The mean field theory of spin glasses: the heuristic replica approach and recent rigorous results

Giorgio Parisi
Dipartimento di Fisica, Sezione INFN, SMC of INFM-CNR,
Università di Roma “La Sapienza”,
Piazzale Aldo Moro 2, 00185 Rome, Italy

December 7, 2009

Abstract

The mathematically correct computation of the spin glasses free energy in the infinite range limit crowns 25 years of mathematic efforts in solving this model. The exact solution of the model was found many years ago by using a heuristic approach; the results coming from the heuristic approach were crucial in deriving the mathematical results. The mathematical tools used in the rigorous approach are quite different from those of the heuristic approach. In this note we will review the heuristic approach to spin glasses in the light of the rigorous results; we will also discuss some conjectures that may be useful to derive the solution of the model in an alternative way.

Keywords

11E95 p-adic theory
46N30 Applications in probability theory and statistics
60F10 Large deviations
82B44 Disordered systems (random Ising models, random Schrödinger operators, etc.)

1 Spin glasses

Spin glasses are interesting for many physical reasons; moreover they are the prototype of a complex system, i.e. a system with many phases. Here we will consider only spin glasses with an infinite range interaction, where the mean field approximation is correct (more of less by definition). In this case \cite{1,2,3,4,5} (i.e. the Sherrington Kirkpatrick model), the Hamiltonian is given by

\[ H_J[\sigma] = \frac{1}{2} \sum_{i,k=1,N} J_{i,k} \sigma_i \sigma_k, \]  \hspace{1cm} (1)

where the \( N \) variables \( \sigma \) are Ising spins, i.e. \( \sigma_i = \pm 1 \). The \( J \) are random variables: they are Gaussian distributed with zero average:

\[ E[J_{i,k}] = 0, \quad E[J_{i,k}^2] = N^{-1}, \]  \hspace{1cm} (2)

where \( E[\cdot] \) denotes the expectation value respect to the random couplings \( J \).

For each instance of the system (i.e. a choice of the couplings \( J \)) we can compute the ground state energy density that is defined as

\[ e^*_N = \frac{\min_{\sigma} H_J[\sigma]}{N}. \]  \hspace{1cm} (3)
The normalization factors have been chosen in such a way that the energy density is a number of order 1, when $N \to \infty$. In the same way for each choice of the variables $J$ we can define the free energy density $f_J^N(\beta)$ and the partition function $Z_J^N(\beta)$ as

$$f_J^N(\beta) = -\frac{\log(Z_J^N(\beta))}{\beta N}, \quad Z_J^N(\beta) = \sum_{\sigma} \exp(-\beta H_J[\sigma]). \quad (4)$$

When $N \to \infty$, it can be proved that the fluctuations in the energy density and in the free energy density go to zero with $N$:

$$E[(e_J^N)^2] - (E[e_J^N])^2 \to 0, \quad E[(f_J^N)^2] - (E[f_J^N])^2 \to 0. \quad (5)$$

Therefore, with probability 1 when $N$ goes to infinity, all systems have the same free energy density. We want to compute

$$\tilde{f}(\beta) = \lim_{N \to \infty} E[f_J^N(\beta)], \quad \tilde{e} = \lim_{N \to \infty} E[e_J^N]. \quad (6)$$

In the limit $N \to \infty$ the probability distribution of $f_J(\beta)$ becomes a delta function and the average value becomes equal to the most likely value.

The limit $N \to \infty$ and $\beta \to \infty$ can be exchanged and therefore we have that

$$\tilde{e} = \lim_{\beta \to \infty} \tilde{f}(\beta). \quad (7)$$

The mean values are not the only interesting quantities. We would like to control the whole probability distribution of the free energy ($f$) at fixed and large $N$ (i.e. $P_N(f)$). We can consider various problems:

- The behaviour of the typical fluctuations around the average value. For example it has been conjectured that in the region where the probability is concentrated we have \[6, 7\]:

$$P_N(f) \approx N^{-5/6} p \left( N^{6/5} (f - \tilde{f}) \right), \quad (8)$$

where $p$ is an unknown function and the difference between the r.h.s and the l.h.s. is negligible for large $N$ (unfortunately we do not know how to compute heuristically the function $p$).

- Large deviations. For values of the free energy that is smaller than the most likely value (i.e. $f < \tilde{f}$), the probability is exponential small in $N$ \[8, 9\]; the following limit exists (for not too large $f$):

$$L(f) = \lim_{N \to \infty} -\frac{\log(P_N(f))}{N}. \quad (9)$$

- Very large deviations. For values of the free energy that is bigger than the most likely value (i.e. $f > \tilde{f}$), it has been conjectured \[8, 9\] that the probability decrease faster than any exponential of $N$ and that it is exponential small in $N^2$ \[7\]. The following limit should exist (for not too large $f$) and it should be non-trivial:

$$\mathcal{L}(f) = \lim_{N \to \infty} -\frac{\log(P_N(f))}{N^2}. \quad (10)$$

Similar questions are well posed also for the ground state energy $e$.

