



# A Note on the Finite Time Behaviour of Simulated Annealing

Andreas Nolte<sup>†</sup> and Rainer Schrader<sup>‡</sup>

Revised Version

March 1999

Simulated Annealing has proven to be a very successful heuristic for various combinatorial optimization problems. It is a randomized algorithm that attempts to find the global optimum with high probability by local exchanges. In this paper we give a new proof of the convergence of Simulated Annealing by applying results about rapidly mixing Markov chains. With this proof technique it is possible to obtain better bounds for the finite time behaviour of Simulated Annealing than previously known.

## 1 Introduction

Simulated Annealing was first introduced by KIRKPATRICK ET AL. [8] as a random local search heuristic to solve large combinatorial optimization problems. With respect to many excellent experimental results [7] it is a general method that yields most times good performance results even when dealing with NP-hard problems.

An instance of a combinatorial optimization problem can be formalized as a pair  $(\Omega, c)$ .  $\Omega$  is the set of configurations,  $c : \Omega \rightarrow \mathbb{R}_+$  is the cost function, that assigns costs to each element of  $\Omega$ . We assume that we are dealing with a minimization problem, i.e. we are looking for a configuration with minimal cost. The set of global minima is called  $\Omega_{\min}$ .

---

<sup>\*</sup>Supported by DFG-grant (Schr 390/2-1)

<sup>†</sup>Universität zu Köln, Weyertal 80, 50931 Köln, Germany; email: anolte@zpr.uni-koeln.de

<sup>‡</sup>Universität zu Köln, Weyertal 80, 50931 Köln, Germany; email: rs@zpr.uni-koeln.de

Simulated Annealing explores the configuration space starting from an arbitrary chosen configuration and generates a new one according to a certain neighborhood structure  $N \subset \Omega^2$ . According to the initial implementation of Simulated Annealing we assume that the neighborhood structure is symmetric and the probability to generate a certain neighboring configuration is  $\alpha \in [0, 1]$ . Furthermore, we take it for granted that the underlying graph induced by the neighborhood structure is strongly connected. This implies that for every pair  $f, g \in \Omega$  there exist a path  $f = v_1, v_2, \dots, v_l = g$  with  $(v_i, v_{i+1}) \in N$ .  $d(f, g)$  denotes the length of the shortest of these paths. Every time a new solution is generated, its cost is evaluated and the new solution is either accepted or rejected according to an acceptance rule.

The following algorithm shows the general structure of Simulated Annealing.

```

simulated annealing( $X_0, T_0$ )
(given: an initial state  $X_0$  and an initial value  $T_0$ )
 $t = 0$ ;
 $k = 0$ ;
while (outer-loop criterion not satisfied)
  while (inner-loop criterion not satisfied)
     $X = \text{generate}(X_k)$ ;
    if  $\text{accept}(X, X_k, T_t)$   $X_{k+1} = X$ 
    else  $X_{k+1} = X_k$ ;
     $k = k + 1$ ;
  end(while);
   $T_{t+1} = \text{update}(T_t)$ ;
   $t = t + 1$ ;
end(while);
end;
```

The acceptance rule is defined in such a way that it is also possible to accept worse configurations with higher cost than the current configuration. This is necessary in order not to get stuck in a local optimum. But the probability to accept worse configurations is lowered during the algorithm. This probability is driven by a so called cooling schedule, i.e. a monotonically decreasing sequence of temperatures  $(T_t)_{t \in \mathbb{N}}$  with  $T_t \rightarrow 0$  as  $t \rightarrow \infty$ . The acceptance function has the following structure:

```

accept( $f, g, T_t$ )
   $r = \text{random number in } [0, 1]$ ;
  if  $r \leq e^{(c(f) - c(g))/T_t}$  return 'accepted'
  else return 'not accepted';
end;
```

Coming to a rigorous mathematical model, we take the variable  $X_k$  in the algorithm as a random variable of the configuration space in the natural

way. Since the next configuration depends only on the current configuration, the best suitable mathematical model to examine Simulated Annealing is an inhomogeneous Markov chain.

