

Some Convergence Results for Probabilistic Tabu Search

ULRICH FAIGLE *University of Twente, NL 7500 AE Enschede, P.O. Box 217, The Netherlands, EMAIL: faigle@math.utwente.nl*

WALTER KERN *University of Twente, NL 7500 AE Enschede, P.O. Box 217, The Netherlands, EMAIL: kern@henut5*

(Received: April 1990; final revision received: April 1991; accepted: May 1991)

During recent years, much work has gone into the exploration of general fundamental principles underlying local search strategies for combinatorial optimization. Many of these strategies can be subsumed under the general framework of Tabu Search, which introduces mechanisms of guidance and control based on flexible memory processes, broadening the range of strategic possibilities beyond those incorporated in memoryless search heuristics such as Simulated Annealing. We consider some examples of such memory based strategies for modifying both the generation and acceptance probabilities and investigate their impact on convergence results. It turns out that several Tabu Search ideas can be subjected to mathematical analyses similar to those applied to Simulated Annealing, making it possible to establish corresponding convergence properties based on a broader foundation.

Tabu search is a general framework for a variety of local search strategies in combinatorial optimization. During recent years, impressive practical success has been reported in applications, ranging from scheduling, cluster analysis, space planning, and even such classical problems as traveling salesman (TSP) or graph coloring. We refer to the survey articles [5, 6, 8] for more information and further references to empirical results.

In order to briefly describe the main ideas of tabu search, we start with ordinary local search. Suppose we are given

- a set $S = \{1, \dots, n\}$ of “feasible solutions”
- a function $N: S \rightarrow 2^S$, assigning to every $i \in S$ a set $N(i) \subseteq S$ of “neighbors”
- an objective function $c: S \rightarrow \mathbb{R}$ to be minimized, say.

Typically, in applications, the size of neighborhoods is small compared with the size of S . For example in a traveling salesman problem, two tours may be defined to be “neighbors” if one arises from the other by a single 2-exchange step of the well-known Kernighan-Lin heuristic. Thus, if r denotes the number of cities, then the sizes of the neighborhoods are bounded by $O(r^2)$ whereas the total number of solutions is $O(r!)$. Ordinary local search, as is well known, starts by choosing (somehow) an initial solution and subsequently replaces any current solution by one of its neighbors with a smaller objective function value until this is no longer possible, i.e., a “local optimum” is reached.

Simulated annealing (which is described in more detail in Section 1) may be seen as a probabilistic version of local search. In order to prevent the local search procedure from getting trapped in local optima, one also allows moves from a current solution to one of its neighbors with strictly larger objective function value. Such moves, however, are allowed to be performed only with a certain “acceptance probability,” depending on the amount of deterioration and the number of iterations performed so far. (If the number of iterations is high, the acceptance probabilities are set close to zero in order to eventually achieve stability).

From the preceding informal description, ordinary local search may be seen as a special version of simulated annealing where acceptance probabilities (i.e., the probabilities for accepting moves which lead to solutions with strictly larger objective function values) are equal to zero. Thus, simulated annealing generalizes ordinary local search in some sense. However, the simulated annealing approach is surely still not the most general local search method one can imagine. In particular, the following characteristic properties appear in some cases as being “too restrictive”:

1. Move decisions are primarily made only with respect to the resulting change in the objective function value. (Of course, the number of iterations, resp. the external temperature parameter $T > 0$ —see Section 2—also plays a role in the decision, but this parameter cannot be considered a “problem specific” value).

Subject classifications: Mathematics: combinatorics.

Other key words: Tabu search, simulated annealing, Markov process, convergence.

2. The procedure is completely “memoryless,” i.e., move decisions are made disregarding previous move decisions or previously obtained intermediate results.

