = q_i(t) - (\epsilon)\frac{p_i(t+\epsilon/2)}{m_i} \tag{8}$$

$$p_i(t+\epsilon) = p_i(t+\epsilon/2) - (\epsilon/2)\frac{\partial U}{\partial q_i}(\mathbf{q}(t+\epsilon)) \tag{9}$$

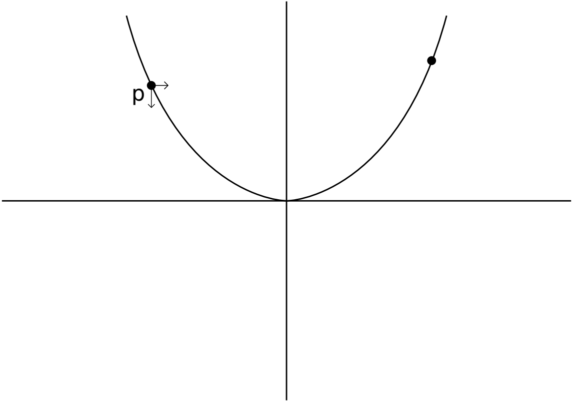
Iterating over this process L times simulates the dynamics for a time of $L\epsilon$.

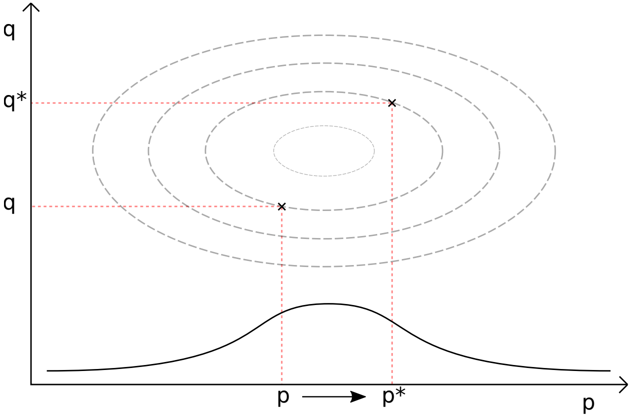
HMC Algorithm

Use Metropolis to correct the approximation error made in Leapfrog.

1. Select an initial $\mathbf{q}$.
2. $\mathbf{p} \sim \mathcal{MN}(\mathbf{0}, \mathbf{M})$
3. Given $(\mathbf{q}, \mathbf{p})$, simulate Hamiltonian dynamics using the leapfrog for L steps with ϵ to find $(\mathbf{q}^*, \mathbf{p}^*)$.
4. Negate $\mathbf{p}^*$ to ensure reversibility.
5. Accept $(\mathbf{q}^*, \mathbf{p}^*)$ as the next step in the Markov chain with probability M given below, otherwise accept $(\mathbf{q}, \mathbf{p})$ as next state.

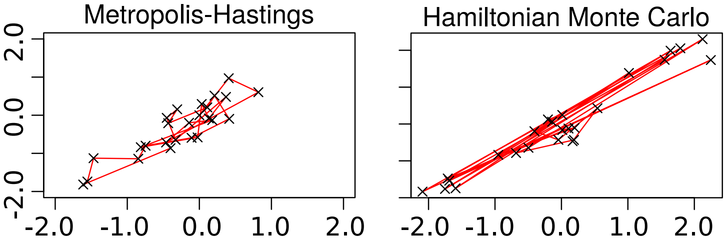
$$\begin{aligned} M &= \min\{1, \exp(-H(\mathbf{q}^*, \mathbf{p}^*) + H(\mathbf{q}, \mathbf{p}))\} \\ &= \min\{1, \exp(-U(\mathbf{q}^*) + U(\mathbf{q}) - K(\mathbf{p}^*) + K(\mathbf{p}))\} \end{aligned} \tag{10}$$


p


q
q*
q
p
p*
p

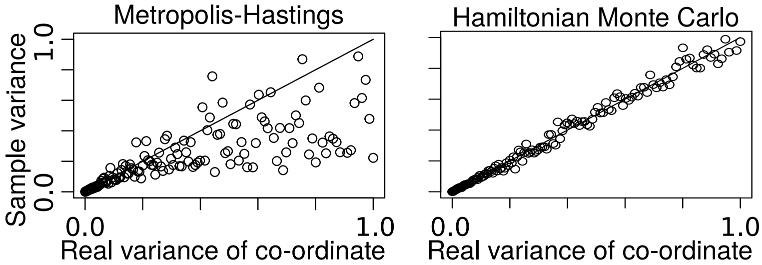
Comparison between HMC and MH

- 25 samples simulated from a bivariate Gaussian with marginal mean 0, $\sigma = 1$ and correlation of 0.95
- HMC: $\epsilon = 0.20$ and $L = 25$. Rejection Rate$= 0$
- MH: uniform proposal with $U[-0.25, 0.25]$ around the current state with thinning of 25 samples. Rejection Rate$= 0.4$.



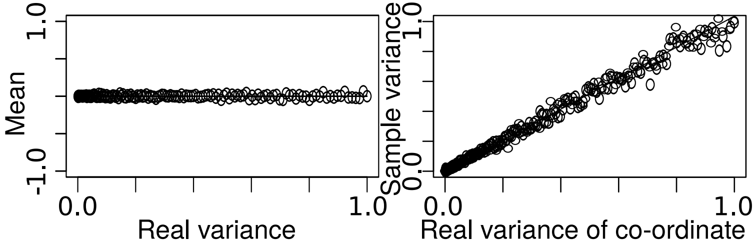
Multivariate Gaussian (150 dimensions): HMC, MH

- n samples from a 150-dimensional Gaussian
- HMC: $\epsilon = 0.20$ and $L = 25$
- MH: uniform proposal with $U[-0.25, 0.25]$ around the current state with thinning of 25 samples.



NUTS

- No U-Turn sampling modelled through *RStan* package
- 150-dimensional Gaussian simulated over 5000 samples



Graphical presentation: abcHMC

- Simulate samples from 2 bivariate Gaussian mixtures whose densities represent the letters of the alphabet
- Mixture models fitted automatically using EM algorithm and an image dataset for 52 letters and '!' and '?'.
- Equal weighting on each letter of a word.
- Simulate from the samples using using HMC and MH.
- abcHMC package can be found on Github.