Journal of Statistical Physics, Vol. 50, Nos. 1/2, 1988

A Quantitative Analysis of the Simulated Annealing Algorithm: A Case Study for the Traveling Salesman Problem

Emile H. L. Aarts,[1] Jan H. M. Korst,[1] and Peter J. M. van Laarhoven[1]

Received May 22, 1987; revision received September 1, 1987

A quantitative study is presented of the typical behavior of the simulated annealing algorithm based on a cooling schedule presented previously by the authors. The study is based on the analysis of numerical results obtained by systematically applying the algorithm to a 100-city traveling salesman problem. The expectation and the variance of the cost are analyzed as a function of the control parameter of the cooling schedule. A semiempirical average-case performance analysis is presented from which estimates are obtained on the expectation of the average final result obtained by the simulated annealing algorithm as a function of the distance parameter, which determines the decrement of the control parameter.

KEY WORDS: Combinatorial optimization; simulated annealing; traveling salesman problem; performance analysis.

1. INTRODUCTION

Ever since Kirkpatrick *et al.*[8] and Černy[4] introduced the concepts of annealing into the field of combinatorial optimization, much effort has been devoted to investigating the theory of the simulated annealing algorithm[2] and many applications to a wide variety of problems in various disciplines have been presented. For an extensive treatment of the theory and the applications the reader is referred to Ref. 10. The annealing algorithm is based on Monte Carlo techniques applying the Metropolis algorithm from statistical physics[13] and can be modeled mathematically

[1] Philips Research Laboratories, 5600 JA Eindhoven, the Netherlands.

[2] Other names used to denote the method are statistical cooling[1] and probabilistic hill climbing.[15]

0022-4715/88/0100-0187$06.00/0

by using concepts of the theory of Markov chains. It was proved by a number of authors that under certain conditions the algorithm converges asymptotically to an optimal solution (see also Section 2). Thus, asymptotically the algorithm is an optimization algorithm. In practical applications, however, asymptoticity is never attained and thus convergence to an optimal solution is no longer guaranteed. Consequently, in practice the algorithm is an approximation algorithm.

The quality of the final solution obtained by the algorithm is determined by the convergence of the algorithm, which is governed by a set of parameters, called the cooling schedule. In the literature the behavior of the simulated annealing algorithm as an approximation algorithm is usually analyzed in an empirical way. This involves the analysis of computation times and quality of final solutions obtained by running the algorithm on a (large) set of problem instances, e.g., by comparing the results with those found by other approximation and optimization algorithms.

However, also for a *fixed* problem instance, it is interesting to analyze the aforementioned parameters because, even for a fixed instance, the computation time and the quality of the final solution are random variables, due to the probabilistic nature of the algorithm. It is for this reason that in this paper we take one (representative) instance of the well-known traveling salesman problem as a starting point so as to present a (semiempirical) average-case analysis of the algorithm by running it a (large) number of times on the same instance. We show that it is possible to reproduce the observed behavior by using standard techniques from statistical physics and some assumptions on the configuration density. We emphasize that it is not our intention to contrast the performance of the algorithm with that of other algorithms nor other implementations of the simulated annealing algorithm (i.e., the simulated annealing algorithm with another cooling schedule).

Grest *et al.*[5] report on an experimental study of spin-glass problems from which they conclude that the ground-state energy obtained by the simulated annealing algorithm depends logarithmically on the cooling rate. To our knowledge, however, a more systematic investigation of the typical behavior and the average-case performance of the simulated annealing algorithm has not been presented in the literature. This paper serves as a first approach to such an investigation.

2. MATHEMATICAL MODEL OF THE SIMULATED ANNEALING ALGORITHM

A combinatorial optimization problem can be characterized by the configuration space $\mathcal{R}$, denoting the set of all possible configurations i,

and a cost function $C: \mathscr{R} \to \mathbf{R}$, which assigns a real number $C(i)$ to each configuration i. We assume that C is defined such that the lower the value of C, the better the corresponding configuration (with respect to the optimization criteria). This can be done without loss of generality. The objective is to find an optimal configuration i_o for which

$$C(i_o) = C_{\min} = \min\{C(i) \mid i \in \mathscr{R}\} \tag{1}$$

where $C_{\min}$ denotes the minimum cost. To apply the simulated annealing algorithm, a mechanism is used to generate a new configuration (a neighbor) from a given one by a small perturbation. A neighborhood $\mathscr{R}_i$ is defined as the set of configurations that can be reached from configuration i by a single perturbation. The simulated annealing algorithm starts off with a given initial configuration and continuously tries to transform a current configuration into one of its neighbors by applying a perturbation mechanism and an acceptance criterion. The acceptance criterion allows for deteriorations in the cost function, thus enabling the algorithm to escape from local minima.

2.1. Asymptotic Convergence

The simulated annealing algorithm can be formulated as a sequence of Markov chains, each Markov chain being a sequence of trials, where the outcome of a given trial only depends on the outcome of the previous trial (the outcomes of the trials are the configurations). The probability that a trial transforms configuration i into configuration j is defined as[1,10]

$$P_{ij}(c) = \begin{cases} A_{ij}(c)\, G_{ij}(c) & \text{if} \quad i \neq j \\ 1 - \sum\limits_{k \in \mathscr{R}, k \neq i} A_{ik}(c)\, G_{ik}(c) & \text{if} \quad i = j \end{cases} \tag{2}$$

where $P_{ij}(c)$ denotes the transition probability, $G_{ij}(c)$ the perturbation probability, i.e., the probability of generating configuration j from configuration i, $A_{ij}(c)$ the acceptance probability, i.e., the probability of accepting configuration j given the configurations i and j, and c denotes a value of the control parameter ($c \in \mathbf{R}^+$), which plays the role of the temperature in the physical annealing process.

