HAMILTONIAN MONTE CARLO METHOD

Nando Tezoh Franky kévin

AMMI-Ghana

December 17, 2019

OUTLINE

- Hamiltonian

$$H(p, q) = K(p) + V(q). \tag{1}$$

- Example

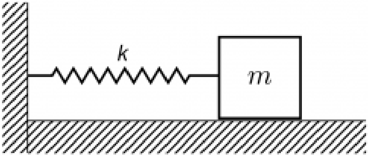


Figure: (https://physics.stackexchange.com/questions/340740/damped-harmonic-oscillator-scenario)

$$H(x, p_x) = \frac{p_x^2}{2m} + \frac{1}{2}kx^2, \qquad Where \quad x = q, p_x = p.$$

- Hamiltonian

$$H(p, q) = K(p) + V(q). \tag{1}$$

- Example

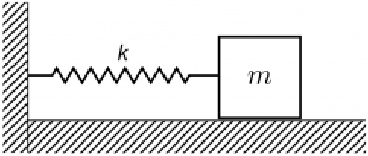


Figure: (https://physics.stackexchange.com/questions/340740/damped-harmonic-oscillator-scenario)

$$H(x, p_x) = \frac{p_x^2}{2m} + \frac{1}{2}kx^2, \qquad Where \quad x = q, p_x = p.$$

Equations

(2) represents the equations of Hamilton in the case of one-dimensional space.

$$p = m\frac{dq}{dt} \quad , \frac{dp}{dt} = -\frac{\partial V}{\partial q} \tag{2}$$

Equations of Hamilton of the First figure

$$p = m\frac{dx}{dt} \quad , \frac{dp}{dt} = -\frac{\partial V}{\partial x} = -kx \tag{3}$$

- Canonical ensemble= System in equilibrium with a heat bath at Temperature T.
- A canonical distribution is given by:

$$\mathbb{P}(x) = \frac{1}{Z} \exp(-\beta U(x)). \tag{4}$$

- **Example of Canonical System**: A bottle of water inside the sea.

Figure: http://www.apeuk.org/effects-oceans-plastics/

What is MCMC

- Markov Chain Monte Carlo (MCMC) are tools used for sampling from a Posteriori distribution.
- MCMC = Hamiltonian Monte Carlo (HMC) and Metropolis Monte Carlo (MMC).

Why do we need to sample?

$$\mathbb{E}_p[f(x)] = \int f(x)\,\mathbb{P}(x)dx \tag{5}$$

$$\mathbb{E}_p[f(x)] = \sum_x f(x)\,\mathbb{P}(x) \tag{6}$$

Equations (5) and (6) are sometimes difficult to solve analytically or to get an exact solution.

HMC Steps

- Suppose $q \sim f(q)$ distribution.
- By adding a momentum variable p, we then have the join distribution $f(q, p)$.
- Using Maximum a posterior Estimation, one can write $f(q, p)$ as:

$$f(q, p) = f(p|q)f(q). \tag{7}$$

- By applying the log onto equation (7), one has:

$$\begin{aligned} H(q, p) &= -\log(f(\mathbf{q}, \mathbf{p})) \\ &= -logf(\mathbf{p}|\mathbf{q}) - \log f(q) \\ &= K(p) + V(q), \end{aligned}$$

Contd

- $K(p)$ is generally considered to follow a normal distribution $\mathcal{N}(0, M)$, with M the mass diagonal Matrix which is positive definite.
- **We should note as well that $f(q, p)$ follows a canonical distribution.**
- Kinetic energy is defined as:

$$K(p) = \frac{\mathbf{p}^T M^{-1} \mathbf{p}}{2},$$

Contd

- To compute p and q, we use Hamiltonian equations described as follows (Here we are in multi-dimensional space):

$$p = M\frac{dq}{dt} \quad , \tfrac{dp}{dt} = -\nabla_q V(q). \tag{8}$$

- **HMC simulates the Hamiltonian dynamics in order to get samples.**
- For non-physical systems, the position corresponds to the variable of interests. Potential energy is $-\log$ the probability density function and the momentum is introduced artificially.

Properties

- **Conservation of Hamiltonian.** It is invariant during the dynamic, that means: $\frac{dH}{dt} = 0$.
- **Volume conservation.** The Hamiltonian dynamics preserves its volume in (q, p) space.
- **Reversibility.**

Contd

Implications on HMC

- Reversibility enables during the MCMC updates to leave the distribution invariant.
- Volume conservation for MCMC since we don't need change of the volume in the acceptance probability for Metropolis update.
- The acceptance probability is one if the Hamiltonian does not change.

