# Local Minima of p-Spin Models

**Estados metaestables para los Hamiltonianos de p espines**

PETER F. STADLER[a,b,c,*] AND BÄRBEL KRAKHOFER[a]

[a]Institut f. Theoretische Chemie, Univ. Wien, Austria

[b]The Santa Fe Institute, Santa Fe, New Mexico, U.S.A.

[c]Instituto de Física, Universidad Nacional Autónoma de México, México D.F.

[*]Mailing Address: Peter F. Stadler
Inst. f. Theoretische Chemie, Universität Wien
Währingerstr. 17, A-1090 Wien, Austria
Phone: **43 1 40480 665 Fax: **43 1 40480 660
E-Mail: studla@tbi.univie.ac.at or stadler@santafe.edu

**Abstract**

Numerical estimates for the number of local optima of $p$-spin Hamiltonians with $p$=3,4,5 and 6 are reported. The data can be explained by means of a simple estimate in terms of the "correlation length" of the Hamiltonian.

**Resumen**

Las estimaciones numéricas para el número de estados metaestables para los Hamiltonianos de $p$ espines con $p$=3,4,5 y 6 son reportados. Los datos pueden ser explicados usando una aproximación simple en términos de "la longitud de correlación" del Hamiltoniano.

**PACS Classification:** 75.10.Nr

# 1. Introduction

A (combinatory) *landscape* [8] is a real valued function defined on the vertex set $V$ of a finite, but usually huge graph $\Gamma$. The Hamiltonian of a spin glass can be viewed as a landscape, provided a neighborhood relation is introduced on the set of all possible spin configurations. In this contribution we shall be concerned exclusively with Ising spins $\sigma_i = \pm 1$. Let $n$ denote the number of spins. The configuration space $V$ therefore consists of the set of the $2^n$ spin vectors $\sigma = (\sigma_1, \dots, \sigma_n)$. Defining neighborhood by single spin flips arranges $V$ as the vertices of the hypercube graph $\mathcal{Q}_2^n$. A local minimum of a landscape $f : V \to \mathbb{R}$ is defined as a configuration $\hat{x} \in V$ such that $f(\hat{x}) \leq f(y)$ for all $y$ that are neighbors of $x$ in the graph $\Gamma$. The use of $\leq$ instead of $<$ is conventional [10] and will not affect our conclusions. A local minimum is hence a spin configuration that is stable against single spin flips.

The number of local optima is a most important characteristic of a landscape because it can be used to measure its *ruggedness* [13], which is, for instance, of crucial importance for the performance of optimization heuristics [9]. In this contribution we report the results of a numerical survey of the local minima of long-range $p$-spin models [5] which are defined by the Hamiltonian

$$\mathcal{H}_p(\sigma) \stackrel{\text{def}}{=} \frac{1}{S} \sum_{i_1 i_2 \ldots i_p} J_{i_1 i_2 \ldots i_p} \sigma_{i_1} \sigma_{i_2} \ldots \sigma_{i_p} \,. \tag{1}$$

where the coupling constants $J_{i_1 i_2 \ldots i_p}$ are i.i.d. random variables with Gaussian distribution. The normalization constant is chosen such that the variance of the energy is unity. This normalization is irrelevant for our purposes.

## 2. Some Exact Results

We are interested in the probability $P(n;p)$ that a given configuration $\sigma$ is a local minimum of the $p$-spin Hamiltonian defined in eqn.(1). This quantity is easily evaluated in some extreme cases.

$p = 1$. In this case the Hamiltonian

$$\mathcal{H}(\sigma) = \sum_{i=1}^{n} J_i \sigma_i \tag{2}$$

describes a paramagnet. The only local minimum is the configuration at which $J_i\sigma_i < 0$ for all $i$, since the energy of all other configurations can obviously be decreased by flipping a spin $\sigma_i$ for which $J_i\sigma_i > 0$. Since the degenerate cases $J_i = 0$ occur with probability 0, we have $P(n;1) = 2^{-n}$.