With respect to the above defined strongly connected neighborhood structure  $N \subset \Omega^2$  we get the following transition probabilities. For technical reasons, we also assume that we stay at each state with a probability of at least  $1/2$ , so that we obtain:

$$\forall (f, g) \in \Omega^2$$

$$p_{fg}^{(t)} = \begin{cases} \min\{1, e^{\frac{c(f)-c(g)}{T_t}}\} \frac{\alpha}{2} & \text{for } f \neq g, (f, g) \in N \\ 0 & \text{for } f \neq g, (f, g) \notin N \\ 1 - \sum_{f' \neq f} p_{ff'}^{(t)} & \text{for } f = g \end{cases}$$

Clearly  $p_{ff}^{(t)} \geq 1/2$ , so that the chain is aperiodic. Assuming fixed transition probabilities  $p_{fg}^{(t)}$  (fixed temperature  $T_t$ ) the resulting homogeneous chain is also ergodic and

$$\pi(t)_f = \frac{e^{-\frac{c(f)}{T_t}}}{\sum_g e^{-\frac{c(g)}{T_t}}}$$

is the stationary distribution. Let  $c_{\min} = \min\{c(f) \mid f \in \Omega\}$ . Obviously

$$\pi_f^* = \lim_{t \rightarrow \infty} \pi(t)_f = \lim_{t \rightarrow \infty} \frac{1}{\sum_g e^{\frac{c(f)-c(g)}{T_t}}} = \begin{cases} 0, & \text{if } c(f) > c_{\min} \\ \frac{1}{|\{f \mid c(f) = c_{\min}\}|}, & \text{if } c(f) = c_{\min} \end{cases}.$$

The limit distribution is therefore a constant probability vector, which probability charges are concentrated on the global optima of the solution space. Many asymptotic convergence results, already published in the mid 80's, are concerned with the choice of an appropriate cooling schedule. This schedule should guarantee the convergence of the probability distribution to  $\pi^*$  as  $t \rightarrow \infty$ . One of the most remarkable is that of HAJEK [5]:

The depth of a local minimum  $f$  is taken to be the smallest number  $d(f)$  such that there is a configuration  $g$  with  $c(f) > c(g)$  reachable at height  $c(f) + d(f)$  from  $f$ . This means that there is a path from  $f$  to  $g$  using a configuration  $h$  with  $c(h) = c(f) + d(f)$  and the maximal cost of a configuration in every path is at least  $c(f) + d(f)$ . This implies

$$\lim_{t \rightarrow \infty} P(X_t \in \Omega_{\min}) = 1 \quad \Leftrightarrow \quad \sum_{t=1}^{\infty} \exp\left(-\frac{D}{T_t}\right) = \infty, \quad (1)$$

where  $D$  is the maximum of all depth  $d(f)$  of all configurations  $f$ , that are local but not global minima. Therefore, HAJEK suggested a logarithmic cooling schedule  $T_t = \frac{\Gamma}{\log(t)}$  with  $\Gamma \geq D$ .

In addition, many other authors (see e.g. [9] for an overview) were able to prove asymptotic convergence with a logarithmic cooling schedule and different, most times easier accessible values of  $\Gamma$ . But, infinitely many steps are quite difficult to perform. Therefore, finite type bounds of the proximity of the probability distribution  $P(t)$  of the configuration space after generation of  $t$  transitions to the uniform distribution on the set of optimal states  $\pi^*$  are of special interest.

ANILY and FEDERGRUEN [1], GIDAS [4] and MITRA ET AL. [13] have obtained independently similar deterministic upper bounds of this proximity using a logarithmic cooling schedule  $T_t = \Gamma / \ln(t)$  with  $\Gamma$  chosen according to HAJEK'S result. To the best of our knowledge the bound of MITRA ET AL. [13] is the best published so far:

Let  $\Omega^*$  be the set of local minima and  $r = \min_{f \in \Omega \setminus \Omega^*} \max_{g \in \Omega} d(f, g)$  be the radius of the underlying graph. Then we define  $a = \frac{1}{r} (\alpha)^r$  and  $b = \frac{\delta}{r\Delta}$ , where  $\delta$  is the difference between the minimal cost and the next to least cost value and  $\Delta$  is the maximal cost difference between two neighboring configurations. Then there exists  $c_1 \in \mathbb{N}$  with

$$t \geq c_1 \left( \frac{1}{\epsilon} \right)^{\frac{1}{\min\{a,b\}}} \Rightarrow \sum_{f \in \Omega} |P(t)_f - \pi_f^*| \leq \epsilon$$

For the  $m$ -city travelling salesman problem this bound would be of size  $\epsilon^{-m^{2m+1}}$  [9].

