Huge Random Structures and Mean Field Models for Spin Glasses

Michel Talagrand

Abstract. To explain (at least qualitatively) the unconventional magnetic behavior of certain materials, the physicists have been led to formulate and to study simple mathematical models. The concepts and methods they developed in this process appear to apply to a number of important random combinatorial optimization problems, for which they have proposed remarkable formulas. Their discoveries point towards a new branch of probability theory. Finding rigorous arguments to support their conjectures is a formidable challenge and a long range program, some steps of which are described in the present paper.

Introduction

The research presented in this paper has largely been influenced by the book of M. Mézard, G. Parisi, M. A. Virosoro "Spin glass theory and beyond" [M-P-V]. This book is remarkable in many respects, and first of all its topic, which is the study of what are canonical, and even fundamental mathematical objects by physicist's methods. The book is an attempt by physicists to explain to other physicists the new concepts they have discovered about "spin glasses" and their relevance to a number of fundamental random structures. This could make difficult (and *did* in the case of the author) for an unprepared mathematician to get any idea of what this is all about. The book contains no rigorous results, and it is not obvious at all to even give a precise mathematical content to many of the statements made there. The existence of the topic of spin glasses appears to be known to quite a few mathematicians (see e.g. the recent book [B-P]), but overall it has been considered as an area where rigorous results are notoriously difficult to obtain. One must keep in mind however that it is rather unreasonable to attempt to directly attack the problems that the physicists (who use much less stringent methods) find challenging themselves (see however [N-S 2, 3]). We believe that there is no chance to make advances on the difficult issues until the easier ones have been clarified, and that only a systematic program to investigate the entire circle of ideas can lead to progress. The present paper reports the current status of this program. Beside attempting to provide an introduction to the topic, its main objective is to explain the author's contributions and point of view, and it should be kept in mind that some of the opinions expressed below are personal and might not be shared by others. All the results presented are fully rigorous; complete proofs can be found in [T5] to [T13].

There seems to be no better way to introduce the topic than to mimic the introduction of [M-P-V]. Consider a large population of individuals, numbered from 1 to N. Consider independent identically distributed (i.i.d) random variables (r.v.) $(g_{ij})_{1 \leq i < j \leq N}$. (The choice of Gaussian distribution is the simplest one, and is not believed to be essential). The variables g_{ij} represents the interaction of individuals $i < j$ (so, the larger g_{ij}, the more friendly i and j are towards each other). The independence requirement implies that for many (actually about 1/2) of the triples i, j, k, then $g_{ij} g_{jk} g_{ik} < 0$, so that we have unpleasant situations such as i friend of j and k ($g_{ij}, g_{ik} > 0$) but j and k enemies ($g_{jk} < 0$). In order to improve upon this tense situation, one tries to split the population in two parts, putting as far as possible friends together and enemies apart. This is done by assigning to each individual a number $\sigma_i \in \{-1, 1\}$, and each configuration $\boldsymbol{\sigma} = (\sigma_i)_{i \leq N}$ defines a splitting of the population in two. How successful this splitting is can be measured by the quantity

$$\sum_{1 \leq i < j \leq N} g_{ij} \sigma_i \sigma_j . \tag{1.1}$$

This adds the interactions between each pair of individuals in the same group, and subtracts the interactions between different groups. We are interested in the maximum of (1.1) over all $\boldsymbol{\sigma}$.

The reason why this maximum is very hard to find is that, for a given typical realization of the (g_{ij}), the function of $\boldsymbol{\sigma}$ given by (1.1) has apparently very many "near maxima" at locations that are not simply related to each other. Computer simulations seem to show that for large N

$$N^{-3/2} \max_{\boldsymbol{\sigma}} \sum_{1 \leq i < j \leq N} g_{ij} \sigma_i \sigma_j \simeq 0.7366 \tag{1.2}$$

with overwhelming probability. It is simple to prove that $N^{3/2}$ is the correct normalization factor and that the left hand side of (1.2) is essentially independent of the realization of the randomness (more precisely has random fluctuations of order $N^{-1/2}$); but the proof of the existence of the limit as $N \to \infty$, or its rigorous computation are nowhere in sight.

Faced with a very difficult optimization problem such as (1.2), the answer of statistical mechanics is to introduce a parameter $T \geq 0$ called temperature, and try to recover the case $T = 0$ as a limit case $T \to 0$. We consider the Hamiltonian (i.e. energy function)

$$H_N(\boldsymbol{\sigma}) = -\frac{1}{\sqrt{N}} \sum_{1 \leq i < j \leq N} g_{ij} \sigma_i \sigma_j - h \sum_{i \leq N} \sigma_i . \tag{1.3}$$

The factor $\sqrt{N}$ is the correct normalization to ensure that H_N/N remains bounded; The minus sign follows physics convention (the system is attracted to *low* energy configurations) and h is an "external field" that favors the choice $\sigma_i = 1$ over $\sigma_i = -1$. The previous optimization problem was the search (when $h = 0$) for the ground state (configuration with lowest energy) of $H_N(\boldsymbol{\sigma})$, which is the Hamiltonian of the famous Sherrington-Kirkpatrick (SK) model for spin glasses. This Hamiltonian was introduced to model disordered interaction between magnetic impurities ("spins") in metals. As a first approximation (mean field approximation), the geometric location of atoms is forgotten, and it is assumed that all pairs interact in the same way. The disorder of the system is then modeled by the random interactions g_{ij}. If the system is in thermal equilibrium at inverse temperature $\beta = 1/T$, statistical mechanics asserts that the probability of observing the system in configuration $\boldsymbol{\sigma}$ is given by Gibbs' measure

$$G_N(\boldsymbol{\sigma}) = \frac{1}{Z_N} \exp -\beta H_N(\boldsymbol{\sigma}) \tag{1.4}$$

where Z_N is the normalization factor (called the partition function)

$$Z_N = Z_N(\beta, h) = \sum \exp -\beta H_N(\boldsymbol{\rho}) \tag{1.5}$$

for a summation over all configurations. The problem is then to understand the structure of Gibbs' measure for the typical realization of the disorder (that is, of the r.v. g_{ij}). The mathematical difficulty is that it is very unclear what is the value of Z_N, which is a sum of 2^N quantities of wildly different orders of magnitude (the more so at large β, i.e. low temperature). Of particular interest is the "free energy"

$$F_N = F_N(\beta, h) = \log Z_N \tag{1.6}$$

(a physicist would rather use $-\frac{1}{\beta}\log Z_N$) the importance of which can be understood by the fact that its derivation with respect to the various parameters are physically measurable quantities, e.g.

$$\frac{1}{\beta}\frac{\partial F_N}{\partial h} = \langle \sum_{i \leq N} \sigma_i \rangle \tag{1.7}$$

is the global magnetization. In (1.7), as well as in the rest of the paper $\langle \cdot \rangle$ denotes thermal average, that is integration with respect to Gibbs' measure. Of course F_N is a random quantity (it depends of the disorder). However, it follows from general principles (the "concentration of measure phenomenon" [I-S-T]) that the random fluctuations of F_N are of order $N^{1/2}$, while F_N is of order N, so that much of the information about F_N is captured by EF_N. There and throughout the paper we denote by E the average with respect to the disorder. The main difficulty is that the typical value of Z_N is very different from (and of course smaller than) EZ_N, and that the bound $E \log Z_N \leq \log EZ_N$ given by Jensen's inequality is not an equality in the interesting cases.

It was soon realized that the first attempt to study the SK model [S-K] had serious flaws, and that the solution proposed there was correct only at high temperature. After several trials G. Parisi has proposed a very intricate picture ("the

Parisi solution") that is believed to be correct. The remarkable objects invented by Parisi start to draw attention from mathematicians [R], [B-S], [A-C1], [A-C2]. Unfortunately, the mathematical study of structures related to Parisi's solution is distinct from the more important issue as to whether these structures are really relevant to the SK model, and at the present time there is very little rigorous evidence that this is the case. It is of course fascinating that a simple and canonical energy function such as (1.3) can give rise to such extreme subtlety. But, beside the intrinsic interest of the SK model, the great discovery made by the physicists is that the behavior exhibited by this system appears to be rather universal, and to be present in a number of other situations involving random structures, several typical examples of which will be considered here.

Let us now try to draw a very rough picture of the situation. A main feature of the physicist's prediction is that given h, above a certain temperature, the system " is in a pure state" while below this temperature it spontaneously decomposes in many "pure states". The later statement can intuitively be understood by saying that if one studies the system at (extremely) long intervals, it looks like different objects. As we work in a disordered mean field model, it is unfortunately not obvious a priori how to formulate a meaningful definition of a pure state, and even less how to decompose a system in pure states. In standard statistical mechanics, say, on a finite subset S_N of an infinite lattice S, this is done by taking "infinite volume limit", $N \to \infty$. The set of configurations is then $\{-1,1\}^S$. The set of Gibbs' measures form a convex compact set, the extreme points of which are the pure states. In the present case, if one selects a sequence $(g_{i,j})_{i,j\in\mathbb{N}}$, then the structure of the Gibbs' measure of the N-spin system defined using $(g_{ij})_{i,j\leq N}$ varies wildly with N, the chaotic size dependence of [N-S1]. Despite the many statements of [M-P-V] starting by "in the thermodynamical limit...", it is not clear how to define a useful limit of the system as $N \to \infty$, that is, a satisfactory set of Gibbs' measures on $\{-1,1\}^N$. (See [AW],[N-S2] for the most interesting tentatives towards infinite volume limits in the lattice case. These attempts unfortunately still require taking subsequences, an operation that goes somewhat against the very goal of the theory, which is the ability to describe finite samples of matter.) This absence of infinite volume limit makes the topic of spin glasses distinctively different from main stream classical statistical mechanics. In Section 2, we will present a set of equivalent conditions that mean that the system is in a pure state, and for all the systems that we shall study we will define the high temperature region as the set of parameters where these conditions hold. (By definition the low temperature region consists of the other values.) The high temperature region is much simpler than the low temperature region and thereby is the natural starting point of a rigorous investigation. The results of Section 3 to 6 of the present paper assert that for four rather different models, the physicists magic formulas are indeed correct at high enough temperature, and a look at these formulas (such as (3.11)) should convince the reader that non trivial phenomenon occur there.

The author is keenly aware that it is a very risky endeavor to attempt rigorous proofs of results that are "known" by another community, in particular when most of his results bear on situations considered easy (if not trivial) by the physicists. It is thereby necessary to say a few words about the physicists methods, even though this might spoil some of the excitement the reader might otherwise have felt when discovering them in [M-P-V]. These methods are extremely creative and brilliant, but their purpose is very different from ours. It is not to provide proofs, but rather to discover what happens with reasonable certainty. The favorite method, the replica method, attempts to compute directly the limiting expected free energy density $\lim_{N\to\infty} N^{-1}EF_N$, which, as we mentioned, captures much information about F_N. The annoying logarithm is disposed of by the formula

$$\log x = \lim_{n\to 0} \frac{x^n - 1}{n} \tag{1.8}$$

and the issue is then to compute EZ_N^n, which can be done for n integer using n copies ("replicas") of the system. One then makes an analytic continuation at $n \to 0$. Besides a few lesser problems, the computation of EZ_N^n in the case of the SK model is done by a saddle point method requiring to minimize a function of $n(n-1)/2$ variables. To quote [M-P-V], p. 12, "$n(n-1)/2$ becomes negative for $0 < n < 1$, and it is not clear how to give a precise definition of the minimum of a function which depends on a negative number of variables". As G. Parisi so nicely puts it "the replica method is yet to be put on firm mathematical ground".

The computations using the replica method involve a tricky issue (the real meaning of which is not clear to me) as to whether the n $(0 < n < 1)$ replicas involved can be assumed to be equivalent ("replica-symmetry", the easiest case) or not ("replica symmetry breaking"). It seems that the case where the system is in a pure state (as defined in Section 2) corresponds to the case of replica-symmetry. The physicists seem to have absolute faith in the replica method, at least in the replica symmetric case. Typing the words "replica symmetry" on a data base such as INSPEC brings in dozens of papers that rely upon this method. More often than not, these papers "solve" a problem by writing down formulas provided by the replica method (sometimes using Parisi's scheme of replica-symmetry breaking) and optimizing over the various parameters. These theoretical results are then supplemented by computer studies for large N, where (due to extreme computational difficulties), "large" means typically of order 100. But despite the fact that it is not clear what the replica method really does (even in the replica symmetric case) it is an amazing tool to discover complicated formulas in a very compact way.

As the mathematical and even the physical contents of the replica method are obscure, physicists have developed an alternative method, the cavity method, which is essentially induction upon N. (This is the method we will use, even though our computations are very different). A possible reason why the physicists find the high temperature case easy is that they assume from the start "on natural physical grounds" that at high temperature the system is in a pure state (see eg. [M]). At the philosophical level, it requires some faith to believe that a mathematical object such as (1.3), that has very little claim to be a realistic model for matter will obey physical principles. At the mathematical level, once one assumes that the system is in a pure state, the magical formula (2.9) below allows all kinds of computations that readily lead to a rather complete picture of the system. On the other hand the mathematician, when faced with the system with no a priori information has at first great difficulties to prove anything at all.

