Available online at www.sciencedirect.com
SCIENCE @ DIRECT®

C. R. Acad. Sci. Paris, Ser. I 337 (2003) 111–114

COMPTES RENDUS MATHEMATIQUE

Probability Theory/Mathematical Physics

The generalized Parisi formula

Michel Talagrand

Équipe d'analyse de l'institut mathématique, 4, place Jussieu, 75230 Paris cedex 05, France

Received 26 May 2003; accepted 27 May 2003

Presented by Yves Meyer

Abstract

For a large class of Gaussian Hamiltonians, we use Guerra's interpolation scheme to show that at any temperature the free energy of the corresponding spin-glass system is asymptotically given by Parisi's formula. This class includes the Sherrington–Kirkpatrick model, the p-spin interaction model for even p, and many others. ***To cite this article: M. Talagrand, C. R. Acad. Sci. Paris, Ser. I 337 (2003).***

Résumé

La formule de Parisi généralisée. Pour une large classe d'Hamiltoniens Gaussiens, nous utilisons la méthode d'interpolation de Guerra pour montrer qu'à toute température l'énergie libre du verre spin correspondant est donnée par la formule de Parisi. Cette classe comprend le modèle de Sherrington–Kirkpatrick, le modèle à p-spin pour p pair, et bien d'autres. ***Pour citer cet article : M. Talagrand, C. R. Acad. Sci. Paris, Ser. I 337 (2003).***

1. Statement of the results

We consider a function $\xi:\mathbb{R}\to\mathbb{R}$ with $\xi(0)=0$ and we assume that ξ is convex, even, and that $\xi''(x)>0$ for $x>0$. For $N\geqslant 1$, denote by $\boldsymbol{\sigma}$ the generic element of $\Sigma_N=\{-1,1\}^N$, and consider an Hamiltonian H_N on Σ_N such that the family $(H_N(\boldsymbol{\sigma}))_{\boldsymbol{\sigma}}$ is jointly Gaussian and

$$\forall\boldsymbol{\sigma}^1,\boldsymbol{\sigma}^2\in\Sigma_N,\quad \left|\frac{1}{N}EH_N(\boldsymbol{\sigma}^1)H_N(\boldsymbol{\sigma}^2)-\xi(R_{1,2})\right|\leqslant c(N),\tag{1}$$

where $\lim_{N\to\infty}c(N)=0$ and $R_{1,2}=N^{-1}\sum_{i\leqslant N}\sigma_i^1\sigma_i^2$ is the overlap of two configurations. We set $\theta(q)=q\xi'(q)-\xi(q)$ so that by convexity $\xi(x)-x\xi'(q)+\theta(q)\geqslant 0$. An important example is the case of the p-spin interaction model when p is even in which case $\xi(x)=\beta^2x^p/2$, $\theta(x)=\beta^2(p-1)x^p/2$, β being the inverse temperature. The most important case is $p=2$, the famous Sherringtom–Kirkpatrick (SK) model.

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

1631-073X/$ – see front matter © 2003 Académie des sciences. Published by Éditions scientifiques et médicales Elsevier SAS. All rights reserved.
doi:10.1016/S1631-073X(03)00268-1

We consider an integer k, and numbers

$$m_0 = 0 < m_1 \leqslant m_2 \leqslant \cdots \leqslant m_k = 1; \qquad q_0 = 0 \leqslant q_1 \leqslant q_2 \leqslant \cdots \leqslant q_{k+1} = 1.$$

For $0 \leqslant \ell \leqslant k$, we set $a_\ell = \sqrt{\xi'(q_{\ell+1}) - \xi'(q_\ell)}$. We set $\psi_{k+1}(x) = \log \operatorname{ch} x$, and for $\ell \geqslant 1$, we define recursively

$$\psi_\ell(x) = \frac{1}{m_\ell} \log E \exp m_\ell \psi_{\ell+1}(x + a_\ell z),$$

where z is standard Gaussian. We define $\psi_0(x) = E\psi_1(x + a_0 z)$. We consider a number h fixed once and for all, and, writing $\mathbf{m} = (m_1, \ldots, m_{k-1})$ and $\mathbf{q} = (q_1, \ldots, q_k)$, we set

$$\mathcal{P}_k(\mathbf{m}, \mathbf{q}) = \log 2 + \psi_0(h) - \frac{1}{2} \sum_{1 \leqslant \ell \leqslant k} m_\ell \big(\theta(q_{\ell+1}) - \theta(q_\ell)\big).$$

Theorem 1.1 (The Parisi formula). *We have*

$$\lim_{N\to\infty} \frac{1}{N} E \log \sum_{\boldsymbol{\sigma}} \exp\left(-H_N(\boldsymbol{\sigma}) + h \sum_{i \leqslant N} \sigma_i\right) = \mathcal{P} := \inf \mathcal{P}_k(\mathbf{m}, \mathbf{q}), \tag{2}$$

where the infimum is computed over all values of the parameters.

