International Workshop on Statistical-Mechanical Informatics 2007 (IW-SMI 2007) IOP Publishing
Journal of Physics: Conference Series **95** (2008) 012021 doi:10.1088/1742-6596/95/1/012021

Mathematical aspects of quantum annealing

Hidetoshi Nishimori and Satoshi Morita

Department of Physics, Tokyo Institute of Technology, Tokyo 152-8551, Japan

Abstract. Sufficient conditions for convergence of quantum annealing are derived for optimization problems represented by the Ising model. Three different types of time evolution are considered; the real-time Schrödinger equation, the path-integral Monte Carlo method and the Green's function Monte Carlo method. It is proved that the system under each dynamics reaches the target solution in the limit of infinite time if the transverse field, representing the strength of quantum fluctuations, decreases inversely proportionally to the power of time in the asymptotic region.

1. Introduction

Development of efficient algorithms for optimization problems is one of the central subjects of information science [1]. Approaches to this task are classified into two categories: Either one focuses on a specific problem and aims to achieve highest efficiency, or seeks generic algorithms applicable to an arbitrary problem. A standard method of the latter is simulated annealing (SA), which was proposed using the analogy to statistical mechanics [2, 3]. In SA, the solution is obtained by numerical simulations of an annealing process, where the cost function to be minimized is identified with the potential energy and an artificial temperature variable is introduced to the system as a control parameter. The temperature is decreased gradually from an initial high value towards zero as the simulation proceeds. Thermal fluctuations enable the system to escape from local minima.

Quantum annealing (QA) was proposed as an alternative generic algorithm for optimization problems [4, 5]. QA searches the optimal state by quantum fluctuations instead of thermal effects used in SA. A fictitious kinetic energy of quantum nature is introduced to the system and is gradually reduced in magnitude. If the initial state is the ground state of the initial Hamiltonian, the system is expected to evolve adiabatically under a slow decrease of quantum fluctuations, keeping track of the ground state of the instantaneous Hamiltonian to finally reach the desired non-trivial optimal state. A number of numerical calculations demonstrate that QA is more efficient in solving optimization problems than SA in most cases [6, 7].

An important problem is the rate of change of quantum fluctuations; how quickly to decrease the quantum fluctuations keeping the system in the ground state of the instantaneous Hamiltonian. A very slow decrease of the kinetic energy would certainly lead the system to the ground state of the potential energy adiabatically. However, such a slow process is not very useful practically. On the other hand, too rapid a change of the Hamiltonian may induce non-adiabatic transitions and the system would miss the optimal solution.

Indeed, the same problem exists for SA, the rate of temperature decrease. A theorem proved by Geman and Geman provides a generic answer to this problem [8]: An inverse-log decrease of

the temperature is sufficient to guarantee convergence to the optimal state in the limit of infinite time. Our goal is to establish similar criteria for QA to the result of Geman and Geman for SA.

In the present contribution, we report on convergence theorems of QA for the transverse field Ising model [9, 10]. We consider three dynamics, the real-time Schrödinger equation, the path-integral Monte Carlo method and the Green's function Monte Carlo method. Although these three types of dynamics are quite different from each other, a common condition for convergence has been derived to our surprise: A power-law decrease of the transverse field guarantees convergence to the optimal solution. It is remarkable that this power-law schedule is faster than the inverse-log law in the Geman-Geman theorem for SA.

This paper is organized as follows. We first consider QA with the real-time Schrödinger dynamics. We derive a convergence condition using the adiabatic theorem on the basis of the idea of Somma, Batista and Ortiz [11]. Next, we discuss the convergence theorem of QA with the path-integral Monte Carlo method, which is often used in numerical simulations of large systems. Here, we introduce definitions and concepts for the inhomogeneous Markov chain to describe Monte Carlo dynamics. In section 4, another simulation method, the Green's function Monte Carlo method is investigated. This approach corresponds to the imaginary-time Schrödinger dynamics. The last section is devoted to summary.

