Hamiltonian Monte Carlo and Langevin Monte Carlo

Kuhwan Jeong [1]

[1] Department of Statistics, Seoul National University, South Korea

July, 2018

Introduction

- *Hamiltonian Monte Carlo* (HMC) suppresses the local random walk behavior in the Metropolis-Hastings algorithm, thus allowing it to move much more rapidly through the target distribution.
- The target density $P(x)$ is augmented by an independent distribution $P(p)$ where p is a momentum variable, thus defining a joint distribution $P(x,p) = P(x)P(p)$.
- Auxiliary variable p is introduced only to enable the algorithm to move faster through the parameter space.
- HMC is a Metropolis-Hastings algorithm which proposes (x^*, p^*) satisfying $P(x,p) \approx P(x^*, p^*)$ using Hamiltonian dynamics.

Hamilton's Equations

- Let $x(t)$ be a location and $p(t)$ be a momentum at time t.
- For each location the object takes potential energy $U(x)$.
- For each momentum there is associated kinetic energy $K(p)$.
- The total energy of the system is constant and known as the Hamiltonian

$$H(x,p) = U(x) + K(p).$$

- The time evolution of the system is uniquely defined by Hamilton's equations:

$$\begin{aligned} \frac{dp}{dt} &= -\frac{\partial H}{\partial x} = -\frac{\partial U(x)}{\partial x}, \\ \frac{dx}{dt} &= \frac{\partial H}{\partial p} = \frac{\partial K(p)}{\partial p}. \end{aligned}$$

Discretizing Hamilton's Equations

- The Hamiltonian equations describe an object's motion in time.
- In order to simulate Hamiltonian dynamics numerically on a computer, it is necessary to approximate the Hamiltonian equations by discretizing time.

(1) Euler's Method

$$\begin{aligned} p_i(t+\delta) &= p_i(t) + \delta \frac{dp_i}{dt}(t) = p_i(t) - \delta \frac{\partial U}{\partial x_i}(x(t)), \\ x_i(t+\delta) &= x_i(t) + \delta \frac{dx_i}{dt}(t) = x_i(t) + \delta \frac{\partial K}{\partial p_i}(p(t)). \end{aligned}$$

(2) Modified Euler's Method

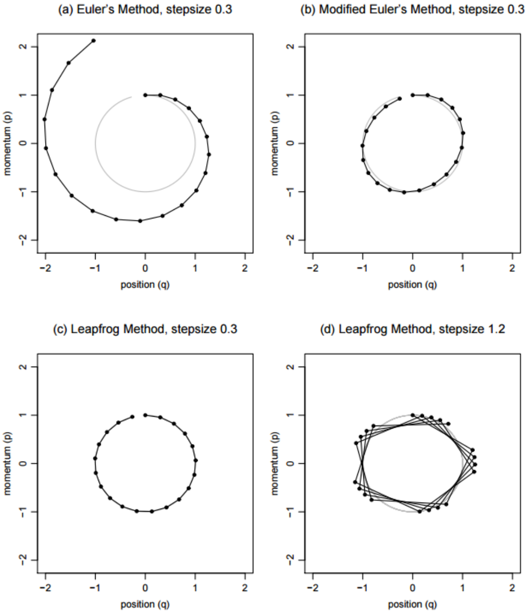
$$\begin{aligned} p_i(t+\delta) &= p_i(t) - \delta \frac{\partial U}{\partial x_i}(x(t)), \\ x_i(t+\delta) &= x_i(t) + \delta \frac{\partial K}{\partial p_i}(p(t+\delta)). \end{aligned}$$

Discretizing Hamilton's Equations

(3) The Leapfrog Method

$$
\begin{aligned}
p_i(t+\delta/2) &= p_i(t) - (\delta/2)\frac{\partial U}{\partial x_i}(x(t)),\\
x_i(t+\delta) &= x_i(t) + \delta\,\frac{\partial K}{\partial p_i}(p(t+\delta/2)),\\
p_i(t+\delta) &= p_i(t+\delta/2) - (\delta/2)\frac{\partial U}{\partial x_i}(x(t+\delta)).
\end{aligned}
$$

Discretizing Hamilton's Equations



Hamiltonian Monte Carlo

- We wish to sample d-dimensional x from

$$P(x) = \frac{1}{Z}\exp(-U(x)).$$

- We introduce a d-dimensional auxiliary variable p such that

$$p \sim \mathcal{N}(\mathbf{0}, M)$$

  where M is a symmetric, positive-definite matrix.
- The joint distribution of x and p is proportional to

$$\begin{aligned} P(x,p) &\propto \exp(-U(x))\exp\big(-p^T M^{-1} p\big) \\ &= \exp\big(-U(x) - p^T M^{-1} p\big). \end{aligned}$$

