Correlation structure of landscapes of NP-complete optimization problems at finite temperatures

Wolfgang Gehring, Peter Salamon and Roger E. Whitney
Dept. of Mathematical Sciences
San Diego State University
San Diego, California 92182

Paolo Sibani
Fysisk Institut
Odense Universitet
Campusvej 55, DK-5230 Odense M
Denmark

November 4, 2003

Abstract

We analyze the autocorrelation function of a time series of energy values sampled in the energy landscapes of four different combinatorial optimization problems. The sampling is performed by Metropolis random walks. The temperature of the walk and the size of the investigated problems are systematically varied.

We find that, in a suitably defined high temperature region, i.e., above the freezing transition, the autocorrelation decays in an exponential fashion. We extract the temperature and system size dependence of the corresponding correlation time, which turns out to be of the Arrhenius form. Energetic and entropic contributions to the correlation time (barriers) are identified and shown to be asymptotically independent of system size.
pacs: 05.40.+j ; 02.50.Ga ; 02.70.Lq

1 Introduction

A number of important biological, physical and combinatorial optimization applications e.g. evolution, thermalization of glassy systems and simulated

annealing are described in terms of energy landscapes[1]. Technically, we define a landscape to be a (large but finite) graph, where each node x represents a configuration of the system, and where edges connect nearest neighbors. The distance d between two nodes is the number of edges in the shortest path from one node to the other. In addition, a real valued function E, the energy, is defined on the nodes; the values of E can be visualized as giving the 'height' of the landscape.

Whether one's interest in landscapes is rooted in physical or biological applications or in optimization *per se*, a good characterization of the landscape geometry is of paramount importance for understanding the dynamics. One successful effort in this direction is a series of studies by Stadler and coworkers[2, 3, 4] which classifies landscapes based on the properties of energy-energy correlation functions[5]. One of their tools is to sample energy values by an unbiased random walk. In the framework of a Metropolis scheme this corresponds to walking at infinite temperature. These authors analyzed several examples and showed that, in each case, the energy-energy correlation function sampled by a random walk is exponential. The dependence of the corresponding correlation times on the system size was also studied and related to the correlation length of the landscape, a quantity which gauges the statistical similarity of energies of configurations a distance d apart.

One motivation for such studies comes from the expectation that classifying systems using their correlation times can select between alternative implementations of optimization algorithms. In particular, Stadler et al. suggested using the correlation time to choose between alternative move classes. The preferred move class is the one with the largest correlation length. The argument is based on Schuster's conjecture [**?**] that each correlated volume of the landscape contains one 'good' minimum. It then follows that larger correlation times will lead to fewer local minima.

It is well known that relaxation properties of hard combinatorial problems, as well as other complex systems, are simply described at high temperatures, but very intricate below the freezing temperature, where the system becomes non-ergodic and multiple relaxational time scales become important. To investigate these relaxation properties, we extend the approach of Stadler and coworkers by sampling the energy landscape with a Metropolis algorithm at large but finite T and by studying the correlation function of the ensuing time series of energy values. We consider four different systems, and find in each case that the correlation function is well described by a single exponential and the temperature dependence of the correlation time

has an Arrhenius form, with energetic and entropic barriers asymptotically independent of system size.

2 Background

We recall that the Metropolis rule accepts a randomly chosen neighbor x' of the current state x with probability $P_{x' \leftarrow x} = \min(1, \exp((E(x) - E(x'))/T)$. The correlation function of the walk of length M is

$$R(t) = \frac{< E(x_i)E(x_{i+t}) > - < E(x_i) >< E(x_{i+t}) >}{\sqrt{(< E(x_i)^2 > - < E(x_i) >^2)(< E(x_{i+t})^2 > - < E(x_{i+t} >^2)}}, \quad (1)$$

where $E(x_i)$ is the energy value sampled at the i-th step of the walk and $\langle\rangle = \frac{1}{M-t}\sum_{i=1}^{M-t}()$. The quantity $R(t)$ has considerable physical interest as, for example, it determines the (linear) response of the system to an arbitrary thermal disturbance [**?**].

