Statistical Science
1993, Vol. 8, No. 1, 10–15

Simulated Annealing

Dimitris Bertsimas and John Tsitsiklis

Abstract. Simulated annealing is a probabilistic method proposed in Kirkpatrick, Gelett and Vecchi (1983) and Cerny (1985) for finding the global minimum of a cost function that may possess several local minima. It works by emulating the physical process whereby a solid is slowly cooled so that when eventually its structure is "frozen," this happens at a minimum energy configuration.

We restrict ourselves to the case of a cost function defined on a finite set. Extensions of simulated annealing to the case of functions defined on continuous sets have also been introduced in the literature (e.g., Geman and Hwang, 1986; Gidas, 1985a; Holley, Kusuoka and Stroock, 1989; Jeng and Woods, 1990; Kushner, 1985). Our goal in this review is to describe the method, its convergence and its behavior in applications.

Key words and phrases: Markov chains, randomized algorithms, simulated annealing.

1. THE METHOD

The basic elements of simulated annealing (SA) are the following:

1. A finite set S.
2. A real-valued cost function J defined on S. Let $S^* \subset S$ be the set of global minima of the function J, assumed to be a proper subset of S.
3. For each $i \in S$, a set $S(i) \subset S - \{i\}$, called the set of neighbors of i.
4. For every i, a collection of positive coefficients q_{ij}, $j \in S(i)$, such that $\Sigma_{j \in S(i)} q_{ij} = 1$. It is assumed that $j \in S(i)$ if and only if $i \in S(j)$.
5. A nonincreasing function T: $N \to (0, \infty)$, called the *cooling schedule*. Here N is the set of positive integers, and $T(t)$ is called the *temperature* at time t.
6. An initial "state" $x(0) \in S$.

Given the above elements, the SA algorithm consists of a discrete-time inhomogeneous Markov chain $x(t)$, whose evolution we now describe. If the current state $x(t)$ is equal to i, choose a neighbor j of i at random; the probability that any particular $j \in S(i)$ is selected is equal to q_{ij}. Once j is chosen, the next state $x(t+1)$ is determined as follows:

If $J(j) \le J(i)$, then $x(t+1) = j$.
If $J(j) > J(i)$, then
$x(t+1) = j$
with probability $\exp[-(J(j) - J(i))/T(t)]$
$x(t+1) = i$ otherwise.

Formally,

$$P[x(t+1) = j \mid x(t) = i] = q_{ij} \exp\left[-\frac{1}{T(t)} \max\{0, J(j) - J(i)\}\right] \quad \text{if } j \neq i, j \in S(i). \tag{1}$$

If $j \neq i$ and $j \notin S(i)$, then $P[x(t+1) = j \mid x(t) = i] = 0$.

The rationale behind the SA algorithm is best understood by considering a homogeneous Markov chain $x_T(t)$ in which the temperature $T(t)$ is held at a constant value T. Let us assume that the Markov chain $x_T(t)$ is irreducible and aperiodic and that $q_{ij} = q_{ji}$ for all i, j. Then $x_T(t)$ is a reversible Markov chain, and its invariant probability distribution is given by

$$\pi_T(i) = \frac{1}{Z_T} \exp\left[-\frac{J(i)}{T}\right], \qquad i \in S, \tag{2}$$

where Z_T is a normalizing constant. (This is easily shown by verifying that the detailed balance equations hold.) It is then evident that as $T \downarrow 0$, the probability distribution π_T is concentrated on the set S^* of global minima of J. This latter property remains valid if the condition $q_{ij} = q_{ji}$ is relaxed (Faigle and Kern, 1989).

The probability distribution (2), known as the Gibbs distribution, plays an important role in statistical mechanics. In fact, statistical physicists have been interested in generating a sample element of S, drawn

Dimitris Bertsimas is Associate Professor at the Sloan School of Management and John Tsitsiklis is Associate Professor in the Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

according to the probability distribution π_T. This is accomplished by simulating the Markov chain $x_T(t)$ until it reaches equilibrium, and this method is known as the Metropolis algorithm (Metropolis et al., 1953). In the optimization context, we can generate an optimal element of S with high probability if we produce a random sample according to the distribution π_T, with T very small. One difficulty with this approach is that when T is very small, the time it takes for the Markov chain $x_T(t)$ to reach equilibrium can be excessive. The SA algorithm tries to remedy this drawback by using a slowly decreasing "cooling schedule" $T(t)$.

