Convergence of Simulated Annealing with Feedback Temperature Schedules

Corrected version,

original paper appeared in Prob. in the Engineering and Informational Sciences, 11,279-304, 1997

(corrections have been made to Theorem 1 (iv) and the proof of Theorem 2)

M. Kolonko and M.T. Tran

Institut für Mathematik, Universität Hildesheim

Marienburger Platz 22, D-31141 Hildesheim

Germany

Tel.: ++49 5121 883713

e-mail kolonko@informatik.uni-hildesheim.de

10.12.96

Abstract: It is well-known that the standard simulated annealing optimization method converges in distribution to the minimum of the cost function if the probability α for accepting an increase in costs goes to zero. α is controlled by the 'temperature' parameter, which in the standard setup is a fixed sequence of values converging slowly to 0.

We study a more general model in which the temperature may depend on the state of the search process. This allows to adapt the temperature to the landscape of the cost function. The temperature may temporarily rise such that the process can leave a local optimum more easily. We give weak conditions on the temperature schedules such that the process of solutions finally concentrates near the optimal solutions.

We also briefly sketch computational results for the job shop scheduling problem.

KEYWORDS: SIMULATED ANNEALING, ADAPTIVE TEMPERATURE SCHEDULE, INHOMOGENEOUS MARKOV CHAINS, JOB SHOP SCHEDULING

1 Introduction

Simulated Annealing is a stochastic local search technique to approximate the minimum of a cost function $c: S \to \mathbb{R}$ over a finite set S. It is an iterative method that randomly chooses elements y from a neightbourhood $N(x)$ of the present solution. The candidate y is either accepted as the new solution or rejected. It may be accepted with a positive probability even if $c(y) > c(x)$, i.e. even if it is worse than the present solution. The search process can thus 'climb uphill' to get out of local minima. Simulated annealing has proven quite successful in many applications, see [9] for an overview. Its theoretical analysis is concentrated on modelling it as an inhomogeneous Markov chain and to find conditions for its convergence to optimal solutions, see e.g. [2], [7], [12].

The long run behaviour of the search process depends critically on $\alpha(x, t, y)$, the probability to accept a candidate y with a present solution x. $\alpha(x, t, y)$ is controlled by the parameter t which is called 'temperature' from an analogy to physical cooling processes. To make the process of solutions an inhomogeneous Markov chain, the temperature values are chosen independently of the process as a fixed sequence $(t_n)_{n\in\mathbb{N}}$, the 'temperature schedule'. Usually the so-called Metropolis acceptance probability is used, that is for $t > 0$

$$\alpha(x, t, y) := \begin{cases} 1 & \text{if } c(y) \le c(x) \\ \exp\left(-\frac{c(y)-c(x)}{t}\right) & \text{if } c(y) > c(x) \end{cases} . \tag{1}$$

The candidate solutions are chosen according to a generating probability $G(x, \cdot)$ which is often the uniform distribution on the neighbourhood $N(x)$. If in addition G fulfills certain symmetry conditions, e.g. if $G(x, y) = G(y, x)$ in the simplest case, then the process converges in distribution (is strongly ergodic) to the uniform distribution on the optimal solutions if $(t_n)_{n\in\mathbb{N}}$ fulfills

$$\lim_{n\to\infty} t_n = 0 \quad \text{and} \quad t_n \ \ge \ \frac{C}{\log n} \text{ for all } n \in \mathbb{N} \tag{2}$$

where C is a constant depending on the variation of the cost function (see e.g. [2], [7], [13]). In most cases, there will only be rough estimates for C.

Temperature schedules with (2) lead to processes which converge very slowly, in particular if C is overestimated. Therefore, in many practical applications temperature schedules are used that go to zero faster but for which convergence to optimal solutions cannot be shown or that do not even lead to Markov chains (see e.g. [6], [17]). For a certain class of randomized schedules as used in [6], a 'piecewise Markovian' model was given in [8].

The present paper introduces a random temperature schedule that may depend on the most recent history of the process and that allows for a temporary rise in temperature. This increases the ability of the process to leave local minima. Though the solution process itself is not Markovian, we can give simple conditions for the convergence of the process to 'good' solutions which are arbitrarily close to the optima. Differing from the usual approach we assume that the temperature process converges to a small but positive minimal temperature $\vartheta_0 > 0$. This will be in accordance to most real implementations of the process and allows to use simple Markov chain theory in the analysis. This part of the present paper generalizes results from [18].

Asymptotic properties of a search procedure over a finite set are of limited practical value, therefore we briefly sketch some computational results of test runs. We applied our method to the well-known job shop scheduling problem in which an optimal order for the processing of jobs on machines has to be found. Over a library of test problems our method proved very successful.

The paper is organized as follows. In Section 2 we define a general stochastic process that covers simulated annealing with random temperature variables. In Section 3 we show that for any stochastically convergent sequence of temperatures the distribution of the solution process converges to the steady state probability corresponding to the minimal temperature ϑ_0. Thus, in the long run, the solutions will concentrate around nearly global optima if ϑ_0 is small enough. In Section 4 we introduce recursive temperature schemes that are both quadratically and almost surely convergent and which have appealing heuristical motivations. In Section 5 we give some empirical results for the job shop scheduling problem. A short summary is given in Section 6. For better readability, the formal proofs and some auxiliary results are collected in an appendix.

2 A Generalized Simulated Annealing Scheme

In this Section we define the stochastic model for a generalized simulated annealing procedure in which the temperature parameter may depend on the most recent history of the process. Note that we restrict ourselves to finite state spaces and do not discuss measurability questions in this paper.

Definition

(i) Let the following be stochastic processes defined on a joint probability space :

- $(X_n)_{n\in\mathbb{N}}$ takes on values in the finite *state space* S. X_n denotes the *state* (solution) at time n.
- $(Y_n)_{n\in\mathbb{N}}$ also takes on values in S. Y_n denotes the *candidate* at time n.
- $(Z_n)_{n\in\mathbb{N}}$ takes on values in $\{0,1\}$. Z_n denotes the *decision* at time n, where 0 stands for rejection and 1 for acceptance of the candidate.
- $(T_n)_{n\in\mathbb{N}}$ takes on values in $[0,\infty)$. T_n denotes the *temperature* at time n.

(ii) We say that $(Y_n, Z_n, T_n, X_n)_{n\in\mathbb{N}}$ forms a *simulated annealing process* if:

- $X_0 = x_0$ and $T_0 = t_0$ for given starting values $x_0 \in S$ and $t_0 > 0$,
- there are suitably defined functions G, α and τ_n $(n \in \mathbb{N})$ such that for $n = 1$

$$\begin{aligned} \mathbf{P}[Y_1 = y \mid X_0, T_0] &:= G(X_0, T_0, y) \qquad \text{for } y \in S \\ \mathbf{P}[Z_1 = 1 \mid X_0, T_0, Y_1] &:= \alpha(X_0, T_0, Y_1) \\ X_1 &:= \begin{cases} Y_1 & \text{if } Z_1 = 1 \\ X_0 & \text{if } Z_1 = 0 \end{cases}, \\ T_1 &:= \tau_1(X_0, T_0, Y_1, Z_1), \end{aligned}$$

and such that for all $n \in \mathbb{N}$ with the history

$$\mathcal{H}_n := (X_0, T_0, Y_1, Z_1, X_1, T_1, \ldots, Y_n, Z_n, X_n, T_n)$$

we have

$$\mathbf{P}[Y_{n+1} = y \mid Y_n, Z_n, X_n, T_n, \mathcal{H}_{n-1}] := G(X_n, T_n, y) \qquad \text{for } y \in S,$$

$$
\begin{aligned}
\mathbf{P}\left[Z_{n+1}=1 \mid Y_n, Z_n, X_n, T_n, Y_{n+1}, \mathcal{H}_{n-1}\right] &:= \alpha(X_n, T_n, Y_{n+1}), \\
X_{n+1} &:= \begin{cases} Y_{n+1} & \text{if } Z_{n+1}=1 \\ X_n & \text{if } Z_{n+1}=0 \end{cases}, \\
T_{n+1} &:= \tau_{n+1}(X_n, T_n, Y_{n+1}, Z_{n+1}).
\end{aligned}
$$

This formal definition has the following interpretation: a simulated annealing process starts in state x_0 with temperature t_0. If at time n state $X_n = x \in S$ is reached and the temperature is $T_n = t$, then a new candidate $Y_{n+1} = y \in S$ is chosen according to the generating probability $G(x,t,y)$. $G(x,t,\cdot)$ is probability measure on S. The candidate y is accepted (i.e. $Z_{n+1} = 1$) with probability $\alpha(x,t,y)$ and becomes the new state X_{n+1} at time $n+1$. With probability $1-\alpha(x,t,y)$ the candidate y is rejected (i.e. $Z_{n+1} = 0$) and x will be the state at time $n+1$.

Finally the new temperature is chosen as $T_{n+1} := \tau_{n+1}(x,t,y,z)$. In its standard version simulated annealing uses $\tau_{n+1} \equiv t_n$, where t_n is a predefined constant, i.e. there is no feedback from the process to the temperature. We shall give a more sophisticated scheme for τ_n in Section 4 below.

It is easy to show from the above definition that the whole process $(Y_n, Z_n, X_n, T_n)_{n\in\mathbb{N}}$ as well as its component $(X_n, T_n)_{n\in\mathbb{N}}$ are inhomogeneous Markov processes. Note that $(X_n)_{n\in\mathbb{N}}$ itself need not be Markovian as the 'future' X_{n+1} conditioned on the 'present' X_n may depend on the 'past' $X_{n-1}, X_{n-2}, \ldots$ via the temperature T_n.

