Cost distributions in large combinatorial optimisation problems

1989 J. Phys. A: Math. Gen. 22 4599

(http://iopscience.iop.org/0305-4470/22/21/022)

J. Phys. A: Math. Gen. **22** (1989) 4599-4609. Printed in the UK

Cost distributions in large combinatorial optimisation problems

Neil Burgess and M A Moore

Department of Theoretical Physics, The University, Manchester M13 9PL, UK

Received 1 March 1989

Abstract. We have examined the cost distribution of locally optimal solutions in certain combinatorial optimisation problems. This distribution is found to be peaked about a value characteristic of the algorithm involved, with a width that decreases with the system size N. It is shown that the distribution of costs ε is of the form $\exp(Ng(\varepsilon))$ and that $g(\varepsilon)$ is self-averaging. Consequently, estimation of the optimal cost is best achieved by curve fitting to $g(\varepsilon)$. Possible forms for $g(\varepsilon)$ are proposed.

1. Introduction

Combinatorial optimisation problems have a long history of mathematical study. Recently, techniques borrowed from statistical physics have provided useful insights, e.g. [1-3]. The difficulty in solving these problems is due to the size of the configuration space (which usually grows very rapidly with the number N of degrees of freedom of the system) and to the conflicting nature of the constraints involved. These constraints tend to give rise to a large number of local minima in the cost functions for the problem. Thus the solution of minimum cost lies in a vast configuration space containing many local minima whose numbers often increase exponentially with N.

In large systems, locating the globally optimal solution is impractical and only locally optimal solutions can be generated in a reasonable amount of computer time. However, in many practical applications a solution only has to be 'good enough'; it does not have to be the best possible. Locally optimal solutions to a problem with respect to a given algorithm are attractors in the configuration space of the problem, under the dynamics of the algorithm. Where the cost function is actually the energy of the system we will refer to metastable states rather than locally optimal solutions. Associated with each attractor is a basin of attraction or gathering bin of configurations for which the attractor is a fixed point under the dynamics of the algorithm. When an algorithm is run from an initial configuration the particular solution in which it terminates depends only on which gathering bin the initial configuration was in.

We argue below that in large systems the distribution of locally optimal costs generated by an algorithm will be very sharply peaked about a cost characteristic of the algorithm; running the algorithm from many different initial configurations will not usually give a solution significantly better than the result of one run. The running time of an algorithm increases rapidly with the quality of the solutions it gives (many of these problems are 'NP complete' i.e. finding the globally optimal solution in general requires a number of computational steps which is an exponential of N). Efficient estimation of the optimal cost is, however, possible by extrapolation from the distribution of easier-to-generate locally optimal costs [4, 5]. Here we look at the likely cost

0305-4470/89/214599+11$02.50

distribution of locally optimal solutions, and how best to use it to extrapolate to the globally optimal cost.

2. The shape of the cost distribution

Let us consider the cost distributions in two typical examples of these problems: (i) finding the ground-state energy of a short-range Ising spin glass, and (ii) the travelling salesman problem (TSP).

Problem (i) involves minimising the Edwards-Anderson Hamiltonian [6]

$$H = -\sum_{\langle ij \rangle} J_{ij} S_i S_j \tag{1}$$

where the spins S_i can be ± 1, the exchange energies J_{ij} are typically taken from a Gaussian distribution with zero mean and the sum is over nearest neighbours. A general set of moves for the dynamics of the system is to reverse the sign of up to m neighbouring spins. A state energetically stable with respect to these dynamics is termed 'm-spin flip' metastable. We simulated a short-range Ising spin glass on a two-dimensional square lattice of side L for $L = 10, 20, \ldots, 50$ with periodic boundary conditions and bonds J_{ij} chosen randomly from a Gaussian distribution with mean 0 and variance 1. A single-spin flip metastable state was generated from a random initial state by sequentially reversing spins so as to lower the total energy until no further reduction was possible in this way. This was done from one hundred different initial spin configurations. We found that in these states the mean energy per spin $\varepsilon = -1.07 \pm 0.04$ (compared to the ground-state energy for which it is ~ -1.3 [7]) and is virtually independent of N. The standard deviation of the metastable state energy per spin $\delta\varepsilon = \sqrt{\langle \varepsilon^2 \rangle - \langle \varepsilon \rangle^2} = (0.59 \pm 0.02) N^{-1/2}$.