2. Real-time Schrödinger evolution

Let us suppose that the optimization problem one wishes to solve can be represented as the ground-state search of an Ising model of general form

$$H_{\rm pot} \equiv -\sum_{i=1}^{N} J_i \sigma_i^z - \sum_{ij} J_{ij} \sigma_i^z \sigma_j^z - \sum_{ijk} J_{ijk} \sigma_i^z \sigma_j^z \sigma_k^z - \cdots, \tag{1}$$

where σ_i^z denotes the z component of the Pauli matrix at site i. Quantum annealing is realized typically by the addition of a time-dependent transverse field

$$H_{\rm kin}(t) \equiv -\Gamma(t) \sum_{i=1}^{N} \sigma_i^x, \tag{2}$$

which may be regarded as the quantum kinetic energy to be compared with the potential energy (1). Initially the coefficient of the kinetic term $\Gamma(t)$ is chosen to be very large, and the total Hamiltonian for the transverse field Ising model (TFIM)

$$H(t) = H_{\rm pot} + H_{\rm kin}(t) \tag{3}$$

is dominated by the second kinetic term. The initial ground state is trivial. The coefficient $\Gamma(t)$ is then decreased gradually toward 0, leaving eventually only the potential term. Accordingly the state vector $|\psi(t)\rangle$, which follows the real-time Schrödinger equation

$$\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = H(t)|\psi(t)\rangle, \tag{4}$$

is expected to evolve from the trivial initial ground state of the transverse field (2) to finally the non-trivial ground state of (1). The problem of central concern in this section is how slowly we should decrease $\Gamma(t)$ to keep the state vector arbitrarily close to the instantaneous ground state of the total Hamiltonian (3), namely the adiabaticity condition, to achieve the goal of minimization of (1). A similar idea has been discussed in the context of quantum information [12]. Our following argument [9] is essentially a zero-temperature version of [11].

The adiabatic theorem [13] provides the excitation probability at time t as

$$|\langle n(t)|\psi(t)\rangle|^2 \simeq \left| \frac{\left\langle n(t) \left| \frac{\partial H(t)}{\partial t} \right| 0(t) \right\rangle}{(\varepsilon_n(t) - \varepsilon_0(t))^2} \right|^2, \tag{5}$$

where $|n(t)\rangle$ is the nth instantaneous eigenstate of $H(t)$ with the eigenvalue $\varepsilon_n(t)$. We assume that $|0(t)\rangle$ is the ground state of $H(t)$ and $|\psi(0)\rangle = |0(0)\rangle$. The probability (5) needs to be arbitrarily small for the success of QA. We therefore evaluate an upper bound of

$$\frac{\left| \left\langle n(t) \left| \frac{\partial H(t)}{\partial t} \right| 0(t) \right\rangle \right|}{(\varepsilon_n(t) - \varepsilon_0(t))^2}. \tag{6}$$

For this purpose we estimate the numerator and the denominator of (6). As for the numerator, it is straightforward to see

$$\left| \left\langle n(t) \left| \frac{\partial H(t)}{\partial t} \right| 0(t) \right\rangle \right| \leq -N \frac{\mathrm{d}\Gamma}{\mathrm{d}t}, \tag{7}$$

since the time dependence of $H(t)$ lies only in the kinetic term, which has N terms. Note that $\mathrm{d}\Gamma/\mathrm{d}t$ is negative.

