

# 用 Velocity Verlet 积分器 改进 HMC 抽样方法

李婉荧, 唐亚勇

(四川大学数学学院, 成都 610064)

**摘要:** Hamilton Monte Carlo (HMC)方法是一种常用的快速抽样方法. 在对哈密顿方程进行抽样时, HMC方法使用 Leapfrog 积分器, 这可能造成方程的位置及动量的迭代值在时间上不同步, 其产生的误差会降低抽样效率及抽样结果的稳定性. 为此, 本文提出了 IHMC (Improved HMC)方法, 该方法用 Velocity Verlet 积分器替代 Leapfrog 积分器, 每次迭代时都计算两变量在同一时刻的值. 为验证方法的效果, 本文进行了两个实验, 一个是将该方法应用于非对称随机波动率模型(RASV模型)的参数估计, 另一个是将方法应用于方差伽马分布的抽样, 结果显示: IHMC方法比 HMC方法的效率更高、结果更稳定.

**关键词:** HMC方法; Velocity Verlet 积分器; RASV模型; 方差伽马分布

**中图分类号:** O29 **文献标志码:** A **DOI:** 10.19907/j.0490-6756.2024.011004

## Improving the HMC sampling method by Velocity Verlet integrator

LI Wan-Ying, TANG Ya-Yong

(School of Mathematics, Sichuan University, Chengdu 610064, China)

**Abstract:** Hamilton Monte Carlo (HMC) method is a fast sampling method. To sampling a Hamiltonian equation, the HMC method uses the Leapfrog integrator. As a result, unsynchronized values of position and momentum variables of the equation are possibly generated during the iterative process and resulting estimation error can seriously damage the sampling efficiency and stability of sampling result. To solve this problem, here we propose an improved HMC (IHMC) method by replacing the Leapfrog integrator in HMC method with Velocity Verlet integrator. The IHMC method calculates the variables simultaneously in each iteration. Two numerical examples are implemented to check the performance of the IHMC method, one is the problem of estimating the parameters of realized and stochastic volatility (RASV) model with asymmetric effects, the other is the problem of sampling the variance Gamma distribution. It is shown that the IHMC method has higher sampling efficiency and more stable sampling result compared with the HMC method.

**Keywords:** Hamiltonian Monte Carlo method; Velocity Verlet integrator; RASV model; Variance Gamma distribution

收稿日期: 2022-10-17

作者简介: 李婉荧(1999-), 山西临汾人, 博士研究生, 主要研究方向为金融数学.

通讯作者: 唐亚勇. E-mail: yayongtang@scu.edu.cn

## 1 Introduction

The Hamilton Monte Carlo (HMC) method is a fast sampling method combining the Markov chain Monte Carlo (MCMC) method and Hamiltonian dynamics. To sampling a Hamiltonian equation, the HMC method uses Leapfrog integrator. This method is firstly introduced in the late 1980s by Duane *et al.*<sup>[1]</sup>, known as the hybrid Monte Carlo method. The HMC method is mainly applied in physics until Neal<sup>[2]</sup> realized that it can also be applied to Bayesian neural network. In 2011, Neal<sup>[3]</sup> further provided a detailed introduction to this method and made it widely noticed and applied. Nowadays the HMC method has more and more applications in simulations of various chemical systems, physical systems and statistical inferences.

Compared with the MCMC method, the HMC method has higher efficiency and stability. But there is still some room for improvement. Recently, combined with some important sampling methods, new modified Hamiltonian Monte Carlo (MHMC) methods are proposed<sup>[4-6]</sup>. Notably, Radivojević *et al.*<sup>[4]</sup> used the splitting integrator instead of the Leapfrog integrator to get better conservation properties. Remarkably, unsynchronized values of the position and momentum variables of the Hamiltonian equation are possibly generated during the iterative process and the resulting estimation error can seriously damage the sampling efficiency and stability of sampling result. In this paper, we try to use the Velocity Verlet integrator to improve the HMC method, since the Velocity Verlet integrator calculates the position and momentum variables simultaneously with accuracy similar to the Leapfrog integrator.

The rest of the paper is organized as follows. Section 2 and Section 3 describe the HMC method and the IHMC method. In Section 4, two numerical examples are used to show the performance of the IHMC method. Finally, we summarize this paper in Section 5.

## 2 The HMC method

The HMC method combines the MCMC method with Hamiltonian dynamics. For computer simulation, the Hamiltonian equation is firstly discretized by using a small time-step  $\epsilon$ . Then the Leapfrog integrator is used to carry out the iterative sampling.

Hamiltonian dynamics operates on a  $d$ -dimensional position vector  $q$  and a  $d$ -dimensional momentum vector  $p$ , thus the full state space is  $2d$ -dimensional. The system can be described by a function of  $q$  and  $p$  known as the Hamiltonian,  $H(q, p)$ , written as

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

here  $U(q)$  is called potential energy,  $K(p)$  is called kinetic energy, and the functions are defined as,

$$U(q) = -\ln f(q), K(p) = \frac{1}{2} |p|^2,$$

where  $f(q)$  is the probability density of  $q$ . The position variable  $q$  is obtained by solving the Hamilton's equation given by the following equations

$$\begin{cases} \frac{d q_t}{d t} = \frac{\partial H}{\partial p_t} = p_t, \\ \frac{d p_t}{d t} = -\frac{\partial H}{\partial q_t} = -\nabla U(q_t) \end{cases} \quad (1)$$

respectively, where  $t$  represents time,  $q_t$  and  $p_t$  can also be written as  $q(t)$  and  $p(t)$ . The Leapfrog integrator has local error of order  $\epsilon^3$  and global error of order  $\epsilon^2$ . The Leapfrog integrator guarantees the symplectic and volume preservation properties by performing the following steps in each iteration:

$$\begin{cases} p\left(t + \frac{\epsilon}{2}\right) = p(t) - \frac{\epsilon}{2} \nabla U(q(t)), \\ q(t + \epsilon) = q(t) + \epsilon p\left(t + \frac{\epsilon}{2}\right), \\ \dot{p}(t + \epsilon) = p\left(t + \frac{\epsilon}{2}\right) - \frac{\epsilon}{2} \nabla U(q(t + \epsilon)) \end{cases} \quad (2)$$

