Applied Probability Trust (February 11, 2000)

RESTARTING SEARCH ALGORITHMS WITH APPLICATIONS TO SIMULATED ANNEALING

F. MENDIVIL, R. SHONKWILER, AND M. C. SPRUILL,* *Georgia Institute of Technology*

Abstract

Some consequences of restarting stochastic search algorithms are studied. It is shown under reasonable conditions that restarting when certain patterns occur yields probabilities that the goal state has not been found by the n^{th} epoch which converge to zero at least geometrically fast in n. These conditions are shown to hold for (RSA), restarted simulated annealing, employing a local generation matrix, a cooling schedule $T_n \sim c/n$, and restarting after a fixed number $r+1$ of duplications of energy levels of states when r is sufficiently large. For (SA), simulated annealing, with logarithmic cooling these probabilities cannot decrease to zero this fast. Numerical comparisons between (RSA) and several modern variations on (SA) are also presented and in all cases (RSA) performs better.

RANDOM SEARCH; STOCHASTIC PROCESS RENEWAL; CODING SCHEDULE; MINIMIZATION

AMS 1991 SUBJECT CLASSIFICATION: PRIMARY 60K20; 60G35; 60J20; 65K10
SECONDARY

1. Introduction

Evidence that stochastic algorithms can spend excessive time in states other than the goal comes most frequently and easily from simulations and more rarely from theoretical arguments. For example, an "optimal" cooling schedule (see Hajek (1988)) for simulated annealing (SA) guarantees that the probability the search process is in the goal state tends to 1 as the number of epochs n tends to infinity; the expected

* Postal address: School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332.

time taken by (SA) to hit the goal can however be infinite (see Shonkwiler and Van Vleck (1994) or Fox (1995)). Our main results are Theorems 3.1 and 4.1 and they show, among other things, that under certain conditions, restarted search processes, particularly simulated annealing with the proper cooling schedule, has finite expected hitting time. Restarting (SA) based on different criteria from ours has been studied numerically as in Atkinson (1992) and Nakakuki and Sadeh (1994). The undesirably slow convergence of (SA) has also motivated some research like that in Kolonko (1995) and Bélisle (1992) on the random adjustment of the cooling schedule, on non-random adjustments as reported, for example, in van Laarhoven and Aarts (1987) or Andresen and Nourani (1998), and the thorough theoretical treatment of simulating direct self-loop sequences in Fox (1995) and the truncated version in Fox and Heine (1995).

Most of the literature to date has been devoted to the investigation of the limiting probability the process is found in the goal states and sometimes (see for example Chiang and Chow (1988)) to the rate of convergence of these probabilities. Following Shonkwiler and Van Vleck (1994), we take a somewhat different approach. The best an algorithm has done up to epoch n is easily recorded and accordingly our results are stated in terms of the probabilities of not yet seeing the goal by the n^{th} epoch rather than in terms of the probability of being in the goal state at the n^{th} epoch. We call the probabilities of not yet seeing the goal by the n^{th} epoch tail probabilities (not to be confused with measure-theoretic notions of the same name as found for example in Niermiro (1995)). Theorem 3.1 provides sufficient conditions for general search processes to guarantee geometric convergence of the tail probabilities to zero under restarting after user specified patterns and Theorem 4.1 specializes these to the case of restarting simulated annealing. Of course, in both cases expected hitting times must be finite.

Fix an arbitrary stochastic algorithm for minimizing a function over the finite set C, a function f to be minimized, and let G be the goal subset determined by f. Let X_n be the candidate point in C dictated by the algorithm at epoch n for what van Laarhoven and Aarts (1987) would term this problem instance. The sequence of random variables X_n is a C-valued stochastic process and its law of motion, which of

course should depend upon f, can be quite general. The process X_n is not assumed to be Markov, just adapted to a filtration $\{\mathcal{F}_n : n \geq 1\}$. One could study alternative restarting rules but the one analyzed here involves repeated instances of energy levels $f(X_n)$ of states and hence hitting times of the vector process of $r+1$ states with values in $C^{r+1} = C \times C \times \cdots \times C$ called the r-process. Upon the event of the r-process lying in a certain fixed subset of C^{r+1} the algorithm is restarted at a randomly selected point, always using the same distribution as the one used in the initial choice.

The analysis of this scheme is shown to be amenable to techniques from discrete time renewal theory. Hu et al. (1997) had previously utilized renewal equations to analyze restarting from specific subsets of the domain C. There, they showed that under some conditions, the tail probabilities decrease geometrically to 0. Their results extended and related earlier results, based on the Perron-Frobenius theory, which Shonkwiler and Van Vleck (1994) obtained for Markov algorithms. The restarting process studied by Hu et al. is implementable in practice only when there is special knowledge of the function to be minimized and the algorithm. Restarting from an arbitrary state can be detrimental to the progress of the algorithm towards the goal and one may not know, a priori, which states are the correct ones to initiate restarting. Restarting on repeated energy values alleviates this problem.

Fox's work on self looping in (SA) and especially the truncated version in Fox and Heine (1995) is closely related to the method of restarting (SA) on the diagonal, but there are some key differences. Foremost perhaps is that Fox's calculation of the objective function at each neighbor of the points visited is not required.

Andresen and Nourani (1998) propose a constant thermodynamic speed cooling schedule for (SA) and compare its performance for solving a permanent problem with a variety of other cooling schedules. We present a comparison of our method with these several methods on a permanent problem also.

The paper is organized by section as follows. In section 2 the restarted process and the r-process are defined and the tail probabilities are shown to satisfy a renewal equation. In section 3 rate of convergence to zero of the tail probabilities is studied. Section 4 deals with simulated annealing where sufficient conditions for geometric convergence of the tail probabilities are presented. The outcomes of simulations

comparing (SA) with (RSA) are presented in section 5 along with some comments on the selection of the parameter r.

2. The restart and r-processes

Restarting when a sequence of states lies in a subset D of C^{r+1} defines a new process on the original search process. Details of this definition and the fact that the tail probabilities for the r-process satisfy a renewal equation are presented here.

