Convergence theorems for quantum annealing

Satoshi Morita and Hidetoshi Nishimori

Department of Physics, Tokyo Institute of Technology, Oh-okayama, Meguro-ku, Tokyo 152-8551, Japan

Abstract. We prove several theorems to give sufficient conditions for convergence of quantum annealing, which is a protocol to solve generic optimization problems by quantum dynamics. In particular the property of strong ergodicity is proved for the path-integral Monte Carlo implementation of quantum annealing for the transverse Ising model under a power decay of the transverse field. This result is to be compared with the much slower inverse-log decay of temperature in the conventional simulated annealing. Similar results are proved for the Green's function Monte Carlo approach. Optimization problems in continuous space of particle configurations are also discussed.

1. Introduction

One of the central problems in computer science is to develop efficient algorithms for hard optimization problems [1]. A standard approach is to propose a new algorithm for a given specific problem by improving existing methods or by devising new approaches. Simulated annealing (SA) presents a different perspective, which provides a *generic* algorithm to be applicable in principle to an arbitrary problem [2, 3]. The basic idea is to numerically simulate a physical annealing process by the introduction of a temperature variable under the identification of the cost function to be minimized with the energy of the system. One decreases the temperature from a very high initial value toward zero as the simulation time proceeds with the hope to reach the optimal state (ground state) at the end of the process.

The efficiency of SA is determined by the choice of the annealing schedule, the rate of temperature decrease. A very slow decrease would certainly lead the system to the ground state because the system stays close to equilibrium at each temperature. However, such a slow process is not very useful practically. On the other hand, when the temperature is decreased too quickly, the system may be trapped in a local minimum. It is therefore important to establish criteria on how fast one can decrease the temperature to reach the optimal state avoiding local minima.

A theorem by Geman and Geman [4] gives a generic answer to this problem. Any system is guaranteed to converge to the optimal state in the limit of infinite time if the temperature is decreased in proportion to $N/\log t$ or slower, where N is the system size and t denotes simulation steps. This result is highly non-trivial since the system reaches

the equilibrium state (ground state) after a long non-equilibrium process in which the temperature changes with time at a finite, non-vanishing, rate.

Quantum annealing (QA) is a relatively new alternative to SA, which uses quantum fluctuations, instead of thermal effects, to search the phase space of the system for the optimal state [5, 6, 7, 8, 9, 10]. An artificial term of kinetic energy of quantum nature is introduced, by which the system moves around in the phase space. The cost function is regarded as the potential energy. A slow decrease of the kinetic energy is expected to bring the system towards the optimal state. A related method of quantum adiabatic evolution [11] is based on essentially the same idea.

A remarkable fact is that QA has been found to be more effective in solving optimization problems than SA in most cases numerically investigated so far, including the ground state search of random spin systems [12, 13, 14, 15], protein folding [16], the configuration of molecules in a Lennard-Jones cluster [17], travelling salesman problem [18], simple potentials [19, 20] and a kinetically constrained system [21]. It has also been observed experimentally that a QA-like process leads to equilibrium more efficiently than a thermal process [22]. In contrast, in the instance of 3-SAT, a hard optimization problem, QA has been found not to outperform SA [23]. It is therefore a very interesting problem to establish when and how QA converges to the ground state, preferably with a comparison with SA in mind.

In the present paper we report on a solution of this problem by proving several theorems which give sufficient conditions for convergence of QA. In many numerical studies of QA, stochastic processes are used in the forms of path-integral and Green's function Monte Carlo simulations mainly due to difficulties in directly solving the Schrödinger equation for large systems. Our approach reflects such developments, and we derive convergence conditions for Monte Carlo implementations of QA using the idea of Geman and Geman for convergence conditions for SA.

This paper consists of five sections. Various definitions of an inhomogeneous Markov chain are given in the next section. Convergence of QA is proved for the path-integral and the Green function Monte Carlos in section 3 and section 4, respectively. The last section is devoted to discussions.

2. Ergodicity of inhomogeneous Markov chain

Since we use the theory of stochastic processes, it is useful to recall various definitions and theorems for inhomogeneous Markov processes [3]. We denote the space of discrete states by $\mathcal{S}$ and assume that the size of $\mathcal{S}$ is finite. A Monte Carlo step is characterized by the *transition probability* from state $x(\in \mathcal{S})$ to state $y(\in \mathcal{S})$ at time step t:

$$G(y,x;t) = \begin{cases} P(y,x)A(y,x;t) & (x \neq y) \\ 1 - \sum_{z \in \mathcal{S}} P(z,x)A(z,x;t) & (x = y) \end{cases}, \tag{2.1}$$

where $P(y,x)$ and $A(y,x;t)$ are called the *generation probability* and the *acceptance probability*, respectively. The former is the probability to generate the next candidate

state y from the present state x. We assume that this probability does not depend on time and satisfies the following conditions:

$$\forall x, y \in \mathcal{S} : P(y, x) = P(x, y) \geq 0, \tag{2.2}$$

$$\forall x \in \mathcal{S} : P(x, x) = 0, \tag{2.3}$$

$$\forall x \in \mathcal{S} : \sum_{y \in \mathcal{S}} P(y, x) = 1, \tag{2.4}$$

$$\forall x, y \in \mathcal{S}, \exists n > 0, \exists z_1, \cdots, z_{n-1} \in \mathcal{S} : \prod_{k=0}^{n-1} P(z_{k+1}, z_k) > 0, z_0 = x, z_n = y. \tag{2.5}$$

The last condition represents irreducibility of $\mathcal{S}$, that is, any state in $\mathcal{S}$ can be reached from any other state in $\mathcal{S}$.

We define $\mathcal{S}_x$ as the neighbourhood of x, i.e., the set of states that can be reached by a single step from x:

$$\mathcal{S}_x = \{y \,|\, y \in \mathcal{S}, P(y, x) > 0\}. \tag{2.6}$$

The acceptance probability $A(y, x; t)$ is the probability to accept the candidate y generated from state x. The matrix $G(t)$, whose (y, x) component is given by (2.1), $[G(t)]_{y,x} = G(y, x; t)$, is called the *transition matrix*.

Let $\mathcal{P}$ denote the set of probability distributions on $\mathcal{S}$. We regard a probability distribution $p(\in \mathcal{P})$ as the column vector with the component $[p]_x = p(x)$. The probability distribution at time t, started from an initial distribution $p_0(\in \mathcal{P})$ at time t_0, is written as

$$p(t, t_0) = G^{t,t_0} p_0 \equiv G(t-1)G(t-2)\cdots G(t_0)p_0. \tag{2.7}$$

A Markov chain is called *inhomogeneous* when the transition probability depends on time. In sections 3 and 4, we will prove that inhomogeneous Markov chains associated with QA are ergodic under appropriate conditions. There are two kinds of ergodicity, weak and strong. *Weak ergodicity* means that the probability distribution becomes independent of initial conditions after a sufficiently long time:

$$\forall t_0 \geq 0 : \lim_{t \to \infty} \sup\{\|p(t, t_0) - p'(t, t_0)\| \,|\, p_0, p'_0 \in \mathcal{P}\} = 0, \tag{2.8}$$

where $p(t, t_0)$ and $p'(t, t_0)$ are the probability distributions with different initial distributions p_0 and p'_0. The norm is defined by

$$\|p\| = \sum_{x \in \mathcal{S}} |p(x)|. \tag{2.9}$$

Strong ergodicity is the property that the probability distribution converges to a unique distribution irrespective of initial state:

$$\exists r \in \mathcal{P}, \forall t_0 \geq 0 : \lim_{t \to \infty} \sup\{\|p(t, t_0) - r\| \,|\, p_0 \in \mathcal{P}\} = 0. \tag{2.10}$$

The following two theorems provide conditions for weak and strong ergodicity of an inhomogeneous Markov chain [3].