**Step 1** Input the step size  $\epsilon$ , the number of iterations  $L$  and the initial value  $q_0$  for  $q$ .

**Step 2** Draw  $p$  from  $N(0, I)$ .

**Step 3** Simulate discretization of the Ham-

iltonian dynamics by (2). Then we do

(i) half-step update  $p$ :  $p \leftarrow p - \frac{\epsilon}{2} \nabla U(q)$ ;

(ii) alternating update  $p$  and  $q$  for  $L$  times,

say,  $\begin{cases} q \leftarrow q + \epsilon p, \\ p \leftarrow p - \epsilon \nabla U(q); \end{cases}$

(iii) half-step update  $p$ :  $p \leftarrow p - \frac{\epsilon}{2} \nabla U(q)$ .

**Step 4** Let  $(q, p)$  be the value taken before the update,  $(\hat{q}, \hat{p})$  be the value taken after the update. Calculate the acceptance probability

$$P = \exp\{-H(\hat{q}, \hat{p}) + H(q, p)\}.$$

**Step 5** Draw  $u \sim \text{Uniform}[0, 1]$ , then

$$q = \begin{cases} q, & \text{if } u \geq P, \\ \hat{q}, & \text{if } u < P. \end{cases}$$

When the HMC method runs, the momentum vector is first updated once with half a step, then the position and momentum vectors are updated a specified number of times with one step, and finally the momentum vector is updated once again with half a step. Specifically, the obtained position and momentum vectors are  $p\left(\frac{\epsilon}{2}\right), q(\epsilon), p\left(\frac{3\epsilon}{2}\right), q(2\epsilon), p\left(\frac{5\epsilon}{2}\right), q(3\epsilon), \dots$  in order. A noteworthy drawback of the HMC method is that the position and momentum vectors are not synchronized in time, which may result in errors during iteration and affect the sampling efficiency and stability of sampling result.

### 3 The IHMC Method

#### 3.1 The Velocity Verlet integrator

To avoid the errors caused by the HMC method, here we introduce the Velocity Verlet integrator<sup>[5]</sup>. Similar to the Leapfrog integrator, the Velocity Verlet integrator also has order  $\epsilon^3$  local error and order  $\epsilon^2$  global error and can preserve the symplectic and volume<sup>[6]</sup> besides that the momentum and position vectors can be calculated simultaneously.

The target variable  $(q_i, p_i)$  discretized by the Velocity Verlet integrator with step size  $\epsilon$  is the solution of the following differential system:

$$\begin{cases} \frac{dq_i}{dt} = p_{\lfloor t \rfloor} - \frac{\epsilon}{2} \nabla U(q_{\lfloor t \rfloor}), \\ \frac{dp_i}{dt} = -\frac{1}{2} (\nabla U(q_{\lfloor t \rfloor}) + \nabla U(q_{\lceil t \rceil})) \end{cases} \quad (3)$$

where  $\lfloor t \rfloor = \max\{s \in \epsilon \mathbf{Z}; s \leq t\}$ ,  $\lceil t \rceil = \min\{s \in \epsilon \mathbf{Z}; s \geq t\}$ . Compared with (1), the gradient term in equation (3) is used to reduce the error of the numerical integration. The Velocity Verlet integrator updates the position and momentum variables as follows.

$$\begin{cases} q(t+2\epsilon) = q(t) + 2\epsilon p(t) - 2\epsilon^2 \nabla U(q(t)), \\ p(t+\epsilon) = p(t) - \frac{\epsilon}{2} [\nabla U(q(t)) + \nabla U(q(t+2\epsilon))], \\ q(t+\epsilon) = q(t) + \epsilon p(t+\epsilon) - \epsilon^2 \nabla U(q(t)) \end{cases} \quad (4)$$

Then, by using (4) to replace (2), Step 3 in the HMC method should be replaced by

**Step 3\*** Simulate discretization of Hamiltonian dynamics by (4). Then we do alternating update  $p$  and  $q$  for  $L$  times:

$$\begin{cases} q_v \leftarrow q + 2\epsilon p - 2\epsilon^2 \nabla U(q), \\ p \leftarrow p - \frac{\epsilon}{2} [\nabla U(q) + \nabla U(q_v)], \\ q \leftarrow q + \epsilon p - \epsilon^2 \nabla U(q). \end{cases}$$

Compared with (2), (4) contains more computations on the premise of obtaining  $p(t+\epsilon)$  and  $q(t+\epsilon)$  with the same  $\epsilon$ . However, (4) can reduce the estimate errors and solve the time synchronization problem with higher stability. That is to say, good result still can be expected even for larger step sizes. We will confirm this by some simulations in the next section.

#### 3.2 Convergence of the IHMC method

To investigate the convergence of the IHMC method, suppose that  $U$  is a quadratic differentiable function satisfying that  $\|\nabla U(x)\| \leq L \|x\|$ , which can be easily satisfied by most functions. Consider that  $T \in [0, \infty)$ ,  $\epsilon \in [0, 1]$  satisfying  $T/\epsilon \in \mathbf{Z} (\epsilon \neq 0)$  and assume that  $L(T^2 + \epsilon T) \leq 1$  which can be satisfied by choosing a suitable  $T$  for a given  $L$ . We have the following two theorems.

