Local Minima in the Graph Bipartitioning Problem

By

Bärbel Krakhofer[a] and Peter F. Stadler[a,b,*]

[a]Institut für Theoretische Chemie, Universität Wien
Währingerstraße 17, A-1090 Wien, Austria
*Phone: ++43 1 40480 665   Fax: ++43 1 40480 660
Email: studla@tbi.univie.ac.at or stadler@santafe.edu

[b]Santa Fe Institute
1399 Hyde Park Rd., Santa Fe, NM 87501, USA

Abstract

We report numerical simulations on the number of local minima in the landscape of the Graph Bipartitioning Problem and provide an explanation in terms of the correlation length of its landscape.

PACS Classification

02.70.Lq, 75.50.Lk

Introduction

The cost function f of a combinatorial optimization problem can be regarded as a mapping of the vertex set V of a (usually huge but finite) graph Γ into the real numbers. The vertex set V corresponds to the set of all possible configurations. The edges of the graph Γ are introduced by defining a "move set" that allows to inter-convert "neighboring" configurations, see e.g. [5, 23, 16]. A mapping $f : V \to \mathbb{R}$ has been termed *landscape* following a picture of evolutionary optimization originally proposed by Sewall Wright [25].

The most important characteristic of a landscape is its *ruggedness* which is either quantified by means of a correlation measure [23, 10] or in terms of the number of local optima or peaks in the landscape [14]. Ruggedness is intimately related to the hardness of an optimization problem for heuristic algorithms [13]. We expect a close relationship between both characterizations of ruggedness, although there cannot be a simple functional relationship: different Ising models with Hamiltonians of the form $\mathcal{H}(\sigma) = \sum_{i<j} J_{ij}\sigma_i\sigma_j$ with $\sigma_k = \pm 1$ have the same correlation measure but somewhat different numbers of local optima depending on the coupling constants J_{ij}, see [2, 21].

It is conjectured in ref. [22] that there is about one local optimum in each ball with a radius that is determined by the correlation length of the landscape. Numerical studies on Derrida's p-spin models [4] support this "correlation length conjecture" [21]. In this letter we report that the same estimate is also in excellent agreement with numerical simulations for the graph bipartitioning problem.

Correlation Length and Elementary Landscapes

Weinberger [23] suggested to characterize a landscape $f : V \to \mathbb{R}$ by means of the autocorrelation function $r(s)$ of the time-series $\{f(x_0), f(x_1), \ldots\}$, which is obtained by sampling the cost function f along a simple random walk[1] $\{x_0, x_1, \ldots\}$ in configuration space Γ. We shall be concerned only with regular graphs Γ, i.e., all vertices of Γ have the same number D of neighbors. In this case it can be shown

[1] A random walk of a graph is simple if the probability for moving to any one of the vertices adjacent to the current position of the random walk equals one over the number of adjacent vertices.

that $r(s)$ is exponential if and only if f is an *elementary landscape*, that is, if f is, up to an additive constant, an eigenvector of the graph Laplacian $-\Delta \stackrel{\text{def}}{=} \mathbf{I} - D\mathbf{A}$, where $\mathbf{I}$ denotes the identity matrix and $\mathbf{A}$ is the adjacency matrix of the graph Γ. If $(\Delta f)(x) = \Lambda[f(x) - \bar{f}]$ for all $x \in V$, then $r(s) = (1 - \Lambda/D)^s$. The *correlation length* $\ell := \sum_{s=0}^{\infty} r(s)$ is given by $\ell = D/\Lambda$ in the case of an elementary landscape (for a proof and more details see [18, 17]).

A random walk of s steps reaches on average a distance $R(s) \leq s$ from its origin. The conjecture in [22] suggests that there should be about one local optimum in a ball of radius[2] $R(\ell)$ around any vertex of the graph Γ. The argument is the following: The size of the big mountains will be determined by the pair correlation in "typical" landscapes. More precisely, the expected radius of a mountain massif should be comparable to the average distance $R(\ell)$ because ℓ determines the size of the large scale structures as seen along the random walk. Thus we expect at least one local optimum in each ball in Γ with radius $R(\ell)$. On the other hand, the high dimensionality of the configuration space Γ makes local optima in the "foothills" an unlikely phenomenon — there will (almost) always be at least a few directions to walk uphill [10]. Since the overwhelming majority of configurations in a mountain is located in the foothills (again due to the high dimensionality of Γ) the number of local optima within a single mountain must be tiny as compared to its size. Hence assuming only a single one (the mountain's summit) is probably not a bad approximation. Denoting the number of vertices in a ball of radius R by $B(R)$ we expect therefore

$$\text{Prob}\{\text{loc.opt.}\} \approx 1/B(R(\ell)) \tag{1}$$

for a "typical" landscape with correlation length ℓ. The meaning of "typical" will be considered in the discussion section.

