
Classification of Acceptance Criteria for the Simulated Annealing Algorithm

Author(s): Peter C. Schuur

Source: *Mathematics of Operations Research*, Vol. 22, No. 2 (May, 1997), pp. 266-275

Published by: INFORMS

Stable URL: <http://www.jstor.org/stable/3690264>

Accessed: 01/06/2010 15:54

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=informs>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



INFORMS is collaborating with JSTOR to digitize, preserve and extend access to *Mathematics of Operations Research*.

CLASSIFICATION OF ACCEPTANCE CRITERIA FOR THE SIMULATED ANNEALING ALGORITHM

PETER C. SCHUUR

We present a complete and explicit description of the class of all acceptance criteria for the simulated annealing algorithm that uniformly depend on the cost of the current and the candidate configuration and that lead to detailed balance when combined with a symmetric generation matrix. Here the cost dependence is called uniform if it is expressed by one and the same function for all cost functions. As an important special case we identify the subclass consisting of those acceptance criteria that depend uniformly on the difference in cost. Furthermore, we present a simple characterization of the Metropolis and the Barker criterion.

1. Introduction. In applications of the simulated annealing algorithm the Metropolis acceptance criterion is widely used. What could be the reason for its popularity? Could it be that the Metropolis criterion arises in a natural way being the only acceptance criterion fulfilling certain plausible requirements?

In this paper we give an affirmative answer. We establish that essentially the Metropolis criterion can be characterized as the unique acceptance criterion such that

- (i) For a fixed value of the control parameter the acceptance criterion depends only on the difference in cost.
- (ii) This cost difference dependence is uniform in the sense that it is expressed by one and the same function for all cost functions.
- (iii) All improvements in cost are accepted.
- (iv) For any symmetric generation matrix the detailed balance conditions hold.

More generally, the aim of this paper is to identify the class of all acceptance criteria that uniformly depend on the cost of the current and the candidate configuration and that lead to detailed balance when combined with a symmetric generation matrix. In particular we present a complete characterization of the subclass consisting of those acceptance criteria that depend uniformly on the difference in cost.

The paper runs as follows. In §2 we give a rough sketch of the simulated annealing algorithm. We introduce two well-known acceptance criteria, named after Metropolis and Barker. Furthermore, we discuss some requirements that a bona fide acceptance criterion should fulfill. In §3 we present a full and explicit description of the class of all uniformly and locally cost dependent acceptance criteria that yield detailed balance for any symmetric generation matrix. We give a simple expression for the associated stationary distribution. In §4 we identify the class of all uniformly cost difference dependent acceptance criteria that produce detailed balance for any symmetric generation matrix. We show where the exponential function familiar from the standard annealing algorithm comes from. In fact, we show that combining

Received December 19, 1994; revised May 2, 1996.

AMS 1991 subject classification. Primary: 90C29; Secondary: 60J10.

OR/MS Index 1978 subject classification. Primary: Programming/Theory; Secondary: Probability/Markov processes.

Key words. Simulated annealing, randomization techniques, combinatorial optimization, reversible Markov chains.

uniform cost difference dependence with detailed balance directly leads to the functional equation of the exponential function. Finally, we present a simple characterization of the Metropolis and the Barker criterion.

2. Outline of the annealing algorithm. Simulated annealing (Kirkpatrick, Gelatt and Vecchi 1983) is a versatile heuristic optimization technique based on the analogy between simulating the physical annealing process of solids and solving large-scale combinatorial optimization problems. For a detailed explanation of the method we refer to Aarts and Korst (1989).

Quite generally, a combinatorial optimization problem may be characterized by a finite set \mathcal{S} consisting of all system configurations and a cost function f assigning a real number to each configuration $i \in \mathcal{S}$. Here we choose the sign of the cost function in such a way that the lower the value the better the corresponding configuration.

The problem is to find an $i_{\text{opt}} \in \mathcal{S}$ satisfying

$$\forall i \in \mathcal{S}: f(i_{\text{opt}}) \leq f(i).$$

Any such solution i_{opt} is called a (global) optimum. Throughout we shall write \mathcal{S}_{opt} for the set of optimal solutions. To exclude trivialities, we shall assume that \mathcal{S} has at least three elements and that any cost function under consideration is such that $\mathcal{S} \neq \mathcal{S}_{\text{opt}}$.