Simulation of an SK spin glass (which has infinite-range interactions, so that the sum in (1) is over all pairs and the standard deviation of J_{ij} is $N^{-1/2}$) by Henkel and Kinzel [8] shows the metastable states with respect to single-spin flips (at zero temperature) to have similar behaviour with $\langle \varepsilon \rangle \sim 0.92 \varepsilon_0$ and $\delta\varepsilon \sim N^{-1/2}$ where ε_0 is the ground-state energy.

Problem (ii) involves minimising the length of a closed path through N points (or 'cities') i.e. minimising

$$\sum_{i=1}^{N} l_{P(i),P(i+1)}$$

with respect to $P(i)$, which is a cyclic permutation of $1, 2, \ldots, N$ giving the order in which cities are visited. l_{ij} are the distances between cities i and j. A general set of moves for the dynamics of the system is to swap up to m intercity bonds in the tour for bonds not in the tour such that the tour is still closed. A tour that cannot be shortened by such a move is said to be m-optimal or 'm-opt' [9].

For the TSP with N cities scattered with uniform distribution on $[0, 1]^d$ the optimal path length $l \simeq C_d N^{(d-1)/d}$ with standard deviation $\delta l \sim N^{(d-2)/2d}$ [10] so that $\delta l / l \sim N^{-1/2}$ where the averaging is over city positions. C_d is a constant depending on the dimensionality of the problem. In two dimensions $0.765 \leqslant C_2 \leqslant 0.765 + 4/N$ [11].

We simulated a TSP on a square with N cities for $N = 100, 120, \ldots, 400$. A 2-opt tour was generated from a random initial tour by sequentially swapping pairs of intercity bonds so as to reduce its length until no further reduction was possible. This

was done from one hundred different initial tours. We found that $l = (0.82 \pm 0.02)\sqrt{N}$ and $\delta l = (0.25 \pm 0.04)$.

Thus both in the spin glass and TSP problems we have $\delta C / C \sim 1/\sqrt{N}$ where C is the locally optimal cost. That is, the locally optimal costs become increasingly closely gathered about a cost typical of the algorithm as the size of the system increases.

We can understand this behaviour if we suppose that the number of local minima N_s grows exponentially in N, say $N_s = \exp \alpha N$ (this is true for single-spin flip metastable states in spin glasses [12-14] and for 'reasonable' paths in the TSP city where each city is connected to cities at most k minimal spacings away [1]). Then we expect the energy distribution of metastable states to be given by $N_s(\varepsilon) = \exp(Nh(\varepsilon))$. Associated with each local minimum in the cost function is a 'gathering bin' of states attracted to the minimum under the dynamics of the particular algorithm. If the gathering bins were all the same size then the probability of an algorithm (started from a random configuration) terminating in a local minimum of cost ε would be given by $P(\varepsilon) = N_s(\varepsilon)/N_s = \exp[N(h(\varepsilon) - \alpha)]$. Since the gathering bins are not generally the same size then

$$P(\varepsilon) = \exp[N(h(\varepsilon) - f(\varepsilon) - \alpha)]$$

where the average fraction of configuration space in the gathering bin associated with a metastable state of energy ε is $\exp(-Nf(\varepsilon))$.

If $g(\varepsilon) = h(\varepsilon) - f(\varepsilon) - \alpha$ has a maximum at ε_1 then to first approximation the probability distribution of the locally optimal costs will be

$$P(\varepsilon) = \exp[N(g(\varepsilon_1) + \tfrac{1}{2}g''(\varepsilon_1)(\varepsilon - \varepsilon_1)^2)] \tag{2}$$

near to the peak at ε_1 with $g''(\varepsilon_1) < 0$. Hence the spread of locally optimal costs obtained by running an algorithm from random initial configurations is given by $\sqrt{\langle(\varepsilon - \varepsilon_1)^2\rangle} \sim N^{-1/2}$.

Thus in these types of problems the distribution of costs which are local minima under given algorithm dynamics will have a Gaussian-like peak of width $N^{-1/2}$ near to the characteristic cost ε_1 of the algorithm. Solutions that are locally optimal under certain algorithm dynamics tend to become fewer in number and lower in cost as a more powerful algorithm is used (e.g. using 2-spin flip dynamics in spin glasses or 3-opt dynamics for the TSP) [15].