[2]In its original application to a traveling salesman problem this conjecture was stated using ℓ instead of $R(\ell)$. It is not hard to see, however, that $R(s) \approx s$ for a Cayley graph of the symmetric group with neighborhood defined by transpositions, as long as s is smaller than the number of cities.

The Graph Bipartitioning Problem

One of the combinatorial optimization problems which have been studied in great detail is the Graph Bipartitioning Problem (GBP) [1, 6, 7, 12, 19, 24]. Given a graph with an even number n of vertices and an associated matrix H of edge weights, the task is to find a partition $\mathfrak{X}$ of the vertex set W of this graph into two equal-sized subsets X and X^* such that the total edge weight

$$f_{GBP}(\mathfrak{X}) \stackrel{\text{def}}{=} \sum_{i\in X}\sum_{j\in X^*} h_{ij} \tag{2}$$

connecting the two subsets is minimized. As usual, we shall assume that the edge weights h_{ij} are i.i.d. random variables.

The cost function of the GBP may be viewed as a Sherrington-Kirkpatrick Hamiltonian [15] with the additional constraint of vanishing total spin [6]. The main difference between the two models, as far as optimization heuristics are concerned, is the different topology of the configuration spaces. The SK Hamiltonian is defined on the hypercube graph $\mathcal{Q}_2^n$ of dimension n. Two equi-partitions are neighbors in a GBP iff they differ by a single pair of vertices; the resulting graph is the Johnson graph $\mathcal{J}(n, n/2)$, see, e.g., [3]. Its most important parameters are the following: There are a total of $N = \binom{n}{n/2}$ configurations; the diameter (maximum distance) is $n/2$, the vertex degree is $D = n^2/4$, and the number of vertices in distance d from an arbitrary reference point is $\partial(d) = \binom{n/2}{d}^2$. The relation between the distance within $\mathcal{J}(n, n/2)$ and the Hamming distance [9] for a string representation of the partitions is discussed in [19].

Random walks on $\mathcal{J}(n, n/2)$ were considered in [19]. Here we only need the expected distance $R(s)$ that is reached after s steps of a simple random walk:

$$R(s) = \frac{n}{4}\left[1 - \left(1 - \frac{4}{n}\right)^s\right] . \tag{3}$$

The landscape of a GBP with arbitrary choice of the edge weights h_{ij} is elementary [8], fulfilling:

$$\Delta f_{GBP}(\mathfrak{X}) + 2(n-1)\left[f_{GBP}(\mathfrak{X}) - \frac{n^2}{4}\overline{h}\right] = 0 \tag{4}$$

for all equi-partitions $\mathfrak{X} \in \mathcal{J}(n, n/2)$. The constant $\overline{h} \stackrel{\text{def}}{=} \dfrac{2}{n(n-1)} \sum_{i<j} h_{ij}$ is the average edge weight. The constant $\Lambda = 2(n-1)$ is the third eigenvalue (second excited state) of the Laplacian of the Johnson graph $\mathcal{J}(n, n/2)$, see e.g. [3].

Theorem 1 in [17] implies that the random walk correlation function of the GBP, see [19] or [17, Thm.1], is given by

$$r(s) = \left(1 - \frac{\Lambda}{D}\right)^s = \left(1 - \frac{8}{n} + \frac{8}{n^2}\right)^s . \tag{5}$$

The correlation length[3] is therefore $\ell = D/\Lambda = (n-1)/8 + \mathcal{O}(1/n)$ irrespective of the choice of the edge weights h_{ij}.

Local Optima

With $\mu := 1/\lim_{n\to\infty} \sqrt[n]{B(R(\ell))}$ we conjecture Prob{loc.opt.} $\sim \mu^n$. In order to evaluate μ it is useful to introduce the scaled correlation length $\xi := \ell/n = 1/8 + \mathcal{O}(1/n)$, and the scaled radius

$$\hat{\xi} := \lim_{n\to\infty} \frac{1}{n} R(\ell) = \frac{1}{4}\left[1 - \frac{1}{\sqrt{\mathrm{e}}}\right] = 0.0983673\ldots . \tag{6}$$