3 Convergence of the Solution Process

We now examine the behaviour of the solution process $(X_n)_{n\in\mathbb{N}}$ under the assumption that the temperature T_n converges stochastically to a minimal value $\vartheta_0 > 0$. It turns out that the distribution of X_n converges to the steady state distribution that X_n would approach if the temperature was fixed to ϑ_0 from the start. Here it is assumed that for fixed temperature ϑ_0 $(X_n)_{n\in\mathbb{N}}$ forms an ergodic homogeneous Markov chain. The theorem is formulated without reference to simulated annealing; it applies to an arbitrary Markov chain $(X_n, T_n)_{n\in\mathbb{N}}$ with state space $S \times \mathbb{R}$, where S is finite . In an example below it is shown that the conditions of the theorem are easily fulfilled in the simulated annealing process with Metropolis acceptance probabilities. Convergent temperature schedules will be presented in the next Section.

Theorem 1

Let $(X_n, T_n)_{n\in\mathbb{N}_0}$ be a Markov chain with state space $S\times\mathbb{R}$ and with starting state $x_0 \in S,\ t_0 \in \mathbb{R}$. Assume that the following conditions hold:

(i) $q_t(x,y) := \mathbf{P}[X_{n+1} = y \mid X_n = x, T_n = t]$ is independent of $n \in \mathbb{N}_0$ for all $x, y \in S,\ \ t \in \mathbb{R}$,

(ii) there is $\vartheta_0 \in \mathbb{R}$ such that $T_n \to \vartheta_0$ stochastically, i.e.

$$\lim_{n\to\infty} \mathbf{P}\left[\ |T_n - \vartheta_0| > \varepsilon \mid X_0 = x_0, T_0 = t_0\right] = 0$$

for all $\varepsilon > 0$,

(iii) $t \mapsto q_t(x,y)$ is continuous in ϑ_0 for all $x, y \in S$, and

(iv) $q_{\vartheta_0}(\cdot,\cdot)$ is an aperiodic and irreducible finite Markov matrix with a limiting distribution $\pi_{\vartheta_0}(y) := \lim_{n\to\infty} q_{\vartheta_0}(x,y)$ for $x, y \in S$.

Then $(X_n)_{n\in\mathbb{N}_0}$ behaves asymptotically as the homogeneous Markov chain with matrix q_{ϑ_0}, more precisely

$$\lim_{n\to\infty} \mathbf{P}\left[X_n = x \mid X_0 = x_0, T_0 = t_0\right] \;=\; \pi_{\vartheta_0}(x) \tag{3}$$

for all $x \in S$.

The proof is given in the appendix below.

For the simulated annealing process $(Y_n, Z_n, X_n, T_n)_{n\in\mathbb{N}}$ as defined in Section 2 we have

$$\begin{aligned} q_t(x,x') &= \mathbf{P}\left[X_{n+1} = x' \mid X_n = x, T_n = t\right] \\ &= \begin{cases} G(x,t,x')\alpha(x,t,x') & \text{if } x \neq x' \\ 1 - \sum_{y\in S-\{x'\}} G(x,t,y)\alpha(x,t,y) & \text{if } x = x' \end{cases} \end{aligned}$$

which is independent of n, i.e. Condition (i) holds. Condition (iii) of Theorem 1 will be fulfilled if

$$t \mapsto G(x,t,x') \quad \text{and} \quad t \mapsto \alpha(x,t,x')$$

are continuous. In most applications, $G(x,t,x') = G(x,x')$ does not depend on the temperature t and for the Metropolis acceptance probability (1) we have indeed that $t \mapsto \alpha(x,t,y)$ is continuous.

Condition (iv) is simplified if $\min_{x,x'\in S} \alpha(x,\vartheta_0,x') > 0$, as then properties as aperiodicity or ergodicity hold for q_{ϑ_0} iff they hold for G. For the Metropolis acceptance probability (1) we have for $\vartheta_0 > 0$

$$\alpha(x,t,y) \;\geq\; \underline{\alpha} := \exp(-\overline{c}/\vartheta_0) \;>\; 0 \tag{4}$$

where $\overline{c} := \max\{c(y) - c(x) \mid x, y \in S\}$ is the largest possible change in the cost function. Note that this relation will not hold for $\vartheta_0 = 0$.

Now let us assume that we have the Metropolis acceptance scheme and a generating probability $G(x,y)$ which does not depend on t. Then for $t \geq \vartheta_0$ we have $\alpha(x,t,y) \geq \underline{\alpha}$ and the recurrent classes of q_t are the classes of G and hence do not depend on $t \geq \vartheta_0$. As a consequence (see [16]), the limiting distribution π_t depends continously on $t \geq \vartheta_0$. If, in addition, $G(x,y)$ is symmetrical in x, y, i.e. if $G(x,y) = G(y,x)$ then

$$\pi_{\vartheta_0}(x) = \frac{\exp\left(-\frac{c(x)}{\vartheta_0}\right)}{\sum_{y\in S} \exp\left(-\frac{c(y)}{\vartheta_0}\right)}$$

for $x \in S$ (see [12]). π_{ϑ_0} gives highest probability to the optimal solutions and approaches the uniform distribution on the optimal states for $\vartheta_0 \to 0$. Here, symmetry of G can be replaced by 'weak reversibility', see [5].

Hence we obtain from Theorem 1 that under the above conditions for any temperature schedule $(T_n)_{n\in\mathbb{N}}$ that is stochastically convergent against ϑ_0 it holds that

$$\lim_{n\to\infty} \mathbf{P}(X_n = x) \;=\; \pi_{\vartheta_0}(x)$$

will concentrate around the optimal states for small values of ϑ_0. In this sense $(X_n)_{n\in\mathbb{N}}$ converges (in distribution) to the optimal states.

Though we cannot prove at present that $(X_n)_{n\in\mathbb{N}}$ converges to π_{ϑ_0} faster than the Markov chain with fixed transition probability q_{ϑ_0} (i.e. with a fixed low temperature), the results in Section 5 show the empirical advantage of the simulated annealing scheme.

4 Feedback Temperature Schedules

We use the following recursive scheme for the control of the temperature T_n :

$$T_0 \;:=\; t_0 \tag{5}$$

$$T_{n+1} \;:=\; \max\{\vartheta_0, T_n - D_{n+1}\} \quad \text{for } n \geq 0, \tag{6}$$

where $t_0 \geq 0$ is a starting temperature, $\vartheta_0 \geq 0$ is the *minimal temperature* and $(D_n)_{n\in\mathbb{N}}$ are real-valued random variables denoting the random change in temperature at time $n \in \mathbb{N}$. D_n may take positive and negative values making $(T_n)_{n\in\mathbb{N}}$ a random walk on $[\vartheta_0, \infty)$.

The following theorem gives conditions on the D_n such that $(T_n)_{n\in\mathbb{N}}$ converges to the minimal temperature ϑ_0 in quadratic mean and almost surely (which implies stochastic convergence as needed in Theorem 1). It is formulated as a general theorem on random variables. The application to simulated annealing where D_n depends explicitly on the history of the process is given below.

Theorem 2

Let $\vartheta_0 > 0$ be a given (minimal) temperature and let $(T_n)_{n\in\mathbb{N}}$ and $(D_n)_{n\in\mathbb{N}}$ be sequences of random variables fulfilling (5) and (6). Let $(\Delta_n)_{n\in\mathbb{N}}$ with $\Delta_n > 0$ be a sequence of real numbers fulfilling

(i) $\sum_{n=1}^{\infty} \Delta_n \;=\; \infty$ and

(ii) $\sum_{n=1}^{\infty} \Delta_n^2 \;<\; \infty$,

and such that for all $n \in \mathbb{N}$

(iii) $|D_n| \;\leq\; \Delta_n$ $(\mathbf{P}\text{-}a.s.)$ and

(iv) $\mathbf{E}\left[D_n \mid T_0, T_1, \ldots, T_{n-1}\right] \;\geq\; K \cdot \Delta_n$ for some constant $K > 0$.

Then

(a) $\lim_{n\to\infty} T_n = \vartheta_0$ $(\mathbf{P}\text{-}a.s.)$ and

(b) $\lim_{n\to\infty} \mathbf{E}\,(T_n - \vartheta_0)^2 \;=\; 0$.

Condition (iii) requires the temperature changes to be bounded by a sequence of constants Δ_n which behave similar to $\frac{1}{n^\gamma}$ with $0.5 < \gamma \leq 1$. By (iv), the temperature decrease D_n is positive in the mean, hence T_n has a tendency towards zero in the mean. Note that T_n itself need not be bounded, i.e. (b) is not a simple consequence of (a). The proof is given in the appendix below. It shows that part b) holds under the slightly weaker condition

$$\text{(iv}'\text{)} \quad \mathbf{E}\left[D_n \mid T_{n-1}\right] \;\geq\; K \cdot \Delta_n \qquad \text{for some constant } K > 0$$

instead of (iv).

The next two theorems give examples for the D_{n+1} in the simulated annealing context such that the definiton of Section 2 is fulfilled for T_n. The temperature is allowed to react to the evolution of the process $(X_n)_{n\in\mathbb{N}}$ in a heuristically (and empirically) reasonable way that ensures convergence of T_n to ϑ_0 in quadratic mean and almost surely.

Theorem 3

Let $\vartheta_0 > 0$ be a given minimal temperature and $0 \leq \delta < 1$ be a constant. Let $(\Delta_n)_{n\in\mathbb{N}}$ be a sequence as in Theorem 2 (i) and (ii). Let $(Y_n, Z_n, X_n, T_n)_{n\in\mathbb{N}}$ be a simulated annealing process as in Section 2 with starting values x_0, t_0 and temperature functions

$$\tau_{n+1}(x,t,y,z) \;:=\; \begin{cases} \max\{\vartheta_0,\ t-\Delta_{n+1}\} & \text{if } c(y) \leq c(x) \\ \max\left\{\vartheta_0,\ t+\delta\Delta_{n+1}\alpha(x,t,y)\right\} & \text{if } c(y) > c(x),\ z=0 \\ \max\left\{\vartheta_0,\ t-\Delta_{n+1}\Big(1-\delta\alpha(x,t,y)\Big)\right\} & \text{if } c(y) > c(x),\ z=1 \end{cases} \tag{7}$$

for all $n \geq 0,\ x,y \in S,\ z \in \{0,1\}$ and $t \geq 0$. Assume that there is $\underline{\alpha} > 0$ such that

$$\inf\{\alpha(x,t,y) \mid x,y \in S, t \geq \vartheta_0\} \;\geq\; \underline{\alpha}. \tag{8}$$

Then $(T_n)_{n\in\mathbb{N}}$ fulfills (5) and (6) with

$$D_{n+1} \;:=\; \begin{cases} \Delta_{n+1} & \text{if } c(Y_{n+1}) \leq c(X_n) \\ \Delta_{n+1}\Big(Z_{n+1} - \delta\alpha(X_n,T_n,Y_{n+1})\Big) & \text{if } c(Y_{n+1}) > c(X_n) \end{cases} \tag{9}$$

and

$$\lim_{n\to\infty} T_n = \vartheta_0 \ (\mathbf{P}-a.s.) \quad \text{and} \quad \lim_{n\to\infty} \mathbf{E}\,(T_n-\vartheta_0)^2 \;=\; 0.$$

Again the proof is given in the appendix below.