Even though the high temperature phase of disordered systems is considered easy in physics, it still has some interest even at this level, in particular because in some important cases (such as that of Section 4) the high temperature region is believed to extend all the way to zero temperature. The real long term challenge is however the low temperature region. The very complicated structure of the predicted low temperature behavior of the SK model does not make it a good place to start from, so we have rather considered the p-spin interaction model, a model closely related to the SK model and for which the predicted low temperature behavior is much simpler (and, actually the simplest possible). We did succeed in this case to prove (at the "edge" of the low temperature region) the main feature of the Parisi's prediction, the spontaneous decomposition of the system in pure states "far from each other". This is the content of Section 7.

ACKNOWLEDGMENT The author will be forever grateful to Erwin Bolthausen for having introduced him to such an immensely enjoyable topic. The author thanks G. Choquet, M. Mézard and many others for useful comments.

2. SYSTEMS IN A PURE STATE

We denote throughout the paper the set $\{-1,1\}^N$ of all configurations by Σ_N. Since (Σ_N, G_N) is a probability space it is natural to consider powers of it, $(\Sigma_N^m, G_N^{\otimes m})$ (often $m \leq 4$). These are called replicas. The word "replicas" simply means that we consider several copies of (Σ_N, G_N). Then copies are taken for the *same* realization of the disorder. A generic point in Σ_N^m is denoted by $(\boldsymbol{\sigma}^1, \cdots, \boldsymbol{\sigma}^m)$. These replicas are often called in physics "real replicas" to distinguish them for the n replicas ($n \to 0$) of the replica formalism (would these be unreal?) and needless to say that we will use only "real" replicas. Replicas are very useful to transform product of integrals for the Gibbs measure into multiple integrals, such as in the formula

$$\langle f_1(\boldsymbol{\sigma})\rangle\langle f_2(\boldsymbol{\sigma})\rangle = \langle f_1(\boldsymbol{\sigma}^1) f_2(\boldsymbol{\sigma}^2)\rangle.$$

There, as well as in the rest of the paper, $\langle \cdot \rangle$ denotes thermal average in Σ_N^n as well as in Σ_N, so that the bracket on the right is a double integral with respect to Gibbs measure.

The overlap between two configurations $\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2$ is defined by

$$\frac{1}{N}\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2 = \frac{1}{N}\sum_{i\leq N} \sigma_i^1 \sigma_i^2, \tag{2.1}$$

a good measure of their distance. The simplest (although not the most intuitive) way to define a system in a pure state is to say that the function $\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2 \to \frac{1}{N}\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2$ is nearly constant on Σ_N^2, as is formalized in (2.2) below. This idea is apparent in [M-V-P]. It is likely that, at least at the intuitive level, many other ideas of this section can also be found there, but of course the point was to identify precise statements that are amenable to proof.

Consider a sequence $(G_N)_{N\geq 1}$, where G_N is an exchangeable random probability measure on Σ_N, that is, such that its distribution is invariant under any permutation of the coordinates (a crucial assumption). Typically (G_N) will be a sequence of Gibbs' measures such as (1.4), at a given value of the parameters β, h, so it should be self evident what we mean by $\langle\,\cdot\,\rangle$, replicas, etc. Of course E will denote expectation with respect to the randomness of G_N. In the next theorem, all brackets $\langle\,\cdot\,\rangle$ are for G_N. This seemingly complicated statement is a result of the author's (unreasonable?) attempt not only to give a description of the results, but also some ideas of what are the obstacles to reach them. The reader who is interested only in getting an overview of the results should only read conditions (2.2), (2.7), (2.8). After reading the comments next to Definition 2.2, he should then skip the sketch of proof of Theorem 2.1.

THEOREM 2.1. *For a sequence G_N of exchangeable random probability measures the following properties are equivalent.*

$$\lim_{N\to\infty} E\langle \frac{1}{N}|\boldsymbol{\sigma}^1\cdot\boldsymbol{\sigma}^2 - \langle\boldsymbol{\sigma}^1\cdot\boldsymbol{\sigma}^2\rangle|\rangle = 0 \tag{2.2}$$

$$\lim_{N\to\infty} E\langle (\frac{1}{N}(\boldsymbol{\sigma}^1-\boldsymbol{\sigma}^2)\cdot(\boldsymbol{\sigma}^3-\boldsymbol{\sigma}^4))^2\rangle = 0 \tag{2.3}$$

$$\lim_{N\to\infty} E\langle (\frac{1}{N}(\boldsymbol{\sigma}^1-\boldsymbol{\sigma}^2)\cdot\boldsymbol{\sigma}^3)^2\rangle = 0 \tag{2.4}$$

$$\lim_{N\to\infty} E(\langle\sigma_1\sigma_2\rangle - \langle\sigma_1\rangle\langle\sigma_2\rangle)^2 = 0 \tag{2.5}$$

$$\lim_{N\to\infty} E\langle(\sigma_1^1-\sigma_1^2)(\sigma_2^1-\sigma_2^2)\rangle\langle\sigma_1^3\sigma_2^3\rangle = 0 \tag{2.6}$$

$$\forall n,\ \lim_{N\to\infty} E(\langle\sigma_1\cdots\sigma_n\rangle - \langle\sigma_1\rangle\cdots\langle\sigma_n\rangle)^2 = 0 \tag{2.7}$$

(2.8) *For each n, the expected total variation distance of the law of $(\sigma_1,\cdots,\sigma_n)$ under G_N to the set of a product measures on $\{-1,1\}^n$ goes to zero.*
(2.9) *For any continuous bounded function f on $\mathbb{R}^{n\times m}$, independent $N(0,1)$ variables $(\xi_i^j)_{i\leq N, j\leq n}, (h^{j\ell})_{j\leq n,\ell\leq m}$ that are independent of the variables g_{ij}, we have*

$$\lim_{N\to\infty} E|\langle f((\frac{\boldsymbol{\xi}^j\cdot\boldsymbol{\sigma}^\ell}{\sqrt{N}})_{\substack{j\leq n\\ \ell\leq m}})\rangle - E_h f((\frac{\boldsymbol{\xi}^j\cdot\boldsymbol{b}}{\sqrt{N}} + h^{j\ell}\sqrt{1-\|\boldsymbol{b}\|^2/N})_{\substack{j\leq n\\ \ell\leq m}})| = 0$$

where $\boldsymbol{\xi}^j\cdot\boldsymbol{\sigma}^\ell = \sum_{i\leq N}\xi_i^j\sigma_i^\ell, \boldsymbol{b} = (\langle\sigma_i\rangle)_{i\leq N} = (\langle\sigma_i^\ell\rangle)_{i\leq N}$; and where E_h denotes expectation in $(h^{j\ell})_{\substack{j\leq n\\ \ell\leq m}}$ only.

DEFINITION 2.2. *We say that the sequence G_N of random measures is in a pure state if the equivalent conditions of Theorem 2.1 hold.*

Conditions (2.2) to (2.4) are global "geometric" conditions. The idea of (2.3) and (2.4) is that the centering in (2.2) is better replaced by symmetrization. The reason for considering these two similar but different expressions will be apparent when we try to prove them. Conditions (2.5) and (2.6) are "local" reformulations of (2.3), (2.4) respectively, that involve only two spins, and are better adapted to induction over N. The least expected, and the most useful fact is (2.9), which means that in practice integral $\langle f(\frac{\boldsymbol{\xi}^j\cdot\boldsymbol{\sigma}^\ell}{\sqrt{N}})\rangle$ depends upon G_N only through $\boldsymbol{b}$, a fact that is at the root of the magic formulas of Sections 3 to 5.

Theorem 2.1 provides conceptual clarification, and is very easy to prove, because we do not relate the rates at which the various quantities go to zero and need only "soft" estimates. This is why, when a physicist assumes "on physical grounds" that, say (2.5) holds, he then correctly feels that the problem is easy. On the other hand, *proving* that the conditions of Theorem 2.1 hold require much more precise estimates.

To sketch the proof of Theorem 2.1 let us set $a_i = (\sigma_i^1 - \sigma_i^2)(\sigma_i^3 - \sigma_i^4)$ so $|a_i| \leq 4$, and (2.3) means that $E(\langle (N^{-1}\sum a_i)^2\rangle) \to 0$, which, by symmetry among the coordinates, is equivalent to $E\langle a_1 a_2\rangle \to 0$. Now, by independence of the replicas

$$\begin{aligned}\langle a_1a_2\rangle &= \langle(\sigma_1^1-\sigma_1^2)(\sigma_2^1-\sigma_2^2)\rangle^2 \\ &= 4(\langle\sigma_1\sigma_2\rangle - \langle\sigma_1\rangle\langle\sigma_2\rangle)^2,\end{aligned} \tag{2.10}$$

which proves the equivalence of (2.3) and (2.5). The equivalence of (2.4) and (2.6) is similar. It is obvious that (2.4)$\Rightarrow$(2.3) and (2.5)$\Rightarrow$(2.6), using (2.10) and Cauchy-Schwarz. The equivalence of (2.2) and (2.4) is easy, since $|(\boldsymbol{\sigma}^1-\boldsymbol{\sigma}^2)\cdot\boldsymbol{\sigma}^3| \leq 2N$. It is obvious that (2.6)$\Rightarrow$(2.8)$\Rightarrow$(2.5). To prove the more surprising fact that (2.5)$\Rightarrow$(2.7), we observe that, since $|\sum\limits_{i\leq N} a_i| \leq 4N$, then (2.5) implies $E\langle(N^{-1}\sum_{i\leq N} a_i)^n\rangle \to 0$, and proceeding as before $\lim\limits_{N\to\infty} E\langle a_1\cdots a_n\rangle \to 0$, which means $\lim\limits_{N\to\infty} E\langle\prod\limits_{i\leq n}(\sigma_i^1-\sigma_i^2)\rangle^2 = 0$, from which (2.7) follows easily. Thus we have the equivalence of (2.2) to (2.8). We will not use that (2.9) implies the other conditions, so we just prove that it is a consequence of (2.2). Setting

$$X = \langle f((\frac{\boldsymbol{\xi}^j\cdot\boldsymbol{\sigma}^\ell}{\sqrt{N}}))\rangle; Y = E_h f((\frac{\boldsymbol{\xi}^j\cdot\boldsymbol{b}}{\sqrt{N}} + h^{j\ell}\sqrt{1-\|\boldsymbol{b}\|^2/N})),$$

the proof consists of showing that $E(X-Y)^2 \to 0$, by showing that $EX^2 - EY^2 \to 0$ and $EXY - EY^2 \to 0$. We will (to avoid complicated notations) prove only that $EX - EY \to 0$. The argument to prove that (2.2) implies (2.9) is the same. If $(w^{j\ell})$ is a jointly gaussian family, the quantity $Ef((w^{j\ell}))$ is determined by the joint law of $(w^{j\ell})$, that is by the numbers $E(w^{j\ell}w^{j'\ell'})$, and this dependence is of course continuous. Denoting by E_ξ expectation in the variables $\boldsymbol{\xi}^j$ only, $E_{\boldsymbol{\xi}} f((\frac{\boldsymbol{\xi}^j\cdot\boldsymbol{\sigma}^\ell}{\sqrt{N}}))$ depends only upon the numbers $E(\frac{\boldsymbol{\xi}^j\cdot\boldsymbol{\sigma}^\ell}{\sqrt{N}}\frac{\boldsymbol{\xi}^{j'}\cdot\boldsymbol{\sigma}^{\ell'}}{\sqrt{N}}) = \delta_{jj'}(\boldsymbol{\sigma}^\ell\cdot\boldsymbol{\sigma}^{\ell'}/N)$. For the generic point $\boldsymbol{\sigma}^1,\cdots,\boldsymbol{\sigma}^m$ of the m-replica, (2.2) says that all products $\boldsymbol{\sigma}^\ell\cdot\boldsymbol{\sigma}^{\ell'}/N(\ell\neq\ell')$ are about $\langle\boldsymbol{\sigma}^\ell\cdot\boldsymbol{\sigma}^{\ell'}\rangle/N = \|\boldsymbol{b}\|^2/N$ (and 1 if $\ell=\ell'$). Now if we set $w^{j\ell} = \frac{\boldsymbol{\xi}^j\cdot\boldsymbol{b}}{\sqrt{N}} + h^{j\ell}\sqrt{1-\|\boldsymbol{b}\|^2/N}$ we see that this jointly gaussian family of r.v. satisfies $E(w^{j\ell}w^{j'\ell'}) = \delta_{jj'}\|\boldsymbol{b}\|^2/N$ for $\ell\neq\ell'$ and $\delta_{jj'}$ for $\ell=\ell'$. Thus, for the generic point $\boldsymbol{\sigma}^1,\cdots,\boldsymbol{\sigma}^m$ we have $E_\xi f((\frac{\boldsymbol{\xi}^j\cdot\boldsymbol{\sigma}^\ell}{\sqrt{N}})) \simeq E_\xi Y$, and the result follows. $\square$

The reader has observed that Theorem 2.1 does not say that "G_N resembles G_{N+1}" or that quantities such as $E\langle\boldsymbol{\sigma}^1\cdot\boldsymbol{\sigma}^2\rangle/N$ converge as $N\to\infty$. Proving this is a different question.

