Reproducing the paper: *Stochastic Gradient Hamiltonian Monte Carlo* by Tianqi Chen, Emily B. Fox and Carlos Guestrin

Sam Adam-Day, Alexander Goodall, Theo Lewy and Fanqi Xu

Abstract

We reproduce the experiments contained in 'Stochastic Gradient Hamiltonian Monte Carlo' [CFG14] by Chen, Fox and Guestrin.

FiXme: Give more details in abstract

List of Corrections

1 Introduction

- Overview of paper and its context.
- Which experiments replicated, and rationale for this choice.
- Target questions of paper.
- Experimental methodology.
- Implementation details.
  - Integration with Pyro.
  - Which parts are new, and which are from publicly available code?
  - Details about how key aspects were implemented.
- Link to repository.
- New aspects?

2 Background

Hamiltonian Monte Carlo (HMC) ([Dua+87; Nea11]) is a Markov Chain Monte Carlo (MCMC) sampling algorithm. Given a target probability distribution — in our case the posterior distribution of a set of variables θ given independent observations $x \in \mathcal{D}$ — it produces samples by carrying out a random walk over the parameter space using Hamiltonian dynamics.

To begin with prior distribution $p(\theta)$ and likelihood $p(x \mid \theta)$. Using these we define the *potential energy* function U:

$$U := -\sum_{x \in \mathcal{D}} \log p(x \mid \theta) - \log p(\theta)$$

Note that, using Bayes' rule, we have that the posterior $p(\theta \mid \mathscr{D}) \propto \exp(-U)$. Hamiltonian dynamics introduces an auxiliary set of momentum variables r. These dynamics have a physical interpretation in which an object moves about a landscape determined by U. We let this object have *mass matrix* M. Then $U(\theta)$ represents the potential energy of the object, and its kinetic energy is given by $\frac{1}{2}r^\top M^{-1}r$. The total energy of the system is a quantity known as the *Hamiltonian function*:

$$H(\theta, r) = U(\theta) + \frac{1}{2}r^\top M^{-1}r$$

The development of the system is governed by the following equations.

$$\begin{aligned} \mathrm{d}\theta &= M^{-1}r\,\mathrm{d}t \\ \mathrm{d}r &= -\nabla U(\theta)\,\mathrm{d}t \end{aligned}$$

To simulate these continuous dynamics in practice, we must use a discretised version of these equations. To correct for the inaccuracies introduced by doing so, it is necessary to make a *Metropolis-Hastings correction step*. A simple algorithm is given in Algorithm 1.

Algorithm 1 A simple HMC algorithm

```
for t = 1, 2, ... do
    r ~ N(0, 1)                                        ▷ Resample momentum
    (θ_0, r_0) = (θ, r)
    for i = 1 to m do
        θ ← θ + εM^{-1} r
        r ← r − ε∇U(θ)
    end for
    u ~ Uniform[0, 1]
    ρ = exp(H(θ, r) − H(θ_0, r_0))                     ▷ Acceptance probability
    if u > min(1, ρ) then              ▷ Only accept new state with probability ρ
        θ = θ_0
    end if
end for
```

In practice, the dataset $\mathscr{D}$ may be large, and so running Algorithm 1 may be computationally expensive. One idea to combat this is to simulate the Hamiltonian system using only a subset of the data at a time, in analogy with stochastic gradient descent. Unfortunately, such a dynamical system can diverge quite rapidly from the true posterior distribution [Nea11], which necessitates frequent Metropolis-Hastings steps. Such steps must be carried out using the whole dataset.

The method 'Stochastic Gradient Hamiltonian Monte Carlo' proposed in [CFG14] addresses this shortcoming. The idea is to incorporate friction into the dynamical system, which works to counteract the noise introduced by selecting a subset of the data.

3 Implementation Details

- Describe choice made and how we implemented the algorithms

4 Experiments

- Describe experiments and compare with results in the paper.

4.1 Simulated examples

4.2 Bayesian Neural Networks for Classification

5 Conclusion

- Analysis and discussion of findings.
- Suggest what could have been done with more time.

References

[CFG14] Tianqi Chen, Emily Fox and Carlos Guestrin. 'Stochastic Gradient Hamiltonian Monte Carlo'. In: *Proceedings of the 31st International Conference on Machine Learning*. Ed. by Eric P. Xing and Tony Jebara. Vol. 32. Proceedings of Machine Learning Research 2. Bejing, China: PMLR, June 2014, pp. 1683–1691. URL: `https://proceedings.mlr.press/v32/cheni14.html`.

[Dua+87] Simon Duane, A.D. Kennedy, Brian J. Pendleton and Duncan Roweth. 'Hybrid Monte Carlo'. In: *Physics Letters B* 195.2 (1987), pp. 216–222. ISSN: 0370-2693. DOI: `https://doi.org/10.1016/0370-2693(87)91197-X`. URL: `https://www.sciencedirect.com/science/article/pii/037026938791197X`.

[Nea11] Radford M. Neal. 'Handbook of Markov Chain Monte Carlo'. In: ed. by BySteve Brooks, Andrew Gelman, Galin Jones and Xiao-Li Meng. Chapman and Hall/CRC, 2011. Chap. MCMC Using Hamiltonian Dynamics, pp. 113–162. ISBN: 9780429138508.