As we shall see below, there are rigorous explicit expressions for $\tilde{f}$ \[9\] and for the large deviations function $L(f)$ \[10\]. On the contrary for the very large deviations function $\mathcal{L}(f)$, there are only heuristic evaluations \[7\]. Nothing analytic is know (only numerical evaluations) on the function $p$ that enters in the scaling form of the probability $P_N(f)$ (eq. \[8\]). The form in eq. \[8\] has been conjectured in order to match with the known behaviour of the large deviation function, i.e.

$$L(f) \propto (\tilde{f} - f)^{6/5} \quad (11)$$

\[1\] As suggested by a referee, the exchange of limits is permitted by the elementary inequality $|E[f_J^N(\beta)] - E[e_J^N]| \leq \ln(2)/\beta$.

\[2\] From time to time, when this does not lead to ambiguities we will not indicate in an explicit way the $\beta$ dependence of the various quantity: the correct, but heavier, notation would be $P_N(f, \beta)$.
for $f$ near (and smaller than) $\tilde{f}$ and with the supposed behavior of the very large deviation function, i.e.

$$\mathcal{L}(f) \propto (\tilde{f} - f)^{12/5}. \quad (12)$$

The reader should notice that the same kind of questions have a known answer for the ground state energy in the linear case where the variable $\sigma$ are continuous variables:

$$e_J^N = \min_{\sigma} \frac{H_J[\sigma]}{N}, \quad \sum_{i=1}^{N} \sigma_i^2 = N. \quad (13)$$

Indeed in this case the ground state coincides with the largest eigenvalue of a Gaussian random matrix where nearly everything is known \[11\] \[12\]. Here we are interested in the more difficult case,

$$\sigma_i = \pm 1, \quad (14)$$

where $e_J^N$ is not a function of only the eigenvalues of $J$.

### 2 Some heuristic considerations

Let us start with a simple theorem \[13\] that gives interesting information on what we should not assume. For a given system we introduce the connected correlation functions at temperature $\beta$

$$C_{i,k} = \langle \sigma_i \sigma_k \rangle_{Gibbs} - \langle \sigma_i \rangle_{Gibbs} \langle \sigma_k \rangle_{Gibbs}, \quad (15)$$

where the bracket denotes the average with respect to the Gibbs distribution\[3\].

$$\langle A[\sigma] \rangle_{Gibbs} = \lim_{N \to \infty} \sum_{\sigma} P[\sigma] A[\sigma], \quad P[\sigma] = \frac{\exp(-\beta H_J[\sigma])}{\sum_{\sigma} \exp(-\beta H_J[\sigma])}. \quad (16)$$

It would be natural to assume that the connected correlations go to zero when $N \to \infty$, more precisely to assume that

$$\lim_{N \to \infty} \sum_{i,k} \frac{C_{i,k}^2}{N^2} = 0. \quad (17)$$

If the previous assumption\[4\] would be valid, we could derive a simple formula for $f(\beta)$ that implies that

$$e = \lim_{\beta \to \infty} f(\beta) = -\sqrt{\frac{2}{\pi}}. \quad (18)$$

The results is reasonable, but in definite disagreement with numerical estimates for $N$ up to $O(10^3)$. However the situation is much worse: the same argument would imply that the entropy $s(\beta)$ satisfies the relation

$$\lim_{\beta \to \infty} s(\beta) = -\frac{1}{(2\pi)} < 0. \quad (19)$$

The entropy $s(\beta)$ is non-negative by definition. We can thus conclude by absurdum that

$$\lim_{N \to \infty} \sum_{i,k} \frac{C_{i,k}^2}{N^2} \neq 0. \quad (20)$$

In order to understand the importance of the result (20), it is convenient to recall some know facts on clustering states \[14\] \[15\] \[16\]. Let us consider an infinite translational invariant (non-random) system. In this

---

\[3\] In order to lighten the notation I have not indicated the dependence of both the Gibbs expectation values ($\langle \cdot \rangle$) and of the $P[\sigma]$ on $\beta$ and $J$.

\[4\] The previous assumption should be formulated more carefully in order to take into account the invariance of the Gibbs measure when we change sign to all the spin simultaneously; I will not discuss here this technical detail.
case one can introduce the general concept of an equilibrium state, i.e. a state that satisfies local equilibrium conditions (e.g. the DLR equations [14]). If the Gibbs probability distribution has a limit when the volume becomes infinite, the Gibbs probability distribution is an example of equilibrium state, but it may be not the unique equilibrium state.

