Markov Chain Monte Carlo using Hamiltonian Dynamics

Joshua James MacDonald

August 22, 2018

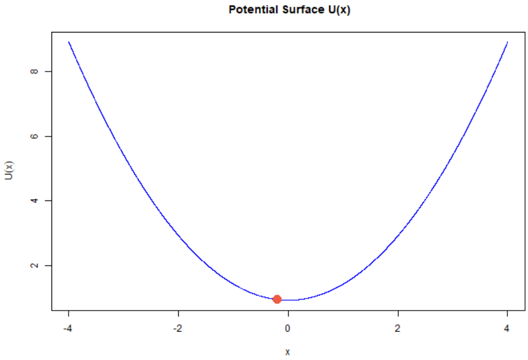
Aims

- Understand why Hamiltonian Monte Carlo works
- Tuning the HMC Algorithm
- Investigating Scaling with Dimension and Optimal Acceptance Rate

Hamiltonian Parameter Space

- A ball rolling around a potential surface
- At any time the ball has position x and momentum p
- Also has a mass m

A Simple Potential Surface



The Hamiltonian

Potential Energy,

$$U(x) = -\log\{\pi(x)\}$$

Kinetic Energy,

$$K(p; m) = \frac{1}{2}\frac{p^2}{m}$$

The Hamiltonian,

$$H(x, p; m) = U(x) + K(p; m)$$

Hamiltonian Equations

$$\frac{dx}{dt} = \frac{p}{m}$$

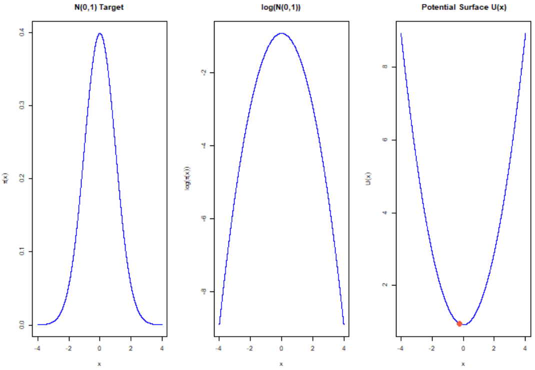
$$\frac{dp}{dt} = -\frac{dU}{dx}$$

Can be generalised to higher dimensions

$$\frac{d\mathbf{x}}{dt} = M^{-1}\mathbf{p}$$

$$\frac{d\mathbf{p}}{dt} = -\nabla U(\mathbf{x})$$

1-dimensional Gaussian Example



Approximating Hamiltonian Dynamics

- Hamiltonian Dynamics are approximated using a numerical discretisation.
- Simulate over time T, using L steps and stepsize ε
- When used in MCMC, gives acceptance probability

$$\alpha(x^*, p^*; x, p) = \min\left\{1, \exp(-(H(x^*, p^*; M) - H(x, p; M))\right\}$$

HMC Algorithm

- We have some position x.
- Draw a new momentum $p \sim \mathcal{N}(0, M)$.
- Approximate Hamiltonian Dynamics for time T in L steps, using stepsize ε.
- Proposes new position and momentum (x^*, p^*).
- Accept proposal with probability

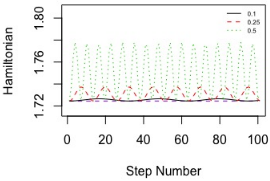
$$\alpha(x^*, p^*; x, p) = \min\left\{1, \exp(-(H(x^*, p^*; M) - H(x, p; M))\right\}$$

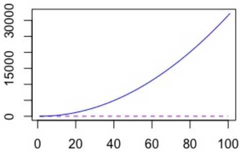
Choosing the Integration Time

- The distance moved in a prosposal can be directly controlled by T
- This doesn't affect the acceptance probability
- Hard to know what values of T will be sensible

Choosing a Stepsize

- Error in the Approximation is controlled by the stepsize.
- Approximation can become unstable if the chosen stepsize is too large





Choosing the Number of Steps

-
- Find a range of stepsizes which perform suitably

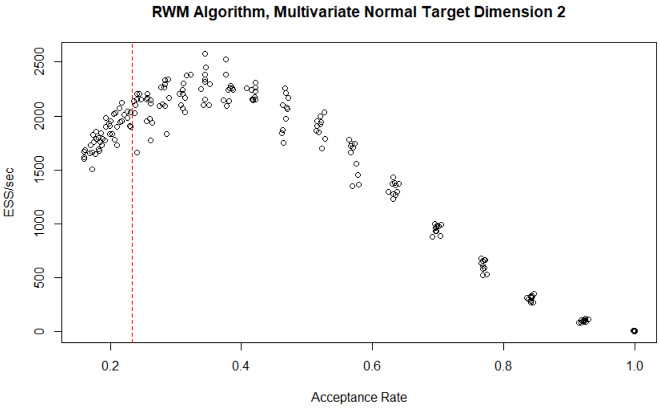
Scaling and Optimal Acceptance Rate

- The Performance of MCMC in high dimensions has become important with the rise of Big Data
- How far can we move with each proposal while keeping a reasonable acceptance rate?

