

A method that reveals the multi-level ultrametric tree hidden in p -spin-glass-like systems

R Baviera^{1,2} and M A Virasoro^{1,3}

¹ Dipartimento di Fisica, *Sapienza* Università di Roma, Piazzale Aldo Moro, 5, 00185 Roma, Italy

² Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milano, Italy

³ Instituto de Ciencias, Universidad Nacional General Sarmiento, Juan María Gutiérrez 1150, 1613 Los Polvorines, Buenos Aires, Argentina

E-mail: roberto.baviera@gmail.com and mvirasoro@gmail.com

Abstract.

In the study of disordered models like spin glasses the key object of interest is the rugged energy hypersurface defined in configuration space. The statistical mechanics calculation of the Gibbs–Boltzmann partition function gives the information necessary to understand the equilibrium behavior of the system as a function of the temperature but is not enough if we are interested in the more general aspects of the hypersurface: it does not give us, for instance, the different degrees of ruggedness at different scales. In the context of the replica symmetry breaking (RSB) approach we discuss here a rather simple extension that can provide a much more detailed picture. The attractiveness of the method relies on the fact that it is conceptually transparent and the additional calculations are rather straightforward. We think that this approach reveals an ultrametric organisation with many levels in models like p -spin glasses when we include saddle points. In this first paper we present detailed calculations for the spherical p -spin glass model where we discover that the corresponding decreasing Parisi function $q(x)$ codes this hidden ultrametric organisation.

Keywords: cavity and replica method, disordered systems (theory), energy landscapes (theory), spin glasses (theory)

1. Introduction

Some time ago we proposed [1–3] a method to introduce more control parameters (in addition to temperature) in order to analyse the detailed ruggedness of an energy hypersurface of a disordered complex system.

In so doing we were following a long line of research that considers systems composed of several real replicas (in addition to the virtual replicas of the replica trick) satisfying various constraints [4–8].

Our proposal here pushes the assumption of ultrametricity to new levels as we will be investigating the distribution of saddle points consistent with ultrametricity. The skeptical reader may perceive this as an unintended and unjustified consequence of past successes with the hypothesis. We hope that the pragmatic reader will instead appreciate the elegance and simplicity of the calculations and join us in hoping that this extension be proved correct. In any case the cost of configurations that do not comply with ultrametricity can be analysed by other methods (see for instance [5, 6, 9, 10]).

In pursuing this approach we were inspired by Derrida’s *generalised random energy model* [11] where a multilevel ultrametric organisation built in can be revealed or hidden in the Gibbs–Boltzmann measure depending on the parameters of the model.

Let us then imagine a configuration space partitioned in a hierarchical way along an ultrametric tree with $K + 1$ layers labeled $(0, 1 \dots K)$ (K replica symmetry levels;

hereinafter RSB levels). The leaves of the tree are the ergodic pure states and define regions in the configuration space limited by the Edward–Anderson overlap parameter $q_K = q_{\text{EA}}$. Each region is characterised by some kind of aggregate energy and by its size. These regions are then grouped in clusters limited by the overlap q_{K-1} that defines a larger scale partition. These clusters are themselves grouped in superclusters in a scheme that continues along the tree. At each level we will have some kind of generalised free energy for every cluster/supercluster. The q_k satisfy $q_K > q_{K-1} \cdots > q_1 > q_0$. Then the generic question at any level of the tree is how many k th clusters of overlap q_k and generalised free energy F are contained in a higher-level cluster of overlap q_{k-1} and another particular generalised free energy F' . This information is relevant, for instance, to the relaxation dynamics of the system or to the design of an efficient search algorithm of the minima and cannot, in general, be derived from the partition function

$$Z_J = \sum_{\{C\}} e^{-\beta H_J[C]}$$

where there is a single parameter $\beta()$ because the sum aggregates information that one needs to keep apart. As usual the subscript J reminds us that in the definition of H_J there are an infinite number of random quenched parameters and that we are dealing with an ensemble of hypersurfaces. The results take the form of analytic expressions for average quantities including correlations among local saddle points at variable distances in the configuration space. From now on we omit the J subscript and unless explicitly stated otherwise we will be referring to quantities averaged over J .

The remainder of the paper is organized as follows: in section 2, subsection 2.1, we first show how to introduce another parameter in addition to the temperature to disaggregate the contributions coming from states with different free energies and configurational complexities so as to ‘see’ smaller or shallower states that would not contribute to Z because of their small weight and larger or deeper ones that do not contribute because of their reduced number. At this point our method turns out to be equivalent to Monasson’s proposal [8]. The advantage of our approach becomes apparent in section 2.2 where a generalisation that adds up to K parameters is considered.