Abiding by the framework of simulated annealing, as dictated by its theory, poses evident shortcomings. Due to (1), simulated annealing does not allow strategic decision making rules to be incorporated which might be based on our knowledge about the global problem structure. Furthermore, due to (2), it is not possible to incorporate some kind of “learning procedure,” i.e., to make effective use of information gained in previous iterations. In [5] is proposed a more general procedure, called “probabilistic tabu search,” where move decisions are derived from three different strategy considerations (cf. Section 9 in [5]):

- (S1) the change in the objective function value;
- (S2) “tabu probabilities,” reflecting our knowledge about the global problem structure;
- (S3) “aspiration levels,” related to the objective function values obtained in previous iterations.

To give a concrete example, one may broaden the ordinary simulated annealing approach for the traveling salesman problem based on Kernighan-Lin 2-exchange, say, in the following way:

- ($\hat{=}$ S2): Subdivide the whole set of cities into different regions and run a “two phase” simulated annealing algorithm. In the first phase, a certain number of iterations is performed in the usual way. In the second phase, 2-exchanges replacing edges that connect different regions are performed with smaller probability than “local exchanges.” The idea behind this approach is that the “global” structure of the solution obtained in phase 1 is already likely to be optimal. In practice, this amounts to changing the “generation probabilities” (cf. Section 1) during the process, i.e., to make the generation probabilities time dependent.
- ($\hat{=}$ S3): Modify the “acceptance probabilities” (cf. Section 1) during the process in such a way that escaping from a local optimum becomes easier. (Note that in ordinary simulated annealing it is presumably possible to check whether or not a given current solution is a local optimum. However, we will not be able to tell whether it is a true local optimum, i.e., non-global, unless we keep track of previously obtained objective function values.) How such modification may be performed will be discussed in more detail in Section 2.

Our proposal to change the generation probabilities during the process corresponds precisely to the introduction of “tabu probabilities” as proposed in [5].

The question, of course, arises whether it is possible to perform such modifications and, at the same time, keep the desirable convergence properties normally associated with the simulated annealing approach. The answer will turn out to be affirmative. Indeed, we will see that probabilistic Tabu Search converges with the following tactics.

- Make the generation probabilities time (resp. temperature) dependent, and
- Modify the acceptance probabilities (in fact modify the objective function values) during run time in a certain way.

This allows us to make effective use of information gained in previous steps by redefining the move decision rules during run time.

The purpose of this article is to propose a suitable mathematical model for probabilistic tabu search that extends the simulated annealing model in a natural way. Within this model, the above ideas for probabilistic tabu search can be described in a formally rigorous way. Moreover, this setting also provides an appropriate framework of simulated annealing to a convergence analysis of probabilistic tabu search. In fact, we will see that the convergence proof of [3] for simulated annealing yields a convergence result for probabilistic tabu search when interpreted within this framework.

The paper is organized in the following way. First, we briefly review the mathematical model for simulated annealing. We then sketch the main ideas for the convergence proof presented in [3]. We conclude by discussing applications to tabu search.

1. The Mathematical Model

In this section, we will briefly review how simulated annealing may be modeled mathematically in terms of Markov processes. We will then embed this model into a general model for probabilistic tabu search.

Let $S = \{1, \dots, n\}$ be a finite set and let $c: S \rightarrow \mathbb{R}$ be an objective function. Suppose furthermore that we are given a stochastic $(n \times n)$ -matrix $A = (a_{ij})$, i.e.,

$$a_{ij} \geq 0 \quad \text{for all } i, j \quad \text{and} \\ \sum_j a_{ij} = 1 \quad \text{for all } i.$$

The a_{ij} 's will be referred to as **generating probabilities**. In the simulated annealing process, if i is the current solution, a_{ij} is the probability for considering j as a potential successor of i . Thus, the **neighborhood** of i is given by $N(i) = \{j \mid a_{ij} > 0\}$. In the simulated annealing process, if i is the current solution, its neighbor j (once it

has been chosen to be a candidate for a successor) will be accepted as the new current solution with probability:

$$f_{ij}(T) = \begin{cases} 1 & \text{if } c_j < c_i \\ e^{-(c_j - c_i)/T} & \text{otherwise} \end{cases} \quad (j \neq i),$$

where c_i and c_j denote the objective function values of i and j , resp. and $T > 0$ is an external temperature parameter.