**Theorem 3.1** For given initial values  $q_0, p_0 \in \mathbf{R}^d$ , the following inequalities hold:

$$\max_{m \leq T} \{ \|q_m\| \} \leq \frac{1}{1 - \frac{L}{2}(T^2 + \epsilon T)} \cdot$$

$$\begin{aligned} & \max \|q_0\|, \|q_0 + p_0 T\|, \\ & \max_{m \in [0, T]} \{ \|p_m\| \} \leq \|p_0\| + \frac{2}{T + \epsilon} \cdot \\ & \max \|q_0\|, \|q_0 + p_0 T\|. \end{aligned}$$

**Proof** From the equations (3), for all  $s \in [0, T]$  we have

$$\begin{aligned} \int_0^s \frac{d}{dt} q_t &= q_s - q_0 = \\ \int_0^s p_{[t]} dr &- \frac{\epsilon}{2} \int_0^s \nabla U(q_{[t]}) dr, \end{aligned}$$

where

$$p_{[t]} = p_0 - \frac{1}{2} \int_0^{[t]} (\nabla U(q_{[u]}) + \nabla U(q_{[u]})) du,$$

$[r] = \max\{t \in \epsilon \mathbf{Z}; t \leq r\}$ ,  $\lceil r \rceil = \min\{t \in \epsilon \mathbf{Z}; t \geq r\}$ .

Combining the above two equations gives

$$\begin{aligned} q_s &= q_0 + p_0 s - \\ & \frac{1}{2} \int_0^s \int_0^{[r]} (\nabla U(q_{[u]}) + \nabla U(q_{[u]})) dudr - \\ & \frac{\epsilon}{2} \int_0^s \nabla U(q_{[r]}) dr \end{aligned} \tag{5}$$

Then we have

$$\begin{aligned} \|q_s - q_0 - p_0 s\| &= \left\| \frac{1}{2} \int_0^s \int_0^{[r]} (\nabla U(q_{[u]}) + \right. \\ & \nabla U(q_{[u]})) dudr - \frac{\epsilon}{2} \int_0^s \nabla U(q_{[r]}) dr \left. \right\| \leq \\ & \frac{1}{2} \int_0^s \int_0^{[r]} (\|\nabla U(q_{[u]})\| + \|\nabla U(q_{[u]})\|) dudr + \\ & \frac{\epsilon}{2} \int_0^s \|\nabla U(q_{[r]})\| dr \leq \frac{L}{2} \int_0^s \int_0^{[r]} (\|q_{[u]}\| + \\ & \|q_{[u]}\|) dudr + \frac{\epsilon L}{2} \int_0^s \|q_{[r]}\| dr \leq \\ & L \int_0^s \int_0^r \max_{m \in [0, T]} \{ \|q_m\| \} dudr + \\ & \frac{\epsilon L}{2} \int_0^s \max_{m \in [0, T]} \{ \|q_m\| \} dr \leq \\ & \frac{L}{2} (s^2 + \epsilon s) \max_{m \in [0, T]} \{ \|q_m\| \} \leq \\ & \frac{L}{2} (T^2 + \epsilon T) \max_{m \in [0, T]} \{ \|q_m\| \}. \end{aligned}$$

In the above equation  $\max_{m \in [0, T]} \{ \|q_m\| \}$  can be re-written as

$$\begin{aligned} \max_{m \in [0, T]} \{ \|q_m\| \} &\leq \max_{m \in [0, T]} \{ \|q_m - q_0 - p_0 m\| \} + \\ & \max\{ \|q_0\|, \|q_0 + p_0 T\| \} \end{aligned} \tag{6}$$

It is further obtained that

$$\begin{aligned} \max_{m \in [0, T]} \{ \|q_m - q_0 - p_0 m\| \} &\leq \\ & \frac{L}{2} (T^2 + \epsilon T) \\ & \frac{1}{1 - \frac{L}{2} (T^2 + \epsilon T)} \max\{ \|q_0\|, \|q_0 + p_0 T\| \} \end{aligned}$$

(7)

Substituting (7) into (6) again gives

$$\max_{m \in [0, T]} \{ \|q_m\| \} \leq \frac{1}{1 - \frac{L}{2} (T^2 + \epsilon T)} \cdot$$

$$\max\{ \|q_0\|, \|q_0 + p_0 T\| \} \tag{8}$$

From the equations (3), we have

$$p_s = p_0 - \frac{1}{2} \int_0^s (\nabla U(q_{[r]}) + \nabla U(q_{[r]})) dr \tag{9}$$

From (8) and (9), we get

$$\max_{m \in [0, T]} \{ \|p_m\| \} \leq \|p_0\| + \frac{LT}{1 - \frac{L}{2} (T^2 + \epsilon T)} \cdot$$

$$\max\{ \|q_0\|, \|q_0 + p_0 T\| \}.$$

From  $L(T^2 + \epsilon T) \leq 1$ , the above equation can be reduced to

$$\begin{aligned} \max_{m \in [0, T]} \{ \|p_m\| \} &\leq \|p_0\| + \frac{2}{T + \epsilon} \cdot \max\{ \|q_0\|, \\ & \|q_0 + p_0 T\| \} \end{aligned} \tag{10}$$

The proof is completed.

