On the Replica Symmetric Equations for the Hopfield Model

L.Pastur[*][†] M.Shcherbina[*] B.Tirozzi[‡]

Abstract

The paper continues our paper [1]. We locate the domain in which the overlap parameters of the Hopfield model assume their values with probability asymptotically close to 1. This allows us to justify the Gaussian form of the probability distribution of the molecular (cavity) field of non condensing patterns and to present in a somewhat different form the replica symmetric self consistent equations for the order parameters of the Hopfield model.

1 Introduction

This paper is a continuation and an addition of the paper [1], that was devoted to the derivation of a system of self consistent equations of the replica symmetric solution of the Hopfield model based on the assumption that the fluctuations of the Edwards-Anderson order parameter vanish in the thermodynamic limit. The model is defined by the Hamiltonian

$$H = -\frac{1}{2}\sum_{i,j=1}^{N} I_{ij}\sigma_i\sigma_j + \varepsilon\sum_{i=1}^{N} h_i\sigma_i, \tag{1.1}$$

where

$$I_{ij} = \frac{1}{N}\sum_{\mu=1}^{p}\xi_i^{\mu}\xi_j^{\mu}, \tag{1.2}$$

ξ_i^{μ} $(i = 1,...N,\ \mu = 1,..,p)$ are independent random variables assuming values ± 1 with probability $\frac{1}{2}$ and h_i are i.i.d. Gaussian random variables with zero mean and the unit variance. Random vectors $\xi^{\mu} = \{\xi_i^{\mu}\}_{i=1}^{N}$, $\mu = 1,..,p$ are known as patterns or modes of interaction.

The model was introduced in [2] for a finite (N-independent) number of patterns p in the spin glass (thermodynamic) context and in [3] in the neural networks (dynamic) context. Subsequent physical studies of the model showed that properties of the model in the case of macroscopic number of patterns

$$p \to \infty,\ N \to \infty, \quad \frac{p}{N} \to \alpha > 0, \tag{1.3}$$

are much more interesting and diverse but also much more hard to study. We refer the reader to books [4],[5] for physical results obtained by the widely accepted in the theoretical physics replica method. Interesting rigorous results on the model were recently obtained by A.Bovier et al. (see their contributions to [6]) and by M.Talagrand [7].

Our approach to the rigorous study of the thermodynamics of the Hopfield model is based on the analysis of the dependence of the order parameter of the model on the total number of spins N by careful control of changes of the parameters induced by the addition of $(N+1)$th spin. The method

[*]Institute for Low Temperature Physics, 310164, Kharkov, Ukraine
[†]University Paris-7, 75251, Paris, France
[‡]Department of Physics of Rome University "La Sapienza", 5, p-za A.Moro, Rome, Italy

(that could be called the rigorous cavity method) was proposed in [8] for the Sherrington-Kirkpatrick model, that can be viewed as $\alpha = \infty$ limit of the model (1.1)-(1.3). However, the technical form of the method used in [8] was rather involved. A considerable simpler version of the method was proposed in [10]. In [1] we applied the latter version to the model (1.1)-(1.3).

To explain the subject of this paper let us recall the simple identity

$$\langle \sigma_1 \rangle = \langle \tanh \beta (\sum_{j=2}^{N} I_{1j}\sigma_i + \varepsilon h_1) \rangle \tag{1.4}$$

valid for the Ising model with any interaction I_{ij} (here and below the symbol $\langle \ldots \rangle$ denotes the Gibbs expectation). The mean field approximation is based on the assumption that the thermodynamic correlations between spins vanish in the macroscopic limit

$$|\langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle| \to 0, \quad N \to \infty. \tag{1.5}$$

This allows us to replace (1.4) by the relation

$$\langle \sigma_1 \rangle = \tanh \beta (\sum_{i=2}^{N} I_{ij} \langle \sigma_i \rangle + \varepsilon h_1). \tag{1.6}$$

that can be regarded as a system of equations for the local magnetization $\langle \sigma_i \rangle$ and leads to the corresponding self consistent equations for the order parameters of the model.

The vanishing of thermodynamic correlations and validity of the mean field approximation has been rigorously proven in various asymptotic regimes of translational invariant models (see e.g. [11],[12]) and in the Hopfield model with finite p [13]. One of the ways to facilitate the proof of this property in the whole temperature regime is to add to the Hamiltonian a properly chosen external field. In the translational invariant case one chooses the field that breaks the symmetry of an infinite system and provides uniqueness of its state for all temperatures. In disordered systems where we often do not know a "genuine" breaking symmetry field, the addition of the external field is at least a rather convenient technical mean, playing the role of "source" term and allowing one to write quantities of interest in a convenient form by using the differentiation with respect to the field, the integration by parts and similar technical tools.

In work [8], using some special vanishing as $N \to \infty$ field it was shown that for the SK model [9] the relation (1.5) is equivalent to the self averaging property of the Edwards-Anderson order parameter q_N

$$E\{(q_N - E\{q_N\})^2\} \to 0, \quad N \to \infty, \tag{1.7}$$

where

$$q_N = \frac{1}{N} \sum_{i=1}^{N} \langle \sigma_i \rangle^2. \tag{1.8}$$

Deriving (1.5) from (1.7) and studying the moments of the random variable $\langle \sigma_1 \rangle$, it was proved that if (1.7) is true, then the following version of (1.6) is valid in distribution

$$\langle \sigma_1 \rangle = \tanh \beta (\sum_{i=2}^{N} I_{1j} \langle \sigma_j \rangle_0 + \varepsilon h_1). \tag{1.9}$$

Here and below the symbol $\langle \ldots \rangle_0$ denotes the Gibbs measure, corresponding to the Hamiltonian (1.1) with $\sigma_1 = 0$.

This idea was developed in [10], where the infinitesimal field was replaced by an ordinary Gaussian one, that, in particular, allowed us to derive on the basis of the stochastic Griffith lemma (see Lemma 3 below) an important relation, valid for almost all values of β and ε

$$\frac{1}{N} \sum_{i=1}^{N} h_i(\sigma_i - \langle \sigma_i \rangle) \to 0, \quad N \to \infty,$$

in the Gibbs measure and probability. This relation simplifies considerably the method of [8] and leads to a more natural proof of the equivalence of (1.7) and (1.9). The further development of this method showed that (1.7) and (1.9) for the SK model are valid for a large region of parameters, including the low temperatures [14].

Now let us recall, that in the case of the SK model the interaction in (1.1) have the form

$$I_{ij} = \frac{1}{\sqrt{N}} J_{ij},$$

where $\{J_{ij}\}_{i,j=1}^{n}$ are independent (modulo the symmetry condition $J_{ij} = J_{ji}$) random variables with zero mean and the variance J^2. Since $\{\langle\sigma_j\rangle_0\}_{j=2}^{N}$ do not depend on $\{J_{1j}\}_{j=2}^{N}$, the cavity field

$$\frac{1}{\sqrt{N}} \sum J_{1j} \langle\sigma_j\rangle_0 \tag{1.10}$$

in (1.9) will be the Gaussian random variable with zero mean and the variance

$$w_N = \frac{J^2}{N} \sum \langle\sigma_j\rangle_0^2 \tag{1.11}$$

if $\{J_{ij}\}_{i,j=1}^{n}$ are Gaussian. Thus (1.9) leads to the relation

$$\overline{q}_N = E\{\langle\sigma_1\rangle_0^2\} = \int \tanh^2 \beta(J\sqrt{\overline{q}_N}u + \varepsilon h_1) \frac{e^{-u^2/2}du}{\sqrt{2\pi}} d\mu(h_1) + o(1) \quad (\overline{q}_N = E\{q_N\}), \tag{1.12}$$

provided the variance that depends on $\{J_{ij}\}_{i,j=2}^{N}$ has the property

$$w_N = J^2\overline{q}_N + R_N, \quad E\{R_N^2\} \to 0, \quad N \to \infty. \tag{1.13}$$

This relations follows from assumption (1.7) if one takes into account that the expectations $E\{q_N\}$, $E\{w_N\}$ and the variances of the random order parameters q_N and w_N can be expressed via the first derivatives with respect to ε and J of the Hamiltonians H (1.1) and $H\Big|_{\sigma_1=0}$. Then, by using the stochastic version of the Griffith's lemma on the convergence of a sequence of convex functions (see Lemma 3 below), we obtain (1.13) for almost all ε, J with respect to the Lebesgue measure. Thus, (1.12) and (1.13) lead to the self consistent equation known as the replica symmetric equation for the order parameter $\overline{q} = \lim_{N\to\infty} \overline{q}_N$ of the SK model.

In the more general case of the i.i.d. but not necessary Gaussian interactions $\{J_{ij}\}_{i,j=1}^{n}$ we obtain the asymptotically Gaussian form of the cavity field (1.10) by using the central limit theorem whose applicability is guaranteed by the relation

$$E\{N^{-2}\sum_{j=2}^{N} \langle\sigma_j\rangle_0^4\} \to 0, \quad as \quad N \to \infty, \tag{1.14}$$

(because $|\langle\sigma_j\rangle| \le 1$ and the r.h.s. is obviously $O(N^{-1})$).

