

Stochastic dynamics without detailed balance condition connecting simple gradient method and Hamiltonian Monte Carlo

Akihisa Ichiki^{1,*} and Masayuki Ohzeki^{2,3,4,†}

¹*Institutes of Innovation for Future Society, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan*

²*Graduate School of Information Sciences, Tohoku University, Sendai 980-8579, Japan*

³*Institute of Innovative Research, Tokyo Institute of Technology, Oh-okayama, Meguro-ku, Tokyo 152-8550, Japan*

⁴*Sigma-i, Co. Ltd., Konan, Minato-ku, Tokyo 108-0075, Japan*

(Dated: December 18, 2021)

Sampling occupies an important position in theories of various scientific fields, and Markov chain Monte Carlo (MCMC) provides the most common technique of sampling. In the progress of MCMC, a huge number of studies have aimed the acceleration of convergence to the target distribution. Hamiltonian Monte Carlo (HMC) is such a variant of MCMC. In the recent development of MCMC, another approach based on the violation of the detailed balance condition has attracted much attention. Historically, these two approaches have been proposed independently, and their relationship has not been clearly understood. In this paper, the two approaches are seamlessly understood in the framework of generalized Monte Carlo method that violates the detailed balance condition. Furthermore we propose an efficient Monte Carlo method based on our framework.

I. INTRODUCTION

Recently, sampling techniques have become of increasing importance in various fields of science and engineering. The sampling methods have been developed to numerically examine the equilibrium behaviors of complex systems such as macromolecules like proteins [1, 2], spin glasses [3], and glass transitions [4]. In addition to these traditional applications, with the background of the recent development of machine learning, sampling has become widely used for various purposes such as model training and its evaluation, and stochastic inference [5, 6].

The most common technique for sampling is provided by the Markov chain Monte Carlo (MCMC) method. MCMC is required to quickly sample random variables that follow an arbitrary target distribution starting from a given initial state. Since Metropolis *et al.* successfully introduced MCMC to investigate complex systems [7], many variants have been proposed to accelerate the convergence to the target distribution. The speed-up techniques have been constructed mainly based on two concepts. One is called the extended ensemble method [8]. In the extended ensemble method, the state space is extended by introducing auxiliary variables, and the convergence is accelerated by a proposal of a path in higher dimension allowing a rapid transition to the target distribution. The techniques of extended ensemble are roughly categorized into three groups: the exchange Monte Carlo [9], the simulated tempering [10, 11], and the multicanonical method [12] with the help of the Wang-Landau algorithm [13]. Hamiltonian Monte Carlo (HMC), which introduces momenta as auxiliary variables [14], is also classified as an extended ensemble

method. The alternative concept for acceleration is based on an efficient proposal of candidates for the updated state. Such efficient candidates are generated via the concept of the coarse-graining. The Swendsen-Wang algorithm [15] makes efficient state updates by using clusters of spins in the Ising model. This algorithm was later extended by Wolff to XY model [16], and is now extended to be applied to an arbitrary target distribution [17].

Recently, in addition to the above mentioned two concepts for acceleration, the possibility of detailed balance violation has been intensively investigated [18–21]. Conventional acceleration algorithms have been developed within the range of the detailed balance condition. However, it has been shown that the violation of the detailed balance accelerates convergence to the target distributions [22]. Based on this result, Ohzeki and Ichiki proposed a systematic construction of detailed balance-violating dynamics that converges to any target distribution in a continuous system [23]. The Ohzeki-Ichiki method duplicates the original system and introduces a probability current between the two systems. The driving force producing the probability current causes the rotational evolution of state in the duplicated state space. This is similar to the symplectic behavior of the Hamiltonian dynamics. In this paper, the Ohzeki-Ichiki method will be generalized, and it will be explained that the generalized Ohzeki-Ichiki method is indeed seamlessly connected to the Hamiltonian dynamics.

The generalized Ohzeki-Ichiki method provides a family of dynamics including the gradient method and the HMC. To show this fact, after reviewing the gradient method in section II, the HMC in section III, and the Ohzeki-Ichiki method in section IV, respectively, we will see that the generalized Ohzeki-Ichiki method contains the gradient method and the HMC as specific limits. In section V, the generalized Ohzeki-Ichiki method is numerically compared to other methods with respect to the speed of convergence to the target distribution. Sec-

* ichiki@chem.material.nagoya-u.ac.jp

† masayuki.ohzeki.a4@tohoku.ac.jp

tion VI is devoted to a summary and discussion.