The SA algorithm can also be viewed as a local search algorithm in which (unlike the usual deterministic local search algorithms) there are occasional "upward" moves that lead to a cost increase. One hopes that such upward moves will help escape from local minima.

2. CONVERGENCE ANALYSIS

2.1 Basic Results

Having defined the algorithm, we now address its performance. The main questions are:

1. Does $x(t)$ "converge" to the optimal set S^*?
2. How fast does the convergence to S^* take place?

The first question has been more or less completely answered, and we will now survey the main results. Much less is known about the second, as we will discuss later.

From now on, we assume that for some fixed T (and therefore for all T) the Markov chain $x_T(t)$ is irreducible and aperiodic. We say that the SA algorithm converges if $\lim_{t\to\infty} P[x(t) \in S^*] = 1$. (Note that this is convergence in probability, rather than almost sure convergence.) A fair amount of work has been concerned with finding necessary and sufficient conditions for convergence, in the above sense. The main result, due to Hajek, is presented next, following some definitions.

THEOREM 1 (Hajek, 1988). *We say that state i communicates with S^* at height h if there exists a path in S (with each element of the path being a neighbor of the preceding element) that starts at i and ends at some element of S^* and such that the largest value of J along the path is $J(i) + h$. Let d^* be the smallest number such that every $i \in S$ communicates with S^* at height d^*. Then, the SA algorithm converges if and only if $\lim_{t\to\infty} T(t) = 0$ and*

$$\sum_{t=1}^{\infty} \exp[-d^*/T(t)] = \infty. \tag{3}$$

The most popular cooling schedules (in theory, at least) are of the form

$$T(t) = \frac{d}{\log t}, \tag{4}$$

where d is some positive constant. Theorem 1 states that SA converges if and only if $d \geq d^*$.

The constant d^* is a measure of the difficulty for $x(t)$ to escape from a local minimum and go from a nonoptimal state to S^*. We are primarily interested in problems where $d^* > 0$, which will be assumed from now on. Such problems have local minima that are not optimal. Some understanding of Theorem 1 is provided by the following argument. Consider a local minimum whose "depth" is d^*. The SA makes an infinite number of trials to escape from it, and the probability of success at each trial is of the order of $\exp[-d^*/T(t)]$. Then condition (3) amounts (by the Borel-Cantelli lemma) to saying that an infinite number of these trials will be successful. Indeed, the proof in Hajek (1988) proceeds by estimating the statistics of the exit times from certain neighborhoods of local minima.

Let $\pi(i; t) = P[x(t) = i]$. If $T(t)$ decreases very slowly, as is the case in (4), then $x(t)$ behaves, over fairly long time intervals, like a homogeneous Markov chain, and it can be reasonably expected that the difference between $\pi_{T(t)}(i)$ and $\pi(i; t)$ is small at all times. Indeed, one of the very first convergence proofs (Geman and Geman, 1984) was based on this idea, although the results therein were less sharp than Theorem 1.

In order to acquire more intuition about the interpretation of Theorem 1, we will carry the connection between SA and the corresponding family of homogeneous Markov chains further. For this goal we consider the cooling schedule $T(t) = d/\log t$. In general, the statistics of the Markov chain $x(t)$ under a slowly varying cooling schedule $T(t)$ remain pretty much unchanged if a related cooling schedule is used in which the temperature is held constant for fairly long time periods. In our case, the schedule $T(t) = d/\log t$ can be approximated as follows. Let $t_1 = 1$ and $t_{k+1} = t_k + \exp(kd)$. Then let $\hat{T}(t) = 1/k$, for $t_k \leq t < t_{k+1}$. Consider the kth segment $[t_k, t_{k+1}]$ of the piecewise constant schedule $\hat{T}(t)$. We are dealing with the reversible homogeneous chain $x_{1/k}(t)$, and we will now investigate how fast it reaches steady state.

