Journal of Statistical Physics, Vol. 58, Nos. 1/2, 1990

Simulated Annealing and Quantum Detailed Balance

Alberto Frigerio[1]

Received April 4, 1989

The analogue of simulated annealing is considered for time-inhomogeneous evolutions of a von Neumann algebra of operators, whose instantaneous generator at each time t satisfies the quantum detailed balance condition with respect to a faithful normal state which depends on time through a suitable cooling schedule. Convergence to the (nonfaithful) limiting state is proved under the usual kinds of assumptions. The approach is interesting in view of possible applications to stochastic Ising models and to Boltzmann machines.

KEY WORDS: Simulated annealing; quantum detailed balance; time-inhomogeneous evolutions; Boltzmann machines.

1. INTRODUCTION

Simulated annealing is a stochastic algorithm to drive a fictitious physical system to a state of minimum energy; it has been applied as a general-purpose method to various kinds of global optimization problems.[5,11,20,23,29] If the system is in state x at time t, a new state y is chosen in the state space X with probability $q_0(x, y) = q_0(y, x)$. The energy difference $U(y) - U(x)$ is computed, and the system actually moves from state x to state y with probability (conditional upon y being chosen) 1 if $U(y) - U(x) \leqslant 0$, and with (conditional) probability $\exp\{-\beta(t)[U(y) - U(x)]\}$ if $U(y) - U(x) > 0$. If $\beta(t)$ were constant $= \beta$, iteration of this procedure would drive the system to its Gibbs distribution corresponding to the energy function U at inverse temperature β. By letting $\beta(t) \to \infty$ as $t \to \infty$ one hopes that the system will be eventually driven to a state of minimum energy.

If one were to set $\beta = +\infty$ from the beginning, then only energy-decreasing transitions would be allowed, and the system could end up in a

[1] Dipartimento di Matematica e Informatica, Universita' di Udine, 33100 Udine, Italy.

0022-4715/90/0100-0325$06.00/0

local minimum of U which is not a global minimum. One may also intuitively think that the probability of escaping (in a single attempt) from a local minimum of (suitably defined) depth d when the inverse temperature is β is of the order of $\exp(-\beta d)$, and results of this kind can be actually proved.[17,19] Then one must not reduce the temperature $1/\beta(t)$ too quickly (the system must be annealed, not quenched). In several papers[10,11,13,17,19] it has been shown that a cooling schedule of the form $\beta(t) = (1/c)\log(t+1)$, where c is a suitable constant (which may depend on the paper) is sufficient to obtain approach to the set of states of minimum energy; in ref. 17 a necessary and sufficient condition has been given.

The present paper grew out of an attempt to understand the paper by Holley and Stroock.[19] There the proofs of convergence of the annealing algorithm are based entirely on consideration of the Dirichlet forms associated with the generators L_β $(\beta > 0)$ of the Markov semigroups determined by the transition rates

$$q_\beta(x, y) = q_0(x, y)\exp\{-\beta[U(y) - U(x)]_+\} \qquad (y \neq x) \tag{1.1}$$

In this attempt it was discovered that those proofs (at least those based on estimates in L^2-norm) carry over to the situation where the algebra of functions on a finite state space X is replaced by a von Neumann algebra $\mathcal{M}$ of operators on a separable Hilbert space $\mathcal{H}$ with a cyclic and separating vector ϕ; the Gibbs distribution at inverse temperature β is replaced by the state μ_β on $\mathcal{M}$ determined by $\mu_\beta(A) = \langle v_\beta, Av_\beta\rangle$, where v_β is given by

$$v_\beta = \{\langle \phi, \exp(-\beta H)\,\phi\rangle\}^{-1/2} J\exp(-\beta H/2)\,\phi, \qquad H = H^* \in \mathcal{M}$$

J being the modular involution associated with the pair $(\mathcal{M}, \phi)$; and, finally, L_β $(\beta > 0)$ becomes the generator of a (quantum) dynamical semigroup[16,22] on $\mathcal{M}$ satisfying the (quantum) detailed balance condition[1,21,25] with respect to μ_β. This allows us to prove some results on approach to equilibrium for time-inhomogeneous irreversible quantum evolutions, thus extending the results of refs. 7, 8, 27, and 28 for quantum dynamical semigroups.

It may be noted that those proofs are quite similar to the ones existing in the literature concerning the *Langevin algorithm*,[2,6,12,14] a variant of simulated annealing involving diffusion in $\mathbb{R}^d$ with decreasing temperature.

The potential interest of having a version of simulated annealing for quantum dynamical evolutions is the following. What is now called a *Boltzmann machine*[18] is essentially a (finite) Ising model with Hamiltonian of the form

$$H = -\tfrac{1}{2}\sum_{i,j} w_{ij}\sigma_i\sigma_j + \sum_j \theta_j\sigma_j \tag{1.2}$$

undergoing a sort of time-dependent Glauber dynamics.[15] Ising models and their Glauber dynamics have been often imbedded into quantum systems[4,26] evolving under quantum dynamical semigroups,[24] and it is hoped that the viewpoint of time-inhomogeneous quantum irreversible evolutions may prove useful in this connection. It should be noted that, in spite of the similarity of the name, this approach has nothing to do with "quantum annealing,"[3] where the quantum tunnel effect is used in place of thermal fluctuations to allow the state of a system to escape from local minima.

The paper by Holley and Stroock[19] deals with a continuous-time version of simulated annealing, with a cooling schedule $t \mapsto \beta(t)$ given by a differentiable function; the same is true for the papers on the Langevin algorithm.[2,6,12,14] I have found it more convenient to consider a piecewise constant function $\beta(t)$ with jumps. As a bonus, the method becomes applicable to simulated annealing in discrete time, which is generally used in applications. Potentially more important than the proof of approach to a state of minimum energy is then an upper bound on the probability of being away from the minimum after n iterations.

The paper is organized as follows. Some preliminary notions are recalled in Section 2, where also the necessary notations are developed. The general theory is described in Section 3 in abstract form. In Section 4 I apply this theory to the classical case, both in continuous and in discrete time; one of the proofs, being technical, is deferred to an Appendix. I describe the quantum case in Section 5; after a brief general discussion, I concentrate on a special class of time-inhomogeneous evolutions of a finite-dimensional matrix algebra, for which a complete parallel of the discussion of the classical case can be given.

Some partial results on simulated annealing for time-inhomogeneous quantum evolutions have been already described in ref. 9. Applications to Ising-type models (or, more ambitiously, to Boltzmann machines) will be given in future publications.

2. NOTATION AND PRELIMINARIES

Let $\mathscr{M}$ be a von Neumann algebra of operators on a separable Hilbert space $\mathscr{K}$, with a cyclic and separating vector ϕ. Denote by μ the faithful normal state on $\mathscr{M}$ defined by $\mu(A) := \langle \phi, A\phi \rangle$, $A \in \mathscr{M}$, by $\{\sigma_t : t \in \mathbb{R}\}$ the associated modular automorphism group, and by J the modular involution on $\mathscr{K}$ associated with the pair $(\mathscr{M}, \phi)$. For the sake of notational clarity in applications, the norm of $\mathscr{M}$ will be denoted by $\|\cdot\|$, whereas the norm of $\mathscr{K}$ will be denoted by $\|\cdot\|_2$. Although $\mathscr{M}$ will be finite-dimensional and μ

will be a tracial state in most applications, the proofs in this and the following section do not depend on these assumptions.

Let H be a self-adjoint element of $\mathcal{M}$. In order to avoid trivialities, it will be assumed that H is not a multiple of the identity. Define a family $\{\mu_\beta : \beta \in [0, \infty)\}$ of faithful normal states on $\mathcal{M}$ as follows. Let $\{V(\beta) : \beta \in [0, \infty)\}$ be the semigroup of self-adjoint bounded operators on $\mathcal{K}$ defined by

$$V(\beta) := J \exp(-\beta H) J, \qquad \beta \in [0, \infty) \tag{2.1}$$

Note that $V(\beta)$ belongs to the commutant $\mathcal{M}'$ of $\mathcal{M}$. Let also

$$Z(\beta) := \mu(\exp(-\beta H)) = \langle \phi, \exp(-\beta H)\, \phi \rangle = \langle \phi, V(\beta)\, \phi \rangle \tag{2.2}$$

for all β in $[0, \infty)$. Then define μ_β on $\mathcal{M}$ by

$$\begin{aligned} \mu_\beta(A) &:= Z(\beta)^{-1} \langle V(\beta)\, \phi, A\phi \rangle \\ &= Z(\beta)^{-1} \langle V(\beta/2)\, \phi, V(\beta/2)\, A\phi \rangle \\ &= Z(\beta)^{-1} \langle V(\beta/2)\, \phi, A V(\beta/2)\, \phi \rangle, \qquad A \in \mathcal{M} \end{aligned} \tag{2.3}$$

In the special case that H is invariant under σ_t, μ_β reduces to

$$\mu_\beta(A) = \mu(A \exp(-\beta H))/\mu(\exp(-\beta H)) \tag{2.4}$$

Note that $\mu_0 = \mu$.

The GNS representation of $\mathcal{M}$ associated with μ_β can be identified with the identity representation of $\mathcal{M}$ acting on $\mathcal{K}$, with cyclic and separating vector

$$v_\beta := Z(\beta)^{-1/2}\, V(\beta/2)\, \phi \tag{2.5}$$

A normal state μ' on $\mathcal{M}$ is majorized by a scalar multiple of μ_β if and only if it can be expressed in the form

$$\mu'(A) = \langle JBJv_\beta, Av_\beta \rangle, \qquad A \in \mathcal{M} \tag{2.6}$$

where B is a positive element of $\mathcal{M}$ satisfying

$$\langle JBJv_\beta, v_\beta \rangle = Z(\beta)^{-1} \langle \exp(-\beta H/2)\, \phi, B \exp(-\beta H/2)\, \phi \rangle = 1 \tag{2.7}$$

By adding to H a suitable constant, which drops out in the expression of μ_β, we may (and do) assume, without loss of generality, that H is nonnegative and that the infimum $\inf \sigma(H)$ of its spectrum $\sigma(H)$ is 0. Then, upon letting $h := \|H\|$, we have

$$\{0, h\} \subseteq \sigma(H) \subseteq [0, h] \tag{2.8}$$

We shall also assume that the limit

$$\mu_\infty(A) := \lim_{\beta\to\infty} \mu_\beta(A) \tag{2.9}$$

exists for all A in $\mathcal{M}$, so that it defines a state μ_∞ on $\mathcal{M}$, not faithful (in general), and not necessarily normal. A special case in which the limit (2.6) exists and defines a nonfaithful, normal state is the situation in which $0 = \inf \sigma(H)$ is in the point spectrum $\sigma_p(H)$ of H; then we have

$$\mu_\infty(A) = \langle P_0\phi, AP_0\phi\rangle / \langle P_0\phi, P_0\phi\rangle \tag{2.10}$$

where P_0 is the projection onto the eigenspace of H corresponding to the eigenvalue 0.