In the situation of Theorem 2.1, since the law of $(\sigma_1,\cdots,\sigma_n)$ under Gibbs' measure is asymptotically close to a product measure, it is close to the product measure ν on $\{-1,1\}^n$ such that $\int\sigma_i d\nu(\sigma_1,\cdots,\sigma_n)=\langle\sigma_i\rangle$. In the cases we will consider, the quantities $\langle\sigma_i\rangle$ are asymptotically i.i.d. r.v. (and converge in law), thereby providing a precise picture of the finite projections of Gibbs' measure.

3. The Sherrington-Kirkpatrick model

The most studied case is when $h=0$. It was proved in [A-L-R] that the system is in a pure state if $\beta<1$. An easy consequence of the result of [C] is that there exists values of $\beta>1$, arbitrarily close to one, for which this is not the case (one expects that this is never the case if $\beta>1$). There is a very special phenomenon happening in the case $\beta<1$, namely that

$$\lim_{N\to\infty}E\frac{1}{N}\log Z_N=\lim_{N\to\infty}\frac{1}{N}\log EZ_N(=\frac{\beta^2}{4})\tag{3.1}$$

(by Jensen's inequality there is always inequality $\leq$). This apparently makes things much simpler. There are several very interesting methods (such as use of stochastic calculus to prove central limit theorems [C-N]) that seem to work for this case only. Even though some nagging questions remain, there is a rather complete picture of this case ([T5 Section 2]). Unfortunately, a behavior such as (3.1) is exceptional and we will concentrate upon the more challenging case $h>0$. The formula corresponding to (3.1) is then given by (3.13) below, and is remarkable enough to make one wonder how such a formula is possible, and moreover can be proved. It turns out that the proof of (3.13) for small β is rather easy. This proof is also very instructive because the other cases considered in Theorem 3.1 below, as well as the results of Sections 4 to 6, although technically very much more involved do follow the same global strategy, so we will outline the main steps. The central issue is always to prove that the system is in a pure state (after which use of (2.9) allows all kinds of computations). In the present case at high enough temperature, that was actually done in [F-Z] as a special case of a powerful (and complicated) approach that handles much more general cases (such as finite range interactions), but it is very instructive to give here a simple direct argument.

We start with the inequality (implicitly proved in Section 2 by expansion of $(\tilde{\boldsymbol{\sigma}}\cdot\boldsymbol{\sigma}^*)^2$)

$$C_N=C_N(\beta,h):=E\langle(\frac{1}{N}\tilde{\boldsymbol{\sigma}}\cdot\boldsymbol{\sigma}^*)^2\rangle\leq\frac{4}{N}+E\langle\tilde{\sigma}_N\sigma^*_N\tilde{\sigma}_{N-1}\sigma^*_{N-1}\rangle$$

where $\tilde{\boldsymbol{\sigma}}=\boldsymbol{\sigma}^1-\boldsymbol{\sigma}^2,\boldsymbol{\sigma}^*=\boldsymbol{\sigma}^3-\boldsymbol{\sigma}^4$. To compute the last term we use the cavity method; we compute the bracket by regrouping in the Hamiltonian the terms not containing σ_N or σ_{N-1} and we find

$$E\langle\tilde{\sigma}_N\sigma^*_N\tilde{\sigma}_{N-1}\sigma^*_{N-1}\rangle\approx E\frac{1}{Z}\langle Av\tilde{\sigma}_N\sigma^*_N\tilde{\sigma}_{N-1}\sigma^*_{N-1}\mathcal{E}\rangle_{N-2}.\tag{3.2}$$

There $\langle\,\cdot\,\rangle_{N-2}$ denotes Gibbs measure for an $(N-2)$ spin system at inverse temperature $\beta' = \beta\sqrt{1-2/N}$, external field $h' = h(1-2/N)^{-1/2}$, and Av the average over all values of $\sigma_N^\ell, \sigma_{N-1}^\ell (\ell \le 4) = \pm 1$,

$$\mathcal{E} = \exp \beta \sum_{\ell \le 4} \sum_{0 \le j \le 1} \sigma_{N-j}^\ell \Big(\frac{1}{\sqrt{N}} \sum_{i \le N-2} \sigma_i^\ell g_{i,N-j} + h \Big)$$

and $Z = \langle Av\mathcal{E}\rangle_{N-2}$. The formidable looking formula (3.2) is actually almost an algebraic identity, except that we have neglected the lower order interaction terms between σ_N^ℓ and σ_{N-1}^ℓ (hence the small error acknowledged by the $\approx$). The slight change of parameters β, h into β', h' turns out to be a secondary detail, and will be ignored from now on. The difficulty with the cavity method is that we do not know more about $\langle\,\cdot\,\rangle_{N-2}$ than about $\langle\,\cdot\,\rangle$, so that it is hard to use (3.2). An easy way out is provided by the observation that $Z \ge 1$ by Jensen's inequality and that $\langle Av\tilde{\sigma}_N \sigma_N^* \tilde{\sigma}_{N-1}^2 \sigma_{N-1}^* \mathcal{E}\rangle_{N-2} \ge 0$, because it can be written (using thermal independence of the variables with a * from those with a $\tilde{}$) as a square. Then the right hand side of (3.2) can be bounded by $E\langle Av\tilde{\sigma}_n \sigma_N^* \tilde{\sigma}_{N-1} \tilde{\sigma}_{N-1}^* \mathcal{E}\rangle_0$. (This argument to dispose of the denominator will be referred to later as the positivity argument). This later quantity is much easier to evaluate. Integrating first in the $g_{i,N-j} (j = 0,1)$ one obtains after a few lines of straightforward estimates a bound $\beta^2 L(\beta) C_{N-2}(\beta', h')$, where $L(\beta)$ remains bounded with β. This yields the relation $C_N(\beta, h) \le \frac{1}{2} C_{N-2}(\beta', h') + o(1)$ if β is small enough, which implies $\lim_{N\to\infty} C_N(\beta, h) = 0$.

The positivity argument used above does not take advantage of the fact that often the denominator is much larger than 1, and as the result of this loss of a constant factor, we cannot expect to reach this way the entire high temperature region. The merit of the positivity argument is that it is the simplest approach we know, and thus it is particularly useful in complicated situations. Unfortunately this argument itself often runs into a serious difficulty (which does not exist in the case of the SK model) namely that the estimation of C_N usually involves D_{N-2} (at slightly different parameters), where $D_N = E\langle (\frac{1}{N}\tilde{\boldsymbol{\sigma}} \cdot \boldsymbol{\sigma}^3)^2 \rangle$. It seems a posteriori true that C_N and D_N are of the same order, but unfortunately we do not see a priori how to prove better than $C_N \le \sqrt{D_N}$ (almost proved in Section 2), and this leads to useless relations such as $C_N \le \theta\sqrt{C_{N-2}} + o(1)$ where $\theta < 1$. Because of this a priori difficulty in relating C_N and D_N, one sees that a better strategy is to study D_N. But then, in the right hand side of (3.2) the numerator has to be replaced by $\langle Av\tilde{\sigma}_N \tilde{\sigma}_{N-1} \sigma_N^3 \sigma_{N-1}^3 \mathcal{E}\rangle_{N-2}$ which has no reason to be positive, and the positivity argument does not work. This unfortunate state of affairs is largely responsible for the great technicality of many proofs, even at a very high temperature.

Now that we have proved that for small β the system is in a pure state, we observe that brackets involving $\mathcal{E}$ resemble the brackets of (2.9). The requirement there that f was bounded was made only to avoid a technical statement; reasonable growth suffices. This means that we can now use (2.9) to make all sorts of computations, of which we now give a typical example. Proceeding as in 3.2, we have

$$\langle \sigma_N \rangle = \frac{1}{Z} \langle Av \sigma_N \mathcal{E} \rangle_{N-1}$$

where $Z = \langle Av\mathcal{E} \rangle_{N-1}$ and $\mathcal{E} = \exp \beta\sigma_N (N^{-1/2} \sum_{i \le N-1} \sigma_i g_{i,N} + h)$. It follows from (2.9) (used for $m = n = 1, f(x) = \exp \beta\sigma_N x$) that

$$\langle \sigma_N \rangle \simeq \operatorname{th} \beta \big(N^{-1/2} \sum_{i \le N-1} g_{i,N} \langle \sigma_i \rangle_{N-1} + h \big). \tag{3.3}$$

If we set $r_{N-1} = N^{-1} \sum_{i \le N-1} \langle \sigma_i \rangle^2_{N-1}$, we then have

$$E r_N \simeq E \langle \sigma_N \rangle^2 \simeq E \operatorname{th}^2 \beta (g \sqrt{r_{N-1}} + h) \tag{3.4}$$

where g is standard normal independent of r_{N-1}.

To make full use of this, it would be very nice to know that r_N is essentially non random, which amounts to show that $(E\langle \sigma_N \rangle^2)^2 \simeq E\langle \sigma_N \rangle^2 \langle \sigma_{N-1} \rangle^2$. The right-hand side can be estimated as in (3.4) using cavity and (2.9), and only a few lines of computations are required to get a relation of the type Var $r_N \le \beta^2 L(\beta)$Var $r_{N-2} + o(1)$, so that for small β we have Var $r_N \to 0$, and (3.4) leads to $q_N \simeq E \operatorname{th}^2 \beta(g\sqrt{q_{N-1}} + h)$ where $q_N = E\langle \sigma_N \rangle^2$ and to $q_N \to q$ where q satisfies (3.10) below. To calculate F_N, we fix $h\beta = h'$ and we write

$$\frac{\partial F_n}{\partial \beta} = \frac{1}{\sqrt{N}} \sum_{i<j} g_{ij} \langle \sigma_i \sigma_j \rangle. \tag{3.5}$$

We then (following [A-L-R]) apply the (extremely useful) integration by parts formula

$$E(gf(g)) = Ef'(g) \tag{3.6}$$

valid when g is standard normal and f smooth enough, to obtain

$$E \frac{1}{N} \frac{\partial F_N}{\partial \beta} = \beta \frac{N-1}{2N} (1 - E\langle \sigma_N \sigma_{N-1} \rangle^2) \simeq \frac{\beta}{2} (1 - q_N^2) \tag{3.7}$$

where we use that $E\langle \sigma_N \sigma_{N-1} \rangle^2 \simeq E\langle \sigma_N \rangle^2 \langle \sigma_{N-1} \rangle^2 \simeq (E\langle \sigma_N \rangle^2)^2 = q_N^2$.

To prove (3.13), one simply checks that it is true for $\beta = 0$, and the (miraculous) fact that $\frac{\partial SK}{\partial \beta} = \frac{\beta}{2}(1 - q^2)$. Concerning the structure of the r.v. $\langle \sigma_i \rangle$, we proceed as in (3.3) to obtain that, for any fixed n, as $N \to \infty$, for $0 \le k \le n-1$

$$\langle \sigma_{N-k} \rangle \simeq \operatorname{th} \beta (g_k \sqrt{q_{N-n}} + h)$$

where $(g_k)_{k \le n-1}$ are i.i.d. $N(0,1)$, so that (3.12) below is obvious.

To go beyond this first round of results, that is to be able to handle cases where the positivity argument does not work, and to perform the previous computations with a better control of the error terms, we need to develop another technique to estimate EU/Z, when Z is a quantity such as in (3.2). The basic procedure is to replace Z by the quantity $\hat{Z}$ provided by (2.9) (even when we have not yet proved that the system is in a pure state) and to write

$$\frac{U}{Z} = \frac{2U}{\hat{Z}} - \frac{UZ}{\hat{Z}^2} + \frac{U(Z-\hat{Z})^2}{Z\hat{Z}^2}. \tag{3.8}$$

The idea is that the last term has a tendency to be small because of the factor $(Z-\hat{Z})^2$, factor which is not affected when one takes bounds and uses absolute values. On the other hand, the first two terms on the right have a denominator where the dependence in the $\boldsymbol{g}^j$ is only through $\boldsymbol{g}^j \cdot \boldsymbol{b}$, so that they can be evaluated by conditioning upon these. Carrying out that program results in extremely long computations but once the arguments are properly organized these allow to gain a very precise picture of the model. The largest domain in which we know how to control the model is

$$D = \{(\beta,h) : \text{ either } \beta < \beta_0, \text{ or } h \geq h_1(\beta) \text{ or } 0 < \beta < 1 \text{ and } \ h \leq h_2(\beta)\} \tag{3.9}$$

where $h_1(\beta), h_2(\beta)$ are certain specific positive functions. We consider $q = q(\beta,h)$, the root of the equation

$$q = E\,\mathrm{th}^2\,\beta(g\sqrt{q}+h) \tag{3.10}$$

(that is well defined on D) and the function

$$SK(\beta,h) = \frac{\beta^2}{4}(1-q)^2 + E\log\mathrm{ch}\,\beta(g\sqrt{q}+h). \tag{3.11}$$

In the following statement, K denotes a number depending upon β, h only.