Notice that the underlying shape of the distribution of metastable state energies is given by $g(\varepsilon) = \ln(P(\varepsilon))/N$. The fluctuations in the curve $g(\varepsilon)$ between different realisations of a problem will be much less than the fluctuations in the distribution of metastable states $N_s(\varepsilon)$ (or $P(\varepsilon)$ itself) since $N_s(\varepsilon)$ grows like $\exp(N)$ rather than N. We can see why this is with a simple example. Suppose the value of an extensive variable f (i.e. one which grows like N) as measured in samples of size N has Gaussian fluctuations of size $\sqrt{N}$ from sample to sample, i.e. $f = \alpha N + \Delta$, with

$$P(\Delta) = \frac{\exp(-\Delta^2/N)}{\sqrt{\pi N}}\,\mathrm{d}\Delta.$$

Then the fluctuations $\delta g = (\langle g^2\rangle_\Delta - \langle g\rangle_\Delta^2)^{1/2}$ in a variable $g = \exp(f)$ are given by

$$\left[\int \exp[2(\alpha N + \Delta) - \Delta^2/N]\frac{\mathrm{d}\Delta}{\sqrt{\pi N}} - \left(\int \exp(\alpha N + \Delta - \Delta^2/N)\frac{\mathrm{d}\Delta}{\sqrt{\pi N}}\right)^2\right]^{1/2}$$

$$\sim \exp[N(\alpha + \tfrac{1}{2})].$$

Although $\delta f \sim \sqrt{N}$; $\ln(\delta g) \sim N$. Notice that $\langle g\rangle_\Delta = \exp(\langle f\rangle_\Delta + N/2)$.

It has been observed in statistical physics that the densities of extensive quantities (e.g. the ground-state energy per spin or the susceptibility per spin) tend to be 'self-averaging' in large systems (i.e. their sample-to-sample fluctuation tends to zero as the sample size tends to infinity). A simple plausibility argument for this was given by Brout [15]. A similar argument in the context of the energy and number of metastable states in combinatorial optimisation problems is as follows.

If a large system can be divided into k independent subsystems such that metastable states of the system corresponds to each subsystem being in a metastable state, then (a) the metastable state energy E_s is given by

$$E_s = \sum_{i=1}^{k} E_i$$

where E_i is the energy of the ith subsystem metastable state, and (b) the number of metastable states N_s is given by

$$N_s = \prod_{i=1}^{k} N_i \qquad \text{i.e.} \quad \ln(N_s) = \sum_{i=1}^{k} \ln(N_i)$$

where N_i is the number of metastable states in the ith subsystem. Thus, if E_i and N_i are independent of E_j and N_j, respectively ($j \neq i$) the central limit theorem implies that $(1/N)(\langle E_s^2\rangle - \langle E_s\rangle^2)^{1/2} \sim N^{-1/2}$ and $(1/N)[(\langle(\ln(N_s))^2\rangle - \langle\ln(N_s)\rangle^2]^{1/2} \sim N^{-1/2}$ as $k \to \infty$ (if the distributions of E_i and N_i have finite moments). Generally such subsystems are not truly independent of their neighbours because of boundary effects.

We thus expect that the distribution $g(\varepsilon) = \ln(P(\varepsilon))/N$ will have the typical shape $\langle\ln(P(\varepsilon))\rangle/N$ for different examples of the same problem (e.g. for a spin glass with different sets of exchange energies $\{J_{ij}\}$ drawn from the same distribution), up to fluctuations which tend to zero as $N \to \infty$.

3. Estimating the optimal cost by curve fitting

We see from the above that the most efficient way of producing a good solution to a large combinatorial optimisation problem is with a few runs of the best algorithm possible in the available computer time; see also [17]. Judging the quality of a solution requires an estimation of the optimal cost. This can be done by running a relatively fast algorithm from many different initial configurations to generate locally minimum costs. If one assumes a form for the distribution $P(\varepsilon)$ of these costs then this theoretical form can be fitted to the data on the frequency of metastable state energies so as to extrapolate to the globally minimum cost ε_0.

This has been done for the TSP by Golden [18], see also [4, 5, 19], on the assumption that locally optimal costs ε will have a 'Weibull' distribution i.e.

$$P(\varepsilon) = \frac{c}{b}\left(\frac{\varepsilon - 1}{b}\right)^{c-1} \exp\left[-\left(\frac{\varepsilon - a}{b}\right)^{c}\right]$$

or

$$P(\varepsilon \leqslant \varepsilon') = F(\varepsilon') = 1 - \exp\left[-\left(\frac{\varepsilon' - a}{b}\right)^{c}\right] \tag{3}$$

for $\varepsilon \geqslant a \geqslant 0$, $b > 0$, $c \geqslant 0$. The three parameters a, b and c are estimated so as to best fit the $\{\varepsilon\}$ generated by some quick algorithm run from random initial paths. Finding the best fit can be done several ways [4, 5, 18], the simplest being to determine b and c for a range of values of a by a least-squares fit on $\ln\{-\ln[1-F(\varepsilon')]\} = c\ln(\varepsilon'-a) - c\ln(b)$ where $F(\varepsilon')$ is the experimentally generated cumulative distribution of energies. Then the value of a is an estimate of the optimal solution, and the interval $[\nu - b, b]$ contains the optimal cost with an estimated probability of $1-\exp(-s)$ [4], where s locally optimal path lengths were generated and ν was the shortest one.

