

Criteria for an optimum simulated annealing schedule for problems of the travelling salesman type

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

1987 J. Phys. A: Math. Gen. 20 1239

(<http://iopscience.iop.org/0305-4470/20/5/032>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 129.252.86.83

The article was downloaded on 01/06/2010 at 05:25

Please note that [terms and conditions apply](#).

Criteria for an optimum simulated annealing schedule for problems of the travelling salesman type

Stephen Rees and Robin C Ball

Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK

Received 20 May 1986

Abstract. The simulated annealing algorithm for optimisation problems such as the travelling salesman problem is reviewed. The concept of the autocorrelation function for cost functions is introduced and it is shown how numerical experiments to measure this quantity can provide criteria as to how rapidly a system can be annealed close to equilibrium. From this we obtain an optimum annealing schedule of general applicability.

1. Introduction: review of the method

Simulated annealing has its origins in the work of Metropolis *et al* [1] who invented a computational scheme for simulating a collection of physical particles in thermal equilibrium at some temperature T , for which the probability of finding the system in microstate i with energy E_i is given by the Boltzmann distribution

$$P(E_i) = \frac{\exp(-E_i/k_B T)}{\sum_i \exp(-E_i/k_B T)}. \quad (1.1)$$

They did this by generating on their computer a collection of particles with a random configuration and random momenta and calculating the energy of the ensemble.

Random changes were then made to the coordinates of one of the particles and the new energy evaluated. They showed that a sufficient criteria for (1.1) to be obtained for long times was to accept the change unconditionally if it decreased the energy and with a probability

$$P = \exp(-\Delta E/k_B T) \quad (1.2)$$

if the energy was increased.

To attain thermal equilibrium at lower temperatures it is best not to apply the Metropolis algorithm straight away at the temperature considered. Instead, one should attempt to simulate annealing, the process in which the system is cooled slowly, being allowed to approach thermal equilibrium at a given temperature before the temperature is lowered by some small amount. Slow cooling is particularly important in such systems as spin glasses [2] where the Hamiltonian has a large number of global and local minima. Cooling too quickly means that the disorder encountered at higher temperatures gets frozen in as the temperature is lowered, corresponding to the system sticking in a local minimum of the Hamiltonian.

Kirkpatrick *et al* [3] pointed out that this simulated annealing algorithm could be applied to minimise any complicated function with numerous local and global minima. He termed such a general function the cost function and used as an example the

travelling salesman problem. This consists of finding the shortest route around a collection of points ('cities') on a two-dimensional grid, the route terminating at its starting point. The cost function is simply the length of the route

$$C = \sum_{i=1}^N Lp(i)p(i+1) \quad (1.3)$$

where $P(i)$ is some permutation of N cities and L_{ij} is a matrix of inter-city distances. To solve the problem exactly takes a time which grows factorially with N .

2. The problem

2.1. Deficiencies of the annealing algorithm

Two further specifications must be added to the basic idea. The first is the set of random changes to be attempted or the 'dynamics' of the system. Clearly, the better one understands the nature of an individual problem, the more astutely these can be chosen.

We deliberately chose a relatively 'dumb' form of change for our study to avoid the simplicity of the travelling salesman problem proving too much a special case. Similarly, we did not concentrate our attention on starting with a 'greedy' initial state, that is where the next city visited on the route is the closest.

The second specification is that of the annealing schedule.

Since Kirkpatrick a number of papers on simulated annealing have appeared [4-7]. None of these however has addressed systematically the question of how fast the simulation should be 'cooled' and by how large steps the temperature may be dropped to optimise the configuration achieved with limited computer time.

To get a feel for the problem at hand, consider figure 1. The lower curve represents the equilibrium curve we might expect to obtain by making an arbitrarily large number of alterations to the cost function at each temperature and making the temperature drops effectively infinitesimal. What happens in any practical simulated annealing schedule is that the cost function initially situated at point A finds itself at B when