Assume that the infimum in (2) is a minimum, and consider $\mathbf{m}$ and $\mathbf{q}$ such that this minimum is attained at $\mathcal{P}_k(\mathbf{m}, \mathbf{q})$, where k is as small as possible. The physicists say that in this situation there is $k-1$ steps of replica symmetry breaking. Only the case $k = 1$ (high temperature behavior) or $k = 2$ are described in the physics literature, but building on the ideas of [2] one can see that all the values of k are possible. The most interesting case however is when the infimum in (2) is not attained. This is expected to be the case for the SK model when β is large enough.

2. Elements of proof

The proof relies on Guerra's interpolation scheme. Given $m_1, m_2, \ldots, m_{k-1}$ and $q_1, q_2, \ldots, q_k$ as above, consider for $i \leqslant N$, $0 \leqslant \ell \leqslant k$ independent standard Gaussian r.v. $z_{i,\ell}$, independent of the randomness of H_N, and for $0 \leqslant t \leqslant 1$ consider

$$-H_t(\boldsymbol{\sigma}) = -\sqrt{t} H_N(\boldsymbol{\sigma}) + \sqrt{1-t} \sum_{i \leqslant N} \sigma_i \sum_{0 \leqslant \ell \leqslant k} z_{i,\ell} a_\ell + h \sum_{i \leqslant N} \sigma_i.$$

Set $F_{k+1,t} = \log \sum_{\boldsymbol{\sigma}} \exp(-H_N(\boldsymbol{\sigma}))$ and define recursively

$$F_{\ell,t} = \frac{1}{m_\ell} \log E_\ell \exp m_\ell F_{\ell+1,t}$$

for $\ell \geqslant 1$, where E_ℓ denote expectation in the r.v. $z_{i,p}$, $p \geqslant \ell$. Set $\phi(t) = N^{-1} E F_{1,t}$, where E denotes expectation in the r.v. $z_{i,0}$ and the randomness of H_N. For $1 \leqslant \ell \leqslant k$, define $W_\ell = \exp m_\ell (F_{\ell+1,t} - F_{\ell,t})$ and for a function f on Σ_N, let

$$\gamma_\ell(f) = E_\ell\big(W_\ell \cdots W_k \langle f \rangle_t\big),$$

where $\langle\cdot\rangle_t$ denote averaging for the Gibbs measure with Hamiltonian H_t. It is a probability on Σ_N; we denote by $\gamma_\ell^{\otimes 2}$ its square on Σ_N^2. Guerra [1] proves the identity

$$\phi'(t) = -\frac{1}{2}\sum_{1\leqslant \ell\leqslant k} m_\ell\big(\theta(q_{\ell+1}) - \theta(q_\ell)\big)$$
$$- \frac{1}{2}\sum_{1\leqslant \ell\leqslant k} (m_\ell - m_{\ell-1}) E\big(W_1\cdots W_{\ell-1}\gamma_\ell^{\otimes 2}\big(\xi(R_{1,2}) - R_{1,2}\xi'(q_\ell) + \theta(q_\ell)\big)\big) + \mathcal{R}, \tag{3}$$

where $|\mathcal{R}| \leqslant c(N)$, and thus obtains the "upper bound" part of Theorem 1.1. (Observe that the terms of the last summation in (3) are $\geqslant 0$.) The natural approach, that we will follow, is to show that there is near equality in (3). This amounts to show that for each ℓ, we have $E(W_1\cdots W_{\ell-1}\gamma_\ell^{\otimes 2}(|\mathcal{R}_{1,2} - q_\ell|)) \simeq 0$. "The overlaps being constant" is a typical "high-temperature" behavior. The fantastic power of Guerra's scheme is that (as already observed in [1]) it breaks low temperature behavior into a series of "high-temperature" problems. One has however no hope to use this idea unless one has solved easier problems, such as controlling the entire high-temperature region of the SK model. This is, in itself, a non-trivial problem, that was solved only very recently [3]. The present work generalizes the techniques of [3]. Theorem 1.1 is an immediate consequence of the following.