In its usual form, simulated annealing can be summarized as follows. The algorithm starts off from an arbitrary initial configuration. In each iteration, by slightly perturbing the current configuration i , a new configuration j is generated. The difference in cost, given by

$$\Delta f = f(j) - f(i),$$

is compared with an acceptance criterion which tends to accept improvements but also admits, in a limited way, deteriorations in cost.

Initially, the acceptance criterion is taken such that deteriorations are accepted with a high probability. As the optimization process proceeds the acceptance criterion is modified such that the probability for accepting deteriorations decreases. At the end of the process this probability tends to zero. In this way the optimization process may be prevented from getting stuck in a local optimum. The process comes to a halt when during a prescribed number of iterations no further improvement in the best value found so far occurs.

2.1. A mathematical model. The simulated annealing process can be modelled mathematically in terms of a one-parameter family of homogeneous Markov chains (see Aarts and Korst 1989). The states of each Markov chain correspond with the configurations $i \in \mathcal{S}$. The transition probabilities depend on the value of the control parameter $c > 0$, the analog of the temperature in the physical annealing process. Thus, if c is kept constant, the corresponding Markov chain is homogeneous and its transition matrix $\mathbf{P}(c)$ can be defined as

$$(2.1) \quad P_{ij}(c) = \begin{cases} G_{ij}(c) A_{ij}(c) & \text{if } i \neq j, \\ 1 - \sum_{k \in \mathcal{S}, k \neq i} P_{ik}(c) & \text{if } i = j, \end{cases}$$

where $G_{ij}(c)$ denotes the *generation probability*, i.e., the probability of generating configuration j from configuration i , and $A_{ij}(c)$ denotes the *acceptance probability*, i.e., the probability of accepting configuration j , once it has been generated from i .

Let us emphasize that for fixed \mathcal{S} the acceptance matrix $A(c)$ depends on the cost function f . For brevity of writing, however, we shall suppress this f -dependence in our notation.

The generation matrix $G(c)$ is called a *standard generation matrix* if

$$(2.2) \quad (i) \quad G(c) \text{ is symmetric, i.e., } \forall c > 0, \forall i, j \in \mathcal{S}: G_{ij}(c) = G_{ji}(c),$$

$$(2.3) \quad (ii) \quad \text{the Markov chain associated with } G(c) \text{ is irreducible.}$$

Furthermore, the acceptance matrix $A(c)$ is called a *standard acceptance matrix* if

$$(2.4) \quad (i) \quad \forall i, j \in \mathcal{S}: A_{ij}(c) > 0,$$

$$(2.5) \quad (ii) \quad A_{ij}(c) < 1 \quad \text{for } i, j \in \mathcal{S} \text{ with } f(i) < f(j).$$

Reasoning as in Aarts and Korst (1989, pp. 39, 40) it is readily verified that for standard $G(c)$ and $A(c)$ the Markov chain associated with $P(c)$ is irreducible and aperiodic. Hence, there exists a unique stationary distribution $q(c)$.

For future reference let us recall an important concept from Markov theory (see Kelly 1979). The Markov chain associated with $P(c)$ is called *reversible* if

$$(2.6) \quad \forall i, j \in \mathcal{S}: q_i(c)P_{ij}(c) = q_j(c)P_{ji}(c).$$

The conditions (2.6) are known as the *detailed balance conditions*. Kolmogorov has shown (Kolmogorov 1936) that a Markov chain is reversible if and only if given a starting point $i \in \mathcal{S}$ any path in the state space \mathcal{S} which ultimately returns to i has the same probability of occurrence whether this path is traced in one direction or the other. In other words, (2.6) holds if and only if for any finite sequence of states i, i_1, \dots, i_n :

$$(2.7) \quad P_{ii_1}(c)P_{i_1i_2}(c) \cdots P_{i_{n-1}i_n}(c)P_{i_ni}(c) = P_{ii_n}(c)P_{i_ni_{n-1}}(c) \cdots P_{i_2i_1}(c)P_{i_1i}(c).$$

We shall refer to (2.7) as the *reversibility conditions*. They play a crucial role in our discussion.