All the considerations of the present section could be extended to the case where H is a nonnegative, self-adjoint, unbounded operator affiliated with $\mathcal{M}$; however, we shall need the boundedness of H in the next section.

For each β in $[0, \infty)$, let L_β be the generator of a uniformly continuous semigroup $\{T_t^\beta : t \in \mathbb{R}^+\}$ of completely positive, identity-preserving normal linear maps of $\mathcal{M}$ into itself (a *dynamical semigroup* on $\mathcal{M}$). Assume that the *detailed balance* condition[1,21,25]

$$\mu_\beta(AL_\beta(B)) = \mu_\beta(L_\beta(A)\, B), \qquad A, B \in \mathcal{M} \tag{2.11}$$

holds. Then we have also

$$\mu_\beta(AT_t^\beta(B)) = \mu_\beta(T_t^\beta(A)\, B), \qquad A, B \in \mathcal{M}; \quad t \in \mathbb{R}^+ \tag{2.12}$$

The above equality is equivalent to

$$\langle V(\beta/2)\, A^*\phi, V(\beta/2)\, T_t^\beta(B)\, \phi\rangle = \langle V(\beta/2)\, T_t^\beta(A^*)\, \phi, V(\beta/2)\, B\phi\rangle \tag{2.13}$$

In particular, μ_β is invariant under T_t^β.

We define the operator S_t^β on $\mathcal{H}$ by

$$S_t^\beta[V(\beta/2)\, A\phi] := V(\beta/2)\, T_t^\beta(A)\, \phi, \qquad A \in \mathcal{M} \tag{2.14}$$

which extends to a contraction on $\mathcal{H}$, since $\overline{V(\beta/2)\, \mathcal{M}\phi} = \mathcal{H}$ and

$$\begin{aligned} &\|S_t^\beta[V(\beta/2)\, A\phi]\|_2^2 \\ &\quad = \langle V(\beta/2)\, T_t^\beta(A)\, \phi, V(\beta/2)\, T_t^\beta(A)\, \phi\rangle \\ &\quad = Z(\beta)\, \mu_\beta(T_t^\beta(A^*)\, T_t^\beta(A)) \\ &\quad \leqslant Z(\beta)\, \mu_\beta(T_t^\beta(A^*A)) \\ &\quad = Z(\beta)\, \mu_\beta(A^*A) \\ &\quad = \|V(\beta/2)\, A\phi\|_2^2 \end{aligned} \tag{2.15}$$

where we have used the Kadison–Schwarz inequality for the completely positive map T_t^β and the invariance of μ_β under T_t^β.

Equation (2.13) tells us that S_t^β is self-adjoint in $\mathscr{K}$. It is clear from the definition (2.14) that $\{S_t^\beta : t \in \mathbb{R}^+\}$ is a semigroup, and also its strong continuity in t is easily proved. Then there exists a nonnegative self-adjoint operator G_β in $\mathscr{K}$ such that

$$S_t^\beta = \exp(-G_\beta t), \qquad t \in \mathbb{R}^+ \tag{2.16}$$

The dense subset $V(\beta/2)\,\mathscr{M}\phi$ of $\mathscr{K}$ is a core for G_β, and we have

$$G_\beta V(\beta/2)\, A\phi = -V(\beta/2)\, L_\beta(A)\, \phi, \qquad A \in \mathscr{M} \tag{2.17}$$

The Dirichlet form $E_\beta(\cdot,\cdot)$ associated with G_β is given by

$$\begin{aligned} E_\beta(A, B) &= -Z(\beta)^{-1} \langle V(\beta/2)\, A\phi, V(\beta/2)\, L_\beta(B)\, \phi \rangle \\ &= -\mu_\beta(A^* L_\beta(B)) \\ &= \tfrac{1}{2}\mu_\beta(D_\beta(A, B)), \qquad A, B \in \mathscr{M} \end{aligned} \tag{2.18}$$

where $D_\beta(\cdot,\cdot)$ is the dissipation function[22] of L_β, defined by

$$D_\beta(A, B) := L_\beta(A^*B) - L_\beta(A^*)\, B - A^* L_\beta(B) \tag{2.19}$$

Let us make the following *spectral gap* assumption.[19] Let

$$\Gamma(\beta) := \inf\{E_\beta(A, A) : A \in \mathscr{M},\ \mu_\beta(A) = 0,\ \mu_\beta(A^*A) = 1\} \tag{2.20}$$

so that

$$E_\beta(A, A)) \geqslant \Gamma(\beta)\, \mu_\beta([A - \mu_\beta(A)]^* [A - \mu_\beta(A)]), \qquad A \in \mathscr{M} \tag{2.21}$$

It will be assumed that $\Gamma(\beta)$ *is strictly positive.* This implies that the eigenspace of G_β corresponding to the eigenvalue 0 is one-dimensional [the multiples of $V(\beta/2)\,\phi$], and that the rest of the spectrum of G_β is contained in $[\Gamma(\beta), +\infty)$.

It is often the case in applications that L_β has a limit L_∞ as $\beta \to \infty$, and that L_∞ is the generator of a dynamical semigroup on $\mathscr{M}$, admitting μ_∞ as an invariant state; however, in general no spectral gap condition will hold for L_∞ and the invariant state for L_∞ will not be unique, in general.

3. SIMULATED ANNEALING

In the present context, *simulated annealing* is the following procedure. Let $\{\beta_k : k = 1, 2,...\}$ be an increasing sequence of positive numbers (inverse temperatures), diverging to $+\infty$, let $\{t_k : k = 1, 2,...\}$ be a sequence of positive numbers (times), and consider, for all positive integers n, the expression

$$\langle JBJv_{\beta_1}, T_{t_1}^{\beta_1} T_{t_2}^{\beta_2} \cdots T_{t_n}^{\beta_n}(A)\, v_{\beta_1} \rangle, \qquad A \in \mathcal{M} \tag{3.1}$$

where B is a positive element of $\mathcal{M}$ such that

$$\langle JBJv_{\beta_1}, v_{\beta_1} \rangle = 1 \tag{3.2}$$

(3.1) is the expectation value of the observable $A \in \mathcal{M}$ in the state evolved from the initial state $\langle JBJv_{\beta_1}, (\cdot)\, v_{\beta_1} \rangle$ under the time-inhomogeneous evolution determined by L_{β_k} in the time interval

$$(s_{k-1}, s_k], \text{ where } s_k := \sum_{j=1}^{k} t_j, \text{ for } k = 1, 2,..., n.$$

The question is whether the limit of (3.1) as $n \to \infty$ exists and equals $\mu_\infty(A)$.

The problem will be rephrased in the Hilbert space $\mathcal{K}$, introducing some shorthand notation. For each positive integer k, define self-adjoint operators S_k, $R_{k,k+1}$ on $\mathcal{K}$ by

$$S_k := S_{t_k}^{\beta_k} \tag{3.3}$$

$$R_{k,k+1} := Z(\beta_k)^{-1/2}\, Z(\beta_{k+1})^{1/2}\, V((\beta_{k+1} - \beta_k)/2)^{-1} \tag{3.4}$$

and vectors v_k, u_k in $\mathcal{K}$ by

$$v_k := Z(\beta_k)^{-1/2}\, V(\beta_k/2)\, \phi = v_{\beta_k} \tag{3.5}$$

$$u_k := S_k R_{k-1,k} \cdots S_2 R_{1,2} S_1 JBJv_1 \tag{3.6}$$

$$(u_1 := S_1 JBJv_1)$$

Lemma 3.1. With the above notation, (3.1) can be rewritten as

$$\langle S_n R_{n-1,n} \cdots S_2 R_{1,2} S_1 JBJv_1, Av_n \rangle = \langle u_n, Av_n \rangle \tag{3.7}$$

Proof. For $j = 1,..., n-1$ let

$$A_j := T_{t_{j+1}}^{\beta_{j+1}} T_{t_{j+2}}^{\beta_{j+2}} \cdots T_{t_n}^{\beta_n}(A) \in \mathcal{M} \tag{3.8}$$

Then

$$
\begin{aligned}
(3.1) &= \langle JBJv_{\beta_1}, T_{t_1}^{\beta_1}(A_1)\, v_{\beta_1}\rangle \qquad [\text{by } (3.8)] \\
&= Z(\beta_1)^{-1/2} \langle JBJv_1, T_{t_1}^{\beta_1}(A_1)\, V(\beta_1/2)\, \phi\rangle \qquad [\text{by } (2.5)] \\
&= Z(\beta_1)^{-1/2} \langle JBJv_1, V(\beta_1/2)\, T_{t_1}^{\beta_1}(A_1)\, \phi\rangle \\
&\qquad [\text{since } V(\beta_1/2) \in \mathcal{M}'] \\
&= Z(\beta_1)^{-1/2} \langle JBJv_1, S_{t_1}^{\beta_1} V(\beta_1/2)\, A_1 \phi\rangle \qquad [\text{by } (2.13)] \\
&= Z(\beta_1)^{-1/2} \langle V(\beta_1/2)\, S_1 JBJv_1, A_1\phi\rangle \qquad [\text{by } (3.3)] \\
&= Z(\beta_2)^{-1/2} \langle V(\beta_2/2)\, R_{1,2} S_1 JBJv_1, A_1\phi\rangle \qquad [\text{by } (3.4)] \\
&= Z(\beta_2)^{-1/2} \langle R_{1,2} S_1 JBJv_1, V(\beta_2/2)\, T_{t_2}^{\beta_2}(A_2)\, \phi\rangle
\end{aligned}
$$

[by (3.5) and (3.8)]. Now we are ready to apply (2.13) once again. The argument can be repeated as many times as needed, to be concluded as follows:

$$
\begin{aligned}
(3.1) &= Z(\beta_n)^{-1/2} \langle R_{n-1,n} \cdots S_2 R_{1,2} S_1 JBJv_1, V(\beta_n/2)\, T_{t_n}^{\beta_n}(A)\, \phi\rangle \\
&= Z(\beta_n)^{-1/2} \langle R_{n-1,n} \cdots S_2 R_{1,2} S_1 JBJv_1, S_{t_n}^{\beta_n} V(\beta_n/2)\, A\phi\rangle \\
&= Z(\beta_n)^{-1/2} \langle S_n R_{n-1,n} \cdots S_2 R_{1,2} S_1 JBJv_1, V(\beta_n/2)\, A\phi\rangle \\
&= \langle S_n R_{n-1,n} \cdots S_2 R_{1,2} S_1 JBJv_1, Av_n\rangle
\end{aligned}
$$

as claimed. ∎

It is clear that we have to control the differences

$$
\begin{aligned}
&|\langle u_n, Av_n\rangle - \langle v_n, Av_n\rangle| \\
&\quad \leqslant \|u_n - v_n\|_2\, \|Av_n\|_2 \\
&\quad = \|u_n - v_n\|_2\, \mu_{\beta_n}(A^*A)^{1/2}, \qquad A \in \mathcal{M} \qquad (3.9)
\end{aligned}
$$