- We define Hamiltonian function $H(x,p)$ and the kinetic energy $K(p)$ as

$$H(x,p) = U(x) + K(p), K(p) = p^T M^{-1} p/2.$$

Hamiltonian Monte Carlo

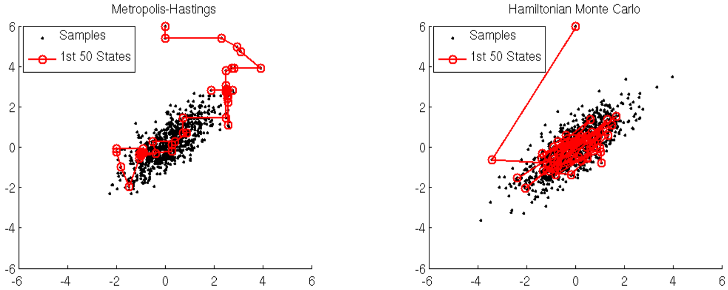
1. Set $m = 0$.
2. Generate an initial position $x^{(0)}$.
3. Repeat until $m = M$:
    1. Set $m = m + 1$.
    2. Sample $p_0 \sim \mathcal{N}(\mathbf{0}, M)$.
    3. Set $x_0 = x^{(m-1)}$.
    4. Starting from (x_0, p_0), do Leapfrog updates for L steps with stepsize δ to obtain (x^*, p^*).
    5. Calculate the Metropolis acceptance probability :

    $$\alpha = \min\left(1, \exp\big(-H(x^*, p^*) + H(x_0, p_0)\big)\right).$$

    6. Sample $u \sim U(0, 1)$:
        - If $u \leq \alpha$, set $x^m = x^*$.
        - Otherwise, set $x^m = x^{m-1}$.

Simulation : Bivariate Normal distribution

- Bivariate normal distribution with $\rho = 0.8$.



Simulation : 100-dimensional Normal distribution

- 100-dimensional

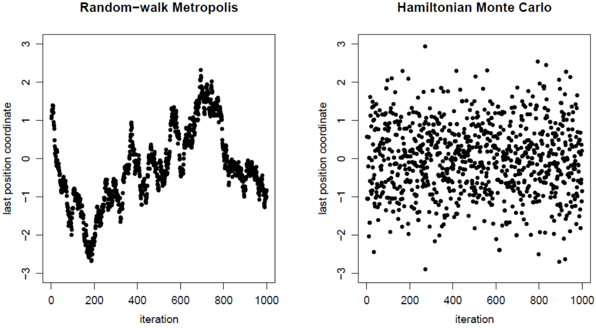


Figure 6: Values for the variable with largest standard deviation for the 100-dimensional example, from a random-walk Metropolis run and an HMC run with $L = 150$. To match computation time, 150 updates were counted as one iteration for random-walk Metropolis.

Langevin Monte Carlo

- A special case of HMC arises when the trajectory used to propose a new state consists of only a single leapfrog step.
- Suppose that $K(p) = \frac{1}{2}p^T p$.
- Given the current x, we sample $p \sim \mathcal{N}(\mathbf{0}, I)$, and then propose x^* and p^* as follows:

$$\begin{aligned} x_i^* &= x_i - \frac{\delta^2}{2}\frac{\partial U}{\partial x_i}(x) + \delta p_i, \\ p_i^* &= p_i - \frac{\delta}{2}\frac{\partial U}{\partial x_i}(x) - \frac{\delta}{2}\frac{\partial U}{\partial q_i}(q^*). \end{aligned} \quad (1)$$

- We accept x^* with a probability

$$\min\left[1, \exp\left\{-U(q^*) + U(q) - \frac{1}{2}\sum_i (p_i^* - p_i^2)\right\}\right].$$

- We can rewrite (1) as

$$\Delta x = \frac{\delta^2}{2}\frac{\partial \log P(x)}{\partial x} + \delta p,$$

  which is the gradient accent update with an additive Gaussian noise.
- Langevin Monte Carlo (LMC) without an accept/resject step is also used.
- LMC explore the distribution via an inefficient random walk, just like random-walk Metropolis updates.

Reference

- https://theclevermachine.wordpress.com/2012/11/18/mcmc-hamiltonian-monte-carlo-a-k-a-hybrid-monte-carlo/

- Neal, Radford M. "MCMC using Hamiltonian dynamics." Handbook of Markov Chain Monte Carlo 2 (2011): 113-162.

- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). Bayesian data analysis. Chapman and Hall/CRC.