II. GRADIENT METHOD

The simplest dynamics converging to the target distribution is given by a gradient method. The gradient method satisfies the so-called detailed balance condition. Physically, the dynamics with the detailed balance condition is relaxed to a steady state in which no macroscopic heat is generated. Such a special steady state is called an equilibrium state. By the gradient method, the Gibbs distribution

$$\pi(x) = \exp[-U(x)/T]/Z \quad (1)$$

with a partition function Z is achieved with the balance between the energy gradient and the diffusion due to noise. The following dynamics gives the simplest gradient method in which the N -dimensional continuous state x converges to the Gibbs distribution:

$$dx_i(t) = -\frac{\partial U}{\partial x_i} dt + \sqrt{2T} dW_i(t), \quad (2)$$

where, dx_i is the displacement of x_i during an infinitesimal time dt , and $U(x)$ and T correspond to the potential and temperature, respectively. $W_i(t)$ is a standard Wiener process that satisfies

$$\langle dW_i(t) \rangle = 0, \quad (3)$$

$$\langle dW_i(t) dW_j(t') \rangle = \delta_{ij} \delta(t - t') dt, \quad (4)$$

where δ_{ij} and $\delta(t)$ denote Kronecker and Dirac delta functions, respectively, and $\langle \cdot \rangle$ represents an expectation. The Fokker-Planck equation corresponding to the Langevin equation (2) is given as

$$\frac{\partial P(x, t)}{\partial t} = -\sum_i \frac{\partial}{\partial x_i} \left[-\frac{\partial U(x)}{\partial x_i} - T \frac{\partial}{\partial x_i} \right] P(x, t). \quad (5)$$

It is straightforwardly confirmed that the Gibbs distribution (1) is the steady solution satisfying the Fokker-Planck equation (5).

It is guaranteed by the H-theorem that the dynamics (2) converges to a unique steady distribution (1) as an equilibrium distribution regardless of an initial condition. Therefore, the target Gibbs distribution can be obtained by providing $U(x)$ and T in the simple gradient dynamics (2). However, since the simple gradient method updates the state along the gradient of the potential U , the update becomes inefficient when the state is trapped in a local minimum of the potential, where the gradient vanishes. To escape from such a local minimum, noise is exploited in MCMC algorithms. However, if the potential around the local minimum is steep, it takes a long time to escape from the local minimum. In the history of MCMC studies, various techniques have been proposed to avoid such a bottleneck restricting the relaxation to the target distribution.

III. HAMILTONIAN MONTE CARLO

We have seen that, in the simple gradient method, the state is updated in the direction along the gradient of the potential, which is normal to the energy surface. With such a method, it is difficult to avoid to be trapped in the local minimum of the potential. To overcome this difficulty, it has been proposed to add extra degrees of freedom to the original system to make new directions to escape from the local minimum of the potential. This idea is called an extended ensemble method. A method called Hamiltonian Monte Carlo (HMC) is one of the realizations of the extended ensemble methods. In the HMC, in addition to the original state variable x , a momentum p is introduced as an auxiliary variable. By introducing the momentum, the dimension of the dynamical system doubles, and it becomes easier to escape from the local minimum of the potential. In other words, when the kinetic energy exceeds the energy gap between the local minimum and the local maximum of the potential $U(x)$, the state can escape from the local minimum of the potential. The basic concept of the HMC is that the Gibbs distribution

$$\pi_{x,p}(x, p) = \exp[-H(x, p)/T]/Z_{x,p}, \quad (6)$$

$$H(x, p) = U(x) + \sum_i \frac{p_i^2}{2m_i} \quad (7)$$

is invariant under the Hamiltonian dynamics

$$\dot{x}_i = \frac{p_i}{m_i}, \quad (8)$$

$$\dot{p}_i = -\frac{\partial U(x)}{\partial x_i}, \quad (9)$$

where $Z_{x,p} := \int dx dp \exp[-H(x, p)/T]$ is a partition function. Here, m_i represents the mass of the i -th degree of freedom. The target Gibbs distribution $\pi(x) = \exp[-U(x)/T]/Z$ is acquired as a marginal distribution $\pi(x) = \int dp \pi_{x,p}(x, p)$ via the Gibbs distribution (6).