Lemma 3.2. For all $n = 1, 2,\ldots$ we have

$$\langle u_n, v_n\rangle = 1 \qquad (3.10)$$

and for all $n = 2, 3,\ldots$ we have also

$$\langle R_{n-1,n} u_{n-1}, v_n\rangle = 1 \qquad (3.11)$$

Proof. For $n=1$ we have $\langle u_1, v_1\rangle = \langle S_1 JBJv_1, v_1\rangle = 1$, since S_1 is self-adjoint and leaves v_1 invariant, and (3.2) holds. For $n>1$ we have $u_n = S_n R_{n-1,n} u_{n-1}$, so that

$$\langle u_n, v_n\rangle = \langle u_{n-1}, R_{n-1,n} S_n v_n\rangle = \langle u_{n-1}, R_{n-1,n} v_n\rangle = \langle u_{n-1}, v_{n-1}\rangle$$

Then (3.10) follows by induction. We have also

$$\langle R_{n-1,n} u_{n-1}, v_n\rangle = \langle u_{n-1}, R_{n-1,n} v_n\rangle = \langle u_{n-1}, v_{n-1}\rangle = 1 \quad \blacksquare$$

Lemma 3.3. For all $n=2, 3, \ldots$ we have

$$\|u_n - v_n\|_2 \leqslant \exp[-\Gamma(\beta_n)\, t_n](\|R_{n-1,n}\| \cdot \|u_{n-1} - v_{n-1}\|_2 + b_{n-1}) \tag{3.12}$$

where

$$\begin{aligned} b_{n-1} &:= \|(R^2_{n-1,n} - 1)\, v_n\| \\ &= \{\|R_{n-1,n} v_{n-1}\|_2^2 - 1\}^{1/2} \\ &= \{Z(2\beta_{n-1} - \beta_n)\, Z(\beta_n)\, Z(\beta_{n-1})^{-2} - 1\}^{1/2} \end{aligned} \tag{3.13}$$

We have also

$$\|u_1 - v_1\|_2 \leqslant \exp[-\Gamma(\beta_1)\, t_1]\, \|JBJv_1 - v_1\|_2 \tag{3.14}$$

Proof. We have

$$u_n - v_n = S_n R_{n-1,n} u_{n-1} - v_n = S_n(R_{n-1,n} u_{n-1} - v_n)$$

By Lemma 3.2, $R_{n-1,n} u_{n-1} - v_n$ is orthogonal to v_n. Then, by the spectral gap assumption, we get

$$\|u_n - v_n\|_2 \leqslant \exp[-\Gamma(\beta_n)\, t_n]\, \|R_{n-1,n} u_{n-1} - v_n\|_2$$

In turn, we have

$$\begin{aligned} &\|R_{n-1,n} u_{n-1} - v_n\|_2 \\ &\quad \leqslant \|R_{n-1,n}(u_{n-1} - v_{n-1})\|_2 + \|R_{n-1,n} v_{n-1} - v_n\|_2 \\ &\quad \leqslant \|R_{n-1,n}\| \cdot \|u_{n-1} - v_{n-1}\|_2 + b_{n-1} \end{aligned}$$

where

$$b_{n-1} = \|R_{n-1,n} v_{n-1} - v_n\|_2 = \|(R^2_{n-1,n} - 1)\, v_n\|_2$$

It remains to prove that b_{n-1} can be rewritten in the alternative forms appearing in (3.13). This can be obtained by a straightforward computation of b_{n-1}^2, recalling that $\langle v_n, v_n \rangle = 1$ for all n and using the explicit expression of $R_{n-1,n}$.

Finally, the proof of (3.14) is immediate from the observations that $u_1 = S_1 JBJv_1$, $v_1 = S_1 v_1$, and $\langle (JBJv_1 - v_1), v_1 \rangle = 0$. ▮

Let us now introduce some additional notation (we recall that $h := \|H\|$). Let

$$a_1 := \Gamma(\beta_1)\, t_1 \tag{3.15}$$

$$a_k := \Gamma(\beta_k)\, t_k - (\beta_k - \beta_{k-1})\, h/2, \qquad k = 2, 3,... \tag{3.16}$$

$$b_0 := \|JBJv_1 - v_1\|_2 \tag{3.17}$$

$$b_r := \{Z(2\beta_r - \beta_{r+1})\, Z(\beta_{r+1})\, Z(\beta_r)^{-2} - 1\}^{1/2}, \qquad r = 1, 2,...$$

[the same as (3.13)]

$$d_n := \|u_n - v_n\|_2, \qquad n = 1, 2,... \tag{3.18}$$

Lemma 3.4. With the above notation, we have, for all $n = 1, 2,...,$

$$d_n \leqslant \sum_{r=0}^{n-1} \exp\left(-\sum_{k=r+1}^{n} a_k\right) b_r \tag{3.19}$$

Proof. Since H is assumed to be nonnegative, $Z(\beta)$ is a nonincreasing function of β, so that $Z(\beta_{n-1})^{-1/2} Z(\beta_n)^{1/2} < 1$ and

$$\|R_{n-1,n}\| \leqslant \|\exp[(\beta_n - \beta_{n-1})\, H/2]\| = \exp[(\beta_n - \beta_{n-1})\, h/2]$$

Hence (3.12) of Lemma 3.3 can be rewritten as

$$d_n \leqslant \exp(-a_n)\, d_{n-1} + \exp[-\Gamma(\beta_n)\, t_n]\, b_{n-1}$$

which implies also

$$d_n \leqslant \exp(-a_n)(d_{n-1} + b_{n-1}) \tag{3.20}$$

since $a_n \leqslant \Gamma(\beta_n)\, t_n$.

For $n = 1$, (3.19) holds, since it reduces to (3.14) of Lemma 3.3. Using (3.20), we can easily prove (3.19) by induction. ▮

Theorem 3.5. Suppose that there exists a constant C, $0 < C < +\infty$, such that

$$\sum_{r=0}^{n-1} \exp\left(-\sum_{k=r+1}^{n} a_k\right) b_r \leqslant C \qquad \text{for } n \text{ sufficiently large} \tag{3.21}$$

Let $A \in \mathcal{M}$ be such that $\mu_\infty(A^*A) = 0$. Then

$$\lim_{n \to \infty} \langle JBJv_{\beta_1}, T_{t_1}^{\beta_1} T_{t_2}^{\beta_2} \cdots T_{t_n}^{\beta_n}(A)\, v_{\beta_1} \rangle = 0 \tag{3.22}$$

Proof. By Lemma 3.1, Lemma 3.4, (3.9), and (3.21), we have

$$|\langle JBJv_{\beta_1}, T_{t_1}^{\beta_1} T_{t_2}^{\beta_2} \cdots T_{t_n}^{\beta_n}(A)\, v_{\beta_1} \rangle - \mu_{\beta_n}(A)| \leqslant C\mu_{\beta_n}(A^*A)^{1/2} \tag{3.23}$$

By assumption, $\mu_{\beta_n}(A^*A)$ [hence also $\mu_{\beta_n}(A)$] converges to 0 as $n \to \infty$. The result follows. ∎

The above theorem corresponds to a result of Holley and Stroock,[19] implying that, under assumption (3.21), the probability that the time-evolved state has energy strictly above the global minimum vanishes in the limit as $t \to \infty$.

It is important that C can be chosen as small as possible. In this respect, note that the b_r for $r \geqslant 1$ are small if the inverse temperature is increased slowly, whereas b_0 depends on the initial state and not on the cooling schedule, and can be large. For example, we shall see in the following section that in the classical case, where one tries to minimize the energy function on a finite state space with N points, b_0 is of the order of $N^{1/2}$. Hence, in order to have the possibility of choosing C small, we must have[10,13,14,17]

$$\sum_{k=1}^{n} a_k \to +\infty \qquad \text{as } n \to +\infty \tag{3.24}$$

Under some stronger assumptions on the sequences $\{a_k\}$ and $\{b_r\}$, one can show that the annealing algorithm converges to the uniform distribution on the set of global minima of H. As an illustration of the kind of results that can be obtained, one can prove the following.

Theorem 3.6. Assume that

$$a_k \geqslant a > 0 \qquad \text{for all } k \text{ sufficiently large} \tag{3.25}$$

$$b_r \to 0 \qquad \text{as } r \to \infty \tag{3.26}$$

Then

$$\lim_{n \to \infty} \langle JBJv_{\beta_1}, T_{t_1}^{\beta_1} T_{t_2}^{\beta_2} \cdots T_{t_n}^{\beta_n}(A)\, v_{\beta_1} \rangle = \mu_\infty(A) \tag{3.27}$$

for all $A \in \mathcal{M}$ and for all B such that $\langle JBJv_{\beta_1}, v_{\beta_1} \rangle = 1$.