A lower bound on the denominator of (6) can be evaluated using an inequality for a strictly positive operator [14]. First we recall that the Perron-Frobenius theorem states that a non-negative square matrix M has a real eigenvalue λ_0 satisfying $|\lambda| \leq \lambda_0$ for any other eigenvalue λ. If all the elements of M are strictly positive, $M_{ij} > 0$, its eigenvalues satisfy the stronger inequality,

$$|\lambda| \leq \frac{\kappa - 1}{\kappa + 1} \lambda_0, \tag{8}$$

where κ is defined by

$$\kappa = \max_{i,j,k} \frac{M_{ik}}{M_{jk}}. \tag{9}$$

We apply the above inequality (8) to the operator $M \equiv (E_{\max} - H(t))^N$, where $E_{\max}$ is the largest eigenvalue of H_{pot}. All the elements of the matrix M are strictly positive in the representation that diagonalizes σ_i^z because $E_{\max} - H(t)$ is non-negative and irreducible (that is, any state can be reached from any other state within N steps at most). In the asymptotic region $t \gg 1$ where $\Gamma(t) \ll 1$, the minimum element of M, which is between two states having all spins in mutually opposite directions, is equal to $N!\Gamma(t)^N$, where $N!$ comes from the ways of permutation to flip spins. Replacement of H_{kin} by $-N$ shows that the maximum matrix element of M has the upper bound $(E_{\max} - E_{\min} + N)^N$, where $E_{\min}$ is the lowest eigenvalue of H_{pot}. Thus we have

$$\kappa \leq \frac{(E_{\max} - E_{\min} + N)^N}{N!\Gamma(t)^N}. \tag{10}$$

Since the eigenvalue of $H(t)$ is denoted by $\varepsilon_n(t)$, (8) is rewritten as

$$(E_{\max} - \varepsilon_n(t))^N \leq \frac{\kappa - 1}{\kappa + 1} (E_{\max} - \varepsilon_0(t))^N. \tag{11}$$

Substitution of (10) into the above inequality yields

$$\varepsilon_n(t) - \varepsilon_0(t) \geq \frac{2(E_{\max} - \varepsilon_0(t))N!}{N(E_{\max} - E_{\min} + N)^N} \Gamma(t)^N \equiv A\Gamma(t)^N, \tag{12}$$

where we used $\kappa \gg 1$ in the asymptotic time region where $\Gamma(t)$ is very small. The coefficient A is estimated using the Stirling formula as

$$A \simeq \frac{2\sqrt{2\pi N}(E_{\max} - \varepsilon_0(t))}{N\mathrm{e}^N} \left(\frac{N}{E_{\max} - E_{\min} + N} \right)^N, \tag{13}$$

which demonstrates that A is exponentially small for large N.

Now, by the combination of the above estimates (7) and (12), we find that the sufficient condition for convergence is that the upper bound of (6)

$$-\frac{N}{A^2 \Gamma(t)^{2N}} \frac{\mathrm{d}\Gamma}{\mathrm{d}t} \tag{14}$$

is arbitrarily small. By equating (14) to a small constant δ and integrating the resulting differential equation, we find

$$\Gamma(t) = (\alpha\, t)^{-1/(2N-1)}, \tag{15}$$

where α is exponentially small for large N and is proportional to δ. The transverse field should be decreased following this functional form or slower. Therefore the asymptotic power decay of the transverse field guarantees that the excitation probability is bounded by the arbitrarily small constant δ^2 at each time.

The same discussions apply to quantum annealing using transverse ferromagnetic interactions in addition to a transverse field,

$$\tilde{H}_{\mathrm{kin}}(t) = -\tilde{\Gamma}(t) \left(\sum_{i=1}^{N} \sigma_i^x + \sum_{ij} \sigma_i^x \sigma_j^x \right). \tag{16}$$

A recent study showed the effectiveness of this type of quantum kinetic energy [15]. A modification of the strictly positive operator to $(E_{\max} - H(t))^{N/2}$ in the above argument leads to a lower bound of the energy gap as a quantity proportional to $\tilde{\Gamma}(t)^{N/2}$. The resulting asymptotic annealing schedule is

$$\tilde{\Gamma}(t) \propto t^{-1/(N-1)}, \tag{17}$$

which is faster than the case with the transverse field only, equation (15). This result implies that the additional non-zero off-diagonal elements of $H(t)$ would widen the energy gap and accelerate the convergence of QA.