Now the set of equilibrium states is convex; by the Alaoglu theorem any state can be written as the linear sum of the extremal (pure) states. Therefore we can write

$$\langle \cdot \rangle_{\text{Gibbs}} = \sum_\alpha w_\alpha \langle \cdot \rangle_\alpha, \quad \sum_\alpha w_\alpha = 1,$$

where the sum runs over the pure states (for simplicity I am assuming that the relevant set of pure states is discrete). A well-known theorem states that in a pure state connected correlation functions vanish at infinity (at large distances), i.e. pure states are clustering states, and vice versa. Intensive quantities (i.e. average on the whole system of local quantities) do not fluctuate in pure states.

The ferromagnetic Ising at zero field at low temperature provides a well-known case: there are two phases (+ and -):

$$\langle \sigma_i \rangle = \pm m, \quad \langle \sigma_i \sigma_k \rangle \approx m^2, \quad w_+ = w_- = \frac{1}{2}, \quad \langle \cdot \rangle_{\text{Gibbs}} = w_+ \langle \cdot \rangle_+ + w_- \langle \cdot \rangle_-, \quad (22)$$

where, for $i$ and $k$ far away,

$$\langle \sigma_i \sigma_k \rangle_{\text{Gibbs}} \approx \langle \sigma_i \rangle_{\text{Gibbs}} \langle \sigma_k \rangle_{\text{Gibbs}} = m^2; \quad \langle \sigma_i \sigma_k \rangle_+ - \langle \sigma_i \rangle_+ \langle \sigma_k \rangle_+ \approx 0. \quad (23)$$

The situation is more complex for disordered systems, where the Gibbs measure may not have a limit when the volume goes to infinity. Although the appropriate and sophisticated tools have been forged to deal with this situation, it may be convenient to consider what happens in large, but finite systems, without putting the $\delta$’s and $\epsilon$’s needed to make the statements sharp.

Let us consider a system where the number of degrees of freedom ($N$) is large, but finite. We suppose [17, 18] that for large $N$ we can write

$$\langle \cdot \rangle_{\text{Gibbs}} \approx \sum_\alpha w_\alpha \langle \cdot \rangle_\alpha, \quad \sum_\alpha w_\alpha \approx 1. \quad (24)$$

The states are approximately clustering if

$$\frac{\sum_{i,k} (C_{i,k})^2}{N^2} = o(1). \quad (25)$$

The previous theorem (see eq. (20)) tells us that for spin glasses more than one $w$ should be different from zero and therefore the decomposition in eq. (24) is non-trivial. In order to describe better the Gibbs decomposition is convenient to introduce additional concepts.

The first interesting quantity is the probability distribution of the $w$’s (i.e. $P(\{w\})$); however the weights $w$’s of the states do not convey the whole information. We would like to know how the states differ one from the others: at this end we can introduce the distance and the overlap between two states:

$$d_{\alpha,\gamma} = \frac{\sum_{i=1,N} \left( \langle \sigma(i) \rangle_\alpha - \langle \sigma(i) \rangle_\gamma \right)^2}{N}; \quad q_{\alpha,\gamma} = \frac{\sum_{i=1,N} \langle \sigma(i) \rangle_\alpha \langle \sigma(i) \rangle_\gamma}{N}. \quad (26)$$

Obviously the distance $d$ and the overlap $q$ are related. We have

$$d_{\alpha,\gamma} = 2(q_{EA} - q_{\alpha,\gamma}), \quad (27)$$

5As a physicist I would also remark that all the experimental systems are finite, and a statement on the behaviour of large finite systems can be experimentally tested much better than a statement on the behaviour of an infinite systems.

6More precisely more than two $w$’s.
where \( q_{EA} \equiv q_{\alpha,\alpha} = q_{\gamma,\gamma} \) (we are assuming that \( q_{\alpha,\alpha} \) is independent form \( \alpha \)). Usually in literature one studies the properties of the overlap, the distances have been introduced here because their definition is more natural.

Now for a given system \( J \), we introduce the descriptor \( D_J \), that is given by the following set

\[
\{w_\alpha, q_{\alpha,\gamma}\},
\]

where the index \( \alpha \) runs over a numerable set\(^7\).

The descriptor \( D_J \) does depend on \( J \). The functional \( \mathcal{P}(D) \) is the probability distribution of the descriptors. One can construct heuristic arguments \(^1\) to compute the probability distribution \( \mathcal{P}(D) \), the free energy \( \hat{f} \) and the large deviation function \( L(f) \). We will describe these results later on in section (4). We now proceed into the description of some rigorous results.

### 3 Rigorous results

In this section we will concentrate our attention on some theorems. They are obtained using two different approaches, the first is based on probabilistic techniques, the second heavily relies on Guerra interpolation technique.