In sections 3 and 4 we present in full detail the case of a p -spin spherical model for a hierarchical tree with 3 levels ($K = 2$ RSB levels). The generalisation for any K is straightforward. In this example we will use what is known about the solutions to the Thouless–Anderson–Palmer (TAP) equations. We will show that the corresponding ultrametric tree is revealed by the monotonically decreasing $x(q)$ function [6]. The presentation will show that in general, for any system, in a particular region of the constraint parameters (when the number of real replicas is large enough) the free energy of the system *per replica* becomes equal to a saddle point solution of the unconstrained system.

In section 5 we discuss some preliminary conclusions.

2. The proposal: real replicas organised ultrametrically

To simplify the presentation we divide the proposal into 2 subsections and assume $K = 2$. It will become obvious how to generalise it for arbitrary K .

2.1. Revealing the states when their statistical weight is too small

In this case we consider the partition function of a system composed of R real replicas, all of which have mutual overlap q . Specifically:

$$Z(\beta, R, q) = \sum_{\{\mathcal{C}_a; a=1, \dots, R\}} e^{-\beta \sum_{a=1}^R H[\mathcal{C}_a]} \prod_{\{a, a'; a \neq a'\}} \delta(\mathcal{C}_a \cdot \mathcal{C}_{a'} - q) \quad (1)$$

where $\mathcal{C}_a \cdot \mathcal{C}_{a'}$ is the overlap between the configurations \mathcal{C}_a and $\mathcal{C}_{a'}$ and both a, a' run from 1 to R . At this moment R is larger than 1 and is an integer but, as usual, after the calculations we will derive an analytic expression and treat R as a real number. Now q roughly delimits a region: due to the properties of the energy hypersurface if a configuration has a finite weight there will be a relatively large number of configurations nearby that will also contribute to Z . Ultrametricity implies that these configurations lie inside a ball with maximum overlap q_{EA} . If $q \geq q_{\text{EA}}$ all the R replicas will be inside that ball and if $q = q_{\text{EA}}$ the entropy per replica will be maximal. The calculation is particularly simple in this case. Therefore we choose:

$$\frac{\partial \log Z(\beta, R, q)}{\partial q} = 0 \Rightarrow q = q_{\text{EA}} \quad (2)$$

We can also write

$$\log Z(\beta, R) = -\beta R E_S + R S_S + \log \mathcal{N}_S \quad (3)$$

$$F_S = E_S - \frac{1}{\beta} S_S = \frac{1}{\beta} \frac{\partial \log Z(\beta, R)}{\partial R} \quad (4)$$

$$\log \mathcal{N}_S = \log Z(\beta, R) + \beta R F_S \quad (5)$$

where we identify E_S as the internal energy, S_S as the state entropy and \mathcal{N}_S as the number of states.

These equations, equal to those in [8], separate the number of configurations per state and the number of states. They should be read in general as a definition of F_S , the free energy per replica, to be identified with previous definitions when the dependence of F_S on R is sufficiently small. Mathematically they reveal a multidimensional Legendre transformation allowing a change of variables from R to F_S and therefore the number of states expressed as a function of the free energy follows and the Legendre transform properties imply:

$$\frac{\partial \log \mathcal{N}_S}{\partial F_S} = \beta R \quad (6)$$

Parenthetically we observe that the lowest energy configuration in a state (the bottom of the corresponding valley) is determined implicitly by the equation:

$$S_S(E_{\text{bottom}}) = 0 \quad (7)$$

It is then evident that

$$E_{\text{bottom}} \approx E_S - \frac{S_S(E_S)}{\frac{\partial S_S(E_S)}{\partial E_S}} = F_S \quad (8)$$

So F_S is the linear extrapolation estimate of E_{bottom} from an expansion of S_S around E_S . Similar arguments will apply going up along the tree; the free energy of a cluster will be a good estimator of the lowest free energy state inside the cluster. This is relevant to the feasibility of a hierarchical search, and is true in practically any system. In the last section with the conclusions we will discuss what goes wrong in p -spin glass systems: why in this case of hierarchical search, for instance, simulated annealing will not work in spite of this general property.

2.2. Clusters of states

We now consider R_1 replicas of groups of R_2 configurations such that the latter are constrained to have an overlap q_2 , while replicas belonging to different groups have an overlap $q_1 < q_2$.

Now the number of states defined by q_2 is decomposed in two components: the number of states belonging to a cluster $\mathcal{N}_{S \in \mathcal{V}}$ and the number of clusters $\mathcal{N}_{\mathcal{V}}$.