3. Path-integral Monte Carlo method

In the previous section, we discussed convergence of QA with the Schrödinger evolution. However, directly solving the Schrödinger equation is difficult on the classical computer because of exponentially large computational resourced needed. Thus, in many numerical studies of QA, stochastic processes are used in the forms of the path-integral and Green's function Monte Carlo simulations. In this section, we derive convergence conditions for Monte Carlo implementations of QA using the idea of Geman and Geman for convergence conditions for SA. First we consider the path-integral Monte Carlo method and introduce various definitions and theorems for inhomogeneous Markov chain [3]. In the next section we describe convergence of QA with Green's function Monte Carlo.

3.1. Inhomogeneous Markov chain associated with path-integral Monte Carlo method
Let us first discuss convergence conditions for the implementation of QA by the path-integral Monte Carlo method (PIMC) [16]. The basic idea of PIMC is to apply the Monte Carlo method to the classical system obtained from the original quantum system using the well-known quantum-classical mapping: The d-dimensional TFIM is mapped to a $(d+1)$-dimensional classical Ising system. In numerical simulations, the Suzuki-Trotter formula [17] is usually employed to express the partition function of the resulting classical system as

$$Z(t) \approx \sum_{\{S_i^{(k)}\}} \exp\left(\frac{\beta}{M} \sum_{k=1}^{M} H_{\rm pot}(\{S_i^{(k)}\}) + \gamma(t) \sum_{k=1}^{M} \sum_{i=0}^{N} S_i^{(k)} S_i^{(k+1)} \right), \tag{18}$$

$$H_{\rm pot}(\{S_i\}) = -\sum_{i=1}^{N} J_i S_i - \sum_{\langle ij \rangle} J_{ij} S_i S_j - \sum_{ijk} J_{ijk} S_i S_j S_k - \cdots, \tag{19}$$

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,

$$\gamma(t) = \frac{1}{2} \log \left(\coth \frac{\beta \Gamma(t)}{M} \right), \tag{20}$$

is ferromagnetic. This approximation (18) becomes exact in the limit $M \to \infty$ for a fixed $\beta = 1/k_B T$. The magnitude of this interaction (20) 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 (possibly 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 (18), 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 (18),

$$Z(t) = \sum_{x \in \mathcal{S}} \exp\left(-\frac{F_0(x)}{T_0} - \frac{F_1(x)}{T_1(t)} \right). \tag{21}$$

Here $F_0(x)$ is the cost function whose global minimum is the desired solution of the combinatorial optimization problem. We assume that states are discrete and the space $\mathcal{S}$ of states is finite. 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 (18), and the second term $-F_1(x)/T_1(t)$ generalizes the transverse-field term in (18).

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:

$$G(y, x; t) = \begin{cases} P(y, x) A(y, x; t) & (x \neq y) \\ 1 - \sum_{z \in \mathcal{S}} P(z, x) A(z, x; t) & (x = y) \end{cases}, \tag{22}$$

where $P(y, x)$ and $A(y, x; t)$ are called the *generation probability* and the *acceptance probability*, respectively. The matrix $G(t)$, whose (y, x) component is given by (22), $[G(t)]_{y,x} = G(y, x; t)$, is called the *transition matrix*.

The generation probability is the probability to generate the next candidate state y from the present state x. For example, that for single-spin flip process is defined by

$$P(x,y) = \begin{cases} \dfrac{1}{N} & \text{(single-spin flip)} \\ 0 & \text{(otherwise).} \end{cases} \tag{23}$$

We assume that the generation probability does not depend on time.