2.2. Acceptance criteria. The standard and original (Kirkpatrick, Gelatt and Vecchi 1983) choice for the acceptance matrix $A(c)$ corresponds to the Metropolis criterion (Metropolis, Rosenbluth, Rosenbluth, Teller and Teller 1953) and is given by

$$(2.8) \quad A_{ij}(c) = \begin{cases} \exp\left(\frac{-\Delta f}{c}\right) & \text{if } \Delta f > 0, \\ 1 & \text{if } \Delta f \leq 0. \end{cases}$$

Another acceptance criterion—arising naturally in the context of Boltzmann machines (see Aarts and Korst (1989, pp. 133, 134))—is the Barker criterion (Barker 1965) given by

$$(2.9) \quad A_{ij}(c) = \frac{1}{1 + \exp\left(\frac{\Delta f}{c}\right)}.$$

It can be proven (Aarts and Korst 1989) that if $\mathbf{G}(c)$ is any standard generation matrix and $\mathbf{A}(c)$ is given either by the Metropolis criterion (2.8) or by the Barker criterion (2.9) then for $c > 0$ fixed, the Markov chain associated with $\mathbf{P}(c)$ has an equilibrium distribution $\mathbf{q}(c)$, whose components are given by

$$(2.10) \quad q_i(c) = \frac{1}{N_0(c)} \exp\left(\frac{-f(i)}{c}\right)$$

with

$$(2.11) \quad N_0(c) = \sum_{j \in \mathcal{S}} \exp\left(\frac{-f(j)}{c}\right).$$

Thus, in this case, after a sufficiently large number of transitions at a fixed value of c the simulated annealing algorithm will find a solution $i \in \mathcal{S}$ with a probability approximately equal to (2.10). From (2.10–2.11) one can derive

$$(2.12) \quad \lim_{c \downarrow 0} q_i(c) = \begin{cases} 0 & \text{if } i \notin \mathcal{S}_{\text{opt}}, \\ \frac{1}{|\mathcal{S}_{\text{opt}}|} & \text{if } i \in \mathcal{S}_{\text{opt}}. \end{cases}$$

This result is very important, since it guarantees *asymptotic* convergence of the annealing algorithm to the set of globally optimal solutions under the condition that equilibrium is obtained at each value of c .

On intuitive and practical grounds one may argue that, in addition to (2.4) and (2.5), an acceptance criterion $A_{ij}(c)$ should fulfill the following requirements:

(i) For $c > 0$ fixed $A_{ij}(c)$ is *locally cost dependent*, i.e., only depending on the cost $f(i)$ of the current configuration and the cost $f(j)$ of the candidate configuration.

(ii) This local cost dependence is *uniform*, i.e., expressed by one and the same function L for all cost functions f (see (3.2–3.3) below for a precise definition of uniform local cost dependence).

(iii) Given any standard generation matrix $\mathbf{G}(c)$ the transition matrix $\mathbf{P}(c)$ satisfies the reversibility conditions (2.7).

(iv) For any standard generation matrix $\mathbf{G}(c)$ the stationary distribution $\mathbf{q}(c)$ associated with $\mathbf{P}(c)$ has the asymptotic convergence property (2.12).

The desirability of (iii) stems from the fact that given a standard generation matrix $\mathbf{G}(c)$ one can derive a simple and explicit expression for the stationary distribution $\mathbf{q}(c)$ associated with $\mathbf{P}(c)$. This will be shown in Lemma 3.1 below. To see why (i) is plausible consider the globally cost dependent acceptance criterion

$$A_{ij}(c) = q_j(c)$$

with $q_j(c)$ as in (2.10–2.11). Clearly then (iii) is fulfilled, since (2.6) holds with (2.10–2.11) as the stationary distribution. However, calculation of $A_{ij}(c)$ is impracticable.

Of course, one may go one step further and require in (i) that $A_{ij}(c)$ depends on the difference in cost, say $A_{ij}(c) = D(c, \Delta f)$. Criteria of this type are extensively used. They are discussed in §4.

In the next section we identify the class of all uniformly and locally cost dependent acceptance criteria that satisfy (iii).

3. Uniform cost dependence. From now on we consider a fixed configuration space \mathcal{S} . We start with a helpful lemma.