Consider now the Hopfield model (1.1)-(1.3). In this case the cavity field in (1.9) has the form

$$\frac{1}{\sqrt{N}} \sum_{\mu=2}^{p} \xi_1^\mu \langle t_1^\mu\rangle_0, \quad t_1^\mu = \frac{1}{\sqrt{N}} \sum_{i=2}^{N} \xi_1^\mu \sigma_i, \tag{1.15}$$

i.e. in the Hopfield model the role of $\langle\sigma_i\rangle_0$ $(i = 2, ..., N)$ of the SK model play $\langle t_1^\mu\rangle_0$ $(\mu = 1, ..., p)$. As in the SK case they are independent of $\{\xi_1^\mu\}$ $(\mu = 1, ..., p)$, but unlike the SK case $\langle t_1^\mu\rangle_0$ are not bounded, because generally we have only the bound $|\langle t_1^\mu\rangle_0| \le N^{1/2}$ that is not sufficient to guarantee the analog

$$E\{N^{-2}\sum_{\mu=2}^{p}\langle t_1^\mu\rangle_0^4\} \to 0, \quad N\to\infty \tag{1.16}$$

of (1.14).

The goal of this paper is to prove (1.16) and to discuss related facts that were used in [1], in particular, the use of the stochastic analogue of the Griffith's lemma on a sequence of random convex functions.

2 The Model and the Main Results

Define H as a sum of two terms

$$H = H_0 + H_1, \tag{2.1}$$

where

$$\begin{aligned} H_0 &= -\frac{J}{2N}\sum_{\mu=s+1}^{p}\sum_{i,j=1}^{N}\xi_i^\mu\xi_j^\mu\sigma_i\sigma_j - \varepsilon_1\sum_{\mu=s+1}^{p}\gamma^\mu t^\mu + \varepsilon\sum_{i=1}^{N}h_i\sigma_i, \\ H_1 &= -\frac{J(1+d_N\zeta)}{2N}\sum_{\nu=1}^{s}\sum_{i,j=1}^{N}\xi_i^\nu\xi_j^\nu\sigma_i\sigma_j - h^1\sum_{i=1}^{N}\xi_i^1\sigma_i - \sum_{\nu=2}^{s}\gamma^\nu t^\nu, \end{aligned} \tag{2.2}$$

$s = [\log^{1/2} N]$ is the number of the patterns which are expected to be condensed, J, h^1, ε_1, and ε are positive parameters, $d_N = s^{-2/3}$, ζ is an independent random variable uniformly distributed in the interval $(1, 2)$, variables γ^μ, h_i are independent Gaussian random variables with zero mean and variance 1.

The Hamiltonian H_0 contains the contribution of the non condensed patterns and the Hamiltonian H_1 includes terms due to the condensed patterns. The random variables γ^μ, h_i play the role of "symmetry breaking fields" (or rather the source terms because we need them for technical reasons) and after thermodynamic limit $N\to\infty$ one can send $\varepsilon_1, \varepsilon \to 0$. The term in H_1 containing $d_N\zeta$ is necessary to single out the condensed patterns. It would be more naturally to have a finite number s of these patterns (even the one would suffice), but we can only prove that singling out $s = O(\log^{1/2} N)$ patterns we shall have all the others "non condensed".

The variable t^μ is just a convenient notation for the following linear combination of spins:

$$t^\mu \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_i^\mu\sigma_i \qquad \mu = 1,\ldots,p\ . \tag{2.3}$$

We will use also notations:

$$\begin{array}{lll} m^\mu \equiv \frac{1}{\sqrt{N}}t^\mu, & r_N = p^{-1}\sum_{\mu=s+1}^{p}\langle t^\mu\rangle^2, & U_N = N^{-1}\sum_{\mu=s+1}^{p}(t^\mu)^2, \\ q_N = N^{-1}\sum_{i=1}^{N}\langle\sigma_i\rangle^2, & \overline{r}_N = E\{r_N\} & \overline{q}_N = E\{q_N\}. \end{array} \tag{2.4}$$

The main result of the paper is the lemma, which allows us to overcome one of the serious technical difficulties, arising in the Hopfield model, if we try to generalize to it the methods proposed in [8] and [14] for the SK model. This difficulty is connected to the fact, that the variables t^μ, which play here the role of "spins" of the SK model (cf.(1.14) and (1.16)), are unbounded.

Lemma 1 *Consider the set*

$$\mathcal{M} \equiv \left\{\mathbf{m} = (m^1, .., m^p) : \max_{\nu\ge s+1}|m^\nu| \ge 4\delta_N\right\},$$

where $\delta_N = s^{-1/3} = d_N^{1/2}$. *Let* $\chi_{\mathcal{M}}(\mathbf{m})$ *be the characteristic function of the set* $\mathcal{M}$

$$\chi_{\mathcal{M}}(\mathbf{m}) = \begin{cases} 1, \mathbf{m} \in \mathcal{M}, \\ 0, otherwise. \end{cases}$$

Then for any $n > 0$ *there exists a quantity* C_n *independent of* N *such that*

$$Prob\{\langle \chi_{\mathcal{M}}(\mathbf{m})\rangle \leq e^{-\beta J N d_N^2/4}\} \geq 1 - C_n N^{-n}. \tag{2.5}$$

Moreover, if we add to the Hamiltonian H *of (2.1)-(2.2) any Hamiltonian* $\tilde{H}$ *which is symmetric with respect to the variables* $\{\xi_1^\mu\}_{\mu>s}$, $\{\xi_2^\mu\}_{\mu>s}$,...,$\{\xi_N^\mu\}_{\mu>s}$ *and the free energy of the sum* $H + \tilde{H}$ *satisfies the large deviation bounds of the type*

$$Prob\{|f(H+\tilde{H}) - E\{f(H+\tilde{H})|\zeta\}| > \varepsilon\} \leq D_n N^{-n},$$

then the estimate of the probability (2.5) is valid for Gibbs averages with respect to $H + \tilde{H}$.

Here and below the symbol $E\{...|\zeta\}$ means the average with respect to all random variables of the problem except ζ. The proof of Lemma 1 is given in the next section.

This lemma allows us to treat m^ν $(\nu > s)$ as random variables satisfying inequalities

$$|m^\nu| \leq 4 d_N^{1/2}, \tag{2.6}$$

because the Gibbs measure of m^ν satisfying the opposite inequality decays exponentially with probability 1, as $N \to \infty$, and so those m^ν can add only exponentially small contributions in our estimates. In particular, denoting by $\mathcal{J}$ the $N \times N$ matrix defined by (1.2) and using the bound

$$\text{Prob}\{||\mathcal{J}|| \geq (1+\sqrt{\alpha})^2 + \tilde{d}\} \leq e^{-N^{2/3}\tilde{d}^{4/3}\,\text{const}} \tag{2.7}$$

proved in [15], we obtain

$$E\{\sum_{\nu=s+1}^{p} \langle (m^\nu)^4\rangle\} \leq 16 d_N E\{\sum_{\nu=s+1}^{p} \langle (m^\nu)^2\rangle\} \leq 16 d_N E\{||\mathcal{J}||\} = \text{const}\, d_N \to 0,\ N \to \infty. \tag{2.8}$$

Thus, the l.h.s. of (1.16) is bounded by $O(s^{-2/3}) = O(\log^{-1/3} N)$.

Another important corollary of Lemma 1 is the formula, analogous to the formula of integration by parts for Gaussian variables.

Proposition 1 *Let* t_1^μ *be defined by (1.15). Then for any* $\mu_1, \ldots, \mu_k > s$, $\mu_i \neq \mu_j$ *and any bounded function* $F_1(\{\sigma\},\{\xi_i^\mu\})$, ...,$F_l(\{\sigma\},\{\xi_i^\mu\})$ *which do not depend on* ξ_1^μ

$$\begin{aligned} &N^{k/2} E\{\xi_1^{\mu_1}\ldots\xi_1^{\mu_k}\langle t_1^{\mu_1}\ldots t_1^{\mu_{k_1}} F_1(\{\sigma\},\{\xi_i^\mu\})\rangle\ldots\langle t_1^{\mu_{k_{l-1}+1}}\ldots t_1^{\mu_k} F_l(\{\sigma\},\{\xi_i^\mu\})\rangle\} = \\ &N^{k/2} E\{\tfrac{\partial^k}{\partial\xi_1^{\mu_1}\ldots\partial\xi_1^{\mu_k}}\langle t_1^{\mu_1}\ldots t_1^{\mu_{k_1}} F_1(\{\sigma\},\{\xi_i^\mu\})\rangle\ldots\langle t_1^{\mu_{k_{l-1}+1}}\ldots t_1^{\mu_k} F_l(\{\sigma\},\{\xi_i^\mu\})\rangle\} + R_N, \end{aligned} \tag{2.9}$$

with

$$|R_N| \leq d_N^{1/2} \cdot (2l)^{k+1} \max|F_1|\ldots\max|F_l| \cdot (2+\sqrt{\alpha})^k.$$

