Simulated Annealing with Extended Neighbourhood[1]

Xin Yao
Computer Sciences Laboratory
Research School of Physical Sciences and Engineering
The Australian National University
GPO Box 4, Canberra, ACT 2601
AUSTRALIA
e-mail: xin@cslab.anu.edu.au
Fax: (+61 6)/(06)249 1884

[1]Published in **International Journal of Computer Mathematics, 40**:169–189, 1991.

Abstract

Simulated Annealing (SA) is a powerful stochastic search method applicable to a wide range of problems for which little prior knowledge is available. It can produce very high quality solutions for hard combinatorial optimization problems. However, the computation time required by SA is very large. Various methods have been proposed to reduce the computation time, but they mainly deal with the careful tuning of SA's control parameters. This paper first analyzes the impact of SA's neighbourhood on SA's performance and shows that SA with a larger neighbourhood is better than SA with a smaller one. The paper also gives a general model of SA, which has both dynamic generation probability and acceptance probability, and proves its convergence. All variants of SA can be unified under such a generalization. Finally, a method of extending SA's neighbourhood is proposed, which uses a discrete approximation to some continuous probability function as the generation function in SA, and several important corollaries of the general model are given.

KEYWORDS — Simulated Annealing, Stochastic Search, Algorithm Analysis, Combinatorial Optimization.

1 Introduction

SA is a powerful stochastic search method applicable to a wide range of problems for which little prior knowledge is available. It has been widely used in computer-aided VLSI design [1], combinatorial optimization [2, 3, 4], neural network training [5, 6], image processing [7, 8], code design [9], function optimization [10], etc. The most important reason for SA's popularity is its ability to produce high quality solutions for hard problems.

The basic idea of SA comes from condensed matter physics. It is well known in condensed matter physics that a good way to find minimum energy states, called ground states, of complex systems such as solids is to use the annealing technique, in which the system (solid) is first heated to some high temperature, and then slowly cooled down. The system (solid) will reach a ground state if the cooling rate around the freezing point of the system is sufficiently slow. This process can be simulated on computers via abstract models, such as systems of interacting particles with many degrees of freedom.

At each step of the simulation, a new state of the system is generated from the current state by giving a random displacement to a randomly selected particle. The new state will be accepted as the current one if the energy of the new state is no greater than that of the current state, otherwise, it will only be accepted with probability

$$\exp\left(-\frac{E_{new\_state} - E_{current\_state}}{T}\right)$$

where E stands for the energy of the system and T is the temperature. This step can be repeated as many as necessary with a slow decrease of temperature in order to find a minimum energy state. Of course, only finite steps are taken in practical simulations. This simulation procedure was proposed by N. Metropolis et al. [11] and is called the Metropolis procedure.

The analogy between finding minimum energy states in a physical system and finding minimum cost configurations in a combinatorial optimization problem was independently observed by S. Kirkpatrick et al. [2] and V. Černy [12]. They proposed a new algorithm for solving combinatorial optimization problems, SA, based on the Metropolis procedure. Since then, research on SA has been booming in both theory and applications. However, SA often suffers from the long computation time, sometimes even severely. Various methods have been proposed to cope with this problem, but they mainly deal with careful tuning of SA's control parameters and the resulting improvement of SA's performance is rather limited. This paper analyzes the impact of the neighbourhood on SA's performance and then gives a general model of SA with both dynamic generation and acceptance probabilities, which unifies the variants of SA under the same general model.

Section 2 briefly reviews SA. Section 3 analyzes the impact of the neighbourhood on SA's performance and shows an important result that the performance of SA with large neighbourhoods is better than that with small neighbourhoods. Section 4 gives a general model of SA as well as its convergence conditions, which removes the restriction that the generation probability of SA be independent of the control parameter, temperature. This generalization not only unifies various SA's with different parameters and functions, but can also derive many useful results which reduce the computation time of SA. Section 5 is devoted to the SA with dynamic generation and acceptance probabilities. A method of using a discrete approximation of some continuous probability function as the generation function is proposed and a set of corollaries is given. Finally, some conclusions are drawn in Section 6.

2 A Brief Review of Simulated Annealing

The combinatorial optimization problems considered herein can be described as: Given a finite configuration space $S = \{X \mid X = (x_1, x_2, \cdots, x_m)\}$, where m is called the dimension of the space, and a cost function $c : S \longrightarrow R^1$, we want to find an optimum configuration $X \in S$, such that $\forall Y \in S, c_X \leq c_Y$. (We only deal with minimization problems here. Maximization problems can be treated similarly.) The neighbourhood system of a combinatorial optimization problem can be defined as

$$\mathcal{N} = \{N_X \mid X \in S, N_X \subseteq S\} \tag{1}$$

where $X \notin N_X$ and $Y \in N_X \Longleftrightarrow X \in N_Y$. Obviously, elements in N_X represent the configurations next to X in the configuration space. They are called neighbours of X.

The standard SA (SSA) [2] starts with an initial configuration generated at random. At each step, it selects the next configuration Y from the neighbourhood N_X of the current configuration X. The next configuration will be accepted as the current one if its cost is no greater than that of the current configuration, otherwise it will only be accepted with probability

$$\exp\left(-\frac{c_Y - c_X}{T}\right)$$

This procedure is repeated with a slow decrease of the control parameter T, called temperature, until a sufficiently good solution has been found. SSA can be summarized as Figure 1.

```
select initial configuration X at random;
select initial temperature T;
REPEAT
   REPEAT
      randomly select Y from N_X with uniform distribution;
      IF c_Y ≤ c_X
         THEN accept Y as the new configuration
         ELSE accept Y as the new configuration with probability
              exp (− (c_Y − c_X) / T)
   UNTIL 'inner-loop stop criterion' satisfied;
   decrease temperature T
UNTIL 'outer-loop stop criterion' satisfied
```

Figure 1: Main steps of standard simulated annealing.

The performance of SSA can be analyzed by nonstationary Markov chain theory [13]. Suppose the probability of generating the next configuration Y from the current one X is g_{XY}, which is independent of temperature T, and the probability of accepting Y as the new current configuration is $a_{XY}(T)$. They satisfy (2) and (3) respectively.

$$g_{XY} = \begin{cases} \frac{1}{|N_X|}, & X \in S, Y \in N_X \\ 0, & X \in S, Y \notin N_X \end{cases} \tag{2}$$

$$a_{XY}(T) = \min\left\{1, \exp\left(-\frac{c_Y - c_X}{T}\right)\right\} \tag{3}$$

The one-step transition probabilities of the nonstationary Markov chain associated with SSA are

$$p_{XY}(T) = \begin{cases} \frac{1}{|N_X|} \min\left\{1, \exp\left(-\frac{c_Y - c_X}{T}\right)\right\}, & X \in S, Y \in N_X \\ 0, & X \in S, Y \notin N_X, X \neq Y \\ 1 - \sum_{Z \in N_X} \frac{1}{|N_X|} \min\left\{1, \exp\left(-\frac{c_Z - c_X}{T}\right)\right\}, & X \in S, Y \notin N_X, X = Y \end{cases}$$

The convergence of SSA has been proved by several researchers [14, 15, 16]. Their results can be formulated as follows:

Given that SSA's generation and acceptance probabilities satisfy (2) and (3) respectively, then SSA converges to global minima, i.e.,

$$\lim_{n \to \infty} Prob\{X \in S^*\} = 1 \tag{4}$$

$$S^* = \{X \mid X \in S, c_X \leq c_Y \forall Y \in S\} \tag{5}$$

if the rate of temperature decrease is

$$T_n = \frac{h}{\ln(n+1)}, \qquad n = 1, 2, \cdots \tag{6}$$

where h is a problem dependent constant and T_n stands for the temperature at time n.