The temperature schedule given in (7) resp. (9) has the following heuristic justification : if the new candidate solution Y_{n+1} is at least as good as the present solution X_n (i.e. if $c(Y_{n+1}) \leq c(X_n)$) then this is regarded as a success in the search and the temperature is decreased by the fixed amount Δ_{n+1}. If the candidate solution is worse than the present solution and is rejected then this is regarded as a failure and the temperature is increased by the amount $\Delta_{n+1}\delta\alpha(X_n,T_n,Y_{n+1})$ (which is smaller than Δ_{n+1}) allowing the process to climb more freely. In the remaining case, if a worse candidate solution is accepted (i.e. $Z_{n+1} = 1$) then the temperature is lowered by the (small) amount $\Delta_{n+1}(1 - \delta\alpha(X_n,T_n,Y_{n+1}))$ in order to avoid too much uphill climbing. The acceptance probability α is explicitly included into the temperature control to make the temperature changes depend on the steepness of the 'hill' which is accepted or rejected. The constant $\delta < 1$ ensures a tendency towards the minimum temperature in the mean.

As an example consider the Metropolis acceptance probability as given in (1) where α depends directly on the increase of costs $c(Y_{n+1}) - c(X_n)$. E.g. with $\Delta_n := 1/n$ we obtain (cp. Fig.1)

$$T_{n+1} := \begin{cases} \max\{\vartheta_0,\ T_n - \frac{1}{n+1}\} & \text{if } c(Y_{n+1}) \le c(X_n) \\ \max\left\{\vartheta_0,\ T_n - \frac{1}{n+1}\left(Z_{n+1} - \delta \exp\left(-\frac{c(Y_{n+1})-c(X_n)}{T_n}\right)\right)\right\} & \text{if } c(Y_{n+1}) > c(X_n) \end{cases}.$$

Again we can use $\underline{\alpha} := \exp(-\overline{c}/\vartheta_0)$ as given in (4).

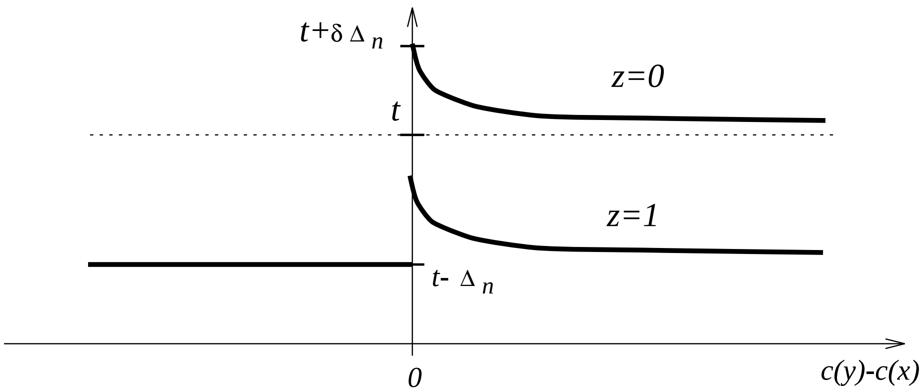

Figure 1: The temperature function $\tau_n(x, t, y, z)$ as given in (9) for a fixed t as funtion of $c(y) - c(x)$.

Another more involved temperature scheme leads to even better practical results. We use a temperature function that is partly linear in $c(y) - c(x)$. Let

$$\tilde{c}(x, y) := \frac{c(y) - c(x)}{c_{97}}$$

be a normed cost difference where c_{97} is a constant which will be a rough estimate of the 97 % -quantile of $c(Y_{n+1}) - c(X_n)$ in the application below. We use another constant c_l which will later be an estimate of the smallest positive cost difference.

Theorem 4

Let $c_l, c_{97} > 0,\ \ t_0 > \vartheta_0 > 0$ be given. Let $(Y_n, Z_n, X_n, T_n)_{n\in\mathbb{N}}$ be a simulated annealing process as in Section 2 with starting values x_0, t_0, the Metropolis acceptance scheme as in (1) and temperature functions

$$\tau_{n+1}(x, t, y, z) := \begin{cases} \max\{\vartheta_0,\ t + \Delta_{n+1}\tilde{c}(x, y)\} & \text{if } c(y) < c(x) \\ \max\left\{\vartheta_0,\ t - \Delta_{n+1}\frac{c_l}{c_{97}}\right\} & \text{if } c(y) = c(x) \\ \max\left\{\vartheta_0,\ t - \Delta_{n+1}\left(\frac{1}{\alpha(x,t,y)} - 1 + \tilde{c}(x, y)\right)\right\} & \text{if } c(y) > c(x), z = 1 \\ t\ +\ \Delta_{n+1}\left(1 - \tilde{c}(x, y)\right)^+ & \text{if } c(y) > c(x), z = 0 \end{cases} \tag{10}$$

for all $n \ge 0,\ x, y \in S,\ z \in \{0, 1\}$ and $t \ge 0$, where $(\Delta_n)_{n\in\mathbb{N}}$ fulfills (i) and (ii) of Theorem 2.

Then

$$\lim_{n\to\infty} T_n = \vartheta_0\ (\mathbf{P} - a.s.) \quad \text{and} \quad \lim_{n\to\infty} \mathbf{E}\,(T_n - \vartheta_0)^2 = 0.$$

In the proof in the appendix it is shown that the D_n contained in (10) fulfills Theorem 2.

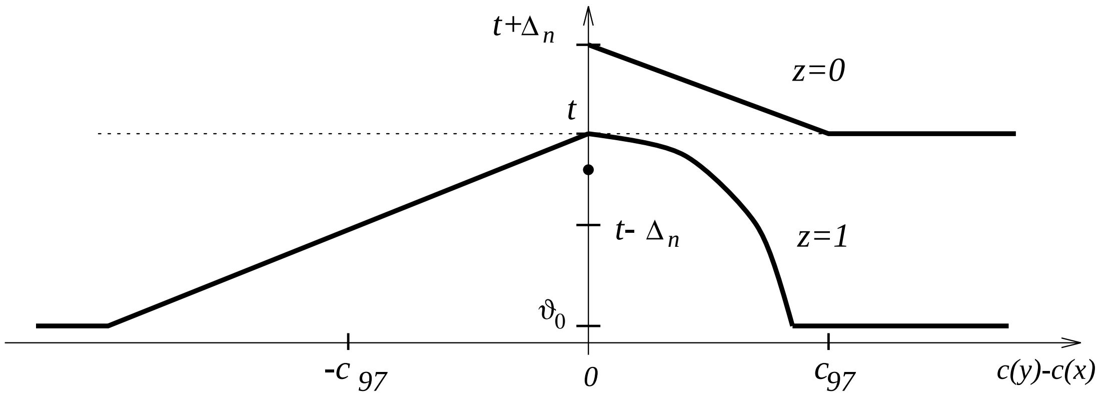

Figure 2: The temperature function $\tau_n(x,t,y,z)$ as given in (12) for a fixed t as funtion of $c(y)-c(x)$. The dot represents $t-\Delta_n c_l/c_{97}$.

The heuristic motivation of the scheme (10) is the following (see also Fig. 2): the temperature changes consist of a temperature step $\pm\Delta_n$ scaled by (a function of) the normed cost difference $\tilde{c}(x,y)$. More precisely : if there is an improvement of y over x then the temperature is decreased by $-\Delta_n\tilde{c}(x,y)\geq 0$ (note that in this case $\tilde{c}$ is negative). If the candidate y has equal costs then the temperature is lowered by the (estimated) smallest normalized quantity $\Delta_n c_l/c_{97}$. If a candidate y with higher costs is accepted ($z=1$) then the temperature is lowered by the (possibly huge) amount

$$\Delta_{n+1}\left(\frac{1}{\alpha(x,t,y)}-1+\tilde{c}(x,y)\right).$$

Finally, if a candidate y with higher costs is rejected ($z=0$) , the temperature is increased by the amount $\Delta_{n+1}(1-\tilde{c}(x,y))^+\geq 0$ depending linearly on small values of the normed cost difference $\tilde{c}(x,y)$.

Fig. 2 shows the behaviour of $\tau_n(x,t,y)$ as a function of the cost difference $c(y)-c(x)$. If c_{97} is chosen such that a large portion of all (X_n,Y_{n+1})-pairs has $|\tilde{c}(X_n,Y_{n+1})|\leq 1$ then the control will move mainly in the central part of the figure.

5 Computational Results for the Job Shop Scheduling Problem

In this Section we sketch some computational results of the scheme (10) applied to the job shop scheduling problem. A detailed study and comparison of several schemes will be published in a separate paper. The job shop scheduling problem is a well-known difficult problem of combinatorial optimization, see e.g. [10] and chapter 6 in [4] for details.

There are given n *jobs* , each consisting of m *operations* which have to be performed on m *machines*. The order in which the operations of one job have to be performed is fixed and also there is precisely one machine for each operation of a job on which this operation must be performed. Each operation needs a certain amount of *processing time*.The aim is to find an order x of the operations on each machine such that the finishing time $c(x)$ of the last job is minimized (the overall production time for all jobs).

