MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

Hamiltonian Monte Carlo with Explicit, Reversible, and Volume-preserving Adaptive Step Size Control

Michiko OKUDO and Hideyuki SUZUKI

(Communicated by Takayasu MATSUO)

METR 2016–18 December 2016

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.keisu.t.u-tokyo.ac.jp/research/techrep/index.html

Hamiltonian Monte Carlo with Explicit, Reversible, and Volume-preserving Adaptive Step Size Control

Michiko OKUDO[1] and Hideyuki SUZUKI[2]

[1]Graduate School of Information Science and Technology
The University of Tokyo
[2]Graduate School of Information Science and Technology
Osaka University

December, 2016

Abstract

Hamiltonian Monte Carlo is a Markov chain Monte Carlo method that uses Hamiltonian dynamics to efficiently produce distant samples. It employs geometric numerical integration to simulate Hamiltonian dynamics, which is a key of its high performance. We present a Hamiltonian Monte Carlo method with adaptive step size control to further enhance the efficiency. We propose a new explicit, reversible, and volume-preserving integration method to adaptively set the step sizes, which does not violate the detailed balance condition or require a large increase in computational time.

1 Introduction

Generating random samples from a prescribed distribution is an important task in physics, statistics, and machine learning. For example, sampling from a posterior distribution is often necessary in Bayesian modeling.

Given a probability density function

$$P(x) \propto \exp(-V(x)),\ x \in \mathbb{R}^d$$

up to a multiplicative constant, the task is to generate unbiased random samples from $P(x)$. Markov chain Monte Carlo (MCMC) of the random-walk Metropolis type is generally employed in such situations. The Metropolis algorithm [1] generates unbiased samples through its acceptance process, but unfortunately the samples are highly correlated owing to the random

walk proposals. Hence it requires a large number of samples to generate effectively independent samples.

Hamiltonian Monte Carlo (HMC) [2] uses Hamiltonian dynamics to avoid the random walk behavior of the Metropolis algorithm, and is able to generate distant samples with a high acceptance rate. The success of HMC is largely due to geometric numerical integration of Hamiltonian dynamics. It preserves the good structure of Hamiltonian dynamics that is critical for constructing MCMC, and also achieves high accuracy and stability with relatively short computational time.

Here, we present an HMC method with adaptive step size control to further enhance the efficiency. In the existing methods, the Hamiltonian dynamics has been solved only with constant time steps, whereas in general numerical analysis it is often typical to employ some adaptive step size control to increase computational efficiency, especially with some time transformation. In fact, this research direction was suggested in an analysis of geometric ergodicity of HMC [3]. Nevertheless, no complete answer has been given so far, possibly due to the following reasons. First, the good natures in geometric numerical integration are generally lost by the adaptive step size control. Second, even if we succeed in the task with very careful implementation, that is still not enough in HMC context which requires severer properties (see the discussions in Section 4.1). So far, only a partial answer has been given in [4] to the best of the authors' knowledge. In this paper, we give a first complete answer to this problem, where the key is to modify the dynamics by regarding the step size as a member of random variables.

2 Hamiltonian Monte Carlo

In this section, we briefly review Hamiltonian dynamics and HMC. The details are explained in, e.g., [5].

2.1 Hamiltonian dynamics

Hamiltonian dynamics operates on a d-dimensional position variable q and a d-dimensional momentum variable p. The dynamics for the Hamiltonian $H(q,p)$ is described by the following differential equations:

$$\frac{dq}{dt} = \frac{\partial H}{\partial p}, \quad \frac{dp}{dt} = -\frac{\partial H}{\partial q}.$$

Let Φ_T be a mapping defined by $\Phi_T(q(0),p(0)) = (q(T),p(T))$. The following two properties of Φ_T are important for the theoretical justification of HMC:

Reversibility. Let ρ be the mapping $\rho(q,p) = (q,-p)$. Then, Φ_T is ρ-reversible; i.e., $\rho \circ \Phi_T = \Phi_T^{-1} \circ \rho$.

