The optimal simulated annealing schedule for a simple model

1990 J. Phys. A: Math. Gen. 23 3511

(http://iopscience.iop.org/0305-4470/23/15/023)

J. Phys. A: Math. Gen. **23** (1990) 3511-3523. Printed in the UK

The optimal simulated annealing schedule for a simple model

Karl Heinz Hoffmann and Peter Salamon†

Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 19, D-6900 Heidelberg, Federal Republic of Germany

Received 23 October 1989

Abstract. Used as a tool for large scale global optimisation, simulated annealing incurs heavy computational costs. Therefore, choosing an optimal cooling schedule is of great scientific and economic importance. For the first time an analytic as well as a numeric solution to this problem is presented, albeit only for a small example system. The example shows the role of optimal control theory for this problem.

1. Introduction

Simulated annealing is a technique [1, 2] for finding optimal and near-optimal solutions to difficult optimisation problems. It has been especially useful in the context of NP-complete combinatorial optimisation problems [3] and it exploits an analogy between these problems and physical systems. Careful annealing of a real physical system should bring it into its state of equilibrium with the ambient temperature T and thus for $T \to 0$ the system moves into its ground state(s). Similarly, the proper simulation of this procedure treating an optimisation problem as a physical system should result in the simulation finding the optimal solution.

Simulated annealing has been applied to a wide range of problems characterised by having a complex state space structure often due to various constraints. The complexity of the structure is due to numerous local minima whose numbers are often exponential in the number of degrees of freedom. The problem of finding the ground state of a spin glass is a problem of this kind from condensed matter physics [4]. Simulated annealing has also proved a useful tool in the design of integrated circuits [1, 5, 6], for partitioning, routing, and placement [7]. It has been applied to many other problems including the travelling salesman [8, 9], graph partitioning [10], restoration of images [11], and parameter estimations [12]. While this list is far from exhaustive, it shows that the problems attacked by simulated annealing are of great scientific and industrial importance.

Simulated annealing requires large amounts of computer time. Lowering the computational costs is a natural way to improve the algorithm [6, 11, 13-16]. Apart from changing the algorithm [17] or choosing special move classes [18], the central question is: what is the optimal annealing schedule?

Attempted answers introduce some additional assumptions [10, 19] or proceed in a purely heuristic way [16]. While suggested schedules can be very valuable, we would like to consider the problem of finding the optimal annealing schedule without additional assumptions. We succeed in finding the optimal schedule, albeit only for a very

† Permanent address: Department of Mathematical Sciences, San Diego State University, San Diego, CA 92182, USA.

0305-4470/90/153511+13$03.50

simple example. The example system consists of only three states and shows how the crossing of a single barrier is optimised. As the state space of the problems optimised by simulated annealing have many barriers, this paper thus provides only a first step towards the general case. The aim of our analysis is to obtain insight into the mechanisms governing the optimal schedule and to determine its connection to system parameters. Should such efforts succeed, methods exist [20] for determining the values of the system parameters using data gathered during an annealing run for actual optimisation problems.

2. Simulated annealing—a Markov process

Simulated annealing is based on the Monte Carlo simulation of physical systems. As such it uses the Metropolis algorithm [21] which involves a biased random walk through the state space of the system. Let $\Omega = \{\omega\}$ represent this finite state space, let $E : \Omega \to R$ be the cost function (energy) defined on this state space, and let T be the adjustable parameter in the algorithm representing the temperature of the heat bath in which the corresponding physical system is immersed. At each step of the algorithm, a neighbour ω' of the current state ω_k is selected at random to be a candidate for becoming the next state. It actually becomes the next state only with probability

$$P_{\text{acceptance}} = \begin{cases} 1 & \text{if } \Delta E \leqslant 0 \\ \exp(-\Delta E/T) & \text{if } \Delta E > 0 \end{cases} \tag{2.1}$$

where $\Delta E = E(\omega') - E(\omega_k)$. If this candidate is accepted, then $\omega_{k+1} = \omega'$, or else the next state is the same as the old state, $\omega_{k+1} = \omega_k$. Thus to complete the definition of the dynamics for the algorithm, we must specify (i) the schedule of temperatures T_k as a function of time (i.e. the number of Metropolis steps performed), and (ii) a definition of which states are to be considered neighbours.