It can be proved under a number of conditions on the probabilities $G_{ij}(c)$ and $A_{ij}(c)$ that asymptotically (i.e., for infinitely long Markov chains and $c \downarrow 0$) the algorithm finds an optimal configuration with probability equal to 1.[1,10] The proof is based on the existence of an equilibrium distribution. Let $\mathbf{X}(k)$ denote the outcome of the kth trial of a Markov chain; then under the condition that the Markov chains are irreducible, aperiodic,

and recurrent, there exists a unique equilibrium distribution given by the $|\mathscr{R}|$-vector $\mathbf{q}(c)$. The components $q_i(c)$ denote the probability that a configuration i will be found after an infinite number of trials and are given by the following expression:

$$q_i(c) = \lim_{k \to \infty} \Pr\{\mathbf{X}(k) = i \mid c\} \tag{3}$$

$$= \lim_{k \to \infty} ([P^k(c)]^T \mathbf{a})_i \tag{4}$$

where $\mathbf{a}$ denotes the initial probability distribution of the configurations and $P(c)$ the transition matrix, whose entries are given by the $P_{ij}(c)$ of Eq. (2). Under certain additional conditions on the probabilities $G_{ij}(c)$ and $A_{ij}(c)$ the algorithm converges as $c \downarrow 0$ to a uniform distribution on the set of optimal configurations, i.e.,[1,10]

$$\lim_{c \downarrow 0} \lim_{k \to \infty} \Pr\{\mathbf{X}(k) = i \mid c\} = \lim_{c \downarrow 0} q_i(c)$$

$$= \pi_i \tag{5}$$

and

$$\pi_i = \begin{cases} |\mathscr{R}_{\text{opt}}|^{-1} & \text{if} \quad i \in \mathscr{R}_{\text{opt}} \\ 0 & \text{elsewhere} \end{cases} \tag{6}$$

where $\mathscr{R}_{\text{opt}}$ denotes the set of optimal configurations.

Here, we apply the standard form of the simulated annealing algorithm, i.e., the perturbation probability $G_{ij}(c)$ is chosen independent of c and uniformly over the neighborhood of a given configuration i. The acceptance probability is chosen as

$$A_{ij}(c) = \begin{cases} \exp(-\Delta C_{ij}/c) & \text{if} \quad \Delta C_{ij} > 0 \\ 1 & \text{if} \quad \Delta C_{ij} \leqslant 0 \end{cases} \tag{7}$$

where $\Delta C_{ij} = C(j) - C(i)$. For this choice the components of the equilibrium distribution take the form

$$q_i(c) = \frac{\exp\{[C_{\min} - C(i)]/c\}}{\sum_{j \in \mathscr{R}} \exp\{[C_{\min} - C(j)]/c\}} \tag{8}$$

2.2. The Cooling Schedule

Commonly one resorts to an implementation of the simulated annealing algorithm in which a sequence of Markov chains of finite length is

generated at decreasing values of the control parameter. Optimization is carried out by starting off at a start value of the control parameter c_0 and repeatedly generating a Markov chain for decreasing values of c until c approaches 0. This procedure is governed by the cooling schedule. The parameters determining the cooling schedule are (1) the *start value* c_0 of the control parameter, (2) the *decrement function* f of the control parameter, (3) the *length* L of the individual Markov chains, and (4) the *stop criterion* to terminate the algorithm.

Here we apply the cooling schedule proposed by Aarts and van Laarhoven. For an extensive treatment see Refs. 1 and 2. Here, we briefly summarize the salient features of this cooling schedule.

Start value c_0: Let η be the acceptance ratio (the ratio between the number of accepted transitions and the number of proposed transitions in a Markov chain). The value of c_0 is calculated from the requirement that initially the acceptance ratio should be large (close to 1). The value of c_0 is obtained by generating a number of trials (say m_0) and applying the following expression:

$$c_0 = \overline{\Delta C}^{(+)} \left(\ln \frac{m_2}{m_2 \eta_0 - (1-\eta_0) m_1} \right)^{-1} \tag{9}$$

where m_1 and m_2 denote the numbers of perturbations obtained with $\Delta C_{ij} \leqslant 0$ and $\Delta C_{ij} > 0$, respectively ($m_1 + m_2 = m_0$), $\overline{\Delta C}^{(+)}$ the average value of those ΔC_{ij} values for which $\Delta C_{ij} > 0$, and η_0 the initial acceptance ratio.

Stop criterion Z: The algorithm is terminated at a value of c for which extrapolation of the smoothed value of the average cost $\overline{C_s}(c)$ satisfies a lower bound. This can be expressed by the following criterion:

$$Z: \quad \left| \frac{\partial \overline{C_s}(c)}{\partial c} \frac{c}{\bar{C}(c_0)} \right| < \varepsilon_s \tag{10}$$

where ε_s is a small, positive, real number called the stop parameter and $\bar{C}(c_0)$ is the average value of the cost function at c_0.

The decrement of the control parameter and the Markov-chain length are obtained from the requirement that the annealing process should stay in quasiequilibrium.[1] From this requirement the following expressions are obtained:

Decrement function f: The next value of the control parameter is calculated from

$$f(c) = c \left(1 + \frac{\ln(1+\delta)\, c}{3\sigma(c)} \right)^{-1} \tag{11}$$

where $\sigma(c)$ denotes the standard deviation of the values of the cost function of the configurations of the Markov chain at c and δ is a small, positive, real number called the distance parameter.

Markov-chain length L: All Markov chains are chosen equally long and the chain length is given by the following expression:

$$L = \max\{|\mathscr{R}_i|: i \in \mathscr{R}\} \tag{12}$$

Remark 1. It has been shown[1] that execution of the simulated annealing algorithm using the cooling schedule described above requires a total number of steps of order $L \ln |\mathscr{R}|$, which can be chosen polynomial in the size of the problem, thus resulting in a polynomial-time execution of the simulated annealing algorithm.

Remark 2. As a consequence of the asymptotic convergence of the simulated annealing algorithm it is intuitively clear that the slower the "cooling" is carried out, the larger the probability is that the final configuration is (close to) an optimal configuration. Thus, the deviation of the final configuration from an optimal configuration can be made as small as desired by investing more computational effort. The literature has not elaborated on the probabilistic dependence of the aforementioned deviation on the parameters of the cooling schedule. In this paper we present semiempirical results on this subject. A more theoretical treatment is still considered an open research topic.

