ELSEVIER
Theoretical Computer Science 265 (2001) 69–77
Theoretical Computer Science
www.elsevier.com/locate/tcs

Rigorous results for mean field models for spin glasses

Michel Talagrand

Equipe d' Analyse, Tour 46, Boite 186, Université Paris 6, 4 Place Jussieu, 75230 Paris Cedex 05, France

Abstract

In this paper I explain why I believe that it is important to prove rigorous results about mean field models for spin glasses, and why I think it is difficult. I also describe at a high level what mathematicians have been able to prove.

Keywords: Spin glasses; Mean field models

1. Introduction

It is assumed in the paper that the reader has at least some idea about what the title means, and in the more detailed parts of the discussion, that he has heard of the main models. I have already written the simplest introduction to the topic, aimed at a reader with a mathematical background and having never heard of it. It can be found in the "expository" section of my Web page [20]. Other more ambitious introductions are also to be found there. The purpose of the present paper is to point to these, and to provide the reader lacking a heavy mathematical background with a non-technical summary of what has been done. This paper has three more sections, that explain, respectively, the need for rigorous results, what has been proved at high temperature, and what has been proved at low temperature. But first, I would like to say a few words of why and how I came to be interested in spin glasses, despite a complete lack of background in physics. For 20 years, I studied very happily Analysis and Probability. The closest I ever came to an "applied" problem was in studying combinatorial optimization problems with random data, such as Bin Packing, or those considered in [9]. However, in the brand of probability to which I contributed, large independent families of random variables, similar to those used to generate randomness in disordered systems, played a central role. The triggering event occurred during summer 1993, when Erwin Bolthausen wrote

E-mail address: mit@ccr.jussieu.fr (M. Talagrand).

0304-3975/01/$ - see front matter
PII: S0304-3975(01)00150-5

the Hamiltonian of the Sherrigton–Kirkpatrick (SK) model,

$$H_N(\sigma_1, ..., \sigma_N) = \frac{1}{\sqrt{N}} \sum_{i<j} g_{ij}\sigma_i\sigma_j, \tag{1}$$

where $(g_{i,j})_{i<j}$ are i.i.d. independent standard normal random variables. He asked for a simple proof of the Aizenmann–Ruelle–Lebovitch theorem that at inverse temperature $\beta < 1$, we have

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

where

$$Z_N = \sum \exp(-\beta H_N(\sigma_1, ..., \sigma_N))$$

for a summation over all values of $\sigma_1, ..., \sigma_N = \pm 1$. The simple proof I eventually found will be sketched in the beginning of Section 3. It relies upon a general principle called concentration of measure, which has been very successful for all kinds of questions of probability (see [13] and the references therein). By the time I understood that this early success was rather accidental, I was hooked for good by the topic.

2. What is the point of proving rigorous results ?

First, I should probably apologize for all the commonplace comments I am going to make, but I am not totally sure that they are completely useless. Also, I should say of course that the opinions expressed below are just that, opinions, and that I certainly should not be sued for them.

What is the point of proving rigorous results when there is at the same time consensus in the Physics community and overwhelming evidence (at least at high temperature) that the "solutions have been found"?

Even though the SK model is motivated by physical phenomenon, it is difficult to argue that it is a realistic model for matter. In my eyes at least, it is a mathematical object, and a rather canonical one at that. Thereby it should be studied by mathematical methods, rather than by methods from physics. This is of course all the more true for other problems, such as the assignment problem, that are purely mathematical, yet have been successfully studied in [9].

In fact, one of the main motivations to obtain rigorous results is precisely that these are so difficult to obtain. This difficulty means that we do not understand well these rather canonical mathematical objects, and it would be dangerous not to try to correct this. It is largely irrelevant that the solution found by physicists to the SK model and others is almost certainly correct. Somehow, for the future developments of mathematics, the tools that allow reaching the solution of a problem are more important than the solution itself. And a good way to have a small chance to find powerful tools is in attacking problems with simple statements, that are yet very difficult.

One must also add that mathematicians like to think that future generations will continue their work. This can be safely done over the long term only by leaving behind rock-solid constructions.

Now, is it not true that the difference between the arguments of the physicists and a mathematically rigorous proof lies only in some hair-splitting details? I will explain that, while this might be the case in some situations, it is definitely not the case concerning spin glasses. It is certainly correct that mathematicians do spend a lot of energy to rule out with certainty possibilities that are extremely (and even sometimes absurdly) unlikely in the first place, and this can indeed be viewed by others as hair-splitting details. This typically makes mathematical papers hardly readable by anybody else. Mathematicians feel that this work is necessary, but it might be hard to convince other communities of this, so I will not try, because this is not the issue here.