Theorem 3.1. *If (β,h) belongs to D, the following occurs*

$$E\langle \exp\frac{1}{KN}(\boldsymbol{\sigma}^1\cdot\boldsymbol{\sigma}^2 - q)^2\rangle \leq K \tag{3.12}$$

$$\lim_{N\to\infty} N^{-1}EF_N = SK(\beta,h) \tag{3.13}$$

(3.14) *Given any n, the r.v. $\langle\sigma_1\rangle,\cdots,\langle\sigma_n\rangle$ are asymptotically i.i.d., and their limiting law is the law of* $\mathrm{th}\,\beta(g\sqrt{q}+h)$ *where g is $N(0,1)$*

(3.15) *Given replicas $\boldsymbol{\sigma}^1,\cdots,\boldsymbol{\sigma}^p$, for any expression f that is the product of k quantities of the type $\boldsymbol{\sigma}^\ell\cdot\boldsymbol{\sigma}^{\ell'} - E\langle\boldsymbol{\sigma}^1\cdot\boldsymbol{\sigma}^2\rangle(\ell\neq\ell')$ then $\lim_{N\to\infty} N^{-k/2}E\langle f\rangle$ exists.*

Comment 1. The validity of (3.13) was also investigated by M. Shcherbina. In her remarkable recent paper [Sh2] she proves in particular that it holds whenever $\beta < 1$.

Comment 2. Condition (3.12) is a very precise improvement of (2.1) and moreover it contains the information that $\langle\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2\rangle$ is nearly q. In the case $h = 0, (\beta < 1)$ it is much easier to prove [T5, Section 2]. We feel that the greatest importance of exponential equalities such as (3.12) is that they carry information that can be transferred (with loss) to *different* (but close) values of the parameters. Specifically, one can show that given $\epsilon > 0$, if β', h' are close enough (depending upon ϵ) of β, h, then, if (3.12) holds at β, h, then at β', h' the overlap $N^{-1}\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2$ will essentially take values only in an interval of length $\leq \epsilon$. This is shown in [T10], and this is how the part $0 < \beta < 1$ and $h < h(\beta)$ of D is controlled, building upon the case $0 < \beta < 1, h = 0$ that is obtained through special arguments. We believe that such a "transfer principle" has to be a part of a proof that would extend Theorem 3.1 to the entire region (3.12).

Comment 3. The limits of (3.15) contain in principle all the information on the (random) joint law under Gibb's measure of the maps $(\boldsymbol{\sigma}^1, \cdots, \boldsymbol{\sigma}^p) \to N^{-1/2}(\boldsymbol{\sigma}^\ell \cdot \boldsymbol{\sigma}^{\ell'} - E\langle\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2\rangle)$, and (3.15) is obtained by an explicit method, allowing in principle computation of the limits. We did check that asymptotically, the law under Gibbs measure of the "symmetrized overlaps" $N^{-1/2}(\boldsymbol{\sigma}^1 - \boldsymbol{\sigma}^2) \cdot (\boldsymbol{\sigma}^3 - \boldsymbol{\sigma}^4)$ is asymptotically gaussian (independent of the disorder) of variance $A/(1-\theta)$, where

$$A = 4E\frac{1}{\mathrm{ch}^2\, \beta(g\sqrt{q} + h)} \tag{3.16}$$

and where θ is the quantity (3.17) below. It seems to us that the joint laws under Gibbs measure of the maps $N^{-1/2}(\boldsymbol{\sigma}^\ell \cdot \boldsymbol{\sigma}^{\ell'} - \langle\boldsymbol{\sigma}^\ell \cdot \boldsymbol{\sigma}^{\ell'}\rangle)$ should be asymptotically gaussian, independent of the disorder, and that the joint laws of the maps $N^{-1/2}\langle\boldsymbol{\sigma}^\ell \cdot \boldsymbol{\sigma}^{\ell'}\rangle$ should also be gaussian. Checking this is in principle elementary, but the algebraic formalism needed to write nicely such a result remains to be found.

What is the high temperature region? The physicists believe that it is the region defined by

$$\theta = \beta^2 E\frac{1}{\mathrm{ch}^4\, \beta(g\sqrt{q} + h)} < 1 \tag{3.17}$$

(where q is as in (3.10)). This conclusion was first obtained in [A-T], literally, by analyzing the eigenvalues of matrices of dimension 0×0. In order to estimate the size of the cultural gap (and, in particular why the notion of "triviality" is very relative) it is instructive to outline the derivation of this using the cavity method from [M-P-V]. After conducting some computations that seem based upon the a priori assumption that most of the conclusions of Theorem 2.1 are valid the authors reach the relation $C_N = \theta C_N + A/N$ where A is given by (3.16) and θ by (3.17) and conclude "thus we must have $\theta < 1$ to have C_N positive". In other words physicists do not mind purely formal computations, and what is amazing is how well this works.

The predicted structure of the low temperature region involves the mysterious phenomenon of "replica-symmetry breaking' and I am much grateful to

M. Mézard for having explained it to me outside the replica formalism (see [To]). The idea is simply that an arbitrarily small coupling between two replicas has big consequences. Consider, on Σ_N^2, the Hamiltonian H_N given by $H_N(\boldsymbol{\sigma}^1)+H_N(\boldsymbol{\sigma}^2)+t\boldsymbol{\sigma}^1\cdot\boldsymbol{\sigma}^2/N$, and the corresponding Gibbs measure $\langle\cdot\rangle_t$ on Σ_N^2 (which is NOT a product measure). Consider the function $\varphi_{N,\beta,h}(t)=\varphi_N(t)=\langle N^{-1}\boldsymbol{\sigma}^1\cdot\boldsymbol{\sigma}^2\rangle_t$. Replica symmetry breaking means that "$\lim\limits_N \varphi_N(t)$ is discontinuous at zero", which, as we cannot prove the existence of the limit, we formulate as follows.

DEFINITION 3.2. *We say that there is replica symmetry breaking (RSB) for the parameters (β,h) if the sequence φ_N is* not *uniformly equicontinuous as $N\to\infty$ at the point $t=0$.*

This means that there is an $\epsilon>0$ such that there are arbitrarily large values of N and arbitrarily small values of t for which $|\varphi_N(t)-\varphi_N(0)|\geq\epsilon$.

THEOREM 3.3. *There is replica symmetry breaking at the generic point (in the sense of Baire category) of the region $\theta>1$.*

One of course expects that there is RSB at each point of the region $\theta>1$. It is a simple consequence of (3.8) that there is no RSB in the region D of Theorem 3.1. The status of the other points of the region (3.9) is unknown.

The proof of Theorem 3.3 relies upon the basic observation that if there is no RSB, then (2.2) holds. We then know how to make computations and we can make the physicist's relation $C_N\cong\theta C_N+A/N$ rigorous. It is worthwhile to detail a bit what happens here, as this touches what seems to us to be the central obstacle in proving that under (3.17) the system is in a pure state. If there is no RSB, one shows that for any $\epsilon>0$, when N is large, we essentially always have $|(\boldsymbol{\sigma}^1-\boldsymbol{\sigma}^2)\cdot\boldsymbol{\sigma}^3|\leq\epsilon N$. This implies that if we set $D_{n,N}=E\langle(N^{-1}(\boldsymbol{\sigma}^1-\boldsymbol{\sigma}^2)\cdot\boldsymbol{\sigma}^3)^{2n}\rangle$ then for large $N, D_{2,N}\ll D_{1,N}$. This is extremely valuable because when one tries to compute $D_N=D_{1,N}$ by the cavity method using an order 2 expansion, we find terms involving $D_{2,N}$, and we now know that these are indeed smaller order terms. In contrast, when we try to prove that $D_{1,N}$ is small under (3.17) we do not know a priori that $D_{2,N}\ll D_{1,N}$. This is not a trivial issue. For a related model (to be considered in Section 7) we did prove rigorously that there exist situations where $D_{2,N}$ and $D_{1,N}$ are of the same order (and of order at most $1/\sqrt{N}$). Moreover, this issue does not seem to have been considered by the physicists. They seem to ignore it when using either the cavity method or the "stability analysis" of the replica formalism (a personal impression based on the fact that, in particular for the model of Section 7 a wrong solution, that roughly speaking "would be true if $D_{2,N}\ll D_{1,N}$", is found to be stable in this sense.) To control $D_{1,N}$ close enough to the low temperature region without a priori assumptions, the most natural way seems to control $D_{2,N}$; but this in turn requires to control $D_{3,N}$, etc., leading naturally to the consideration of exponential inequalities such as in Theorem 3.1.

One of the striking and easily formulated predictions of the Parisi solution is that at low temperature certain quantities depend upon the realization of randomness. For different Hamiltonians (that make matters easier) it is shown in [P-S], [Sh1] that the quantity $q_N=N^{-1}\sum\limits_{i\leq N}\langle\sigma_i\rangle^2$ essentially depends upon the randomness when β is large enough (its variance does not go to zero as $N \to \infty$). For the Hamiltonian (1.3) it is shown in [T5] that when $h = 0$ and β is large enough the quantity $q' = N^{-2} \sum_{i<j} \langle \sigma_i \sigma_j \rangle^2$ must essentially depend upon the randomness. (In contrast, in the region D, this quantity is asymptotically close to $q^2/2$.) The basis for the argument is the fact that the random convex function F_N/N has fluctuations of order $N^{-1/2}$, so that for most values of β, $F'_N(\beta)/N$ has only small fluctuations. One then computes the variance of this quantity using integration by parts, assuming that q' has vanishing fluctuations, and this yields the information that $E\langle |(\frac{\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2}{N})^2 - \langle (\frac{\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2}{N})^2 \rangle | \rangle \to 0$. This is not as good as (2.2), but is sufficient to prove that (3.13) would hold, which is known to be wrong for large β as proved in [C].

The idea for the first part of the above argument is in germ in [A-L-R] . This line of arguments is exquisitely developed by F. Guerra [Gu2]. Using only integration by parts and the fact that $EF''_N(\beta)/N$ is non negative and of order 1 for most β, he shows that for most values of β,

$$E\langle (\frac{\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2}{N})^4 \rangle - 4E\langle (\frac{\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2}{N})^2 (\frac{\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^3}{N})^2 \rangle + 3E\langle (\frac{\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2}{N})^2 \rangle^2 \simeq 0. \tag{3.18}$$

It is explained in [A-C1] why this is less miraculous than it seems at first sight.

4. The Hopfield Model

The Hopfield model was introduced by Pastur and Figotin [P-F] in the spin glass context, but became famous only after Hopfield interpreted it as a model for memory. We will refer the reader to [H1], [H2], [H-K-P], [T-D-C] for this aspect of the model, and we will directly turn towards the underlying mathematics. The model involves N spins, and M configurations "to be memorized" $(\boldsymbol{\eta}_k)_{k \le M}$, where $\boldsymbol{\eta}_k = (\eta_{i,k})_{i \le N}$. These configurations are called the *prototypes* and are chosen at random in the simplest possible manner, independently, with $P(\eta_{i,k} = 1) = P(\eta_{i,k} = -1) = \frac{1}{2}$. The object of interest is the function on Σ_N defined by

$$H_{N,M}(\boldsymbol{\sigma}) = -\frac{N}{2} \sum_{k \le M} m_k(\boldsymbol{\sigma})^2 \tag{4.1}$$

where

$$m_k(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i \le N} \sigma_i \eta_{i,k} \tag{4.2}$$

is the overlap between $\boldsymbol{\sigma}$ and $\boldsymbol{\eta}_k$. The normalizing factor $N/2$ will be pleasant when we will use a temperature; one way to look at H_N is that it is among the simplest functions one can write that is a candidate to take a large negative value when $\boldsymbol{\sigma} = \boldsymbol{\eta}_k$ (since $m_k(\boldsymbol{\eta}_k) = 1$). We will study the Hopfield model only at $N \to \infty$. There are different regimes of growth of $M = M(N)$ that are of interest; we will consider here only the most challenging one, when $M = [\alpha N]$ is a proportion of

N. (We consider α as fixed and no longer write the dependence in M.) Not surprisingly, the smaller α is, the easier the model is.

Even though this is not our main line of interest, we will say a few words about the "zero temperature case", that is the study of the function H_N itself. The rigorous results concerning that case are not sharp, and often obtained by ad-hoc methods that cannot yield optimal results; but at least they exist.

It is believed that for large N (and with an overwhelming probability) for $\alpha \leq .13$ there is an "energy barrier" around each prototype; that is, for some $\delta > 0$, $\epsilon > 0$

$$\inf_{d(\boldsymbol{\sigma},\boldsymbol{\eta}_k)=\delta N} H_N(\boldsymbol{\sigma}) > H_N(\boldsymbol{\eta}_k) + \Sigma N$$

where $d(\boldsymbol{\sigma},\boldsymbol{\eta}_k)$ is the number of indexes $i \leq N$ such that $\sigma_i \neq \eta_{i,k}$. This was proved for $\alpha \leq 0.05$ by C. Newman [N], $\alpha \leq 0.07$ by D. Loukianova [Lou 1] and can be further improved [T6, Section 9]. Let us say that a configuration $\boldsymbol{\sigma}$ is a local minimum if the value of $H_{N,M}(\boldsymbol{\sigma})$ cannot be decreased by changing the sign of one single spin. (The importance of these is that they can be thought as the configurations "memorized" by $H_{N,M}$.) Possibly the prettiest proof is due to Loukianova [Lou 2], who shows, that, as $\alpha \to \infty$, the function H_N cannot have a local minimum anywhere close to a prototype. However, nagging questions remain. In particular, it is believed that for $\alpha = .1$, $H_{N,M}$ has a local minimum near each prototype, but a lower global minimum. (This lower global minimum is believed not to be simply related to any prototype, and does not seem to be accessible by any explicit algorithm.) Our inability to deal rigorously with this question takes its root in the fact that, while we know, at least in principle, how to calculate the order of (the expected value of) the supremum of a gaussian process (see [T1]) we do not know how to do this, say, within 10% (or even a factor 2).