Without any reference to the physical meaning of the descriptor and to the Gibbs phase decomposition we can introduce the descriptor \( D \) that is defined as a normalized vector and a matrix, satisfying certain inequalities \(^{19}\). Neglecting technical details, a descriptor \( (D) \) is given by the set of \( w_\alpha \) and of \( q_{\alpha,\gamma} \), where \( \alpha \) belong to a numerable set and \( \sum_\alpha w_\alpha = 1 \). In this context one would naturally associate to a descriptor a decomposition of the Gibbs measure into states, however for the sake of proving theorems one can define a descriptor in an abstract way. Moreover we can also introduce a probability distribution on the space of the descriptors \( (\mathcal{P}(D)) \). The following theorems hold.

- One can define a free energy functional \( \mathcal{F}[\mathcal{P}] \) \( (\mathcal{P}(D) \text{ being the probability of descriptors}) \) such \(^{19}\) that the true free energy is given by

\[
\hat{f}(\beta) = \max_{\mathcal{P}} \mathcal{F}[\mathcal{P}].
\]

The explicit form of the free energy functional \( \mathcal{F}[\mathcal{P}] \) is rather complex \(^{19}\) and it will be not described here.

- Using arguments, which do not refer to descriptors and are based on Guerra interpolation, it is possible to compute in an explicit way the value of \( \hat{f} \) and consequently (from the previous theorem) \( \max_{\mathcal{P}} \mathcal{F}[\mathcal{P}] \). By inspection one verifies that the maximum is reached for the probability distribution predicted by the heuristic approach \( (\mathcal{P}^*) \): also the value of the free energy \( f(\beta) \) coincides with the one computed by the heuristic approach \(^{20,21}\). Therefore we know a probability distribution that maximize \( \mathcal{F}[\mathcal{P}] \).

- Under some assumptions, i.e. if stochastic stability and ultrametricity (see the definitions later) hold \(^{22}\), the function

\[
P(q) = \sum_{\alpha,\gamma} w_\alpha w_\gamma \delta(q - q_{\alpha,\gamma})
\]

determines the probability distribution \( \mathcal{P}(D) \). Therefore in the stochastic stable ultrametric case the probability distribution is essentially unique, apart from a redefinition of the overlap.

The bad news is that we do not know if the true \( \mathcal{P}(D) \) has some relation with the \( \mathcal{P}^*(D) \) that maximizes the free energy; for the moment the probability distribution \( \mathcal{P}(D) \) does not have a definite meaning. It also not clear if the maximum of the free energy functional is unique, also if we consider only those probability that are stochastically stable\(^8\).

\(^7\)The numbers of states is obviously finite for finite \( N \), however in order to consider the limit \( N \to \infty \) it is useful to consider a numerable set of states.

\(^8\)The reader can find the definition of stochastic stability and a discussion on its consequences in \([23, 24, 25, 26]\).
Figure 1: A taxonomic tree: the leaves denote the states; they are at the bottom. The distance among the states corresponds to the minimum height we have to cross for going from a state to an other state. This definition of the distance is ultrametric.

It may be interesting to note the maximum is reached for the probability an ultrametric $P^*$: i.e. the probability matrix is concentrated on distances that satisfy the ultrametric inequalities:

$$d_{\alpha,\gamma} \leq \max(d_{\alpha,\beta}, d_{\beta,\gamma}) \forall \beta.$$  \hspace{1cm} (31)

Ultrametricity implies that the probability distribution of the distance among three random configurations is supported, in the limit of very large systems, only on equilateral and isosceles triangles with no contributions coming from scalene triangles. In an ultrametric space if two spheres have one point in common, their union coincides with the sphere having maximum radius: in other words a random walk with steps of length $l$ would never arrive at a distance larger than $l$ from the origin.

Ultrametricity is a very striking property: it implies that the equilibrium configurations of a large system can be classified in a taxonomic (hierarchical) way (as animals in different taxa). When ultrametricity holds, the space of the descriptors reduces to the space of infinite weighted trees (each leave has a weight). Ruelle gives a precise mathematical definition of a probability distribution over these trees \cite{27, 28, 29} (in his paper one can also find interesting comments on the pruning of infinite trees.)

3.1 Some details on the solution

In this section we present the explicit form for the free energy $\tilde{f}(\beta)$. We firstly introduce an non-decreasing function $q(x)$ that is defined in the interval $[0 - 1]$. In the replica approach the inverse function $q'[x]$ is related to the function $P(q)$ (defined in eq. (30)) by the equation:

$$x(q) = \int_0^q d q' P(q').$$  \hspace{1cm} (32)

Let us consider a function $g(x, h)$, that is defined in the strip $0 \leq x \leq 1$. The function $g$ satisfies the following boundary condition at $x = 1$:

$$g(1, h) = \log(\cosh(\beta h)).$$  \hspace{1cm} (33)

Moreover (if we assume for simplicity that $q(x)$ is differentiable), the function $g(x, h)$ must be the unique solution of the following antiparabolic equation

$$\frac{\partial g}{\partial x} = -\frac{d q}{d x} \left( \frac{\partial^2 g}{\partial h^2} + x \left( \frac{\partial g}{\partial h} \right)^2 \right).$$  \hspace{1cm} (34)
We now introduce a functional $F[q]$:

$$F[q] = \beta \int_{0}^{1} dx \left(1 - q(x)^2\right) - g(0,0) .$$

(35)

Talagrand theorem tell us that the free energy $F$ is given by

$$F = \max_{q(x)} F[q] .$$

(36)

Moreover, using Ruelle construction [27], we can associate to each function $q(x)$ a probability distribution over the descriptors, i.e. $P(D)$. For the probability distribution over the descriptors $P(D)$ that corresponds to a function $q(x)$, the functional $F[P]$ (introduced in [19]) gives the same result of eq. (35).