LEMMA 3.1. *Let $A(c)$ be a standard acceptance matrix. Then the following assertions are equivalent:*

(i) *Given any standard generation matrix $G(c)$ the transition matrix $P(c)$ satisfies the reversibility conditions (2.7).*

(ii) *$A(c)$ satisfies the reversibility conditions (2.7).*

If (i) holds then for any standard generation matrix $G(c)$ the stationary distribution $q(c)$ associated with $P(c)$ is given by

$$(3.1) \quad \forall i \in \mathcal{S}: q_i(c) = \frac{1}{\sum_{j \in \mathcal{S}} A_{ij}(c)/A_{ji}(c)}.$$

PROOF. The equivalence of (i) and (ii) is obvious. Suppose (i) is true. Consider first a standard generation matrix $G(c)$ with positive nondiagonal entries. In view of the detailed balance conditions (2.6) the stationary distribution $q(c)$ associated with $P(c)$ satisfies

$$\begin{cases} q_i(c) A_{ij}(c) = q_j(c) A_{ji}(c), \\ \sum_{j \in \mathcal{S}} q_j(c) = 1. \end{cases}$$

Solving for $q_i(c)$ we immediately obtain (3.1). Consider next a general standard generation matrix $G(c)$. Clearly, the distribution $q(c)$ given by (3.1) satisfies the detailed balance conditions (2.6). Hence $q(c)$ is the unique stationary distribution associated with $P(c)$. \square

The acceptance matrix $A(c)$ will be called *uniformly cost dependent and generated by L* if there exists a function $L: (0, \infty) \times \mathbb{R} \times \mathbb{R} \rightarrow (0, 1]$ such that for fixed $c > 0$ one has

$$(3.2) \quad (i) \quad \text{For all cost functions } f: \mathcal{S} \rightarrow \mathbb{R}:$$

$$A_{ij}(c) = L(c, f(i), f(j)) \quad \forall i, j \in \mathcal{S}$$

$$(3.3) \quad (ii) \quad \forall x, y \in \mathbb{R}: x < y \Rightarrow L(c, x, y) < 1.$$

Condition (3.3) is included to ensure that $A(c)$ is standard.

The next theorem characterizes all uniformly cost dependent acceptance criteria that satisfy the reversibility conditions (2.7).

THEOREM 3.1. *Let the acceptance matrix $A(c)$ be uniformly cost dependent and generated by L . Then the following assertions are equivalent:*

(i) *For any standard generation matrix $G(c)$ and for any cost function $f: \mathcal{S} \rightarrow \mathbb{R}$ the transition matrix $P(c)$ associated with $G(c)$ and $A(c)$ satisfies the detailed balance conditions (2.6).*

(ii) *There exist functions $\phi: (0, \infty) \times \mathbb{R} \rightarrow (0, \infty)$ and $H: (0, \infty) \times \mathbb{R} \times \mathbb{R} \rightarrow (0, 1]$ such that for $c > 0$ and $x, y \in \mathbb{R}$:*

$$(3.4) \quad L(c, x, y) = H(c, x, y) \min\left(1, \frac{\phi(c, y)}{\phi(c, x)}\right),$$

$$(3.5) \quad H(c, x, y) = H(c, y, x).$$

PROOF. Suppose (i) holds. From Lemma 3.1 it is clear that $A(c)$ satisfies the reversibility conditions (2.7). Let us restrict ourselves to the case $n = 2$ in (2.7). In terms of L this means that for any cost function $f: \mathcal{S} \rightarrow \mathbb{R}$ one has for $c > 0$ and $i, j, k \in \mathcal{S}$,

$$\begin{aligned} & L(c, f(i), f(j))L(c, f(j), f(k))L(c, f(k), f(i)) \\ &= L(c, f(i), f(k))L(c, f(k), f(j))L(c, f(j), f(i)). \end{aligned}$$

Since by assumption $|\mathcal{S}| \geq 3$ this obviously yields

$$\begin{aligned} (3.6) \quad & \forall c > 0 \forall x, y, z \in \mathbb{R}: L(c, x, y)L(c, y, z)L(c, z, x) \\ &= L(c, x, z)L(c, z, y)L(c, y, x). \end{aligned}$$

Taking $z = 0$ we obtain

$$(3.7) \quad \forall c > 0 \forall x, y \in \mathbb{R}: \frac{L(c, 0, x)}{L(c, x, 0)}L(c, x, y) = \frac{L(c, 0, y)}{L(c, y, 0)}L(c, y, x).$$

