Classification of acceptance criteria for the simulated annealing algorithm

Schuur, P.C.

Published: 01/01/1989

Document Version
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 24. Nov. 2018
Classification of acceptance criteria for the simulated annealing algorithm

P.C. Schuur

Eindhoven, November 1989
The Netherlands
Classification of acceptance criteria for the simulated annealing algorithm

Peter C. Schuur

Eindhoven University of Technology
Department of Mathematics and Computing Science
P.O. Box 513, 5600 MB Eindhoven, the Netherlands

Abstract

We present a complete and explicit description of the class of all acceptance criteria for the simulated annealing algorithm that are uniformly and locally cost dependent and that lead to reversibility when combined with a symmetric generation matrix. In particular 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. As an interesting special case 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 section 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 section 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. Conversely, given such a stationary distribution we construct a large family of Markov chains having this stationary distribution in common. In section 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 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 [6] 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 as well as the origin of our notation we refer to [1].

Quite generally, a combinatorial optimization problem may be characterized by a finite set \( S \) consisting of all system configurations and a cost function \( f \) assigning a real number to each configuration \( i \in S \). 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 \( i_{\text{opt}} \in S \) satisfying

\[
\forall i \in S: \quad f(i_{\text{opt}}) \leq f(i).
\]

Any such solution \( i_{\text{opt}} \) is called a (global) optimum. Throughout we shall write \( f_{\text{opt}} = f(i_{\text{opt}}) \) for the optimal cost and \( S_{\text{opt}} \) for the set of optimal solutions. To exclude trivialities, we shall assume that \( S \) has at least 3 elements and that any cost function under consideration is such that \( S \neq S_{\text{opt}} \).

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 optimum 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 [1]). The states of each Markov chain correspond with the configurations \( i \in S \). 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 \( P(c) \) can be defined as

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

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 \( S \) 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

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

(ii) The Markov chain associated with \( G(c) \) is irreducible. (2.3)

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

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

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

Reasoning as in [1], 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 [5]).

The Markov chain associated with \( P(c) \) is called reversible if

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

The conditions (2.6) are known as the detailed balance conditions. Kolmogorov has shown [7] that a Markov chain is reversible if and only if given a starting point \( i \in S \) any path in the state space \( S \) 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 \( i, i_1, \ldots, i_n \):

\[
P_{i_1}(c)P_{i_1i_2}(c)\ldots P_{i_{n-1}i_n}(c)P_{i_n}(c) = P_{i_n}(c)P_{i_{n-1}i_n}(c)\ldots P_{i_1i_2}(c)P_{i_1}(c).
\]

In section 3 we shall present a slight modification of Kolmogorov's result that will turn out to be quite useful in our discussion.
2.2 Acceptance criteria

The standard and original [6] choice for the acceptance matrix $A(c)$ corresponds to the Metropolis criterion [8] and is given by

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

Another acceptance criterion - arising naturally in the context of Boltzmann machines (see [1], pp. 133, 134) - is the Barker criterion [3] given by

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

It can be proven [1] that if $G(c)$ is any standard generation matrix and $A(c)$ 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 $P(c)$ has an equilibrium distribution $q(c)$, whose components are given by

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

with

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

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 $i \in S$ with a probability approximately equal to (2.10). From (2.10-11) one can derive

$$\lim_{c \to 0} q_i(c) = \begin{cases} 0 & \text{if } i \notin S_{\text{opt}} \\ \frac{1}{|S_{\text{opt}}|} & \text{if } i \in S_{\text{opt}}. \end{cases} \quad (2.12)$$

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 expressed by one and the same function $L$ for all cost functions $f$.

(iii) Given a standard generation matrix $G(c)$ an explicit expression for the stationary distribution $q(c)$ associated with $P(c)$ can be easily obtained.

(iv) The stationary distribution $q(c)$ has the asymptotic convergence property (2.12).
In lemma 3.1 below it will be shown that a sufficient condition for (iii) is:

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

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 (2.10-11). Clearly then (iii) is fulfilled with (2.10-11) as 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 section 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 $S$.

