Chapter 10

THE THEORY AND PRACTICE OF SIMULATED ANNEALING

Darrall Henderson
Department of Mathematical Sciences
United States Military Academy
West Point, NY 10996-1786, USA
E-mail: darrall@stanfordalumni.org

Sheldon H. Jacobson
Department of Mechanical and Industrial Engineering
University of Illinois at Urbana-Champaign
1206 West Green Street, MC-244
Urbana, Illinois 61801-2906, USA
E-mail: shj@uiuc.edu

Alan W. Johnson
Department of Mathematical Sciences
United States Military Academy
West Point, NY 10996-1786, USA
E-mail: aa2895@usma.edu

Abstract   Simulated annealing is a popular local search meta-heuristic used to address discrete and, to a lesser extent, continuous optimization problems. The key feature of simulated annealing is that it provides a means to escape local optima by allowing hill-climbing moves (i.e., moves which worsen the objective function value) in hopes of finding a global optimum. A brief history of simulated annealing is presented, including a review of its application to discrete and continuous optimization problems. Convergence theory for simulated annealing is reviewed, as well as recent advances in the analysis of finite time performance. Other local search algorithms are discussed in terms of their relationship to simulated annealing. The chapter also presents practical guidelines for the implementation of simulated annealing in terms of cooling schedules, neighborhood functions, and appropriate applications.

Keywords: Local Search Algorithms, Simulated Annealing, Heuristics, Meta-heuristics

1   BACKGROUND SURVEY

Simulated annealing is a local search algorithm (meta-heuristic) capable of escaping from local optima. Its ease of implementation, convergence properties and its use of hill-climbing moves to escape local optima have made it a popular technique over the past two decades. It is typically used to address discrete, and to a lesser extent, continuous optimization problems. Recent survey articles that provide a good overview of simulated annealing's theoretical development and domains of application include Eglese (1990), Fleischer (1995), Koulamas et al. (1994), and Romeo and Sangiovanni-Vincentelli (1991). Aarts and Korst (1989) and van Laarhoven and Aarts (1988) devote entire books to the subject. Aarts and Lenstra (1997) dedicate a chapter to simulated annealing in their book on local search algorithms for discrete optimization problems.

1.1   History and Motivation

Simulated annealing is so named because of its analogy to the process of physical annealing with solids, in which a crystalline solid is heated and then allowed to cool very slowly until it achieves its most regular possible crystal lattice configuration (i.e., its minimum lattice energy state), and thus is free of crystal defects. If the cooling schedule is sufficiently slow, the final configuration results in a solid with such superior structural integrity. Simulated annealing establishes the connection between this type of thermodynamic behavior and the search for global minima for a discrete optimization problem. Furthermore, it provides an algorithmic means for exploiting such a connection.

At each iteration of a simulated annealing algorithm applied to a discrete optimization problem, the objective function generates values for two solutions (the current solution and a newly selected solution) are compared. Improving solutions are always accepted, while a fraction of non-improving (inferior) solutions are accepted in the hope of escaping local optima in search of global optima. The probability of accepting non-improving solutions depends on a temperature parameter, which is typically non-increasing with each iteration of the algorithm.

The key algorithmic feature of simulated annealing is that it provides a means to escape local optima by allowing *hill-climbing* moves (i.e., moves which worsen the objective function value). As the temperature parameter is decreased to zero, hill-climbing moves occur less frequently, and the solution distribution associated with the inhomogeneous Markov chain that models the behavior of the algorithm converges to a form in which all the probability is concentrated on the set of globally optimal solutions (provided that the algorithm is convergent; otherwise the algorithm will converge to a local optimum, which may or not be globally optimal).

1.2   Definition of Terms

To describe the specific features of a simulated annealing algorithm for discrete optimization problems, several definitions are needed. Let Ω be the *solution space* (i.e., the set of all possible solutions). Let $f : \Omega \rightarrow \Re$ be an *objective function* defined on the solution space. The goal is to find a global minimum, ω^* (i.e., $\omega^* \in \Omega$ such that $f(\omega) \geq f(\omega^*)$ for all $\omega \in \Omega$). The objective function must be bounded to ensure that ω^* exists. Define $\mathsf{N}(\omega)$ to be the *neighborhood function* for $\omega \in \Omega$. Therefore, associated with every solution, $\omega \in \Omega$, are neighboring solutions, $\mathsf{N}(\omega)$, that can be reached in a single iteration of a local search algorithm.

Simulated annealing starts with an initial solution $\omega \in \Omega$. A neighboring solution $\omega' \in N(\omega)$ is then generated (either randomly or using some pre-specified rule). Simulated annealing is based on the Metropolis acceptance criterion (Metropolis et al., 1953), which models how a thermodynamic system moves from the current solution (state) $\omega \in \Omega$ to a candidate solution $\omega' \in N(\omega)$, in which the energy content is being minimized. The candidate solution, ω', is accepted as the current solution based on the acceptance probability

$$P\{\text{Accept } \omega' \text{ as next solution}\} = \begin{cases} \exp[-(f(\omega') - f(\omega))/t_k] & \text{if } f(\omega') - f(\omega) > 0 \\ 1 & \text{if } f(\omega') - f(\omega) \le 0. \end{cases} \tag{1}$$

Define t_k as the temperature parameter at (outer loop) iteration k, such that

$$t_k > 0 \quad \text{for all } k \quad \text{and} \quad \lim_{k \to +\infty} t_k = 0 \tag{2}$$

This acceptance probability is the basic element of the search mechanism in simulated annealing. If the temperature is reduced sufficiently slowly, then the system can reach an equilibrium (steady state) at each iteration k. Let $f(\omega)$ and $f(\omega')$ denote the energies (objective function values) associated with solutions $\omega \in \Omega$ and $\omega' \in N(\omega)$, respectively. This equilibrium follows the Boltzmann distribution, which can be described as the probability of the system being in state $\omega \in \Omega$ with energy $f(\omega)$ at temperature T such that

$$P\{\text{System is in state } \omega \text{ at temperature } T\} = \frac{\exp(-f(\omega)/t_k)}{\sum_{\omega'' \in \Omega} \exp(-f(\omega'')/t_k)}. \tag{3}$$

If the probability of generating a candidate solution ω' from the neighbors of solution $\omega \in \Omega$ is $g_k(\omega, \omega')$, where

$$\sum_{\omega' \in N(\omega)} g_k(\omega, \omega') = 1, \quad \text{for all } \omega \in \Omega, \ k = 1, 2, \ldots, \tag{4}$$

then a non-negative square stochastic matrix $\mathbf{P}_k$ can be defined with transition probabilities

$$\mathbf{P}_k(\omega, \omega') = \begin{cases} g_k(\omega, \omega') \exp(-\Delta_{\omega,\omega'}/t_k) & \omega' \in N(\omega), \omega' \ne \omega \\ 0 & \omega' \notin N(\omega), \omega' \ne \omega \\ 1 - \sum_{\substack{\omega'' \in N(\omega) \\ \omega'' \ne \omega}} \mathbf{P}_k(\omega, \omega'') & \omega' = \omega \end{cases} \tag{5}$$

for all solutions $i \in \Omega$ and all iterations $k = 1, 2, \ldots$ and $\Delta_{\omega,\omega'} \equiv f(\omega') - f(\omega)$. These transition probabilities define a sequence of solutions generated from an inhomogeneous Markov chain (Romeo and Sangiovanni-Vincentelli, 1991). Note that boldface type indicates matrix/vector notation, and all vectors are row vectors.

1.3   Statement of Algorithm

Simulated annealing is outlined in pseudo-code (Eglese, 1990).

Select an initial solution $\omega \in \Omega$
Select the temperature change counter $k=0$
Select a temperature cooling schedule, t_k
Select an initial temperature $T = t_0 \geq 0$
Select a repetition schedule, M_k that defines the number of iterations executed at each
    temperature, t_k
Repeat
Set repetition counter $m = 0$
    Repeat
    Generate a solution $\omega' \in N(\omega)$
    Calculate $\Delta_{\omega,\omega'} = f(\omega') - f(\omega)$
    If $\Delta_{\omega,\omega'} \leq 0$, then $\omega \leftarrow \omega'$
    If $\Delta_{\omega,\omega'} > 0$, then $\omega \leftarrow \omega'$ with probability $\exp(-\Delta_{\omega,\omega'}/t_k)$
    $m \leftarrow m + 1$
    Until $m = M_k$
$k \leftarrow k + 1$
Until stopping criterion is met

This simulated annealing formulation results in $M_0 + M_1 + \cdots + M_k$ total iterations being executed, where k corresponds to the value for t_k at which the stopping criteria is met. In addition, if $M_k = 1$ for all k, then the temperature changes at each iteration.

1.4   Discrete versus Continuous Problems

The majority of the theoretical developments and application work with simulated annealing has been for discrete optimization problems. However simulated annealing has also been used as a tool to address problems in the continuous domain. There is considerable interest in using simulated annealing for global optimization over regions containing several local and global minima (due to inherent non-linearity of objective functions). Fabian (1997) studies the performance of simulated annealing methods for finding a global minimum of a function and Bohachevsky et al. (1986) presents a generalized simulated annealing algorithm for function optimization for use in statistical applications. Optimization of continuous functions involves finding a candidate location by picking a direction from the current (incumbent) solution and a step size to take in this direction, and evaluating the function at the new (candidate) location. If the function value of this candidate location is an improvement over the function value of the incumbent location, then the candidate becomes the incumbent. This migration through local minima in search of a global minimum continues until the global minimum is found or some termination criteria are reached. Belisle (1992) presents a special simulated annealing algorithm for global optimization, which uses a heuristically motivated cooling schedule. This algorithm is easy to implement and provides a reasonable alternative to existing methods.

Belisle et al. (1993) discusses convergence properties of simulated annealing algorithms applied to continuous functions and applies these results to hit-and-run algorithms used in global optimization. His convergence properties are consistent

with those presented by Hajek (1988) and he provides a good contrast between convergence in probability and the stronger almost sure convergence. Zabinsky et al. (1993) extends this work to an improved hit-and-run algorithm used for global optimization.

Fleischer and Jacobson (1996) proposes cybernetic optimization by simulated annealing as a method of parallel processing that accelerated the convergence of simulated annealing to the global optima. Fleischer (1999) extends the theory of cybernetic optimization by simulated annealing into the continuous domain by applying probabilistic feedback control to the generation of candidate solutions. The probabilistic feedback control method of generating candidate solutions effectively accelerates convergence to a global optimum using parallel simulated annealing on continuous variable problems.

Locatelli (1996) presents convergence properties for a class of simulated annealing algorithms for continuous global optimization by removing the restriction that the next candidate point must be generated according to a probability distribution whose support is the whole feasible set. Siarry et al. (1997) studies simulated annealing algorithms for globally minimizing functions of many-continuous variables. Their work focuses on how high-dimensionality can be addressed using variables discretization, as well as considering the design and implementation of several complementary stopping criteria. Yang (2000) and Locatelli (2000) also provide convergence results and criteria for simulated annealing applied to continuous global optimization problems. Kiatsupaibul and Smith (2000) introduces a general purpose simulated annealing algorithm to solve mixed integer linear programs. The simulated annealing algorithm is constructed using a Markov chain sampling algorithm to generate uniformly distributed points on an arbitrary bounded region of a high dimensional integer lattice. They show that their algorithm converges in probability to a global optimum. Romeijn et al. (1999) also studies a simulated annealing algorithm that uses a reflection generator for mixed integer/continuous global optimization problems.