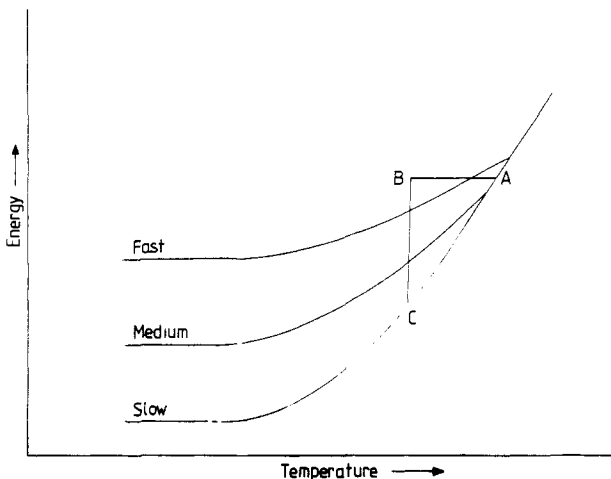


Figure 1. Schematic of equilibrium cooling curve for glasses.

the temperature is dropped suddenly by a finite amount. Application of the Metropolis algorithm is meant to get it back into equilibrium at C.

In the annealing of real glasses, the relaxation time τ taken to fall from B to C at temperature T is governed by a phenomenological law variously known as the Vogel-Fulcher law, Doolittle's equation or the WLF equation [8-10]:

$$\tau = \tau_0 \exp [A/(T - T_g)] \tag{2.1}$$

for appropriate choices of the constants τ_0 , A and T_g , the latter constant being known as the glass transition temperature.

We wished first to establish an algorithm for measuring τ in an annealing and hence the rate at which one can cool, remaining close to equilibrium. Further, we investigated whether it obeyed the form (2.1), in which case determining the parameters τ_0 , A and T_g would suffice to design the annealing schedule.

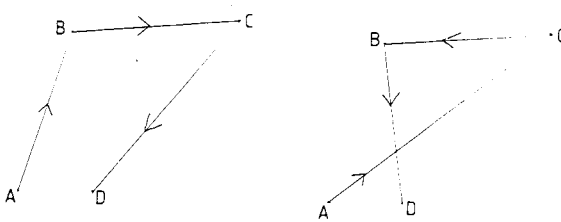


Figure 2. Depiction of 'two-bond' move employed in the annealing schedule. Route ABCD goes to ACBD.

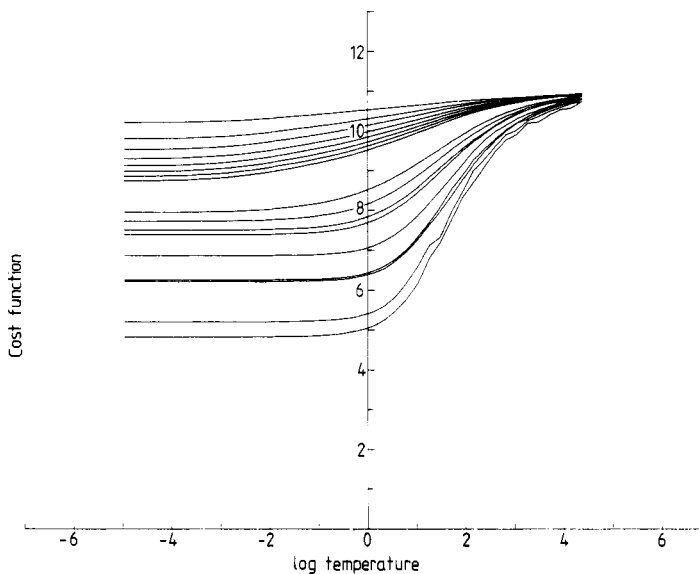


Figure 3. Cooling curves for salesman problem over 50 temperature steps for various cooling rates. The leftmost number is the number of changes made to the cost function per temperature. The right-hand number is the number of runs made for averaging purposes. Cost function is measured in units of cost/city. (In descending order, the labels for the curves are: 1×80000 , 2×40000 , 3×20000 , 5×16000 , 6×13000 , 7×10000 , 8×10000 , 30×2500 , 50×1500 , 80×1000 , 100×800 , 300×200 , 800×100 , 1000×80 , 10000×8 , 80000×1 .)

In figure 1, the two upper curves represent the average energy of the glass for faster cooling rates. The quicker one cools the glass the higher the energy at which it finally levels off becomes.

If an equation analogous to (2.1) held for annealing the salesman problem then we might expect to obtain cooling curves as in figure 1 for this problem also.