We assume that the goal set G is a non-empty subset of the finite set C. Fix $r \geq 1$, let D be a subset of C^{r+1} and define for subsets A of C the sets

$$D_A = \{(x_1, x_2, \ldots, x_{r+1}) \in D : x_1 \in A\}.$$

Denote by E the set $E(t) = \big\{x \in C\backslash G : P\big[(X_n, X_{n+1}, \ldots, X_{n+r}) \in D \mid X_n = x \text{ and all histories up to } n\big] \geq t \text{ for all } n\big\} \neq \emptyset$ for some fixed $t > 0$, and let $U = G \cup E$. Here and below, by informal expressions like $P[(X_n, X_{n+1}, \ldots, X_{n+r}) \in D \mid X_n = x$ and all histories up to $n] \geq t$, we mean the more pedantic

$$P[(X_n, X_{n+1}, \ldots, X_{n+r}) \in D \mid \mathcal{F}_n] \geq t$$

on the event $X_n = x$. Introduce the following two conditions, where $T_E = \min\{n : X_n \in E\}$.

(A1) $1 > P[T_E < T_G] > 0$.

(A2) There is a finite $K \geq 1$ and a number $\phi \in (0,1)$ such that uniformly for $x \in C\backslash G$, and all n,

$$P[T_U > m + n \mid X_n = x \text{ and all histories up to } x] \leq K\phi^m,$$

where the probability on the left hand side is that the first epoch after n at which the X process lies in U is greater than $m + n$.

Let $\{X_n^{(j)} : n \geq 1\}$ be iid copies of the process X and define

$$N(j) = \min\left\{n : (X_n^{(j)}, X_{n+1}^{(j)}, \ldots, X_{n+r}^{(j)}) \in D \text{ or } X_n^{(j)} \in G\right\}$$

or $N(j) = +\infty$ if the set above is empty.

Lemma 2.1 Under (A2), $P[N(j) < \infty] = 1$.

Proof. We have

$$\begin{aligned} &P[N(1) > m \mid X_1^{(1)} = x] \\ &= \sum P[X_{m+r}^{(1)} = x(m+r), \ldots, X_3^{(1)} = x(3),\ X_2^{(1)} = x(2) \mid X_1^{(1)} = x], \end{aligned}$$

where the sum extends over $x(2), \ldots, x(m+r)$ with no segment of length $r+1$ in D. With U as above define

$$K(x(1), x(2), \ldots, x(m)) = \min\{j \geq 1 : x(j) \in U\}$$

or $+\infty$. Then

$$P[N(1) > m \mid X_1^{(1)} = x] = \sum_{j=1}^{m} p_j(x) P[N(1) > m \mid K = j] + q_m(x)$$

Certainly, since there are more paths from x at 1 to $x(m)$ at m which do not touch U than there are which don't touch U and don't have segments in D

$$q_m(x) \leq P[T_U > m \mid X_1^{(1)} = x] \leq K\phi^{m-1}$$

and since given U first hit at $j \leq m$, $N(1) > m$ entails,

$$P[N(1) > m \mid K = j] \leq 1 - t.$$

Therefore,

$$P[N(1) > m \mid X_1^{(1)} = x] \leq K\phi^{m-1} + 1 - t.$$

Choose M so large that $M > r+1$ and $K\phi^{M-1} + 1 - t = \rho < 1$. If A_j is the event of a run of length at least $r+1$ of states in C for epochs $n \in [(M+r)j + 1, (M+r)j + M]$ then conditioning shows that

$$P\left[\bigcap_{i=j+1}^{j+k} A_i^c\right] \leq \rho^k,$$

so that $P[A_j \text{ i. o.}] = 1$.

The restarted process Y_n and the r-process $\Pi(n)$ can now be well defined as follows. Let $u(0) = 0$ and for $j \geq 1$

$$u(j) = \sum_{i=1}^{j} N(i).$$

For $m \geq 1$ let

$$J(m) = \min\{j > 0 : u(j) \geq m\}.$$

The r-process is

$$\Pi(m) = \left(X_{m-u(J(m)-1)}^{(J(m))}, X_{m-u(J(m)-1)+1}^{(J(m))}, \ldots, X_{m-u(J(m)-1)+r}^{(J(m))}\right)$$

and

$$Y_m = X_{m-u(J(m)-1)}^{(J(m))}.$$

The r-process $\Pi(n)$ and restart process Y_n are perhaps best illustrated by the diagram below, where the rightmost entries in a row are in parentheses indicating that a segment of $r+1$ states lying in D has occurred there and that it is the first such run in that row. In the interests of readability of the tabular form, write $x(i,j) = X_j^{(i)}$. Here we take $r = 2$.

x(1,1)	x(1,2)	(x(1,3)	x(1,4)	x(1,5))									
			x(2,1)	x(2,2)	x(2,3)	(x(2,4)	x(2,5)	x(2,6))					
							x(3,1)	(x(3,2)	x(3,3)	x(3,4))			
									x(4,1)	x(4,2)	x(4,3)	x(4.4)	
y(1)	y(2)	y(3)	y(4)	y(5)	y(6)	y(7)	y(8)	y(9)	y(10)	y(11)	y(12)	y(13)	

One has from the realization depicted above the values of the r-process $\Pi(m)$ for $m = 1, \ldots, 4$:

$$\Pi(1) = (x(1,1), x(1,2), x(1,3)), \qquad \Pi(2) = (x(1,2), x(1,3), x(1,4))$$

$$\Pi(3) = (x(1,3), x(1,4), x(1,5)), \qquad \Pi(4) = (x(2,1), x(2,2), x(2,3))$$

Our convention throughout is that if the goal is encountered then the next r are taken as identical (and our interest in the process is terminated) and otherwise the first hitting time τ_U is defined by

$$\tau_U = \min\{n \geq 1 : \Pi(n) \in D_U\}.$$

Writing

$$u_n = P[\tau_G > n], \tag{2.1}$$

one has upon decomposition of the event $\{\tau_G > n\}$ as

$$\{\tau_G > n\} = \{\tau_U > n\} \cup (\{\tau_G > n\} \cap \{\tau_U = 1\}) \\ \cup \cdots \cup (\{\tau_G > n\} \cap \{\tau_U = n\})$$

that

$$u_n = b_n + \sum_{j=1}^{n} P[\{\tau_G > n\} \cap \{\tau_U = j\}] = b_n + \sum_{j=1}^{n} u_{n-j} f_j,$$

where $f_n = P[\tau_G > n,\, \tau_U = n]$ and $b_n = P[\tau_U > n]$. Therefore, the tail probabilities u_n for the r-process hitting times satisfy a renewal equation.