Theorem 1 (Condition for weak ergodicity) *An inhomogeneous Markov chain is weakly ergodic if and only if there exists a strictly increasing sequence of positive numbers* $\{t_i\}, (i = 0, 1, 2, \ldots)$*, such that*

$$\sum_{i=0}^{\infty} \left(1 - \alpha(G^{t_{i+1},t_i})\right) \longrightarrow \infty, \tag{2.11}$$

where $\alpha(G^{t_{i+1},t_i})$ *is the coefficient of ergodicity defined by*

$$\alpha(G^{t_{i+1},t_i}) = 1 - \min\left\{\sum_{z\in\mathcal{S}} \min\{G(z,x), G(z,y)\} \Big| x, y \in \mathcal{S}\right\} \tag{2.12}$$

with the notation $G(z,x) = [G^{t_{i+1},t_i}]_{z,x}$*.*

Theorem 2 (Condition for strong ergodicity) *An inhomogeneous Markov chain is strongly ergodic if the following three conditions hold:*

(i) the Markov chain is weakly ergodic,

(ii) for all t *there exists a stationary state* $p_t \in \mathcal{P}$ *such that* $p_t = G(t)p_t$*,*

(iii) p_t *satisfies*

$$\sum_{t=0}^{\infty} \|p_t - p_{t+1}\| < \infty. \tag{2.13}$$

Moreover, if $p = \lim_{t\to\infty} p_t$*, then* p *is equal to the probability distribution* r *in (2.10).*

3. Quantum annealing with path-integral Monte Carlo method

3.1. Path-integral Monte Carlo method

Let us first discuss convergence conditions for the implementation of quantum annealing by the path-integral Monte Carlo (PIMC) method [24]. The basic idea of PIMC is to apply the Monte Carlo method to the classical system obtained from the original quantum system by the path-integral formula. It is instructive to first consider the example of ground state search of the Ising spin system as a typical combinatorial optimization problem. The Ising system with generic interactions as discussed below covers a wide range of problems in combinatorial optimization. Examples include the ground-state search of spin glasses, travelling salesman problem, neural networks and the satisfiability problem, many of which have been treated in the Ising expression in the literature mentioned in Introduction.

Quantum fluctuations are introduced by adding a transverse field to the usual Ising spin system. The Hamiltonian of the transverse-field Ising model (TFIM) thus obtained is written as

$$H(t) = -\sum_{\langle ij\rangle} J_{ij}\sigma_i^z\sigma_j^z - \Gamma(t)\sum_{i=0}^{N}\sigma_i^x, \tag{3.1}$$

where the σ_i^α $(\alpha = x, y, z)$ are the Pauli matrices, components of the spin $\frac{1}{2}$ operator at site i, and J_{ij} denotes the coupling constant between sites i and j. There is no restriction

in the spatial dimensionality and the lattice structure. It is also to be noted that the existence of arbitrary many-body interactions between z components of Pauli matrix and longitudinal random magnetic field $\sum h_i \sigma_i^z$, in addition to the above Hamiltonian, would not change the following argument.

The first term of the right-hand side of (3.1) is the cost function (or potential) to be minimized. The transverse field $\Gamma(t)$ represents the strength of kinetic energy of quantum nature, which induces spin flips between up and down states measured in the z direction. In the QA, $\Gamma(t)$ is gradually reduced from a very large (or infinitely large) initial value to zero as time proceeds. By starting from the trivial ground state of the initial system composed only of the transverse-field term $-\Gamma(t)\sum_i \sigma_i^x$ and following the time development of the system under a slow decrease of the transverse field, one hopes to eventually reach the non-trivial ground state of the original problem, $-\sum J_{ij}\sigma_i^z\sigma_j^z$, when $\Gamma(t)$ vanishes. An important problem is how slow is sufficiently slow to achieve this goal.

In the path-integral method, the d-dimensional TFIM is mapped to a $(d+1)$-dimensional classical Ising system so that the quantum system can be simulated on classical computer. In numerical simulations, the Suzuki-Trotter formula [25, 26] is usually employed to express the partition function of the resulting classical system,

$$Z(t) \approx \sum_{\{S_i^{(k)}\}} \exp\left(\frac{\beta}{M}\sum_{k=1}^{M}\sum_{\langle ij\rangle} J_{ij} S_i^{(k)} S_j^{(k)} + \gamma(t)\sum_{k=1}^{M}\sum_{i=0}^{N} S_i^{(k)} S_i^{(k+1)}\right), \tag{3.2}$$

where M is the length along the extra dimension (Trotter number) and $S_i^{(k)}(=\pm 1)$ denotes a classical Ising spin at site i on the kth Trotter slice. The nearest-neighbour interaction between adjacent Trotter slices,

$$\gamma(t) = \frac{1}{2}\log\left(\coth\frac{\beta\Gamma(t)}{M}\right), \tag{3.3}$$

is ferromagnetic. This approximation (3.2) becomes exact in the limit $M \to \infty$ for a fixed $\beta = 1/k_B T$. The magnitude of this interaction (3.3) increases with time t and tends to infinity as $t \to \infty$, reflecting the decrease of $\Gamma(t)$. We fix M and β to arbitrary large values, which corresponds to the actual situation in numerical simulations. Therefore the theorem presented below does not directly guarantee the convergence of the system to the true ground state, which is realized only after taking the limits $M \to \infty$ and $\beta \to \infty$. We will rather show that the system converges to the thermal equilibrium represented by the right-hand side of (3.2), which can be chosen arbitrarily close to the true ground state by taking M and β large enough.

With the above example of TFIM in mind, it will be convenient to treat a more general expression than (3.2),

$$Z(t) = \sum_{x\in\mathcal{S}} \exp\left(-\frac{F_0(x)}{T_0} - \frac{F_1(x)}{T_1(t)}\right). \tag{3.4}$$

Here $F_0(x)$ is the cost function whose global minimum is the desired solution of the combinatorial optimization problem. The temperature T_0 is chosen to be sufficiently

small. The term $F_1(x)$ derives from the kinetic energy, which is the transverse field in the TFIM. Quantum fluctuations are tuned by the extra temperature factor $T_1(t)$, which decreases with time. The first term $-F_0(x)/T_0$ corresponds to the interaction term in the exponent of (3.2), and the second term $-F_1(x)/T_1(t)$ generalizes the transverse-field term in (3.2).