**Theorem 3.2** For given initial values  $q_0, p_0 \in \mathbf{R}^d$ , the following inequalities hold:

$$\begin{aligned} \max_{s, t \leq T} \{ \|q_t - q_s\| \} &\leq \|p_0\| T + \\ & \max\{ \|q_0\|, \|q_0 + p_0 T\| \}, \\ \max_{s, t \leq T} \{ \|p_t - p_s\| \} &\leq \\ & \frac{2}{T + \epsilon} \max\{ \|q_0\|, \|q_0 + p_0 T\| \}. \end{aligned}$$

**Proof** Without loss of generality, assume that  $0 \leq s < t \leq T$ . From (5), we have

$$\begin{aligned} \|q_t - q_s\| &\leq \|p_0(t - s)\| + \\ & \frac{1}{2} \int_s^t \int_s^{[r]} (\|\nabla U(q_{[u]})\| + \|\nabla U(q_{[u]})\|) dudr + \\ & \frac{\epsilon}{2} \int_s^t \|\nabla U(q_{[r]})\| dr \leq \|p_0\| T + \frac{L}{2} ((t - s)^2 + \\ & \epsilon(t - s)) \max_{m \in [0, T]} \{ \|q_m\| \} \leq \|p_0\| T + \frac{L}{2} (T^2 + \\ & \epsilon T) \frac{1}{1 - \frac{L}{2} (T^2 + \epsilon T)} \cdot \end{aligned}$$

$$\begin{aligned} \max\{ \|q_0\|, \|q_0 + p_0 T\| \} &\leq \\ & \|p_0\| T + \max\{ \|q_0\|, \|q_0 + p_0 T\| \}. \end{aligned}$$

*i. e.*,

$$\begin{aligned} \max_{s, t \leq T} \|q_t - q_s\| &\leq \|p_0\| T + \\ & \max\{ \|q_0\|, \|q_0 + p_0 T\| \} \end{aligned} \tag{11}$$

From (9) we have

$$\|p_t - p_s\| \leq \frac{1}{2} \int_s^t \|\nabla U(q_{[r]})\| + \|\nabla U(q_{[r']})\| dr \leq$$

$$L(t-s) \max_{m \in [0, T]} \{\|q_m\|\} \leq \frac{LT}{1 - \frac{L}{2}(T^2 + \epsilon T)} \cdot$$

$$\max\{\|q_0\|, \|q_0 + p_0 T\|\} \leq \frac{2}{T + \epsilon} \max\{\|q_0\|, \|q_0 + p_0 T\|\},$$

*i. e.*,

$$\max_{s, t \leq T} \|p_t - p_s\| \leq \frac{2}{T + \epsilon} \max\{\|q_0\|, \|q_0 + p_0 T\|\} \tag{12}$$

asrequired. The proof is end.

**Remark 1** Theorem 3. 1 tells us that in a time interval  $T$ , the values of position variable  $p$  and the velocity variable  $q$  generated by the IHMC method can be controlled by the corresponding constant velocity motion (initial position  $q_0$ , initial velocity  $p_0$ ). Furthermore, Theorem 3. 2 tells us that in a chain generated by IHMC method, the deviations of the variable values at all moments are restricted to a certain range (given by the initial conditions of  $q_0$ ,  $p_0$ , and the upper time limit  $T$ ). Thus the convergence of the IHMC method is proved.

### 4 Numerical examples

**Example 4. 1** Parameter estimation for the RASV model.

Stochastic volatility (SV) model is an essential tool for financial time series volatility. It identifies the time-varying return volatility in financial market by describing volatility as a random function. Nevertheless, the application of SV model is limited by the problem of evaluating their likelihood functions. To handle this problem, some Bayesian-based approach with the HMC method are used to accurately infer the parameters of SV model<sup>[7,8]</sup>. To compare the performance of the HMC and IHMC methods, we establish an SV model with realized volatility (RV) and asymmetric effect, abbreviated as RASV model<sup>[9]</sup>, by utilizing daily rate of return data and RV containing intraday high-frequency rate of return data information.

Suppose that we have  $m$  intraday returns during day  $t$ ,  $\{r_{t,i}\}_{i=1}^m$ , so that a simple RV estimator<sup>[10]</sup> is defined as the sum of squared returns,

*i. e.*,  $RV_t = \sum_{i=1}^m r_{t,i}^2$ . The RASV model is described as

$$\begin{aligned} R_t &= \sigma_\epsilon^2 e^{\frac{1}{2}\alpha_t} \epsilon_t, t=1, \dots, T, \\ Y_t &= c + \alpha_t + \sigma_u u_t, t=1, \dots, T, \\ \alpha_{t+1} &= \varphi \alpha_t + \sigma_\eta \eta_t, t=1, \dots, T-1, \\ \alpha_1 &\sim N\left(0, \frac{\sigma_\eta^2}{1 - \varphi^2}\right), \end{aligned}$$

$$\begin{bmatrix} \epsilon_t \\ u_t \\ \eta_t \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & \rho \\ 0 & 1 & 0 \\ \rho & 0 & 1 \end{bmatrix} \right),$$

where  $R = \{R_t\}_{t=1}^T, Y = \{Y_t\}_{t=1}^T, \alpha = \{\alpha_t\}_{t=1}^T$ . Let  $R$  and  $Y$  be two observation vectors,  $R_t$  is the rate of return over a unit time period, and  $Y_t$  is the logarithm of RV estimator,  $\alpha$  is the latent vector, and  $N(\cdot, \cdot)$  represent a normal distribution. Let  $\theta = (c, \sigma_u, \sigma_\epsilon, \varphi, \sigma_\eta, \rho)$ . The joint posterior density of the model can be written as

$$\begin{aligned} p(\theta, \alpha | R, Y) &= \prod_{t=1}^T p(R_t | \sigma_\epsilon, \alpha_t) \times \\ &\prod_{t=1}^T p(Y_t | c, \sigma_u) \times p(\theta) \times p(\alpha_1 | \varphi, \sigma_\eta) \times \\ &\prod_{t=2}^T p(\alpha_t | \alpha_{t-1}, \sigma_\epsilon, \varphi, \sigma_\eta, \rho). \end{aligned}$$

While estimating the RASV model, both the HMC method and the IHMC method are applied for parameters  $\alpha, \varphi, \sigma_\epsilon, \sigma_\eta, \rho$ , and parameters  $\sigma_u$  and  $c$  are sampled directly from the conditional posterior density.