The issue is that the arguments provided in the papers written by the physicists do not address the facts that are mathematically the hardest to prove. As much as I can tell, these papers rely from the very start on general principles, such as the existence of the thermodynamical limit, and other principles to be mentioned below. There is no doubt that these principles are very reliable, in particular when backed by clever and extensive numerical simulation, and I am certainly not questioning the legitimacy of their use in work of physical nature. But the use of these principles has, by its nature, little to do with a mathematical proof. It is one thing to rely upon such a general principle, but it is an entirely different matter to be able to analyze the situation in enough detail to obtain a mathematical argument. For example, we all agree that if a random variable depends upon the influence of many independent random variables, but not too much upon any of them, it is very likely to be asymptotically normal. Proving explicitly central limit theorems is another story. To mention one of my favorite problems, there is little doubt that the length of the shortest tour through N random points of the unit square is asymptotically normal (not standard) but I do not have any hope to ever prove it.

The amount of work needed to rely upon a general principle and the amount of work needed to provide actual arguments are often not commensurate, but the benefits reaped by the two approaches are not commensurate either. A famous paper of Gardner on the capacity of the Perceptron [5] starts with a one-line assertion that the free energy of a certain Hamiltonian should be self-averaging. This is proved in [16]. Despite the existence of a well-developed set of mathematical tools designed to approach this type of "concentration of measure" results (which is the reason behind this success, and why we chose this example), this rigorous proof does require some work. But what provides a better understanding, this proof, or what was after all an (educated) act of faith?

It seems necessary to insist at length upon this difference between reliance upon general heuristic principles and actual proofs, because mathematicians have the greatest difficulties to prove statements that are taken as granted by physicists. In particular, the high temperature phase of mean field models is considered as extremely easy by physicists who rely upon such general principles but it seems very hard to analyze mathematically. (Think about low temperature then.) Physicists consider the fact that the system is "in a pure state" as the "default" situation, and feel no need for further justification, anymore than they feel the need to prove the existence of limits that "obviously" exist. (On the other hand, most of the work I have done in this area is to prove this "obvious" fact.)

It is probably instructive to dwell in more detail on a particularly striking example of the contrast general principle/proof. It concerns the ground state energy of the SK model, that is the quantity

$$\min H_N(\sigma_1, ..., \sigma_N),$$

where H_N is defined in (1), and where the minimum is over $\sigma_1, ..., \sigma_N = \pm 1$.

Mathematicians can prove that this quantity is self-averaging, and is of order N, so that instead one can study

$$\frac{1}{N} E \min H_N(\sigma_1, ..., \sigma_N).$$

But does the limit of this quantity exist as $N \to \infty$? There is no reason to think that it does not, but I do not think that anybody has any idea of how this existence could be actually proved. I know only two methods to prove that a quantity has a limit. One is to prove that this limit is some actual quantity such as $\sqrt{2}$ or π. Another method is to relate problems of different sizes. This second method can be applied in a very subtle way, such as in the work of Aldous on the assignment problem [1]. It seems hard to use either method here.

A nice feature of this problem is that it is related to other even simpler questions of the same nature, as one observes by diagonalization of the random matrix $(g_{ij})_{i,j \leqslant N}$, where $g_{ij} = g_{ji}$.

I have used one such problem, with a strong geometric flavor, to make the point that there are "obvious" things one does not know how to prove. The problem is as follows. Consider a parameter α . In the Euclidean space $\mathbb{R}^N$, it makes sense to speak of a random subspace of dimension $\lfloor \alpha N \rfloor$ because there is a natural probability measure on the set of these subspaces. We denote by E the corresponding expectation. Let us denote by Q the orthogonal projection upon this random space, and consider the quantity

$$E\left(\frac{1}{\sqrt{N}} \max \|Q(x)\|\right),$$

where $\|.\|$ is the Euclidean norm, and where the maximum is taken over x in the unit cube, that is in the set of vectors that have all their coordinates between -1 and 1. Is it true that the limit

$$\lim_{N\to\infty} E(\frac{1}{\sqrt{N}} \max \|Q(x)\|)$$

exists? There is no reason why it should not, but saying this is simply NOT a valid mathematical argument.