We start with a helpful and general lemma, which is a minor modification of Kolmogorov's result quoted in subsection 2.1.

**Lemma 3.1** Let $G(c)$ and $A(c)$ be standard. Let the unique stationary distribution associated with $P(c)$ be denoted by $q(c)$. Then the following assertions are equivalent.

(i) \[ \forall i, j \in S : \quad q_i(c)A_{ij}(c) = q_j(c)A_{ji}(c) \]  
(ii) \[ \forall i, j, k \in S : \quad A_{ij}(c)A_{jk}(c)A_{ki}(c) = A_{ik}(c)A_{kj}(c)A_{ji}(c). \]

If the equivalent conditions (3.1) and (3.2) hold then the stationary distribution is given by

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

**Proof** Suppose (3.1) holds. Solving $q_j(c)$ from

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

we immediately obtain (3.3). Applying (3.1) three times and multiplying one gets

$$\forall i, j, k \in S : \quad q_i(c)A_{ij}(c)q_j(c)A_{jk}(c)q_k(c)A_{ki}(c) = q_j(c)A_{ji}(c)q_k(c)A_{kj}(c)q_i(c)A_{ik}(c).$$

Cancelling the positive equilibrium probabilities gives us (3.2).

Conversely, suppose (3.2) holds. Then

$$\forall i, j, k \in S : \quad A_{ij}(c)\frac{A_{jk}(c)}{A_{ki}(c)} = A_{ji}(c)\frac{A_{ik}(c)}{A_{kj}(c)}.$$
Hence
\[ \forall i, j \in S : \quad A_{ij}(c) \sum_{k \in S} \frac{A_{jk}(c)}{A_{ki}(c)} = A_{ji}(c) \sum_{k \in S} \frac{A_{ik}(c)}{A_{kj}(c)}. \]

Let
\[ \tilde{q}_i(c) = \frac{1}{\sum_{k \in S} A_{ik}(c)/A_{ki}(c)}. \]

Then the \( \tilde{q}_i(c) \) satisfy
\[ \forall i, j \in S : \quad \tilde{q}_i(c) A_{ij}(c) = \tilde{q}_j(c) A_{ji}(c). \]

Consequently
\[ \sum_{i \in S} \tilde{q}_i(c) = \tilde{q}_j(c) \sum_{i \in S} A_{ji}(c) = \frac{1}{\sum_{k \in S} A_{jk}(c)/A_{kj}(c)} \sum_{i \in S} A_{ij}(c) = 1. \]

Hence \( \tilde{q}(c) \) is the unique distribution satisfying the detailed balance equation (2.6). Therefore \( \tilde{q}(c) \) coincides with \( q(c) \) and (3.1) is proven. \( \square \)

The acceptance matrix \( A(c) \) will be called uniformly cost dependent 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
\begin{align*}
(i) & \quad \text{for all cost functions } f : S \rightarrow \mathbb{R} : \\
& \quad A_{ij}(c) = L(c, f(i), f(j)) \quad \forall i, j \in S \quad (3.4) \\
(ii) & \quad \forall x, y \in \mathbb{R} : \quad x < y \Rightarrow L(c, x, y) < 1. \quad (3.5)
\end{align*}

In this case we shall say that \( A(c) \) is generated by \( L \).

Condition (3.5) is enclosed to ensure that \( A(c) \) is standard.

The next theorem characterizes all uniformly cost dependent acceptance criteria that satisfy (3.2).

**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 \( f : S \rightarrow \mathbb{R} \) the transition matrix \( P(c) \) associated with \( G(c) \) and \( A(c) \) satisfies the detailed balance equation (2.6).