The Weibull distribution is intuitively plausible since if independent samples of size m are taken from a continuous parent population bounded below by a, then the distribution of the smallest member of a sample approaches that of equation (3) (for some b and c) as $m \to \infty$ [18]. This is the third of the three limiting forms of the distribution of extreme values derived by Fisher and Tippett [20]. Because each locally optimal cost x_i is implicitly the smallest of the m costs in the ith basin of attraction we might expect the distribution of x_i to be Weibull.

However, the actual distributions of metastable state energies have been calculated for spin glasses and they are *not* Weibull distributions.

The curve given by $\langle \ln(N_s(\varepsilon))\rangle / N$ over the full range of ε has been found for single-spin flip metastable states in a spin glass in one dimension [21] (see § 4, and also [15] for multi-spin flip metastable states), two dimensions [22] (numerically) and in the SK model [12]. In the region of interest (i.e. for ε near to ε_0) $N_s(\varepsilon)$ for one-, two- and three-dimensional spin glasses has the form (Bray and Moore [13])

$$N_s(\varepsilon) \sim \exp[N\beta(\varepsilon - \varepsilon_0)^\alpha] \tag{4}$$

where $\alpha = \frac{1}{2}$ for $d = 1$. $P(\varepsilon)$ varies from $N_s(\varepsilon)/N_s$ due to the energy dependence of the average size of a gathering bin associated with a local minimum of a particular energy. In the one-dimensional case there is no such energy dependence [21] so that $N_s(\varepsilon)/N_s = P(\varepsilon)$. Also, for the TSP in d dimensions, an argument in [17] implies that $N_s(\varepsilon)$ might also have the form of expression (4) with $\alpha = d/(d-1)$ for ε near ε_0.

The weak point in the intuitive argument for the Weibull distribution is the assumption that each metastable state energy is implicitly the lowest of m independent samples from the population of possible energies (meaning the energies of the m configurations in that gathering bin). The energies of the configurations in a particular gathering bin will not be independent, and in general m varies from bin to bin.

Golden and co-authors [4, 5, 18, 19] did obtain encouraging results for the TSP. It may be that doing a good job with a three-parameter fit using any curve of roughly the right shape will give reasonable results, or that in the TSP the path lengths in a gathering bin happen to be more like a random sample of possible lengths than in the spin glass.

This leaves us with the problem of what curve to fit to the locally minimum costs generated by a quick algorithm. Obviously in problems where the underlying probability distribution of metastable states ($g(\varepsilon)$ in equation (2)) has been calculated then this distribution (rather than, e.g., a Weibull distribution) is the one to try to fit to the experimental data (for any reasonable number of data points). Where there has been no such calculation a curve such as $\beta(\varepsilon - \varepsilon_0)^\alpha$ for ε near ε_0 (from equation (4)) should be fitted to the data on $\ln(P(\varepsilon))/N$ (or their cumulative equivalents can be used if more convenient). Then, if this distribution is 'self-averaging', a good fit for one realisation of the problem (e.g. one set $\{J_{ij}\}$ for the spin glass) will also fit for other realisations of the problem up to an error that tends to zero as the size of the system

N tends to infinity. Consequently, estimation of the globally optimal cost ε_0 in large systems by this method in many realisations of a problem would require only small corrections to the parameters (e.g. α and β) once the curve $\ln(P(\varepsilon))/N$ had been fitted for one realisation.

It is also possible to use the knowledge of the distribution taken by the smallest element of a sample as follows: if the metastable state energies generated by some algorithm are put into batches of size m then the probability distribution $P_B(\varepsilon)$ of the lowest energy ε from each batch will be [13]

$$P_B(\varepsilon) = mP(\varepsilon)\left(\int_\varepsilon^\infty P(x)\,dx\right)^{m-1} = mP(\varepsilon)\left(1-\int_0^\varepsilon P(x)\,dx\right)^{m-1}$$

where $P(\varepsilon)$ is the distribution of metastable-state energies.