It could also be instructive to the reader to have another example of a seemingly obvious statement that probably nobody can prove. It deals with a different issue, the validity of approximations. (Typically, approximations behave much better than mathematicians know how to prove.) The problem concerns again the ground state energy of the SK model. Does this ground state energy change significantly if the variables are replaced by variables ε_{ij} that are independent "coin flipping" variables i.e. $P(\varepsilon_{ij}=1)=P(\varepsilon_{ij}=-1)=1/2$? Most likely it does not, but there does not seem to exist an approach to this problem. The difficulty is that it is a different matter to apply the central limit theorem to a fixed number of random variables, or to have a collection of random variables whose size grows very fast.

The reader might have wondered why I have not appealed to the all-too-obvious fact that the replica method, the most used method in the physics literature, is not yet quite mathematically justified. This is simply because I felt that most readers might be unaware that other much more innocent looking assumptions are already very problematic from a mathematical point of view. Considering the replica method, it would be very nice to understand why it allows to "guess" the correct value for the free energy. But I personally think that trying to obtain mathematical proofs following the lines of the replica method is not going to be a rewarding project in the short term.

All this being said, I see another reason to try to obtain rigorous results about mean field models. I think there is a potential in this topic for a new direction in probability theory. Probability theory has paid a lot of attention to independent random variables. It has then paid attention to more complex structures, which mostly are inspired by the idea of stochastic processes, in the sense of a time-indexed family of random variables. The gaussian random variables $H_N(\sigma_1,\ldots,\sigma_N)$ of (1) however have a global structure rather different from either of these. Their correlation is given by

$$E(H_N(\sigma_1,\ldots,\sigma_N)H_N(\rho_1,\ldots,\rho_N))=\frac{1}{2N}\left(\sum_{i\leqslant N}\sigma_i\rho_i\right)^2-\tfrac{1}{2}.$$

It is nicely smooth, but in a very "high dimensional" fashion. This type of correlation structure had hardly been investigated. The fact that it allows the emergence of very interesting random structures is quite revolutionary, and possibly indicates that entire new areas of probability theory awaits discovery.

3. High temperature results

First I will sketch the already announced proof of (1). The first observation is that

$$EZ_N^2(\beta)\leqslant K(EZ_N(\beta))^2,$$

where K depends only upon β . This is a simple computation. The "second moment method", i.e. the fact that any positive r.v. Y satisfies

$$P(Y\geqslant \tfrac{1}{2}EY)\geqslant\frac{1}{4}\frac{(EY)^2}{EY^2}$$

implies that

$$P(\log Z_N(\beta) \geqslant N(\log 2 + \frac{\beta^2}{4}) - K) \geqslant \frac{1}{K},$$

where K depends upon β only. When one knows that the free energy is self-averaging, (a fact that easily follows from general "concentration of measure" principles discovered long ago [8]), it is enough to obtain (1).

Thus (1) is very easy. The real reason for this is simply that the expected free energy is as small as it can possibly be. The natural approach in the case where the annealed free energy is not the quenched free energy (such as in the SK model with external field) seems to be the cavity method, that is induction upon the number of spins N. For this one expresses averages in the N spin system in function of averages in the $(N-1)$-spin system. One is led to the evaluation of quantities of the type

$< A > / < B >$, where A, B are complicated random quantities, and where $< . >$ denote an average with respect to Gibbs measure. How to evaluate these? Here there is a huge difference between assuming from the start, as physicists do, that at high temperature a lot of nice things happen, and trying to prove them. This is because when attempting to produce proofs, one simply does not know anything at all about Gibbs' measure when the proof starts, and the greatest difficulty is to gain any control at all. In particular, it must be stressed that mathematicians are NOT permitted to use what I will call the "triple dot magic wand". This remarkable tool allows a physicist to write $A = B + ...$ where A is the quantity of interest, B is a much simpler quantity, and the magic ... represents a remainder, about which nothing is usually specified, and that is presumably small. This hypothesis is justified a posteriori by checking that its consequences are reasonable and agree with numerical simulation. Mathematicians have unfortunately no other choice than telling what this remainder is and proving inequalities to control it. This is a different game, with a different purpose, and it is typically very much harder.

