Some mathematical problems of neural networks theory

M.Shcherbina

Institute for Low Temperature Physics,Ukr. Ac. Sci., 47 Lenin ave., Kharkov, Ukraine

Abstract

We discuss some problems of the dynamics of neural networks, in particular, the rigorous results on the critical capacity of the Hopfield model, the cavity method for spin glass and the rigorous solution of the Gardner problem.

1 Critical capacity of neural network models

The spin glass and neural network theories are of considerable importance and interest for a number of branches of theoretical and mathematical physics (see [MPV] and references therein). Among many topics of interest the analysis of the different models of neural network dynamics is one of the most important because of its numerical links with computer applications, in particular, with the models of the associative memory. To discuss these models let us start from a simple example.

Suppose that we have p chosen "patterns" and we want to teach computer to recognize them, when they are slightly modified. Fore example, these patterns can be hand-written letters, some sequences of sounds ("words"), some pictures of people etc. Let us map these patterns into the sequences of ± 1 of length N (usually N is very big): $\boldsymbol{\xi^{(\mu)}} = (\xi_1^{(\mu)}, \dots, \xi_N^{(\mu)})$ $(\xi_i^{(\mu)} = \pm 1)$ $(\mu = 1, \dots, p)$.

Now consider the set $\Sigma_N = \{\boldsymbol{\sigma} \in \mathbb{R}^N, \sigma_i = \pm 1\}$ of all possible sequences of ± 1 of length N with a usual distance:

$$||\boldsymbol{\sigma} - \boldsymbol{\sigma}'||^2 \equiv \sum_{i=1}^{N} (\sigma_i - \sigma_i')^2.$$

Consider also some initial configuration $\boldsymbol{\sigma}(0) = (\sigma_1(0), \dots, \sigma_N(0)) \in \Sigma_N$ which is close enough to one of chosen patterns, e.g. to $\boldsymbol{\xi^{(1)}}$

$$||\boldsymbol{\sigma}(0) - \boldsymbol{\xi^{(1)}}|| \leq \varepsilon_0 N, \tag{1.1}$$

Our goal is to introduce some sequential dynamics on Σ_N in such a way to be sure that, starting from any $\boldsymbol{\sigma}(0)$, satisfying the above condition we arrive at the end into $\boldsymbol{\xi^{(1)}}$. Moreover, we want to have the same property of our dynamics around any of chosen patterns.

One of the most popular ways to introduce the sequential dynamics on Σ_N is the following:

$$\sigma_k(t+1) = \operatorname{sign}\left\{ \sum_{j=1, j\neq k}^{N} J_{kj}\sigma_j(t) \right\}, \tag{1.2}$$

where the matrix of interactions $\{J_{ij}\}$ (not necessary symmetric) depends on the concrete model.

In the mathematical models usually $\{\boldsymbol{\xi^{(\mu)}}\}_{\mu=1}^{p}$ are chosen i.i.d. random vectors with i.i.d. components $\xi_i^{(\mu)}$ $(i = 1, \dots, N)$, assuming values ± 1 and $E\{\xi_i^{(\mu)}\} = 0$. Here and below we denote by $E\{\dots\}$ the averaging with respect to all random variables of the problem.

Now the problem of storing p chosen patterns is transformed into the following mathematical problem:

Problem 1: To introduce interactions $\{J_{ij}\}_{i,j=1}^N$ *in such a way that chosen random independent vectors* $\{\boldsymbol{\xi}^{(\mu)}\}_{\mu=1}^p$ *(patterns) with i.i.d. components* $\xi_i^{(\mu)} = \pm 1$ *are the fixed points of the dynamics (1.2).*

To analyze dynamics (1.2) usually it is useful to consider also the energy function (the Hamiltonian)

$$\mathcal{H}(\boldsymbol{\sigma}) = -\frac{1}{2}\sum_{j\neq k}^{N} J_{jk}\sigma_j\sigma_k.$$

It is easily seen that the function $\mathcal{H}(\boldsymbol{\sigma}(t))$ does not increase in the process of evolution. Thus, the dynamics of the model depends on the "energy landscape" of the function $\mathcal{H}(\boldsymbol{\sigma})$ and the local minima of the function are the fixed points of dynamics.

Hopfield [H] was one of the first who proposed a possible form of the matrix J_{jk}. He introduced the matrix of the form

$$J_{jk} = \frac{1}{N}\sum_{\mu=1}^{p} \xi_j^{(\mu)}\xi_k^{(\mu)}, \tag{1.3}$$

where $\boldsymbol{\xi}^{(\mu)}$ $(\mu = 1, \ldots, p)$ are chosen patterns.

It is easy to check that if p is finite and $N \to \infty$, then with probability 1, starting from any points close enough to any of the patterns $\xi^{(\mu)}$ (see condition (1.1)), the dynamics (1.2) will stop at the point $\xi^{(\mu)}$.