2   CONVERGENCE RESULTS

Convergence results for simulated annealing have typically taken one of two directions; either the algorithm has been modeled as a sequence of homogeneous Markov chains or as a single inhomogeneous Markov chain.

The homogeneous Markov chain approach (see, e.g., Aarts and van Laarhoven, 1985; Faigle and Kern, 1991; Granville et al., 1994; Johnson and Jacobson, 2002a,b; Lundy and Mees, 1986; Mitra et al., 1986; Rossier et al., 1986) assumes that each temperature t_k is held constant for a sufficient number of iterations m such that the stochastic matrix $\mathbf{P}_k$ can reach its stationary (steady state) distribution $\boldsymbol{\pi}_k$. Note that in the interest of simplifying notation, the inner loop index m is suppressed. However, the index k should be interpreted as the double index k,m, where a sequence of $m = 1, 2, \ldots, \mathbf{M}_k$ simulated annealing iterations occur for each fixed k.) The existence of a stationary distribution at each iteration k follows from Theorem 10.1. (Note: to ensure that Theorem 1 is consistent with the simulated annealing algorithm depicted in Section 1.3, without loss of generality, let t_k be a function only of each outer loop iteration k, and let the respective number of inner loop iterations $\mathbf{M}_k$ and outer loop iterations k each approach infinity).

Theorem 10.1.   *Let* $\mathbf{P}_k(\omega, \omega')$ *be the probability of moving from solution ω to solution ω' in one inner iteration at outer loop k, and let* $\mathbf{P}_k^{(m)}(\omega, \omega')$ *be the probability of going from solution ω to solution ω' in m inner loops. If the Markov chain associated with* $\mathbf{P}_k^{(m)}(\omega, \omega')$ *is irreducible and aperiodic with finitely many solutions, then* $\lim_{m \to \infty} \mathbf{P}_k^{(m)}(\omega, \omega') = \pi_k(\omega')$ *exists for all $\omega, \omega' \in \Omega$ and iterations k. Furthermore,* $\pi_k(\omega')$ *is the unique strictly positive solution of*

$$\pi_k(\omega') = \sum_{\omega \in \Omega} \pi_k(\omega)\mathbf{P}_k(\omega, \omega'), \quad \text{for all } \omega \in \Omega, \tag{6}$$

and

$$\sum_{\omega \in \Omega} \pi_k(\omega) = 1. \tag{7}$$

   Proof.   See Cinlar (1974)p. 153. ∎

   The key requirements for the existence of the stationary distributions and for the convergence of the sequence of $\boldsymbol{\pi}_k$ vectors include

1. transition matrix irreducibility (for every finite outer loop *k*, the transition matrix can assign a path of non-zero probability between any two solutions $\omega, \omega' \in \Omega$),

2. aperiodicity (starting at solution $\omega' \in \Omega$, it is possible to return to ω, ω' with period 1. See Isaacson and Madsen (1976),

3. a non-zero stationary probability distribution, as the number of outer loops *k* approaches infinity.

   Note that all simulated annealing proofs of convergence in the literature based on homogeneous Markov chain theory, either explicitly or implicitly, use the sufficient condition of reversibility (also called detailed balance) (Ross, 1996), defined as

$$\pi_k(\omega)\mathbf{P}_k(\omega, \omega') = \pi_k(\omega')\mathbf{P}_k(\omega', \omega), \text{ for all } \omega, \omega' \in \Omega, \text{ and all iterations } k. \tag{8}$$

   Reversibility is sufficient condition for a unique solution to exist for (6) and (7) at each outer loop iteration *k*. Ross (1997) shows that a necessary condition for reversibility is multiplicativity (i.e., for any three solutions $\omega, \omega', \omega'' \in \Omega$ such that $f(\omega) \leq f(\omega') \leq f(\omega'')$, and for all iterations *k*,

$$a_k(\Delta_{\omega, \omega''}) = a_k(\Delta_{\omega, \omega'})a_k(\Delta_{\omega', \omega''}) \tag{9}$$

where $a_k(\Delta_{\omega, \omega'})$ is the probability of accepting the transition from solution ω to solution ω' at outer loop iteration *k*). Reversibility is enforced by assuming conditions of symmetry on the solution generation probabilities and by either directly expressing the acceptance probability using an exponential form, or by requiring the multiplicative condition in (9).

   The homogeneous Markov chain proofs of convergence in the literature (implicitly or explicitly) require the condition in (9) to hold for the acceptance function, and then address the sufficient conditions on the solution generation matrix. For example, the original homogeneous proofs of convergence (Aarts and van Laarhoven, 1985; Lundy and Mees, 1986) require the multiplicative condition for the acceptance function, and then assume that the solution generation function is symmetric and constant for all outer

loop iterations k. Rossier et al. (1986) partitions the solution space into blocks composed of neighboring solutions of equal objective function value, and then requires that only the solution generation probabilities be symmetric between these blocks. Rossier et al. (1986) then expresses the acceptance function as a ratio of the stationary distribution probabilities (discussed in Section 2.1.3). Faigle and Schrader (1988) and Faigle and Kern (1991) use a graph theoretic approach to relax the solution generation function symmetry condition. However, they require that the solution acceptance probability function satisfies (9).

Granville et al. (1994) proposes a simulated annealing procedure for filtering binary images, where the acceptance function is based on the probability of the current solution, instead of the change in objective function value. The probability function that Granville et al. (1994) present for accepting a candidate solution at (outer loop) iteration k is based on the ratio of the stationary probability of the incumbent solution from iteration $k - 1$, versus the stationary probability of an initial solution (which is based on a maximum likelihood estimate). The acceptance probability is

$$\xi_k = (q\pi_0(\omega)/\pi_{k-1}(\omega'))^{\phi(k)} \tag{10}$$

where $q = \inf_{\omega \in \Omega} \pi(\omega) / \sup_{\omega' \in \Omega} \pi(\omega')$ (q must also be estimated), and $\phi(k)$ is a slowly increasing function. Therefore, the probability of a solution transition does not consider the objective function value of the candidate solution. Granville et al. (1994) provides a proof of asymptotic convergence of this approach, but note that the proof methodology does not show that the set of globally optimal solutions are asymptotically uniformly distributed.

Simulated annealing and the homogeneous convergence theory are based on the work of Metropolis et al. (1953), which addresses problems in equilibrium statistical mechanics (Hammersley and Handscomb, 1964). To see this relationship, consider a system in thermal equilibrium with its surroundings, in solution (state) S with energy $F(S)$. The probability density in phase space of the point representing S is proportional to

$$\exp(-F(S)/bT), \tag{11}$$

where b is the Boltzmann constant, and T is the absolute temperature of the surroundings. Therefore the proportion of time that the system spends in solution S is proportional to (11) (Hammersley and Handscomb, 1964), hence the equilibrium probability density for all $S \in \Omega$ is

$$\pi_S = \frac{\exp(-F(S)/bT)}{\int \exp(-F(S)/bT) \, dS}. \tag{12}$$

The expectation of any valid solution function $f(S)$ is thus

$$E[f] = \frac{\int f(S) \exp(-F(S)/bT) \, dS}{\int \exp(-F(S)/bT) \, dS}. \tag{13}$$

Unfortunately, for many solution functions, (13) cannot be evaluated analytically. Hammersley and Handscomb (1964) notes that one could theoretically use naive Monte Carlo techniques to estimate the value of the two integrals in (11). However, this often

fails in practice since the exponential factor means that a significant portion of the integrals is concentrated in a very small region of the solution space Ω. This problem can be overcome using importance sampling (see Bratley et al., 1987; Chapter 2), by generating solutions with probability density (12). This approach would also seem to fail, because of the integral in the denominator of (12). However, Metropolis et al. (1953) solves this problem by first discretizing the solution space, such that the integrals in (12) and (13) are replaced by summations over the set of discrete solutions $\omega' \in \Omega$, and then by constructing an irreducible, aperiodic Markov chain with transition probabilities $\mathbf{P}(\omega, \omega')$ such that

$$\pi(\omega') = \sum_{\omega \in \Omega} \pi(\omega)\mathbf{P}(\omega, \omega') \quad \text{for all } \omega' \in \Omega, \tag{14}$$

where

$$\pi(\omega') = \frac{\exp(-F(\omega')/bT)}{\sum_{\omega \in \Omega} \exp(-F(\omega)/bT)} \quad \text{for all } \omega' \in \Omega. \tag{15}$$

Note that to compute the equilibrium distribution $\boldsymbol{\pi}$, the denominator of (13) (a normalizing constant) does not need to be calculated. Instead, the ratios $\pi(\omega')/\pi(\omega)$ need only be computed and a transition matrix $\mathbf{P}$ defined that satisfies (14). Hammersley and Handscomb (1964) show that Metropolis et al. (1953) accomplishes this by defining $\mathbf{P}$ as the product of symmetric solution generation probabilities $g(\omega, \omega')$, and the equilibrium ratios $\pi(\omega)'/\pi(\omega)$,

$$\mathbf{P}(\omega, \omega')$$
$$= \begin{cases} g(\omega, \omega')\pi(\omega')/\pi(\omega) & \text{if } \pi(\omega')/\pi(\omega) < 1, \omega' \neq \omega \\ g(\omega, \omega') & \text{if } \pi(\omega')/\pi(\omega) \geq 1, \omega' \neq \omega \\ g(\omega, \omega') + \sum_{\substack{\omega'' \in \Omega, \\ \pi(\omega'') < \pi(\omega)}} g(\omega, \omega'')(1 - (\pi(\omega'')/\pi(\omega))) & \text{if } \omega' = \omega \end{cases} \tag{16}$$

where

$$g(\omega, \omega') \geq 0, \sum_{\omega' \in \Omega} g(\omega, \omega') = 1, \text{ and } g(\omega, \omega') = g(\omega', \omega) \quad \text{for all } \omega, \omega' \in \Omega \tag{17}$$

The use of stationary probability ratios to define the solution acceptance probabilities, combined with symmetric solution generation probabilities, enable Metropolis et al. (1953) to use the reversibility condition in (8) to show that (16) and (17) satisfy (14).

Homogeneous proofs of convergence for simulated annealing become more difficult to establish when the reversibility condition is not satisfied. Note that the existence of a unique stationary distribution for each outer loop iteration k is easily shown by specifying that each transition matrix $\mathbf{P}_k$ be irreducible and aperiodic. On the other hand, it becomes very difficult to derive an explicit closed-form expression for each stationary distribution $\boldsymbol{\pi}_k$ that remains analytically tractable as the problem's solution space becomes large. One can no longer use (8) to describe each stationary distribution, because in general, the multiplicative condition is not met. Instead, one must directly solve the system of equations formed with (6) and (7). For example, Davis (1991)

attempts to obtain a closed-form expression for π_k by using Cramer's rule and rewriting (6) and (7) as

$$\pi_k(\mathbf{I} - \mathbf{P}_k) = 0 \tag{18}$$

and

$$\pi_k\mathbf{e}^{\mathbf{T}} = 1, \tag{19}$$

respectively, where boldface type indicates vector/matrix notation, $\mathbf{I}$ is the identity matrix, and $\mathbf{e}^{\mathbf{T}}$ is a column vector of ones. Note that the $\text{card}(\Omega) \times \text{card}(\Omega)$ transition matrix $\mathbf{P}_k$ associated with (18) is of rank $\text{card}(\Omega)\text{-}1$ (Çinlar, 1974). Therefore, by deleting any one equation from (18), and substituting (19), the result is the set of $\text{card}(\Omega)$ linearly independent equations

$$\pi_k\underline{(\mathbf{I} - \mathbf{P}_k)} = \mathbf{e_i} \tag{20}$$

where the square matrix $\underline{(\mathbf{I} - \mathbf{P}_k)}$ is obtained by substituting the ith column of matrix $(\mathbf{I} - \mathbf{P}_k)$ with a column vector of ones. The vector $\mathbf{e_i}$ is a row vector of zeroes, except for a one in the ith position. Since $\underline{(\mathbf{I} - \mathbf{P}_k)}$ is of full rank, then its determinant (written as $\det\underline{(\mathbf{I} - \mathbf{P}_k)}$), is non-zero. Define $\underline{(\mathbf{I} - \mathbf{P}_k)}^{\omega}$ to be the same matrix as $\underline{(\mathbf{I} - \mathbf{P}_k)}$ except that the elements of the ωth row of $\underline{(\mathbf{I} - \mathbf{P}_k)}$ are replaced by the vector $\mathbf{e}_{\omega}$. Therefore, for all iterations k,

$$\pi_k(\omega) = \frac{\det(\mathbf{I} - \mathbf{P}_k)^{\omega}}{\det\underline{(\mathbf{I} - \mathbf{P}_k)}}, \quad \text{for all } \omega \in \Omega \tag{21}$$

Davis (1991) attempts to solve (21) for each $\omega \in \Omega$ via a multivariate Taylor series expansion of each determinant, but is not able to derive a closed-form analytical expression.