A better result, obtained in [17], can be stated as

Given that SSA's generation and acceptance probabilities satisfy (2) and (3) respectively, then SSA converges to global minima if the rate of temperature decrease is

$$T_n = \frac{\lambda}{\sum_{i=1}^{N} \left((\ln)^i \left(\left\lfloor \frac{n}{t} \right\rfloor + n_0 \right) \right)} \tag{7}$$

where

$$\lambda \geq \left(\Delta c^+ \right) r \tag{8}$$

$$\Delta c^+ = \max_{X \in S, Y \in N_X} \{ |c_Y - c_X| \} \tag{9}$$

$$(\ln)^i (f(x)) = \ln((\ln)^{i-1}(f(x))), \; i > 0 \tag{10}$$

$$(\ln)^0 (f(x)) = f(x) \tag{11}$$

$$n_0 > 1 \tag{12}$$

$$N \; (N \geq 1) \; \text{ is an arbitrarily large integer} \tag{13}$$

$$t \; (t < \infty) \; \text{ is the inner} - \text{loop iterations of SA} \tag{14}$$

In practice, the cooling rates of (6) or (7) are too slow to be followed exactly. A frequently used method is to set the cooling rate as

$$T_{n+1} = \rho T_n \tag{15}$$

where ρ is a constant normally selected between 0.80 and 0.99. Although this kind of SSA has been successfully used in many applications [2, 18, 4], it needs a large amount of computation time. Various efforts made to reduce this time can be divided into two classes; one is parallel processing and the other is careful tuning of the SSA's control parameters, such as initial temperature, cooling rate, inner-loop stop criterion and outer-loop stop criterion. Different acceptance probabilities have also been tried [19, 20]. However, little work has done on the impact of SA's generation probabilities, which are directly related to SA's neighbourhood system as well as its search behaviour in the configuration space, on SA's performance. The following sections will analyze this impact and give a general model of SA.

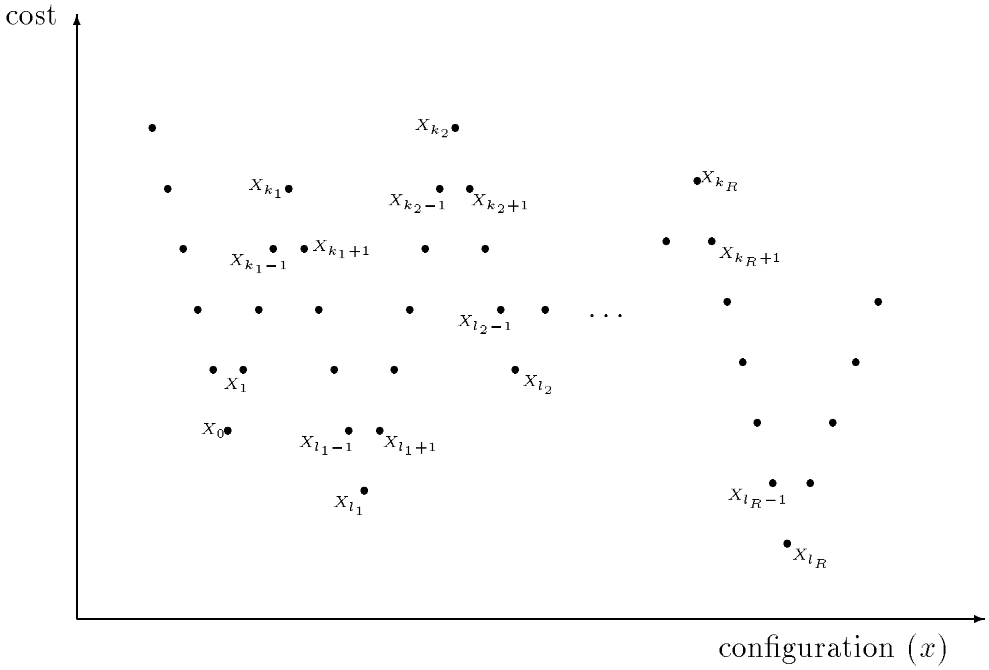

Figure 2: Configuration space of a one-dimensional optimization problem.

3 Analysis of the Neighbourhood System of Simulated Annealing

It is well known that a greedy algorithm, like hill climbing, will stick in a local minimum if the algorithm has reached it, while a stochastic algorithm such as SA can still escape from a local minimum and move to a better region in the configuration space. Obviously, the probability of escaping from a local minimum is a very important quantity affecting SA's performance. A larger value will reduce the time SA stays in a local minimum and speed up its movement towards global minima. This section analyzes the above probability and discusses the impact of the neighbourhood system on it.

In SSA, a neighbour Y of a configuration X in N_X is different from X in only one element, as in the Metropolis procedure. That is, if $X = (x_1, x_2, \cdots, x_i, \cdots, x_m)$, and $Y = (y_1, y_2, \cdots, y_i, \cdots, y_m)$, then $y_i \neq x_i,\ y_j = x_j,\ i \neq j,\ 1 \leq i \leq m,\ j = 1, 2, \cdots, m$. In the following discussion, we say that the 'Hamming' distance between two configurations is μ and they are μ-steps reachable from each other if there are exactly μ elements different between them. For simplicity, we first consider one-dimensional combinatorial optimization problems, i.e., the case of $X = (x)$, which can be represented as in Figure 2.

