Introduction To Hamiltonian Monte Carlo

Kota Takeda

Department of Mathematics

Kyoto University

Program

1. Introduction
2. Preliminary
3. Hamiltonian Monte Carlo
4. Demonstration
5. Discussion

1. Introduction

Introduction to today’s presentation and Hamiltonian Monte Carlo

Introduction

Almost all parts of this presentation is from

"A Conceptual Introduction to Hamiltonian Monte Carlo", Michael Betancourt .

This is **intuitive introduction** of Hamiltonian Monte Carlo.

Motivation

Computing expectations with respect to the posterior distribution in Bayesian inference

$$\mathbb{E}_\pi[f] = \int_Q f(q)\pi(q)dq$$

2. Preliminary

Preparing to constructing Hamiltonian Monte Carlo

Notation

- ✓ Sample space $Q = \mathbb{R}^d$.
- ✓ Denote $M > 0$, if M is symmetric positive definite matrix.
- ✓ $N(0, \Sigma)$; for mean zero covariance Σ Gaussian distribution.
- ✓ $q \sim \pi$; q is distributed by π.
- ✓ Use "$\equiv$" for definition

Monte Carlo Method

Approximate distribution $\pi(q)$ with ***large samples***

$$\mathbb{E}_{\pi}[f] \approx \frac{1}{N} \sum_{n=1}^{N} f(q_n)$$

Sampling Problem

Define sample problem we are interested in

How do we make efficient sampling from given target distribution $\pi(q)$ *?*

Markov Chain Monte Carlo: MCMC

*Construct Markov chain that converges to target distribution by **Random proposal and Acceptance***

Desired transition kernel $\mathbb{T}(q'|q)$ to satisfy reversibility

$$\pi(q)\mathbb{T}(q'|q) = \pi(q')\mathbb{T}(q|q')$$

Random Walk Metropolis: RWM

One simple implementation of Metropolis-Hasting algorithm

- ✓ Require: target $\pi(q)$
- ✓ proposal covariance $\Sigma > 0$

1. Random proposal

   given q_n, draw $q_{prop} \sim N(q_n, \Sigma)$

2. Accept

   accept $q_{n+1} = q_{prop}$ with probability

   $\min\{1, \frac{\pi(q_{prop})}{\pi(q_n)}\}$

Issue of MCMC

Poor performance with ***high dimension*** *and* ***complex target distributions***

Hamilton Dynamics

Hamilton equation on phase space

- ✓ **preserve volume** in phase space(Liouville's Theorem)
- ✓ **preserve total energy** in phase space, which is Hamiltonian
- ✓ **time reversal symmetry**

$$\begin{cases} \dfrac{dq}{dt} = \dfrac{dH}{dp} \\ \dfrac{dp}{dt} = -\dfrac{dH}{dq} \end{cases}$$

3. Hamiltonian Monte Carlo

Constructing Hamiltonian Monte Carlo

Hamiltonian Monte Carlo: HMC

Q. How do we make efficient sampling from given target distribution $\boldsymbol{\pi_U(q)}$ *?*

A. One approach is using **geometric information** of target and constructing **conservative transition kernel** by Hamilton flow.

Information of Gradient

Using geometric information of target density

1. Consider $\pi_U(q) = e^{-U(q)}$, where $U(q) \equiv -\log(\pi_U(q))$

2. But gradient $\frac{\mathrm{dU}}{\mathrm{dq}}$ pulls us the mode of density!

3. $\rightarrow$ Need to introduce momentum p

Expand Sample Space

Expand sample space to phase space,

We can always gain sample q by projection(marginalization).

1. Expand to phase space $q \rightarrow (q, p)$ with p
2. Choose conditional distribution $\pi_K(p|q)$
3. Lift $\pi_U(q)$ to $\pi_H(q, p) \equiv \pi_K(p|q)\pi_U(q)$

Choice of Kinetic Energy

To define conditional distribution of momentum $\pi_K(p|q)$, a user choose kinetic energy.

In simple case, let $K(q,p) = \frac{1}{2} p^T M^{-1} p$.

