A simple approach to time-inhomogeneous dynamics and applications to (fast) simulated annealing

1999 J. Phys. A: Math. Gen. 32 5389

(http://iopscience.iop.org/0305-4470/32/29/301)

J. Phys. A: Math. Gen. **32** (1999) 5389–5407. Printed in the UK PII: S0305-4470(99)01565-6

A simple approach to time-inhomogeneous dynamics and applications to (fast) simulated annealing

G Gielis† and C Maes‡§

† King's College Research Centre and Statistical Laboratory, University of Cambridge, Cambridge CB2 1ST, UK
‡ Instituut voor Theoretische Fysica, KU Leuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium

E-mail: Guy.Gielis@kings.cam.ac.uk and Christian.Maes@fys.kuleuven.ac.be

Received 3 February 1999

Abstract. We study time-inhomogeneous stochastic dynamical systems with vanishing noise. Conditions implying strong ergodicity are proposed; the asymptotic probability distribution is characterized and we estimate the speed of relaxation. Applying these results to simulated annealing enables us to introduce systems that converge faster than a classical annealing dynamics determined by a Gibbs factor.

1. Introduction

There is a long tradition of applying and studying the technique of simulated annealing. This well known method allows to locate the absolute minima of a complicated energy landscape. The main idea is to run a time-inhomogeneous dynamics for which on every (discrete) time step n the Gibbs probability distribution μ_{β_n}, with respect to the considered energy function and at temperature β_n, is invariant for the applied transition. The hope is that when $\beta_n, n = 1, 2, \ldots$, gradually increases this dynamics approaches the probability distribution $\mu_\infty = \lim_{\beta\uparrow\infty} \mu_\beta$. This distribution assigns equal weights to all the states where the energy is minimal.

To limit the computation time, one is inclined to increase the inverse temperature β_n as fast as possible. However, when the system is cooled too quickly, it is possible for it to become stuck in a local minimum and, as a consequence, for it never to reach the desired probability distribution. So, a central question is to find the optimum cooling schedule. A first rigorous result was given by Geman and Geman [3]. They showed that, for a system in which the energy of a configuration σ of N different variables is $U(\sigma)$, the simulated annealing process is successful as soon as

$$\beta_n \leqslant \frac{1}{N[\max_\sigma U(\sigma) - \min_\sigma U(\sigma)]} \log n$$

for large times n. Afterwards, this result was confirmed for similar dynamical systems by Gidas [4], Holley and Stroock [6], Holley *et al* [5], and Deuschel and Mazza in [2] and references therein. Moreover, in several of these papers the result was completed by the calculation of an upper bound for the rate at which the dynamics reaches its asymptotic distribution.

§ Onderzoeksleider FWO Belgium.

0305-4470/99/295389+19$30.00

Recently, some new computer simulations have pushed this discussion in a new direction. Penna and Tsallis [8] and Stariolo [9] proposed new algorithms in which the transition probabilities do still depend on the 'temperature' and the energy, but in a different form than in the standard Gibbs factor. Nevertheless, the simulations suggest that the asymptotic probability distribution is the same and, in addition, it seems that these new systems relax faster than the classical simulated annealing procedures.

In this paper, we first extend the ideas of simulated annealing to a more general class of stochastic dynamics. The energy function will no longer play a central role and the notion of temperature is replaced by a more general notion of noise. More precisely, we consider serial update, discrete time Markov processes in which the noise level is asymptotically extinguished. We introduce conditions on the rate at which this happens, ensuring that the dynamics remain ergodic. The same result allows us to characterize the asymptotic probability distribution and to estimate the speed of convergence. Finally, we propose a criterion that is useful in comparing the asymptotic behaviour of different dynamical systems. All these results are formulated in section 2 of the paper.

In section 3 we apply the theorems to simulated annealing. We not only recover, in a very simple way, some known results, but also propose new dynamical rules that allow a more efficient location of the absolute minima in the energy landscape.

Next, in section 4, we propose a very useful coupling between two dynamics. This coupling plays a key role in several of the proofs collected in section 5. Finally, in the appendix, one can find the necessary, but rather trivial calculations supporting some statements made in the main text.

2. Definitions and main results

We consider N variables $\sigma(i)$, $i = 1, \ldots, N$. For every i, the variable $\sigma(i)$ takes values in the finite set $\Omega_0 = \{1, 2, \ldots, M\}$, $M \in \mathbb{N}$. This gives rise to a configuration space $\Omega = \Omega_0^N$, which is the N-fold Cartesian product of the set Ω_0. For every configuration $\sigma \in \Omega$, we call $\sigma(i)$ the state of the ith variable. The set of real-valued functions on Ω is denoted by $\mathcal{B}(\Omega)$ and $\mathcal{P}(\Omega)$ is the set of all possible probability distributions on Ω.

For $\mu \in \mathcal{P}(\Omega)$ and $f \in \mathcal{B}(\Omega)$, we denote by $\mu(f)$ the expected value of f with respect to μ. $\|f\| = \max_{\eta\in\Omega} |f(\eta)|$ is the (supremum) norm of f and the oscillation at i is defined by $\Delta_i f = \max_{\eta\in\Omega} \max_{\zeta\in\Omega_0} |f(\eta^{i,\zeta}) - f(\eta)|$. Herein is the configuration $\eta^{i,\zeta}$, $\eta \in \Omega$, $\zeta \in \Omega_0$ such that

$$\eta^{i,\zeta}(j) = \begin{cases} \eta(j) & \text{if } j \neq i \\ \zeta & \text{if } j = i. \end{cases}$$

Finally, the semi-norm $|||f||| = \sum_{i=1}^{N} \Delta_i f$ is called the total oscillation.

We introduce a discrete time, serial update stochastic dynamics σ_n, $n = 1, 2, \ldots$, on Ω. σ_n is the configuration at time n. To construct this dynamics we start from a sequence of probability kernels $P_{\beta_n}(\cdot|\cdot)$, $\beta_n \geqslant 0$, $n = 1, 2, \ldots$, from $\Omega \times \Omega$ to $[0, 1]$, which assign to every $\eta \in \Omega$ the probability distributions $P_{\beta_n}(\cdot|\eta) \in \mathcal{P}(\Omega)$. The transition probability $P_{\beta_n}(\sigma|\eta)$ gives the probability to obtain the configuration σ at time n, when the previous configuration was η. The parameters β_n, $n = 1, 2, \ldots$, are supposed to determine the amount of noise in the system. The higher β_n, the lower the noise level. This will be clear from the examples. We imagine that the β_n are taken from a non-decreasing positive function β_t, $t \geqslant 0$.

In the serial update dynamics studied in this paper, the variables $\sigma(i)$, $i = 1, 2, \ldots, N$, are modified one by one, in a random order. This means that on every time step, first an index i is selected from the set $\{1, 2, \ldots, N\}$ with probability $1/N$ and then the configuration is updated from $\sigma_{n-1} = \eta$ to $\sigma_n = \sigma$ according to some individual probability law $p^{(i)}_{\beta_n}(\sigma|\eta)$. This law is such that the probability $p^{(i)}_{\beta_n}(\sigma|\eta)$ differs only from 0 when $\sigma = \eta^{i,\xi}$ for some $\xi \in \Omega_0$. The full transition probability can then be written as

$$P_{\beta_n}(\sigma|\eta) = \frac{1}{N}\sum_{i=1}^{N} p^{(i)}_{\beta_n}(\sigma|\eta). \tag{1}$$

Starting from (1), we define for positive integers n_0, n, with $n_0 \leqslant n$, the transition operators P_{β_n} and $P^{n_0,n}$ acting on $f \in \mathcal{B}(\Omega)$, as follows:

$$\begin{aligned} &P^{n,n} f(\eta) = f(\eta) \\ &P^{n-1,n} f(\eta) = P_{\beta_n} f(\eta) \equiv \sum_{\sigma\in\Omega} f(\sigma) P_{\beta_n}(\sigma|\eta) \end{aligned} \tag{2}$$

and by iteration

$$P^{n_0,n} f(\eta) = P_{\beta_{n_0+1}}[P^{n_0+1,n} f](\eta) \equiv \sum_{\sigma\in\Omega} [P^{n_0+1,n} f(\sigma)] P_{\beta_{n_0+1}}(\sigma|\eta).$$

So, $P^{n_0,n} f(\eta)$ is the expected value of the function f at time n, when the configuration at time n_0 was η. When the noise depends on the time in a non-trivial way, the dynamics becomes time-inhomogeneous. Otherwise we have a time-homogeneous evolution. The latter case, this is when $\beta_n = \beta$ for a fixed number β on every time n, will be emphasized by adding an extra index β to the transition operators, i.e. we write $P^{n_0,n}_\beta$ instead of $P^{n_0,n}$.

It is well known that as soon as the noise is non-zero, in our examples this will correspond to $\beta < \infty$, very mild conditions (e.g., that the dynamics is an irreducible and aperiodic Markov chain) imply that there exists a unique invariant probability distribution μ_β such that for every $n_0 \leqslant n$

$$\mu_\beta(P^{n_0,n}_\beta f) = \mu_\beta(f) \tag{3}$$

and

$$\|P^{n_0,n}_\beta f - \mu_\beta(f)\| \leqslant C\mathrm{e}^{-\lambda(n-n_0)} \tag{4}$$

for some constants $\lambda > 0$ and $C = C(\lambda, f) < \infty$. For this reason, we feel free to assume as a hypothesis the existence of such a probability distribution for the time-homogeneous versions of the dynamics studied in this paper.