LM

- Leapfrog Method is a modified version of Euler method which allow us to get an approximate solution of a differential equations.
- For the case of the HMC, one has:

$$p_i(t+\varepsilon/2) = p_i(t) - (\varepsilon/2)\frac{\partial V}{\partial q_i}(q_i(t)) \tag{9}$$

$$q_i(t+\varepsilon) = q_i + \varepsilon p_i(t+\varepsilon/2) \tag{10}$$

$$p_i(t+\varepsilon) = p_i(t+\varepsilon/2) - \frac{h}{2}\frac{\partial V}{\partial q_i}(q_i(t+\varepsilon)). \tag{11}$$

Contd

- Equations (9), (10) and (11) represent the Leapfrog step that enable to get the values of p and q.
- Other methods that could be used are Runge-Kutta, even Euler itself. Nevertheless, it has been proved that LM is the one performing the best.

Algorithm

Algorithm 1 :HMC Algorithm

Input: Starting position $q^{(1)}$ and step size ε

For $t = 1, 2, \cdots$

Resample momentum p

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

$(q_0, p_0) = (q^{(t)}, p^{(t)})$

Simulate discretization of Hamiltonian dynamics

$p_0 \leftarrow p_0 - \frac{\epsilon}{2}\nabla V(q_0)$.

for $i = 1$ to m do:

$q_i \leftarrow q_i - M^{-1}\epsilon p_{i-1}$

$p_i \leftarrow p_{i-1} - \epsilon\nabla V(q_i)$

end.

$p_m \leftarrow p_m - \frac{\epsilon}{2}\nabla V(q_m)$

$(\hat{q}, \hat{p}) = (q_m, p_m)$

Metropolis-Hastings correction:

$u \sim Uniform[0, 1]$

$\rho = \exp(H(\hat{q}, \hat{p}) - H(q^{(t)}, p^{(t)}))$

if $u < min(1, \rho)$, then $q^{(t+1)} = \hat{q}$

Stochastic HMC

- Stochastic Gradient HMC. It is when we take into consideration the noisy effects, the Gradient becomes:

$$\nabla \tilde{V}(q) \simeq \nabla V(q) + \mathcal{N}(0, v(q)),$$

  with v, the covariance of the stochastic gradient noise which can depend on the current model parameters and sample size.
- Stochastic Gradient HMC with friction, it is version of stochastic HMC where a frictional term has been added to decrease the energy $H(q, p)$, Hence reducing noise influence.

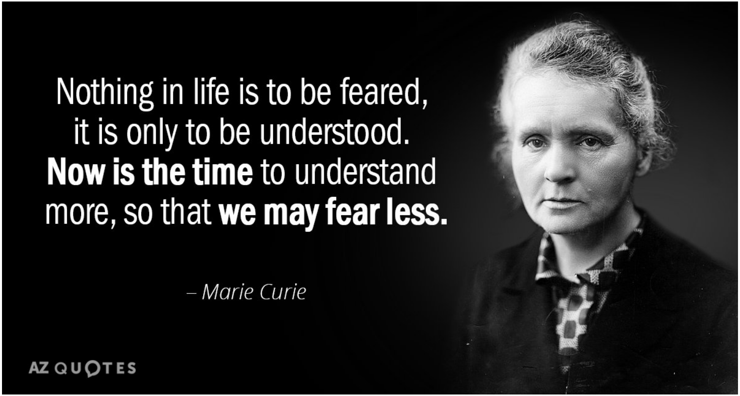
Cons

- Required gradient computation of the potential energy in order to simulate the Hamiltonian dynamics of the systems.
- That gradient computation is infeasible in problems involving large dataset.
- To handle the issues of gradient computation, some methods as SGHMC has been introduced (replace the full gradient by a mini-batch stochastic gradient).
- HMC is inefficient to sample from multimodal distributions.

Different application of HMC

- Computing volume in high dimensions
- Expectation Maximization
- Apply in Neural Networks to determining the set of observed and unobserved random variables for Boltzmann machines.
- Statistical Physics.
- Compute the eigenvalues and eigenvectors of some operators such the Hamiltonian.

Appendix


Nothing in life is to be feared,
it is only to be understood.
Now is the time to understand
more, so that **we may fear less.**
– Marie Curie
AZ QUOTES

Bibliography

[1] Nando de Freitas,Monte Carlo Simulation for Statistical Inference, Model Selection and Decision Making, University of British Columbia.

[2] Neal, Radford M. ”MCMC using Hamiltonian dynamics.” Handbook of Markov Chain Monte Carlo 2 (2011).

[3] Tianqi Chen, Emily Fox, and Carlos Guestrin. Stochastic gradient hamiltonian monte carlo.

[4] Tianqi Chen, Emily Fox, and Carlos Guestrin. Stochastic gradient hamiltonian monte carlo.

[5] Erlend Davidson's. Visualising hamiltonian monte carlo, 2018.

[6] Dieter W Heermann and K Binder. Monte Springer-Verlag Berlin Heidelberg, 2010. Carlo Simulation in Statistical Physics .