The same relations are valid for any perturbed Hamiltonian $H + \tilde{H}$, *if the perturbation* $\tilde{H}$ *is symmetrical with respect to* ξ_1^μ *in the sense of Lemma 1 and for any* k, n

$$E^{1/n}\{\langle (\frac{\partial^k \tilde{H}}{\partial\xi_1^{\mu_1}\ldots\partial\xi_1^{\mu_k}})^{2n}\rangle\} \leq \frac{C_{k,l}}{N^k} \tag{2.10}$$

with the constant $C_{k,n}$ *independent of* N. *Here and below the symbol* $\frac{\partial}{\partial\xi_1^\mu}$ *means the formal derivative, which can be obtained if we replace* ξ_1^μ *by a continuous variable.*

To study the influence of the "condensed" patterns we use

Lemma 2 *Consider the "approximate" Hamiltonian of the form*

$$H^a(\mathbf{c}) = H_0 - J(1+d_N\zeta)\sum_{\nu=1}^{s} c^\nu \sum_{i=1}^{N} \xi_i^\nu \sigma_i - \sum_{\nu=1}^{s}\gamma^\nu t^\nu - h^1\sum_{i=1}^{N}\xi_i^1\sigma_i + J(1+d_N\zeta)\frac{N}{2}\sum_{\nu=1}^{s}(c^\nu)^2, \quad (2.11)$$

where H_0 is defined by formula (2.2) and $\mathbf{c} \equiv (c^1,\ldots,c^s)$. Then the free energies of the initial Hamiltonian H and the "approximate" Hamiltonian H^a satisfy the inequality,

$$0 \le \min_{\mathbf{c}} E\{f(H^a(\mathbf{c}))|\zeta\} - E\{f(H)|\zeta\} \le \frac{const}{\log N} \quad (2.12)$$

and for almost all J, h^1, ε_1, ε

$$\overline{r}_N^* \equiv E\{p^{-1}\sum_{\mu=s+1}^{p}\langle t^\mu\rangle_*^2\} = \overline{r}_N + o(1), \qquad \overline{q}_N^* \equiv E\{N^{-1}\sum_{i=1}^{N}\langle\sigma_i\rangle_*^2\} = \overline{q}_N + o(1). \quad (2.13)$$

Here and everywhere below we use notation

$$H_*^a = H^a(\mathbf{c}_*(\zeta))$$

for the Hamiltonian H^a computed at the point $\mathbf{c}_*(\zeta) \equiv (c_*^1(\zeta),\ldots,c_*^s(\zeta))$ which provides the minimum value of the mean free energy $E\{f(H^a(c^1,\ldots,c^s))|\zeta\}$, and the symbol $\langle\ldots\rangle_*$ for the respective Gibbs average. Thus the symbols $\overline{q}_N^*$, $\overline{r}_N^*$ are the values of order parameters computed by means of this Gibbs measure.

Remarks

1. Lemma 2 allows us to replace the Hamiltonian H by H_*^a, which is linear with respect to the first s patterns.
2. Since the Hamiltonian H_*^a has the form

$$H_*^a = H + \frac{J(1+\zeta d_N)N}{2}\sum_{\nu=1}^{s}(m^\nu - c_*^\nu)^2,$$

it is easy to see, that it satisfies conditions of Lemma 1 and Proposition 1 for the perturbed Hamiltonians, and the estimates (2.9) is also valid for the $\langle\ldots\rangle_*$ averages.

The proof of Lemma 2 is based on an important technical lemma which we call the stochastic Griffith's lemma

Lemma 3 *Consider the sequence of convex random functions $\{f_n(t)\}_{n=1}^\infty$ ($f_n''(t) \ge 0$) in the interval (a,b). If functions f_n are self averaging, i.e. uniformly in t*

$$\lim_{n\to\infty} E\{(f_n(t) - E\{f_n(t)\})^2\} = 0,$$

and bounded ($|E\{f_n(t)\}| \le C$ uniformly in n, $t \in (a,b)$), then for almost all t

$$\lim_{n\to\infty} E\{[f_n'(t) - E\{f_n'(t)\}]^2\} = 0, \quad (2.14)$$

i.e. the derivatives $f_n'(t)$ are also self averaging ones for almost all t.

In addition, if we consider another sequence of convex functions $\{g_n(t)\}_{n=1}^\infty$ ($g_n'' \ge 0$) which are also self averaging uniformly in t

$$\lim_{n\to\infty} E\{(g_n(t) - E\{g_n(t)\})^2\} = 0$$

and

$$\lim_{n\to\infty} |E\{f_n(t)\} - E\{g_n(t)\}| = 0. \tag{2.15}$$

uniformly in t, then for all t, which satisfy (2.14)

$$\lim_{n\to\infty} |E\{f_n'(t)\} - E\{g_n'(t)\}| = 0, \qquad \lim_{n\to\infty} E\{[g_n'(t) - E\{g'(t)\}]^2\} = 0. \tag{2.16}$$

We prove Lemma 3 in Appendix.

Lemmas 1,2 and Proposition 1 allow us to derive a somewhat different than in [1] variant of self consistent equations for the Hopfield model.

Theorem 1 *Consider the Hopfield model of the form (2.1)-(2.2). Set*

$$\Delta_N = E\{(Np)^{-1} \sum_{\mu>s,i\geq 1} \langle (t^\mu - \langle t^\mu\rangle)(\sigma_i - \langle\sigma_i\rangle)\rangle^2\}, \tag{2.17}$$

Then for almost all values of parameters J, h^1, ε_1, ε α and β parameters the parameters $\overline{q}_N$, $\overline{r}_N$ satisfy the system of equations

$$\begin{aligned}\overline{r}_N &= \frac{\{\overline{q}_N + \varepsilon_1^2 J^2\beta^2(1-\overline{q}_N)^2\}}{(1-\beta J(1-\overline{q}_N))^2} + O(\Delta_N^{1/2}) + o(1),\\ \overline{q}_N &= E\left\{\int \frac{dv\exp(-\frac{v^2}{2})}{\sqrt{2\pi}} \tanh^2\beta((\alpha\overline{r}_N(\varepsilon_1))^{1/2}v + h^1\xi_1^1 + J\sum_{\nu=1}^s c^\nu\xi_1^\nu + \varepsilon h_1)\right\}\\ &\quad +O(\Delta_N^{1/2}) + o(1),\end{aligned} \tag{2.18}$$

where

$$\overline{r}_N(\varepsilon_1) = J^2\overline{r}_N + \frac{2J\beta\varepsilon_1^2}{1-\beta J(1-\overline{q}_N)} + \varepsilon_1^2,$$

and the parameters c^ν satisfy the equations

$$c^\nu = E\left\{\int \frac{dv\exp(-\frac{v^2}{2})}{\sqrt{2\pi}} \xi_1^\mu \tanh\beta((\alpha\overline{r}_N(\varepsilon_1))^{1/2}v + h^1\xi_1^1 + J\sum_{\nu=1}^s c^\nu\xi_1^\nu + \varepsilon h_1)\right\} + O(\Delta_N^{1/2}) + o(1).$$

Using Lemmas 1 and 2 and Proposition 1, one can prove Theorem 1 by the method proposed in [1]. One of the main steps in the proof is given by the lemma which in fact is some version of relations (1.7). To formulate this lemma we need some extra definitions.

Define the Hamiltonian $\Phi(\tau)$, interpolating between systems of $N-1$ and N spins

$$\Phi(\tau) = H_1^a - \frac{J\tau}{\sqrt{N}} \sum_{\mu=1}^p \xi_1^\mu t_1^\mu, \tag{2.19}$$

where

$$\begin{aligned}H_1^a = H_1^0 &- \frac{J}{2N}\sum_{\mu=s+1}^p (t_1^\mu)^2 - J(1+\zeta d_N)\sum_{\nu=1}^s c_*^\nu \sum_{i=2}^N \xi_i^\nu\sigma_i -\\ &h^1\sum_{i=2}^N \xi_i^1\sigma_i + \frac{JN(1+\zeta d_N)}{2}\sum_{\nu=1}^s (c_*^\nu)^2 - \sum_{\nu=2}^s \gamma^\nu t_1^\nu,\end{aligned} \tag{2.20}$$

It is easy to see, that if $\tau = \sigma_1$, then $\Phi(\tau)$ coincides with H^a up to the term $\tilde{h}_1\sigma_1$ ($\tilde{h}_1 = h^1\xi_1^1 + \varepsilon h_1 + \varepsilon_1\sum_\mu N^{-1/2}\gamma^\mu\xi_1^\mu$), which add only some constant to the Hamiltonian $\Phi(\tau)$ and therefore does not play an important role.

Consider also the corresponding partition function

$$Z(\tau) = \sum_{\sigma_2,\dots,\sigma_N} e^{-\beta\Phi(\tau)}, \tag{2.21}$$

and define the relative free energy

$$u(\tau) = \ln \frac{Z(\tau)}{Z(0)} \,. \tag{2.22}$$

Lemma 4 *The function $u(\tau)$ can be represented in the form*

$$u(\tau) = \beta J\tau \sum_{\nu=s+1}^{p} \frac{\xi_1^\nu}{\sqrt{N}} \langle t_1^\nu \rangle_0 + \beta J\tau \sum_{\mu=1}^{s} \xi_1^\mu c_*^\mu + (\langle U_N\rangle_0 - \alpha \overline{r}_N)\frac{\tau^2(\beta J)^2}{2} + R_N(\tau), \tag{2.23}$$

where $\langle \dots \rangle_\tau$ is the Gibbs averaging, corresponding to the Hamiltonian $\Phi(\tau)$, U_N is defined by (2.4) and the remainder $R_N(\tau)$ can be estimated as

$$E\{R_N^2(\tau)\} \le \, const \cdot \Delta_N + o(1),\ N \to \infty \tag{2.24}$$

The proof of Lemma 4 is given in [1] (see also [7]).