The algorithm of the HMC consists of the following steps. (i) Sample the momentum p'_i ($i = 1, \dots, N$) from the Gaussian distribution

$$P_G(p'_i) = \frac{1}{\sqrt{2\pi m_i T}} \exp\left[-\frac{p'^2_i}{2m_i T}\right]. \quad (10)$$

This procedure changes the state from (x, p) to (x, p') . (ii) Evolve the state for waiting time τ starting from the initial state (x, p') according to the Hamiltonian dynamics (8) and (9). We denote the obtained state as (x'', p'') . (iii) According to the Metropolis-Hasting rule [7, 24], the state obtained in the step (ii), (x'', p'') , is accepted with the acceptance rate $\min[1, \exp\{-[H(x'', p'') - H(x, p')]/T\}]$. Otherwise, the state remains at (x, p') . The algorithm of the HMC consists of a repetition of these three steps.

Note that the Gibbs distribution (6) is invariant under the Hamiltonian dynamics (8) and (9). In particular, the

Gaussian distribution (10) gives the steady state distribution for the momentum. In step (i), the momentum p is sampled from this invariant distribution. The advantage of the HMC is that the Gaussian random variables can be easily generated in numerical manners. In step (ii), the state update is ballistic on the energy surface. Even if the state is located at the local minimum of the potential $U(x)$, it is possible to escape from it by the effect of kinetic energy. The rejection in step (iii) is exploited to eliminate nonphysical time evolution [25]. Since the total energy is conserved under the Hamiltonian dynamics, the acceptance rate is theoretically always unity. However, naive numerical calculations have been reported to show an increase in total energy. The step (iii) is introduced to eliminate this possibility to guarantee the calculation accuracy. Thus, step (iii) is extra and can be omitted when the time evolution of the Hamiltonian dynamics is calculated with sufficiently high accuracy.

In the simple gradient method (2), the state update in the normal direction of the energy surface is ballistic. The update on the energy surface is diffuse, since the state update on the energy surface is caused only by noise. On the other hand, in the HMC, the update in the normal direction of the energy surface is caused only by the random sampling of momentum. However, the update on the energy surface is ballistic since the state evolves according to the Hamiltonian dynamics. The Gibbs distribution obeys the principle of equal a priori weights for states with equal energy. The HMC is expected to quickly satisfy the principle of equal a priori weights by the ballistic state updates on the energy surface.

IV. OHZEKI-ICHIKI METHOD

The violation of the detailed balance condition was shown to accelerate relaxation to the steady state due to the eigenvalue shift for the Fokker-Planck operator [22]. In order to systematically introduce the violation of the detailed balance condition, Ohzeki and Ichiki have proposed to duplicate the original system to introduce a rotating probability current between the two duplicated systems:

$$dx_i(t) = -\frac{\partial U(x)}{\partial x_i} dt + \gamma \frac{\partial U(y)}{\partial y_i} dt + \sqrt{2T} dW_i^x(t), \quad (11)$$

$$dy_i(t) = -\frac{\partial U(y)}{\partial y_i} dt - \gamma \frac{\partial U(x)}{\partial x_i} dt + \sqrt{2T} dW_i^y(t), \quad (12)$$

where x_i and y_i are degrees of freedom belonging to the original and the replicated system, respectively. W_i^x and W_i^y are independent standard Wiener processes:

$$\langle dW_i^x(t) dW_j^x(t') \rangle = \delta_{ij} \delta(t - t') dt, \quad (13)$$

$$\langle dW_i^y(t) dW_j^y(t') \rangle = \delta_{ij} \delta(t - t') dt, \quad (14)$$

$$\langle dW_i^x(t) dW_j^y(t') \rangle = 0. \quad (15)$$

This system has the steady state distribution of Gibbsian form

$$\pi_{x,y}(x, y) = \exp \{-\beta [U(x) + U(y)]\} / Z_{x,y}, \quad (16)$$

where $\beta = 1/T$, and $Z_{x,y}$ is a partition function. Then, the target distribution $\pi(x) = \exp[-U(x)/T] / Z$ is acquired as the marginal distribution $\pi(x) = \int dy \pi_{x,y}(x, y)$. Note that this system violates the detailed balance condition, but satisfies the balance condition

$$\sum_i \frac{\partial}{\partial x_i} u_i^x \pi(x, y) + \sum_i \frac{\partial}{\partial y_i} u_i^y \pi(x, y) = 0, \quad (17)$$

where the driving force

$$u_i^x = \gamma \frac{\partial U(y)}{\partial y_i}, \quad (18)$$

$$u_i^y = -\gamma \frac{\partial U(x)}{\partial x_i} \quad (19)$$

yields the probability current characteristic to the violation of the detailed balance. The introduction of the driving force satisfying the balance condition remains the Gibbs distribution (16) to be the steady state distribution. Although the two duplicated systems affect each other via the driving force, the steady state distribution for each system is independent.