The master equation[7] describing the dynamics for the Metropolis process [1] has the form

$$\frac{dP}{dt} = WP \quad (2)$$

Here, P is a vector of probabilities, where P_x is the probability of the system being in state x. The stochastic W matrix is obtained from the connectivity matrix of the graph C by $W_{x,x'} = C_{x,x'} \min(1, \exp(-(E(x) - E(x'))/T)$, for $x \neq x'$. The diagonal elements are found by imposing the stochasticity condition that all column sums be zero. By construction W is negative semidefinite. It has a zero eigenvalue, corresponding to the equilibrium solution. The relaxation properties at sufficiently large times are given by the next largest eigenvalue, λ_1. The reciprocal of its absolute value is often denoted by the term 'relaxation time', $\tau = \mid \lambda_1 \mid^{-1}$. Generically (i.e. unless special symmetries of the problem make themselves felt or if the ergodicity of the system is broken) one would expect the relaxation eigenvalue to describe the decay of *every* function on the set of states to their equilibrium value. In particular, the decay of the energy-energy autocorrelation function at sufficiently long times should occur with τ. Since $R(0) = 1$ and $R(t) \approx 0$

[1] While the algorithm itself actually takes place in discrete time, in this paragraph we refer to the description in continuous time since this description is more commmon in the physics literature. The two descriptions correspond closely, with the discrete time transition matrix being given by $exp(W)$ [7].

for t large, it follows that when the autocorrelation time is of the simple exponential form, it must be given by

$$R(t) = \exp(-t/\tau). \tag{3}$$

Thus, the correlation time and the relaxation time are likely to be one and the same quantity.

Little is known about the behavior of τ even at infinite temperature where the W matrix is doubly stochastic and equal to minus the laplacian of the graph. In general, it can be shown that λ_1 must vanish as the system size grows to infinity. There are also some known bounds for λ_1 and related quantities for walks on arbitrary graphs in the limit of large system size.[8] The time dependence of the relaxation time for random walks on a line (which makes W tridiagonal) have been considered by Larsen[9]. No general results for the relaxation time at finite temperatures and for arbitrary graphs appear to be available.

3 Numerical experiments

3.1 Systems considered

We have considered instances of the symmetric and euclidean traveling salesman problem, of the graph bipartitioning problem and of the spin glass problem with long range interactions. We find that the correlation function at high temperatures has in all cases the exponential form in Eq. 3. The single exponential decay is observed through several decades, i.e. until the value of the correlation function R has practically decayed to zero.

While the numerical value of τ is unaffected by the choice of units for the energy, a meaningful comparison of the temperature and size dependence of $\tau(T, N)$ for systems of different sizes requires the use of uniform energy and temperature scales. This has been achieved by ensuring that in each case the energy is an extensive quantity (i.e. proportional to N).

The symmetric traveling salesman problem, henceforth STSP, is defined by a set of N cities and by an $N \times N$ 'distance' matrix with positive elements d_{ij}. The d_{ij}'s are drawn from a uniform distribution on the interval $(0, 1]$. A configuration or tour x is a permutation $\mathcal{P}_x$ of the N cities. The energy of the tour is

$$E(x) = \sum_{i=1}^{N} d_{\mathcal{P}_x(i),\mathcal{P}_x(i+1)}, \tag{4}$$

which, being a sum of N positive terms, scales linearly with N. In other words, the energy is an extensive property, as required. The move class, which defines the neighborhood of each configuration is the so called 'two-bond' move [10]: two positions i and j are picked uniformly at random along the current tour. The corresponding cities are exchanged and the direction of the tour between them is reversed.

In the euclidean traveling salesman problem, henceforth denoted ETSP, both the energy function and the move class are the same as in the STSP problem. The distances between pairs of cities are however 'true' distances, i.e. they are calculated from the positions of the cities according to the usual euclidean metric. The positions are in turn drawn from a uniform distribution in the unit square. In this system most distances are of order one, and most configurations have therefore energies of order N, leading again to an extensive energy function. Note however that the lowest energy of the ETSP scales as $N^{1/2}$[11]. This is due to the fact that low-energy configurations mainly involve neighbor cities. As the density of cities increases linearly with N, the distance between close neighbors must scale as $N^{-1/2}$, whence the scaling follows. The lack of extensivity of the ground state does not concern us here, as we do not probe very low energy configurations during our sampling.

Thirdly, we consider a mean-field spin glass problem, henceforth called SPG: a set of N Boolean variables $\sigma_i = \pm 1$ defines the configuration space. A set of coupling constants J_{ij}, with $J_{ij} = J_{ji}$, is created by independently drawing each J_{ij}, $i < j$, from a symmetric distribution, which in our case is the two-valued distribution $J_{ij} = \pm 1$. For any configuration x, the energy is given by[12]

$$E(x) = \sum_{i<j} J_{ij} \sigma_i^x \sigma_j^x. \tag{5}$$

The move class takes one spin at random and multiplies it by -1. Note that the energy is a sum of N terms, where each term is again a sum of N stochastic variables with zero mean. It follows that, for large N, the distribution of E becomes normal with $\sigma \propto N^{3/2}$. To restore extensivity, we must divide the J_{ij} by $N^{1/2}$.