Overall, the difficulty of explicitly expressing the stationary distributions for large solution spaces, combined with bounding the transition matrix condition number for large k, suggests that it is very difficult to prove asymptotic convergence of the simulated annealing algorithm by treating (5) and (6) as a linear algebra problem.

Lundy and Mees (1986) notes that for each fixed outer loop iteration k, convergence to the solution equilibrium probability distribution vector π_k (in terms of the Euclidean distance between $P_k^{(m)}$ and π_k, as $m \rightarrow +\infty$) is geometric since the solution space is finite, and the convergence factor is given by the second largest eigenvalue of the transition matrix $\mathbf{P}_k$. This result is based on a standard convergence theorem for irreducible, aperiodic homogeneous Markov chains (see Çinlar, 1974). Note that a large solution space precludes practical calculation of this eigenvalue. Lundy and Mees (1986) conjectures that when the temperature t_k is near zero, the second largest eigenvalue will be close to one for problems with local optima, and thus convergence to the equilibrium distribution will be very slow (recall that the dominant eigenvalue for $\mathbf{P}_k$ is one, with algebraic multiplicity one (Isaacson and Madsen, 1976). Lundy and Mees (1986) uses its conjecture to justify why simulated annealing should be initiated with a relatively high temperature. For an overview of current methods for assessing non-asymptotic rates of convergence for general homogeneous Markov chains, see Rosenthal (1995).

The assumption of stationarity for each outer loop iteration k limits practical application of homogeneous Markov chain theory—Romeo and Sangiovanni-Vincentelli (1991) shows that if equilibrium (for a Markov chain that satisfies the reversibility condition) is reached in a finite number of steps, then it is achieved in one step. Thus,

Romeo and Sangiovanni-Vincentelli (1991) conjectures that there is essentially no hope for the most-used versions of simulated annealing to reach equilibrium in a finite number of iterations.

The second convergence approach for simulated annealing is based on inhomogeneous Markov chain theory (Anily and Federgruen, 1987; Gidas, 1985; Mitra et al., 1986). In this approach, the Markov chain need not reach a stationary distribution (e.g., the simulated annealing inner loop need not be infinitely long) for each outer loop k. On the other hand, an infinite sequence of (outer loop) iterations k must still be examined, with the condition that the temperature parameter t_k cool sufficiently slowly. The proof given by Mitra et al. (1986) is based on satisfying the inhomogeneous Markov chain conditions of weak and strong ergodicity (Isaacson and Madsen, 1976; Seneta, 1981). The proof requires four conditions:

1. The inhomogeneous simulated annealing Markov chain must be weakly ergodic (i.e., dependence on the initial solution vanishes in the limit).

2. An eigenvector π_k with eigenvalue one must exist such that (6) and (7) hold for every iteration k.

3. The Markov chain must be strongly ergodic (i.e., the Markov chain must be weakly ergodic and the sequence of eigenvectors π_k must converge to a limiting form), i.e.,

$$\sum_{k=0}^{\infty} \|\pi_k - \pi_{k+1}\| < +\infty. \tag{22}$$

4. The sequence of eigenvectors must converge to a form where all probability mass is concentrated on the set of globally optimal solutions ω^*. Therefore,

$$\lim_{k \to \infty} \pi_k = \pi^{\text{opt}}. \tag{23}$$

where π^{opt} is the equilibrium distribution with only global optima having probabilities greater than zero. (Note that weak and strong ergodicity are equivalent for homogeneous Markov chain theory.)

Mitra et al. (1986) satisfies condition (1) (weak ergodicity) by first forming a lower bound on the probability of reaching any solution from any local minimum, and then showing that this bound does not approach zero too quickly. For example, they define the lower bound for the simulated annealing transition probabilities in (5) as

$$\mathbf{P}^{(m)}(\omega, \omega') \geq w^m \exp(-m\Delta_L/t_{km-1}) \tag{24}$$

where m is the number of transitions needed to reach any solution from any solution of non-maximal objective function value, $w > 0$ is a lower bound on the one-step solution generation probabilities, and Δ_L is the maximum one-step increase in objective function value between any two solutions. Mitra et al. (1986) shows that the Markov chain is weakly ergodic if

$$\sum_{k=k_0}^{\infty} \exp(-m\Delta_L/t_{km-1}) = +\infty \tag{25}$$

Therefore, weak ergodicity is obtained if the temperature t_k is reduced sufficiently slowly to zero such that (25) is satisfied. In general, the (infinite) sequence of temperatures $\{t_k\}, k = 1, 2, \ldots$ must satisfy

$$t_k \geq \frac{\beta}{\log(k)} \tag{26}$$

where $\lim_{k \to \infty} t_k = 0$, β is a problem-dependent constant, and k is the number of iterations. Mitra et al. (1986) shows that conditions (2), (3), and (4) are satisfied by using the homogeneous Markov chain theory developed for the transition probabilities (5), and assuming that the solution generation function is symmetric.

Romeo and Sangiovanni-Vincentelli (1991) notes that while the logarithmic cooling schedule in (26) is a sufficient condition for the convergence, there are only a few values for β which make the logarithmic rule also necessary. Furthermore, there exists a unique choice for β which makes the logarithmic rule both necessary and sufficient for the convergence of simulated annealing to the set of global optima. Hajek (1988) was the first to show that the logarithmic cooling schedule (26) is both necessary and sufficient, by developing a tight lower bound for β, namely the depth of the deepest local minimum which is not a global minimum, under a weak reversibility assumption. (Note that Hajek requires the depth of global optima to be infinitely deep.) Hajek defines a Markov chain to be weakly reversible if, for any pair of solutions $\omega, \omega' \in \Omega$ and for any non-negative real number h, ω is reachable at height h from ω' if and only if ω' is reachable at height h from ω. Note that Hajek (1988) does not attempt to satisfy the conditions of weak and strong ergodicity, but instead uses a more general probabilistic approach to develop a lower bound on the probability of escaping local, but not global optima. Connors and Kumar (1989) substantiate the necessary and sufficient conditions in Hajek (1988) using the orders of recurrence,

$$B_i \equiv \sup \left\{ x \geq 0 : \sum_{k=0}^{\infty} \exp(-x/t_k)\pi_k(\omega) = +\infty \right\} \quad \text{for all } i \in \Omega. \tag{27}$$

Connors and Kumar (1989) shows that these orders of recurrence quantify the asymptotic behavior of each solution's probability in the solution distribution. The key result is that the simulated annealing inhomogeneous Markov chain converges in a Cesaro sense to the set of solutions having the largest recurrence orders. Borkar (1992) improves this convergence result by using a convergence/oscillation dichotomy result for martingales. Tsitsiklis (1989) uses bounds and estimates for singularly perturbed, approximately stationary Markov chains to develop a convergence theory that subsumes the condition of weak reversibility in Hajek (1988). (Note that Tsitsiklis (1989) defines $N(h) \subset \Omega$ as the set of all local minima (in terms of objective function value) of depth $h + 1$ or more. Therefore β is the smallest h such that all local (but not global) minima have depth h or less. Tsitsiklis (1989) conjectures that without some form of reversibility, there does not exist any h such that the global optima are contained in the set of local optima.) Note that Chiang and Chow (1988, 1994), Borkar (1992), Connors and Kumar (1989), Hajek (1988), and Mitra et al. (1986) all require (either explicitly or implicitly) the multiplicative condition (9) for their proofs of convergence.

Anily and Federgruen (1987) uses perturbation analysis techniques (e.g., see Meyer, 1980) to prove convergence of a particular stochastic hill-climbing algorithm, by

bounding the deviations of the sequence of stationary distributions of the particular hill-climbing algorithm against the sequence of known stationary distributions corresponding to a simulated annealing algorithm. In general, this convergence proof approach is only useful for a restrictive class of simulated annealing algorithms, since the transition matrix condition number grows exponentially as the number of iterations k becomes large.

Anily and Federgruen (1987) also presents a proof of convergence for simulated annealing with general acceptance probability functions. Using inhomogeneous Markov chain theory, it proves convergence under the following necessary and sufficient conditions:

1. The acceptance probability function must, for any iteration k, allow any hill-climbing transition to occur with positive probability.

2. The acceptance probability function must be bounded and asymptotically monotone, with limit zero for hill-climbing solution transitions.

3. In the limit, the stationary probability distribution must have zero probability mass for every non-globally optimal solution.

4. The probability of escaping from any locally (but not globally) optimal solution must not approach zero too quickly.

Anily and Federgruen (1987) uses condition (3) to relax the acceptance function multiplicative condition (9). However, in practice, condition (3) would be very difficult to check without assuming that (9) holds. Condition (4) provides the necessary condition for the rate that the probability of hill-climbing transitions approaches zero. Condition (4) is expressed quantitatively as follows: let t_k be defined by (2), and define the minimum one-step acceptance probability as

$$\underline{a}_{t_k} = \min_{\substack{\omega \in \Omega, \\ \omega' \in N(\omega)}} a_{t_k}(\omega, \omega') \tag{28}$$

Define the set of local optima $L \subset \Omega$ such that $\omega \in L$ implies that $f(\omega) \le f(\omega')$ for all $\omega' \in N(\omega)$, and let

$$\bar{a}_{t_k} = \max_{\substack{\omega \in L, \\ \omega' \in N(\omega) \backslash L}} a_{t_k}(\omega, \omega') \tag{29}$$

Finally, let any solution $\omega' \in \Omega$, be reachable from any solution $\omega \in \Omega$, in q transitions or less. Then if (non-globally) locally optimal solutions exist,

$$\sum_{k=1}^{\infty} \left(\underline{a}_{t_k}\right)^q = +\infty \tag{30}$$

and conditions (1), (2), and (3) hold, then the simulated annealing algorithm will asymptotically converge to the set of global optima with probability one. However, if (non-globally) locally optimal solutions exist and

$$\sum_{k=1}^{\infty} \bar{a}_{t_k} < +\infty \tag{31}$$

then the probability of each solution is asymptotically dependent upon the initial solution. Therefore, the simulated annealing algorithm will not always converge to the set of global optima with probability one. Johnson and Jacobson (2002b) relaxes the sufficient conditions found in Anily and Federgruen (1987) by using a path argument between global optima and local (but not global) optima.

