Finite time scaling of energy in simulated annealing

1991 J. Phys. A: Math. Gen. 24 L141

(http://iopscience.iop.org/0305-4470/24/3/008)

J. Phys. A: Math. Gen. 24 (1991) L141–L144. Printed in the UK

LETTER TO THE EDITOR

Finite time scaling of energy in simulated annealing

Shigeru Shinomoto and Yoshiyuki Kabashima

Department of Physics, Kyoto University, Kyoto 606, Japan

Received 6 November 1990

Abstract. The energy of a local minimum obtained by the simulated annealing generally depends on a time τ in which a complex system has been immersed in a heat bath. How the resultant energy $E(\tau)$ scales with a time τ is an interesting question. The diffusion process of a point in a wiggly parabola is analysed to discuss the scaling. The model is exactly solvable and the energy is found to scale as $E(\tau)=\varepsilon+c(\ln\tau)^{-1}$. This scaling is considered rather common to general complex systems. However, the limit ε obtained from practical data is not necessarily the ground state energy of a system.

Simulated annealing is a physical analogy for heuristics to obtain approximate solutions for combinatorial optimization problems (Kirkpatrick *et al* 1983, Černý 1985). The point is to find a fairly low energy state of the imaginary physical system whose energy is the cost function of an optimization problem. In order to obtain a sufficiently low energy state of a physical system, one generally has to anneal the system very slowly. The complex physical system has a tremendous number of local minima and there is no general scheme which enables us to find the ground state in a finite time. Furthermore, the level of the ground state is generally not known in advance, and one does not know the residual energy of the resultant state left for the ground state.

In order to get an estimate of the residual energy, Grest *et al* (1986) proposed an empirical law for how the resultant energy E depends on a time τ spent in annealing. Their proposal is twofold: $E(\tau)=E_0+c(\ln\tau)^{-1}$ for the problems *NP*, and $E(\tau)=E_0+c'\tau^{-\nu}$ for *P*. Their numerical data are not enough, however, to ensure the law. This gave rise to a theoretical issue by Huse and Fisher (1986). They first discussed the annealing process of a two-level system. The average energy of a two-level system scales as $\tau^{-\delta}$, whose exponent δ proportionally depends on the energy difference of the two minima. For a system composed of the independent two-level systems, they got the scaling $E(\tau)=c(\ln\tau)^{-2}$ by integrating the individual energy $\tau^{-\delta}$ over various energy differences δ. By taking into account the small scale low energy excitations in a general physical system, they concluded that the scaling should be of the form $E(\tau)=E_0+c(\ln\tau)^{-\zeta}$, with its exponent $\zeta\leqslant 2$. The exponents 2 and ζ above are obtained under the assumption that the distribution of excited state energies of individual two-level systems does not depend on the energy. The exponents are, however, sensitive to the distribution of excited state energies, and thus the upper bound of ζ is not so clear.

We wish to discuss the exponent ζ from another point of view. We found a solvable model whose energy turned out to scale $E(\tau)=c(\ln\tau)^{-1}$. The exponent 1 here does not depend on the dimensionality of the model system, nor on the details of the rough shape of the energy surface. Thus the exponent 1 appears rather common to various complex systems. However, in the case that there is a singular cusp in the global shape

0305-4470/91/030141+04$03.50

of the energy surface, we have not succeeded in specifying the exponent and the problem is thus still open in this respect.

The system corresponding to a complex optimization problem generally has an exponential number of local minima. In order to describe the circumstance in which a global minimum is hidden among a tremendous number of local minima, we first discuss the random walk of a point in the energy landscape depicted in figure 1. The system has two scales: each local minimum is surrounded by energy barriers of $O(B)$, and they are distributed over a larger basin $E = x^2/2$ for 'macroscopic' distance x. Let us assume that the local minima are arranged at regular intervals of distance a, and the macroscopic position of the ith local minimum is $x = ai$. The master equation describing the hopping between neighbouring local minima is given by

$$dP_i/dt = e^{-B/T}P_{i+1} + e^{-(B+\Delta_{i-1})/T}P_{i-1} - (e^{-B/T} + e^{-(B+\Delta_i)/T})P_i \tag{1}$$

