Appl Math Optim 24:183–195 (1991)

Applied Mathematics and Optimization

On the Convergence of "Threshold Accepting"

Ingo Althöfer and Klaus-Uwe Koschnick

Fakultät für Mathematik, Universität Bielefeld,
Postfach 8640, W-4800 Bielefeld 1, Federal Republic of Germany

Abstract. Simulated Annealing (SA) has become a very popular tool in combinatorial optimization since its introduction in 1982. Recently Dueck and Scheuer proposed another simple modification of local search which they called "Threshold Accepting" (TA). In this paper some convergence results for TA are presented. The proofs are not constructive and make use of the fact that in a certain sense "SA belongs to the convex hull of TA".

1. Introduction

An instance of a combinatorial optimization problem consists of a finite set $\mathscr{R}$ of configurations and a cost function $C: R \to \mathbb{R}$ which assigns a real number to each configuration. The task is to find a configuration for which C takes its minimum value C_{opt}. $\mathscr{R}_{\text{opt}}$ denotes the set of all configurations with this minimum value C_{opt}. When a neighborhood $\mathscr{R}_i \subset \mathscr{R}$ for each configuration $i \in \mathscr{R}$ exists, a *local search* can be performed:

Starting at a given configuration, a sequence of iterations is made, each iteration consisting of a possible transition from the current configuration i to a configuration j selected at random from the neighborhood $\mathscr{R}_i$ of i. If j has a lower cost, i is replaced by j, otherwise another neighbor of i is selected and compared for its cost value. The algorithm terminates when a configuration is obtained whose cost is no worse than any of its neighbors.

Local search (also known as *iterative improvement*) often terminates in a local minimum which is suboptimal. The stochastic optimization algorithm *Simulated Annealing* (SA) overcomes this local-search problem of being trapped in suboptimal local minima by a more liberal acceptance rule:

Let i be the current configuration at time k and let j be the selected neighbor.

Then j becomes the next configuration in the sequence,

$$\begin{cases} \text{with probability 1} & \text{if } C(j) < C(i), \\ \text{with probability } \exp\left(\dfrac{C(i) - C(j)}{c_k}\right) & \text{if } C(j) \geq C(i). \end{cases}$$

Thus, there is a nonzero probability of continuing with a configuration with higher cost than the current configuration. $c_k > 0$ is called the temperature at time k. Typically, the sequence $(c_k)_{k=0}^{K}$ of temperatures is chosen nonincreasingly. SA has become very popular since its introduction by Kirkpatrick *et al.* in 1982 [7], [8]. The book by van Laarhoven and Aarts [9] gives a survey of theory and applications of SA.

Recently Dueck and Scheuer proposed another more simple modification of local search which they called *Threshold Accepting* (TA). As presented in [1], TA finds very-near-to-optimum tours for the famous 442-city traveling-salesman problem of Grötschel [5] within 3 or 4 minutes of CPU time and has been used to construct good error-correcting codes.

TA works as follows: Let i be the current configuration at time k and let j be the selected neighbor. Then j becomes the next configuration in the sequence,

$$\begin{cases} \text{with probability 1} & \text{if } C(j) - C(i) \leq T_k, \\ \text{with probability 0} & \text{if } C(j) - C(i) > T_k. \end{cases}$$

Thus, in contrast to SA, the acceptance rule of TA for new configurations works deterministically at all times k. The threshold sequence $(T_k)_{k=0}^{K}$ consists of nonnegative real numbers and has always been chosen nonincreasingly by Dueck and Scheuer.

In this paper we prove some convergence results for TA. Section 2 contains the description of the theoretical model and the results. In Section 3 the proofs are worked out. Section 4 consists of some concluding remarks.

2. The Theoretical Model and the Results

The theoretical analysis of SA is conveniently done in terms of Markov chains. The conditional probability $P_{ij}(k-1, k)$ denotes the probability that the kth transition is to configuration j, if i is the current configuration at time $k-1$. $X(k)$ denotes the configuration obtained after k iterations. $P_{ij}(k-1, k)$ is called the transition probability and the $|\mathscr{R}| \times |\mathscr{R}|$ matrix $P(k-1, k)$ is called the transition matrix. The entries of the transition matrix depend on the control parameter c_{k-1}. Figure 2.1 shows an example: $\mathscr{R} = \{u, v, x, y, z\}$, $\mathscr{R}_u = \{v\}$, $\mathscr{R}_v = \{u, x\}$, $\mathscr{R}_x = \{v, y\}$, $\mathscr{R}_y = \{x, z\}$, $\mathscr{R}_z = \{y\}$, $C(u) = 0$, $C(v) = 3$, $C(x) = 2$, $C(y) = 4$, $C(z) = 1$, $C_{\text{opt}} = 0$, $\mathscr{R}_{\text{opt}} = \{u\}$. For a fixed parameter c_{k-1} the transition matrix is shown in Table 2.1.