The latter is known as the move class and defines an undirected graph structure on the state space. We will denote by $N(\omega)$ the set of neighbours of a state ω in this graph. As an example we mention the case of an Ising spin glass, where two states (i.e. spin configurations) are defined to be neighbours if they differ by one spin flip.

The matrix $\Pi = (\Pi_{\beta\alpha})$ of infinite-temperature transition probabilities from state α to β is defined by

$$\Pi_{\beta\alpha} = \begin{cases} 0 & \text{if } \beta \notin N(\alpha) \\ 1/|N(\alpha)| & \text{if } \beta \in N(\alpha) \end{cases} \tag{2.2}$$

where $|N(\alpha)|$ is the number of neighbours of α. These are the transition probabilities if the algorithm automatically accepts each attempted move, i.e. if $T = \infty$. At finite temperature the acceptance decision is superimposed on Π in (2.2) to give $\Gamma(T)$ defined by

$$\Gamma_{\beta\alpha} = \begin{cases} \Pi_{\beta\alpha} \exp(-\Delta E/T) & \text{if } \Delta E > 0,\ \alpha \neq \beta \\ \Pi_{\beta\alpha} & \text{if } \Delta E \leqslant 0,\ \alpha \neq \beta \\ 1 - \sum\limits_{\xi \neq \alpha} \Gamma_{\xi\alpha} & \text{if } \alpha = \beta \end{cases} \tag{2.3}$$

where now $\Delta E = E(\beta) - E(\alpha)$.

From a given move class $\{N(\omega)\}$ the stationary distribution of $\Pi = \Gamma(\infty)$ can be determined. Typically, one would like to choose a move class such that the stationary

distribution at infinite temperature is uniform [22]. An easy way to achieve this is to make $|N(\omega)|$ constant over ω, i.e. to make each vertex have the same number of neighbours. This property is expressed in the language of graph theory [23] by saying that the graph is regular. Assuming the graph to be connected, it is a regular and undirected graph if and only if Π is symmetric.

In the discussion below, we will find it convenient to lump [20] states together which are energetically as well as dynamically equivalent. The new states will have a higher degeneracy and this will be represented in the stationary distribution of the lumped Π matrix.

From a technical point of view, the simulated annealing procedure introduced above is thus a discrete time Markov process with time-dependent transition probabilities. After specifying an annealing schedule T_k, $k = 1, \ldots$, the dynamics is given by

$$\boldsymbol{p}(k) = \boldsymbol{\Gamma}(T_k)\boldsymbol{p}(k-1) \tag{2.4}$$

which describes the time evolution of the probability distribution $\boldsymbol{p}(k)$ of a random walker in the state space of the system. Note that in the most general case treated here the temperature T_k can be reset after each Metropolis step.

In this picture the idea behind simulated annealing becomes clear: by construction, $\boldsymbol{p}(k)$ will tend to a Boltzmann distribution at the temperature T_k if the temperature is kept fixed. The probability of being in low energy states is increased for successively lower temperatures and long enough cooling times. The ultimate hope then is to have all the weight of the probability distribution in the ground state(s) for an appropriate cooling schedule in which T goes to zero.

3. Optimisation criteria

Before an optimal annealing schedule can be determined, we must choose an optimality criterion. Indeed several criteria are possible:

(a) the probability to be in the ground state;

(b) the final energy;

(c) the best so far (BSF) energy.

The probability to be in the ground state has mainly been investigated in the context of infinite time schedules [24, 25]. One should note, however, that as an optimising goal for finding near optimal solutions within a limited time it is only useful if it also increases the probability to be in other low lying states.

While the probability to be in the ground state is a number, the final and the BSF energy [12, 26] are stochastic quantities. Consider an actual simulation of the annealing. Starting from a random initial state ω_0, the stochastic sequence ω_k is chosen according to the Metropolis algorithm. The BSF energy of a given path up to step N is given as

$$E_{\mathrm{BSF}} = \mathrm{Min}_{0 \leqslant k \leqslant N}\{E(\omega_k)\} \tag{3.1}$$

and describes the lowest energy found on that path. The final energy is the energy of the last state ω_N. The probability distributions of the final energy and of the BSF energy change with time; their evolution is induced by the underlying random walk in the state space.