1. Choose Kinetic Energy $K(q,p)$

2. Conditional distribution of momentum determined by

   $\pi_K(p|q) \equiv e^{-K(q,p)}$

Hamiltonian

Hamiltonian H and canonical distribution π_H are defined as below.

1. $H(q,p) \equiv K(q,p) + U(q)$

2. $\pi_H(q,p) \equiv \pi_K(p|q)\pi_U(q) = e^{-H(q,p)}$

Symplectic integrator

Scheme exactly preserving volume

- ✓ Assume: $K(q,p) \equiv \frac{1}{2}p^T M^{-1} p$ with Mass matrix $M > 0$,
- ✓ $U(q)$ is differentiable.
- ✓ Require: step size $\varepsilon > 0$

$$\begin{cases} p_{n+\frac{1}{2}} = p_n - \frac{\varepsilon}{2}\frac{dU}{dq}(q_n) \\ q_{n+1} = q_n + \varepsilon M^{-1} p_{n+\frac{1}{2}} \\ p_{n+1} = p_{n+\frac{1}{2}} - \frac{\varepsilon}{2}\frac{dU}{dq}(q_{n+1}) \end{cases}$$

$$\varphi_\varepsilon(q_n, p_n) \equiv q_{n+1}, p_{n+1}$$

Numerical Hamilton Flow

Define numerical Hamilton flow by symplectic integrator on previous page and define L times composition

1. $\varphi_\varepsilon(q_n, p_n) \equiv q_{n+1}, p_{n+1}$

2. ${\varphi_\varepsilon}^L \equiv \varphi_\varepsilon \circ \cdots \circ \varphi_\varepsilon$ for integer L.

HMC Algorithm

Hybrid of deterministic and stochastic transitions

- ✓ $H(q,p) \equiv \frac{1}{2}p^T M^{-1} p + U(q)$
- ✓ Require: $M > 0, L \in \mathbb{N}, \varepsilon > 0$

1. Energy Lift

   given q_n, draw $p_n \sim N(0, M)$

2. Hamilton flow

   $q_{prop}, p_{prop} = \varphi_\varepsilon{}^L(q_n, p_n)$

3. Accept

   accept $q_{n+1} = q_{\text{prop}}$ with probability

   $\min\{1, \exp(H(q_n, p_n) - H(q_{prop}, -p_{prop}))\}$

Conceptual Animation of HMC Algorithm

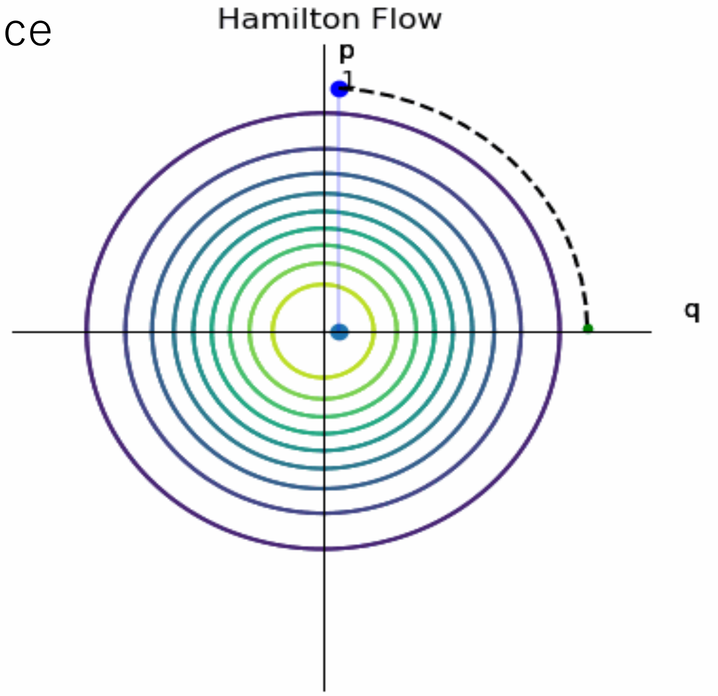
1. *Energy Lift*
2. *Hamilton flow*
3. *(Accept)*