In this paper we focus on the non-homogeneous case. Since for this kind of dynamics, in general, there does not exist any longer a probability distribution that remains unaffected at every time step, it becomes more complicated to predict the asymptotic behaviour of these systems. One can ask whether an expression similar to (4) still holds, knowing that equation (3) cannot be satisfied. To address this problem, we introduce the following two notions.

Definition 1. *A dynamics is called weakly ergodic if for every function $f \in \mathcal{B}(\Omega)$ and every initial time n_0*

$$\lim_{n\uparrow\infty} \sup_{\eta,\eta'} |P^{n_0,n} f(\eta) - P^{n_0,n} f(\eta')| = 0.$$

Weak ergodicity implies that the asymptotic behaviour of the dynamics becomes independent of its initial configuration. This statement does not say anything more about this behaviour. It is for instance not necessary that the dynamics converges and if it does, it is not clear what the limit will be. The following definition allows for these shortcomings.

Denote by μ_β the unique probability distribution that satisfies $\mu_\beta(P_\beta f) = \mu_\beta(f)$ and suppose that $\lim_{n\uparrow\infty} \mu_{\beta_n}$ exists for some sequence $\beta_n, n = 1, 2, \ldots$.

Definition 2. *A dynamics is called strongly ergodic if for every function $f \in \mathcal{B}(\Omega)$ and every initial time n_0*

$$\lim_{n\uparrow\infty} \| P^{n_0,n} f - \mu_\infty(f) \| = 0$$

with

$$\mu_\infty(f) = \lim_{n\uparrow\infty} \mu_{\beta_n}(f).$$

To state our results in this context, we rely on the following parameters:

$$q_{\beta_n} = \max_k \max_{\eta,\eta'} \mathrm{var}(p^{(k)}_{\beta_n}(\eta^{k,\cdot}|\eta), p^{(k)}_{\beta_n}(\eta'^{k,\cdot}|\eta')). \tag{5}$$

Here var denotes the variational distance between probability distributions: for two distributions μ_1 and μ_2 on Ω_0 (or on Ω), this distance is defined as

$$\mathrm{var}(\mu_1, \mu_2) = \frac{1}{2} \sum_\sigma |\mu_1(\sigma) - \mu_2(\sigma)| \leqslant 1.$$

The sum is over Ω_0 (or over Ω).

When the q_{β_n} approach 1, the dynamics becomes strongly dependent on the past, which corresponds to the low noise (high β) regime. The first result says that, in order to have weak ergodicity, $\lim_{n\uparrow\infty} q_{\beta_n} = 1$ is allowed as long as the convergence to 1 is slow enough.

Theorem 1. *When*

$$\lim_{n\uparrow\infty} n^\alpha (1 - q_{\beta_n}) = \infty \tag{6}$$

for some $0 < \alpha < 1/N$, then the dynamics is weakly ergodic. Moreover, there exist constants $C = C(N, \alpha, \{q_{\beta_n}\}_{n\in\mathbb{N}}) < \infty$ and $\lambda = \lambda(N, \alpha) > 0$ such that for every $f \in \mathcal{B}(\Omega)$

$$\sup_{\eta,\eta'} |P^{n_0,n} f(\eta) - P^{n_0,n} f(\eta')| \leqslant C|||f||| \exp(-\lambda(n^{1-N\alpha} - {n_0}^{1-N\alpha})) \tag{7}$$

for every $n_0 \leqslant n$.

Under the extra condition that the transition probabilities are smooth enough functions of the time for large n, we also can prove strong ergodicity.

Suppose that $\Gamma(t) = \max_{\eta,n} \sum_\sigma \left| \frac{\mathrm{d}}{\mathrm{d}t} p^{(n)}_{\beta_t}(\sigma|\eta) \right|$ exists, then the following result holds.

Theorem 2. *When*

$$\lim_{n\uparrow\infty} n^\alpha (1 - q_{\beta_n}) = \infty \tag{8}$$

and

$$\lim_{t\uparrow\infty} t^\delta \Gamma(t) = 0 \tag{9}$$

for some $0 < \alpha < 1/N$ and $\delta > 2\alpha N$, then the dynamics is strongly ergodic. Moreover, there exist a constant $0 < \gamma < \delta - 2\alpha N$ and a constant $C = C(N, M, \alpha, \delta, \gamma, \{q_{\beta_n}\}_{n\in\mathbb{N}}) < \infty$ such that

$$\| P^{0,n} f - \mu_\infty(f) \| \leqslant |\mu_\infty(f) - \mu_n(f)| + C|||f||| n^{-\gamma} \tag{10}$$

for every $f \in \mathcal{B}(\Omega)$.

In a final result the asymptotic probability distributions of different strongly ergodic dynamics are compared. More precisely, we propose conditions that allow to define a set of dynamical systems that are all strongly ergodic and that all have the same asymptotic probability distribution as a given reference dynamics.

Theorem 3. *Suppose that the dynamics with transition operators* $(P^{n_0,n})_{n_0,n}$ *satisfies the conditions (8) and (9). When another dynamics* $(\tilde{P}^{n_0,n})_{n_0,n}$ *satisfies*

$$\lim_{n\uparrow\infty} n^{\delta} \max_{\eta,k} \mathrm{var}(p^{(k)}_{\beta_n}(\eta^{k,\cdot}|\eta), \tilde{p}^{(k)}_{\tilde{\beta}_n}(\eta^{k,\cdot}|\eta)) = 0 \tag{11}$$

for some $\delta > \alpha N$*, then this dynamics also is strongly ergodic and has the same asymptotic probability distribution, i.e.*

$$\tilde{\mu}_\infty(f) = \mu_\infty(f)$$

for every $f \in \mathcal{B}(\Omega)$*.*

Remark 1. Under the general condition (6), theorem 1 is optimal. To see this, we consider for instance the following very simple system on $\Omega = \Omega_0 = \{-1, +1\}$:

$$P_{\beta_n}(\sigma|\eta) = \begin{cases} 1 - \frac{1}{2}\mathrm{e}^{-\beta_n} & \text{if } \sigma = \eta \\ \frac{1}{2}\mathrm{e}^{-\beta_n} & \text{if } \sigma \neq \eta. \end{cases}$$

When $\beta_n = (1+\delta)\log(1+n)$ for some $\delta \geqslant -1$, then

$$q_{\beta_n} = 1 - \mathrm{e}^{-\beta_n} = 1 - \frac{1}{(1+n)^{1+\delta}}.$$

So, condition (6) is only satisfied when $\delta < 0$. In the other case, i.e. when $\delta \geqslant 0$, the inequality (7) is violated. Indeed, let $I_+(\sigma)$ be the function that gives 1 when $\sigma = +1$ and 0 when $\sigma = -1$, then the following identity holds:

$$\begin{aligned} |P^{0,n}I_+(+1) - P^{0,n}I_+(-1)| &\\ = \big|&(1 - \tfrac{1}{2}\mathrm{e}^{-\beta_1})P^{1,n}I_+(+1) + \tfrac{1}{2}\mathrm{e}^{-\beta_1}P^{1,n}I_+(-1) \\ &- \tfrac{1}{2}\mathrm{e}^{-\beta_1}P^{1,n}I_+(+1) - (1 - \tfrac{1}{2}\mathrm{e}^{-\beta_1})P^{1,n}I_+(-1)\big| \\ = (1 &- \mathrm{e}^{-\beta_1})\left|P^{1,n}I_+(+1) - P^{1,n}I_+(-1)\right| \\ = \prod_{i=1}^{n}&(1 - \mathrm{e}^{-\beta_i}) \\ = \prod_{i=1}^{n}&\left(1 - \frac{1}{(i+1)^{1+\delta}}\right). \end{aligned} \tag{12}$$

When $\delta = 0$ this implies that

$$|P^{0,n}I_+(+1) - P^{0,n}I_+(-1)| = (n+1)^{-1}.$$

When $\delta > 0$, the right-hand side of (12) is strictly positive uniformly in n and the dynamics is no longer weakly ergodic.

Remark 2. One of the advantages of theorem 2 is that it does not require knowledge of the invariant probability distributions $\mu_{\beta_n}, n = 1, 2, \ldots$. This is in contrast to some other arguments proving strong ergodicity. For instance, in [3] the following condition appears:

$$\sum_{n=1}^{\infty} \mathrm{var}(\mu_{\beta_n}, \mu_{\beta n+1}) < \infty. \tag{13}$$

Remark 3. It is also interesting to notice that condition (9) becomes weaker as soon as the parameter α in (8) is smaller, which implies that the convergence in (7) is faster. It is not so hard to extend our argument to see that if for every n, $q_{\beta_n} < \epsilon$ for some $\epsilon < 1$, then an upper bound (7) holds which decreases exponentially fast when $n - n_0$ grows. In that case, it is sufficient that $\Gamma(t)$ converges to 0 in order to get a strongly ergodic system.