3 Proofs of Lemmas 1,2 and Proposition 1.

Proof of Lemma 1.
For given $\mu \le s$ and $\nu \ge s+1$ consider the set

$$\mathcal{A}^{\mu\nu} = \Big\{ \mathbf{m} = (m^1, .., m^p) : (1 + d_N\zeta)|m^\mu| \le |m^\nu| - 3\delta_N \Big\}. \tag{3.1}$$

Its Gibbs measure is

$$a^{\mu\nu} = \langle \theta(|m^\nu| - (1 + d_N\zeta)|m^\mu| - 3\delta_N)\rangle, \tag{3.2}$$

where $\theta(m) = \frac{1}{2}(1 + \mathrm{sign} m)$.

Let us assume, that for any n we have proved the estimate

$$\mathrm{Prob}\{a^{\mu\nu} \le e^{-\beta J N d_N^2/5}\} \ge 1 - C_n' N^{-n}. \tag{3.3}$$

Consider the set

$$\begin{gathered} \mathcal{A} \equiv \cup_{\mu=1}^{s} \cup_{\nu=s+1}^{p} \mathcal{A}^{\mu\nu} = \\ \Big\{ \mathbf{m} = (m^1, .., m^p) : (1 + d_N\zeta) \min_{\mu \le s} |m^\mu| \le \max_{\nu \ge s+1} |m^\nu| - 3\delta_N \Big\}. \end{gathered} \tag{3.4}$$

Then it follows from (3.3) that the Gibbs measure $\langle \chi_{\mathcal{A}}(m)\rangle$ of the set $\mathcal{A}$ satisfies the estimate

$$\begin{gathered} \mathrm{Prob}\{\langle \chi_{\mathcal{A}}(m)\rangle \le s(p-s)e^{-\beta J N d_N^2/5} \le e^{-\beta J N d_N^2/6}\} \\ \ge 1 - s(p-s)C_n' N^{-n} \ge 1 - C_{n-2} N^{-(n-2)}, \end{gathered}$$

when N is large enough. Then we use the inequality

$$\langle \chi_{\mathcal{M}}(m)\rangle = \langle \chi_{\mathcal{M}\cap\mathcal{A}}(m)\rangle + \langle \chi_{\mathcal{M}\cap\overline{\mathcal{A}}}(m)\rangle \le \langle \chi_{\mathcal{A}}(m)\rangle + \langle \chi_{\mathcal{M}\cap\overline{\mathcal{A}}}(m)\rangle. \tag{3.5}$$

But

$$\begin{gathered} \mathcal{M} \cap \overline{\mathcal{A}} \equiv \{\mathbf{m} : (1 + d_N\zeta) \min_{\mu \le s} |m^\mu| \ge \max_{\nu \ge s+1} |m^\nu| - 3\delta_N,\ \max_{\nu \ge s+1} |m^\nu| > 4\delta_N\} \subset \\ \{\mathbf{m} : (1 + d_N\zeta) \min_{\mu \le s} |m^\mu| \ge \delta_N\} \subset \{\mathbf{m} : (1 + d_N\zeta)^2 \sum_{\mu \le s} (m^\mu)^2 \ge s\delta_N^2\}. \end{gathered} \tag{3.6}$$

By using the definition (2.4) of m^μ and (2.7), we get that for any $\tilde{d}>0$ (in our case $\tilde{d}=\frac{(\log N)^{1/6}}{(1+d_N\zeta)}-(1+\sqrt{\alpha})^2$) the probability of the event

$$\{\sum_{\mu\le s}(m^\mu)^2\ge(1+\sqrt{\alpha})^2+\tilde{d}\}\subset\{\tfrac{1}{N}\sum_{i,j}J_{ij}\sigma_i\sigma_j\ge(1+\sqrt{\alpha})^2+\tilde{d}\}\subset$$
$$\{||\mathcal{J}||\ge(1+\sqrt{\alpha})^2+\tilde{d}\}$$

is less than $e^{-N^{2/3}\tilde{d}^{4/3}\text{const}}$ (recall that $\mathcal{J}$ is the matrix (1.2)). Therefore the probability to have the last set in (3.6) nonempty is also less then $e^{-N^{2/3}\text{const}}$. Thus (3.5), (3.6) and (3.3) prove (2.5).

Now we are left to prove (3.3).

To this end we use the standard representation

$$\begin{gathered}\exp\{\beta JN((1+d_N\zeta)\frac{(m^\mu)^2}{2}+\frac{(m^\nu)^2}{2})\}=\\(2\pi\beta JN)^{-1}(1+d_N\zeta))^{1/2}\int dxdy\exp\{\beta JN(xm^\mu+ym^\nu-\frac{x^2}{2(1+d_N\zeta)}-\frac{y^2}{2})\}\end{gathered}\tag{3.7}$$

and study

$$a_1^{\mu\nu}=\frac{\int\theta(|y|-|x|-2\delta_N)\exp\{\beta JNF_{N,\mu\nu}(x,y)\}dxdy}{\int\exp\{\beta JNF_{N,\mu\nu}(x,y)\}dxdy},\tag{3.8}$$

where $F_{N,\mu\nu}(x,y)$ is a random function defined by the formulae

$$\begin{aligned}F_{N,\mu\nu}(x,y)&\equiv f_{N,\mu\nu}(Jx,Jy)-\frac{x^2}{2(1+d_N\zeta)}-\frac{y^2}{2},\\f_{N,\mu\nu}(x,y)&\equiv\frac{1}{\beta JN}\log\sum_{\{\sigma\}}\exp\{-\beta H_{\mu\nu}(\sigma;x,y),\}\end{aligned}\tag{3.9}$$

with the Hamiltonian $H_{\mu,\nu}(\sigma;x,y)$ of the form

$$\begin{gathered}H_{\mu\nu}(\sigma;x,y)=-NJ\sum_{\mu'\ne\nu;\mu'>s}(m^{\mu'})^2-NJ(1+\zeta d_N)\sum_{\mu'\ne\mu;\mu'\le s}(m^{\mu'})^2-\sum_{i=1}^N h_i\sigma_i-\\-\varepsilon_1\sqrt{N}\sum_{\nu'>s}\gamma^{\nu'}m^{\nu'}-\sqrt{N}\sum_{\mu'\le s}\gamma^{\mu'}m^{\mu'}-Nxm^\mu-Nym^\nu.\end{gathered}$$

that is linear in m^μ and m^ν.

Then we use the inequality which follows from the Laplace method:

$$\begin{gathered}\frac{\int\theta(|y|-|x|-2\delta_N)\exp\{\beta JN(m^\mu x+m^\nu y-\frac{x^2}{2(1+d_N\zeta)}-\frac{y^2}{2})\}dxdy}{\int\exp\{\beta JN(m^\mu x+m^\nu y-\frac{x^2}{2(1+d_N\zeta)}-\frac{y^2}{2})\}dxdy}\ge\\(\theta(|m^\nu|-(1+d_N\zeta)|m^\mu|-3\delta_N)-e^{-N\beta J\delta_N^2/2(2+d_N\zeta)}).\end{gathered}\tag{3.10}$$

Indeed, if $|m^\nu|-(1+d_N\zeta)|m^\mu|-3\delta_N<0$, this inequality is trivial.

If $|m^\nu|-(1+d_N\zeta)|m^\mu|-3\delta_N\ge0$, then the θ-function in the r.h.s. of (3.10) is equal to 1 and the ratio in the l.h.s. is equal to

$$1-\frac{\int_{\mathcal{D}}\exp\{\beta JN(-\frac{(x-x^*)^2}{2(1+d_N\zeta)}-\frac{(y-y^*)^2}{2})\}dxdy}{\int\exp\{\beta JN(-\frac{(x-x^*)^2}{2(1+d_N\zeta)}-\frac{(y-y^*)^2}{2})\}dxdy},$$

where $\mathcal{D}=\{(x,y):|y|-|x|-2\delta_N\le 0\}$ and $x^*=(1+d_N\zeta)m^\mu$, $y^*=m^\nu$. In our case $(|m^\nu|-(1+d_N\zeta)|m^\mu|-3\delta_N\ge 0)$ and so $(x^*,y^*)\notin\mathcal{D}$. According to the Laplace method, the last ratio in the above formula can be estimated from above as follows

$$\frac{\int_{\mathcal{D}}\exp\{\beta JN(-\frac{(x-x^*)^2}{2(1+d_N\zeta)}-\frac{(y-y^*)^2}{2})\}dxdy}{\int\exp\{\beta JN(-\frac{(x-x^*)^2}{2(1+d_N\zeta)}-\frac{(y-y^*)^2}{2})\}dxdy}\le\exp\{\beta JN\max_{\mathcal{D}}[-\frac{(x-x^*)^2}{2(1+d_N\zeta)}-\frac{(y-y^*)^2}{2}]\}\le e^{-N\beta J\delta_N^2/2(2+d_N\zeta)}.$$