Thus, if the move $i \rightarrow j$ under consideration leads to an improvement (recall that we assume c to be minimized), it is accepted with probability equal to 1.

The **acceptance probabilities** $f_{ij}(T)$ above are defined in such a way that $f_{ij}(T) \rightarrow 0$ as $T \rightarrow 0$ for $c_j > c_i$, i.e., non-improving moves are less likely to be accepted if $T \approx 0$. The choice of exponential functions for the acceptance probabilities is motivated by analogies with concepts from statistical mechanics (cf, e.g., [10]). Other choices for the acceptance probabilities are possible and in fact the convergence analysis presented below can be done for a more general class of accepting probability functions (cf. [3] for more details). Here, we will restrict our attention to the case of exponential accepting probabilities as defined above.

Once the generating probabilities a_{ij} and the acceptance probabilities $f_{ij}(T)$ are fixed, a single move in the simulated annealing process may be described by the transition matrix $P(T)$ defined via

$$P_{ij}(T) = a_{ij} \cdot f_{ij}(T) \quad \forall i \neq j$$

(and $P_{ii}(T) = 1 - \sum_{j \neq i} P_{ij}(T)$).

At this point, two different (but related) directions may be pursued. One is to model the simulated annealing process by choosing a sequence of temperatures $T_1 \geq T_2 \geq \dots \geq T_k \geq \dots > 0$ with $T_k \rightarrow 0$ as $k \rightarrow \infty$ and consider the inhomogeneous Markov chain corresponding to the transition matrices $P_k = P(T_k)$. This corresponds to the simulated annealing method with the temperature parameter T being lowered in each step. Convergence analysis in this case appears to be quite involved (cf, e.g., [7]).

In practice, however, one does not proceed in this way, but rather chooses a few (say, 10) different temperature levels $T_1 \geq \dots \geq T_{10} \approx 0$ and runs a few thousand iterations at each temperature level T_i successively. The idea behind this approach is the following: Assume that $P(T)$ is irreducible and aperiodic for every $T > 0$ (which in fact is a rather weak assumption, see below). Now, for every $T > 0$, $P(T)$ has a unique stationary distribution $\pi(T)$. Running a few thousand iterations at each of the ten temperature levels T_k above, one expects that for every T_k random solutions are generated approximately according to the limit distribution $\pi(T_k)$. Thus, from a practical

point of view, one is interested in what happens with the limit distributions $\pi(T)$ as $T \rightarrow 0$ rather than in the convergence properties of the inhomogeneous Markov process described above. Furthermore, there is also theoretical motivation for restricting one's attention to the limit distribution

$$\pi^* = \lim_{T \rightarrow 0} \pi(T).$$

In fact, there are quite general results about Markov chains (cf. [1] and [2]) showing that if the inhomogeneous chain defined by $P_k = P(T_k)$ does converge, then its limit distribution is identical to π^* . Furthermore, it is shown that, under very mild conditions, the inhomogeneous process converges, provided $T_k \rightarrow 0$ sufficiently slowly as $k \rightarrow \infty$. This also settles the question about existence of π^* : Whenever the transition matrices $P(T)$ are irreducible for every $T > 0$, the limit distribution $\pi^* = \lim \pi(T)$ exists. Hence, in what follows, we will merely be interested in whether π^* has the desirable property of being nonzero only at (globally) optimal solutions $i \in S$.