3. A QUANTITATIVE ANALYSIS

In this section we discuss some quantitative aspects of the simulated annealing algorithm. The discussion is of a semiempirical nature and is based on an extensive set of numerical data obtained by applying the algorithm to a specific instance of the traveling salesman problem (TSP).[11] The description of the problem instance is given in Section 3.1 (see also the Appendix). In Section 3.2 we analyze the typical behavior of the simulated annealing algorithm. The analysis is based on an analytical approach from which the expectation and the variance of the cost function are expressed as a function of the control parameter. Furthermore, a parametrization is given which models the observed behavior. In Section 3.3 a discussion is presented of the probabilistic dependence of the final results obtained by the simulated annealing algorithm as a function of the distance parameter, which determines the convergence speed of the simulated annealing algorithm (see Section 2). The discussion is based on an average-case performance analysis.

3.1. The Problem Instance

All numerical data presented in this paper are obtained by applying the simulated annealing algorithm to an instance of the TSP with 100 cities. This is a medium-size optimization problem and we consider it as a proper representative of a broad class of combinatorial optimization problems to which the analysis presented in Sections 3.2 and 3.3 can be successfully applied, for the following reasons: (1) there are many different values of the cost function, (2) there are many local minima, and (3) the cost function is reasonably smooth (no clustering).

The problem under consideration is a symmetric instance of the TSP defined on a set of 100 large European cities. A full description of this problem instance is given in the Appendix. The minimum tour length in this problem instance is 21134 and was obtained in 59.5 CPU sec on a Cyber-205 in PASCAL. The minimum tour length is calculated using a TSP algorithm developed at the University of Amsterdam by Jonker and Volgenant.[7] The algorithm uses a branch and bound technique based on a one-tree relaxation, applying a mechanism that eliminates nonoptimal edges of the TSP instance. The algorithm is implemented in PASCAL.

Implementation of the simulated annealing algorithm was done in PASCAL on a VAX-11/780. As mentioned before, we used the cooling schedule given in Section 2.2 with the following parameters (unless explicitly mentioned otherwise): $\eta_0 = 0.9$, $\varepsilon_s = 10^{-5}$, and $\delta = 0.1$. For the perturbation mechanism we used 2-opt moves as defined by Lin and Kernighan[12] and for the acceptance probability we used the standard form given by Eq. (7).

3.2. Analysis of the Cost Function

To model the typical behavior of the simulated annealing algorithm, we discuss an analytical approach to calculate the expectation $\langle C \rangle_c$ and the variance σ_c^2 of the cost function. Let $\mathbf{X}$ denote the outcome of a given trial; then $\langle C \rangle_c$ and σ_c^2 can be defined as

$$\langle C \rangle_c = \sum_{i \in \mathscr{R}} \Pr\{\mathbf{X} = i \mid c\}\, C(i) \tag{13}$$

$$\sigma_c^2 = \sum_{i \in \mathscr{R}} \Pr\{\mathbf{X} = i \mid c\}[C(i) - \langle C \rangle_c]^2 \tag{14}$$

In equilibrium we obtain [using Eqs. (3) and (8)]

$$\langle C \rangle_c = \sum_{i \in \mathscr{R}} q_i(c)\, C(i)$$

$$= \frac{\sum_{i \in \mathscr{R}} \exp\{[C_{\min} - C(i)]/c\}\, C(i)}{\sum_{j \in \mathscr{R}} \exp\{[C_{\min} - C(j)]/c\}} \tag{15}$$

$$\sigma_c^2 = \sum_{i \in \mathcal{R}} q_i(c)[C(i) - \langle C \rangle_c]^2$$

$$= \frac{\sum_{i \in \mathcal{R}} \exp\{[C_{\min} - C(i)]/c\}[C(i) - \langle C \rangle_c]^2}{\sum_{j \in \mathcal{R}} \exp\{[C_{\min} - C(j)]/c\}} \tag{16}$$

Next, we introduce the configuration density $\omega(C)$ defined as

$$\omega(C)\, dC = \frac{1}{|\mathcal{R}|} |\{i \in \mathcal{R} \mid C \leqslant C(i) < C + dC\}| \tag{17}$$

Then, in the case of the simulated annealing algorithm employing the acceptance probability of Eq. (7), the *equilibrium*-configuration density $\Omega(C, c)$ at a given value of c is given by

$$\Omega(C, c)\, dC = \frac{\omega(C) \exp[(C_{\min} - C)/c]\, dC}{\int_{-\infty}^{\infty} \omega(C') \exp[(C_{\min} - C')/c]\, dC'} \tag{18}$$

Clearly, $\Omega(C, c)$ is the equivalent of the stationary distribution $\mathbf{q}(c)$ given by Eq. (8). As indicated by the notion "equilibrium," $\Omega(C, c)$ is the configuration density in equilibrium when applying the simulated annealing algorithm. Thus, one obtains

$$\langle C \rangle_c = \int_{-\infty}^{\infty} C' \Omega(C', c)\, dC' \tag{19}$$

$$\sigma_c^2 = \int_{-\infty}^{\infty} [C' - \langle C \rangle_c]^2\, \Omega(C', c)\, dC' \tag{20}$$

Given an analytical expression for the configuration density $\omega(C)$, it is possible to evaluate the integrals of Eqs. (18)–(20). To estimate $\omega(C)$ for a given combinatorial optimization problem is in most cases very hard. Indeed, $\omega(C)$ may vary drastically for different specific problem instances, especially for C values close to $C_{\min}$. Analysis of the configuration space by using the concept of ultrametricity can be of use for some problems.[9,17] However, in the present paper we intend to concentrate on the typical (i.e., average) behavior of the configuration density as observed for many problem instances.