Thus, we obtain that

$$a_1^{\mu\nu}=\frac{\int\theta(|y|-|x|-2\delta_N)\exp\{\beta JNF_{N,\mu\nu}(x,y)\}dxdy}{\int\exp\{\beta JNF_{N,\mu\nu}(x,y)\}dxdy}=$$
$$\frac{\sum_{\{\sigma\}}\exp\{-\beta H_{\mu\nu}(\sigma;0,0)\}\int\theta(|y|-|x|-2\delta_N)e^{\beta JN(m^\mu x+m^\nu y-\frac{x^2}{2(1+d_N\zeta)}-\frac{y^2}{2})}dxdy}{2\pi\beta JN(1+\zeta d_N)^{-1/2}\sum_{\{\sigma\}}e^{-\beta H(\sigma)}}\ge$$
$$\frac{\sum_{\{\sigma\}}\exp\{-\beta H_{\mu\nu}(\sigma;0,0)\}\theta(|m^\nu|-(1+d_N\zeta)|m^\mu|-3\delta_N)\int e^{\beta JN(m^\mu x+m^\nu y-\frac{x^2}{2(1+d_N\zeta)}-\frac{y^2}{2})}dxdy}{2\pi\beta JN(1+\zeta d_N)^{-1/2}\sum_{\{\sigma\}}e^{-\beta H(\sigma)}}-$$
$$-e^{-N\beta J\delta_N^2/2(2+d_N\zeta)}=a^{\mu,\nu}-e^{-N\beta J\delta_N^2/2(2+d_N\zeta)}.$$

So,

$$a^{\mu\nu}\le a_1^{\mu\nu}+e^{-N\beta J\delta_N^2/2(2+d_N\zeta)}. \tag{3.11}$$

Now we apply the Laplace method to the integral in the r.h.s. of (3.8) (let us recall that evidently $|\frac{\partial}{\partial x}f_{N,\mu\nu}(x,y)|$, $|\frac{\partial}{\partial y}f_{N,\mu\nu}(x,y)|\le 1$ and therefore $F_{N,\mu\nu}(x,y)$ has bounded derivatives in the domain of interest). We have

$$a_1^{\mu\nu}=\frac{\int_{|y|-|x|\ge 2\delta_N}\exp\{\beta JNF_{N,\mu\nu}(x,y)\}dxdy}{\int\exp\{\beta JNF_{N,\mu\nu}(x,y)\}dxdy}\le$$
$$\exp\{\beta JN[\max_{|y|-|x|\ge 2\delta_N}F_{N,\mu\nu}(x,y)-\max_{x,y}F_{N,\mu\nu}(x,y)]\}. \tag{3.12}$$

We will show below that

$$\max_{|y|-|x|\ge 2\delta_N}F_{N,\mu\nu}(x,y)-\max_{x,y}F_{N,\mu\nu}(x,y)\le-\zeta\frac{d_N\delta_N^2}{2}. \tag{3.13}$$

¿From this inequality, using (3.12) and the fact that $1\le\zeta\le 2$, we get

$$a_1^{\mu\nu}\le e^{-\beta JNd_N\delta_N^2/2}, \tag{3.14}$$

if N is large enough. This inequality together with (3.11) prove (3.3) and thus Lemma 1.

Now let us show (3.13). Denote by $E_{\mu\nu}\{...\}$ the average with respect to all random parameters of the problem except γ^μ, γ^ν and ζ and rewrite $F_{N,\mu\nu}(x,y)$ as

$$F_{N,\mu\nu}(x,y)=E_{\mu\nu}\{F_{N,\mu\nu}(x,y)\}+R_{N,\mu\nu}(x,y),\quad R_{N,\mu\nu}(x,y)\equiv F_{N,\mu\nu}(x,y)-E_{\mu\nu}\{F_{N,\mu\nu}(x,y)\}.$$

Then we have

$$\max_{|y|-|x|\ge 2\delta_N}F_{N,\mu\nu}(x,y)-\max_{x,y}F_{N,\mu\nu}(x,y)\le$$
$$\max_{|y|-|x|\ge 2\delta_N}E_{\mu\nu}\{F_{N,\mu\nu}(x,y)\}-\max_{x,y}E_{\mu\nu}\{F_{N,\mu\nu}(x,y)\}+2\max_{x,y}|R_{N,\mu\nu}(x,y)|\}. \tag{3.15}$$

To proceed further, we use

Proposition 2 *Let $f(x,y)$ be a smooth function, satisfying the symmetry conditions*

$$f(x,y)=f(y,x), \qquad f(x,y)=f(-x,y). \tag{3.16}$$

Consider the function $F(x,y)$ of the form

$$F(x,y)\equiv f(x,y)-\frac{x^2}{2(1+d)}-\frac{y^2}{2} \qquad (d>0). \tag{3.17}$$

Then

$$\max F(x,y)-\max_{|x|\le|y|-2\delta}F(x,y)\ge\frac{2d\delta^2}{(1+d)}. \tag{3.18}$$

We prove this proposition in Appendix.

Let us remark that

$$\max_{|y|-|x|\ge 2\delta_N}E_{\mu\nu}\{F_{N,\mu\nu}(x,y)\}-\max_{x,y}E_{\mu\nu}\{F_{N,\mu\nu}(x,y)\}\le$$

$$\max_{|y|-|x|\ge 2\delta_N}E_{\mu\nu}\{F_{N,\mu\nu}(x,y)\}\Big|_{\gamma^\mu=\gamma^\nu=0}-\max_{x,y}E_{\mu\nu}\{F_{N,\mu\nu}(x,y)\}\Big|_{\gamma^\mu=\gamma^\nu=0}+2(|\frac{\gamma^\mu}{\sqrt{N}}|+|\frac{\varepsilon_1\gamma^\nu}{\sqrt{N}}|).$$

Now we apply Proposition 2 to the function $E_{\mu\nu}\{F_{N,\mu\nu}(x,y)\}\Big|_{\gamma^\mu=\gamma^\nu=0}$ and use also the estimate

$$2(|\frac{\gamma^\mu}{\sqrt{N}}|+|\frac{\varepsilon_1\gamma^\nu}{\sqrt{N}}|)\le\frac{1-d_N\zeta}{1+\zeta d_N}\zeta d_N\delta_N^2. \tag{3.19}$$

Since γ^μ and γ^ν are Gaussian random variables, this estimate is valid with probability larger than $1-e^{-N\,\mathrm{const}\,d_N^2\delta_N^4}$. Thus,

$$\begin{gathered}\max_{|y|-|x|\ge 2\delta_N}E_{\mu\nu}\{F_{N,\mu\nu}(x,y)\}-\max_{x,y}E_{\mu\nu}\{F_{N,\mu\nu}(x,y)\}\le\\ -\frac{2d_N\zeta\delta_N^2}{1+\zeta d_N}+\frac{d_N\zeta\delta_N^2(1-\zeta d_N)}{1+\zeta d_N}=-\zeta d_N\delta_N^2=-\zeta d_N^2\end{gathered} \tag{3.20}$$

with the same probability.

If also

$$2\max_{x,y}|R_{N,\mu\nu}(x,y)|\}\le\zeta\frac{d_N^2}{2}, \tag{3.21}$$

then (3.15) and (3.20) imply (3.13). Thus, we obtain that the probability of (3.14) is bounded from above by the sum of probabilities of (3.19) and (3.21).

Now we are faced with the problem of estimating the probability of (3.21). To this aim let us remark, that since $|\frac{\partial}{\partial x}f_{N,\mu\nu}|$, $|\frac{\partial}{\partial x}f_{N,\mu\nu}|\le 1$, all the extremal points of $F_{N,\mu,\nu}(x,y)$ are inside the square $\{(x,y):|x|,|y|\le 1\}$. Besides, for any (x,y) from this square there exist i,j, $(|i|,|j|\le M$, $M\equiv[4\sqrt{2}d_N^{-2}])$ such that

$$\begin{gathered}f_{N,\mu\nu}(x,y)-E\{f_{N,\mu\nu}(x,y)|\zeta\}=f_{N,\mu\nu}(\tfrac{i}{M},\tfrac{j}{M})-E\{f_{N,\mu\nu}(\tfrac{i}{M},\tfrac{j}{M})|\zeta\}+\\ f_{N,\mu\nu}(x,y)-f_{N,\mu\nu}(\tfrac{i}{M},\tfrac{j}{M})+E\{f_{N,\mu\nu}(\tfrac{i}{M},\tfrac{j}{M})|\zeta\}-E\{f_{N,\mu\nu}(x,y)|\zeta\}=\\ f_{N,\mu\nu}(\tfrac{i}{M},\tfrac{j}{M})-E\{f_{N,\mu\nu}(\tfrac{i}{M},\tfrac{j}{M})|\zeta\}+R_1+R_2,\end{gathered} \tag{3.22}$$

where

$$|R_{1,2}|\le\sqrt{(x-\frac{i}{M})^2+(y-\frac{j}{M})^2}\le\frac{1}{\sqrt{2}M}\le\frac{d_N^2}{8}.$$

Thus, for our goal it suffices to estimate the probability of the event

$$|f_{N,\mu,\nu}(\frac{i}{M},\frac{j}{M}) - E\{f_{N,\mu,\nu}(\frac{i}{M},\frac{j}{M})|\zeta\}| \le \frac{d_N^2}{4}, \tag{3.23}$$

because this inequality and (3.22) imply (3.21).