Example 1. Let $\Omega_0 = \{-1, +1\}$. Assign to every $i = 1, 2, \ldots, N$ two disjoint subsets A_i and B_i of $\{1, 2, \ldots, N\}$. However, sets with different indices i can overlap. Define

$$p^{(i)}_{\beta_n}(\eta^{i,+1}|\eta) = \begin{cases} 1 - \frac{1}{2}\mathrm{e}^{-\beta_n} & \text{if } \operatorname{sgn}\left[\sum_{j\in A_i}\eta(j) - \sum_{j\in B_i}\eta(j)\right] \geqslant 0 \\ \frac{1}{2}\mathrm{e}^{-\beta_n} & \text{otherwise.} \end{cases}$$

In this model

$$q_{\beta_n} = 1 - \mathrm{e}^{-\beta_n}$$

which means that, given $\alpha < 1/N$, condition (6) is satisfied as soon as, for any $\alpha' < \alpha$ and n large enough, $\beta_n \leqslant (\alpha'/N)\log n$. Moreover, when $\beta_t = \epsilon \log t$ for some $\epsilon > 0$, then $\max_{\sigma,\eta,n} |\frac{\mathrm{d}}{\mathrm{d}t} p^{(n)}_{\beta_t}(\sigma|\eta)| = \frac{1}{2}\epsilon t^{-(1+\epsilon)}$. Hence, condition (9) is satisfied for every $\delta < 1$, uniformly in $\epsilon > 0$. So, taking $\alpha = \frac{1}{4N}$, we see that the dynamics is strongly ergodic for $\beta_n = \frac{1}{5N}\log n$ and the second term in the right-hand side of (10) decays with a rate $0 < \gamma < 1/2$.

3. Simulated annealing

In this section, we apply the above results to simulated annealing. We assume that the energy function to be minimized is a non-negative function $U \in \mathcal{B}(\Omega)$ such that $\min_\sigma U(\sigma) = 0$ and $\max_\sigma U(\sigma) = U_{\max}$, for some finite $U_{\max} < \infty$. We define $\Omega(U)$ to be the set of minima

$$\Omega(U) = \{\sigma \in \Omega : U(\sigma) = 0\}$$

and for a given configuration σ, we define δ_σ as the probability distribution with the following prescription:

$$\delta_\sigma(\eta) = \begin{cases} 1 & \text{if } \eta = \sigma \\ 0 & \text{otherwise.} \end{cases}$$

Finally, $\mu_{\min}$ is the probability distribution that assigns equal weights to all the elements of $\Omega(U)$, i.e. for $f \in \mathcal{B}(\Omega)$

$$\mu_{\min}(f) = \frac{1}{|\Omega(U)|}\sum_{\sigma\in\Omega(U)}\delta_\sigma(f) = \frac{1}{|\Omega(U)|}\sum_{\sigma\in\Omega(U)} f(\sigma). \tag{14}$$

We propose several possible dynamical rules that are all indexed by the energy function U and a time-dependent noise parameter (temperature) β_n. We look to see under which conditions on β_n these dynamics are strongly ergodic, we wonder which of them have $\mu_{\min}$ as asymptotic probability distribution and we investigate the relaxation behaviour.

Example 2. The first example is the analogue of the so called heat bath algorithm and is determined by the Boltzmann–Gibbs distribution, i.e.

$$p^{(k)}_{\beta_n}(\eta^{k,\zeta}|\eta) = \frac{1}{Z_{\beta_n}(\eta)}\exp(-\beta_n[U(\eta^{k,\zeta}) - U(\eta)]) \tag{15}$$

with $\eta \in \Omega$, $\zeta \in \Omega_0$ and $Z_{\beta_n}(\eta)$ the normalization constant. The only probability distribution that solves the equation $\mu_{\beta_n}(P_{\beta_n} f) = \mu_{\beta_n}(f)$ for every $f \in \mathcal{B}(\Omega)$, is given by the Boltzmann–Gibbs distribution:

$$\mu_{\beta_n}(\sigma) = \frac{1}{Z_{\beta_n}}\exp(-\beta_n U(\sigma)). \tag{16}$$

Here $Z_{\beta_n} = \sum_\sigma \exp(-\beta_n U(\sigma))$ is the normalizing partition sum.

The corresponding density q_{β_n} (see equation (5)) can be estimated as follows:

$$\begin{aligned}
q_{\beta_n} &= \max_{k,\eta,\eta'} \mathrm{var}(p^{(k)}_{\beta_n}(\eta^{k,\cdot}|\eta), p^{(k)}_{\beta_n}({\eta'}^{k,\cdot}|\eta')) \\
&\leqslant \max_{k,\eta,\eta'} \sum_{\zeta\neq\zeta'} p^{(k)}_{\beta_n}(\eta^{k,\zeta}|\eta) p^{(k)}_{\beta_n}({\eta'}^{k,\zeta'}|\eta') \\
&\leqslant 1 - \min_{k,\eta,\eta'} \sum_{\zeta} p^{(k)}_{\beta_n}(\eta^{k,\zeta}|\eta) p^{(k)}_{\beta_n}({\eta'}^{k,\zeta}|\eta') \\
&\leqslant 1 - \min_{k,\eta,\zeta} p^{(k)}_{\beta_n}(\eta^{k,\zeta}|\eta) \\
&\leqslant 1 - \frac{1}{|\Omega_0|} \mathrm{e}^{-\beta_n U_{\max}}.
\end{aligned} \tag{17}$$

Hence, given $0 < \alpha < 1/N$, condition (8) is satisfied when, for some $\alpha' < \alpha$,

$$\beta_n \leqslant \frac{\alpha'}{U_{\max}} \log n \tag{18}$$

as soon as n is large enough.

Moreover, we show in lemma A1 that when $\beta_n = \epsilon \log n$, for some $\epsilon > 0$, condition (9) is verified for any $0 < \delta < 1$, independently of ϵ. So, for $0 < 2\alpha N < \delta < 1$ and β_n equal to the right-hand side of (18), theorem 2 applies. This means that for any $0 < \gamma < 1 - 2\alpha N$, there exist constants $C_i = C_i(N, M, \alpha, \delta, \gamma, \{q_{\beta_n}\}_{n\in\mathbb{N}}) < \infty$, $i = 1, 2$ such that

$$\begin{aligned}
\|P^{0,n} f - \mu_\infty(f)\| &\leqslant |\mu_\infty(f) - \mu_{\beta_n}(f)| + C_1 |||f||| n^{-\gamma} \\
&\leqslant C_2 \|f\| (n^{-\alpha' U_{\min}/2U_{\max}} + n^{-\gamma})
\end{aligned} \tag{19}$$

for $U_{\min} = \min\{U(\sigma) : \sigma \notin \Omega(U)\}$, $\alpha' < \alpha$ and n large enough. Finally, we also notice that μ_∞ is indeed equal to $\mu_{\min}$. This follows immediately from (16).

Both equations (18) and (19) confirm earlier results that can be found in [2–6].

For the practical use of simulated annealing it is important to find an algorithm that limits the computation time. For this reason it is interesting to develop a dynamical rule that converges as quick as possible to $\mu_{\min}$. In the previous example this corresponds to increasing β_n as fast as possible. However, when β_n grows too rapidly, the dynamics may no longer be ergodic. Another strategy is to construct a completely different dynamical rule for which the Gibbsian probability distributions may no longer be invariant for the individual steps of the dynamics.

Example 3. A first such alternative example is

$$p^{(k)}_{\beta_n}(\eta^{k,\zeta}|\eta) = \frac{1}{Z_{\beta_n}(\eta)} (1 + \beta_n U(\eta^{k,\zeta}))^{-2} \tag{20}$$

where $\eta \in \Omega$, $\zeta \in \Omega_0$ and $Z_{\beta_n}(\eta)$ is the normalization constant. Recall that for every configuration σ, $U(\sigma) \geqslant 0$, so that (20) is well defined. The corresponding invariant probability distributions μ_{β_n}, $n = 1, 2, \ldots$, assign the following weight to a configuration σ:

$$\mu_{\beta_n}(\sigma) = \frac{(1 + \beta_n U(\sigma))^{-2}}{\sum_{\sigma\in\Omega} (1 + \beta_n U(\sigma))^{-2}}.$$

Using the fact that $\min_\sigma U(\sigma) = 0$, we also find that $\mu_\infty = \mu_{\min}$ in this case.

Analogous to (17), we see that

$$q_{\beta_n} \leqslant 1 - \frac{1}{|\Omega_0|}(1 + \beta_n U_{\max})^{-2}$$

which means that condition (8) is satisfied when, for some $\alpha' < \alpha$,

$$\beta_n \leqslant \frac{1}{2U_{\max}|\Omega_0|^{1/2}} n^{\alpha'/2} \tag{21}$$

as soon as n is large enough.

Furthermore, in lemma A2 it is shown that when $\beta_n = Bn^{\epsilon/2}$ for some $\epsilon > 0$ and $B > 0$, condition (9) is verified for any $0 < \delta < 1$, uniformly in ϵ and B. Hence, when $0 < 2\alpha N < \delta < 1$ and when β_n is equal to the right-hand side of (21), we can conclude from theorem 2 that for any $0 < \gamma < 1 - 2\alpha N$ there exist finite constants $C_i = C_i(N, M, \alpha, \delta, \gamma, \{q_{\beta_n}\}_{n\in\mathbb{N}})$, $i = 1, 2$ such that

$$\begin{aligned} \|P^{0,n} f - \mu_\infty(f)\| &\leqslant |\mu_\infty(f) - \mu_{\beta_n}(f)| + C_1|||f|||n^{-\gamma} \\ &\leqslant C_2\|f\|\left(\frac{U_{\max}^2}{U_{\min}^2} n^{-\alpha'} + n^{-\gamma}\right) \end{aligned} \tag{22}$$

for $U_{\min} = \min\{U(\sigma) : \sigma \notin \Omega(U)\}$, $\alpha' < \alpha$ and n large enough.