It is natural to study the function H_N through the introduction of a temperature $T = 1/\beta$, and to study the corresponding Gibbs measure (that gives weight $Z^{-1} \exp -\beta H_N(\boldsymbol{\sigma})$ to $\boldsymbol{\sigma}$, where Z is the normalizing factor). The results of the study through the replica formalism are presented in [A-G-S].

The region $\beta(1+\sqrt{\alpha}) < 1$ corresponds to the case $h = 0, \beta < 1$ of the SK model. In that region we have

$$\lim_{N\to\infty} \frac{1}{N} E \log Z_N = \lim \frac{1}{N} \log EZ_N (= \frac{\alpha}{2} \log \frac{1}{1-\beta}). \tag{4.3}$$

As we already mentioned, this seems to make things simpler and this region is rather well understood [T6, Section 2]. The situation can be physically described by saying that the temperature is so high that nothing can be learned about the prototypes by studying Gibbs measure.

At $\beta = 1$, there seems to be an instability that has yet to be analyzed, so we will consider directly the case $\beta > 1$. Important rigorous work has been done in that case by A. Bovier and V. Gayrard (sometimes jointly with P. Picco) [B], [B-G1, 2, 3, 4], B-G-P1, 2]. These authors have in particular been interested in the image of $\overline{G}$ on $\mathbb{R}^M$ of Gibbs measure under the map $\boldsymbol{\sigma} \to (m_k(\boldsymbol{\sigma}))_{k\leq M}$. It is very natural to consider this measure since $H_N(\boldsymbol{\sigma})$ is defined in function of the overlaps $m_k(\boldsymbol{\sigma})$ only. They proved that if $\alpha \leq L^{-1}\min(1,(\beta-1)^2)$ (where L is a universal constant) then $\overline{G}$ is essentially supported by the union of $2M$ disjoint balls of $\mathbb{R}^M$. These balls are centered at the points $m^*\boldsymbol{e}_k$, where $(\boldsymbol{e}_k)$ is the canonical basis of $\mathbb{R}^M$ and $m^* = \operatorname{th}\beta m^*$. This spontaneous decomposition of $\overline{G}$ into "states" simply reflects the strong influence of each prototype on the Hamiltonian. Much more precise information on the structure of $\overline{G}$ is contained in Theorem 4.1 below, so we will not state the results of [B-G1] in detail, but beside the intrinsic interest of these results, it must be pointed out that this a priori information of $\overline{G}$ is essential for the use of the cavity method.

Since $\overline{G}$ (and hence G) breaks into rather unrelated pieces, it is quite natural to study these separately. One way (introduced in [A-G-S]) to do this is to replace the Hamiltonian (4.1) by

$$H_N(\boldsymbol{\sigma}) = -\frac{N}{2}\sum_{k\le M} m_k(\boldsymbol{\sigma})^2 - hNm_1(\boldsymbol{\sigma}) \tag{4.4}$$

where $h > 0$ (and small). The effect of the extra term is to favor the part of $\overline{G}$ close to $\boldsymbol{e}_1 m^*$ over the parts close to $-\boldsymbol{e}_1 m^*$ or $\pm\boldsymbol{e}_k m^*, k \ge 2$.

To state our main result, we consider the domain

$$D = \{(\alpha,\beta,h); \beta > 1, \alpha \le \frac{1}{L}\min((\beta-1)^2, \frac{1}{\log\beta}); 0 < h < h(\alpha,\beta)\} \tag{4.5}$$

where L is a (suitably large) number and $h(\alpha,\beta)$ is positive (and suitably small). The condition upon h means that we are interested only in the case of h very small; the results can be extended to the case of any $h > 0$ with some extra effort; on the other hand the requirement on α is essentially the best possible. It should be pointed out that the region D is a part of what is usually called the low temperature region, but the behavior there is typically high temperature ("replica-symmetry").

We consider the system of equations

$$\mu = E\operatorname{th}\beta(g\sqrt{r} + \mu + h) \tag{4.6}$$
$$q = E\operatorname{th}^2\beta(g\sqrt{r} + \mu + h) \tag{4.7}$$
$$r = \alpha q(1-\beta(1-q))^{-2} \tag{4.8}$$

It can be shown that if $(\alpha,\beta,h) \in D$ (and the constant L of (4.5) is large enough), this system of equations has a unique solution. We consider the function $\varphi(x) = \min(x, x^2)$. The somewhat complicated inequalities (4.9) to (4.13) mostly intend to convey the message that great accuracy can be reached, and need not be understood in detail by the casual reader.

THEOREM 4.1. *For each value of (α,β,h) in D there exists a number K indepen-*

dent of N with the following properties

$$E\langle \exp \frac{1}{KN}(\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2 - Nq)^2\rangle \leq K \tag{4.9}$$

$$E\langle \exp \frac{N}{K}\varphi(\sum_{2\leq k\leq M} m_k(\boldsymbol{\sigma}^1)m_k(\boldsymbol{\sigma}^2) - r)\rangle \leq K \tag{4.10}$$

$$E\langle \exp \frac{N}{K}(m_1(\boldsymbol{\sigma}) - \mu)^2\rangle \leq K \tag{4.11}$$

$$\forall k \geq 2, E\langle \exp \frac{N}{K} m_k^2(\boldsymbol{\sigma})\rangle \leq K \tag{4.12}$$

$$E\langle \exp \frac{N}{K}\varphi(\sum_{2\leq k\leq M} m_k^2(\boldsymbol{\sigma}) - \frac{1-\beta(1-q)^2}{(1-\beta(1-q))^2})\rangle \leq K \tag{4.13}$$

$$E\langle \exp \frac{N}{K}\varphi(\sum_{2\leq k\leq M} (m_k(\boldsymbol{\sigma}) - \langle m_k(\boldsymbol{\sigma})\rangle)^2 - \frac{\alpha(1-q)}{(1-\beta(1-q))^2})\rangle \leq K \tag{4.14}$$

Moreover, for any $n > 0$, the r.v. $\langle\sigma_1\rangle, \cdots, \langle\sigma_n\rangle$ are asymptotically independent; their limit law is the law of $\operatorname{th}\beta(g\sqrt{q} + \mu + h)$ *where g is standard normal.*

The use of the function φ rather than x^2 is motivated by problems with the very large values; the reason why $m_1(\boldsymbol{\sigma})$ plays a special role should be obvious from (4.4). The meaning of (4.9) is that the measure $\overline{G}'$ image of Gibbs' measure under the map $\boldsymbol{\sigma} \to (m_k))_{2\leq k\leq M}$ "is in a pure state". The meaning of (4.14) is that this measure is nearly carried by a small shell around the sphere of center $\boldsymbol{b} = (\langle m_k(\boldsymbol{\sigma})\rangle)_{2\leq k\leq M}$ and of radius $(\alpha(1-q))^{1/2}/1-\beta(1-q)$; and (4.13) implies that $\|\boldsymbol{b}\|^2$ is nearly r. We thus have very accurate information on $\overline{G}'$. We consider now the function

$$\begin{aligned} RS(\alpha,\beta,h) = &-\frac{\mu^2\beta}{2} + \frac{\alpha}{2}(\frac{\beta q}{1-\beta(1-q)} - \log(1-\beta(1-q)) \\ &-\beta^2\frac{r}{2}(1-q) + E\log\operatorname{ch}\beta(g\sqrt{r} + \mu + h) \end{aligned} \tag{4.15}$$

where of course r, q, μ are solutions of (4.6) to (4.8).

Theorem 4.2. *If the parameters (α,β,h) belong to D, then*

$$\lim_{N\to\infty} N^{-1}EF_N(\alpha,\beta,h) = RS(\alpha,\beta,h). \tag{4.16}$$

Theorem 4.2 was first proved in [T6] in the smaller domain $D_1 \subset D$ where, for $\beta \geq 2$, the condition $L\alpha\log\beta \leq 1$ is replaced by the stronger constraint $L\alpha\beta \leq 1$. Equality (4.16) extends by continuity to the case $h = 0$. The proof of [T6] uses the (somewhat unsatisfactory) technique of adding an appropriate small perturbation term to the Hamiltonian (4.4), a trick that produces magical and mysterious results. This perturbation term is $\gamma\varphi(N)\sum_{k\leq M} g_k N m_k(\boldsymbol{\sigma})$, where $\varphi(N)$ goes to zero but not too fast (say $\varphi(N) = N^{-1/3}$), where g_k are i.i.d. $N(0,1)$, and where $0 \leq \gamma \leq 1$ (say). As $N \to \infty$, for fixed γ the extra term has a vanishing influence on the expected free energy density (because $\varphi(N) \to 0$). On the other hand, differentiation of $N^{-1}F_N$ with respect to γ as in [Gu2] leads to precious information. This information comes essentially "for free", a miraculous fact that it would be nice to really understand. Upon reading [T6], Bovier and Gayrard [B-G3] discovered a very beautiful approach (in the same smaller domain D_1) that deals directly with the Hamiltonian (4.4), and where the fact that the system is in a pure state follows from a transparent geometric property. Unfortunately this property is not true in the entire domain D of (4.5). One can only hope that their approach can be modified to cover the correct domain D and is not specific to this particular model.

It is possible to explain some of the mystery of the formula (4.15). If we consider the right-hand side of (4.15) as a function of independent variables $\alpha, \beta, h, q, r, \mu$, equations (4.6) to (4.8) mean that the partial derivatives of this function with respect to μ, q, r respectively are zero, so that even though these depend upon α, the partial derivative of $RS(\alpha, \beta, h)$ with respect to α can be computed as if it were not the case. One simply has to check that this partial derivative coincides with the increase of expected free energy when M is replaced by $M+1$, that is

$$E \log\langle \exp \beta \big(\sum_{i \leq N} \eta_{i,M+1} \sigma_i \big)^2 \rangle$$

which is calculated with (2.9) showing first that the variables $\eta_{i,M+1}$ can be replaced by independent gaussian (although non trivial technicalities arise due to lack of boundedness).

THEOREM 4.3. *If a function W on a p replica is a product of finitely many expressions of one of the following types:*

$$N^{-1/2}(\boldsymbol{\sigma}^\ell \cdot \boldsymbol{\sigma}^{\ell'} - E\langle \boldsymbol{\sigma}^\ell \cdot \boldsymbol{\sigma}^{\ell'} \rangle) \tag{4.17}$$

$$N^{1/2}\Big(\sum_{2 \leq k \leq M} m_k(\boldsymbol{\sigma}^\ell) m_k(\boldsymbol{\sigma}^{\ell'}) - E\langle \sum_{2 \leq k \leq M} m_k(\boldsymbol{\sigma}^\ell) m_k(\boldsymbol{\sigma}^{\ell'}) \rangle \Big) \tag{4.18}$$

$$N^{1/2}(m_k(\boldsymbol{\sigma}^\ell) - E\langle m_k(\boldsymbol{\sigma}^\ell) \rangle)(k \geq 1) \tag{4.19}$$

then $\lim_{N \to \infty} E\langle W \rangle$ *exists.*

This theorem is proved by an explicit method allowing in principle explicit computation of the limits. Only remains the uninspiring (and in principle elementary) task to clarify the underlying algebraic structure. Motivated by [B-G4] (that considers the case $\alpha = \alpha(N) \to 0$) we did check that given any n, the laws under Gibbs measure of the maps $\boldsymbol{\sigma} \to N^{1/2} m_k(\boldsymbol{\sigma}) (2 \leq k \leq n)$ are asymptotically i.i.d. gaussian centered, of variance $(1 - \beta(1-q)^2)(1 - \beta(1-q))^{-2}$.

Let us now outline the main aspect in which the proofs differ from the case of the SK model. When one tries to compute a quantity such as $D_N = E\langle (\tilde{\boldsymbol{\sigma}} \cdot \boldsymbol{\sigma}^3)^2 \rangle$ in function of D_{N-2}, by regrouping in the Hamiltonian the terms containing $\sigma_N^\ell, \sigma_{N-1}^\ell (\ell \leq 3)$ one rather finds terms such as A_{N-2}, where

$$A_N = E\langle \big(\sum_{2 \leq k \leq M} \tilde{\boldsymbol{m}}_k \cdot \boldsymbol{m}_k^3 \big)^2 \rangle \tag{4.20}$$

(again with a small change of temperature) where $\tilde{\boldsymbol{m}}_k = m_k(\boldsymbol{\sigma}^1) - m_k(\boldsymbol{\sigma}^2), \boldsymbol{m}_k^3 = m_k(\boldsymbol{\sigma}^3)$, and to obtain a useful relation one would have to relate such terms to terms such as D_{N-2}. Since it is not obvious how to do this, we chose the alternate strategy to show first that A_N is small. To relate A_N with A_{N-1}, the first step is to isolate in $\tilde{m}_k, m_k^3$ the contribution of $\tilde{\sigma}_N, \sigma_N^3$. After expansion, one faces dangerous sums of the type $E \sum_k \eta_{k,N} \langle f_k \rangle$. In these terms f_k is not small, but does not depend upon $\eta_{k,N}$. Cancellation occurs because the bracket $\langle \cdot \rangle$ depends only weakly upon $\eta_{k,N}$; this is expressed by an extension of the integration by parts formula (3.6) to Bernoulli r.v. (with now an error term). After integration by parts the various terms can then be related to a $N-1$ spin situation via the scheme (3.8). The only drawback of this approach is that integration by parts creates numerous terms, and from each of these (3.8) creates numerous new terms, so that the computations soon reach gargantuan proportions; but once one has learned how to identify the leading terms, all it really takes to go through them is a few weeks of patience.