Volume preservation. The determinant of the Jacobian matrix of Φ_T has absolute value one.

Reversibility is necessary for HMC to satisfy the detailed balance (see, e.g., [6]), which is a condition generally used in constructing MCMC to guarantee that the target distribution is invariant for the Markov chain. Volume preservation can be relaxed by including a Jacobian in the acceptance probability of the Metropolis acceptance criterion. However, it is computationally challenging, and often impossible, to compute the Jacobian at every sample proposal. Ideally we wish to use as Φ_T the exact solution map of the Hamiltonian dynamics, but this is intractable except for in highly trivial cases. Thus, Hamiltonian dynamics is usually approximated by discretizing time. For HMC, the numerical integration method of choice is generally the leapfrog method. The leapfrog method preserves the reversibility and volume preservation properties of the Hamiltonian dynamics, as far as the step size is kept constant.

2.2 MCMC from Hamiltonian dynamics

Here, we briefly describe the algorithm of HMC. HMC can be viewed as a variation of the Metropolis algorithm, which uses Hamiltonian dynamics to create proposals. Let

$$P(q) \propto \exp(-V(q)),\ q \in \mathbb{R}^d$$

be the target distribution, where q is regarded as the position variable. To define Hamiltonian dynamics, we introduce a momentum variable

$$p \in \mathbb{R}^d,\ p \sim \mathcal{N}(0, I_d).$$

The Hamiltonian is defined as

$$H(q,p) = V(q) + K(p),$$

where $K(p) = p^\top p/2$. $V(q)$ and $K(p)$ can be viewed as the potential and kinetic energy of the Hamiltonian system, respectively. The probability density function of the joint distribution of (q,p) is proportional to $\exp(-V(q))\exp(-K(p)) = \exp(-H(q,p))$. HMC generates samples $(q_1,p_1),(q_2,p_2),\ldots$ from this joint distribution, and the user can obtain the samples from the target distribution by picking up only $q_1, q_2, \ldots$.

HMC algorithm is summarized as follows:

1. Propose a new sample: $(q^*_{n+1}, p^*_{n+1}) = \Phi_T(q_n, p_n)$.
2. Accept the proposal with the probability

$$\min\left(1, \frac{\exp(-H(q^*_{n+1}, p^*_{n+1}))}{\exp(-H(q_n,p_n))}\right).$$

If accepted, let $(q_{n+1}, p_{n+1}) = (q^*_{n+1}, p^*_{n+1})$. If rejected, let $(q_{n+1}, p_{n+1}) = (q_n, p_n)$.

3. Resample $p_{n+1} \sim \mathcal{N}(0, I_d)$.

Step 3 is necessary because without this, (q, p) will only move within a limited area where $H(q, p)$ is close to a constant. If the numerical error of $H(q, p)$ is small, the acceptance rate is close to 1. Under these steps, $P(q, p)$ remains invariant.

3 HMC with adaptive time step

In this section, we describe adaptive step size control, and consider how to adapt it to HMC.

In the field of numerical analysis, adaptive step size control methods aim to control numerical error by adaptively varying the step size. A considerable amount of work has been dedicated to this and various adaptive integrators have been studied for Hamiltonian dynamics. Here, we are interested in adaptive step size control using time transformations, where the step size is adaptively varied by performing time transformation $t \leftrightarrow \tau$, and applying a constant step size integration to the transformed system. The transformed Hamiltonian system can be written using τ as

$$\frac{dq}{d\tau} = \sigma(q, p)\frac{\partial H}{\partial p}, \quad \frac{dp}{d\tau} = -\sigma(q, p)\frac{\partial H}{\partial q},$$

where $\sigma(q, p) = dt/d\tau$.

By discretizing the above equations with a constant step size, it is possible to obtain an approximation with adaptive step sizes. However, the problem is that the scheme needs to be reversible for constructing HMC, and it is also strongly desirable that it is volume-preserving and explicit. For explicit schemes with adaptive step size control, reversibility is usually defined not for (q, p), but for the triple of q, p, and another variable which controls step sizes. This reversibility requires careful treatment when applied to HMC. We explain about this difficulty of reversibility in Section 4.1.