Figure 1 shows (a) the average cost $\bar{C}(c)$ and (b) the spreading $\sigma(c)$ of the cost as a function of the control parameter c obtained by applying the simulated annealing algorithm to the TSP instance described in Section 3.1. The numerical data are obtained from the following expressions:

$$\bar{C}(c) = L^{-1} \sum_{i=1}^{L} C_i(c) \tag{21}$$

$$\sigma(c) = \left\{ L^{-1} \sum_{i=1}^{L} [C_i(c) - \bar{C}(c)]^2 \right\}^{1/2} \tag{22}$$

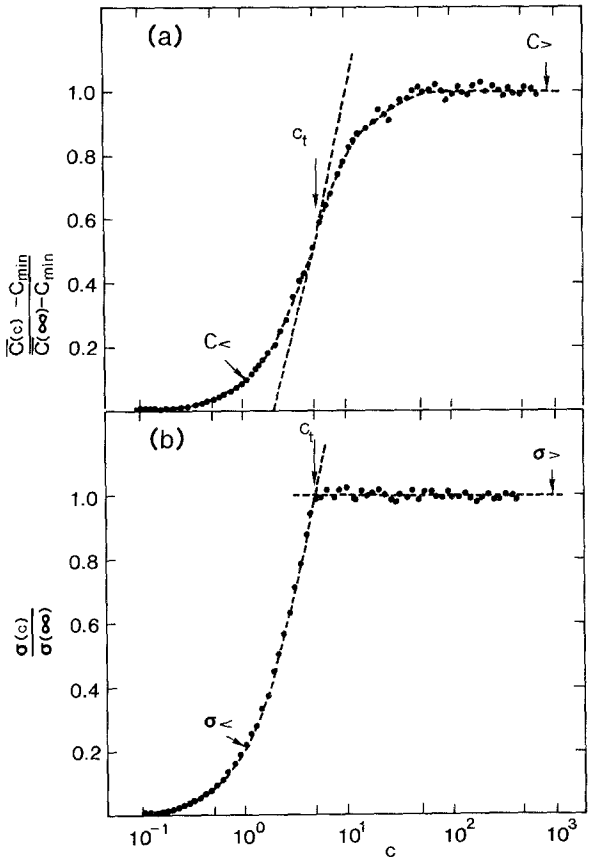

Fig. 1. (a) Normalized average cost $[\bar{C}(c)-C_{\min}]/[\bar{C}(\infty)-C_{\min}]$ and (b) spreading $\sigma(c)/\sigma(\infty)$ as a function of the control parameter c obtained for the 100-European-city TSP. The dashed lines are calculated according to Eqs. (38)–(41).

where the average is taken over the values of the cost function $C_i(c)$ $(i=1,\ldots, L)$ of a Markov chain generated at a given value of the control parameter c. The typical behavior shown in Figs. 1a and 1b is observed for many different problem instances and is reported by a number of authors.[1,6,8,10,18] From the figures we can deduce some characteristic features of the expectation $\langle C\rangle_c$ and the variance σ_c^2 of the cost function. First, it is observed that for large values of c the average and the spreading of the cost are about constant and equal to $\bar{C}(\infty)$ and $\sigma(\infty)$. This behavior is directly explained from Eqs. (15) and (16), or Eqs. (18)–(20), namely

$$\langle C\rangle_\infty = \lim_{c\to\infty}\langle C\rangle_c = \frac{1}{|\mathscr{R}|}\sum_{i\in\mathscr{R}} C(i) \tag{23}$$

$$\sigma_\infty^2 = \lim_{c\to\infty}\sigma_c^2 = \frac{1}{|\mathscr{R}|}\sum_{i\in\mathscr{R}}[C(i)-\langle C\rangle_\infty]^2 \tag{24}$$

Second, we observe that there exists a threshold value c_t of the control parameter for which

$$\langle C \rangle_{c_t} \approx \tfrac{1}{2}(\langle C \rangle_\infty + C_{\min}) \tag{25}$$

and

$$\begin{aligned} \sigma_c^2 &= \sigma_\infty^2 \qquad \text{if} \quad c \geqslant c_t \\ \sigma_c^2 &< \sigma_\infty^2 \qquad \text{if} \quad c < c_t \end{aligned} \tag{26}$$

The threshold value c_t marks the transition of the observed configurations from one region to another. We argue that these regions correspond to different (dominant) contributions of the configuration density $\omega(C)$ to the observed configuration density $\Omega(C, c)$. This leads to the following postulate.

Postulate. Let R_1 and R_2 be two regions of the value of the cost function, where R_1 denotes the region of a few standard deviations σ_∞ around $\langle C \rangle_\infty$ and R_2 the region close to C_{min}. Then, for a typical combinatorial optimization problem, $\omega(C)$ is given by a normal distribution $\omega_{\mathcal{N}}(C)$ in the region R_1 and by an exponential distribution $\omega_{\mathcal{E}}(C)$ in the region R_2. Furthermore, we state that the number of configurations in R_1 is much larger than the number of configurations in R_2. □

At the end of this section we discuss some arguments to support the type of distribution functions given in the postulate. From the postulate and Eq. (18) it follows that for large values of c, $\Omega(C, c)$ is dominated by $\omega_{\mathcal{N}}(C)$, whereas for small values of c, $\omega_{\mathcal{E}}(C)$ is dominant.

Thus, writing $\omega_{\mathcal{N}}(C)$ as

$$\omega_{\mathcal{N}}(C) \propto \exp\left(-\frac{(C - \langle C \rangle_\infty)^2}{2\sigma_\infty^2}\right) \tag{27}$$

one obtains for large values of c

$$\Omega(C, c) \approx \frac{\exp(-[C - (\langle C \rangle_\infty - \sigma_\infty^2/c)]^2/2\sigma_\infty^2)}{\int_{C_{\min}}^{C_{\max}} \exp(-[C' - (\langle C \rangle_\infty - \sigma_\infty^2/c)]^2/2\sigma_\infty^2)\, dC'} \tag{28}$$

$$= N^{-1}(c) \exp\left(-\frac{[C - (\langle C \rangle_\infty - \sigma_\infty^2/c)]^2}{2\sigma_\infty^2}\right) \tag{29}$$

where $C_{\max}$ denotes the maximum value of the cost function and

$$N(c) = \frac{1}{2}\sigma_\infty (2\pi)^{1/2} \left\{ \operatorname{erf}\left(\frac{C_{\max} - (\langle C \rangle_\infty + \sigma_\infty^2/c)}{\sigma_\infty \sqrt{2}}\right) - \operatorname{erf}\left(\frac{C_{\min} - (\langle C \rangle_\infty + \sigma_\infty^2/c)}{\sigma_\infty \sqrt{2}}\right) \right\} \tag{30}$$