In the one-dimensional case, $|N_{X_i}| = 2$ for all $X_i \in S$. Hence we can get, from (2) and (3), that the probability of moving from configuration X_0 to the local maximum X_{k_1} at temperature T using no greater than k_1 steps is

$$\begin{aligned}
& Prob\,(X_0, X_{k_1}, k_1, T, |N_{X_i}| = 2) \\
= & \prod_{i=1}^{k_1} \left(\frac{1}{2}\right) \exp\left(-\frac{c_{X_i} - c_{X_{i-1}}}{T}\right) \\
= & \left(\frac{1}{2}\right)^{k_1} \exp\left(-\frac{c_{X_{k_1}} - c_{X_0}}{T}\right) \qquad (16)
\end{aligned}$$

and the probability of moving from X_{k_1} to the local minimum X_{l_1} at temperature T using no greater than $l_1 - k_1$ steps is

$$Prob\,(X_{k_1}, X_{l_1}, l_1 - k_1, T, |N_{X_i}| = 2) = \left(\frac{1}{2}\right)^{l_1 - k_1} \qquad (17)$$

Similarly, we can obtain the probability of finding the global minimum X_{l_R} from configuration X_0 using no greater than l_R steps:

$$Prob\,(X_0, X_{l_R}, l_R, T, |N_{X_i}| = 2) = \left(\frac{1}{2}\right)^{l_R} \exp\left(-\frac{1}{T}\sum_{i=1}^{R}\left(c_{X_{k_i}} - c_{X_{l_{i-1}}}\right)\right) \qquad (18)$$

where $l_0 = 0$.

It is clear from (18) that there are three factors which affect the probability of reaching a global minimum from a non-global minimum: (1) the total height of the hills which have to be surmounted; (2) the temperature; and (3) the total probability of generating the global minimum from the non-global minimum, i.e.,

$$\left(\frac{1}{2}\right)^{l_R} \qquad (19)$$

The heights of the hills in the configuration space are solely decided by the problem. Hence, they are unlikely to be altered for a particular problem. However, we can certainly raise the temperature to make escaping from the local minimum easier. This is exactly what SSA does in the early stages of the annealing. Another way to accelerate SSA's escape from local minima is to make (19) larger. This goal can be achieved by extending the neighbourhood of SSA's search.

Suppose we have extended the neighbourhoods from those of Hamming distance 1 to those of Hamming distance ≤ 2, that is, instead of having $N_{X_i} = \{X_{i-1}, X_{i+1}\}$, we have $N_{X_i} = \{X_{i-2}, X_{i-1}, X_{i+1}, X_{i+2}\}$. Let $Prob\,(X_0, X_{k_1}, \nu, \kappa, T, |N_{X_i}| = 4)$ be the probability of arriving at X_{k_1} from X_0 for the κ-th time in exact ν steps. Then we get $|N_{X_i}| = 4$, and

$$Prob\,(X_0, X_{k_1}, k_1, T, |N_{X_i}| = 4)$$

$$
\begin{aligned}
&= \sum_{\nu=1}^{k_1} \sum_{\kappa=1}^{\nu} Prob\,(X_0, X_{k_1}, \nu, \kappa, T, |N_{X_i}| = 4) \\
&= \sum_{\nu=\frac{k_1}{2}}^{k_1} \sum_{\kappa=1}^{\nu} Prob\,(X_0, X_{k_1}, \nu, \kappa, T, |N_{X_i}| = 4) \\
&> \sum_{\nu=\frac{k_1}{2}}^{k_1} Prob\,(X_0, X_{k_1}, \nu, 1, T, |N_{X_i}| = 4) \\
&= \left(\frac{1}{4}\right)^{\frac{k_1}{2}} \exp\left(-\frac{c_{X_{k_1}} - c_{X_0}}{T}\right) \\
&\quad + \left(\frac{k_1}{2} + \frac{k_1 - 2}{2} + \cdots + 2 + 1\right) \left(\frac{1}{4}\right)^{\frac{k_1}{2}+1} \exp\left(-\frac{c_{X_{k_1}} - c_{X_0}}{T}\right) \\
&\quad \vdots \\
&\quad + [(k_1 - 3) + (k_1 - 4) + \cdots + 2 + 1] \left(\frac{1}{4}\right)^{k_1 - 2} \exp\left(-\frac{c_{X_{k_1}} - c_{X_0}}{T}\right) \\
&\quad + (k_1 - 1) \left(\frac{1}{4}\right)^{k_1 - 1} \exp\left(-\frac{c_{X_{k_1}} - c_{X_0}}{T}\right) \\
&\quad + \left(\frac{1}{4}\right)^{k_1} \exp\left(-\frac{c_{X_{k_1}} - c_{X_0}}{T}\right) \\
&> \left(\frac{1}{4}\right)^{\frac{k_1}{2}} \exp\left(-\frac{c_{X_{k_1}} - c_{X_0}}{T}\right) && (20) \\
&= \left(\frac{1}{2}\right)^{k_1} \exp\left(-\frac{c_{X_{k_1}} - c_{X_0}}{T}\right) && (21) \\
&= Prob\,(X_0, X_{k_1}, k_1, T, |N_{X_i}| = 2) && (22)
\end{aligned}
$$

That is,

$$
Prob\,(X_0, X_{k_1}, k_1, T, |N_{X_i}| = 4) > Prob\,(X_0, X_{k_1}, k_1, T, |N_{X_i}| = 2) \qquad (23)
$$

In the above analysis, we have assumed that k_1 (≥ 2) is even, the odd case can similarly be shown. It is obvious from the above analysis that the probability of moving from X_0 to X_{k_1} is controlled by the total probability of generating X_{k_1} from X_0; the exponential part is the same. The total generation probability is independent of the starting configuration X_0 and the ending configuration X_{k_1}; it only depends on the total Hamming distance between them. Let $G\,(k, \nu, \kappa, T, |N_{X_i}| = 4)$ be the total probability of generating the ending configuration which is k distance from the starting configuration for the κ-th ($1 \leq \kappa \leq \nu$) time in exact ν ($\frac{k}{2} \leq \nu \leq k$) steps. Then the following recurrence holds.

$$
G\,(k, \nu, 1, T, |N_{X_i}| = 4) = \sum_{j=1}^{\nu-(k-\nu)+1} \left(\frac{1}{4}\right)^{j} G\,(k, \nu - 1, 1, T, |N_{X_i}| = 4) \qquad (24)
$$

Notice that

$$G\left(k,k,1,T,|N_{X_i}|=4\right)=\left(\frac{1}{4}\right)^k \tag{25}$$

Using a similar analysis to (16) – (25) we get the following result:

$$Prob\left(X_0,X_{l_R},l_R,T,|N_{X_i}|=4\right)>Prob\left(X_0,X_{l_R},l_R,T,|N_{X_i}|=2\right) \tag{26}$$

where $l_R \geq 2$.

(26) indicates that SA with large neighbourhoods is easier to escape from a local minimum than that with smaller neighbourhoods. In general, if we extend neighbourhoods from those of Hamming distance 1 to those of Hamming distance $\leq m$ $(m \geq 2)$, we can get

$$Prob\left(X_0,X_l,l,T,|N_{X_i}|=2m\right)>Prob\left(X_0,X_l,l,T,|N_{X_i}|=2\right) \tag{27}$$

where $l \geq m$.

Although the above conclusion is derived for the case of one-dimensional problems, it also holds for multi-dimensional problems. We can summarize this result in the following theorem.