From the viewpoint of the analogy with statistical mechanics, the mean of the final energy is the most natural criterion of merit. It is computationally simple as it depends only on the distribution of the system at a single time.

The BSF energy is more complicated as it depends on the whole path of a random walker. On the other hand it is a more natural choice for simulated annealing, where one is seeking the lowest energy state.

Which objective to optimise is a subjective decision. The criterion of merit for commercial applications should go beyond the above criteria and should use an objective which includes the price of computer time and the financial gain per unit improvement in the value of the minimal energy found by the simulated annealing procedure.

Due to its greater simplicity we have in this paper chosen the mean final energy as a criterion for the optimal schedule problem.

4. The optimal annealing schedule—an optimisation problem

Using the energy vector $\boldsymbol{E} = (\ldots, E_\alpha, \ldots)$, with E_α being the energy of state α, the formal problem of finding an optimal schedule to minimise the mean final energy in a finite number of Metropolis steps N can be written as

$$\text{Min } \bar{E} = \min_{\{T_k\}_{k=1,\ldots,N}} \boldsymbol{E} \prod_{k=1}^{N} \boldsymbol{\Gamma}(T_k)\boldsymbol{p}(0) \tag{4.1}$$

with $\boldsymbol{p}(0)$ being the initial distribution.

For known energy function and move class the solution of this optimisation problem can be addressed. However the transition probability matrices for real problems are far too large for explicit calculation and moreover only little is known about the systems. Nonetheless we investigate the optimal annealing schedule for a completely known system. The immediate goal is to establish optimal schedules for the example system for different input parameters. Those schedules can be compared with other frequently used schedules like the linear and the exponential schedule thus giving indications of the possible gains. Moreover it allows us to decide whether the constant thermodynamic speed schedule [10, 19] discussed in more detail in section 8 is always optimal. But apart from this short term goal there is also a long term goal connected with our approach.

By calculating the optimal annealing schedule explicitly for a number of systems we hope to get some insight into which features of the system are relevant for the optimal schedule. As pointed out above, many of the already used schedules do depend on certain system parameters—our approach should shed some light on whether these are the correct ones. Once the relevant quantities are known, the question of how they can be determined for an actual system under consideration has to be attacked. The lumping procedure described in [20] is an example of such an effort.

5. The example system

Our example is a special four-state system which can be reduced by lumping to a three-state system. It is the simplest system with a move class represented by a regular graph which possesses one local and one global minimum. Figure 1 shows the four-state

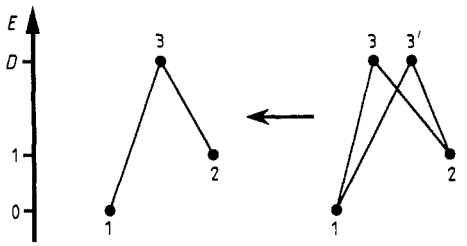

Figure 1. The four-state system shown on the right is the simplest system with a move class represented by a regular graph and with one local and one global minimum. As the energies of states 3 and 3′ are chosen to be equal, one can lump the system to the three-state system shown on the left.

system and its lumped three-state version. The state space consists of the states 1, 2, 3 and 3′ with energies:

$$\begin{aligned} E(1) &= 0 \\ E(2) &= 1 \\ E(3) &= E(3') = D \geqslant 1. \end{aligned} \tag{5.1}$$

As states 3 and 3′ are energetically equivalent and as they are connected to the same other states with the same transition probabilities, they can be combined to one new state with degeneracy 2. Using the abbreviation $x = \mathrm{e}^{-1/T}$ the transition probability matrix $\boldsymbol{\Gamma}$ implementing the Metropolis algorithm is given by

$$\boldsymbol{\Gamma}(x) = \begin{pmatrix} 1-x^D & 0 & \frac{1}{2} \\ 0 & 1-x^{D-1} & \frac{1}{2} \\ x^D & x^{D-1} & 0 \end{pmatrix}. \tag{5.2}$$

The dynamics of the probability distribution $\boldsymbol{p}$ in the state space is then described by

$$\boldsymbol{p}(k) = \boldsymbol{\Gamma}(x_k)\boldsymbol{p}(k-1). \tag{5.3}$$

Using the energy vector $\boldsymbol{E} = (0, 1, D)$ we have to determine

$$\operatorname{Min} \bar{E} = \operatorname*{Min}_{\{x_k\}_{k=1,\ldots,N}} \boldsymbol{E} \prod_{k=1}^{N} \boldsymbol{\Gamma}(x_k)\boldsymbol{p}(0). \tag{5.4}$$

As initial distribution $\boldsymbol{p}(0)$ we used Boltzmann distributions at different initial temperatures. We investigated this problem first numerically and then analytically, using two different approaches.