I have now succeeded to prove the validity of the so-called replica symmetric (RS) solutions at a high enough temperature for the SK model (with external field), the Hopfield model (with external field), the model for the K-sat problem in the version of Monasson and Zecchina [10], the perceptron capacity model of Gardner [5] and Gardner–Derrida [6], and for a simpler version of the assignment problem studied in [1, 9]. (This project occupied the best of the last five years.) Typically the regions where I can do this are "proportion" of the region where the replica-symmetric behavior is expected. (So, for the Hopfield model, this includes a low temperature region.) With the exception of the SK model, where a kind of miracle happens, the current proofs of these results are really difficult. (An alternative approach by Shcherbina [12] for the SK model gives a slightly better region but is very much harder. An alternative approach of Bovier and Gayrard [3] for the Hopfield model is somewhat simpler than the author's approach for this model, but unfortunately seems specific to that case and gives a worse region.) The papers I wrote are available on my Web page, but I will take the unusual step to urge the reader NOT to look at any proof there, because this would only discourage him. Rather, he should wait for the mini-course [19] that will eventually be available, and where the proofs, having been better understood, and written one more time, are simpler and better explained. This mini-course treats only the case of the SK model, and its natural extension to the p-spins interaction model of [4], but should give a fairly good idea of the techniques used overall. For the SK model and the Hopfield model I have been able to gain control at the level of the fluctuations around the mean field, and to obtain very precise results (at high enough temperature). I hope to be able to do this soon for the other models.

What about proving the validity of the RS solution in the entire high temperature region? A first observation here is that the arguments of "stability" provided by physicists, either with the cavity method, or the more mysterious replica method, simply show (heuristically) something like "if the system is close to the RS solution" for N spins, this remains true for $N+1$ spins. But how can one show in the first place that the system is close to the RS solution? There exists a pretty mathematical method that solves this conceptual problem. It allows transferring information from a given temperature to a slightly lower temperature. It is to show, at a certain temperature, that the overlap of two independent configurations can hardly ever (i.e. with exponentially small probability) fall outside a certain small interval. This then remains true at a slightly lower temperature, and allows to show at this slightly lower temperature that the overlaps are nearly constant, from which one has a chance to show that the system is close to the RS solution at this lower temperature. This is the motivation behind the exponential inequalities that I proved in [17]. Unfortunately, this does not make progress on the main difficulty. This main difficulty is the possibility of "discontinuous" transitions, where the overlaps start to take really different values below a certain temperature, as occurs in the case of the p-spin interaction model with no external field. After trying for years to find a way to rule these out, I believe that I have finally succeeded (April 2000), using in an essential way an idea of Guerre [7].

4. Low temperature results

It is at first hard to believe that one could prove anything at all about low temperature without first having been able to control the entire high temperature region. This is fortunately not entirely true. But before we discuss some of the results, we must discuss some of the issues. The celebrated "Parisi solution" describes the organization of the "states" of the SK model at low temperature. But what are really these? To paraphrase a famous saying, when I do not think about them, I know what they are, but when I think about them, I do not know anymore. And, for a start, why should these states exist at all in the SK model? Certainly for a physicist, this is a natural a priori assumption, but, as we discussed earlier, finding a mathematical justification is another story. The very existence of these "states" implies a strong property, namely that (somewhat imprecisely) the overlaps take essentially only a few values that depend upon the randomness. Why should this be true?

One of the first low temperature results is due to Pastur and Shcherbina [11]. It asserts that at sufficiently low temperature, the order parameter

$$\frac{1}{N}\sum_{i\leqslant N} <\sigma_i>^2$$

is not self-averaging. Unfortunately, they cannot prove it using Hamiltonian (1), so that these authors introduce a somewhat unnatural symmetry-breaking term. This result is essentially a high temperature result "turned upside down" because it consists in proving that if the parameter were self-averaging, then the free energy would have the value predicted by the RS solution, which is wrong at low temperature. Somewhat of the same nature is the "replica symmetry breaking" result of [17]. It asserts that at (essentially) every point of the low temperature region, "an infinitesimal coupling between two replicas has a macroscopic effect", and can be seen as a rigorous identification of the low temperature region.