Then

$$Z(\beta, R_1, R_2, q_1, q_2) = \sum_{\{C_{ab}; a=1 \dots R_1; b=1 \dots R_2\}} e^{-\beta \sum_{a,b} H[C_{ab}]} \prod_{a,b \neq b'} \delta(C_{ab} \cdot C_{ab'} - q_2) \prod_{a \neq a', b, b'} \delta(C_{ab} \cdot C_{a'b'} - q_1) \quad (9)$$

where the indices a, a' run from 1 to R_1 while b, b' run from 1 to R_2 . We will again choose q_1, q_2 such that

$$\frac{\partial \log Z(\beta, R_1, R_2, q_1, q_2)}{\partial q_1} = \frac{\partial \log Z(\beta, R_1, R_2, q_1, q_2)}{\partial q_2} = 0 \quad (10)$$

Then $\log Z$ is a function of only R and β and as in equation (3) can be decomposed into:

$$\log Z(\beta, R_1, R_2) = -\beta R_1 R_2 E_S + R_1 R_2 S_S + R_1 \log \mathcal{N}_{S \in \mathcal{V}} + \log \mathcal{N}_{\mathcal{V}} \quad (11)$$

This equation, as in equation (3), should be read as a multidimensional Legendre transform and thus be complemented with

$$\begin{aligned} \frac{\partial \log Z(\beta, R_1, R_2)}{\partial R_2} &= R_1 \beta F_S \\ \frac{\partial \log Z(\beta, R_1, R_2)}{\partial R_1} &\equiv \beta F_{\mathcal{V}} = R_2 \beta F_S + \log \mathcal{N}_{S \in \mathcal{V}}, \end{aligned} \quad (12)$$

the meaning of which is rather transparent: the first equation reads as another definition of F_S , for a replica interacting with $R_1 R_2 - 1$ other replicas while the 2nd can be read as a definition of the free energy of a cluster. The equation corresponding to equation (6) in this case is:

$$\frac{\partial \log \mathcal{N}_{\mathcal{V}}}{\partial F_{\mathcal{V}}} = \beta R_1 \quad (13)$$

In these formulae $\log \mathcal{N}$ means $\log \overline{\mathcal{N}}$ rather than $\overline{\log \mathcal{N}}$. We will be using them where they are positive and extensive or at least where the sum is positive and extensive. In the latter case, for instance, if

$$\log \mathcal{N}_\gamma > 0; \log \mathcal{N}_{S \in \gamma} < 0; \log \mathcal{N}_\gamma > \log \mathcal{N}_{S \in \gamma}$$

then we can read $\log \mathcal{N}_{S \in \gamma} \mathcal{N}_\gamma$ as the total number of states with free energy F_S contained in clusters of a certain free energy F_γ .

We observe that configurations inside a state are weighed by the parameter β , while states inside a cluster use the parameter βR_2 and the clusters themselves $\beta R_1 R_2$.

The generalisation of this approach to a generic ultrametric tree with K total levels is now straightforward so we spare the reader a cumbersome notation. A warning however is worth repeating: as we have stressed we will choose all the overlaps q_k of the constraints to coincide with the saddle points of the system. This is the key to the simplicity of the calculations and implies that all relevant triangles among both real and virtual replicas satisfy ultrametricity.

3. Calculations for the spherical p -spin glass

In this section we detail the calculations for the p -spin spherical model. The Hamiltonian is [12]

$$H = - \sum_{i_1 > i_2 > \dots > i_p} J_{i_1 i_2 \dots i_p} s_{i_1} s_{i_2} \dots s_{i_p} \quad (14)$$

where the s_i are real variables subject to a spherical constraint, $\sum_{i=1, N} s_i^2 = 1$, and we have to calculate equation (9).

We apply the canonical Parisi trick ([6, 12]) to derive the free energy per spin as a function of an $n R_1 R_2 \times n R_1 R_2$ matrix $\mathbf{O}^{\alpha, \beta}$:

$$nF = -\frac{\beta}{4} \sum_{\alpha\beta} (\mathbf{O})^p - \frac{1}{2\beta} \text{Tr}[\ln \mathbf{O}] \quad (15)$$

We can represent \mathbf{O} as a matrix of matrices. The natural way to consider the \mathbf{O} matrix is to order the columns (rows) lexicographically with 3 indices: the slowest one a indexing the clusters and running from 1 to R_1 , then b indexing the states and running from 1 to R_2 and finally j indexing the n virtual replicas. This generalises the parametrisation proposed in [6].