3. Finite expected hitting time and geometric decrease of tail probabilities

Rapid convergence of the tail probabilities $u_n = P[\tau_G > n]$ is established in this section. It follows from (A1) that

$$\sum_{j \geq 1} P[\tau_G > j, \tau_U = j] < 1,$$

for $\sum_{j \geq 1} P[\tau_G > j, \tau_U = j] = P[\tau_E < \tau_G]$ and the latter probability can be seen to be in $(0, 1)$ as follows. The differences in the times at which Π hits D are clearly iid and each such time difference is the result of one of precisely three types of events; either the X process has a succession of $r + 1$ states for which the Π process is in D_{U^c}, or a succession of $r + 1$ states for which Π is in D_E, or one of the states in G is hit (once) at that epoch. Let q, p_E, and p_G denote the respective probabilities of these three types of hits, with $q = 1 - p_E - p_G$. These are numbers defined by the X process. The conditional probability the stop was of type E given that the stop was of type U is

$$P[\tau_E < \tau_G] = \frac{p_E}{p_E + p_G}$$

which under the condition (A1) is clearly not zero or one.

Theorem 3.1 Under the conditions (A1) *and* (A2), $E[\tau_U] < \infty$, $E[\tau_G] = \frac{E[\tau_U]}{1 - P[\tau_E < \tau_G]} < \infty$, *and there is a* $\gamma \in (0, 1)$ *and a finite constant* c *such that* $\gamma^{-n} u_n \to c$ *as* $n \to \infty$, *with* u_n *given by* (2.1).

Proof. To begin, we prove that $1 < \liminf_{k\to\infty} f_k^{-1/k}$. Since the event $(\tau_G > n) \cap (\tau_U = n)$ implies the event $(\tau_U > n-1)$, one has $f_n \leq b_{n-1}$ for all n and $\liminf_{k\to\infty} f_k^{-1/k} \geq \liminf_{k\to\infty} b_k^{-1/k}$. We will show below that $\liminf_{k\to\infty} b_k^{-1/k} > 1$. Assuming this for the moment, it would follow that both of the power series $\Psi_b(z) = \sum_{n\geq 0} b_n z^n$ and $\Psi_f(z) = \sum_{n\geq 1} f_n z^n$ have radius of convergence greater than 1. On a sufficiently small neighborhood of 0 one has

$$\Psi_u(z) = \Psi_b(z) + \Psi_f(z)\Psi_u(z) \tag{3.1}$$

where $\Psi_u(z) = \sum_{n\geq 0} u_n z^n$ and u_n is given by (2.1). Since the power series $\Psi_f(z)$ has radius of convergence greater than 1 and $\sum_{n\geq 1} f_n < 1$, the function $\Psi_b(z)/(1-\Psi_f(z))$ is holomorphic in an open set containing the unit disc. Therefore, $\Psi_u(z)$ has radius of convergence greater than 1 and it follows that there is a $\gamma \in (0,1)$ and a finite constant c such that $\gamma^{-n} u_n \to c$ as $n \to \infty$. Furthermore, (see also Feller (1968))

$$E[\tau_G] = \sum_{n\geq 0} u_n = \frac{b}{(1-f)}.$$

To complete the proof decompose the event $(\tau_U > n)$, using $H = \min\{j : \Pi(j) \in D_{C\setminus U}\}$, as

$$(\tau_U > n) = \bigcup_{j=1}^{n} ((\tau_U > n) \cap (H = j)) \cup (H > n) \cap (\tau_U > n).$$

Then

$$\begin{aligned} P[\tau_U > n] &= \sum_{j=1}^{n} P[(\tau_U > n) \cap (H = j)] + P[(H > n) \cap (\tau_U > n)] \\ &= \sum_{j=1}^{n} P[\tau_U > n \mid (H = j) \cap (\tau_U > j)] P[(H = j) \cap (\tau_U > j)] + c_n \\ &= \sum_{j=1}^{n} P[\tau_U > n - j] k_j + c_n. \end{aligned}$$

Since $\sum_{j\geq 1} k_j = P[\Pi \text{ hits } D_{C\setminus U} \text{ before } D_U] < 1$, if we can show $\liminf_{k\to\infty} c_k^{-1/k} > 1$ then we'll have by application to $P[\tau_U > n]$ of the arguments above applied to u_n, that $\liminf_{k\to\infty} b_k^{-1/k} > 1$. But c_n is the probability of no restarts in the first n epochs so consulting Lemma 2.1, for $n = k(M+r)+d$, $0 \leq d < M+r-1$, $c_n \leq \rho^k \leq K\eta^n$, where $K = \rho^{1/(M+r)-1} > 1$ and $\eta = \rho^{1/(M+r)} < 1$.

The constant c in Theorem 3.1 may be zero and the rate of convergence could be faster than geometric. Sometimes the series $\Psi_b(z)$ and $\Psi_f(z)$ satisfy conditions under which the rate can be established precisely as follows.

Corollary 3.2 If (A1), *and* (A2) *hold, if* f_n *is not periodic and there is a real solution* $\theta > 1$ *to* $\Psi_f(\theta) = 1$ *satisfying* $\Psi_b(\theta) < \infty$ *then there is a* $\rho \in (0,1)$ *and a finite positive constant* c *such that* $\rho^{-n} u_n \to c$ *as* $n \to \infty$.

Proof. Taking $\rho = 1/\theta$, setting $v_n = \theta^n u_n$, the v_n satisfy a renewal equation

$$v_n = B_n + \sum_{j=1}^{n} v_{n-j} F_j,$$

with $B_n = \theta^n b_n$ and $F_n = \theta^n f_n$. Under aperiodicity of the f_n, Theorem XIII.10.1 of Feller (1968) yields the conclusion and the precise rate of convergence if $\Psi_b(\theta) = \sum_n B_n < \infty$.

By restarting, the expected time to goal of a search process can be transformed from infinite to finite. Multistart, (see Schoen (1991)) where under no restarting the hitting time is infinite with positive probability, is an obvious example. Under simple conditions like restarting according to a distribution which places positive mass on each state, multistart trivially satisfies the conditions (A1) and (A2) with $t = 1$. Furthermore the conditions of Corollary 3.2 hold and it provides an interesting formula for the Perron-Frobenius eigenvalue as the reciprocal of the root of a low degree polynomial (see Hu et al. (1997)). More subtle is the (SA) Example 4.3 below.

Of course it is easy to construct, as in the next example, a search process which makes a restarting scheme perform, unless it is done correctly, worse than the original algorithm.

Example 3.3 Let the state space be $\{0,1,2\}$ with 0 as the goal. Suppose for all $n \geq 2$

$$P[X_{n+1} = i \mid X_n = j, X_{n-1} = k] = \begin{cases} 1 & \text{if } i = 0 \text{ and } j = k = 1 \\ 1 & \text{if } i = 1 \text{ and } j = 1 \text{ and } k \neq 1 \\ 1 & \text{if } i = 1 \text{ and } j \neq 1 \end{cases}$$

and $P[X_2 = i \mid X_1 = j] = 1$ if $i = 1$ and 0 otherwise. Restarting whenever a repeated state occurs is clearly detrimental in this case since the original process goes deterministically directly to the goal as quickly as possible, but must repeat the state 1 before doing so. Restarting after 3 the same however does not hinder the original.