In this example we assumed that for a current configuration i with $|\mathscr{R}_i| = n$ every neighbor has the same *a priori* probability $1/n$ to become the candidate for the next transition. Our results below, however, hold for any other type of transition probabilities as well.

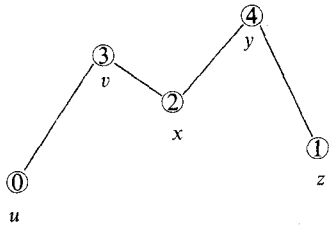

Figure 2.1

The starting configuration $X(0)$ is chosen at random, thus $\Pr\{X(0) = i\} = 1/|\mathscr{R}|$ for every $i \in \mathscr{R}$. SA is said to obtain a global minimum at time k if $\Pr\{X(k) \in \mathscr{R}_{\text{opt}}\} = 1$. SA is said to obtain a global minimum asymptotically if

$$\lim_{k \to \infty} \Pr\{X(k) \in \mathscr{R}_{\text{opt}}\} = 1. \tag{2.1}$$

Gelfand and Mitter [3], [4] gave a set of conditions which are sufficient to ensure the convergence of SA to a set $I \subset \mathscr{R}$ (of course we can choose $I = \mathscr{R}_{\text{opt}}$). We reformulate their result as we need a slightly modified version of it to prove the theorem below:

For any pair of configurations, i and j, we define $\mathscr{T}_{ij}^{(d)}$ as the set of all chains of transitions $i = l_0 \to l_1 \to \cdots \to l_d = j$ of length d, for which

$$\left.\begin{array}{l} l_{\alpha+1} \in \mathscr{R}_{l_\alpha} \\ \text{or} \\ l_{\alpha+1} = l_\alpha \text{ and there exists some} \\ l' \in \mathscr{R}_{l_\alpha} \text{ with } C(l') > C(l_\alpha) \end{array}\right\} \quad \text{for} \quad \alpha = 0, 1, \ldots, d-1.$$

For $i, j \in \mathscr{R}$, $d \in \mathbb{N}$, and $\tau \in \mathscr{T}_{ij}^{(d)}$ let

$$\Gamma(\tau) = \sum_{\alpha=0}^{d-1} \max(0, C(l_{\alpha+1}) - C(l_\alpha)) \tag{2.2}$$

and

$$\Gamma_{ij}^{(d)} = \begin{cases} \min_{\tau \in \mathscr{T}_{ij}^{(d)}} \Gamma(\tau) & \text{if } \mathscr{T}_{ij}^{(d)} \neq \varnothing, \\ \infty & \text{otherwise,} \end{cases} \tag{2.3}$$

Table 2.1

	k				
$k-1$	u	v	x	y	z
u	$1 - \exp(-3/c_{k-1})$	$\exp(-3/c_{k-1})$	0	0	0
v	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0
x	0	$\frac{1}{2}\exp(-1/c_{k-1})$	$1 - \cdots$	$\frac{1}{2}\exp(-2/c_{k-1})$	0
y	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$
z	0	0	0	$\exp(-3/c_{k-1})$	$1 - \exp(-3/c_{k-1})$

and, finally,

$$\Gamma_{ij} = \min_{d \in \mathbb{N}} \Gamma_{ij}^{(d)}. \tag{2.4}$$

For a fixed subset $I \subset \mathscr{R}$ let J be the complement, i.e., $J = \mathscr{R} - I$. If an index in (2.2), (2.3), or (2.4) is replaced by a set, then an additional minimization is to be performed over the elements of the set, for instance, $\Gamma_{iJ}^{(d)} = \min_{j \in J} \Gamma_{ij}^{(d)}$.

Result (Gelfand and Mitter). The following conditions are sufficient to ensure $\lim_{k \to \infty} \Pr\{X(k) \in I\} = 1$:

1. There exists some $d \in \mathbb{N}$ such that, for all $j \in J$, $\Gamma_{jI}^{(d)} = \Gamma_{jI}$. (2.5)
2. $\max_{j \in J} \Gamma_{jI} < \infty$. (2.6)
3. $\Gamma_{jI} < \Gamma_{Ij}$ for every $j \in J$. (2.7)
4. $\lim_{k \to \infty} c_k = 0$ and $c_k \geq \Gamma^*/(\log k)$ for all k large enough, with $\Gamma^* = \max_{j \in J} \Gamma_{jI}$.

A set $I \subset \mathscr{R}$ is called *nice with respect to C* if conditions (2.5)–(2.7) hold.