Consequently, since the expression given by Eq. (29) again corresponds to a normal distribution, one obtains

$$\langle C\rangle_c \approx \langle C\rangle_\infty - \sigma_\infty^2/c \tag{31}$$

and

$$\sigma_c^2 \approx \sigma_\infty^2 \tag{32}$$

The configuration density $\omega_{\mathscr{E}}(C)$ is given by an exponential distribution of the form

$$\omega_{\mathscr{E}}(C) \propto \exp[(C - C_{\min})\,\gamma] \tag{33}$$

for some constant γ $(0 < \gamma < c^{-1})$. Thus, for small values of c one obtains

$$\begin{aligned}\Omega(C, c) &\approx \frac{\exp[(C_{\min} - C)(1-\gamma c)/c]}{\int_{C_{\min}}^{C_{\max}} \exp[(C_{\min} - C')(1-\gamma c)/c]\, dC'} \\ &= M(c)^{-1}\left(\frac{1-\gamma c}{c}\right)\exp\left[(C_{\min} - C)\left(\frac{1-\gamma c}{c}\right)\right]\end{aligned} \tag{34}$$

where

$$M(c) = 1 - \exp\left[(C_{\min} - C_{\max})\left(\frac{1-\gamma c}{c}\right)\right] \tag{35}$$

Consequently, assuming that $C_{\max} \gg C_{\min}$, we obtain

$$\langle C\rangle_c - C_{\min} \propto \frac{c}{1-\gamma c} = c[1 + \gamma c + (\gamma c)^2 + \cdots] \tag{36}$$

$$\sigma_c^2 \propto \left(\frac{c}{1-\gamma c}\right)^2 = c^2[1 + 2\gamma c + 3(\gamma c)^2 + \cdots] \tag{37}$$

From these derivations it follows that (1) for large values of c, $\langle C\rangle_c$ is linear in c^{-1} and σ_c^2 is constant, and (2) for small values of c, $\langle C\rangle_c$ is proportional to c, and σ_c^2 to c^2.

Using the threshold value c_t and the results obtained above, a simple parametrization is suggested of the form

$$\langle C\rangle_c = \begin{cases} C_< = C_{\min} + N_t\left(\langle C\rangle_\infty - C_{\min} - \dfrac{\sigma_\infty^2}{c_t}\right)\left(\dfrac{c}{1-\gamma c}\right) & \text{if } c \leqslant c_t \\ C_> = \langle C\rangle_\infty - \dfrac{\sigma_\infty^2}{c} & \text{if } c > c_t \end{cases} \tag{38}$$

and

$$\sigma_c^2 = \begin{cases} \sigma_<^2 = N_t^2 \sigma_\infty^2 \left(\dfrac{c}{1-\gamma c} \right)^2 & \text{if} \quad c \leqslant c_t \\ \sigma_>^2 = \sigma_\infty^2 & \text{if} \quad c > c_t \end{cases} \tag{39}$$

where

$$N_t = (1 - \gamma c_t)/c_t \tag{40}$$

and

$$c_t = 2\sigma_\infty^2/(\langle C \rangle_\infty - C_{\min}) \tag{41}$$

The latter equation is obtained from Eqs. (25) and (38).

The curves corresponding to the parametrization of Eqs. (38)–(41) applied to the 100 European-city TSP instance of Section 3.1 are drawn in Figs. 1a and 1b with $\gamma = 0.01$ and $c_t = 4.9$. The value of c_t was obtained from Eq. (41) using $\sigma_\infty = 53$ and $\langle C \rangle_\infty = 1353$. The value of γ was determined experimentally. The curves fit the data very well (the correlation coefficient $\chi^2 = 0.91$). The parametrization formulas (38)–(41) have also been applied successfully to a number of other problem instances and we conclude that they represent very well the gross features of the expectation and the variance of the cost function obtained by the simulated annealing algorithm applied to these problem instances.

Remark 1. The normal distribution $\omega_{\mathcal{N}}(C)$ is also reported by other authors.[6,18] It corresponds to a distribution that maximizes the entropy subject to the observed values of $\langle C \rangle_\infty$ and σ_∞. More formally stated, $\omega_{\mathcal{N}}(C)$ is given by the solution of the following constrained optimization problem:

$$\text{maximize} \quad \int_{-\infty}^{\infty} \omega_{\mathcal{N}}(C) \ln \omega_{\mathcal{N}}(C)\, dC \tag{42}$$

$$\text{subject to} \quad \int_{-\infty}^{\infty} \omega_{\mathcal{N}}(C)\, dC = 1 \tag{43}$$

$$\int_{-\infty}^{\infty} C' \omega_{\mathcal{N}}(C')\, dC' = \langle C \rangle_\infty \tag{44}$$

$$\int_{-\infty}^{\infty} (C' - \langle C \rangle_\infty)^2\, \omega_{\mathcal{N}}(C')\, dC' = \sigma_\infty^2 \tag{45}$$

Using standard techniques from the theory of constrained optimization, it is straightforward to show that $\omega_{\mathcal{N}}(C)$ is given by a normal distribution of the form given by Eq. (27).