4. Application to simulated annealing

The generality of the conditions in Theorem 3.1 suggests the possibility of discovering particular algorithms or classes of algorithms exhibiting these general underlying features and hence also ones for which restarting results in the rapid decrease of the tail probabilities. One instance of this is simulated annealing. Let the generation matrix Q be fixed (see any standard reference on (SA) such as van Laarhoven and Aarts for terminology) and introduce the following assumption (A3) which we will invoke at the appropriate time. Let $\{N(x) : x \in C\}$ be a collection of subsets of C, one for each distinct x, which we call neighborhoods of x, having the property that for any x, y in C there is a finite sequence $z(1), z(2), \ldots, z(j)$ of points in C such that, setting $x = z(0)$ and $y = z(j+1)$, $z(i+1) \in N(z(i))$ for all $i = 0, 1, \ldots, j$. We assume also that $y \in N(x)$ entails $x \in N(y)$.

(A3) - The generation matrix Q allows only transitions to the immediate neighbors so that $q_{xy} = 0$ unless $y \in N(x)$, assigns positive mass to each of them so that $q_{xy} \neq 0$ for $y \in N(x) \backslash \{x\}$ and $q_{xx} = 0$.

Consider the minimization of a function f defined on the finite set C. Let

$$a = \min_{x \in C} \min_{y \in N(x)} \{f(y) - f(x) : f(y) - f(x) > 0\},$$

where $N(x)$ is the collection of immediate neighbors in C of x. Since C is finite, $\alpha > 0$. Let the acceptance rule be the usual with

(A4) the probability of accepting a transition from x to $y \neq x$ at epoch n being given by $\min\{1, e^{-(f(y)-f(x))/c(n)}\}$, where $c(n)$ is strictly decreasing to 0 with increasing n and the probability of a transition from x to x is one minus the sum of the probabilities of transitions from x to all other states.

The binary relation $x \Delta y$ if $f(x) = f(y)$ defined on $C \times C$ is an equivalence relation so

that C can be partitioned into disjoint equivalence classes $C_1, \ldots, C_k$. Furthermore, the relation $x \sim y$ defined on C_j by $x \sim y$ if x and y are in C_j and there is a path from x to y, $x = z_0$, $z_i \in N(z_{i-1})$, $i = 1, \ldots, k$, $z_k = y$, some $k < \infty$, and $z_i \in C_j$ for every i is an equivalence relation. The set C is thereby partitioned into disjoint subsets B_{ij}. Such a subset B is a *basin bottom* if for every $x \in B$ and $y \notin B$ every path $z_0, \ldots, z_k$ from x to y has $f(z_j) > f(z_{j-1})$ for some j.

In the following theorem the subset D is the set of points in C^{r+1} such that $f(x_1) = \cdots = f(x_{r+1})$.

Theorem 4.1 Under conditions (A3)–(A4) *above and if there is* $\beta > 1$ *such that*

$$\sum_{n \geq 1} \beta^n e^{-\alpha/c(n)} < \infty \tag{4.1}$$

then under restarting (SA) *by a distribution which places positive probability on each point in* C *for* r, $1 \leq r < \infty$, *sufficiently large there is a* $\gamma \in (0,1)$, *and a finite constant* c *such that* $\gamma^{-n} u_n \to c$ *as* $n \to \infty$.

Proof. The method of proof is simply to employ the structure of this particular algorithm and the assumptions made in the hypotheses to check that the appropriate conditions of Theorem 3.1 are met.

We turn first to the set E. Let W (a candidate for E which we will show works for an appropriate r) denote the union of the disjoint basin bottoms associated with non-global minima. Let $V = G \cup W$.

For any $x \in B$, a basin bottom

$$\begin{aligned} P[f(X_{n+1}) = f(X_n) \mid X_n = x] &= 1 - \sum_{f(z)>f(x)} q_{xz} e^{-(f(z)-f(x))/c(n)} \\ &\geq 1 - \sum_{f(z)>f(x)} q_{xz} e^{-\alpha/c(n)} \\ &= \sum_{f(z)=f(x)} q_{xz} + \sum_{f(z)>f(x)} q_{xz}(1 - e^{-\alpha/c(n)}) \\ &\geq \sum_{z \in N(x)} q_{xz}(1 - e^{-\alpha/c(n)}) = 1 - e^{-\alpha/c(n)}. \end{aligned}$$

Furthermore, if $x \in B$, $y \in N(x)$, and $f(y) = f(x)$, then $y \in B$ also so, letting for $x \in W$, $a(r,x) = P[f(X_{1+r}) = \cdots = f(X_1) \mid X_1 = x]$, it is clear that

$$\begin{aligned}\ln(P[f(X_{n+r}) = \cdots = f(X_n) \mid X_n = x]) &\geq \ln(a(r,x)) \\ &\geq \sum_{j=1}^{r+1} \ln(1 - e^{-\alpha/c(j)}) \geq \sum_{j=1}^{r+1} \frac{e^{-\alpha/c(j)}}{1 - e^{-\alpha/c(j)}} \\ &\geq -\frac{1}{1 - e^{-\alpha/c(1)}}\, e^{-\alpha/k}\, \frac{1 - e^{-\alpha(r+1)/k}}{1 - e^{-\alpha/k}}\,.\end{aligned}$$

Therefore, for each $x \in W$ there is a $c(x) > 0$ such that $a(r,x) \geq c(x)$ for all r. Now $\nu = \min_{x \in W} c(x)$ must be positive since it is the minimum over a (non-empty) finite set of positive numbers. For any $t \in (0, \nu)$ one will have for all n

$$P[(X_n, X_{n+1}, \ldots, X_{n+r}) \in D \mid X_n = x \text{ and all histories up to } n] \geq t$$

for all $x \in W$. We will see below that for all x in $C \backslash G$ which are not in W this fails for some n and therefore we can take $W = E(t)$.