The acceptance probability $A(y,x;t)$ is the probability to accept the candidate y generated from state x. For the partition function (21), we define it as

$$A(y,x;t) = g\left(\frac{q(y;t)}{q(x;t)}\right), \tag{24}$$

$$q(x;t) = \frac{1}{Z(t)} \exp\left(-\frac{F_0(x)}{T_0} - \frac{F_1(x)}{T_1(t)}\right). \tag{25}$$

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 $0 \leq g(u) \leq 1$ and $g(1/u) = g(u)/u$ for $u \geq 0$. For instance, for the heat bath and the Metropolis methods, we have

$$g(u) = \frac{u}{1+u}, \tag{26}$$

$$g(u) = \min\{1, u\}, \tag{27}$$

respectively. The conditions mentioned above for $g(u)$ guarantee that $q(x;t)$ satisfies the detailed balance condition, $G(y,x;t)q(x;t) = G(x,y;t)q(y;t)$. Thus, $q(x;t)$ 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.

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

$$p(t,t_0) = G^{t,t_0}p_0 \equiv G(t-1)G(t-2)\cdots G(t_0)p_0. \tag{28}$$

Markov chains associated with QA are inhomogeneous, which means that the transition probability depends on time. Convergence of an inhomogeneous Markov chain is described in terms of ergodicity. 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:

$$\forall t_0 \geq 0 : \lim_{t\to\infty} \sup\{\|p(t,t_0) - p'(t,t_0)\| \,|\, p_0, p_0' \in \mathcal{P}\} = 0, \tag{29}$$

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

$$\|p\| = \sum_{x\in\mathcal{S}} |p(x)|. \tag{30}$$

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

$$\exists r \in \mathcal{P}, \forall t_0 \geq 0 : \lim_{t\to\infty} \sup\{\|p(t,t_0) - r\| \,|\, p_0 \in \mathcal{P}\} = 0. \tag{31}$$

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*

$$\sum_{i=0}^{\infty} \left(1 - \alpha(G^{t_{i+1},t_i})\right) \longrightarrow \infty, \tag{32}$$

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

$$\alpha(G^{t_{i+1},t_i}) = 1 - \min\left\{\sum_{z\in\mathcal{S}} \min\{G(z,x), G(z,y)\}\Big| x, y \in \mathcal{S}\right\} \tag{33}$$

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

The coefficient of ergodicity is a measure of the variety of the transition probability. For example, if $G(z,x)$ is independent of a state x, $\alpha(G)$ is equal to zero.

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*

$$\sum_{t=0}^{\infty} \|p_t - p_{t+1}\| < \infty. \tag{34}$$

Moreover, if $p = \lim_{t\to\infty} p_t$*, then* p *is equal to the probability distribution* r *in (31).*

3.2. Convergence theorem for QA-PIMC

A difference between QA by the PIMC and SA is the existence of $F_0(x)/T_0$ in (21): In SA the whole Hamiltonian is divided by a time-dependent temperature whereas in QA an additional time-independent term $F_0(x)/T_0$ exists. We obtain a sufficient condition for strong ergodicity of the system (21) by a similar discussion as in the Geman-Geman theorem [10].

Theorem 3 (Strong ergodicity of the system (21)) *The inhomogeneous Markov chain generated by (24) and (25) is strongly ergodic and converges to the equilibrium state corresponding to the first term of the right-hand side of (25),* $\exp(-F_0(x)/T_0)$*, if*

$$T_1(t) \geq \frac{RL_1}{\log(t+2)}, \tag{35}$$

where R *is the minimum number of maximum steps needed to reach any other state from an arbitrary state and* L_1 *stands for the maximum change of* $F_1(x)$ *in a single step.*

Application of this theorem to the PIMC implementation of QA represented by (18) 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 (18) is strongly ergodic and converges to the equilibrium state corresponding to the first term on the right-hand side of (18) if*

$$\Gamma(t) \geq \frac{M}{\beta} \tanh^{-1} \frac{1}{(t+2)^{2/RL_1}}. \tag{36}$$

For sufficiently large t, the above inequality reduces to

$$\Gamma(t) \geq \frac{M}{\beta}(t+2)^{-2/RL_1}. \tag{37}$$

This result implies that, in the asymptotic region, a power decay of the transverse field is sufficient to guarantee the convergence of QA by the PIMC. In most cases, the power $2/RL_1$ is inversely proportional to the system size. For example, R is equal to NM and L_1 is of order one in the single-spin flip process.