In TA the transition matrices are somewhat simpler, for instance, in the case with equal *a priori* transition probability $1/|\mathscr{R}_i|$ for all neighbors of i.

$$P_{ij}(k-1, k) = \begin{cases} 0 & \text{if } j \notin \{\mathscr{R}_i \cup \{i\}\}, \\ 1/n & \text{if } j \in \mathscr{R}_i,\ j \neq i,\ |\mathscr{R}_i| = n, \\ & \quad \text{and } C(j) - C(i) \leq T_{k-1}, \\ 0 & \text{if } j \in \mathscr{R}_i \text{ and } C(j) - C(i) > T_{k-1}, \\ 1 - \sum_{l \neq i} P_{il}(k-1, k) & \text{if } j = i. \end{cases}$$

Table 2.2

$Q_0 =$

$k-1$ \ k	u	v	x	y	z
u	1	0	0	0	0
v	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0
x	0	0	1	0	0
y	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$
z	0	0	0	0	1

$Q_1 =$

$k-1$ \ k	u	v	x	y	z
u	1	0	0	0	0
v	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0
x	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0
y	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$
z	0	0	0	0	1

$Q_2 =$

$k-1$ \ k	u	v	x	y	z
u	1	0	0	0	0
v	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0
x	0	$\frac{1}{2}$	0	$\frac{1}{2}$	0
y	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$
z	0	0	0	0	1

$Q_3 =$

$k-1$ \ k	u	v	x	y	z
u	0	1	0	0	0
v	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0
x	0	$\frac{1}{2}$	0	$\frac{1}{2}$	0
y	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$
z	0	0	0	1	0

Table 2.2 shows the relevant TA-transition matrices Q_j for the example in Figure 2.1. Any threshold value other than 0, 1, 2, 3 also yields one of these matrices. This example shows that a convergence condition like (2.1) cannot hold for TA.

Fact 2.1. Every sequence $(T_k)_{k=0}^{\infty}$ of thresholds yielding

$$\limsup_{k\to\infty} \Pr\{X(k) = u\} = 1$$

also yields

$$\limsup_{k\to\infty} \Pr\{X(k) = v) = 1.$$

This stems from the fact that any such sequence must have $T_k = 3$ for infinitely many $k \in \mathbb{N}$. Moreover, the example shows that the restriction to monotone sequences $(T_k)_{k=0}^{\infty}$ does not suffice.

Fact 2.2. Let $(T_k)_{k=0}^{\infty}$ be monotonically decreasing. Then

$$\lim_{k\to\infty} \Pr\{X(k) \in \{y, z\}\} \geq \tfrac{1}{4}.$$

However, we were able to prove some "convergence" results for threshold sequences of finite lengths. For simplicity we have restricted ourselves to the case where $(\mathscr{R}, (\mathscr{R}_i)_{i\in\mathscr{R}})$ is a finite undirected connected graph $G(V, E)$ with $V = \mathscr{R}$ and $\mathscr{R}_i = \{j | \{i, j\} \in E\}$. The example in Figure 2.1 is of this type. In the following the graph $G(V, E)$ is always assumed to be finite, undirected, and connected, and C: $V \to \mathbb{R}$ is a cost function. We use the short notation (G, C) for such a pair.

Proposition 2.3. *Let (G, C) be given. If for some $\varepsilon > 0$ there exists a sequence $(c_k)_{k=0}^{K-1}$ of temperatures such that $\Pr\{X(K) \in \mathscr{R}_{\text{opt}}\} = 1 - \varepsilon$ in the SA algorithm, then there also exists a sequence $(T_k)_{k=0}^{K-1}$ of thresholds such that $\Pr\{X(K) \in \mathscr{R}_{\text{opt}}\} \geq 1 - \varepsilon$ in the TA algorithm.*

Proposition 2.4. *Let $\{(G_1, C_1), \ldots, (G_r, C_r)\}$ be a finite set of (G, C) pairs. For every $\varepsilon > 0$ there exists a constant $K = K(\varepsilon, (G_1, C_1), \ldots, (G_r, C_r)) \in \mathbb{N}$ and a threshold sequence $(T_k)_{k=0}^{K-1}$ of length K such that the application of TA to (G_s, C_s) yields*

$$\Pr\{X(K) \in \mathscr{R}_{\text{opt}}(s)\} \geq 1 - \varepsilon \qquad \text{for} \quad s = 1, \ldots, r.$$

Here $\mathscr{R}_{\text{opt}}(s)$ denotes the set of optimal vertices in (G_s, C_s).

Theorem. *For every tuple (ε, a, n), with $\varepsilon > 0$, $a > 0$, and $n \in \mathbb{N}$, there exists a threshold sequence $(T_k)_{k=0}^{K-1}$ of natural numbers such that an application of TA with this sequence to any pair $(G(V, E), C)$, with $|V| \leq n$ and a real-valued function C: $V \to [0, a]$, yields*

$$\Pr\{C(X_K) - C_{\text{opt}} \leq (n-2)(n-1)\} \geq 1 - \varepsilon.$$

The length K of the threshold sequence depends on ε, a, and n.

Our theorem can be formulated in a slightly different way by introducing a multiplicative factor $\delta/((n-2)(n-1)) > 0$.