If the lowest energy ε per batch is near enough to ε_0 and $P(\varepsilon)\sim\beta(\varepsilon-\varepsilon_0)^{c-1}$, say, then we recover the Weibull distribution

$$P_B(\varepsilon) = m\beta(\varepsilon-\varepsilon_0)^{c-1}\left(1-\frac{\beta(\varepsilon-\varepsilon_0)^c}{c-1}\right)^{m-1}$$

$$\to m\beta(\varepsilon-\varepsilon_0)^{c-1}\exp\left(\frac{-m\beta(\varepsilon-\varepsilon_0)^c}{c-1}\right)$$

as $m\to\infty$. So that in the Weibull distributon (3) we know that the parameter $b\sim(m/(c-1))^{1/c}$.

More accurately for the spin glass (i.e. for a larger range of ε), we know that $P(\varepsilon)\sim\exp[\beta N(\varepsilon-\varepsilon_0)^\alpha]$ for some α, β so that

$$P_B(\varepsilon)\to m\exp[\beta N(\varepsilon-\varepsilon_0)^\alpha - mI(\varepsilon)]$$

where

$$I(\varepsilon) = \int_0^\varepsilon \exp[\beta N(x-\varepsilon_0)^\alpha]\,dx \sim \exp(\beta N(\varepsilon-\varepsilon_0)^\alpha)$$

for large n.

From our earlier argument we see that a more powerful algorithm must be used as the size of the problem increases to achieve low enough metastable-state energies, whatever fitting method is used.

4. Example problem: the one-dimensional Ising spin glass

The one-dimensional short-range Ising spin glass, i.e.

$$H = -\sum_{i=1}^{N} J_i S_i S_{i+1} \qquad S_i = \pm 1 \qquad P(J_i) = \frac{1}{\sqrt{2\pi}}\exp(-\tfrac{1}{2}J_i^2)$$

has the characteristic large configuration space and exponential number of degrees of freedom of many harder combinatorial optimisation problems, without being too hard to solve analytically. Also, since the size of the basins of attraction under single-spin flips do not vary with the energy of the attractor [21], the probability of locating a metastable state of energy ε from a random initial spin configuration is $P(\varepsilon) = N_s(\varepsilon)/N_s$. We can demonstrate the difference in the distribution of $N_s(\varepsilon)$ and $g(\varepsilon) = \ln(N_s(\varepsilon))/N$ from equation (2). The typical distribution ($\langle \ln N_s(\varepsilon)\rangle_J/N$) has been

calculated by Ettalaie and Moore [21] (see figure 1). We will now calculate $\langle N_s(\varepsilon)\rangle_J$ for comparison.

In a single-spin flip metastable state each spin is aligned with the local field $h_i = J_{i-1}S_{i-1} + J_iS_{i+1}$ so that $\lambda_i = h_iS_i \geq 0$ for all i. The metastable-state energy per spin $\varepsilon = -(1/2N)\,\Sigma_i\lambda_i$. Hence

$$N_s(\varepsilon) = \operatorname*{Tr}_{\{S_i\}} \int_0^\infty \prod_{i=1}^N d\lambda_i \delta\left(2N\varepsilon + \sum_i \lambda_i\right) \prod_{i=1}^N \delta\left[\lambda_i - S_i(J_{i-1}S_{i-1} + J_iS_{i+1})\right]. \tag{5}$$

Using the integral representation of the delta function and averaging over the J_i (which leaves no S_i dependence so that we can take the trace) gives

$$\langle N_s(\varepsilon)\rangle_J = 2^N \int_{-i\infty}^{i\infty} \frac{du}{2\pi i} \exp(-2Nu\varepsilon) \int_0^\infty \prod_{i=1}^N d\lambda_i \times \int_{-i\infty}^{i\infty} \prod_{i=1}^N \frac{dx_i}{2\pi i} \prod_{i=1}^N \exp[-\lambda_i u - x_i\lambda_i + \tfrac{1}{2}(x_i + x_{i+1})^2]. \tag{6}$$

Putting

$$\exp[\tfrac{1}{2}(x_i + x_{i+1})^2] = \int_{-\infty}^{\infty} \frac{dt_i}{\sqrt{2\pi}} \exp[-\tfrac{1}{2}t_i^2 + (x_i + x_{i+1})t_i]$$

and recognising the x_i integrals as delta functions, gives for the right-hand side of (6)

$$2^N \int_{-i\infty}^{i\infty} \frac{du}{2\pi i} \exp(-2Nu\varepsilon) \int_0^\infty \prod_{i=1}^N d\lambda_i \exp(-\lambda_i u) \times \int_{-\infty}^{\infty} \prod_{i=1}^N \frac{dt_i}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}t_i^2)\delta[\lambda_i - (t_i + t_{i-1})]$$

$$= 2^n \int_{-i\infty}^{i\infty} \frac{du}{2\pi i} \exp(-2Nu\varepsilon) \int_0^\infty \prod_{i=1}^N d\lambda_i \exp\left[-\lambda_i u - \frac{1}{2}\left(\sum_{j=1}^{i} (-1)^{j+1}\lambda_j\right)^2\right].$$