Yao and Li (1991) and Yao (1995) also discuss simulated annealing algorithms with general acceptance probabilities, though their primary contribution is with respect to general neighborhood generation distributions. Schuur (1997) provides a description of acceptance functions ensuring the convergence of the associated simulated annealing algorithm to the set of global optima.

The inhomogeneous proof concept is stronger than the homogeneous approach in that it provides necessary conditions for the rate of convergence, but its asymptotic nature suggests that practical implementation may not be feasible. Romeo and Sangiovanni-Vincentelli (1991) notes "there is no reason to believe that truncating the logarithmic temperature sequence would yield a good configuration, since the tail of the sequence is the essential ingredient in the proof." In addition, the logarithmic cooling schedule dictates a very slow rate of convergence. Therefore, most recent work has focused on methods of improving simulated annealing's finite-time behavior and modifying or blending the algorithm with other search methods such as genetic algorithms (Liepins and Hilliard, 1989), tabu search (Glover, 1994), or both (Fox, 1993).

3 RELATIONSHIP TO OTHER LOCAL SEARCH ALGORITHMS

The hill-climbing strategy inherent in simulated annealing has lead to the formulation of other such algorithms (e.g., threshold accepting, the noising method). Moreover, though different in how they traverse the solution space, both tabu search and genetic algorithm share with simulated annealing the objective of using local information to find global optima over solution spaces contaminated with multiple local optima.

3.1 Threshold Accepting

Questioning the very need for a randomized acceptance function, Dueck and Scheuer (1990), and independently, Moscato and Fontanari (1990) propose the threshold accepting algorithm, where the acceptance function is defined as

$$a_k(\Delta_{\omega,\omega'}) = \begin{cases} 1 & \text{if } Q_k \geq \Delta_{\omega,\omega'} \\ 0 & \text{otherwise} \end{cases}$$

with Q_k defined as the threshold value at iteration k. Q_k is typically set to be a deterministic, non-increasing step function in k. Dueck and Scheuer (1990) reports computational results that suggest dramatic improvements in traveling salesman problem solution quality and algorithm run-time over basic simulated annealing. Moscato and Fontanari (1990) reports more conservative results—they conjecture that simulated annealing's probabilistic acceptance function does not play a major role in the search for near-optimal solutions.

Althofer and Koschnick (1991) develops a convergence theory for threshold accepting based on the concept that simulated annealing belongs to the convex hull of threshold

accepting. The idea presented in Althofer and Koschnick (1991) is that (for a finite Q_k threshold sequence) there can exist only finitely many threshold accepting transition matrices; but simulated annealing can have infinitely many transition matrices because of the real-valued nature of the temperature at each iteration. However, every simulated annealing transition matrix for a given problem can be represented as a convex combination of the finitely many threshold accepting transition matrices. Althofer and Koschnick (1991) is unable to prove that threshold accepting will asymptotically reach a global minimum, but it does prove the existence of threshold schedules that provide convergence to within an ε-neighborhood of the optimal solutions. Jacobson and Yücesan (2002a) proves that if the threshold value approaches zero as k approaches infinity, then the algorithm does not converge in probability to the set of globally optimal solutions.

Hu et al. (1995) modifies threshold accepting to include a non-monotonic, self-tuning threshold schedule in the hope of improving the algorithm's finite-time performance. Hu et al. (1995) allows the threshold Q_k to change dynamically (either up or down), based on the perceived likelihood of being near a local minimum. These changes are accomplished using a principle they call *dwindling expectation*—when the algorithm fails to move to neighboring solutions, the threshold Q_k is gradually increased, in the hope of eventually escaping a local optimum. Conversely, when solution transitions are successful, the threshold is reduced, in order to explore local optima. The experimental results based on two traveling salesman problems presented in Hu et al. (1995) showed that the proposed algorithm outperformed previous hill-climbing methods in terms of finding good solutions earlier in the optimization process.

Threshold accepting's advantages over simulated annealing lie in its ease of implementation and its generally faster execution time, due to the reduced computational effort in avoiding acceptance probability computations and generation of random numbers (Moscato and Fontanari, 1990). However, compared to simulated annealing, relatively few threshold accepting applications are reported in the literature (Lin et al., 1995; Scheermesser and Bryngdahl, 1995; Nissen and Paul, 1995).

3.2   Noising Method

Charon and Hudry (1993) advocates a simple descent algorithm called the *noising method.* The algorithm first perturbs the solution space by adding random noise to the problem's objective function values. The noise is gradually reduced to zero during the algorithm's execution, allowing the original problem structure to reappear. Charon and Hudry (1993) provides computational results, but does not prove that the algorithm will asymptotically converge to the set of globally optimal solutions. Charon and Hudry (2001) shows how the noising method is a generalization of simulated annealing and threshold accepting.

Storer et al. (1992) proposes an optimization strategy for sequencing problems, by integrating fast, problem-specific heuristics with local search. Its key contribution is to base the definition of the search neighborhood on a heuristic problem pair (h,p), where h is a fast, known, problem-specific heuristic and p represents the problem data. By perturbing the heuristic, the problem, or both, a neighborhood of solutions is developed. This neighborhood then forms the basis for local search. The hope is that the perturbations will cluster good solutions close together, thus making it easier to perform local search.

3.3 Tabu Search

Tabu search (Glover, 1994) is a general framework for a variety of iterative local search strategies for discrete optimization. Tabu search uses the concept of *memory* by controlling the algorithm's execution via a dynamic list of forbidden moves. This allows the tabu search algorithm to intensify or diversify its search of a given problem's solution space in an effort to avoid entrapment in local optima. Note that a criticism of simulated annealing is that it is completely memoryless (i.e., simulated annealing disregards all historical information gathered during the algorithm's execution). On the other hand, no proofs of convergence exist in the literature for the general tabu search algorithm.

Faigle and Kern (1992) proposes a particular tabu search algorithm called *probabilistic tabu search* (Glover, 1989) as a meta-heuristic to help guide simulated annealing. Probabilistic tabu search attempts to capitalize on both the asymptotic optimality of simulated annealing and the memory feature of tabu search. In probabilistic tabu search, the probabilities of generating and accepting each candidate solution are set as functions of both a temperature parameter (as in simulated annealing) and information gained in previous iterations (as for tabu search). Faigle and Kern (1992) are then able to prove asymptotic convergence of their particular tabu search algorithm by using methods developed for simulated annealing (Faigle and Kern, 1991; Faigle and Schraeder, 1988). Note that the results of Faigle and Kern (1992) build upon Glover (1989) where probabilistic tabu search was first introduced and contrasted with simulated annealing.

3.4 Genetic Algorithms

Genetic algorithms (Liepens and Hilliard, 1989) emulate the evolutionary behavior of biological systems. They generate a sequence of populations of qcandidate solutions to the underlying optimization problem by using a set of genetically inspired stochastic solution transition operators to transform each population of candidate solutions into a descendent population. The three most popular transition operators are reproduction, cross-over, and mutation (Davis, 1991). Davis and Principe (1991) and Rudolph (1994) attempt to use homogeneous finite Markov chain techniques to prove convergence of genetic algorithms (Cerf, 1998), but are unable to develop a theory comparable in scope to that of simulated annealing.

Mühlenbein (1997) presents a theoretical analysis of genetic algorithms based on population genetics. He counters the popular notion that models that mimic natural phenomenon are superior to other models. The article argues that evolutionary algorithms can be inspired by nature, but do not necessarily have to copy a natural phenomenon. He addresses the behavior of transition operators and designs new genetic operators that are not necessarily related to events in nature, yet still perform well in practice.

One criticism of simulated annealing is the slow speed at which it sometimes converges. Delport (1998) combines simulated annealing with evolutionary algorithms to improve performance in terms of speed and quality of solution. He improves this hybrid system of simulated annealing and evolutionary selection by improving the cooling schedule based on fast recognition of the thermal equilibrium in terms of selection intensity. This technique results in much faster convergence of the algorithm.

Sullivan and Jacobson (2000) links genetic algorithms with simulated annealing using generalized hill climbing algorithms (Jacobson et al., 1998). They first link genetic algorithms to ordinal hill climbing algorithms, which can then be used, through

its formulation within the generalized hill climbing algorithm framework, to form a bridge with simulated annealing. Though genetic algorithms have proven to be effective for addressing intractable discrete optimization problems, and can be classified as a type of hill-climbing approach, its link with generalized hill climbing algorithms (through the ordinal hill climbing formulation) provides a means to establish well-defined relationships with other generalized hill climbing algorithms (like simulated annealing and threshold accepting). They also present two formulations of genetic algorithms that provide a first step towards developing a bridge between genetic algorithms and other local search strategies like simulated annealing.

4   PRACTICAL GUIDELINES

Implementation issues for simulated annealing can follow one of two paths—that of specifying problem-specific choices (neighborhood, objective function, and constraints), and that of specifying generic choices (generation and acceptance probability functions, and the cooling schedule) (Eglese, 1990). The principal shortcoming of simulated annealing is that it often requires extensive computer time. Implementation modifications generally strive to retain simulated annealing's asymptotic convergence character, but at reduced computer run-time. The methods discussed here are mostly heuristic.

Problem-Specific Choices
Objective Functions   One problem-specific choice involves the objective function specification. Stern (1992) recommends a heuristic temperature-dependent penalty function as a substitute for the actual objective function for problems where low cost solutions have neighbors of much higher cost, or in cases of degeneracy (i.e., large neighborhoods of solutions of equal, but high costs). The original objective function surfaces, as the penalty and the temperature are gradually reduced to zero. This technique is similar to the noising method presented by Charon and Hudrey (1993), where the penalty function is described as noise and is reduced at each outer loop iteration of the algorithm. One speed-up technique is to evaluate only the difference in objective functions, $\Delta_{\omega,\omega'}$ instead of calculating both $f(\omega)$ and $f(\omega')$. Tovey (1988) suggests several methods of approximating $\Delta_{\omega,\omega'}$ by using surrogate functions (that are faster to evaluate than $\Delta_{\omega,\omega'}$, but not as accurate) probabilistically for cases when evaluation of $\Delta_{\omega,\omega'}$ is expensive; this technique is referred to as the *surrogate function swindle*.

Straub et al. (1995) improves the performance of simulated annealing on problems in chemical physics by using the classical density distribution instead of the molecular dynamics of single point particles to describe the potential energy landscape. Ma and Straub (1994) reports that using this distribution has the effect of smoothing the energy landscape by reducing both the number and depth of local minima.

Yan and Mukai (1992) considers the case when a closed-form formula for the objective function is not available. They use a probabilistic simulation (termed the stochastic ruler method) to generate a sample objective function value for an input solution, and then accept the solution if the sample objective function value falls within a predetermined bound. They also provide a proof of asymptotic convergence by extrapolating the convergence proofs for simulated annealing, and analyze the rate of convergence.