4. 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 [16]. The basic idea is to solve the imaginary-time Schrödinger equation by stochastic processes. It has been reported that the Schrödinger dynamics with imaginary time has an extra advantage that one can reach the optimal state more efficiently than by real-time dynamics for simple cases [18]. In the present section we describe sufficient conditions for strong ergodicity to hold in GFMC.

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

$$|\psi(t)\rangle = \mathrm{T}\exp\left(-\int_0^t \mathrm{d}t' H(t')\right)|\psi_0\rangle, \tag{38}$$

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

$$|\psi(t)\rangle = \lim_{n\to\infty} \hat{G}_0(t_{n-1})\hat{G}_0(t_{n-2})\cdots\hat{G}_0(t_1)\hat{G}_0(t_0)|\psi_0\rangle, \tag{39}$$

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 (39) by a stochastic method, we rewrite this equation in a recursive form,

$$\psi_{k+1}(y) = \sum_x \hat{G}_1(y, x; t_k)\psi_k(x), \tag{40}$$

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

$$\hat{G}_1(y, x; t) = \langle y|1 - \Delta t(H(t) - E_T)|x\rangle. \tag{41}$$

Equation (40) looks similar to a Markov process but is significantly different in several ways. An important difference is that $\hat{G}_1$ is not normalized, $\sum_y \hat{G}_1(y, x; t) \neq 1$. In order to avoid this problem, we decompose the Green's function into a normalized probability G_1 and a weight w:

$$\hat{G}_1(y, x; t) = G_1(y, x; t)w(x; t), \tag{42}$$

where

$$G_1(y,x;t) \equiv \frac{\hat{G}_1(y,x;t)}{\sum_y \hat{G}_1(y,x;t)}, \quad w(x;t) \equiv \frac{\hat{G}_1(y,x;t)}{G_1(y,x;t)}. \tag{43}$$

Thus, using (40), the wave function at time t is written as

$$\begin{aligned}\psi_n(y) = &\textstyle\sum_{\{x_k\}} \delta_{y,x_n} w(x_{n-1};t_{n-1}) w(x_{n-2};t_{n-2}) \cdots w(x_0;t_0) \\ &\times G_1(x_n,x_{n-1};t_{n-1}) G_1(x_{n-1},x_{n-2};t_{n-2}) \cdots G_1(x_1,x_0;t_0) \psi_0(x_0).\end{aligned} \tag{44}$$

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 (44), the wave function $\psi_n(y)$ is approximated by the distribution of walkers at the final step weighted by W_n,

$$\psi_n(y) = \lim_{M\to\infty} \frac{1}{M} \sum_{i=1}^{M} W_n^{(i)} \delta_{y,x_n^{(i)}}, \tag{45}$$

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 Green's function are immediately calculated as

$$\hat{G}_1(y,x;t) = \begin{cases} 1 - \Delta t(E_0(x) - E_T) & (x = y) \\ \Delta t\, \Gamma(t) & (x \text{ and } y \text{ differ by a single-spin flip}) \\ 0 & (\text{otherwise}), \end{cases} \tag{46}$$

where $E_0(x) = \langle x|H_{\text{pot}}|x\rangle$. 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

$$w(x;t) = 1 - \Delta t(E_0(x) - E_T) + N\Delta t\, \Gamma(t). \tag{47}$$

One can decompose this transition probability into the generation probability and the acceptance probability as in (22):

$$P(y,x) = \begin{cases} \dfrac{1}{N} & (\text{single-spin flip}) \\ 0 & (\text{otherwise}) \end{cases} \tag{48}$$

$$A(y,x;t) = \frac{N\Delta t\, \Gamma(t)}{1 - \Delta t(E_0(x) - E_T) + N\Delta t\, \Gamma(t)}. \tag{49}$$

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

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 [10].