The next problem which it is natural to study is the possibility for p to increase together with N, so that our dynamical system has the same property as $N, p \to \infty$. How big can be p comparing with N? This problem was solved by Mac Eliece et al[McEPRV]. They proved that if $p < N/4\log N$, then $\{\boldsymbol{\xi}^{(\mu)}\}_{\mu=1}^p$ are the fixed points of the dynamics (1.2) for the Hopfield model. If $p >> N/\log N$, $\{\boldsymbol{\xi}^{(\mu)}\}_{\mu=1}^p$ are not fixed points of (1.2).

Let us transform a bit *Problem 1.* We assume now that the fixed points of dynamics do not coincide exactly with $\boldsymbol{\xi}^{(\mu)}$ but are situated in such a small neighborhoods of $\boldsymbol{\xi}^{(\mu)}$ that anyway we can distinguish different patterns.

Numerical computations shows that such a picture is observed if $p/N \to \alpha$, as $p, N \to \infty$, and $\alpha < \alpha_c$, $\alpha_c \sim 0.138....$ For these α the dynamics (1.2), starting from any initial point $\boldsymbol{\sigma}(0)$ close enough to one of the patterns $\boldsymbol{\xi}^{(\mu)}$ converges quickly to some fixed point, which is very close to $\boldsymbol{\xi}^{(\mu)}$. And if $\alpha > \alpha_c$ the dynamics (1.2) becomes chaotic.

The first rigorous result for the Hopfield model with $p/N \to \alpha$, as $p, N \to \infty$ was obtained by Newman [N]. He proved that for $\alpha \leq 0.056...$, an "energy barrier" exists with probability 1 around every pattern $\boldsymbol{\xi}^{(\mu)}$, i.e. there exist some positive numbers $\delta < 1/2$ and ε, such that for any $\boldsymbol{\sigma}$, belonging to

$$\Omega_\delta^\mu \equiv \{\boldsymbol{\sigma} : ||\boldsymbol{\sigma} - \boldsymbol{\xi}^{(\mu)}||^2 = 4[\delta N]\}$$

the inequality holds

$$\mathcal{H}(\boldsymbol{\sigma}) - \mathcal{H}(\boldsymbol{\xi}^{(\mu)}) \geq \varepsilon N$$

To illustrate the methods of the field let us explain in a few lines the idea of the proof. We denote $\mathcal{A} = \left\{ \min_{\boldsymbol{\sigma}\in\Omega_\delta^1} (\mathcal{H}(\boldsymbol{\sigma}) - \mathcal{H}(\boldsymbol{\xi}^{(\mu)})) \geq \varepsilon N \right\}$ and show that

$$\text{Prob}\{\overline{\mathcal{A}}\} \leq e^{-NC}. \tag{1.4}$$

It is easy to see that

$$\begin{aligned}\mathrm{Prob}\{\overline{\mathcal{A}}\} = \mathrm{Prob}\{\cup_{\boldsymbol{\sigma}\in\Omega_\delta^1}\{\mathcal{H}(\boldsymbol{\sigma}) - \mathcal{H}(\boldsymbol{\xi^{(1)}}) \le \varepsilon N\}\} &\le \sum_{\boldsymbol{\sigma}\in\Omega_\delta^1} \mathrm{Prob}\{\mathcal{H}(\boldsymbol{\sigma}) - \mathcal{H}(\boldsymbol{\xi^{(1)}}) \le \varepsilon N\} \\ &= C_N^{[\delta N]}\mathrm{Prob}\{\mathcal{H}(\boldsymbol{\sigma}^{(1,\delta)}) - \mathcal{H}(\boldsymbol{\xi^{(1)}}) \le \varepsilon N\}\end{aligned}$$

where $\boldsymbol{\sigma}^{(1,\delta)}$ is some fixed point in Ω_δ^μ, e.g.

$$\sigma_k^{(1,\delta)} = -\xi_k^1, \ \ (k = 1, \dots, [\delta N]), \quad \sigma_k^{(1,\delta)} = \xi_k^1, \ \ (k = 1 + [\delta N], \dots, N)$$

According to the Chebyshev inequality

$$\begin{aligned}&\mathrm{Prob}\{\mathcal{H}(\boldsymbol{\sigma}^{(1,\delta)}) - \mathcal{H}(\boldsymbol{\xi^{(1)}}) \le \varepsilon N\} \\ &\qquad \le \min_{\lambda>0} E\left\{\exp\left\{\lambda\left(-\mathcal{H}(\boldsymbol{\sigma}^{(1,\delta)}) + \mathcal{H}(\boldsymbol{\xi^{(1)}}) + \varepsilon N\right)\right\}\right\} \\ &= \min_{\lambda>0}\prod_{\mu=1}^{p} E\left\{\exp\left\{\lambda\left(N^{-1/2}\sum_{i=1}^{[\delta N]}\xi_i^{(\mu)}\right)\left(N^{-1/2}\sum_{j=[\delta N]+1}^{N}\xi_i^{(\mu)}\right) + \lambda\varepsilon\right\}\right\} \\ &\qquad = \exp\{-N(\phi(\delta,\varepsilon) + o(1))\}\end{aligned}$$