It remains to compute the number of vertices contained in a ball of radius $R = n\hat{\xi}$. For the Johnson graph $\mathcal{J}(n, n/2)$ we have

$$B(R) = \sum_{q=0}^{R} \binom{n/2}{q}^2 \approx \binom{n/2}{R}^2 . \tag{7}$$

The approximation holds as long as $R \le n/4$, the error being at most a factor $\mathcal{O}(n)$. Writing $R = n\zeta$ and replacing all factorials in the binomial coefficient with Stirling's formula yields

$$B(n\zeta) \approx \left[\left(\frac{1-2\zeta}{2\zeta}\right)^{2\zeta} \frac{1}{1-2\zeta}\right]^n =: \mu(\zeta)^{-n} , \tag{8}$$

where we have again neglected all non-exponential contributions. The "correlation length conjecture" thus becomes $\mu = \mu(\hat{\xi}) = 0.609058\ldots$

We have performed numerical simulations based on graphs with random edge weights $h_{ij} = h_{ji} \in [0, 1)$ in order to check the prediction of the conjecture (Figure 1). Linear regression analysis yields $\mu_{\text{num.}} = 0.594 \pm 0.007$ with a correlation coefficient of $\varrho \approx 0.9998$.

[3]The definition of the correlation length in ref. [19] amounts to $-1/\ln(1 - \Lambda/D) = D/\Lambda + \mathcal{O}(1)$.

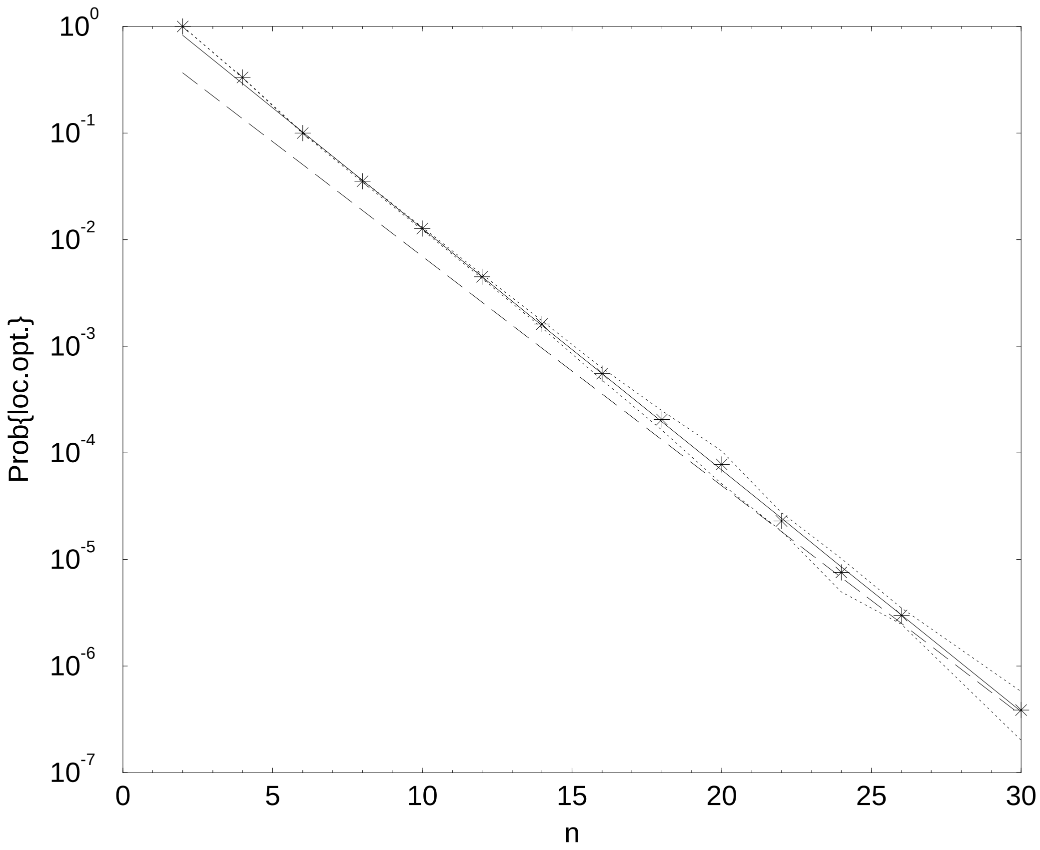

Figure 1: Numerical estimates for the probability of finding a local optimum in a random graph bipartitioning problem (stars) are obtained by determining the number of local optima among 1000 random configurations in at least 1000 independently generated random graphs. The dotted lines mark a ±3sdv interval. A least square fit (full line) to the data yields $\mu \approx 0.594 \pm 007$, compared to the estimate from the "correlation length conjecture" $\mu \approx 0.609$ (dashed line).