$p = n$. In this case the Hamiltonian reduces to the single term

$$\mathcal{H}_n(\sigma) = J_{123\ldots n} \sigma_1 \sigma_2 \ldots \sigma_n \,. \tag{3}$$

Flipping a single spin, $\sigma \to \sigma'$, implies $\mathcal{H}_n(\sigma') = -\mathcal{H}_n(\sigma')$, and thus each configuration is either a local minimum or a local maximum. Therefore $P(n;n) = 1/2$.

$n = 3$. We already know $P(3;1) = 1/8$ and $P(3;3) = 1/2$. The remaining probability $P(3;2)$ can be obtained by a combinatorial argument. We observe the symmetry $\mathcal{H}_2(\sigma) = \mathcal{H}_2(-\sigma)$ and consider the ground state configuration $\hat{\sigma}$. It is a local minimum by definition, and in the absence of degeneracies its three neighbors cannot be local minima. The symmetry of the Hamiltonian implies that $-\hat{\sigma}$ is also an isolated local minimum with probability 1. The sets of neighbors of $\hat{\sigma}$ and $\hat{\sigma}$, respectively, are disjoint and contain three configurations each. Hence we know that for any generic choice of the coupling constants 2 out of the possible 8 configurations are minima, while the remaining 6 configurations are not. Therefore $P(3;2) = 1/4$. Unfortunately, combinatorial arguments of this type seem to fail for larger $n$.

$p = 2$ corresponds to the Sherrington-Kirkpatrick model [14]. It received considerable attention around 1980; at least three groups have computed the number of local minima of the SK model by means of what are now considered standard methods in Statistical Mechanics. Tanaka and Edwards [21] computed the expected number of local optima $\langle g_0 \rangle$, while Bray and Moore [1] and De Dominicis *et al.* [3] used a replica approach to evaluate $\langle \ln g_0 \rangle$. These papers provide also a detailed analysis of the distribution of local minima as a function of their energies. The common result of the three groups is

$$\lim_{n\to\infty} \frac{1}{n} \ln\langle g_0 \rangle = \lim_{n\to\infty} \frac{1}{n} \langle \ln g_0 \rangle = \alpha(2) \approx 0.19923 \tag{4}$$

The numerical value is obtained as the solution of a set of coupled algebraic equations. For the case of short range spin glasses, in which only a small number $z$ of coupling constants $J_{ij}$ are non-zero for any given spin $i$, a slightly larger number of local optima has been found

$$\lim_{n\to\infty} \frac{1}{n} \ln\langle g_0 \rangle = \lim_{n\to\infty} \frac{1}{n} \langle \ln g_0 \rangle = \alpha(2) + \frac{\alpha'}{z} + \mathcal{O}(z^{-2}) \tag{5}$$

where $\alpha' \approx 0.0656$ [2, 21]. The only known case in which the logarithmic average deviates from the direct average is the linear spin chain. Derrida and Gardner [7] found $\ln\langle g_0 \rangle/n \to \ln(4/\pi) \approx 0.2416$ and $\langle \ln g_0 \rangle/n \to (\ln 2)/3 \approx 0.2310$ for this example.

## 3. Numerical Simulations

In this contribution we are interested exclusively in the long range case, for which we expect no difference between the direct and the logarithmic average. Numerical simulations for small $n$ in fact do not show a difference. Since the direct average can be estimated from much smaller samples it has been used to generate the data reported in this section.

We have used two models for the coupling constants. Following the original definition of the $p$-spin models [5] we choose the coupling constants $J_{i_1 i_2 \ldots i_p}$ to be i.i.d. Gaussian random variables with mean 0 and variance 1. As an alternative model we used i.i.d. random variables distributed uniformly between $-1$ and 1. We find that the number of local minima is the same for both distributions. There are no trends in the data that would suggest that the number of local minima depends on the the choice of the distribution function.