It is evident that if for some α there exist δ, ε such that $\phi(\delta,\varepsilon) > N^{-1}\log C_N^{(\delta N)} \sim \delta\log\delta + (1-\delta)\log(1-\delta)$, then the the inequality (1.4) holds and so, according to the Borrel-Cantelli lemma, the energy barrier exists with probability 1. One can show, that if such a "barrier" exists, then inside each open ball

$$B_\delta^\mu \equiv \{\boldsymbol{\sigma} : ||\boldsymbol{\sigma} - \boldsymbol{\xi^{(\mu)}}||^2 < 2[\delta N]\}, \quad (\mu = 1, \dots, p)$$

there exists a point of local minimum of the function $\mathcal{H}(\boldsymbol{\sigma})$, which, as it was mentioned above, is the fixed point of dynamics (1.2).

This result was improved by Loukianova [L]. Using a similar method, she proved the existence of the energy barriers for $\alpha \le 0.071$, so $\alpha_c \ge 0.071$.... Then this result was improved a little by Talagrand [T1].

In the paper [FST] a novel approach of study α_c was introduced. This approach is based upon analysis of the Fourier transform of the joint distribution of the effective fields

$$\tilde{x}_k \equiv N^{-1}\sum_{\mu=1}^{p}\sum_{j=1}^{p}\xi_k^{(\mu)}\xi_j^{(\mu)}, \tag{1.5}$$

It enables us to obtain a new bound for the critical capacity ($\alpha_c \ge 0.113..$) and also allows us to find an asymptotic (for small α) behavior of the distance between the fixed points and the nearest patterns.

The idea of the method is to study the probability that the point $\boldsymbol{\sigma}^{(1,\delta)} \in \Omega_\delta^1$ is a local minimum of the function $\mathcal{H}(\boldsymbol{\sigma})$ on Ω_δ^1. This means that

$$\mathcal{H}(\boldsymbol{\sigma}^{(1,\delta)}) - \mathcal{H}(\boldsymbol{\sigma}^{(1,\delta,j,k)}) \le 0.$$

for any $\boldsymbol{\sigma}^{(1,\delta,j,k)} \in \Omega_\delta^1$ which is the "nearest neighbor" in Ω_δ^1, where

$$\sigma_i^{(1,\delta,k,j)} = \begin{cases} \sigma_i^{(1,\delta)}, & i \ne j,k \\ -\sigma_i^{(1,\delta)}, & otherwise \end{cases}$$

($k = 1, \dots, [\delta N]$ and $j = [\delta N] + 1, \dots, N$).

After some transformations we are faced with the problem to find

$$P_N(q, q'; \alpha, \delta) \equiv E \left\{ \prod_{k=1}^{[\delta N]} \theta(\tilde{x}_k - a_1) \prod_{k=1+[\delta N]}^{N} \theta(\tilde{x}_k - a_2) \right\},$$

where $\tilde{x}_k$ are the effective fields defined by (1.5), $\theta(x) = \begin{cases} 1, & x \geq 0 \\ 0 & x < 0 \end{cases}$ - is the Heaviside function, and

$$a_1 \equiv \alpha + 1 - 2\delta + q, \quad a_2 \equiv \alpha - 1 + 2\delta + q'$$

The following theorem proven in [FST] gives us the motivation to study the asymptotic behavior as $N \to \infty$ of the function $P_N(q, q'; \alpha, \delta)$:

Theorem 1. *Denote by $\mathcal{A}$ the event that there exist some $\delta, \varepsilon > 0$ and some point $\boldsymbol{\sigma}^0$: $||\boldsymbol{\sigma}^0 - \boldsymbol{\xi}^{(1)}||^2 < 4\delta N$, such that*
$\min_{\boldsymbol{\sigma} \in \Omega_\delta^1} \mathcal{H}(\boldsymbol{\sigma}) - \mathcal{H}(\boldsymbol{\sigma}^0) > \varepsilon^2 N$.
Then if for some α and δ

$$\max_{q>0} \limsup_{N \to \infty} \frac{1}{N} \left(\log P_N(q, -q; \alpha, \delta) + \log C_N^{[\delta N]} \right) < 0, \tag{1.6}$$

then there exists some $C(\alpha) > 0$ such that $Prob\{\overline{\mathcal{A}}\} \leq e^{-NC(\alpha)}$.