Theorem 2.1. *Given $t_0 < 1$, there exists $\varepsilon > 0$ (depending only on ξ and t_0) such that if*

$$\mathcal{P}_k(\mathbf{m},\mathbf{q}) \leqslant \mathcal{P} + \varepsilon, \tag{4}$$

$$\mathcal{P}_k(\mathbf{m},\mathbf{q}) \text{ is minimum over the choices of } m_1,\ldots,m_{k-1};\ q_1,\ldots,q_k \tag{5}$$

then

$$\forall t \leqslant t_0,\quad \lim_{N\to\infty} \phi(t) = \phi(0) - \frac{t}{2}\sum_{1\leqslant \ell\leqslant k} m_\ell\big(\theta(q_{\ell+1}) - \theta(q_\ell)\big). \tag{6}$$

Let us denote by $\psi(t)$ the right-hand side of (6). To prove (6) we will show that

$$(u - q_\ell)^2 \geqslant K\big(\psi(t) - \phi(t)\big) \Rightarrow E\big(W_1\cdots W_{\ell-1}\gamma_\ell^{\otimes 2}(1_{\{R_{1,2}=u\}})\big) \leqslant K\exp(-N/K). \tag{7}$$

Here and below, K denotes a constant depending on ξ, t_0, k, $\mathbf{m}$ and $\mathbf{q}$, but not on N or $t \leqslant t_0$. Since $\psi(0) = \phi(0)$, (3) implies a differential inequality that proves (6).

We fix once and for all $1 \leqslant \ell_0 \leqslant k$, and we consider couples of standard normal r.v. $(z_{i,\ell}^1, z_{i,\ell}^2)$ for $\ell \leqslant k$, $i \leqslant N$. We assume that $z_{i,\ell}^1 = z_{i,\ell}^2$ if $\ell < \ell_0$ while these two variables are independent if $\ell \geqslant \ell_0$. We define $m'_\ell = m_\ell$ if $\ell \geqslant \ell_0$ and $m'_\ell = m_\ell/2$ if $\ell < \ell_0$. We define

$$T_{k+1} = \log\sum_{R_{1,2}=u}\exp\bigg(-\sqrt{t}H_N(\boldsymbol{\sigma}^1) - \sqrt{t}H_N(\boldsymbol{\sigma}^2) + \sum_{j=1,2}\sum_{i\leqslant N}\sigma_i^j\bigg(h + \sum_{\ell\leqslant k} a_\ell z_{i,\ell}^j\bigg)\bigg) \tag{8}$$

and inductively $m'_\ell T_\ell = \log E_\ell \exp m'_\ell T_{\ell+1}$, where E_ℓ denotes expectation in the r.v. $z_{i,p}^j$ for $p \geqslant \ell$.

Lemma 2.2. *If for some $\varepsilon' > 0$ we have $\eta(u) := N^{-1}ET_1 \leqslant 2\phi(t) - \varepsilon'$ then we have*

$$E\big(W_1\cdots W_{\ell_0-1}\gamma_{\ell_0}^{\otimes 2}(1_{\{R_{1,2}=u\}})\big) \leqslant K\exp(-N/K).$$

This is not difficult once one observes that if in (8) we remove the restriction $R_{1,2} = u$ in the summation we obtain $2\phi(t)$ rather than $\eta(u)$. To prove (7) it then suffices to show that

$$\eta(u) \leqslant 2\psi(t) - (u - q_{\ell_0})^2/K. \tag{9}$$

The key tool is a principle that iterates Guerra's construction, but for a subset of two coupled copies of the system, in the spirit of [3], Theorem 2.11.15.

Theorem 2.3. *Consider $\eta = \pm 1$, numbers $n_0 = 0 < n_1 \leqslant \cdots \leqslant n_{k'} = 1$ and $0 \leqslant \rho_1 \leqslant \cdots \leqslant \rho_{k'} \leqslant \rho_{k'+1} = 1$. Set $b_\ell = \sqrt{\xi'(\rho_{\ell+1}) - \xi'(\rho_\ell)}$. Consider an Hamiltonian such as in (1), Gaussian r.v. $(w_{i,\ell}^j)$ for $i \leqslant N, \ell \leqslant k'$, and* *standard Gaussian r.v.* $(y_{i,\ell}^j)$. *These are independent of* H_N *and the* $(w_{i,\ell}^j)$. *The couples* $(y_{i,\ell}^1, y_{i,\ell}^2)$ *are independent of each other, and* $y_{i,\ell}^1 = \eta y_{i,\ell}^2$ *if* $\ell < \ell_1$, *while these variables are independent if* $\ell \geqslant \ell_1$. *Define*

$$T_{k'+1} = \log \sum_{R_{1,2}=\eta\rho_{\ell_1}} \exp\biggl(-\sqrt{v} H_N\bigl(\boldsymbol{\sigma}^1\bigr) - \sqrt{v} H_N\bigl(\boldsymbol{\sigma}^2\bigr) + \sum_{j=1,2,i\leqslant N} \sigma_i^j \biggl(h + \sum_{\ell\leqslant k'} \bigl(w_{i,\ell}^j + \sqrt{1-v}\, b_\ell y_{i,\ell}^j\bigr)\biggr)\biggr).$$