Corollary. *For every tuple* $(\varepsilon, a, n, \delta)$, *with* $\varepsilon > 0$, $a > 0$, $n \in \mathbb{N}$, $n \geq 3$, *and* $\delta > 0$, *there exists a threshold sequence* $(T_k)_{k=0}^{K-1}$ *in* $\{i \cdot \delta/((n-2)(n-1)) | i \in \mathbb{N}\}$ *of some length* $K = K(\varepsilon, a, n, \delta)$ *such that an application of TA to any pair* $(G(V, E), C)$, *with* $|V| \leq n$ *and a real-valued function* $C: V \to [0, a]$, *yields*

$$\Pr\{C(X(K)) - C_{\text{opt}} \leq \delta\} \geq 1 - \varepsilon.$$

The corollary cannot be improved to get

$$\Pr\{C(X(K)) = C_{\text{opt}}\} \geq 1 - \varepsilon$$

for *all* functions $C: V \to [0, a]$. Namely, let $(T_k)_{k=0}^{K}$ be a threshold sequence of finite length. This sequence is outwitted by the simple graph with $V = \{0, 1, 2\}$, $E = \{\{0, 1\}, \{1, 2\}\}$, and $C(0) = 0$, $C(1) = 2\delta$, $C(2) = \delta$, where

$$0 < 2\delta < \min_{0 \leq k \leq K}\{T_k | T_k > 0\}.$$

3. The Proofs

For a fixed pair (G, C) there exist only $m + 1 \leq |E| + 1$ many different transition matrices $Q_0, Q_1, \ldots, Q_m$ in TA, namely at most one for each threshold $T = |C(i) - C(j)|$, $\{i, j\} \in E$ and Q_0 for the trivial threshold $T = 0$. In contrast, the transition matrices P_{c_k} of SA for (G, C) depend on the real-valued temperature parameters $c_k > 0$, where $P_c \neq P_{c'}$ for $c \neq c'$. Hence there are infinitely many different SA matrices. Yet all these P_c's can be written as convex combinations of the TA transition matrices for the same pair (G, C).

Fact 3.1. For every $c > 0$ there exist unique positive real numbers $\lambda_0(c), \ldots, \lambda_m(c)$, with $\sum_{j=0}^{m} \lambda_j(c) = 1$, such that

$$P_c = \sum_{j=0}^{m} \lambda_j(c) Q_j.$$

Example. For the graph in Figure 2.1 we get

$$\lambda_3(c) = \exp\left(\frac{-3}{c}\right), \qquad \lambda_2(c) = \exp\left(\frac{-2}{c}\right) - \exp\left(\frac{-3}{c}\right),$$

$$\lambda_1(c) = \exp\left(\frac{-1}{c}\right) - \exp\left(\frac{-2}{c}\right), \qquad \lambda_0(c) = 1 - \exp\left(\frac{-1}{c}\right).$$

In the general case let $0 = T_0 < T_1 < \cdots < T_m$ be the thresholds relevant to (G, C). The corresponding weights are

$$\lambda_m(c) = \exp\left(\frac{-T_m}{c}\right)$$

and

$$\lambda_i(c) = \exp\left(\frac{-T_i}{c}\right) - \exp\left(\frac{-T_{i+1}}{c}\right) \qquad \text{for} \quad i = 0, 1, \ldots, m-1.$$

Hence all P_c's for $c > 0$ belong to the convex hull of $\{Q_0, \ldots, Q_m\}$.

Proof of Proposition 2.3. Fix a pair (G, C) and let $(c_k)_{k=0}^{K-1}$ be a sequence for SA such that

$$\Pr\{X(K) \in \mathscr{R}_{\text{opt}}\} = 1 - \varepsilon. \tag{3.1}$$

Let $P_{c_k} = \sum_{j=0}^{m} \lambda_j(c_k) Q_j$ be a representation of P_{c_k} as a convex combination of $\{Q_0, \ldots, Q_m\}$ for every $k \in \{0, \ldots, K-1\}$, $V = \{1, \ldots, n\}$, $(p_1(0), \ldots, p_n(0)) = (1/n, \ldots, 1/n)$ the starting distribution of SA, and $(p_1(K), \ldots, p_n(K))$ the distribution at time K, where $p_i(k) \triangleq \Pr\{X(k) = i\}$. Thus

$$(p_1(K), \ldots, p_n(K)) = (p_1(0), \ldots, p_n(0)) \cdot \left[\prod_{k=0}^{K-1} P_{c_k}\right]$$

with

$$\sum_{i \in \mathscr{R}_{\text{opt}}} p_i(K) = 1 - \varepsilon \qquad \text{by (3.1).} \tag{3.2}$$