The main technical result of the paper [FST] is the theorem which describes the asymptotic behavior of the function $P_N(q, q'; \alpha, \delta)$, as $N \to \infty$ is given by Theorem 2 below. We would like to stress that the proofs for the cases of Gaussian $\xi_i^{(\mu)}$ and nongaussian $\xi_i^{(\mu)}$ are very different. In the Gaussian case the computations are not very simple, but they are straightforward. We express the function $P_N(q, q'; \alpha, \delta)$ in terms of the joint Fourier transform $F(\boldsymbol{\zeta})$ ($\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_N)$) of the distribution of the effective fields $\tilde{x}_k$ (1.5).

$$F(\boldsymbol{\zeta}) \equiv (2\pi)^{-N/2} \langle \exp\{i \sum_{k=1}^{N} \tilde{x}_k \zeta_k\} \rangle = (2\pi)^{-N/2} \langle e^{i(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}})} \rangle,$$

where we denote $\tilde{\boldsymbol{u}} = (\tilde{u}^1, \dots, \tilde{u}^p)$, $\tilde{\boldsymbol{v}} = (\tilde{v}^1, \dots, \tilde{v}^p)$ with

$$\tilde{u}^\mu \equiv N^{-1/2} \sum_{k=1}^{N} \xi_k^{(\mu)} \zeta_k, \quad \tilde{v}^\mu \equiv N^{-1/2} \sum_{j=1}^{N} \xi_j^{(\mu)}.$$

It is easy to see that

$$\langle e^{i\tilde{u}^\mu \tilde{v}^\mu} \rangle = (2\pi)^{-1} \int du^\mu dv^\mu \langle e^{i(u^\mu \tilde{u}^\mu + v^\mu \tilde{v}^\mu)} \rangle e^{-iu^\mu v^\mu}.$$

Thus, using the inverse Fourier transform for the function $F(\boldsymbol{\zeta})$, we get

$$P_N(q, q'; \alpha, \delta) = \frac{1}{(2\pi)^{N/2}} \int \prod_{k=1}^{N} \theta(x_k - a_k) dx_k \int d\boldsymbol{\zeta} \exp\{-i \sum_{k=1}^{N} x_k \zeta_k\} F(\boldsymbol{\zeta})$$

$$= \frac{1}{(2\pi)^{(N+p)}} \int e^{-i(\boldsymbol{u}, \boldsymbol{v})} d\boldsymbol{u} d\boldsymbol{v} \prod_{k=1}^{N} \int dx_k \theta(x_k - a_k) \int d\zeta_k \langle \exp\{-i\zeta_k x_k + i(\tilde{\boldsymbol{u}}, \boldsymbol{u}) + i(\tilde{\boldsymbol{v}}, \boldsymbol{v})\} \rangle,$$

where we denote for simplicity $a_k = \begin{cases} a_1, & k \le [\delta N], \\ a_2, & k > [\delta N]. \end{cases}$

But, since $\xi_i^{(\mu)}$ are independent normal variables,

$$\langle \exp\{-i\zeta_k x_k + i(\tilde{\boldsymbol{u}}, \boldsymbol{u}) + i(\tilde{\boldsymbol{v}}, \boldsymbol{v})\}\rangle$$
$$= \int (\prod_{\mu=1}^{p} \frac{e^{-(\xi_k^{(\mu)})^2/2}}{\sqrt{2\pi}}) \exp\{i(N^{-1/2}\sum_{\mu=1}^{p} u^\mu \xi_k^{(\mu)}\zeta_k + N^{-1/2}\sum_{\mu=1}^{p} v^\mu \xi_k^{(\mu)})\}$$
$$= \prod_{\mu=1}^{p} \exp\{-\frac{(u^\mu\zeta_k + v^\mu)^2}{2N}\} \quad (1.7)$$

Therefore

$$P_N(q, q'; \alpha, \delta) = \frac{1}{(2\pi)^{(\frac{N}{2}+p)}} \int d\boldsymbol{u} d\boldsymbol{v} \exp\{-i(\boldsymbol{u}, \boldsymbol{v}) - \frac{1}{2}(\boldsymbol{v}, \boldsymbol{v})\}$$
$$\cdot \prod_{k=1}^{N} \int dx_k \frac{\theta(x_k - a_k)}{U} \exp\{\frac{(ix_k + N^{-1}(\boldsymbol{u}, \boldsymbol{v}))^2}{2U^2}\},$$

where $U \equiv (\boldsymbol{u}, \boldsymbol{u})^{1/2} N^{-1/2}$. Integrating with respect to x_k, we get

$$P_N(q, q'; \alpha, \delta) = (2\pi)^{-p} \int d\boldsymbol{u} d\boldsymbol{v} \exp\{-i(\boldsymbol{u}, \boldsymbol{v}) - \frac{1}{2}(\boldsymbol{v}, \boldsymbol{v})\} \prod_{k=1}^{N} H(\frac{a_k - iN^{-1}(\boldsymbol{u}, \boldsymbol{v})}{U}).$$