The observation that (22) decreases faster than (19) confirms [7–9]. In these papers algorithms for simulated annealing different from the Metropolis algorithm, but similar to (20) are investigated using computer simulations. Also there a faster relaxation is found. Of course, since (19) and (22) are only upper bounds, we cannot conclude here that the dynamics determined by (20) indeed converges quicker. This leads us to our last result: theorem 4 allows to show that the dynamics (15) is at least not faster than the one with transition probabilities (20).

We first notice that for every potential U, we can always choose β_n, $n = 1, 2, \ldots$, such that the probabilities (20) verify condition (8) for, for instance, $\alpha = 1/3N$. Hence, setting $\alpha' = 1/4N$ and $\gamma = \frac{1}{4}$ in the upper bound (22), we observe that it is possible to decrease the noise so that the dynamics determined by (20) relaxes faster than the function $n^{-1/4N}$ when time increases.

So, in order to prove that the dynamics (15) is in general not faster, it is sufficient to find one energy function U for which the corresponding simulated annealing procedure relaxes slower than $n^{-1/4N}$, and this for all possible cooling schedules β_n, $n = 1, 2, \ldots$.

Therefore, we take $\Omega_0 = \{-1, +1\}$ and we consider the following potential:

$$U(\sigma) = \begin{cases} 0 & \text{if } \sigma(i) = +1, i = 1, 2, \ldots, N \\ U_{\min} & \text{if } \sigma(i) = -1, i = 1, 2, \ldots, N \\ U_{\max} & \text{otherwise.} \end{cases} \tag{23}$$

We show that for this potential and for the indicator function $I_{-1}(\sigma)$, which is 1 when $\sigma(i) = -1$, $i = 1, \ldots, N$ and 0 for all the other configurations, the following theorem is true.

Theorem 4. *Suppose that the potential (23) satisfies*

$$(2\delta^{-1} + 1)U_{\min} < U_{\max} \tag{24}$$

for some $\delta > 0$. *Then, there exists* $\beta_c = \beta_c(U_{\min}, U_{\max}, N) > 0$ *such that for every increasing and diverging sequence* β_n, $n = 1, 2, \ldots$, *for which the dynamics determined by (15) is strongly ergodic and that satisfies* $\beta_1 > \beta_c$,

$$|\mu_{\beta_1}(P^{0,n} I_{-1}) - \mu_\infty(I_{-1})| = \mu_{\beta_1}(P^{0,n} I_{-1}) \geqslant n^{-\delta} \tag{25}$$

as soon as n *is large enough.*

Using this theorem for $\delta < \frac{1}{4N}$, it is straightforward to conclude that, for all possible choices $\beta_n, n = 1, 2, \ldots$, the transition probabilities (15) determine, in general, not the most efficient dynamics for simulated annealing. Indeed, we found an upper bound on the rate of convergence that, by choosing the appropriate sequence $\beta_n, n = 1, 2, \ldots$, can be satisfied for every possible energy function when the transition probabilities (20) are applied. However, there exist potentials so that for all cooling schedules the dynamics (15) is still slower than this bound.

By the next example we illustrate that the transition rule proposed in (20) is only one possible choice out of a large class of dynamics.

Example 4. Take any positive, differentiable function F on $\mathbb{R}^+$ such that $F(0) = 1$, $\lim_{x\uparrow\infty} F(x) = 0$ and that is strictly decreasing. Then, we can define the transition probabilities

$$p_{\beta_n}^{(k)}(\eta^{k,\zeta}|\eta) = \frac{1}{Z_{\beta_n}(\eta)} F(\beta_n U(\eta^{k,\zeta})) \tag{26}$$

with again the same meaning for $Z_{\beta_n}(\eta)$. The corresponding invariant probability distributions are

$$\mu_{\beta_n}(\sigma) = \frac{F(\beta_n U(\sigma))}{\sum_\sigma F(\beta_n U(\sigma))} \qquad n = 1, 2, \ldots$$

and

$$q_{\beta_n} \leqslant 1 - \frac{1}{|\Omega_0|} F(\beta_n U_{\max}).$$

This implies that (8) is satisfied when, for some $\alpha' < \alpha$,

$$\beta_n \leqslant \frac{1}{U_{\max}} F^{-1}\left(\frac{|\Omega_0|}{n^{\alpha'}}\right) \tag{27}$$

as soon as n is large enough.

Hence, when we can find out for what $\beta_n, n = 1, 2, \ldots$, satisfying (27), there exists a $\delta > 2\alpha N$ so that condition (9) also holds, we obtain a whole set of strongly ergodic dynamical systems for which $\mu_\infty = \mu_{\min}$. Moreover, for any $0 < \gamma < \delta - 2\alpha N$ there exist finite constants $C_i = C_i(N, M, \alpha, \delta, \gamma, \{q_{\beta_n}\}_{n\in\mathbb{N}}, F), i = 1, 2$ such that

$$\begin{aligned}\|P^{0,n} f - \mu_\infty(f)\| &\leqslant |\mu_\infty(f) - \mu_{\beta_n}(f)| + C_1 |||f||| n^{-\gamma} \\ &\leqslant C_2 \|f\| \left(F\left(\frac{U_{\min}}{U_{\max}} F^{-1}\left(\frac{|\Omega_0|}{n^{\alpha'}}\right)\right) + n^{-\gamma}\right)\end{aligned}$$

for $U_{\min} = \min\{U(\sigma) : \sigma \notin \Omega(U)\}, \alpha < \alpha'$ and n large enough.

In lemma A2 it is shown that, for instance, for the functions $F(x) = (1+x)^{-\lambda}, \lambda > 0$, and $F(x) = (1 + \log(1+x))^{-1}$ there exists a sequence $\beta_n, n = 1, 2, \ldots$, that verifies the condition (8) and the condition (9) for $0 < \delta < 1$. None of these functions, however, yields an upper bound for the decay that is considerably faster than (22).

A disadvantage of the previous rules is that they only are well-defined and of interest for positive energy functions whose minimal value equals 0. The latter is necessary to obtain the wished asymptotic probability distribution $\mu_{\min}$. In contrast to example 1, where the dynamical rule is determined by the Gibbs factor $\exp(-\beta U(\sigma))$, this represents no longer the general case in which $\min_\sigma U(\sigma) \neq 0$. For practical applications of simulated annealing, however, we do not explicitly know the full energy function, and we do not know *a priori* whether $\min_\sigma U(\sigma) = 0$ is true or not. For that reason it is interesting to modify the above transition probabilities to a more general form in which $\min_\sigma U(\sigma) = 0$ is no longer crucial. In [7–9] the authors introduce systems similar to the following proposal.

Example 5. For any $\lambda > 0$, take

$$p^{(k)}_{\beta_n}(\eta^{k,\zeta}|\eta) = \begin{cases} \dfrac{1}{Z_{\beta_n}(\eta)}\big(1+\beta_n[U(\eta^{k,\zeta})-U(\eta)]\big)^{-\lambda} & \text{when } \beta_n[U(\eta^{k,\zeta})-U(\eta)] > -\dfrac{1}{2} \\ \dfrac{2^\lambda}{Z_{\beta_n}(\eta)} & \text{otherwise} \end{cases} \tag{28}$$

where $\eta \in \Omega$, $\zeta \in \Omega_0$ and $Z_{\beta_n}(\eta)$ is the normalization constant.

In this case we see that

$$q_{\beta_n} \leqslant 1 - \frac{1}{2^\lambda|\Omega_0|}(1+\beta_n U_{\max})^{-\lambda}.$$

This system can be analysed in a manner similar to that used in the previous examples, but we are not able to verify that the asymptotic probability distribution equals $\mu_{\min}$. This leads us to theorem 3, which enables us to find μ_∞ by comparing the dynamics with a reference dynamics for which we do know the asymptotic distribution.

Example 6. Let $g(n) = \mathrm{e}^{3NU_{\max}\tilde{\beta}_n}$; then we can apply theorem 3 to the following transition probabilities:

$$\tilde{p}^{(k)}_{\tilde{\beta}_n}(\eta^{k,\xi}|\eta) = \begin{cases} \dfrac{1}{Z_{\tilde{\beta}_n}(\eta)}\big(1+g(n)^{-1}[\tilde{\beta}_n(U(\eta^{k,\xi})-U(\eta)]\big)^{-g(n)} \\ \qquad\qquad \text{when } g(n)^{-1}\tilde{\beta}_n(U(\eta^{k,\xi})-U(\eta)) > -\dfrac{1}{2} \\ \dfrac{2^{g(n)}}{Z_{\tilde{\beta}_n}(\eta)} \qquad \text{otherwise.} \end{cases} \tag{29}$$

In lemma A3 we use the dynamics of example 2, with $\beta_n = (\alpha/2U_{\max})\log n$, as a reference to show that when $\tilde{\beta}_n = \beta_n$ this dynamics asymptotically approaches $\mu_{\min}$.