In the Ohzeki-Ichiki dynamics (11) and (12), the same form of the potential in the original x -system is chosen as that in the duplicated y -system. However, there is arbitrariness in the choice of the potential in the y -system, since y is an auxiliary variable and the target distribution is given as the marginal distribution $\pi(x) = \int dy \pi_{x,y}(x, y)$. Therefore, the potential in the y -system does not have to be the same as that of the x -system. Consider the following dynamics:

$$dx_i(t) = \left[-\frac{\partial H_x(x)}{\partial x_i} + \gamma \frac{\partial H_y(y)}{\partial y_i} \right] dt + \sqrt{2T} dW_i^x(t), \quad (20)$$

$$dy_i(t) = \left[-\frac{\partial H_y(y)}{\partial y_i} - \gamma \frac{\partial H_x(x)}{\partial x_i} \right] dt + \sqrt{2T} dW_i^y(t), \quad (21)$$

where $H_x(x) = U(x)$ is the potential in the original x -system, and the energy $H_y(y)$ in the y -system can be in the form of an arbitrary function. This system has the following steady state distribution independent of the value of γ :

$$\pi_{x,y}(x, y) = \exp \{-\beta [H_x(x) + H_y(y)]\} / Z_{x,y}. \quad (22)$$

Therefore, the target distribution is obtained as a marginal distribution $\pi(x) = \int dy \pi_{x,y}(x, y)$ for an arbitrary form of H_y .

Consider the change of variables in dynamics (20) and (21) as $\gamma = \tilde{\gamma}T$, $\tilde{t} = \tilde{\gamma}Tt$. Then the dynamics

$$dx_i(\tilde{t}) = \frac{\partial H_y(y)}{\partial y_i} d\tilde{t}, \quad (23)$$

$$dy_i(\tilde{t}) = -\frac{\partial H_x(x)}{\partial x_i} d\tilde{t} \quad (24)$$

is obtained in the limit of $\tilde{\gamma} \rightarrow \infty$. Note that H_x and H_y play the roles of potential and kinetic energies in this dynamics, respectively. In fact, the choice of $H_y(y) = \sum_i y_i^2/2m_i$ reproduces the Hamiltonian dynamics (8) and (9). In dynamics (20) and (21), the driving force proportional to γ causes the violation of the detailed balance condition. The case of $\gamma = 0$ corresponds to the simple gradient method. On the other hand, the dynamics in the limit $\gamma \rightarrow \infty$ corresponds to the Hamiltonian dynamics. Thus, it is concluded that the dynamics (20) and (21) seamlessly connects the gradient method and the Hamiltonian dynamics that is the basis of the HMC.

V. HYBRID USE OF GRADIENT METHOD AND HAMILTONIAN DYNAMICS

In the previous section, we have introduced the dynamics, which incorporates the simple gradient method and the Hamiltonian dynamics. By the simple gradient method, the state update on the energy surface is realized diffusely, and it takes a long time to satisfy the principle of equal a priori weights. On the other hand, in the HMC, the state update on the energy surface is so ballistic that the principle of equal a priori weights is quickly satisfied. However, since the total energy is conserved under the Hamiltonian dynamics, transitions between energy surfaces are prohibited. For this reason, the HMC requires resampling of momentum from the Gaussian distribution (10) which is realized in the steady state.

Consider the case of finite γ in the dynamics (20) and (21) with harmonic H_y that connects the simple gradient method and the Hamiltonian dynamics. In such a dynamics, the state update on the energy surface, which has been a bottleneck of relaxation to the steady state in the simple gradient method, is realized to become ballistic. In addition, the effects of gradients and noise automatically enhance transitions between energy surfaces. Therefore, it is not required to resample the momentum, unlike the case of conventional HMC.