Generic Choices

Generation Probability Functions   Generation probability functions are usually chosen as uniform distributions with probabilities proportional to the size of the neighborhood. The generation probability function is typically not temperature-dependent. Fox (1993) suggests that instead of *blindly* generating neighbors uniformly, adopt an *intelligent* generation mechanism that modifies the neighborhood and its probability distribution to accommodate search intensification or diversification, in the same spirit of the tabu search meta-heuristic. Fox (1993) also notes that simulated annealing convergence theory does not preclude this idea. Tovey (1988) suggests an approach with a similar effect, called the *neighborhood prejudice swindle.*

Acceptance Probability Functions   The literature reports considerable experimentation with acceptance probability functions for hill-climbing transitions. The most popular is the exponential form (1). Ogbu and Smith (1990) considers replacing the basic simulated annealing acceptance function $a_k(\Delta_{\omega,\omega'})$ with a geometrically decreasing form that is independent of the change in objective function value. They adopt a *probabilistic-exhaustive* heuristic technique in which randomly chosen neighbors of a solution are examined and all solutions that are accepted are noted, but only the last solution accepted becomes the new incumbent. The hope is that this scheme will explore a broader area of the solution space of a problem. Their acceptance probability function is defined for all solutions $\omega, \omega' \in \Omega$ and for $k = 1, 2, \ldots, K$ as

$$a_k(\omega, \omega') = a_k = \begin{cases} a_1 x^{k-1} & \text{if } f(\omega') > f(\omega) \\ 1 & \text{otherwise} \end{cases}$$

where a_1 is the initial acceptance probability value, $x \in (0,1)$ is a reducing factor, and K is the number of stages (equivalent to a temperature cooling schedule). They also experiment with this method (and a neighborhood of large cardinality) on a permutation flow shop problem, and reports that its approach found comparable solutions to the basic simulated annealing algorithm in one-third the computation time.

4.1   Choice of Cooling Schedule

The simulated annealing cooling schedule is fully defined by an initial temperature, a schedule for reducing/changing the temperature, and a stopping criterion. Romeo and Sangiovanni-Vincentelli (1991) notes that an effective cooling schedule is essential to reducing the amount of time required by the algorithm to find an optimal solution. Therefore much of the literature on cooling schedules (e.g., Cardoso et al., 1994; Fox and Heine, 1993; Nourani and Andersen, 1998; and Cohn and Fielding, 1999) is devoted to this topic.

   Homogeneous simulated annealing convergence theory has been used to design effective cooling schedules. Romeo and Sangiovanni-Vincentelli (1991) suggests the following procedure for designing a cooling schedule:

   1. Start with an initial temperature t_0 for which a good approximation of the stationary distribution π_{t_0} is quickly reached.

2. Reduce t_0 by an amount $\delta(t)$ small enough such that π_{t_0} is a good starting point to approximate $\pi_{t_0-\delta(t)}$.

3. Fix the temperature at a constant value during the iterations needed for the solution distribution to approximate $\pi_{t_0-\delta(t)}$.

Repeat the above process of cooling and iterating until no further improvement seems possible.

Cooling schedules are grouped into two classes: *static* schedules, which must be completely specified before the algorithm begins; and *adaptive* schedules, which adjust the temperature's rate of decrease from information obtained during the algorithm's execution. Cooling schedules are almost always heuristic; they seek to balance moderate execution time with simulated annealing's dependence on asymptotic behavior.

Strenski and Kirkpatrick (1991) presents an exact (non-heuristic) characterization of finite-length annealing schedules. They consider extremely small problems that represent features (local optima and smooth/hilly topologies), and solve for solution probabilities after a finite number of iterations to gain insights into some popular assumptions and intuition behind cooling schedules. Their experiments suggest that optimal cooling schedules are *not* monotone decreasing in temperature. They also show that for the test problem (a white noise surface), geometric and linear cooling schedules perform better than inverse logarithmic cooling schedules, when sufficient computing effort is allowed. Moreover, their experiments do not show measurable performance differences between linear and geometric cooling schedules. They also observe that geometric cooling schedules are not greatly affected by excessively high initial temperatures. The results presented suggest that the even the most robust adaptive cooling schedule "produces annealing trajectories which are never in equilibrium" (Strenski and Kirkpatrick, 1991). However, they also conclude that the transition acceptance rate is not sensitive to the degree of closeness to the equilibrium distribution.

Christoph and Hoffmann (1993) also attempts to characterize optimal cooling schedules. They derive a relationship between a finite sequence of optimal temperature values (i.e., outer loops) and the number of iterations (i.e., inner loops) at each respective temperature for several small test problems to reach optimality (i.e., the minimal mean final energy). They find that this *scaling behavior* is of the form

$$x_m = a_m v^{-b_m} \qquad (32)$$

where a and b are scaling coefficients, $x_m = e^{-1/t_k}$ is referred to as the temperature, v is the number of inner loop iterations at temperature x_m, and m is the number of outer loops at which the temperature x_m is reduced. The proposed approach is to solve for the coefficients a and b based on known temperature and iteration parameter values for an optimal schedule based on several replications of the algorithm using $(m \times v)$ iterations for each replication, and then use (32) to interpolate the optimal cooling schedule for intermediate iterations. They however do not make any suggestions on how to efficiently solve for the necessary optimal cooling schedules for a (typically large) problem instance.

Romeo and Sangiovanni-Vincentelli (1991) presents a theoretical framework for evaluating the performance of the simulated annealing algorithm. They discuss annealing schedules in terms of initial temperature T, the number of inner loops for each value of T, the rate that the temperature T decreases (i.e., cooling schedule) and the criteria for stopping the algorithm. They conclude that the theoretical results obtained thus far

have not been able to explain why simulated annealing is so successful even when a diverse collection of static cooling schedule heuristics is used. Many heuristic methods are available in the literature to find optimal cooling schedules, but the effectiveness of these schedules can only be compared through experimentation. They conjecture that the neighborhood and the corresponding topology of the objective function are responsible for the behavior of the algorithm.

Conn and Fielding (1999) conducts a detailed analysis of various cooling schedules and how they affect the performance of simulated annealing. Convergent simulated annealing algorithms are often too slow in practice, whereas a number of non-convergent algorithms may be preferred for good finite-time performance. They analyze various cooling schedules and present cases where repeated independent runs using a non-convergent cooling schedule provide acceptable results in practice. They provide examples of when it is both practically and theoretically justified to use a very high, fixed temperature, or even fast cooling schedules which have a small probability of reaching global minima and apply these cooling schedules to traveling salesman problems of various sizes. Fielding (2000) computationally studies fixed temperature cooling schedules for the traveling salesman problem, the quadratic assignment problem, and the graph partitioning problem, and demonstrates that a fixed temperature cooling schedule can yield superior results in locating optimal and near-optimal solutions. Orosz and Jacobson (2002a,b) present finite-time performance measures for simulated annealing with fixed temperature cooling schedules. They illustrate their measures using randomly generated instances of the traveling salesman problem.

Another approach to increasing the speed of simulated annealing is to implement a two-staged simulated annealing algorithm. In two-staged simulated annealing algorithms, a fast heuristic is used to replace simulated annealing at higher temperatures, with a traditional simulated annealing algorithm implemented at lower temperatures to improve on the fast heuristic solution. In addition to implementing an intelligent cooling schedule, finding the initial temperature t_0 to initialize the traditional simulated annealing algorithm is important to the success of the two-staged algorithm. Varanelli and Cohoon (1999) proposes a method for determining an initial temperature t_0 for two-staged simulated annealing algorithms using traditional cooling schedules. They note that if t_0 is too low at the beginning of the traditional simulated annealing phase, the algorithm can get trapped in an inferior solution, while if the initial temperature t_0 is too high, the algorithm can waste too many iterations (hence computing time) by accepting too many hill-climbing moves.

4.2   Choice of Neighborhood

A key problem-specific choice concerns the neighborhood function definition. The efficiency of simulated annealing is highly influenced by the neighborhood function used (Moscato, 1993). The choice of neighborhood serves to enforce a topology—Eglese (1990) reports that "a neighborhood structure which imposes a 'smooth' topology where the local minima are shallow is preferred to a 'bumpy' topology where there are many deep local minima." Solla et al. (1986) and Fleischer and Jacobson (1999) report similar conclusions. This also supports the result in Hajek (1988) that shows that asymptotic convergence to the set of global optima depends on the depth of the local minima.

Another factor to consider when choosing neighborhood functions is the neighborhood size. No theoretical results are available, other than the necessity of reachability (in a finite number of steps) from any solution to any other solution. Cheh et al. (1991) reports that small neighborhoods are best, while Ogbu and Smith (1990) provides evidence that larger neighborhoods result in better simulated annealing performance. Goldstein and Waterman (1988) conjectures that if the neighborhood size is small compared to the total solution space cardinality, then the Markov chain cannot move around the solution space fast enough to find the minimum in a reasonable time. On the other hand, a very large neighborhood has the algorithm merely sampling randomly from a large portion of the solution space, and thus, is unable to focus on specific areas of the solution space. It is reasonable to believe that neighborhood size is heavily problem-specific. For example, problems where the smoothness of its solution space topology is moderately insensitive to different neighborhood definitions may benefit from larger neighborhood sizes.

Fleischer (1993) and Fleischer and Jacobson (1999) use concepts from information theory to show that the neighborhood structure can affect the *information rate* or total uncertainty associated with simulated annealing. Fleischer (1993) shows that simulated annealing tends to perform better as the entropy level of the associated Markov chain increases, and thus conjectures that an entropy measure could be useful for predicting when simulated annealing would perform well on a given problem. However, efficient ways of estimating the entropy are needed to make this a practical tool.

Another issue on neighborhood function definition addresses the solution space itself. Chardaire et al. (1995) proposes a method for addressing 0–1 optimization problems, in which the solution space is progressively reduced by fixing the value of *strongly persistent* variables (which have the same value in all optimal solutions). They isolate the persistent variables during simulated annealing's execution by periodically estimating the expectation of the random variable (a vector of binary elements) that describes the current solution, and fixing the value of those elements in the random variable that meet threshold criteria.

4.3   Domains—Types of Problems with Examples

Simulated annealing has developed into a popular tool for optimization in the last decade. It has been used to address numerous discrete optimization problems as well as continuous variable problems. Several application and computational survey articles have been published on simulated annealing. Johnson et al. (1989, 1991) present a series of articles on simulated annealing applied to certain well-studied discrete optimization problems. The first in the series of articles uses the graph partitioning problem to illustrate simulated annealing and highlight the effectiveness of several modifications to the basic simulated annealing algorithm. The second in the series focuses on applying lessons learned from the first article to the graph coloring and number partitioning problems. Local optimization techniques were previously thought to be unacceptable approaches to these two problems. Johnson et al. (1991) also observes that for long run lengths, simulated annealing outperforms the traditional techniques used to solve graph coloring problems. However, simulated annealing did not compare well with traditional techniques on the number partitioning problem except for small problem instances. The third article in the series (not yet published) uses simulated annealing to approach the well-known traveling salesman problem.

Koulamas et al. (1994) focuses specifically on simulated annealing applied to applications in production/operations management and operations research. They discuss traditional problems such as single machine, flow shop and job shop scheduling, lot sizing, and traveling salesman problems as well as non-traditional problems to include graph coloring and number partitioning. They conclude that simulated annealing is an effective tool for solving many problems in operations research and that the degree of accuracy that the algorithm achieves can be controlled by the practitioner, in terms of number of iterations and neighborhood functions (i.e., an increased number of iterations (outer loops) combined with increased number of searches at each iteration (inner loops) can result in solutions with a higher probability of converging to the optimal solution). Fleischer (1995) discusses simulated annealing from a historical and evolutionary point of view in terms of solving difficult optimization problems. He summarizes on-going research and presents an application of simulated annealing to graph problems.

The simulated annealing algorithm has proved to be a good technique for solving difficult discrete optimization problems. In engineering optimization, simulated annealing has emerged as an alternative tool to address problems that are difficult to solve by conventional mathematical programming techniques. The algorithm's major disadvantage is that solving a complex system problem may be an extremely slow, albeit convergent process, using much more processor time than some conventional algorithms. Consequently, simulated annealing has not been widely embraced as an optimization algorithm for engineering problems. Attempts have been made to improve the performance of the algorithm either by reducing the annealing length or changing the generation and the acceptance mechanisms. However, these faster schemes, in general, do not inherit the property of escaping local minima. A more efficient way to reduce the processor time and make simulated annealing a more attractive alternative for engineering problems is to add parallelism (e.g., see Hamma et al., 2000). However, the implementation and efficiency of parallel simulated annealing algorithms are typically problem-dependent. Leite et al. (1999) considers the evaluation of parallel schemes for engineering problems where the solution spaces may be very complex and highly constrained, with function evaluations varying from medium to high cost. In addition, they provide guidelines for selecting appropriate schemes for engineering problems. They also present an engineering problem with relatively low fitness evaluation cost and strong time constraints to demonstrate the lower bounds of applicability of parallel schemes.