(ii) There exist functions \( \phi, \psi : (0, \infty) \times \mathbb{R} \rightarrow (0, \infty) \) and \( H : (0, \infty) \times \mathbb{R} \times \mathbb{R} \rightarrow (0, 1] \) such that for \( c > 0 \) and \( x, y \in \mathbb{R} \):
\begin{align*}
L(c, x, y) &= H(c, x, y) \min \left( 1, \frac{\phi(c, y)}{\phi(c, x)} \right), \quad (3.6) \\
H(c, x, y) &= H(c, y, x). \quad (3.7)
\end{align*}

Let these equivalent conditions be 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
\[ \forall i \in S : \quad q_i(c) = \frac{\phi(c, f(i))}{\sum_{j \in S} \phi(c, f(j))}. \quad (3.8) \]
For this stationary distribution the asymptotic convergence property

\[
\lim_{c \to 0} q(c) = \begin{cases} 
0 & \text{if } i \notin S_{opt} \\
\frac{1}{|S_{opt}|} & \text{if } i \in S_{opt}
\end{cases}
\]  

(3.9)

is valid for all cost functions \( f : S \to \mathbb{R} \) if and only if

\[
\forall x, y \in \mathbb{R} : \quad x < y \Rightarrow \lim_{c \to 0} \frac{\phi(c,y)}{\phi(c,x)} = 0.
\]

(3.10)

**Proof** Suppose (i) holds. Since, given a cost function \( f \) and any symmetric \( G(c) \), the transition matrix \( P(c) \) satisfies the detailed balance equation (2.6), it is clear that (3.1) or equivalently (3.2) holds. In terms of \( L \) this means that for any cost function \( f : \mathbb{S} \to \mathbb{R} \) one has for \( c > 0 \) and \( i,j,k \in \mathbb{S} \)

\[
L(c, f(i), f(j)) = L(c, f(i), f(k))L(c, f(k), f(j))L(c, f(j), f(i)).
\]

Since by assumption \( |\mathbb{S}| \geq 3 \) this obviously yields

\[
\forall c > 0 \forall x, y, z \in \mathbb{R} : \quad L(c, x, y)L(c, y, z)L(c, z, x) = L(c, x, z)L(c, z, y)L(c, y, x).
\]

(3.11)

Taking \( z = 0 \) we obtain

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

(3.12)

Now, put

\[
K(c, x, y) = \frac{L(c, 0, x)}{L(c, x, 0)}L(c, x, y),
\]

(3.13)

\[
\phi(c, x) = \frac{L(c, 0, x)}{L(c, x, 0)}.
\]

(3.14)

Then one has

\[
L(c, x, y) = \frac{K(c, x, y)}{\phi(c, x)},
\]

(3.15)

\[
K(c, x, y) = K(c, y, x),
\]

(3.16)

where (3.16) is merely a transcription of (3.12). Since \( 0 < L(c, x, y) \leq 1 \) we find from (3.15) that

\[
K(c, x, y) \leq \phi(c, x).
\]

(3.17)

From the symmetry relation (3.16) it follows that (3.17) is still valid if we replace \( \phi(c, x) \) by \( \phi(c, y) \). Hence, defining

\[
H(c, x, y) = \frac{K(c, x, y)}{\min(\phi(c, x), \phi(c, y))}
\]

(3.18)

we find that \( 0 < H(c, x, y) \leq 1 \). Combining (3.15) and (3.16) with (3.18) we obtain the desired relations (3.6) and (3.7).
Conversely, suppose (ii) holds. Let $f$ be an arbitrary cost function. Then it is straightforward to show that $A(c)$ satisfies (3.2), which in turn implies the detailed balance equation (2.6) for any standard $G(c)$.

Finally, let us assume that the conditions (i) and (ii) hold. Using (3.15) in the representation (3.3) and exploiting the symmetry of $K$ we easily arrive at (3.8).