$$\mathbf{O} = \begin{pmatrix} O^{11,11} & O^{11,12} & \dots & O^{11,1R_2} & O^{11,21} & \dots & \dots & O^{11,R_1 R_2} \\ O^{12,11} & O^{12,12} & \dots & O^{12,1R_2} & O^{12,21} & \dots & \dots & O^{12,R_1 R_2} \\ \vdots & & & & & & & \\ O^{R_1 R_2, 11} & O^{R_1 R_2, 12} & \dots & O^{R_1 R_2, 1R_2} & O^{R_1 R_2, 21} & \dots & \dots & O^{R_1 R_2, R_1 R_2} \end{pmatrix} \quad (16)$$

where $O^{ab, a'b'}$ is an $n \times n$ matrix that encodes the overlap distribution of (a, b) with the (a', b') real replicas. The diagonal terms of these matrices satisfy the constraints:

$$\begin{aligned}
&\text{Diagonal terms of } O^{ab,ab} = 1; \quad \forall a, b \\
&\text{Diagonal terms of } O^{ab,ab'} = q_2; \quad \forall a, b \neq b' \\
&\text{Diagonal terms of } O^{ab,a'b'} = q_1; \quad \forall a \neq a', b, b'
\end{aligned} \tag{17}$$

Clearly up to this point nothing is imposed on off-diagonal terms that relate different virtual replicas. Then we make the natural ansatz:

$$\begin{aligned}
O^{ab,ab} &= Q & \forall a, b \\
O^{ab,ab'} &= P_2 & \forall a, b \neq b' \\
O^{ab,a'b'} &= P_1 & \forall a \neq a', b \neq b'
\end{aligned} \tag{18}$$

with the Q, P_1, P_2 $n \times n$ Parisi matrices having the same block sizes $m_1 \times m_1, m_2 \times m_2$ as in the standard RSB approach for virtual replicas with 2 RSB levels. Once the limit $n \rightarrow 0$ is taken the corresponding functions $q(x), p_1(x), p_2(x)$ will have the usual probabilistic interpretation. For instance the probability distribution of $\mathcal{C}_{11} \cdot \mathcal{C}_{12}$ will be encoded into $p_2(x)$. Therefore the 3 functions must be monotonously increasing for the formalism to make sense.

We now restrict our consideration to the case where

$$q(x) = q_2 \theta(x - m_2) + q_1 \theta(m_2 - x) \theta(x - m_1) \tag{19}$$

$$p_2(x) = q(x) \tag{20}$$

$$p_1(x) = q_1 \theta(x - m_1). \tag{21}$$

where $\theta(\cdot)$ is the Heaviside function and $m_2 > m_1$. There are 2 different levels of assumptions behind these ansatzes, as stated in the previous section. The first one has to do with the diagonal terms of the matrices being chosen equal to the saddle point of $\log Z(\beta, R_1, R_2)$. The second is that *all* triplets of configurations $\mathcal{C}_{ab}, \mathcal{C}_{a'b'}, \mathcal{C}_{a''b''}$ obey ultrametricity. We conjecture, and we have checked it in some examples, that this 2nd statement is a consequence of the 1st one.

It is basically this second assumption that simplifies the calculations because if all triangles are ultrametric then there must be a way to reshuffle the rows and columns of \mathbf{O} to write a new matrix \mathbf{O}^U explicitly ultrametric as a Parisi $nR_1R_2 \times nR_1R_2$ matrix with 2 levels of replica symmetry breaking. It is easy to check that this matrix will have block sizes $m_2^U \times m_2^U, m_1^U \times m_1^U$ with parameters ($m_2^U = m_2R_2, m_1^U = m_1R_1R_2$). *The calculation of $\log Z(\beta, R_1, R_2)$ is then identical to the calculation for the unconstrained case just replacing the m_i variables by the m_i^U ones.*

From equation (57) in [6] we write:

$$\begin{aligned}
lz \equiv \frac{\log Z(\beta, R_1, R_2)}{R_1R_2} &= -\frac{\beta^2}{4} \{ q_1^p(m_2^U - m_1^U) + q_2^p(1 - m_2^U) - 1 \} \\
&+ \frac{1}{2} \left\{ \frac{1}{m_1^U} \ln(\Sigma_1) + \left(\frac{1}{m_2^U} - \frac{1}{m_1^U} \right) \ln(\Sigma_2) + \left(1 - \frac{1}{m_2^U} \right) \ln(\Sigma_3) \right\}
\end{aligned} \tag{22}$$

where:

$$\begin{aligned}
\Sigma_1 &= 1 - q_2 + m_2^U(q_2 - q_1) + m_1^U q_1 \\
\Sigma_2 &= 1 - q_2 + m_2^U(q_2 - q_1) \\
\Sigma_3 &= 1 - q_2
\end{aligned} \tag{23}$$

The saddle point (SP) equations will obviously be *identical* to the equations for the unconstrained system with two subtle differences:

- The investigation of the fluctuations around the saddle point is complicated by the fact that some of the matrix elements are fixed by the constraints.
- The saddle point values for the breaking locations apply to $m_2^U = m_2 R_2$, $m_1^U = m_1 R_1 R_2$ while the inequalities necessary for the probabilistic interpretation of the solution apply to m_2, m_1 .