Many signal processing applications create optimization problems with multimodal and non-smooth cost functions. Gradient methods are ineffective in these situations because of multiple local minima and the requirement to calculate gradients. Chen and Luk (1999) proposes an adaptive simulated annealing algorithm as a viable optimization tool for addressing such difficult non-linear optimization problems. The adaptive simulated annealing algorithm maintains the advantages of simulated annealing, but converges faster. Chen and Luk demonstrate the effectiveness of adaptive simulated annealing with three signal processing applications: maximum likelihood joint channel and data estimation, infinite-impulse-response filter design and evaluation of minimum symbol-error-rate decision feedback equalizer. They conclude that the adaptive simulated annealing algorithm is a powerful global optimization tool for solving such signal processing problems.

Abramson et al. (1999) describes the use of simulated annealing for solving the school time tabling problem. They use the scheduling problem to highlight the performance of six different cooling schedules: the basic geometric cooling schedule, a scheme that uses multiple cooling rates, geometric reheating, enhanced geometric reheating, non-monotonic cooling, and reheating as a function of cost. The basic geometric cooling schedule found in van Laarhoven and Aarts (1987) is used as the baseline schedule for comparison purposes. Experimental results suggest that using multiple cooling rates for a given problem yields better quality solutions in less time than the solutions produced by a single cooling schedule. They conclude that the cooling scheme that uses the phase transition temperature (i.e., when sub-parts of the combinatorial optimization problem are solved) in combination with the best solution to date produces the best results.

Emden-Weinert and Proksch (1999) presents a study of a simulated annealing algorithm for the airline crew-pairing problem based on an algorithm run-cutting formulation. They found that the algorithm run-time can be decreased and solution quality can be improved by using a problem-specific initial solution, relaxing constraints, combining simulated annealing with a problem-specific local improvement heuristic, and by conducting multiple independent runs.

There is no question that simulated annealing can demand significant computational time to reach global minima. Recent attempts to use parallel computing schemes to speed up simulated annealing have provided promising results. Chu et al. (1999) presents a new, efficient, and highly general-purpose parallel optimization method based upon simulated annealing that does not depend on the structure of the optimization problem being addressed. Their algorithm was used to analyze a network of interacting genes that control embryonic development and other fundamental biological processes. They use a two-stage procedure which monitors and pools performance statistics obtained simultaneously from all processors and then mixes states at intervals to maintain a Boltzman-like distribution of costs. They demonstrate that their parallel simulated annealing approach leads to nearly optimal parallel efficiency for certain optimization problems. In particular, the approach is appropriate when the relative effort required to compute the cost function is large compared to the relative communication effort between parallel machines for pooling statistics and mixing states.

Alrefaei and Andradottir (1999) presents a modified simulated annealing algorithm with a constant temperature to address discrete optimization problems and use two approaches to estimate an optimal solution to the problem. One approach estimates the optimal solution based on the state most visited versus the state last visited, while the other approach uses the best average estimated objective function value to estimate the optimal solution. Both approaches are guaranteed to converge almost surely to the set of global optimal solutions under mild conditions. They compare performance of the modified simulated annealing algorithm to other forms of simulated annealing used to solve discrete optimization problems.

Creating effective neighborhood functions or neighborhood generation mechanisms is a critical element in designing efficient simulated annealing algorithms for discrete optimization problems. Tian et al. (1999) investigates the application of simulated annealing to discrete optimization problems with a permutation property, such as the traveling salesman problem, the flow-shop scheduling problem, and the quadratic assignment problems. They focus on the neighborhood function

of the discrete optimization problem and in particular the generation mechanism for the algorithm used to address the problem. They introduce six types of perturbation scheme for generating random permutation solutions and prove that each scheme satisfies asymptotic convergence requirements. The results of the experimental evaluations on the traveling salesman problem, the flow-shop scheduling problem, and the quadratic assignment problem suggest that the efficiencies of the perturbation schemes are different for each problem type and solution space. They conclude that with the proper perturbation scheme, simulated annealing produces efficient solutions to different discrete optimization problems that possess a permutation property.

Research continues to focus on the idea of simulated annealing applied to optimization of continuous functions. Continuous global optimization is defined as the problem of finding points on a bounded subset of $\Re^n$ where some real valued function f assumes its optimal (maximal or minimal) value. Application of simulated annealing to continuous optimization generally falls into two classes. The first approach closely follows the original idea presented by Kirkpatrick et al. (1983) in that the algorithm mimics the physical annealing process. The second approach describes the annealing process with Langevin equations, where the global minimum is found by solving stochastic differential equations (see Aluffi-Pentini et al., 1985). Geman and Hwang (1986) proves that continuous optimization algorithms based on Langevin equations converge to the global optima. Dekkers and Aarts (1991) proposes a third stochastic approach to address global optimization based on simulated annealing. Their approach is very similar to the formulation of simulated annealing as applied to discrete optimization problems. They extend the mathematical formulation of simulated annealing to continuous optimization problems, and prove asymptotic convergence to the set of global optima based on the equilibrium distribution of Markov chains. They also discuss an implementation of the proposed algorithm and compares its performance with other well-known algorithms on a standard set of test functions from the literature.

5   FUTURE DIRECTIONS

5.1   Generalized Hill Climbing Algorithms

Generalized Hill Climbing algorithms (GHC) (Jacobson et al., 1998) provide a framework for modeling local search algorithms used to address intractable discrete optimization problems. All generalized hill climbing algorithms have the same basic structure, but can be tailored to a specific instance of a problem by changing the hill-climbing random variable (which is used to accept or reject inferior solutions) and neighborhood functions. Generalized hill climbing algorithms are described in pseudo-code form:

Select an initial solution $\omega \in \Omega$
Set the outer loop counter bound K and the inner loop counter bounds
$\quad$ $M_k, k = 1, 2, \ldots, K$
Define a set of hill-climbing (random) variables $R_k : \Omega \times \Omega \rightarrow \Re \cup \{-\infty, +\infty\}$,
$\quad$ $k = 1, 2, \ldots, K$

Set the iteration indices $k = m = 1$

> *Repeat while $k \leq K$*
> > *Repeat while $m \leq \mathrm{M}_k$*
> > Generate a solution $\omega' \in \mathrm{N}(\omega)$
> > Calculate $\Delta_{\omega,\omega'} = f(\omega') - f(\omega)$
> > If $\mathrm{R}_k(\omega, \omega') \geq \Delta_{\omega,\omega'}$, then $\omega \leftarrow \omega'$
> > If $\mathrm{R}_k(\omega, \omega') < \Delta_{\omega,\omega'}$, then $\omega \leftarrow \omega$
> > $m \leftarrow m + 1$
> > *Until $m = \mathrm{M}_k$*
> > $m \leftarrow 1, k \leftarrow k + 1$
> *Until $k = K$*

Note that the outer and inner loop bounds, K and $\mathrm{M}_k, k = 1, 2, \ldots, K$, respectively, may all be fixed, or K can be fixed with the $\mathrm{M}_k, k = 1, 2, \ldots, K$, defined as random variables whose values are determined by the solution at the end of each set of inner loop iterations satisfying some property (e.g., the solution is a local optima).

Generalized hill climbing algorithms can be viewed as sampling procedures over the solution space Ω. The key distinction between different generalized hill climbing algorithm formulations is in how the sampling is performed. For example, simulated annealing produces biased samples, guided by the neighborhood function, the objective function, and the temperature parameter. More specifically, simulated annealing can be described as a generalized hill climbing algorithm by setting the hill-climbing random variable, $\mathrm{R}_k(\omega, \omega') = -t_k \ln(\mu_k), \omega \in \Omega, \omega' \in \mathrm{N}(\omega), k = 1, 2, \ldots, K$, and the $\{\mu_k\}$ are independent and identically distributed $U(0,1)$ random variables. To formulate Monte Carlo search as a generalized hill climbing algorithm, set $\mathrm{R}_k(\omega, \omega') = +\infty, \omega \in \Omega, \omega' \in \mathrm{N}(\omega), k = 1, 2, \ldots, K$. Deterministic local search accepts only neighbors of improving (lower) objective function value and can be expressed as a generalized hill climbing algorithm with $\mathrm{R}_k(\omega, \omega') = 0, \omega \in \Omega, \omega' \in \mathrm{N}(\omega), k = 1, 2, \ldots, K$. Other algorithms that can be described using the generalized hill climbing framework include threshold accepting (1990) some simple forms of tabu search (1997), and Weibull accepting. Jacobson et al. (1998), Sullivan and Jacobson (2001), and Johnson and Jacobson (2002b) provide a complete discussion of these algorithms and a description of how these algorithms fit into the generalized hill climbing algorithm framework.

5.2   Convergence versus Finite-Time Performance

The current literature focuses mainly on asymptotic convergence properties of simulated annealing algorithms (Section 2 outlines and describes several of these results); however, considerable work on finite-time behavior of simulated annealing has been presented over the past decade. Chiang and Chow (1989) and Mazza (1992) investigate the statistical properties of the first visit time to a global optima which provides insight into the time-asymptotic properties of the algorithm as the outer loop counter, $k \to +\infty$. Catoni (1996) investigates optimizing a finite-horizon cooling schedule to maximize the number of visits to a global optimum after a finite number of iterations. Desai (1999) focuses on finite-time performance by incorporating size-asymptotic information supplied by certain eigenvalues associated with the transition matrix. Desai provides some quantitative and qualitative information about the performance of simulated annealing after a finite number of steps by observing the quality of solutions related to the number of steps that the algorithm has taken.

Srichander (1995) examines the finite-time performance of simulated annealing using spectral decomposition of matrices. He proposes that an annealing schedule on the temperature is not necessary for the final solution of the simulated annealing algorithm to converge to the global minimum with probability one. Srichander shows that initiating the simulated annealing algorithm with high initial temperatures produces an inefficient algorithm in the sense that the number of function evaluations required to obtain a global minima is very large. A modified simulated annealing algorithm is presented with a low initial temperature and an iterative schedule on the size of the neighborhood sets that leads to a more efficient algorithm. The new algorithm is applied to a real-world example and performance is reported.

Fleischer and Jacobson (1999) uses a reverse approach to establish theoretical relationships between the finite-time performance of an algorithm and the characteristics of problem instances. They observe that the configuration space created by an instance of a discrete optimization problem determines the efficiency of simulated annealing when applied to that problem. The entropy of the Markov chain embodying simulated annealing is introduced as a measure that captures the entire topology of the configuration space associated with the particular instance of the discrete optimization problem. By observing the expected value of the final state in a simulated annealing algorithm as it relates to the entropy value of the underlying Markov chain, the article presents measures of performance that determine how well the simulated annealing algorithm performs in finite-time. Their computational results suggest that superior finite-time performance of a simulated annealing algorithm are associated with higher entropy measures.