To prove the last assertion note that (3.9) is obviously equivalent with

$$\forall i \in S_{opt} : \lim_{c \to 0} \frac{1}{q_i(c)} = |S_{opt}|. \quad (3.19)$$

Now (3.8) tells us

$$\frac{1}{q_i(c)} = |S_{opt}| + \sum_{j \in S \setminus S_{opt}} \frac{\phi(c, f(j))}{\phi(c, f(i))}. \quad (3.20)$$

Hence for (3.19) to hold for all cost functions it is necessary and sufficient that

$$\forall f : S \to \mathbb{R} : \ i \in S_{opt}, j \in S \setminus S_{opt} \Rightarrow \lim_{c \to 0} \frac{\phi(c, f(j))}{\phi(c, f(i))} = 0,$$

which is evidently equivalent with (3.10). \(\square\)

Note that functions $\phi$ satisfying (3.10) are easily found. To illustrate this let $U : \mathbb{R} \to (0, \infty)$ be strictly increasing. Let $R : (0, \infty) \to (0, \infty)$ belong to the class of rapidly varying functions [2] of index $-\infty$, i.e.:

$$\forall \lambda > 1 : \lim_{x \to \infty} \frac{R(\lambda x)}{R(x)} = 0.$$

Then the function

$$\phi(c, x) = R \left( \frac{U(x)}{c} \right)$$

clearly satisfies (3.10).

Now, consider any $\phi : (0, \infty) \times \mathbb{R} \to (0, \infty)$ such that (3.10) holds. Then $q(c)$ given by (3.8) is a bona fide stationary distribution, since it satisfies (3.9). To simplify the reasoning let us make the plausible assumption that for $c > 0$ fixed the function $x \mapsto \phi(c, x)$ is strictly decreasing on $\mathbb{R}$, which in view of (3.8) is equivalent to assuming that the higher the cost value of a particular configuration the lower its equilibrium probability.

The preceding theorem is helpful in constructing a suitable Markov chain having $q(c)$ as its stationary distribution. To this end we keep the standard generation matrix $G(c)$ fixed and modify $A(c)$ in accordance with (3.4-6).

Obviously, all of the acceptance probabilities obtained in this way are majorated by

$$A_{ij}(c) = \begin{cases} \frac{\phi(c, f(j))}{\phi(c, f(i))} & \text{if } f(j) > f(i) \\ 1 & \text{if } f(j) \leq f(i) \end{cases}, \quad (3.20)$$

which corresponds to setting $H(c, x, y)$ equal to 1 in (3.6). Criterion (3.20) can be regarded as a generalization of the Metropolis acceptance criterion to arbitrary stationary distributions of type (3.8).
More generally, motivated by (3.6) we might look for an acceptance criterion such that \( L(c, x, y) \) can be written as a function of \( \Phi(c, y)/\Phi(c, x) \), say

\[
L(c, x, y) = F \left( \frac{\Phi(c, y)}{\Phi(c, x)} \right).
\]  

(3.21)

To satisfy (3.6-7) it is sufficient (and also necessary if we stipulate uniformity in \( \Phi \)) that \( F : (0, \infty) \to (0, 1] \) has the property

\[
\forall x > 0 : \quad F(x) = \frac{1}{x} F \left( \frac{1}{x} \right).
\]  

(3.22)

It is not hard to see that any function \( F : (0, \infty) \to (0, 1] \) satisfying (3.22) has the form

\[
F(x) = \min \left( 1, \frac{1}{x} \right) B \left( \min \left( x, \frac{1}{x} \right) \right)
\]  

(3.23)

with \( B : (0, 1] \to (0, 1] \) arbitrary. Consequently, by choosing \( H \) in (3.6) as

\[
H(c, x, y) = B \left( \min \left( \frac{\Phi(c, y)}{\Phi(c, x)}, \frac{\Phi(c, x)}{\Phi(c, y)} \right) \right),
\]

we obtain a large class of acceptance matrices, all satisfying the requirements (i) and (ii) of the preceding theorem and - when combined with a standard generation matrix - having \( q(c) \) in common as stationary distribution. Note that this class contains also a generalization of the Barker criterion (2.9), namely:

\[
A_{ij}(c) = \frac{\phi(c, f(j))}{\phi(c, f(i)) + \phi(c, f(j))}.
\]  

(3.24)

This criterion results from taking \( B(x) = 1/(1 + x) \) in (3.23). For the sake of completeness, let us remark that the class derived above can be shown to be a subclass of the one mentioned in [4], p. 100.