Theorem 3.1 *SA with larger neighbourhoods as extended above has greater probability of arriving at a global optimum than SSA has if the other conditions, i.e., the initial configuration, initial temperature and temperature decreasing rate, are the same.*

<u>*Proof:*</u> Suppose SSA's neighbourhood is N_{X_SSA}, i.e., those of Hamming distance 1, and we extend it to N_{X_m}, i.e., those of Hamming distance $\leq m$ $(m \geq 2)$, in an n-dimensional configuration space. Because of the symmetry of the neighbour relationship, we have, for $m \geq 2$,

$$|N_{X\_m}| \leq \left(|N_{X\_SSA}|\right)^m \tag{28}$$

For each path leading from a non-global minimum X_0 to a global minimum X_{l_R} in the n-dimensional configuration space, the probability of moving from X_0 to X_{l_R} along this path using no greater than k steps can be represented as

$$\begin{aligned} & Prob\left(X_0,X_{l_R},k,T,|N_{X_i}|,0\leq i\leq l_R\right) \\ = \; & G\left(X_0,X_{l_R},k,T,|N_{X_i}|,0\leq i\leq l_R\right)\cdot A\left(X_0,X_{l_R},k,T,|N_{X_i}|,0\leq i\leq l_R\right) \end{aligned} \tag{29}$$

For SSA we have

$$G\left(X_0,X_{l_R},l_R,T,|N_{X\_SSA}|,0\leq i\leq l_R\right) = \left(\frac{1}{|N_{X\_SSA}|}\right)^{l_R} \tag{30}$$

$$A\left(X_0,X_{l_R},l_R,T,|N_{X\_SSA}|,0\leq i\leq l_R\right) = \exp\left(-\frac{1}{T}\sum_{i=1}^{R}\left(c_{X_{k_i}}-c_{X_{l_{i-1}}}\right)\right) \tag{31}$$

For SA with neighbourhood N_{X_m}, as indicated before, function A is decided by the problem and is the same as (31),

$$A\left(X_0, X_{l_R}, l_R, T, |N_{X\_m}|, 0 \le i \le l_R\right) = \exp\left(-\frac{1}{T}\sum_{i=1}^{R}\left(c_{X_{k_i}} - c_{X_{l_{i-1}}}\right)\right) \tag{32}$$

while from (28) and with analyses similar to before, we can get

$$\begin{aligned} & G\left(X_0, X_{l_R}, l_R, T, |N_{X\_m}|, 0 \le i \le l_R\right) \\ > & \left(\frac{1}{|N_{X\_m}|}\right)^{\frac{l_R}{m}} \\ \ge & \left(\frac{1}{(|N_{X\_SSA}|)^m}\right)^{\frac{l_R}{m}} \\ = & \left(\frac{1}{|N_{X\_SSA}|}\right)^{l_R} \end{aligned} \tag{33}$$

where

$$l_R \ge m \tag{34}$$

Hence we get the final result: For $l_R \ge m$,

$$\begin{aligned} & Prob\left(X_0, X_{l_R}, l_R, T, |N_{X\_m}|, X_i \in S, 0 \le i \le l_R\right) \\ > & Prob\left(X_0, X_{l_R}, l_R, T, |N_{X\_SSA}|, X_i \in S, 0 \le i \le l_R\right) \end{aligned}$$

Since we have the above inequality for each path leading from a non-global minimum X_0 to a global minimum X_{l_R} in the n-dimensional configuration space, it is easy to show that the total probability of SA with neighbourhood $|N_{X\_m}|$ of moving from X_0 to X_{l_R} by various paths is greater than that of SSA with neighbourhood $|N_{X\_SSA}|$.

Q.E.D.

It seems from Theorem 3.1 that the larger the neighbourhood is, the better. But there is a precondition to this assertion, i.e., condition (34), which says that the assertion is true if the global minimum is far enough from the current configuration. Fortunately, the probability that a randomly generated initial configuration is very close to a global minimum is quite small in practice. It is certainly worth extending SA's neighbourhood in the early stages of the annealing search. However, with the progress of search, the current configuration is likely to move closer and closer towards a global minimum. So it is necessary to restrict SA's neighbourhood in the later stages of the annealing search. Section 5 will describe an SA with dynamic generation and acceptance probabilities, which can gradually reduce its neighbourhood with the decrease of temperature.

$X := X_0$; { X_0 is the initial configuration. }
$n := 0$; { X is the current configuration. }
$T_n := T_0$; { T_0 is the initial temperature }
REPEAT
  REPEAT
    $Y := generate(X, T_n)$;
    IF $accept(X, Y, T_n)$ THEN $X := Y$
  UNTIL 'inner-loop stop criterion' satisfied;
  $T_{n+1} := update\,(T_n)$;
  $n := n + 1$
UNTIL 'outer-loop stop criterion' satisfied

Figure 3: General simulated annealing.

4 A General Model of Simulated Annealing

Section 3 has shown the benefit of extending SA's neighbourhood, but in doing this can we not only reduce the computation time but also guarantee the convergence of SA? Instead of assuming that SA's generation probability is independent of temperature, we make no restrictions on SA's generation and acceptance probabilities in this section and analyze the convergence conditions of SA.

A general SA can be described by Figure 3. Function $generate\,(X, T_n)$ is described by the generation probability $g_{XY}\,(T_n)$, which is the probability of generating configuration Y from configuration X at temperature T_n. Function $accept\,(X, Y, T_n)$ is described by the acceptance probability $a_{XY}\,(T_n)$, which is the probability of accepting configuration Y after it has been generated at temperature T_n. Function $update\,(T_n)$ represents the rate of the temperature decrease. The convergence of SA is fully determined by the above three functions, although parameters such as initial temperature, initial configuration, inner-loop stop criterion, and outer-loop stop criterion can sometimes have significant practical effects on the finite-time behaviour of SA.

4.1 Preliminaries

Before looking at the conditions which make SA, shown in Figure 3, converge to global minima, we first introduce some definitions and theorems. The results presented in this subsection can be found in [21, 13, 22].

Definition 4.1 *A class $F \subset C^1$ of functions defined on $(0,1]$ is a closed class of asymptotically monotone functions (CAM) if*

(1) $f \in F \implies f' \in F$ and $-f \in F$;

(2) $f, g \in F \implies (f+g)$ and $(f \cdot g) \in F$; and

(3) all $f \in F$ change signs finitely often on $(0,1]$.

Definition 4.2 *A class F of functions defined on $(0,1]$ is a rationally closed class of bounded variation (RCBV) if*

(1) $f \in F \implies f$ is of bounded variation on $(0,1]$;

(2) $f \in F \implies -f \in F$;

(3) $f, g \in F \implies (f+g)$ and $(f \cdot g) \in F$; and

(4) $f, g \in F$ with (f/g) bounded on $(0,1] \implies f/g$ is of bounded variation.