5.3   Extensions

The popularity of simulated annealing has spawned several new annealing-like algorithms. Pepper et al. (2000) introduce demon algorithms and test them on the traveling salesman problem. Ohlmann and Bean (2000) introduce another variant of simulated annealing termed compressed annealing. They incorporate the concepts of pressure and volume, in addition to temperature, to address discrete optimization problems with relaxed constraints. They also introduce a primal/dual meta-heuristic by simultaneously adjusting temperature and pressure in the algorithm.

Much work continues in the area of convergence and comparing the performance of simulated annealing algorithms to other local search strategies. Jacobson and Yücesan (2002b) presents necessary and sufficient (asymptotic) convergence conditions for generalized hill climbing algorithms that include simulated annealing as a special case. They also introduce new performance measures that can be used to evaluate and compare both convergent and non-convergent generalized hill climbing algorithms with random restart local search (Jacobson, 2002). Such a comparison provides insights into both asymptotic and finite-time performance of discrete optimization algorithms. For example, they use the global visit probability to evaluate the performance of simulated annealing using random restart local search as a benchmark. These results suggest that random restart local search may outperform simulated annealing provided that a sufficiently large number of restarts are executed. Ferreira and Zerovnik (1993) develop bounds on the probability that simulated annealing obtains an optimal (or near-optimal) solution, and use these bound to obtain similar results for random restart local search and simulated annealing. Fox (1994) notes that this result is only true if both the number

of accepted and rejected moves are counted. Fox (1994) also provides a clever example to illustrate this point, and notes that comparing random restart local search and simulating annealing may not be prudent. Fox (1993, 1995) presents modifications of simulated annealing that circumvent this counting issue, hence yielding superior performing simulated annealing algorithm implementations. The primary value of using simulated annealing may therefore be for finite-time executions that obtain near-optimal solutions reasonably quickly. This, in turn, suggests that one should focus on the finite-time behavior of simulated annealing rather than the asymptotic convergence results that dominate the literature.

ACKNOWLEDGEMENTS

BIBLIOGRAPHY

Aarts, E.H.L. and Korst, J. (1989) *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing.* John Wiley & Sons, Chichester, England.

Aarts, E.H.L. and Lenstra, J.K. (1997) *Local Search in Combinatorial Optimization.* John Wiley & Sons, Chichester, England.

Aarts, E.H.L. and van Laarhoven, P.J.M. (1985) Statistical cooling: A general approach to combinatorial optimization problems. *Phillips Journal of Research,* **40**, 193–226.

Abramson, D., Krishnamoorthy, M. and Dang, H. (1999) Simulated annealing cooling schedules for the school timetabling problem. *Asia-Pacific Journal of Operational Research,* **16**, 1–22.

Alrefaei, M.H. and Andradottir, S. (1999) A simulated annealing algorithm with constant temperature for discrete stochastic optimization. *Management Science,* **45**, 748–764.

Althofer, I. and Koschnick, K.U. (1991) On the convergence of threshold accepting. *Applied Mathematics and Optimization,* **24**, 183–195.

Aluffi-Pentini, F., Parisi, V. and Zirilli, F. (1985) Global optimization and stochastic differential equations. *Journal of Optimization Theory and Applications,* **47**, 1–16.

Anily, S. and Federgruen, A. (1987) Simulated annealing methods with general acceptance probabilities. *Journal of Applied Probability,* **24**, 657–667.

Belisle, C.J.P. (1992) Convergence theorems for a class of simulated annealing algorithms on R^D. *Journal of Applied Probability,* **29**, 885–895.

Belisle, C.J.P, Romeijn, H.E. and Smith, R.L. (1993) Hit-and-run algorithms for generating multivariate distributions. *Mathematics of Operations Research,* **18**, 255–266.

Bohachevsky, I.O., Johnson, M.E. and Stein, M.L. (1986) Generalized simulated annealing for function optimization. *Technometrics,* **28**, 209–217.

Borkar, V.S. (1992) Pathwise recurrence orders and simulated annealing. *Journal of Applied Probability,* **29**, 472–476.

Bratley, P., Fox, B.L. and Schrage, L. (1987) *A Guide to Simulation,* Springer-Verlag, New York.

Cardoso, M.F., Salcedo, R.L. and de Azevedo, S.F. (1994) Nonequilibrium simulated annealing: a faster approach to combinatorial minimization. *Industrial Engineering and Chemical Research,* **33**, 1908–1918.

Catoni, O. (1996) Metropolis, simulated annealing, and iterated energy transformation algorithms: theory and experiments. *Journal of Complexity,* **12**, 595–623.

Cerf, R. (1998) Asymptotic convergence of genetic algorithms. *Advances in Applied Probability,* **30**, 521–550.

Chardaire, P., Lutton, J.L. and Sutter, A. (1995) Thermostatistical persistency: a powerful improving concept for simulated annealing algorithms. *European Journal of Operational Research,* **86**, 565–579.

Charon, I. and Hudry, O. (1993) The noising method—a new method for combinatorial optimization. *Operations Research Letters,* **14**, 133–137.

Charon, I. and Hudry, O. (2001) The noising methods—a generalization of some metaheuristics. *European Journal of Operational Research,* **135**, 86–101.

Cheh, K.M., Goldberg, J.B. and Askin, R.G. (1991) A note on the effect of neighborhood-structure in simulated annealing. *Computers and Operations Research,* **18**, 537–547.

Chen, S. and Luk, B.L. (1999) Adaptive simulated annealing for optimization in signal processing applications. *Signal Processing,* **79**, 117–128.

Chiang, T.S. and Chow, Y.S. (1988) On the convergence rate of annealing processes. *SIAM Journal on Control and Optimization,* **26**, 1455–1470.

Chiang, T.S. and Chow, Y.Y. (1989) A limit-theorem for a class of inhomogeneous markov-processes. *Annals of Probability,* **17**, 1483–1502.

Chiang, T.S. and Chow, Y.Y. (1994) The asymptotic-behavior of simulated annealing processes with absorption. *SIAM Journal on Control and Optimization,* **32**, 1247–1265.

Christoph, M. and Hoffmann, K.H. (1993) Scaling behavior of optimal simulated annealing schedules. *Journal of Physics A—Mathematical and General,* **26**, 3267–3277.

Chu, K.W., Deng, Y.F. and Reinitz, J. (1999) Parallel simulated annealing by mixing of states. *Journal of Computational Physics,* **148**, 646–662.

Çinlar, E. (1974) *Introduction to Stochastic Processes.* Prentice-Hall, Englewood Cliffs, New Jersey.

Cohn, H. and Fielding, M. (1999) Simulated annealing: searching for an optimal temperature schedule. *SIAM Journal on Optimization,* **9***, 779–802.

Connors, D.P. and Kumar, P.R. (1989) Simulated annealing type markov-chains and their order balance-equations. *SIAM Journal on Control and Optimization,* **27**, 1440–1461.

Davis, T.E. (1991) *Toward an Extrapolation of the Simulated Annealing Convergence Theory onto the Simple Genetic Algorithm* (Doctoral Dissertation), University of Florida, Gainesville, Florida.

Davis, T.E. and Principe, J.C. (1991) A simulated annealing like convergence theory for the simple genetic algorithm. In: *Fourth Conference on Genetic Algorithm,* pp. 174–181.

Dekkers, A. and Aarts, E. (1991) Global optimization and simulated annealing. *Mathematical Programming,* **50**, 367–393.

Delport, V. (1998) Parallel simulated annealing and evolutionary selection for combinatorial optimisation. *Electronics Letters,* **34**, 758–759.

Desai, M.P. (1999) Some results characterizing the finite time behaviour of the simulated annealing algorithm. *Sadhana-Academy Proceedings in Engineering Sciences,* **24**, 317–337.

Dueck, G. and Scheuer, T. (1990) Threshold accepting—a general-purpose optimization algorithm appearing superior to simulated annealing. *Journal of Computational Physics,* **90**, 161–175.

Eglese, R.W. (1990) Simulated annealing: a tool for operational research. *European Journal of Operational Research,* **46**, 271–281.

Emden-Weinert, T. and Proksch, M. (1999) Best practice simulated annealing for the airline crew scheduling problem. *Journal of Heuristics,* **5**, 419–436.

Fabian, V. (1997) Simulated annealing simulated. *Computers and Mathematics with Applications,* **33**, 81–94.

Faigle, U. and Kern, W. (1991) Note on the convergence of simulated annealing algorithms. *SIAM Journal on Control and Optimization,* **29**, 153–159.

Faigle, U. and Kern, W. (1992) Some convergence results for probabilistic tabu search. *ORSA Journal on Computing,* **4**, 32–37.

Faigle, U. and Schrader, R. (1988) On the convergence of stationary distributions in simulated annealing algorithms. *Information Processing Letters,* **27**, 189–194.

Faigle, U. and Schrader, R. (1988) Simulated annealing-—a case-study. *Angewandte Informatik,* 259–263.

Fielding, M. (2000) Simulated annealing with an optimal fixed temperature. *SIAM Journal of Optimization,* **11**, 289–307.

Fleischer, M.A. (1995) *Assessing the Performance of the Simulated Annealing Algorithm Using Information Theory* (Doctoral Dissertation), Department of Operations Research, Case Western Reserve University, Clevelend, Ohio.

Fleischer, M.A. (1995) Simulated annealing: past, present, and future. In: C. Alexopoulos, K. Kang, W.R. Lilegdon and D. Goldsman (eds.), *Proceedings of the 1995 Winter Simulation Conference,* IEEE Press, pp. 155–161.

Fleischer, M.A. (1999) Generalized cybernetic optimization: solving continuous variable problems. In: S. Voss, S. Martello, C. Roucairol, H. Ibrahim, and I.H. Osman (eds.), *Meta-heuristics: Advances and Trends in Local Search Paradigms for Optimization,* Kluwer Academic Publishers, pp. 403–418.

Fleischer, M.A. and Jacobson, S.H. (1996) Cybernetic optimization by simulated annealing: an implementation of parallel processing using probabilistic feedback control. In: I.H. Osman and J.P. Kelly (eds.), *Meta-Heuristics: Theory and Applications,* Kluwer Academic Publishers, pp. 249–264.

Fleischer, M.A. and Jacobson, S.H. (1999) Information theory and the finite-time behavior of the simulated annealing algorithm: experimental results. *INFORMS Journal on Computing,* **11,** 35–43.

Fox, B.L. (1993) Integrating and accelerating tabu search, simulated annealing, and genetic algorithms. *Annals of Operations Research,* **41,** 47–67.

Fox, B.L. (1994) Random restarting versus simulated annealing. *Computers and Mathematics with Applications,* **27**, 33–35.

Fox, B.L. (1995) Faster simulated annealing. *Siam Journal of Optimzation,* **5**, 485–505.

Fox, B.L. and Heine, G.W. (1993) Simulated annealing with overrides, technical, Department of Mathematics, University of Colorado, Denver, Colorado.

Gemen, S. and Hwang, C.R. (1986) Diffusions for global optimization. *SIAM Journal on Control and Optimization,* **24,** 1031–1043.

Gidas, B. (1985) Nonstationary markov chains and convergence of the annealing algorithm. *Journal of Statistical Physics,* **39,** 73–131.

Glover, F. (1989) Tabu search—Part I. *ORSA Journal on Computing,* **1**, 190–206.

Glover, F. (1994) Tabu search for nonlinear and parametric optimization (with links to genetic algorithms). *Discrete Applied Mathematics,* **49,** 231–255.

Glover, F. and Laguna, M. (1997) *Tabu Search,* Kluwer Academic Publishers, Boston, Massachusetts.

Goldstein, L. and Waterman, M. (1988) Neighborhood size in the simulated annealing algorithm. *American Journal of Mathematical and Management Sciences,* **8**, 409–423.