Proof. By (3.9) and (3.18), it suffices to prove that $d_n \to 0$ as $n \to \infty$. Indeed, let

$$r_0 := \max\{k: a_k < a\}$$

$$A_0 := \max\left\{\sum_{k=r+1}^{r_0} (a - a_k): r < r_0, 0\right\}$$

Then, for all $n = 1, 2,\ldots,$ and for all $r = 0,\ldots, n-1$ we have

$$\exp\left(-\sum_{k=r+1}^{n} a_k\right) \leqslant \exp(A_0) \exp[-(n-r)\,a]$$

and, by Lemma 3.4,

$$d_n \leqslant \exp(A_0) \sum_{r=0}^{n-1} \exp[-(n-r)\,a]\, b_r$$

Since $b_r \to 0$ as $n \to \infty$, for each positive δ there exists n_δ such that $(0 \leqslant)\, b_r \leqslant \delta$ for $n > n_\delta$. Then, for $n > n_\delta$ we have

$$\begin{aligned}
d_n &\leqslant \exp(A_0) \left\{\sum_{r=0}^{n_\delta - 1} \exp[-(n-r)\,a]\, b_r \right. \\
&\quad \left. + \delta \sum_{r=n_\delta}^{n-1} \exp[-(n-r)\,a]\right\} \\
&= \exp(A_0) \left\{\exp[-(n-n_\delta)\,a] \sum_{r=0}^{n_\delta - 1} \exp[-(n_\delta - r)\,a]\, b_r \right. \\
&\quad \left. + \delta \sum_{k=1}^{n-n_\delta} \exp(-ka)\right\} \\
&< \exp(A_0) \left\{\exp[-(n-n_\delta)\,a] \sum_{r=0}^{n_\delta - 1} \exp[-(n_\delta - r)\,a]\, b_r \right. \\
&\quad \left. + \delta/[1 - \exp(-a)]\right\}
\end{aligned} \tag{3.28}$$

In the limit as $n \to \infty$ for fixed n_δ, (3.28) tends to $\delta \exp(A_0)/[1 - \exp(-a)]$. Since δ was arbitrary, the result follows. ∎

Remark. It seems that the above approach can be adapted to a more general situation in which H is not fixed, but changes at each step (say, $v_k := \exp(-\beta_k H_k/2)\,\phi/\langle\phi, \exp(-\beta_k H_k)\,\phi\rangle$, $\{H_k\}$ being a sequence of self-adjoint elements of $\mathscr{M}$), provided suitable assumptions hold. This should provide a technique for proving convergence of "estimation and annealing" algorithms, as in ref. 31. This problem, which is of interest for applications to Boltzmann machines, is currently under investigation.

4. APPLICATIONS: THE CLASSICAL CASE

Let $\mathcal{M}$ be the algebra of all complex-valued functions defined on a set X having a finite number N of points, equipped with the supremum norm $\|f\| := \sup_{x\in X} |f(x)|$, $f \in \mathcal{M}$. Let μ be the normalized counting measure on X, $\mu(x) := 1/N$ for all x in X, and let $\mathcal{H}$ be the Hilbert space $L^2(X, d\mu)$. $\mathcal{H}$ coincides with $\mathcal{M}$ as a set, and its norm $\|\cdot\|_2$ is given by

$$\|f\|_2^2 = \langle f, f\rangle = \int_X |f|^2 \, d\mu = \frac{1}{N} \sum_{x\in X} |f(x)|^2, \qquad f \in \mathcal{H}$$

We denote again by μ the faithful state on $\mathcal{M}$ defined by

$$\mu(f) := \int_X f \, d\mu = \frac{1}{N} \sum_{x\in X} f(x) = \langle \phi, f\phi\rangle, \qquad f \in \mathcal{M}$$

where $\phi(x) = 1$ for all x. The modular automorphism group is trivial and the modular involution is just complex conjugation.

Let $U: X \to \mathbb{R}$ be a function, and let $H \in \mathcal{M}$ be defined as $U - \min U$, so that $h := \|H\| = \max U - \min U$. For all β in $\mathbb{R}^+$, we have, from Section 2,

$$[V(\beta) f](x) = f(x) \exp[-\beta H(x)/2] \tag{4.1a}$$

$$Z(\beta) = \frac{1}{N} \sum_{x\in X} \exp[-\beta H(x)] \tag{4.1b}$$

$$\mu_\beta(f) = \frac{1}{NZ(\beta)} \sum_{x\in X} f(x) \exp[-\beta H(x)] \tag{4.1c}$$

$$v_\beta(x) = Z(\beta)^{-1/2} \exp[-\beta H(x)/2] \tag{4.1d}$$

$$\mu_\infty(f) = \frac{1}{NZ(\infty)} \sum_{x\in S^0} f(x) \tag{4.1e}$$

where

$$S^0 := \{x \in X: U(x) = \min U\} \tag{4.2}$$

is the set of global minima of U, and where $NZ(\infty) := \lim_{\beta\to\infty} NZ(\beta)$ is the cardinality of S^0.

As initial state, one usually considers a situation in which the system is in some arbitrary configuration x_0 with probability 1; this can be described by

$$\mu'(f) := f(x_0) = \langle JBJv_1, fv_1 \rangle, \qquad f \in M \tag{4.3}$$

where B is the multiplication operator defined by

$$B(x) := \delta_{x,x_0} NZ(\beta_1) \exp[\beta_1 H(x)] \tag{4.4}$$

The $\|\cdot\|_2$-norm of the vector $u_0 := JBJv_1$ $(= Bv_1)$ in $\mathscr{K}$ is given by

$$\|u_0\|_2 = [Z(\beta_1)\, N]^{1/2} \exp[\beta_1 H(x_0)/2] \tag{4.5}$$

For each positive, real β, let $q_\beta(\cdot,\cdot)$ be a transition probability function on X, i.e.,

$$q_\beta(x, y) \geqslant 0, \quad x, y \in X, \qquad \sum_{y \in X} q_\beta(x, y) = 1, \quad x \in X$$

Let us make the following assumptions:

(i) Irreducibility. For each $\beta \in (0, \infty)$, $q_\beta(\cdot, \cdot)$ is irreducible, i.e., for each $x, y \in X$, there exists a positive integer n such that the matrix element $q_\beta^{(n)}(x, y)$ of the nth power of the matrix $q_\beta(\cdot, \cdot)$ is strictly positive;

(ii) Detailed Balance. For each $\beta \in (0, \infty)$,

$$\exp[-\beta H(x)]\, q_\beta(x, y) = \exp[-\beta H(y)]\, q_\beta(y, x), \quad x, y \in X \tag{4.6}$$

(iii) "Uniformity". For any $x, y \in X$, either $q_\beta(x, y) = 0$ for all $\beta \in (0, \infty)$, or $\alpha_1 \leqslant \max\{q_\beta(x, y), q_\beta(y, x)\} \leqslant \alpha_2$, where α_1, α_2 are strictly positive constants, independent of $x, y \in X$, $\beta \in (0, \infty)$;

(iv) Continuity. The function $\beta \mapsto q_\beta(\cdot, \cdot)$ is continuous and has a limit $q_\infty(\cdot, \cdot)$ as $\beta \to \infty$. [Note that $q_\infty(\cdot, \cdot)$ is not irreducible if U has local minima which are not global minima.]

Two examples satisfying all the above assumptions are given below.

Example 1 (The usual one). Let a symmetric, irreducible transition probability function $q_0(\cdot, \cdot)$ be given, and let

$$q_\beta(x, y) := q_0(x, y) \exp[-\beta(U(y) - U(x))_+], \quad x \neq y \in X \tag{4.7}$$

where, for any real-valued function g, $(g(x))_+ := \max\{g(x), 0\}$;

$$q_\beta(x, x) := 1 - \sum_{y:\, y \neq x} q_\beta(x, y) \tag{4.8}$$

Example 2 (As in refs. 11 and 18). Let $q_0(\cdot, \cdot)$ be a symmetric, irreducible transition probability function with $q_0(x, x) = 0$ for all x, and let

$$q_\beta(x, y) := q_0(x, y)\{1 + \exp[\beta(U(y) - U(x))]\}^{-1} \qquad x \neq y \in X \tag{4.9}$$

$$\begin{aligned} q_\beta(x, x) &:= 1 - \sum_{y:\, y \neq x} q_\beta(x, y) \\ &= \sum_{y \in X} q_0(x, y)\{1 + \exp[-\beta(U(y) - U(x))]\}^{-1} \end{aligned} \tag{4.10}$$

For each β in $(0, \infty]$, define the maps Q_β and L_β on $\mathcal{M}$ by

$$(Q_\beta f)(x) := \sum_{y \in X} q_\beta(x, y) f(y) \qquad f \in \mathcal{M}, \quad x \in X \tag{4.11}$$

$$(L_\beta f)(x) := \sum_{y \in X} q_\beta(x, y)[f(y) - f(x)], \qquad f \in \mathcal{M}, \quad x \in X \tag{4.12}$$

Note that the term with $y = x$ drops from the summation (4.12) and that $L_\beta = Q_\beta - I$, where I is the identity map.

By (4.6), L_β satisfies the detailed balance condition with respect to μ_β. The corresponding Dirichlet form is most easily obtained through (2.19); it is given by

$$E_\beta(f, g) = \sum_{x, y \in X} e_\beta(x, y)[\bar{f}(x) - \bar{f}(y)][g(x) - g(y)] \tag{4.13}$$

where

$$e_\beta(x, y) := \frac{1}{2NZ(\beta)} \exp[-\beta H(x)]\, q_\beta(x, y) = e_\beta(y, x), \quad x, y \in X \tag{4.14}$$

The simulated annealing procedure of Section 3 can be described as follows. Let $\{s_k : k = 0, 1, 2,...\}$ be an increasing sequence of positive numbers, with $s_0 = 0$, let $t_k := s_k - s_{k-1}$, $k = 1, 2,...$, and let $\beta(s)$ be a monotonically nondecreasing function (the *cooling schedule*) defined on $(0, \infty)$, with $\beta(s) \to \infty$ as $s \to \infty$, which takes on a constant value β_k on each interval $(s_{k-1}, s_k]$ of length t_k. Let $\{P_{s,t} : 0 \leqslant s \leqslant t \in \mathbb{R}\}$ be the solution of

$$\frac{d}{dt} P_{s,t} f = P_{s,t} L_{\beta(t)}(f), \qquad f \in \mathcal{M}\ (s, t \neq s_k) \tag{4.15}$$

with $\lim_{\varepsilon \to 0} P_{s,s+\varepsilon} = P_{s,s} = I$ for all s [including the s_k where $s \mapsto \beta(s)$ is discontinuous]. Let also $p_{s,t}(x, y)$, $x, y \in X$, be such that

$$(P_{s,t} f)(x) = \sum_{y \in X} p_{s,t}(x, y) f(y), \qquad f \in \mathcal{M}, \quad x \in X \tag{4.16}$$

Then the simulated annealing algorithm can be described by a Markov process $\{Y_t: t \in \mathbb{R}^+\}$ such that, for all $0 < t_1 < \cdots < t_n \in \mathbb{R}$ and $A_1,..., A_n$ subsets of X, one has

$$\text{Prob}[Y_{t_1} \in A_1, Y_{t_2} \in A_2,..., Y_{t_n} \in A_n] = \sum_{y_1 \in A_1} \sum_{y_2 \in A_2} \cdots \sum_{y_n \in A_n} p_{0,t_1}(x_0, y_1)\, p_{t_1,t_2}(y_1, y_2) \cdots p_{t_{n-1},t_n}(y_{n-1}, y_n) \tag{4.17}$$