Evidently, the structure of a combinatorial optimization problem differs from one problem to another and the configurations in the region around $\langle C \rangle_\infty$ are not by definition normally distributed. However, if the number of configurations is extremely large and the values of the cost function are distributed sufficiently uniformly (no clustering), then the number of degrees of freedom is large and the structure only plays a minor role. Consequently, the configuration density is approximately given by a normal distribution typical of disordered systems.[3]

Remark 2. In the region close to $C_{\min}$ the structure of a combinatorial optimization problem is important and strongly determines the configuration density $\omega(C)$. Therefore, the exponential distribution $\omega_{\mathscr{E}}(C)$ given in Eq. (33) should be treated with caution, since it does not hold by definition. It can be argued to hold for a number of examples.[6,14] Furthermore, it reproduces the behavior of the simulated annealing algorithm very well, as was shown above. However, more research into the analysis of combinatorial optimization problems is required to obtain insight into the behavior of the configuration density near $C_{\min}$. With respect to this aspect, the work on configuration space analysis and ultrametricity may prove useful.[9,17]

3.3. Probabilistic Performance Analysis

In this section we present a semiempirical average-case analysis of the final result obtained by the simulated annealing algorithm as a function of the distance parameter δ, which governs the decrement of the control parameter of the cooling schedule (Section 2.2). Let $\mathbf{X}_{\text{fin}}$ denote the outcome of the last trial of the simulated annealing algorithm. Then the expectation and spreading of the final deviation of the cost function are defined as

$$\langle \Delta C_{\text{fin}} \rangle_\delta = \sum_{i \in \mathscr{R}} \Pr\{\mathbf{X}_{\text{fin}} = i \mid \delta\}[C(i) - C_{\min}] \tag{46}$$

and

$$(\sigma^2_{\text{fin}})_\delta = \sum_{i \in \mathscr{R}} \Pr\{\mathbf{X}_{\text{fin}} = i \mid \delta\}[(C(i) - C_{\min}) - \langle \Delta C_{\text{fin}} \rangle_\delta]^2 \tag{47}$$

Figure 2a shows the average deviation of the final value of the cost function from the minimum value as a function of the distance parameter δ

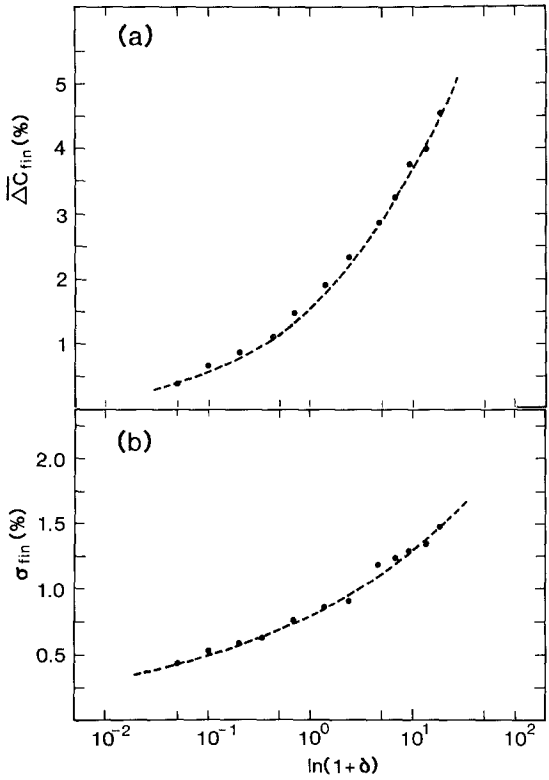

Fig. 2. (a) Normalized average deviation of the final cost $\overline{\Delta C}_{fin}$ and (b) spreading σ_{fin} as a function of the distance parameter δ obtained for the 100-European-city TSP. The dashed lines are given by Eqs. (49) and (50).

($\overline{\Delta C}_{fin}(\delta) = \bar{C}_{fin}(\delta) - C_{min}$, where $\bar{C}_{fin}(\delta)$ denotes the average final value of the cost function obtained by the algorithm for a given value of δ). The corresponding spreading $\sigma_{fin}(\delta)$ is shown in Fig. 2b. The data are calculated from a set of final values of the cost function obtained by applying the simulated annealing algorithm (using the cooling schedule described in Section 2.2) n times with different initial configurations to the 100 European-city TSP instance described in Section 3.1. The values of n ranged from 10 for small values of δ to 20 for large values of δ. As mentioned in Remark 2 of Section 2.2, it is intuitively clear that the slower the cooling process is carried out (small values of δ), the larger is the probability that the final result is close to an optimum. Indeed, this behavior is observed in Figs. 2a and 2b, i.e.,

$$\lim_{\delta \downarrow 0} \overline{\Delta C}_{fin}(\delta) = 0 \qquad \text{and} \qquad \lim_{\delta \downarrow 0} \sigma_{fin}(\delta) = 0 \tag{48}$$

From the behavior observed in Figs. 2a and 2b the following parametrizations for the expectation and the variance of $\Delta C_{\text{fin}}(\delta)$ can be deduced:

$$\langle \Delta C_{\text{fin}} \rangle_\delta = [\alpha(\delta)]^a \tag{49}$$

$$(\sigma_{\text{fin}})^2_\delta = [\alpha(\delta)]^a \tag{50}$$

where $\alpha = b \ln(1+\delta)$. The values of the parameters a and b depend on the problem instance and are experimentally determined to be $a = 0.42$ and $b = 1.54$ ($\chi^2 = 0.91$).

Figure 3 shows the average running time $\bar{T}_{\text{CPU}}$ required by the simulated annealing algorithm to obtain a final result deviating (on the average) ΔC_{fin} from the optimum cost. The expectation of the running time is parametrized as

$$\langle T_{\text{CPU}} \rangle = t_0 \left(\frac{C_{\min}}{\Delta C_{\text{fin}}} \right)^t \tag{51}$$

where $\langle T_{\text{CPU}} \rangle$ denotes the expectation of the running time. The parameters t_0 and t are experimentally determined to be $t_0 = 4.91 \times 10^{-3}$ sec on a VAX-11/780 and $t = 2.447$ ($\chi^2 = 0.877$).