These results extend to large deviations. Indeed a theorem of Talagrand states that for $1 > n > 0$ we can define in the large $N$ limit a large deviation function $\hat{f}(n)$:

$$\int dP_N(f) \exp(-N n \beta f) \approx \exp(-N \beta \hat{f}(n)) .$$

(37)

The functions $\hat{f}(n)$ and $L(f)$ are simple related by a Legendre transform at usual. Talagrand has proved that the explicit value of $\hat{f}(n)$ is given by

$$\hat{f}(n) = \max_{q(x)} \left[ \beta \int_{n}^{1} dx \left(1 - q(x)^2\right) - g(n,0) \right] \text{ for } n < 1 .$$

(38)

It is evident that for $n = 0$ we have that $\hat{f}(0) = \tilde{f}$, as it should be.

It is possible to perform analytic and numerical evaluations of the function $\hat{f}(n)$ following [6, 10], where the properties of the function $\hat{f}(n)$ are discussed in details.

When $n < 0$, there are no theorems. There is a conjecture [8]:

$$\lim_{N \to \infty} \frac{\log \left( \int dP_N(f) \exp(-N n \beta f) \right)}{N n \beta} = \hat{f}(0) \equiv \tilde{f} .$$

(39)

In other words the probability $P_N(f)$ should be smaller than an exponential of $N$ if $f > \tilde{f}$.

What about very large deviations? In [7] it has been conjectured that

$$\int dP_N(f) \exp \left(-N^2 n \beta (f - \tilde{f})\right) \approx \exp \left(-N^2 \beta \hat{f}_{LL}(n)\right) .$$

(40)

Moreover some heuristic computations of the function $\hat{f}_{LL}(n)$ have been performed and one arrives the following formula:

$$P_N(f) \approx \exp \left(-C N^2 |f - \tilde{f}|^{12/5} + \ldots\right) ,$$

(41)

where the value of $C$ is approximately evaluated [7] and it should be near to $2/3$.

4 \hspace{1cm} The algebraic replica approach

In the original heuristic approach the computations were done using the algebraic replica method [11]. Later it was found that the replica method was equivalent to a probabilistic approach [11] that inspired the rigorous results [20, 19, 21]. However it may worthwhile to try to understand the basis of the replica method in order to see if we can give it a rigorous basis. This is not easy because the algebraic replica approach is based on very strange mathematics.

The mean field theory was solved by replica approach. The basic idea is quite simple and essentially it goes back to Nicola d’Oresme. Our goal is to compute

$$\tilde{f}(\beta) \equiv f_{\infty}(\beta) = \lim_{N \to \infty} f_N(\beta) , \hspace{0.5cm} f_N(\beta) = -\frac{E \ln \left( Z_N^N(\beta) \right)}{\beta N} .$$

(42)
However in our case it is much simpler to compute (for integer $n$)

$$f_N(n, \beta) = -\frac{\log \left( E \left[ (Z_N^\beta(\beta))^n \right] \right)}{\beta N n}, \quad (43)$$

and at the end performing the limit

$$f_\infty(n, \beta) = \lim_{N \to \infty} f_N(n, \beta). \quad (44)$$

Although the computations are simple for integer $n$ the construction can be done also for non-integer $n$ and we have that

$$\lim_{n \to 0} f(n, \beta) = f_\infty(\beta). \quad (45)$$

The function $f_\infty(n, \beta)$ is interesting because it is the one relevant in the large deviations regime.

A simple computations can be done for integer $n$, while we need the function $f_\infty(n, \beta)$ also for non-integer $n$. While $f_N(n, \beta)$ is an analytic function $n$, at low temperatures (high $\beta$) $f_\infty(n, \beta)$ is not an analytic function of $n$ so that we cannot obtain its value at $n = 0$ by analytic continuation in $n$ from integer $n$.

In order to bypass this problem physicists were lead to an unusual mathematics: as we shall see below there are many ways in which a $0 \times 0$ matrix may be constructed. At the end of the day in the replica approach the matrix $Q$ is parameterized in terms of a function $F(\beta)$ that is analytic in $\beta$ and one gets.

$$\exp(-\beta n N f_N(n, \beta)) = \log \left( E \left[ (Z_N^\beta(\beta))^n \right] \right) = \int dQ \exp(-\beta N n F_\beta(Q)), \quad (46)$$

where $Q$ are $n \times n$ symmetric matrices (that are zero on the diagonal) and the integral is done over all these matrices (the function $F(\beta)$ has a simple expression).