As

$$\prod_{k=0}^{K-1} P_{c_k} = \prod_{k=0}^{K-1} \left[\sum_{j=0}^{m} \lambda_j(c_k) Q_j\right] = \sum_{0 \le j_0, \ldots, j_{K-1} \le m} \left[\prod_{k=0}^{K-1} \lambda_{j_k}(c_k)\right] \cdot \left[\prod_{k=0}^{K-1} Q_{j_k}\right],$$

we have

$$\begin{aligned}(p_1(K), \ldots, p_n(K)) &= (p_1(0), \ldots, p_n(0)) \cdot \sum_{0 \le j_0, \ldots, j_{K-1} \le m} \left[\prod_{k=0}^{K-1} \lambda_{j_k}(c_k)\right] \left[\prod_{k=0}^{K-1} Q_{j_k}\right] \\ &= \sum_{0 \le j_0, \ldots, j_{K-1} \le m} \left[\prod_{k=0}^{K-1} \lambda_{j_k}(c_k)\right] \cdot \left[(p_1(0), \ldots, p_n(0)) \left[\prod_{k=0}^{K-1} Q_{j_k}\right]\right].\end{aligned} \tag{3.3}$$

For $(j_0, \ldots, j_{K-1}) \in \{0, \ldots, m\}^K$ we define

$$(q_1(j_0, \ldots, j_{K-1}), \ldots, q_n(j_0, \ldots, j_{K-1})) = (p_1(0), \ldots, p_n(0)) \cdot \left[\prod_{k=0}^{K-1} Q_{j_k}\right].$$

Combining (3.2) and (3.3), the method of first moments (see, for instance, [2] as a reference), there exists at least one sequence $(j_0, \ldots, j_{K-1})$ with

(i) $\prod_{k=0}^{K-1} \lambda_{j_k}(c_k) > 0$ and

(ii) $\sum_{i \in \mathscr{R}_{\text{opt}}} q_i(j_0, \ldots, j_{K-1}) \ge 1 - \varepsilon$.

$(Q_{j_0}, \ldots, Q_{j_{K-1}})$ is the desired sequence of transition matrices for TA. □

Proof of Proposition 2.4. Let $(c_k)_{k=0}^{K-1}$ be a temperature sequence of SA such that

$$\Pr\{X_s(K) \in \mathscr{R}_{\text{opt}}(s)\} \geq 1 - \frac{\varepsilon}{r} \qquad \text{for} \quad s = 1, \ldots, r. \tag{3.4}$$

Such a sequence exists for some K large enough, for instance, by the result of Gelfand and Mitter.

For $s = 1, \ldots, r$, let $G_s = (V_s, E_s)$ with $V_s = \{1, \ldots, n_s\}$, let $P_{c_k}^{(s)}$ be the kth transition matrix of SA for the pair (G_s, C_s), and let $Q_0^{(s)}, \ldots, Q_{\bar{m}}^{(s)}$ with $\bar{m} \leq \sum_{t=1}^{r} |E_r|$ be the finitely many different transition matrices resulting from the thresholds $T_0 = 0$ and $\{T_1, \ldots, T_{\bar{m}}\} = \{|C_t(i) - C_t(j)| \,|\, \{i, j\} \in E_t \text{ for some } 1 \leq t \leq r\}$. Not all of the matrices $Q_j^{(s)}$ for a fixed s need to be different. Namely, $Q_j^{(s)} = Q_{j'}^{(s)}$ if $T_j < T_{j'}$ and if there is no edge $\{i, i'\}$ in G_s with $T_j < |C_s(i) - C_s(i')| \leq T_{j'}$. Then there exist unique weights $\lambda_j(c_k)$, for $j = 0, \ldots, \bar{m}$ and $k = 0, \ldots, K-1$, such that

$$P_{c_k}^{(s)} = \sum_{j=0}^{\bar{m}} \lambda_j(c_k) Q_j^{(s)} \qquad \text{for all} \quad s = 1, \ldots, r.$$

Note that the coefficients $\lambda_j(c_k)$ do not depend on s. Thus

$$\prod_{k=0}^{K-1} P_{c_k}^{(s)} = \sum_{0 \leq j_0, \ldots, j_{K-1} \leq \bar{m}} \left[\prod_{k=0}^{K-1} \lambda_{j_k}(c_k)\right] \cdot \left[\prod_{k=0}^{K-1} Q_{j_k}^{(s)}\right] \qquad \text{for} \quad s = 1, \ldots, r.$$

Analogously to the proof of Proposition 2.3 we define, for $(j_0, \ldots, j_{K-1}) \in \{0, \ldots, \bar{m}\}^K$,

$$(q_1^{(s)}(j_0, \ldots, j_{K-1}), \ldots, q_{n_s}^{(s)}(j_0, \ldots, j_{K-1})) = (p_1(0), \ldots, p_{n_s}(0)) \cdot \left[\prod_{k=0}^{K-1} Q_{j_k}^{(s)}\right].$$

Set $q_{\text{opt}}^{(s)}(j_0, \ldots, j_{K-1}) \triangleq \sum_{i \in R_{\text{opt}}(s)} q_i^{(s)}(j_0, \ldots, j_{K-1})$ and denote by M_s the subset of all $(j_0, \ldots, j_{K-1}) \in \{0, \ldots, \bar{m}\}^K$ which satisfy $q_{\text{opt}}^{(s)}(j_0, \ldots, j_{K-1}) < 1 - \varepsilon$.