In the following we give a new proof for the convergence of the distribution of Simulated Annealing using a logarithmic cooling schedule  $T_t = \gamma / \ln(t)$  to the distribution  $\pi^*$  that is concentrated on the optimal states. Our proof technique enables us to give considerably better finite time bounds than established so far. Our bound for the  $n$ -city travelling salesman problem for example would be  $\frac{2c_1 m^2 \log(m)}{\epsilon^{c_2 m}}$  and thus the super-exponent in the result of MITRA is removed. As an easy consequence we can derive that Simulated Annealing is with high probability in a global minimum after exponentially many steps.

## 2 Convergence Bounds

The main inequality of this section is stated in the following theorem:

**Theorem 1** *The distribution  $P(t)$  of the inhomogeneous Markov chain, that corresponds to Simulated Annealing with logarithmic cooling schedule, satisfies for sufficiently large  $t$ :*

$$\|P(t+1) - \pi(t+1)\|_{t+1} \leq \frac{1}{\sqrt{t}} c(G),$$

where  $c(G)$  is a problem-dependent parameter and  $\|x\|_t = \sum_f \frac{x_f^2}{\pi(t)_f}$ .

The proof is technical and uses a recent result of MIHAIL [12] about *homogeneous* Markov chains. It relates the difference of the distribution at time  $t$  and the stationary distribution to the difference of the distribution at time  $t - 1$  and the stationary distribution. The main tool is SINCLAIR and JERRUM's [6] conductance

$$\Phi_p = \min_{A \subset \Omega, \sum_{u \in A} \pi_u \leq 1/2} \frac{\sum_{u \in A} \sum_{v \notin A} \pi_u p_{uv}}{\sum_{u \in A} \pi_u} \quad (2)$$

of an ergodic Markov chain with transition probabilities  $p_{uv}$  and stationary distribution  $\pi$ . The mathematics behind this approach dates back to CHEEGER [2] and LAWLER and SOKAL [10]. SINCLAIR and JERRUM's [6] used CHEEGER's inequality to derive a result for the running time of random walks on graphs based on their conductance properties. Later MIHAIL [12] proved a version of this running time bound basically without linear algebra tools and this version is especially helpful here.

*Proof:*

$$\begin{aligned} & \sqrt{\|P(t) - \pi(t+1)\|_{t+1}} \\ &= \sqrt{\sum_u \frac{(P(t)_u - \pi(t+1)_u)^2}{\pi(t+1)_u}} \\ &= \left\| \left( \frac{P(t)_u - \pi(t)_u + \pi(t)_u - \pi(t+1)_u}{\sqrt{\pi(t+1)_u}} \right)_u \right\|_2 \\ &\leq \left\| \left( \frac{P(t)_u - \pi(t)_u}{\sqrt{\pi(t+1)_u}} \right)_u \right\|_2 + \left\| \left( \frac{\pi(t)_u - \pi(t+1)_u}{\sqrt{\pi(t+1)_u}} \right)_u \right\|_2 \\ &\leq \sqrt{\sum_u \frac{(P(t)_u - \pi(t)_u)^2}{\pi(t+1)_u}} + \sqrt{\sum_u \frac{(\pi(t)_u - \pi(t+1)_u)^2}{\pi(t+1)_u}}. \end{aligned} \quad (3)$$

