CSIAM Trans. Appl. Math.
doi: 10.4208/csiam-am.SO-2021-0031

Vol. **4**, No. 1, pp. 41-73
March 2023

A Splitting Hamiltonian Monte Carlo Method for Efficient Sampling

Lei Li[1,2,3], Lin Liu[1,2,3,4] and Yuzhou Peng[1,*]

[1] *School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, P.R. China.*
[2] *Institute of Natural Sciences and MOE-LSC, Shanghai Jiao Tong University, Shanghai 200240, P.R. China.*
[3] *Shanghai Artificial Intelligence Laboratory, Shanghai, P.R. China.*
[4] *CMA-Shanghai and SJTU-Yale Joint Center for Biostatistics and Data Science, Shanghai Jiao Tong University, Shanghai 200240, P.R. China.*

Received 6 June 2021; Accepted 17 June 2022

Abstract. We propose a splitting Hamiltonian Monte Carlo (SHMC) algorithm, which can be computationally efficient when combined with the random mini-batch strategy. By splitting the potential energy into numerically nonstiff and stiff parts, one makes a proposal using the nonstiff part of U, followed by a Metropolis rejection step using the stiff part that is often easy to compute. The splitting allows efficient sampling from systems with singular potentials (or distributions with degenerate points) and/or with multiple potential barriers. In our SHMC algorithm, the proposal only based on the nonstiff part in the splitting is generated by the Hamiltonian dynamics, which can be potentially more efficient than the overdamped Langevin dynamics. We also use random batch strategies to reduce the computational cost to $\mathcal{O}(1)$ per time step in generating the proposals for problems arising from many-body systems and Bayesian inference, and prove that the errors of the Hamiltonian induced by the random batch approximation is $\mathcal{O}(\sqrt{\Delta t})$ in the strong and $\mathcal{O}(\Delta t)$ in the weak sense, where Δt is the time step. Numerical experiments are conducted to verify the theoretical results and the computational efficiency of the proposed algorithms in practice.

AMS subject classifications: 62M05, 65C05
Key words: Markov chain Monte Carlo, potential splitting, random-batch method, many body systems, Bayesian inference.

1 Introduction

Markov chain Monte Carlo (MCMC) [4, 22, 27, 47, 53, 58] methods are nowadays routinely used in a variety of scientific computing problems, including computing statistics

*Corresponding author. *Email addresses:* leili2010@sjtu.edu.cn (L. Li), linliu@sjtu.edu.cn (L. Liu), pengyz@sjtu.edu.cn (Y. Peng)

for many-body systems [1, 21], sampling from log-concave distributions [11, 13, 39, 43], parameter estimation in Bayesian statistics [2, 22, 25, 57] and Bayesian inverse problems [24, 50], just to name a few. Among MCMC methods, Hamiltonian Monte Carlo (HMC) [3, 7, 16, 49] has recently garnered a lot of attention in practice due to its scalability and efficiency in high-dimensional settings [6, 49]. Nonetheless, there are several situations where HMC can encounter difficulties. The first such scenario might be sampling from the Gibbs distribution

$$\mu(q) \propto \exp[-\beta U(q)] \tag{1.1}$$

of a many-body interacting particle system, where $\beta > 0$ is the (dimensionless) inverse temperature. It takes $\mathcal{O}(N^2)$ operations to compute the total potential energy

$$U(q) = \sum_{i=1}^{N} w_i V(q_i) + \sum_{i,j:i<j} w_i w_j \phi(q_i - q_j), \tag{1.2}$$

where $q_i \in \mathbb{R}^d$ is the position of the i-th particle and w_i denotes the weight. If one moves one particle per step, which is preferred in some applications [15, 21, 42], the computational cost of evolving the Hamiltonian dynamics and the Metropolis-Hastings correction step in HMC are both $\mathcal{O}(N)$. This fact makes HMC computationally expensive when sampling from the Gibbs distribution. Moreover, interaction potentials ϕ such as the Coulomb potential or the Lennard-Jones potential are usually singular [21]. Singularity in ϕ can introduce stiffness to the Hamiltonian system, which makes the numerical simulations difficult [62], and possibly leads to low acceptance rates [42], thus deteriorating the sampling efficiency of HMC.

As for another well-known example, let us consider Bayesian inference of a parameter θ based on its posterior distribution $p_{\text{post}}(\theta|\mathcal{D}_N)$ given the observed data $\mathcal{D}_N$

$$p_{\text{post}}(\theta|\mathcal{D}_N) \propto p_{\text{prior}}(\theta) \prod_{i=1}^{N} p(y_i;\theta), \tag{1.3}$$

where $\mathcal{D}_N = \{y_1, \ldots, y_N\}$ is a sample of size N drawn from the probabilistic model $p(\cdot;\theta)$. When $p(\cdot;\theta)$ is posited to be a mixture model, which is often the case in clustering and density estimation problems [20], the corresponding posterior distribution $p_{\text{post}}(\theta|\mathcal{D}_N)$ may be multimodal. One main reason for multimodality in mixture models is the non-identifiability of the parameters due to label switching [29]. When the target distribution (the posterior distribution in this case) is multimodal, most MCMC algorithms have difficulty moving between isolated modes and are therefore prone to generate biased samples. As a result, HMC could fail to explore the entire state space and lead to even worse performance than the simple Random Walk Metropolis (RWM) algorithm [44]. Indeed, the Hamiltonian simulation heavily relies on the gradient information of the potential U, easily resulting in samples trapped in one single well of U when U has multiple modes. A popular strategy of sampling from multimodal distributions is running multiple chains with over-dispersed initializations in parallel. Other methods include parallel