Whether or not π^* has this property turns out to depend on the neighborhood structure defined on S implicitly by means of the generation probabilities a_{ij} , as we will see below. Let $G = G(A)$ denote the graph with vertex set S , whose edges are those pairs (i, j) of vertices such that

$$a_{ij} > 0.$$

Thus G is the “(directed) neighborhood” graph associated with our problem at hand. Assume that G is strongly connected, i.e., for any two vertices i and j there exists a directed path from i to j . It is easy to see that this property implies that $P(T)$ is irreducible and aperiodic for every $T > 0$ (in fact, this is equivalent to strong connectedness of G). Thus, if G is strongly connected (which we will assume in what follows without further explicit mention), we can be sure that every $P(T)$ has a unique stationary distribution $\pi(T)$ and—as we noted already above—the limit $\pi^* = \lim \pi(T)$ exists for $T \rightarrow 0$.

In order to assure that π^* has the desirable property of being nonzero only at optimal solutions $i \in S$, it is sufficient to know that the neighborhood graph G is in a certain sense “symmetric.” A first result in this direction has been proved by Lundy and Mees:

Proposition 1.1^[9]. *If A is numerically symmetric, i.e.,*

$$a_{ij} = a_{ji} \quad \text{for all } i, j \in S,$$

then $\pi^ = \lim \pi(T)$ is such that $\pi_i^* > 0$ only if $i \in S$ is a (global) optimum.*

A first generalization of this is the following result of Faigle and Schrader:

Proposition 1.2^[4]. *If A is combinatorially symmetric, i.e.*

$$a_{i,j} > 0 \Leftrightarrow a_{j,i} > 0 \quad \forall i, j \in S,$$

then $\pi^ = \lim \pi(T)$ is such that $\pi_i^* > 0$ only if $i \in S$ is a (global) optimum.*

A yet more general notion of ‘‘symmetry’’ has been introduced by Hajek (cf. [7]): Say that the neighborhood graph $G = G(A)$ is **weakly reversible**, if for every $c \in \mathbb{R}$, the connected components of the connected components of the restriction of G to the set

$$S(c) = \{i \in S \mid c_i < c\}$$

are strongly connected. In other words, G is weakly reversible if for every $c \in \mathbb{R}$ and any two vertices $i, j \in S$ the following is true: vertex j can be reached from i along a directed path without exceeding weight c along the way if and only if i can be reached from j along such a path. From this it is immediate that weak reversibility is a weaker condition than combinatorial symmetry as defined above. Let us now pass to a more general model, which will turn out to have its appropriate setting in the context of tabu search (see Section 2). The model allows us to consider generation matrices that may vary with the temperature parameter T (within certain bounds). Simulating annealing thereby corresponds to the special case of constant generation matrices. The desired convergence of simulated annealing in the presence of weak reversibility thus follows from Theorem 1.3 below, which guarantees convergence to optimum of probabilistic tabu search under mild assumptions.

We consider a family $A(T) = (a_{i,j}(T))$, $T > 0$ of stochastic matrices satisfying the following ϵ -condition

(EC) There exists an $\epsilon > 0$ such that for each $T > 0$

$$a_{i,j}(T) > 0 \quad \text{implies} \quad a_{i,j}(T) \geq \epsilon$$

whenever $i \neq j$, and all $A(T)$ define the same neighborhood graph G .

For example, $A(T) := A$ for all $T > 0$ gives rise to such a family. From this is obvious that Theorem 1.3 below in fact generalizes Proposition 1.2.

Theorem 1.3. *Let $A(T)$, $T > 0$, be a family of stochastic matrices satisfying (EC) above and such that their neighborhood graph is weakly reversible. Then $\pi^* = \lim \pi(T)$ is such that $\pi_i^* > 0$ only for optimal solutions $i \in S$.*

To make this paper self-contained, we will outline the proof of [3] for Theorem 1.3.

Proof. We will in fact prove the following stronger result.