The data used in this example are the 1-minute high-frequency data of Shanghai Composite Index from January 4, 2021 to February 21, 2022. We take the opening and closing prices of each day to calculate the daily return rate, and get the series  $R = \{R_t\}_{t=1}^T$ . The RV estimator was calculated by using the daily 1-minute high-frequency data, and the logarithm was taken to obtain the sequence  $Y = \{Y_t\}_{t=1}^T$ . We estimate the RASV model by running the corresponding 25 000 iterations of the HMC and IHMC methods and taking 5000 iterations as burn-in period. The prior dis-

tributions required in the joint prior density are set as

$$\begin{aligned} \sigma_\epsilon^2 &\sim IG(5, 4), \sigma_u^2 \sim IG(2.5, 0.025), \\ \sigma_\eta^2 &\sim IG(2.5, 0.025), \frac{\varphi+1}{2} \sim B(20, 1.5), \\ \frac{\rho+1}{2} &\sim B(1.5, 5), c \sim N(0, 10), \end{aligned}$$

where  $B(\cdot, \cdot)$  and  $IG(\cdot, \cdot)$  represent the Beta and inverse Gamma distributions, respectively. The choice of these prior distributions ensures that all parameters make sense. Particularly, the Beta prior for  $\varphi$  and  $\rho$  ensures that  $-1 < \varphi, \rho < 1$ . For initial parameter values, we set

$$\begin{aligned} \sigma_\epsilon^{(0)} &= 1, c^{(0)} = 0, \sigma_u^{(0)} = 0.1, \varphi^{(0)} = 0.95, \\ \sigma_\eta^{(0)} &= 0.1, \rho^{(0)} = -0.3, \alpha^{(0)} = \{\alpha_i^{(0)}\}_{i=1}^T, \end{aligned}$$

where

$$\alpha_1^{(0)} \sim N(0, \frac{\sigma_\eta^2}{1-\varphi^2}), \alpha_i^{(0)} \sim N(\varphi \alpha_{i-1}, \sigma_\eta^2).$$

Tuning parameters for the HMC method have been reported to be more difficult than tuning the parameters for other MCMC methods, because the performance of the HMC-based method is highly sensitive to two specified parameters, *i. e.*, the step size and the number of iterations. For its larger computation in the IHMC method, we use a larger step size to compensate. In the following simulation, we set iteration number 100 for both the HMC method and the IHMC method. For the parameter  $\alpha$ , we set step size 0.005 for the HMC method, while 0.01 for the IHMC method, and for the parameters  $\varphi, \sigma_\epsilon, \sigma_\eta, \rho$ , we set step size 0.006 for the HMC method, while 0.012 for the IHMC method. The estimation is implemented by R software.

Tab. 1 Numerical result of Example 4. 1 for different methods

Parameter	Mean	SD	Standard Error	95% interval	IF	Acceptance rates
Panel A: the HMC method						
$\varphi$	0.9971	0.0023	$1.61 \times 10^{-5}$	[0.9926, 0.9998]	54.9148	1
$\sigma_\epsilon$	0.1915	0.0060	$4.22 \times 10^{-5}$	[0.1822, 0.2018]	2640.993	0.824 80
$\sigma_\eta$	0.3897	0.0304	$2.15 \times 10^{-4}$	[0.3450, 0.4445]	22.6592	0.999 88
$\rho$	0.0846	0.0649	$4.59 \times 10^{-4}$	[-0.023 28, 0.190 30]	3.1630	0.999 44
$\sigma_u$	0.0977	0.0214	$1.51 \times 10^{-4}$	[0.063 74, 0.136 78]	206.3357	—
$c$	0.1393	0.1152	$8.14 \times 10^{-4}$	[-0.052 84, 0.332 29]	3217.453	—
Panel B: the IHMC method						
$\varphi$	0.9968	0.0023	$1.64 \times 10^{-5}$	[0.9924, 0.9997]	24.0917	1
$\sigma_\epsilon$	0.2131	0.0027	$1.91 \times 10^{-5}$	[0.220 91, 0.2177]	26.1327	0.998 94
$\sigma_\eta$	0.3611	0.0016	$1.12 \times 10^{-5}$	[0.3588, 0.3640]	3.3032	1
$\rho$	0.0822	0.0035	$2.44 \times 10^{-5}$	[0.077 02, 0.088 32]	7.5342	1
$\sigma_u$	0.0135	0.0006	$4.06 \times 10^{-6}$	[0.012 57, 0.014 45]	2.2205	—
$c$	0.8634	0.0624	$4.41 \times 10^{-4}$	[0.7701, 0.9679]	24.0893	—

Tab. 1 presents the result obtained by using the HMC method and the IHMC method. The IF is defined as  $IF = 1 + 2 \sum_{j=1}^{\infty} \rho_j$ , where  $\rho_j$  is the autocorrelation at lag  $j$ , which is used to illustrate the efficiency of the method and reflect the convergence rate. This quantity can be interpreted as the number of correlated samples with the same variance-reducing power as one independent sam-

ple. An IF value of 1 indicates that the draws are uncorrelated while large values indicate a slow convergence as well as a bad mixing. Note that Nugroho and Morimoto<sup>[11]</sup> have proven that the IFs are relatively not sensitive to the initial values. It can be seen that the resulting parameter estimates are quite close. By comparison, the SD value and standard error of the IHMC method is smaller, indicating that the fluctuation of variance

of the sampling result obtained by the IHMC method is smaller, and the sampling results are more reliable. Meanwhile, the IFs of the IHMC method are all lower than those of the HMC method, indicating that the convergence rate is

faster and the mixing effect is better. The acceptance rate of the IHMC method is obviously higher than that of the HMC method, and the acceptance rate is very close to 1, which shows that the IHMC method has a higher sampling efficiency.