A problem instance can be given as a directed graph where the vertices are the operations (plus a start and a stop node). Arcs are between operations of one job indicating their required order and between successive operations on each machine as given by a solution x. The cost function $c(x)$ is obtained as the length of a longest path in the graph where the length (duration) of a path is determined by the processing times of its nodes. As in [10] a candidate y is obtained from a solution x by reversing one arc that connects two operations on the same machine on a longest path.

For our test runs we used the Metropolis acceptance probability (1) and the temperature function as given in Theorem 4. To find reasonable values for c_{97} and c_l we took a sample $(x_1, y_1), \ldots, (x_N, y_N)$ of solutions x_i and candidates y_i, $i = 1 \ldots N$ by setting $\alpha \equiv 1$ during the first N steps. Then we put

$$c_l \;:=\; \min\{|c(x_i) - c(y_i)| \mid c(x_i) \neq c(y_i), i = 1 \ldots N\} \tag{11}$$

$$c_{97} \;:=\; 97\%\text{-quantile of the sample } |c(x_1) - c(y_1)|, \ldots, |c(x_N) - c(y_N)| \tag{12}$$

These are very rough estimates only. Analysing their quality is difficult, as they are based on dependent trials and will depend on the part of the unknown cost landscape that is reachable from the starting solution within N steps. Note however, that the convergence as shown in Theorem 4 does not depend on the precise values of c_{97} and c_l.

We used a small sample size $N := 0.003*$ no. of trials which, as was to be expected, needed only about 0.3% of the run time in our experiments.

The starting temperature t_0 was chosen such that the estimated mean cost increase of

$$\mu := \frac{1}{N} \sum_{i=1}^{N} |c(x_i) - c(y_i)|$$

would be accepted with probability 0.85 and similarly, under ϑ_0 an increase μ would be accepted with probability 0.05. A suitable temperature step was

$$\Delta_n \;:=\; \frac{t_0 - \vartheta_0}{n^{0.51}}$$

scaling Δ_n to the temperature range.

Table 1 lists our results for the test examples used in [10] from five runs with 1.000.000 trials each. The CPU times are for a Pentium 120 PC.

In the last two columns of Table 1 we also summarize the results from [10]. The schedule used in [10] allows different parameter settings. In column $\overline{C}_{min}$ we give the range of average best values obtained from different parameters and in column $\overline{t}$ we give the different run times (lower costs have higher run times). Though it is difficult to compare run times (the results in [10] were obtained on a VAX-785 and used a different stopping criterion) the table shows that our schedule yields very good results in a moderate time. In 24 of the 42 examples (marked by "*") the optimal value was found in each run and in only two instances (marked by "†") out of the 42 there occured mean relative errors larger than 2 %. In general, the performance of our algorithm gets worse for the larger problems. Here the the solution space and the local neighbourhood increase exponentially and the evaluation of a schedule takes more time. Nevertheless, there are hard and easy instances. E.g. law31 - law35 are large but easy problems, whereas law26 - law30 have less operations but are harder to solve. In particular, law29 is the only problem in this list not yet solved to optimality. Our algorithms has its weakest result with this problem (with comparatively little variance over the runs). Note however,

			Results with feedback schedule					Van Laarhoven et al.	
problem	jobs × mach.	optimal value resp. lower/ upper bound	$\overline{C}_{min}$	MRE	σ_{min}	$\overline{t}_{min}$	t_{tot}	$\overline{C}_{min}$	$\overline{t}$
law01*	10×5	666	666.0	0.000	0.0	0.3	45.0	666 - 707	6 - 123
law02*	10×5	655	655.0	0.000	0.0	2.5	44.4	663 - 719	6 - 117
law03*	10×5	597	597.0	0.000	0.0	4.7	47.8	617 - 689	5 - 129
law04*	10×5	590	590.0	0.000	0.0	4.2	47.7	593 - 665	6 - 121
law05*	10×5	593	593.0	0.000	0.0	0.1	44.1	593 - 594	5 - 118
law06*	15×5	926	926.0	0.000	0.0	0.3	61.6	926 - 937	16 - 286
law07*	15×5	890	890.0	0.000	0.0	0.6	66.8	890 - 948	15 - 376
law08*	15×5	863	863.0	0.000	0.0	0.4	62.6	863 - 905	16 - 292
law09*	15×5	951	951.0	0.000	0.0	0.3	62.0	951 - 965	13 - 283
law10*	15×5	958	958.0	0.000	0.0	0.2	60.9	958 -958	14 - 243
law11*	20×5	1222	1222.0	0.000	0.0	0.4	82.8	1222 - 1229	32 - 627
law12*	20×5	1039	1039.0	0.000	0.0	0.4	80.8	1039 - 1042	34 - 655
law13*	20×5	1150	1150.0	0.000	0.0	0.4	78.1	1150 - 1154	32 - 564
law14*	20×5	1292	1292.0	0.000	0.0	0.3	77.8	1292 -1292	27 - 462
law15*	20×5	1207	1207.0	0.000	0.0	1.3	84.9	1207 - 1299	34 - 736
law16†	10×10	945	979.8	0.037	3.4	4.1	92.5	966 - 1023	26 - 686
law17	10×10	784	786.4	0.003	4.3	12.9	89.2	787 - 861	23 - 720
law18	10×10	848	850.6	0.003	3.3	34.8	81.3	861 - 902	23 - 673
law19	10×10	842	846.2	0.005	3.8	46.5	77.5	853 -950	24 -830
law20	10×10	902	903.0	0.001	2.0	13.1	79.1	908 - 1021	30 - 667
law21	15×10	1046	1053.8	0.007	3.3	76.6	112.8	1067 - 1176	69 - 1991
law22	15×10	927	933.8	0.007	3.4	55.3	116.4	944 - 1125	65 - 2163
law23*	15×10	1032	1032.0	0.000	0.0	1.8	123.7	1032 - 1155	63 - 2093
law24	15×10	935	945.6	0.011	3.0	32.8	115.0	966 - 1101	71 - 2098
law25	15×10	977	984.2	0.007	4.4	59.9	126.2	1004 - 1114	77 - 2133
law26*	20×10	1218	1218.0	0.000	0.0	14.6	173.0	1219 - 1397	139 - 4342
law27	20×10	1235	1259.0	0.019	5.2	120.4	176.5	1273 - 1434	139 - 4535
law28	20×10	1216	1221.8	0.005	6.1	136.9	170.6	1244 - 1414	135 - 4354
law29†	20×10	1130/1153	1193.6	0.056	3.2	91.0	176.4	1226 - 1387	138 - 4408
law30*	20×10	1355	1355.0	0.000	0.0	12.0	178.2	1355 - 1539	145 - 3956
law31*	30×10	1784	1784.0	0.000	0.0	13.0	298.5	1784 - 1882	442 - 1517
law32*	30×10	1850	1850.0	0.000	0.0	11.9	298.4	1850 - 1921	492 - 1752
law33*	30×10	1719	1719.0	0.000	0.0	16.3	298.6	1726 - 1761	433 - 1880
law34*	30×10	1721	1721.0	0.000	0.0	25.5	297.8	1775 - 1816	470 - 1886
law35*	30×10	1888	1888.0	0.000	0.0	16.3	327.2	1890 - 2011	434 - 1668
law36	15×15	1268	1283.4	0.012	8.5	81.1	237.2	1300 - 1487	152 - 5346
law37	15×15	1397	1421.4	0.017	3.6	125.7	205.9	1442 - 1580	173 - 5287
law38	15×15	1196	1214.4	0.015	5.3	108.7	194.4	1227 - 1422	173 - 5480
law39	15×15	1233	1245.2	0.010	4.4	126.3	205.1	1258 - 1408	186 - 5766
law40	15×15	1222	1235.0	0.011	4.8	99.7	196.0	1247 - 1399	162 - 5373
mt06*	6 ×6	55	55.0	0.000	0.0	0.1	36.5	55 -56	8 - 52
mt10	10×10	930	944.2	0.015	3.9	66.4	86.0	933 - 1039	113 - 57772
mt20	20×5	1165	1172.6	0.007	4.4	46.9	79.6	1173 - 1354	123 - 62759

Table 1: Results from 5 runs of 1.000.000 trials each. $\overline{C}_{min}$ is the average value of the minimal costs in the five runs, MRE denotes the mean of the relative error, σ_{min} is the standard deviation of the minimal costs, $\overline{t}_{min}$ is the average time spent to find the minimal value and t_{tot} is the time (CPU secs) for one run.

that the relative error is calculated with respect to the known lower bound of the optimal value. With respect to the upper bound the MRE of our algorithm for law29 would be 0.035 or 3.5 %.

Vaessens, Aarts and Lenstra [19] give a detailed computational study of 35 different methods for the job shop scheduling problem including standard simulated annealing and taboo search. They use 13 test examples : mt10, law02, law19, law21, law24, law25, law29 and law36 - law40. Their paper lists the mean relative error for each method averaged over all 13 instances. These error values vary from 0.54% to 8.41% with most values larger than 2%. Our mean error averaged over these 13 examples in Table 1 is 1.4%. Also our method seems to be among the fastest. The only methods from [19] that seem comparable are taboo search algorithms. Their disadvantage is that they need more problem specific programming effort e.g. to administrate the taboo lists. Again, this comparison shows that our algorithm competes very well with other deterministic and randomized procedures on a broad range of test cases.

6 Summary

The paper presents analytical and empirical results on a new method to control the temperature parameter in simulated annealing. It allows to choose the temperature depending on the most recent history of the process. Using a general result on inhomogeneous Markov chains, it is shown that under mild bounding conditions on the temperature changes the process of solutions converges to a distribution that is concentrated near the optimal states. Within these mathematical limits the precise temperature control may be chosen rather arbitrarily.

We defined a temperature schedule that reacts to the process in a plausible way: it raises the temperature if the process seems to get stuck in a local optimum and lowers it if the process seems to move too fast. This heuristically motivated scheme proved very successful in the application to the job shop scheduling problem, where it competes very well with other heuristic search methods and with problem specific methods.