For the partition function (3.4), we define the acceptance probability of PIMC as

$$A(y,x;t) = g\left(\frac{q(y;t)}{q(x;t)}\right), \tag{3.5}$$

$$q(x;t) = \frac{1}{Z(t)}\exp\left(-\frac{F_0(x)}{T_0} - \frac{F_1(x)}{T_1(t)}\right). \tag{3.6}$$

This $q(x;t)$ is the equilibrium Boltzmann factor at a given fixed $T_1(t)$. The function $g(u)$ is the *acceptance function*, a monotone increasing function satisfying $0 \le g(u) \le 1$ and $g(1/u) = g(u)/u$ for $u \ge 0$. For instance, for the heat bath and the Metropolis methods, we have

$$g(u) = \frac{u}{1+u}, \tag{3.7}$$

$$g(u) = \min\{1, u\}, \tag{3.8}$$

respectively. The conditions mentioned above for $g(u)$ guarantee that $q(x;t)$ is the stationary distribution of the homogeneous Markov chain defined by the transition matrix $G(t)$ with a fixed t. In other words, $q(x;t)$ is the right eigenvector of $G(t)$ with eigenvalue 1.

3.2. Convergence theorem for QA-PIMC

We first define a few quantities. The set of local maximum states of F_1 is written as $\mathcal{S}_m$,

$$\mathcal{S}_m = \{x \,|\, x \in \mathcal{S},\ \forall y \in \mathcal{S}_x,\ F_1(y) \le F_1(x)\}. \tag{3.9}$$

We denote by $d(y,x)$ the minimum number of steps necessary to make a transition from x to y. Using this notation we define the minimum number of maximum steps needed to reach any other state from an arbitrary state in the set $\mathcal{S} \setminus \mathcal{S}_m$,

$$R = \min\Big\{\max\{d(y,x) \,|\, y \in \mathcal{S}\} \,\Big|\, x \in \mathcal{S} \setminus \mathcal{S}_m\Big\}. \tag{3.10}$$

Also, L_0 and L_1 stand for the maximum changes of $F_0(x)$ and $F_1(x)$, respectively, in a single step,

$$L_0 = \max\{|F_0(x) - F_0(y)| \,|\, P(y,x) > 0,\ x, y \in \mathcal{S}\}, \tag{3.11}$$

$$L_1 = \max\{|F_1(x) - F_1(y)| \,|\, P(y,x) > 0,\ x, y \in \mathcal{S}\}. \tag{3.12}$$

Our main results are summarized in the following theorem and its corollary.

Theorem 3 (Strong ergodicity of the system (3.4)) *The inhomogeneous Markov chain generated by (3.5) and (3.6) is strongly ergodic and converges to the equilibrium state corresponding to the first term of the right-hand side of (3.6),* $\exp(-F_0(x)/T_0)$*, if*

$$T_1(t) \ge \frac{RL_1}{\log(t+2)}. \tag{3.13}$$

Application of this theorem to the PIMC implementation of QA represented by (3.2) immediately yields the following corollary.

Corollary 1 (Strong ergodicity of QA-PIMC for TFIM) *The inhomogeneous Markov chain generated by the Boltzmann factor on the right-hand side of (3.2) is strongly ergodic and converges to the equilibrium state corresponding to the first term on the right-hand side of (3.2) if*

$$\Gamma(t) \geq \frac{M}{\beta} \tanh^{-1} \frac{1}{(t+2)^{2/RL_1}}. \tag{3.14}$$

Remark. For sufficiently large t, the above inequality reduces to

$$\Gamma(t) \geq \frac{M}{\beta} (t+2)^{-2/RL_1}. \tag{3.15}$$

This result implies that a power decay of the transverse field is sufficient to guarantee the convergence of quantum annealing of TFIM by the PIMC.

To prove strong ergodicity it is necessary to prove weak ergodicity first. The following lemma is useful for this purpose. The proof of this lemma is given in Appendix A.

Lemma 1 (Lower bound on the transition probability) *The elements of the transition matrix defined by (2.1), (3.5) and (3.6) have the following lower bound:*

$$P(y,x) > 0 \Rightarrow \forall t > 0 : G(y,x;t) \geq w\, g(1) \exp\left(-\frac{L_0}{T_0} - \frac{L_1}{T_1(t)} \right), \tag{3.16}$$

and

$$\exists t_1 > 0, \forall x \in \mathcal{S} \setminus \mathcal{S}_m, \forall t \geq t_1 : G(x,x;t) \geq w\, g(1) \exp\left(-\frac{L_0}{T_0} - \frac{L_1}{T_1(t)} \right). \tag{3.17}$$

Proof of weak ergodicity implied in Theorem 3. Let us introduce the following quantity

$$x^* = \arg\min \Big\{ \max \{ d(y,x) \,|\, y \in \mathcal{S} \} \,\Big|\, x \in \mathcal{S} \setminus \mathcal{S}_m \Big\}. \tag{3.18}$$

Comparison with the definition of R in (3.10) implies that the state x^* is reachable by at most R transitions from any states. Also, w stands for the minimum non-vanishing value of $P(y,x)$,

$$w = \min \{ P(y,x) \,|\, P(y,x) > 0,\ x, y \in \mathcal{S} \}. \tag{3.19}$$

Now, consider the transition probability from an arbitrary state x to x^*. From the definitions of R and x^*, there exists at least one transition route within R steps:

$$x \equiv x_0 \neq x_1 \neq x_2 \neq \cdots \neq x_l = x_{l+1} = \cdots = x_R \equiv x^*.$$

Then Lemma 1 yields that, for sufficiently large t, the transition probability at each time step has the following lower bound:

$$G(x_{i+1}, x_i; t-R+i) \geq w g(1) \exp\left(-\frac{L_0}{T_0} - \frac{L_1}{T_1(t-R+i)} \right). \tag{3.20}$$

Thus, by taking the product of (3.20) from $i = 0$ to $i = R - 1$, we have

$$\begin{aligned} G^{t,t-R}(x^*, x) &\geq G(x^*, x_{R-1}; t-1)G(x_{R-1}, x_{R-2}; t-2)\cdots G(x_1, x; t-R) \\ &\geq \prod_{i=0}^{R-1} w\, g(1) \exp\left(-\frac{L_0}{T_0} - \frac{L_1}{T_1(t-R+i)}\right) \\ &\geq w^R g(1)^R \exp\left(-\frac{RL_0}{T_0} - \frac{RL_1}{T_1(t-1)}\right), \end{aligned} \tag{3.21}$$

where we have used monotonicity of $T_1(t)$. Consequently, it is possible to find an integer $k_0 \geq 0$ such that, for all $k > k_0$, the coefficient of ergodicity satisfies

$$1 - \alpha(G^{kR,kR-R})) \geq w^R g(1)^R \exp\left(-\frac{RL_0}{T_0} - \frac{RL_1}{T_1(kR-1)}\right). \tag{3.22}$$

We now substitute the annealing schedule (3.13). Then weak ergodicity is immediately proved from Theorem 1 because we obtain

$$\sum_{k=1}^{\infty}(1 - \alpha(G^{kR,kR-R})) \geq w^R g(1)^R \exp\left(-\frac{RL_0}{T_0}\right) \sum_{k=k_0}^{\infty} \frac{1}{kR+1} \longrightarrow \infty. \tag{3.23}$$