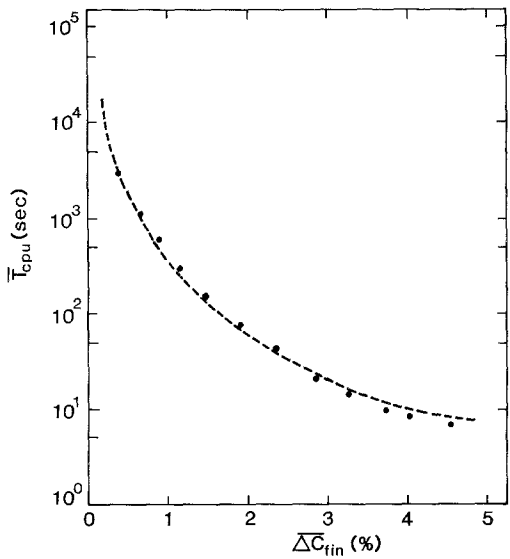

Fig. 3. Average running time $\bar{T}_{\text{CPU}}$ as a function of the average deviation of the final cost $\overline{\Delta C}_{\text{fin}}$ obtained for the 100-European-city TSP. The dashed line is given by Eq. (51).

The typical behavior of the average and the variance of the final value of the cost function obtained by the simulated annealing algorithm can be explained by assuming that the values of ΔC_{fin} are distributed according to a gamma distribution of the form

$$G_p(\Delta C_{\text{fin}}) = \frac{1}{\Gamma(p)} (\Delta C_{\text{fin}})^{p-1} e^{-\Delta C_{\text{fin}}} \tag{52}$$

where $p = [\alpha(\delta)]^a$. In fact, the gamma distribution is the only relatively simple distribution function having a mean and a variance corresponding to (49) and (50). From Eq. (52) an expression is derived for the probability Q of finding a solution whose cost deviates less than a small positive amount ε from the minimal cost, i.e.,

$$\begin{aligned} Q = \Pr\{\Delta C_{\text{fin}} \leq \varepsilon\} &= \int_0^{\varepsilon} G_p(C')\, dC' \\ &= P(p, \varepsilon) \end{aligned} \tag{53}$$

where $P(p, \varepsilon)$ denotes the incomplete gamma function. This function can be written as a series expansion in the following way:

$$P(p, \varepsilon) = \frac{\varepsilon^p}{\Gamma(p)} \sum_{n=0}^{\infty} \frac{(-\varepsilon)^n}{(p+n)\, n!} \tag{54}$$

The probability Q can be straightforwardly evaluated from this expression, since the series expansion converges rapidly. Figure 4 shows results of

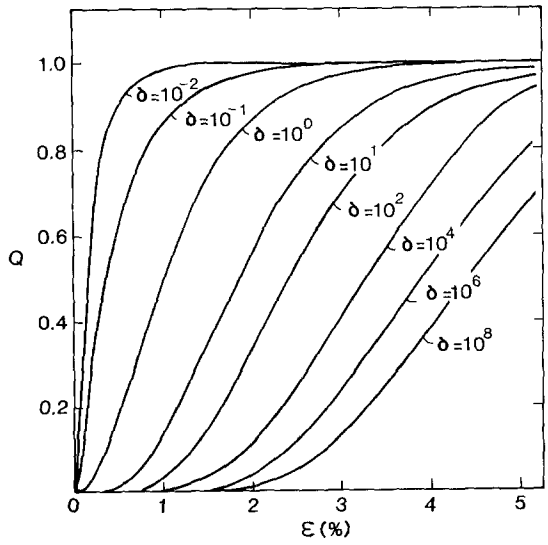

Fig. 4. Probability Q of obtaining a final result deviating less than a small amount ε from the minimum cost. The curves are calculated according to Eqs. (53) and (54).

numerical calculations of Q for different values of δ. The curves are calculated for the same set of parameters obtained from the 100-European-city TSP discussed above. From Fig. 4 it is possible to determine a value of the distance parameter δ to obtain a given expected final result with a given probability. Moreover, Eq. (51) gives an estimate of the expected running time. In practice this would be very useful, since it enables the user of the simulated annealing algorithm to control the expected performance of the algorithm. However, the parameters used in our approach depend on the problem instance under consideration and can only be obtained from an analysis as described above. Clearly, this limits the practical use of the approach to those problem instances for which the parameters are known. On the other hand, we speculate that the global behavior presented in this paper is representative for a large class of problem instances. We therefore conclude that the analysis might prove useful for deducing global estimates for the average-case performance of the algorithm when applied to this class of problem instances.

More detailed estimates of the average-case performance of the simulated annealing algorithm can only be deduced from a rigorous performance analysis which takes into account the detailed structure of the optimization problem at hand. To our knowledge such a theoretical average-case performance analysis is not known in the literature[3] and is therefore considered as an open problem.

4. CONCLUSIONS

The typical behavior of the simulated annealing algorithm is discussed by analyzing the expectation and the variance of the cost function as a function of the control parameter for a given instance of the traveling salesman problem. The observed behavior can be modeled by assuming that the configuration density contains two components, each component being dominant in a different region of the control parameter. In the region of a few standard deviations around the average cost (over all configurations) the configuration density is approximated by a normal distribution. This component dominates the observed configuration density at large values of the control parameter. In the region close to the optimum cost the configuration density is approximated by an exponential distribution. This component becomes dominant at small values of the control parameter. The expectation and variance of the cost function can be effectively parametrized by simple functions.

The average-case performance of the simulated annealing algorithm

[3] A theoretical upper bound on the quality of the final solution is only known for the maximum matching problem.[16]

for the problem instance at hand is investigated by analyzing the expectation and the variance of the final cost obtained by the algorithm as a function of the distance parameter that governs the decrement of the cooling control parameter. It is concluded that the average-case performance of the simulated annealing algorithm can be explained by assuming that the deviation of the final cost from the optimum cost is distributed according to a gamma distribution.

For computational reasons it has not been possible to check whether the obtained results allow one to predict accurately the expected performance of the algorithm for larger instances than the one studied in this paper (the experiments described here took a few hundred hours of CPU time on a VAX-780). However, since the performance of the simulated annealing algorithm is reported by many to be more or less independent of the problem to which it is applied (see Ref. 10), we are confident that our results can be used for a wide class of problems represented by the problem instance used in the analysis.