Before we explain our proposed algorithm, we mention an existing work. Nishimura and Dunson [4] proposed an HMC method that employs adaptive step size control using time transformations. This was the first one of its kind to satisfy detailed balance to our knowledge. However, its numerical integration scheme is linearly implicit and not volume-preserving, and it can only deal with time transformations in the form of $\sigma(q)$; that is, time transformations that do not depend on p.

4 Proposed algorithm

In this section, we explain our proposed method. Our proposed method enables HMC to employ adaptive step size control in a reversible, volume-preserving, and explicit manner, and hence it does not violate the detailed balance or significantly increase the computational time. First, we explain about the adaptive integration method that our proposed method is based on, and then consider its adaptation for HMC.

4.1 Reversibility in HMC

An adaptive, reversible, and explicit (but not volume-preserving) scheme for Hamiltonian systems is proposed in [7]. A variable is introduced that defines the step sizes adaptively, given by $z = 1/\sigma(q,p)$, and the solution of (q,p,z) is numerically approximated. Let $G(q,p) = dz/d\tau$. Then the integration proceeds as follows:

$$\begin{aligned} z(\tau+\epsilon/2) &= z(\tau) + (\epsilon/2)G(q(\tau),p(\tau)), \\ (q(\tau+\epsilon),p(\tau+\epsilon)) &= \Phi_{\epsilon/z(\tau+\epsilon/2)}(q(\tau),p(\tau)), \\ z(\tau+\epsilon) &= z(\tau+\epsilon/2) + (\epsilon/2)G(q(\tau+\epsilon),p(\tau+\epsilon)). \end{aligned}$$

Here, $z(0) = 1/\sigma(q(0),p(0))$ and let $\Phi_{\epsilon/z(\tau+\epsilon/2)}$ be the mapping by the one step of leapfrog integrator.

Let $\hat{\Phi}_\epsilon^L$ be a mapping $(q(0),p(0),z(0)) \mapsto (q(T'),p(T'),z(T'))$ by the above scheme, where L is the number of integration steps and T' is the integration time, i.e. the sum of L adaptive time steps. $\hat{\Phi}_\epsilon^L$ is reversible regarding the mapping $\rho' : (q,p,z) \mapsto (q,-p,z)$. This means that when $\hat{\Phi}_\epsilon^L(q_1,p_1,z_1) = (q_2,p_2,z_2)$, $\hat{\Phi}_\epsilon^L(q_2,-p_2,z_2) = (q_1,-p_1,z_1)$ holds. This definition of reversibility, as dealt in [7], is sufficient in usual numerical computation, but here we emphasize an important fact that it is not the case for HMC. The difference comes from the fact that while the usual computation considers *one orbit*, HMC considers *a (infinitely many) collection of orbits*. Recall the step 3 of the HMC algorithm; there, we randomly resample the moment p_{n+1}, by which we randomly hop to a different orbit. In the usual computation, we move back and forth on only one orbit without hopping, so the ρ'-reversibility implies if we go from (q_1,p_1) to (q_2,p_2), we come from $(q_2,-p_2)$ to $(q_1,-p_1)$ by preserving z_2 in the reversing step. But what happens if we hopped to $(q_2,-p_2)$ from another orbit with time stepping variable $z_2' \neq z_2$ by resampling the moment? We can no longer come to $(q_1,-p_1)$ because the point we arrive depends on z. This results in violating the detailed balance of HMC. Here, note that we do not have any clue to reaching $(q_1,-p_1)$ from $(q_2,-p_2)$, as we do not know the value of z_2 and in general z_2 does not coincide with $1/\sigma(q_2,-p_2)$.

Thus the "reversibility" in HMC demands that the overall dynamics including the random hopping is "reversible" in some sense. Without the

reversibility regarding (q, p), the detailed balance regarding (q, p) cannot be ensured. Note that such a problem does not occur in HMC with constant step sizes because z is fixed to one.