Proof of Theorem 3. To prove strong ergodicity, we refer to Theorem 2. The condition (i) has already been proved. As has been mentioned, the Boltzmann factor (3.6) satisfies $q(t) = G(t)q(t)$, which is the condition (ii). Thus the proof will be complete if we prove the condition (iii) by setting $p_t = q(t)$. As shown in Appendix B, $q(x;t)$ is monotonically increasing for large t:

$$\forall t \geq 0, \forall x \in \mathcal{S}_1^{\min} : q(x; t+1) \geq q(x; t), \tag{3.24}$$

$$\exists t_1 > 0, \forall t \geq t_1, \forall x \in \mathcal{S} \setminus \mathcal{S}_1^{\min} : q(x; t+1) \leq q(x; t), \tag{3.25}$$

where $\mathcal{S}_1^{\min}$ denotes the set of global minimum states of F_1. Consequently, for all $t > t_1$, we have

$$\begin{aligned} \|q(t+1) - q(t)\| &= \sum_{x \in \mathcal{S}_1^{\min}} \{q(x; t+1) - q(x; t)\} - \sum_{x \notin \mathcal{S}_1^{\min}} \{q(x; t+1) - q(x; t)\} \\ &= 2 \sum_{x \in \mathcal{S}_1^{\min}} \{q(x; t+1) - q(x; t)\}, \end{aligned} \tag{3.26}$$

where we used $\|q(t)\| = \sum_{x \in \mathcal{S}_1^{\min}} q(x; t) + \sum_{x \notin \mathcal{S}_1^{\min}} q(x; t) = 1$. We then obtain

$$\sum_{t=t_1}^{\infty} \|q(t+1) - q(t)\| = 2 \sum_{x \in \mathcal{S}_1^{\min}} \{q(x; \infty) - q(x; t_1)\} \leq 2. \tag{3.27}$$

Therefore $q(t)$ satisfies the condition (iii):

$$\begin{aligned} \sum_{t=0}^{\infty} \|q(t+1) - q(t)\| &= \sum_{t=0}^{t_1-1} \|q(t+1) - q(t)\| + \sum_{t=t_1}^{\infty} \|q(t+1) - q(t)\| \\ &\leq 2t_1 + 2 < \infty, \end{aligned} \tag{3.28}$$

which completes the proof of strong ergodicity.

3.3. Remarks

Remark 1. In the above analyses we treated systems with discrete degrees of freedom. Theorem 3 does not apply directly to a continuous system. Nevertheless, by discretization of the continuous space we obtain the following result.

Let us consider a system of N distinguishable particles in a continuous space of finite volume with the Hamiltonian

$$H = \frac{1}{2m(t)} \sum_{i=1}^{N} \boldsymbol{p}_i^2 + V(\{\boldsymbol{r}_i\}). \tag{3.29}$$

The mass $m(t)$ controls the magnitude of quantum fluctuations. The goal is to find the minimum of the potential term, which is achieved by a gradual increase of $m(t)$ to infinity according to the prescription of QA. After discretization of the continuous space (which is necessary anyway in any computer simulations with finite precision) and an application of the Suzuki-Trotter formula, the equilibrium partition function acquires the following expression in the representation to diagonalize spatial coordinates

$$Z(t) \approx \mathrm{Tr} \exp\left(-\frac{\beta}{M} \sum_{k=1}^{M} V\left(\{\boldsymbol{r}_i^{(k)}\}\right) - \frac{Mm(t)}{2\beta} \sum_{i=1}^{N} \sum_{k=1}^{M} \left| \boldsymbol{r}_i^{(k+1)} - \boldsymbol{r}_i^{(k)} \right|^2 \right), \tag{3.30}$$

where we choose the unit $\hbar = 1$. Theorem 3 is applicable to this system under the identification of $T_1(t)$ with $m(t)^{-1}$. We therefore conclude that a logarithmic increase of the mass suffices to guarantee strong ergodicity of the potential-minimization problem under spatial discretization.

The coefficient corresponding to the numerator of the right-hand side of (3.13) is estimated as

$$RL_1 \approx M^2 N L^2 / \beta, \tag{3.31}$$

where L denotes the maximum value of $\left| \boldsymbol{r}_i^{(k+1)} - \boldsymbol{r}_i^{(k)} \right|$. To obtain this coefficient, let us consider two extremes. One is that any states are reachable at one step. By definition, $R = 1$ and $L_1 \approx M^2 N L^2 / \beta$, which yield (3.31). The other case is that only one particle can move to the nearest neighbour point at one time step. With a ($\ll L$) denoting the lattice spacing, we have

$$L_1 \approx \frac{M}{2\beta} \left\{ L^2 - (L-a)^2 \right\} \approx \frac{MLa}{\beta}. \tag{3.32}$$

Since the number of steps to reach any configurations is estimated as $R \approx NML/a$, we again obtain (3.31).

Remark 2. In Theorem 3, the acceptance probability is defined by the conventional Boltzmann form, (3.5) and (3.6). However, we have the freedom to choose any transition (acceptance) probability as long as it is useful to achieve our objective since our goal is not to find finite-temperature equilibrium states but to identify the optimal state. There have been attempts to accelerate the annealing schedule in SA by modifying the transition probability. In particular Nishimori and Inoue [27] have proved weak ergodicity of the inhomogeneous Markov chain for classical simulated annealing using

the probability of Tsallis and Stariolo [28]. There the property of weak ergodicity was shown to hold under the annealing schedule of temperature inversely proportional to a power of time steps. This annealing rate is much faster than the log-inverse law of Geman and Geman for the conventional Boltzmann factor.

A similar generalization is possible for QA-PIMC by using the following modified acceptance probability

$$A(y,x;t) = g\left(u(y,x;t)\right), \tag{3.33}$$

$$u(y,x;t) = \mathrm{e}^{-(F_0(y)-F_0(x))/T_0}\left\{1+(q-1)\frac{F_1(y)-F_1(x)}{T_1(t)}\right\}^{1/(1-q)}, \tag{3.34}$$

where q is a real number. In the limit $q \to 1$, this acceptance probability reduces to the Boltzmann form. Similarly to the discussions leading to Theorem 3, we can prove that the inhomogeneous Markov chain with this acceptance probability is weakly ergodic if

$$T_1(t) \geq \frac{b}{(t+2)^c}, \qquad 0 < c \leq \frac{q-1}{R}, \tag{3.35}$$

where b is a positive constant. We have to restrict ourselves to the case $q > 1$ for a technical reason as was the case previously [27]. We do not reproduce the proof here because it is quite straightforward to generalize the discussions for Theorem 3 in combination with the argument of [27]. The result (3.35) applied to the TFIM is that, if the annealing schedule asymptotically satisfies

$$\Gamma(t) \geq \frac{M}{\beta}\exp\left(-\frac{2(t+2)^c}{b}\right), \tag{3.36}$$

the inhomogeneous Markov chain is weakly ergodic. Notice that this annealing schedule is faster than the power law of (3.15). We have been unable to prove strong ergodicity because we could not identify the stationary distribution for a fixed $T_1(t)$ in the present case.

4. Quantum annealing with Green's function Monte Carlo method

The path-integral Monte Carlo simulates only the equilibrium behaviour at finite temperature because its starting point is the equilibrium partition function. Moreover, it follows an artificial time evolution of Monte Carlo dynamics, not the natural Schrödinger dynamics. An alternative approach to improve these points is the Green's function Monte Carlo (GFMC) method [24, 29, 30]. The basic idea is to solve the imaginary-time Schrödinger equation by stochastic processes. The Schrödinger dynamics with imaginary time has an extra advantage that one can reach the optimal state more efficiently than by real-time dynamics [31]. Thus, for our purpose to solve optimization problems, it is more important to discuss imaginary-time Schrödinger equation than the "natural" real-time evolution.

In the present section we derive sufficient conditions for strong ergodicity to hold in GFMC.