Granville, V., Krivanek, M. and Rasson, J.P. (1994) Simulated annealing—a proof of convergence. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **16,** 652–656.

Hajek, B. (1988) Cooling schedules for optimal annealing. *Mathematics of Operations Research,* **13,** 311–329.

Hamma, B., Viitanen, S. and Torn, A. (2000) Parallel continuous simulated annealing for global optimization. *Optimization Methods and Software,* **13,** 95–116.

Hammersley, J.M. and Handscomb, D.C. (1964) *Monte Carlo Methods,* Methuen, John Wiley & Sons, London, New York.

Hu, T.C., Kahing, A.B. and Tsao, C.W.A. (1995) Old bachelor acceptance: a new class of non-monotone threshold accepting methods. *ORSA Journal on Computing,* **7**, 417–425.

Isaacson, D.L. and Madsen, R.W. (1976) *Markov Chains, Theory and Applications.* John Wiley & Sons, New York.

Jacobson, S.H. (2002) Analyzing the performance of local search algorithms using generalized hill climbing algorithms, pp. 441–467. (Chapter 20 in *Essays and Surveys*

on Metaheuristics, P. Hansen and C.C. Ribeiro (eds.), Kluwer Academic Publishers, Norwell, Massachusetts.

Jacobson, S.H., Sullivan, K.A. and Johnson, A.W. (1998) Discrete manufacturing process design optimization using computer simulation and generalized hill climbing algorithms. *Engineering Optimization,* 31, 247–260.

Jacobson, S.H. and Yücesan, E. (2002a) A performance measure for generalized hill climbing algorithms. Technical Report, Department of Mechanical and Industrial Engineering, University of Illinois, Urbana, Illinois.

Jacobson, S.H. and Yücesan, E. (2002b) On the effectiveness of generalized hill climbing algorithms. Technical Report, Department of Mechanical and Industrial Engineering, University of Illinois, Urbana, Illinois.

Johnson, A.W. and Jacobson, S.H. (2002a) A class of convergent generalized hill climbing algorithms. *Applied Mathematics and Computation,* 125(2–3), 359–373.

Johnson, A.W. and Jacobson, S.H. (2002b) On the convergence of generalized hill climbing algorithms. *Discrete Applied Mathematics* (*To Appear*).

Johnson, D.S., Aragon, C.R., McGeoch, L.A. and Schevon, C. (1989) Optimization by simulated annealing—an experimental evaluation; part 1, graph partitioning. *Operations Research,* 37, 865–892.

Johnson, D.S., Aragon, C.R., McGeoch, L.A. and Schevon, C. (1991) Optimization by simulated annealing—an experimental evaluation; part 2, graph-coloring and number partitioning. *Operations Research,* 39, 378–406.

Kiatsupaibul, S. and Smith, R.L. (2000) A general purpose simulated annealing algorithm for integer linear programming. Technical Report, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan.

Kirkpatrick, S., Gelatt, Jr., C.D. and Vecchi, M.P. (1983) Optimization by simulated annealing. *Science,* 220, 671–680.

Koulamas, C., Antony, S.R. and Jaen, R. (1994) A survey of simulated annealing applications to operations-research problems. *OMEGA-International Journal of Management Science,* 22, 41–56.

Leite, J.P.B. and Topping, B.H.V. (1999) Parallel simulated annealing for structural optimization. *Computers and Structures,* 73, 545–564.

Liepins, G.E. and Hilliard, M.R. (1989) Genetic algorithms: foundations and applications. *Annals of Operations Research,* 21, 31–58.

Lin, C.K.Y., Haley, K.B. and Sparks, C. (1995) A comparative study of both standard and adaptive versions of threshold accepting and simulated annealing algorithms in three scheduling problems. *European Journal of Operational Research,* 83, 330–346.

Locatelli, M. (1996) Convergence properties of simulated annealing for continuous global optimization. *Journal of Applied Probability,* 33, 1127–1140.

Locatelli, M. (2000) Simulated annealing algorithms for continuous global optimization: convergence conditions. *Journal of Optimization Theory and Applications,* 104, 121–133.

Lundy, M. and Mees, A. (1986) Convergence of an annealing algorithm. *Mathematical Programming,* 34, 111–124.

Ma, J. and Straub, J.E. (1994) Simulated annealing using the classical density distribution. *Journal of Chemical Physics,* **101**, 533–541.

Mazza, C. (1992) Parallel simulated annealing. *Random Structures and Algorithms,* 3, 139–148.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953) Equation of state calculations by fast computing machines. *Journal of Chemical Physics,* **21**, 1087–1092.

Meyer, C.D. (1980) The condition of a finite markov chain and perturbation bounds for the limiting probabilities. *SIAM Journal of Algebraic and Discrete Methods,* **1**, 273–283.

Mitra, D., Romeo, F. and Sangiovanni-Vincentelli, A.L. (1986) Convergence and finite time behavior of simulated annealing. *Advances in Applied Probability,* **18**, 747–771.

Moscato, P. (1993) An introduction to population approaches for optimization and hierarchical objective functions: a discussion on the role of tabu search. *Annals of Operations Research,* **41**, 85–121.

Moscato, P. and Fontanari, J.F. (1990) Convergence and finite-time behavior of simulated annealing. *Advances in Applied Probability,* **18**, 747–771.

Muhlenbein, H. (1997) Genetic algorithms. In: E, Aarts and J.K. Lenstra (eds.), *Local Search in Combinatorial Optimization.* John Wiley & Sons, New York, New York, pp. 137–172.

Nissen, V. and Paul, H. (1995) A modification of threshold accepting and its application to the quadratic assignment problem. *OR Spektrum,* **17**, 205–210.

Nourani, Y. and Andresen, B. (1998) A comparison of simulated annealing cooling strategies. *Journal of Physics A—Mathematical and General,* **31**, 8373–8385.

Ogbu, F.A. and Smith, D.K. (1990) The application of the simulated annealing algorithm to the solution of the N/M/Cmax flowshop problem. *Computers and Operations Research,* **17**, 243–253.

Ohlmann, J.W. and Bean, J.C. (2000) Compressed annealing: simulated annealing under pressure. Technical Report, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan.

Orosz, J.E. and Jacobson, S.H. (2002a) Finite-time performance analysis of static simulated annealing algorithms. *Computational Optimization and Applications,* **21**, 21–53.

Orosz, J.E. and Jacobson, S.H. (2002b) Analysis of static simulated annealing algorithms. *Journal of Optimization Theory and Application* (to appear).

Pepper, J.W., Golden, B.L. and Wasil, E.A. (2000) Solving the traveling salesman problem with demon algorithms and variants. Technical Report, Smith School of Business, University of Maryland, College Park, Maryland.

Romeijn, H.E., Zabinsky, Z.B., Graesser, D.L. and Noegi, S. (1999) New reflection generator for simulated annealing in mixed-integer/continuous global optimization. *Journal of Optimization Theory and Applications,* **101**, 403–427.

Romeo, F. and Sangiovanni-Vincentelli, A. (1991) A theoretical framework for simulated annealing. *Algorithmica,* **6**, 302–345.

Rosenthal, J.S. (1995) Convergence rates for markov chains. *SIAM Review,* **37**, 387–405.

Ross, S.M. (1996) *Stochastic Processes.* John Wiley & Sons, New York, New York.

Ross, S.M. (1997) *Introduction to Probability Models.* Academic Press, San Diego, California.

Rossier, Y., Troyon, M. and Liebling, T.M. (1986) Probabilistic exchange algorithms and euclidean traveling salesman problems. *OR Spektrum,* **8**, 151–164.

Rudolph, G. (1994) Convergence analysis of cononical genetic algorithms. *IEEE Transactions on Neural Networks, Special Issue on Evolutional Computing,* **5**, 96–101.

Scheermesser, T. and Bryngdahl, O. (1995) Threshold accepting for constrained half-toning. *Optics Communications,* **115**, 13–18.

Schuur, PC. (1997) Classification of acceptance criteria for the simulated annealing algorithm. *Mathematics of Operations Research,* **22**, 266–275.

Seneta, E. (1981) *Non-Negative Matrices and Markov Chains.* Springer-Verlag, New York, New York.

Siarry, P., Berthiau, G., Durbin, F. and Haussy, J. (1997) Enhanced simulated annealing for globally minimizing functions of many-continuous variables. *ACM Transactions On Mathematical Software,* **23**, 209–228.

Solla, S.A., Sorkin, G.B. and White, S.R. (1986) Configuration space analysis for optimization problems. *NATO ASI Series, Disordered Systems and Biological Organization,* **F20**, 283–293.

Srichander, R. (1995) Efficient schedules for simulated annealing. *Engineering Optimization,* **24**, 161–176.

Stern, J.M. (1992) Simulated annealing with a temperature dependent penalty function. *ORSA Journal on Computing,* **4**, 311–319.

Storer, R.H., Wu, S.D. and Vaccari, R. (1992) New search spaces for sequencing problems with application to job shop scheduling. *Management Science,* **38**, 1495–1509.

Straub, J.E., Ma, J. and Amara, P. (1995) Simulated annealing using coarse grained classical dynamics: smouuchowski dynamics in the gaussian density approximation. *Journal of Chemical Physics,* **103**, 1574–1581.

Strenski, P.N. and Kirkpatrick, S. (1991) Analysis of finite length annealing schedules. *Algorithmica,* **6**, 346–366.

Sullivan, K.A. and Jacobson, S.H. (2000) Ordinal hill climbing algorithms for discrete manufacturing process design optimization problems. *Discrete Event Dynamical Systems,* **10**, 307–324.

Sullivan, K.A. and Jacobson, S.H. (2001) A convergence analysis of generalized hill climbing algorithms. *IEEE Transactions on Automatic Control,* **46**, 1288–1293.

Tian, P., Ma, J. and Zhang, D.M. (1999) Application of the simulated annealing algorithm to the combinatorial optimisation problem with permutation property: an investigation of generation mechanism. *European Journal of Operational Research,* **118**, 81–94.

Tovey, C.A. (1988) Simulated simulated annealing. *American Journal of Mathematical and Management Sciences,* **8**, 389–407.

Tsitsiklis, J.N. (1989) Markov chains with rare transitions and simulated annealing. *Mathematics of Operations Research,* **14,** 70–90.

van Laarhoven, P.J.M. (1988) *Theoretical and Computational Aspects of Simulated Annealing,* Centrum voor Wiskunde en Informatica, Amsterdam, Netherlands.

van Laarhoven, P.J.M. and Aarts, E.H.L. (1987) *Simulated Annealing: Theory and Applications,* D. Reidel, Kluwer Academic Publishers, Dordrecht, Boston, Norwell, Massachusetts.

Varanelli, J.M. and Cohoon, J.P. (1999) A fast method for generalized starting temperature determination in homogeneous two-stage simulated annealing systems. *Computers and Operations Research,* **26,** 481–503.

Yan, D. and Mukai, H. (1992) Stochastic discrete optimization. *SIAM Journal on Control and Optimization,* **30,** 594–612.

Yang, R.L. (2000) Convergence of the simulated annealing algorithm for continuous global optimization. *Journal of Optimization Theory and Applications,* **104,** 691–716.

Yao, X. (1995) A new simulated annealing algorithm. *International Journal of Computer Mathematics,* **56**, 161–168.

Yao, X. and Li, G. (1991) General simulated annealing. *Journal of Computer Science and Technology,* **6**, 329–338.

Zabinsky, Z.B., Smith, R.L., McDonald, J.F., Romeijn, H.E. and Kaufman, D.E. (1993) Improving hit-and-run for global optimization. *Journal of Global Optimization,* **3**, 171–192.