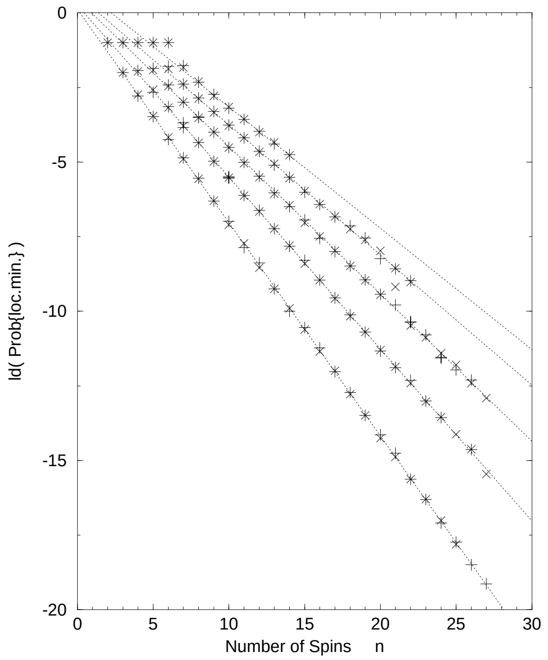


**Figure 1:** Logarithm (of base 2) of the frequencies of local minima in $p$-spin Hamiltonians versus the number $n$ of spins.
The symbol $\times$ refers to the classical $p$-spin models with Gaussian coupling constants, while $+$ denotes a variant with uniformly distributed coupling constants. The coupling order $p$ increases from bottom to top.

Numerical estimates have been obtained by testing a moderate number (usually 1000) of configurations for being local optima of the Hamiltonian with a given (random) choice of the coupling constants. The result was then averaged over $10^3$ to $10^5$ different sets of coupling constants. The reason for testing about 1000 configurations on the same landscape is that the assignment of the coupling constants requires substantial computational resources for large $p$. Due to the memory requirements for storing the coupling constants we had to restrict ourselves to $p \leq 6$.

**Table 1.** Best fits for the parameters $\psi(p)$ from the data in figure 1.

| $p$ | $\psi \pm$ sdv | corr. | $k$ |
|---|---|---|---|
| 2 | $-0.7144 \pm 0.0011$ | $-0.999945$ | 1 |
| 3 | $-0.5761 \pm 0.0011$ | $-0.999914$ | 2 |
| 4 | $-0.4937 \pm 0.0028$ | $-0.999261$ | 1 |
| 5 | $-0.4363 \pm 0.0021$ | $-0.999631$ | 2 |
| 6 | $-0.4061 \pm 0.0036$ | $-0.999521$ | *2 |

* The data suggest that $\psi(6)$ is slightly (about a standard deviation) underestimated. Unfortunately, the memory required for storing the coupling constants does not allow for the investigation of larger systems.

We use the following notation for the constants governing the exponential scaling of the number of local minima

$$P(n;p) \stackrel{\text{def}}{=} \frac{\langle g_0 \rangle}{N} \sim \mu^n = 2^{\psi\, n} = \exp(\alpha\, n)/N \tag{6}$$

where $N$ is the total number of configurations, in our case $N = 2^n$. Therefore we have $\psi = \alpha/\ln 2 - 1 = \ln\mu/\ln 2$. The raw data are collected in figure 1. $P(n;p)$ shows an exponential decay with $n$ for constant values of the parameter $p$. There are minor deviations from an exponential function for small $n$, roughly $n \leq p+3$. In order to minimize the impact of these finite size effects we used the following procedure. Omitting the first $k$ entries from the data set we accept the estimate of the linear regression analysis for $\psi(p)$ that maximizes the correlation coefficient when $k$ is varied. We have performed the analysis separately for the data with the uniform and the Gaussian distribution of the coupling constants, respectively. Since the two data sets are consistent with the hypothesis that $P(n;p)$

is independent of which of the two probability distributions are chose, we have combined them for producing our best estimates, which are displayed in Table 2.

Our numerical value for the SK model, $\psi(2) = -0.714 \pm 1$ is consistent with the analytical value $\psi(2) \approx -0.7126$ given in refs. [1, 21, 3]. As expected, the number of local minima increases with the order $p$ of the coupling.