Therefore any solution to the saddle point equations can be interpreted probabilistically if:

$$\frac{m_1^{U,SP}}{R_1} \leq m_2^{U,SP} \quad \text{and} \quad \frac{m_2^{U,SP}}{R_2} \leq 1 \tag{24}$$

which is always possible for sufficiently large R_1, R_2 . When this happens $\log Z(\beta, R_1, R_2)$ becomes equal to $R_1 R_2$ times the $\log Z$ of the unconstrained system and both $\log \mathcal{N}_V$ and $\log \mathcal{N}_{S \in V}$ are zero. In fact when $R_1(R_2)$ grows we probe deeper clusters (states in the cluster) and eventually we hit the corresponding bottom configuration where $\log \mathcal{N}_V$ ($\log \mathcal{N}_{S \in V}$) are zero. It is interesting to notice here that contrary to what happens in the SK model, in systems like p -spin glasses the lowest state in the lowest cluster lies higher in free energy than the absolute ground state. We will discuss this further in the last section.

3.1. R_1 and/or R_2 below their critical values

But even if we consider the decreasing values of R_1, R_2 reaching and surpassing their critical values $R_1^c = m_1^{u,SP}/m_2^{u,SP}$; $R_2^c = m_2^{u,SP}$ we can still take advantage of the solutions of the free unconstrained system.

In fact when R_2 (R_1) approaches R_2^c (R_1^c) the probability of 2 states having overlap q_2 (q_1) goes to zero as can be seen from the matrices' Q and P . Physical arguments imply that the same must be true for smaller values. Therefore

$$\begin{aligned}
R_1 \leq \frac{m_1^{u,SP}}{m_2^{u,SP}} &\Rightarrow m_1 = m_2 \Rightarrow m_1^U = R_1 m_2^U \\
R_2 \leq m_2^{u,SP} &\Rightarrow m_2 = 1 \Rightarrow m_2^U = R_2
\end{aligned} \tag{25}$$

This means that m_1 or m_2 hits a boundary and $\log Z(\beta, R_1, R_2)$ is no longer stationary on them. The system behaviour is then described by the equations of the

unconstrained system but with m_1^U, m_2^U now as possible control parameters. There are 4 possible scenarios:

1. $R_1 > R_1^c, R_2 > R_2^c$.
2. $R_1 > R_1^c, R_2 \leq R_2^c$
3. $R_1 \leq R_1^c, R_2 > R_2^c$
4. $R_1 \leq R_1^c, R_2 \leq R_2^c$

In all 4 cases, defining lz as in equation (21) we have:

$$\begin{aligned} \frac{\partial lz}{\partial q_1} &= \frac{\partial lz}{\partial q_2} = 0 \quad \text{Saddle Point eqs in } q_1, q_2 \\ \frac{\partial lz}{\partial 1/R_2} &= \log \mathcal{N}_{S \in V} \\ \frac{\partial lz}{\partial 1/(R_1 R_2)} &= \log \mathcal{N}_V \end{aligned} \tag{26}$$

In case 1 in addition we have 2 more SP equations:

$$\begin{aligned} \frac{\partial lz}{\partial m_1^U} &= \frac{\partial lz}{\partial m_2^U} = 0 \\ \Rightarrow \log \mathcal{N}_{S \in V} &= \log \mathcal{N}_V = 0 \end{aligned} \tag{27}$$

that detect the lowest free energy state in the lowest free energy cluster.

In case 2 we have one more SP equation:

$$\frac{\partial lz}{\partial m_1^U} = 0 \Rightarrow \log \mathcal{N}_V = 0 \tag{28}$$

and a new control parameter $m_2^U = R_2$ that allows us to probe states in the lowest free energy cluster with varying free energies above the minimum one.

In case 3 again we have one additional SP equation

$$\begin{aligned} \frac{\partial lz}{\partial m_2^U} + R_1 \frac{\partial lz}{\partial m_1^U} &= 0 \\ \Rightarrow \log \mathcal{N}_{S \in V}^{R_1} \mathcal{N}_V &= 0 \end{aligned} \tag{29}$$

and R_1 as a control parameter to explore the lowest free energy states in higher free energy clusters.

In case 4, finally, there are no additional SP equations and instead we have 2 control parameters $m_1^U = R_1 R_2$ and $m_2^U = R_2$ that explore higher free energy states lying in higher free energy clusters.