4.2 Satisfying reversibility and volume preservation

Here, we present a natural way to incorporate adaptive step size control in HMC without violating the detailed balance condition. The key is to treat the variable z in the integration scheme [7], which controls the step sizes, as a random variable, and construct the Markov chain $(q_1, p_1, z_1), (q_2, p_2, z_2), \ldots$. Actually, the ρ'-reversibility can work in our proposed method.

We modify the dynamics in HMC by regarding z as a member of random variables, and consider satisfying the detailed balance regarding (q, p, z). Thus, we can make use of the ρ'-reversibility to ensure the detailed balance. Moreover, the scheme is volume-preserving regarding (q, p, z), although it is not volume-preserving for (q, p). Therefore, we can avoid computing the Jacobian of the mapping in computing the acceptance ratio. Thus, we have obtained a reversible, volume-preserving, and explicit scheme for HMC.

We treat z as a random variable to be sampled, so we have to set the target distribution of z, which can be set arbitrarily. In the following numerical experiments, we use the uniform distribution $U(C_1, C_2)$, where C_1 and C_2 are constants.

Our proposed algorithm is summarized as follows:

1. Propose a new sample; $(q^*_{n+1}, p^*_{n+1}, z^*_{n+1}) = \hat{\Phi}^L_\epsilon(q_n, p_n, z_n)$.
2. Accept the proposal with the probability

$$\min\left(1, \frac{\exp(-H(q^*_{n+1}, p^*_{n+1}))P(z^*_{n+1})}{\exp(-H(q_n, p_n))P(z_n)}\right).$$

   If accepted, let $(q_{n+1}, p_{n+1}, z_{n+1}) = (q^*_{n+1}, p^*_{n+1}, z^*_{n+1})$. If rejected, let $(q_{n+1}, p_{n+1}, z_{n+1}) = (q_n, p_n, z_n)$.
3. Resample $p_{n+1} \sim \mathcal{N}(0, I_d)$.

$P(z)$ is the probability density function of the target distribution of z. In Step 1, we cannot set $z(0) = 1/\sigma(q(0), p(0))$, because this violates the detailed balance. $z(0)$ is set as the value of the previous sample. That is, $\hat{\Phi}^L_\epsilon$ computes the approximate path of z for a different initial value.

Note that our proposed algorithm can deal with time transformations $\sigma(q, p)$ which depend on both q and p, which were out of the scope of [4].

4.3 Proof of stationarity

Now, we show that our proposed algorithm satisfies stationarity. This proof is based on the proof of stationarity given in [8]. Let $x = (q, p, z)$, and let

$P(x)$ be the probability density function of the target distribution. Since P clearly remains invariant under Step 3 in the proposed algorithm, it is sufficient to show that it remains invariant under Step 1 and 2.

First, we show that our proposed method satisfies the modified detailed balance

$$P(x)P(x \to x') = P(\rho'(x'))P(\rho'(x') \to \rho'(x))$$

for each x and x'. Here, $P(x \to x')$ denotes the conditional distribution of the next sample x', conditional on the present sample x. The left and right hand sides can be rewritten as:

$$\begin{aligned}
&P(x)P(x \to x') \\
&= P(x)\delta(x' - \hat{\Phi}_\epsilon^L(x)) \min(1, P(x')/P(x)), \\
&P(\rho'(x'))P(\rho'(x') \to \rho'(x)) \\
&= P(\rho'(x'))\delta(\rho'(x) - \hat{\Phi}_\epsilon^L(\rho'(x'))) \min(1, P(\rho'(x))/P(\rho'(x'))) \\
&= P(x')\delta(\rho'(x) - \hat{\Phi}_\epsilon^L(\rho'(x'))) \min(1, P(x)/P(x')).
\end{aligned}$$