2.2. Cooling curves

The set up for our travelling salesman problem was as follows. 100 cities were distributed randomly on a flat surface whose X and Y ranges both stretched from -10 to 10 units. The coordinates were otherwise completely unrestricted real numbers. An inter-city distance look-up table was formed to provide quick and easy reference to the Euclidean distance between any two cities.

Making lots of random changes to the cost function we found the average change. This would be equivalent to the average change at infinite temperature when all moves would be accepted. A change consisted of the 'two-bond move' depicted in figure 2. This average change was about 10 units and since the Boltzmann factor k_B was set to 1 a suitable choice of starting temperature was $T \gg 10$ and we took $T = 100$.

Cooling curves for this problem are shown in figure 3. In each case the temperature was lowered through 50 decrements of 20%, for various number of attempted moves per decrement. Averaging was taken over repetitions with different random initial routes but the same set of city locations.

The cooling curves do have a structure similar to those of glasses, the slower cooling rate curves levelling off at a lower valued final cost function than those for faster cooling.

3. Relaxation times

3.1. Autocorrelation functions and power spectra

To obtain relaxation times we considered the thermal equilibrium autocorrelation function defined by

$$\langle C(t_1)C(t_1+t) \rangle = A + BF(t) \quad (3.1)$$

where $C(t_1)$ is the time-dependent cost function, the average being taken over all t_1 . One can show that if F is normalised such that

$$F(0) = 1 \quad (3.2)$$

then A and B are given by

$$A = \langle C(t_1) \rangle^2 \quad (3.3a)$$

$$B = \langle C^2(t_1) \rangle - \langle C(t_1) \rangle^2 = \sigma^2(T). \quad (3.3b)$$

We make the assumption that the time averages can be replaced by averages over an ensemble of repetitions of the simulation, taken at some particular time $t = 0$, and further we can write the autocorrelation function in terms of deviations from the equilibrium average ΔC_i

$$\langle \Delta C_i(0) \Delta C_i(t) \rangle_i = \sigma^2(T)F(t). \quad (3.4)$$

At large times, $\Delta C(0)$ and $\Delta C(t)$ will be unrelated so that the ensemble average will be zero. Thus the autocorrelation function decays and one may be tempted to write

$$\langle \Delta C_i(0) \Delta C_i(t) \rangle_i = \sigma^2(T) \exp(-t/\tau(T)) \quad (3.5)$$

although there is, of course, no *a priori* reason for assuming that the decay is simple exponential. The well known fluctuation dissipation theorem indicates that the time dependence of the fluctuations which we impose is the same as the time dependence of thermal fluctuations.

The power spectrum of the cost function $C(t)$ is given by

$$P(\omega) = \text{Re} \int \langle C(t_1) C(t_1+t) \rangle \exp(i\omega t) dt \quad (3.6)$$

and in the simplest case of equation (3.5) we would have at a particular temperature

$$\sigma^2(T)/P(\omega) = \omega^2 \tau + 1/\tau. \quad (3.7)$$

We make the more general assumption that the time-dependent part of the autocorrelation function, $f(t)$, can be written as a sum over exponentials

$$F(t) = \int A(\tau_1) \exp(-t/\tau_1) d\tau_1. \quad (3.8)$$

If this assumption holds and we calculate the gradient and intercept of a general $1/P$ against ω^2 curve close to $\omega = 0$, we obtain

$$\lim_{\omega \rightarrow 0} \frac{1}{P(\omega)} = \left(\int A(\tau_1) \tau_1 d\tau_1 \right)^{-1} = \frac{1}{\langle \tau_1 \rangle} \quad (3.9)$$

$$\lim_{\omega \rightarrow 0} \frac{d}{d(\omega^2)} \left(\frac{1}{P(\omega)} \right) = \frac{\langle \tau_1^3 \rangle}{\langle \tau_1^2 \rangle}. \quad (3.10)$$

Each of these measures, in some sense, give the overall relaxation time for the problem. The gradient measure gives more bias towards the longer time components of the process.

3.2. Numerical experiments to obtain relaxation times

In order to take power spectra from a time sequence of data, it is necessary in the initial instance to start with cost function which is in thermal equilibrium. To do this we assume that the lower curve in figure 3 represents a sequence of equilibrium states. Thus an annealing was repeated as in § 2.2 for the cooling rate corresponding to the lower curve and the complete route or microstate recorded at the end of each temperature step. Twenty of the states sampled at various points over the range $T = 100$ to $T = 0$ were then used for analysis as below.