Discussion

The comparison of the "correlation length conjecture" for the number of local optima with numerical simulations yields an excellent agreement for the graph bipartitioning problem with random edge weights. This lends further credibility that analogous results for Derrida's p-spin models [21] and for symmetric traveling salesman problems [22] are not accidental.

All elementary landscapes for which the "correlation length conjecture" yields a good prediction are typical in the sense that they fulfill a kind of "maximum entropy" condition: they have the most general form of a cost function that is consistent with a given correlation length: Given a basis $\{\psi_k^\Lambda\}$ of the eigenspace belonging to Λ one can represent the cost function in the form $f = \bar{f} + \sum_k a_k \psi_k^\Lambda$ with i.i.d. coefficients a_k. The edge weights h_{ij} are by construction i.i.d. coefficients in our case. The corresponding eigenbasis is discussed in more detail in the appendix; it is not quite orthogonal, but the deviations approach 0 for large n.

The notion of *(statistical) isotropy* was introduced from a purely geometric point of view [23, 18]. It is interesting to note that our notion of a "typical" landscape coincides with the definition of *-isotropy if one requires the basis $\{\psi_k\}$ to be orthonormal and the distribution of the coefficients to be Gaussian [20, Thm.3]. Using the methods in this paper it can be shown that the GBP is in fact *-isotropic, i.e., that one can construct an ONB of the Λ-eigenspace such that the coefficients are uncorrelated with common variance and mean zero.

Landscapes that are known to deviate from the conjecture, on the other hand, have strongly constrained coefficients. In short-range Ising spin glasses most of the coupling coefficients are zero [2], and the graph matching problem can be treated as a TSP with a severely constrained distance matrix [11].

We have as yet no good explanation why the "correlation length conjecture" works so well. The increasing number of examples for its validity on graphs with quite different topologies and the fact that deviations seem to be related to violating a maximum-entropy (isotropy) condition render it a non-trivial relation between correlation-based and purely geometrical measures of ruggedness.

Acknowledgments

This work was partially supported by the Austrian *Fonds zur Förderung der Wissenschaftlichen Forschung*, Proj. No. 10578-MAT. Stimulating discussions with Ricardo García-Pelayo, Wim Hordijk, and Josef Leydold are gratefully acknowledged.

Appendix

In this appendix we list a few algebraic properties of the cost function of the GBP. Let $\eta^{pq}(\mathfrak{X}) = 1$ if the vertices p and q are in different sets of the equipartition $\mathfrak{X}$, and $\eta^{pq}(\mathfrak{X}) = 0$ otherwise. Clearly we have $f_{GBP} = \sum_{p<q} h_{pq}\eta^{pq}$. It is straightforward to check that $\sum_{\mathfrak{X}} \eta^{pq}(\mathfrak{X}) = \frac{1}{2}\frac{n}{n-1}\binom{n}{n/2}$ for all $p \neq q$, and thus the vectors η^{pq} are not eigenvectors of the graph Laplacian of the Johnson graph because they are not orthogonal to the trivial solution $(1, 1, \ldots, 1)$ belonging to $\Lambda_0 = 0$. With the definition

$$\varphi^{pq}(\mathfrak{X}) \stackrel{\text{def}}{=} \eta^{pq}(\mathfrak{X}) - \frac{1}{2}\frac{n}{n-1} = \begin{cases} \frac{1}{2}\frac{n-2}{n-1} & \text{if } p \text{ and } q \text{ are in different sets,} \\ -\frac{1}{2}\frac{n}{n-1} & \text{if } p \text{ and } q \text{ are in the same set,} \end{cases}$$

however, we obtain the representation

$$f_{GBP}(\mathfrak{X}) = \frac{n^2}{4}\overline{h} + \sum_{p<q} h_{pq}\varphi^{pq}(\mathfrak{X})\,. \tag{A1}$$

The vectors φ^{pq} form a basis of the eigenspace of the graph Laplacian of the Johnson graph $\mathcal{J}(n, n/2)$ belonging to the eigenvalue $\Lambda = 2(n-1)$. They are not orthogonal, however. It is not hard to check that $\sum_{q\neq p}\varphi^{pq} = 0$ which accounts for the fact that the $\binom{n}{2}$ vectors φ^{pq}, $p < q$, span an eigenspace of $\mathcal{J}(n, n/2)$ which is known to be only $\binom{n}{2} - n$ dimensional [3]. For completeness we note the explicit expressions for the scalar products

$$\|\varphi^{pq}\|^2 = \frac{1}{4}\left(1 - \frac{1}{(n-1)^2}\right), \qquad \langle\varphi^{pq}, \varphi^{pr}\rangle = -\frac{n}{4(n-1)^2}, \qquad \langle\varphi^{pq}, \varphi^{rs}\rangle = \frac{n}{2(n-3)(n-1)^2} \tag{A3}$$

where p, q, r, s are pairwise distinct integers. The basis is almost orthogonal for large n and can be easily normalized.