where T is the temperature and Δ_i denotes the energy difference between adjacent local minima, $\Delta_i = a^2((i+1)^2 - i^2)/2$. We are able to derive from (1) the Fokker–Planck equation for the coarse-grained distribution function of the macroscopic distance x:

$$\partial P/\partial t = \gamma(T)\partial(xP)/\partial x + D(T)\partial^2 P/\partial x^2 \tag{2}$$

where both the coefficient γ and D should be functions of the temperature T, of the form $\gamma = a^2 T^{-1} e^{-B/T}$ and $D = T\gamma$. In order to obtain (2) from (1), we have assumed that $a^2/T \ll 1$. The approximation is not valid until zero temperature but it holds until $T \sim \sqrt{a}$, in which a can be taken sufficiently small.

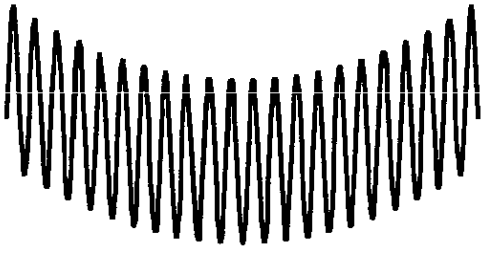

Figure 1. Characteristic shape of the energy landscape of current interest.

Let us consider the annealing process of an ensemble of the systems whose average energy is initially high. We wish to establish how the average energy, $y(t) = \langle E \rangle(t) = \int dx\, x^2/2P(x, t)$, relaxes with a time spent in annealing. The evolution equation for the average energy is

$$dy/dt = -2\gamma(T)y + D(T) \tag{3}$$

where the temperature T is then a function of time t. If the average residual energy y is known, we are able to obtain an optimal annealing schedule by choosing $T(t)$ so as to keep minimizing the RHS of (3). In this optimal schedule, the residual energy y obeys the equation,

$$dy/dt = -y\, e^{-1/y} \tag{4}$$

where we have rescaled the time $t \to 2t/Be$ and the energy $y \to 2y/B$, respectively. In the case where y is sufficiently small compared with unity, the RHS of (4) satisfies the inequalities,

$$-y^{(2+\epsilon)/(1+\epsilon)} \exp(-y^{-1/(1+\epsilon)}) < -y \exp(-y^{-1}) < -y^2 \exp(-y^{-1}) \tag{5}$$

where ϵ is of the order of y. By integrating these two bounds for the phase flow, the solution of (4) is found to satisfy the inequalities,

$$(\ln \tau)^{-1-\epsilon} < y(\tau) < (\ln \tau)^{-1} \tag{6}$$

where ϵ is of the order of y. Thus in the original units, the energy scales as

$$y(\tau) = \langle E \rangle(\tau) \sim B/\ln \tau \tag{7}$$

in the asymptotic limit. The optimal annealing schedule is then

$$T(t) \sim B/\ln t. \tag{8}$$

Geman and Geman (1984) obtained the general bounds of temperature schedules in which one can eventually get the ground state with probability unity in the limit, $\tau \to \infty$. Equation (8) is in accordance with the lower bound for this.

The asymptotic form of the residual energy, equation (7) is, however, not sensitive to the details of a temperature schedule, $T(t)$. For most temperature schedules in which the temperature is smoothly lowered from $T \gtrsim B$ to $T = 0$, one gets the same asymptotic form, equation (7).

In addition the qualitative aspect of the result is also not sensitive to the dimensionality of the system. The high-dimensional system can be decomposed into independent systems of the form of (2) and the total energy is the sum of the individual energies. Thus the change of dimensionality induces a simple modification of the factor of the result. On the other hand, the change of dimensionality induces drastic change in the 'density of states'. The invariance of the power, -1, for $\ln \tau$ against the change of the distribution is thus rather remarkable and should be contrasted with the result of Huse and Fisher, which can be sensitive to the distribution of the excited state energies of two-level systems.