We want to study the convergence of the chain $x_{1/k}(t)$. The eigenvalues of its transition-probability matrix are real. Its relaxation time is determined by its second-largest eigenvalue λ_2 for which good estimates are available, at least in the limit as $k \to \infty$ (e.g., Chiang and Chow, 1988; Holley and Stroock, 1988; and Ventcel, 1972). In particular, if the cost function J has a unique global minimum, the relaxation time is well approximated by $\exp(kd^*)$. Interestingly enough, this is the same as the constant d^* defined in Theorem 1, something that is far from obvious. This yields another interpretation of the convergence condition $d \geq d^*$ for the schedule $\hat{T}(t)$. If $d < d^*$, then at each temperature $1/k$ we are running $x_{1/k}(t)$ for a negligible fraction of its relaxation time, and this is not enough for $\pi(i; t)$ to

stay close to $\pi_T(i)$. On the other hand, if $d > d^*$, then the interval $[t_k, t_{k+1}]$ corresponds to $\exp[k(d^* - d)]$ relaxation times of $x_{1/k}(t)$, which implies that $\pi(i; t_{k+1})$ is very close to $\pi_{1/k}(i)$, as $k \to \infty$.

One can also pursue this approximation by piecewise constant schedules directly, without introducing eigenvalue estimates. The main idea is that, at low temperatures, the statistics of a homogeneous chain can be accurately bounded by viewing it as a singularly perturbed Markov chain and using rather crude large-deviation bounds (Tsitsiklis, 1988, 1989). For other convergence proofs, see Connors and Kumar (1988), Gidas (1985b), Holley and Stroock (1988), and Mitra, Romeo and Sangiovanni-Vincentelli (1986).

The convergence of SA (in the sense defined earlier in this section) is a reassuring property, but it is far from enough for SA to be a useful algorithm. We also need to know the speed of convergence. It can be shown that for any schedule $T(t) = d/\log t$, and for all t,

$$\max_{x(0)} P[x(t) \notin S^* \mid x(0)] \geq A/t^a, \tag{5}$$

where A and a are positive constants depending on the function J and the neighborhood structure. If we wish $x(t)$ to be outside S^* with probability less than ε, we need $t \geq (A/\varepsilon)^{1/a}$.

We now turn our attention to the practical relevance of some of the convergence results. From a more practical perspective, while the algorithm is being run, one should keep track of the best state i encountered so far and its associated cost $J(i)$. If the algorithm is to be run for t^* time steps, we are not really interested in the value of $P(x(t^*) \notin S^*)$. Rather, we are interested in the probability that no state in S^* is visited during the execution of the algorithm. Given a cooling schedule of the form $T(t) = d/\log t$, with $d > d^*$, it can be shown that this probability is at most $A/(t^*)^a$ for some positive constants A and a. It vanishes as $t \to \infty$, which might seem encouraging. On the other hand, if the temperature is fixed at any positive value (or even at infinity, corresponding to a random walk), the probability that no state in S^* is visited in t^* time units is at most $Be^{-t^*/b}$ for suitable positive constants B and b. So, for very large times, the performance of a random walk would seem to be better than the performance guarantees of SA. The key point is that the above analyses, based on time asymptotics, involve extremely large constants and are largely irrelevant. Indeed the constants in the above estimates are often much larger than the cardinality of the state space S. In particular, the above analysis cannot even establish that SA is preferable to exhaustive search.

2.2 Taking the Instance Size into Account

One can trace the inadequacy of the analytical methods mentioned so far to the fact that they deal with one instance at a time. A more realistic approach would be to consider a family of instances, together with a notion of instance size, and then study the statistics of the first hitting time of the set S^* as a function of the instance size. (We refer to such results as *complexity results*.) This would then provide a meaningful yardstick for comparing SA to alternative methods. As an extension, if we are only interested in approximately optimal solutions to a given problem, we can define a set $\hat{S}$ of approximately optimal solutions and study the statistics of the first hitting time of $\hat{S}$. Unfortunately, such results have proved very difficult to derive. We review briefly some of the progress that has been made along this direction.

All available positive complexity results we are aware of use a schedule $T(t)$ in which the temperature is held constant in time, although it may depend on the instance size. The only nontrivial combinatorial problem for which SA has been rigorously analyzed is the maximum matching problem. It has been shown by Sasaki and Hajek (1988) that, for bounded-degree graphs with n nodes, the expected time until SA hits an optimal configuration is $O(n^5)$. This is slower than other available algorithms, but the result is encouraging. On the other hand, no algorithm of the SA type (even with time-varying temperature) can solve the matching problem in polynomial expected time when the degree bound is relaxed (see Sasaki and Hajek, 1988).