The transition matrices P_n converge to the matrix P and the xy element of $P_n - P$ is

$$(P_n - P)_{xy} = q_{xy} e^{-(f(y)-f(x))/c(n)}$$

if $f(y) > f(x)$ and 0 otherwise. The limiting matrix P has xy element q_{xy} if $y \in N(x)$ and $f(y) \leq f(x)$ and 0 otherwise. The submatrix S of P which corresponds to the states in $G^c \cap W^c$ has spectral radius less than 1 since every state communicates with at least one state, say x, therein for which $f(y) < f(x)$ for some $y \in N(x)$ and y is in a basin bottom. Therefore, in the limiting chain, with positive probability at least q_{xy} one leaves the set $G^c \cap W^c$ never to return. With S_n the submatrix of P_n corresponding to the states in $G^c \cap W^c$ one has for the matrix norm $\|A\| = \max_i \sum_j |a_{ij}|$

$$\|S_n - S\| \leq \max_x \sum_{y \in N(x)} q_{xy} e^{-(f(y)-f(x))/c(n)} \leq e^{-\alpha/c(n)}$$

so by the assumption of Theorem 4.1, $\sum_{n \geq 1} \tau^n \|S_n - S\| < \infty$, for some $\tau > 1$. By Lemma (A1) it follows that for some $\psi \in (0,1)$ and all m and n,

$$\|S_m S_{m+1} \cdots S_{n+m-1}\| \leq E\psi^n.$$

Since for $x \in G^c \cap W^c$

$$e_x' S_m S_{m+1} \cdots S_{n+m-1} 1 = P[(X_{n+m-1} \in G^c \cap W^c) \cap \cdots \cap (X_m \in G^c \cap W^c) \mid X_{m-1} = x]$$

with $U = G \cup W$ one has for $x \in G^c \cap W^c$

$$P[T_U > m + n \mid X_m = x \text{ and all histories up to } x] \leq K\psi^n.$$

If $x \in W$ then a transition will either occur immediately to W or to W^c so (A2) holds.

Furthermore, as promised above, W satisfies the requirements of E, for if $x \in G^c \cap W^c$ then

$$P[(X_n, X_{n+1}, \ldots, X_{n+r}) \in D \mid X_n = x] \leq e_x' S_n S_{n+1} \cdots S_{n+r-1} 1 \leq K\psi^r$$

since a transition to $G \cup W$ would entail a decrement in the value of $f(X)$. Therefore, for r sufficiently large all elements $x \in G^c \cap W^c$ will fail the test of being in W.

Therefore, for the r we chose above, restarting results in a finite expected first hitting time and convergence to 0 of $P[\tau_G > n]$ at least geometrically quickly.

Corollary 4.2 The (RSA) *algorithm which uses the local generation scheme satisfying* (A3)*, the standard acceptance scheme of* (A4)*, and a cooling schedule of* $c(n) = 1/n$ *will, for sufficiently large* r*, under* (A5) *have tail probabilities which converge to* 0 *at least geometrically fast in* n*.*

The conditions of Theorem 4.1 are not necessary for the geometric convergence of the tail probabilities. In the following example, geometric rate of decrease of the tail probabilities is shown for a restarted simulated annealing which uses the usual logarithmic cooling schedule and a generation matrix which does not satisfy (A3). This example is one for which the (SA) satisfies the conditions of Hajek's theorem but which, without restarting, has an infinite expected hitting time of the goal.

Example 4.3 The Sandia Mountain example provided by Shonkwiler and Van Vleck (1994) as an illustration that independent identical parallel processing can make the expected time to hit the goal go from infinite to finite is presented here from the

perspective of restarting when a state is repeated; we show, by showing that the conditions of Theorem 3.1 are met, that the expected time to goal can be made finite simply by restarting on the diagonal. The Sandia Mountain function, for $N = 4$, is piecewise linear from $(0, -1)$ to $(1, 1)$ and from there to $(0, 4)$. In their example simulated annealing is applied to the search for the global minimum of the function $f(x)$ defined on $\{0, 1, 2\}$ by $f(0) = -1$, $f(1) = 1$, $f(2) = 0$, Sandia 2. They showed that the expected hitting time for the simulated annealing process whose generation matrix is

$$Q = \begin{bmatrix} 0.5 & 0.5 & 0.0 \\ 0.5 & 0.0 & 0.5 \\ 0.0 & 0.5 & 0.5 \end{bmatrix}$$

and which results, under the cooling schedule T, in the Markov transition matrices

$$P = \begin{bmatrix} 1 - \frac{1}{2}e^{-2/T} & \frac{1}{2}e^{-2/T} & 0 \\ 1/2 & 0 & 1/2 \\ 0 & \frac{1}{2}e^{-1/T} & 1 - \frac{1}{2}e^{-1/T} \end{bmatrix}$$

is infinite when the traditional cooling schedule $T = \frac{C}{\ln(n+1)}$, $C = 1$, n being the epoch number, is used. The generation matrix here does not satisfy assumption (A3) but we show that nevertheless the tail probabilities do converge to zero geometrically quickly even using the logarithmic schedule. First, under equally likely restarting (A1) is clearly satisfied. For (A2), note that $U = \{0, 2\}$, $G = \{0\}$, $E = \{2\}$, so that $P[T_U > m + n \mid X_n = 1] = 0$ for any $m \geq 1$ and $n \geq 1$. If $X_n = 2$ then either there is an immediate transition back to U or one to U^c followed immediately by one to U so that for $m = 1$

$$P[T_U > m + n \mid X_n = 2] = \frac{1}{2}e^{-1/T}$$

and otherwise is this is 0 so (A2) is satisfied with any $\phi \in (0, 1)$. For $N = 2$ of Example 4.3 the non-restart case was terminated if it took longer than 40 epochs. Median search times were 3 for restarting and 40 for not.

Obviously, geometric or faster decrease to 0 of the tail probabilities $P[\tau_G > n]$ under (RSA) or otherwise entails a finite expected hitting time of the goal states G, but by itself geometric decrease of the tail probabilities is not a strong recommendation.

Although a direct comparison of expected times to goal for the two processes (RSA) and (SA) seems unattainable, Example 4.3 shows that this geometric convergence of the tail probabilities is not a feature of (SA). We make some informal observations under the assumption that both processes use the same generation matrix with (SA) using a logarithmic schedule $T_n = c/\ln(n+1)$, $c \geq d^*$, and (RSA) using a linear schedule $T_n = 1/n$. Assume a common position of the two algorithms at epoch n. Since the r-process is defined on a different space the two processes will be taken to have a common position at epoch n if the most recent coordinate of the r-process at epoch n coincides with that of the (SA) process. At any instant of time at which the two processes happen to reside at the same "location" the cooling schedule of one, which is logarithmic, should be compared with that of the other, which is linear in the (random) age of the process, for this will indicate the relative tendencies of going downhill. If r is small then the clock will likely have been reset for (RSA), but if r is large then very likely the r-process will not have restarted at all and the epoch number will also be the current age. It is the latter instance which is of interest since (RSA) is assumed to have r "large." At a location which is not a local minimum the (SA) process will have, as the epochs tick away, an ever increasing tendency in comparison with (RSA) to proceed in uphill directions. Thus (RSA) should proceed more rapidly downhill than (SA) at points which are not local minima.