4.1. Green's function Monte Carlo method

The evolution of states by the imaginary-time Schrödinger equation starting from an initial state $|\psi_0\rangle$ is expressed as

$$|\psi(t)\rangle = \mathrm{T}\exp\left(-\int_0^t \mathrm{d}t' H(t')\right)|\psi_0\rangle, \tag{4.1}$$

where T is the time-ordering operator. The right-hand side can be decomposed into a product of small-time evolutions,

$$|\psi(t)\rangle = \lim_{n\to\infty} \hat{G}_0(t_{n-1})\hat{G}_0(t_{n-2})\cdots\hat{G}_0(t_1)\hat{G}_0(t_0)|\psi_0\rangle, \tag{4.2}$$

where $t_k = k\Delta t$, $\Delta t = t/n$ and $\hat{G}_0(t) = 1 - \Delta t \cdot H(t)$. In the GFMC, one approximates the right-hand side of this equation by a product with large but finite n and replaces $\hat{G}_0(t)$ with $\hat{G}_1(t) = 1 - \Delta t(H(t) - E_T)$, where E_T is called the reference energy to be taken approximately close to the final ground-state energy. This subtraction of the reference energy simply adjusts the standard of energy and changes nothing physically. However, practically, this term is important to keep the matrix elements positive and to accelerate convergence to the ground state as will be explained shortly.

To realize the process of (4.2) by a stochastic method, we rewrite this equation in a recursive form,

$$\psi_{k+1}(y) = \sum_x \hat{G}_1(y, x; t_k)\psi_k(x), \tag{4.3}$$

where $\psi_k(x) = \langle x|\psi_k\rangle$ and $|x\rangle$ denotes a basis state. The matrix element of Green's function is given by

$$\hat{G}_1(y, x; t) = \langle y|1 - \Delta t(H(t) - E_T)|x\rangle. \tag{4.4}$$

Equation (4.3) looks similar to a Markov process but is significantly different in several ways. An important difference is that the Green's function is not normalized, $\sum_y \hat{G}_1(y, x; t) \neq 1$. In order to avoid this problem, one decomposes the Green's function into a normalized probability G_1 and a weight w:

$$\hat{G}_1(y, x; t) = G_1(y, x; t)w(x; t), \tag{4.5}$$

where

$$G_1(y, x; t) \equiv \frac{\hat{G}_1(y, x; t)}{\sum_y \hat{G}_1(y, x; t)}, \quad w(x; t) \equiv \frac{\hat{G}_1(y, x; t)}{G_1(y, x; t)}. \tag{4.6}$$

Thus, using (4.3), the wave function at time t is written as

$$\begin{aligned}\psi_n(y) = \sum_{\{x_k\}} &\delta_{y,x_n} w(x_{n-1}; t_{n-1})w(x_{n-2}; t_{n-2})\cdots w(x_0; t_0) \\ &\times G_1(x_n, x_{n-1}; t_{n-1})G_1(x_{n-1}, x_{n-2}; t_{n-2})\cdots G_1(x_1, x_0; t_0)\psi_0(x_0).\end{aligned} \tag{4.7}$$

The algorithm of GFMC is based on this formula and is defined by a weighted random walk in the following sense. One first prepares an arbitrary initial wave function $\psi_0(x_0)$, all elements of which are non-negative. A random walker is generated, which sits initially ($t = t_0$) at the position x_0 with a probability proportional to

$\psi_0(x_0)$. Then the walker moves to a new position x_1 following the transition probability $G_1(x_1, x_0; t_0)$. Thus this probability should be chosen non-negative by choosing parameters appropriately as described later. Simultaneously, the weight of this walker is updated by the rule $W_1 = w(x_0; t_0)W_0$ with $W_0 = 1$. This stochastic process is repeated to $t = t_{n-1}$. One actually prepares M independent walkers and let those walkers follow the above process. Then, according to (4.7), the wave function $\psi_n(y)$ is approximated by the distribution of walkers at the final step weighted by W_n,

$$\psi_n(y) = \lim_{M\to\infty} \frac{1}{M} \sum_{i=1}^{M} W_n^{(i)} \delta_{y,x_n^{(i)}}, \tag{4.8}$$

where i is the index of a walker.

As noted above, $G_1(y, x; t)$ should be non-negative, which is achieved by choosing sufficiently small Δt (i.e. sufficiently large n) and selecting E_T within the instantaneous spectrum of the Hamiltonian $H(t)$. In particular, when E_T is close to the instantaneous ground-state energy of $H(t)$ for large t (i.e. the final target energy), $\hat{G}_1(x, x; t)$ is close to unity whereas other matrix components of $\hat{G}_1(t)$ are small. Thus, by choosing E_T this way, one can accelerate convergence of GFMC to the optimal state in the last steps of the process.

If we apply this general framework to the TFIM with the σ^z-diagonal basis, the matrix elements of Green's function are immediately calculated as

$$\hat{G}_1(y, x; t) = \begin{cases} 1 - \Delta t(E_0(x) - E_T) & (x = y) \\ \Delta t\, \Gamma(t) & (x \text{ and } y \text{ differ by a single-spin flip}) \\ 0 & (\text{otherwise}), \end{cases} \tag{4.9}$$

where $E_0(x) = \langle x| \left(-\sum_{ij} J_{ij}\sigma_i^z\sigma_j^z\right) |x\rangle$. One should choose Δt and E_T such that $1 - \Delta t(E_0(x) - E_T) \geq 0$ for all x. Since $w(x, t) = \sum_y \hat{G}_1(y, x; t)$, the weight is given by

$$w(x; t) = 1 - \Delta t(E_0(x) - E_T) + N\Delta t\, \Gamma(t). \tag{4.10}$$

One can decompose this transition probability into the generation probability and the acceptance probability as in (2.1):

$$P(y, x) = \begin{cases} \dfrac{1}{N} & (\text{single-spin flip}) \\ 0 & (\text{otherwise}) \end{cases} \tag{4.11}$$

$$A(y, x; t) = \frac{N\Delta t\, \Gamma(t)}{1 - \Delta t(E_0(x) - E_T) + N\Delta t\, \Gamma(t)}. \tag{4.12}$$

We shall analyze the convergence properties of stochastic processes under these probabilities for TFIM.

4.2. Convergence theorem for QA-GFMC

Similarly to the QA by PIMC, it is necessary to reduce the strength of quantum fluctuations slowly enough in order to find the ground state in the GFMC. The following theorem provides a sufficient condition in this regard.

Theorem 4 (Strong ergodicity of QA-GFMC) *The inhomogeneous Markov process of random walker for the QA-GFMC of TFIM, (2.1), (4.11) and (4.12), is strongly ergodic if*

$$\Gamma(t) \geq \frac{b}{(t+1)^c}, \qquad 0 < c \leq \frac{1}{N}. \tag{4.13}$$

The lower bound of the transition probability given in the following lemma will be used in the proof of Theorem 4.