The Metropolis algorithm was applied to each sample in turn at the relevant temperature and the value of the cost function monitored as a function of time. One does not expect significant correlations in the changes of cost function over less than 1 'sweep' or 1 attempted change per city, so we adopted 1 sweep as our unit of time.

Every time unit the value of the cost function was recorded until we had 1024 values whose power spectrum was then computed from the Fourier transform of the time sequence data. In order to get reasonable statistics, the process was repeated 50 times for each temperature and the average of the power spectra obtained. An example

of such an average power spectrum is shown in figure 4. In figure 5, a plot of $1/P$ against ω^2 is shown for a temperature $T = 80$ and $T = 10$. All power spectra have been normalised by σ^2 .

Even with the averaging, there is still too much noise associated with the data for accurate measurement of either gradient of intercept, so various smoothing operations were performed. A smoothed version of the $1/P$ against ω^2 curve, henceforth called

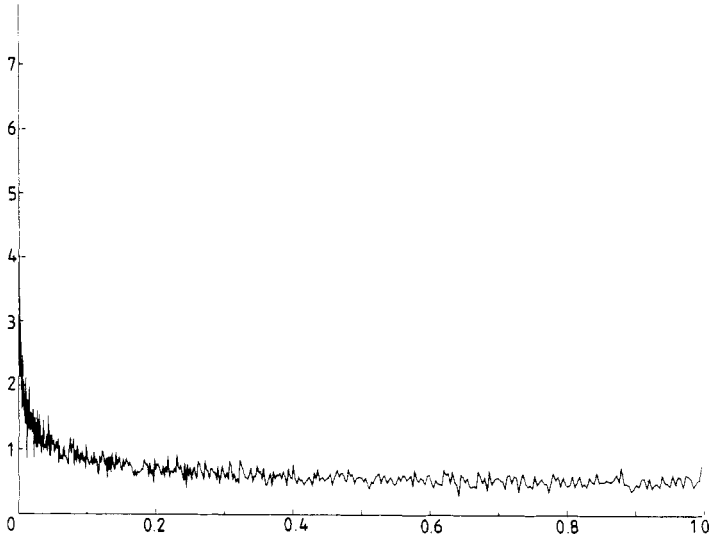


Figure 4. A power spectrum obtained at temperature $T = 80$.

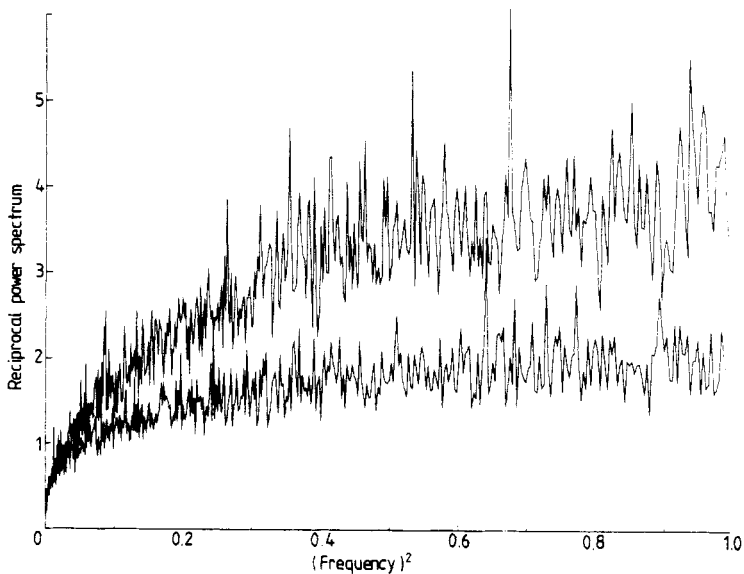


Figure 5. $1/\text{power}$ plotted against frequency squared for temperature 80 (upper curve) and 10 (lower curve).