What happens when (SA) and (RSA) are at a local minimum at the same epoch? Very likely the (RSA) will be out of this "cup" (see Hajek) in r steps whereas the (SA) will take some time. Since the goal cannot be reached until the process gets out of the cup this is a crucial quantity in determining the relative performance of the two methods when there are prominent or numerous local minima. The (RSA) will have an immediate chance of finding the cup containing the goal whereas, depending upon the proximity of the present cup to the one containing the goal, (SA) may be forced to negotiate many more cups.

It follows from Fox and Heine (1995) and Fox (1995) that the enriched neighborhood version of QUICKER-j has tail probabilities converging geometrically quickly to 0. An interesting question to which we have no answer is how these two compare. In our obviously completely unbiased opinion, (RSA) should perform better; only small

prescribed numbers of function values in small neighborhoods need be computed and progress downhill is faster.

5. Numerical comparisons

Some numerical results are presented comparing the performance of various forms of (SA) to (RSA). The comparisons were carried out for three types of problems, minimization of a univariate function, minimization of tour length for some TSP's, and finding the maximum value of the permanent of a matrix. In each case a parameter enters which has considerable influence on the performance of the method; for standard (SA) it is the constant C in the cooling schedule $\frac{C}{\ln(n+1)}$ and for (RSA) it is the number $r+1$ of energies duplicated before restarting. We attempted to choose good values of the parameters. When the univariate function or the TSP had a known optimal tour length, the constant C for (SA) was set appropriately according to the results of Hajek. More precisely in the terminology of Fox (1995), we chose his "$\{T_k\}$" schedule. For the other cases of (SA) an attempt was made to estimate the correct depth to use Fox's schedule. For setting the parameter r of (RSA) we employed an heuristic argument based upon the following.

In (RSA) it seems desirable to proceed as quickly as possible to points where the function has a local minimum and then, if necessary, to restart. Rushing to restart is undesirable however for local information about the function is indispensable in charting a course to a local minimum; by prematurely restarting, this information is lost. Therefore one should take care to stay sufficiently long in a location to examine a large enough collection of "directions" from the current point to ensure that paths to lower values are discovered. For functions on the line there are only two directions so one would expect to require very few duplications before the decision to restart is made. Were the selection of new directions deterministic, clearly at most two would be required, but the algorithm chooses these stochastically. In contrast, for a TSP on a reasonable number of cities, if the neighborhood system arises from a 2-change (see Aarts and Korst (1989)) then one should presumably wait for a fairly large number of duplications to make sure enough "directions" have been examined.

As a rough guide we note that in (SA) as long as the state has not changed, the generation matrix yields a sequence of iid "directions." Assuming the proportion of directions downhill is p and uniform probability spread over those directions by the generation matrix, the probability the generation matrix has not yielded a downhill after m generations is simply $(1-p)^m$. To make this quantity small, say less than β, m should be approximately $\ln(\beta)/\ln(1-p)$. On the line the most interesting places are where one direction is up and the other down so $p \sim 1/2$ seems reasonable. Furthermore, the consequences of restarting are minimal so a large α, say $1/2$, also seems reasonable. Thus one should take r around 1. In a TSP with 100 cities restarting can be costly since the considerable time it takes to get downhill will likely be wasted upon restarting. Thus we take α small, say .01. It is not clear what p should be. Presumably the "surface" represented by the tour lengths could be rather rough so we'll take $p = .05$ to ensure a thorough although perhaps too lengthy examination of directions. This translates to run lengths of $r \sim 100$ and an examination of a fairly small proportion of the 4851 "directions" available under 2-change.

Example 5.1 For a randomly generated function the median number of epochs required to find the global minimum by (SA) under optimal cooling, with the stipulation that the search was terminated at 221 epochs if the minimum had not yet been found, was 221. For (RSA) with $r = 1$ the median number of epochs required to find the global minimum of the function was 21.

Example 5.2 In this example an optimal 100 city tour was sought using 2-change as the neighborhood system with equally likely probabilities for the generation matrix. The locations were scaled from a TSP instance known as kroA100 taken from a data base located on the WEB at
`http://softlib.rice.edu/softlib/catalog/tsplib.html`.
Each of (SA) and (RSA) was run for 1000 epochs. The median best tour length found by (SA) was of 35.89 with a minimum of 34.06. For (RSA) the median best tour length found in 1000 epochs was 14.652 with a minimum of 13.481.

Example 5.3 A 24-city TSP instance known as gr24 obtained from the same data base as kroA above was analyzed again using 2-change for the neighborhood system

and equally likely choices for the generation matrix. Each of (SA) and (RSA) was run for 500 epochs. The best tour lengths found by (SA) had a median of 2350.5 and a minimum of 1943. (RSA) with $r + 1 = 24$ had a median best tour length after 500 epochs of 1632.5 with a minimum of 1428. The optimal length is 1272. A similar result on 24 cities was obtained by running the two for 1000 epochs. Under (SA) the median was 2202 with a minimum of 1852 while for (RSA) the median best tour length was 1554.5 and minimum 1398.

Example 5.4 Performance of (SA) with depth 40 and (RSA) with $r = 100$ was compared on a randomly generated 100-city TSP. Median best tour length after 1000 epochs for (SA) was 43.14 and the minimum was 40.457. For (RSA) the median best was 19.177 and the minimum best was 17.983.

Example 5.5 We also tested the performance of our method in comparison with several other (SA) cooling schedules employed by Nourani and Andresen (1998) on the problem of optimizing the permanent of 0-1 matrices. The permanent of an $n \times n$ matrix M is defined to be

$$\text{perm}(M) = \sum_{\sigma} \prod_{i=1}^{n} m_{i,\sigma(i)}$$

where the sum extends over all permutations σ of the first n integers. The permanent is simply the determinant without the alternating signs. Allowing the matrix elements to be only 0 or 1, for a given matrix size n and number d of 1's, $0 < d < n^2$, the problem is to find the matrix having maximum permanent.