Lemma 2 *The transition probability of random walk in the GFMC defined by (2.1), (4.11) and (4.12) has the lower bound:*

$$P(y,x) > 0 \Rightarrow \forall t > 0 : G_1(y,x;t) \geq \frac{\Delta t\, \Gamma(t)}{1 - \Delta t(E_{\min} - E_T) + N\Delta t\, \Gamma(t)}, \tag{4.14}$$

$$\exists t_1 > 0, \forall t > t_1 : G_1(x,x;t) \geq \frac{\Delta t\, \Gamma(t)}{1 - \Delta t(E_{\min} - E_T) + N\Delta t\, \Gamma(t)}, \tag{4.15}$$

where $E_{\min}$ *is the minimum value of* $E_0(x)$

$$E_{\min} = \min\{E_0(x) | x \in \mathcal{S}\}. \tag{4.16}$$

Proof of Lemma 2. The first part of Lemma 2 is trivial because the transition probability is an increasing function with respect to $E_0(x)$ when $P(y,x) > 0$ as seen in (4.12). Next, we prove the second part of Lemma 2. According to (4.9) and (4.10), $G_1(x,x;t)$ is written as

$$G_1(x,x;t) = 1 - \frac{N\Delta t\, \Gamma(t)}{1 - \Delta t(E_0(x) - E_T) + N\Delta t\, \Gamma(t)}. \tag{4.17}$$

Since the transverse field $\Gamma(t)$ decreases to zero with time, the second term on the right-hand side tends to zero as $t \to \infty$. Thus, there exists $t_1 > 0$ such that $G_1(x,x;t) > 1-\varepsilon$ for $\forall \varepsilon > 0$ and $\forall t > t_1$. On the other hand, the right-hand side of (4.15) converges to zero as $t \to \infty$. We therefore have (4.15).

Proof of Theorem 4. We show that the condition (4.13) is sufficient to satisfy the three conditions of Theorem 2.

(i) From Lemma 2, we obtain a bound on the coefficient of ergodicity for sufficiently large k as

$$1 - \alpha(G_1^{kN,kN-N}) \geq \left\{ \frac{\Delta t\, \Gamma(kN-1)}{1 - \Delta t(E_{\min} - E_T) + N\Delta t\, \Gamma(kN-1)} \right\}^N, \tag{4.18}$$

in the same manner as we derived (3.22), where we used $R = N$. Substituting the annealing schedule (4.13), we can prove weak ergodicity from Theorem 1 because

$$\sum_{k=1}^{\infty} \left(1 - \alpha(G_1^{kN,kN-N})\right) \geq \sum_{k=k_0}^{\infty} \frac{b^N}{(kN)^{cN}} \tag{4.19}$$

which diverges when $0 < c \leq 1/N$.

(ii) As shown in Appendix C, the stationary distribution of the instantaneous transition probability $G_1(y,x;t)$ is

$$q(x;t) \equiv \frac{w(x;t)}{\sum_{x\in\mathcal{S}} w(x;t)} = \frac{1}{2^N} - \frac{\Delta t\, E_0(x)}{2^N\left\{1+\Delta t\, E_T + N\Delta t\,\Gamma(t)\right\}}. \tag{4.20}$$

(iii) Since the transverse field $\Gamma(t)$ decreases monotonically with t, the above stationary distribution $q(x;t)$ is an increasing function of t if $E_0(x)<0$ and is decreasing if $E_0 \geq 0$. Consequently, using the same procedure as in (3.26), we have

$$\|q(t+1)-q(t)\| = 2\sum_{E_0(x)<0}\{q(x;t+1)-q(x;t)\}, \tag{4.21}$$

and thus

$$\sum_{t=0}^{\infty}\|q(t+1)-q(t)\| = 2\sum_{E_0(x)<0}\{q(x;\infty)-q(x;0)\} \leq 2. \tag{4.22}$$

Therefore the sum $\sum_{t=0}^{\infty}\|q(t+1)-q(t)\|$ is finite, which completes the proof of the condition (iii).

Remark. Theorem 4 asserts convergence of the distribution of random walkers to the equilibrium distribution (4.20) with $\Gamma(t)\to 0$. This implies that the final distribution is not delta-peaked at the ground state with minimum $E_0(x)$ but is a relatively mild function of this energy. The optimality of the solution is achieved after one takes the weight factor $w(x;t)$ into account: The repeated multiplication of weight factors as in (4.7), in conjunction with the relatively mild distribution coming from the product of G_1 as mentioned above, leads to the asymptotically delta-peaked wave function $\psi_n(y)$ because $w(x;t)$ is larger for smaller $E_0(x)$ as seen in (4.10).

4.3. Alternative choice of Green's function

So far we have used the Green's function defined in (4.4), which is linear in the transverse field, allowing single-spin flips only. It may be useful to consider another type of Green's function which accommodates multi-spin flips. Let us try the following form of Green's function,

$$\hat{G}_2(t) = \exp\left(\Delta t\,\Gamma(t)\sum_i \sigma_i^x\right)\exp\left(\Delta t\sum_{ij} J_{ij}\sigma_i^z\sigma_j^z\right), \tag{4.23}$$

which is equal to $\hat{G}_0(t)$ to the order Δt. The matrix element of $\hat{G}_2(t)$ in the σ^z-diagonal basis is

$$\hat{G}_2(y,x;t) = \cosh^N(\Delta t\,\Gamma(t))\tanh^{\delta}(\Delta t\,\Gamma(t))\,\mathrm{e}^{-\Delta t\,E_0(x)}, \tag{4.24}$$

where δ is the number of spins in different states in x and y. According to the scheme of GFMC, we decompose $\hat{G}_2(y,x;t)$ into the normalized transition probability and the weight:

$$G_2(y,x;t) = \left\{\frac{\cosh(\Delta t\,\Gamma(t))}{\mathrm{e}^{\Delta t\,\Gamma(t)}}\right\}^N \tanh^{\delta}(\Delta t\,\Gamma(t)), \tag{4.25}$$

$$w_2(x;t) = \mathrm{e}^{\Delta t\, N\Gamma(t)}\mathrm{e}^{-\Delta t\, E_0(x)}. \tag{4.26}$$

It is remarkable that the transition probability G_2 is independent of $E_0(x)$. Thus, the stationary distribution of random walk is uniform. This property is lost if one interchanges the order of the two factors in (4.23).

The property of strong ergodicity can be shown to hold in this case as well:

Theorem 5 (Strong ergodicity of QA-GFMC 2) *The inhomogeneous Markov chain generated by (4.25) is strongly ergodic if*

$$\Gamma(t) \geq -\frac{1}{2\Delta t}\log\left\{1 - 2b(t+1)^{-1/N}\right\}. \tag{4.27}$$

Remark. For sufficiently large t, the above annealing schedule is reduced to

$$\Gamma(t) \geq \frac{b}{\Delta t\,(t+1)^{1/N}}. \tag{4.28}$$

Since the proof is quite similar to the previous cases, we just outline the idea of the proof. The transition probability $G_2(y,x;t)$ becomes smallest when $\delta = N$. Consequently, the coefficient of ergodicity is estimated as

$$1 - \alpha(G_2^{t+1,t}) \geq \left\{\frac{1-\mathrm{e}^{-2\Delta t\,\Gamma(t)}}{2}\right\}^N.$$

We note that R is equal to 1 in the present case because any states are reachable from an arbitrary state in a single step. From Theorem 1, the condition

$$\left\{\frac{1-\mathrm{e}^{-2\Delta t\,\Gamma(t)}}{2}\right\}^N \geq \frac{b'}{t+1} \tag{4.29}$$

is sufficient for weak ergodicity. From this, one obtains (4.27). Since the stationary distribution of $G_2(y,x;t)$ is uniform as mentioned above, strong ergodicity readily follows from Theorem 2.