In order to estimate the difference between the two stationary distributions we consider

$$\frac{\pi(t)_f}{\pi(t+1)_f} = \frac{\sum_g (t+1)^{\frac{c(f)-c(g)}{\gamma}}}{\sum_g t^{\frac{c(f)-c(g)}{\gamma}}} \leq (1 + 1/t)^{c_{\max}/\gamma} \quad (4)$$

due to

$$\frac{(t+1)^{\frac{c(f)-c(g)}{\gamma}}}{t^{\frac{c(f)-c(g)}{\gamma}}} = (1 + 1/t)^{\frac{c(f)-c(g)}{\gamma}} \leq (1 + 1/t)^{c_{\max}/\gamma},$$

where  $c_{\max} = \max\{c(f)|f \in \Omega\}$ . Applying the main result of MIHAIL [12] we get

$$\|P(t+1) - \pi(t+1)\|_{t+1} \leq (1 - \Phi(t+1)^2)\|P(t) - \pi(t+1)\|_{t+1}$$

( $\Phi(t+1)$  is the conductance of the homogeneous chain with constant transition matrix  $p^{t+1}$  and stationary distribution  $\pi(t+1)$ ) and therefore from (3):

$$\begin{aligned} & \sqrt{\|P(t+1) - \pi(t+1)\|_{t+1}} \\ & \leq \sqrt{(1 - \Phi(t+1)^2)} \left( \sqrt{(1 + 1/t)^{c_{\max}/\gamma} \|P(t) - \pi(t)\|_t} \right. \\ & \quad \left. + \sqrt{\sum_u \frac{(\pi(t)_u - \pi(t+1)_u)^2}{\pi(t+1)_u}} \right). \end{aligned} \quad (5)$$

The next step is to relate the time-dependent conductance  $\Phi(t)$  to the conductance  $\Phi$  in the uniform case (uniform distribution:  $\pi_u = 1/|\Omega|$ ):

We get for all  $t \in \mathbb{N}$ :

$$\begin{aligned} \Phi(t) &= \frac{\sum_{f \in A} \sum_{g \notin A} \pi(t)_f p_{fg}^{(t)}}{\sum_{h \in A} \pi(t)_h} \\ &= \frac{\sum_{f \in A, g \notin A, (f,g) \in N} \min\{t^{-c(f)/\gamma}, t^{-c(g)/\gamma}\}}{2n \sum_{h \in A} t^{-c(h)/\gamma}} \\ &\geq t^{-c_{\max}/\gamma} \frac{\#\{(f,g) \in N | f \in A, g \notin A\}}{2n|A|} \\ &\geq \Phi t^{-c_{\max}/\gamma} \end{aligned}$$

and therefore

$$\sqrt{1 - \Phi(t+1)^2} \leq 1 - \frac{\Phi^2}{2t^{\frac{3c_{\max}}{\gamma}}}. \quad (6)$$

By choosing  $\gamma = 48c_{\max}$ , we get with  $t \geq \left(\frac{2}{\Phi^2}\right)^{16}$ :

$$1 - \frac{\Phi^2}{2t^{3c_{\max}/\gamma}} \leq 1 - \frac{1}{t^{1/8}}. \quad (7)$$

So we can obtain an estimation of the first factor in (5):

$$\sqrt{(1 - (\Phi(t+1)^2)(1 + \frac{1}{t})^{c_{\max}/\gamma}} \leq 1 - \frac{1}{2t^{1/4}} \quad (8)$$

Next we consider the second part of the sum in (5):

$$\begin{aligned} \sqrt{\sum_u \frac{(\pi(t)_u - \pi(t+1)_u)^2}{\pi(t+1)_u}} &= \sqrt{\sum_u \left( \frac{\pi(t)_u^2}{\pi(t+1)_u} - 2\pi(t)_u + \pi(t+1)_u \right)} \\ &\stackrel{(4)}{\leq} \sqrt{(1 + 1/t)^{c_{\max}/\gamma} - 1} \\ &\leq \sqrt{1/t}. \end{aligned} \quad (9)$$