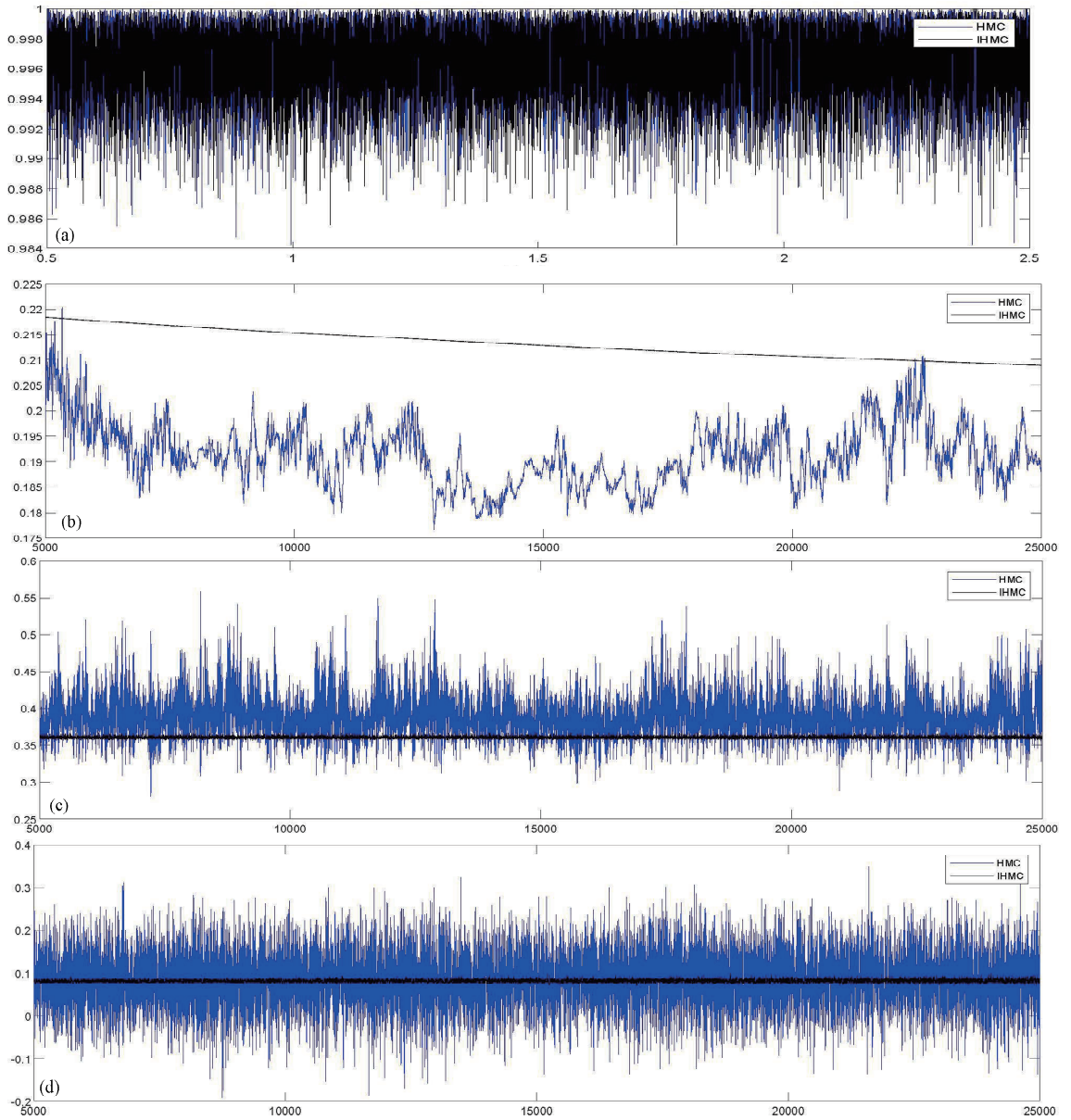


Fig. 1 Trace plots for the parameters  $\varphi, \sigma_\epsilon, \sigma_\eta, \rho$  of RASV model (from top to bottom) estimated by using the HMC method and the IHMC method

The trace plots and the autocorrelation plots for posterior samples of two methods are displayed in Figs. 1 and 2, respectively. The trace plots show that the sampling result of the IHMC is more stable. The autocorrelation plots exhibit quick decay of the autocorrelation as time lag be-

tween samples increases, indicating that the process is stationary. As observed in the autocorrelation plots, the attenuation of the autocorrelation of two HMC methods are at a good pace and have the similar performance. For two HMC methods, the autocorrelation attenuation and ac-

ceptance rate of  $\sigma_\epsilon$  is slower than other parameters, it may be that the  $\sigma_\epsilon$  parameter is highly sen-

sitive to the model, which needs further study.