Similarly to the case of PIMC, we can discuss the convergence condition of QA-GFMC in systems with continuous degrees of freedom. The resulting sufficient condition is a logarithmic increase of the mass as will be shown now. The operator $\hat{G}_2$ generated by the Hamiltonian (3.29) is written as

$$\hat{G}_2(t) = \exp\left(-\frac{\Delta t}{2m(t)}\sum_{i=1}^N \boldsymbol{p}_i^2\right)\mathrm{e}^{-\Delta t V(\{\boldsymbol{r}_i\})}. \tag{4.30}$$

Thus, the Green's function is calculated in a discretized space as

$$\hat{G}_2(y,x;t) \propto \exp\left(-\frac{m(t)}{2\Delta t}\sum_{i=1}^N |\boldsymbol{r}_i' - \boldsymbol{r}_i|^2 - \Delta t V(\{\boldsymbol{r}_i\})\right), \tag{4.31}$$

where x and y represent $\{\boldsymbol{r}_i\}$ and $\{\boldsymbol{r}_i'\}$, respectively. Summation over y, i.e., integration over $\{\boldsymbol{r}_i'\}$, yields the weight $w(x;t)$, from which the transition probability is obtained:

$$w(x;t) \propto \mathrm{e}^{-\Delta t V(\{\boldsymbol{r}_i\})}, \tag{4.32}$$

$$G_2(y,x;t) \propto \exp\left(-\frac{m(t)}{2\Delta t}\sum_{i=1}^N |\boldsymbol{r}_i' - \boldsymbol{r}_i|^2\right). \tag{4.33}$$

The lower bound for the transition probability depends exponentially on the mass: $G_2(y,x;t) \geq \mathrm{e}^{-Cm(t)}$. Since $1-\alpha(G_2^{t+1,t})$ has the same lower bound, the sufficient condition for weak ergodicity is $\mathrm{e}^{-Cm(t)} \geq (t+1)^{-1}$, which is rewritten as

$$m(t) \leq C^{-1}\log(t+1). \tag{4.34}$$

The constant C is proportional to $NL^2/\Delta t$, where L denotes the maximum value of $|\boldsymbol{r}'-\boldsymbol{r}|$. The derivation of C is similar to (3.31), because $G_2(t)$ allows any transition to arbitrary states at one time step.

5. Discussion

We have proved strong ergodicity of the inhomogeneous Markov chains associated with QA-PIMC and QA-GFMC, mainly with the application to the TFIM in mind, which covers a wide range of combinatorial optimization problems. Our proof is quite general in the sense that it does not depend on the spatial dimensionality or the lattice structure of the system. The convergence of QA is guaranteed if the transverse field decreases as $\Gamma(t) \approx \mathrm{const}/t^c$ asymptotically. This annealing schedule for the transverse field is faster than the temperature-annealing schedule, the log-inverse law, found by Geman and Geman for SA. Moreover, the generalized transition probability in PIMC accelerates the annealing schedule to $\Gamma(t) \approx \exp(-t^c)$ (although we could not prove strong ergodicity in this case). Since the constant c appearing in these formulas depends on the system size as $1/N$ and is therefore very small for large systems, our result may not provide practically useful guidelines to anneal the transverse field. This is the same situation as in SA, in which the temperature annealing should be $N/\log t$ or slower to converge. Nevertheless our Theorems and Corollary represent quite non-trivial results because they assure eventual convergence of the system to the ground state (or a state near the ground state for PIMC) after non-stationary processes without being trapped in local minima.

Let us write a few words on computational complexity. Although the annealing schedule of QA, the power-law dependence on t, is much faster than the log-inverse law for SA, this does not mean that QA provides an algorithm to solve NP problems in polynomial time. The time for $\Gamma(t)$ to reach a sufficiently small value δ is estimated from (3.15) as

$$t_1 \sim \exp\left(\frac{RL_1}{2}\log\frac{M}{\beta\delta}\right). \tag{5.1}$$

Since RL_1 is of the order of N, the QA needs a time exponential in N to converge. An important point is that the coefficient of N in the exponent, $\mathcal{O}(\log\delta^{-1})$, is much smaller than that for SA, in which the coefficient is $\mathcal{O}(1/\delta)$ as can be seen from $T(t) \approx N/\log t \approx \delta$. The situation is the same in QA-GFMC. If one uses the generalized transition probability, the corresponding time is

$$t_2 \sim \exp\left(N\log\left(\log\frac{1}{\delta}\right)\right), \tag{5.2}$$

which again shows exponential dependence on N with a much smaller coefficient.

Acknowledgments

This work was partially supported by CREST, JST. One of the authors (S.M.) is supported by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists.

Appendix A. Proof of Lemma 1

The first part of Lemma 1 is proved straightforwardly. Equation (3.16) follows directly from the definition of the transition probability and the property of the acceptance function g. When $q(y;t)/q(x;t) < 1$, we have

$$G(y,x;t) \geq w\, g\left(\frac{q(x;t)}{q(y;t)}\right)\frac{q(y;t)}{q(x;t)} \geq w\, g(1)\exp\left(-\frac{L_0}{T_0}-\frac{L_1}{T_1(t)}\right). \tag{A.1}$$

On the other hand, if $q(y;t)/q(x;t) \geq 1$,

$$G(y,x;t) \geq w\, g(1) \geq w\, g(1)\exp\left(-\frac{L_0}{T_0}-\frac{L_1}{T_1(t)}\right), \tag{A.2}$$

where we used the fact that both L_0 and L_1 are positive.

Next, we prove (3.17). Since x is not a member of $\mathcal{S}_m$, there exists a state $y \in \mathcal{S}_x$ such that $F_1(y) - F_1(x) > 0$. For such a state y,

$$\lim_{t\to\infty} g\left(\exp\left(-\frac{F_0(y)-F_0(x)}{T_0}-\frac{F_1(y)-F_1(x)}{T_1(t)}\right)\right) = 0, \tag{A.3}$$

because $T_1(t)$ tends to zero as $t \to \infty$ and $0 \leq g(u) \leq u$. Thus, for all $\varepsilon > 0$, there exists $t_1 > 0$ such that

$$\forall t > t_1 : g\left(\exp\left(-\frac{F_0(y)-F_0(x)}{T_0}-\frac{F_1(y)-F_1(x)}{T_1(t)}\right)\right) < \varepsilon. \tag{A.4}$$

We therefore have

$$\begin{aligned}\sum_{z\in\mathcal{S}} P(z,x)A(z,x;t) &= P(y,x)A(y,x;t) + \sum_{z\in\mathcal{S}\setminus\{y\}} P(z,x)A(z,x;t)\\ &< P(y,x)\varepsilon + \sum_{z\in\mathcal{S}\setminus\{y\}} P(z,x)\\ &= 1-(1-\varepsilon)P(y,x),\end{aligned} \tag{A.5}$$

and consequently,

$$G(x,x;t) > (1-\varepsilon)P(y,x) > 0. \tag{A.6}$$

Since the right-hand side of (3.17) can be arbitrarily small for sufficiently large t, we obtain the second part of Lemma 1.