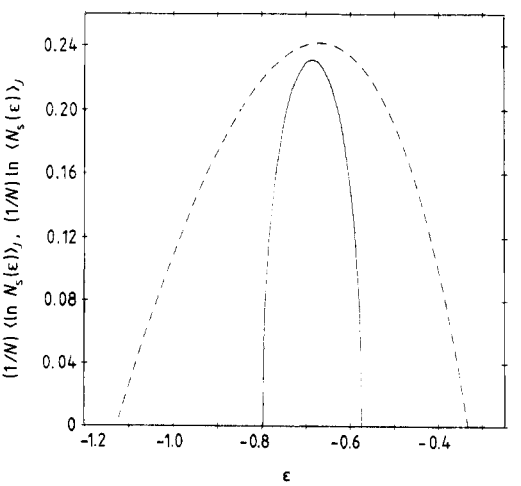

Figure 1. Plots of $(1/N)\ln\langle N_s(\varepsilon)\rangle_J$ (broken curve) and $(1/N)\langle\ln N_s(\varepsilon)\rangle_J$ (full curve) against ε for the one-dimensional Ising spin glass.

Making the transformation $y_i = \Sigma_{j=1}^{i} (-1)^{j+1}\lambda_j$, i.e. $\lambda_i = y_i + y_{i+1}$, for which the Jacobian is unity, gives

$$\langle N_s(\varepsilon)\rangle_J = 2^N \int_{-i\infty}^{i\infty} \frac{du}{2\pi i} \exp(-2Nu\varepsilon) \int_0^\infty dy_1 \times \exp(-2y_1u - \tfrac{1}{2}y_1^2) \int_{-y_1}^\infty dy_2 \ldots \int_{-y_{N-1}}^\infty dy_N \exp(-2y_Nu - \tfrac{1}{2}y_N^2). \tag{7}$$

To solve equation (7) for $\langle N_s(\varepsilon)\rangle_J$ we let

$$g_1(y) = 1$$

$$g_j(y) = \int_{-y}^\infty \frac{dx}{\sqrt{2\pi}} \exp(-2xu - \tfrac{1}{2}x^2) g_{j-1}(x) \qquad j = 2, 3, \ldots, N$$

then as $j \to \infty$, $g_j(y) \to g(y)$, where

$$\lambda(u) g(y) = \int_{-y}^\infty \frac{dx}{\sqrt{2\pi}} \exp(-2ux - \tfrac{1}{2}x^2) g(x) \tag{9}$$

which is equivalent to the differential equation

$$g''(y) + (y - 2u)g'(y) + \frac{\exp(-y^2)}{2\pi\lambda^2} g(y) = 0 \tag{10}$$

with boundary conditions

$$\frac{g(0)}{g'(0)} = \sqrt{2\pi\lambda} \qquad g(-\infty) = 0.$$

Hence, for large N, equation (7) becomes

$$\langle N_s(\varepsilon)\rangle_J = 2^N \int_{-i\infty}^{i\infty} \frac{du}{2\pi i} \exp(-2Nu\varepsilon)(\lambda_{max}(u))^N g_{max}(0) \tag{11}$$

where $g_{max}(y)$ is the solution of equation (10) with the largest eigenvalue $\lambda_{max}(u)$.

We can solve equation (10) exactly if we put $u = 0$. This corresponds to removing the constraint on the energy of a metastable state in equation (5) so that equation (11) with $u = 0$ gives the average number of metastable states

$$\langle N_s\rangle_J = 2^N(\lambda_{max}(0))^N g_{max}(0). \tag{12}$$

If we make the change of variables

$$z(x) = \int_{-x}^\infty \frac{dv}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}v^2)$$

in equation (10) with $u = 0$, then we get

$$g''(z) + \frac{1}{\lambda^2} g(z) = 0 \qquad \frac{g(z = \frac{1}{2})}{g'(z = \frac{1}{2})} = \lambda \qquad g(z = 0) = 0$$

which has solution $g_n(y) = A \sin(z(y)/\lambda_n)$, $\lambda_n = (-1)^n/[(n + \frac{1}{2})\pi]$, for $n = 0, 1, \ldots$. Thus equation (12) gives the average number of metastable states as

$$\langle N_s\rangle_J = \frac{1}{\sqrt{2}} \left(\frac{4}{\pi}\right)^N$$

for large N, in agreement with [23].