Claim.

$$\mu_s \triangleq \sum_{(j_0, \ldots, j_{K-1}) \in M_s} \prod_{k=0}^{K-1} \lambda_{j_k}(c_k) < \frac{1}{r} \qquad \text{for} \quad s = 1, \ldots, r. \tag{3.5}$$

Inequality (3.5) holds since, because of (3.4),

$$\mu_s \cdot (1 - \varepsilon) + (1 - \mu_s) \cdot 1 > \Pr\{X_s(K) \in R_{\text{opt}}(s)\} \geq 1 - \frac{\varepsilon}{r}$$

which is equivalent to $\mu_s < 1/r$.

Let $\mu = \sum_{(j_0, \ldots, j_{K-1}) \in M} \prod_{k=0}^{K-1} \lambda_{j_k}$, where $M \triangleq \bigcup_{s=1}^{r} M_s$. By (3.5)

$$\mu \leq \sum_{s=1}^{r} \mu_s < r \cdot \frac{1}{r} = 1.$$

As in the proof of Proposition 2.3, there exists at least one sequence $(j_0, \ldots, j_{K-1})$ with $\prod_{k=0}^{K-1} \lambda_{j_k} > 0$ and

$$q_{\text{opt}}^{(s)}(j_0, \ldots, j_{K-1}) \geq 1 - \varepsilon \qquad \text{for all} \quad s = 1, \ldots, r.$$

The corresponding thresholds $T_{j_0}, \ldots, T_{j_K - 1}$ form the desired sequence for TA. Observe that its length K may be much greater than the length L of the shortest SA sequence $(c_k)_{k=0}^{L-1}$, yielding

$$\Pr\{X(L) \in \mathscr{R}_{\text{opt}}(s)\} \geq 1 - \varepsilon \qquad \text{for} \quad s = 1, \ldots, r. \qquad \square$$

Our theorem cannot be proved by the same argument as Proposition 2.4 because there are *infinitely* many possible choices for the real-valued function C: $V \to [0, a]$ for each graph $G(V, E)$.

Proof of the Theorem. Consider a fixed graph $G(V, E)$ and the set $\mathscr{C}_a$ of all functions C: $V \to [0, a]$. The key idea is to find a partition of $\mathscr{C}_a$ in finitely many equivalence classes, which allows us to apply the proof of Proposition 2.4. Before doing this we present a modified version of Theorem 4.2 of [4]: the cost function C: $\mathscr{R} \to \mathbb{R}$ is substituted by an *arc cost function* f, which assigns real values to all neighboring (directed) pairs (i, j) of configurations.

In this arc cost model SA works as follows: Let i be the current configuration at time k and let j be the selected neighbor of i. Then j becomes the next configuration in the sequence

$$\begin{cases} \text{with probability 1} & \text{if} \quad f(i,j) < 0, \\ \text{with probability } \exp\left(\dfrac{-f(i,j)}{c_k}\right) & \text{if} \quad f(i,j) \geq 0. \end{cases}$$

Now we adopt the notion of the reformulated result of Gelfand and Mitter in Section 2, except the definition of $\Gamma(\tau)$, which becomes

$$\Gamma(\tau) = \sum_{\alpha=0}^{d-1} \max\{0, f(l_\alpha, l_{\alpha+1})\},$$

and see that the result, as formulated in Section 2, remains valid in the arc cost model.

For $G(V, E)$ and C: $V \to \mathbb{R}$ we define integral arc costs f by

$$F(i,j) = \begin{cases} 0 & \text{if} \quad i = j, \\ 0 & \text{if} \quad \{i,j\} \in E \quad \text{and} \quad C(i) \geq C(j), \\ \lceil C(j) - C(i) \rceil & \text{if} \quad \{i,j\} \in E \quad \text{and} \quad C(i) < C(j), \\ \text{undefined} & \text{otherwise.} \end{cases} \tag{3.6}$$

Here $\lceil c \rceil$ denotes the smallest integer not smaller than $c \in \mathbb{R}$.

Lemma 3.2. *Let $G(V, E)$ with $|V| = n$ be given, let C: $V \to \mathbb{R}$, and let f be the corresponding arc cost function. Then there exists a nice subset $I \subset V$ with respect to f such that*

$$C(i) - C_{\text{opt}} < (n-2)(n-1)$$

for all $i \in I$.

Proof of Lemma 3.2. Assume $V = \{v_1, \ldots, v_n\}$ and $C_{\text{opt}} = C(v_1) \leq C(v_2) \leq \cdots \leq C(v_n)$.

Case 1. $C(v_{i+1}) - C(v_i) < n-2$ for $i = 1, \ldots, n-1$. Then $I = V$ is an appropriate choice.

Case 2. Let $i \in \{1, \ldots, n\}$ be minimal with $C(v_{i+1}) - C(v_i) \geq n-2$.