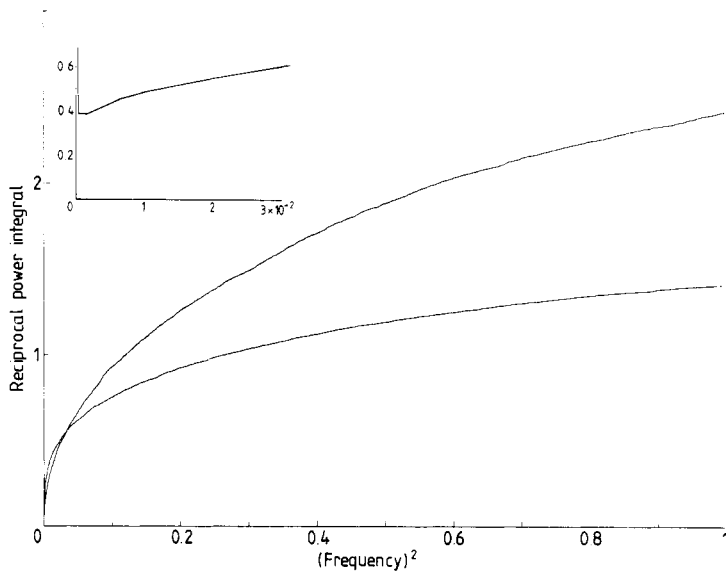


Figure 6. Binned reciprocal power integral plotted against frequency squared for temperatures 80 (upper curve) and 10 (lower curve). The inset shows the $T = 80$ cure expanded in the region $(0, 0.03)$.

a reciprocal power integral, was defined according to the formula

$$S(\omega) = \frac{1}{\omega} \int_0^\omega \frac{1}{P(\omega)} d\omega. \tag{3.11}$$

This introduces accumulation of the data whilst still being dominated by the frequency of interest, since $1/P(\omega)$ increases with ω .

The noise associated with the region of interest, the low frequency region, was still unacceptable so $S(\omega)$ was further refined by a binning procedure whereby the average of 10 consecutive values of $S(\omega)$ was found to yield a new variable $S_1(\omega)$ defined by

$$S_1(\omega_i) = \frac{1}{10} \sum_{j=i-10}^i S(\omega_j). \tag{3.12}$$

One can easily show that $S_1(\omega)$ close to $\omega = 0$ still has a gradient and intercept given by equations (3.9) and (3.10) to within appropriate numerical factors.

Figure 6 shows, superimposed, the binned functions $S_1(\omega)$ obtained for temperatures $T = 80$ and $T = 10$. The inset shows the binned function $S_1(\omega)$ in the range $\omega^2 = (0, 0.03)$ for temperature $T = 80$ providing an example of the data we used for measuring gradients and intercepts.

4. Results

The results which we obtained for the relaxation times derived respectively from measuring the intercept and gradient are shown in figures 7(a) and (b), displayed on an Arrhenius plot.

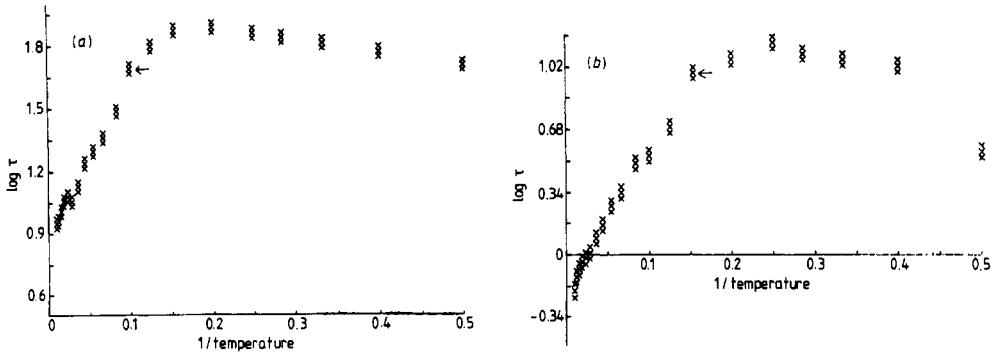


Figure 7. An Arrhenius plot of $\ln(\text{relaxation time})$ against reciprocal temperature, (a) obtained from intercept and (b) gradient measurements, respectively.

Generally, we can write an expression for τ as

$$\tau = \tau_0 \exp(A(T)/T) \quad (4.1)$$

where $A(T)$ is the temperature-dependent activation energy.