Appendix B. Proof of (3.24) and (3.25)

We use the following notations:

$$A(x) = \exp\left(-\frac{F_0(x)}{T_0}\right), \quad B = \sum_{x\in\mathcal{S}_1^{\min}} A(x), \tag{B.1}$$

$$\Delta(x) = F_1(x) - F_1^{\min}. \tag{B.2}$$

If $x \in \mathcal{S}_1^{\min}$, the Boltzmann distribution can be rewritten as

$$q(x;t) = \frac{A(x)}{B + \sum_{y\in\mathcal{S}\setminus\mathcal{S}_1^{\min}} \exp\left(-\frac{\Delta(y)}{T_1(t)}\right) A(y)}. \tag{B.3}$$

Since $\Delta(y) \geq 0$ by definition, the denominator decreases with time. Thus, we obtain (3.24).

To prove (3.25), we consider the derivative of $q(x;t)$ with respect to $T_1(t)$,

$$\frac{\partial q(x;t)}{\partial T_1(t)} = \frac{A(x)\left\{B\Delta(x) + \sum_{y\in\mathcal{S}\setminus\mathcal{S}_1^{\min}} (F_1(x) - F_1(y)) \exp\left(-\frac{\Delta(y)}{T_1(t)}\right) A(y)\right\}}{T(t)^2 \exp\left(\frac{\Delta(x)}{T_1(t)}\right)\left[B + \sum_{y\in\mathcal{S}\setminus\mathcal{S}_1^{\min}} \exp\left(-\frac{\Delta(y)}{T_1(t)}\right) A(y)\right]^2}. \tag{B.4}$$

Only $F_1(x) - F_1(y)$ in the numerator has the possibility of being negative. However, the first term $B\Delta(x)$ in the curly brackets is larger than the second one for sufficient large t because $\exp(-\Delta(y)/T_1(t))$ tend to zero as $T_1(t) \to \infty$. Thus there exists $t_1 > 0$ such that $\partial q(x;t)/\partial T(t) > 0$ for all $t > t_1$. Since $T_1(t)$ is a decreasing function of t, we have (3.25).

Appendix C. Proof of (4.20)

The transition probability defined by (2.1), (4.11) and (4.12) is rewritten in terms of the weight (4.10) as

$$G_1(y,x;t) = \begin{cases} 1 - \dfrac{N\Delta t\,\Gamma(t)}{w(x;t)} & (x = y) \\ \dfrac{\Delta t\,\Gamma(t)}{w(x;t)} & (x \in \mathcal{S}_y;\ \text{single-spin flip}) \\ 0 & (\text{otherwise}). \end{cases} \tag{C.1}$$

Thus, we have

$$\begin{aligned} \sum_{x\in\mathcal{S}} G_1(y,x;t) q(x;t) &= \left(1 - \frac{N\Delta t\,\Gamma(t)}{w(y;t)}\right)\frac{w(y;t)}{A} + \sum_{x\in\mathcal{S}_y} \frac{\Delta t\,\Gamma(t)}{w(x;t)}\frac{w(x;t)}{A} \\ &= q(y;t) - \frac{N\Delta t\,\Gamma(t)}{A} + \frac{\Delta t\,\Gamma(t)}{A}\sum_{x\in\mathcal{S}_y} 1, \end{aligned} \tag{C.2}$$

where A denotes the normalization factor,

$$\sum_{x\in\mathcal{S}} w(x;t) = \mathrm{Tr}\left\{1-\Delta t\left(-\sum_{\langle ij\rangle} J_{ij}\sigma_i^z\sigma_j^z - E_T\right) + N\Delta t\,\Gamma(t)\right\}$$
$$= 2^N\left\{1+\Delta t\,E_T + N\Delta t\,\Gamma(t)\right\}. \qquad \text{(C.3)}$$

Since the volume of $\mathcal{S}_y$ is N, (C.2) indicates that $q(x;t)$ is the stationary distribution of $G_1(y,x;t)$. The right-hand side of (4.20) is easily derived from the above equation.

References

[1] Garey M R and Johnson D S 1979 *Computers and Intractability: A Guide to the Theory of NP Completeness* (San Francisco: Freeman)
[2] Kirkpatrick S, Gelett S D and Vecchi M P 1983 *Science* **220** 671
[3] Aarts E and Korst J 1984 *Simulated Annealing and Boltzmann Machines: a Stochastic Approach to Combinatorial Optimization and Neural Computing* (New York: Wiley) ch 3
[4] Geman S and Geman D 1984 *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-6** 721
[5] Amara P, Hsu D and Atraub J E 1993 *J. Phys. Chem.* **97** 6715
[6] Finnila A B, Gomez M A, Sebenik C, Stenson C and Doll J D 1994 *Chem. Phys. Lett.* **219** 343
[7] Tanaka K and Horiguchi H 1997 *Trans. Inst. Electron. Inform. Commun. Eng.* **J80** 2117 (in Japanese); Tanaka K and Horiguchi H 2000 Electron. Commun. Jpn. Pt. 3 **83** 84 (English translation)
[8] Kadowaki T and Nishimori H 1998 *Phys. Rev. E* **58** 5355
[9] Kadowaki T 1999 *Thesis* (Tokyo Institute of Technology) quant-ph/0205020
[10] Das A and Chakrabarti B K (eds) 2005 *Quantum Annealing and Related Optimization Methods (Lecture Notes in Physics 679)* (Berlin Heidelberg: Springer)
[11] Fahhi E, Goldstone J, Gutmann S and Sipser M 2000 quant-ph/0001106
[12] Santoro G E, Martoňàk R, Tosatti E and Car R 2002 *Science* **295** 2427
[13] Martoňàk R, Santoro G E and Tosatti E 2002 *Phys. Rev. E* **66** 094203
[14] Sarjala M, Petäjä V and Alava M 2006 *J. Stat. Mech.* P01008
[15] Suzuki S and Okada M 2005 *J. Phys. Soc. Jpn* **74** 1649
[16] Lee Y-H and Berne B J 2000 *J. Phys. Chem. A* **104** 86
[17] Liu P and Berne B J 2003 *J. Chem. Phys.* **118** 2999
[18] Martoňàk R, Santoro G E and Tosatti E 2004 *Phys. Rev. E* **70** 057701
[19] Stella L, Santoro G E and Tosatti E 2005 *Phys. Rev. B* **72** 014303
[20] Stella L, Santoro G E and Tosatti E 2006 *Phys. Rev. E* **73** 144302
[21] Das A, Chakrabarti B K and Stinchcombe R B 2005 *Phys. Rev. E* **72** 026701
[22] Brooke J, Bitko D, Rosenbaum T F and Aeppli G 1999 *Science* **284** 779
[23] Battaglia D A, Santoro G E and Tosatti E 2005 *Phys. Rev. E* **71** 066707
[24] Landau D P and Binder K 2000 *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge: Cambridge University Press) ch 8
[25] Trotter H F 1959 *Proc. Am. Math. Soc.* **10** 545
[26] Suzuki M 1971 *Prog. Theor. Phys.* **46** 1337
[27] Nishimori H and Inoue J 1998 *J. Phys. A: Math. Gen.* **31** 5661
[28] Tsallis C and Stariolo D A 1996 *Physica* **233A** 395
[29] Ceperley D M and Alder B J 1980 *Phys. Rev. Lett.* **45** 566
[30] Trivedi N and Ceperley D M 1990 *Phys. Rev. B* **41** 4552
[31] Stella L, Santoro G E and Tosatti E 2005 *Phys. Rev. B* **72** 014303