6. Results

Optimal schedules were determined numerically for a varying number of steps N, for different activation energies D, and for different initial distributions.

Figure 2 shows the optimal annealing schedules for different initial distributions. As the probability distribution for the three-state system is normalised, every distribution corresponds to one point in the p_1, p_2 plane. The bold curve in figure 2 corresponds to the family of Boltzmann equilibrium distributions at different temperatures T. For

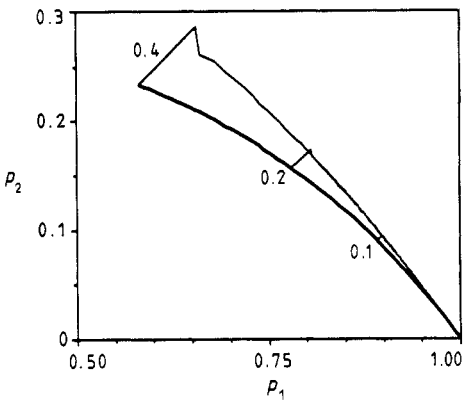

Figure 2. For the three-state system of figure 1, the paths created by the optimal cooling schedules for different initial temperatures ($e^{-1/T} = 0.1$, 0.2 and 0.4) are shown. The bold curve marks the equilibrium states. Note the turnpike behaviour of the optimal paths. After the first few steps the optimal schedules lead the system always along the same curve.

a given initial distribution the optimal annealing schedule determines a sequence of probability distributions as a function of the step number. In figure 2 we have plotted three of these sequences for different initial distributions ($e^{-1/T} = 0.1$, 0.2 and 0.4). For each of these sequences, successive points have been connected by straight lines.

The most remarkable and unexpected feature is the 'turnpike' behaviour seen in the figure. By 'turnpike' we mean that the optimal path—after a short transition period lasting only a few steps—is independent of the initial distribution. Moreover it turns out that it is independent of the duration N of the annealing schedule. This means that the optimal schedule drives the system in a few steps onto the turnpike. Thereafter any additional time (number of steps) is used to proceed further along the turnpike towards that probability distribution which has all its weight in state 1. Only at the very end of an annealing schedule does the system leave the turnpike again.

We now discuss the optimal path and how it can be characterised in more detail. The numerical results show that the optimisation policy for the last step should be $x = 0$, i.e. zero temperature. This can easily be understood: while all of the probability in state 3 is transported to states 1 and 2 independent of the temperature, any finite temperature will shift some probability from states 1 and 2 to state 3. This increases the final energy as can analytically be seen from

$$\bar{E} = \boldsymbol{E}\boldsymbol{\Gamma}\boldsymbol{p}(0). \tag{6.1}$$

Also the policy for the next to the last step can be analytically determined:

$$\mathrm{Min}_x\, \bar{E} = \mathrm{Min}_x\, \boldsymbol{E}\boldsymbol{\Gamma}(0)\boldsymbol{\Gamma}(x)\boldsymbol{p}(0). \tag{6.2}$$

This results in the following optimal x

$$x = \frac{D-1}{D}\,\frac{p_2(0)}{p_1(0)}. \tag{6.3}$$

Note that x depends only on the ratio p_2/p_1.

A careful inspection of the numerical results led us to make the following assumptions for the remaining schedule.

(i) At any given point along the optimal path, the optimal policy for the next step is only a function of the current state.

(ii) The probabilities enter the optimal policy only as the ratio p_2/p_1.