Now let us fix $\boldsymbol{u}$ and change variables in the integral with respect to $\boldsymbol{v}$

$$v_1 = \frac{1}{\sqrt{N}}(\mathbf{e}_1, \boldsymbol{v}),\ v_2 = (\mathbf{e}_2, \boldsymbol{v}), ..., v_p = (\mathbf{e}_p, \boldsymbol{v}),$$

where $\{\mathbf{e}_i\}_{i=1}^p$ is the orthonormal system of vectors in $\mathbf{R}^p$ such that $e_1^\mu = (U\sqrt{N})^{-1} u^\mu$. Then, integrating with respect $v_2, ..., v_p$, we obtain

$$P_N(q, q'; \alpha, \delta) = (2\pi)^{-(p-1)/2} \int (\prod_{\mu=1}^{p} du^\mu) \int dv_1 \exp\{-iNUv_1 - \frac{N}{2}(v_1)^2$$
$$+ [N\delta] \log H(\frac{a_1}{U} - iv_1) + (N - [N\delta]) \log H(\frac{a_2}{U} - iv_1)\}.$$

Using the spherical coordinates in the integral with respect to $\boldsymbol{u}$ and integrating with respect to angular variables, we get

$$P_N(q, q'; \alpha, \delta) = \Gamma(p) \int_0^\infty dU \int dv_1 \exp\{(p-1)\log U - iNUv_1 - \frac{N}{2}(v_1)^2$$
$$+ [N\delta] \log H(\frac{a_1}{U} - iv_1) + (N - [N\delta]) \log H(\frac{a_2}{U} - iv_1)\}.$$

Then, using the saddle point method we obtain the asymptotic expression (1.8).

The difference of non-Gaussian case from the Gaussian one is that we have, in (1.7), $\prod_{\mu=1}^{p} \cos\frac{u^\mu\zeta_k + v^\mu}{\sqrt{N}}$ instead of $\prod_{\mu=1}^{p} \exp\{-\frac{(u^\mu\zeta_k + v^\mu)^2}{2N}\}$. To replace the former term by the latter one we have to estimate the difference between them for different $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{\zeta}$. Besides, since most of integrals do not converge absolutely, hence the estimates of the absolute values

(differently from the Newman work) do not work. This produces so many technical difficulties that in the nongaussian case we are able to prove only the upper bound for $P_N(q,q';\alpha,\delta)$. Till now there are some doubts that the true asymptotic for $P_N(q,q';\alpha,\delta)$ for all values of parameters q,q',α,δ coincides with that for the Gaussian case. But the remarkable fact is that in the field of parameters which we need to study in order to apply Theorem 1 we can prove that the upper bound for $P_N(q,q';\alpha,\delta)$ coincides with the asymptotic expression for $P_N(q,q';\alpha,\delta)$ in the case of normal $\xi_i^{(\mu)}$.

Theorem 2. *For the Gaussian i.i.d.* $\xi_i^{(\mu)}$

$$\lim_{N\to\infty} N^{-1}\log P_N(q,q';\alpha,\delta) = \max_{U>0}\min_{V}\mathcal{F}_0(U,V;\alpha,\delta,q,q') - \frac{\alpha}{2}\log\alpha + \frac{\alpha}{2}. \tag{1.8}$$

where

$$\mathcal{F}_0(U,V;\alpha,\delta,q,q') \equiv \delta\log H(\frac{a_1^*}{U} - V) + (1-\delta)\log H(\frac{a_2^*}{U} - V)$$
$$-UV + \frac{1}{2}V^2 + \alpha\log U, \quad \left(H(x) \equiv \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt\right)$$

For the Bernulli i.i.d. $\xi_i^{(\mu)}$*:*

$$\limsup_{N\to\infty} N^{-1}\log P_N(q,q';\alpha,\delta) \le \max_{U>0}\min_{V}\mathcal{F}_0^{(D)}(U,V;\alpha,\delta,q,q') - \frac{\alpha}{2}\log\alpha + \frac{\alpha}{2}. \ ,$$

where $\mathcal{F}_0^{(D)}(U,V;\alpha,\delta,q,q') \ge \mathcal{F}_0(U,V;\alpha,\delta,q,q')$ *(see [FST] for the exact expression of* $\mathcal{F}_0^{(D)}(U,V;\alpha,\delta,q,q')$*).*

As it was already mentioned above, in the field of interest

$$\mathcal{F}_0^{(D)}(U,V;\alpha,\delta,q,q') = \mathcal{F}_0(U,V;\alpha,\delta,q,q')$$

.

Remark 1. *Numerical calculations show that condition (1.6) is fulfilled for any* $\alpha \le \alpha_c^* =$ 0.113...