## 4. Local Optima and Correlation Length

Weinberger [22] suggested to characterize the ruggedness of (fitness) landscapes by means of correlation functions. He proposed to sample energy values along a simple unbiased random walk and to use the autocorrelation function of this "time-series", $r(s) \stackrel{\mathrm{def}}{=} (\langle f(x_t)f(x_{t+s})\rangle - \langle f\rangle^2)/(\langle f^2\rangle - \langle f\rangle^2)$, or rather the corresponding correlation length, $\ell \sum_{s=0}^{\infty} r(s)$, as a measure of ruggedness.

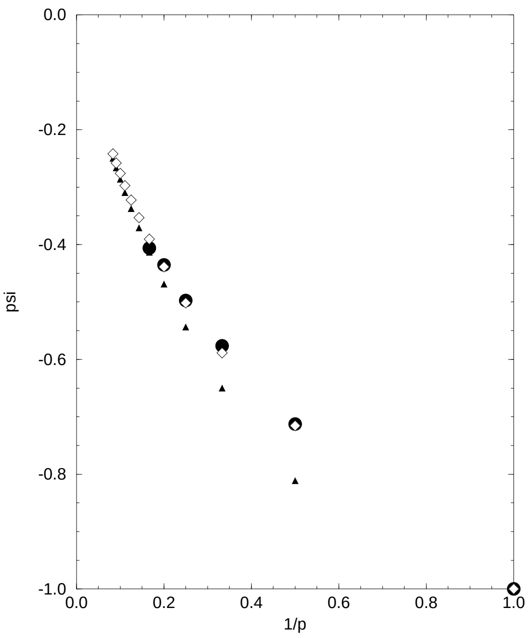


**Figure 2:** Dependence of the exponent $\psi$, equ.(6), on the coupling order $p$. Numerical simulations are indicated by $\bullet$, the estimates obtained by assuming one local optimum in a ball with radius $R$ are shown as $\diamond$. The full triangles are obtained by assuming one local optimum in a ball with radius $\ell$.

The function $r(s)$ can be readily computed for the $p$-spin models [23]. The same result can be obtained more easily by means of a general algebraic approach [17, 15, 16, 19] which is based on the observation that the $p$-spin Hamiltonian, eqn.(1), is an eigenfunction of the graph Laplacian $-\Delta$ [12] associated with the hypercube $\mathcal{Q}_2^n$. For an regular graph $\Gamma$ the Laplacian is defined as

$$-\Delta = \mathbf{A} - D\mathbf{I} \tag{7}$$

where $\mathbf{I}$ is the identity matrix, $D$ is the common number of neighbors of each configuration and $\mathbf{A}$ is the *adjacency matrix* of the configuration space with entries $\mathbf{A}_{xy} = 1$ if configurations $x$ and $y$ are neighbors and 0 otherwise. One finds explicitly

$$-\Delta\, \mathcal{H}_p = 2p\mathcal{H}_p \tag{8}$$

for $1 \leq p \leq n$ which implies

$$r(s) = \left(1 - \frac{2p}{n}\right)^s , \tag{9}$$

and a correlation length $\ell = n/2p$. The special properties of eigenfunctions of graph Laplacians are discussed in [16].

Since both the number of local optima and the correlation length have been used to quantify "ruggedness" it seems natural to look for a connection between these quantities. Since $\ell$ does not depend on the distribution of the $J_{ij}$ while $\alpha$ depends on the number $z$ of non-zero entries for a given spin $i$, there cannot be a simple functional relationship. Nevertheless one can look for such a connection at least in the long range case, which corresponds to a "maximum entropy" assumption *given* the correlation length.

Stadler and Schnabl [20] conjectured that there is on the order of one local minimum within a ball in the configuration space graph $\Gamma$, the radius $R$ of which is given by the correlation length $\ell$ of the landscape. This conjecture was tested for traveling salesman problems, and was found to be remarkably accurate. It seems more natural to require that the radius $R$ is not $\ell$ but the average distance traveled in $\ell$ steps along the simple random walk that is used to define the autocorrelation function $r(s)$. In the TSP example considered in [20] the difference between $\ell$ and $R$ is very small, and hence this distinction was not discussed there.