Now, put

$$(3.8) \quad K(c, x, y) = \frac{L(c, 0, x)}{L(c, x, 0)}L(c, x, y),$$

$$(3.9) \quad \phi(c, x) = \frac{L(c, 0, x)}{L(c, x, 0)}.$$

Then one has

$$(3.10) \quad L(c, x, y) = \frac{K(c, x, y)}{\phi(c, x)},$$

$$(3.11) \quad K(c, x, y) = K(c, y, x),$$

where (3.11) is merely a transcription of (3.7). Since $0 < L(c, x, y) \leq 1$ we find from (3.10) that

$$(3.12) \quad K(c, x, y) \leq \phi(c, x).$$

From the symmetry relation (3.11) it follows that (3.12) is still valid if we replace $\phi(c, x)$ by $\phi(c, y)$. Hence, defining

$$(3.13) \quad H(c, x, y) = \frac{K(c, x, y)}{\min(\phi(c, x), \phi(c, y))}$$

we find that $0 < H(c, x, y) \leq 1$. Combining (3.10) and (3.11) with (3.13) we obtain the desired relations (3.4) and (3.5).

Conversely, suppose (ii) holds. Let f be an arbitrary cost function. Then it is straightforward to show that $A(c)$ satisfies the reversibility conditions (2.7), which in turn imply the detailed balance conditions (2.6) for any standard $G(c)$. \square

The preceding theorem provides a nice representation of the stationary distribution in terms of the function ϕ :

COROLLARY 3.1. *Let $A(c)$ be generated by L . Suppose condition (i) of Theorem 3.1 is fulfilled. Then, given a cost function f and a standard generation matrix $G(c)$, the stationary distribution $q(c)$ associated with the transition matrix $P(c)$ is given by*

$$(3.14) \quad \forall i \in \mathcal{S}: q_i(c) = \frac{\phi(c, f(i))}{\sum_{j \in \mathcal{S}} \phi(c, f(j))}$$

with ϕ as in (3.4).

For this stationary distribution the asymptotic convergence property (2.12) is valid for all cost functions $f: \mathcal{S} \rightarrow \mathbb{R}$ if and only if

$$(3.15) \quad \forall x, y \in \mathbb{R}: x < y \Rightarrow \lim_{c \downarrow 0} \frac{\phi(c, y)}{\phi(c, x)} = 0.$$

PROOF. Using (3.10) in the representation (3.1) and exploiting the symmetry of K we easily arrive at (3.14).

To prove the second assertion note that (2.12) is obviously equivalent with

$$(3.16) \quad \forall i \in \mathcal{S}_{\text{opt}}: \lim_{c \downarrow 0} \frac{1}{q_i(c)} = |\mathcal{S}_{\text{opt}}|.$$

Now (3.14) tells us

$$\frac{1}{q_i(c)} = |\mathcal{S}_{\text{opt}}| + \sum_{j \in \mathcal{S} \setminus \mathcal{S}_{\text{opt}}} \frac{\phi(c, f(j))}{\phi(c, f(i))}.$$

Hence for (3.16) to hold for all cost functions it is necessary and sufficient that

$$\forall f: \mathcal{S} \rightarrow \mathbb{R}: i \in \mathcal{S}_{\text{opt}}, j \in \mathcal{S} \setminus \mathcal{S}_{\text{opt}} \Rightarrow \lim_{c \downarrow 0} \frac{\phi(c, f(j))}{\phi(c, f(i))} = 0,$$

which is evidently equivalent with (3.15). \square

4. Uniform cost difference dependence. In this section we examine an important special case of the uniform cost dependence discussed in the previous section.

The acceptance matrix $A(c)$ will be called *uniformly cost difference dependent and generated by D* if there exists a function $D: (0, \infty) \times \mathbb{R} \rightarrow (0, 1]$ such that for fixed $c > 0$ one has

$$(4.1) \quad \text{(i) For all cost functions } f: \mathcal{S} \rightarrow \mathbb{R}:$$

$$A_{ij}(c) = D(c, f(j) - f(i)) \quad \forall i, j \in \mathcal{S},$$

$$(4.2) \quad \text{(ii) } \forall x > 0: D(c, x) < 1,$$

$$(4.3) \quad \text{(iii) } \exists x_0 > 0: D(c, x_0) < D(c, -x_0).$$

$$(4.4) \quad \text{(iv) The function } x \mapsto D(c, x)/D(c, -x) \text{ is Lebesgue measurable on } \mathbb{R}.$$

When comparing this definition with that of uniform cost dependence one may notice that we have added the conditions (iii) and (iv). The first of these requires that there is at least one amount of improvement in cost that is more likely to be accepted than

precisely the same amount of deterioration in cost. This plausible requirement is added to facilitate the discussion somewhat. For instance it excludes such impracticable choices as $D(c, x) = p$ with p some constant. Though a property of type (iv) is essential in the proof of the next theorem, it is such a weak regularity property that it has no practical consequences at all. Clearly, condition (ii) corresponds to (3.3) and it has the same function, namely to guarantee that $A(c)$ is standard.