4 Uniform cost difference dependence

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

\[
(i) \quad \text{for all cost functions } f : S \to \mathbb{R} : \quad A_{ij}(c) = D(c, f(j) - f(i)) \quad \forall i, j \in S
\]  

(4.1)

\[
(ii) \quad \forall x > 0 : \quad D(c, x) < 1
\]  

(4.2)

\[
(iii) \quad \exists x_0 > 0 : \quad D(c, x_0) < D(c, -x_0)
\]  

(4.3)

\[
(iv) \quad \text{the function } x \mapsto D(c, x)/D(c, -x) \text{ is Lebesgue measurable on } \mathbb{R}.
\]  

(4.4)

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 (3.5) 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. Let the dependence be expressed by the function \( D : (0, \infty) \times \mathbb{R} \rightarrow (0, 1] \) satisfying (4.1) to (4.4). Then the following assertions are equivalent:

(i) For any standard generation matrix \( G(c) \) and for any cost function \( f : S \rightarrow \mathbb{R} \) the transition matrix \( P(c) \) associated with \( G(c) \) and \( A(c) \) satisfies the detailed balance equation (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} : 
\[
D(c, x) = E(c, x) \min \left( 1, \exp \left( \frac{-x}{\gamma(c)} \right) \right), \\
E(c, x) = E(c, -x).
\]

Let these equivalent conditions be fulfilled. Then, given a cost function \( f \) and a standard generation matrix \( G(c) \), the stationary distribution associated with the transition matrix \( P(c) \) is given by

\[
\forall i \in S : \quad q_i(c) = \frac{\exp \left( \frac{-f(i)}{\gamma(c)} \right)}{\sum_{j \in S} \exp \left( \frac{-f(j)}{\gamma(c)} \right)}.
\]

Furthermore, the stationary distribution (4.7) has for all cost functions \( f : S \rightarrow \mathbb{R} \) the asymptotic convergence property

\[
\lim_{c \to 0} q_i(c) = \begin{cases} 
0 & \text{if } i \notin S_{opt} \\
\frac{1}{|S_{opt}|} & \text{if } i \in S_{opt}
\end{cases}
\]

if and only if

\[
\lim_{c \to 0} \gamma(c) = 0.
\]

**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 (3.11), which we can rewrite in terms of \( D \) as

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

10
Clearly, the function $\phi$ introduced in (3.14) can be expressed in terms of $D$ as

$$\phi(c,x) = \frac{D(c,x)}{D(c,-x)}. \tag{4.11}$$

Combining (4.10) and (4.11) we obtain

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

or equivalently

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

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 [9] to conclude that there exists a function $\alpha : (0, \infty) \to \mathbb{R}$ such that

$$\forall c > 0 \ \forall x \in \mathbb{R} : \quad \phi(c, x) = \exp(\alpha(c)x). \tag{4.12}$$

From (4.3) and (4.11) it is clear that

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

Hence

$$\forall c > 0 : \quad \gamma(c) \equiv -\frac{1}{\alpha(c)} > 0. \tag{4.13}$$

Using (4.12) and (4.13) it is readily verified that the function $H$ introduced in (3.18) 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 (3.7), 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.

Finally, inserting the explicit representation $\phi(c, x) = \exp(-x/\gamma(c))$ in (3.8) and (3.10) one easily finds the desired representation (4.7) as well as the equivalence of (4.8) (for all cost functions) and (4.9).

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

**Corollary 4.1** Let $A(c)$ be an acceptance matrix. Assume there exists a function $D : (0, \infty) \times \mathbb{R} \to (0, 1]$ such that for fixed $c > 0$ one has

(i) for all cost functions $f : \quad A_{ij}(c) = D(c, f(j) - f(i)) \quad \forall i, j \in S$

(ii) the function $x \mapsto D(c, x)/D(c, -x)$ is Lebesgue measurable on $\mathbb{R}$.  