By irreducibility, the Markov semigroup T_t^β generated by L_β has only the constant functions as fixed points. Since $\mathscr{K}$ is of finite dimension, the spectral gap assumption holds. An estimate of the constant $\Gamma(\beta)$ has been given by Holley and Stroock[19] in the following way. For each pair x, $y \in X$, a *path* p^{xy} from x to y is a finite sequence $\{p_0^{xy}, p_1^{xy},..., p_n^{xy}\}$ $[n = n(x, y)]$ such that $p_0^{xy} = x$, $p_n^{xy} = y$, and $q_\beta(p_{k-1}^{xy}, p_k^{xy}) > 0$, $k = 1,..., n$ [by assumption (iii) the last inequality holds for all $\beta \in (0, \infty)$ or for no β]. By irreducibility, the set of paths p^{xy} from x to y is nonempty for any x, $y \in X$. Let, for x, $y \in X$,

$$m(x, y) := \min\{(\max_k \{H(p_k^{xy})\}) - H(x) - H(y): \text{all paths } p^{xy}\} \tag{4.18}$$

and let

$$m := \max\{m(x, y): x, y \in X\} \tag{4.19}$$

If x, y are such that $m(x, y) = m$, then either x or y belongs to the set S^0 of global minima of U.[19] Then m coincides with the maximum depth of a local minimum which is not a global minimum, as defined in ref. 17. The following theorem, proved in the Appendix, is a slight generalization of Theorem 2.1 of Holley and Stroock.[19]

Theorem 4.1.[19] Under assumptions (i)–(iv), there exist two strictly positive constants Γ_1 and Γ_2 such that

$$\Gamma_1 e^{-\beta m} \leqslant \Gamma(\beta) \leqslant \Gamma_2 e^{-\beta m} \tag{4.20}$$

In the following it will be assumed that m is strictly positive, otherwise there would be no point in using simulated annealing. In analogy with the literature, I shall consider a cooling schedule of the form

$$\beta(s_k) = (1/c) \log(s_k + 1), \qquad k = 1, 2,... \tag{4.21}$$

c being a (large) positive constant, which can be read in at least two convenient ways:

$$\text{(a)} \quad s_k = k, \qquad t_k = 1$$

$$\beta_k = \beta(s_k) = (1/c)\log(k+1), \qquad k = 1, 2, \ldots \tag{4.22}$$

$$\text{(b)} \quad \beta_k = \beta(s_k) = k, \qquad s_k = \exp(ck) - 1$$

$$t_k = s_k - s_{k-1} = \exp(ck)\,[1 - \exp(-c)], \qquad k = 1, 2, \ldots \tag{4.23}$$

Taking into account (4.15) and writing just Γ for Γ_1, we have in case (a)

$$\begin{aligned} a_k &\geqslant \Gamma k^{-m/c} - (h/2c)[\log(k+1) - \log k] \\ &\geqslant \Gamma k^{-m/c} - (h/2c)\,k^{-1} \end{aligned} \tag{4.24}$$

and in case (b)

$$a_k \geqslant \Gamma e^{k(c-m)}(1 - e^{-c}) - h/2 \tag{4.25}$$

Theorem 4.2. Suppose that assumptions (i)–(iv) hold. Then in case (a) the assumptions of Theorem 3.5 hold if $c > m$, so that, for any initial state x_0, we have

$$\lim_{t \to \infty} \text{Prob}[Y_t \in X - S^0] = 0 \tag{4.26}$$

Proof. If $c > m$, then $k^{1-m/c}$ diverges as $k \to +\infty$. Fix a constant $a > 0$, and choose k large enough to have

$$k^{1-m/c} \geqslant (a + h/2c)/\Gamma$$

so that

$$a_k \geqslant a/k \qquad \text{for} \quad k \text{ large enough} \tag{4.27}$$

On the other hand, from Lemma 3.3 we have the bound

$$\begin{aligned} b_r &\leqslant \|R^2_{r,r+1} - 1\| \\ &= Z(\beta_r)^{-1} Z(\beta_{r+1}) \exp[(\beta_{r+1} - \beta_r)\,h] - 1 \\ &\leqslant \exp[(\beta_{r+1} - \beta_r)\,h] - 1 \end{aligned}$$

since $Z(\beta)$ is a nonincreasing function of β (see the proof of Lemma 3.4). By (4.24) we have $\beta_{r+1} - \beta_r \leqslant 1/[(r+1)\,c] < 1/(rc)$, so that, letting $b := 2h/c$, we obtain the bound

$$b_r \leqslant \exp(h/rc) - 1 \leqslant br^{-1} \qquad \text{for} \quad r \text{ large enough} \tag{4.28}$$

From (4.27) and (4.28) one can obtain the bound of the form (3.21) which is needed for Theorem 3.5. I present the argument as it would be if (4.27) and (4.28) were known to hold for all positive k and r; the general case can be handled with some extra complication. We have

$$\sum_{r=0}^{n-1} \exp\left(-\sum_{k=r+1}^{n} a_k\right) b_r$$
$$\leqslant b_0 \exp\left(-a \sum_{k=1}^{n} k^{-1}\right) + b \sum_{r=1}^{n-1} r^{-1} \exp\left(-a \sum_{k=r+1}^{n} k^{-1}\right) \tag{4.29}$$

The first term on the rhs of (4.29) tends to 0 in the limit as $n \to \infty$ since $\sum_{k=1}^{\infty} k^{-1}$ is divergent (note that this removes the dependence of the error on the initial condition). The second term is bounded, by Lemma 2.1 of ref. 10, or by the following explicit evaluation: Since

$$\sum_{k=r+1}^{n} k^{-1} \geqslant \int_r^n x^{-1}\, dx = \log(n/r)$$

we have

$$b \sum_{r=1}^{n-1} r^{-1} \exp\left(-a \sum_{k=r+1}^{n} k^{-1}\right)$$
$$\leqslant (b/n) \sum_{r=1}^{n-1} (n/r)^{-1} \exp[-a \log(n/r)] = (b/n) \sum_{r=1}^{n-1} (r/n)^{a-1}$$
$$\underset{n\to\infty}{\approx} b \int_0^1 x^{a-1}\, dx = b/a = 2h/ac \quad \blacksquare \tag{4.30}$$

Theorem 4.3. Suppose that assumptions (i)–(iv) hold. Then in case (b) the assumptions of Theorem 3.6 hold if $c > m$, so that, for any subset A of X and for any initial state x_0, we have

$$\lim_{t\to\infty} \text{Prob}[Y_t \in A] = \text{card}(A \cap S^0)/\text{card}(S^0) \tag{4.31}$$

Proof. Fix a constant $a > 0$ and choose k large enough to have

$$e^{k(c-m)} \geqslant (a + h/2c)/\Gamma(1 - e^{-c})$$

Then $a_k \geqslant a$ for k large enough. To prove that $b_r \to 0$, we consider, from (3.13),

$$b_r = [Z(2\beta_r - \beta_{r+1})\, Z(\beta_{r+1})\, Z(\beta_r)^{-2} - 1]^{1/2}$$
$$= [Z(r-1)\, Z(r+1)\, Z(r)^{-2} - 1]^{1/2}$$

which tends to 0 as $r \to \infty$ since $Z(r)$ tends to a finite limit $[\text{card}(S^0)/N]$ as $r \to \infty$. ∎

I conclude this section by sketching how the foregoing arguments have to be modified to become applicable to the case of discrete time. Consider

$$\lim_{n \to \infty} \mu'(Q_{\beta_1}^{t_1} Q_{\beta_2}^{t_2} \cdots Q_{\beta_n}^{t_n}(f)) \tag{4.32}$$

where now $\{t_k\}$ is a sequence of positive integers, and where Q_β is given by (4.11). The annealing algorithm for discrete time is described by a Markov chain $\{Y_t: t = 0, 1, 2,...\}$ with $Y_0 = x_0$ and with transition (i.e., conditional) probabilities given by

$$\text{Prob}[Y_t = y \mid Y_{t-1} = x] = q_{\beta(t)}(x, y), \qquad x, y \in X \tag{4.33}$$

Since now we are dealing with iteration of maps at discrete time instants, we must get rid of possible periodicities. To this end, let us assume, in addition to (i)–(iv), the following condition:

(v) Aperiodicity. For each fixed β in $(0, \infty]$, the Markov chain determined by the transition map Q_β is aperiodic.

In the cases of Examples 1 and 2 above, condition (v) may be already implied by assumptions (i)–(iv); however, I do not investigate this matter here.

Theorem 4.4. Suppose that assumptions (i)–(v) hold, and consider the simulated annealing algorithm for discrete time, with a cooling schedule of the form (4.21), with $c > m$. Then, for any initial state x_0, we have in case (a)

$$\lim_{t \to \infty} \text{Prob}[Y_t \in X - S^0] = 0 \tag{4.34}$$

and in case (b), with t_k given by the smallest integer which is not smaller than $\exp(ck)[1 - \exp(-c)]$,

$$\lim_{t \to \infty} \text{Prob}[Y_t \in A] = \text{card}(A \cap S^0)/\text{card}(S^0) \tag{4.35}$$

Proof. We may view Q_β ($\beta \in (0, \infty]$) as a positivity-preserving linear operator on $\mathscr{M}$. Since $Q_\beta(1) = 1$, the spectrum of Q_β is contained in the closed disk of unit radius, and the only eigenvalue having absolute value 1 is $+1$ (if not, the corresponding Markov chain would not be aperiodic). If we now define

$$S_k(f \exp(-\beta_k H/2)) := Q_{\beta_k}^{t_k}(f) \exp(-\beta_k H/2) \tag{4.36}$$

we note that, as before, S_k is self-adjoint by (4.6). The eigenvalues of S_k are the same as the eigenvalues of $Q_{\beta_k}^{t_k}$: then the only one having absolute value 1 is $+1$, which is simple. The spectral gap $\Gamma(\beta)$ is the second eigenvalue of $-L_\beta = I - Q_\beta$, and tends to 0 as $\beta \to \infty$. Eigenvalues of Q_β cannot accumulate toward -1 as $\beta \to \infty$; if not, Q_∞ would have the eigenvalue -1. Then, for β large enough the eigenvalue of S_k with absolute value nearest to 1 is $[1-\Gamma(\beta_k)]^{t_k}$, which is not larger than $[1-\Gamma\exp(-\beta_k m)]^{t_k}$.