Thus we get the following conclusion:

$$\begin{aligned} \sqrt{\|P(t) - \pi(t)\|_t} &\geq 4t^{-1/4} \\ \Rightarrow -\frac{1}{4t^{1/4}}\sqrt{\|P(t) - \pi(t)\|_t} + \sqrt{\sum_u \frac{(\pi(t)_u - \pi(t+1)_u)^2}{\pi(t+1)_u}} &\leq 0. \end{aligned} \quad (10)$$

Finally we can establish the last estimation:

**Lemma 2** *Let  $t_0 = \left(\frac{2}{\Phi^2}\right)^{16}$  and  $\gamma = 48c_{\max}$ . This implies the following inequality for all  $t \geq t_0$ :*

$$\sqrt{\|P(t+1) - \pi(t+1)\|_{t+1}} \leq \frac{1}{\sqrt[4]{t}} \max \left\{ 5, \frac{16}{\Phi^8} \sqrt{\|P(t_0+1) - \pi(t_0+1)\|_{t_0+1}} \right\}.$$

*Proof:* For  $t = t_0$ , the claim is obvious. For  $t > t_0$  we distinguish two cases.

(i)  $\sqrt{\|P(t) - \pi(t)\|_t} \geq 4t^{-1/4}$ . The proof follows with (5) and (10) by induction.

$$\begin{aligned} \sqrt{\|P(t+1) - \pi(t+1)\|_{t+1}} &\leq \left(1 - \frac{1}{4t^{1/4}}\right) \sqrt{\|P(t) - \pi(t)\|_t} \\ &\leq \left(1 - \frac{1}{t}\right) \sqrt{\|P(t) - \pi(t)\|_t} \\ &\leq \left(1 - \frac{1}{t}\right) \frac{1}{\sqrt[4]{t-1}} \max \left\{ 5, \frac{16}{\Phi^8} \sqrt{\|P(t_0+1) - \pi(t_0+1)\|_{t_0+1}} \right\} \\ &\leq \frac{1}{\sqrt[4]{t}} \max \left\{ 5, \frac{16}{\Phi^8} \sqrt{\|P(t_0+1) - \pi(t_0+1)\|_{t_0+1}} \right\} \end{aligned}$$

(ii)  $\sqrt{\|P(t) - \pi(t)\|_t} \leq 4t^{-1/4}$ . Straightforward calculation with the help of (9).

$$\begin{aligned} \sqrt{\|P(t) - \pi(t)\|_t} + \sqrt{\sum_u \frac{(\pi(t)_u - \pi(t+1)_u)^2}{\pi(t+1)_u}} &\leq \frac{4}{t^{1/4}} + \sqrt{\frac{1}{t}} \\ &\leq \frac{5}{t^{1/4}} \end{aligned}$$

The main theorem follows with

$$c(G) = \max^2 \left\{ 5, \frac{16}{\Phi^8} \|P(t_0+1) - \pi(t_0+1)\|_{t_0+1} \right\}$$

immediately from the last lemma.  $\square$

Now it is easy to derive the following corollary:



**Corollary 3** *The distribution  $P(t)$  of the Simulated Annealing Markov chain converges to the constant probability vector  $\pi^*$ , of which probability charges are concentrated uniformly on the global optima.*

*Proof:*

$$\begin{aligned} \lim_{t \rightarrow \infty} \|P(t) - \Pi^*\|_2 &\leq \lim_{t \rightarrow \infty} \|P(t) - \pi(t)\|_2 + \lim_{t \rightarrow \infty} \|\pi(t) - \pi^*\|_2 \\ &\leq \lim_{t \rightarrow \infty} \sqrt{\|P(t) - \pi(t)\|_t} + 0 \\ &\leq 0 \end{aligned}$$

□

With Lemma 2 we can also derive finite time bounds for Simulated Annealing:

**Proposition 4** *Let  $\delta$  be the difference between the minimal cost and the next to least cost value and  $c_{\max}$  the maximal value of the cost function. Then there exist constants  $c_1, c_2 \in \mathbb{N}$  with*

$$t \geq \frac{1}{\Phi^{c_2}} \left( \frac{|\Omega|}{\epsilon} \right)^{-c_1 \frac{c_{\max}}{\delta}} \Rightarrow \sum_f |P(X_t = f) - \pi_f^*| \leq \epsilon.$$