✓ Sample space $Q = \mathbb{R}^1$

✓ target $\pi_U(q) = e^{-\frac{1}{2}q^2}, \pi_K(p|q) = e^{-\frac{1}{2}p^2}$

$$\Longrightarrow H(q,p) = \frac{1}{2}q^2 + \frac{1}{2}p^2$$

Phase space



Advantage of Hamiltonian Monte Carlo

- ✓ Rich Theoretical Support

  effective for wider class of target than non-gradient method

- ✓ Computational Efficiency

  Fast exploration and large acceptance probability

4. Demonstration

Demonstration of efficient Hamiltonian Monte Carlo compared with Random Walk Metropolis

Strongly Nonlinear Banana Gaussian

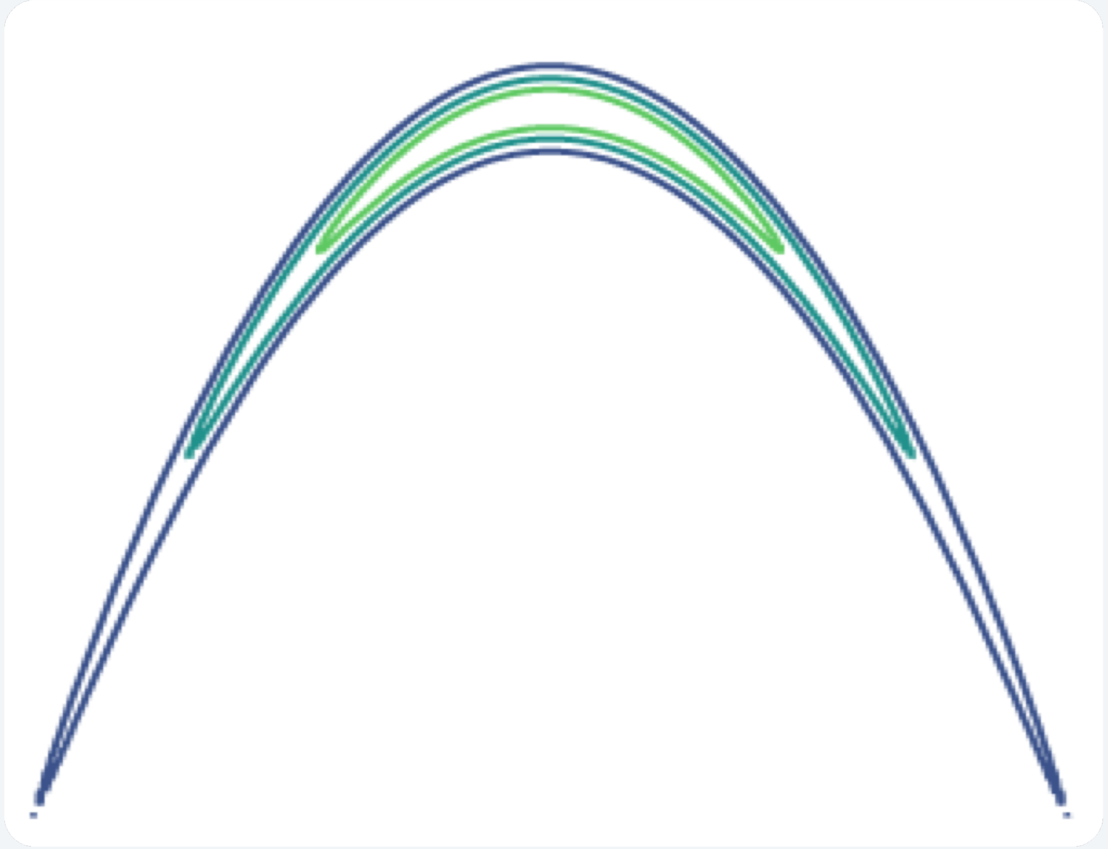
Test Target distribution is

Strongly banana gaussian

- ✓ Sample space $Q = \mathbb{R}^2$
- ✓ target $\pi_U(q_1, q_2) = g \circ \psi_{b=0.1}(q_1, q_2)$