4. The saddle point equations and their interpretation in terms of $\log \mathcal{N}_{S \in V}$ and \mathcal{N}_V

In principle one should first solve q_1, q_2 in terms of m_1^U, m_2^U and then use equations (11), (12) and (26) to derive the expression of other functions in terms of m_1^U, m_2^U . But as the derivatives of $\log Z(\beta, R_1, R_2)$ with respect to the q are zero it is simpler to take the first derivatives with respect to the m at constant q and afterwards use the SP equations in q to choose any pair of independent variables to express our results. We borrow from section 4 of [6] the saddle point equations for $K = 2$ of the spherical p -spin glass and a set of variables that are convenient because their SP values are independent of β :

1. Variables:

$$w_1 = \frac{q_1}{q_2}; \quad y_1 = \frac{\Sigma_2}{\Sigma_1}; \quad y_2 = \frac{\Sigma_3}{\Sigma_2}; \quad Y = \frac{p\beta^2}{2} q_2^{p-2} (1 - q_2)^2 \quad (30)$$

2. SP equations in q_1, q_2

$$\frac{p\beta^2}{2} (q_s^{p-1} - q_{s-1}^{p-1}) = \frac{q_s - q_{s-1}}{[\Sigma_{s+1}\Sigma_s]} \quad s = 1, 2 \quad q_0 = 0 \quad (31)$$

that imply one equation among the new variables:

$$w_1^{1-p} - 1 = \frac{w_1^{-1} - 1}{y_1 y_2} \quad (32)$$

and another equation that is β dependent;

$$\frac{p\beta^2}{2} q_1^{p-2} = \frac{1}{\Sigma_2 \Sigma_1} \Rightarrow y_1 = \frac{p\beta^2}{2} q_1^{p-2} \Sigma_2^2 \quad (33)$$

that implies a 2nd equation for Y

$$Y = y_1 w_1^{2-p} y_2^2 \quad (34)$$

3. Variations with respect to m_1^U, m_2^U

$$\begin{aligned} -(m_1^U)^2 \frac{\partial \log \mathcal{N}_V}{\partial m_1^U} &= \log \mathcal{N}_V \\ &= -\frac{1}{2} \left(1 - y_1 + \log y_1 + \frac{(1 - y_1)^2}{p y_1} \right) \end{aligned} \quad (35)$$

$$\begin{aligned}
-(m_2^U)^2 \frac{\partial l_z}{\partial m_2^U} &= \log \mathcal{N}_{S \in V} \\
&= -\frac{1}{2} \left(1 - y_2 + \log y_2 + \frac{(1 - y_2)^2}{py_2} \frac{(1 - p(1 - w_1)w_1^{p-1} - w_1^p)}{(1 - w_1)(1 - w_1^{p-1})} \right)
\end{aligned} \tag{36}$$

We highlight the equality of the expression for $\log \mathcal{N}_V$ and the corresponding one for $\log \mathcal{N}_S$ derived in [13], both of them independent of the temperature when expressed in terms of the corresponding variables: y_1 and $Y = (p\beta^2/2)q_2^{p-1}(1 - q_2)^2$ respectively. This fact is evidence that the multiplicity of states and clusters is connected to the multiplicity of solutions at $T = 0$, a fact that could be checked by studying the TAP equations for the clusters derived from the cavity method. From [6, 13] we know that in the range of variations of Y the $\log \mathcal{N}_S$ is 0 when $Y = 0.354\,993$ for $p = 3$ while at $Y = 1/(p - 1)$ all states become unstable. The lower bound for y_1 is obviously the same but the higher one would depend on a stability analysis that has not been done; we are not distinguishing between stable and unstable clusters⁴.

5. Conclusions and discussion

In this section we first present the picture derived from these calculations that explains some idiosyncratic properties of the p -spin glass and in the next subsection we relate them to the roughness properties of the rugged energy hypersurface at different length scales as a way to connect this analysis to a general disordered system.

5.1. Some results for the p -spin glass system

Using equations (32) and (34) we express $\log \mathcal{N}_{S \in V}$ in terms of Y and y_1 . In figure 1(a) we show the contour plot of this function. We observe that the boundary of the region where there are an exponentially large number of states inside clusters has a positive derivative. This means that if we choose two clusters with different y_1 the one with the largest value contains states with larger Y , an unsurprising result if as expected y_1 and Y are related to the free energies of the clusters and states. The same result is expected in a model like the Sherrington–Kirkpatrick model and should in general be true. In a cluster with lower free energy we expect to find the lowest free energy state because as proved by definition the cluster free energy is an estimate of the free energy of the lowest states inside it.