Claim. *Then* $\{v_1, \ldots, v_i\}$ *is nice with respect to* f.

Proof of the claim. Conditions (2.5) and (2.6) obviously hold, so only (2.7) remains to be proved. Let $j \in J$ and let $I \ni i = u_0 \to \cdots \to u_r = j$ be a cheapest path from I to j of minimal length. Thus $u_1, \ldots, u_{r-1} \notin I$ and $r - 1 \leq n - 2$. We define

$$\Delta_{ij} = \sum_{\substack{0 \leq \alpha \leq r-1: \\ C(u_{\alpha+1}) > C(u_\alpha)}} (C(u_{\alpha+1}) - C(u_\alpha))$$

and

$$\Delta_{ji} = \sum_{\substack{0 \leq \alpha \leq r-1: \\ C(u_{\alpha+1}) < C(u_\alpha)}} (C(u_\alpha) - C(u_{\alpha+1})).$$

Reversing the path we see that

$$\Gamma_{ji} \leq \sum_{\substack{0 \leq \alpha \leq r-1: \\ C(u_{\alpha+1}) < C(u_\alpha)}} \lceil C(u_\alpha) - C(u_{\alpha+1}) \rceil$$

$$< (r-1) + \Delta_{ji}.$$

The term $(r-1)$ instead of r stems from the fact that $C(j) > C(i)$. Thus there must be at least one α with $C(u_\alpha) < C(u_{\alpha+1})$.

On the other hand,

$$\Gamma_{Ij} = \Gamma_{ij} \geq \Delta_{ij},$$

$$\Delta_{ij} - \Delta_{ji} = C(j) - C(i),$$

thus by the case condition

$$\Delta_{ij} - \Delta_{ji} \geq n - 2.$$

Hence

$$\Gamma_{Ij} - \Gamma_{jI} \geq \Gamma_{ij} - \Gamma_{ji} > \Delta_{ij} - \Delta_{ji} - (r-1) \geq n - 2 - (r-1) \geq 0,$$

which completes the proof of the claim and also the proof of Lemma 3.2. □

Conjecture. *Lemma* 3.2 *can be improved to have the condition* $C(i) - C_{\text{opt}} < n - 2$ *instead of* $C(i) - C_{\text{opt}} < (n-2)(n-1)$.

That we cannot hope for more than this may be seen from the example in Figure 3.1 where $\varepsilon > 0$ is arbitrarily small. The names of the (some) nodes are

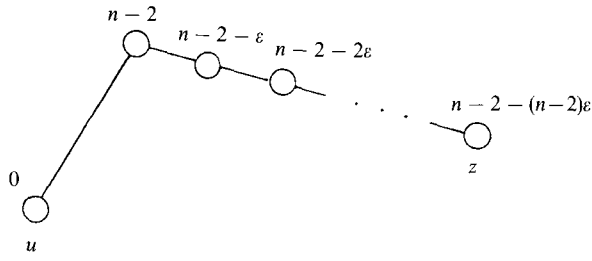

Figure 3.1

given below the node and the C-values are given above the nodes. The minimal nice set is $I = \{u, z\}$ and

$$C(z) - C(u) = n - 2 - (n-2)\varepsilon \approx n - 2.$$

Two functions $C_1, C_2: V \to \mathbb{R}$ are called equivalent, if the corresponding arc cost functions from (3.6) coincide.

Fact 3.3. Let $G(V, E)$ be given. For every real number $a > 0$ there exist only finitely many nonequivalent functions $C: V \to [0, a]$.

Fact 3.4. Let $C_1, C_2: V \to \mathbb{R}$ be two equivalent functions on $G(V, E)$ and let $(T_k)_{k=0}^K$ be an *integral* sequence of thresholds. Then an application of TA with $(T_k)_{k=0}^K$ to (G, C_1) yields the same distribution on V as an application to (G, C_2).

Now we combine Fact 3.3, Fact 3.4, Lemma 3.2, and the arc cost version of the result of Gelfand and Mitter. As there are only finitely many different graphs $G(V, E)$ with $|V| \leq n$, a proof like that of Proposition 2.4 yields the result of our theorem. □

4. Concluding Remarks

1. Our proofs are only results of existence and do not tell us anything about how to construct optimal (or even good) threshold sequences. Therefore the results will not have a direct impact on TA practice. But this missing relevance for practice is (partly by other reasons) also true for most of the theoretical results on SA.

2. In applications we are often not interested in getting convergence to the set $\mathscr{R}_{\text{opt}}$. Instead it is sufficient to have visited a member of $\mathscr{R}_{\text{opt}}$ or even only a near-to-optimum configuration in $\mathscr{R}$ once along the tour of transitions. In these models analogs of our results (for instance, in Proposition 2.3 replace the condition "$X(K) \in R_{\text{opt}}$" by "$X(k) \in R_{\text{opt}}$ for some $k \in \{0, \ldots, K\}$") can also be proved by the convex hull argument. We omit the straightforward elaborations. See also the recent paper by Hajek and Sasaki [6] whose Proposition 1a is closely related to our Proposition 2.3.