*Proof:* Let  $f \in \Omega_{\min}$ . Then we get:

$$|\pi_f(t) - \pi_f^*| \leq \frac{1}{|\Omega_{\min}|} - \frac{1}{|\Omega_{\min}| + t^{-\delta/\gamma} |\Omega|}$$

and

$$t \geq \left( \frac{4|\Omega|}{|\Omega_{\min}| \epsilon} \right)^{\gamma/\delta} \Rightarrow \frac{1}{|\Omega_{\min}|} - \frac{1}{|\Omega_{\min}| + t^{-\delta/\gamma} |\Omega|} \leq \frac{\epsilon}{4|\Omega_{\min}|}.$$

Now let  $f \notin \Omega_{\min}$ . Here we get:

$$t \geq \left( \frac{4|\Omega|}{|\Omega_{\min}| \epsilon} \right)^{\gamma/\delta} \Rightarrow |\pi_f(t) - \pi_f^*| \leq \frac{\epsilon}{4|\Omega|}.$$

Hence

$$\sum_{f \in \Omega} |\pi_f(t) - \pi_f^*| \leq \epsilon/2.$$

Furthermore we get with Lemma 2:

$$\sum_f |P(X_t = f) - \pi_f(t)| \leq \sqrt{|\Omega|} \frac{1}{\sqrt[4]{t-1}} \max \left\{ 5, \frac{16}{\Phi^8} \sqrt{\|P(t_0 + 1) - \pi(t_0 + 1)\|_{t_0+1}} \right\}$$

and

$$\|P(t_0 + 1) - \pi(t_0 + 1)\|_{t_0+1} \leq |\Omega|(t_0 + 1)^{c_{\max}/\gamma}.$$

Therefore we get

$$t \geq \left( \frac{32|\Omega|^{3/2}(t_0 + 1)}{\epsilon \Phi^8} \right)^4 \Rightarrow \sum_f |P(X_t = f) - \pi_f(t)| \leq \epsilon/2.$$

and the proposition follows from the above calculations.  $\square$

In the following we will describe an application of Simulated Annealing to the traveling salesman problem, a well known strongly NP-hard combinatorial optimization problem [3]. We consider only a restricted, but still NP-hard version of this problem with special distance values to get a proper estimation for  $\delta$  and  $c_{\max}$ .

We are given a set  $C$  of  $m$  cities and distances  $d(c_i, c_j) \in \{1, 2\}$  for each pair of cities  $c_i, c_j \in C$ . We are looking for the shortest tour that connects all these cities. Given a starting tour, we consider the generation method for neighboring tours known as 2opt-transitions. We select two cities in the present tour uniformly at random with probability  $\frac{1}{m(m-1)}$  and reverse the order in which these cities in between this pair of cities are visited. Therefore we get with Proposition 4:  $\exists c_1, c_2 \in \mathbb{N}$

$$t \geq \frac{2^{c_1} m^2 \log(m)}{\epsilon^{c_2 m}} \Rightarrow \sum_f |P(X_t = f) - \pi_f(t)| \leq \epsilon.$$

This bound is considerably smaller than the best bound  $\epsilon^{-m^{2^{m+1}}}$  published so far by MITRA ET AL. [13]. A similar result yields the application of Simulated Annealing to graph coloring [15], where the super-exponent can also be removed.

As an easy consequence we could estimate the probability to be in an optimal state after sufficiently many steps:

**Corollary 5** *The Markov chain is in an optimal state (i.e. optimal tour) with probability of at least 3/4 after exponentially many steps.*

*Proof:*

$$\begin{aligned} P(X_t \in \Omega_{\min}) &\geq \sum_{f \in \Omega_{\min}} \pi_f^* - |P(X_t \in \Omega_{\min}) - \sum_{f \in \Omega_{\min}} \pi_f^*| \\ &\geq 1 - \sum_{f \in \Omega_{\min}} |P(X_t = f) - \pi_f^*| \end{aligned}$$