Hajek (1985) studied an example in which the state space is $S = \{1, 2, 3\}^N$ and the cost of a typical state $s = (s_1, \ldots, s_N)$ is equal to $\sum_{i=1}^N V(s_i)$. This corresponds to a minimization in N identical uncoupled problems, and the optimal solution is evident. It is argued that if the temperature is suitably chosen as a function of N, the expected first hitting time of the global minimum grows polynomially with N. This is better than exhaustive search, because the cardinality of S is 3^N. It is also better than the "local search with multistart" method, whereby a state in S is chosen repeatedly at random and each time a local search (pure descent) algorithm is run until a local minimum is reached (Hajek, 1985). Of course, in this example a much more efficient algorithm exists. What is encouraging is that SA does not really use the information that the cost is separable, that is, that there is no coupling between s_i and s_j. For this reason, one may hope that SA would run in polynomial time even if the problem was perturbed by introducing some sufficiently weak coupling between components. No such extensions are currently available.

One reason why SA works well in the preceding example is that if a given state is far from optimal, then there exists a large number of paths that lead to another state with lower cost. The probability that the state $x(t)$ escapes from a local minimum of depth d along any particular path, and in a single trial, is at most $\exp(-d/T)$. On the other hand, if the number of

candidate paths is very large, the probability of escape is substantial. It is here that the combinatorial structure of the state space S and its neighborhood system should become significant.

In interesting combinatorial settings the state space is usually exponential in the instance size. So constant-temperature SA would work in polynomial time if the relaxation time were polylogarithmic in the size of the state space ("rapid mixing"). There are a number of available positive results on rapid mixing in Markov chains, but they deal mostly with the case where $T = \infty$, corresponding to a random walk. (This is a much simpler special case, because the cost function J has no effect.) Unfortunately, SA becomes interesting at the opposite end, when T is very small. Proving rapid mixing for large SA Markov chains at small temperatures is a challenging task.

One class of problems for which there is some hope of obtaining positive complexity results arises in the context of image processing. Here, we have an $N \times N$ grid. To each gridpoint (i, j), we associate a variable s_{ij} taking values in a finite set A. We thus obtain a configuration space $S = A^{N^2}$. Many image-processing and pattern-recognition problems lead to a cost function $J : S \to \mathbf{R}$ of the form

$$J(s) = \sum_{(i,j)} f_{ij}(s_{ij}) + \sum_{(i,j),(k,\ell)} g_{ij,k\ell}(s_{ij}, s_{k\ell}),$$

where $g_{ij,k\ell}$ is identically zero unless (i, j) and (k, ℓ) are neighboring gridpoints. Starting with Geman and Geman (1984), SA has become a very popular method for such problems. Here one defines two states (configurations) to be neighbors if they differ only at a single gridpoint. Note that when a configuration change is contemplated (i.e., a change of some s_{ij}), the cost difference (which determines the probability of accepting the change) depends only on the gridpoints neighboring (i, j). For this reason, the evolution of the configuration can be viewed as the time evolution of a Markov random field. The relaxation times of Markov random fields have been extensively studied (e.g., Liggett, 1988), but under rather special assumptions on the functions f_{ij} and $g_{ij,k\ell}$. Thus, the available results are not yet applicable to the cost functions that arise in image processing.

As far as theory is concerned, there is at present a definite lack of rigorous results justifying the use of SA. Even if SA is accepted, there are no convincing theoretical arguments favoring the use of time-varying (decreasing) cooling schedules, as opposed to the use of a constant temperature. [This latter question is partially addressed in Hajek and Sasaki (1989).]

3. BEHAVIOR IN PRACTICE

Despite the lack of a rigorous theoretical justification of its speed of convergence, researchers have used SA extensively in the last decade. There are numerous papers discussing applications of SA to various problems. We have already mentioned that SA is extensively used in image processing. In order to give an indication of its performance, we will review some of the work concerning the application of SA to combinatorial optimization problems.