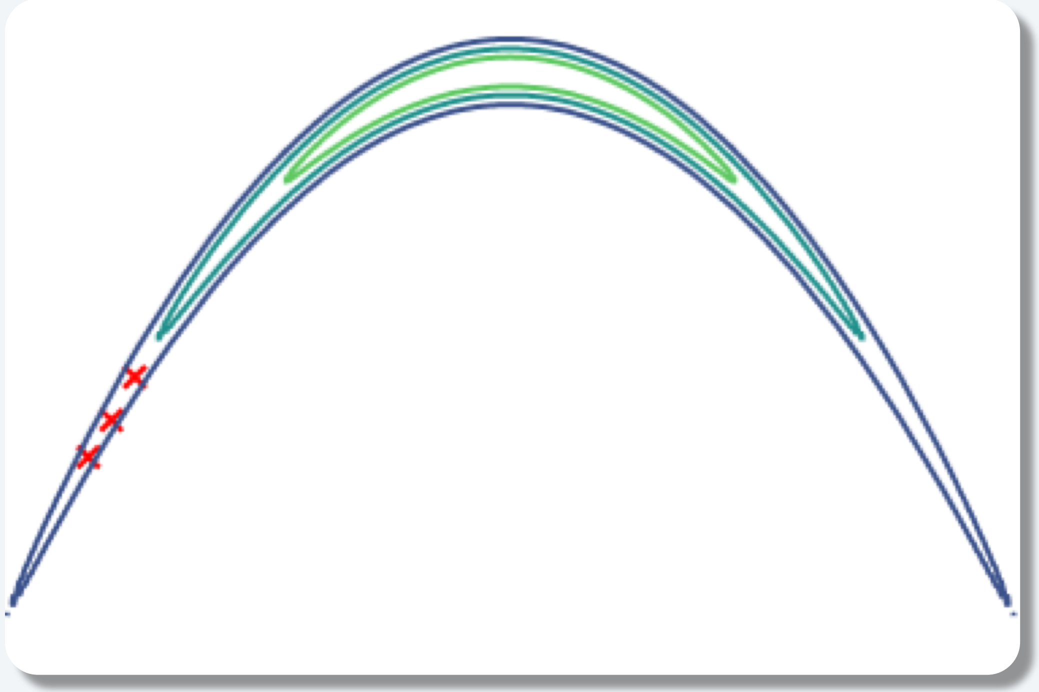
where

- ✓ $g(q_1, q_2) = e^{-\frac{1}{200}q_1^2 - \frac{1}{2}q_2^2}$
- ✓ $\psi_b : (q_1, q_2) \mapsto (q_1, q_2 + \boldsymbol{bq_1^2} - 100b)$

