Eigenvalue analysis of an irreversible random walk with skew detailed balance conditions

Yuji Sakai\textsuperscript{1,1} and Koji Hukushima\textsuperscript{1,2,1}\textsuperscript{†}

\textsuperscript{1}Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan
\textsuperscript{2}Center for Materials Research by Information Integration, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan

(Dated: December 3, 2015)

An irreversible Markov-chain Monte Carlo (MCMC) algorithm with skew detailed balance conditions originally proposed by Turitsyn et al. is extended to general discrete systems on the basis of the Metropolis-Hastings scheme. To evaluate the efficiency of our proposed method, the relaxation dynamics of the slowest mode and the asymptotic variance are studied analytically in a random walk on one dimension. It is found that the performance in irreversible MCMC methods violating the detailed balance condition is improved by appropriately choosing parameters in the algorithm.

PACS numbers: 02.50.-r, 05.10.Ln, 02.70.Tt, 05.70.Ln

I. INTRODUCTION

Markov-chain Monte Carlo (MCMC) methods have already been applied to numerous problems in various fields such as physics, biochemistry, information sciences, and economics \cite{1, 2}. Ever since Metropolis et al. invented the MCMC method in 1953 \cite{3}, many kinds of improved MCMC methods have been proposed and some of them have contributed to the development of sciences, especially to the understanding of phase transitions and critical phenomena in statistical physics.

Applying an MCMC method to a problem requires the preparation of a target distribution and a transition matrix in a Markov chain. For ascertaining the efficiency of MCMC methods, it is important to investigate whether the distribution in a Markov chain converges to the stationary target distribution rapidly and whether the correlation between samples from the Markov chain is sufficiently small. The multicanonical method \cite{4} and the replica-exchange Monte Carlo method \cite{5} improve the convergence rate quantitatively by extending the target distribution. Cluster algorithms, such as the Swendsen-Wang algorithm \cite{6} and the Wolff algorithm \cite{7}, yield a remarkable reduction in the critical slowing down of spin systems by choosing an appropriate transition matrix with a multi-spin update. The correlation in a Markov chain, which is worth considering in addition to the convergence rate, is characterized by an asymptotic variance. This asymptotic variance can be reduced by decreasing the rejection rate in a Markov chain in MCMC methods with the detailed balance condition (DBC); this is known as Peskun’s theorem \cite{8}. As mentioned in the following sections, both the convergence rate and the asymptotic variance are closely related to the second-largest eigenvalue of the transition matrix. Thus, improving the efficiency in MCMC methods is equivalent to reducing the second-largest eigenvalue of the transition matrix in the corresponding Markov chain.

The detailed balance condition is usually imposed upon the transition matrix to guarantee that our target distribution is exactly the stationary distribution in the Markov chain. Conventional MCMC methods, such as the Metropolis-Hastings algorithm \cite{9} and heat-bath method \cite{10}, have been developed within the framework of the DBC. However, it is possible to construct such a Markov chain without imposing the DBC because the DBC is not always necessary but sufficient to make the MCMC method work correctly. Furthermore, it is unclear whether MCMC methods with the DBC are more efficient than those without the DBC. In fact, several studies have shown numerically and partly analytically that the efficiency of MCMC methods can be improved by violating the DBC in some cases \cite{12, 22}. Hence MCMC methods without the DBC have been eagerly studied recently.

Strategies used to violate the DBC are roughly divided into two types. One is to reduce the rejection rate in the Markov chain, as proposed by Suwa and Todo \cite{13, 14}. They constructed a Markov chain without the DBC by using a geometric allocation approach, showing numerically that their method reduces the autocorrelation time by a factor of more than 6 for four- and eight-state Potts models at the transition temperature. The other strategy is to extend the state space and the target distribution \cite{15, 23}. Such a strategy is called “lifting,” and an irreversible Markov chain on the extended state space is referred to as a “lifted” Markov chain. In Refs. \cite{15} and \cite{16} it is shown that the convergence toward the target distribution is accelerated by applying the methodology of a “lifted” Markov chain to a simple random walk. Especially in Refs. \cite{17, 19}, the authors have reported that the dynamical critical exponent in a two-dimensional Ising model and fully connected Ising model can be reduced by their methods. The event-chain Monte Carlo (ECMC) algorithm \cite{22} is constructed for systems of continuous
degree of freedom, such as hard-sphere, more general particle systems [23], and continuous spin systems [24, 25]. Recently, Nishikawa et al. have shown numerically that the ECMC algorithm for a three-dimensional Heisenberg model reduces the dynamical critical exponent down to \( z \approx 1 \) [23]. These results imply that the violation of the DBC could change the relaxation dynamics of physical quantities qualitatively.

The efficiency of the violation of the DBC is partially confirmed theoretically. Ichiki and Ohzeki have revealed that the real part of the second-largest eigenvalue of a transition matrix decreases by violating the DBC, compared to that of a symmetrized transition matrix satisfying the DBC [24]. The asymptotic variance is also reduced in comparison with that of the corresponding symmetrized matrix [27]. It should be noted that the symmetrized transition matrix is not always equivalent to that before violating the DBC. Hence, the relation between the efficiency of MCMC methods and the violation of the DBC in general is not yet well understood.

In this paper, we focus on the skew detailed balance condition (SDBC) originally proposed by Turitsyn et al. [17], which belongs to the latter strategy mentioned above. We develop an irreversible Metropolis-Hastings algorithm with the SDBC so that it can be applied to any general system. In general, it is quite difficult to evaluate the second-largest eigenvalue of the transition matrix of a Markov chain even with the DBC. It is worth evaluating the efficiency of the algorithm in a toy model where all the eigenvalues of the matrix is explicitly written down under both the DBC and the SDBC. Here by applying the proposed algorithm to a simple random walk on one dimension the efficiency of the algorithm is studied. Then, we show analytically that the second-largest eigenvalue in absolute value and the asymptotic variance of the corresponding transition matrix for the random-walk problem can be reduced by imposing the SDBC. Our results imply that the relaxation dynamics in a Markov chain can be qualitatively changed from diffusive to ballistic by introducing the violation of the DBC, and that the violation does not always improve the efficiency.

The paper is organized as follows. Section II introduces the theoretical foundation of a Markov chain with the DBC. In Sec. III, a Markov chain with the SDBC is constructed. In Sec. IV an algorithm to realize the Markov chain with the SDBC is proposed. In Sec. V the efficiency of the proposed algorithm is discussed by analyzing a random walk in one dimension. Section VI summarizes this study.

II. MARKOV CHAIN WITH THE DETAILED BALANCE CONDITION

The Markov chain with the SDBC is constructed based on the Markov chain with the DBC. We briefly review the theoretical aspects of the Markov chain satisfying the DBC [28]. Note that all the vectors in the paper are defined as row vectors.

A. Setup

Throughout the paper, we discuss a system of discrete degree of freedom. Let \( I = \{1, \ldots, \Omega\} \) be a state space of the system with \( \Omega \) being the total number of states. Suppose that a target probability distribution is given as \( \pi = (\pi_1, \ldots, \pi_\Omega) \) with \( \pi_i > 0 \) and \( \sum_{i=1}^{\Omega} \pi_i = 1 \).

We need to (i) generate sampling states according to the target distribution and (ii) calculate the expectation of a quantity \( \hat{f} \) with respect to the target distribution, defined as

\[
\langle \hat{f} \rangle_\pi \equiv \sum_{i=1}^{\Omega} \pi_i f_i = \pi f^\top,
\]

where \( f_i \) depending on the state \( i \) denotes the realization of \( \hat{f} \) and \( f \equiv (f_1, \ldots, f_\Omega) \). It is, however, difficult to evaluate \( \langle \hat{f} \rangle_\pi \) analytically for systems with high-dimensional state space in general. Thus, MCMC methods are often employed to achieve our goals numerically for sufficiently large \( \Omega \).

B. Transition matrix and master equation

A discrete-time Markov chain is a fundamental stochastic process in which the transition probability from the current state \( i \) to a new state \( j \) is independent of the history of the transition. Let \( T = (T_{ij})_{i,j \in I} \) be the transition matrix of the Markov chain. An element \( T_{ij} \) denotes the transition probability from state \( i \) to \( j \) in a unit time of the Markov chain (Fig. 1). Note that \( T \) is the stochastic matrix, i.e., \( T_{ij} \geq 0 \) for all \( i, j \in I \) and \( \sum_{j=1}^{\Omega} T_{ij} = 1 \) for all \( i \in I \).

Let \( p^{(n)} \equiv (p_1^{(n)}, \ldots, p_\Omega^{(n)}) \) be a probability distribution after \( n \) steps in the Markov chain. \( p_i^{(n)} \) denotes the probability for finding a state \( i \) at time \( n \). Then, the time evolution of the probability distribution is described by the master equation expressed as

\[
p^{(n+1)} = p^{(n)} T,
\]
and, consequently, we obtain
\[ p^{(n)} = p^{(0)}T^n, \]
which shows that the distribution at arbitrary time is completely characterized by an initial distribution and the transition matrix.