It is evident that for positive integer $n$ the integrals in equation (46) can be evaluated by the saddle point method and one gets.

$$f_\infty(n, \beta) = \min_{Q} F_\beta(Q) = F_\beta(Q^*). \quad (47)$$

Unfortunately, at high $\beta$ the function $f_\infty(n)$ that is defined for all $n$, is non-analytic in $n$. So in order to evaluate it in the interesting region and to find out the value of $Q^*$ one has to consider integral representation for non-integer $n$. In this way one is lead to consider the saddle point over $0 \times 0$ matrices.

There are many ways in which a $0 \times 0$ matrix may be constructed. At the end of the day in the replica approach the matrix $Q$ is parameterized in terms of a function $q(x)$ defined on the interval $0 - 1$, so that the space of $0 \times 0$ matrices becomes infinite dimensional. The game consists in computing everything by analytic continuation of the integral representation.

There are many way equivalent to do the construction and they lead to the same result. We choose one that is particularly simple [30][31].

Let us take a prime number $p$ and let us assume that there is an integer $m$ such that $p^m = n$. We look for saddle point of $F_\beta(Q)$, where the matrix $Q$ has the form:

$$Q_{a,b} = q \left( |a-b|_p^{-1} \right). \quad (48)$$

Here $|a|_p$ is the $p$-adic norm: (the $p$-adic norm is defined as follows: if $p^k$ divides $a$ and $p^{k+1}$ does not divides $a$, $|a|_p = p^{-k}$).

We write $F_\beta(Q)$ as function of $p$ and $m$ we make an analytic continuation in $p$ and $m$ of the result. Everything is computed by making analytic continuations from the region where the construction make sense. We will also write down the equations for the critical points in an explicit way for the function $q(x)$. The final recipe consists in assuming that the free energy is given by the analytic continuation of the free energy that is evaluated at the analytic continuation of the saddle point. There are no serious mathematical justifications for this procedure. The heuristic justifications are:

- At the present moment in the framework of the replica approach there is nothing else that you can try. You have two choices: or you accept this approach, or you change framework (or you invent something new).
• It works, it gives results that have sense and that have been confirmed by a rigorous mathematical analysis.

• It is quite likely that the rigorous results would have not been derived without the heuristic analysis of the replica method.

Now we can proceed to the limit $n \to 0$: at this point nothing is anymore an integer. Let us suppose that the prime number $p$ satisfies the condition $0 < p < 1$; this is quite possible because $p$ is not anymore a prime number, but it is the analytic continuation of a prime number. Now if we send $m \to \infty$ the quantity $n$ goes to zero, as it should be ($p$ is less than 1).

As a final step we perform the limit $p \to 1^{-}$: the range of $|a - b|_p^{-1}$ becomes the interval $[0 - 1]$ and the matrix $Q$ depends on the function $q(x)$ defined on the interval $[0 - 1]$. In this way the function $F[Q]$ is promoted to a functional of $q(x)$ and one finds the previous formulae for $F[q]$. The replica approach gives the exact result for the free energy and the functions $q(x)$ is the same of the one of the mathematical approach. It is remarkable that integers are an ultrametric space with respect to the $p$-adic distance and this ultrametricity property implies the ultrametricity of the distances in the descriptor.

Later one it was a surprise to discover was that all the formulas of the replica approach can be translated into probabilistic statements [1]. The matrix $Q$ of the replica approach is a very compact to code the probability $P(D)$. It looks strange that a matrix codes a probability distribution, so that let us explain how it works.

One can prove the following formula

$$E [\delta(Q_{a,b} - q)] = E \left[ \sum_{\alpha,\beta} w_{\alpha} w_{\beta} \delta(q_{\alpha,\beta} - q) \right],$$

where the $E$ at the l.h.s. denotes the average over the indices $a$ and $b$ (with $a \neq b$) and $E$ at the r.h.s. denotes the average over the probability distribution of the descriptors. In a similar way one can prove

$$E [\delta(Q_{a,b} - q_{1,2}) \delta(Q_{b,c} - q_{2,3}) \delta(Q_{c,a} - q_{3,1})] =
E \left[ \sum_{\alpha,\beta,\gamma} w_{\alpha} w_{\beta} w_{\gamma} \delta(q_{\alpha,\beta} - q_{1,2}) \delta(q_{\beta,\gamma} - q_{2,3}) \delta(q_{\gamma,\alpha} - q_{3,1}) \right].$$

Going on one can write an infinite set of relations that should uniquely determine the probability distribution over the descriptors. It is quite remarkable that a $0 \times 0$ matrix is used to code the probability distribution over a space of infinite dimensional matrices and that the ultrametricity of the integers with respect to the $p$-adic norm implies the ultrametricity of the matrix of distances among states.