Definition 4.3 *A real-valued function f is an exponential sum in $1/x$ if it is of the form $\sum_{j=1}^{n} Q_j(1/x)\, e^{\lambda_j/x}$, with λ_j a given real number and $Q_j(\bullet)$ a given polynomial $(j = 1, 2, \cdots, n)$.*

Definition 4.4 *A real-valued function f is an exponential rational in $1/x$ if it is the ratio of two exponential sums in $1/x$.*

Proposition 4.1 *The following classes of functions defined on $(0,1]$ are RCBV and CAM:*

(1) the polynomials (RCBV) and (CAM);

(2) the rational functions (RCBV) and (CAM);

(3) the piecewise rationals (RCBV), and

(4) the exponential rationals (CAM).

Definition 4.5 *Let $\{u(n)\}_{n=1}^{\infty}$ be a sequence with $u(n) \in R^m$ for some $m \geq 1$. The (vector) function $\vec{u}(v) : (0,1] \to R^m$ is an extension of the sequence if $\vec{u}(v_n) = u(n)$ for some sequence $\{v_n\}_{n=1}^{\infty}$, with $\lim_{n\to\infty} v_n = 0$.*

Definition 4.6 *A nonstationary Markov chain $\{P(n)\}_{n=1}^{\infty}$ is said to have the regular extension $\overline{P}(\bullet)$ if a real number $v_0 > 0$ exists such that the collection of subchains of $\overline{P}(v)$ is identical for all $v < v_0$.*

Proposition 4.2 *A nonstationary Markov chain $\{P(n)\}_{n=1}^{\infty}$ is said to have the regular extension $\overline{P}(\bullet)$ if a real number $v_0 > 0$ exists such that $\left\{(i,j) \mid \overline{p}_{ij}(v) > 0\right\}$ is identical for all $v < v_0$.*

Theorem 4.1 *Let $\{P(n)\}_{n=1}^{\infty}$ be a weakly ergodic nonstationary Markov chain and $\overline{P}(v)$ a regular extension such that all entry functions $\overline{p}_{ij}(v)$ $(1 \leq i, j \leq N)$ belong to CAM or RCBV, then $\{P(n)\}_{n=1}^{\infty}$ is strongly ergodic. Moreover, for sufficiently large n, each $P(n)$ has a unique quasi-stationary state distribution $\pi(n)$ with $\lim_{n\to\infty} \pi(n) = \pi^*$ and $\lim_{n\to\infty} p_{ij}(m,n) = \pi_j^*$ for all $1 \leq i, j \leq N$ and $m \geq 1$.*

Theorem 4.2 *Let $\{X_n\}$ be a nonstationary Markov chain with transition matrices $\{P_n\}_{n=1}^{\infty}$. The chain $\{X_n\}$ is weakly ergodic if and only if there exists a subdivision of $P_1 \cdot P_2 \cdot P_3 \cdots$ into blocks of matrices $[P_1 \cdot P_2 \cdots P_{n_1}] \cdot [P_{n_1+1} \cdot P_{n_1+2} \cdots P_{n_2}] \cdots \left[P_{n_j+1} \cdot P_{n_j+2} \cdots P_{n_{j+1}}\right] \cdots$ such that*

$$\sum_{j=0}^{\infty} \alpha\left(P\left(n_j, n_{j+1}\right)\right) = \infty$$

where $n_0 = 0$.

4.2 Convergence of the general simulated annealing

This subsection considers the conditions under which the general SA converges. For simplicity, we abbreviate generation probability $g_{XY}(T_n)$ as $g_{XY}(n)$ and acceptance probability $a_{XY}(T_n)$ as $a_{XY}(n)$. The method used to prove the convergence of SA is straightforward. We first prove the weak ergodicity of the nonstationary Markov chain associated with the general SA, and then prove the strong ergodicity and the convergence. Denote

$$g^-(n) = \min_{X\in S, Y\in N_X} \{g_{XY}(n)\} \tag{35}$$

$$a^-(n) = \min_{X\in S, Y\in N_X} \{a_{XY}(n)\} \tag{36}$$

The following theorem pertains to the weak ergodicity of the nonstationary Markov chain associated with SA.

Theorem 4.3 *Suppose the generation probability $g_{XY}(n)$, $X \in S, Y \in N_X$ of SA satisfies:*

(g1) $g_{XY}(n)$ is a non-increasing function of n for sufficiently large n and $\lim_{n\to\infty} g_{XY}(n)$ exists;

and the acceptance probability $a_{XY}(n)$, $X, Y \in S$ satisfies:

(a1) For $c_X < c_Y$, $a_{XY}(n)$ is a decreasing function of n for sufficiently large n and $\lim_{n\to\infty} a_{XY}(n) = 0$;

(a2) For $c_X \geq c_Y$, $a_{XY}(n) = 1$, $n = 0, 1, 2, \cdots$;

then the nonstationary Markov chain associated with SA is weakly ergodic if

$$\sum_{k=n_1}^{\infty} \left[g^{-}(kr)a^{-}(kr)\right]^{r} = \infty \tag{37}$$

where n_1 $(n_1 > 1)$ is an integer, r is the radius of the graph $G = (S, E)$, S is the set of configurations, $E = \{(X, Y) \mid X \in N_Y, X, Y \in S\}$, i.e.,

$$r = \min_{X \in S - S_{lm}} \left\{ \max_{Y \in S} \{d_{XY}\} \right\} \tag{38}$$

$$S_{lm} = \{X \mid X \in S,\ c_X \geq c_Y \forall Y \in N_X\} \tag{39}$$

where d_{XY} stands for the length of the shortest path from node X to node Y in graph G. The center of the graph, denoted by X_c, is the node which obtains the minimum in (38).

Proof: According to conditions (g) and (a) in Theorem 4.3 and (35) and (36), for sufficiently large n and $\forall X \in S, Y \in N_X$,

$$p_{XY}(n) = g_{XY}(n)a_{XY}(n) \geq g^{-}(n)a^{-}(n) \tag{40}$$

and $\forall X \in S - S_{lm}, Y \in N_X$,

$$p_{XY}(n) = \begin{cases} g_{XY}(n), & c_Y \leq c_X \\ g_{XY}(n)a_{XY}(n), & c_Y > c_X \end{cases} \tag{41}$$

It is obvious from (41) that $\forall X \in S - S_{lm}, \exists Y \in N_X$, such that $p_{XY}(n)$ is the decreasing function of n for sufficiently large n. Since $\forall X \in S$,

$$p_{XX}(n) = 1 - \sum_{Z \in N_X} p_{XZ}(n)$$

we know that $p_{XX}(n)$ is an increasing function of n for sufficiently large n. Because $g^{-}(n)a^{-}(n)$ is a decreasing function of n, we get, for sufficiently large n and $\forall X \in S - S_{lm}$,

$$p_{XX}(n) \geq g^{-}(n)a^{-}(n) \tag{42}$$

From (40) and (42) we have, $\forall X \in S, \exists n_1 > 1$, such that $\forall n \geq n_1 r$,

$$\begin{aligned} & p_{XX_c}(n-r, n) \\ = & \prod_{k=n-r+1}^{n} p_{XX_c}(k) \\ \geq & \prod_{k=n-r+1}^{n} \left[g^{-}(k)a^{-}(k)\right] \\ \geq & \left[g^{-}(n)a^{-}(n)\right]^{r} \end{aligned} \tag{43}$$