4. A coupling argument

A useful way to compare two dynamical systems is connecting the systems via a coupling. This means that, given the dynamics $(\sigma_n)_{n=1,2,\ldots}$ and $(\sigma'_n)_{n=1,2,\ldots}$ on Ω, with transition operators $(P^{n_0,n})_{n_0,n}$ and $(\tilde{P}^{n_0,n})_{n_0,n}$ respectively, we consider a new evolution $(\sigma_n,\sigma'_n)_{n=1,2,\ldots}$ on $\Omega\times\Omega$, with transition operators $(\mathrm{Prob}^{n_0,n})_{n_0,n}$ such that

$$\sum_\sigma \mathrm{Prob}^{n_0,n}[\sigma,\sigma'|\eta,\eta'] = \tilde{P}^{n_0,n}(\sigma'|\eta') \qquad \text{and} \qquad \sum_{\sigma'} \mathrm{Prob}^{n_0,n}[\sigma,\sigma'|\eta,\eta'] = P^{n_0,n}(\sigma|\eta).$$

The coupling that we have in mind is constructed using individual transition probabilities $p^{(k)}_{\beta,\tilde{\beta}}(\sigma,\sigma'|\eta,\eta')$, $k = 1,2,\ldots,N$, such that, for each k, the probability $p^{(k)}_{\beta,\tilde{\beta}}(\sigma,\sigma'|\eta,\eta')$ only differs from 0 when $\sigma = \eta^{k,\xi}$ and $\sigma' = \eta'^{k,\xi'}$ for some $\xi,\xi' \in \Omega_0$.

In a first step of the coupled dynamics, an index k is selected from the set $\{1,2,\ldots,N\}$, each choice having a probability $1/N$. Then, the variable $(\eta(k),\eta'(k))$ is updated according to the following transition probabilities:

$$p^{(k)}_{\beta,\tilde\beta}(\eta^{k,\xi}, \eta'^{k,\xi}|\eta,\eta') = \begin{cases} \min\{p^{(k)}_\beta(\eta^{k,\xi}|\eta), \tilde p^{(k)}_{\tilde\beta}(\eta'^{k,\xi'}|\eta')\} & \text{if } \xi = \xi' \\ \left[\mathrm{var}(p^{(k)}_\beta(\eta^{k,\cdot}|\eta), \tilde p^{(k)}_{\tilde\beta}(\eta'^{k,\cdot}|\eta'))\right]^{-1} & \\ \quad \times (p^{(k)}_\beta(\eta^{k,\xi}|\eta) - p^{(k)}_{\beta,\tilde\beta}(\eta^{k,\xi}, \eta'^{k,\xi}|\eta,\eta')) & \\ \quad \times (\tilde p^{(k)}_{\tilde\beta}(\eta'^{k,\xi'}|\eta') - p^{(k)}_{\beta,\tilde\beta}(\eta^{k,\xi'}, \eta'^{k,\xi'}|\eta,\eta')) & \text{if } \xi \neq \xi'. \end{cases} \tag{30}$$

The full transition probability is

$$\mathrm{Prob}_{\beta,\beta'}(\sigma,\sigma'|\eta,\eta') = \frac{1}{N}\sum_{i=1}^{N} p^{(i)}_{\beta,\tilde\beta}(\sigma,\sigma'|\eta,\eta'). \tag{31}$$

Notice that the individual transition probabilities (30) have the following property:

$$p^{(k)}_{\beta,\tilde\beta}(\sigma(k) \neq \sigma'(k)|\eta,\eta') = \mathrm{var}(p^{(k)}_\beta(\eta^{k,\cdot}|\eta), \tilde p^{(k)}_{\tilde\beta}(\eta'^{k,\cdot}|\eta')). \tag{32}$$

Starting from (31), the prescription in (2) allows us to construct the coupled dynamics $(\sigma_n, \sigma'_n)_{n=1,2,\ldots}$ on $\Omega \times \Omega$.

Applying this coupling, we connect two copies of the same dynamical system, $(\sigma_n)_{n=1,2,\ldots}$ and $(\sigma'_n)_{n=1,2,\ldots}$, started with different initial configurations, η and η', at time n_0. Next, we assign to this dynamics the following $\{0,1\} \times \{1,2,\ldots,N\}$-valued process $(s(n), S(n))_{n\in\mathbb{Z}}$.

For each $i = 1,\ldots,N$ we set $(s(n_0+i-N), S(n_0+i-N)) = (1,i)$ if $\eta(i) \neq \eta'(i)$ and $(s(n_0+i-N), S(n_0+i-N)) = (0,i)$ otherwise.

For each $n > n_0$, we put $S(n) = j_n$, when $\sigma(j_n)$ is the variable that is updated at time n, and we set $s(n) = 1$ as soon as $\sigma(j_n) \neq \sigma'(j_n)$. When $\sigma_n(j_n)$ and $\sigma'_n(j_n)$ are equal, then $(s(n), S(n)) = (0, j_n)$.

Now, it is important to notice that, due to (32), $s(n) = 1$ can only happen when there is at least one site j for which $\sigma_{n-1}(j) \neq \sigma'_{n-1}(j)$ at the time step $n-1$. If for this j, $k < n$ is the last time that $S(k) = j$, then this implies that $s(k) = 1$. In other words, the event $s(n) = 1$ can only happen when there exists a sequence of integers $n = m_0 > m_1 > \cdots > m_k$, $k > 1$ so that $m_k \leqslant n_0$, $s(m_i) = 1$, $i = 1,\ldots,k$, and

$$\{S(j) : m_{i-1} < j < m_i\} \neq \{1,2,\ldots,N\}.$$

We denote the event that there exists such a sequence by $\{n \longrightarrow n_0\}$. So, when $\{n \longrightarrow n_0\}$ is not true, then $s(k) = 0$ and $\sigma_k = \sigma'_k$ for all $k \geqslant n$.

Notice that $\{n \longrightarrow n_0\}$ is only a function of the variables $(s(k), S(k))$, $k < n$, for k strictly smaller than n.

It is now straightforward to see from (32) that the probability that $s(n) = 1$ is not larger than

$$q_{\beta_n} = \begin{cases} \max\limits_k \max\limits_{\eta,\eta'} \mathrm{var}(p^{(k)}_\beta(\eta^{k,\cdot}|\eta), p^{(k)}_\beta(\eta^{k,\cdot}|\eta')) & \text{when } n > n_0 \\ 1 & \text{otherwise.} \end{cases} \tag{33}$$

Finally, we assign to every $n \in \mathbb{Z}$ Bernoulli random variables $\tilde s(n)$ with density q_{β_n} as defined above. Denote by ν_q the joint probability distribution of the variables $\tilde s(n)$ and $S(n)$. Then, for every finite $\Lambda \subset \mathbb{Z}$ and any set $(j_1,\ldots,j_{|\Lambda|}) \in \{1,\ldots,N\}^{|\Lambda|}$,

$$\nu_q((\tilde s(n), S(n)) = (1, j_n), \forall n \in \Lambda) = \left(\frac{1}{N}\right)^{|\Lambda|} \prod_{n\in\Lambda} q_{\beta_n}.$$

Combining all these ingredients, we can conclude that for the coupled process

$$\mathrm{Prob}[\sigma_n(i) \neq \sigma'_n(i) | \sigma_{n_0}, \sigma'_{n_0}] \leqslant \nu_q(n \longrightarrow n_0). \tag{34}$$

Lemma 1. *When*

$$\lim_{n\uparrow\infty} n^\alpha (1 - q_{\beta_n}) = \infty \tag{35}$$

for some $0 < \alpha < 1/N$, *then there exist constants* $C = C(N, \alpha, \{q_{\beta_n}\}_{n\in\mathbb{N}}) < \infty$ *and* $\lambda = \lambda(N, \alpha) > 0$ *such that*

$$\nu_q(m \longrightarrow n) \leqslant C \exp(-\lambda(m^{1-N\alpha} - n^{1-N\alpha})).$$

Proof. $\{m \longrightarrow n\}$ does not happen as soon as there is a set

$$\{k+1, k+2, \ldots, k+N\} \subset \mathbb{Z}$$

with

$$k \in \{n, \ldots, m - N - 1\}$$

so that

$$\{S(i), i = k+1, \ldots, k+N\} = \{1, \ldots, N\}$$

and

$$s(i) = 0 \qquad \text{for } \ k+1 \leqslant i \leqslant k+N.$$

Denote by $\lfloor x \rfloor$, $x \in \mathbb{R}$, the largest integer that is not larger than x, then

$$\begin{aligned}
\nu_q(m+1 \longrightarrow n) \leqslant \prod_{i=1}^{\lfloor (m-n)/N \rfloor} & \nu_q\big(\{S(j) : j = m - N(i-1), \ldots, m - Ni + 1\} \neq \{1, \ldots, N\} \\
& \text{or } \exists k \in \{m - N(i-1), \ldots, m - Ni + 1\} \text{ such that } s(k) = 1\big) \\
\leqslant \prod_{i=1}^{\lfloor (m-n)/N \rfloor} & \left(1 - \frac{N!}{N^N} \prod_{k=m-Ni+1}^{m-N(i-1)} (1 - q_k)\right).
\end{aligned} \tag{36}$$

Condition (35) implies that there exists a time $n' = n'(\alpha, \{q_{\beta_n}\}_{n\in\mathbb{N}})$ such that $q_{\beta_n} \leqslant 1 - 1/n^\alpha$ as soon as $n > n'$. Hence, for large n,

$$\begin{aligned}
\nu_q(m+1 \longrightarrow n) &\leqslant \prod_{i=1}^{\lfloor (m-n)/N \rfloor} \left(1 - \frac{N!}{N^N} \prod_{k=m-Ni+1}^{m-N(i-1)} \frac{1}{k^\alpha}\right) \\
&\leqslant \prod_{i=1}^{\lfloor (m-n)/N \rfloor} \left(1 - \frac{N!}{N^N} \frac{1}{(m - N(i-1))^{N\alpha}}\right) \\
&\leqslant \exp\left(-\frac{N!}{N^N} \sum_{i=1}^{\lfloor (m-n)/N \rfloor} \frac{1}{(m - N(i-1))^{N\alpha}}\right)
\end{aligned}$$

since $1 - \epsilon \leqslant \mathrm{e}^{-\epsilon}$ for sufficiently small ϵ.