$\square$

### 3 Concluding remarks

We can improve the finite time bounds of the proximity of the probability distribution  $P(t)$  after  $t$  transitions to the uniform distribution on the set of optimal states  $\pi^*$ . Thus, we improve the bound for the finite time behaviour of Simulated Annealing considerably. But, this bound is rather poor for practical purposes. This is broadly in line with practical experiences [9] with logarithmic cooling schedules that ensure the convergence, however they are far too slow for a practical implementation. Corollary 5 states that Simulated Annealing is in an optimal state with high probability after a number of steps that exceeds the cardinality of the configuration space. Therefore, it would be less time consuming to enumerate the whole configuration space. On the other hand we need not to bound the difference  $|P(t) - \pi_f^*|$  summarized over the whole configuration space to get Corollary 5. This implies that the exponential size of the bound is possibly not necessary, when one does not summarize over the whole configuration space of exponential size. As a consequence, despite the technical difficulties, considering the maximum norm is possibly better to obtain results improving those in Corollary 5. But there is no hope at least in the general case to prove sub-exponential bounds for the performance of Simulated Annealing. The reason is the following: It is possible to prove for an application of Simulated Annealing to a specially constructed instance of the 3-coloring problem and arbitrary cooling schedule [15]:

$$P(X_t \in \Omega_{\text{opt}}) \leq te^{-n},$$

where  $n$  is the the number of vertices to be colored.

### 4 Acknowledgements

The authors are grateful to the anonymous referees for their helpful comments and suggestions.

### References

- [1] Anily, S. and Federgruen, A.: *Ergodicity in Parametric Nonstationary Markov Chains: An Application to Simulated Annealing Methods*, Oper. Res. 35, 1987
- [2] Cheeger, J.: *A lower bound for the smallest eigenvalue of the Laplacian*, Problems in Analysis, Princeton University Press, New Jersey, 1970

- [3] Garey, M.R.; Johnson, S.J.: *Computers and Intractability*, W.H. Freeman and Company, 1979
- [4] Gidas, B.: *Nonstationary Markov Chains and Convergence of the Annealing Algorithm*, J. Statist. Phys. 39, 1985
- [5] Hajek, B.: *Cooling Schedules for Optimal Annealing*, Mathematics of Operations Research 13, 1988
- [6] Jerrum, J; Sinclair, A.: *Approximating the permanent*, Proceedings of the 20th ACM-Symposium on Theory of Computing, 1988.
- [7] Johnson, D.S. et al.: *Optimization by Simulated Annealing: An Experimental Evaluation*, Operations Research 39, 1991
- [8] Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P.: *Optimization by Simulated Annealing*, Science 220, 1983.
- [9] Laarhoven, P.J.M.; Aarts, E.H.L.: *Simulated Annealing: Theory and Applications*, Kluwer Academic Publishers, 1989.
- [10] Lawler, G.F.; Sokal, A.D.: *Bounds on the  $L^2$  Spectrum for Markov Chains and Markov Processes, A Generalization of Cheeger's Inequality*, Trans. Amer. Math. Soc. 309, 1988
- [11] Metropolis, N. et al.: *Equations of state calculations by fast computer machines*, Journal of Chemical Physics 21, 1953
- [12] Mihail, M.: *Conductance and Convergence of Markov chains- A combinatorial treatment of expanders*, Proceedings of the 30th Annual Symposium on Foundations of Computer Science, 1989.
- [13] Mitra, D.; Romeo, F.; Sangiovanni-Vincentelli, A.L.: *Convergence and Finite Time Behaviour of Simulated Annealing*, Adv. Appl. Prob. 18, 1986
- [14] Nolte, A., Schrader, R.: *A Note on the Finite Time Behaviour of Simulated Annealing*, Operations Research Proceedings 1996 (SOR 1996), Springer
- [15] Nolte, A.; Schrader, R.: *Simulated Annealing and its problems to color graphs*, Proceedings of the ESA 96, Lecture Notes in Computer Science 1136, Springer
- [16] Tsitsiklis, J.: *Markov Chains with Rare Transitions and Simulated Annealing*, Mathematics of Operations Research 14, 1989