5. Intersecting random half spaces: The capacity of the Perceptron

The problem to be discussed in this section originates in the theory of neural networks. Its basic nature makes it however of interest well beyond this theory, and the reader interested in neural networks is referred to [G2], [H-K-P]. We will consider random half spaces in $\mathbb{R}^N$ that are at a given distance from the origin. The random direction will be modeled by a sequence $\boldsymbol{\xi} = (\xi_i)_{i \leq N}$ of r.v. with $P(\xi_i = 1) = P(\xi_i = -1) = 1/2$. This choice (rather than the most canonical choice of gaussian r.v.) is motivated by the origin of the problem. The same result (often quite easier) can be obtained in the Gaussian case. Given a number κ, we consider the half space $H(\boldsymbol{\xi}) = \{\boldsymbol{x} \in \mathbb{R}^N ; \boldsymbol{\xi} \cdot \boldsymbol{x} \geq \kappa\sqrt{N}\}$: Given independent choices $\boldsymbol{\xi}^1, \cdots, \boldsymbol{\xi}^M$ of random directions, we would like to know whether typically $\bigcap_{k \leq M} H(\boldsymbol{\xi}^k)$ meets Σ_N. If λ_N denotes the homogeneous probability on Σ_N, when $\kappa = 0$ (the most important case) and (to avoid minor complications) N is odd, it is trivial that $E\lambda_N(\bigcap_{k \leq M} H(\boldsymbol{\xi}^k)) = 2^{-M}$, and this shows that if $M > (1+\epsilon)N$ the answer is no. It is proved in [K-R] that there is $\epsilon > 0$ such that for large N the set $\bigcap_{k \leq M} H(\boldsymbol{\xi}^k)$ typically meets Σ_N if $M \leq \epsilon N$, but not if $M \geq (1-\epsilon)N$, a result that is somewhat streamlined and improved in [T13]. It is conjectured in [K-M] that the critical value of M is about $M = .83N$. One would like to compute exactly (in the limit) the "typical value" of $N^{-1} \log \lambda_N(\bigcap_{k \leq M} H(\boldsymbol{\xi}^k))$ (the mean is not defined since $\bigcap_{k \leq M} H(\boldsymbol{\xi}^k)$ can be empty). There is an obstacle to the study of a quantity such as $\bigcap_{k \leq M} H(\boldsymbol{\xi}^k)$, namely that the size of this set is extremely dependent upon *each* direction $\boldsymbol{\xi}^k$ (e.g. the set is empty if $\boldsymbol{\xi}^M = -\boldsymbol{\xi}^1, \kappa > 0$). Of course one expects that "in general configurations" this is not the case, but showing this requires works. It does not seem even trivial to show that the random quantity

$N^{-1}\log\lambda_N(\bigcap_{k\le M} H(\boldsymbol{\xi}^k))$ has small fluctuations around its median value, and this despite a well developed machinery that has been constructed to handle such problems [T2], [T4]. It is currently not known how to show that these fluctuations are of order $N^{-1/2}$, as one should expect (see [T9] for a weaker result). It is thus natural to study first a version of the problem "with temperature", by considering the Hamiltonian

$$H_{N,M}(\boldsymbol{\sigma}) = -\sum_{k\le M}\theta\Big(\frac{\boldsymbol{\xi}^k\cdot\boldsymbol{\sigma}}{\sqrt{N}}\Big) \tag{5.1}$$

where $\theta(x) = 1_{\{x\ge\kappa\}}$. One will then consider the corresponding Gibbs measure G_N at inverse temperature β. When $\theta(x) = x^2/2$ (5.1) is the Hamiltonian of the Hopfield model; but the fact that θ is now bounded suppresses the strong attraction of the system towards the configurations $\boldsymbol{\xi}^k$.

Given a function θ, and $\beta>0$, we consider the function (defined for $y<1$)

$$\Phi(x,y) = \frac{1}{\sqrt{1-y}}\frac{Eg\exp\beta\theta(x+g\sqrt{1-y})}{E\exp\beta\theta(x+g\sqrt{1-y})} \tag{5.2}$$

where g is $N(0,1)$. In the next statement, z also denotes a $N(0,1)$ variable independent of g, and E_g denotes integration in g only.

THEOREM 5.1. *Given $\beta>0$, there exists a number $\alpha_0(\beta)>0$ with the following property. Consider a nondecreasing function $\theta:\mathbb{R}\to[-1,1]$, and the function Φ given by (5.2). Then, if $\alpha\le\alpha_0(\beta)$ the system of equations*

$$q = E\,\mathrm{th}^2(z\sqrt{\hat q});\, \hat q = \alpha E\Phi^2(z\sqrt{q},q) \tag{5.3}$$

has a unique solution $q = q(\alpha,\theta,\beta), \hat q = \hat q(\alpha,\theta,\beta)$. Moreover, if $Z_{N,M}$ denotes the partition function of the system governed by the Hamiltonian (5.1) at inverse temperature β, we have

$$\lim_{N\to\infty}\frac{1}{N}E\log Z_{N,M} = RS(\alpha,\beta) \tag{5.4}$$

when $M=[\alpha N]$ and

$$\begin{aligned} RS(\alpha,\beta) = &-\frac{1}{2}\hat q(1-q) + E\log 2\,\mathrm{ch}\, z\sqrt{\hat q} \\ &+ \alpha E\log E_g\exp\beta\theta(z\sqrt{q}+g\sqrt{1-q}). \end{aligned} \tag{5.5}$$

It is of interest to compare this formula with the corresponding formula for the Hopfield model. When $\theta(x) = x^2, \Phi$ is well defined for $\beta(1-y)<1$, and the second equation of (5.3) becomes $\hat q = \alpha\beta^2 q(1-\beta(1-q))^{-2}$. Then (5.5) gives the formula (4.15) in the case $\mu = h = 0$.

The reader has noted that Theorem 5.1 does *not* require that θ be smooth. On the other hand, we do not know how to relate an N spin system with an

$(N-1)$ spin system unless θ is smooth and we can make power expansions. To prove Theorem 5.1, we first assume that θ is smooth, and we use the monotonicity of $Z_{N,M}$ in θ. With this approach, it is not clear how to prove (2.2), or even whether this is true when θ is an indicator function. The difficulty is a problem of interversion of limits. The useful estimates when θ is smooth require N large, where "large" seems to depend on how large the derivative of θ can be.

When relating an N spin system with an $(N-1)$ spin system, the role that was played by the quantities m_k in the case of the Hopfield model is now played by $\theta'(s_k)/\sqrt{N}$, where $s_k = N^{-1/2} \sum_{i \le N} \xi_i^k \sigma_i$. A first observation is that $\sum_{k \le M} (\theta'(s_k)/\sqrt{N})^2$ (among other quantities) will not be bounded by a quantity depending upon $\|\theta\|_\infty = \sup |\theta|$ only. In order to be able to prove Theorem 5.1, we must make estimates that (for large N) do not depend on $\|\theta'\|_\infty$ but only on $\|\theta\|_\infty$; not surprisingly, the main tool for that purpose is integration by parts. A second observation is that we no longer benefit as in the Hopfield case from the fact that $\theta'(x) = \beta x$ is a very simple function. This made possible (through integration by parts) to relate quantities such (4.20) (quantifying that the image of Gibbs measure under the map $\boldsymbol{\sigma} \to (\theta'(s_k)/\sqrt{N})_{k \le M}$ is nearly in a pure state) with quantities such as $E\langle(\frac{1}{N}\tilde{\boldsymbol{\sigma}} \cdot \boldsymbol{\sigma}^3)^2\rangle$ that involve only configurations. As a substitute to these explicit evaluations, we use another version of the cavity method (that we learned in [M]), which relies on the simple observation that for any function f,

$$\langle f(s_k) \rangle = \frac{\langle f(s_k) \exp \beta\theta(s_k) \rangle_1}{\langle \exp \beta\theta(s_k) \rangle_1} \tag{5.6}$$

where $\langle \cdot \rangle_1$ denotes Gibbs relative to the Hamiltonian $H_{N,M-1}$ of (5.1) (thus the summation is over $k \le M-1$). In order to compute expectation of the right hand side of (5.6) (and of the similar quantities required to work with several replicas) one first integrates in $\boldsymbol{\xi}^k = (\xi_i^k)_{i \le N}$. To do this one shows first that we can replace the ξ_i^k by i.i.d. $N(0,1)$ variables. One then uses a decomposition of the type (3.8), where now $Z = \langle \exp \beta\theta(s_k) \rangle_1$ and where $\hat{Z}$ (motivated by (2.9)) is

$$\hat{Z} = E_g \exp \beta\theta(\boldsymbol{\xi}^k \cdot \boldsymbol{b} + g\sqrt{1 - \|\boldsymbol{b}\|^2})$$

for $\boldsymbol{b} = (\langle \sigma_i \rangle_1 / \sqrt{N})_{i \le N}$. In these computations, we are not dealing with explicit functions (and thus cannot make explicit computations); instead we obtain estimates through comparison theorems for Gaussian processes.

6. The random p-sat problem

Consider independent Boolean variables $x_1, \cdots, x_N$. A p-clause is a Boolean function $y_{i_1} \vee y_{i_2} \vee \cdots \vee y_{i_p}$ where $i_1 < \cdots < i_p$, and where, for each $\ell \le p$, either $y_{i_\ell} = x_{i_\ell}$ or $y_{i_\ell} = \overline{x}_{i_\ell}$. Thus there is exactly one truth assignment of the variables $(x_{i_\ell})_{\ell \le p}$ that does not satisfy the clause. Given M clauses, the satisfiability problem is the question of whether or not there is a truth assignment of the variables that satisfies them all. It is a fundamental problem of computer science. In the random model of the p-sat problem the set of M clauses is chosen independently uniformly among all sets of Mp-clauses. The question is then to decide whether in the typical case these M random clauses can be simultaneously satisfied, and, more generally, what is the typical proportion of truth assignments that will satisfy them all. To see the relation with previous sections, we replace "true" by 1 and "false" by -1. We denote by $[N]^p$ the collection of subsets of $\{1, \cdots, N\}$ of cardinal p. Given $I \in [N]^p$ and $\boldsymbol{\rho} \in \Sigma_N$, we consider the set

$$A_{I,\boldsymbol{\rho}} = \{\boldsymbol{\sigma} \in \Sigma_N; \exists i \in I, \sigma_i \neq \rho_i\} \tag{6.1}$$

and the problem is now to find the typical proportion of configurations that belong to M random sets $A_{I,\boldsymbol{\rho}}$. This problem is formally very close to the perceptron capacity problem of Section 6. The big difference is that the random sets depend only upon finitely many coordinates; but as previously the important case is when $M = \lfloor \alpha N \rfloor$.

In order to introduce a temperature, we consider (following [M-Z]) the Hamiltonian

$$H_N(\boldsymbol{\sigma}) = -\sum_{k \leq M} 1_{\{\boldsymbol{\sigma} \in A_k\}} \tag{6.2}$$

where $A_k, k = 1, \cdots, M$ are M sets of the type (6.1) chosen uniformly among all possibilities.

In order not to be hypnotized by the specific form of (6.2), we consider a more general setting, as follows. Consider a function $f : [0,1] \times \{-1,1\}^p \to [-1,1]$. For each set $I \in [N]^p$ consider the random function $f_I(\boldsymbol{\sigma}) = f(X_I, \sigma_{i_1}, \cdots, \sigma_{i_p})$ where $I = \{i_1 < \cdots < i_p\}$ and where the collection $(X_I)_{I \in [N]^p}$ is independently uniform over $[0,1]$. We then consider the more general form of (6.2)

$$H_N(\boldsymbol{\sigma}) = -\sum_{k \leq M} f_{I_k}(\boldsymbol{\sigma}) \tag{6.3}$$

and the corresponding random Gibbs measure G_N on Σ_N. The expected number of intervals I_k that contain N is pM/N, so that the conditional distribution of σ_N (for Gibbs' measure) given $\sigma_1, \cdots, \sigma_{N-1}$ depends of $(\sigma_1, \cdots, \sigma_{N-1})$ through only finitely many components, a fact that is expressed in physics by saying that the N^{th} site interacts with finitely many other sites. We cannot expect the central limit theorem to come into effect, and the gaussian r.v. that were ubiquitous in the previous sections will not appear here. This makes the situation more complicated. The formal computations of the physicists that lead then to (e.g.) (4.15) make it natural for them to think of the Hopfield model as depending upon these parameters (μ, q, r), that are determined by the relations (4.6) to (4.8). They say that the system "depends on the order parameters μ, q, r". The situation is more complex here, and the central object is the limiting distribution of $\langle \sigma_1 \rangle$, a fact expressed in physics by saying that "the order parameter of the system is a function". (In that case the replica formalism involves yet another arbitrary step. Namely, one has to look for the extremum of a certain functional over a very large function space, and one restricts a priori the search to a more manageable very small subspace.)