Define $n_\ell T_\ell = \log E_\ell \exp n_\ell T_{\ell+1}$, *where* E_ℓ *denotes expectation in the r.v.* $y_{i,p}^j$ *and* $w_{i,p}^j$ *for* $p \geqslant \ell$, *and define* $\varphi(v) = N^{-1} E T_1$. *Then we have*

$$\varphi(1) \leqslant \varphi(0) - 2 \sum_{\ell\leqslant\ell_1} n_\ell\bigl(\theta(\rho_{\ell+1}) - \theta(\rho_\ell)\bigr) - \sum_{\ell_1<\ell\leqslant k} n_\ell\bigl(\theta(\rho_{\ell+1}) - \theta(\rho_\ell)\bigr). \tag{10}$$

The proof of (9) is obtained by a suitable application of (10) (to $\sqrt{t} H_N$ rather than H_N). The most delicate part of the proof is to show that $\eta(u) < 2\psi(t)$ when $u \neq q_{l_0}$. The details depend on the position of u with respect to the numbers q_ℓ. We will sketch only the case where $q_{\ell_0} < u \leqslant q_{\ell_0+1}$. In that case we take $k' = k+1$ and the sequence $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_{k'})$ is the sequence $(q_1, \ldots, q_{\ell_0}, u, q_{\ell_0+1}, \ldots)$. For the sequence $(n_1, \ldots, n_{k'})$ we take the sequence $(m_1/2, \ldots, m_{\ell_0-1}, m/2, m_{\ell_0}, \ldots)$ where m is some number with $m_{\ell_0-1} < m < 2m_{\ell_0}$. (It is a crucial ingredient of this construction that m can be bigger than m_{ℓ_0}.) We take $w_{i,p}^j = \sqrt{1-t}\, a_p z_{i,p}^j$ if $p \leqslant \ell_0$, $w_{i,\ell_0}^j = 0$ and $w_{i,p}^j = \sqrt{1-t}\, a_{p-1} z_{i,p-1}^j$ if $p > \ell_0$. The bound provided by (10) is then a function $\Phi(t,m)$.

Crucial to the construction is the fact that $\Phi(t, m_{\ell_0}) = 2\psi(t)$ and that the partial derivatives in m of $\Phi(1,m)$ and $2\mathcal{P}_{k'}(\mathbf{m}(m), \boldsymbol{\rho})$ coincide at $m = m_{\ell_0}$, where $\mathbf{m}(m) = (m_1, \ldots, m_{\ell_0-1}, m, m_{\ell_0}, \ldots, m_k)$. A (non-trivial) consequence of (4) is that these partial derivatives must be $\leqslant \alpha$, where α is a small positive number. Calculus then shows that when we decrease t a little, then the partial derivative of $\Phi(t,m)$ at $m = m_0$ decreases enough to ensure that it is <0 if $t \leqslant t_0$, and thus that we can find m with $\Phi(t,m) < 2\psi(t)$. The argument is somewhat different when $\ell_0 = 1$ and $u < q_1$ since in that case there is no room to vary $m_0 = 0$, and different again when $u < 0$.

Once we know that $\eta(u) < 2\psi(t)$ when $u \neq q_{\ell_0}$, to prove (9) it suffices to consider the case where $|u - q_{\ell_0}|$ is very small, and this case is much easier because everything can be computed by power expansions.

In conclusion, even though there is quite a distance between Theorem 1.1 and a complete mathematical understanding of the SK model, one can now hope that it is only a matter of time and work until this understanding is reached.

The real challenge has then moved to the other important spin glass models, such as the Hopfield model and the Perceptron model, where one has yet to find arguments that generalize Guerra' approach. (More accurately, interpolating Hamiltonians are easy to write, but the miracle of positivity upon which our arguments rely does not occur in an obvious manner.)

References

[1] F. Guerra, Replica broken bounds in the mean field spin glass model, Comm. Math. Phys. 233 (2003) 1–12.
[2] M. Talagrand, Multiple levels of symmetry breaking, Probab. Theory Related Fields 117 (2000) 449–466.
[3] M. Talagrand, Spin Glasses: A Challenge to Mathematicians, Springer-Verlag, 2003.