Then all the arguments of Section 3 can be repeated, with the only modification that now a_k is given by

$$a_k := t_k\,|\log[1-\Gamma(\beta_k)]| - (\beta_k - \beta_{k-1})\,h/2 \tag{4.37}$$

In the limit as $\beta \to \infty$, $\Gamma(\beta)$ tends to 0 and $|\log[1-\Gamma(\beta)]|$ is asymptotic to $\Gamma(\beta)$; then we can repeat essentially the same arguments as in Theorems 4.2 and 4.3, to obtain the desired result. ∎

The present results (the same as in ref. 19) are weaker than the conclusion of Hajek[17] that a necessary and sufficient condition for (4.26) to hold is that (in the present notation) $\sum_k \exp(-\beta_k m) = +\infty$, which is implied by (3.24). In particular, $c = m$ ensures convergence. Hajek does not even need detailed balance, but only a much weaker reversibility condition. However the present approach, following ref. 19, has the advantages of providing bounds on the errors and of being easily adaptable to other situations (actually, it is quite analogous to the approach of the authors who have worked on the Langevin algorithm[2,6,12,14]; see in particular Gidas[14] for a discussion which is very similar to the present one).

5. APPLICATIONS: FINITE QUANTUM SYSTEMS

Let $\mathcal{M}$ be the von Neumann algebra $\mathcal{B}(\mathcal{H}_0)$ of all bounded linear operators on a separable Hilbert space $\mathcal{H}_0$. Let $\mathcal{K}$ be the Hilbert space of all Hilbert–Schmidt operators on $\mathcal{H}_0$, with scalar product

$$\langle u, v\rangle_{\mathcal{K}} := \mathrm{Tr}_{\mathcal{H}_0}[u^* v], \qquad u, v \in \mathcal{K}$$

Then $\mathcal{M}$ may be identified with its left regular representation $\pi(\mathcal{M})$ acting on $\mathcal{K}$ as

$$\pi(A)\,u := Au, \qquad A \in \mathcal{M}, \quad u \in \mathcal{K}$$

A faithful normal state μ on $\mathcal{M}$ is determined by a positive trace-class operator on $\mathscr{H}_0$, with unit trace and with densely defined inverse. Denoting by ϕ ($\in \mathscr{K}$) the positive square root of this operator, we have

$$\mu(A) = \mathrm{Tr}_{\mathscr{H}_0}[\phi^2 A] = \langle \phi, A\phi \rangle_{\mathscr{K}}, \qquad A \in \mathscr{M}$$

The modular automorphism group σ_t is determined by

$$\sigma_t(A) = \phi^{2it} A \phi^{-2it}, \qquad A \in \mathscr{M}, \quad t \in \mathbb{R}$$

the modular involution J is the involution (*) on $\mathscr{K}$, so that

$$JAJB\phi = B\phi A^*, \qquad A, B \in \mathscr{M}$$

In the following, all scalar products will be understood in $\mathscr{K}$ and all traces on $\mathscr{H}_0$.

Let H be a self-adjoint element of $\mathscr{M}$, let $h := \|H\|$, and assume that $\{0, h\} \subseteq \sigma(H) \subseteq [0, h]$. From Section 2 we have, for all positive, real β,

$$V(\beta)\, u = u \exp(-\beta H/2), \qquad u \in \mathscr{K} \tag{5.1a}$$

$$Z(\beta) = \langle \phi, \exp(-\beta H)\, \phi \rangle = \mathrm{Tr}[\phi^2 \exp(-\beta H)] \tag{5.1b}$$

$$\begin{aligned} \mu_\beta(A) &= Z(\beta)^{-1} \langle \phi \exp(-\beta H), A\phi \rangle \\ &= Z(\beta)^{-1}\, \mathrm{Tr}[\phi \exp(-\beta H)\, \phi A], \qquad A \in \mathscr{M} \end{aligned} \tag{5.1c}$$

$$v_\beta = Z(\beta)^{-1/2}\, \phi \exp(-\beta H/2) \tag{5.1d}$$

For the sake of simplicity, we shall assume that ϕ and H commute, so that $v_\beta = v_\beta^*$ $(= Jv_\beta)$, $\mu_\beta(A) = \mathrm{Tr}[v_\beta^2 A]$ for all A, and the modular automorphism group σ_t^β associated with $\mu_\beta = \langle v_\beta, \cdot\, v_\beta \rangle$ is given by

$$\sigma_t^\beta(A) = v_\beta^{2it} A v_\beta^{-2it}, \qquad A \in \mathscr{M}, \quad t \in \mathbb{R} \tag{5.2}$$

If μ' is any normal state on $\mathscr{M}$, we have $\mu'(A) = \mathrm{Tr}(pA)$, $A \in \mathscr{M}$, where p is a positive trace-class operator with unit trace (a density operator) on $\mathscr{H}_0$. We may formally write

$$\begin{aligned} \mu'(A) &= \mathrm{Tr}(v_\beta^{-1} p A v_\beta) = \langle p v_\beta^{-1}, A v_\beta \rangle \\ &= \langle JBJ v_\beta, A v_\beta \rangle, \qquad A \in \mathscr{M} \end{aligned} \tag{5.3}$$

where $B := v_\beta^{-1} p v_\beta^{-1}$; for a dense set of normal states, B is indeed a bounded operator (an element of $\mathscr{M}$).

We recall from refs. 1 and 21 the general form of the generator L_β of a quantum dynamical semigroup T_t^β on $\mathscr{M} = \mathscr{B}(\mathscr{H}_0)$ satisfying the quantum detailed balance condition with respect to the state μ_β with density operator $v_\beta^2 = Z(\beta)^{-1}\,\phi\exp(-\beta H)\,\phi$, omitting the skew-symmetric Hamiltonian part (as in ref. 25). We have

$$L_\beta(A) = \sum_j \{(V_j^* A V_j - \tfrac{1}{2}[V_j^* V_j, A]_+) + \exp[-\alpha_j(\beta)](V_j A V_j^* - \tfrac{1}{2}[V_j V_j^*, A]_+)\}, \qquad A \in \mathscr{M} \quad (5.4)$$

where $[B, A]_+ := BA + AB$, and where the $V_j \in \mathscr{M}$, $\alpha_j(\beta) \in \mathbb{R}$, satisfy

$$v_\beta^{-2} V_j v_\beta^2 = \exp[-\alpha_j(\beta)]\, V_j, \qquad j = 1, 2,... \quad (5.5)$$

The series in (5.4) are convergent in the ultraweak topology. The corresponding Dirichlet form can be most easily computed through the dissipation function D_β of (2.19). Recalling the expression of D_β from ref. 22, we get

$$\begin{aligned} E_\beta(A, B) &= \tfrac{1}{2}\sum_j \{\mu_\beta([V_j, A]^*\,[V_j, B]) + \exp[-\alpha_j(\beta)]\,\mu_\beta([V_j^*, A]^*\,[V_j^*, B])\} \\ &= \tfrac{1}{2}\sum_j \{\mu_\beta([V_j, A]^*\,[V_j, B]) + \mu_\beta([v_\beta V_j^* v_\beta^{-1}, A]^*\,[v_\beta V_j^* v_\beta^{-1}, B])\} \end{aligned} \quad (5.6)$$

It is also interesting to give the explicit form of the nonnegative self-adjoint operator G_β on $\mathscr{K}$ such that $S_t^\beta = \exp(-G_\beta t)$. For all u in the dense subset $\mathscr{M} v_\beta$ of $\mathscr{K}$ we may write

$$G_\beta u = G_\beta^0 u - G'_\beta u \quad (5.7)$$

where

$$\begin{aligned} G_\beta^0 u &:= \tfrac{1}{2}\sum_j (V_j^* V_j + J v_\beta^{-1} V_j v_\beta^2 V_j^* v_\beta^{-1} J \\ &\quad + v_\beta^{-1} V_j v_\beta^2 V_j^* v_\beta^{-1} + J V_j^* V_j J)\, u \\ &= \tfrac{1}{2}\sum_j \{(V_j^* V_j u + u V_j^* V_j) + \exp[-\alpha_j(\beta)](V_j V_j^* u + u V_j V_j^*)\} \end{aligned} \quad (5.8)$$

and where

$$\begin{aligned} G'_\beta u &:= \tfrac{1}{2}\sum_j (V_j J v_\beta^{-1} V_j v_\beta J + V_j^* J v_\beta V_j^* v_\beta^{-1} J \\ &\quad + v_\beta^{-1} V_j v_\beta J V_j J + v_\beta V_j^* v_\beta^{-1} J V_j^* J)\, u \\ &= \sum_j \exp[-\alpha_j(\beta)/2](V_j^* u V_j + V_j u V_j^*) \end{aligned} \quad (5.9)$$

The state μ_β is invariant under T_t^β. Recall from ref. 7 that it is the unique T_t^β-invariant normal state and that

$$\text{weak-}\lim_{t\to\infty} \mu' \cdot T_t^\beta = \mu_\beta \quad \text{for all normal states } \mu' \text{ on } \mathcal{M} \tag{5.10}$$

if and only if

$$\{V_j, V_j^*: j = 1, 2,..., \}'' = \mathcal{M} \tag{5.11}$$

This condition amounts to saying that the eigenspace of G_β corresponding to the eigenvalue 0 is one-dimensional (spanned by v_β), and is a necessary prerequisite for the existence of a spectral gap $\Gamma(\beta) > 0$ as in (2.20). In the present paper I shall not be concerned with the general problem of finding conditions ensuring the existence of $\Gamma(\beta) > 0$ and estimates on its dependence on β. When $\mathcal{H}_0$ is finite-dimensional, we have $\Gamma(\beta) > 0$ if and only if (5.11) holds; then the theory of Section 3 can be applied.[9] Estimates for the spectral gap of a generator satisfying the quantum detailed balance condition for some models of infinite quantum systems have been considered in ref. 30 and references quoted therein in connection with the problem of *critical slowing down.*

In the remainder of this section I shall describe how a (fictitious) quantum system can be associated to a problem of classical simulated annealing as in Section 4, and I shall estimate the spectral gap for the corresponding class of generators.