Still some problem instances seem to be very hard even for the adaptive temperature schedule. In our future research we shall therefore include parallel techniques. First results indicate that this improves the empirical quality of our algorithm drastically. Also we shall examine other applications from combinatorial optimization.

Appendix

6.1 Proof of Theorem 1

Proof : 1. From the stationarity of limiting distributions we have $\pi_{\vartheta_0} q_{\vartheta_0} = \pi_{\vartheta_0}$. Hence we have for any $n, m \in \mathbb{N}$ and arbitrary event sets A:

$$
\begin{aligned}
&\Big| \mathbf{P}[X_{n+m} = x \mid X_0 = x_0, T_0] \;-\; \pi_{\vartheta_0}(x) \Big| \\
&\qquad = \; \Big| \mathbf{E}\Big[\mathbf{P}\left[X_{n+m} = x \mid X_n, T_n\right] \Big| X_0 = x_0, T_0 \Big] \; - \pi_{\vartheta_0}(x) \Big| \\
&\qquad \leq \; \mathbf{E}\Big[\mid \; \mathbf{P}\left[X_{n+m} = x \mid X_n, T_n\right] \; - \; \pi_{\vartheta_0}(x) \mid \Big| X_0 = x_0, T_0 \Big]
\end{aligned} \tag{13}
$$

$$
\begin{aligned}
&= \mathbf{E}\Big[\mathbf{1}_A \cdot |\, \mathbf{P}[X_{n+m} = x \mid X_n, T_n] \;-\; \pi_{\vartheta_0}(x) \,| \;\Big|X_0 = x_0, T_0\Big] \\
&\quad + \mathbf{E}\Big(\mathbf{1}_{A^c} \cdot |\, \mathbf{P}[X_{n+m} = x \mid X_n, T_n] \;-\; \pi_{\vartheta_0}(x) \,| \;\Big|\; X_0 = x_0, T_0\Big].
\end{aligned}
$$

Now let $A = A_{\varepsilon,n} := [\, |T_n - \vartheta_0| < \varepsilon \,]$ for some $\varepsilon > 0$. For better readability we shall drop the condition $[X_0 = x_0, T_0]$ from our notation.

2. For the second term in (13) we have from (ii) that for any $\varepsilon > 0$ and all $m \in \mathbb{N}$

$$
\begin{aligned}
&\lim_{n\to\infty} \mathbf{E}\Big(\mathbf{1}_{A^c} \cdot \;|\, \mathbf{P}[X_{n+m} = x \mid X_n, T_n] \;-\; \pi_{\vartheta_0}(x) \,| \;\Big) \le \lim_{n\to\infty} \mathbf{E}\,(\mathbf{1}_{A^c} \cdot 2\,) \\
&\quad = \;\; 2 \cdot \lim_{n\to\infty} \mathbf{P}(\; |\; T_n - \vartheta_0| \ge \varepsilon\;) = 0.
\end{aligned}
$$

3. Now let $\delta > 0$ be arbitrary. We show that there are $m_0, n_0 \in \mathbb{N}$ such that for all $n \ge n_0$

$$
\mathbf{E}\Big(\mathbf{1}_A \cdot \;|\, \mathbf{P}[X_{n+m_0} = x \mid X_n, T_n] \;-\; \pi_{\vartheta_0}(x) \,| \;\Big) \;<\; \delta \tag{14}
$$

then (3) follows from (13). Let q_t^m denote the m-fold product of the matrix q_t. We have for any $\varepsilon > 0$ and all $n, m \in \mathbb{N}$

$$
\begin{aligned}
&\mathbf{E}\Big(\mathbf{1}_A \cdot \;|\, \mathbf{P}[X_{n+m} = x \mid X_n, T_n] \;-\; \pi_{\vartheta_0}(x) \,| \;\Big) \\
&\quad \le \;\; \mathbf{E}\Big(\mathbf{1}_A \cdot \;|\, \mathbf{P}[X_{n+m} = x \mid X_n, T_n] \;-\; q_{T_n}^m(X_n, x) \,| \;\Big) \\
&\qquad\qquad + \mathbf{E}\Big(\mathbf{1}_A \cdot \;|\, q_{T_n}^m(X_n, x) \;-\; q_{\vartheta_0}^m(X_n, x) \,| \;\Big) \\
&\qquad\qquad + \mathbf{E}\Big(\mathbf{1}_A \cdot \;|\, q_{\vartheta_0}^m(X_n, x) \;-\; \pi_{\vartheta_0}(x) \,| \;\Big) \\
&\quad \le \;\; \mathbf{E}\Big(\mathbf{1}_A \cdot \;|\, \mathbf{P}[X_{n+m} = x \mid X_n, T_n] \;-\; q_{T_n}^m(X_n, x) \,| \;\Big) \\
&\qquad\qquad + \mathbf{E}\Big(\mathbf{1}_{[\;|\;T_n - \vartheta_0| < \varepsilon]} \;\cdot\; |\, q_{T_n}^m(X_n, x) \;-\; q_{\vartheta_0}^m(X_n, x) \,| \;\Big) \\
&\qquad\qquad + \mathbf{E}\Big(\;|\, q_{\vartheta_0}^m(X_n, x) \;-\; \pi_{\vartheta_0}(x) \,| \;\Big).
\end{aligned}
\tag{15}
$$

4. As $\lim_{m\to\infty} q_{\vartheta_0}^m(y, x) = \pi_{\vartheta_0}(x)$ by (iv) and as S is finite we can find $m_0 \in \mathbb{N}$ such that

$$
\max_{x,y\in S} \;|\, q_{\vartheta_0}^{m_0}(y, x) \;-\; \pi_{\vartheta_0}(x) \,| \;\; < \;\delta/3
$$

and hence in (15)

$$
\mathbf{E}\Big(\;|\, q_{\vartheta_0}^{m_0}(X_n, x) \;-\; \pi_{\vartheta_0}(x) \,| \;\Big) \;<\; \delta/3. \tag{16}
$$

5. As $t \mapsto q_t^{m_0}(y, x)$ is continuous we can find $\varepsilon > 0$ such that for all n

$$
\mathbf{1}_{[\;|\;T_n - \vartheta_0| < \varepsilon]} \;\cdot\; \max_{x,y\in S} \;|\, q_{T_n}^{m_0}(y, x) \;-\; q_{\vartheta_0}^{m_0}(y, x) \,| \;\; < \;\delta/3 \qquad (\mathbf{P}\text{-}a.s.)
$$

hence in (15)

$$
\mathbf{E}\Big(\mathbf{1}_{[\;|\;T_n - \vartheta_0| < \varepsilon]} \;\cdot\; |\, q_{T_n}^{m_0}(X_n, x) \;-\; q_{\vartheta_0}^{m_0}(X_n, x) \,| \;\Big) \;<\; \delta/3. \tag{17}
$$

6. It remains to bound the first term in (15) for m_0 and ε as found in 4. and 5. . In the following formulae sums are supposed to run over S.

6.1 We have for arbitrary $n, m \in \mathbb{N}$ and $k := n+m$

$$\begin{aligned}
&\sum_{x_k} \Big| \mathbf{P}\left[X_k = x_k \mid X_n = x_n, T_n = t_n\right] \; - \; q_{t_n}^m(x_n, x_k) \; \Big| \\
&\quad = \; \sum_{x_k} \Big| \sum_{x_{k-1}} \int_0^\infty \Big(q_{t_{k-1}}(x_{k-1}, x_k) \; \mathbf{P}\left[X_{k-1} = x_{k-1}, T_{k-1} = dt_{k-1} \mid X_n = x_n, T_n = t_n\right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad - \; q_{t_n}(x_{k-1}, x_k) \; q_{t_n}^{m-1}(x_n, x_{k-1}) \Big) \; \Big| \\
&\quad = \; \sum_{x_k} \Big| \sum_{x_{k-1}} \int_0^\infty \\
&\qquad\quad \Big(q_{t_n}(x_{k-1}, x_k) \cdot \Big(\mathbf{P}\left[X_{k-1} = x_{k-1}, T_{k-1} = dt_{k-1} \mid X_n = x_n, T_n = t_n\right] \; - \; q_{t_n}^{m-1}(x_n, x_{k-1}) \Big) \\
&\qquad\quad + \mathbf{P}\left[X_{k-1} = x_{k-1}, T_{k-1} = dt_{k-1} \mid X_n = x_n, T_n = t_n\right] \cdot \Big(q_{t_{k-1}}(x_{k-1}, x_k) - q_{t_n}(x_{k-1}, x_k) \Big) \Big) \; \Big| \\
&\quad \le \; \sum_{x_{k-1}} \Big| \mathbf{P}\left[X_{k-1} = x_{k-1} \mid X_n = x_n, T_n = t_n\right] \; - \; q_{t_n}^{m-1}(x_n, x_{k-1}) \Big| \\
&\qquad\qquad + \mathbf{E}\left[\sum_{x_k} \Big| q_{T_{k-1}}(X_{k-1}, x_k) - q_{T_n}(X_{k-1}, x_k) \Big| \mid X_n = x_n, T_n = t_n \right].
\end{aligned}$$

Repeating this argument for $k-1, k-2, \ldots, k-m+1$ i.e. for $n+m-1, n+m-2, \ldots, n+1$ we obtain

$$\begin{aligned}
&\sum_{x' \in S} \Big| \mathbf{P}\left[X_{n+m} = x' \mid X_n = x_n, T_n = t_n\right] \; - \; q_{t_n}^m(x_n, x') \; \Big| \\
&\qquad \le \; \sum_{i=1}^{m-1} \mathbf{E}\left[\sum_{x' \in S} \Big| q_{T_{n+i}}(X_{n+i}, x') - q_{T_n}(X_{n+i}, x') \Big| \mid X_n = x_n, T_n = t_n \right]
\end{aligned}$$

and in particular for any $x \in S$

$$\begin{aligned}
&\Big| \mathbf{P}\left[X_{n+m} = x \mid X_n = x_n, T_n = t_n\right] \; - \; q_{t_n}^m(x_n, x) \; \Big| \\
&\qquad \le \; \sum_{i=1}^{m-1} \mathbf{E}\left[\sum_{x' \in S} \Big| q_{T_{n+i}}(X_{n+i}, x') - q_{T_n}(X_{n+i}, x') \Big| \mid X_n = x_n, T_n = t_n \right].
\end{aligned} \tag{18}$$