Thus we make the ansatz

$$x(k) = b_1 \frac{p_2(k-1)}{p_1(k-1)} + b_2 \left(\frac{p_2(k-1)}{p_1(k-1)} \right)^2 . \tag{6.4}$$

In addition we assume that the initial distribution has most of its weight in the low energy state

$$p_2 = \gamma p_1 \tag{6.5}$$

$$p_3 = \gamma_3 \gamma^2 p_1 \tag{6.6}$$

with $\gamma \ll 1$ and γ_3 of order one. The unknown constants b_1 and b_2 were determined by minimising the average final energy $\bar{E}$ analytically for different number of steps N. Expanding $\bar{E}$ in powers of γ and taking only terms up to order M gave the results displayed in table 1 for the various activation energies. Note that b does not depend on N. As these calculations become quite lengthy for a larger number of steps, the results have been obtained by using the analytic calculation package REDUCE.

The examples in the table show that

$$b_1 = \frac{D-1}{D} \tag{6.7}$$

which matches the coefficient of p_2/p_1 in the policy for the last step. For higher orders in γ, b depends on the initial distribution.

Table 1. The coefficients b_1 and b_2 of the optimal policy for the annealing of the example system in figure 1 are shown for various activation energies D. Note that b_1 equals $(D-1)/D$.

D	b_1	b_2	M
2	$\frac{1}{2}$	$-\frac{1}{4}$	3
3	$\frac{2}{3}$	0	4
4	$\frac{3}{4}$	0	5
5	$\frac{4}{5}$	0	6

We determined the thermodynamic speeds for the various optimal schedules. We found that the thermodynamic speed was not constant along the schedules. Up to now we have not been able to detect any systematic relation between the thermodynamic speed and other properties of the optimised schedules. So far we can draw only one conclusion: constant thermodynamic speed schedules cannot always represent the optimal schedules as we have found explicit counterexamples.

7. The optimal annealing schedule as an optimal control problem

In this section we show how optimal control theory [27-30] can be employed to determine the optimal annealing schedule for this example system. A short review of

the technique is given in the appendix. First we change the discrete time description of the dynamics equation (5.3) into a continuous one [24]:

$$\dot{\boldsymbol{p}} = \boldsymbol{A}\boldsymbol{p} \tag{7.1}$$

with

$$\boldsymbol{A} = \boldsymbol{\Gamma} - \mathbf{1}. \tag{7.2}$$

We note that apart from the stationary mode $\boldsymbol{A}$ has a slow mode describing the flow of probability over the barrier and a fast mode which brings p_3 into equilibrium with given p_1 and p_2. To simplify the control problem further we eliminate p_3 adiabatically. Taking into account the normalisation of $\boldsymbol{p}$ we find for p_2

$$\dot{p}_2 = -p_2 x^{D-1}\left(1 - \frac{1}{2(1+x^D)}\right) + \tfrac{1}{2}x^D(1-p_2)\frac{1}{1+x^D}. \tag{7.3}$$

For small temperatures $x \ll 1$ and with a rescaling of time by a factor 2, this can be simplified to

$$\dot{p}_2 = -p_2 x^{D-1} + x^D(1-p_2). \tag{7.4}$$

This is the usual dynamics for a two-level system with Arrhenius factors for jumping over the separating barrier of height D [31].

In this model minimising the mean energy is equivalent to minimising the final p_2. This can be written as

$$\text{Min} \int_0^\tau \mathrm{d}t\, \dot{p}_2 \tag{7.5}$$

where τ is now the total time (= total number of steps) for the process. The dynamics is given by (7.4). Applying optimal control theory we find the Hamiltonian:

$$H = \dot{p}_2 + \lambda \dot{p}_2 = (1+\lambda)\dot{p}_2. \tag{7.6}$$

The optimal policy is obtained by extremising H

$$\frac{\partial H}{\partial x} = \frac{\partial \dot{p}_2}{\partial x} = 0. \tag{7.7}$$

For any dynamics $\dot{p}_2$, the optimal x can only depend on p_2 and not on λ. For our $\dot{p}_2$, we find

$$x = \frac{D-1}{D}\frac{p_2}{1-p_2} = \frac{D-1}{D}\frac{p_2}{p_1} \tag{7.8}$$

which is the policy from our other analytical and numerical results.

Note how quickly the optimal feedback control was obtained by using optimal control theory. This solution of the optimal annealing schedule for the two-level model might be also of interest as a building block for the optimal schedule of much larger systems. Huse and Fischer [31] for instance used a number of independent two-level systems to describe the relaxation in disordered Ising systems. However, it is an open question whether such an approximation is useful for a *global* path optimisation.