Using the fact that $(n+3N)^{1-N\alpha} \leqslant n^{1-N\alpha} + (3N)^{1-N\alpha}$, the proof of the lemma can be completed as follows:

$$\begin{aligned}
\nu_q(m+1 \longrightarrow n) &\leqslant \exp\left(-\frac{N!}{N^N} \int_1^{\lfloor (m-n)/N \rfloor + 1} \frac{\mathrm{d}x}{(m - N(x-2))^{N\alpha}}\right) \\
&\leqslant \exp\big(-\lambda((m+N)^{1-N\alpha} - (n+3N)^{1-N\alpha})\big) \\
&\leqslant \exp\big(\lambda(3N)^{1-N\alpha}\big) \exp\big(-\lambda(m+1)^{1-N\alpha}\big) \exp\big(\lambda n^{1-N\alpha}\big)
\end{aligned}$$

with $\lambda = (N!/N^N)(N(1 - \alpha N))^{-1}$. $\square$

5. Proofs

Proof of theorem 1. Using (34) and lemma 1 we see that

$$\begin{aligned}|P^{n_0,n}f(\eta)-P^{n_0,n}f(\eta')| &\leqslant \sum_{\sigma,\sigma'}|f(\sigma)-f(\sigma')|\,\mathrm{Prob}^{n_0,n}[\sigma,\sigma'|\eta,\eta']\\ &\leqslant \sum_{i=1}^{N}\Delta_i f\,\mathrm{Prob}^{n_0,n}[\sigma(i)\neq\sigma'(i)|\eta,\eta']\\ &\leqslant |||f|||\max_i \mathrm{Prob}^{n_0,n}[\sigma(i)\neq\sigma'(i)|\eta,\eta']\\ &\leqslant |||f|||\nu_q(n\longrightarrow n_0)\\ &\leqslant C|||f|||\exp(-\lambda(n^{1-N\alpha}-n_0^{1-N\alpha}))\end{aligned}$$

for some $C=C(N,\alpha,\{q_{\beta_n}\}_{n\in\mathbb{Z}})<\infty$ and $\lambda=\lambda(N,\alpha)>0$. □

Proof of theorem 2. For any $n_0<n$

$$\begin{aligned}\|P^{0,n}f-\mu_\infty(f)\| &= \|P^{0,n_0}[P^{n_0,n}f]-\mu_\infty(f)\|\\ &\leqslant |\mu_\infty(f)-\mu_{\beta_n}(f)|+\|\mu_{\beta_{n_0+1}}(P^{n_0,n}f)-P^{n_0,n}f\|\\ &\quad+|\mu_{\beta_n}(f)-\mu_{\beta_{n_0+1}}(P^{n_0,n}f)|.\end{aligned}\tag{37}$$

By the definition of μ_∞, the first term vanishes when n tends to infinity. Since the dynamics verifies the conditions of theorem 1, also the second term goes to 0 when n grows. This happens with a rate given by (7). To estimate the last term we observe $\mu_\beta(P_\beta f)=\mu_\beta(f)$ for every $f\in\mathcal{B}(\Omega)$. So

$$\begin{aligned}&|\mu_{\beta_n}(f)-\mu_{\beta_{n_0+1}}(P^{n_0,n}f)|\\ &\quad\leqslant |\mu_{\beta_n}(f)-\mu_{\beta_{n_0+2}}(P^{n_0+1,n}f)|+|\mu_{\beta_{n_0+2}}(P^{n_0+1,n}f)-\mu_{\beta_{n_0+1}}(P^{n_0,n}f)|\\ &\quad\leqslant |\mu_{\beta_n}(f)-\mu_{\beta_{n_0+2}}(P^{n_0+1,n}f)|+|\mu_{\beta_{n_0+2}}(P^{n_0+1,n}f)-\mu_{\beta_{n_0+1}}(P^{n_0+1,n}f)|.\end{aligned}$$

We can repeat this procedure until we get that

$$|\mu_{\beta_n}(f)-\mu_{\beta_{n_0}+1}(P^{n_0,n}f)|\leqslant\sum_{k=n_0+1}^{n-1}|\mu_{\beta_{k+1}}(P^{k,n}f)-\mu_{\beta_k}(P^{k,n}f)|.\tag{38}$$

Since μ_β is the unique invariant probability distribution of the time homogeneous dynamics $P_\beta^{0,n}$, we know that for every $g\in\mathcal{B}(\Omega)$ and all values β and β'

$$|\mu_\beta(g)-\mu_{\beta'}(g)|=\lim_{k\uparrow\infty}\|P_\beta^{0,k}g-P_{\beta'}^{0,k}g\|.\tag{39}$$

We first consider the right-hand side for a finite time n and use the coupling (31) to see that for every $\eta\in\Omega$

$$\begin{aligned}&|P_\beta^{0,n}g(\eta)-P_{\beta'}^{0,n}g(\eta)|\\ &\quad\leqslant\sum_{k=0}^{n-1}|P_\beta^{0,n-k}[P_{\beta'}^{n-k,n}g](\eta)-P_\beta^{0,n-k-1}[P_{\beta'}^{n-k-1,n}g](\eta)|\\ &\quad\leqslant\sum_{k=0}^{n-1}\max_\eta|(P_\beta-P_{\beta'})P_{\beta'}^{n-k,n}g(\eta)|\end{aligned}$$

$$\leqslant \sum_{k=0}^{n-1} \max_{\eta} \sum_{\sigma,\sigma'} |P_{\beta'}^{n-k,n} g(\sigma) - P_{\beta'}^{n-k,n} g(\sigma')| \,\mathrm{Prob}_{\beta,\beta'}[\sigma,\sigma'|\eta,\eta]$$

$$\leqslant \frac{1}{N} \sum_{j=1}^{N} \sum_{k=0}^{n-1} (\Delta_j P_{\beta'}^{n-k,n} g) \max_{\eta} p_{\beta,\beta'}^{(j)}(\sigma(j) \neq \sigma'(j)|\eta,\eta)$$

$$\leqslant \max_{j,\eta} \mathrm{var}(p_{\beta}^{(j)}(\eta^{j,\cdot}|\eta), p_{\beta'}^{(j)}(\eta^{j,\cdot}|\eta)) \sum_{k=0}^{n-1} |||P_{\beta'}^{n-k,n} g|||. \tag{40}$$

To estimate the total oscillation of the function $P_{\beta'}^{n-k,n} g$, we need an argument similar to that used in the proof of theorem 1. The only difference is that here we consider a time homogeneous dynamics. Hence

$$|||P_{\beta'}^{n_0,n} g||| = \sum_{i=1}^{N} \max_{\eta} \max_{\zeta} |P_{\beta'}^{n_0,n} g(\eta) - P_{\beta'}^{n_0,n} g(\eta^{i,\zeta})|$$

$$\leqslant N|||g|||\nu_q^{\mathrm{hom}}(n \longrightarrow n_0). \tag{41}$$

Here ν_q^{hom} is the joint probability distribution of the independently distributed Bernoulli random variables $\tilde{s}(k)$, $k \in \mathbb{Z}$ with density

$$q_{\beta'} = \max_{n} \max_{\eta,\eta'} \mathrm{var}(p_{\beta'}^{(n)}(\eta^{n,\cdot}|\eta), p_{\beta'}^{(n)}(\eta'^{n,\cdot}|\eta'))$$

and the variables $S(k)$, $k \in \mathbb{Z}$, as we defined in the previous paragraph. Analogous to (36) this can be bounded from above by

$$N|||g|||\left(1 - \frac{N!}{N^N}(1-q_{\beta'})^N\right)^{\lfloor (n-n_0-1)/N \rfloor}.$$

So

$$\lim_{n\uparrow\infty} \sum_{k=0}^{n-1} |||P_{\beta'}^{n-k,n} g||| \leqslant N|||g||| \frac{(1-(N!/N^N)(1-q_{\beta'})^N)^{-(1+1/N)}}{1-(1-(N!/N^N)(1-q_{\beta'})^N)^{1/N}}$$

$$\leqslant 2\frac{N^{N+1}}{N!}|||g|||(1-q_{\beta'})^{-N} \tag{42}$$

when β' is large enough.