6.2 From condition (iii) we also have that $t \mapsto q_t(x,y)$ is uniformly continuous on the compact set $\overline{U_\varepsilon} := \{t \in \mathbb{R} \mid |t - \vartheta_0| \le \varepsilon\}$ for the ε chosen in 5., i.e.

$$\forall\, x, x' \in S \quad \forall\, \delta' > 0 \quad \exists\, \eta > 0 \quad \forall\, t, t' \in \overline{U_\varepsilon} \quad \Big(|t - t'| < \eta \Rightarrow |q_t(x,x') - q_{t'}(x,x')| \; < \; \delta' \Big).$$

As S is finite we may conclude that for m_0 and ε as in 4. and 5. above

$$\begin{aligned}
&\forall\, \delta' > 0 \quad \exists\, \eta > 0 \quad \forall\, t, t_1, \ldots, t_{m_0 - 1} \in \overline{U_\varepsilon} \\
&\qquad\qquad \Big(\forall i = 1, \ldots, m_0 - 1 \;\; |t - t_i| < \eta \Rightarrow \max_{x \in S} \sum_{i=1}^{m_0 - 1} \sum_{x' \in S} |q_t(x,x') - q_{t'}(x,x')| \; < \; \delta' \Big).
\end{aligned}$$

Hence with $\delta' = \delta$ we have for

$$B = B_\eta := \bigcap_{i=1}^{m_0-1} \Big[\ |T_{n+i} - \vartheta_0| < \varepsilon \ \text{ and } \ |T_{n+i} - T_n| < \eta\Big]$$

we have that for m_0 and ε as above there is an η such that

$$\mathbf{1}_A \cdot \mathbf{1}_B \cdot \sum_{i=1}^{m-1} \sum_{x' \in S} \Big| q_{T_{n+i}}(X_{n+i}, x') - q_{T_n}(X_{n+i}, x')\Big| \ < \ \delta/6 \quad (\mathbf{P}\text{-}a.s.). \tag{19}$$

6.3 Returning to the first summand in (15) we see from (19)and (18) that with m_0, ε and η as above

$$\begin{aligned}
&\mathbf{E}\Big(\ \mathbf{1}_A \cdot \ |\ \mathbf{P}[X_{n+m_0} = x \mid X_n, T_n] \ - \ q_{T_n}^{m_0}(X_n, x)\ |\ \Big) \\
&\quad \le \ \mathbf{E}\Big(\ \mathbf{1}_A \cdot \sum_{i=1}^{m_0-1} \mathbf{E}\Big[\sum_{x' \in S} \ |\ q_{T_{n+i}}(X_{n+i}, x') - q_{T_n}(X_{n+i}, x')\ |\ \Big|\ X_n, T_n\Big]\ \Big) \\
&\quad = \ \mathbf{E}\Big(\ \mathbf{1}_A \cdot \sum_{i=1}^{m_0-1} \sum_{x' \in S} \ |\ q_{T_{n+i}}(X_{n+i}, x') - q_{T_n}(X_{n+i}, x')\ |\ \Big) \\
&\quad = \ \mathbf{E}\Big(\ \mathbf{1}_A \cdot \mathbf{1}_B \cdot \sum_{i=1}^{m_0-1} \sum_{x' \in S} \ |\ \ q_{T_{n+i}}(X_{n+i}, x') - q_{T_n}(X_{n+i}, x')\ |\ \ \Big) \\
&\qquad\quad + \mathbf{E}\Big(\ \mathbf{1}_A \cdot \mathbf{1}_B^c \cdot \sum_{i=1}^{m_0-1} \sum_{x' \in S} \ |\ \ q_{T_{n+i}}(X_{n+i}, x') - q_{T_n}(X_{n+i}, x')\ |\ \ \Big) \\
&\quad \le \ \ \delta/6 \ + \ 2(m_0 - 1) \cdot \mathbf{P}\,(A \ \cap \ B^c) \\
&\quad \le \ \ \delta/6 \ + \ 2(m_0 - 1) \cdot \mathbf{P}\Big(\exists i = 1, \ldots, m_0 - 1 \ \ |T_{n+i} - \vartheta_0| \ge \varepsilon \ \text{ or } \ |T_{n+i} - T_n| \ge \eta\Big) \\
&\quad \le \ \ \delta/6 \ + \ 2(m - 1) \cdot \sum_{i=1}^{m_0-1} \Big(\mathbf{P}(\ |\ T_{n+i} - \vartheta_0| \ge \varepsilon\ ) \ + \ \mathbf{P}(\ |\ T_{n+i} - T_n| \ge \eta\ )\Big)
\end{aligned}$$

as for Markov matrices

$$\sum_{x' \in S} \ |\ q_t(x, x') - q_{t'}(x, x')\ | \ \ \le \ 2.$$

From

$$\begin{aligned}
&\mathbf{P}(\ |\ T_{n+i} - T_n| \ge \eta) \\
&\quad \le \ \mathbf{P}(\ |\ T_{n+i} - \vartheta_0| \ge \eta/2 \ \text{ or } \ |T_n - \vartheta_0| \ge \eta/2\ ) \\
&\quad \le \ \mathbf{P}(\ |\ T_{n+i} - \vartheta_0| \ge \eta/2\ ) \ + \ \mathbf{P}(\ |\ T_n - \vartheta_0| \ge \eta/2\ )
\end{aligned}$$

and condition (ii) we see that we can find $n_0 \in \mathbb{N}$ such that for all $n \ge n_0$ we have

$$\mathbf{P}(\ |\ T_n - \vartheta_0| \ge \varepsilon) \ < \frac{\delta}{36(m_0 - 1)^2}$$

and

$$\mathbf{P}(\ |\ T_n - \vartheta_0| \ge \eta) \ < \frac{\delta}{36(m_0 - 1)^2}.$$

But then for $n \ge n_0$

$$\mathbf{E}\Big(\ \mathbf{1}_A \cdot \ |\ \mathbf{P}[X_{n+m_0} = x \mid X_n, T_n] \ - \ q_{T_n}^{m_0}(X_n, x)\ |\ \Big) \tag{20}$$

$$
\begin{aligned}
&\leq \delta/6 \;+\; 2(m_0-1)\cdot\sum_{i=1}^{m_0-1} 3\cdot\frac{\delta}{36(m_0-1)^2}\\
&= \delta/3.
\end{aligned}
$$

Now (20), (17) and (16) prove (14) which completes the proof.

■

6.2 Proof of Theorem 2

For the proof of Theorem 2b) we need the following lemma. The first part of it represents a generalized Cauchy convergence criterion and is a slight generalization of Lemma 4.1 from [15].

Lemma 1

a) Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of real numbers and let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence with $\limsup_{n\to\infty}\varepsilon_n = 0$. Assume that for n sufficiently large and all $n' \geq n$

$$a_{n'} \;\leq\; a_n + \varepsilon_n.$$

Then $\lim_{n\to\infty} a_n$ exists with a possible value $-\infty$.

b) If, in addition, there is a non-negative real-valued sequence $(b_n)_{n\in\mathbb{N}}$ such that

$$\sum_{n=1}^{\infty} b_n \;=\; \infty \qquad\text{and}\qquad \sum_{n=1}^{\infty}|a_n|b_n \;<\; \infty$$

then $\lim_{n\to\infty} a_n \;=\; 0.$

Proof of the Lemma: For a) see [15] : we have $\sup_{n'\geq n} a_{n'} - \varepsilon_n \leq a_n$ and therefore

$$
\begin{aligned}
\liminf_{n\to\infty} a_n \;&\geq\; \liminf_{n\to\infty}\left(\sup_{n'\geq n} a_{n'} - \varepsilon_n\right)\\
&\geq\; \lim_{n\to\infty}\sup_{n'\geq n} a_{n'} \;-\; \limsup_{n\to\infty}\varepsilon_n\\
&=\; \limsup_{n\to\infty} a_n.
\end{aligned}
$$

Hence $\lim_{n\to\infty} a_n$ exists.

To prove b) assume that $a := \lim_{n\to\infty} a_n \;\neq\; 0$. Let n_0 be such that for $n \geq n_0$ we have $|a_n| > |a|/2$. Then we obtain a contradiction by

$$\infty \;>\; \sum_{n=1}^{\infty}|a_n|b_n \;\geq\; \sum_{n=n_0}^{\infty}|a_n|b_n \;\geq\; |a|/2\cdot\sum_{n=n_0}^{\infty} b_n \;=\; \infty.$$

■

Proof of Theorem 2:

a) **Almost sure convergence**

Note that the proof follows the lines of a convergence proof for autoregressive estimation processes in [3] B.5.