C. Detailed balance condition

A transition matrix \( T \) is called ergodic if there exists an integer \( n > 0 \) such that all elements of the \( n \)th power of the transition matrix are positive. A Markov chain characterized by an ergodic transition matrix is ensured to have a unique stationary distribution. It is guaranteed that \( p^{(n)} \) converges to our desired distribution \( \pi \) as \( n \to \infty \) for arbitrary initial distributions by imposing the ergodicity and the balance condition (BC), which is expressed as \( \pi = \pi T \). In practice, the DBC, which is widely imposed as a sufficient condition of the BC. For instance, one can easily find the transition probability satisfying the DBC such as the Metropolis-Hastings type \([10]\) and the heat-bath type \([4]\). The DBC is also called a reversibility and a Markov chain satisfying the DBC is referred to as a reversible Markov chain. In contrast, a Markov chain without the DBC is called an irreversible Markov chain.

D. Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm \([10]\), one of the most famous MCMC algorithms, numerically performs the reversible Markov chain explained in the previous subsection. In this subsection, we describe the Metropolis-Hastings algorithm to fix our notation.

First, the transition matrix \( T = (T_{ij})_{i,j \in I} \) satisfying the DBC with respect to \( \pi \) is decomposed as
\[ \pi_i T_{ij} = \pi_j T_{ji}, \]
is widely imposed as a sufficient condition of the BC. For instance, one can easily find the transition probability satisfying the DBC such as the Metropolis-Hastings type \([10]\) and the heat-bath type \([4]\). The DBC is also called a reversibility and a Markov chain satisfying the DBC is referred to as a reversible Markov chain. In contrast, a Markov chain without the DBC is called an irreversible Markov chain.

E. Eigenvalues of the transition matrix and efficiency of MCMC methods

In this subsection, we survey the relationship between the efficiency of MCMC methods and the eigenvalues of the corresponding transition matrix in the Markov chain.

1. Convergence rate and second-largest eigenvalue

Let \( T \) be the transition matrix satisfying the DBC with respect to \( \pi \) and let \( B \equiv \text{diag}(\pi_1, \ldots, \pi_I) \) be the diagonal matrix in \( \mathbb{R}^{I \times I} \). The matrix \( B \) is positive-definite, real-symmetric, and invertible because we have assumed that \( \pi_i > 0 \) for all \( i \in I \). Thus, it is well defined that \( B^{1/2} = \text{diag}(\pi_1^{1/2}, \ldots, \pi_I^{1/2}) \) and \( B^{-1/2} = \text{diag}(\pi_1^{-1/2}, \ldots, \pi_I^{-1/2}) \).

A similarity transformation of \( T \) with respect to \( B^{-1/2} \) is defined as
\[ S \equiv B^{1/2}T B^{-1/2}. \]
The eigenvalues of \( S \) coincide with those of \( T \) including multiplicity because the similarity transformation does not change the characteristic polynomial of \( T \). In addition, \( S \) is a real-symmetric matrix and thus all the eigenvalues are ensured to be real because \( T \) is reversible with respect to the target distribution \( \pi \). From the Perron-Frobenius theorem, it is guaranteed that the largest eigenvalue is equal to 1 with multiplicity 1 and the absolute value of other eigenvalues is less than 1 if and only if the transition matrix is ergodic. Therefore, when we denote a set of eigenvalues of the reversible transition matrix \( T \) as \( \{\lambda_k\}_{k=1}^K \), the eigenvalues can be rearranged as \( 1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_K \geq -1 \) without loss of generality. A set of left eigenvectors \( \{x_{k,\sigma}\} \) of \( S \) can be set as an orthonormal basis in \( \mathbb{R}^I \) described as follows:
\[ x_{k,\sigma} S = \lambda_k x_{k,\sigma}, \quad x_{k,\sigma} \in \mathbb{R}^I, \quad x_{k,\sigma} x_{l,\rho}^\top = \delta_{kl} \delta_{\sigma \rho}, \]
where $\sigma \in \{1, \ldots, m_k\}$ is an index for multiplicity with $m_k$ being the multiplicity of $\lambda_k$. Left eigenvectors $\{u_{k,\sigma}\}$ and right eigenvectors $\{v_{k,\sigma}\}$ of $T$ are given as $u_{k,\sigma} = x_{k,\sigma} B^{1/2}$ and $v_{k,\sigma} = x_{k,\sigma} B^{-1/2}$, respectively. In particular, the Perron-Frobenius theorem ensures that the left and right eigenvectors associated with $\lambda_1 = 1$ are given as $u_{1,1} = \pi$ and $v_{1,1} = 1 \equiv (1, \ldots, 1)$, respectively.

Consequently, the reversible transition matrix $T$ can be decomposed into

$$ T = A + \sum_{k=2}^{K} \lambda_k \left( \sum_{\sigma=1}^{m_k} v_{k,\sigma}^T u_{k,\sigma} \right), $$

where $A \equiv 1^T \pi$ is the so-called limiting matrix. From the master equation one can derive a formal solution,

$$ p^{(n)} = \pi + \sum_{k=2}^{K} (\lambda_k)^n \left[ \sum_{\sigma=1}^{m_k} (p^{(0)} v_{k,\sigma}^T) u_{k,\sigma} \right]. $$

$p^{(n)}$ converges to the stationary distribution $\pi$ as $n \to \infty$ because $|\lambda_k| < 1$ for $2 \leq k \leq K$. Moreover, this indicates that the convergence rate is determined by the second-largest eigenvalue in absolute value, denoted by

$$ \eta \equiv \max_{2 \leq k \leq K} |\lambda_k|. $$

Thus, the relaxation time, defined by

$$ \tau_{\text{relax}} = -\frac{1}{\log \eta}, $$

reflects the convergence rate.

Correspondingly, the expectation value of $\hat{f}$ at time $n$ is also defined as

$$ \langle \hat{f} \rangle_n \equiv \sum_{i=1}^{\Omega} p_i^{(n)} f_i = p^{(n)} f^T. $$

From Eq. (11), it can be rewritten as

$$ \langle \hat{f} \rangle_n = \langle \hat{f} \rangle \pi + \sum_{k=2}^{K} (\lambda_k)^n \left[ \sum_{\sigma=1}^{m_k} (p^{(0)} v_{k,\sigma}^T) (u_{k,\sigma} f^T) \right], $$

which may indicate that the convergence rate of the expectation is also determined by $\eta$. However, the eigenvalue $\lambda_k$ does not affect the relaxation dynamics of $\langle \hat{f} \rangle_n$ if $u_{k,\sigma} f^T = 0$ holds for all $\sigma = 1, \ldots, m_k$. In particular, the convergence rate of $\langle \hat{f} \rangle_n$ is not determined by $\lambda_2$ if $\hat{f}$ is orthogonal to all the left eigenvectors associated with the second-largest eigenvalue.

2. Correlations and second-largest eigenvalue

If the central limit theorem holds, an effective variance of the estimator in Eq. (7) is given by $v(\hat{f}, \pi, T)/M$ for sufficiently large $M$, where $v(\hat{f}, \pi, T)$ denotes the asymptotic variance defined by

$$ v(\hat{f}, \pi, T) \equiv \lim_{M \to \infty} M \text{var} \left[ \frac{1}{M} \sum_{n=1}^{M} f^{(n)} \right]. $$

The asymptotic variance is often enlarged by the correlation between samples in the Markov chain. It is convenient to measure the correlation by an integrated autocorrelation time, defined as

$$ \tau_{\text{int}, \hat{f}} = \sum_{n=1}^{\infty} C^{(n)}_{\hat{f}}, $$

where

$$ C^{(n)}_{\hat{f}} = \frac{\langle f \hat{f} \rangle_n - \langle \hat{f} \rangle_n^2}{\text{var}_{\pi}[\hat{f}]} $$

denotes the autocorrelation function, $\text{var}_{\pi}[\hat{f}] \equiv \langle \hat{f}^2 \rangle - \langle \hat{f} \rangle^2$, is the variance of $\hat{f}$ for an independent sampling, and $\langle f \hat{f} \rangle_n \equiv f B^{1/2} f^T$. The relationship between the asymptotic variance and the integrated autocorrelation time is explicitly given by

$$ v(\hat{f}, \pi, T) = (1 + 2 \tau_{\text{int}, \hat{f}}) \text{var}_{\pi}[\hat{f}], $$

indicating that strong correlation results in poor estimation.

Here, we define the fundamental matrix $Z$ as

$$ Z \equiv (1 - T + A)^{-1} = 1 + \sum_{n=1}^{\infty} (T^n - A). $$

Using the matrix $Z$, we can rewrite the asymptotic variance as

$$ v(\hat{f}, \pi, T) = f Z f^T = f (B Z + (B Z)^T - B - B A) f^T $$

$$ = \sum_{k=2}^{K} \frac{1 + \lambda_k}{1 - \lambda_k} \left[ \sum_{\sigma=1}^{m_k} (f B v_{k,\sigma}) (u_{k,\sigma} f^T) \right]. $$

It turns out that the asymptotic variance depends on all the eigenvalues of the transition matrix. However, the upper bound of the ratio between the variance and the asymptotic variance is determined only by the second-largest eigenvalue of the transition matrix as

$$ \max_{\hat{f} \neq 0} \frac{v(\hat{f}, \pi, T)}{\text{var}_{\pi}[\hat{f}]} = \frac{1 + \lambda_2}{1 - \lambda_2}, $$

where the equality is attained by $\hat{f}$ given by the linear combination of $\{v_{2,\sigma}\}_{\sigma=1}^{m_2}$.