Since, according to the result [16], the probability of the last event for fixed (i,j) can be estimated as

$$\text{Prob}\left\{|f_{N,\mu\nu}(x_k,y_k) - E\{f_{N,\mu\nu}(x_k,y_k)|\zeta\}| \le \frac{d_N^2}{4}\right\} \ge 1 - D_n N^{-n},$$

and the number of these events is $4M^2$, the probability of inequality (3.21) is more than $1-4M^2D_nN^{-n} \ge 1-C_n'N^{2-n}$. On the other hand, since γ^ν,γ^μ are Gaussian random variables, the probability of (3.19) is more then $1-e^{-N\,\text{const}\,d_N^2}$. Therefore the inequalities (3.11)-(3.23) and the last conclusions prove (3.3), that, as it was mentioned above, implies (2.5).

For the perturbed Hamiltonian the proof is the same.

Proof of Proposition 1

To simplify formulae we prove formula (2.9) in the case $l=1$. The general case for this formula and also its modification for the perturbed Hamiltonian can be proved similarly.

Let us introduce the vector $\overline{\theta}\in\mathbf{R^k}$, $\overline{\theta}\equiv(\theta_1,\ldots,\theta_k)$, and define $H(\overline{\theta})$ to be equal to the Hamiltonian H_*^a , if we substitute in the latter $\xi_1^{\mu_1},\ldots,\xi_1^{\mu_k}$ by $\theta_1\xi_1^{\mu_1},\ldots,\theta_k\xi_1^{\mu_k}$. Set

$$\phi(\theta_1,\ldots,\theta_k)\equiv N^{k/2}\langle t_1^{\mu_1}\ldots t_1^{\mu_k}F_1(\{\sigma\},\{\xi_i^\mu\})\rangle_{H(\overline{\theta})}.$$

and consider

$$\begin{array}{c} E\{\int_0^1\ldots\int_0^1 d\theta_1\ldots d\theta_k\xi_1^{\mu_1}\ldots\xi_1^{\mu_k}\frac{\partial^k}{\partial\theta_1\ldots\partial\theta_k}\phi(\theta_1\ldots\theta_k)\} = \\ E\{\int_0^1\ldots\int_0^1 \theta_2\ldots d\theta_k\xi_1^{\mu_1}\ldots\xi_1^{\mu_k}\frac{\partial^{(k-1)}}{\partial\theta_2\ldots\partial\theta_k}(\phi(1,\theta_2\ldots\theta_k)-\phi(0,\theta_2\ldots\theta_k))\}. \end{array}$$

Since $\phi(0,\theta_2\ldots\theta_k)$ does not depend on $\xi_1^{\mu_1}$, the second term in the l.h.s. of the last relation is zero after averaging with respect to $\xi_1^{\mu_1}$. Repeating this procedure $k-1$ times more, we get

$$E\{\int_0^1\cdots\int_0^1 d\theta_1\ldots d\theta_k\xi_1^{\mu_1}\ldots\xi_1^{\mu_k}\frac{\partial^k}{\partial\theta_1\ldots\partial\theta_k}\phi(\theta_1\ldots\theta_k)\} = E\{\xi_1^{\mu_1}\ldots\xi_1^{\mu_k}\phi(1,\ldots,1)\}.$$

Therefore

$$\begin{array}{c} E\{\xi_1^{\mu_1}\ldots\xi_1^{\mu_k}\phi(1,\ldots,1)\} = E\{\xi_1^{\mu_1}\ldots\xi_1^{\mu_k}\frac{\partial^k}{\partial\theta_1\ldots\partial\theta_k}\phi(1\ldots1)\} + R_{\mu_1\ldots\mu_k}(\phi) = \\ N^{k/2}E\{\frac{\partial^k}{\partial\xi_1^{\mu_1}\ldots\partial\xi_1^{\mu_k}}\langle t_1^{\mu_1}\ldots t_1^{\mu_k}F_1(\{\sigma\},\{\xi_i^\mu\})\rangle_{H(\overline{1})}\} + R_{\mu_1\ldots\mu_k}(\phi), \end{array} \tag{3.24}$$

where

$$\begin{array}{c} |R_{\mu_1\ldots\mu_k}(\phi)| \le N^{k/2}E\{\max_{|\theta_1|,\ldots,|\theta_k|\le1}\frac{\partial^{k+1}}{\partial^2\theta_1\ldots\partial\theta_k}\phi(\theta_1\ldots\theta_k)\} + \ldots \\ N^{k/2}E\{\max_{|\theta_1|,\ldots,|\theta_k|\le1}\frac{\partial^{k+1}}{\partial\theta_1\ldots\partial^2\theta_k}\phi(\theta_1\ldots\theta_k)\} \end{array} \tag{3.25}$$

Now we should remark that since the Hamiltonian H_*^a depends on ξ_1^μ via the expression $N^{-1/2}\xi_1^\mu\sigma_1(Jt_1^\mu+\varepsilon_1\gamma^\mu)$, after differentiation with respect to ξ_1^μ this expression will appear in several places. Thus, we obtain

$$\begin{array}{c} R_{\mu_1,\ldots,\mu_k} \le \text{ const}\,E\{N^{-1/2}\langle|t_1^{\mu_1}|(|t_1^{\mu_1}|+\varepsilon_1|\gamma^{\mu_1}|)^2\cdot \\ |t_1^{\mu_2}|(|(t_1^{\mu_2}|+\varepsilon_1|\gamma_{\mu_2}|)\ldots|t_1^{\mu_k}|(|(t_1^{\mu_k}|+\varepsilon_1|\gamma_{\mu_k}|)|F_1(\{\sigma\},\{\xi_i^\mu\})|\rangle_{H_*^a+S^1_{\mu_1,\ldots,\mu_k}}\}+\ldots+ \\ E\{N^{-1/2}\langle|t_1^{\mu_1}|(|(t_1^{\mu_1}|+\varepsilon_1|\gamma_{\mu_1}|)\cdot \\ |t_1^{\mu_2}|(|(t_1^{\mu_2}|+\varepsilon_1|\gamma_{\mu_2}|)\ldots|t_1^{\mu_k}|(|(t_1^{\mu_k}|+\varepsilon_1|\gamma_{\mu_k}|)^2|F_1(\{\sigma\},\{\xi_i^\mu\})|\rangle_{H_*^a+S^k_{\mu_1,\ldots,\mu_k}}\} \end{array} \tag{3.26}$$

with $S^i_{\mu_1,\dots,\mu_k}$ of the form

$$S^i_{\mu_1,\dots,\mu_k}=\frac{\theta_1(\mu_1,\dots,\mu_k)}{\sqrt{N}}t_1^{\mu_1}\sigma_1+...+\frac{\theta_k(\mu_1,\dots,\mu_k)}{\sqrt{N}}t_1^{\mu_k}\sigma_1$$

By using the inequalities $|S^i_{\mu_1,\dots,\mu_k}|\le k$ and then (2.6) and (2.7), we get the estimate

$$\begin{gathered}|R_{\mu_1,\dots,\mu_k}|\le \text{ const } d_N E\{\langle((t_1^{\mu_1})^2+\varepsilon_1^2\gamma_{\mu_1}^2)((t_1^{\mu_2})^2+\varepsilon_1^2\gamma_{\mu_2}^2)\dots((t_1^{\mu_k})^2+\varepsilon_1^2\gamma_{\mu_k}^2)\rangle\}=\\ \text{const } d_N^{1/2}E\{p^{-k}\textstyle\sum_{\mu_1,\dots,\mu_k=s+1}^p\langle((t_1^{\mu_1})^2+\varepsilon_1^2\gamma_{\mu_1}^2)((t_1^{\mu_2})^2+\varepsilon_1^2\gamma_{\mu_2}^2)\dots((t_1^{\mu_k})^2+\varepsilon_1^2\gamma_{\mu_k}^2)\rangle\}\le\\ \text{const } d_N^{1/2}E\{(\alpha||\mathcal{J}||+\alpha\varepsilon_1^2)^k\}=O(d_N^{1/2}).\end{gathered}\tag{3.27}$$

Proposition 1 is proved.