References

[1] J. R. Banavar, D. Sherrington, and N. Sourlas. Graph bipartioning and statistical mechanics. *J. Phys. A: Math. Gen.*, 20:L1–L8, 1987.

[2] A. J. Bray and M. A. Moore. Metastable states in spin glasses with short-ranged interactions. *J.Phys.C*, 14:1313–1327, 1981.

[3] A. Brouwer, A. Cohen, and A. Neumaier. *Distance-regular Graphs.* Springer Verlag, Berlin, New York, 1989.

[4] B. Derrida. Random energy model: Limit of a family of disordered models. *Phys.Rev.Lett.*, 45:79–82, 1980.

[5] M. Eigen and P. Schuster. The hypercycle A: A principle of natural self-organization : Emergence of the hypercycle. *Naturwissenschaften*, 64:541–565, 1977.

[6] Y. Fu and P. W. Anderson. Application of statistical mechanics to NP-complete problems in combinatorial optimization. *J.Phys.A:Math.Gen.*, 19:1605–1620, 1986.

[7] Y. Fu and P. W. Anderson. The use and abuse of statistical mechanics in computational complexity. In D. Stein, editor, *Lectures in the Science of Complexity*, SFI Studies in Science of Complexity. Addison-Wesley, Longman, 1989.

[8] L. Grover. Local search and the local structure of NP-complete problems. *Oper.Res.Lett.*, 12:235–243, 1992.

[9] R. W. Hamming. Error detecting and error correcting codes. *Bell Syst. Tech. J.*, 29:147–160, 1950.

[10] S. Kauffman. *The Origin of Order.* Oxford University Press, New York, Oxford, 1993.

[11] B. Krakhofer. Local optima in landscapes of combinatorial optimization problems. Master's thesis, University of Vienna, Dept. of Theoretical Chemistry, 1995.

[12] W. Liao. Graph bipartitioning problem. *Phys. Rev. Lett.*, 59:1625–1628, 1987.

[13] B. Manderick, M. de Weger, and P. Spiessen. The genetic algorithm and the structure of the fitness landscape. In R. K. Belew and L. B. Booker, editors, *Proceedings of the 4th International Conference on Genetic Algorithms.* Morgan Kaufmann Inc., 1991.

[14] R. Palmer. Optimization on rugged landscapes. In A. S. Perelson and S. A. Kauffman, editors, *Molecular Evolution on Rugged Landscapes: Proteins, RNA, and the Immune System*, pages 3–25. Addison Wesley, Redwood City, CA, 1991.

[15] D. Sherrington and S. Kirkpatrick. Solvable model of a spin-glass. *Physical Review Letters*, 35(26):1792 – 1795, 1975.

[16] P. F. Stadler. Correlation in landscapes of combinatorial optimization problems. *Europhys. Lett.*, 20:479–482, 1992.

[17] P. F. Stadler. Landscapes and their correlation functions. Santa Fe Institute Preprint 95-07-067, 1995.

[18] P. F. Stadler. Towards a theory of landscapes. In R. López Peña, editor, *Complex Systems and Binary Networks*. Springer-Verlag, New York, 1995. in press, Santa Fe Institute Preprint 95-03-30.

[19] P. F. Stadler and R. Happel. Correlation structure of the landscape of the graph-bipartitioning-problem. *J. Phys. A.: Math. Gen.*, 25:3103–3110, 1992.

[20] P. F. Stadler and R. Happel. Random field models for fitness landscapes. SFI preprint 95-07-069, 1996.

[21] P. F. Stadler and B. Krakhofer. Local minima of p-spin models. Santa Fe Institute Preprint 95-07-076, 1995.

[22] P. F. Stadler and W. Schnabl. The landscape of the traveling salesman problem. *Phys. Letters A*, 161:337–344, 1992.

[23] E. D. Weinberger. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biol.Cybern.*, 63:325–336, 1990.

[24] W. Wiethege and D. Sherrington. Bipartioning of random graphs with fixed extensive valence. *J. Phys. A: Math. Gen.*, 20:L9–L11, 1987.

[25] S. Wright. The roles of mutation, inbreeding, crossbreeeding and selection in evolution. In D. F. Jones, editor, *Int. Proceedings of the Sixth International Congress on Genetics*, volume 1, pages 356–366, 1932.