To calculate $\langle N_s(\varepsilon)\rangle_J$ we numerically calculate $\lambda_{max}(u)$ from equation (10) with $u \neq 0$ by an iterative procedure which started from the $u = 0$ solution. Finally, equation (11) tells us that for large N

$$\langle N_s(\varepsilon)\rangle_J = \int_{-i\infty}^{i\infty} \frac{du}{2\pi i} \exp\{-N[2u\varepsilon - \ln(2\lambda_{max}(u))]\}. \tag{13}$$

The integral (13) is evaluated by steepest descent.

The results are displayed in figure (1). The 'average' $\ln\langle N_s(\varepsilon)\rangle_J/N$ curve is wider than the 'typical' curve $\langle \ln N_s(\varepsilon)\rangle_J/N$, and predicts the wrong ground-state energy per spin. In the ground state each bond contributes $-|J_i|$ to the energy so that the average ground-state energy per spin is $\langle \varepsilon_0\rangle_J = \langle -\Sigma_i |J_i|\rangle_J/N = -\sqrt{2/\pi} = -0.798$, in agreement with the typical curve $\exp(\langle \ln N_s(\varepsilon)\rangle_J)$, whereas $\langle N_s(\varepsilon)\rangle_J = 1$ for $\varepsilon = -0.113$. That $\ln\langle N_s(\varepsilon)\rangle_J \geqslant \langle \ln(N_s(\varepsilon))\rangle_J$ follows trivially from the inequality $\exp(\langle x\rangle) \leqslant \langle \exp(x)\rangle$. Unusual metastable state energies from rare J_i congfigurations are responsible for the atypical behaviour of $\langle N_s(\varepsilon)\rangle_J$.

The peaks in the distributions also occur at slightly different energies; -0.686 for the typical curve and -0.671 for the average curve. Notice that a graph of $\langle N_s(\varepsilon)\rangle_J$ rather than the $(1/N)\ln\langle N_s(\varepsilon)\rangle_J$ shown in figure 1 would be *very* sharply peaked about ε_1, with $N_s(\varepsilon) \sim \exp(N(\varepsilon - \varepsilon_1)^2)$ near to ε_1. The typical curve also has this behaviour, in agreement with our earlier prediction for $P(\varepsilon)$ in equation (2). The behaviour of the typical curve near to ε_0 is given by $\langle \ln(N_s(\varepsilon))\rangle_J/N \simeq 1.1456(\varepsilon - \varepsilon_0)^{1/2}$ [15], in agreement with equation (4). The peak in the average curve corresponds to $u = 0$ in the steepest-descent calculation of (13), giving the average (total) number of metastable states to leading order. The height of the average curve at the peak is $\ln(4/\pi)$, i.e. $\langle N_s\rangle \sim \exp(0.241N)$, in agreement with [23], compared to the typical number of metastable states $\exp(\langle \ln(N_s)\rangle) \sim \exp(0.231N)$ [21, 23, 24].

The fluctuation with $\{J_i\}$ of both the typical and the average number of metastable states can be calculated using results from Derrida and Gardner [23]. They showed that

$$\lim_{N\to\infty} \frac{1}{N} \ln\langle N_s^p\rangle_J = \ln(\mu(p \ln 2))$$

$$\mu(x) = \frac{\sqrt{\exp(x)-1}}{\tan^{-1}(\sqrt{\exp(x)-1})}$$

so that the fluctuation δN_s of N_s (in the large-N limit) is given by

$$\frac{1}{N} \ln\sqrt{\langle N_s^2\rangle_J - \langle N_s\rangle_J^2} = \tfrac{1}{2} \ln\left(\frac{3\sqrt{3}}{\pi}\right).$$

This can be viewed as an 'error bar' at the peak in the average distribution.

By comparison, the fluctuation in $\ln(N_s)$ (in the large-N limit) is given by

$$\frac{1}{N}\sqrt{\langle(\ln(N_s))^2\rangle_J - \langle \ln(N_s)\rangle_J^2}$$

$$= \frac{1}{N} \lim_{p\to 0} \left(\left\langle \left(\frac{N_s^p - 1}{p} \right)^2 \right\rangle_J - \left(\left\langle \frac{N_s^p - 1}{p} \right\rangle_J \right)^2 \right)^{1/2}$$

$$= \frac{1}{N} \lim_{p\to 0} \frac{1}{p} \{(\mu(2p \ln 2))^N - 2(\mu(p \ln 2))^N - 1 - [(\mu(p \ln 2))^N - 1]^2\}^{1/2} = 0$$

since $\lim_{p\to 0} \ln(\mu(p \ln 2))/p = (\ln 2)/3$. That is, the typical number of metastable states is self-averaging. The same behaviour for the complete distribution $\ln(N_s(\varepsilon))/N$ is expected.