The reader should note that not all $Q$ matrices codes for a probability: the probability must a positive function and there are many matrices $Q$ where the l.h.s. of eq. (49) cannot be interpreted as a probability. It is also not clear if there are probabilities $P(D)$ that cannot be coded using the matrix $Q$: a negative result that is difficult to prove and there is no constructive procedure to get the matrix $Q$ (if any) from the probability distribution of the descriptors.

In conclusions the heuristic version of the descriptor approach (i.e. the cavity method) and the replica approach are equivalent [17]. The functions $q(x)$ of both approaches are equal. The replica approach gives the exact result for the free energy.

The reader may be perplexed by this procedure that was invented about thirty years ago. However it has a very strong heuristic value and both the heuristic probabilistic approach and the rigorous results have strongly beneficed by the fact that the answer to many question was known especially in this context that is rather unfamiliar.

### 5 Two conjectures

Can the original replica derivation made rigorous? The distance from conventional mathematics seems to be very far so that a direct assault may be not successful. However it is clear that the subject of our discussions
is the set of critical points of the function $F[Q]$ for all $n$, in the same way as the analytic continuation of a function on the integers is related to the value that the function takes on all integers.

Following [32] we would like to present some conjectures, whose proof would be a step forward a rigorous understanding of the replica approach.

Let us consider a function $Z(n, N)$ that can be simultaneously written as

$$Z(n, N) = \int_0^{+\infty} dz \mu_N(z) z^n = \int dQ \exp(-N F(Q)),$$

(51)

where $\mu(z)$ is a positive function and the integral runs over symmetric $n \times n$ matrices; $F(Q)$ is an analytic function of $n$, in the sense that it has a fixed form for all $n$. The existence of functions, which can be written in both ways, is non-trivial, however the partitions function of spin glasses and generalized spin glasses provide an explicit example.

We know define a new function $\tilde{Z}(n, N)$, i.e.

$$\tilde{Z}(n, N) = \sum_k \exp(-N F(Q_k))$$

(52)

and the sum goes over all the critical points of the function $F$, i.e. those satisfying the condition

$$\left. \frac{\partial F}{\partial Q} \right|_{Q_k} = 0.$$

(53)

An alternative definition could be $\tilde{Z}(n, N) = \sum_k s_k \exp(-N F(Q_k))$, where the factors $s_k$ are given by $s_k = \text{sign} \left( \det \left( \frac{\partial^2 F}{\partial Q^2} \right) \right)_{Q_k}$. It not clear which of the two definitions is the most appropriate.

Our aim would by to approximate $Z(n, N)$ with $\tilde{Z}(n, N)$ also for non-integer $n$. At this end we need to have some $a$ priori control on the analytic properties of both functions. In this context we put forward the following conjectures.

- The function $\tilde{Z}(n, N)$ satisfies for integer $n$ the following integral representation.

$$\tilde{Z}(n, N) = \int_0^{+\infty} dz \tilde{\mu}_N(z) z^n.$$

(54)

The previous formula implies the existence of a natural (and under certain conditions unique) analytic continuation of $\tilde{Z}(n, N)$.

- The function $\tilde{Z}(n, N)$ is a good and uniform approximation to $Z(n, N)$ for large $N$: we should have:

$$\lim_{N \to \infty} \frac{\log \left( \tilde{Z}(n, N) \right)}{N} = \lim_{N \to \infty} \frac{\log \left( Z(n, N) \right)}{N}.$$

(55)

The previous formula should not only for integer positive $n$, where it is trivial, but also for non-integer $n$ and in particular for $n$ in the interval $[0 - 1]$. In the best of possible words one should have:

$$\lim_{N \to \infty} \frac{\log \left( \tilde{\mu}_N(x^N) \right)}{N} = \lim_{N \to \infty} \frac{\log \left( \mu_N(x^N) \right)}{N}.$$

(56)

These conjectures are interesting also because not many results are known in this direction (as far as I can tell). However in the contest of this paper they are very relevant because in some cases these conjectures have been used to prove the result of coming from the replica approach [32].

---

9Examples of allowed functions are $\text{Tr}(Q^4)$ or $\sum_{a,b} Q_{a,b}^4/n$, the factor $1/n$ is needed in order to have a non zero limit at $n = 0$ in interesting cases.
References

[1] M. Mézard, G. Parisi and M. A. Virasoro, *Spin glass theory and beyond*, World Scientific (Singapore 1987).

[2] G. Parisi in *Les Houches Summer School - Session LXXVII: Slow relaxation and non equilibrium dynamics in condensed matter*, ed. by J.-L. Barrat, M. V. Feigelman, J. Kurchan, and J. Dalibard, Elsevier 2003.