One can see from the plots that in the high-temperature regime A is independent of temperature which one would expect even if a Vogel-Fulcher law held. Vogel-Fulcher behaviour, however, would result in the curve bending away from linear behaviour and tending towards infinity as the glass transition temperature is approached. However, close to T_g , the divergence of the relaxation time would imply that a prohibitively large amount of CPU time would need to be used to produce an equilibrium microstate of the cost function. One can see from the lower cooling curve in figure 3, which depicts the sample we assumed to be in equilibrium, that it is in the region below about $T = 5$ that the cost function begins to quench out corresponding to thermal equilibrium being lost. This corresponds to the region on the Arrhenius plots at which the relaxation time begins to level out.

One is tempted to imagine that the single points labelled by arrows on the Arrhenius plots represent the initial stages of approach towards a glass transition. However, we do not believe that we have sufficient evidence to claim that a Vogel-Fulcher law is obeyed for this problem.

We assume that the behaviour is Arrhenius and measurements conducted on the linear portions of the plots give expressions for the relaxation time for the intercept and gradient measures, respectively, as being

$$\tau = (2.46 \pm 0.04) \exp[(6.85 \pm 1.05)/T] \quad (4.2a)$$

$$\tau = (0.809 \pm 0.003) \exp[(7.91 \pm 0.59)/T]. \quad (4.2b)$$

Note that the activation energies agree within error bars.

5. An optimum annealing schedule

We took

$$\tau = 2.5 e^{7/T}. \quad (5.1)$$

If our philosophy is correct then to accomplish an optimum annealing schedule, we apply the Metropolis algorithm at temperature T for a number of sweeps τ before

the next temperature drop is made. There is a problem, however, about the size of the temperature drops. Reference to equation (3.4) shows that τ is the time required for the time-dependent part of the autocorrelation function to decay from $\sigma^2(T)$ to around zero. Using the formula for the specific heat

$$\frac{d\langle c \rangle}{dT} = \frac{\sigma^2(T)}{T^2} \quad (5.2)$$

we can, assuming that the temperature drop ΔT is small and the corresponding change in the mean cost $\sigma(T - \Delta T)$ is almost equal to $\sigma(T)$, derive an expression for ΔT as

$$\Delta T = T^2 / \sigma(T). \quad (5.3)$$

However, particularly at high temperatures when fluctuations in the cost are likely to be large, the assumption of ΔT being small is not a valid one. An alternative way to use the relaxation time to optimise the CPU time is to cool continuously according to the formula

$$\delta T = \frac{\epsilon T}{\tau(T)} \delta t \quad (5.4)$$

for a suitable choice of ϵ and $\delta t = 1$ sweep. Thus, the system is given a time τ to respond to the temperature dropping by a constant fraction ϵ .

We did various annealings to determine how this strategy compared with other types of annealing schedule. In each of the schemes considered, the annealing was to take in a total of 4×10^5 changes and results obtained from averaging over 10 runs.

In the first set of annealings, the temperature was dropped according to the formula

$$T_n = \chi_T T_{n-1} \quad (5.5)$$

for various values of χ_T , the changes being divided equally between the temperatures. In the second set, the temperature was simply lowered by a constant decrement, D :

$$T_n = T_{n-1} - D \quad (5.6)$$

the changes again being equally spread between the temperatures.

Thirdly, the scheme outlined in equation (5.4), henceforth termed 'continuous cooling', was tried for various values of ϵ .

Values of the final cost function C_f obtained together with the final temperature T_f are shown in table 1. One can see that for appropriate choice of the parameters χ_T , D , ϵ all methods reach a final value of around 4.7. This fact, coupled with the evidence of the cooling curves (figure 3) suggests that the global minimum of the cost function is around 4.7. To determine the efficiency or otherwise of continuous cooling as an annealing schedule, it was necessary to determine after what proportion of the 400 000 moves the global minimum was attained. Table 2 shows the proportion of the moves needed for the cost function to get below a value of 5. Note that we are considering the number of moves for the cost function to get below 5 on average. Thus, although in principle it is possible for the average cost function to move off into less favourable parts of the phase space after this value is achieved, it in fact does not do so. Continuous cooling with $\epsilon = \frac{1}{50}$ achieves this most quickly.