We now explore the asymptotic form of our optimal schedule. On substituting x in (7.8) into (7.4) and integrating p_2 for $p_2 \ll 1$ we obtain

$$p_2(t) = \left[t \cdot \left(\frac{D-1}{D} \right)^D + p_2(0)^{-(D-1)} \right]^{-1/(D-1)} \tag{7.9}$$

which for long times

$$t \gg p_2(0)^{-(D-1)} \cdot \left(\frac{D}{D-1} \right)^D$$

reduces to

$$p_2(t) = \left(\frac{D}{D-1} \right)^{D/(D-1)} \cdot t^{-1/(D-1)}. \tag{7.10}$$

Within the approximations made above we find for the optimal schedule

$$x(t) = \frac{D-1}{D} \cdot p_2(t) = \left(\frac{D}{D-1} \right)^{1/(D-1)} \cdot t^{-1/(D-1)} \tag{7.11}$$

and thus for long times

$$T(t) \sim \frac{D-1}{\ln t}. \tag{7.12}$$

As has been previously shown [11, 24], this schedule guarantees convergence to the ground state with probability 1. Note that minimising the mean final energy is equivalent to maximising the likelihood $p_1(\tau)$ of occupying the ground state at the final time and thus the optimal schedules for these two objectives are identical.

8. Comparison with other annealing schedules

In this section we compare the average final energy of a two-level system obtained by the optimal schedule with those obtained by other frequently used schedules. These are the linear schedule

$$T(t) = T_0 - \varepsilon t \tag{8.1}$$

and the exponential schedule

$$T(t) = T_0 \alpha^t. \tag{8.2}$$

We determined the mean final energy by integrating (7.4) numerically for different values of the total available annealing time τ (= number of available Metropolis steps) and for different values of ε and α. In all cases the initial distribution was an equilibrium distribution at $T = \infty$, i.e. $p_2(0) = 0.5$, and $D = 1.2$. The initial temperature T_0 was always taken to equal the initial temperature of the optimal schedule.

Figure 3 shows the mean final energy as a function of ε for the linear schedule. The available annealing time $\tau = 1000$. The distinct minimum corresponds to $\varepsilon = T_0/\tau$. Thus here the temperature should be lowered from its initial value down to zero. The full line corresponds to the energy obtained by the optimal schedule for the same total annealing time.

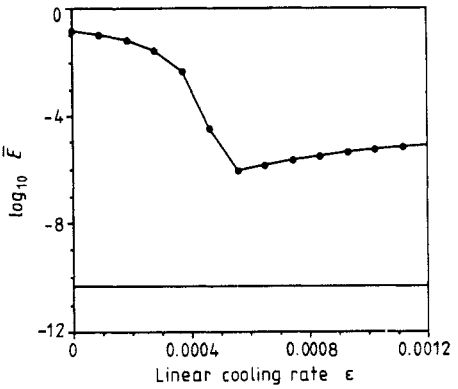

Figure 3. The mean final energy of a two-level system is shown for a linear annealing schedule $T(t) = T_0 - \varepsilon t$ as a function of ε. The full line corresponds to the energy obtained by the optimal schedule for the same total annealing time.

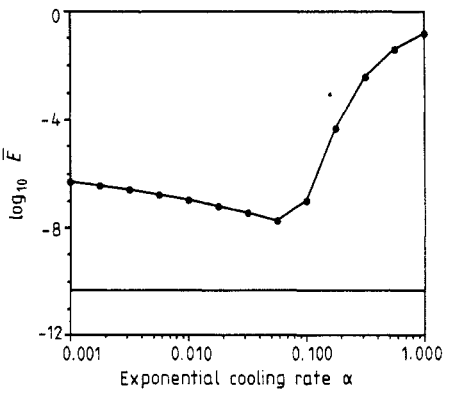

Figure 4. The mean final energy of a two-level system is shown for an exponential annealing schedule $T(t) = T_0\alpha^t$ as a function of α. The full line corresponds to the energy obtained by the optimal schedule for the same total annealing time.