THEOREM 6.1. *Given the integer p, and $\alpha > 0$, there exists a number $\beta(p, \alpha) > 0$ with the following property. Whenever $\beta < \beta(p, \alpha)$, the system governed by (6.3) at inverse temperature β is in a pure state. Given any n, the r.v. $\langle\sigma_1\rangle, \cdots, \langle\sigma_n\rangle$ are asymptotically identically distributed and the expected free energy density converges as $N \to \infty$.*

The limit law $\nu = \nu(f, \beta, \alpha)$ of $\langle\sigma_1\rangle$ appears as the fixed point of a certain operator (in the spirit of the previous sections). The limiting expected free energy density can be in principle computed in function of ν (see [M-Z] for a rather formal expression, obtained through the replica formalism, in the case of (6.2)).

To prove Theorem 6.1, the main difficulty is to prove the conditions of Theorem 2.1. The positivity argument is very precious here because, if one tries an approach along the lines of (3.8), the natural candidate for $\hat{Z}$ is complicated enough so that it is not clear how to estimate simply $EU/\hat{Z}$. The statement about the limiting behavior of $\langle\sigma_1\rangle, \cdots, \langle\sigma_n\rangle$, which, as we explained, is an essentially obvious consequence of the conditions of Theorem 1.2 in the previous examples lies somewhat deeper here. The basic idea is however simple. The last spin σ_N interacts with only finitely many other spins. Each of these in turn interacts only with finitely many other spins, etc. The key point is that the (global) influence upon σ_N of the finitely many spins obtained at the k-stage decreases with k, so that the behavior of σ_N is essentially controlled by a finite set of other spins. When applying the same principle to σ_{N-1}, another finite set of spins is involved, that is generically disjoint of the previous one, and this creates independence. The reader has noticed that the role of α and β are reversed in Theorem 6.1 compared to Theorem 5.1. It is true that given f, β, the conclusion of Theorem 6.1 does hold for small α, but for uninteresting reasons. In fact if $\alpha(p-1) < 1$, with high probability the interactions "die out" and the set $\{1, \cdots, N\}$ decomposes in small pieces that do not interact with each other.

We have given Theorem 6.1 as an illustration of the fact that even the case of "functional order parameter" is amenable to rigorous results because it relates to a known famous problem. There are, however, simpler situations of the same nature. One of them is the diluted SK model, where the Hamiltonian (1.3) is replaced by

$$H_N(\boldsymbol{\sigma}) = -\sum_{i<j} \eta_{ij} g_{ij} \sigma_i \sigma_j - h \sum_{i<j} \sigma_i. \tag{6.4}$$

There, the r.v. η_{ij} are independent among themselves and of the g_{ij}, and satisfy $P(\eta_{ij} = 1) = \gamma/N$, $P(\eta_{ij} = 0) = 1 - \gamma/N$ so that each spin interacts with an average number of γ other spins. A result similar to Theorem 6.1 can be proved at high temperature. The proof is much easier because (2.2) can be obtained through an immediate adaptation of the argument we outlined in Section 3.

7. THE p-SPIN INTERACTION MODEL: LOW TEMPERATURE

The p-spin model is a generalization of the SK model. If p is an integer ≥ 2, the

Hamiltonian is

$$H(\boldsymbol{\sigma}) = -\big(\frac{p!}{2N^{p-1}}\big)^{1/2} \sum_{1\le i_1<\cdots<i_p\le N} g_{i_1\cdots i_p}\sigma_{i_1}\cdots\sigma_{i_p}. \tag{7.1}$$

The summation is over all possible choices of $i_1,\cdots,i_p$ and the $g_{i_1\cdots i_p}$ are i.i.d. standard normal. A basic observation is that (neglecting terms of order one)

$$2EH(\boldsymbol{\sigma})H(\boldsymbol{\sigma}') = \frac{p!}{N^{p-1}} \sum_{i_1<\cdots<i_p} \sigma_{i_1}\cdots\sigma_{i_p}\sigma'_{i_1}\cdots\sigma'_{i_p} \simeq N\big(\frac{\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}'}{N}\big)^p. \tag{7.2}$$

Thus the complicated covariance structure of the gaussian variables $(H(\boldsymbol{\sigma}))_{\boldsymbol{\sigma}}$ (that is responsible for the difficulty of the problem) simplifies as $p\to\infty$ and the r.v. $(H(\boldsymbol{\sigma}))_{\boldsymbol{\sigma}}$ become independent, a situation that can be analyzed in great detail [D]. We are however not interested in having N fixed, $p\to\infty$, but rather p fixed, $N\to\infty$. Still, (7.2) indicates that the larger p, the easier the model should be. Physics predicts that for low (but not too low) temperature, the behavior of the model is non trivial, yet much simpler than the conjectured behavior of the SK model [G-M], [G1]. The different behavior starts at $p=3$ for reasons that will soon be obvious.

The basic idea to obtain information about the low temperature region is to use the "transfer principle" outlined in Section 3. This principle allows only a small change of inverse temperature, so that in order to reach the low temperature region we must first be able to control most of the high temperature region, which we know best how to do when there is no external field (our results can be extended to small external field, say h of order 2^{-p}, but we do not see how to handle the case where h is not small, say, $h=1$). But what is the high temperature region? Let us define the critical number β_p as the supremum of the numbers β for which

$$\lim_{N\to\infty}\frac{1}{N}E\log Z_N = \lim_{N\to\infty}\frac{1}{N}\log EZ_N \big(=\frac{\beta^2}{4}\big). \tag{7.3}$$

We do not know the exact value of β_p if $p>2$, but we proved that $2\sqrt{\log 2}-2^{-p} < \beta_p < 2\sqrt{\log 2}$ for large p. To obtain information about the range of the overlaps at high temperature, the idea is as follows. We write, for an interval I,

$$G_N^2(\{(\boldsymbol{\sigma}^1,\boldsymbol{\sigma}^2);\boldsymbol{\sigma}^1\cdot\boldsymbol{\sigma}^2/N\in I\}) = Z_N^{-2}\sum_{\boldsymbol{\sigma}^1,\boldsymbol{\sigma}^2}\exp -\beta(H_N(\boldsymbol{\sigma}^1)+H_N(\boldsymbol{\sigma}^2))$$

where the summation is over $\boldsymbol{\sigma}^1\cdot\boldsymbol{\sigma}^2/N\in I$. We control EZ_N from (6.3), and we use that from general principles (concentration of measure [I-S-T]) Z_N is very close from its expectation so that it can be controlled from below. To control the summation from above, we then estimate

$$E\sum\exp -\beta(H(\boldsymbol{\sigma}^1)+H(\boldsymbol{\sigma}^2)) = \sum\exp\frac{\beta^2}{2}E(H(\boldsymbol{\sigma}^1)+H(\boldsymbol{\sigma}^2))^2$$

using (7.2), where now appears the importance of the exponent p, contrasted with the fact that the proportion of configurations $\boldsymbol{\sigma}^1,\boldsymbol{\sigma}^2$ for which $\boldsymbol{\sigma}^1\cdot\boldsymbol{\sigma}^2 = Nt$ is about $\exp -Nt^2/2$. (In practice some technicalities like truncation are required). To avoid complications, let us give a typical result (which is not the best we can prove).

Theorem 7.1. *There exists a number L with the following property. If $p \geq L, x \leq 1/L$ and if $\beta \leq \beta_p + x/L$, then the overlap of two replicas essentially never belongs to the set $J \cup -J$ where $J = [x, 1-(x+2^{-p/L})]$.*

There (and below) "essentially never belongs" means

$$EG_N^2(\{\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2; \boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2/N \in J \cup -J\}) \leq \exp -N/K$$

where K does not depend upon N.

Let us now consider a probability ν on the sphere S_N of $\mathbb{R}^N$ of radius $\sqrt{N}$ (so that $\Sigma_N \subset S_N$) such that the overlap $\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2/N$ of two independent configurations belong to $J \cup -J$ only with very small probability where J is, say, the interval $[.01, .99]$. (Thus, for p large enough and $\beta \leq \beta_p + 1/L, G_N$ has this property with overwhelming probability). Then it is intuitively clear, and easy to prove, that almost all the mass ν must be carried by a union $\bigcup_{\alpha \geq 1} C_\alpha$ of sets C_α such that each C_α is the union of two opposite small caps. The decomposition is finite ($C_\alpha = \emptyset$ for α large enough). It is such that when $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ belong to two different sets C_α, then $|\boldsymbol{\sigma} \cdot \boldsymbol{\sigma}'/N|$ is small (say, $\leq 1/10$). The fact that C_α has to be the union of two pieces is clear when ν is invariant by symmetry about zero, as is the case for Gibbs measure when p is even. Moreover, the decomposition is essentially unique "within sets of small measure".

Thus, Theorem 7.1 proves that if $\beta \leq \beta_p + \beta_0$ (where β_0 is a fixed number) and p is large enough, the Gibbs measure is supported by a union of small sets $(C_\alpha)_{\alpha \geq 1}$ that are far apart. The remarkable feature here is that this decomposition is **not**(in contrast with the case of the Hopfield model) apparent from the form of the Hamiltonian. We will call the sets C_α the *lumps*, to avoid the overused word "state". (We will consider later the question of whether they are "pure states".)

Theorem 7.2. *There exists a number L such that if $p \geq L$ and $\beta \leq \beta_p + 1/L$, then*

$$\lim_{N \to \infty} E\langle\Big(\frac{\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2}{N}\Big)^2 1_{\{|\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2| \leq N/2\}}\rangle \to 0.$$

This means that two configurations in different lumps have generically a zero overlap, so that the lumps are as far from each other as they can possibly be. They are also small, since $\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2/N$ is close to 1 for $\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2$ in the same lump, so they are well separated from each other, which of course greatly helped us to construct them. Theorem 7.2 is proved by the cavity method; due to the restriction to integration over the region where overlaps are $\leq 1/2$, it does not seem possible to use a positivity argument, but here again gaussian processes are very useful.

Let us now denote by w_α the weight $G_N(C_\alpha)$ of lump α, and assume that the numbering is such that $w_1 \geq w_2 \geq \cdots$. The random sequence $(w_\alpha)_{\alpha \geq 1}$ is obviously crucial for the understanding of the model; it is unfortunately not easy to obtain information about it. From Theorem 7.2, we have

$$E\langle\Big(\frac{\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2}{N}\Big)^p\rangle \leq \sum w_\alpha^2 + o(1)$$

where $o(1) \to 0$ as $N \to \infty$.

Combining with the relation

$$E\frac{1}{N}\frac{\partial}{\partial\beta}\log Z_N = \frac{\beta}{2}(1-\langle(\frac{\boldsymbol{\sigma}^1\cdot\boldsymbol{\sigma}^2}{N})^p\rangle)$$

(that is obtained as the first equality of (3.7)) one can prove that for $\beta > 2\sqrt{\log 2}$ and large N we have $E\sum w_\alpha^2 > \epsilon(p) > 0$ where $\epsilon(p)$ does not depend upon N. (This ought to be true for all $\beta > \beta_p$ but we do not know how to show this. Note however that the restriction $\beta \le \beta_p + 1/L$ does allow values of $\beta > 2\sqrt{\log 2}$). The condition $E\sum w_\alpha^2 > \epsilon(p)$, together with $\sum w_\alpha = 1$, shows that at least some of the weights w_α are "macroscopic", i.e. of order 1.

The distribution predicted (and reinvented) by the physicists for the weights (w_α) is of interest (see [P-Y] for a modern view and earlier references). Consider a number $m \in (0,1)$ and a Poisson point process on $\mathbb{R}^+$ such that its intensity measure has density mx^{-m-1} with respect to Lebesgue measure. Consider a realization $(x_\alpha)_{\alpha\ge1}$ of this process. Then $S = \sum x_\beta < \infty$. a.s, and it is believed that as $N\to\infty$, the distribution of the weights w_α converges to the distribution of $v_\alpha = x_\alpha/S$, where the parameter $m = m(p,\beta)$ is such that $(1-m)$ is about $(\beta-\beta_p)/\beta_p$ for $\beta-\beta_p$ small and p large. There would be some hope to prove this conjecture [A-C2] if we knew that the distribution of the weights w_α has a limit as $N\to\infty$; but, unfortunately, the best argument to date towards the existence of such a limit seems to be that there is no reason why it should not exist!

In this situation, it makes sense to try to go forward and examine the fundamental question of whether the lumps are "pure states" by assuming as weak as possible unproven properties of the weight distribution. One particularly useful such condition is as follows.

(H) There exists $\delta > 0$, $p_0 > 0$ such that, if $p \ge p_0$, we have for each $\epsilon > 0$

$$\limsup_{N\to\infty} P\Big(\sum_{\alpha\le 200} w_\alpha \ge 1-\epsilon\Big) \le \epsilon^\delta.$$

The number 200 is of course somewhat arbitrary. This condition simply means that it is rare that a few weights carry almost all the mass, and is (of course) satisfied by the conjectured distribution. To simplify the statement of the following result, we consider only the case p even.