In a practical annealing schedule one cannot evaluate the expression for τ using the detailed methods outlined in this paper. Ideally one would like to be able to calculate a running estimate of τ quickly as the annealing proceeds. We can in principle do this at any time using past values of the cost function. From the formalism in § 3, one can show that

Table 1. Table showing final value of cost function and temperature for various annealing schedules. Each result is taken from an average of 10 runs of 400 000 moves.

Schedule									
1	X_T	0.95	0.9	0.8	0.7	0.6			
	$T_n = \chi_T T_{n-1}$	C_f	7.47	4.78	5.01	5.04	5.56		
		T_f	8.10	0.57	0.00	0.00	0.00		
2	D	1	2	3	8	16			
	$T_n = T_{n-1} - D$	C_F	4.74	5.05	4.85	6.11	6.16		
		T_f	0.00	0.00	0.00	4.00	4.00		
3	$1/\epsilon$	1	5	10	50	100	1000	10 000	
	$T_n = T_{n-1} - \frac{\epsilon T_{n-1} \Delta t}{\tau(T_{n-1})}$	C_f	6.36	5.62	5.39	4.74	4.69	9.04	9.91
		T_f	0.74	0.92	1.03	1.46	1.95	24.01	82.93

Table 2. The proportion of 400 000 moves required for the average cost over 10 repetitions to get below $C_f = 5$ for each of the best annealing schedules in table 1.

Method	1 $\chi_T = 0.9$	2 $D = 1$	3 $\epsilon = 1/50$	3 $\epsilon = 1/100$
Proportion of moves to reach global minimum	0.76	0.98	0.39	0.57

$$\begin{aligned}
 \tau &= \int_0^\infty \left(\frac{\langle c(t_1)C(t_1+t) \rangle - \langle c(t_1) \rangle^2}{\langle C^2(t_1) \rangle - \langle C(t_1) \rangle^2} \right) dt \\
 &\approx - \int_0^\alpha \left(\frac{\langle C(t_1)C(t_1-t) \rangle - \langle c(t_1) \rangle^2}{\langle C^2(t_1) \rangle - \langle C(t_1) \rangle^2} \right) dt
 \end{aligned}
 \tag{5.7}$$

where the time period α must be at least of the order of τ .

However, since the various moments in this expression are meant to be calculated from a sequence of cost functions sampled at one temperature and yet we change the temperature every time a new cost function is recorded, equation (5.7) is not of much use in practice.

A number of schemes can be devised in which equation (5.7) can be used to estimate τ from cost functions sampled across a range of temperatures but we have not attempted to implement any of these.

6. Conclusion

We have shown that for the process of simulated annealing of a travelling salesman cost function, the relaxation time as a function of temperature can be measured and is given, within our resolution, by an Arrhenius type law. We have shown that we can utilise this relaxation time to devise an annealing schedule in which the time for the

global minimum to be reached is smaller than in previously used *ad hoc* methods by a factor of two in one case study.

Acknowledgments

We would like to thank Professor Sir Sam Edwards for his criticism and advice. S Rees would like to thank the SERC and BP research, Sunbury upon Thames, for the provision of a CASE award.

References

- [1] Metropolis N, Rosenblum A W, Rosenblum M N and Teller A H 1953 *J. Chem. Phys.* **21** 1087
- [2] Edwards S F and Anderson P W 1975 *J. Phys. F: Met. Phys.* **5** 965
- [3] Kirkpatrick S, Gelatt C D and Vecchi M P 1983 *Science* **220** 671
- [4] Edwards S F 1983 *Inst. Math. Appl.* **19** 162
- [5] Vannimenus J and Mezard M 1984 *J. Physique Lett.* **45** 1145
- [6] Lundy M and Mees A 1986 *Convergence of the annealing algorithm* to be submitted
- [7] Kirkpatrick S and Toulouse G 1985 *Configuration space analysis of travelling salesman problem, IBM Res. Rep.* RC 10972 (no 42918)
- [8] Vogel H 1921 *Phys. Z.* **22** 645
- [9] Fulcher G S 1925 *Am. Ceram. Soc.* **6** 339
- [10] Williams M, Landel R and Ferry J D 1955 *J. Am. Chem. Soc.* **77** 3701
- [11] Doolittle A K 1951 *J. Appl. Phys.* **22** 1471