APPENDIX

The 100 European-city TSP is a symmetric instance of the traveling salesman problem defined on the following set of cities:

1 Amsterdam	26 Dortmund	51 London	76 Roma
2 Antwerpen	27 Dresden	52 Luxembourg	77 Rostock
3 Athinai	28 Dublin	53 Lyon	78 Rotterdam
4 Barcelona	29 Düsseldorf	54 Madrid	79 Sarajevo
5 Basel	30 Edinburgh	55 Magdeburg	80 Sevilla
6 Belfast	31 Gdansk	56 Malaga	81 Sheffield
7 Beograd	32 Genova	57 Malmö	82 Skopje
8 Bergen	33 Glasgow	58 Manchester	83 Smolensk
9 Berlin	34 Göteborg	59 Marseille	84 Sofija
10 Bern	35 Granada	60 Milano	85 Southampton
11 Bilbao	36 Graz	61 Minsk	86 Split
12 Birmingham	37 Hamburg	62 Monaco	87 Stockholm
13 Bonn	38 Hannover	63 Moskva	88 Strasbourg
14 Bordeaux	39 Helsinki	64 München	89 Stuttgart
15 Bratislava	40 Istanbul	65 Napoli	90 Thessaloniki
16 Bremen	41 Köln	66 Nice	91 Torino
17 Brno	42 København	67 Odessa	92 Toulouse
18 Bruxelles	43 Krakow	68 Oslo	93 Trieste
19 Bucuresti	44 Leeds	69 Palermo	94 Turku
20 Budapest	45 Leipzig	70 Paris	95 Uppsala
21 Burgas	46 Leningrad	71 Plovdiv	96 Valencia
22 Constanta	47 Liège	72 Plzen	97 Warszawa
23 Cork	48 Lisboa	73 Pôrto	98 Wien
24 Craiova	49 Liverpool	74 Praha	99 Zagreb
25 Den Haag	50 Lódz	75 Riga	100 Zürich

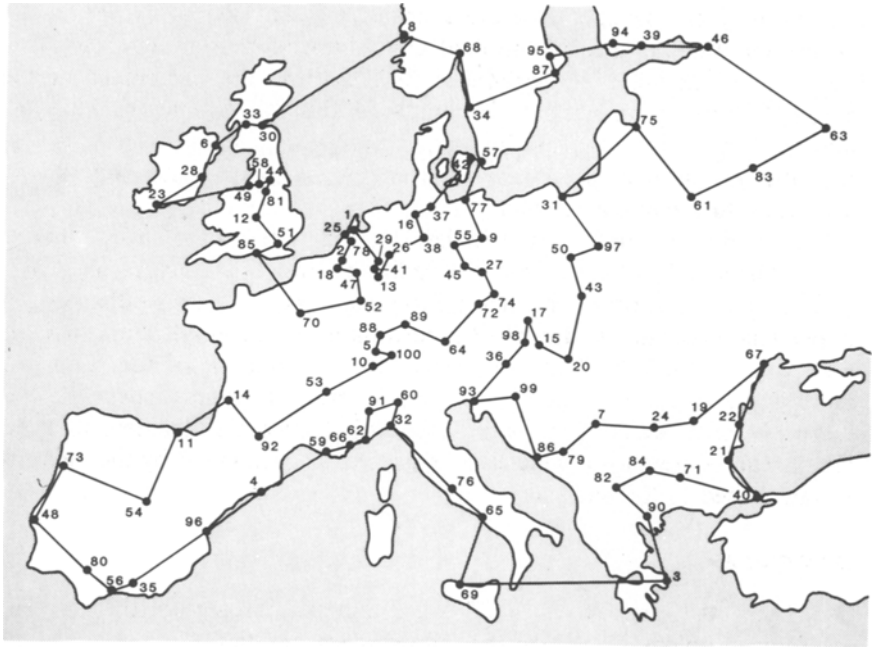

Fig. 5. The 100-European-city traveling salesman problem. The solid lines indicate an optimal tour.

A listing of the distance matrix that goes with the problem instance is available from the authors. Figure 5 shows the location of the 100 cities together with a minimal tour whose length equals 21,134 (km).

REFERENCES

1. E. H. L. Aarts and P. J. M. van Laarhoven, *Philips J. Res.* **40**:193–226 (1985).
2. E. H. L. Aarts and P. J. M. van Laarhoven, in *Proceedings International Conference on Computer-Aided Design* (Santa Clara, November 1985), pp. 206–208.
3. K. Binder, *Monte Carlo Methods in Statistical Physics* (Springer-Verlag, Berlin, 1986).
4. V. Černy, *J. Optim. Theory Appl.* **45**:41–51 (1985).
5. G. S. Grest, C. M. Soukoulis, and K. Levin, *Phys. Rev. Lett.* **56**:1148–1151 (1986).
6. B. Hajek, in *Proceedings 24th Conference on Decision and Control* (Ft. Lauderdale, December 1985), pp. 755–760.
7. R. Jonker and T. Volgenant, *Oper. Res.* **32**:837–846 (1984).
8. S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Science* **220**:671–680 (1983).
9. S. Kirkpatrick and G. Toulouse, *J. Phys. (Paris)* **46**:1277–1292 (1985).
10. P. J. M. van Laarhoven and E. H. L. Aarts, *Simulated Annealing: Theory and Applications* (Kluwer, Dordrecht, 1987).

11. E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, and D. B. Shmoys, *The Traveling Salesman Problem* (Wiley, Chichester, 1985).
12. S. Lin and B. W. Kernighan, *Oper. Res.* **21**:498–516 (1973).
13. M. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, *J. Chem. Phys.* **21**:1087–1092 (1953).
14. E. Palmer, *Graphical Evolution* (Wiley Interscience, New York, 1985).
15. F. Romeo and A. L. Sangiovanni-Vincentelli, in *Proceedings 1985 Chapel Hill Conference on VLSI* (May 1985), pp. 393–417.
16. G. H. Sasaki and B. Hajek, *J. ACM*, to appear.
17. S. Solla, G. Sorkin, and S. White, in *Disordered Systems and Biological Organization* (Springer-Verlag, Berlin, 1985).
18. S. R. White, in *Proceedings International Conference Computer Design* (Port Chester, November 1984), pp. 646–651.