To demonstrate the performance of our proposed method, i.e., the dynamics with harmonic H_y , we first deal with a toy model of a one-dimensional double-well potential:

$$U(x) = \frac{1}{4}x^4 - \frac{1}{2}x^2. \quad (25)$$

The initial condition is set to be in one of the potential wells at $x = 1$. Thus, the system must go beyond the potential hill at $x = 0$ to realize the steady state. In our numerical calculations, we set the temperature as $T = 1.0$. The infinitesimal time-step is set to be $dt = 1.0 \times 10^{-4}$. We compare the performance of the simple gradient method, the conventional HMC, the conventional Ohzeki-Ichiki method, namely, the dynamics (20) and (21) with $H_x(x) = H_y(x) = U(x)$, and the proposed dynamics (20) and (21) with $H_x(x) = U(x)$

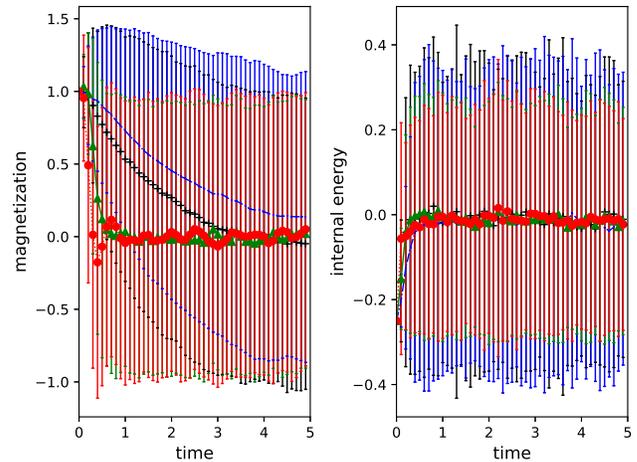


FIG. 1. (Color online) Time evolution of the state $\langle x \rangle$ (left panel) and the internal energy (right panel). The black cross marks, blue dots, green triangles, and red circles indicate the results of the simple gradient method, conventional HMC, conventional Ohzeki-Ichiki method, and the proposed method, respectively. The error bars indicate variances.

and $H_y(y) = y^2/2m$. The time evolution of the Langevin equations is calculated by applying the Heun scheme [26]. The time evolution of the Hamiltonian dynamics in the HMC is calculated using the leapfrog method [25]. Other parameters are set as follows: In the HMC, the particle mass is set as $m = 1$. In the algorithm of the HMC, it is necessary to evolve the Hamiltonian dynamics by a certain waiting time τ_{wait} before resampling the momentum. We set the waiting time as $\tau_{\text{wait}} = 0.01$. In the Ohzeki-Ichiki method, the parameter γ characterizing the violation of the detailed balance condition is set as $\gamma = 10.0$. In the generalized Ohzeki-Ichiki method where $H_y(y)$ is harmonic, the particle mass is set as $m = 1.0$. The value of $\gamma = 10.0$ is also chosen in this dynamics. Figure 1 shows the numerical results averaged over $N_{\text{sample}} = 1000$ independent runs taking time average during $\Delta t = 0.1$. The Ohzeki-Ichiki method shows faster convergence to the steady state than the simple gradient method because of the detailed balance violation. Furthermore, it can be seen that the convergence of the proposed dynamics with harmonic potential for y is faster than the Ohzeki-Ichiki dynamics, since the potential of the y -system is complicated in the conventional Ohzeki-Ichiki method. In the HMC, relaxation depends on the waiting time τ_{wait} . The larger τ_{wait} , the smaller the number of Monte Carlo steps is required for convergence. However, as seen in Fig. 1, it requires a longer calculation time, which is given by the product of τ_{wait} and the Monte Carlo steps in HMC, than other methods. As seen in the previous section, the timescale conversion in the Ohzeki-Ichiki dynamics reproduces the Hamiltonian dynamics. Due to the limit of this timescale conversion, it is difficult to make a direct comparison between the HMC and the Ohzeki-Ichiki method. In fact, in the limit of $\tilde{\gamma} \rightarrow \infty$, $d\tilde{t}$ corresponding

to the infinitesimal time step dt diverges. This means that one Monte Carlo step in the Ohzeki-Ichiki method should be compared with the result of the HMC with the limit of long waiting time $\tau_{\text{wait}} \rightarrow \infty$.

We also evaluate the integrated auto-correlation time $\tau_{\text{int}} := \int_0^\infty dt' [\langle x(t)x(t+t') \rangle - \langle x \rangle^2] / [\langle x^2 \rangle - \langle x \rangle^2]$. The integrated auto-correlation time for each dynamics is evaluated by the empirical average after the convergence to the steady state. We obtain $\tau_{\text{int}} = 2.00$ for the simple gradient method, which corresponds to the dynamics with $\gamma = 0$, $\tau_{\text{int}} = 0.19$ for the conventional Ohzeki-Ichiki method with $\gamma = 10.0$, and $\tau_{\text{int}} = 0.14$ for the proposed hybrid use of the gradient method and the Hamiltonian dynamics with $\gamma = 10.0$, respectively. In addition to the convergence of $\langle x \rangle$ and $\langle U(x) \rangle$ shown in Fig. 1, these results imply that the proposed method leads the significant reduction of the relaxation time to the steady state.

