Non-Canonical Hamiltonian Monte Carlo Algorithm for Two Particles

Sihyun Yu
Department of Mathematics
KAIST
Yuseong-gu 291, Daejeon
yusihyunc@kaist.ac.kr

Hyunsoo Kim
Department of Physics
KAIST
Yuseong-gu 291, Daejeon
khszone02@kaist.ac.kr

Abstract

Probabilistic inference is one of the important concepts in machine learning. However, many problems still exist while sampling from high dimensional distribution or multimodal distributions. Many researchers try to solve this problem using existing physical concepts. We propose a new algorithm using non-canonical Hamiltonian dynamics and copied and pasted independent probabilistic distribution. In this algorithm, two particles interacts each other while sampling the one target distribution. This interaction might solve existing problems in probabilistic sampling since interaction can make the trace of each particle more broadly.

1 Introduction

2 Research Purpose

3 Theoretical Background

4 Results

4.1 Algorithmic Background

Consider the case we sample q from the distribution $\pi(q)$. From now, we consider the case as sampling the distribution $\pi(q_1, q_2) = \pi(q_1)\pi(q_2)$. (Copy and paste the target distribution) we assume q_1 and q_2 has the momentum, p_1 and p_2, respectively, with $p_1, p_2 \sim N(0, I)$. Then the Hamiltonian is defined by $H(q_1, q_2, p_1, p_2) = U(q_1) + U(q_2) + \frac{1}{2}{p_1}^T p_1 + \frac{1}{2}{p_2}^T p_2$. We observe what happens to this assumption in non-canonical Hamiltonian dynamics with $A \in \mathbb{R}^{4n \times 4n}$. We denote A as

$$A = \left[\begin{array}{c|c} E & F \\ \hline -F^T & G \end{array}\right]$$

where

$$E = \left[\begin{array}{c|c} E_1 & E_2 \\ \hline E_3 & E_4 \end{array}\right] \in \mathbb{R}^{2n \times 2n},$$

$$F = \left[\begin{array}{c|c} F_1 & F_2 \\ \hline F_3 & F_4 \end{array}\right] \in \mathbb{R}^{2n \times 2n},$$

$$G = \left[\begin{array}{c|c} G_1 & G_2 \\ \hline G_3 & G_4 \end{array}\right] \in \mathbb{R}^{2n \times 2n}.$$

Lemma 1. If $G_1 = G_4 = E_1 = E_4 = 0$ and $G_2^T + G_3$, $E_2^T + E_3$ are 0, (in other words, E, G are *anti-symmetric* matrix). Then the *non-canonical* Hamiltonian dynamics preserves H.

proof. The non-canonical hamiltonian dynamics is following:

$$\frac{d}{dt}\begin{bmatrix} q_1 \\ q_2 \\ p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} E_1 & E_2 & F_1 & F_2 \\ E_3 & E_4 & F_3 & F_4 \\ -F_1^T & -F_3^T & G_1 & G_2 \\ -F_2^T & -F_4^T & G_3 & G_4 \end{bmatrix} \begin{bmatrix} \nabla_{q_1} U(q_1) \\ \nabla_{q_2} U(q_2) \\ p_1 \\ p_2 \end{bmatrix}$$

Since $G_1 = G_4 = E_1 = E_4 = 0$,

$$\begin{bmatrix} E_1 & E_2 & F_1 & F_2 \\ E_3 & E_4 & F_3 & F_4 \\ -F_1^T & -F_3^T & G_1 & G_2 \\ -F_2^T & -F_4^T & G_3 & G_4 \end{bmatrix} \begin{bmatrix} \nabla_{q_1} U(q_1) \\ \nabla_{q_2} U(q_2) \\ p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} E_2 \nabla_{q_2} U(q_2) + F_1 p_1 + F_2 p_2 \\ E_3 \nabla_{q_1} U(q_1) + F_3 p_1 + F_4 p_4 \\ -F_1^T \nabla_{q_1} U(q_1) - F_3^T \nabla_{q_2} U(q_2) + G_2 p_2 \\ -F_2^T \nabla_{q_1} U(q_1) - F_4^T \nabla_{q_2} U(q_2) + G_3 p_1 \end{bmatrix}$$

Also, note that the poisson bracket is

$$\frac{dQ}{dt} = \sum_{i=1}^{2n} \frac{\partial Q}{\partial p_i} \frac{dp_i}{dt} + \sum_{i=1}^{2n} \frac{\partial Q}{\partial q_i} \frac{dq_i}{dt}$$