Finally, we look at an instance of the graph bipartitioning problem, henceforth called GBP. One considers a set of $2N$ nodes and creates, with probability p, an edge $J_{ij} = 1$ between nodes i and j. The case of no edge corresponds to $J_{ij} = 0$. The probability of a non zero connection was in our case $p = 0.1$. We again assign to each node i a spin variable $\sigma_i = \pm 1$. If $\sigma_i = 1$ the i'th node is declared to be 'white' otherwise 'black'.

A configuration of the problem is then a binary string of length $2N$ where half of the nodes are white and half black. Its energy is the number of edges joining nodes of different types. The energy of a configuration x can compactly be written as

$$E(x) = \sum_{i<j,j} J_{ij} \frac{(1 - \sigma_i^x \sigma_j^x)}{2} \tag{6}$$

which is quite similar to Eq. 5. Note however that the average graph has pN^2 edges, a finite fraction of which contributes to the energy. Hence, the energy scales with N^2, and all the J_{ij} (or alternatively, the temperature) must be divided by N to restore extensivity. The move class is also different from the spin-glass case. Here a move is performed by taking a pair of nodes and switching their colors. In spin language this corresponds to a dynamics which conserves the total magnetization.

3.2 Results

All simulations were performed on a SUN Sparc 1000E workstation at San Diego State University. For each problem we considered the system sizes $N = 20, 40, 60, 80$ and 100. The Metropolis algorithm was used to generate time series of energy values, each comprising one million updates. Different sets of data were produced, corresponding to different temperatures. Besides the case $T = \infty$, which served as a check, we used a set of temperatures equally spaced on a logarithmic scale, starting at $T = 128$, and ending at a temperature where the correlation function clearly deviated from the exponential behavior. The correlation time was calculated as the reciprocal of the slope of the logarithm of the empirical correlation, plotted as a function of time. Only data with $R > 0.1$ were used in the fits and the goodness of fit measure[2] r^2 well above 0.99. Each temperature and system size were run 100 times and the *tau* values used were an average of the 100 runs.

In all four problems, the dependence of the correlation length on the system size and the temperature is well described by an Arrhenius type formula:

$$\tau_{fit} = N(c_1 + c_2/N) \exp(\frac{c_3 + c_4/N}{T}) \tag{7}$$

The values of the coefficients are summarized in the Table 1.

[2]coefficient of determination

-	c_1	c_2	c_3	c_4
STSP	0.52 ± 0.02 [1/2]	-1.0 ± 0.2	0.39 ± 0.03	(1.1 ± 0.5)
ETSP	0.50 ± 0.01 [1/2]	-1.23 ± 0.04	0.46 ± 0.02	(0.5 ± 0.3)
SPG	0.24 ± 0.01 [1/4]	(-0.33 ± 0.62)	2.0 ± 0.1	(-2 ± 5)
GBP	0.125 ± 0.001 [1/8]	-0.44 ± 0.06 [−3/8]	0.027 ± 0.004	1.45 ± 0.15

Table 1: Parameters for the fitted dependency of the correlation time on temperature and system size, according to Eq. 7. Available theoretical values describing the $T = \infty$ behavior are dispayed in brackets. They agree very well with the values obtained by our empirical fit. Values not significantly different from zero are indicated in parentheses. Omitting them in the fits changes the parameters only slightly, still leaving all values within the reported error bars. All fits used had r^2 values well in excess of 0.99.

The coefficients, c_i, were determined in a two pass operation. During the first pass, least squares lines were fit to $1/T$ versus $\ln(\tau)$ to find

$$\ln(\tau) = \alpha \frac{1}{T} + \beta. \tag{8}$$

These coefficients α and $\exp(\beta)$ were then fit in the second pass to linear functions of N and $1/N$ respectively. Again, as for the extraction of τ values from the time series, all fits had r^2 values well in excess of 0.99.

During the first pass, the line of $1/T$ versus $\ln(\tau)$, we used only high temperature points. The exact choice of what constituted low temperature had a significant effect on the resulting regression coefficients. Since we had 100 τ values for each T and N, we had available the standard deviation in the observed τ values. This standard deviation divided by τ was constant down to a certain temperature below which its value rose dramatically. This temperature was used to define the divide between high T and low T data and small variations around this temperature were explored. The reported error in the values of the c_i was determined as the maximum of the standard estimate of error in the final regression coefficients and the observed standard deviation obtained by varying the exact cutoff for what constituted low temperature points.

As discussed further below, the lack of a fit of our functional form in Eq. (7) for low temperature data is expected, and can be physically explained as freezing. At low temperatures more than one relaxational time scales become

important, whereby the correlation function no longer can be described by a single exponential. Concurrently, the apparent correlation length no longer describes the decay of the slowest relaxation mode. Its value stays well below what formula (7) would predict at the relevant temperature, and, in addition, it becomes almost insensitive to the system size. This is consistent with a broken ergodicity picture, where the system only relaxes in subsets of its state-space during the available observation time.