3. Irreversible transition matrix and additive reversiblization

A transition matrix that has a unique stationary distribution but does not satisfy the DBC with respect
to the stationary distribution is referred to as an irreversible transition matrix. Let us consider the irreversible transition matrix $\mathbf{T}$ that is ergodic and has a unique stationary distribution $\pi$. Although some eigenvalues might be complex and it is not guaranteed that $\mathbf{T}$ is diagonalizable in general, the transition matrix $\mathbf{T}$ is diagonalizable if all the eigenvalues of $\mathbf{\Omega}$ are distinct. In this case, one can rearrange the eigenvalues of $\mathbf{T}$ of the irreversible transition matrix $\mathbf{T}$ even in an irreversible case.

Thus, the formula in Eq. (10) also holds in an irreversible case.

Let us introduce the symmetrized transition matrix of the irreversible transition matrix $\mathbf{T}$, defined as follows [28]:

$$
\mathbf{T}_0 \equiv \frac{1}{2}(\mathbf{T} + \mathbf{B}^{-1}\mathbf{TB});
$$

(23)

this is referred to as the additive reversibilization. Note that the additive reversibilization $\mathbf{T}_0$ satisfies the DBC with respect to $\pi$. In addition, it is theoretically proved that the irreversible transition matrix is always better than that of the corresponding additive reversibilization in terms of the real part of the second-largest eigenvalue [26] and the asymptotic variance [27].

When the irreversible transition matrix $\mathbf{T}$ is constructed by modifying a reversible transition matrix $\mathbf{T}_{\text{rev}}$, we are interested in comparing the efficiency of $\mathbf{T}$ with that of $\mathbf{T}_{\text{rev}}$. Although the relation between the efficiency of $\mathbf{T}$ and that of the additive reversibilization of $\mathbf{T}$ is discussed in Refs. [26,27], the additive reversibilization of $\mathbf{T}$ is not always equivalent to $\mathbf{T}_{\text{rev}}$ and the relation between the efficiency of $\mathbf{T}$ and that of $\mathbf{T}_{\text{rev}}$ in general is not yet well understood. In Sec. [3] the relation is discussed by the eigenvalue analysis of the transition matrix for a specific probabilistic model.

III. MARKOV CHAIN WITH THE SKEW DETAILED BALANCE CONDITION

The Markov chain with the skew detailed balance condition has been proposed by Turitsyn et al. [17]. In this section, we review how to construct an irreversible Markov chain by imposing the SDBC in general.

A. Extension of the stationary distribution

First, we double the state space $I$ by introducing an auxiliary variable $\varepsilon \in \{+,-\}$. The extended state space is given as $\tilde{I} := I \times \{+,-\}$ and a state in $\tilde{I}$ is described by $(i, \varepsilon)$. Let $\pi_{(i,\varepsilon)}$ be the probability of finding the state $(i, \varepsilon)$ and $\pi_{(i,\varepsilon)}$ be uniform with respect to $\varepsilon$. Then, the target distribution is extended to

$$
\tilde{\pi} \equiv (\pi_{(1,+)}, \ldots, \pi_{(\Omega,+)}, \pi_{(1,-)}, \ldots, \pi_{(\Omega,-)}) = \frac{1}{2}(\pi, \pi).
$$

(24)

Let $\hat{f}$ be a quantity defined on the extended state space. $f_{(i,\varepsilon)}$ denotes the realization of $\hat{f}$ at the extended state $(i, \varepsilon)$. Then, the expectation value of $\hat{f}$ with respect to the extended target distribution is defined as

$$
\langle \hat{f} \rangle_{\tilde{\pi}} \equiv \sum_{i=1}^{\Omega} \sum_{\varepsilon = \pm} \pi_{(i,\varepsilon)} f_{(i,\varepsilon)}.
$$

(25)

In the case in which $\hat{f}$ is independent of $\varepsilon$, i.e., $f_{(i,\varepsilon)} = f_i$, we have

$$
\langle \hat{f} \rangle_{\tilde{\pi}} = \sum_{i=1}^{\Omega} \pi_i f_i = \sum_{i=1}^{\Omega} \pi_i f_i = \langle \hat{f} \rangle_{\pi}.
$$

(26)

In other words, the expectation with respect to the extended target distribution corresponds to that with respect to the original target distribution.

B. Skew detailed balance condition

Here, we construct the Markov chain on the extended state space $\tilde{I}$. The transition matrix in the Markov chain on $\tilde{I}$ is given as follows:

$$
\mathbf{\tilde{T}} = \begin{pmatrix} \mathbf{T}^{(+)} & \mathbf{\Lambda}^{(+)} \\ \mathbf{\Lambda}^{(-)} & \mathbf{T}^{(-)} \end{pmatrix},
$$

(27)

where $\mathbf{T}^{(\pm)} = (T_{ij}^{(\pm)})_{i,j \in I}$ and $\mathbf{\Lambda}^{(\pm)} = \text{diag}(\lambda_i^{(\pm)})_{i \in I}$. $T_{ij}^{(\pm)}$ denotes the transition probability from state $(i, \varepsilon)$ to state $(j, \varepsilon)$ and $\lambda_i^{(\pm)}$ is the eigenvalue of $\mathbf{T}_{ij}$ that is from state $(i, \varepsilon)$ to state $(i, -\varepsilon)$ (Fig. [2]). Conservation of probability in the transition matrix is expressed as

$$
\sum_{j \in I} T_{ij}^{(\pm)} + \lambda_i^{(\pm)} = 1,
$$

(28)

for all $i \in I$.

Let us assume that $\mathbf{\tilde{T}}$ is ergodic. Then, the BC $\tilde{\pi} = \tilde{\pi} \mathbf{\tilde{T}}$ ensures that the stationary distribution of $\mathbf{\tilde{T}}$ is exactly $\tilde{\pi}$. To construct an irreversible Markov chain, we impose the SDBC [17] described as

$$
\pi_i T_{ij}^{(+)} = \pi_j T_{ji}^{(-)}.
$$

(29)

This condition means that the stochastic flow with a transition from state $(i, +)$ to state $(j, -)$ balances with the transition from state $(j, -)$ to state $(i, +)$. In general, the DBC is violated by imposing the SDBC.

By the conservation of probability in Eq. (28) and the SDBC in Eq. (29), the BC can be rewritten as

$$
\lambda_i^{(+)} - \lambda_i^{(-)} = \sum_{j \neq i} (T_{ij}^{(-)} - T_{ij}^{(+)}).
$$

(30)
Consequently, the convergence to the extended stationary distribution is guaranteed by imposing the SDBC and Eq. (30).

IV. IRREVERSIBLE METROPOLIS-HASTINGS ALGORITHM

In this section, we construct the MCMC method on the basis of the SDBC. Although the prototype algorithm for a mean-field Ising model has been proposed by Turitsyn et al. \[17\], we develop the algorithm so as to be applicable to more general systems.

A. Choice of the transition matrix

An example of the transition probability \( T_{ij}^{(\pm)} \) is given in this subsection. To begin, we prepare a transition matrix on \( I \) satisfying the DBC with respect to \( \pi \). Namely, a transition matrix \( T = (T_{ij})_{i,j \in I} \) with \( \pi_i T_{ij} = \pi_j T_{ji} \) is given. The transition probability \( T_{ij}^{(\pm)} \) satisfying the SDBC in Eq. (29) is obtained by modifying \( T_{ij} \) with an arbitrary function \( \Delta_{ij}^{(\pm)} \) of two states \( i, j \in I \) as follows:

\[
T_{ij}^{(\pm)} = \frac{1 + \Delta_{ij}^{(\pm)}}{2} T_{ij}, \tag{31}
\]

where \( \Delta_{ij}^{(\pm)} \) satisfies \( |\Delta_{ij}^{(\pm)}| \leq 1 \) and \( \Delta_{ij}^{(+)} = \Delta_{ji}^{(-)} \) for all \( i, j \in I \). It is straightforward to show that \( T_{ij}^{(\pm)} \) satisfies the SDBC in Eq. (29).

Even if the transition probability \( T_{ij}^{(\pm)} \) is fixed as Eq. (31), there remain several choices of the transition probability \( \Lambda_i^{(\pm)} \). The following transition probabilities satisfy the condition of Eq. (30):

\[
\Lambda_{i,SH_1}^{(\pm)} = \sum_{j \neq i} T_{ij}^{(\mp)} = \sum_{j \neq i} \frac{1 + \Delta_{ij}^{(\mp)}}{2} T_{ij}, \tag{32}
\]

\[
\Lambda_{i,SH_2}^{(\pm)} = \frac{1}{2} + \frac{1}{4} \sum_{j \neq i} (\Delta_{ij}^{(+)} - \Delta_{ij}^{(-)}) T_{ij}, \tag{33}
\]

\[
\Lambda_{i,SH_3}^{(\pm)} = \sum_{j \neq i} \frac{1 - \Delta_{ij}^{(\pm)}}{2} T_{ij}, \tag{34}
\]

and

\[
\Lambda_{i,TCV}^{(\pm)} = \max \left\{ 0, -\frac{1}{2} \sum_{j \neq i} (\Delta_{ij}^{(+)} - \Delta_{ij}^{(-)}) T_{ij} \right\}. \tag{35}
\]

They are referred to as the Sakai-Hukushima 1 (SH1) type, the Sakai-Hukushima 2 (SH2) type \[20\], the Sakai-Hukushima 3 (SH3) type, and the Turitsyn-Chertkov-Vucelja (TCV) type \[17\], respectively.