For a Boolean hypercube it is not hard to explicitly compute

$$R \stackrel{\text{def}}{=} \sum_{d} \varphi_{\ell d} d \,, \tag{10}$$

where $\varphi_{\ell d}$ is the probability that the random walk is at distance $d$ from its initial point after $\ell$ steps. Along the lines of [18] one finds

$$R = \frac{n}{2}\left[1 - \left(1 - \frac{2}{n}\right)^{\ell}\right] \approx \frac{n}{2}\left[\,1 - \exp(-2\ell/n)\,\right]. \tag{11}$$

Since in our case the correlation length $\ell$ is proportional to $n$, it will be convenient to use the scaled variables $\xi = \ell/n$ and $\hat{\xi} = R/n$, respectively. For the $p$-spin models we have $\xi = 1/(2p)$. The relation between $R$ and $\ell$ becomes

$$\hat{\xi} = \frac{1}{2}\left(1 - \mathrm{e}^{-2\xi}\right) \,. \tag{12}$$

in the limit $n \to \infty$.

The number of configurations inside a ball of radius $r = xn$, $0 < x < 1/2$, is also easily evaluated for the Boolean hypercube graph $\mathcal{Q}_2^n$:

$$B(r) = \sum_{q=0}^{r} \binom{n}{q} \approx \binom{n}{r}. \tag{13}$$

The error introduced by approximating the sum by its largest term is a factor of order at most $\mathcal{O}(n)$ which does not affect the exponential growth of $B(r)$. As long as $x$ is bounded away from 0 we can replace the three Gamma functions in the binomial coefficient by Stirling's formula, yielding

$$B(r) \approx \binom{n}{nx} \approx \left[\left(\frac{1-x}{x}\right)^{x} \frac{1}{1-x}\right]^{n} \frac{1}{\sqrt{2\pi x(1-x)n}}\,. \tag{14}$$

Neglecting all non-exponential terms we have $B(nx) \sim \mu(x)^{-n}$, where

$$\mu(x) \stackrel{\text{def}}{=} (1-x)\left(\frac{x}{1-x}\right)^{x} \,. \tag{15}$$

We conjecture therefore that $P(n,p) \sim \mu(\hat{\xi})^n$. A comparison with the numerical simulation shows an excellent agreement, while using $\xi$ instead of $\hat{\xi}$ consistently underestimates the number of local optima, see table 2 and figure 2.

**Table 2.** Comparison between the numerical data for $\mu = 2^{\psi}$ and the estimates based on the correlation length $\ell = n/(2p)$ of the random walk autocorrelation functions.

| $p$ | $\mu_{\text{num}}$ | $\mu(\hat{\xi})$ | $\mu(\xi)$ |
|---|---|---|---|
| 2 | 0.6095 | 0.6091 | 0.5699 |
| 3 | 0.6708 | 0.6649 | 0.6372 |
| 4 | 0.7102 | 0.7062 | 0.6861 |
| 5 | 0.7391 | 0.7379 | 0.7222 |
| 6 | 0.7546 | 0.7628 | 0.7506 |

## 5. Discussion

We have presented numerical data showing that the number of local minima of a $p$-spin Hamiltonian increases exponentially with the number of spins. The fraction of configurations that are local minima decreases exponentially as $2^{\psi(p)n}$. The parameter $\psi$ depends strongly on the interaction order $p$. As expected, the number of local optima increases with $p$, consistent with the expectation that more rugged landscapes have a larger number of local minima.

The values of $\psi(p)$ can be explained by assuming that there is about one local minimum contained in a ball whose radius is determined by the correlation length $\ell$ of the landscape. We have at present **no** explanation why this hypothesis yields such an excellent description of the numerical data reported in this study as well as of the TSP data reported in [20]. Its applicability to a number of different landscapes suggests that it is more than a numerical coincidence.