Let $s := \sum_{n=1}^{\infty} \Delta_n^2 < \infty$ and $C_n := T_n - \vartheta_0$ for $n \geq 1$. From the definition of T_n we have for all $n \geq 0$

$$\begin{aligned} C_{n+1}^2 &= \left(\max\{\vartheta_0, T_n - D_{n+1}\} - \vartheta_0\right)^2 = \left(\max\{0, C_n - D_{n+1}\}\right)^2 \\ &\leq (C_n - D_{n+1})^2 = C_n^2 - 2C_n D_{n+1} + D_{n+1}^2, \end{aligned} \tag{21}$$

hence with $\mathcal{F}_n = (T_0, T_1, \ldots, T_n)$

$$\mathbf{E}\,[C_{n+1}^2 \mid \mathcal{F}_n] \leq C_n^2 - 2C_n \mathbf{E}\,[D_{n+1} \mid \mathcal{F}_n] + \mathbf{E}\,[D_{n+1}^2 \mid \mathcal{F}_n].$$

With (iv) and with $C_n \geq 0$ ($\mathbf{P}$-a.s.) we have

$$0 \leq \mathbf{E}\,[C_{n+1}^2 \mid \mathcal{F}_n] \leq C_n^2 + \mathbf{E}\,[D_{n+1}^2 \mid \mathcal{F}_n]. \tag{22}$$

For

$$B_n := C_n^2 + s - \sum_{k=1}^{n-1} \mathbf{E}\,[D_{k+1}^2 \mid \mathcal{F}_k]$$

we obtain from (ii) and (iii) that $B_n \geq 0$. From (22) we have

$$\begin{aligned} \mathbf{E}\,[B_{n+1} \mid \mathcal{F}_n] &= \mathbf{E}\,[C_{n+1}^2 \mid \mathcal{F}_n] + s - \sum_{k=1}^{n} \mathbf{E}\,[\mathbf{E}\,[D_{k+1}^2 \mid \mathcal{F}_k] \mid \mathcal{F}_n] \\ &\leq C_n^2 + \mathbf{E}\,[D_{n+1}^2 \mid \mathcal{F}_n] + s - \sum_{k=1}^{n} \mathbf{E}\,[D_{k+1}^2 \mid \mathcal{F}_k] \\ &= C_n^2 + s - \sum_{k=1}^{n-1} \mathbf{E}\,[D_{k+1}^2 \mid \mathcal{F}_k] \\ &= B_n. \end{aligned}$$

Hence B_n is a non-negative supermartingale. Therefore $\lim_{n\to\infty} B_n$ exists a.s. by the submartingale convergence theorem (see e.g. [11] 29.A). But as also

$$s - \sum_{k=1}^{n-1} \mathbf{E}\,[D_{k+1}^2 \mid \mathcal{F}_k]$$

is a non-negative supermartingale we may conclude that $\lim_{n\to\infty} C_n^2$ and hence $\lim_{n\to\infty} C_n$ exist $\mathbf{P}$-a.s.. In part b) it is shown independently that $\lim_{n\to\infty} \mathbf{E}\, C_n^2 = 0$, hence it follows that $\lim_{n\to\infty} C_n = 0$ ($\mathbf{P}$-a.s.).

b) **Convergence in quadratic mean**

2. From (21) we obtain for all $m \geq 0$

$$C_{n+m}^2 \leq C_n^2 + \sum_{i=n}^{n+m-1} \Big(D_{i+1}^2 - 2D_{i+1}C_i \Big). \tag{23}$$

As (iv) implies

$$\mathbf{E}\Big(2D_{i+1}C_i\Big) = 2\mathbf{E}\Big(\mathbf{E}\,[D_{i+1} \mid \mathcal{F}_i] C_i\Big) \geq \mathbf{0}.$$

we have for any $m \geq 0$

$$\mathbf{E}\, C_{n+m}^2 \;\leq\; \mathbf{E}\, C_n^2 \;-\; 2\sum_{i=n}^{n+m-1} \mathbf{E}\Big(D_{i+1}C_i\Big) \;+\; \sum_{i=n+1}^{n+m} \Delta_i^2 \tag{24}$$
$$\leq \;\; \mathbf{E}\, C_n^2 \;+\; \sum_{i=n+1}^{n+m} \Delta_i^2$$
$$\leq \;\; \mathbf{E}\, C_n^2 \;+ \varepsilon_n$$

where

$$\varepsilon_n := \sum_{i=n+1}^{\infty} \Delta_{i+1}^2.$$

Note that $\lim_{n\to\infty} \varepsilon_n = 0$.

3. On the other hand we obtain from (24)

$$\mathbf{E}\, C_{n+1}^2 \;\leq EC_0^2 \;+\; \sum_{i=0}^{n} \Delta_{i+1}^2 \;-\; 2\sum_{i=0}^{n} \mathbf{E}\, C_i D_{i+1}$$

and hence

$$2\sum_{i=0}^{n} \mathbf{E}\, C_i D_{i+1} \;\leq\; \mathbf{E}\, C_0^2 + \sum_{i=0}^{n} \Delta_{i+1}^2 < \infty \tag{25}$$

4. We apply Lemma 1 to (24) with $a_n := \mathbf{E}\, C_n^2$

and

$$b_n := \frac{K\Delta_{n+1}}{\sum_{i=0}^{n+1} \Delta_i}$$

where we put $\Delta_0 := |C_0| = |t_0 - \vartheta_0|$. From

$$\lim_{n\to\infty} \frac{\Delta_n}{\sum_{i=0}^{n} \Delta_i} \;=\; 0$$

we may conclude (see I.76 in [14]) that

$$\lim_{N\to\infty} \frac{\sum_{n=0}^{N} \frac{\Delta_n}{\sum_{i=0}^{n} \Delta_i}}{\log \sum_{i=0}^{N} \Delta_i} \;=\; 1.$$

But as $\lim_{N\to\infty} \log \sum_{i=0}^{N} \Delta_i = \infty$ by (i) we have

$$\sum_{n=0}^{\infty} b_n \;=\; K\sum_{n=0}^{\infty} \frac{\Delta_{n+1}}{\sum_{i=0}^{n+1} \Delta_i} \;=\; K \lim_{n\to\infty} \log \sum_{i=0}^{n+1} \Delta_i \;=\; \infty.$$

From (iii) we obtain

$$|C_n| \;=\; \Big|\max\{\vartheta_0, T_{n-1} - D_n\} - \vartheta_0\Big| \tag{26}$$
$$\leq \;\; \Big|C_{n-1} - D_n\Big| \;\leq\; |C_0| \;+\; \sum_{i=1}^{n} |D_i|$$
$$\leq \;\; |C_0| \;+\; \sum_{i=1}^{n} \Delta_i \quad (\mathbf{P}\text{-}a.s.)$$
$$\leq \;\; \sum_{i=0}^{n+1} \Delta_i$$

For the second condition in Lemma 1b) we obtain from (26) as $C_n \geq 0$

$$
\begin{aligned}
\sum_{n=0}^{\infty} b_n \; \mathbf{E}\, C_n^2 \;&=\; \sum_{n=0}^{\infty} \frac{K\Delta_{n+1}}{\sum_{i=0}^{n+1}\Delta_i} \; \mathbf{E}\, C_n^2 \\
&\leq\; \sum_{n=0}^{\infty} \frac{K\Delta_{n+1}}{\sum_{i=0}^{n+1}\Delta_i} \; \mathbf{E}\Big(C_n \sum_{i=0}^{n+1} \Delta_i\Big) \\
&\leq\; \sum_{n=0}^{\infty} K\Delta_{n+1} \; \mathbf{E}\, C_n \\
&\leq\; \sum_{n=0}^{\infty} \mathbf{E}\Big(C_n \mathbf{E}\;[D_{n+1} \mid \mathcal{F}_n]\Big) \\
&=\; \sum_{n=0}^{\infty} \mathbf{E}\Big(D_{n+1} C_n\Big) \\
&<\; \infty
\end{aligned}
$$

Here we used (iv) and (25) in the last three lines. Now the assertion of the theorem follows from Lemma 1 b). ■

6.3 Proof of Theorem 3

Proof : We only have to show that D_n as given in (9) fulfills Theorem 2 (iii) and (iv). As we have $|z - \delta\alpha(x,t,y)| \leq 1$ condition (iii) follows immediately from (9). Let again $\mathcal{F}_n = (T_0, T_1, \ldots T_n)$. To prove (iv) we first examine $\mathbf{E}\,[D_n \mid X_{n-1}, \mathcal{F}_{n-1}, Y_n]$ which is equal to Δ_n if $c(Y_n) \leq c(X_{n-1})$. If $c(Y_n) > c(X_{n-1})$ we have

$$
\begin{aligned}
\mathbf{E}\,[D_n \mid X_{n-1}, T_{n-1}, Y_n] \;&=\; \Delta_n \cdot \mathbf{E}\,[Z_n - \delta\alpha(X_{n-1}, \mathcal{F}_{n-1}, Y_n) \mid X_{n-1}, \mathcal{F}_{n-1}, Y_n] \\
&=\; \Delta_n \cdot \bigg(\Big(1 - \delta\alpha(X_{n-1}, T_{n-1}, Y_n)\Big) \cdot \mathbf{P}\,[Z_n = 1 \mid X_{n-1}, \mathcal{F}_{n-1}, Y_n] \\
&\qquad -\; \delta\alpha(X_{n-1}, T_{n-1}, Y_n) \cdot \mathbf{P}\,[Z_n = 0 \mid X_{n-1}, T_{n-1}, Y_n]\bigg) \\
&=\; \Delta_n \cdot \bigg(\Big(1 - \delta\alpha(X_{n-1}, T_{n-1}, Y_n)\Big)\alpha(X_{n-1}, T_{n-1}, Y_n) \\
&\qquad -\; \delta\alpha(X_{n-1}, T_{n-1}, Y_n)\Big(1 - \alpha(X_{n-1}, T_{n-1}, Y_n)\Big)\bigg) \\
&=\; \Delta_n \cdot \alpha(X_{n-1}, T_{n-1}, Y_n)(1 - \delta) \\
&\geq\; K \cdot \Delta_n.
\end{aligned}
$$

with $K := (1-\delta)\underline{\alpha} \in (0,1)$. But then

$$
\mathbf{E}\;[D_n \mid \mathcal{F}_{n-1}] \;=\; \mathbf{E}\Big[\mathbf{E}\;[D_n \mid \mathcal{F}_{n-1}, Y_n, Z_n]\Big|\mathcal{F}_{n-1}\Big] \;\geq\; K \cdot \Delta_n.
$$

■

6.4 Proof of Theorem 4

Proof : Let $\underline{c} := \min\{|c(x) - c(y)| \mid c(x) \neq c(y),\; x \in S, y \in N(x)\}$ and $\overline{c} = \max\{c(x) - c(y) \mid x, y \in S\}$

and $\underline{\alpha} = \exp\left(-\frac{\overline{c}}{\vartheta_0}\right)$ as before. Put

$$\Delta'_n \;:=\; \Delta_n \left(1 \;+\; \frac{1}{\underline{\alpha}} + \frac{\overline{c}}{c_{97}}\right)$$

for $n \in \mathbb{N}$. Then the sequence $(\Delta'_n)_{n\in\mathbb{N}}$ fulfills (i) and (ii) of Theorem 2.