B. Irreversible Metropolis-Hastings algorithm

Let us decompose the transition probability \( T_{ij} \) as \( T_{ij} = q_{ij} w_{ij} \) \( (i \neq j) \) as was done in the previous section. \( \hat{X}^{(n)} \) denotes the state in \( I \) after \( n \) steps in a Markov chain. Then, the irreversible Metropolis-Hastings (IMH) algorithm is described as follows:

(i) Set an initial state \( \hat{X}^{(0)} \) chosen arbitrarily.

(ii) Suppose that \( \hat{X}^{(n)} = (i, \varepsilon) \) and propose a new state \( (j, \varepsilon) \) by using the probability distribution \( q_{ij} \) \( j \in I \).

(iii) Accept the proposed state as \( \hat{X}^{(n+1)} = (j, \varepsilon) \) with the probability \( (1 + \Delta_{ij}^{(\varepsilon)}) w_{ij} / 2 \).

(iv) If the proposed state is rejected in step (iii), set \( \hat{X}^{(n+1)} = (i, -\varepsilon) \) with the probability \( p \) given as

\[
p = \frac{\Delta_{ij}^{(\varepsilon)}}{1 - \sum_{j \neq i} T_{ij}^{(\varepsilon)}}. \tag{36}
\]

If also rejected, set \( \hat{X}^{(n+1)} = \hat{X}^{(n)} \).

By repeating the above procedures (ii)–(iv) \( M \) times, one can obtain the Markov chain \( (\hat{X}^{(n)})_{n=0,1,2,\ldots,M} \) generated by the transition matrix \( \hat{T} \), verified in Appendix A.

The expectation \( \langle f \rangle_{\hat{\mu}} \) is estimated in the same way as in the original Metropolis-Hastings algorithm.
V. PERFORMANCE EVALUATION

In this section, we discuss a random walk on a circle as a toy model. By specifying all the eigenvalues and eigenvectors of the corresponding transition matrices, we discuss the efficiency of the irreversible MCMC method we have proposed in the previous sections.

A. Random walk on a circle

Suppose that there are states \( i = 1, 2, \ldots, \Omega \) on a circle under a periodic boundary condition. Then, we give the transition matrix on the state space as

\[
T = (1 - \alpha)I_{\Omega} + J_{\Omega} \left( \frac{\alpha}{2}, \frac{\alpha}{2} \right),
\]

where \( 0 < \alpha < 1 \) is a transition rate, \( I_{\Omega} \) is the \( \Omega \)th identity matrix, and

\[
J_{\Omega}(a, b) \equiv \begin{pmatrix} 0 & a & b \\ b & 0 & a \\ \vdots & \ddots & \ddots \\ a & b & 0 \end{pmatrix},
\]

respectively. Figure 3 illustrates a transition graph on \( I \).

The stationary distribution of \( T \) is given by the uniform distribution as \( \pi_i = 1/\Omega \) since the transition matrix is doubly stochastic. Notice that the transition matrix satisfies the DBC with respect to the uniform distribution.

All the eigenvalues of the transition matrix for the toy model are derived as

\[
\lambda_k = 1 - \alpha + \alpha \cos \theta_k,
\]

where \( \theta_k \equiv 2\pi(k - 1)/\Omega \) for \( 1 \leq k \leq \lceil \Omega/2 \rceil + 1 \) and \( [x] \) denotes the maximum integer that does not exceed \( x \in \mathbb{R} \). The multiplicities of eigenvalues for even \( \Omega \) are given by \( m_1 = 1 \), \( m_k = 2 \) for \( 2 \leq k \leq \lfloor (\Omega + 1)/2 \rfloor \), and \( m_{\Omega/2+1} = 1 \), respectively. The eigenvectors of \( T \) are the same as those of \( J_{\Omega}(\alpha/2, \alpha/2) \), described in Appendix B in detail.

B. Irreversible random walk with the SDBC

Some previous works addressed the effectiveness of the violation of the DBC. Diaconis et al. have analyzed the convergence rate of the total variation and \( \chi^2 \) distance of a nonreversible random walk [13]. Chen et al. have shown that the mixing time of the random walk is reduced by violating the DBC [14]. In this subsection, we derive all the eigenvalues and eigenvectors of the extended transition matrix with the SDBC. By using them, we reveal that the convergence rate and the worst evaluation of the asymptotic variance are improved by imposing the SDBC. We also find that the violation of the DBC is not always superior to the original method with the DBC.

Let us apply the methodology of the SDBC to the random walk described in the previous subsection. By adding the auxiliary variable \( \epsilon \in \{+,-\} \), the state space is doubled and the target distribution is extended as the uniform distribution on the extended state space. The transition matrix on the extended state space is given as

\[
\tilde{T} = \left[ 1 - \left( 1 + \frac{\delta - \delta'}{2} \right) \frac{\alpha}{2} - \gamma \right] I_{2\Omega}
+ \tilde{J}_{2\Omega} \left( \frac{1 + \delta}{2}, \frac{1 - \delta'}{2}, \gamma \right),
\]

where

\[
\tilde{J}_{2\Omega}(a, b; c) \equiv \begin{pmatrix} J_{\Omega}(a, b) & c_{\Omega}^T \\ c_{\Omega} & J_{\Omega}(b, a) \end{pmatrix},
\]

with \(|\delta| \leq 1 \) and \(|\delta'| \leq 1 \) being parameters to control the violation of the DBC, and where \( \gamma \) is a transition rate for \( \epsilon \) flip. The allowed range of \( \gamma \), \( 0 \leq \gamma < 1 - (2 + \delta - \delta')\alpha/4 \), includes the particular transition rates such as \( \gamma_{\text{SH}_1} = (2 + \delta - \delta')\alpha/4 \), \( \gamma_{\text{SH}_2} = 1/2 \), \( \gamma_{\text{SH}_3} = (2 - \delta + \delta')\alpha/4 \), and \( \gamma_{\text{TCV}} = 0 \), respectively. The transition graph on the extended state space is given as Fig. 4.
All the eigenvalues of $\tilde{T}$ are obtained with Eq. (39) as
\[
\tilde{\lambda}_k^\pm = 1 - \left(1 + \frac{\delta - \delta'}{2}\right) \alpha \sin^2 \frac{\theta_k}{2} - \gamma 
\pm \sqrt{\gamma^2 - \left(\frac{\delta + \delta'}{2}\right)^2 \left(\frac{\alpha}{2} \sin \theta_k\right)^2},
\]
for $1 \leq k \leq \lceil \Omega/2 \rceil + 1$. The multiplicities of eigenvalues, depending only on the label $k$, are given by $m_1 = 1$, $m_2 = 2$ for $2 \leq k \leq \lceil (\Omega + 1)/2 \rceil$, and $m_{\Omega/2+1} = 1$ if $\Omega$ is even, respectively. We should note that some eigenvalues are degenerate or might be complex depending on $\delta$, $\delta'$, and $\gamma$. In particular, the extended transition matrix $\tilde{T}$ is not diagonalizable when $\gamma = |\delta + \delta'|(\alpha/4) \sin \theta_k$ for some $2 \leq k \leq \lceil (\Omega + 1)/2 \rceil$. All the eigenvectors of the extended transition matrix $\tilde{T}$ are the same as those of $\tilde{T}_{\Omega/2}$ with $a = (1 + \delta)\alpha/4$, $b = (1 - \delta')\alpha/4$, and $c = \gamma$, as described in Appendix C in detail.

C. Comparison of efficiency

In the present random-walk problem, the extended transition matrix $\tilde{T}$ with $\delta = \delta' = 0$ is equivalent to the additive reversibilization of $\tilde{T}$ with $\delta = \delta' (\neq 0)$. Thus, it is theoretically guaranteed from Refs. [26, 27] that the real part of the second-largest eigenvalue and the asymptotic variance of $\tilde{T}$ with $\delta = \delta' (\neq 0)$ are reduced in comparison with those of $\tilde{T}$ with $\delta = \delta' = 0$. However, it is unclear whether the convergence rate and the asymptotic variance of $\tilde{T}$ are reduced in comparison with those of the original reversible transition matrix $T$. In this subsection, we study two particular cases: (A) $\delta = \delta'$ and (B) $\delta' = 1$. Note that in the case (B), a path from state $(i, \varepsilon)$ to $(i - \varepsilon, \varepsilon)$ in the corresponding transition graph vanishes for all $i$. We assume that $\alpha = 1/2$ and that $\Omega$ is a sufficiently large even number for simplicity. By using the explicit expression of eigenvalues and eigenvectors derived in the previous subsections and appendices, we show analytically that the irreversible Markov chain with the SDBC for the random walk is more efficient than the reversible one in terms of the relaxation time and the asymptotic variance.