Then we can get a lower bound on the ergodic coefficient (see Definition V.2.1 in [13]) of the nonstationary Markov chain associated with SA,

$$\begin{aligned}
& \alpha(P(n-r,n)) \\
= \; & \min_{X,Z}\left\{\sum_{Y\in S}\min\left(p_{XY}(n-r,n),p_{ZY}(n-r,n)\right)\right\} \\
\geq \; & \min_{X,Z}\left\{\min\left(p_{XX_c}(n-r,n),p_{ZX_c}(n-r,n)\right)\right\} \\
\geq \; & \left[g^-(n)a^-(n)\right]^r
\end{aligned} \tag{44}$$

The above result (44) and condition (37) of the Theorem lead to

$$\sum_{k=1}^{\infty}\alpha\left(P\left(kr-r,kr\right)\right)\geq\sum_{k=n_1}^{\infty}\left[g^-(kr)a^-(kr)\right]^r=\infty \tag{45}$$

Now the weak ergodicity of the nonstationary Markov chain associated with SA can easily be seen by Theorem 4.2.

Q.E.D.

The strong ergodicity and the convergence of general SA are described by Theorem 4.4.

Theorem 4.4 *Suppose the conditions specified in Theorem 4.3 are satisfied, and* $\{P(n)\}_{n=1}^{\infty}$ *are the transition matrices of the nonstationary Markov chain associated with SA. If* $P(n)$ *has a regular extension* $\overline{P}(x)$*, where* $\forall X,Y\in S$*,* $\overline{p}_{XY}(x)$ *belongs to CAM or RCBV, then* $\{P(n)\}_{n=1}^{\infty}$ *is strongly ergodic. Moreover, for sufficiently large* n*, each* $P(n)$ *has a unique quasi-stationary state distribution* $\pi(n)$ *with* $\lim_{n\to\infty}\pi(n)=\pi^*$ *and* $\forall X,Y\in S, m\geq 1$, $\lim_{n\to\infty}p_{XY}(m,n)=\pi_Y^*$*. The general SA converges if* π^* *satisfies,* $\forall X\in S$*,*

$$\pi_X=\begin{cases}\sigma_X\ (\sigma_X>0), & X\in S^*\\ 0, & X\in S-S^*\end{cases} \tag{46}$$

where S^* *is defined by (5), and*

$$\sum_{X\in S^*}\sigma_X=1$$

<u>*Proof:*</u> By Theorem 4.1 and Theorem 4.3.

Q.E.D.

The main difference between Theorem 4.4 and the earlier convergence theorems is that this theorem allows the generation probability to change dynamically according to different temperatures. The earlier convergence theorems can all be treated as corollaries of Theorem 4.4. Further important corollaries will be given in Section 5.

The conditions (g) and (a) in Theorem 4.3 are easy to verify in practice. There are many functions which satisfy these conditions and belong to CAM or RCBV as well. In some sense, condition (37) determines the convergence of the general SA. A too rapid cooling rate will lead to convergence of the sum in (37). This means that the probability of the uphill movement decreases quickly and the search is more likely to stick in a local minimum. Hence, condition (37) actually sets an upper bound on how fast generation and acceptance probabilities can decrease, or in other words, how fast the temperature can decrease.

5 Simulated Annealing with Dynamic Generation and Acceptance Probabilities

As indicated in Section 3, the larger an SA's neighbourhood is, the easier it is to arrive at a global minimum, provided the global minimum is far enough from the current configuration. In practice, the probability that a randomly selected initial configuration is very close to a global minimum is quite small. Hence, it is beneficial to have large neighbourhoods in the initial stages of SA's search, i.e., at high temperatures. With the progress of the search, i.e., the decrease of temperature, the current configuration will move closer and closer to a global minimum, and SA's neighbourhood should be reduced accordingly. This section gives a method of changing SA's neighbourhoods according to the temperature. An important feature of the method described here is that it enjoys the global convergence, and at the same time increases the efficiency of SA's search.

According to Theorem 4.4, many real-valued probability distribution functions can be used as the generation probability of SA, but for combinatorial optimizations they have to be discrete as we are dealing with a discrete configuration space. Suppose we are going to use a continuous density function $f(x)$ to generate the Hamming distance ξ between the current configuration and the next one. Denote the set of configurations which are ξ distant from current configuration X as $S_X(\xi)$,

$$S_X(\xi) = \{Y \mid Y \in S, d_{XY} = \xi\} \tag{47}$$

The probability of generating configuration Y, which is d_{XY} distant from configuration X, could be defined as

$$\begin{aligned} g_{XY}(T_n) &= \frac{1}{|S_X(d_{XY})|} Prob\left\{d_{XY} - \frac{1}{2} < \xi \leq d_{XY} + \frac{1}{2}\right\} \\ &= \frac{1}{|S_X(d_{XY})|} \int_{d_{XY}-\frac{1}{2}}^{d_{XY}+\frac{1}{2}} f(x)dx \end{aligned}$$

$$\approx \frac{f(d_{XY})}{|S_X(d_{XY})|} \tag{48}$$

Suppose the maximum Hamming distance allowed for one move is d_{max}, then the normalized generation function is

$$g_{XY}(T_n) = \frac{f(d_{XY}) / |S_X(d_{XY})|}{F_X(T_n)} \tag{49}$$

where

$$F_X(T_n) = \sum_{d_{XZ}=1}^{d_{max}} \sum_{Z \in S} \frac{f(d_{XZ})}{|S_X(d_{XZ})|} \tag{50}$$

Corollary 5.1 *Suppose the generation and acceptance probabilities of SA are (49) and (3) respectively, and $f(x)$ is the Normal function $N(0, T_n)$. Then the SA converges to global minima if the cooling rate is*

$$T_n = \frac{\lambda_G}{\sum_{i=1}^{N} \left[(\ln)^i \left(\left\lfloor \frac{n}{t} \right\rfloor + n_0\right)\right]}, \qquad n = 0, 1, 2, \cdots \tag{51}$$

where

$$\lambda_G \geq \left[\frac{1}{2}(d_{max})^2 + \Delta c^+\right] r \tag{52}$$

and other definitions are the same as (9) – (14).