Let

$$d_n(x,t,y) \;:=\; \begin{cases} -\;\; \Delta_n \;\tilde{c}(x,y) & \text{if } c(y) < c(x) \\ \quad\; \Delta_n \;\frac{c_l}{c_{97}} & \text{if } c(y) = c(x) \\ \quad\; \Delta_n \;\left(\frac{1}{\alpha(x,t,y)} - 1 + \tilde{c}(x,y)\right) & \text{if } c(y) > c(x), z = 1 \\ -\;\; \Delta_n \;\left(1 - \tilde{c}(x,y)\right)^+ & \text{if } c(y) > c(x), z = 0 \end{cases} \tag{27}$$

then we have from (10) that $T_{n+1} = \max\{\vartheta_0, T_n - D_{n+1}\}$, i.e. (6) holds and

$$\begin{aligned} |d_n(x,t,y)| \quad &\leq \quad \begin{cases} \Delta_n \;|\tilde{c}(x,y)| & \text{if } c(y) < c(x) \\ \Delta_n \;\frac{c_l}{c_{97}} & \text{if } c(y) = c(x) \\ \Delta_n \;\left(\frac{1}{\alpha(x,t,y)} + |\tilde{c}(x,y)|\right) & \text{if } c(y) > c(x), z = 1 \\ \Delta_n \;\left(1 - \tilde{c}(x,y)\right)^+ & \text{if } c(y) > c(x), z = 0 \end{cases} \\ &\leq \quad \begin{cases} \Delta_n \;\frac{\overline{c}}{c_{97}} & \text{if } c(y) < c(x) \\ \Delta_n \;\frac{\overline{c}}{c_{97}} & \text{if } c(y) = c(x) \\ \Delta_n \;\left(\frac{1}{\underline{\alpha}} + \frac{\overline{c}}{c_{97}}\right) & \text{if } c(y) > c(x), z = 1 \\ \Delta_n & \text{if } c(y) > c(x), z = 0 \end{cases} \\ &\leq \quad \Delta_n \left(1 \;+\; \frac{1}{\underline{\alpha}} + \frac{\overline{c}}{c_{97}}\right) \;\;=\;\; \Delta'_n \end{aligned}$$

hence with $D_n := d_n(X_n, T_n, Y_{n+1})$ condition (iii) of Theorem 2 holds. To show (iv) let us assume that $c(Y_{n+1}) - c(X_n) < 0$, then we obtain from (27)

$$\mathbf{E}\left[D_{n+1} \mid X_n, \mathcal{F}_n, Y_{n+1}\right] \;=\; \frac{-\Delta_{n+1}}{c_{97}} \mathbf{E}\left[c(Y_{n+1}) - c(X_n) \mid X_n, T_n, Y_{n+1}\right] \;\geq\; \Delta_{n+1}\,\frac{\underline{c}}{c_{97}}.$$

If $c(Y_{n+1}) - c(X_n) = 0$ we have

$$\mathbf{E}\left[D_{n+1} \mid X_n, \mathcal{F}_n, Y_{n+1}\right] \;=\; \Delta_{n+1}\frac{c_l}{c_{97}} \;\geq\; \Delta_{n+1}\,\frac{\underline{c}}{c_{97}}.$$

Now assume that $c(Y_{n+1}) - c(X_n) > 0$ then

$$\begin{aligned} &\mathbf{E}\left[D_{n+1} \mid X_n, \mathcal{F}_n, Y_{n+1}\right] \\ &\quad= \;\Delta_{n+1}\;\mathbf{E}\left[\frac{1}{\alpha(X_n,T_n,Y_{n+1})} \;-\; 1 \;+\; \frac{c(Y_{n+1}) - c(X_n)}{c_{97}} \mid X_n, \mathcal{F}_n, Y_{n+1}, Z_{n+1} = 1\right] \\ &\qquad\qquad \cdot\mathbf{P}\left[Z_{n+1} = 1 \mid X_n, \mathcal{F}_n, Y_{n+1}\right] \\ &\qquad -\Delta_{n+1}\mathbf{E}\left[\left(1 - \frac{c(Y_{n+1}) - c(X_n)}{c_{97}}\right)^+ \mid X_n, \mathcal{F}_n, Y_{n+1}, Z_{n+1} = 0\right] \cdot \mathbf{P}\left[Z_{n+1} = 0 \mid X_n, \mathcal{F}_n, Y_{n+1}\right] \\ &\quad= \;\Delta_{n+1}\Bigg(\left(\frac{1}{\alpha(X_n,T_n,Y_{n+1})} \;-\; \left(1 \;-\; \frac{c(Y_{n+1}) - c(X_n)}{c_{97}}\right)\right) \cdot \alpha(X_n,T_n,Y_{n+1}) \end{aligned}$$

$$
\begin{aligned}
& \quad -\Big(1 - \frac{c(Y_{n+1}) - c(X_n)}{c_{97}}\Big)^+ \cdot (1 - \alpha(X_n, T_n, Y_{n+1}))\Big) \\
\geq \ & \Delta_{n+1}\Big(1 \ - \ \Big(1 - \frac{c(Y_{n+1}) - c(X_n)}{c_{97}}\Big)^+ \alpha(X_n, T_n, Y_{n+1}) \\
& \quad -\Big(1 - \frac{c(Y_{n+1}) - c(X_n)}{c_{97}}\Big)^+ \cdot (1 - \alpha(X_n, T_n, Y_{n+1}))\Big) \\
= \ & \Delta_{n+1}\ \Big(1 \ - \ \Big(1 - \frac{c(Y_{n+1}) - c(X_n)}{c_{97}}\Big)^+\Big) \\
\geq \ & \Delta_{n+1}\ \Big(1 \ - \ \Big(1 - \frac{\underline{c}}{c_{97}}\Big)\Big) \\
= \ & \Delta_{n+1}\ \frac{\underline{c}}{c_{97}}.
\end{aligned}
$$

Hence we have

$$
\mathbf{E}\left[D_{n+1} \mid X_n, \mathcal{F}_n, Y_{n+1}\right] \ \geq \ \Delta_{n+1}\ \frac{\underline{c}}{c_{97}}
$$

and

$$
\mathbf{E}\left[D_{n+1} \mid \mathcal{F}_n\right] \ \geq \ \Delta_{n+1}\ \frac{\underline{c}}{c_{97}} \ = \ \Delta'_{n+1} \cdot K
$$

with

$$
K \ := \ \frac{\frac{\underline{c}}{c_{97}}}{\Big(1 \ + \ \frac{1}{\underline{\alpha}} + \frac{\overline{c}}{c_{97}}\Big)} \ > \ 0
$$

which completes the proof. ■

References

[1] Aarts, E. H. L., P. J. M. van Laarhoven, J. K. Lenstra and N. L. J. Ulder (1994) A Computational Study of Local Search Algorithms for Job Shop Scheduling. *ORSA Journal on Computing* **6**, 118–125

[2] Anily, S. and Federgruen, A. (1987) Simulated Annealing Methods with General Acceptance Probabilities. *J. Appl. Prob.* **24**, 657–667 .

[3] Bouleau, N. and D. Lépingle (1984) *Numerical Methods for Stochastic Processes.* J. Wiley, New York.

[4] P. Brucker (1995) *Scheduling Algorithms.* Springer Verlag, Berlin.

[5] Faigle, U. and W. Kern (1989) Note on the Convergence of Simulated Annealing Algorithms. *SIAM J. of Control and Optimization* **29**, 153-159.

[6] Greene, J. W. and Supowit, K. J. (1986) Simulated Annealing without Rejecting Moves. *IEEE Trans. on CAD* **5**, No.1, 221–228 .

[7] Hajek, B. (1988) Cooling Schedules for Optimal Annealing. *Math. Op. Res.* **13**, 311–329.

[8] Kolonko, M. (1994) A piecewise Markovian Model for Simulated Annealing with Stochastic Cooling Schedules. *J. Appl. Prob.* **32**, 649 – 658.

[9] van Laarhoven, P.J.M. and Aarts, E.H.L. (1987) *Simulated Annealing: Theory and Applications.* D. Reidel Publ. Comp., Dordrecht NL.

[10] van Laarhoven, P. J. M., E. H. L. Aarts and J. K. Lenstra (1992) Job Shop Scheduling by Simulated Annealing. *Operations Research* **40**, 113 – 125.

[11] Loève, M. (1978) *Probability Theory II.* Fourth Edition, Springer Verlag, Berlin.

[12] Lundy, M. and Mees, A. (1986) Convergence of an Annealing Algorithm. *Math. Progr.* **34**, 111–124 .

[13] Mitra, D., F. Romeo and A. Sangivanni-Vincentelli (1986) Convergence and Finite-Time Behaviour of Simulated Annealing. *Adv. Appl. Prob.* **18**, 747 – 771.

[14] Pólya, G. and G. Szegö (1970) "Aufgaben und Lehrsätze aus der Analysis I", 4th edition, Springer Verlag, Berlin.

[15] Schäl, M. (1975) Conditions for Optimality in Dynamic Programming and for the Limit of n-Stage Optimal Policies to Be Optimal. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **32**, 179–196.

[16] Schweitzer, P.J. (1968) Pertubation Theory and Finite Markov Chains. *J. Appl. Prob.* **5**, 401-413.

[17] Shahookar, K. and Mazumder, P. (1991) VLSI Cell Placement Techniques. *ACM Computing Surveys* **23**, 143–220.

[18] Tran, M.T. (1995) "Simulated Annealing: Konstruktion und Analyse einer Auswahl stochastischer Optimierungsverfahren mit Hilfe inhomogener Markoff-Ketten", Dissertation, Universität Hildesheim.

[19] Vaessens, R. J. M., E. H. L Aarts and J. K. Lenstra (1994) Job Shop Scheduling by Local Search. *Memorandum COSOR 94-05, Eindhoven University of Technology, February 1994.*