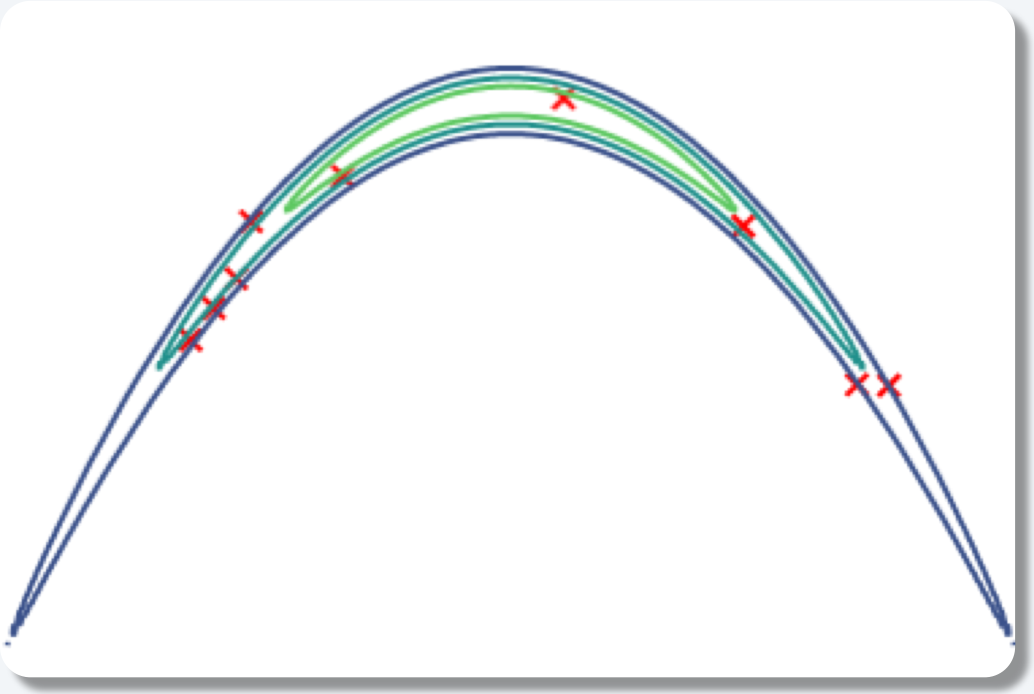


RWM vs HMC after 10 iterations

RWM: $\Sigma = 2I$



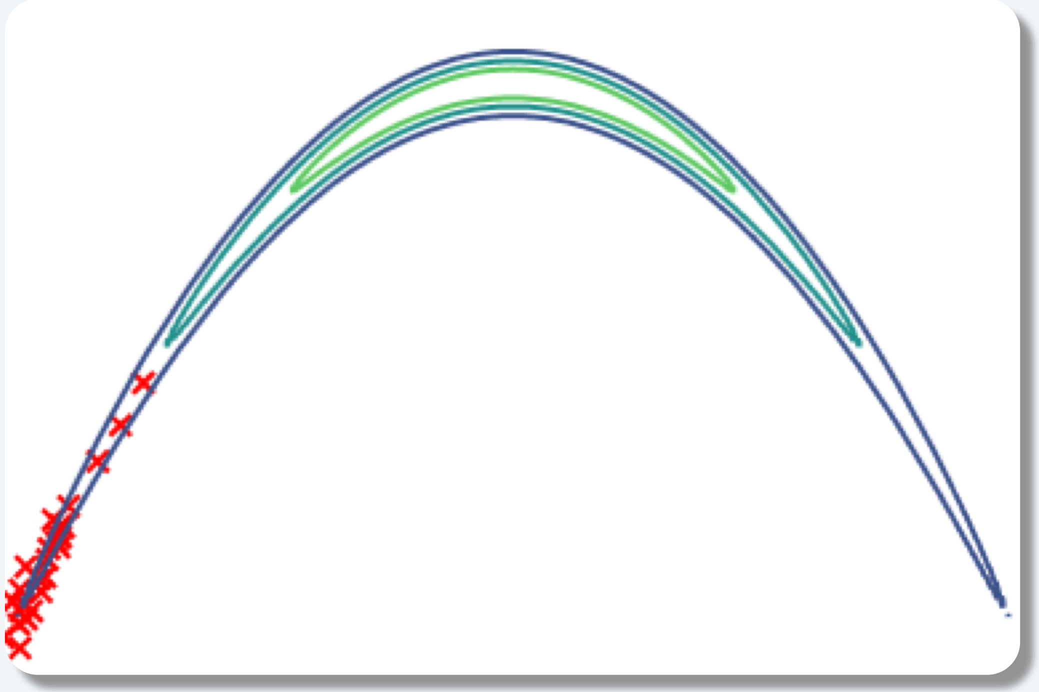
HMC: $\varepsilon = 0.5, L = 10, M = I$

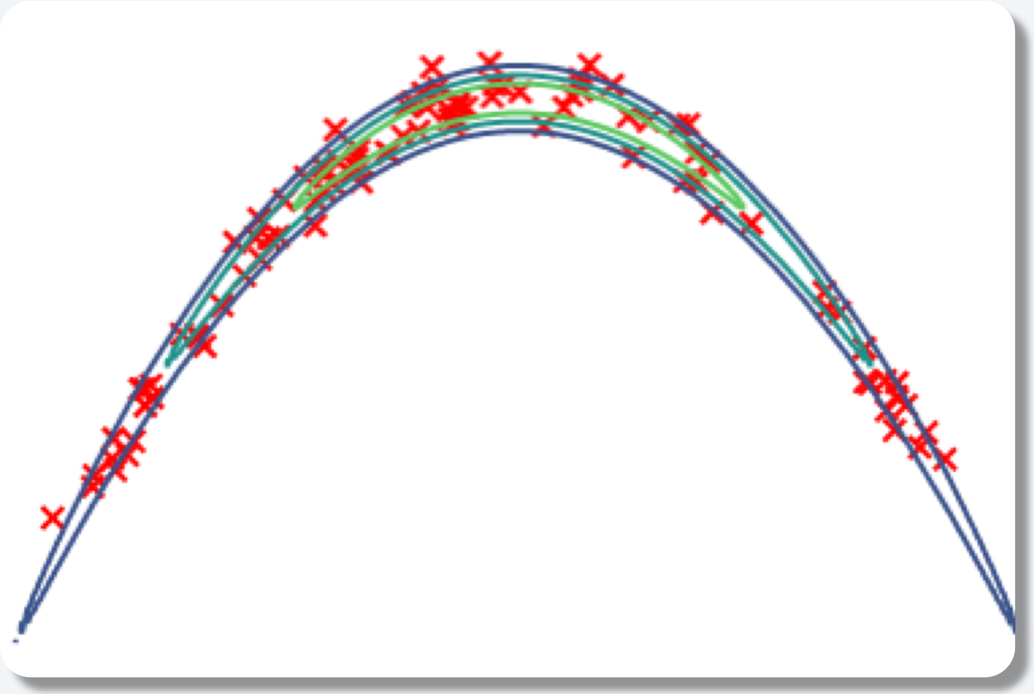


RWM vs HMC after 100 iterations

RWM: $\Sigma = 2I$

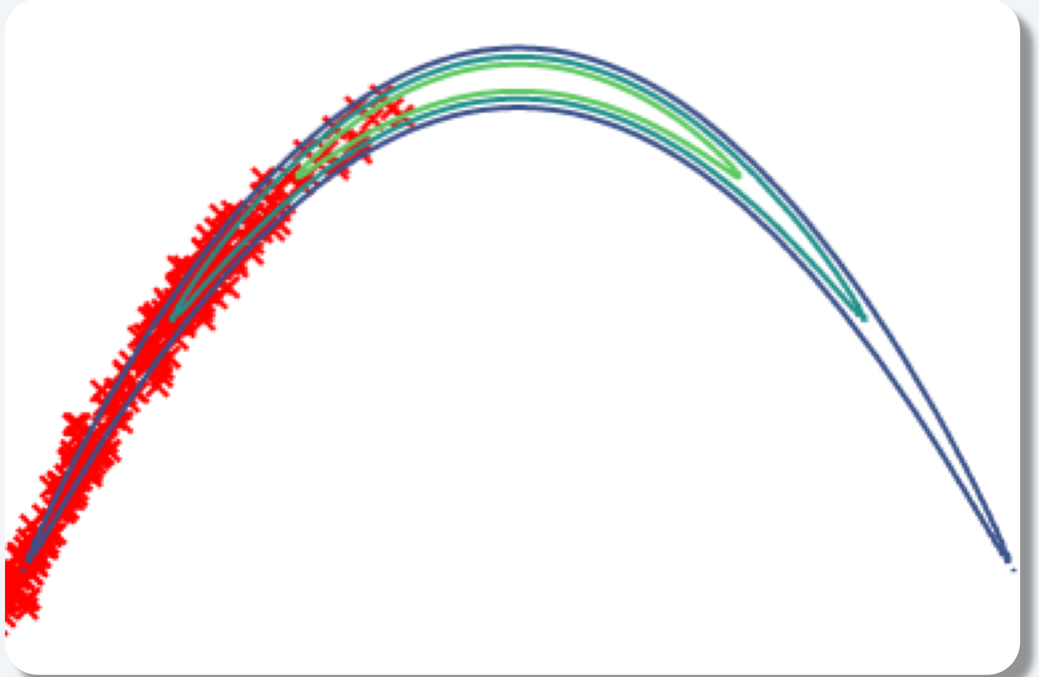
HMC: $\varepsilon = 0.5, L = 10, M = I$



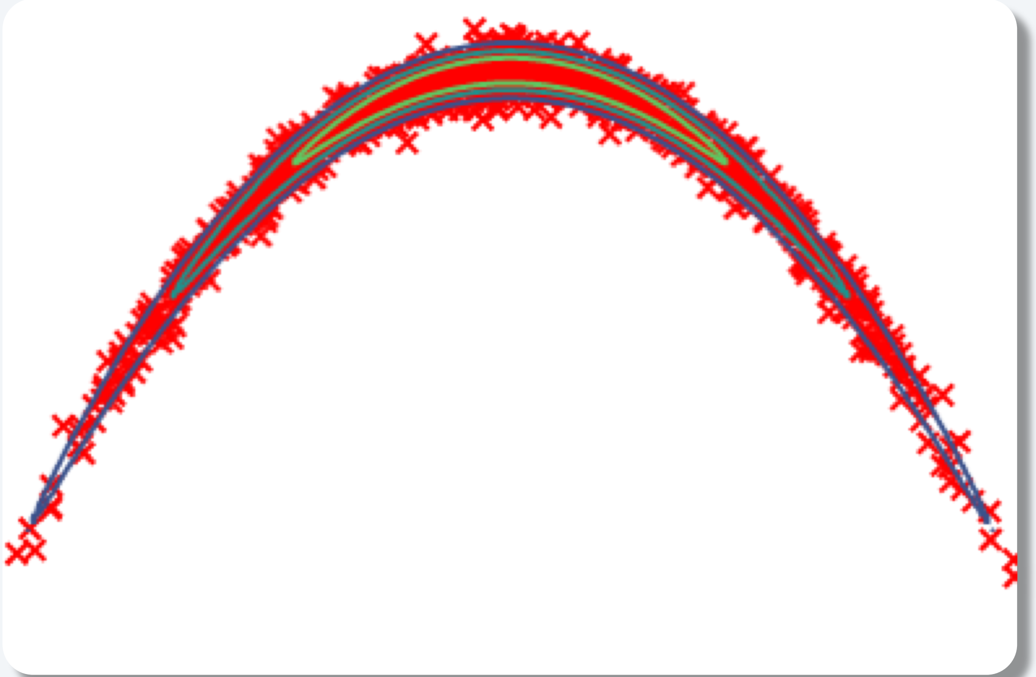


RWM vs HMC after 1000 iterations

RWM: $\Sigma = 2I$



HMC: $\varepsilon = 0.5, L = 10, M = I$



Stats*

RWM: $\Sigma = 2I$

Try Sample	Time*(ms)	Acceptance Probability
10	3.02	0.300
100	25.8	0.260
1000	179	0.288

HMC: $\varepsilon = 0.5, L = 10, M = I$

Try Sample	Time*(ms)	Acceptance Probability
10	7.43	0.900