1. Convergence rate

First, let us discuss the relaxation time of the random-walk problem. In the case of the reversible random walk in Sec. V A, the relaxation time is obtained as
\[
\tau_{\text{relax}} = \frac{1}{\log |\lambda_2'|},
\]
In contrast, in the irreversible random walk of cases (A) and (B), the relaxation time is obtained as
\[
\tilde{\tau}_{\text{relax}} = -\frac{1}{\log \tilde{\eta}},
\]
where
\[
\tilde{\eta} \equiv \max_{2 \leq k \leq \lceil \Omega/2 \rceil + 1} |\tilde{\lambda}_k^\pm|
\]
is the second-largest eigenvalue in absolute value and the candidates of $\tilde{\eta}$ are $|\tilde{\lambda}_{1}^\pm|$, $|\tilde{\lambda}_{2}^\pm|$, and $|\tilde{\lambda}_{\Omega/2+1}^\pm|$. Then, $\tilde{\eta}$ is identified as follows:

\[
\begin{cases}
\tilde{\eta} = |\tilde{\lambda}_{1}^\pm| & \text{if } 0 \leq \gamma \leq \min \left[\frac{1}{4} \left(\sin^2 \frac{\pi}{\Omega} + \delta^2 \cos^2 \frac{\pi}{\Omega}\right), \frac{\pi}{4} \left(1 - \frac{1}{2} \sin^2 \frac{\pi}{\Omega}\right) - \frac{1}{4} \sqrt{(1 - \frac{1}{2} \sin^2 \frac{\pi}{\Omega}) \left(1 - \frac{1}{4} \sin^2 \frac{\pi}{\Omega}\right) + \frac{\delta^2}{4} \sin^2 \frac{\pi}{\Omega}}\right], \\
\tilde{\eta} = |\tilde{\lambda}_{\Omega/2+1}^\pm| & \text{if } \frac{\pi}{4} \left(1 - \frac{1}{4} \sin^2 \frac{\pi}{\Omega}\right) - \frac{1}{4} \sqrt{(1 - \frac{1}{4} \sin^2 \frac{\pi}{\Omega})^2 - \frac{\delta^2}{4} \sin^2 \frac{2\pi}{\Omega}} \leq \gamma \leq \frac{3}{4},
\end{cases}
\]
for case (A) and

\[
\begin{cases}
\tilde{\eta} = |\tilde{\lambda}_{1}^\pm| & \text{if } 0 \leq \gamma \leq \frac{1}{4} \left(1 + \frac{1 + \delta}{4} \sin^2 \frac{\pi}{\Omega}\right) - \frac{1}{4} \sqrt{(1 + \frac{1 + \delta}{4} \sin^2 \frac{\pi}{\Omega})^2 - \left(\frac{1 + \delta}{4}\right)^2 \frac{\delta^2}{4} \sin^2 \frac{2\pi}{\Omega}}, \\
\tilde{\eta} = |\tilde{\lambda}_{\Omega/2+1}^\pm| & \text{if } \frac{3}{4} \left(1 - \frac{1}{8} \left(1 + \sin^2 \frac{\pi}{\Omega}\right)\right) + \frac{1}{4} \sqrt{\left[1 - \frac{1 + \delta}{8} \left(1 + \sin^2 \frac{\pi}{\Omega}\right)\right]^2 - \frac{(1 + \delta)^2}{32} \sin^2 \frac{2\pi}{\Omega}} \leq \gamma \leq \frac{7 - \delta}{8},
\end{cases}
\]
for case (B), respectively. By comparing $\tau_{\text{relax}}$ with $\tilde{\tau}_{\text{relax}}$, we find that $\tau_{\text{relax}} \geq \tilde{\tau}_{\text{relax}}$ in the parameter region of $(\delta, \gamma)$, explicitly given by
\[
\frac{1}{2} \sin^2 \frac{\pi}{\Omega} + \frac{1}{16} \frac{\sin^2 \frac{2\pi}{\Omega}}{1 + \cos^2 \frac{\pi}{\Omega}} \left(\delta^2 - \tan^2 \frac{\pi}{\Omega}\right) \leq \gamma \leq \frac{1}{4} \left(\sin^2 \frac{\pi}{\Omega} + \delta^2 \cos^2 \frac{\pi}{\Omega}\right).
\]
FIG. 5. (Color online) Maps of the second-largest eigenvalue of the extended transition matrix $\tilde{T}$ in absolute value. The left and right panels represent cases (A) $\delta = \delta'$ and (B) $\delta' = 1$, respectively. In each panel, the shaded portion represents the parameter region $(\delta, \gamma)$ where the relaxation time is reduced by imposing the skew detailed balance condition. The thick dashed line represents the minimum of the relaxation time of the irreversible random walk for a fixed $\delta$ and the star symbol provides the lowest value of the relaxation time in the parameter region. Note that $\tilde{\lambda}_2$ is a complex number below the thick dashed line.

for case (A) and

$$\frac{4 \sin^2 \frac{\pi}{\bar{\Omega}}}{4 - (1 + \delta) \sin^2 \frac{\pi}{\bar{\Omega}}} \left[ \cos^2 \frac{\pi}{\bar{\Omega}} + \left( \frac{3 - \delta}{4} \right)^2 \right] \leq \gamma \leq \frac{3 - \delta}{8} \left[ \sin^2 \frac{\pi}{\bar{\Omega}} + \left( \frac{1 + \delta}{3 - \delta} \right) \cos^2 \frac{\pi}{\bar{\Omega}} \right]$$

(49)

for case (B), respectively. The parameter region of $(\delta, \gamma)$ is shown by the shaded areas in Fig. 5.

In particular, $\tilde{\tau}_{\text{relax}}$ is minimized by setting $|\delta| = 1$ and $\gamma = (1/4) \sin(2\pi/\bar{\Omega})$ in both cases. This is the best choice of parameters in terms of the relaxation time. Then, the $\Omega$ dependence of the relaxation time is qualitatively improved from $O(\Omega^2)$ in the reversible case to $O(\Omega)$ in the irreversible case, meaning that the violation of the DBC yields the qualitative change of the relaxation dynamics in the random walk from diffusive to ballistic. However, the relaxation time increases from the minimum value if the transition rate $\gamma$ is set as the SH$_1$, SH$_2$, SH$_3$, and TCV types previously attained [17, 20]. This result implies that the efficiency is not always improved even if the DBC is violated.

One may consider that the irreversible Markov chains can possibly have complex eigenvalues. In fact, as shown in Fig. 3 there is a finite region where the second-largest eigenvalue is complex. Interestingly, the parameter set with the minimum relaxation time is located on the boundary of the region. However, the emergence of the complex eigenvalues does not always involve the efficiency of the irreversible MCMC method.

2. Asymptotic variance

From the general result discussed in Sec. II the ratio of the asymptotic variance and the variance for any quantity is upper-bounded by $(1 + \lambda_2)/(1 - \lambda_2)$ for the reversible Markov chain. We study the corresponding upper bound of the ratio in the case of an irreversible random walk with the SDBC. It is reasonable to consider the asymptotic variance for the extended quantity defined as $f' : (i, \varepsilon) \mapsto f_i$ because we are interested in the quantities that are independent of the auxiliary variable $\varepsilon$. Then, we obtain the asymptotic variance of $f'$ as

$$v(f', \tilde{\pi}, \tilde{T}) = \sum_{k=2}^{[\bar{\Omega}/2]+1} \left[ \sum_{\varepsilon=\pm} \frac{1}{(A_k^{\varepsilon})^2} \frac{1 + \tilde{\lambda}_k^{\varepsilon}}{1 - \tilde{\lambda}_k^{\varepsilon}} \sum_{\sigma=1}^{m_k} (fBv_{k,\sigma})^T(u_{k,\sigma}f^T) \right].$$

(50)
FIG. 6. (Color online) The parameter region \((\delta, \gamma)\) where the asymptotic variance is reduced by imposing the skew detailed balance condition (shaded portion). The left and right panels represent cases (A) \(\delta = \delta'\) and (B) \(\delta' = 1\), respectively. The star symbol indicates the lowest value in the parameter region.

The ratio of the asymptotic variance to the variance is found to be maximized when \(\hat{f}' = \hat{v}_2^{\prime,\sigma}\):

\[
\max_{f' \neq 0} \frac{\nu(\hat{f}', \hat{\pi}, \hat{T})}{\var_n[\hat{f}']} = \frac{\nu(\hat{v}_2^{\prime,\sigma}, \hat{\pi}, \hat{T})}{\var_n[\hat{v}_2^{\prime,\sigma}]} = \frac{1}{1 - (A_2^{\prime})^2} \frac{1 + \lambda_2^{\prime}}{1 - \lambda_2^{\prime}}.
\]  

(51)

By comparing the upper bound of the asymptotic variance, we show that the worst evaluation of the asymptotic variance of \(\tilde{T}\) is improved from that of \(T\) when the parameters \((\delta, \gamma)\) satisfy the following inequalities:

\[
\gamma \leq \frac{1}{4} \left( \delta^2 \cos^2 \frac{\pi}{\Omega} - \sin^2 \frac{\pi}{\Omega} \right)
\]  

(52)

for case (A) and

\[
\gamma \leq \frac{1 + \delta}{8} \left( \frac{1 + \delta}{3 - \delta} \cos^2 \frac{\pi}{\Omega} - \sin^2 \frac{\pi}{\Omega} \right)
\]  

(53)

for case (B), respectively. The parameter region is illustrated in Fig. 6. This result indicates that the upper bound of the asymptotic variance can never be improved by changing the value of \(\delta\) for the SH1, SH2, or SH3 type of transition rate. However, by choosing the TCV type of transition rate, it is improved when \(\delta\) is sufficiently large and the limit of \(|\delta| = 1\) provides the most efficient algorithm in the parameter region. The most efficient point is different from that in the sense of the relaxation time.