Proof: We first show the weak ergodicity of the nonstationary Markov chain associated with SA. Under the assumption of the Normal function,

$$f(d_{XY}) = \frac{1}{\sqrt{2\pi T_n}} \exp\left(-\frac{d_{XY}^2}{2T_n}\right)$$

According to (35) and (49), for $k = 0, 1, 2, \cdots$,

$$g^-(kr) = \frac{\prod_{i=0}^{N-1} \left[(\ln)^i \left(\left\lfloor \frac{kr}{t} \right\rfloor + n_0\right)\right]^{-(d_{max})^2/(2\lambda_G)} \sqrt{\sum_{i=1}^{N} \left[(\ln)^i \left(\left\lfloor \frac{kr}{t} \right\rfloor + n_0\right)\right]}}{\sqrt{2\pi\lambda_G}\ |S^+(d)|\ F^+(T_n)} \tag{53}$$

where

$$\begin{aligned} \left|S^+(d)\right| &= \max_{X \in S} \{|S_X(\xi)|, \xi = 1, 2, \cdots, d_{max}\} && (54) \\ F^+(T_n) &= \max_{X \in S} \{F_X(T_n)\} && (55) \end{aligned}$$

According to (36) and (3), for $k = 0, 1, 2, \cdots$,

$$\begin{aligned} & a^-(kr) \\ \geq\ & \exp\left(-\frac{\Delta c^+}{T_{kr}}\right) \\ =\ & \prod_{i=0}^{N-1} \left[(\ln)^i \left(\left\lfloor \frac{kr}{t} \right\rfloor + n_0\right)\right]^{-(\Delta c^+)/\lambda_G} && (56) \end{aligned}$$

Hence, for $k = 0, 1, 2, \cdots$,

$$\sum_{k=0}^{\infty}\left[g^{-}(kr)a^{-}(kr)\right]^{r}$$

$$\geq \frac{\sum_{k=0}^{\infty}\left[\prod_{i=0}^{N-1}\left[(\ln)^{i}\left(\left\lfloor\frac{kr}{t}\right\rfloor+n_0\right)\right]^{-\left[\frac{1}{2}d_{max}^{2}+\Delta c^{+}\right]r/\lambda_G}\left[\sum_{i=1}^{N}\left[(\ln)^{i}\left(\left\lfloor\frac{kr}{t}\right\rfloor+n_0\right)\right]\right]^{r/2}\right]}{\left[\sqrt{2\pi\lambda_G}\ |S^{+}(d)|\ F^{+}(T_n)\right]^{r}}$$

We know from (52) that the above sum is infinite. By Theorem 4.3 we obtain the weak ergodicity. To show the strong ergodicity and convergence, let us first examine the one-step transition probabilities of the nonstationary Markov chain,

$$p_{XY}(T_n) = \begin{cases} \frac{\exp\left(-\frac{d_{XY}^2}{2T_n}\right)\ \min\left\{1, \exp\left(-\frac{c_Y - c_X}{T_n}\right)\right\}}{\sqrt{2\pi T_n}\,|S_X(d_{XY})|\,F_X(T_n)}, & Y \in N_X \\ 0, & Y \notin N_X, Y \neq X \\ 1 - \sum_{Z\in S} p_{XZ}(T_n), & Y \notin N_X, Y = X \end{cases} \tag{57}$$

Substituting v for T_n in (57), we can get functions $\overline{p}_{XY}(v)$, which satisfies $\overline{p}_{XY}(v) > 0$ for all $X \in S, Y \in N_X$ and $Y = X, v > 0$. That is, $\{(X,Y) \mid \overline{p}_{XY}(v) > 0\}$ is the same for all $0 < v < v_0$, where v_0 (> 0) can be any real number. Hence, $\overline{P}(v)$ is a regular extension of the nonstationary Markov chain $\{P(n)\}_{n=0}^{\infty}$ according to Proposition 4.2. It is easy to see that the sequence $\{T_n\}_{n=0}^{\infty}$ satisfies $\lim_{n\to\infty} T_n = 0$ from (51). Furthermore, we know, by Definition 4.3 and 4.4 and Proposition 4.1, that $p_{XY}(T_n)$ belongs to CAM for all $X, Y \in S$. Now the strong ergodicity can be arrived at by Theorem 4.4, and at the same time we conclude that there exists a unique quasi-stationary distribution $\pi(T_n)$ for each $P(T_n)$ for sufficiently large n, i.e.,

$$\pi(T_n) = \pi(T_n)\,P(T_n) \tag{58}$$

(58) implies that $\forall X, Y \in S$,

$$\pi_X(T_n)\,p_{XY}(T_n) = \pi_Y(T_n)\,p_{YX}(T_n)$$

Hence we get, $\forall X \in S$,

$$\pi_X(T_n) = \frac{\exp(-c_X / T_n)}{\sum_{Z\in S}\exp(-c_Z / T_n)} \tag{59}$$

Obviously,

$$\lim_{n\to\infty}\pi_X(T_n) = q_X^{*} = \begin{cases} \frac{1}{|S^{*}|}, & X \in S^{*} \\ 0, & X \in S - S^{*} \end{cases} \tag{60}$$

where

$$\mathbf{q}^{*} = \left(q_{X_1}^{*}, q_{X_2}^{*}, \cdots, q_{X_{|S|}}^{*}\right), \qquad X_i \in S \tag{61}$$

is called the optimum distribution of configurations in the configuration space.

Q.E.D.

The same technique can be used to derive a set of corollaries, including the following relatively important ones.

Corollary 5.2 *Suppose the generation and acceptance probabilities of SA are (49) and (3) respectively, and $f(x)$ is the exponential function $E(0, T_n)$. Then the SA converges to global minima if the cooling rate is*

$$T_n = \frac{\lambda_E}{\sum_{i=1}^{N}\left[(\ln)^i\left(\left\lfloor\frac{n}{t}\right\rfloor + n_0\right)\right]}, \qquad n = 0, 1, 2, \cdots \tag{62}$$

where

$$\lambda_E \geq \left(d_{max} + \Delta c^+\right) r \tag{63}$$

and other definitions are the same as (9) – (14).

<u>Proof:</u> The same as Corollary 5.1. Note that $\forall X, Y \in S$,

$$f\left(d_{XY}\right) = \frac{1}{T_n}\exp\left(-\frac{d_{XY}}{T_n}\right)$$

Q.E.D.

Corollary 5.3 *Suppose the generation and acceptance probabilities of SA are (49) and (3) respectively, and $f(x)$ is Cauchy function $C(0, T_n)$. Then the SA converges to global minima if the cooling rate is*

$$T_n = \frac{\lambda_C}{\sum_{i=1}^{N}\left[(\ln)^i\left(\left\lfloor\frac{n}{t}\right\rfloor + n_0\right)\right]}, \qquad n = 0, 1, 2, \cdots \tag{64}$$

where

$$\lambda_C \geq \left(\Delta c^+ + 1\right) r \tag{65}$$

and other definitions are the same as (9) – (14).

<u>Proof:</u> The same as Corollary 5.1. Note that $\forall X, Y \in S$,

$$f\left(d_{XY}\right) = \frac{1}{\pi}\frac{T_n}{d_{XY}^2 + T_n^2}$$

Q.E.D.