To demonstrate the removal of the critical slowing down in our method, we next deal with the two-dimensional XY model on a square lattice:

$$U(x) = - \sum_{\langle i,j \rangle} \cos(x_i - x_j), \quad (26)$$

where the sum is taken over all pairs of the nearest neighboring sites. The two-dimensional XY model exhibits the Kosterlitz-Thouless transition at $T_c = 0.89213(10)$ [27]. At temperatures below T_c , magnetization $m = \sum_{i=1}^N \sin x_i / N$ exhibits slow relaxation following the power law decay [28]. Since the critical slowing down is a bottleneck for convergence to the targeted steady state, it is preferred to avoid such slowing down behaviors.

We compare the convergence performance of the gradient method, the Ohzeki-Ichiki method, and the proposed method, in which the potential of the y -system is given by $H_y(y) = \sum_{i=1}^N y_i^2 / 2m_i$. In our numerical calculations, the number of spins is set to be $N = 10 \times 10$. According to the finite size correction, the effective critical temperature for this system is evaluated as $T_{\text{ceff}} \sim 0.975$ [29]. To demonstrate the removal of the critical slowing down, the temperature is set to be $T = 0.5 < T_{\text{ceff}}$. Other parameters are set as follows: the mass in the proposed dynamics is set as $m_i = 1.0$ for each $i = 1, \dots, N$. The parameter $\gamma = 5.0$ is chosen for both cases of $H_y(y) = U(y)$ and $H_y(y) = \sum_{i=1}^N y_i^2 / 2m_i$. The state with all spins in up-state, i.e., $x_i = \pi/2$ for all $i = 1, \dots, N$ is chosen as the initial state. The infinitesimal time step is set to be $dt = 1.0 \times 10^{-4}$. The Langevin equations are integrated by the Heun scheme.

Figure 2 shows the results averaged over $N_{\text{sample}} = 1000$ independent runs taking time average during $\Delta t = 0.1$. Although the simple gradient dynamics exhibits the critical slowing down, the Ohzeki-Ichiki and our proposed dynamics show faster convergence. The conventional Ohzeki-Ichiki method provides faster relaxation than the gradient method, and the significant improve-

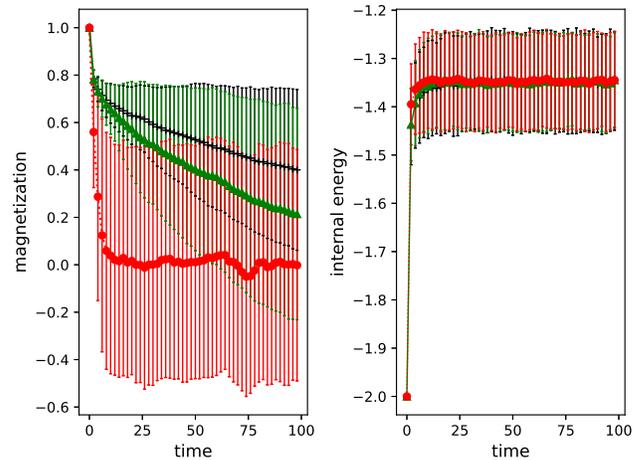


FIG. 2. (Color online) Time evolution of the magnetization (left panel) and internal energy (right panel) of XY model. The black cross marks, green triangles, and red circles indicate the results of the gradient method, conventional Ohzeki-Ichiki method, and the proposed method, respectively. The error bars indicate variances.

ment is achieved by the proposed method. In the proposed method, both magnetization and internal energy rapidly converge to the steady state values, and the critical slowing down appears to be eliminated.

VI. SUMMARY AND DISCUSSION

We have seen that the Ohzeki-Ichiki method seamlessly connects the simple gradient method with the Hamiltonian dynamics. The Hamiltonian dynamics corresponds to a specific limit of the generalized Ohzeki-Ichiki method. The HMC does not satisfy the detailed balance condition in general. In the HMC, the candidate of the updated state depends on the waiting time, which defines a leapfrog operator \hat{L} . Even if the updated state $(x', p') = \hat{L}(x, p)$ is proposed starting from the current state (x, p) by the leapfrog operator, the reverse transition $(x', p') \rightarrow (x, p)$ is not necessarily proposed. In other words, $(x, p) = \hat{L}(x', p')$ is not satisfied in general. The HMC with the detailed balance condition can be realized using a leapfrog operator \hat{L} adaptively defined with an appropriate waiting time [30]. In contrast, our proposed method does not constrain any leapfrog operator. The detailed balance condition is not satisfied, but the balance condition is in any timescale.