Let X be a finite state space with N points, and let $U: X \to \mathbb{R}$ be a function to be interpreted as energy. We may suppose that points x in X are arranged in a definite order, such that $U(x) < U(y)$ implies $x < y$. Associate to X the Hilbert space $\mathcal{H}_0 = \mathbb{C}^N$, spanned by the canonical orthonormal basis, denoted as $\{e_x : x \in X\}$. Then $\mathcal{M} = \mathcal{B}(\mathcal{H}_0) = M(N, \mathbb{C})$ is spanned by the matrix units $\{e_{xy} : x, y \in X\}$ such that $e_{xy} e_z = \delta_{yz} e_x$. Let $\mathcal{K}$ be $M(N, \mathbb{C})$ equipped with the scalar product $\langle u, v \rangle = \mathrm{Tr}[u^* v]$, and let $\phi := N^{-1/2} 1$, so that $\mu = \langle \phi, \cdot \phi \rangle$ is the normalized trace. The algebra of complex-valued functions defined on X can be embedded isomorphically into the subalgebra $\mathcal{D}$ of $\mathcal{M}$ consisting of all diagonal matrices; the embedding, denoted by $D(\cdot)$, is given by

$$D(f) := \sum_{x \in X} f(x)\, e_{xx} \tag{5.12}$$

Define H in $\mathcal{M}$ by

$$H := D(U - \min U) \tag{5.13}$$

Then the state μ_β is given by

$$\mu_\beta(A) = \mathrm{Tr}[\exp(-\beta H)\, A]/\mathrm{Tr}[\exp(-\beta H)], \qquad A \in \mathcal{M} \tag{5.14}$$

It is clear that, for any function f on X,

$$\mu_\beta(D(f)) = \mu_{\beta,cl}(f) \tag{5.15}$$

where $\mu_{\beta,cl}$ is the state defined by (4.1c). In the limit as $\beta \to \infty$ we have $\mu_\beta \to \mu_\infty$, where

$$\mu_\infty(A) = \mathrm{Tr}[P_0 A]/\mathrm{Tr}[P_0], \qquad A \in \mathcal{M} \tag{5.16}$$

where $P_0 := \sum_{x \in S^0} e_{xx}$ is the projection onto the eigenspace of $\mathcal{H}_0$ corresponding to the eigenvalue 0.

A classical simulated annealing algorithm is determined by a family of transition probability functions $q_\beta(\cdot, \cdot)$, $\beta \in (0, \infty)$, satisfying assumptions (i)–(iv) and by a cooling schedule $\beta(\cdot)$. The corresponding quantum simulated annealing will be determined by the same cooling schedule $\beta(\cdot)$ and by a family $\{V_j\}$ of elements of $\mathcal{M}$ defined as follows. Let

$$\mathcal{J} := \{j := (x, y)\colon x, y \in X, x < y, q_\beta(x, y) \neq 0\} \tag{5.17}$$

[independent of β by (iii)], and define

$$V_j := q_\beta(y, x)^{1/2}\, e_{xy} \qquad \text{if} \quad j = (x, y) \tag{5.18}$$

Note that if $q_\beta(\cdot, \cdot)$ is as in Example 1, then V_j is independent of β. Then we have

$$[H, V_j] = -w_j V_j, \qquad j \in \mathcal{J} \tag{5.19}$$

where

$$w_j = U(y) - U(x) \quad (\geqslant 0) \qquad \text{if} \quad j = (x, y) \tag{5.20}$$

For each positive β, define L_β by

$$\begin{aligned} L_\beta(A) := \sum_{j \in \mathcal{J}} \{&(V_j^* A V_j - \tfrac{1}{2}[V_j^* V_j, A]_+) \\ &+ \exp(-\beta w_j)(V_j A V_j^* - \tfrac{1}{2}[V_j V_j^*, A]_+)\} \end{aligned} \tag{5.21}$$

L_β satisfies the quantum detailed balance condition with respect to μ_β.

Proposition 5.1. For any complex-valued function f on X, we have

$$L_\beta(D(f)) = D(L_{\beta,cl}(f)) \tag{5.22}$$

where L_β is given by (5.21) above and where $L_{\beta,cl}$ is the classical generator given by (4.12).

Proof. We have

$$L_\beta(D(f)) = \sum_{j\in\mathscr{I}} \sum_{z\in X} \{(V_j^* e_{zz} V_j - \tfrac{1}{2}[V_j^* V_j, e_{zz}]_+)$$
$$+ \exp(-\beta w_j)(V_j e_{zz} V_j^* - \tfrac{1}{2}[V_j V_j^*, e_{zz}]_+)\} f(z)$$
$$= \sum_{x,y:x<y} \sum_{z\in X} \{q_\beta(y,x)\, e_{yy}(\delta_{zx} - \delta_{zy})$$
$$+ \exp[-\beta(U(y) - U(x))]\, q_\beta(y,x)\, e_{xx}(\delta_{zy} - \delta_{zx})\} f(z)$$

Changing x into y and y into x in the first line and recalling (4.6), we obtain

$$L_\beta(D(f)) = \sum_{x\in X} \Bigg\{ \sum_{y:y<x} q_\beta(x,y)[f(y) - f(x)]$$
$$+ \sum_{y:y>x} \exp[-\beta(U(y) - U(x))]\, q_\beta(y,x)[f(y) - f(x)]\Bigg\} e_{xx}$$
$$= \sum_{x\in X} \sum_{y:y\neq x} q_\beta(x,y)[f(y) - f(x)]\, e_{xx}$$
$$= D(L_{\beta,cl}(x)) \quad \blacksquare$$

The generator G_β on K corresponding to $-L_\beta$ can be written as $G_\beta = G_\beta^0 - G_\beta'$, where

$$G_\beta^0 u = \tfrac{1}{2} \sum_{x,y:x\neq y} q_\beta(x,y)[e_{xx}, u]_+ \tag{5.23}$$

$$G_\beta' u = \sum_{x,y:x<y} \{\exp[-\beta(U(y) - U(x))/2]\, q_\beta(x,y)$$
$$\times (e_{yx} u e_{xy} + e_{xy} u e_{yx})\} \tag{5.24}$$

The classical generator $G_{\beta,cl}$ corresponding to $-L_{\beta,cl}$ can be obtained by $D(G_{\beta,cl}(f)) = G_\beta(D(f))$ for all f.

The estimates on the spectral gap for G_β can be obtained as follows. Let Γ_1, Γ_2, m be the constants such that the spectral gap $\Gamma_{cl}(\beta)$ for $G_{\beta,cl}$ satisfies

$$\Gamma_1 e^{-\beta m} \leqslant \Gamma_{cl}(\beta) \leqslant \Gamma_2 e^{-\beta m} \tag{5.25}$$

Let l be the maximum energy gain in one step:

$$l := \max\{U(y) - U(x) : x, y\in X, x<y, q_\beta(x,y)>0\} \tag{5.26}$$

Theorem 5.2. Let

$$m^* := \max\{m, l\} \tag{5.27}$$

$$\Gamma^* := \min\{\Gamma_1, \alpha_1\} \tag{5.28}$$

where $\alpha_1 > 0$ has been introduced in assumption (iii) of Section 4. Then the spectral gap $\Gamma(\beta)$ for G_β satisfies the bound

$$\Gamma^* e^{-\beta m^*} \leqslant \Gamma(\beta) \leqslant \Gamma_2 e^{-\beta m} \tag{5.29}$$

Proof. Let F be the orthogonal projection of $\mathscr{K}$ onto the subspace of diagonal matrices, given by

$$Fu = \sum_{x \in X} e_{xx} u e_{xx}, \qquad u \in \mathscr{K} \tag{5.30}$$

We have, clearly,

$$G_\beta^0 F = F G_\beta^0, \qquad G'_\beta F = G'_\beta = F G'_\beta \tag{5.31}$$

so that both $F\mathscr{K}$ and $(1-F)\mathscr{K}$ are globally invariant under $S_t^\beta = \exp(-G_\beta t)$. On $F\mathscr{K}$ the generator G_β coincides with the classical generator $G_{\beta,cl}$. On $(1-F)\mathscr{K}$, G_β reduces to G_β^0. Note that G_β^0 is strictly positive and satisfies the bound

$$\begin{aligned} G_\beta^0 &\geqslant \min_{x \in X} \left\{ \sum_{y:\, y \neq x} q_\beta(x, y) \right\} \\ &\geqslant \min_{x \in X} \left\{ e^{-\beta l} \sum_{y:\, y \neq x} \max\{q_\beta(x, y), q_\beta(y, x)\} \right\} \geqslant e^{-\beta l} \alpha_1 \end{aligned} \tag{5.32}$$

since, for any given x, at least one $q_\beta(x, y)$ $(y \neq x)$ must be nonzero, by irreducibility of $q_\beta(\cdot, \cdot)$. The conclusion follows. ∎

From the above discussion we can conclude that the same results as in Theorems 4.1 and 4.2 [cases (a) and (b), respectively] hold also in the present quantum framework, provided m is replaced by m^*. This conclusion is formulated as follows.

Theorem 5.3. Suppose that assumptions (i)–(iv) of Section 4 hold. Let $\beta(s_k) = (1/c) \log(s_k + 1)$, $k = 1, 2, \ldots$, where $c > m^*$. Let L_β be given by (5.21). Then, for any state μ' on $\mathscr{M} = M(N, \mathbb{C})$ we have in case (a)

$$\lim_{n \to \infty} \mu'(T_{t_1}^{\beta_1} \cdots T_{t_n}^{\beta_n}(A)) = 0 \qquad \text{for all } A \text{ such that } \mu_\infty(A^* A) = 0 \tag{5.33}$$

and in case (b)

$$\lim_{n \to \infty} \mu'(T_{t_1}^{\beta_1} \cdots T_{t_n}^{\beta_n}(A)) = \mu_\infty(A) \qquad \text{for all } A \tag{5.34}$$

APPENDIX

Here I prove the inequality

$$\Gamma_1 e^{-\beta m} \leqslant \Gamma(\beta) \leqslant \Gamma_2 e^{-\beta m}$$

Since this is a slight generalization of Lemmas 2.3 and 2.7 of ref. 19, I shall assume that the reader is familiar with those proofs and shall insist only on the (minor) changes that I have introduced. Recall that

$$E_\beta(f, f) = \sum_{x, y \in X} e_\beta(x, y) \, |f(x) - f(y)|^2, \qquad f \in \mathscr{M}$$

where

$$\begin{aligned} e_\beta(x, y) &:= \frac{1}{2NZ(\beta)} \exp[-\beta H(x)] \, q_\beta(x, y) \\ &= e_\beta(y, x), \qquad x, y \in X \end{aligned}$$

Upon defining (as in ref. 19)

$$\begin{aligned} \mathrm{Var}_\beta(f) &:= \mu_\beta(|f - \mu_\beta(f)|^2) \\ &= [2N^2 Z(\beta)^2]^{-1} \sum_{x, y \in X} \exp[-\beta(H(x) + H(y))] \, |f(x) - f(y)|^2 \end{aligned}$$

we have

$$\Gamma(\beta) = \inf\{E_\beta(f, f)/\mathrm{Var}_\beta(f) \colon f \in \mathscr{M}, \mathrm{Var}_\beta(f) \neq 0\}$$

Then we have to prove the following inequalities:

$$\mathrm{Var}_\beta(f) \leqslant C e^{\beta m} E_\beta(f, f), \qquad f \in \mathscr{M}, \quad \beta \in (0, \infty) \tag{A.1}$$

for some constant C (to be identified with $1/\Gamma_1$), and

$$E_\beta(F, F) \leqslant \Gamma_2 e^{-\beta m} \mathrm{Var}_\beta(F), \qquad \beta \in (0, \infty) \tag{A.2}$$

for some function $F \in \mathscr{M}$.