Consequently the curve $(1/N)(\ln(N_s(\varepsilon)))_J$ will fit experimentally generated data on $\ln(N_s(\varepsilon))/N$ in systems specified by different sets $\{J_i\}$, up to fluctuations which tend to zero as $N \to \infty$. Thus, for large systems, a curve (e.g. $\beta(\varepsilon - \varepsilon_0)^{1/2}$ for ε near ε_0) which gives a good fit to the data on $\ln(N_s(\varepsilon))/N$ for one realisation of the problem (i.e. one $\{J_i\}$) will also give a good fit for other realisations of the problem with only small corrections to the parameters (β and ε_0). Hence estimation of the ground-state energy by this method in many realisations of the problem would require very little effort once the curve $\ln(N_s(\varepsilon))/N$ had been found for one realisation.

5. Conclusions

We have studied the cost distribution of locally optimal solutions in the two-dimensional spin glass and the TSP. We found that this distribution has a Gaussian-like peak about a value characteristic of the algorithm involved, with a width that decreases with $N^{-1/2}$. We expect this form of distribution to be general for combinatorial optimisation problems of this type. Thus, in the large-system limit an algorithm run from different initial states will only sample states of a similar cost, the best of which will probably be worse than the result of the run of a more powerful algorithm.

The form of the cost distribution $P(\varepsilon)$ indicates that estimation of the optimal cost from many locally optimal costs should be done by extrapolation of $g(\varepsilon) = \ln(P(\varepsilon))/N$ (or the equivalent for the cumulative distribution). This curve contains the typical behaviour of a problem and may be sample indendent in the large-system limit (thus saving a lot of effort where many realisations of a problem need solving). We have indicated likely forms for $g(\varepsilon)$ for spin glasses and the TSP for ε close to the globally optimal cost.

Acknowledgments

We should like to thank A J Bray, J L Shapiro, M Birse and A J McKane for many helpful discussions.

References

[1] Mezard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)
[2] Fu Y T and Anderson P W 1986 *J. Phys. A: Math. Gen.* **19** 1605
[3] Kirkpatrick S, Gelatt C D and Vecchi M P 1983 *Science* **220** 671
[4] Golden B L and Alt F B 1979 *Naval Res. Logist. Quart.* **26** 67
[5] Golden B L 1978 *Comm. Statist.* B **7** 361
[6] Edwards S F and Anderson P W 1975 *J. Phys. F: Met. Phys.* **5** 965
[7] Morgenstern I and Binder K 1980 *Phys. Rev.* B **22** 288
[8] Henkel R D and Kinzel W 1987 *J. Phys. A: Math. Gen.* **20** L727
[9] Lin S 1965 *Bell. Syst. Tech. J.* **44** 2245
[10] Beardwood J, Halton J H and Hammersley J M 1959 *Proc. Camb. Phil. Soc.* **55** 299
[11] Stein D 1977 *Doctoral dissertation* Harvard University
[12] Bray A J and Moore M A 1980 *J. Phys. C: Solid State Phys.* **13** L469

[13] Bray A J and Moore M A 1987 *Proc. 1986 Heidelberg Coll. on Glassy Dynamics and Optimisation* (*Lecture Notes in Physics* **275**) (Berlin: Springer)
[14] Tanaka F and Edwards S F 1980 *J. Phys. F: Met. Phys.* **10** 2769
De Dominicis C, Gabay M, Garel T and Orland H 1980 *J. Physique* **41** 933
[15] Ettelaie R and Moore M A 1987 *J. Physique* **48** 1255
[16] Brout R 1959 *Phys. Rev.* **115** 824
[17] Moore M A 1987 *Phys. Rev. Lett.* **58** 1703
[18] Golden B L 1977 *Networks* **7** 209
[19] Golden B L and Stewart W R 1985 *The Travelling Salesman Problem* ed E L Lawler, J K Lenstra, A H G Rinnooy Kan and D B Shmoys (New York: Wiley) p 244
[20] Fisher R A and Tippett L H C 1928 *Proc. Camb. Phil. Soc.* **24** 180
[21] Ettelaie R and Moore M A 1985 *J. Physique Lett.* **46** L893
[22] Masui S, Southern B W and Jacobs A E 1989 *Phys. Rev.* B to be published
[23] Derrida B and Gardner E 1986 *J. Physique* **47** 959
[24] Li T 1981 *Phys. Rev.* B **24** 6579