The result of Theorem 2 also enables us to obtain a rather simple upper bound for the probability to have a fixed point of the dynamics (1.2) at the distance δ from the first pattern:

Theorem 3. $P_N^*(\delta,\alpha)$ *- the probability to have a fixed point of the dynamics of the Hopfield model at the distance* δ *from the first pattern has an upper bound of the form:*

$$P_N^*(\delta,\alpha) \le \exp\{N[-\delta\log\delta - (1-\delta)\log(1-\delta) + \delta\log H(\frac{1-2\delta}{\sqrt{\alpha}})$$
$$+(1-\delta)\log H(-\frac{1-2\delta}{\sqrt{\alpha}}) + O(e^{-1/\alpha}) + o(\delta\log\alpha^{-1}) + o(1)]\}.$$

It is shown in [FST] that this bound becomes asymptotically exact for small α ($\alpha \to 0$). Moreover, Theorem 3 implies very important corollary:

Corollary 1. *It follows from Theorem 3, that* $\delta_c(\alpha)$*- the minimal* δ *for which* $P_N^*(\delta,\alpha)$ *does not decay exponentially in* N*, as* $N\to\infty$*, has the asymptotic behaviour*

$$\delta_c(\alpha) \sim \frac{\sqrt{\alpha}}{\sqrt{2\pi}}e^{-1/2\alpha}.$$

This result coincides with the formula found by Amit et al [AGS] with replica calculations.

2 The Hopfield model of spin glasses

Now we discuss another method to study the Hopfield model - so-called statistical mechanics approach. This approach is based on the observation that if we take some positive parameter β (usually β is called the inverse temperature) and introduce the Gibbs measure on $\Sigma_N = \{\boldsymbol{\sigma} \in \mathbb{R}^N, \sigma_i = \pm 1\}$ of the form

$$\langle \dots \rangle = Z_N^{-1} \sum_{\boldsymbol{\sigma} \in \Sigma_N} (\dots) e^{-\beta \mathcal{H}(\boldsymbol{\sigma})}, \quad Z_N = \sum_{\boldsymbol{\sigma} \in \Sigma_N} e^{-\beta \mathcal{H}(\boldsymbol{\sigma})},$$

then this measure is an invariant measure of the so called Glauber dynamics for fixed β. The Glauber dynamics is some special kind of stochastic dynamics. And the neural network dynamics (1.2) is the limiting case of the Glauber dynamics for $\beta \to \infty$. So the idea is to study the Gibbs measure for fixed β and then make some conclusions about its behavior as $\beta \to \infty$.

The key role in studies of the Gibbs measure plays the free energy

$$f_N(\beta) = -\frac{1}{\beta N} \log Z_N,$$

because the most important characteristics of the Gibbs measure can be obtained as derivatives of the free energy with respect to the different parameters.

Consider the Hopfield model with additional parameters τ, ε which correspond some additional terms (fields) in the energy function:

$$\mathcal{H}(\boldsymbol{\sigma}) = -\sum_{i,j=1}^{N} J_{ij} \sigma_i \sigma_j + \tau \sum_{i}^{N} \xi_i^{(1)} \sigma_i + \varepsilon \sum_{i}^{N} h_i \sigma_i, \quad J_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^{(\mu)} \xi_j^{(\mu)}$$

with h_i-i.i.d. normal variables.

This model for the case $p = \text{const}$ was introduced initially by Pastur and Figotin [PF] as an exactly solvable model of spin glasses. They have shown that the free energy of the Hopfield model with the finite number of patterns in the limit $N \to \infty$ coincides with that for the Curie-Weiss model. This result means, in particular, that the Gibbs measure for the finite β in the limit $N \to \infty$ is concentrated on some spheres around the patterns $\boldsymbol{\xi}^{(\mu)}$ and the radius of these spheres tends to zero, as $\beta \to \infty$.

Similar result was obtained by Koch and Piasko [KP] in the case as $p \sim \log N$ when $N \to \infty$. And finally in the work [ST1] this result was generalized on the case when $p, N \to \infty$, $p/N \to 0$.

The Hopfield model with extensively many patterns ($p, N \to \infty$, $p/N \to \alpha$) was widely discussed in the physical literature. By using so called replica calculations, which are not rigorous from mathematical point of view but sometimes very efficient, a lot of results on the Hopfield model were found. But most of them only wait for their mathematical proof. Let us discuss briefly this method and results.