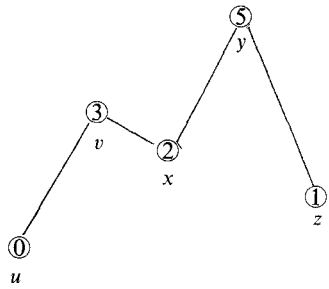

Figure 4.1

3. In some sense TA is a half-deterministic skeleton of SA. The only remaining probabilistic part is the random choice of a neighbor j of the current configuration i. Dueck and Scheuer have also successfully applied variants of TA, which make even these choices deterministically [1]. A theoretical modeling of this can be done as follows: For each configuration i the neighbors are arranged in a linear list $j_1, \ldots, j_{\deg(i)}$. For every time k not only a threshold T_k but also a parameter γ_k,

Table 4.1

K	Optimal threshold sequence of length K	$\Pr\{X(K) = u\}$
0		0.2000
1	0	0.3000
2	10	0.3500
3	110	0.4250
4	1110	0.4750
5	11110	0.5188
6	111110	0.5531
7	1111110	0.5831
8	11111110	0.6039
9	311111110	0.6273
10	3111111110	0.6508
11	01311111110	0.6723
12	013111111110	0.6967
13	0131111111110	0.7164
14	01311111111110	0.7324
15	301311111111110	0.7455
16	3013111111111110	0.7583
17	13013111111111110	0.7708
18	013013111111111110	0.7833
19	0130131111111111110	0.7937
20	01301311111111111110	0.8020
21	013110131111111111110	0.8103
22	0131101311111111111110	0.8187
23	01311101311111111111110	0.8263
24	013111101311111111111110	0.8335

$0 < \gamma_k \leq 1$, is given. γ_k determines which neighbor of i has to be selected, namely, for instance, $j_{\lceil \gamma_k \cdot \deg(i) \rceil}$ in the case where each neighbor has the same probability of becoming the candidate for the next transition. It is a simple exercise to prove the analogs of our results for this completely deterministic version of TA. The key idea is that "usual" TA belongs to the convex hull of "deterministic" TA.

4. Our results are not restricted to undirected graphs. By some more notional and technical efforts it is possible to yield, for example, a TA analogon to the sufficiency result of Gelfand and Mitter for SA.

5. In contrast to SA it is by TA possible to get convergence in a suboptimal configuration, even in case of undirected neighborhoods. Figure 4.1 shows an example which is formally similar to that of Figure 2.1. The unique global minimum is in u, yet it is possible to achieve $\lim_{k \to \infty} \Pr\{X(k) = z\} = 1$ by the constant threshold sequence 3, 3, 3, ...

Open Problem: For each instance $(\mathscr{R}, (\mathscr{R}_i)_{i \in \mathscr{R}}, C)$ characterize all those subsets $I \subset \mathscr{R}$, to which a TA "convergence" (lim or limsup) can be achieved by an appropriate threshold sequence.

6. By computer enumeration we have found the optimal threshold sequences $(T_k)_{k=0}^{K-1}$ for the graph in Figure 2.1 for all $K \leq 24$, where the optimization criterion was to maximize $\Pr\{X(K) = u\}$. The starting distribution was $(\frac{1}{5}, \ldots, \frac{1}{5})$. For all $K \geq 2$, sequences with $T_{K-2} \in \{1, 2\}$ and $T_{K-1} \in \{0, 1, 2\}$ are equioptimal. The sequences shown in Table 4.1 are the lexicographic smallest among all equioptimal sequences.

References

1. G. Dueck and T. Scheuer, Threshold accepting: A general purpose optimization algorithm appearing superior to simulated annealing, Journal of Computational Physics 90 (1990), 161–175.
2. P. Erdös and J. Spencer, Probabilistic Methods in Combinatorics, Academic Press, New York, 1974.
3. S. B. Gelfand and S. K. Mitter, Analysis of simulated annealing for optimization, Technical Report LIDS-P-1495, August 1985.
4. S. B. Gelfand and S. K. Mitter, Analysis of simulated annealing for optimization, Proc. 24th Conf. on Decision and Control, Ft Lauderdale, December 1985, pp. 779–786.
5. M. Grötschel, Polyedrische Kombinatorik und Schnittebenenverfahren, Preprint No. 38, Universität Augsburg, 1984.
6. B. Hajek and G. Sasaki, Simulated annealing—to cool or not, Systems & Control Letters 12 (1989), 443–447.
7. S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, Optimization by simulated annealing, IBM Research Report RC 9355, 1982.
8. S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, Optimization by simulated annealing, Science 220 (1983), 671–680.
9. P. J. M. van Laarhoven and E. H. L. Aarts, Simulated Annealing: Theory and Applications, Reidel, Dordrecht, 1987.

Accepted 20 September 1990