Some rigorous low temperature concerning the p-spin interaction model [G2] with no external field are given in [18]. These are made possible by a kind of a lucky accident. It is possible to show rigorously that if p is large, just below the critical temperature, the overlap of two configurations is essentially never between 0.01 and 0.99, or between -0.01 and -0.99. This is obtained by transferring information on the system just above the critical temperature. It is then an automatic consequence of this fact that Gibbs measure creates "lumps", amounts of mass that are supported by a small subset of the configuration space. This is the identification of the predicted "states", which are easier to find in this situation because they are far from each other. (Here one should pay homage to the prophetic intuition of Gardner, who, at the end of her paper [5], mentions that this model might be somewhat accessible to rigorous study.) From this point the hard work starts. The cavity method allows to show that the overlap of configurations in different lumps is essentially zero. The next question is whether the lumps are pure states, that is, whether the overlap of two configurations in the same lump is essentially independent of these configurations. This question is closely related to the behavior of the sequence of the weights of the states for Gibbs'measure. It is shown in [18] that if one knows a mild property of this sequence of weights, the lumps are indeed pure states.

After being stopped at that point for several years, I got a new idea in December 1999. For large (but fixed) p, I can prove the asymptotic validity of "one level of symmetry breaking" solution for almost all values of the temperature in an interval that grows with p, and (allowing for technical reason a suitable infinitesimal perturbation of the Hamiltonian) I can even treat the case where one adds a small external field. The only question that remains is the actual computation of the various parameters. I could not prove the physicist's belief that "the true solution maximizes the free energy".

Despite this somewhat unexpected success, knowing when, or even whether, rigorous and significant structure results will be obtained for the low temperature phase of the SK model is for anybody to guess.

Acknowledgements

I am grateful for the organizers to have accepted this paper in this special issue, thereby displaying interest for the objectives I have pursued.

References

[1] D. Aldous, Asymptoties in the random assignment problem, Probab. Theory Related Fields 93 (1992) 507–534.
[2] M. Aizenman, J.L. Lebowitz, D. Ruelle, Some rigorous results on the Sherrington-Kirkpatrick model, Comm. Math. Phys. 112 (1987) 3–20.
[3] A. Bovier, V. Gayrard, Hopfield models as generalized random field models, in: A. Bovier and G. Picco (eds.), Mathematical Aspects of Spin Glasses and Neural Networks, Progress in Probability, Vol. 41, Birkhauser, Boston, 1997.
[4] E. Gardner, Spin glasses with p-spin interaction. Nuclear Phys. B 257 (1985) 747–765.
[5] E. Gardner, The space of interactions in neural network models, J. Phys. A 21 (1988) 257–270.
[6] E. Garner, B. Derrida, Optimal storage properties of neural network models, J. Phys. A 21 (1988) 257–270.
[7] F. Guerra, Conference delivered in les Houches, January 2000.
[8] I. Ibragimov, V. Sudakov, B. Tsirelson, Norms of Gaussian sample functions, Proc. 3rd Japan–USSR Symp. on probability Theory, Springer Lecture Notes in Mathematics, Vol. 550, Springer, Berlin, 1976, pp. 20–41.
[9] M. Mezard, G. Parisi, M. Virasiro, Spin Glass Theory and Beyond, World Scientific, Singapore, 1987.
[10] R. Monasson, R. Zechina, Statistical mecanics of the random K- satisfiability model, Phys. Rev. 56 (1997) 1357–1370.
[11] L. Pastur, M. Shcherbina, Absence of self-averaging of the order parameter in the Sherrington-Kirkpatrick model, J. Statist. Phys. 62 (1991) 1–19.
[12] M. Shcherbina, On the replica-symmetric solution for the Sherrington-Kirkpatrick model, Helv. Phys. Acta. 70 (1997) 838–853.
[13] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, Publ. Math. I.H.E.S. 81 (1998) 73–205.
[14] M. Talagrand, The Sherrington-Kirkpatrick model : A challenge for mathematiciens, Probab. Theory Related Fields 110 (1998) 109–176.
[15] M. Talagrand, Rigorous results for the Hopfield model with many patterns, Probab. Theory Related Fields 110 (1998) 177–276.
[16] M. Talagrand, Self-averaging and the space of interactions in neural networks, Random Struct. Algorithms 14 (1999) 199–213.
[17] M. Talagrand, Replica symmetry breaking and exponential inequalities for the Sherrington–Kirkpatrick model, Ann. Probab. 28 (2000) 1018–1062.
[18] M. Talagrand, Rigorous low temperature results for the p-spins interaction model, Probab. Theory Related Fields 117 (2000) 303–360.
[19] M. Talagrand, A first course on spin glasses, Proc. of the Saint Flour Summer School in Probability, Summer 2000, Lecture Notes in Math., Springer, Berlin, to appear.
[20] M. Talagrand, La page de Michel Talagrand, http://www.proba.jussieu.fr.