VI. SUMMARY AND DISCUSSION

In this paper, the irreversible Metropolis-Hastings algorithm satisfying the SDBC is generalized so that it is applicable to any system with discrete degrees of freedom. In this algorithm, the violation of the DBC is characterized by the function \(\Delta_0^{(s)}\), referred to as the skewness function. In addition, there are four different transition probabilities for \(\varepsilon\) flip \(\Lambda_{i}^{(\pm)}\) referred to as the SH1, SH2, SH3, and TCV types. This algorithm has already been applied to several Ising spin systems and the relaxation dynamics of magnetization density has been discussed [20, 21].

To acquire further knowledge about this algorithm, it is applied to the random walk in one dimension as a benchmark. According to the general procedure mentioned in Sec. III the irreversible transition matrix \(\hat{T}\), which is characterized by the parameters \(\delta, \delta',\) and \(\gamma\), is constructed. Then, all the eigenvalues and eigenvectors of \(\hat{T}\) are derived analytically by the explicit diagonalization, and the parameter \((\delta, \gamma)\) dependence of the relaxation rate and the asymptotic variance are discussed in the two particular cases \(\delta = \delta'\) and \(\delta' = 1\). As a result, it is found that the relaxation rate and the asymptotic variance in the irreversible MCMC method are improved by selecting appropriate values of parameters \((\delta, \gamma)\), in comparison with those in the corresponding reversible one. In particular, the relaxation rate is qualitatively improved by the appropriate choice of the parameter \(\gamma\). Therefore, it is theoretically confirmed that the violation of the DBC by imposing the SDBC can improve the efficiency of MCMC methods. From the present theoretical analysis of the toy model, it should be noticed that the efficiency of the MCMC method depends on how its efficiency is evaluated, and the violation of the DBC does not always improve the efficiency.

As discussed in this paper, the efficiency of our proposed algorithm depends on the transition probability
\section*{Acknowledgments}

YS is supported by a Grant-in-Aid for the Japan Society for Promotion of Science (JSPS) Fellows (Grant No. 267868). KH is supported by Grants-in-Aid for Scientific Research from MEXT, Japan (Nos. 25610102, and 25120010).

\section*{Appendix A: Verification of the irreversible Metropolis-Hastings algorithm}

In this appendix, we verify that a sequence \( \{ \hat{X}(n) \}_{n=0,1,2,...} \) generated with the IMH algorithm in Sec. IV B is a Markov chain characterized by the transition matrix \( \mathbf{T} \). It is obvious that the sequence \( \{ \hat{X}(n) \}_{n=0,1,2,...} \) is a homogeneous Markov chain. Thus, we only have to examine whether a conditional probability \( \text{Prob}[\hat{X}(n+1) = (j, \varepsilon)|\hat{X}(n) = (i, \varepsilon)] \) is identical with the corresponding element of \( \mathbf{T} \). For \( i \in I, j \neq i, \varepsilon = \pm, \) and \( n = 0, 1, 2, ... \), each correspondence is verified by the following calculations:

\begin{equation}
\text{Prob}[\hat{X}(n+1) = (j, \varepsilon)|\hat{X}(n) = (i, \varepsilon)] = 0, \quad (A1)
\end{equation}

\begin{equation}
\text{Prob}[\hat{X}(n+1) = (j, \varepsilon)|\hat{X}(n) = (i, \varepsilon)] = q_{ij} \frac{1 + \Delta^{(\varepsilon)}_{ij}}{2} w_{ij} = T_{ij}^{(\varepsilon)}, \quad (A2)
\end{equation}

\section*{Appendix B: Spectral decomposition of \( J_\Omega(a,b) \)}

In this appendix, the matrix \( J_\Omega(a,b) \in \mathbb{R}^{\Omega \times \Omega} \), defined by

\begin{equation}
J_\Omega(a,b) = \begin{pmatrix}
0 & a & b \\
b & 0 & a \\
\cdot & \cdot & \cdot \\
a & b & 0
\end{pmatrix}, \quad (B1)
\end{equation}

is considered. By specifying all the eigenvalues and eigenvectors, the spectral decomposition of \( J_\Omega(a,b) \) is derived.

The eigenvalues of \( J_\Omega(a,b) \) with \( \mu = a \) are given by

\begin{equation}
\mu_k(a,b) = (a + b) \cos \theta_k + i(a - b) \sin \theta_k, \quad (B2)
\end{equation}

where \( \theta_k = 2\pi(k - 1)/\Omega \) for \( k = 1, 2, \ldots, \Omega \). The imaginary unit is denoted by \( i \). The left (right) eigenvectors \( \mathbf{u}_k (\mathbf{v}_k) \) associated with the eigenvalue \( \mu_k(a,b) \) are obtained as

\begin{equation}
\mathbf{u}_k = (\frac{1}{\Omega} e^{-i n \theta_k})_{n=1}^{\Omega}, \quad \mathbf{v}_k = (e^{i n \theta_k})_{n=1}^{\Omega}, \quad (B3)
\end{equation}

respectively. They satisfy the orthonormal relation and the complete relation described as

\begin{equation}
\mathbf{u}_k^\top \mathbf{v}_l = \delta_{kl}, \quad (B4)
\end{equation}

and

\begin{equation}
\sum_{k=1}^{\Omega} \mathbf{v}_k^\top \mathbf{u}_k = I_{\Omega}, \quad (B5)
\end{equation}
respectively. Thus, the spectral decomposition of $J_\Omega(a, b)$ is derived as

$$J_\Omega(a, b) = \sum_{k=1}^{\Omega} \mu_k(a, b) v_k^\top u_k. \quad (B6)$$

For $a = b$, all the eigenvalues of $J_\Omega(a, a)$ are real and the eigenvalues $\mu_k(a, a)$ and $\mu_{\Omega + 2 - k}(a, a)$ in Eq. (12) are degenerate for $2 \leq k \leq [(\Omega + 1)/2]$ because $\theta_{\Omega + 2 - k} = 2\pi - \theta_k$. Thus, the eigenvalue of $J_\Omega(a, a)$ is obtained as

$$\mu_k(a, a) = 2a \cos \theta_k, \quad (B7)$$

for $k = 1, 2, \ldots, [(\Omega/2) + 1]$, and the multiplicity of $\mu_k(a, a)$ is given as

$$m_k = \begin{cases} 
1 & \text{for } k = 1, \\
2 & \text{for } 2 \leq k \leq [(\Omega + 1)/2], \\
\Omega/2 + 1 & \text{for } k = \Omega/2 + 1 \text{ with even } \Omega.
\end{cases} \quad (B8)$$

In this case, all the eigenvectors can be chosen as real vectors because all the eigenvalues are real numbers. Let $\sigma = 1, \ldots, m_k$ be an index for multiplicity and $1 \equiv (1, \ldots, 1) \in \mathbb{R}^\Omega$. Then, the left eigenvector $u_{k, \sigma}$ and the right eigenvector $v_{k, \sigma}$ associated with $J_\Omega(a, a)$'s eigenvalue $\mu_k(a, a)$ are given as follows:

$$u_{1, 1} = \frac{1}{\Omega} \mathbf{1}, \quad v_{1, 1} = \mathbf{1}, \quad (B9)$$

$$u_{k, 1} = \left(\frac{\sqrt{2}}{\Omega} \cos n\theta_k\right)^{n=1}_n, \quad v_{k, 1} = \left(\frac{\sqrt{2}}{\Omega} \cos n\theta_k\right)^{\Omega=1}_n, \quad (B10)$$

$$u_{k, 2} = \left(\frac{\sqrt{2}}{\Omega} \sin n\theta_k\right)^{n=1}_n, \quad v_{k, 2} = \left(\frac{\sqrt{2}}{\Omega} \sin n\theta_k\right)^{\Omega=1}_n, \quad (B11)$$

for $2 \leq k \leq [(\Omega + 1)/2]$, and

$$u_{\Omega/2 + 1, 1} = (\frac{1}{\Omega} (-1)^n)^{n=1}_n, \quad v_{\Omega/2 + 1, 1} = ((-1)^n)^{\Omega=1}_n, \quad (B12)$$

for $k = \Omega/2 + 1$ with even $\Omega$. These eigenvectors satisfy the orthonormal relation

$$u_{k, \sigma} v_{l, \rho}^\top = \delta_{k l} \delta_{\sigma \rho}. \quad (B13)$$

and the complete relation

$$\sum_{k=1}^{[\Omega/2]+1} \sum_{\sigma=1}^{m_k} v_{k, \sigma}^\top u_{k, \sigma} = I_\Omega. \quad (B14)$$

Thus, the spectral decomposition of $J_\Omega(a, a)$ is derived as

$$J_\Omega(a, a) = \sum_{k=1}^{[\Omega/2]+1} \mu_k(a, a) \left(\sum_{\sigma=1}^{m_k} v_{k, \sigma}^\top u_{k, \sigma}\right). \quad (B15)$$

### Appendix C: Spectral decomposition of $\tilde{J}_{2\Omega}(a, b; c)$

The matrix $\tilde{J}_{2\Omega}(a, b; c) \in \mathbb{R}^{2\Omega \times 2\Omega}$, defined as

$$\tilde{J}_{2\Omega}(a, b; c) = \left(\begin{array}{cc} J_{2\Omega}(a, b) & c I_{\Omega} \\ c I_{\Omega} & J_{2\Omega}(b, a) \end{array}\right), \quad (C1)$$

is considered. In this appendix, all the eigenvalues and eigenvectors of $\tilde{J}_{2\Omega}(a, b; c)$ are explicitly given. It should be noted that the matrix $\tilde{J}_{2\Omega}(a, b; c)$ is not diagonalizable for $c = \pm (a - b) \sin \theta_k$ with $k = 2, 3, \ldots, [(\Omega + 1)/2]$.