In a comprehensive study of SA, Johnson et al. (1990, 1991, 1992) discuss the performance of SA on four problems: the traveling salesman problem (TSP), graph partitioning problem (GPP), graph coloring problem (GCP) and number partitioning problem (NPP). Johnson et al. apply SA to these NP-hard problems using a cooling schedule in which the temperature decreases geometrically, namely, $T(t + 1) = rT(t)$. In general, the performance of SA was mixed: in some problems, it outperformed the best known heuristics for these problems, and, in other cases, specialized heuristics performed better. More specifically:

1. In the graph partitioning problem, a graph $G = (V, E)$ is given, and we are asked to partition the set of vertices V into two subsets V_1 and V_2 so as to minimize

$$|X| + \lambda(|V_1| - |V_2|)^2,$$

where X is the set of edges joining a node in V_1 with a node in V_2, and λ is a weighting factor. For the GPP, SA obtains final solutions that are at best some 5 percent better than those obtained by the best of the more traditional algorithms [e.g., the Kernighan and Lin (1970) heuristic] if the latter are allowed the same amount of computation time as SA. For sparse graphs, SA was better than repeated applications of the Kernighan-Lin heuristic, which is based on ideas of local optimization, whereas for some structured graphs the Kernighan-Lin heuristic was better.
2. In the graph coloring problem, a graph $G = (V, E)$ is given, and we are asked to partition the set of vertices into a minimal number of subsets, such that no edge has both endpoints in the same subset. For the graph coloring problem, SA produces final solutions that are competitive with those obtained by a tailored heuristic [the one by Johri and Matula (1982)], which is considered the best one for this problem. However, computation times for SA are considerably longer than those of the specialized heuristic.
3. For the traveling salesman problem, SA consistently outperforms solutions found by repeated application of iterative improvement, based on 2-opt or 3-opt transitions, but it is a consistent loser when compared with the well-known algorithm of Lin and Kernighan (1973). The latter is based on k-opt transitions, and at each iteration it decides dynamically the value of k.

Another interesting point is that the choice of the

cooling schedule influences the quality of solution obtained. In van Laarhoven and Aarts (1987) the authors compare three different cooling schedules for the graph partitioning problem, and they observe that the quality of the solution found by the different cooling schedules can differ as much as 10 percent. Another observation is that the computation times can be excessive for some problems.

In addition to the above mentioned developments in image processing, SA and various alternative versions based roughly on it have been used in statistical applications. Bohachevsky, Johnson and Stein (1986) proposed a "generalized" SA procedure for continuous optimization problems and applied their method to an optimal design problem. Many researchers have considered SA as a tool in the development of optimal experimental designs. Recent examples include Currin et al. (1991), Meyer and Nachtsheim (1988) and Sacks and Schiller (1988). Variants of SA based on Bayesian ideas have been proposed by Laud, Berliner and Goel (1989) and van Laarhoven et al. (1989).

Overall, SA is a generally applicable and easy-to-implement probabilistic approximation algorithm that is able to produce good solutions for an optimization problem, even if we do not understand the structure of the problem well. We believe, however, that more research, both theoretical and experimental, is needed to assess further the potential of the method.

REFERENCES

BOHACHEVSKY, I., JOHNSON, M. E. and STEIN, M. L. (1986). Generalized simulated annealing for function optimization. *Technometrics* **28** 209–217.

CERNY, V. (1985). A thermodynamic approach to the traveling salesman problem: An efficient simulation. *J. Optim. Theory Appl.* **45** 41–51.

CHIANG, T.-S. and CHOW, Y. (1988). On eigenvalues and optimal annealing rate. *Math. Oper. Res.* **13** 508–511.

CONNORS, D. P. and KUMAR, P. R. (1988). Balance of recurrence order in time-inhomogeneous Markov chains with applications to simulated annealing. *Probability in the Engineering and Informational Sciences* **2** 157–184.

CURRIN, C., MITCHELL, T., MORRIS, M. and YLVISAKER, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *J. Amer. Statist. Assoc.* **86** 953–963.

FAIGLE, U. and KERN, W. (1989). Note on the convergence of simulated annealing algorithms. Memorandum 774, Univ. Twente.

GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs' distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6** 721–741.

GEMAN, S. and HWANG, C.-R. (1986). Diffusions for global optimization. *SIAM J. Control Optim.* **24** 1031–1043.

GIDAS, B. (1985a). Global optimization via the Langevin equations. In *Proceedings of the 24th IEEE Conference on Decision and Control* 774–778. IEEE, New York.

GIDAS, B. (1985b). Nonstationary Markov chains and convergence of the annealing algorithm. *J. Statist. Phys.* **39** 73–131.

HAJEK, B. (1985). A tutorial survey of theory and applications of simulated annealing. In *Proceedings of the 24th IEEE Conference on Decision and Control* 755–760. IEEE, New York.

HAJEK, B. (1988). Cooling schedules for optimal annealing. *Math. Oper. Res.* **13** 311–329.