Theorem 4 (Strong ergodicity of QA-GFMC) *The inhomogeneous Markov process of random walker for the QA-GFMC of TFIM, (22), (48) and (49), is strongly ergodic if*

$$\Gamma(t) \geq \frac{b}{(t+1)^c}, \qquad 0 < c \leq \frac{1}{N}. \tag{50}$$

This theorem asserts convergence of the distribution of random walkers to the equilibrium distribution

$$q(x;t) \equiv \frac{w(x;t)}{\sum_{x \in \mathcal{S}} w(x;t)} = \frac{1}{2^N} - \frac{\Delta t\, E_0(x)}{2^N \left\{1 + \Delta t\, E_T + N \Delta t\, \Gamma(t)\right\}}, \tag{51}$$

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 (44), 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 (47).

5. Summary

We have discussed convergence conditions of QA for TFIM under three types of dynamics. For QA following the Schrödinger equation, the asymptotic power-law decrease of the transverse field $\Gamma(t) \sim 1/t^c$ guarantees the adiabatic evolution during the annealing process. The same schedule also assures the strong ergodicity of the inhomogeneous Markov chains associated with QA by PIMC and GFMC. It is remarkable that the different types of dynamics share the same asymptotic condition for convergence. This power-law condition is faster than the inverse-log schedule of temperature found by Geman and Geman. Since the constant c depends on the system size as $1/N$, our result may not provide a practically useful guideline for large systems. We nevertheless emphasize that the system size N is kept finite in the present paper because our purpose is to study optimization problems in which the number of elements is always finite in contrast to genuine statistical-mechanical problems.

It is an interesting future problem to investigate the relation between the present result and the asymptotic decay rate of the residual energy for annealing within finite time [19, 20].

Acknowledgments

We thank S Suzuki, G E Santoro and G Ortiz for useful discussions and comments. This work was supported by the Grand-in-Aid for Scientific Research on the Priority Area "Deepening and Expansion of Statistical Mechanical Informatics" by the Ministry of Education, Culture, Sports, Science and Technology of Japan and by the CREST, JST.

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] Kadowaki T and Nishimori H 1998 *Phys. Rev.* E **58** 5355

[5] Finnila A B, Gomez M A, Sebenik C, Stenson C and Doll J P 1994 *Chem. Phys. Lett.* **219** 343
[6] Das A and Chakrabarti B K (eds) 2005 *Quantum Annealing and Related Optimization Methods (Lecture Notes in Physics 679)* (Berlin Heidelberg: Springer)
[7] Santoro G E and Tossati E 2006 *J. Phys. A: Math. Gen.* **39** R393
[8] Geman S and Geman D 1984 *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-6** 721
[9] Morita S and Nishimori H 2007 *J. Phys. Soc. Jpn.* **76** 064002
[10] Morita S and Nishimori H 2006 *J. Phys. A: Math. Gen.* **39** 13903
[11] Somma R D, Batista C D and Ortiz G 2007 *Phys. Rev. Lett.* **99** 030603
[12] Farhi E, Goldstone J, Gutmann S and Sipser M 2000 quant-ph/0001106
[13] Messiah A 1976 *Quantum Mechanics* (New York: Wiley)
[14] Hopf E 1963 *J. Math. Mech.* **12** 683
[15] Suzuki S, Nishimori H and Suzuki M 2007 *Phys. Rev.* E **75** 051112
[16] Landau D P and Binder K 2000 *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge: Cambridge University Press) ch 8
[17] Suzuki M 1976 *Prog. Theor. Phys.* **56** 1454
[18] Stella L, Santoro G E and Tosatti E 2005 *Phys. Rev.* B **72** 014303
[19] Suzuki S and Okada M 2005 *J. Phys. Soc. Jpn.* **74** 1649
[20] Morita S 2007 *J. Phys. Soc. Jpn.* **76** 104001