If $c = 0$, the matrix $\tilde{J}_{2\Omega}(a, b; 0)$ is equivalent to a block diagonal matrix $J_{2\Omega}(a, b) \oplus J_{2\Omega}(b, a)$. Thus, the eigenvalue of $\tilde{J}_{2\Omega}(a, b; 0)$ is given as

$$\tilde{\mu}_k(a, b; 0) = (a + b) \cos \theta_k + i(a - b) \sin \theta_k, \quad (C2)$$

for $k = 1, 2, \ldots, \Omega$, and all the eigenvalues are doubly-degenerate. Let $0 \equiv (0, \ldots, 0) \in \mathbb{R}^\Omega$, $u_k$ and $v_k$ be the vectors defined in Appendix B, and an asterisk denote the complex conjugate. Then, the left eigenvector $\tilde{u}_{k, \sigma}$ and the right eigenvector $\tilde{v}_{k, \sigma}$ associated with $\tilde{\mu}_k(a, b; 0)$ are straightforwardly obtained as

$$\tilde{u}_{k, 1} = (u_{k, 0}, 0), \quad \tilde{v}_{k, 1} = (v_{k, 0}, 0) \quad (C3)$$

and

$$\tilde{u}_{k, 2} = (0, u_{k, 0}^\top), \quad \tilde{v}_{k, 2} = (0, v_{k, 0}^\top), \quad (C4)$$

for $1 \leq k \leq \Omega$, respectively. They satisfy the orthonormal and complete relation and thus the spectral decomposition of $\tilde{J}_{2\Omega}(a, b; 0)$ is obtained as

$$\tilde{J}_{2\Omega}(a, b; 0) = \sum_{k=1}^{\Omega} \tilde{\mu}_k(a, b; 0) \left(\tilde{v}_{k, 1}^\top \tilde{u}_{k, 1} + \tilde{v}_{k, 2}^\top \tilde{u}_{k, 2}\right). \quad (C5)$$

When $c \neq 0$ and $a = b$, the eigenvalue of $\tilde{J}_{2\Omega}(a, a; c)$ is given as

$$\tilde{\mu}_k^\pm(a, a; c) = 2a \cos \theta_k \pm c, \quad (C6)$$

for $k = 1, 2, \ldots, [(\Omega/2) + 1]$. The multiplicity of $\tilde{\mu}_k^\pm(a, a; c)$, which depends only on the label $k$, is obtained as

$$m_k = \begin{cases} 
1 & \text{for } k = 1, \\
2 & \text{for } 2 \leq k \leq [(\Omega + 1)/2], \\
\Omega/2 + 1 & \text{for } k = \Omega/2 + 1 \text{ with even } \Omega.
\end{cases} \quad (C7)$$

Let $u_{k, \sigma}$ and $v_{k, \sigma}$ be the vectors defined in Appendix B, then, the left eigenvector $\tilde{u}_{k, \sigma}$ and the right eigenvector $\tilde{v}_{k, \sigma}$ associated with $\tilde{\mu}_k^\pm(a, a; c)$ are given as follows:

$$\tilde{u}_{k, 1}^\pm = \frac{1}{2} \left(\begin{array}{c} 1 \\ \mp 1 \end{array}\right), \quad \tilde{v}_{k, 1}^\pm = \left(\begin{array}{c} 1 \\ \mp 1 \end{array}\right), \quad (C8)$$