Here, we have

$$\begin{aligned}
\delta(\rho'(x) - \hat{\Phi}_\epsilon^L(\rho'(x'))) &= \delta(x - \rho'(\hat{\Phi}_\epsilon^L(\rho'(x')))) \\
&= \delta(\hat{\Phi}_\epsilon^L(x) - \hat{\Phi}_\epsilon^L(\rho'(\hat{\Phi}_\epsilon^L(\rho'(x'))))) \\
&= \delta(\hat{\Phi}_\epsilon^L(x) - x'),
\end{aligned}$$

because ρ' and $\hat{\Phi}_\epsilon^L$ are volume preserving and $\hat{\Phi}_\epsilon^L$ is ρ'-reversible. Therefore, we obtain the modified detailed balance. Stationarity can be proven as follows:

$$\begin{aligned}
\int P(x)P(x \to x')dx &= \int P(\rho'(x'))P(\rho'(x') \to \rho'(x))dx \\
&= \int P(x')P(\rho'(x') \to \rho'(x))dx \\
&= P(x').
\end{aligned}$$

5 Numerical experiments

In this section, we present numerical experiment to demonstrate our proposed algorithm does in fact work and adaptive step size control is effective in HMC.

Our method is tested on a high-dimensional mixture of two Gaussians. The settings are based on the experiments described in Section IV.A.1 in [9]. The difference is that here the distribution is set to be 200-dimensional, and the two Gaussians of the first dimension are located at ± 3.5 to make the sampling more difficult. The step size ϵ was set for the acceptance

rate to be about 95% and the number of integration steps $L = 5$. In our method, step sizes are adaptively varied by time transformation with $\sigma(q, p) = \exp(-p_1^2/2)^\alpha$, where p_1 is the first component of p, and α is $1/14$. It is expected that the step sizes will increase only when $K(p_1)(= -p_1^2/2)$ is large, that is $V(q_1)(= -\log(P(q_1)))$ is small, and that the samples will effectively travel between the two modes, while the algorithm uses small step sizes in other areas and keeps high acceptance rate. The constants for the uniform distribution of z are $C_1 = 0.7$ and $C_2 = 6$.

We confirmed empirically that our proposed algorithm preserves the target distribution invariant by checking the histogram of the samples. We checked the change of step sizes during integrations to verify our proposed method does in fact change the step size adaptively according to the time transformation. Figure 1 illustrates the step sizes and $\exp(-p_1^2/2)^\alpha$ during one integration. Here, the number of steps L was altered from 5 to 20 in order to illustrate the change in step sizes over a longer period. The step sizes are varied adaptively so that they are large when $\exp(-p_1^2/2)^\alpha$ is large. This shows that our proposed method does in fact work and varies the step size adaptively according to the step size function $\sigma(q, p)$. In addition, we compared our proposed method with standard HMC in terms of the effective sample size (ESS) of the observables $A(q_1) = 1/(1 + \exp(-q_1))$. ESS describes the number of samples that can be regarded as being sampled independently from the target distribution. We computed 10^4 samples 100 times, and took the average of ESS. The average ESS values of our proposed method and standard HMC were 15.0 and 9.2, respectively. Thus, the ESS of our proposed method was about 1.6 times greater. This means that samples of our proposed method have smaller correlation. Figure 2 illustrates the history of q_1. In the results of our proposed method, the samples are not caught in one mode for a long time, so q_1 travels between the two modes of the Gaussian mixture more frequently than in the results of standard HMC. The average computational time of our method and standard HMC for 10^4 samples were 14.0 seconds and 13.0 seconds, respectively. This shows that our proposed method did not largely increase computational time in this setting.

6 Concluding remarks

In this paper, we have proposed a framework to incorporate adaptive step size control into HMC using a reversible, volume-preserving, and explicit integration scheme. We demonstrated that this method works effectively through numerical experiments.