There exists a constant $K > 0$ such that for every $T > 0$ and any two solutions i, j with $c_i < c_j$

$$\pi_j(T)/\pi_i(T) \leq K\bar{\epsilon}^{(c_j - c_i)/T}. \quad (*)$$

From this the Theorem follows immediately. The proof of (*) will be by induction on

$$n + |\{c_i \mid i \in S\}|,$$

i.e., the number of solutions plus the number of level sets. Let S^0 denote the set of optimal solutions. We consider two cases.

Case 1. $a_{i,j}(T) > 0$ for some $i, j \in S^0$, $i \neq j$. First note that, in the presence of (EC), this property does not depend on T . Indeed, we must have $a_{i,j}(T) > \epsilon$ for every T . The stationary distribution $\pi(T)$ satisfies

$$\pi(T)P(T) = \pi(T).$$

From this we get

$$\pi_j(T) \geq p_{i,j}(T)\pi_i(T) = a_{i,j}(T)\pi_i(T) \geq \epsilon\pi_i(T).$$

(Note that $p_{i,j}(T) = a_{i,j}(T)$ since $c_i = c_j$ for any two $i, j \in S^0$.) On the other hand, since G is weakly reversible, there must be a path from j to i containing only optimal solutions $k \in S^0$. Multiplying the transition probabilities along this path, we get

$$\pi_i(T) \geq \epsilon^{n-1}\pi_j(T).$$

Summarizing, we conclude

$$\pi_i(T) \geq \epsilon^{n-1}\pi_j(T) \geq \epsilon^n\pi_i(T).$$

Now define

$$\lambda(T) := \frac{\pi_i(T)}{\pi_i(T) + \pi_j(T)}.$$

From the above inequalities we deduce that $\lambda(T)$ can be bounded away from 0 and 1 by some value independent of T , say

$$\bar{\epsilon} \leq \lambda(T) \leq 1 - \bar{\epsilon} \quad \text{for some } \bar{\epsilon} \leq 1/2.$$

This allows us to ‘‘glue’’ i and j together to a new optimal solution by carrying out the following construction for each $A(T)$:

- multiply the i th row by $\lambda(T)$
- multiply the j th row by $1 - \lambda(T)$
- add the new j th row to the new i th row
- delete the j th row
- add the new j th column to the new i th column
- delete the j th column.

This gives rise to a family $\hat{A}(T)$ of $(n-1) \times (n-1)$ -matrices. These are clearly again stochastic matrices and satisfy an ϵ -condition (EC) with ϵ replaced by $\hat{\epsilon} := \epsilon\bar{\epsilon}$. Moreover, the underlying neighborhood graph \hat{G} obviously arises from contracting the edge (i, j) in G . Hence \hat{G} is also weakly reversible. Observe, furthermore, that the unique steady state distribution $\hat{\pi}(T)$ for the corresponding transition matrix $\hat{P}(T)$ must be related to $\pi(T)$ via

$$\begin{aligned}\hat{\pi}_i(T) &= \pi_i(T) + \pi_j(T) \\ \hat{\pi}_k(T) &= \pi_k(T) \quad k \neq i, j.\end{aligned}$$

Thus, by induction, there exists a constant \hat{K} such that

$$\hat{\pi}_{k/\hat{\pi}_i} \leq \hat{K}\bar{\epsilon}^{(c_k - c_i)/T}$$

whenever $c_k > c_i$. Whence it follows that (*) holds with $K = \hat{K}/\bar{\epsilon}$.

Case 2. $a_{ij}(T) = 0$ for any two $i, j \in S^0$, $i \neq j$. In this case we use induction on the number of level sets. Let

$$\delta := \min\{c_j - c_i \mid c_j \in S \setminus S^0, i \in S^0\}.$$

and consider the modified objective values

$$\hat{c}_i := \begin{cases} c_i + \delta & \text{if } i \in S^0 \\ c_i & \text{if } i \in S \setminus S^0. \end{cases}$$

By induction, we know that for the transition matrices $\hat{P}(T)$ corresponding to the modified problem and their stationary distributions $\hat{\pi}(T)$ there is a constant $\hat{K} > 0$ such that

$$\hat{\pi}_k(T) / \hat{\pi}_i(T) \leq \hat{K}\bar{\epsilon}^{(\hat{c}_k - \hat{c}_i)/T}$$

whenever $\hat{c}_k > \hat{c}_i$.