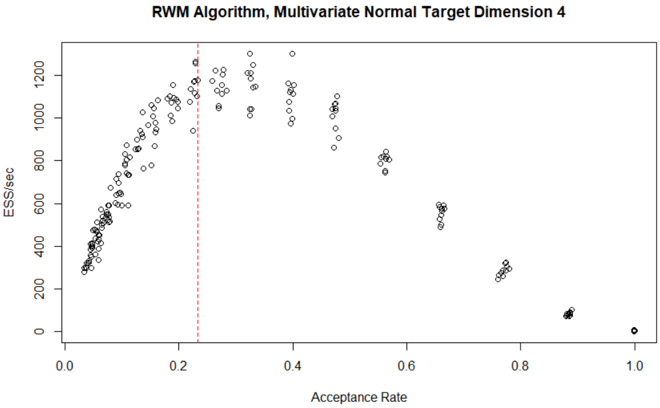
Scaling with Dimension

Algorithm	Scaling	Optimal Acceptance Rate ($d \to \infty$)
RWM	$d^{-1/2}$	23.4% (Roberts & Rosenthal 2001)
MALA	$d^{-1/3}$	57.4% (Roberts & Rosenthal 2001)
HMC	$d^{-1/4}$	65% (Beskos et. al. 2010)

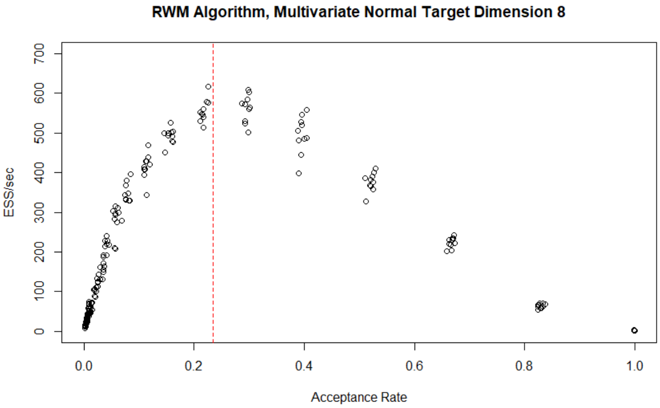
Optimal Acceptance Rate: RWM



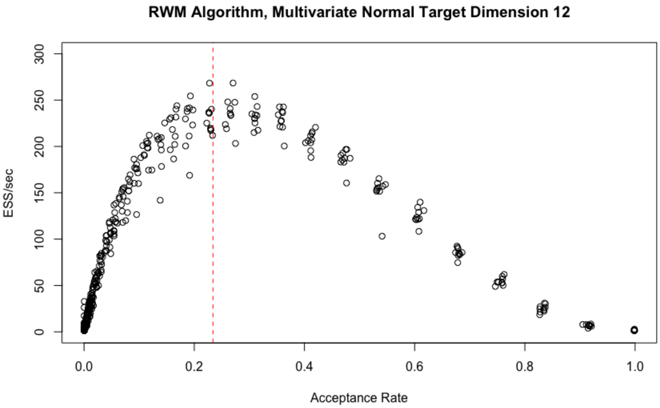
Optimal Acceptance Rate: RWM



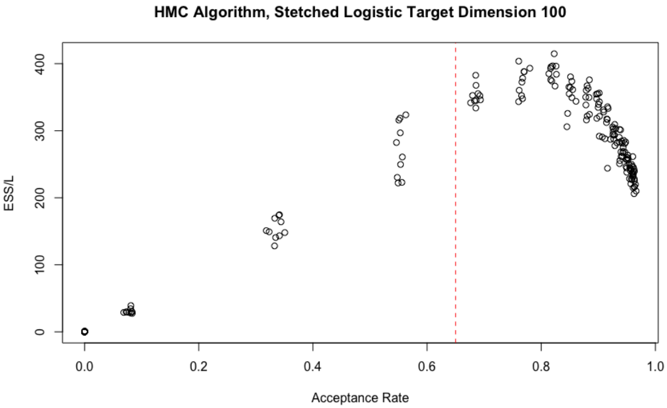
Optimal Acceptance Rate: RWM



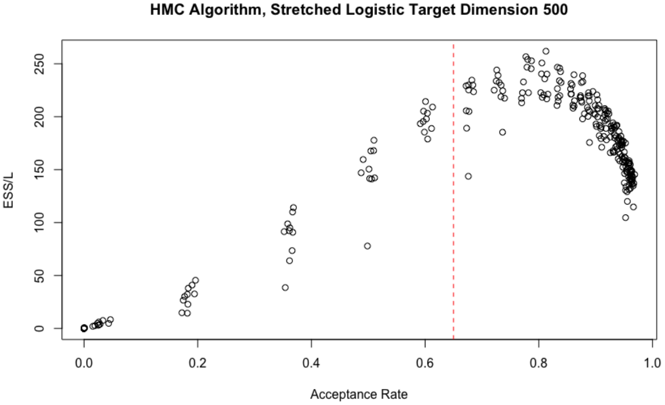
Optimal Acceptance Rate: RWM



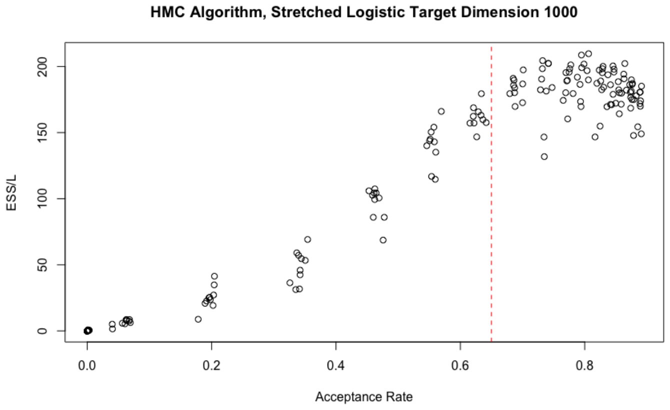
Optimal Acceptance Rate: HMC



Optimal Acceptance Rate: HMC



Optimal Acceptance Rate: HMC



Further Topics

- Reversibility
- Störmer-Verlet Approximation
- Partial Momentum Refreshment
- Relativistic Monte Carlo

Thank you for listening.

Any Questions?