11
Assume furthermore that the function $c \mapsto D(c, 1)$ is a strictly increasing mapping of $(0, \infty)$ onto $(0, 1)$.

Finally, assume that for any standard generation matrix $G(c)$ and for any cost function $f : S \rightarrow \mathbb{R}$ the transition matrix $P(c)$ associated with $G(c)$ and $A(c)$ satisfies the detailed balance equation (2.6).

Then one has the following:

(a) If for $c > 0$

\begin{align*}
(a1) \quad \forall x \leq 0 : \quad D(c, x) &= 1, \\
(a2) \quad \forall x > 0 : \quad D(c, x) < 1,
\end{align*}

then there exists an increasing bijective function $\tilde{c} : (0, \infty) \rightarrow (0, \infty)$ such that $A(c)$ after a reparametrisation 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 for $c > 0$

\begin{align*}
(b1) \quad \forall x \in \mathbb{R} : \quad D(c, x) + D(c, -x) &= 1, \\
(b2) \quad \exists x_0 > 0 : \quad D(c, x_0) < D(c, -x_0),
\end{align*}

then there exists an increasing bijective function $\tilde{c} : (0, \infty) \rightarrow (0, \infty)$ such that $A(c)$ after a reparametrisation 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 (4.5-6). In case (a) we obtain from (a1) that $E \equiv 1$. In case (b), note that (b1) and (4.5-6) imply 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 $\gamma$ is a strictly increasing mapping of $(0, \infty)$ onto itself. Now take $\tilde{c} = \gamma^{-1}$ and the result follows.