Let us now see how $\hat{\pi}(T)$ relates to the stationary distribution $\pi(T)$. Consider first the vector $\bar{\pi}(T)$ defined by

$$\bar{\pi}_i(T) = \begin{cases} \pi_i(T) & \text{if } i \in S \setminus S^0 \\ e^{-\delta/T} \pi_i(T) & \text{if } i \in S^0 \end{cases}$$

One can now show that $\bar{\pi}(T)\hat{P}(T) = \bar{\pi}(T)$. For details, we refer to [3]. Consequently, the stationary distribution $\hat{\pi}$ of \hat{P} equals $\bar{\pi}$ up to scaling, and the claim follows.

2. Applications to Tabu Search

In this section we will consider some applications to tabu search and state open problems. As mentioned in the introduction, one of the major disadvantages of simulated annealing may be seen in the fact that it does not allow the incorporation of learning processes (e.g., modifications of the transition probabilities which take into account previously acquired knowledge about the problem at hand).

However, our results on time variant generation probabilities from Section 1 can be used to guarantee convergence even when the generation probabilities are modified arbitrarily within certain bounds $[\epsilon, 1 - \epsilon]$. (Of course, setting generation probabilities to zero during the process is inadmissible since otherwise we would not even be able to check whether the resulting neighborhood graph is still connected. But what is the difference between 0 and an arbitrarily small $\epsilon > 0$ in practice?) Furthermore, the proof of Theorem 1.3 shows also that the acceptance probabilities may be changed during the process in a certain manner as discussed below. For illustration, let us look at a concrete example, say, TSP with 2-exchange and discuss some possible strategies.

Modification of Generation Probabilities

The general idea behind this approach is precisely the same as proposed in [5]. There, one introduces lists of ‘‘tabu probabilities’’ used in order to premultiply the transition probabilities (cf. [5] for details). A first example of that kind has been discussed in the introduction, where we considered a 2-phase modification of the simulated annealing approach with a local exchange phase following a global exchange phase. Another idea could be to keep track of the total number of occurrences of edges in the tours obtained during the process, possibly weighted with respect to the length of the tours containing them. An edge which only rarely occurred in (good) solutions obtained in previous iterations may be considered to be a ‘‘tabu’’ edge in subsequent iterations, i.e., one may decide to lower the generation probability of tours containing this edge by a factor of, say $1/r$ or $1/r!$, where r denotes the number of cities.

In general, a tabu search based amendment of the simulated annealing procedure, with modified generation probabilities, would consist in choosing a sequence $T_1 \geq T_2 \geq \dots$ of temperatures, running a few thousand iterations at each temperature level T_k and using acceptance probabilities

$$a_{ij}(T_k)$$

which, in fact, do depend not only on T_k , but also on information gained in previous iterations. The results from Section 1 show that this procedure still converges with $T \rightarrow 0$ to the set of optimum solution.

Modification of Acceptance Probabilities

Suppose that during the search process we arrive at a particular solution i which is locally optimal. Usually, the preceding type of algorithmic implementation will not recognize local optimality but of course, at least theoretically, we can perform a local optimality check in each iteration. A more practical manner to achieve this goal might consist in counting the number of subsequent

iterations in which no move is performed. If that count exceeds a certain threshold value, we are likely to be in a local optimum and thus decide to perform a (comparatively expensive) local optimality check. Assume furthermore, that the local optimum we are in is a strict local optimum i.e., every neighbor has a strictly larger objective function value. This will be the case with a 99% chance in practice, but may also be assured in advance by using perturbation techniques. Finally, assume we know that we are not in a global optimum (e.g., because we have encountered a better solution already in previous iterations). In this case it follows as in the proof of Theorem 1.3 that we may set the current solution's objective function value to the new value

$$\hat{c}_i = \min\{c_j \mid j \text{ is a neighbor of } i\}$$

without affecting the conditions (weak reversibility) necessary to ensure convergence. This operation obviously results in gradually "lifting deep valleys," i.e., flattening the landscape associated with the cost function of the problem at hand. As a consequence, local optima will be easier to escape from in subsequent iterations.