Usually, broken ergodicity is described in terms of energetic or entropic barriers which diverge as the size of the system grows large. In our present context, due to the Arrhenius form of the correlation time, the quantity $c_3 + c_4/N$ plays the role an energy barrier separating regions of configuration spaces which must be sampled before the energy can decay to its average value. We find this energy barrier to be asymptotically independent of system size. An entropic barrier can be defined as the logarithm of the relaxation time scale at $T = \infty$. This quantity is found to have a linear dependence on N. This linear dependence is usually trivially removed by using N Metropolis updates (a Monte Carlo step) as the unit of time. Our data lead to the conclusion that provided one properly scales the energy (so as to be extensive) and the time (measured in N Metropolis steps), the energetic and entropic barriers probed by high temperature relaxation processes are basically independent of system size.

4 Summary and conclusions

The energy-energy correlation function for four different combinatorial problems was studied by Monte Carlo sampling at different temperatures. For sufficiently high temperatures, the correlation function decays exponentially and correlation times depend on the temperature in a simple Arrhenius fashion. The entropic and energetic barriers are found to be independent of the system size.

We conclude with some speculations regarding the reasons for our findings. One possibility is that there exist barriers which diverge with system size, but separate regions of the landscape with identical statistical properties. A second possibility — and one which we believe to be the case — is that the landscape is fraught with basins of local minima which get deeper as the system grows larger. In this case, the basins themselves must be invisible to Monte Carlo sampling at high T. The reason for this invisibility [13] is due to the fact that the local density of states within each basin

grow exponentially as $\exp(E/T_c)$. In this case no local Boltzmann equilibration can exist within the basins for $T > T_c$ and the basins cannot be detected at high temperature. Basically this amounts to an entropic effect: the overwhelming weight of configurations located at the rim of the basins prevents the Metropolis algorithm from visiting any low-energy states deep inside. An approximately exponential behavior of the local density of states has been found by exhaustive enumeration of states within single basins for a number of different complex landscapes[13, 14, 15, 16]. In systems with quenched randomness, such as those considered here, it seems unlikely that the landscapes should possess the degree of symmetry required for the first possibility to apply. Thus, we see our results as indications that the landscape is rather 'flat' at high temperatures, with the complex behavior appearing in a sudden fashion when exponential basins open up at low temperatures.

Aknowledgments Paolo Sibani is grateful to Ken Dykema for useful conversations at an early stage of this work. This project was partly supported by a block grant from Statens Naturvidenskabelige Forskningsråd.

References

[1] Landscape paradigms in Physics and Biology, Hans Frauenfelder, Alan R. Bishop, Angel Garcia, Alan Perelson, Peter Schuster, David Sherrington and Peter J. Swart Eds. Physica D 107, Nos. 2-4, 1996

[2] Peter F. Stadler. *Europhysics Letters*, 20:479–482, 1992.

[3] Peter F. Stadler and Robert Happel. *J. Phys. A: Math. Gen*, 25:3103–3110, 1992.

[4] Peter F. Stadler and Wolfgang Schnabl. The landscape of the traveling salesman problem. *Physics Letters A*, 161:337–344, 1992.

[5] E. D. Weinberger. *Biol. Cybern.*, 63:325, 1990.

[6] Bärbel Krakhofer and Peter F. Stadler. *Europhysics Letters*,34:85, 1996.

[7] N. G. Van Kampen. Stochastic processes in physics and chemistry. North-Holland, Amsterdam , 1992.

[8] Lawrence E. Thomas and Zhong Yin. *J. Math. Phys.* 27:2475–2477, 1986.

[9] Ulf Larsen. *J. Phys. A: Math. Gen*, 17:267–276, 1984.

[10] S. Lin and B.W. Kernighan. *Operations Research*, 21:498–516, 1973.

[11] Allon G. Percus and Olivier C. Martin *Phys. Rev. Lett.*, 76:1188–1191, 1996.

[12] S. F. Edwards and P. W. Anderson. *J. Phys. F* 5, (1975), 89

[13] P. Sibani, J. C. Schön, P. Salamon and J.-O. Andersson. *Europhys. Lett.* 22, (1993), 479.

[14] P. Sibani and P. Schriver. *Phys. Rev B* 49:6667, 1994.

[15] P. Sibani. Physica A 258: 249, 1998.

[16] J. C. Schön and P. Sibani. J. Phys. A: Math. Gen31: 8165, 1998.

[17] P. Sibani and K. H. Hoffmann. *Physica* A 234, (1997), 751