The scaling behaviour of the residual energy, (7), appears rather general with respect to the dimensionality. However, we have restricted consideration to the quadratic global potential. In order to consider the dependence of the scaling behaviour on the shape of the potential surface, we discuss here non-quadratic systems. Consider a one-dimensional system $E = x^\alpha/\alpha$, $x \geq 0$. In general cases, $\alpha \neq 2$, we cannot obtain the closed form, such as (3), for the average energy, $y = \int_0^\infty \mathrm{d}x\, x^\alpha/\alpha P(x, t)$, but

$$\alpha^{-1}\, \mathrm{d}\langle x^\alpha \rangle/\mathrm{d}t = -2\gamma\langle x^{2\alpha-2}\rangle + (\alpha - 1)\, \mathrm{D}\langle x^{\alpha-2}\rangle. \tag{9}$$

This equation (9) is obtained via partial integration, and is valid for $\alpha > 1$. If the distribution function is sufficiently compact, equation (9) can be read as

$$\mathrm{d}y/\mathrm{d}t = -2\gamma(\alpha y)^{2-2/\alpha} + \mathrm{D}(\alpha - 1)(\alpha y)^{1-2/\alpha}. \tag{10}$$

At this stage, the optimal temperature schedule can be obtained, giving

$$\mathrm{d}y/\mathrm{d}t = -y^{2-2/\alpha}\, \mathrm{e}^{-1/y} \tag{11}$$

where we have rescaled the time and the energy. Solutions of this equation can also be bounded by the inequalities (6), which give the same asymptotic form, $y \sim 1/\ln \tau$. The scaling (7) thus holds even in the non-quadratic potential, $E = x^\alpha/\alpha$, as far as there is no cusp, or $\alpha > 1$. The energy scaling in the case where the global potential has a cusp is still open here, as equation (9) is not valid for $0 < \alpha \leq 1$.

We have analysed a few model systems, which turned out to exhibit the same scaling behaviour. However, we have assumed here some kind of continuity in the global structure of the wiggly surface, and have not succeeded in analysing the case

in which singular structures are present in the global shape. There is no general scheme to classify wiggly surfaces into 'continuous' and 'singular' ones. However, some kind of continuity would be expected if the cost function is expressed by a low-order polynomial of elements. Bernasconi (1987) showed a pathological case in which a system has 'golf-hole-like' structure. The system consists of four-body interaction terms. Even in such a case, the system exhibits the same scaling behaviour, $E(\tau) = \varepsilon + c/\ln \tau$, for an intermediate time scale τ which is already sufficiently long for practical numerical simulation. We cannot conclude that the limit ε is the ground state energy. The limit ε can be some 'global' minimum over the restricted region where the singular parts are excluded.

In addition to this, we must keep in mind that there should be model systems in which we cannot assume continuity in any sense. A white noise potential configuration proposed by Strenski and Kirkpatrick (1990) in another context exemplifies the circumstance, where the energy scaling is yet to be clarified.

We would like to thank Yoshiki Kuramoto and Tsuyoshi Chawanya for helpful discussion. The present study was extended further via discussion with Daniel Fisher and Joachim Buhmann, when one of us (SS) joined the Aspen workshop.

References

Bernasconi J 1987 Low auto correlation binary sequences: statistical mechanics and configuration space analysis *J. Physique* **48** 559–67

Černý V 1985 Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm *J. Opt. Theor. Appl.* **45** 41–52

Geman S and Geman D 1984 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images *IEEE Trans. Patt. Anal. Machine Intel.* **6** 721–41

Grest G S, Soukoulis C M and Levin K 1986 Cooling-rate dependence for spin-glass ground-state energy: implications for optimization by simulated annealing *Phys. Rev. Lett.* **56** 1148–51

Huse D A and Fisher D S 1986 Residual energies after slow cooling of disordered systems *Phys. Rev. Lett.* **57** 2203–6

Kirkpatrick S, Gelatt C D and Vecchi M P 1983 Optimization by simulated annealing *Science* **220** 671–80

Strenski P N and Kirkpatrick S 1990 *Preprint*