A family of very rugged landscapes is obtained by choosing $p = \mathcal{O}(n)$ instead of a constant. It has been shown recently that a random energy model [4, 6] with a Gaussian distribution of values can be represented [16] as a superposition of $p$-spin models in the form

$$REM(\sigma) = \sum_{p=1}^{n} 2^{-n} \binom{n}{p} \mathcal{H}_p(\sigma). \tag{16}$$

The dominating terms on this expansion are those with $p \approx n/2$. Consequently, one might expect that $\mathcal{H}_{n/2}$ behaves roughly like the random energy model. In particular this would imply $P(n; n/2) \sim 1/n$, since $P(REM(n)) = 1/(n+1)$, see e.g. [11]. The numerical data reported in Figure 1 are consistent with this assumption, although they are not very accurate because we have data only up to $n = 12$. The "correlation length hypothesis" leads to power laws for the Hamiltonians $\mathcal{H}_{yn}$ with $0 < y \leq 1/2$: The correlation length is $\ell = 1/(2y)$ and hence

$$B(\ell) \approx B(R) \sim \binom{n}{1/(2y)} \sim n^{1/(2y)} \,. \tag{17}$$

Again we arrive at the prediction $P(n, n/2) \sim 1/n$. For $y > 1$ the autocorrelation function $r(s)$ has alternating signs and the definition of a correlation length becomes ambiguous.

**Acknowledgments**

The main part of this study was performed while P.F.S. was a guest of UNAM in Mexico City. Thanks to the colleagues at the Instituto de Física and the Instituto de Ciencias Nucleares for their hospitality, and special thanks to Ricardo García-Pelayo for the stimulating discussion we had.

## References

[1] A. J. Bray and M. A. Moore. Metastable states in spin glasses. *J.Phys.C*, 13:L469–L476, 1980.

[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] C. De Dominicis, M. Gabay, T. Garel, and H. Orland. White and weighted averages over solutions of Thouless Anderson Palmer equations for the Sherrington Kirkpatrick spin glass. *J.Physique*, 41:923–930, 1980.

[4] B. Derrida. The random energy model. *Phys.Rep.*, 67:29–35, 1980.

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

[6] B. Derrida. Random-energy model: An exactly solvable model of disordered systems. *Phys. Rev. B*, 24(5):2613–2626, 1981.

[7] B. Derrida and E. Gardner. Metastable states of a spin glass chain at 0 temperature. *J.Physique*, 47:959–965, 1986.

[8] W. Fontana, P. F. Stadler, E. G. Bornberg-Bauer, T. Griesmacher, I. L. Hofacker, M. Tacker, P. Tarazona, E. D. Weinberger, and P. Schuster. RNA folding and combinatory landscapes. *Phys. Rev. E*, 47(3):2083 – 2099, 1993.

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

[10] W. Kern. On the depth of combinatorial optimization problems. *Discr.Appl.Math.*, 43:115–129, 1993.

[11] C. Macken, P. Hagan, and A. Perelson. Evolutionary walks on rugged landscapes. *SIAM J.Appl.Math.*, 51:799–827, 1991.

[12] B. Mohar. The laplacian spectrum of graphs. In Y. Alavi, G. Chartrand, O. Ollermann, and A. Schwenk, editors, *Graph Theory, Combinatorics, and Applications*, pages 871–897, New York, 1991. John Wiley & Sons.

[13] 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.

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

[15] P. F. Stadler. Linear operators on correlated landscapes. *J.Physique*, 4:681–696, 1994.

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

[17] P. F. Stadler. Random walks and orthogonal functions associated with highly symmetric graphs. *Disc. Math.*, 145:229–238, 1995.

[18] 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.

[19] P. F. Stadler and R. Happel. Random field models for fitness landscapes. Santa Fe Institute Preprint 95-07-069, 1995.

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

[21] F. Tanaka and S. F. Edwards. Analytic theory of ground state properties of a spin glass: I. Ising spin glass. *J.Phys.F*, 10:2769–2778, 1980.

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

[23] E. D. Weinberger and P. F. Stadler. Why some fitness landscapes are fractal. *J. Theor. Biol.*, 163:255–275, 1993.