It should be noted that the key aim of this paper was to propose the framework itself, and more careful consideration regarding its practical employment is required, along with further numerical experiments. For exam-

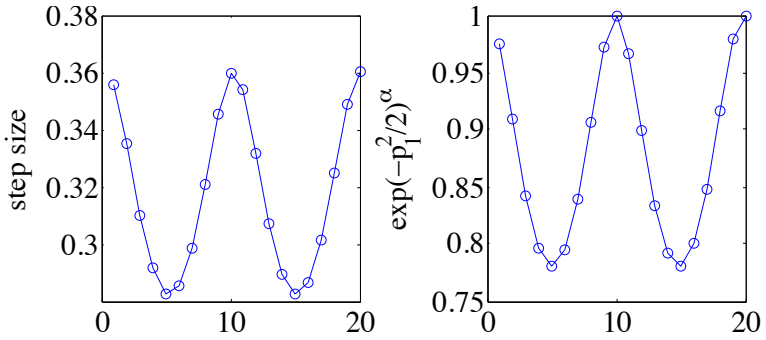

Figure 1: Step sizes and $\exp(-p_1^2/2)^\alpha$ during one integration. The horizontal axis shows the integration step.

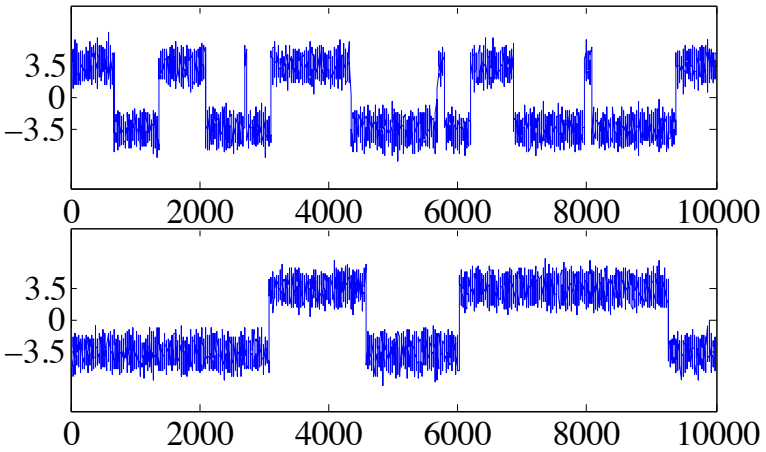

Figure 2: History of q_1. The horizontal axis is the sample index and the vertical axis is q_1. The upper panel represents our proposed method, and the lower is for standard HMC.

ple, these could involve exploring other time transformations and performing experiments on other distributions.

Acknowledgments

The authors are grateful to Takayasu Matsuo and Shun Sato for valuable comments. This study was supported in part by ImPACT Program of Council for Science, Technology and Innovation (Cabinet Office, Government of Japan).

References

[1] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, Equation on state calculations by fast computing machines, J. Chem. Phys., **21** (1953), 1087–1092.

[2] S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth, Hybrid Monte Carlo, Phys. Lett. B, **195** (1987), 216–222.

[3] S. Livingstone, M. Betancourt, S. Byrne and M. Girolami, On the geometric ergodicity of Hamiltonian Monte Carlo, arXiv:1601.08057 [stat.CO].

[4] A. Nishimura and D. Dunson, Geometrically tempered Hamiltonian Monte Carlo, arXiv:1604.00872 [stat.CO].

[5] R. M. Neal, MCMC using Hamiltonian dynamics, In Handbook of Markov Chain Monte Carlo, Chapman & Hall/CRC Press, Boca Raton, 2010, pp. 113–162.

[6] J. M. Sanz-Serna, Markov chain Monte Carlo and numerical differential equations, in: Current challenges in stability issues for numerical differential equations, L. Dieci and N. Guglielmi eds., LNM, Vol. 2082, pp. 39–88, Springer, Cham, 2014.

[7] E. Hairer and G. Söderlind, Explicit, time reversible, adaptive step size control, SIAM J. Sci. Comput., **26** (2005), 1838–1851.

[8] C. M. Campos and J. M. Sanz-Serna, Extra chance generalized hybrid Monte Carlo, J. Comput. Phys., **281** (2015), 365–374.

[9] Y. Fang, J. M. Sanz-Serna and R. D. Skeel, Compressible generalized hybrid Monte Carlo, J. Chem. Phys., **140** (2014), 174108.