Therefore, putting previous evaluations,

$$\frac{dQ}{dt} = (\nabla_{p_1} Q)^T (-F_1^T \nabla_{q_1} U(q_1) - F_3^T \nabla_{q_2} U(q_2) + G_2 p_2) + (\nabla_{p_2} Q)^T (-F_2^T \nabla_{q_1} U(q_1) - F_4^T \nabla_{q_2} U(q_2)$$
$$+G_3 p_1) + (\nabla_{q_1} Q^T)(E_2 \nabla_{q_2} U(q_2) + F_1 p_1 + F_2 p_2) + (\nabla_{q_2} Q^T)(E_3 \nabla_{q_1} U(q_1) + F_3 p_1 + F_4 p_4)$$

This can be re-expressed as

$$\begin{gathered} \frac{dQ}{dt} = (\nabla_{p_1} Q)^T (-F_1^T \nabla_{q_1} U(q_1) - F_3^T \nabla_{q_2} U(q_2)) + (\nabla_{p_2} Q)^T (-F_2^T \nabla_{q_1} U(q_1) - F_4^T \nabla_{q_2} U(q_2)) \\ +(\nabla_{q_1} Q^T)(F_1 p_1 + F_2 p_2) + (\nabla_{q_2} Q^T)(F_3 p_1 + F_4 p_4) + (\nabla_{p_1} Q)^T G_2 p_2 + (\nabla_{p_2} Q)^T G_3 p_1 \\ +(\nabla_{q_1} Q^T) E_2 \nabla_{q_2} U(q_2) + (\nabla_{q_2} Q^T) E_3 \nabla_{q_1} U(q_1) \end{gathered}$$

.

Now, letting $Q = H$, we know

$$\begin{gathered} (\nabla_{p_1} H)^T (-F_1^T \nabla_{q_1} U(q_1) - F_3^T \nabla_{q_2} U(q_2)) + (\nabla_{p_2} H)^T (-F_2^T \nabla_{q_1} U(q_1) - F_4^T \nabla_{q_2} U(q_2)) \\ +(\nabla_{q_1} H^T)(F_1 p_1 + F_2 p_2) + (\nabla_{q_2} H^T)(F_3 p_1 + F_4 p_4) = 0 \end{gathered}$$

.

Therefore, it is enough to find the condition for

$$(\nabla_{p_1} H)^T G_2 p_2 + (\nabla_{p_2} H)^T G_3 p_1 + (\nabla_{q_1} H^T) E_2 \nabla_{q_2} U(q_2) + (\nabla_{q_2} H^T) E_3 \nabla_{q_1} U(q_1) = 0$$

.

By assumption of H, this is equivalent to

$${p_1}^T G_2 p_2 + {p_2}^T G_3 p_1 + (\nabla_{q_1} U(q_1))^T E_2 \nabla_{q_2} U(q_2) + (\nabla_{q_2} U(q_2))^T E_3 \nabla_{q_1} U(q_1)$$

$$= {p_2}^T(G_2^T + G_3)p_1 + (\nabla_{q_2}U(q_2))^T(E_2^T + E_3)\nabla_{q_1}U(q_1) = 0$$

Since $G_2^T + G_3$ and $E_2^T + E_3$ are zero, we completes the proof.

Lemma 2. We can make *leapfrog like integrator* in this dynamic system.

proof. We split our Hamiltonian into sub-Hamiltonians as follow :

$$H(q_1, q_2, p_1, p_2) = \frac{1}{2}U(q_1) + \frac{1}{2}U(q_2) + (\frac{1}{2}{p_1}^T p_1 + \frac{1}{2}{p_2}^T p_2) + \frac{1}{2}U(q_2) + \frac{1}{2}U(q_1)$$

,

with

$$H_1 = \frac{1}{2}U(q_1)$$

$$H_2 = \frac{1}{2}U(q_2)$$

$$H_3 = \frac{1}{2}{p_1}^T p_1 + \frac{1}{2}{p_2}^T p_2$$

Then, the corresponding non-canonical dynamics for each Hamiltonian are following :

$$H_1 : \frac{d}{dt}\begin{bmatrix} q_1 \\ q_2 \\ p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 0 & E_2 & F_1 & F_2 \\ E_3 & 0 & F_3 & F_4 \\ -F_1^T & -F_3^T & 0 & G_2 \\ -F_2^T & -F_4^T & G_3 & 0 \end{bmatrix} \begin{bmatrix} \nabla_{q_1}U(q_1) \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ E_3\nabla_{q_1}U(q_1) \\ -F_1^T\nabla_{q_1}U(q_1) \\ -F_2^T\nabla_{q_1}U(q_1) \end{bmatrix}$$