If we combine (39), (40) and (42), we obtain the result that, for sufficiently large n_0, (38) is not larger than

$$2\frac{N^{N+1}}{N!} \sum_{k=n_0+1}^{n-1} (1-q_{\beta_k})^{-N} \max_{j,\eta} \mathrm{var}(p_{\beta_k}^{(j)}(\eta^{j,\cdot}|\eta), p_{\beta_{k+1}}^{(j)}(\eta^{j,\cdot}|\eta))|||P^{k,n} f|||$$

$$\leqslant C\frac{N^{N+2}}{N!}|\Omega_0|\,|||f|||\exp(-\lambda(n^{1-N\alpha})) \sum_{k=n_0+1}^{n-1} k^{\alpha N-\delta} \exp(\lambda(k^{1-N\alpha})) \tag{43}$$

for some $C < \infty$ and $\lambda > 0$.

In the last inequality we applied condition (9) and theorem 1 to estimate $|||P^{k,n} f|||$ and $(1-q_{\beta_n})^N$ and we used the fact that condition (9) allows us to compute the following upper bound:

$$\mathrm{var}(p_{\beta_{k+1}}^{(n)}(\eta^{n,\cdot}|\eta), p_{\beta_k}^{(n)}(\eta^{n,\cdot}|\eta)) = \frac{1}{2} \sum_{\xi\in\Omega_0} \left| p_{\beta_{k+1}}^{(n)}(\eta^{n,\xi}|\eta) - p_{\beta_k}^{(n)}(\eta^{n,\xi}|\eta) \right|$$

$$
\begin{aligned}
&= \frac{1}{2}\sum_{\xi\in\Omega_0}\left|\frac{\mathrm{d}}{\mathrm{d}t}p^{(n)}_{\beta_t}(\eta^{n,\xi}|\eta)\right|_{t=t_k} \\
&\leqslant \tfrac{1}{2}|\Omega_0|k^{-\delta}.
\end{aligned}
$$

for some $k \leqslant t_k \leqslant k+1$, when k is large enough.

For sufficiently large n_0, the sum in (43) is smaller than

$$
\begin{aligned}
&\int_{n_0+1}^{n} (k+1)^{\alpha N-\delta}\exp(\lambda(k+1)^{1-\alpha N}) \\
&\qquad \leqslant \frac{1}{\lambda(1-\alpha N)}(n_0+2)^{2\alpha N-\delta}\left[\exp(\lambda(n+1)^{1-\alpha N}) - \exp(\lambda(n_0+2))^{1-\alpha N}\right]
\end{aligned}
$$

and since $(n+1)^{1-\alpha N} \leqslant n^{1-\alpha N}+1$, this is not larger than

$$
\frac{\mathrm{e}^{\lambda}}{\lambda(1-\alpha N)}(n_0+2)^{2\alpha N-\delta}\exp(\lambda n^{1-\alpha N}).
$$

We insert this in (43) and use this upper bound to estimate (37). Since $\delta > 2\alpha N$, we can conclude that the dynamics is strongly ergodic with asymptotic probability distribution μ_∞ by first taking the limit $n \uparrow \infty$ and then the limit $n_0 \uparrow \infty$. To obtain the rate (10), it suffices to take $n_0 = n^{\chi}$ everywhere in the previous proof, for some $0 < \chi < 1$ sufficiently large. □

Proof of theorem 3.

$$
\|\mu_\infty(f) - \tilde{P}^{0,n}f\| \leqslant \|\mu_\infty(f) - P^{0,n}f\| + \|\tilde{P}^{0,n}f - P^{0,n}f\|.
$$

Since the dynamics $P^{0,n}$ is strongly ergodic with μ_∞ the asymptotic probability distribution, the first term on the right-hand side vanishes when n tends to infinity.

Using the coupling (31) and lemma 1, the second term can be bounded as follows:

$$
\begin{aligned}
&\sum_{k=0}^{n-1}\|\tilde{P}^{0,k}[P^{k,n}f] - \tilde{P}^{0,k+1}[P^{k+1,n}f]\| \\
&\qquad \leqslant \sum_{k=0}^{n-1}\|(P_{\beta_{k+1}} - \tilde{P}_{\tilde{\beta}_{k+1}})P^{k+1,n}f\| \\
&\qquad \leqslant \sum_{k=0}^{n-1}\sum_{\sigma,\sigma'}|P^{k+1,n}f(\sigma) - P^{k+1,n}f(\sigma')|\max_{\eta}\mathrm{Prob}_{\beta_{k+1},\tilde{\beta}_{k+1}}[\sigma,\sigma'|\eta,\eta] \\
&\qquad \leqslant \frac{1}{N}\sum_{k=0}^{n-1}\sum_{j=1}^{N}\Delta_j P^{k+1,n}f\max_{\eta}p^{(j)}_{\beta_{k+1},\tilde{\beta}_{k+1}}[\sigma(j)\neq\sigma'(j)|\eta,\eta] \\
&\qquad \leqslant \sum_{k=0}^{n-1}|||P^{k+1,n}f|||\max_{j,\eta}\mathrm{var}(p^{(j)}_{\beta_{k+1}}(\eta^{j,\cdot}|\eta), \tilde{p}^{(j)}_{\tilde{\beta}_{k+1}}(\eta^{j,\cdot}|\eta)).
\end{aligned} \tag{44}
$$

If we now apply theorem 1 to estimate $|||P^{k+1,n}f|||$ and use condition (11), we see that the last sum is not larger than

$$
C\exp(-\lambda n^{1-N\alpha})\sum_{k=0}^{n-1}(k+1)^{-\delta}\exp(\lambda(k+1)^{1-N\alpha})
$$

for some finite $C < \infty$ and $\lambda > 0$, both independent of n.

Writing the sum as

$$\sum_{k=1}^{\lfloor n^{1/2}\rfloor} k^{-\delta} \exp(\lambda k^{1-N\alpha}) + \sum_{k=\lfloor n^{1/2}\rfloor}^{n-1} k^{-\delta} \exp(\lambda k^{1-N\alpha})$$

we can use the same argument as in the previous proof to see that the upper bound (44) tends to 0 as soon as $\delta > N\alpha$. □

Proof of theorem 4. Consider the energy function (23). Then, when β is large enough, the dynamics with (individual) transition probabilities (15) is attractive. Furthermore, $\mu_\beta \leqslant \mu_{\beta'}$ as soon as $\beta \leqslant \beta'$. So, for large β_1 and for the decreasing function I_{-1}, we see that

$$\begin{aligned}\mu_{\beta_n}(I_{-1}) &= \mu_{\beta_n}(P_{\beta_n} I_{-1})\\ &\leqslant \mu_{\beta_{n-1}}(P_{\beta_n} I_{-1})\\ &= \mu_{\beta_{n-1}}(P_{\beta_{n-1}} P_{\beta_n} I_{-1})\\ &= \mu_{\beta_{n-1}}(P^{n-2,n} I_{-1}).\end{aligned}$$

If we repeat this procedure, we obtain the result that

$$\mu_{\beta_n}(I_{-1}) \leqslant \mu_{\beta_1}(P^{0,n} I_{-1}).$$

In other words,

$$\mu_{\beta_1}(P^{0,n} I_{-1}) \leqslant n^{-\delta}$$

can only hold when

$$\mu_{\beta_n}(I_{-1}) \leqslant n^{-\delta}$$

or when

$$\frac{\mathrm{e}^{-\beta_n U_{\min}}}{1 + \mathrm{e}^{-\beta_n U_{\min}} + (2^N - 2)\mathrm{e}^{-\beta_n U_{\max}}} \leqslant n^{-\delta}.$$

For large n, this inequality can only be satisfied when

$$\beta_n \geqslant \frac{\delta}{2U_{\min}} \log n. \tag{45}$$

This choice for β_n, however, is in contradiction with the assumption that the dynamics is ergodic. Indeed, when (45) is true for large n, then there exist constants $n_0 > 0$ and $C = C(n_0) > 0$ such that for $n > n_0$

$$\begin{aligned}&P^{0,n}\big(\sigma_n = \sigma_{n-1} = \cdots = \sigma_0 | \sigma_0(i) = -1, i = 1, \ldots, N\big)\\ &\qquad = \prod_{i=1}^{n} \frac{1}{1 + (N-1)\mathrm{e}^{-\beta_i (U_{\max} - U_{\min})}}\\ &\qquad \geqslant C \prod_{i=n_0}^{n} \frac{1}{1 + (N-1) i^{-\delta(U_{\max} - U_{\min})/2U_{\min}}} > 0\end{aligned}$$

uniformly in n.

This implies that when the dynamics is started with the initial configuration i.e. the all minus configuration, then, with a strictly positive probability, it will stay in this configuration, and never reach the wished asymptotic probability distribution. Hence theorem 4 holds. □

Appendix

Lemma A1. *The dynamics with transition probabilities (15) and with* $\beta_n = \epsilon \log n$ *satisfies condition (9) for any* $0 < \delta < 1$*, independently of* $\epsilon > 0$.