2.1 Replica trick

The replica trick was proposed initially by Parisi to study the free energy of the other very popular model of spin glasses - the Sherrington -Kirkpatrick model (see [MPV] and references therein). The method is based on a simple observation that

$$E \log Z_N = \lim_{l \to 0} \frac{d}{dl} E Z_N^l$$

So the idea is to find for $n \in \mathbb{N}$

$$E Z_N^n = \exp\{N(\phi(n) + o(1))\}$$

Then we construct the analytical function: $\phi(\zeta) \to \phi(n)\Big|_{\zeta=n}$ and find $\phi'(0)$. Then $\lim\limits_{N\to\infty} f_N = -\frac{1}{\beta}\phi'(0)$. These scheme for the Hopfield model was realized by Amit, Gutfreund and Sompolinsky [AGS]. They found that there exists some $\alpha_c(\beta)$ such that for $\alpha < \alpha_c(\beta)$ the order parameter of the problem (so called Edwards-Anderson order parameter)

$$q_N = N^{-1}\sum_{i=1}^{N}\langle\sigma_i\rangle^2, \tag{2.1}$$

possess the self averaging property (his variance vanishes as $N \to \infty$) and his limiting mean value is the solution of so-called replica symmetric equations:

$$\begin{aligned} m^\nu &= \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} E\{\xi^\nu \tanh\beta(\sqrt{\alpha r}z + \sum_{\nu=1}^{s}(m^\nu + h^\nu)\xi_1^\nu)\} \\ q &= \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} E\{\tanh^2\beta(\sqrt{\alpha r}z + \sum_{\nu=1}^{s}(m^\nu + h^\nu)\xi_1^\nu)\}. \\ r &= \frac{q}{(1-\beta(1-q))^2} \end{aligned} \tag{2.2}$$

where

$$\begin{aligned} &q = \lim_{N\to\infty} q_N, \quad m^\nu = \lim_{N\to\infty} m_N^\nu, \quad r = \lim_{N\to\infty} r_N, \\ &m_N^\nu = N^{-1}\textstyle\sum_{i=1}^{N}\xi_i^{(\mu)}\langle\sigma_i\rangle, \quad r_N = \sum_{\mu=s+1}^{p}(m_N^\nu)^2. \end{aligned} \tag{2.3}$$

And the mean value of the free energy has the limit

$$\begin{aligned} f = \min_{m^1\dots,m^s,r,q}\{&\frac{1}{2}\alpha + \frac{1}{2}\sum_{\nu=1}^{s}(m^\nu)^2 + \frac{\alpha\beta r(1-q)}{2} \\ &+\frac{\alpha}{2\beta}\left[\ln(1-\beta(1-q)) - \frac{\beta q}{1-\beta(1-q)}\right] + \\ &-\frac{1}{\beta}\int\frac{dz}{\sqrt{2\pi}}e^{-z^2/2}E\{\ln[2\cosh\beta(\sqrt{\alpha r}z + \sum_{\nu=1}^{s}(m^\nu + h^\nu)\xi_1^\nu)]\}\}. \end{aligned} \tag{2.4}$$

It is easy to check that equations (2.2) can be obtained as the extremum conditions of the for the l.h.s. of (2.4).

And for $\alpha > \alpha_c(\beta)$ the Edward-Anderson order parameter is a random variable even in the limit $N \to \infty$ and its distribution is a solution of some rather complicated nonlinear partial differential equation of the second order. The most important for the neural networks dynamics result here is that for $\alpha < \alpha_c(\beta)$ the Gibbs measure is concentrated around the patterns $\boldsymbol{\xi}^{(\mu)}$. And it was shown that $\alpha_c(\beta) \to 0.138...$ as $\beta \to \infty$.

Till now there are not so many rigorous results for the Hopfield model with extensively many patterns ($p, N \to \infty$, $p/N \to \alpha$). Self averaging property of the free energy, i.e. that the variance of the free energy vanishes as $N \to \infty$

$$\lim_{N\to\infty} E\left\{(f_N - Ef_N)^2\right\} = 0$$

was proven in [ST1]. This result was generalized by Bovier et al [BGP] who proved the large deviation type bounds for $(f_N - Ef_N)$.

The most interesting rigorous results on the Hopfield model of spin glass (see [PST1], [PST2], [BG], [T1]) were obtained by using some version of the cavity method, which we are going to discuss now.

2.2 Cavity method

In the spin glass theory this method is used mainly to study the replica symmetric field (for $\alpha < \alpha_c(\beta)$).

Recall the simple identity

$$\langle \sigma_i \rangle = \langle \tanh \beta (\sum_{j \neq i}^{N} J_{ij} \sigma_j + \varepsilon h_1) \rangle \tag{2.5}$$

valid for the Ising model ($\sigma_i = \pm 1$) with any interaction $J_{i,j}$.