On the other hand we observe that the lowest value of Y in figure 1(a) lies around 0.43, a value much larger than the ground state value 0.354 993 found in [6]. So the question is, where is the ground state? This is explained in figure 1(b) where we draw the contour plot of $\log \mathcal{N}_V \mathcal{N}_{S \in V}$, i.e. the log of the total number of states labeled by Y contained in all clusters labeled by y_1 and we witness a dramatic change of the derivative of the contour line 0. What we are finding is that the number of clusters grows

⁴In a work in progress by Parisi, Ricci-Tersenghi and MAV we have shown that the Plefka stability criterion implies $y_1 \leq 1/(p - 1)$.

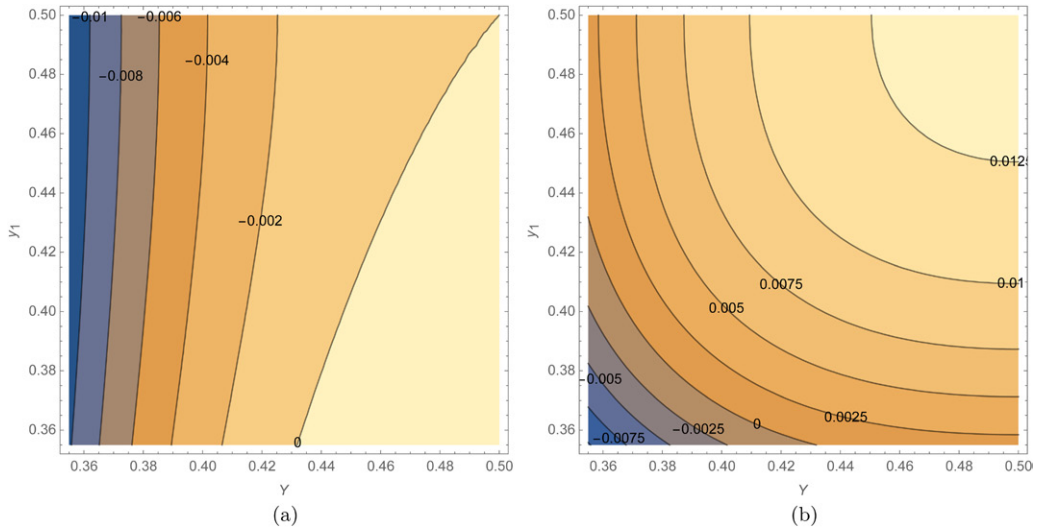


Figure 1. (a) Contour plot of $\log \mathcal{N}_{S \in V}$ as a function of y_1 and Y . (b) Contour plot of $\log \mathcal{N}_{S \in V} N_S$ as a function of y_1 and Y .

exponentially with y_1 at such a rate that even if the probability of finding a state with lower Y in a typical cluster goes exponentially to zero there remain a fraction of clusters that contain such states. This is exactly the picture suggested by the GREM and we believe it is common to all systems where one finds a decreasing Parisi $x(q)$. In fact in the linear approximation around the lowest state in the lowest cluster:

$$\begin{aligned} \log \mathcal{N}_V(F_V) &\cong \beta x(q_1)(F_V - F_{V,0}) \\ \log \mathcal{N}_{S \in V}(F_S) &\cong \beta x(q_2)(F_S - F_V) \end{aligned} \quad (37)$$

so the contour line 0 of figure 1(b)

$$\log \mathcal{N}_V N_{S \in V} = 0 \Rightarrow x(q_2)F_S + (x(q_1) - x(q_2))F_V = \text{constant}. \quad (38)$$

Therefore $x(q_1) > x(q_2)$ for $q_1 < q_2$ is telling us that an increase in F_V allows for a decrease in F_S .

The calculation for the spherical p -spin determines the value of y_1 (in figure 1(b) for $p = 3$ around 0.41) that labels the lowest free energy clusters that contain the ground state. Unfortunately, and contrary to what we argue in a previous paper [2] it is not easy to use this additional information to improve on a possible hierarchical search of the ground state. At that value of y_1 the number of clusters is exponentially large even if with a smaller exponent and we have to visit all of them since only a very small fraction (in the limit, just one) will contain the ground state. Furthermore there is no signal at that level of the search that reveals which are the good clusters. If the search inside a cluster could be done using gradient descent there would be some advantage but we do not think that even this happens.