Proof of Lemma 2
It is easy to see that for any $\mathbf{c}\in\mathbf{R^s}$

$$H^a(\mathbf{c})-H=\frac{J(1+\zeta d_N)N}{2}\sum_{\nu=1}^s(m^\nu-c^\nu)^2.$$

Then, on the basis of the Bogolyubov inequality

$$\frac{1}{N}\langle H_N^{(2)}-H_N^{(1)}\rangle_{H_2}\le f_N(H_2)-f_N(H_1)\le\frac{1}{N}\langle H_N^{(2)}-H_N^{(1)}\rangle_{H_1},\tag{3.28}$$

we have for any $\mathbf{c}\in\mathbf{R^s}$

$$\frac{J(1+\zeta d_N)}{2}\sum_{\nu=1}^s\langle(m^\nu-c^\nu)^2\rangle_{H^a(\mathbf{c})}\le f_N(H^a(\mathbf{c}))-f_N(H)\le\frac{J(1+\zeta d_N)}{2}\sum_{\nu=1}^s\langle(m^\nu-c^\nu)^2\rangle.\tag{3.29}$$

Taking the minimum with respect to all $\mathbf{c}$ and averaging with respect to all random parameters of the problem, except ζ, we get

$$\begin{gathered}\frac{J(1+\zeta d_N)}{2}E\{\langle\sum_{\nu=1}^s(m^\nu-\tilde{c}^\nu)^2\rangle_{H^a(\tilde{\mathbf{c}})}|\zeta\}\le E\{\min_{c^\nu}f_N(H^a(\mathbf{c}))|\zeta\}-E\{f_N(H)|\zeta\}\le\\ \frac{J(1+\zeta d_N)}{2}E\{\langle\sum_{\nu=1}^s(m^\nu-\langle m^\nu\rangle)^2\rangle|\zeta\},\end{gathered}\tag{3.30}$$

where $\tilde{\mathbf{c}}$ is a random minimum point of the function $f_N(H^a(\mathbf{c}))$. Integrating by parts with respect to the variables $\{\gamma^\nu\}_{\nu=1}^s$, it is easy to obtain that

$$\begin{gathered}E\{\langle\frac{1}{2}\sum_{\nu=1}^s(m^\nu-\langle m^\nu\rangle)^2\rangle|\zeta\}=E\{\frac{1}{2\beta\sqrt{N}}\sum_{\nu=1}^s\gamma^\mu\langle m^\nu\rangle|\zeta\}\le\\ E^{1/2}\{\frac{1}{4N}\sum_{\nu=1}^s(\gamma^\nu)^2\}E^{1/2}\{\frac{1}{N}\sum_{\nu=1}^s\langle t^\nu\rangle^2\}\le(\frac{s}{p})^{1/2}E^{1/2}\{||\mathcal{J}||\}.\end{gathered}$$

Substituting this estimate in (3.30), we get

$$\frac{J(1+\zeta d_N)}{2}E\{\langle\sum_{\nu=1}^s(m^\nu-\tilde{c}^\nu)^2\rangle_{H^a(\tilde{\mathbf{c}})}|\zeta\}\le E\{\min_{c^\nu}f_N(H^a(\mathbf{c}))|\zeta\}-E\{f_N(H)|\zeta\}\le\text{ const }(\frac{s}{p})^{1/2}.\tag{3.31}$$

Now to prove Lemma 2 we are left to prove that

$$|E\{\min_{c^\nu}f_N(H^a(\mathbf{c}))|\zeta\}-\min_{c^\nu}E\{f_N(H^a(\mathbf{c}))|\zeta\}|=o(1),\quad N\to\infty.\tag{3.32}$$

Since for fixed N $f_N(H^a(\mathbf{c}))$ and $E\{f_N(H^a(\mathbf{c}))|\zeta\}$ tend evidently to infinity, as $\mathbf{c} \to \infty$, these functions assume their minimal values at the finite extremal points. But the conditions of extremum for these functions have the form $c^\nu = \langle m^\nu \rangle_{H^a(\mathbf{c})}$ and $c^\nu = E\{\langle m^\nu \rangle_{H^a(\mathbf{c})}|\zeta\}$ and since $|m^\nu| \le 1$ to prove (3.32) it is enough to prove that

$$\text{Prob}\{X\} \equiv \text{Prob}\left\{ \sup_{|c^\nu|\le 1} |f_N(H^a(\mathbf{c})) - E\{f_N(H^a(\mathbf{c}))|\zeta\}| \ge \frac{2J}{\log N} \right\} = o(1). \tag{3.33}$$

By using once more the fact that the derivatives of the functions $f_N(H^a(\mathbf{c}))$ and $E\{f_N(H^a(\mathbf{c}))|\zeta\}$ with respect to c^ν are bounded, we obtain that

$$|f_N(H^a(\mathbf{c}_1)) - f_N(H^a(\mathbf{c}_2))| \le J\sqrt{s}(\sum (c_1^\nu - c_2^\nu)^2)^{1/2}$$

and so, if $|c_1^\nu - c_2^\nu| \le (2k)^{-1}$ ($k \equiv [s \log N]$), then

$$|f_N(H^a(\mathbf{c}_1)) - f_N(H^a(\mathbf{c}_2))| \le \frac{Js}{2k} \le \frac{J}{2\log N}.$$

Therefore

$$\sup_{|c^\nu|\le 1} |f_N(H^a(\mathbf{c})) - E\{f_N(H^a(\mathbf{c}))|\zeta\}| \le$$
$$\sup_{|j_1|,..,|j_s|\le k} |f_N(H^a(\frac{j_1}{k},\dots,\frac{j_s}{k})) - E\{f_N(H^a(\frac{j_1}{k},\dots,\frac{j_s}{k}))|\zeta\}| + \frac{J}{\log N}.$$

Thus,

$$\text{Prob}\{X\} \le \sum_{j_1,\dots,j_s} \text{Prob}\{X_{j_1,\dots,j_s}\}, \tag{3.34}$$

where $j_1, \dots . j_s$ are integer numbers from the interval $(-k,k)$ and $X_{j_1,\dots,j_s}$ is the notation of the event

$$|f_N(H^a(\frac{j_1}{k},\dots,\frac{j_s}{k})) - E\{f_N((\frac{j_1}{k},\dots,\frac{j_s}{k}))|\zeta\}| \ge \frac{J}{\log N}.$$

According to [16], the probability of the last event can be estimated by $D_1 N^{-1}\log^2 N$. Thus, on the basis of (3.34),

$$\text{Prob}\{X\} \le D_1 N^{-1}(2k+1)^s \le D_1 \exp\{s\log(2k+1) + 2\log\log N - logN\} =$$
$$D_1 \exp\{\,\text{const}\,[\log N]^{1/2}(\log\log N) - \log N\} \le N^{-1/2},$$

and we obtain (3.32), that joined with (3.31) proves (2.12).

Now let us prove relations (2.13). To this end we use Lemma 3. Using this lemma and (2.12), we obtain

$$\beta E\{(1-q_N^*)|\zeta\} = -E\{\tfrac{\partial f_N(H^a)}{\partial \varepsilon}|\zeta\} = -E\{\tfrac{\partial f_N(H)}{\partial \varepsilon}|\zeta\} + o(1) = \beta E\{(1-q_N)|\zeta\} + o(1)$$
$$\beta E\{(\langle U\rangle_* - \alpha r_N^*)|\zeta\} = -E\{\tfrac{\partial f_N(H^a)}{\partial \varepsilon_1}|\zeta\} = -E\{\tfrac{\partial f_N(H)}{\partial \varepsilon_1}\} + o(1) = \beta E\{(\langle U\rangle - \alpha r_N)|\zeta\} + o(1)$$
$$E\{\langle U_N\rangle_*|\zeta\} + \sum_{\mu=1}^{s}\int_1^2 d\zeta (c_*^\nu(\zeta))^2 = -2E\{\frac{\partial f_N(H^a)}{\partial J}|\zeta\} =$$
$$-2E\{\frac{\partial f_N(H)}{\partial J}|\zeta\} + o(1) = E\{\langle U_N\rangle + \sum_{\mu=1}^{s}\langle (m^\mu)^2\rangle\} + o(1)$$

(3.35)

Remark also, that

$$\sum_{\mu=1}^{s}\int_1^2 d\zeta (c_*^{\nu}(\zeta)^2) = -2\frac{1}{d_N}\int_1^2 d\zeta E\{\frac{\partial f_N(H^a)}{\partial\zeta}|\zeta\} = -2\frac{1}{d_N}E\{(f_N(H^a)|_{\zeta=2} - f_N(H^a)|_{\zeta=1})|\zeta\} =$$
$$-2\frac{1}{d_N}E\{(f_N(H)|_{\zeta=2} - f_N(H)|_{\zeta=1})|\zeta\} + o(1) = -2\frac{1}{d_N}\int_1^2 d\zeta E\{\frac{\partial f_N(H)}{\partial\zeta}|\zeta\} + O(\frac{1}{\log N d_N})$$
$$= E\{\sum_{\mu=1}^{s}\langle(m^{\mu})^2\rangle\} + o(1)$$

This relations, joint to (3.35) prove (2.13).

4 Appendix

Proof of Proposition 2

Due to the symmetry of the problem we can restrict ourselves by the case when x and y are positive.

If $x > 0$, $y > x + 2\delta$ consider $x' = y$, $y' = x$. Then

$$\begin{aligned} F(x', y') - F(x, y) = f(y, x) - \tfrac{y^2}{2(1+d)} - \tfrac{x^2}{2} - \\ -f(x, y) + \tfrac{x^2}{2(1+d)} + \tfrac{y^2}{2} = \tfrac{d}{2(1+d)}(y^2 - x^2) \geq 4\delta^2\tfrac{d}{2(1+d)} \end{aligned} \tag{4.1}$$

Proposition 2 is proved.