[3] T. Castellani and A. Cavagna, *Spin glasses for pedestrian*, cond-mat/0505032

[4] G. Parisi, in *Les Houches Summer School - Session LXXXV: Complex Systems*, ed. by P. Bouchaud, M. Mézard, J. Dalibard, Elsevier 2007.

[5] G. Parisi, *Field Theory, Disorder and Simulations*, World Scientific, (Singapore 1992).

[6] I. Kondor, *Parisi’s mean-field solution for spin glasses as an analytic continuation in the replica number*, J. Phys. A 16, L127 (1983).

[7] G. Parisi and T. Rizzo *Universality and Deviations in Disordered Systems*, cond-mat 0901.1100 (2009).

[8] V Dotsenko, S Franz and M Mézard, *Partial annealing and overfrustration in disordered systems* J. Phys. A 27, 2351 (1994).

[9] M. Talagrand *Large Deviations, Guerras and ASS Schemes, and the Parisi Hypothesis*, J. Stat. Phys. 126, 837 (2007).

[10] G. Parisi and T. Rizzo, *Large deviations in the free energy of mean-field spin glasses*, Phys. Rev. Lett. 101, 117205 (2008).

[11] C.A. Tracy and H. Widom *Level-spacing distributions and the Airy kernel*, Commun. Math. Phys. 159, 151 (1994); *On orthogonal and symplectic matrix ensembles*, ibid. 177, 727 (1996).

[12] D.S. Dean and S.N. Majumdar *Large Deviations of Extreme Eigenvalues of Random Matrices*, Phys. Rev. Lett. 97, 160201 (2006); *Extreme value statistics of eigenvalues of Gaussian random matrices*, Phys. Rev. E, 77, 041108 (2008).

[13] L. Pastur and M. Shcherbina, *The absence of self-averaging of the order parameter in the Sherrington-Kirkpatrick model*, J. Stat. Phys 62 1 (1992).

[14] R. Haag and D. Kastler *An Algebraic Approach to Quantum Field Theory*, J. Math. Phys. 5, 848 (1964).

[15] D. Kastler and D. W. Robinson, *Invariant states in statistical mechanics*, Comm. Math. Phys. 3, 151 (1966).

[16] D. Ruelle, *Statistical Mechanics* (Benjamin, Reading 1969).

[17] G. Parisi, *Facing complexity*, Physica Scripta 35, 123 (1987).

[18] E. Marinari, G. Parisi, F. Ricci-Tersenghi, J. Ruiz-Lorenzo and F. Zuliani, *Replica symmetry breaking in short-range spin glasses: Theoretical foundations and numerical evidences*, J. Stat. Phys. 98, 973 (2000).

[19] M. Aizenman, R. Sims and S. L. Starr, *Extended variational principle for the Sherrington-Kirkpatrick spin-glass model* (2003) Phys. Rev. B 68, 214403.

[20] F. Guerra, *Broken Replica Symmetry Bounds in the Mean Field Spin Glass Model*, Comm. Math. Phys. 233, 1 (2002).

[21] Talagrand M., *The Parisi formula*, Ann. of Math., 163 221 (2006).
[22] Iniguez D., Parisi G. and Ruiz-Lorenzo J.J. Simulation of three-dimensional Ising spin glass model using three replicas: Study of Binder cumulants, J. Phys. A29 4337 (1996).

[23] F. Guerra, About the overlap distribution in mean field spin-glass model, Int. J. Phys. B, 10, 1675 (1997).

[24] S. Ghirlanda and F. Guerra, General properties of overlap probability distributions in disordered spin systems. Toward Parisi ultrametricity, J. Phys. A: Math. Gen. 31 9149 (1998).

[25] M. Aizenman and P. Contucci, On the stability of the quenched state in mean field spin glass models, cond-mat 9712129.

[26] G. Parisi, On the probabilistic formulation of the replica approach to spin glasses, cond-mat/9801081.

[27] D. Ruelle A mathematical reformulation of Derridas REM and GREM, Commun. Math. Phys. 48, 351 (1988).

[28] J.P. Bouchaud and D.S. Dean Aging on Parisi’s tree, J. Phys. I France 5, 265 (1995).

[29] Y. V. Fyodorov and J.-P. Bouchaud, Statistical mechanics of a single particle in a multiscale random potential: Parisi landscapes in finite dimensional Euclidean spaces, J. Phys.A 41 324009 (2008).

[30] G. Parisi and N. Sourlas P-adic numbers and replica symmetry breaking, Europ. Phys. J. B 14, 535 (2000).

[31] A.Yu. Khrennikova and S.V. Kozyrev, Replica symmetry breaking related to a general ultrametric space II: RSB solutions and the n → 0 limit, Physica A 359, 241 (2006).

[32] M. Campellone, G.Parisi and M. Virasoro Replica method and finite volume corrections, arXiv:0911.1313.