**References**


**List of COSOR-memoranda - 1989**

<table>
<thead>
<tr>
<th>Number</th>
<th>Month</th>
<th>Author</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>M 89-01</td>
<td>January</td>
<td>D.A. Overdijk</td>
<td>Conjugate profiles on mating gear teeth</td>
</tr>
<tr>
<td>M 89-02</td>
<td>January</td>
<td>A.H.W. Geerts</td>
<td>A priori results in linear quadratic optimal control theory</td>
</tr>
<tr>
<td>M 89-03</td>
<td>February</td>
<td>A.A. Stoorvogel, H.L. Trentelman</td>
<td>The quadratic matrix inequality in singular $H_\infty$ control with state feedback</td>
</tr>
<tr>
<td>M 89-04</td>
<td>February</td>
<td>E. Willekens, N. Veraverbeke</td>
<td>Estimation of convolution tail behaviour</td>
</tr>
<tr>
<td>M 89-05</td>
<td>March</td>
<td>H.L. Trentelman</td>
<td>The totally singular linear quadratic problem with indefinite cost</td>
</tr>
<tr>
<td>M 89-06</td>
<td>April</td>
<td>B.G. Hansen</td>
<td>Self-decomposable distributions and branching processes</td>
</tr>
<tr>
<td>M 89-07</td>
<td>April</td>
<td>B.G. Hansen</td>
<td>Note on Urbanik's class $L_\infty$</td>
</tr>
<tr>
<td>M 89-08</td>
<td>April</td>
<td>B.G. Hansen</td>
<td>Reversed self-decomposability</td>
</tr>
<tr>
<td>M 89-09</td>
<td>April</td>
<td>A.A. Stoorvogel</td>
<td>The singular zero-sum differential game with stability using $H_\infty$ control theory</td>
</tr>
<tr>
<td>M 89-10</td>
<td>April</td>
<td>L.J.G. Langenhoff, W.H.M. Zijm</td>
<td>An analytical theory of multi-echelon production/distribution systems</td>
</tr>
<tr>
<td>M 89-11</td>
<td>April</td>
<td>A.H.W. Geerts</td>
<td>The Algebraic Riccati Equation and Singular Optimal Control</td>
</tr>
<tr>
<td>Number</td>
<td>Month</td>
<td>Author</td>
<td>Title</td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>-----------------------------------------</td>
<td>-----------------------------------------------------------------------</td>
</tr>
<tr>
<td>M 89-12</td>
<td>May</td>
<td>D.A. Overdijk</td>
<td>De geometrie van de kroonwieloverbrenging</td>
</tr>
<tr>
<td>M 89-13</td>
<td>May</td>
<td>I.J.B.F. Adan J. Wessels W.H.M. Zijm</td>
<td>Analysis of the shortest queue problem</td>
</tr>
<tr>
<td>M 89-14</td>
<td>June</td>
<td>A.A. Stoorvogel</td>
<td>The singular $H_\infty$ control problem with dynamic measurement feedback</td>
</tr>
<tr>
<td>M 89-15</td>
<td>June</td>
<td>A.H.W. Geerts M.L.J. Hautus</td>
<td>The output-stabilizable subspace and linear optimal control</td>
</tr>
<tr>
<td>M 89-16</td>
<td>June</td>
<td>P.C. Schuur</td>
<td>On the asymptotic convergence of the simulated annealing algorithm in the presence of a parameter dependent penalization</td>
</tr>
<tr>
<td>M 89-17</td>
<td>July</td>
<td>A.H.W. Geerts</td>
<td>A priori results in linear-quadratic optimal control theory (extended version)</td>
</tr>
<tr>
<td>M 89-18</td>
<td>July</td>
<td>D.A. Overdijk</td>
<td>The curvature of conjugate profiles in points of contact</td>
</tr>
<tr>
<td>M 89-19</td>
<td>August</td>
<td>A. Dekkers J. van der Wal</td>
<td>An approximation for the response time of an open CP-disk system</td>
</tr>
<tr>
<td>M 89-20</td>
<td>August</td>
<td>W.F.J. Verhaegh</td>
<td>On randomness of random number generators</td>
</tr>
<tr>
<td>M 89-21</td>
<td>August</td>
<td>P. Zwietering E. Aarts</td>
<td>Synchronously Parallel: Boltzmann Machines: a Mathematical Model</td>
</tr>
<tr>
<td>M 89-22</td>
<td>August</td>
<td>I.J.B.F. Adan J. Wessels W.H.M. Zijm</td>
<td>An asymmetric shortest queue problem</td>
</tr>
<tr>
<td>M 89-23</td>
<td>August</td>
<td>D.A. Overdijk</td>
<td>Skew-symmetric matrices in classical mechanics</td>
</tr>
<tr>
<td>M 89-24</td>
<td>September</td>
<td>F.W. Steutel J.G.F. Thiemann</td>
<td>The gamma process and the Poisson distribution</td>
</tr>
<tr>
<td>M 89-25</td>
<td>September</td>
<td>A.A. Stoorvogel</td>
<td>The discrete time $H_\infty$ control problem: the full-information case</td>
</tr>
<tr>
<td>Number</td>
<td>Month</td>
<td>Author</td>
<td>Title</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>--------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>M 89-26</td>
<td>October</td>
<td>A.H.W. Geerts</td>
<td>Linear-quadratic problems and the Riccati equation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M.L.J. Hautus</td>
<td></td>
</tr>
<tr>
<td>M 89-27</td>
<td>October</td>
<td>H.L. Trentelman</td>
<td>Completion of the squares in the finite Horizon $H$ control problem</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A.A. Stoorvogel</td>
<td>by measurement feedback</td>
</tr>
<tr>
<td>M 89-28</td>
<td>November</td>
<td>P.J. Zwietering</td>
<td>A Note on the Convergence of a Synchronously parallel Boltzmann</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E.H.L. Aarts</td>
<td>machine for the Knapsack Problem</td>
</tr>
<tr>
<td>M 89-29</td>
<td>November</td>
<td>P.C. Schuur</td>
<td>Classification of acceptance criteria for the simulated annealing algo-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rithm</td>
</tr>
</tbody>
</table>