Proof of Lemma 3.

Denote

$$d_n^{(1)} \equiv \max_{a\le t\le b} E\{[f_n(t) - E\{f_n(t)\}]^2\}, \quad d_n^{(2)} \equiv \max_{a\le t\le b} E\{[g_n(t) - E\{g_n(t)\}]^2\},$$
$$d_n^{(3)} \equiv \max_{a\le t\le b} |E\{f_n(t)\} - E\{g_n(t)\}|, \quad \varepsilon_n = [\max\{d_n^{(1)}, d_n^{(2)}, d_n^{(3)}\}]^{1/3}.$$

Then, using the convexity of $f_n(t)$, we have

$$\begin{aligned} f_n'(t) - E\{f_n'(t)\} \leq \frac{f_n(t+\varepsilon_n) - f_n(t)}{\varepsilon_n} - E\{f_n'(t)\} = -\frac{f_n(t) - E\{f_n(t)\}}{\varepsilon_n} + \\ \frac{f_n(t+\varepsilon_n) - E\{f_n(t+\varepsilon_n)\}}{\varepsilon_n} + \Big[\frac{E\{f_n(t+\varepsilon_n)\} - E\{f_n(t)\}}{\varepsilon_n} - E\{f_n'(t)\}\Big]. \end{aligned} \tag{4.2}$$

Denote also

$$R_n^+(t) \equiv \frac{E\{f_n(t+\varepsilon_n)\} - E\{f_n(t)\}}{\varepsilon_n} - E\{f_n'(t)\}$$

and prove that $R_n^+(t) \to 0$ for almost all t. To this end we study

$$\int_{a_1}^{b_1} R_n^+(t)dt = \Big[\frac{F_n(b_1+\varepsilon_n) - F_n(b_1)}{\varepsilon_n} - F_n'(b_1)\Big] - \Big[\frac{F_n(a_1+\varepsilon_n) - F_n(a_1)}{\varepsilon_n} - F_n'(a_1)\Big], \tag{4.3}$$

where $F_n(t) \equiv \int_a^t E\{f_n(\tau)\}d\tau$ and (a_1, b_1) is some subinterval of (a, b). It is evident, that

$$0 \leq F_n''(t) = E\{f_n'(t)\} \leq 2(b - b_1)^{-1}[E\{f_n(b)\} - E\{f_n(\frac{b+b_1}{2})\}].$$

Therefore if n is large enough to provide $\varepsilon_n \leq \frac{b-b_1}{2}$, then, according to the Taylor formula, the r.h.s. of (4.3) is of order $O(\varepsilon_n)$. Thus, since $R_n^+(t) \geq 0$, it follows from (4.3) that $R_n^+(t) \to 0$ for almost all t.

On the other hand, similarly to (4.2) we get that

$$f_n'(t) - E\{f_n'(t)\} \geq -\frac{f_n(t-\varepsilon_n) - E\{f_n(t-\varepsilon_n)\}}{\varepsilon_n} + \frac{f_n(t) - E\{f_n(t)\}}{\varepsilon_n} + R_n^-(t)$$

and $R_n^-(t) \equiv \frac{E\{f_n(t)\}-E\{f_n(t-\varepsilon_n)\}}{\varepsilon_n} - E\{f_n'(t)\} \to 0$ for almost all t. Then

$$E\{[f_n'(t) - E\{f'(t)\}]^2\} \le 2E\left\{\left(\frac{f_n(t+\varepsilon_n) - E\{f_n(t+\varepsilon_n)\}}{\varepsilon_n}\right)^2\right\} +$$
$$2E\left\{\left(\frac{f_n(t-\varepsilon_n) - E\{f_n(t-\varepsilon_n)\}}{\varepsilon_n}\right)^2\right\} + 4E\left\{\left(\frac{f_n(t) - E\{f_n(t)\}}{\varepsilon_n}\right)^2\right\}+$$
$$+(R_n^-(t))^2 + (R_n^+(t))^2 \le 8\varepsilon_n + (R_n^-(t))^2 + (R_n^+(t))^2 \to 0$$

for almost all t.

By the same way,

$$E\{[g_n'(t) - E\{f_n'(t)\}]^2\} \le 2E\left\{\left(\frac{g_n(t+\varepsilon_n) - E\{f_n(t+\varepsilon_n)\}}{\varepsilon_n}\right)^2\right\} +$$
$$2E\left\{\left(\frac{g_n(t-\varepsilon_n) - E\{f_n(t-\varepsilon_n)\}}{\varepsilon_n}\right)^2\right\} + 4E\left\{\left(\frac{g_n(t) - E\{f_n(t)\}}{\varepsilon_n}\right)^2\right\} + (R_n^-(t))^2 + (R_n^+(t))^2 =$$
$$2E\left\{\left(\frac{g_n(t+\varepsilon_n) - E\{g_n(t+\varepsilon_n)\}}{\varepsilon_n}\right)^2\right\} + 2E\left\{\left(\frac{g_n(t-\varepsilon_n) - E\{g_n(t-\varepsilon_n)\}}{\varepsilon_n}\right)^2\right\} +$$
$$4E\left\{\left(\frac{g_n(t) - E\{g_n(t)\}}{\varepsilon_n}\right)^2\right\} + 2\left(\frac{E\{g_n(t+\varepsilon_n)\} - E\{f_n(t+\varepsilon_n)\}}{\varepsilon_n}\right)^2 +$$
$$2\left(\frac{E\{g_n(t-\varepsilon_n)\} - E\{f_n(t-\varepsilon_n)\}}{\varepsilon_n}\right)^2 + 4\left(\frac{E\{g_n(t)\} - E\{f_n(t)\}}{\varepsilon_n}\right)^2 +$$
$$(R_n^-(t))^2 + (R_n^+(t))^2 \le 16\varepsilon_n + (R_n^-(t))^2 + (R_n^+(t))^2. \tag{4.4}$$

On the other hand,

$$E\{[g_n'(t) - E\{f_n'(t)\}]^2\} = E\{[g_n'(t) - E\{g_n'(t)\}]^2\} + [E\{g_n'(t)\} - E\{f_n'(t)\}]^2.$$

Therefore (4.4) proves (2.15) and (2.16)

References

[1] L. Pastur, M. Shcherbina, B. Tirozzi. 'The Replica-Symmetric Solution Without Replica Trick for the Hopfield Model'. *J. Stat. Phys.*, **74**, 5/6, 1161-1183, (1994).

[2] L.Pastur, A.Figotin. 'Exactly Soluble Model of the Spin Glass'. *Soviet.Phys.J.E.T.P.*, **25**, 348-353, (1977).

[3] J. Hopfield. 'Neural Networks and Physical Systems with Emergent Collective Computational Abilities'. *Proc.Nat.Ac.Sci.*, **79**, 2554-2558, (1982).

[4] M. Mezard, G. Parisi, M. A. Virasoro. 'Spin Glass Theory and Beyond', World Scientific, (1987).

[5] D.J.Amit 'Modelling Brain Functions' Cambridge Univ. Press, (1989).

[6] Mathematical Aspects of Spin Glasses and Neural Networks. A. Bovier, P.Picco (Eds), Birkhauser, 1998.

[7] M. Talagrand. 'Rigorous Results for the Hopfield Models with Many Patterns. *Prob. Theor. Rel. Fields*, **110**, 109-176, (1998)

[8] L. Pastur, M. Shcherbina. 'Absence of Self-Averaging of the Order Parameter in the Sherrington-Kirkpatrick Model'. *J.Stat.Phys.*, **62**, 1-26, (1991).

[9] D. Sherrington, S. Kirkpatrick. 'Infinite Models of Spin Glasses' *Phys. Rev. B*, **17**, 11, 4384-4403, (1978).

[10] M. Shcherbina. 'More About the Absence of Self averaging of the Order Parameter for the Sherrington-Kirkpatrick Model'. *CARR Preprint* 3/91, Rome, (1991).

[11] P.C.Hemmer, J.L.Lebowitz. 'Systems with Weak Long-Range Potential' In: Phase Transition and Critical Phenomena, vol.2, M.S.Green, C.Domb (Eds.), Academic Press, New York (1973).

[12] L.Pastur, M.Shcherbina. 'Long Range Limit for the Correlation Functions of Lattice Systems'. *Teor.Mat.Fiz.* **61**, 955-964, (1984).

[13] L.Pastur, A.Figotin. 'Infinite Range Limit for a class of Disordered Spin Systems'. *Teor.Mat.Fiz.*, **51**, 564-589, (1982).

[14] Shcherbina M.. On the Replica Symmetric Solution for the Sherrington-Kirkpatrick Model. Helvetica Physica Acta, **70**, p.772-797, (1997).

[15] M. Shcherbina, B. Tirozzi. 'The Free Energy of a Class of Hopfield Models'. *J. of Stat. Phys.*, **72** 1/2, 113-125, (1993).

[16] J. Feng, B. Tirozzi. 'The SLLN for Free-Energy of a Class of Neural Works'. *Helvetica Pysica Acta*, **68**, 4, 367-379, (1995).