Open Problems

Note that in our examples for modifying the generation probabilities we have changed the generation probabilities of whole classes of feasible solutions simultaneously (e.g., the class of solutions containing a certain edge) rather than dealing with individual solutions. In fact, individual treatment of particular solutions seems to be a difficult task from a technical point of view. Although we did not mention it in our above example for modifying acceptance probabilities, the problem, of course, also arises there: Note that, usually, the set of solutions and their corresponding objective function values are defined implicitly, as by specifying a set of r cities and their relative distances. Thus an operation like redefining a particular objective function value c_i to be some new value \hat{c}_i seems to be difficult to implement.

As a matter of fact, the very same problem occurs in many other situations. For example, [6] proposes to compute "tabu probabilities," i.e., to modify the transition

probabilities, in order to prevent the process from "cycling." This is to say, if after a certain number of iterations a solution i is encountered which has been visited before, one might want to decrease the transition probability a_{ij} to the particular neighbor j that was selected the previous time. But how can we even know whether a particular solution already has been visited once and which move was performed then? Keeping track of all previous moves is practically impossible. One possible way out of this problem might consist in using hashing functions (cf. [11]). Another approach to resolving this dilemma might consist in developing "associative memory" based techniques (which would possibly lead also to a less strict definition of "objective function values").

REFERENCES

1. S. ANILY AND A. FEDERGRUEN, 1985. Ergodicity in Parametric Nonstationary Markov Chains: Applications to Simulated Annealing Methods, Working Paper, Columbia University, New York.
2. S. ANILY AND A. FEDERGRUEN, 1987. Simulated Annealing Methods with General Acceptance Probabilities, *Journal of Applied Probability* 24, 657-667.
3. U. FAIGLE AND W. KERN, 1991. Note on the Convergence of Simulated Annealing Algorithms, *SIAM Journal of Control and Optimization* 29, 153-159.
4. U. FAIGLE AND R. SCHRADER, 1988. On the Convergence of Stationary Distributions in Simulated Annealing Algorithms, *Information Processing Letters* 27, 189-194.
5. F. GLOVER, 1989. Tabu Search—Part I, *ORSA Journal on Computing* 1:3, 190-204.
6. F. GLOVER, 1989. Tabu Search: A Tutorial, Working Paper, University of Colorado, to appear in *Interfaces*.
7. B. HAJEK, 1988. Cooling Schedules for Optimal Annealing, *Mathematics of Operations Research* 13, 311-329.
8. A. HERTZ AND D. DE WERRA, 1989. The Tabu Search Metaheuristic: How We Used It, Working Paper, Swiss Federal Institute of Technology, to appear in *Annals of Mathematics and Artificial Intelligence*.
9. M. LUNDY AND A. MEES, 1986. Convergence of the Annealing Algorithm, *Mathematical Programming* 34, 111-124.
10. N. METROPOLIS, A.W. ROSENBLUTH, M.N. ROSENBLUTH, A.H. TELLER, AND E. TELLER, 1953. Equations of State Calculations by Fast Computing Machines, *Journal of Chemical Physics* 21, 1087-1092.
11. D. WOODRUFF AND E. ZEMEL, 1990. Hashing Vectors for Tabu Search, Technical Report 90-08, Department of Industrial Engineering, Northwestern University, Evanston, IL.

Copyright of ORSA Journal on Computing is the property of INFORMS: Institute for Operations Research and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.