and

$$\tilde{u}_{k, \sigma}^\pm = \frac{1}{\sqrt{2}} \left(\begin{array}{c} u_{k, \sigma} \\ \pm v_{k, \sigma} \end{array}\right), \quad \tilde{v}_{k, \sigma}^\pm = \frac{1}{\sqrt{2}} \left(\begin{array}{c} u_{k, \sigma} \\ \pm v_{k, \sigma} \end{array}\right). \quad (C9)$$
for $2 \leq k \leq \lfloor \Omega/2 \rfloor + 1$. They satisfy the orthonormal relation described as
\[
\tilde{u}_{k,\sigma}^\dagger \tilde{v}_{l,\rho} = \delta_{kl} \delta_{\sigma\rho} \delta_{\epsilon l}
\]  
(C10)
and the relation
\[
\sum_{k=1}^{\lfloor \Omega/2 \rfloor + 1} \sum_{l=\pm} m_k \tilde{v}_{k,\sigma}^\dagger \tilde{u}_{k,\sigma} = 1_{2\Omega}.
\]  
(C11)
Thus, the spectral decomposition of $\tilde{J}_{2\Omega}(a; a; c)$ is derived as
\[
\tilde{J}_{2\Omega}(a; b; c) = \sum_{k=1}^{\lfloor \Omega/2 \rfloor + 1} \left( \sum_{\sigma=\pm} \tilde{u}_{k,\sigma}^\dagger \tilde{v}_{k,\sigma} \tilde{u}_{k,\sigma}^\dagger \right).
\]  
(C12)
If $c \neq 0$, $a \neq b$, and $c \neq \pm (a - b) \sin \theta_k$ for all $k = 2, 3, \ldots, \lfloor (\Omega + 1)/2 \rfloor$, the eigenvalue of $\tilde{J}_{2\Omega}(a; b; c)$ is obtained as
\[
\tilde{\mu}_k^\pm(a; b; c) = (a + b) \cos \theta_k \pm \sqrt{c^2 - (a - b)^2 \sin^2 \theta_k},
\]  
(C13)
for $k = 1, 2, \ldots, \lfloor \Omega/2 \rfloor + 1$. Note that $\tilde{\mu}_k^\pm(a; b; c)$ might be complex when $|c| < |a - b| \sin \theta_k$. The multiplicity of $\tilde{\mu}_k^\pm(a; b; c)$, depending only on the label $k$, is given as
\[
\begin{align*}
& m_1 = 1 & \text{for } k = 1, \\
& m_k = 2 & \text{for } 2 \leq k \leq \lfloor (\Omega + 1)/2 \rfloor, \\
& m_{\Omega/2 + 1} = 1 & \text{for } k = \Omega/2 + 1 \text{ with even } \Omega.
\end{align*}
\]  
(C14)
The left eigenvector $\tilde{u}_{k,\sigma}^\pm$ and the right eigenvector $\tilde{v}_{k,\sigma}^\pm$ associated with $\tilde{\mu}_k^\pm(a; b; c)$ are given as follows:
\[
\tilde{u}_{1,1}^\pm = \frac{1}{2\Omega}(1, \pm 1), \quad \tilde{v}_{1,1}^\pm = (1, \pm 1),
\]  
(C15)
\[
\begin{align*}
& \tilde{u}_{k,1}^\pm = \frac{1}{\sqrt{2}(1 - (A_k^\pm)^2)}(u_{k,1} + A_k^\pm u_{k,2}, u_{k,1} - A_k^\pm u_{k,2}), \\
& \tilde{v}_{k,1}^\pm = \frac{1}{\sqrt{2}}(v_{k,1} - A_k^\pm v_{k,2}, v_{k,1} + A_k^\pm v_{k,2}),
\end{align*}
\]  
(C16)
\[
\begin{align*}
& \tilde{u}_{k,2}^\pm = \frac{1}{\sqrt{2}(1 - (A_k^\pm)^2)}(u_{k,2} - A_k^\pm u_{k,1}, u_{k,2} + A_k^\pm u_{k,1}), \\
& \tilde{v}_{k,2}^\pm = \frac{1}{\sqrt{2}}(v_{k,2} + A_k^\pm v_{k,1}, v_{k,2} - A_k^\pm v_{k,1}),
\end{align*}
\]  
(C17)
for $2 \leq k \leq \lfloor (\Omega + 1)/2 \rfloor$, where
\[
A_k^\pm = \frac{c \pm \sqrt{c^2 - (a - b)^2 \sin^2 \theta_k}}{(a - b) \sin \theta_k}
\]  
(C18)
and
\[
\begin{align*}
& \tilde{u}_{\Omega/2 + 1,1}^\pm = \frac{1}{\sqrt{2}}(u_{\Omega/2 + 1,1}, \pm u_{\Omega/2 + 1,1}), \\
& \tilde{v}_{\Omega/2 + 1,1}^\pm = \frac{1}{\sqrt{2}}(v_{\Omega/2 + 1,1}, \pm v_{\Omega/2 + 1,1}),
\end{align*}
\]  
(C19)
for $k = \Omega/2 + 1$ with even $\Omega$. They satisfy the orthonormal and complete relation described as
\[
\tilde{u}_{k,\sigma}^\dagger \tilde{v}_{l,\rho}^\dagger = \delta_{kl} \delta_{\sigma\rho} \delta_{\epsilon l}
\]  
(C20)
and
\[
\sum_{k=1}^{\lfloor \Omega/2 \rfloor + 1} \sum_{l=\pm} m_k \tilde{v}_{k,\sigma}^\dagger \tilde{u}_{k,\sigma} = 1_{2\Omega},
\]  
(C21)
respectively. Thus, the spectral decomposition of $\tilde{J}_{2\Omega}(a; b; c)$ is obtained as
\[
\tilde{J}_{2\Omega}(a; b; c) = \sum_{k=1}^{\lfloor \Omega/2 \rfloor + 1} \left( \sum_{\sigma=\pm} \tilde{u}_{k,\sigma}^\dagger \tilde{v}_{k,\sigma} \tilde{u}_{k,\sigma}^\dagger \right).
\]  
(C22)
Let us consider the case in which $a \neq b$ and there exists an integer $q \in \{2, 3, \ldots, \lfloor (\Omega + 1)/2 \rfloor\}$ such that $c = (a - b) \sin \theta_q \equiv c_q$ with $\epsilon = \pm 1$. In this case, the eigenvalues $\tilde{\mu}_k^\pm(a; b; c_q)$ and $\tilde{\mu}_k^\pm(a; b; c_q)$ obtained from Eq. (C13) are degenerate. Therefore, the multiplicity of the eigenvalue $\tilde{\mu}_k \equiv \tilde{\mu}_k^\pm(a; b; c_q)$ is 4. However, the dimension of the eigenspace corresponding to $\tilde{\mu}_k$ is 2 and thus $\tilde{J}_{2\Omega}(a; b; c_q)$ is not diagonalizable. In this case, one can transform $\tilde{J}_{2\Omega}(a; b; c_q)$ into a Jordan normal form by considering the set of vectors defined as
\[
\begin{align*}
& \tilde{u}_q^{(i)} = B_q^{(i)}(u_{q,1} - c u_{q,2}, u_{q,1} + c u_{q,2}), \\
& \tilde{u}_q^{(ii)} = B_q^{(ii)}(u_{q,1} + c u_{q,2}, u_{q,1} - c u_{q,2}), \\
& \tilde{u}_q^{(iii)} = B_q^{(iii)}(u_{q,2} + c u_{q,1}, u_{q,2} - c u_{q,1}), \\
& \tilde{u}_q^{(iv)} = B_q^{(iv)}(u_{q,2} - c u_{q,1}, u_{q,2} + c u_{q,1}),
\end{align*}
\]  
(C23)
\[
\begin{align*}
& \tilde{v}_q^{(i)} = C_q^{(i)}(v_{q,1} + c v_{q,2}, v_{q,1} - c v_{q,2}), \\
& \tilde{v}_q^{(ii)} = C_q^{(ii)}(v_{q,2} + c v_{q,1}, v_{q,2} - c v_{q,1}), \\
& \tilde{v}_q^{(iii)} = C_q^{(iii)}(v_{q,2} - c v_{q,1}, v_{q,2} + c v_{q,1}), \\
& \tilde{v}_q^{(iv)} = C_q^{(iv)}(v_{q,1} - c v_{q,2}, v_{q,1} + c v_{q,2}),
\end{align*}
\]  
(C24)
where the coefficients $B_q^{(x)}$ and $C_q^{(x)}$ ($x = i, ii, iii, iv$) satisfy
\[
B_q^{(x)} C_q^{(x)} = \frac{1}{4}
\]  
(C25)
and
\[
B_q^{(ii)} C_q^{(i)} = B_q^{(iv)} C_q^{(ii)} = \frac{1}{2} c_q.
\]  
(C26)
Note that they satisfy the orthonormal relation as
\[
\tilde{u}_q^{(x)} \tilde{v}_q^{(y)\top} = \delta_{xy}.
\]  
(C27)
Let $U, V \in \mathbb{R}^{4 \times 2\Omega}$ be
\[
U \equiv \begin{pmatrix}
\tilde{u}_q^{(i)} & \tilde{u}_q^{(ii)} & \tilde{u}_q^{(iii)} & \tilde{u}_q^{(iv)} \\
\tilde{v}_q^{(i)} & \tilde{v}_q^{(ii)} & \tilde{v}_q^{(iii)} & \tilde{v}_q^{(iv)}
\end{pmatrix}, \quad
V \equiv \begin{pmatrix}
\tilde{u}_q^{(i)} & \tilde{u}_q^{(ii)} & \tilde{u}_q^{(iii)} & \tilde{u}_q^{(iv)} \\
\tilde{v}_q^{(i)} & \tilde{v}_q^{(ii)} & \tilde{v}_q^{(iii)} & \tilde{v}_q^{(iv)}
\end{pmatrix}.
\]  
(C28)
Then, the orthonormal relation in Eq. (C27) can be rewritten as
\[ \mathbf{U} \mathbf{V}^T = \mathbf{I}_4. \]  
(C29)

Moreover, by replacing the term
\[ \sum_{\varepsilon = \pm} \tilde{\mu}_q(a,b; c) \left( \sum_{\sigma = 1, 2} \tilde{\nu}^\varepsilon_{q,\sigma}^T \tilde{u}^\varepsilon_{q,\sigma} \right) \]  
(C30)
in Eq. (C22) with
\[ V^T \begin{pmatrix} \tilde{\mu}_q & 1 \\ 0 & \tilde{\mu}_q \\ \tilde{\mu}_q & 1 \\ 0 & \tilde{\mu}_q \end{pmatrix} U = \tilde{\mu}_q \sum_x \tilde{v}_q^{(x)} \tilde{u}_q^{(x)} + \tilde{v}_q^{(i)} \tilde{u}_q^{(i)} + \tilde{v}_q^{(ii)} \tilde{u}_q^{(ii)} + \tilde{v}_q^{(iii)} \tilde{u}_q^{(iii)}, \]  
(C31)
the spectral decomposition in Eq. (C22) holds even in this case.

[1] D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2014), 4th ed.
[2] J. S. Liu, *Monte Carlo Strategies in Scientific Computing* (Springer-Verlag, New York, 2001).
[3] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. 21, 1087 (1953).
[4] B. A. Berg and T. Neuhaus, Phys. Rev. Lett. 68, 9 (1992).
[5] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. 65, 1604 (1996).
[6] R. H. Swendsen and J. S. Wang, Phys. Rev. Lett. 58, 86 (1987).
[7] U. Wolff, Phys. Rev. Lett. 62, 361 (1989).
[8] P. H. Peskun, Biometrika 60, 307 (1973).
[9] L. Pollet, S. M. A. Rombouts, K. Van Houcke, and K. Heyde, Phys. Rev. E 70, 056705 (2004).
[10] W. K. Hastings, Biometrika 57, 97 (1970).
[11] A. A. Barker, Aust. J. Phys. 18, 119 (1965).
[12] R. Ren and G. Orkoulas, J. Chem. Phys. 124, 064109 (2006).
[13] H. Suwa and S. Todo, Phys. Rev. Lett. 105, 120603 (2010).
[14] S. Todo and H. Suwa, J. Phys.: Conf. Ser. 473, 012013 (2013).
[15] P. Diaconis, S. Holmes, and R. M. Neal, Technical Report BU-13850-M, Biometrics Unit, Cornell University (1997); Ann. Appl. Probab. 10, 726 (2000).
[16] F. Chen, L. Lovasz, and I. Pak, In: Proc. 17th Annual ACM Symposium on Theory of Computing, 275 (1999).
[17] K. S. Turitsyn, M. Chertkov, and M. Vucelja, Physica D 240, 410 (2011).
[18] H. C. Fernandes and M. Weigel, Comput. Phys. Commun. 182, 1856 (2011).
[19] R. D. Schram and G. T. Barkema, Physica A 418, 88 (2014).
[20] Y. Sakai and K. Hukushima, J. Phys. Soc. Jpn. 82, 064003 (2013).
[21] K. Hukushima and Y. Sakai, J. Phys.: Conf. Ser. 473, 012012 (2013).
[22] E. P. Bernard, W. Krauth, and D. B. Wilson, Phys. Rev. E 80, 056704 (2009).
[23] M. Michel, S. C. Kapfer, and W. Krauth, J. Chem. Phys. 140, 054116 (2014).
[24] M. Michel, J. Mayer, and W. Krauth, arXiv:1508.06541 (2015).
[25] Y. Nishikawa, M. Michel, W. Krauth, and K. Hukushima, arXiv:1508.05661 (2015).
[26] A. Ichiki and M. Ohzeki, Phys. Rev. E 88, 020101(R) (2013).
[27] Y. Sun, F. Gomez, and J. Schmidhuber, In: NIPS, 2235 (2010).
[28] P. Brémaud, *Markov Chains, Gibbs Field, Monte Carlo Simulation, and Queues* (Springer, New York, 1999).