THEOREM 7.3. (informal version). *There exists a constant L with the following property. If (H) is true, then for p large enough, and $\beta \le \beta_p + 1/pL$, the lumps (C_α) are in the limit the union of two* pure *states related by a global symmetry around zero.*

Thus, we will have $C_\alpha = \Sigma_\alpha \cup (-\Sigma_\alpha)$, where Σ_α is a "pure state ". A physicist would define a pure state by saying that the overlap of two independent configurations belonging to Σ_α are generically constant (which is a way to express that (2.2) holds for the restriction of G_N to Σ_α). How to express this mathematically is a bit more tricky. One way to do this is to introduce the quantity

$$E_N = E\langle(N^{-1}(\boldsymbol{\sigma}^1-\boldsymbol{\sigma}^2)\cdot(\boldsymbol{\sigma}^3-\boldsymbol{\sigma}^4))^2 1_A\rangle \tag{7.4}$$

where $A = \{\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2, \boldsymbol{\sigma}^3, \boldsymbol{\sigma}^4; \forall i, j \leq 4, \boldsymbol{\sigma}^i \cdot \boldsymbol{\sigma}^j \geq N/2\}$. Restricting the thermal integral to A essentially means that we force $\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2, \boldsymbol{\sigma}^3, \boldsymbol{\sigma}^4$ to belong simultaneously to a set of the type Σ_α or to a set of the type $-\Sigma_\alpha$. The final statement of Theorem 7.3 is that $\lim_{N \to \infty} E_N = 0$, which essentially means that "(2.3) holds in each Σ_α". The proof again relies upon relating E_N and E_{N-1} via the cavity method.

Thus Theorem 7.3 asserts that if $\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2 \in \Sigma_\alpha$ then (generically) $\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2 = \pm N q_\alpha$, where q_α is a certain (possibly random) quantity depending possibly upon α. Physicists believe that for each $\alpha, q_\alpha = q$, where q is non random; but it unfortunately seems to be difficult to gather evidence in this direction unless one has a much better control of the weights distribution.

References

[A-C1] M. Aizenman, P. Contucci, *On the stability of the quenched state in Mean Field Spin Glass Models*, Manuscript, 1997.

[AC2] ———, *The Indy 500 model*, Manuscript, 1998.

[A-L-R] M. Aizenman, J. L. Lebowitz, D. Ruelle, *Some rigorous results on the Sherrington-Kirkpatrick model*, Commun Math. Phys., 112 (1987), 3-20.

[A-W] M. Aizenmam, J. Wehr, *Rounding effects of quenched randomness of first-order phase transitions*, Commun. Math. Phys. 130, 1990, 485-528.

[A-G-S] D.J. Amit, H. Gutfreund, H.Sompolinsky, *Statistical Mechanics of Neural Networks near Saturation*, Annals of Physics 173, 1987, 30-67.

[A-T] J. R. L. Almeida, D. T. Thouless, *Stability of the Sherrington-Kirkpatrick solution of a spin glass model*, J. Phys. A: Math, Gen., II (1978), 983-990.

[B] A. Bovier, *Comment on "Capacity of Hopfield model"*, J. Phys. A 30, 1997, 7993-7996.

[B-G1] A. Bovier, V. Gayrard, *The retrival phase of the Hopfield Model*, Prob. Theor. Relat. Fields 107, 1997, 61-98.

[B-G2] ———, *An almost sure large derivation principle for the Hopfield model*, Ann. Probab. 24, 1996, 1444-1475.

[B-G3] ———, *Hopfield models as generalized random mean field models*, Mathematical aspects of spin glasses and neural networks, A. Bovier and P. Picco editors, Progress in Probability, Vol 41, Birkaeuser, Boston, 1997..

[B-G4] ———, *An almost sure central limit theorem for the Hopfield model*, Markov Proc. Rel. Fields 3, 1997, 151-173.

[B-G-P1] A. Bovier, V. Gayrard, P. Picco, *Gibbs states for the Hopfield model in the regime of perfect memory*, Prob. Theory. Relat. Fields 100, 1994, 329-363.

[B-G-P2] ———, *Gibbs states of the Hopfield model with extensively many patterns*, J. Stat. Phys. 79, 1995, 395-414.

[B-P] A. Bovier, P. Picco (editors), *Mathematical aspects of Spin Glasses and Neural Networks*, Progress in Probability, Vol. 41, Birkhäuser, Boston, 1997.

[B-S] E. Bolthausen, A. S. Sznitman, *On Ruelle's probability casades and an astract cavity method*, Manuscript, 1997.

[C] F. Comets, *A spherical bound for the Sherrington-Kirkpatrick model*, in "Hommage à P.-A. Meyer et J. Neveu", Astérisque 236, 103-108, 1996.

[C-N] F. Comets, J. Neveu, *Sherrington-Kirkpatrick model of spin glasses and stochastic calculus: the high temperature case*, Commun Math. Phys. 166 (1995), 549-564.

[C-S] A. Crisanti, H. J. Sommers, *The spherical p-spin interaction spin glass model: the statics*, Z. Phys. B. Condensed Matter 87, 341-354 (1992).

[D] B. Derrida, *Random energy model: An exactly solvable model of disordered systems*, Phy. Rev. B, 24 # 5, 2613-2626, 1981.

[D-G] B. Derrida, E. Gardner, *Optimal storage properties of Neural Network models*, J. Phys. A. 21, 271-284, 1988.

[F] F. Friedgut, *Necessary and sufficient conditions for sharp threshold for graph properties and the k-sat problem*, Manuscript, 1996.

[F-Z] J. Fröhlich, B. Zegarlinski, *Some comments on the Sherrington-Kirkpatrick model of spin glasses*, Commun. Math. Phys. 112 (1987), 553-566.

[G1] E. Gardner, *Spin glasses with p-spin interactions*, Nuclear Phys. B 257, #6, 747-765, 1985.

[G2] ———, *The space of interactions in neural network models*, J. Phys. A. 21, 257-270, 1988.

[Gu1] F. Guerra, *Fluctuations and thermo dynamic variables in mean field spin glass models in: Stochastic Processes, Physics and Geometry*, S. Albeverio et. al. editors, World Scientific, Singapore, 1995.

[Gu2] ———, *About the overlap distribution in mean field spin glass models*, Int. J. Modern Physics B, 1996, 1675-1684.

[G-M] D. Gross, M. Mézard, *The simplest spin glass*, Nuclear Phys. B 240, [F 512], 1984, 431-452.

[H1] J. J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proc. Natl. Acad. Sci. USA 79, 1982, 1554-2558.

[H2] ———, Proc. Natl. Acad. Sci. USA 81, 1982, 3088.

[H-K-P] J. Hertz, A. Krogh, R. C. Palmer, *Introduction to the theory of Neural computation*, Addison Wesley, 1991.

[I-S-T] A. Ibragimov, V. N. Sudakov, B.S. Tsirelson, *Norms of Gaussian sample functions*, Proceedings of the Third Japan USSR Symposium on Probability theory, Lecture Notes in Math 550, Springer Verlag, 20-41, 1976.

[K-M] Krauth W. and Mézard M., *Storage capacity of memory network with binary couplings*, J. Phys. 50, 1989, 3057-3066.

[K-R] Kim J.H., Roche J.R., *On the Optimal Capacity of Binary Neural Networks: Rigorous Combinatorial approaches*, Preprint,1996.

[L-T] M. Ledoux, M. Talagrand, *Probability in a Banach Space*, Springer-Verlag, 1991.

[Lou1] D. Loukianova, *Capacité de mémoire dans le modèle de Hopfield*, C. R. Acad. Sci. Paris 318, Serie I, 1994, 157-160.

[Lou2] D. Loukianova, *Lower bounds on the restitution error of the Hopfield model*, Probab. Th. Relat. Fields 107, 1997, 161-176.

[M] M. Mézard, *The space of interactions in neural networks: Gardner's computation with the cavity method*, J. Phys. A. 22, 2181-2190, 1988.

[M-P-V] M. Mézard, G. Parisi, M.Virasiro, *Spin glass theory and beyond*, World Scientific, Singapore, 1987.

[M-Z] R. Monasson, R. Zecchina, *Statistical Mechanics of the Random K-Sat Model*, Phys. Rev. E, 1997, 1357-1370.

[N] C. Newman, *Memory capacity in neural network models: Rigorous lower bounds*, Neural Networks I, 1988, 223-238.

[N-S1] C. Newman, D. Stein, *Chaotic size dependence in spin glasses. Cellular automata and cooperative systems*, (Les Houches, 1992), 525-529, NATO. Adv. Sci. Inst. Ser. C Math. Phys. Sci., 396, Kluwer Acad. Publ. Dordrecht, 1993.

[N-S2] ———, *Non-mean-field behavior of realistic spin glasses*, Phys. Rev. Letters 76, 1997, 515-518.

[N-S3] ———, *Spatial inhomogeneity and thermodynamic chaos*, Phys. Rev. Letters 76, 1997, 4821-4824.

[N-S4] ———, *Thermodynamic chaos and the structure of short-range spin glasses, Mathematical aspects of Spin Glasses and Neural networks*, A. Bovier and P. Picco editors, Progress in Probability Vol. 41, Birkhäuser, Boston, 1998, 243-287.

[P-F1] L. Pastur, A. Figotin, *Exactly soluble model of a spin glass*, Sov. J. Low Temp. Phys. 3, 1977, 378-383.

[P-F2] ———, *On the theory of disordered spin systems*, Theor. Math. Phys. 35, 1978, 403-414.

[P-S] L. Pastur, M. Shcherbina, *Absence of Self-Averaging of the order parameter in the Sherrington-Kirkpatrick model*, J. Stat. Phys. 62 (1991), 1-19.

[Pe1] D. Petritris, *Thermodynamic formalism of Neural computing*, Dynamics of Complex interacting systems (Santiago, 1994), 81-146, Nonlinear Phenom. Complex Systems, 2, Kluwer Acad. Publ..

[Pe2] ———, *Equilibrium statistical mechanics of frustrated spin glasses: a survey of mathematical results*, Am. Inst. H. Poincaré Phys. Théor. 64, 1996, 255-288.

[Pi] J. Pitman, *Coalescents with multiple collisions*, Manuscript, 1997.

[P-Y] J. Pitman, M. Yor, *The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator*, Ann. Probab. 25, 1997, 855-900.

[R] D. Ruelle, *A mathematical reformulation of Derrida's REM and GREM*, Comm. Math. Phys. 108, 1987, 225-239.

[Sh1] M. Shcherbina, *More about Absence of Self averages of the order parameter in the Sherrington-Kirkpatrick Model*, CARR Reports in Mathematical Physics, no 3/91, Department of Mathematics, University of Rome "la Sapienza", 1991.

[Sh2] ———, *On the Replica-Symmetric Solution for the Sherrington-Kirkpatrick Model*, Helv. Phys. Acta. 70, 1997, 838-853.

[S-K] D. Sherrington, S. Kirkpatrick, *Solvable model of a spin glass*, Phys. Rev. Lett. 35, 1972, 1792-1796.

[T1] M. Talagrand, *Regularity of Gaussian processes*, Acta. Math. 159, 1987, 99-149.

[T2] ———, *Concentration of measure and isoperimetric inequalities in product spaces*, Publ. Math. I.H.E.S. 81, 1995, 73-205.

[T3] ———, *Rigorous results for the Hopfield model*, C. R. A. S. 321, 1995, 109-112.

[T4] ———, *A New look at independence*, Ann. Probab. 24, 1996, 1-34.

[T5] ———, *The Sherrington-Kirkpatrick model: a challenge to mathematicians*, Prob. Th. Relat. Fields 110, 1998, 109-176.

[T6] ———, *Rigorous results for the Hopfield model with many patterns*, Prob. Th. Relat. Fields 110, 1998, 177-276.

[T7] ———, *Rigorous low temperature results for the p-spin interaction model*, Manuscript, 1997.

[T8] ———, *The high temperature case of the K-sat problem*, Manuscript, 1998.

[T9] ———, *Self averaging and the space of interactions in neural networks*, Manuscript, 1998.

[T10] ———, *Exponential inequalities and replica-symmetry breaking for the Sherrington-Kirkpatrick model*, Manuscript, 1998.

[T11] ———, *Exponential inequalities and convergence of moments in the replica-symmetric regime of the Hopfield model*, Manuscript, 1998.

[T12] ———, *Intersecting random half spaces and the Perception capacity formula*, Manuscript, 1998.

[T13] ———, *Intersecting random half cubes*, Manuscript, 1998.

[To] G. Toulouse, *Frustration and disorder, new problems in statistical mechanics, spin glasses in a historical perspective*, Heidelberg Colloquium on Spin Glasses, J. L. van Hemmen and I. Morgenstern eds., Lecture Notes in Physics 192, Springer Verlag, 1983.

[T-D-C] G. Toulouse, S. Dehaene and J. P. Changeux, *Spin glass model of learning by selection*, Proc. Natl. Acad. Sci. USA 83, 1986, 1695-1698.

Michel Talagrand
Equipe d'Analyse-Tour 46,
E.S.A. au C.N.R.S. no. 7064,
Université Paris VI
Homepage www.proba.jussieu.fr