The convergence performance of the HMC strongly depends on the waiting time. The method called No-U-Turn Sampler (NUTS) adaptively determines the waiting time and efficiently proposes the updated state [31]. This technique accelerates the convergence speed to the steady state with respect to Monte Carlo steps. However, in the HMC, a sufficient number of Monte Carlo steps are required for convergence, since the transition

between energy surfaces occurs only when the momentum is resampled every single Monte Carlo step. Conversely, to shorten the simulation time, it is required to shorten the time of single Monte Carlo step, i.e., the waiting time. Note that the simulation time is proportional to the actual calculation time. Due to the trade-off relationship between the waiting time and Monte Carlo steps in the HMC, it takes a long simulation time to converge to the steady state. In contrast, our proposed method shows faster relaxation in simulation time. Therefore, it is concluded that our method shows better performance than the HMC with respect to actual calculation time.

It is worth mentioning that our method can be used in combination with other methods. For example, it can be used with the exchange Monte Carlo method and coarse-grained dynamics. In addition to this advantage, our proposed method can be easily applied to existing algorithms with detailed balance condition to improve their convergence speed. It is required only to add the driving force generating probability current and momentum dynamics to the original dynamics with the detailed balance condition.

In the conventional Ohzeki-Ichiki method, the steady state distribution for the auxiliary variable y has the same form as that for the original system. Thus, y can be directly used to evaluate expectations for the target distribution. Note that the auxiliary variable y doubles the number of samples in the procedure for evaluating the expectation in empirical manner. The obtained empirical average has less variance than that obtained by the dynamics with $H_y(y) \neq U(y)$. On the other hand, the choice of harmonic H_y exhibits faster convergence,

but the auxiliary variable y cannot be directly used to evaluate the expectations for the target distribution.

Since our method exploits the violation of the detailed balance condition, the probability current is generated in the system. The probability current realizes a biased sampling, resulting in accelerated convergence to the target distribution [32]. It is known in such a system that the convergence of the long-time average of physical quantities, namely, the empirical average, to the ensemble average is accelerated [33].

Note that the choice of H_y still has some arbitrariness. In the conventional Ohzeki-Ichiki method, H_y is chosen as the potential of the original system. In the method proposed in this paper, H_y is chosen as a harmonic one, which is the bridge between the gradient method and the HMC. However, $H_y(y)$ can be a function of an arbitrary form. We have seen that $H_y(y) = U(y)$ and $H_y(y) = \sum_i y_i^2/2m_i$ show different convergence performance to the target distribution. The performance of the dynamics (20) and (21) depends on the choice of H_y . Since the harmonic H_y has only a single energy valley, it is expected to be relaxed quickly. Thus, the relaxation of the variable x belonging to the original system is also expected to be accelerated. However, a detailed discussion of the optimal H_y is a matter for the future. For example, it remains an open problem whether the optimal H_y for convergence depends on H_x .