5.2. Roughness of the energy hypersurface at different scales

The ultrametric hypothesis implies a correspondence between tree levels and length scales (q_i) in configuration space. As argued in [3] this leads to the definition of a

hierarchy of smoothed hypersurfaces obtained by flattening the energy surface in the interior of the i th level clusters. Then we can define the *roughness at the q_i level* as the variance of the fluctuations of the i th surface with respect to the $i - 1$ th surface: in practice the fluctuations of the i th clusters' energies inside an $i - 1$ th cluster; in our example the fluctuations of the q_2 states inside a q_1 cluster, coded in $\mathcal{N}_{S \in V}$ as a function of Y at fixed y_1 and the fluctuations of the q_1 cluster coded in \mathcal{N}_V as a function of y_1 . We observe that the same increment in energies (where again we use Y and y_1 as proxies for the free energies of the states and clusters) implies a much larger variation of \mathcal{N}_V than of $\mathcal{N}_{S \in V}$. This indicates that the distribution of y_1 among the clusters is more peaked than the corresponding distribution of Y among states inside the cluster, that is, the variance of y_1 is less than the variance of Y , which translates into: the roughness at the q_1 level is smaller than the roughness at the q_2 level. This is part of a picture that when first encountered may seem counterintuitive: the larger the p is, and as a consequence the more rapidly the correlation of the energy falls with distance, the smaller the hypersurface roughness at large length scales. But it is this same fact that connects these models to the 'golf course hole' landscape type and therefore 'explains' why it is difficult to find a good minimum in them.

Comparing this analysis with Derrida's study of the GREM [11] we observe that the thermodynamic behaviour of the GREM depends not only on the roughness parametrised by the width of the Gaussian energy distribution but also on the relative phase space available at the different levels, which in our case would correspond to the total number of clusters/states. In fact if we characterise the roughness distribution by a probability distribution of energies in a population of states/clusters the same probability distribution over a larger population will lead to a lower minimum energy where the derivative of the logarithm is larger thus emulating the effects of a larger roughness. In the example of this paper the number of clusters is always larger than the number of states inside a cluster and therefore the effect of the larger phase space considerably reduces the effective roughness at the q_1 level when compared with the roughness at the q_2 level.

5.3. Future work

There are many things that remain to be done in the follow-up to this work. The most immediate is the analysis of the stability of the solutions of the constrained system. Another one is to apply the method to mixtures of different p 's spin glasses. There is also the idea of exploring alternative $q(x)$ for these same systems. Further work on these aspects is the subject of present work (Parisi, Ricci-Tersenghi, MAV, in preparation).

Acknowledgments

One of us (MAV) would like to acknowledge relevant conversations with G Parisi and F Ricci-Tersenghi. Many of the results in this paper were included in Baviera's thesis [1], which is almost 20 years old. The decision to publish them originates basically from Parisi's interest and encouragement.

References

- [1] Baviera R 1995 Metodo cavità vetri spin vincolati (thesis) *Technical Report* Sapienza Università di Roma, Roma
- [2] Virasoro M A 1996 *Simulated Annealing Methods under Analytical Control* (*StatPhys* vol 19) ed B L Hao (Singapore: World Scientific Publishing) pp 205–14
- [3] Baviera R and Virasoro M A 1997 Multiscale analysis of hierarchical landscapes *Phys. D: Nonlinear Phenom.* **107** 151–5
- [4] Parisi G and Virasoro M A 1989 On a mechanism for explicit replica symmetry breaking *J. Phys.* **50** 3317–29
- [5] Franz S, Parisi G and Virasoro M A 1993 Free-energy cost for ultrametricity violations in spin glasses *Europhys. Lett.* **22** 405–11
- [6] Kurchan J, Parisi G and Virasoro M A 1993 Barriers and metastable states as saddle points in the replica approach *J. Phys. I, France* **3** 1819–38
- [7] Franz S and Parisi G 1995 Recipes for metastable states in spin glasses *J. Phys. I, France* **5** 1401
- [8] Monasson R 1995 Structural glass transition and the entropy of the metastable states *Phys. Rev. Lett.* **75** 1–2848
- [9] Cavagna A, Giardina I and Parisi G 1997 Structure of metastable states in spin glasses by means of a three replica potential *J. Phys. A: Math. Gen.* **30** 4449–66
- [10] Cavagna A, Giardina I and Parisi G 1997 An investigation of the hidden structure of states in a mean-field spin-glass model *J. Phys. A: Math. Gen.* **30** 7021
- [11] Derrida B 1985 A generalization of the random energy model which includes correlations between energies *J. Physique Lett.* **46** L-401
- [12] Crisanti A and Sommers H-J 1992 The spherical p -spin interaction spin glass model: the statics *Z. Phys. B: Condens. Matter* **87** 341–54
- [13] Crisanti A and Sommers H-J 1995 Thouless–Anderson–Palmer approach to the spherical p -spin spin glass model *J. Phys. I, France* **5** 805–13