This problem is simple and scalable being completely determined by the two integer values, n and d, the matrix size, and its number of 1's. As n grows the problem becomes harder in two ways. First the difficulty in calculating a permanent grows as $n!$ since the number of possible permutations σ grows as $n!$. Second, the number of possible 0/1 matrices grows as 2^{n^2}. The difficulty also depends on d in that for a given n, smaller values of d yield fewer non-zero terms in the sum in the calculation of perm(M) while for $d \sim n^2$ almost all $n!$ terms need to be computed.

Nourani and Andresen (1998) studied their constant thermodynamic simulated annealing cooling schedules with various cooling schedules.

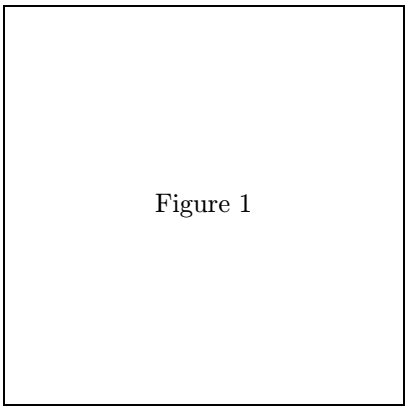

Figure 1. Simulated annealing for the permanent problem with various cooling schedules

To compare the several approaches for solving this problem, we fixed $n = 14$ and $d = 40$. This gives the number of possible arrangements of the 1's, the size of the search space, to be $\binom{196}{40}$ or about 8.3×10^{41}. These values for n and d result in a permanent calculation that is fast enough that millions of permanent calculations could be tried in a few minutes; this was necessary for one of the approaches.

In all the experiments reported below, our neighborhood system for the simulated annealing was defined by allowing any 1 appearing in the matrix to move one position up or down or to the left or to the right. In this, we allowed wrapping, that is, a 1 on the bottom row could swap positions with a 0 on the top row and similarly for the first and last columns. In this way, each solution, or arrangement of d 1's, has approximately $4d$ neighbors.

The "energy" of the annealing, to be minimized, was taken as the negative of the permanent itself and the cooling schedules we tested were geometric, inverse log, inverse linear, linear, and dynamic as described in Andresen and Gordon (1994), Nulton and Salamon (1988), and Nourani and Andresen (1999). Temperature ranged from on the order of 6 down to the order of 0.2. For each different cooling schedule, we tried several temperature ranges until we found one that seemed to work well. Thus, we compared the "best" runs for each cooling schedule.

For the restart method, we also ran experiments with these cooling strategies except

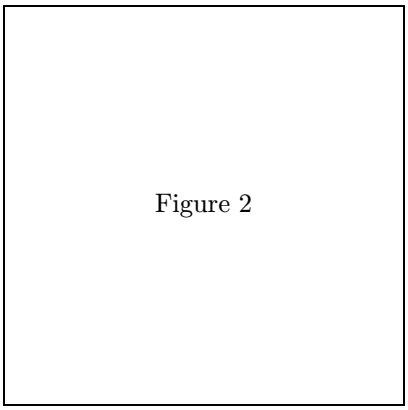

Figure 2. Restart for the permanent problem with various cooling schedules

for the dynamic strategy. We took the restart repeat count, $r + 1$, to be 160.

The results are displayed in Figures 1–3 and are the averages of 10 runs. As seen in the figures, this problem displays a definite "phase transition" in terms of normal annealing. This phase transition occurs for temperatures in the neighborhood of 1. The different schedules reached 1 at different times during their run giving rise to the different times for their precipitous fall. All of the simulated annealing approaches which spent some time above and below the phase transition temperature performed about the same, achieving minima on the order of -1000. As can be seen, all the annealing approaches got stuck and ceased to find improvement when their temperature dropped below the phase transition temperature.

By contrast, the restart algorithm made both very rapid progress at the beginning of the runs and continued to make progress even up to the time the runs were halted. All the annealing runs with restart consistently achieved solutions on the order of -1500. The restarting step was effective in allowing the algorithm to escape from local minima even at temperatures below the critical temperature (of approximately 1) where the phase transition occurs.

We do not know the absolute minimum for this problem, but the best energy seen in any of the runs in all the experiments was -2016.

Of special note is the dynamic cooling schedule (see Andresen (1996) for details).

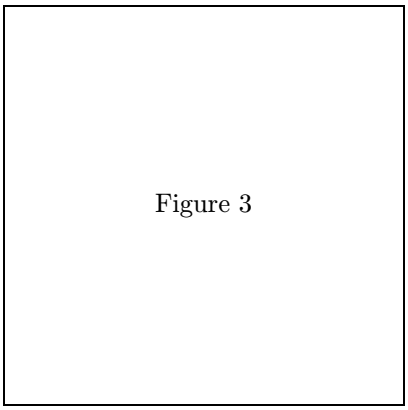

Figure 3. Restart versus the best standard annealer for the permanent problem

In order to use this approach, one must perform preliminary runs in order to gather "infinite temperature" energy density and transition statistical data. We found this to be very hard to do and very costly with essentially no payoff in terms of benefit. Even after forty eight million iterations at the (practically) "infinite temperature" of 100, there was still no statistical data for energies below -400. Furthermore, we found that the statistical technique for estimating the second largest eigenvalue (necessary to compute the internal relaxation time of the statistical system, see Andresen (1996)) entailed so much error as to be practically useless. Instead we directly calculated this eigenvalue from the estimated transition matrix. For this problem, the effort required to gather the (incomplete) statistical data in order to start simulated annealing runs with the dynamic cooling schedule is better spent in annealing using some other cooling schedule.

An alternative, more careful, analysis of the size of r is provided by closer examination of the proof of Theorem 4.1. Under the cooling schedule $c(n) = 1/n$ with an equally likely generation matrix the choice

$$r \geq \frac{-e^{-\alpha}}{(1-e^{-\alpha})^2} \frac{1}{\ln(1-p)}$$

will guarantee the conclusion of the theorem under its other hypotheses, where $p = 1 - \gamma$ is the worst case, smallest probability of a downhill from among the points in

$C\backslash U$. However, this may not help in the determination of a "good" r since one would expect these quantities to be unknown. Theorem 4.1 provides no indication of the consequences of choosing r small, however, the formula

$$E[\tau_G] = \frac{E[\tau_U]}{1 - P[\tau_E < \tau_G]}$$

suggests that choosing r large should make the numerator small. Our simulations indicate that the performance could be fairly robust against the choice of r; the numerical evidence presented above is just a portion of a larger body for which an educated choice of r always resulted in (RSA) performing better than (SA).

Appendix

We use in the proof of Theorem 4.1 the following simple Lemma A.1. Once the problem-states at local minima have been removed from transition matrices the properties we need are easily shown. Contrast this with the delicate results of Tsitiklis (1989) for example. Proof of Lemma A.0 is left to the reader.