HAJEK, B. and SASAKI, G. (1989). Simulated annealing—to cool or not. *Systems Control Lett.* **12** 443–447.

HOLLEY, R. A. and STROOCK, D. W. (1988). Simulated annealing via Sobolev inequalities. *Comm. Math. Phys.* **115** 553–569.

HOLLEY, R. A., KUSUOKA, S. and STROOCK, D. W. (1989). Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.* **83** 333–347.

JENG, F.-C. and WOODS, J. W. (1990). Simulated annealing in compound Gaussian random fields. *IEEE Trans. Inform. Theory* **36** 94–107.

JOHNSON, D., ARAGON, C., MCGEOCH, L. and SCHEVON, C. (1990). Optimization by simulated annealing: An experimental evaluation, Part I: Graph partitioning. *Oper. Res.* **37** 865–892.

JOHNSON, D., ARAGON, C., MCGEOCH, L. and SCHEVON, C. (1991). Optimization by simulated annealing: An experimental evaluation, Part II: Graph coloring and number partitioning. *Oper. Res.* **39** 378–406.

JOHNSON, D., ARAGON, C., MCGEOCH, L. and SCHEVON, C. (1992). Optimization by simulated annealing: An experimental evaluation, Part III: The traveling salesman problem. Unpublished manuscript.

JOHRI, A. and MATULA, D. (1982). Probabilistic bounds and heuristic algorithms for coloring large random graphs. Technical Report 82-CSE-6, Dept. Computer Science and Engineering, Southern Methodist Univ.

KERNIGHAN, B. and LIN, S. (1970). An efficient heuristic procedure for partitioning graphs. *Bell Systems Technical Journal* **49** 291–307.

KIRKPATRICK, S., GELETT, C. D. and VECCHI, M. P. (1983). Optimization by simulated annealing. *Science* **220** 621–630.

KUSHNER, H. J. (1985). Asymptotic global behavior for stochastic approximations and diffusions with slowly decreasing noise effects: Global minimization via Monte Carlo. Technical Report LCDS 85-7, Dept. Applied Mathematics, Brown Univ.

LAUD, P., BERLINER, L. M. and GOEL, P. K. (1989). A stochastic probing algorithm for global optimization. In *Computer Science and Statistics: Proceedings of the 21st Symposium on the Interface* (K. Berk and L. Malone, eds.) 220–229. Amer. Statist. Assoc., Alexandria, VA.

LIGGETT, T. (1988). *Interacting Particle Systems.* Springer, New York.

LIN, S. and KERNIGHAN, B. (1973). An effective heuristic algorithm for the traveling salesman problem. *Oper. Res.* **21** 498–516.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.

MEYER, R. K. and NACHTSHEIM, C. J. (1988). Constructing exact D-optimal experimental designs by simulated annealing. *Amer. J. Math. Management Sci.* **8** 329–359.

MITRA, D., ROMEO, F. and SANGIOVANNI-VINCENTELLI, A. (1986). Convergence and finite-time behaviour of simulated annealing. *Adv. in Appl. Probab.* **18** 747–771.

SACKS, J. and SCHILLER, S. (1988). Spatial designs. In *Statistical Decision Theory and Related Topics* **4** (S. S. Gupta and J. O. Berger, eds.). Academic, New York.

SASAKI, G. and HAJEK, B. (1988). The time complexity of maximum matching by simulated annealing. *J. Assoc. Comput. Mach.* **35** 387–403.

TSITSIKLIS, J. N. (1988). A survey of large time asymptotics of simulated annealing algorithms. In *Stochastic Differential Systems, Stochastic Control Theory and Applications* (W. Fleming and P. L. Lions, eds.) 583–599. Springer, New York.

TSITSIKLIS, J. N. (1989). Markov chains with rare transitions and simulated annealing. *Math. Oper. Res.* **14** 70–90.

VAN LAARHOVEN, P. and AARTS, E. (1987). *Simulated Annealing: Theory and Applications.* Reidel, Dordrecht.

VAN LAARHOVEN, P., BOENDER, C., AARTS, E. and RINNOOY KAN, A. (1989). A Bayesian approach to simulated annealing. *Probability in the Engineering and Informational Sciences* **3** 453–475.

VENTCEL, A. D. (1972). On the asymptotics of eigenvalues of matrices with elements of order $\exp\{-V_{ij}/(2\epsilon^2)\}$. *Dokl. Akad. Nauk SSSR.* **202** 65–68.