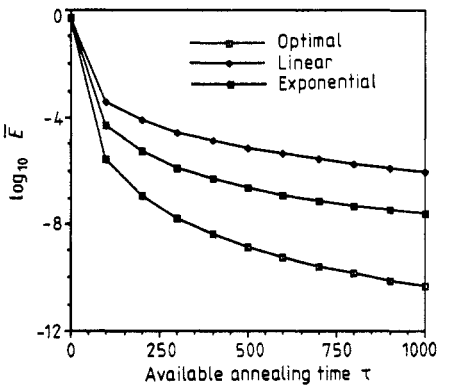

Figure 5. The mean final energies obtained by linear, exponential, and optimal schedules are compared as a function of the total available annealing time τ.

Figure 4 shows the mean final energy as a function of α for the exponential schedule. Again the available annealing time $\tau = 1000$. Here the minimum is attained roughly for an α such that at the end of the schedule the temperature equals $T_0/20$.

In figure 5 we compare the mean final energy for the optimal, the linear, and the exponential schedules as a function of τ. The linear and the exponential schedules use the best values for ε and α as determined above.

One clearly sees that the optimal schedule out performs the two other schedules for all τ. For $\tau = 1000$ the mean final energy for the optimal schedule is smaller by a factor of 10^{-3} than that for the exponential schedule. Note how the decrease of the mean final energy slows down for larger annealing times τ.

Finally using (7.10) we are now able to calculate the thermodynamic speed along the optimal path. The thermodynamic speed v can be conveniently expressed as [10, 19]

$$v = \frac{E(t) - E(T(t))}{\sigma(T(t))} \tag{8.3}$$

where $E(t)$ is the mean energy at time t, $E(T(t))$ is the mean energy of the Boltzmann distribution at temperature $T(t)$ and $\sigma(t)$ is its variance. For our problem

$$v = \frac{p_2 - x/(1+x)}{\sqrt{x/(1+x)^2}}. \tag{8.4}$$

In the long time limit and for $x \ll 1$ this reduces to

$$v = \frac{p_2 - x}{\sqrt{x}} = \left(\frac{D}{D-1} - 1\right)\frac{x}{\sqrt{x}} = \frac{1}{D-1} \cdot \sqrt{x} \propto t^{-1/2(D-1)}. \tag{8.5}$$

Thus the optimal thermodynamic speed, far from being constant, decreases with time. Incidentally, the entropy production rate is also not constant and it also decreases with time.

9. Summary

We have shown how the optimal annealing schedule can be determined for a simple three-state system. Even though it was conceived as an introductory example to demonstrate methods, it already shows a very surprising and unexpected feature: the turnpike. Under what conditions and whether at all this feature can persist for larger and more complex systems remains an open question.

We compared the optimal schedule with the linear and the exponential schedule using the best values of the free parameters in these schedules. The mean final energy for the optimal schedule is smaller by factors up to 10^3 for the annealing times studied. This indicates the potential gains of using optimised schedules for simulated annealing and stresses the importance of further research.

A further result gained from this one example is that the constant thermodynamic speed schedule is not generally optimal. But as our model is very simple compared with the systems usually encountered in simulated annealing it is quite feasible that the constant speed schedule may be still a very good approximation to the optimal one.

We plan to take this approach further and study larger systems. This will lead to a better understanding of which quantities enter the optimal schedule in general. We note that our approach is the only route which can find the best schedule with certainty.

Appendix. Optimal control theory

The optimal control problem is to minimise the functional

$$\min_{u \in U} F[x, u] = \int_0^\tau \mathrm{d}t\, I(x(t), u(t)) \tag{A1}$$

subject to constraints

$$\dot{x} = f(x, u) \tag{A2}$$

and the initial conditions

$$x(0) = x_0. \tag{A3}$$

x are the state variables, u the controls and U the set of allowed controls. With the adjoint variables λ a Hamiltonian is defined by

$$H(x, u, \lambda) = I(x, u) + \lambda^{tr} \cdot f(x, u). \tag{A4}$$

The optimal path is now characterised by

$$\begin{aligned} \dot{x} &= \frac{\partial H}{\partial \lambda} = f(x, u^*) \\ \dot{\lambda} &= -\frac{\partial H}{\partial x}(x, u, \lambda)|_{u=u^*} \end{aligned} \tag{A5}$$

and u^* is obtained by minimising H

$$H(x, u^*, \lambda) \leqslant H(x, u, \lambda) \tag{A6}$$

over all $u \in U$.