Proof of (A.1). For any pair x, y of points in X, choose a path p^{xy} from x to y on which $m(x, y)$ is attained. The length $n(x, y)$ of any path is bounded above by some number (say) n. For z, w in X, let[19]

$$\chi_{z,w}(x, y) := \begin{cases} 1 & \text{if, for some } i, p_i^{xy} = z \text{ and } p_{i+1}^{xy} = w \\ 0 & \text{otherwise} \end{cases}$$

$[\chi_{z,w}(x, y)$ depends also on the path p^{xy} chosen]. Then, as in ref. 19, we have

$$\operatorname{Var}_\beta(f) = \frac{1}{2N^2Z(\beta)^2} \sum_{x,y,z,w \in X} \chi_{z,w}(x, y)\, |f(z) - f(w)|^2 \times \exp[-\beta(H(x) + H(y))] \tag{A.3}$$

Multiply and divide each nonzero summand in (A.3) by $e_\beta(z, w)$. [Note that if $e_\beta(z, w) = 0$, then also $\chi_{z,w}(x, y) = 0$]. Writing $e_\beta(z, w)$ as $[2NZ(\beta)]^{-1} \exp[-\beta H(z)]\, q_\beta(z, w)$ and recalling that, for the path chosen, $H(p_i^{xy}) - H(x) - H(y) \leqslant m(x, y) \leqslant m$, we obtain

$$\frac{\chi_{z,w}(x, y) \exp[-\beta(H(x) + H(y))]}{2N^2Z(\beta)^2\, e_\beta(z, w)} \leqslant \frac{\exp(\beta m)\, \chi_{z,w}(x, y)}{NZ(\beta)\, q_\beta(z, w)} \tag{A.4}$$

Similarly, writing $e_\beta(z, w)$ as $[2NZ(\beta)]^{-1} \exp[-\beta H(w)]\, q_\beta(w, z)$ and recalling that, for the path chosen, $H(p_{i+1}^{xy}) - H(x) - H(y) \leqslant m(x, y) \leqslant m$, we obtain

$$\frac{\chi_{z,w}(x, y) \exp[-\beta(H(x) + H(y))]}{2N^2Z(\beta)^2\, e_\beta(z, w)} \leqslant \frac{\exp(\beta m)\, \chi_{z,w}(x, y)}{NZ(\beta)\, q_\beta(w, z)} \tag{A.5}$$

It follows from assumption (iii) that

$$\min\{1/q_\beta(z, w), 1/q_\beta(w, z) : q_\beta(z, w) \neq 0\} \leqslant 1/\alpha_1$$

Then we insert into (A.3) the more restrictive of the inequalities (A.4) and (A.5), and recall that $Z(\beta) \geqslant \lim_{\beta \to \infty} Z(\beta) = \operatorname{card}(S^0)/N$. So we obtain

$$\operatorname{Var}_\beta(f) \leqslant \frac{ne^{\beta m}}{\operatorname{card}(S^0)\, \alpha_1} \sum_{z,w \in X} \sum_{x,y \in X} \chi_{z,w}(x, y)\, e_\beta(z, w) \times |f(z) - f(w)|^2$$

Since X is finite, $\sum_{x,y \in X} \chi_{z,w}(x, y)$ is bounded by a constant, independent of z, w, and β. Then (A.1) follows, with

$$C := \left[n \max_{z,w} \sum_{x,y} \chi_{z,w}(x, y) \right] \Big/ [\operatorname{card}(S^0)\, \alpha_1] \quad \blacksquare$$

Proof of A.2. Let x, y be two points such that $m(x, y) = m$. Let A be a subset of X containing y and not containing x, and let F be its indicator function. Then

$$\operatorname{Var}_\beta(F) \geqslant [N^2Z(\beta)^2]^{-1} \exp[-\beta(H(x) + H(y))] \tag{A.6}$$

On the other hand, we have

$$E_\beta(F, F) = [NZ(\beta)]^{-1} \sum_{z \notin A} \sum_{w \in A} \exp[-\beta H(z)]\, q_\beta(z, w)$$

$$\leqslant \alpha_2 [NZ(\beta)]^{-1} \sum_{z \notin A,\, w \in A:\, q_\beta(z, w) > 0} \exp[-\beta H(z)] \qquad \text{(A.7)}$$

Recalling that $Z(\beta) \leqslant 1$, we obtain from (A.6) and (A.7)

$$E_\beta(F, F)/\mathrm{Var}_\beta(F) \leqslant \alpha_2 N \sum_{z \notin A,\, w \in A:\, q_\beta(z, w) > 0} \exp[-\beta(H(z) - H(x) - H(y))]$$

It follows from ref. 19, Lemma 2.3, that we can choose A such that

$$H(z) - H(x) - H(y) \geqslant m(x, y) \qquad \text{for} \quad z \notin A,\, w \in A,\, q_\beta(z, w) > 0$$

By assumption, $m(x, y) = m$. Then we obtain

$$E_\beta(F, F)/\mathrm{Var}_\beta(F) \leqslant \alpha_2 N e^{-\beta m} \operatorname{card}\{(z, w): z \notin A,\, w \in A,\, q_\beta(z, w) > 0\} \qquad \text{(A.8)}$$

Since X is finite, the cardinality of the set in (A.8) is some finite constant, which is independent of β by assumption (iii). The claim follows. ∎

ACKNOWLEDGMENTS

This work was supported in part by CNR project "Statistica dei Processi Aleatori." I thank P. Serafini and M. Colombetti for arousing my interest in simulated annealing.

REFERENCES

1. R. Alicki, On the quantum detailed balance condition for non-Hamiltonian systems, *Rep. Math. Phys.* **10**:249–258 (1976).
2. F. Aluffi-Pentini, V. Parisi, and F. Zirilli, Global optimization and stochastic differential equations, *J. Opt. Theory Appl.* **47**:1–16 (1985).
3. B. Apolloni, N. Cesa-Bianchi, and D. de Falco, A numerical implementation of "quantum annealing," Preprint BiBoS Nr. 324/88 (1988); B. Apolloni, C. Carvalho, and D. de Falco, Quantum stochastic optimization, Preprint BiBoS Nr. 327/88 (1988).
4. H. J. Brascamp, Equilibrium states for a classical lattice gas, *Commun. Math. Phys.* **18**:82–96 (1970).
5. V. Cerny, A thermodynamical approach to the travelling salesman problem. An efficient simulation algorithm, *J. Opt. Theory Appl.* **45**:41–51 (1985).
6. T. S. Chiang, K. R. Hwang, and S. J. Sheu, Diffusion for global optimization in R^n, *SIAM J. Control Opt.* **25**:737–753 (1987).
7. A. Frigerio, Stationary states of quantum dynamical semigroups, *Commun. Math. Phys.* **63**:269–276 (1978).
8. A. Frigerio and M. Verri, Long-time asymptotic properties of dynamical semigroups on W^*-algebras, *Math. Z.* **180**:275–286 (1982).

9. A. Frigerio, Time-inhomogeneous and nonlinear quantum evolutions, in *Quantum Probability and Applications V*, L. Accardi and W. von Waldenfels, eds., (Springer-Verlag, Berlin, to appear).
10. S. B. Gelfand and S. K. Mitter, Simulated annealing, in *Stochastics in Combinatorial Optimization*, G. Andreatta, F. Mason, and P. Serafini, eds. (World Scientific, Singapore, 1987).
11. S. Geman and D. Geman, Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.* **6**:721–741 (1984).
12. S. Geman and C. R. Hwang, Diffusions for global optimization, *SIAM J Control Opt.* **24**:1031–1043 (1986).
13. B. Gidas, Non-stationary Markov chains and convergence of the annealing algorithm, *J. Stat. Phys.* **39**:73–131 (1985).
14. B. Gidas, Global optimization via the Langevin equation, in *Proceedings 24th IEEE Conference on Decision and Control* (1985), pp. 774–777.
15. R. J. Glauber, Time-dependent statistics of the Ising model, *J. Math. Phys.* **4**:294–307 (1963).
16. V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, Completely positive dynamical semigroups of N-level systems, *J. Math. Phys.* **17**:821–825 (1976).
17. B. Hajek, Cooling schedules for optimal annealing, *Math. Oper. Res.* **13**:311–329 (1988).
18. G. E. Hinton, T. J. Sejnowski, and D. H. Ackley, Boltzmann machines: Constraint satisfaction machines that learn, Department of Computer Science Technical Report CMU-CS-84-119, Carnegie-Mellon University, Pittsburgh, Pennsylvania (1984).
19. R. Holley and D. Stroock, Simulated annealing via Sobolev inequalities, *Commun. Math. Phys.* **115**:553–569 (1988).
20. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Optimization by simulated annealing, *Science* **220**:621–680 (1983).
21. A. Kossakowski, A. Frigerio, V. Gorini, and M. Verri, Quantum detailed balance and KMS condition, *Commun. Math. Phys.* **57**:97–110 (1977); *Erratum*, *Commun. Math. Phys.* **60**:96 (1978).
22. G. Lindblad, On the generators of quantum dynamical semigroups, *Commun. Math. Phys.* **48**:119–130 (1976).
23. M. Lundy and A. Mees, Convergence of an annealing algorithm, *Math. Progr.* **34**:111–124 (1986).
24. P. A. Martin, Modèles en mecanique statistique des processus irreversibles, *Lecture Notes in Physics*, Vol. 103 (Springer, Berlin, 1979).
25. J. Quaegebeur, G. Stragier, and A. Verbeure, Quantum detailed balance, *Ann. Inst. H. Poincaré* **41**:25–36 (1984).
26. C. Radin, Approach to equilibrium for a simple model, *J. Math. Phys.* **11**:2945–2955 (1970).
27. D. W. Robinson, Strongly positive semigroups and faithful invariant states, *Commun. Math. Phys.* **85**:129–142 (1982).
28. H. Spohn, An algebraic condition for the approach to equilibrium of an open N-level system, *Lett. Math. Phys.* **2**:33–38 (1977).
29. P. J. M. van Laarhoven and E. H. L. Aarts, (eds.) *Simulated Annealing: Theory and Applications* (Reidel, Dordrecht, 1987).
30. A. Verbeure, Detailed balance and critical slowing down, in *Quantum Probability and Applications III*, L. Accardi and W. von Waldenfels, eds. (Springer, Berlin, 1988), pp. 354–362.
31. L. Younes, Estimation and annealing for Gibbsian fields, *Ann. Inst. H. Poincaré B* **24**:269–294 (1988).