Proof. Since for every n, η, ζ,

$$\begin{aligned}\left|\frac{\mathrm{d}}{\mathrm{d}t} p_{\beta_t}^{(n)}(\eta^{n,\zeta}|\eta)\right| &= \left|\frac{\mathrm{d}}{\mathrm{d}\beta_t} p_{\beta_t}^{(n)}(\eta^{n,\zeta}|\eta)\frac{\mathrm{d}}{\mathrm{d}t}\beta_t\right| \\ &= \left|\frac{\mathrm{d}}{\mathrm{d}\beta_t} p_{\beta_t}^{(n)}(\eta^{n,\zeta}|\eta)\frac{\epsilon}{t}\right| \\ &\leqslant 2\epsilon U_{\max} t^{-1}\end{aligned}$$

the lemma follows immediately. □

Lemma A2. *When for every* $\lambda > 0$ *there exists a* $C < \infty$ *such that*

$$\frac{F'(\lambda x)}{F'(x)} < C \tag{A1}$$

uniformly in x*, then the dynamics with transition probabilities (26) and with*

$$\beta_n = \frac{1}{U_{\max}} F^{-1}\left(\frac{|\Omega_0|}{n^\epsilon}\right) \tag{A2}$$

satisfy condition (9) for every $0 < \delta < 1$*, uniformly in* $\epsilon > 0$.

Notice that this lemma covers the examples mentioned in paragraph 3, i.e. $F(x) = (1+x)^{-\lambda}$, $\lambda > 0$ and $F(x) = (1+\log(1+x))^{-1}$.

Proof. For every k, η, ζ,

$$\begin{aligned}&\left|\frac{\mathrm{d}}{\mathrm{d}t} p_{\beta_t}^{(k)}(\eta^{k,\zeta}|\eta)\right| \\ &\quad= \left|\frac{\mathrm{d}}{\mathrm{d}\beta_t} p_{\beta_t}^{(k)}(\eta^{k,\zeta}|\eta)\frac{\mathrm{d}}{\mathrm{d}t}\beta_t\right| \\ &\quad= \Bigg|(U(\eta^{k,\zeta})F'(\beta_t U(\eta^{k,\zeta}))\left[\sum_{\zeta\in\Omega_0} F(\beta_t U(\eta^{k,\zeta}))\right]^{-1} \\ &\qquad - F(\beta_t U(\eta^{k,\zeta}))\left[\sum_{\zeta\in\Omega_0} U(\eta^{k,\zeta})F'(\beta_t U(\eta^{k,\zeta}))\right]\left[\sum_{\zeta\in\Omega_0} F(\beta_t U(\eta^{k,\zeta}))\right]^{-2} \\ &\qquad \times \frac{\Omega_0}{U_{\max}}\epsilon t^{-1-\epsilon}\left[F'\left(F^{-1}\left(\frac{|\Omega_0|}{t^\epsilon}\right)\right)\right]^{-1}\Bigg| \\ &\quad\leqslant \epsilon t^{-1-\epsilon}|\Omega_0|(|\Omega_0|+1)\max_\zeta\left|F'(\beta_t U(\eta^{k,\zeta}))[F(\beta_t U_{\max})F'(\beta_t U_{\max})]^{-1}\right| \\ &\quad\leqslant C|\Omega_0|^2\epsilon t^{-1}\end{aligned}$$

for some $C = C(U) < \infty$. In the last step we used (A1) and (A2) to replace $F(\beta_t U_{\max})$ by $|\Omega_0|/n^\epsilon$ and to estimate the ratio $F'(\beta_t U(\eta^{k,\zeta}))/F'(\beta_t U_{\max})$. □

Lemma A3. *If* $\beta_n = \tilde{\beta}_n = (\alpha/2U_{\max})\log n$, *then the condition (11) of theorem 3 is satisfied for the dynamical systems (15) and (29).*

Proof. Let x, y and β be constants such that $\beta > 0$, $y > 0$ and $e^{-\beta y}\beta x > -1$. Consider then

$$\begin{aligned}
&|e^{-\beta x} - (1 + e^{-\beta y}\beta x)^{-e^{\beta y}}| \\
&\qquad = e^{-\beta x}|1 - e^{\beta x}(1 + e^{-\beta y}\beta x)^{-e^{\beta y}}| \\
&\qquad = e^{-\beta x}|1 - \exp(\beta x - e^{\beta y}\log(1 + e^{-\beta y}\beta x)| \\
&\qquad = e^{-\beta x}\left|1 - \exp\left(\tfrac{1}{2}(\beta x)^2 e^{-\beta y} - \tfrac{1}{3}(\beta x)^3 e^{-2\beta y} + \cdots\right)\right| \\
&\qquad \leqslant \tfrac{1}{2}(\beta x)^2 e^{-\beta x} e^{-\beta y} + e^{-\beta x} o\left(\tfrac{1}{3}(\beta x)^3 e^{-2\beta y}\right).
\end{aligned}$$

So

$$\begin{aligned}
(1 + e^{-\beta y}\beta x)^{-e^{\beta y}} &= e^{-\beta x} + \theta e^{-\beta x}\left(\tfrac{1}{2}(\beta x)^2 e^{-\beta y} + o(\tfrac{1}{3}(\beta x)^3 e^{-2\beta y})\right) \\
&\equiv e^{-\beta x} + R(x)
\end{aligned}$$

in which $\theta = \theta(x, y, \beta)$ can be -1 or $+1$.

Next, we consider a set of M numbers x_i, $i = 1, 2, \ldots, M$, such that $x_1 = 0$ and for every i, $e^{-\beta y}\beta x_i > -1$. Then

$$\begin{aligned}
&\sum_{i=1}^{M}\left|\frac{e^{-\beta x_i}}{\sum_{i=1}^{M} e^{-\beta x_i}} - \frac{(1 + e^{-\beta y}\beta x_i)^{-e^{\beta y}}}{\sum_{i=1}^{M}(1 + e^{-\beta y}\beta x_i)^{-e^{\beta y}}}\right| \\
&\qquad = \sum_{i=1}^{M}\left|\frac{e^{-\beta x_i}}{\sum_{i=1}^{M} e^{-\beta x_i}} - \frac{e^{-\beta x_i} + R(x_i)}{\sum_{i=1}^{M} e^{-\beta x_i} + R(x_i)}\right| \\
&\qquad \leqslant \sum_{i=1}^{M}\left|\frac{e^{-\beta x_i}}{\sum_{i=1}^{M} e^{-\beta x_i}} - \frac{e^{-\beta x_i}}{\sum_{i=1}^{M} e^{-\beta x_i}}\left(1 - \frac{\sum_{i=1}^{M} R(x_i)}{\sum_{i=1}^{M} e^{-\beta x_i}} - \left(\frac{\sum_{i=1}^{M} R(x_i)}{\sum_{i=1}^{M} e^{-\beta x_i}}\right)^2 - \cdots\right)\right| \\
&\qquad\quad + \sum_{i=1}^{M}\left|\frac{R(x_i)}{\sum_{i=1}^{M} e^{-\beta x_i} + R(x_i)}\right|.
\end{aligned}$$

We used the fact that, since $x_1 = 0$, we know that $\sum_{i=1}^{M} e^{-\beta x_i} > \sum_{i=1}^{M} R_i$ as soon as β is large enough. For the same reason $\sum_i e^{-\beta x_i} \geqslant 1$ and $\sum_i (e^{-\beta x_i} + R(x_i)) \geqslant \frac{1}{2}$. Hence, when β is large, the above expression is smaller than

$$C \max_i \left\{\tfrac{1}{2}(\beta x_i)^2 e^{-\beta x_i} e^{-\beta y} + o\left(\tfrac{1}{3} e^{-\beta x_i}(\beta x_i)^3 e^{-2\beta y}\right)\right\}$$

for some constant $C < \infty$, independent of x_i, $i = 1, 2, \ldots M$, y, and β.

To prove the lemma we replace the index set $\{i, i = 1, \ldots, M\}$ by the set Ω_0, the numbers x_i, $i = 1, 2, \ldots, M$,, by the numbers $(U(\eta^{k,\zeta}) - U(\eta))$, $\zeta \in \Omega_0$, for k fixed, and β by β_n. If we take $y > 3NU_{\max}$, we obtain an upper bound for

$$\mathrm{var}(p_{\beta_n}^{(k)}(\eta^{k,\cdot}|\eta), \tilde{p}_{\tilde{\beta}_n}^{(k)}(\eta^{k,\cdot}|\eta))$$

which decreases faster than $n^{-\alpha N}$, independently of η and k. □

References

[1] Aarts E and Korst J 1989 *Simulated Annealing and Boltzmann Machines. A Stochastic Approach to Combinatorial Optimization and Neural Computing* (Chichester: Wiley)

[2] Deuschel J D and Mazza C 1994 L^2-convergence of time non-homogeneous Markov processes: I. Spectral estimates *Annal. Appl. Prob.* **4** 1012–56

[3] Geman S and Geman D 1984 Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images *IEEE Trans. Pattern Anal. Machine Intell.* **6** 721–41

[4] Gidas B 1985 Non-stationary Markov chains and convergence of the simulated annealing algorithm *J. Stat. Phys.* **39** 73–131

[5] Holley R, Kusuoka S and Stroock D 1989 Asymptotics of the spectral gap with applications to the theory of simulated annealing *J. Funct. Anal.* **83** 333–47

[6] Holley R and Stroock D 1988 Simulated annealing via Sobolev inequalities *Commun. Math. Phys.* **115** 553–69

[7] Inoue J and Nishimori H 1997 Convergence of simulated annealing by the generalized transition probability *Preprint*

[8] Penna T J S 1995 Travelling salesman problem and Tsallis statistics *Phys. Rev.* E **51** R1–3

[9] Tsallis C and Stariolo D A 1996 Generalized simulated annealing *Physica* A **233** 395–406