The following theorem is crucial. It tells us where the exponential function occurring in the standard annealing algorithm comes from.

THEOREM 4.1. *Let the acceptance matrix $A(c)$ be uniformly cost difference dependent and generated by D . Then the following assertions are equivalent:*

(i) *For any standard generation matrix $G(c)$ and for any cost function $f: \mathcal{S} \rightarrow \mathbb{R}$ the transition matrix $P(c)$ associated with $G(c)$ and $A(c)$ satisfies the detailed balance conditions (2.6).*

(ii) *There exist functions $\gamma: (0, \infty) \rightarrow (0, \infty)$ and $E: (0, \infty) \times \mathbb{R} \rightarrow (0, 1]$ such that for $c > 0$ and $x \in \mathbb{R}$:*

$$(4.5) \quad D(c, x) = E(c, x) \min \left(1, \exp \left(\frac{-x}{\gamma(c)} \right) \right),$$

$$(4.6) \quad E(c, x) = E(c, -x).$$

PROOF. Obviously, $A(c)$ is uniformly cost dependent and generated by $L(c, x, y) = D(c, y - x)$. Suppose (i) holds. Then the proof of Theorem 3.1 gives us (3.6), which we can rewrite in terms of D as

$$(4.7) \quad \forall c > 0 \forall x, y, z \in \mathbb{R}: \quad D(c, y - x) D(c, z - y) D(c, x - z) \\ = D(c, z - x) D(c, y - z) D(c, x - y).$$

Clearly, the function ϕ introduced in (3.9) can be expressed in terms of D as

$$(4.8) \quad \phi(c, x) = \frac{D(c, x)}{D(c, -x)}.$$

Combining (4.7) and (4.8) we obtain

$$\forall c > 0 \forall x, y, z \in \mathbb{R}: \phi(c, x - y) = \phi(c, x - z) \phi(c, z - y)$$

or equivalently

$$\forall c > 0 \forall x, y \in \mathbb{R}: \phi(c, x + y) = \phi(c, x) \phi(c, y).$$

Thus for fixed $c > 0$ the function $x \mapsto \phi(c, x)$ satisfies the functional equation of the exponential function. Since by assumption $\phi(c, x)$ is Lebesgue measurable in x on \mathbb{R} we may use the classical result from Ostrowski (1929) to conclude that there exists a function $\alpha: (0, \infty) \rightarrow \mathbb{R}$ such that

$$(4.9) \quad \forall c > 0 \forall x \in \mathbb{R}: \phi(c, x) = \exp(\alpha(c)x).$$

From (4.3) and (4.8) it is clear that

$$\forall c > 0 \exists x_0 > 0: \phi(c, x_0) < 1.$$

Hence

$$(4.10) \quad \forall c > 0: \gamma(c) \equiv -\frac{1}{\alpha(c)} > 0.$$

Using (4.9) and (4.10) it is readily verified that the function H introduced in (3.13) can be written as

$$H(c, x, y) = E(c, y - x)$$

with the function E given by

$$E(c, x) = \frac{D(c, x)}{\min\left(1, \exp\left(\frac{-x}{\gamma(c)}\right)\right)}.$$

From H the function E inherits the property $0 < E(c, x) \leq 1$. Since (4.6) is a direct consequence of (3.5), the proof of assertion (ii) is completed.

Conversely, suppose (ii) holds. Then the functions $L(c, x, y) = D(c, y - x)$, $\phi(c, x) = \exp(-x/\gamma(c))$ and $H(c, x, y) = E(c, y - x)$ satisfy condition (ii) of Theorem 3.1. Hence assertion (i), which is the same for that and this theorem, holds. \square

It is a remarkable fact that when combined with a standard generation matrix all of the acceptance criteria featuring in Theorem 4.1 have the Boltzmann distribution in common as stationary distribution:

COROLLARY 4.1. *Let $\mathbf{A}(c)$ be uniformly cost difference dependent. Suppose condition (i) of Theorem 4.1 holds. Then, given a cost function f and a standard generation matrix $\mathbf{G}(c)$, the stationary distribution $\mathbf{q}(c)$ associated with the transition matrix $\mathbf{P}(c)$ is given by*

$$(4.11) \quad \forall i \in \mathcal{S}: q_i(c) = \frac{\exp(-f(i)/\gamma(c))}{\sum_{j \in \mathcal{S}} \exp(-f(j)/\gamma(c))}$$

with γ as in (4.5).