ACKNOWLEDGMENTS

A. Ichiki was supported by JSPS KAKENHI Grants No. JP17H06469.

-
- [1] C. Schütte, A. Fischer, W. Huisinga, and P. Deuffhard, A direct approach to conformational dynamics based on hybrid monte carlo, *J. Comput. Phys.* **151**, 146–168 (1999).
 - [2] A. Mitsutake, Y. Sugita, and Y. Okamoto, Generalized-ensemble algorithms for molecular simulations of biopolymers, *Biopolymers* **60**, 96 (2001).
 - [3] A. T. Ogielski, Dynamics of three-dimensional ising spin glasses in thermal equilibrium, *Phys. Rev. B* **32**, 7384 (1985).
 - [4] R. Yamamoto and W. Kob, Replica-exchange molecular dynamics simulation for supercooled liquids, *Phys. Rev. E* **61**, 5473 (2000).
 - [5] C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan, *Machine Learning* **50**, 5 (2003).
 - [6] D. J. C. MacKay, *Information Theory, Inference & Learning Algorithms* (Cambridge University Press, USA, 2002).
 - [7] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, *The Journal of Chemical Physics* **21**, 1087 (1953).
 - [8] Y. Iba, Extended ensemble Monte Carlo, *International Journal of Modern Physics C* **12**, 623 (2001).
 - [9] K. Hukushima and K. Nemoto, Exchange monte carlo method and application to spin glass simulations, *Journal of the Physical Society of Japan* **65**, 1604 (1996).
 - [10] E. Marinari and G. Parisi, Simulated tempering: A new monte carlo scheme, *Europhysics Letters (EPL)* **19**, 451 (1992).
 - [11] A. P. Lyubartsev, A. A. Martynovskii, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, New approach to monte carlo calculation of the free energy: Method of expanded ensembles, *The Journal of Chemical Physics* **96**, 1776 (1992).
 - [12] B. A. Berg and T. Neuhaus, Multicanonical ensemble: A new approach to simulate first-order phase transitions, *Phys. Rev. Lett.* **68**, 9 (1992).
 - [13] F. Wang and D. P. Landau, Efficient, multiple-range random walk algorithm to calculate the density of states, *Phys. Rev. Lett.* **86**, 2050 (2001).
 - [14] S. Duane, A. Kennedy, B. J. Pendleton, and D. Roweth, Hybrid monte carlo, *Physics Letters B* **195**, 216 (1987).
 - [15] R. H. Swendsen and J.-S. Wang, Nonuniversal critical dynamics in monte carlo simulations, *Phys. Rev. Lett.* **58**, 86 (1987).

- [16] U. Wolff, Collective monte carlo updating for spin systems, *Phys. Rev. Lett.* **62**, 361 (1989).
- [17] A. Barbu and S.-C. Zhu, Generalizing swendsen-wang to sampling arbitrary posterior probabilities, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **37**, 1239 (2015).
- [18] H. Suwa and S. Todo, Markov chain monte carlo method without detailed balance, *Phys. Rev. Lett.* **105**, 120603 (2010).
- [19] K. S. Turitsyn, M. Chertkov, and M. Vucelja, Irreversible monte carlo algorithms for efficient sampling, *Physica D: Nonlinear Phenomena* **240**, 410 (2011).
- [20] H. C. Fernandes and M. Weigel, Non-reversible monte carlo simulations of spin models, *Computer Physics Communications* **182**, 1856 (2011), *computer Physics Communications Special Edition for Conference on Computational Physics Trondheim, Norway, June 23-26, 2010*.
- [21] Y. Sakai and K. Hukushima, Dynamics of one-dimensional ising model without detailed balance condition, *Journal of the Physical Society of Japan* **82**, 064003 (2013), <https://doi.org/10.7566/JPSJ.82.064003>.
- [22] A. Ichiki and M. Ohzeki, Violation of detailed balance accelerates relaxation, *Phys. Rev. E* **88**, 020101 (2013).
- [23] M. Ohzeki and A. Ichiki, Langevin dynamics neglecting detailed balance condition, *Phys. Rev. E* **92**, 012105 (2015).
- [24] W. K. Hastings, Monte carlo sampling methods using markov chains and their applications, *Biometrika* **57**, 97 (1970).
- [25] E. Hairer, C. Lubich, and G. Wanner, Geometric numerical integration illustrated by the störmer-verlet method, *Acta Numerica* **12**, 399–450 (2003).
- [26] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer, Heidelberg, 1992).
- [27] P. Olsson, Monte carlo analysis of the two-dimensional xy model. ii. comparison with the kosterlitz renormalization-group equations, *Phys. Rev. B* **52**, 4526 (1995).
- [28] H. Nishimori and G. Ortiz, *Elements of Phase Transitions and Critical Phenomena* (Oxford University Press, 2010).
- [29] Y. Komura and Y. Okabe, Large-scale monte carlo simulation of two-dimensional classical xy model using multiple gpus, *Journal of the Physical Society of Japan* **81**, 113001 (2012), <https://doi.org/10.1143/JPSJ.81.113001>.
- [30] M. Okudo and H. Suzuki, Hamiltonian monte carlo with explicit, reversible, and volume-preserving adaptive step size control, *JSIAM Letters* **9**, 33 (2017).
- [31] M. D. Hoffman and A. Gelman, The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo, *Journal of Machine Learning Research* **15**, 1593 (2014).
- [32] A. Ichiki and M. Ohzeki, Full-order fluctuation-dissipation relation for a class of nonequilibrium steady states, *Phys. Rev. E* **91**, 062105 (2015).
- [33] F. Coghi, R. Chetrite, and H. Touchette, Role of current fluctuations in nonreversible samplers, *Physical Review E* **103**, 10.1103/physreve.103.062142 (2021).