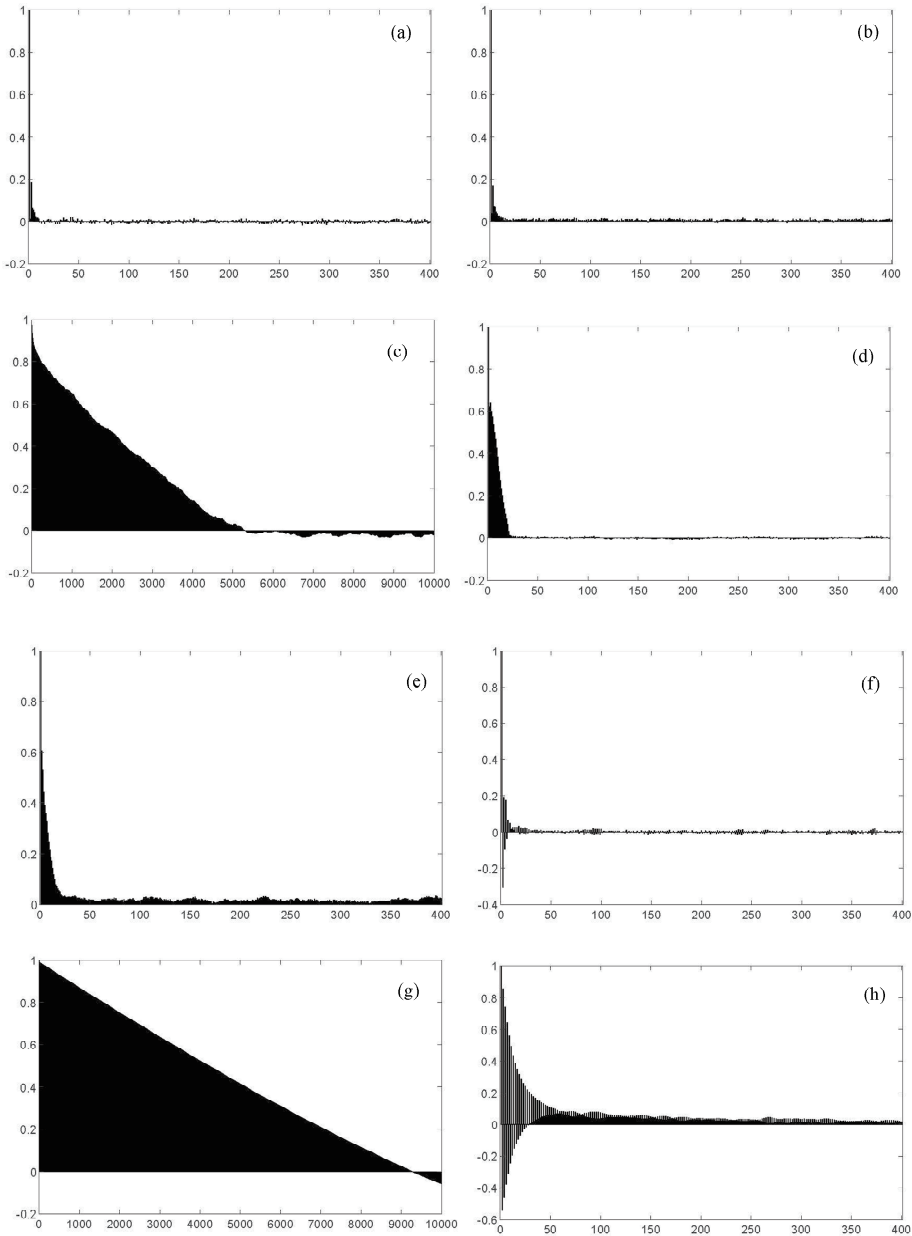


Fig. 2 Autocorrelation function of the parameters  $\varphi, \sigma_\epsilon, \sigma_\gamma, \rho$  (from top to bottom) estimated by the IHMC method (left) and the HMC method (right)

**Example 4.2** Sampling the variance Gamma distribution.

Financial time series noise often exhibits spiky and thick-tailed properties. we use the normal scale mixture distribution—Variance Gamma distribution (VGD) to verify the effectiveness of the IHMC method for sampling distributions that exhibit spiky and thick-tailed properties. The

VGD was proposed by Seneta *et al.*<sup>[12,13]</sup> to describe stock market returns. The density of VGD can be written as

$$\int_0^\infty N(x | \mu, \frac{\sigma^2}{\theta}) IG(\theta | \frac{\vartheta}{2}, \frac{\vartheta}{2}) d\theta.$$

That is to say, the VGD is equivalent to the following hierarchical form:

$$X | \mu, \sigma^2, \theta \sim N(x | \mu, \frac{\sigma^2}{\theta}) \text{ and } \theta | \vartheta \sim$$

$$IG\left(\theta \mid \frac{\vartheta}{2}, \frac{\vartheta}{2}\right).$$

Set the parameter  $(\mu, \sigma, \vartheta) = (6, 2, 5)$  and initial value  $x_0 = 2$  to sample the distribution using the HMC method and the IHMC method. Similar to Example 4.1, we set iteration number 100 for two methods, step size 0.25 for the HMC method and 0.5 for the IHMC method. The sampling re-

sult are shown in Tab. 2, Figs. 3 and 4. From the sampling result, it can be seen that the IHMC method is more stable and has higher acceptance rate. Comparing the IF values and the autocorrelation function plots in Fig. 4 reveals that the sampling result for the IHMC method has better convergence speed and mixing effect.