Finally, if the endpoint conditions $x(\tau)$ are not given, we must require

$$\lambda(\tau) = 0. \tag{A7}$$

Thus the control problem has been reduced to solving a boundary value problem or an initial value problem for $2n$ ordinary coupled differential questions where n is the dimension of the state vector x.

Acknowledgments

We would like to acknowledge the stimulating environment of the Telluride Summer Research Center where part of this work was done as well as partial travel support by the Deutsche Forschungsgemeinschaft (KHH). PS would like to thank the Alexander von Humboldt Stiftung for support during most of this work.

References

[1] Kirkpatrick S, Gelatt Jr C D and Vecchi M P 1983 *Science* **220** 671
[2] Cerny V 1983 *JOTA* **45** 41
[3] Garey M R and Johnson D S 1979 *Computers and Intractability* (San Francisco: Freeman)
[4] Ettelaie R and Moore M A 1985 *J. Physique Lett.* **46** L893
[5] Slarry P and Dreyfus G 1983 *J. Physique Lett.* **45** L39

[6] White S 1984 *Proc. Int. Conf. on Computer Design* (Piscataway, NJ: IEEE) p 646
[7] Vecchi M P and Kirkpatrick S 1983 *IEEE Trans. Computer-Aided Design* **CAD-2** 215
[8] Bonomi E and Lutton J L 1984 *SIAM Rev.* **26** 551
[9] Durbin R and Willshaw D 1987 *Nature* **326** 689
[10] Salamon P, Nulton J, Robinson J, Pedersen J M, Ruppeiner G and Liao L 1988 Simulated annealing with constant thermodynamic speed *Comput. Phys. Commun.* **49** 423
[11] Geman S and Geman D 1984 *IEEE PAMI* **6** 721
[12] Jacobsen M O, Mosegaard K and Pedersen J M 1988 Global model optimisation in reflection seismology by simulated annealing *Model Optimisation in Exploration Geophysics 2* ed A Vogel (Braunschweig: Vieweg-Verlag)
[13] Huang M D, Romeo F and Sangiovanni-Vincentelli A 1986 *Proc. ICAAD* (Piscataway, NJ: IEEE) p 381
[14] Lundy M and Mees A 1984 *Convergence of the Annealing Algorithm* (*Simulated Annealing Workshop*) (New York: IBM Thomas J Watson Research Center)
[15] Aarts E H L and van Laarhoven P J M 1985 *Philips J. Res.* **40** 193
[16] Morgenstern I and Wuertz D 1987 *Z. Phys.* B **67** 397
[17] Greene J W and Supowit K J 1984 *Proc. ICCD* (Piscataway, NJ: IEEE) p 658
[18] Szu H and Hartley R 1987 *Phys. Lett.* **122A** 157
[19] Nulton J and Salamon P 1988 Statistical mechanics of combinatorial optimization *Phys. Rev.* A **37** 1351
[20] Andresen B, Hoffmann K H, Mosegaard K, Nulton J, Pedersen J M and Salamon P 1988 *J. Physique* **49** 1485
[21] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 *J. Chem. Phys.* **21** 1087
[22] Binder K 1979 *Monte Carlo Methods in Statistical Physics* (Berlin: Springer)
[23] Harary F 1969 *Graph Theory* (Reading, MA: Addison-Wesley)
[24] Hajek B 1985 *Int. Conf. on Decision and Control* (Piscataway, NJ: IEEE)
[25] Mitra D, Romeo F and Sangiovanni-Vincentelli A 1985 *Memorandum UCB/ERL M85/23* Electronics Research Laboratory, University of California, Berkeley
[26] Salamon P, Pedersen J M, Sibani P and Hoffmann K H 1990 Ensemble implementations in simulated annealing: an analytic approach, in preparation
[27] Boltyanskii V G 1971 *Mathematical Methods of Optimal Control* (New York: Holt, Reinhart and Winston)
[28] Tolle H 1975 *Optimization Methods* (Berlin: Springer)
[29] Hsu J C and Meyer A U 1968 *Modern Control Principles and Applications* (New York: McGraw-Hill)
[30] Bryson Jr A E and Ho Y-C 1975 *Applied Optimal Control* (New York: Wiley)
[31] Huse D A and Fisher D S 1986 *Phys. Rev. Lett.* **57** 2203