Although we adopted (3) as the acceptance probability in the above analyses, others can also be used, such as [20]

$$a_{XY}\left(T_n\right) = \min\left\{1, \exp\left(\frac{c_X^{\eta}\left(c_X - c_Y\right)}{T_n}\right)\right\}, \qquad \eta \geq 0$$

or those described in [23, 19].

6 Conclusion

This paper gives a probabilistic analysis of the impact of SA's neighbourhood on SA's performance and shows that an SA with large neighbourhoods is easier to arrive at a global minimum than that with smaller ones, thus reduces the computation time. A general model of SA with both dynamic generation and acceptance probabilities is also studied and its convergence conditions are shown. This general model unifies various SA's within the same framework and allows derivation of many useful results with both theoretical and practical importance.

A practical method of using continuous probability density functions as SA's generation functions to extend SA's neighbourhood is given herein. Such SA's reduce the computation time and outperform SSA. This is also shown by our experimental results [24], where we used TSP as an example and compared the result of SSA and that of SA with the Normal function as generation function. A question which arises in this stage is which probabiity density function can give SA best performance. We have done some preliminary experiments on this issue [24]. We compared SA's with five different generation functions, i.e., Normal function, exponential function, Cauchy function, stable function with index $\frac{1}{2}$ and a hybrid function with Cauchy at high temperatures and Normal at low temperatures, where the density function of the stable distribution with index $\frac{1}{2}$ is [25]

$$f(x) = \frac{1}{\sqrt{2\pi x^3}} \exp\left(-\frac{1}{2x}\right), \qquad x > 0 \tag{66}$$

It is easy to show that the expectation and variance of (66) do not exist. Our results indicate that the hybrid function and Cauchy function are better than the other three and the stable function with index $\frac{1}{2}$ is the worst. However, more extensive empirical study and rigorous theoretical analysis are needed to find out which kinds of distribution or combinations of distributions are best for SA's generation and acceptance functions.

Another important research direction is to incorporate the concept of knowledge guided search into SA or other stochastic search methods [26]. It is very difficult for SA to know whether a region in the configuration space has been explored or whether a region is good, even though SA has been wandering in the space for a long time. SA should be able to discover some hints or gain some knowledge about the space through previous searches to guide the subsequent searches. The totally blind stochastic search can hardly achieve very high efficiency from the AI's point of view.

Acknowledgement — The author would like to thank Profs. Richard P. Brent and Guo-Jie Li for clarifying the meaning of Theorem 3.1 and Dr. Iain D. G. Macleod for correcting errors in English .

References

[1] D. F. Wong, H. W. Leong, and C. L. Liu.
Simulated Annealing for VLSI Design.
Kluwer Academic Publishers, Boston, 1988.

[2] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi.
Optimization by simulated annealing.
Science, 220:671–680, 1983.

[3] P. J. M. van Laarhoven and E. H. L. Aarts.
Simulated Annealing: Theory and Applications.
D. Reidel Publishing Co., 1987.

[4] E. H. L. Aarts and P. J. M. van Laarhoven.
Statistical cooling: a general approach to combinatorial optimization problems.
Philips Journal of Research, 40:193–226, 1985.

[5] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski.
A learning algorithm for Boltzmann machine.
Cognitive Science, 9:147–169, 1985.

[6] M. F. Tenorio and W. T. Lee.
Self-organizing neural nets for the identification problem.
In D. S. Touretzky, editor, *Advances in Neural Information Processing Systems (I).* Morgan Kaufmann, San Mateo, CA, 1989.

[7] P. Carnevalli, L. Coletti, and S. Petarnello.
Image processing by simulated annealing.
IBM Journal of Research and Development, 29:569–579, 1985.

[8] E. D. Sontag and H. J. Sussman.
Image restoration and segmentation using the annealing algorithm.
In *Proceedings of the 24th International Conference on Decision and Control*, pages 768–773, Ft. Lauderdale, Florida, 1985.

[9] A. A. El Gemal.
Using simulated annealing to design good codes.
IEEE Trans. on Information Theory, IT-33:116–123, 1987.

[10] D. H. Ackley.
An empirical study of bit vector function optimization.
In L. Davis, editor, *Genetic Algorithms and Simulated Annealing*, chapter 13, pages 170–204. Morgan Kaufmann, San Mateo, CA, 1987.

[11] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller.
Equations of state calculations by fast computing machines.
Journal of Chemical Physics, 21:1087–1091, 1953.

[12] V. Černy.
Thermodynamical approach to the travelling salesman problem: an efficient simulation algorithm.
Journal of Optimization Theory and Applications, 45:41–51, 1985.

[13] D. L. Isaacson and R. W. Madsen.
Markov Chains Theory and Applications.
John Wiley & Sons, 1976.

[14] S. Geman and D. Geman.
Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images.
IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI-6:721–741, 1984.

[15] B. Hajek.
Cooling schedules for optimal annealing.
Mathematics of Operations Research, 13:311–329, 1988.

[16] D. Mitra, F. Romeo, and A. Sangiovanni-Vincentelli.
Convergence and finite-time behavior of simulated annealing.
Advanced Applied Probability, 18:747–771, 1986.

[17] X. Yao and G.-J. Li.
Performance analysis of simulated annealing.
In L.-W. Chen, B.-C. Chen, and Q.-N. Sun, editors, *Proc. of the Third Pan-Pacific Computer Conf.*, pages 972–980, Beijing, 1989. International Academic Publishers and Pergamon Press.

[18] M. P. Vecchi and S. Kirkpatrick.
Global wiring by simulated annealing.
IEEE Trans. on CAD, CAD-2:215–222, 1983.

[19] S. Anily and A. Federgruen.
Simulated annealing methods with general acceptance probabilities.
Journal of Applied Probability, 24:657–667, 1987.

[20] I. O. Bohachevsky, M. E. Johnson, and M. L. Stein.
Generalized simulated annealing for function optimization.
Technometrics, 28:209–217, 1986.

[21] H. L. Royden.

Real Analysis.
Macmillan Publishing Co., Inc., 2nd edition, 1968.

[22] S. Anily and A. Federgruen.
Ergodicity in parametric nonstationary Markov chains: an application to annealing methods.
Operations Research, 35:867–874, 1987.

[23] B. Gidas.
Non-stationary Markov chains and convergence of the annealing algorithm.
Journal of Statistical Physics, 39:73–131, 1985.

[24] X. Yao.
A General Stochastic Search Algorithm — Theory and Applications of Simulated Annealing.
PhD thesis, University of Science and Technology of China, Hefei, Anhui 230026, P.R.China, 1989.

[25] W. Feller.
An Introduction to Probability Theory and Its Applications, volume 2.
John Wiley & Sons, Inc., 2nd edition, 1971.

[26] X. Yao.
Optimization by genetic annealing.
In M. Jabri, editor, *Proc. of Second Australian Conf. on Neural Networks*, pages 94–97, Sydney, Australia, 1991.