Furthermore, the stationary distribution (4.11) has for all cost functions $f: \mathcal{S} \rightarrow \mathbb{R}$ the asymptotic convergence property (2.12) if and only if

$$(4.12) \quad \lim_{c \downarrow 0} \gamma(c) = 0.$$

PROOF. Inserting the explicit representation $\phi(c, x) = \exp(-x/\gamma(c))$ in (3.14) and (3.15) one easily finds the desired representation (4.11) as well as the equivalence of (2.12) (for all cost functions) and (4.12). \square

As a direct consequence of Theorem 4.1 we obtain the following characterization of the Metropolis and the Barker criterion.

COROLLARY 4.2. *Let $\mathbf{A}(c)$ be uniformly cost difference dependent and generated by D . Assume that the function $c \mapsto D(c, 1)$ is a strictly increasing mapping of $(0, \infty)$ onto $(0, 1)$.*

Assume furthermore that for any standard generation matrix $\mathbf{G}(c)$ and for any cost function $f: \mathcal{S} \rightarrow \mathbb{R}$ the transition matrix $\mathbf{P}(c)$ associated with $\mathbf{G}(c)$ and $\mathbf{A}(c)$ satisfies the detailed balance conditions (2.6).

Then one has the following:

(a) If

$$\forall c > 0 \forall x \leq 0: D(c, x) = 1,$$

then there exists an increasing bijective function $\tilde{c}: (0, \infty) \rightarrow (0, \infty)$ such that $A(c)$ after a reparametrization by \tilde{c} is given by the Metropolis acceptance criterion:

$$A_{ij}(\tilde{c}(c)) = \begin{cases} \exp\left(\frac{f(i) - f(j)}{c}\right) & \text{if } f(i) < f(j), \\ 1 & \text{if } f(i) \geq f(j). \end{cases}$$

(b) If

$$\forall c > 0 \forall x \in \mathbb{R}: D(c, x) + D(c, -x) = 1,$$

then there exists an increasing bijective function $\tilde{c}: (0, \infty) \rightarrow (0, \infty)$ such that $A(c)$ after a reparametrization by \tilde{c} is given by the Barker acceptance criterion:

$$A_{ij}(\tilde{c}(c)) = \frac{1}{1 + \exp\left(\frac{f(j) - f(i)}{c}\right)}.$$

PROOF. In both cases Theorem 4.1 tells us that the function D can be represented by (4.5–4.6). In case (a) we immediately obtain that $E \equiv 1$. In case (b), it follows that $E(c, x) = D(c, -|x|)$ with

$$D(c, x) = \frac{1}{1 + \exp\left(\frac{x}{\gamma(c)}\right)}.$$

In both cases, the assumption about $D(c, 1)$ implies that γ is a strictly increasing mapping of $(0, \infty)$ onto itself. Now take $\tilde{c} = \gamma^{-1}$ and the result follows.

References

- Aarts, E. H. L., J. H. M. Korst (1989). *Simulated Annealing and Boltzmann Machines*. Wiley, Chichester.
- Barker, A. A. (1965). Monte Carlo calculations of the radial distribution functions for a proton-electron plasma. *Austral. J. Phys.* **18** 119–133.
- Kelly, F. P. (1979). *Reversibility and Stochastic Networks*. Wiley, New York.
- Kirkpatrick, S., C. D. Gelatt Jr., M. P. Vecchi (1983). Optimization by simulated annealing. *Science* **220** 671–680.
- Kolmogorov, A. (1936). Zur Theorie der Markoffschen Ketten. *Math. Ann.* **112** 155–160.
- Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.
- Ostrowski, A. (1929). Mathematische Miszellen XIV: Ueber die Funktionalgleichung der Exponentialfunktion und verwandte Funktionalgleichungen. *Jahresber. Deutsch. Math.-Verein* **38** 54–62.

P. C. Schuur: School of Management Studies, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands