Internat. J. Math. & Math. Sci.
VOL. 17 NO. 4 (1994) 791-798

SIMULATED ANNEALING ON UNCORRELATED ENERGY LANDSCAPES

BEN GOERTZEL and MALWANE ANANDA

Department of Mathematical Sciences
University of Nevada,
Las Vegas, NV 89154

(Received December 18, 1992 and in revised form November 6, 1993)

ABSTRACT. A function f: $\{0,1,2,\ldots,a\}^n \to R$ is said to be uncorrelated if Prob $[f(x) \le u] = G(u)$. This paper studies the effectiveness of simulated annealing as a strategy for optimizing uncorrelated functions. A recurrence relation expressing the effectiveness of the algorithm in terms of the function G is derived. Surprising numerical results are obtained, to the effect that for certain parametrized families of functions $\{G_c, c \in R\}$, where c represents the "steepness" of the curve $G'(u)$, the effectiveness of simulated annealing increases steadily with c These results suggest that on the average annealing is effective whenever most points have very small objective function values, but a few points have very large objective function values.

KEYWORDS AND PHRASES: Simulated annealing, Evolutionary mutation, uncorrelated functions.

AMS SUBJECT CLASSIFICATION CODE 92A12

1. INTRODUCTION

The simulated annealing algorithm has proved effective for optimizing a variety of complex multiextremal functions[1], such as those involved in VLS1 design[2]. In this paper we explore the application of simulated annealing to optimizing **uncorrelated** functions - functions f so that

$$\text{Prob}\left(f(x) \le u\right) = \begin{cases} 0 & \text{if } u < 0 \\ G(u) & \text{if } u \ge 0, \end{cases} \tag{1}$$

In itself, the study of the optimization of uncorrelated functions is of little practical interest. However, it serves to reveal some interesting mathematical phenomena; and it is furthermore of intuitive relevance to the understanding of the optimization of functions defined in terms of "noisy" data, which often behave in an uncorrelated manner over portions of their domains.

In Section 1 we describe the simulated annealing algorithm and explain its relation to the evolutionary mutation algorithm of Hagen, Macken and Perelson [3]. In section 2 we derive a recurrence

relation describing the effectiveness of the simulated annealing method in terms of the distribution G Finally, in Section 3 we describe some intriguing numerical results, regarding the specific distributions

$$g(u) = G'(u) = (1-c)u^{-c}, \quad 0 < c < 1, \quad 0 < u < 1 \tag{2}$$

$$g(u) = G'(u) = ce^{-cu}, \quad c > 0, \quad u > 0 \tag{3}$$

In both cases, the parameter c determines the "steepness" of g And in each case it is shown that, as c increases, the effectiveness of simulated annealing increases correspondingly

2 EVOLUTIONARY MUTATION AND SIMULATED ANNEALING

The optimization of functions $f\ \{0,\cdots,a\}^n \to R$ is intimately related to statistical physics [4] The function f is interpreted as the energy associated with some physical system, and one is concerned with finding the minimum-energy configuration of the system The graph of f is an "energy landscape " This terminology is often used in discussing simulated annealing, since the origin of the simulated annealing algorithm was in statistical physics [5]

Here, however, we shall find it useful to approach discrete optimization from another point of view that of molecular biology In Hagen, Macken and Perelson a molecule is represented as a sequence of n letters, each letter being chosen from an alphabet of size a. Each molecule x is assigned a **fitness** $f(x)$ according to formula (1) Instead of an "energy landscape" one speaks of a "fitness landscape", instead of minimizing energy one speaks of maximizing fitness.

Evolution is assumed to occur by a simple process of point mutation. A single molecule is considered, at each time step it mutates in exactly one location, meaning that exactly one of its letters is changed If the mutant is fitter than the old molecule (according to f), then the mutant is retained; if not the old molecule is retained. Whichever molecule is retained becomes the object of the next step· it is mutated, and either it or its offspring is retained, etc A sequence of successively retained molecules is called an "evolutionary walk."

This is clearly a caricuture of the evolutionary process. It incorporates mutation and survival of the fittest, but in an oversimplified way: real evolution does not consist of a sequence of winner-take-all contests between parents and their children. However, the advantage of such a simplified model is that an exact mathematical analysis is possible One can give excellent estimates of such quantities as the expected fitness of the local maximum arrived at by this evolutionary process (it is about 38% fitter than the average local maximum). What makes the analysis particularly simple is that the nature of G is irrelevant to much of the behavior of the evolutionary process: what matters most is the ordering induced by G, not the relative magnitudes of the fitnesses of various points.

The simulated annealing algorithm may be understood as a modification of this simple evolutionary procedure. Under simulated annealing, at each time step a molecule creates an offspring (a "one-mutant neighbor") by mutating in one location, and if the mutant is fitter then the old molecule, it is retained. However, even if the mutant is **less** fit than the old molecule, the mutant is **still** retained with probability $p(v,u) = \exp\left(\frac{u-v}{T_k}\right)$, where v is the fitness of the old molecule, u is the fitness of the mutant, k is the number of evolutionary steps so far, and $0 < T_k \to 0$ as $k \to \infty$. Simulated annealing gives an advantage to the mutant simply for being **different**, and the extent of this advantage is determined by the magnitude of T_k. As T_k approaches ∞ the evolutionary process becomes a random search in which the new mutant is always accepted; and as $T_k \to 0$ simulated annealing reduces to the method of Hagen, Macken and Perelson. In practical optimization applications one usually sets $T_k = b^k T_0$, $0 < b < 1$, $k = 1,2,3,\text{L}$; and this rule of thumb has recently been validated by Sorkin's [6] remarkable results regarding simulated annealing on fractal energy landscape

3 A RECURRENCE RELATION

Let $U_k(u_0)$ be the fitness attained on the k'th step of the evolutionary walk by simulated annealing, starting from a fitness u_0 Suppose

$$F_k(u,u_0) = \text{Prob}\left\{U_k(u_0) < u \text{ and the walk has at least } k \text{ steps}\right\}$$

and

$$f_k(u,u_0) = \frac{d}{du} F_k(u,u_0)$$

In order to calculate $f_k(u,u_0)$, let us analyze the evolutionary walk starting from u_0.

First consider the case $u > u_0$ In order to achieve a fitness in the range $(u,u+du)$ within k steps, one of two possibilities must hold It may be that in the first *k-1* steps the walk has achieved a fitness in the range $(u',u'+du')$, $u_0 < u' < u$ with probability $f_{k-1}(u',u_0)du'$, in which case in the next step, the evolutionary walk must achieve a higher fitness in the range $(u,u+du)$ Or on the other hand, in the first *k-1* steps the walk must have achieved a fitness in the range $(v',v'+dv')$, $u_0 < u < v'$ with probability $f_{k-1}(v',u_0)dv'$, in which case in the next step the evolutionary walk must achieve a lower fitness in the range $(u,u+du)$

The probability associated with achieving a higher fitness in the range $(u,u+du)$ starting from a lower fitness in the range $(u',u'+du)$, in one step, is given by

$$\begin{aligned} f_1(u,u') &= \sum_{i=0}^{D-2}\left(\int_0^{u'} g(u'')q(u'',u')du''\right)^i g(u) \\ &= \frac{1-\left(\int_0^{u} g(u'')q(u'',u)du''\right)^{D-1}}{1-\left(\int_0^{u} g(u'')q(u'',u)du''\right)} g(u) \\ &= K(u')g(u) \end{aligned}$$

where $q(v,u) = 1 - p(v,u)$ and $D = (a-1)n$ is the number of one mutant neighbors which a sequence has.

And the probability associated with achieving in one step a lower fitness in the range $(u,u+du)$, starting from a higher fitness in the range $(v',v'+dv')$, is given by

$$f_1(u,v') = \sum_{i=0}^{D-2}\left(\int_0^{v'} g(v'')q(v'',v')dv''\right)^i g(u)p(v',u) = K(v')g(u)p(v',u).$$

By combining these formulas one obtains the recursion relation

$$\begin{aligned} f_k(u,u_0) &= \int_{u_0}^{u} f_{k-1}(u',u_0)g(u)K(u')du' \\ &\quad + \int_{u}^{\infty} f_{k-1}(v',u_0)K(v')g(u)p(v',u)dv' \end{aligned} \tag{4}$$

for all $u \geq u_0$, $k \geq 2$.

Similarly, for $u < u_0$, one obtains

$$\begin{aligned} f_k(u,u_0) &= \int_0^{u} f_k(v',u_0)g(u)K(v')dv' \\ &\quad + \int_{u}^{u_0} f_{k-1}(u',u_0)K(u')g(u)p(u',u)du'. \end{aligned}$$

Next, suppose $h_k(u,u_0) = f_k(u,u_0)/g(u)$. Then for $u \geq u_0$ and $k \geq 2$,

$$\begin{aligned} h_k(u,u_0) &= \int_{u_0}^{u} h_{k-1}(u',u_0)g(u')K(u')du' \\ &\quad + \int_{u}^{\infty} h_{k-1}(v',u_0)K(v')g(v')p(v',u)dv' \end{aligned}$$

Differentiating twice and doing some simple algebra, one obtains

$$h_k'' - \frac{1}{T_k} h_k' + \frac{1}{T_k} g(u) K(u) h_{k-1} = 0 \tag{5}$$

where $h_k = h_k(u, u_0)$, with the initial conditions $h_k(u_0, u_0) = h_k'(u_0, u_0)$ and $h_k(u_0, u_0) = \int_{u_0}^{\infty} h_{k-1}(v', u_0) K(v') g(v') p(v', u_0) dv'$ for all k

Given the solution h_k to this recurrence relation, one may determine the expected fitness of the final solution provided by the simulated annealing algorithm as a function of g. Let $U_k(u_0)$ denote the fitness attained on the k'th step of the walk starting from initial fitness u_0, let $U_{fin}(u_0)$ denote the fitness attained on the final step of a walk starting from initial fitness u_0. Then we may set

$$P_{fin}(u; u_0) = \text{ Prob } \{u < U_{fin} < u + du\},$$

and obtain

$$\text{Prob} \left\{ \begin{matrix} u < U_k < u + du \text{ and path has} \\ \text{exactly } k \text{ steps} \end{matrix} \right\} = f_k(u; u_0) \left(\int_0^u g(u') q(u', u) du' \right)^{D-1} du$$

$$\text{Prob} \left\{ \begin{matrix} u < U_k < u + du \text{ and path has} \\ \text{exactly 0 steps} \end{matrix} \right\} = f_0(u; u_0) \left(\int_0^u g(u') q(u', u_0) du' \right)^{D-1} du$$

Finally, this yields

$$\begin{aligned} P_{fin}(u; u_0) &= f_0(u; u_0) \left(\int_0^u g(u') q(u', u_0) du' \right)^D \\ &+ \left(\sum_{k=1}^{\infty} f_k(u; u_0) \right) \left(\int_0^u g(u') q(u', u) du' \right)^{D-1} \end{aligned} \tag{6}$$

Equation (5) is analytically intractable; and, furthermore, there is no apparent stable method for numerically approximating its solution. Obtaining h_k from h_{k-1} numerically, inaccuracies will tend to compound, leading to meaningless answers for large k -- the only k which are interesting.

However, it may nonetheless be possible to squeeze a little insight into the optimization process out of equation (5). For example, one may explore the manner in which h_k depends on k. Let $h_k(v; v_0) = \frac{h_1(u_0)[z(u)]^{k-1}}{(k-1)!}$. Substituting into (5) yields

$$(k-2) z'^2 + z z'' - \frac{1}{T_k} z z' + c_2 z = 0 \tag{7}$$

where $c_2 = \frac{K(u)}{T_k}$, and the stated initial condition implies $z(u_0) = z'(u)$. This equation is easily seen to have a solution in some neighborhood of u_O. So, locally at least, the dependence of h_k on k is rather simple.

4 SOME PUZZLING NUMERICAL RESULTS

It appears to be very difficult to get accurate numerical results out of equation (5). Therefore, in order to compare the effectiveness of simulated annealing with that of the ascent method of Hagen, Macken and Perelson, we have resorted to direct optimization experiments with particular distributions G. In particular, we have considered the uniform density $g(u) = 1$, and the exponential and power-function densities given by formulas (2) and (3). Results of these tests for various values of c, T_O, b and n are given in Tables 1-3.

For the uniform distribution, simulated annealing and the ascent method tend to provide final answers of approximately equal fitness. However, simulated annealing takes a much longer time arriving

at its final answer In this example, the higher-temperature steps are largely, perhaps even entirely, wasted The surprise is that when one varies g from the uniform, the same conclusion does **not** always hold

This is somewhat counterintuitive Two reasons are generally given for the effectiveness of simulated annealing [2,6] its implicit "divide-and-conquer" methodology, and its ability to climb out of the basins of shallow local optima Both of these properties are of dubious value in an uncorrelated setting For example, what's the point of climbing out of a shallow basin when the fitness of u point on the rim of the basin is drawn from the same distribution as the fitness of an arbitrary point?

However, even uncorrelated functions possess a certain combinatorial structure regarding optimization For instance, as pointed out in [3], fitter local maxima tend to have larger basins This combinatorial structure apparently comes into play when optimizing uncorrelated functions drawn from the exponential and power-function densities (2) and (3) Let $E_{AS}(G,u_0)$ denote the expected fitness of the final answer provided by the ascent algorithm of Hagen, Macken and Perelson, for an uncorrelated function f with distribution G, beginning from an initial point selected from a uniform distibution on $\{0,1,\cdots,a\}^n$ Let $E_{AN}(G,T_0,b)$ denote the expected fitness of the final answer provided by the annealing algorithm with parameters T_O and b, under identical conditions. Finally, let

$$R_E(c)=\frac{E_{AN}(E_c,T_0,b)}{E_{AS}(E_c)},$$

$$R_P(c)=\frac{E_{AN}(P_c,T_0,b)}{E_{AS(P_c)}}.$$

Our numerical results, reported in Tables 1,2 and 3 and more extensively in [7], indicate that for a variety of $a,n,T_0>0$ and $b>0,R_E$ and R_P are **increasing functions** of c

Intuitively, in each case an increase in c corresponds to a "steeper" curve g. The larger c is, the less likely are sequences of relatively high fitness, but the greater is the disparity in fitness between the fittest and the least fit sequences. Annealing pays off better when reasonably fit sequences are rare but nearly always very, very fit If one could solve equation (5), one could explore this phenomenon analytically, using the formula for P_{fin} obtained above. However, as observed above, this equation presents formidable difficulties: it remains a challenge.

In the abscence of an explanation in terms of equation (5), the best one can do is to give an intuitive, heuristic explanation of the phenomenon. For example, one may construct a deterministic function similar to the uncorrelated functions in our experiments Suppose $f\ \{0,1\}^n\to R$ is so that

a) there are P points x in $\{0,1\}^n$ for which $f(x)>L$,

b) for all other points x in $\{0,1\}^n$,

$$f(x)=\left\{\begin{matrix}s \text{ if } x \text{ is of odd parity}\\0 \text{ if } x \text{ is of even parity}\end{matrix}\right\}$$

where $0<s<L$ and $P<2^n$ Then it is easy to see that, for fixed $s,T_0>0$ and $0<b<1$, $R_f(L)=\frac{E_{AN}(f,T_0,b)}{E_{AS}(f)}$ is an increasing function of L . For $E_{AS}^{(L)}=\frac{cL+(2^n-c)s}{2^n}$, where c is the number of points on uphill evolutionary walks leading to an x so that $f(x)>L$ (by elementary combinatoric considerations, $c\le n(n-1)P$) And similarly, $E_{SA}(L)=\frac{c_1L+(2^n-c_1)s}{2^n}$, where $c_1>c$

The case of this contrived function is particularly easy because there are no uphill walks of length greater than 2. But, intuitively, the situation should be basically the same whenever there are P points x

in $\{0,1,\mathrm{L}\ ,a\}^n$ so that $f(x) > L$, and for all other points $f(x) < s < L$ To see why, suppose $s << L$ Then the probability of an *SA* evolutionary walk leaving one of the *P* points is very low; it is less than

$$e^{\frac{s-L}{T_k}} \sim 1 - \frac{L}{T_k} + \frac{s}{T_k}$$

at any one step But on the other hand, the probability of leaving one of the unfit points to go to another is greater than

$$e^{\frac{-s}{T_k}} \sim 1 - \frac{s}{T_k}$$

at any one step So if, say, $s << T_k \approx L$, then *SA* will walk randomly through the unfit points, but when it hits one of the *P* very fit points it will stay there

5 CONCLUSIONS

Finally, let us put these results in perspective. They are of little direct practical utility. They do **not** imply that simulated annealing is necessarily a good strategy for optimizing uncorrelated functions, in any situation. It can yield better answers than simple evolutionary ascent, but it also generally takes more steps. And in any case, optimizing uncorrelated functions is of rather little practical interest.

However, the results do have interesting biological implications As Hagen, Macken and Perelson [3] observe, the fitness landscapes associated with proteins tend to be highly correlated in some regions, uncorrelated in others, and partially correlated in yet others. Our results indicate that, in the uncorrelated and partially correlated regions, proteins might benefit from following a strategy similar to simulated annealing -- by being progressively less liberal in the acceptance of mutations. It is tempting to conjecture that some proteins do follow such a strategy.

And finally, we suggest that the main value of these results is heuristic. We conjecture that the phenomenon observed in the numerical experiments described above is in fact much more general: that, for uncorrelated functions and **noisy** objective functions in general, simulated annealing will be more effective when most points are very unfit, but a few points are **very** fit.

It is not clear how one would go about demonstrating this conjecture, but a comprehensive analysis of (5) would be a good start. In particular, we propose the following

CONJECTURE: Where g is given by (2) or (3), simulated annealing outperforms evolutionary ascent. That is, $E\left[P_{fin}(u;u_0)\right] \geq E\left[\hat{P}_{fin}(u;u_0)\right]$ where P_{fin} is given by (6) and $\hat{P}_{fin}$, the final fitness of evolutionary ascent, is given in [3] as

$$\hat{f}_0(u;u_0)G(u_0)^D + \sum_{k=1}^{\infty} \hat{f}_k(u;u_0)G(u)^{D-1}$$

$$\text{where } \hat{f}_k = \frac{g(u)h_1(u_0)}{(k-1)!}\left[\int_{G(u_0)}^{G(u)} \frac{1-q^{D-1}}{1-q}\right]^k$$

However, when $g = 1$, $E\left[P_{fin}(u;u_0)\right] = E\left[\hat{P}_{fin}(u;u_0)\right]$.

Table 1· Average fitness for uniform distribution
(n = 20, a = 2, b = 0 99)

T_0	$E_{AN}(U,T_0,b)^*$	$E_{AS}(U)^*$	$\frac{E_{AN}(U,T_0,b)}{E_{AS}(U)}$
4	0.9640	0 9683	0.9918
8	0.9631	0.9654	0 9939

* These expected values were calculated using 100 iterations.

Table 2: Average fitness for the exponential function
(n = 20, a = 2, b = 0.99)

T_0	C	$E_{AN}(E_c,T_0,b)^*$	$E_{AS}(E_c)^*$	$R_E(c)$
1	0.1	41.7075	40.4471	1.03
	3.0	1.8669	1.3535	1.37
	5.0	1.1125	0.8069	1.37
4	0.1	47.6544	40.4773	1.17
	3.0	1.8931	1.4070	1.34
	5.0	1.1509	0.8090	1.42
8	0.1	55.1330	40.2780	1.36
	3.0	1.9057	1.3751	1.38
	5.0	1.1343	0.8010	1.41

* Using 500 iterations, these expected values were calculated.

Table 3 Average fitness for the power function
(n = 20, a = 2, b = 0 99)

T_0	C	$E_{AN}(E_c, T_0, b)^*$	$E_{AS}(E_c)^*$	$R_E(c)$
1	0 9	0 8811	0.7484	1 17
	0 99	0 6161	0 2498	2 46
	0 999	0 3110	0.0284	10.96
4	0.9	0 8925	0.7513	1 18
	0 99	0 6471	0 2680	2 41
	0.999	0 2930	0.0288	10.17
8	0 9	0.8823	0 7763	1 13
	0.99	0 6389	0.2584	2 47
	0.999	0.3039	0.0331	9.18

* Using 500 iterations, these expected values were calculated.

REFERENCES

1 ACKLEY, D (1987). An Empirical Study of Bit Vector Optimization, in Genetic Algorithms and Simulated Annealing, Ed by L. Davis, London: Pitman.

2. KIRPATRICK, S., C. GELATT and M. VECCHI (1983). Optimization by Simulated Annealing. Science 220, No. 4598.

3. MACKEN, C., P. HAGAN and A. PERELSON (1991). Evolutionary Walks on Rugged Landscapes. Preprint.

4. SALAMON, P., J. NULTON, J. HARLNAD, J. PEDERSEN, G. RUPPEINER and L. LIAO (1988). Simulated Annealing with Constant Thermodynamic Speed. Computer Physics Communications 4 pp. 423-428.

5. METROPOLIS, N., A. ROSENBLUTH, M. ROSENBLUTH, A. TELLER and E TELLER (1953). Journal of Chemical Physics 21, p 1087.

6 SORKIN, G. (1991). Efficient Simulated Annealing on Fractal Energy Landscapes Algorithmica 6, No. 3

7. ANANDA, M. and B. GOERTZEL (1991). Simulated Annealing on Uncorrelated Energy Landscapes: Numerical Results. Technical report