Tab. 2 Numerical result of Example 4.2 for different methods

method	Mean	SD	Standard Error	95% interval	IF	Acceptance rates
HMC	-0.0045	2.0137	$8.81 \times 10^{-3}$	$[-4.0308, 3.9892]$	3.3186	0.999 04
IHMC	$9.59 \times 10^{-5}$	$5.28 \times 10^{-5}$	$2.36 \times 10^{-4}$	$[-0.0657, 0.0699]$	2.0084	0.999 92

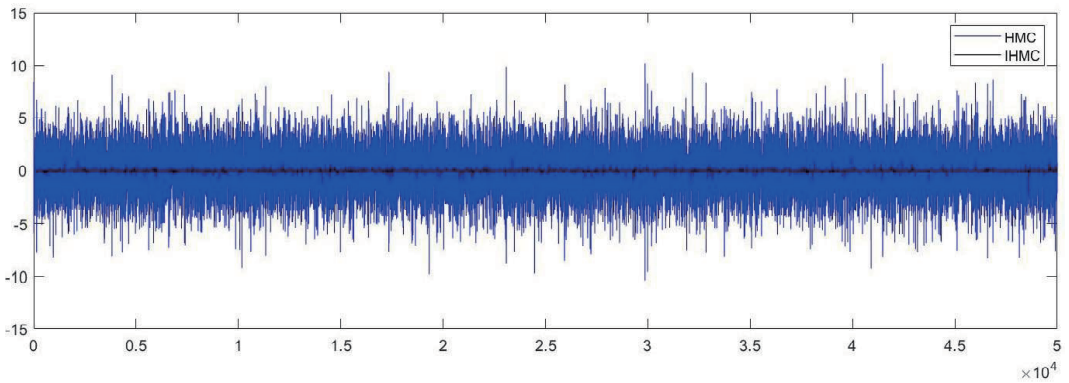


Fig. 3 Trace plots of VGD for different methods

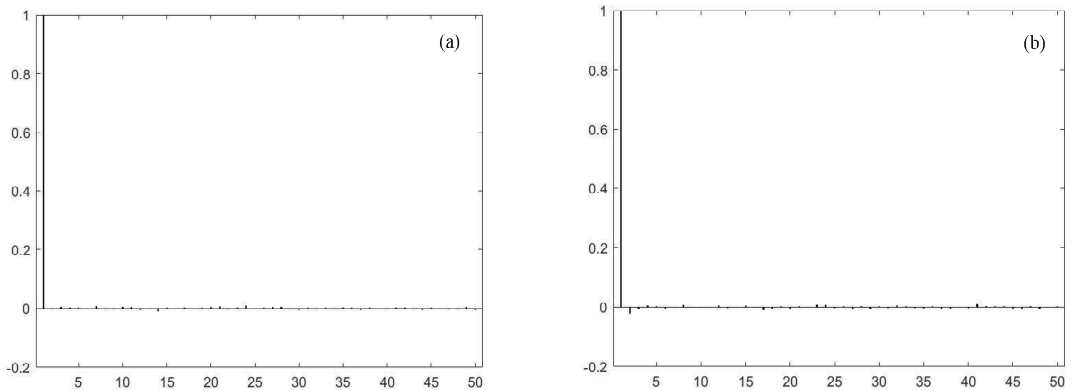


Fig. 4 Autocorrelation function of VGD estimated by the IHMC method (a) and the HMC method (b)

## 5 Conclusions

In this study, we have improved the HMC method by combining the Velocity Verlet integrator and proposed the IHMC method. By using two numerical examples, we show that the sampling results for the IHMC method are more stable, with better SD, standard error, acceptance

rate and IF values. Therefore, we may make the conclusion that the IHMC method has higher sampling efficiency and stability than the HMC method.

It is worth noting that it is also interesting to improve the HMC method by combining the Velocity Verlet integrator mentioned in this paper with the existing MHMC methods, which is the

major topic of our future work.

### References:

- [1] Duane S, Kennedy A D, Pendleton B J, *et al.* Hybrid Monte Carlo [J]. Phys Lett B, 1987, 195: 216.
- [2] Neal R M. Bayesian learning for neural networks [M]. New York: Springer, 1996.
- [3] Neal R M. MCMC using Hamiltonian dynamics [C]// Handbook of Markov chain Monte Carlo. Boca Raton: Chapman & Hall/CRC Press, 2011.
- [4] Radivojević T, Fernández-Pendás M, Sanz-Serna J M, *et al.* Multi-stage splitting integrators for sampling with modified Hamiltonian Monte Carlo methods [J]. J Comput Phys, 2018, 373: 900.
- [5] Swope, W C, Andersen H C, Berens P H, *et al.* A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: application to small water clusters [J]. J Chem Phys, 1982, 76: 637.
- [6] Bou-Rabee N, Sanz-Serna J M. Geometric integrators and the Hamiltonian Monte Carlo method [J]. Acta Numer, 2018, 27: 113.
- [7] Shephard N. Fitting nonlinear time-series models with applications to stochastic variance models [J]. J Appl Econ, 1993, 8: S135.
- [8] Jacquier E, Polson N G, Rossi P E. Bayesian analysis of stochastic volatility models [J]. J Bus Econ Stat, 2002, 20: 69.
- [9] Takahashi M, Omori Y, Watanabe T. Estimating stochastic volatility models using daily returns and realized volatility simultaneously [J]. Comput Stat Data An, 2009, 53: 2404.
- [10] Takahashi M, Watanabe T, Omori Y. Volatility and quantile forecasts by realized stochastic volatility models with generalized hyperbolic distribution [J]. Inter J Forecast, 2016, 32: 437.
- [11] Nugroho D B, Morimoto T. Estimation of realized stochastic volatility models using Hamiltonian Monte Carlo-Based methods [J]. Comput Stat, 2015, 30: 491.
- [12] Madan D B, Seneta E. The variance gamma (VG) model for share market returns [J]. J Bus, 1990: 511.
- [13] Seneta E. Fitting the variance-Gamma model to financial data [J]. J Appl Probab, 2004, 41: 177.