Lemma A.0 *Let for integers $n \geq 1$, $m \geq 1$, let a_{nm} be non-negative and satisfy for all $u \geq 1$*

$$a_{nn+m}a_{n+m+1\,n+m+u} \geq a_{nn+m+u}.$$

If there are finite constants K and M and an $r \in (0,1)$ such that $m \geq M$ entails $a_{mm+n} \leq Kr^n$ for all n then there is a finite constant $K' \geq K$ such that for all m and n, $a_{mm+n} \leq Kr^n$.

Lemma A.1 *If for some $\tau > 1$, $\sum_{n\geq 1} \tau^n \|P_n - P\| < \infty$ and for some $k \geq 1$, P^k has norm $\eta < 1$, then there is a constant $K < \infty$ such that for all n and m*

$$\|P_m P_{m+1} \cdots P_{m+n-1}\| < K\eta^n.$$

Proof. Let $\delta > 0$ and $1/\beta = \tau - \delta > 1$. By our assumptions there is a finite constant A such that $\|D(j)\| < A(\tau - \delta)^{-j}$ for all j, where $D(j) = P_j - P$. Let M be so large

that $m \geq M$ entails $\frac{A\beta^m}{\eta} < 1$. Consider

$$\begin{aligned}
&P_m P_{m+1} \cdots P_{nk+m-1} - P^{nk} \\
&= (P + D(m))(P + D(m+1)) \cdots (P + D(m+nk-1)) - P^{nk} \\
&= P^{nk-1} D(m+nk-1) + \cdots + D(m) P^{nk-1} \\
&\quad + P^{nk-2} D(m+nk-2) D(m+nk-1) \\
&\quad + \cdots + D(m) D(m+1) P^{nk-2} \\
&\quad + \cdots + D(m) D(m+1) \cdots D(m+nk-1).
\end{aligned}$$

The j^{th} term has norm no larger than $K'\eta^n \left(\frac{A\beta^m}{h}\right)^j$ so

$$\|P_m P_{m+1} \cdots P_{nk+m-1} - P^{nk}\| \leq K'\eta^n \sum_{i=1}^{nk} \left(\frac{A\beta^m}{\eta}\right)^i \leq B\eta^n$$

for $m \geq M$ and all n. Now suppose n is arbitrary, $n = dk+\Delta$, where $0 \leq \Delta < k$ so that $\|P^{dk+\Delta} - P_m P_{m+1} \cdots P_{dk+\Delta+m-1}\| = \|AP^{\Delta} - H(P+D(dk+m)) \cdots (P+D(dk+\Delta+m-1))\|$, where $A = P^{dk}$ and $H = (P+D(m))(P+D(m+1)) \cdots (P+D(m+dk-1))$. So

$$\begin{aligned}
&\|P^{dk+\Delta} - P_m P_{m+1} \cdots P_{dk+\Delta+m-1}\| \\
&\quad = \|AP^{\Delta} - HP^{\Delta} + HP^{\Delta} - H(P + D(dk+m)) \cdots \\
&\qquad (P + D(dk+\Delta+m-1))\| \\
&\quad \leq \|A - H\| \|P^{\Delta}\| + \|H\| \|P^{\Delta} - (P + D(dk+m)) \\
&\qquad \cdots (P + D(dk+\Delta+m-1))\| \\
&\quad \leq B\eta^n \|P^{\Delta}\| + B'\eta^n M \leq K\eta^n.
\end{aligned}$$

Acknowledgement

The authors wish to thank the editors and referees for their helpful comments which greatly improved the paper.

References

AARTS, E. AND KORST, J. (1989) *Simulated Annealing and the Boltzmann Machines.* Wiley, Chichester.

ANDRESEN, BJARNE (1996) Finite-time thermodynamics and simulated annealing. *Entropy and Entropy Generation*, ed. J. S. Shinov (Dordrecht Kluwer), 111–127.

ANDRESEN, BJARNE AND GORDON, J. M. (1994) Constant thermodynamic speed for minimizing entropy production in thermodynamic processes and simulated annealing. *Physical Review E,* **50**, no. 6, 4346–4351.

ATKINSON, A. C. (1992) A segmented algorithm for simulated annealing. *Statistics and Computing* **2**, 221–230.

BÉLISLE, CLAUDE (1992) Convergence theorems for a class of simulated annealing algorithms on R^d. *J. Appl. Prob.* **29**, 885–895.

CHIANG, T. AND CHOW, Y. (1988) On the convergence rate of annealing processes. *SIAM J. Control and Optimization* **26**, 1455–1470.

FELLER, W. (1968) *An Introduction to Probability Theory.* Wiley, New York.

FOX, B. (1995) Faster simulated annealing. *SIAM J. Optimization* **5**, 488–505.

FOX, B. AND HEINE, G. (1995) Probabilistic search with overrides. *Annals of Applied Probability* **5**, 1087–1094.

HAJEK, B. (1988) Cooling schedules for optimal annealing. *Math. Operat. Res.* **13**, No. 2, 311–329.

HU, X., SHONKWILER, R., AND SPRUILL, M. (1997) Randomized restarts, submitted.

KOLONKO, M. (1995) A piecewise Markovian model for simulated annealing with stochastic cooling schedules. *J. Appl. Prob.* **32**, 649–658.

NAKAKUKI, Y, AND SADEH, N. (1994) Increasing the efficiency of simulated annealing search by learning to recognize (un)promising runs. *Proceedings of the Twelfth National Conference on Artificial Intelligence* **2**, 1316–1322.

NOURANI, YAGHOUT AND ANDRESEN, BJARNE (1998) A comparison of simulated annealing cooling strategies. *J. Phys. A: Math. Gen.* **31**, 8373–8385.

Nourani, Yaghout and Andresen, Bjarne (1999) Exploration of NP-hard enumeration problems by simulated annealing – the spectrum values of permanents. *Theoretical Computer Science* **215**, 51–68.

Nulton, J. D. and Salamon, P. (1988) Statistical mechanics of combinatorial optimization, *Physical Review A* **37**, no. 4, 1351–1356.

Schoen, F. (1991) Stochastic Techniques for Global Optimization: A Survey of Recent Advances. *Journal of Global Optimization* **1**, 207–228.

van Laarhoven, P. and Aarts, E. (1987) *Simulated Annealing: Theory and Applications.* Reidel, Boston.

Shonkwiler, R. and Van Vleck, E. (1994) Parallel speed-up of Monte Carlo methods for global optimization. *J. Complexity* **10**, 64–95.

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