100	34.3	0.970
1000	474	0.940

*Not guaranteed value, **just a reference**.
*Time is measured by jupyter magic command `%%time`.

5. Discussion

Discussion about future work or application of Hamiltonian Monte Carlo

Future work

- ✓ Studying mathematical guarantee and guideline
- ✓ Adaptive tuning of parameters
- ✓ Selecting Integrators
- ✓ Generalizing to infinite-dimensional sample space
- ✓ Introducing inverse temperature

Discussion

- ✓ Particle Filter
- ✓ Variational method
- ✓ Inverse problem
- ✓ “Estimation of Hydraulic Conductivity from Steady State Seepage Flow Using Hamiltonian Monte Carlo”- 01/02/2021

For Your Information…

This is pinned tweet of Michael Betancourt.

"Remember that using Bayes' Theorem doesn't make you a Bayesian. Quantifying uncertainty with probability makes you a Bayesian." - Michael Betancourt

https://twitter.com/betanalpha/status/817012860643635204

References

- ✓ "A Conceptual Introduction to Hamiltonian Monte Carlo"
- ✓ "The Geometric Foundations of Hamiltonian Monte Carlo"
- ✓ "The Adaptive proposal distribution for Random Walk Metropolis Algorithm" – *only for banana Gaussian*
- ✓ "Estimation of Hydraulic Conductivity from Steady State Seepage Flow Using Hamiltonian Monte Carlo" - http://soil.en.a.u-tokyo.ac.jp/jsidre/search/PDFs/20/%5B1-52%5D.pdf

Documents

- ✓ Detail documents on my site
  https://kotatakeda.github.io/math/2021/01/03/hamiltonian-monte-carlo.html