$$H_2 : \frac{d}{dt}\begin{bmatrix} q_1 \\ q_2 \\ p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 0 & E_2 & F_1 & F_2 \\ E_3 & 0 & F_3 & F_4 \\ -F_1^T & -F_3^T & 0 & G_2 \\ -F_2^T & -F_4^T & G_3 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \nabla_{q_2}U(q_2) \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} E_2\nabla_{q_2}U(q_2) \\ 0 \\ -F_3^T\nabla_{q_2}U(q_2) \\ -F_4^T\nabla_{q_2}U(q_2) \end{bmatrix}$$

$$H_3 : \frac{d}{dt}\begin{bmatrix} q_1 \\ q_2 \\ p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 0 & E_2 & F_1 & F_2 \\ E_3 & 0 & F_3 & F_4 \\ -F_1^T & -F_3^T & 0 & G_2 \\ -F_2^T & -F_4^T & G_3 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} F_1p_1 + F_2p_2 \\ F_3p_1 + F_4p_2 \\ G_2p_2 \\ G_3p_1 \end{bmatrix}$$

which are all tractable. Therefore, for given $\epsilon > 0$ and L, we can construct a leapfrog operator for H

$$\Phi_{H,\epsilon} = \Phi_{H_1,\epsilon} \circ \Phi_{H_2,\epsilon} \circ \Phi_{H_3,\epsilon} \circ \Phi_{H_2,\epsilon} \circ \Phi_{H_1,\epsilon}$$

Theorem 1. There exists an operator for the probabilistic sampling operator in non-canonical Hamiltonian dynamics with two particles interacts each other.

proof. Note that E, G and p must be flipped for each loop in non-canonical Hamiltonian dynamics for time reversibility. Therefore, by adding these operators and the operator earned in **Lemma 2**, we earn the following volume-preserving, self-inverse, time-reversible operator for our algorithm. The operator is given as follows:

$$\tilde{\Phi}_{H,\epsilon} = \Phi_p \circ \Phi_{E,G} \circ (\Phi_{H_1,\epsilon} \circ \Phi_{H_2,\epsilon} \circ \Phi_{H_3,\epsilon} \circ \Phi_{H_2,\epsilon} \circ \Phi_{H_1,\epsilon})^L$$

Therefore, we can make an algorithm for sampling by making a loop with flipping block matrices and momentum (sampled by standard normal), and doing leapfrog steps with given step size ϵ (local) and L (global).

5 Conclusion

Probability density sampling algorithms are very important for probabilistic inference, since it is hard to evaluate the exact value and functions with closed form in most cases. However, pre-existing probabilistic density sampling algorithms for probabilistic inferences still have problems in multimodal distributions or high-dimensional cases. Therefore, in this research, we tried to apply many existing physical theories such as Noether's Theorem, non-canonical Hamiltonian dynamics to propose better probabilistic inference algorithm for these situations.

Applying non-canonical Hamiltonian dynamics with independent probability distribution and two particles, we proposed two particle sampling algorithm for density sampling interacting each other. We copied and pasted the target distribution that we want to sample for making the assumption for independent distribution, and used non-canonical Hamiltonian dynamics to make interaction between these two particles. By interacting each other via non-canonical Hamiltonian dynamics, particles can be sampled in more various situation and get various traces. Setting proper parameters for two particles' interaction by real implementation and experiments, we expect more efficient, broad probabilistic density sampling will be available.

6 Further Research

We theoretically verified sampling algorithms for multiple particles with interaction is possible. Yet, we don't know this algorithm gives better results via interaction in real. Therefore, based on the real implementation of our results, setting various parameters and distribution, we need to check which interaction made by certain parameters could make the sampling more efficiently.

Furthermore, our algorithm can be generalized with more than 2 particles, not just 2 particles. If our algorithm is efficient with multimodal distribution sampling, we can increase more particles by the number of nodes in distributions. By balancing the efficiency earned by interaction and increased computation time via increasing the particle, we can optimize the number for sampling for given distribution to sample.

References

[1] S.P. Meyn & R.L Tweedie (2005) Markov Chains and Stochastic Stability, Springer-Verlag.

[2] Charles J. Geyer D. (2005) Markov Chain Monte Carlo Lecture Notes

[3] Michael Betancourt (2014) A Conceptual Introduction Hamiltonian Monte Carlo

[4] Radford M. Neal (2011) MCMC using Hamiltonian dynamics, Appears as Chapter 5 of the Handbook of Markov Chain Monte Carlo, CRC Press, 2011

[5] Paul Vanetti, George Deligiannidis (2018) Piecewise-Deterministic Markov Chain Monte Carlo

[6] Nilesh Tripuraneni, Mark Rowland, Zoubin Ghahramani, Richard Turner (2017) Magnetic Hamiltonian Monte Carlo