The mean field approximation is based on the assumption that the thermodynamic correlations between spins vanish in the macroscopic limit

$$|\langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle| \to 0, \quad N \to \infty. \tag{2.6}$$

Then if $J_{i,j} \to 0$, as $N \to \infty$, we can replace (2.5) by the relation

$$\langle \sigma_i \rangle = \tanh \beta (\sum_{j \neq i}^{N} J_{ij} \langle \sigma_j \rangle + \varepsilon h_1) + o(1). \tag{2.7}$$

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 rigorous version of the cavity method for the spin glass theory was proposed first in [PS], [S] and the adopted to the Hopfield model in [PST1], [PST2]. It was shown here that vanishing of correlations is equivalent to the self-averaging property of the Edwards-Anderson order parameter

$$E\{(q_N - E\{q_N\})^2\} \to 0, \quad N \to \infty, \tag{2.8}$$

and if for some $\alpha, \beta, \varepsilon, t$ the parameter q_N is s.a., then of (2.7) is valid

$$\langle \sigma_1 \rangle = \tanh \beta (\sum_{j=2}^{N} J_{1j} \langle \sigma_j \rangle_0 + \varepsilon h_1) + r_{1,N}, \quad E r_{1,N}^2 \to 0. \tag{2.9}$$

Here $\langle \ldots \rangle_0$ is the Gibbs measure, corresponding to the $\mathcal{H}(\boldsymbol{\sigma})\Big|_{\sigma_1 = 0}$. From the last relation it is straightforward to derive the replica symmetric equations (2.2) for the order parameters.

Thus, the key point of the cavity method is the proof of some analog of (2.6). As soon as we establish (2.6) for some model, then we can derive some kind of self-consistent equations.

There are a few works, where (2.6) is obtained for the Hopfield model and then equations (2.2) are derived (see, e.g., [BG] and [T1]). But unfortunately all of them deal with $\alpha << 1$, so they cannot be used for the purposes of the neural networks dynamics.

3 The Gardner problem

Now let us come back to the neural networks dynamics (1.2) and recall that the main problem here was to introduce an interaction $\{J_{ij}\}_{i,j=1}^{N}$ (not necessary symmetric) in such a way that some chosen vectors $\{\boldsymbol{\xi}^{(\mu)}\}_{\mu=1}^{p}$ (patterns) are the fixed points of the dynamics (1.2). The choice of matrix $\{J_{ij}\}_{i,j=1}^{N}$ depends on the concrete model, but one can see easily that multiplication of all coefficients in the same line $\{J_{ij}\}_{j=1}^{N}$ by some positive constant λ_i does not change the dynamics (1.2). So it is natural to consider the matrices whose lines satisfies some kind of

normalization conditions. For most popular models of neural networks dynamics (e.g. for the Hopfield model) these conditions have the form

$$\sum_{j=1,j\neq i}^{N} J_{ij}^2 = NR \quad (i=1,\dots,N), \tag{3.1}$$

where R is some fixed number which could be taken equal to 1.

It is obvious also that if $\boldsymbol{\xi}^{(\boldsymbol{\mu})}$ are the fixed points of (1.2), then the interactions matrix $\{J_{ij}\}$ satisfies also conditions

$$\xi_i^{(\mu)} \sum_{j=1,j\neq i}^{N} J_{ij}\xi_j^{(\mu)} > 0 \quad (i=1,\dots,N), \quad (\mu=1,\dots,p). \tag{3.2}$$

Sometimes condition (3.2) is not sufficient to have $\boldsymbol{\xi}^{(\boldsymbol{\mu})}$ as the end points of the dynamics. To have some "basin of attraction" (that is some neighborhood of $\boldsymbol{\xi}^{(\boldsymbol{\mu})}$, starting from which we for sure arrive in $\boldsymbol{\xi}^{(\boldsymbol{\mu})}$) one should introduce some positive parameter k and impose the conditions:

$$\xi_i^{(\mu)} \sum_{j=1,j\neq i}^{N} J_{ij}\xi_j^{(\mu)} > k \quad (i=1,\dots,N), \quad (\mu=1,\dots,p). \tag{3.3}$$

Gardner (see [G]) was the first who solved a kind of inverse problem.

Problem 2. For which $\alpha = \frac{p}{N}$ the interaction $\{J_{ij}\}$, satisfying (3.1) and (3.3) exists? What is the ratio of the total Lebesque measure of the interactions satisfying (3.3) and (3.1) to the measure of all interactions, satisfying (3.1) (she called this quantity the typical fractional volume of the interactions)?

Since all conditions (3.1) and (3.3) are factorized with respect to i, this problem after a simple transformation should be replaced by the following. For the system of $p \sim \alpha N$ i.i.d. random patterns $\{\boldsymbol{\xi}^{(\boldsymbol{\mu})}\}_{\mu=1}^{p}$ with i.i.d. $\xi_i^{(\mu)}$ $(i=1,\dots,N)$ assuming values ± 1 with probability $\frac{1}{2}$, consider

$$\Theta_{N,p}(k) = |S_N|^{-1} \int_{(\boldsymbol{J},\boldsymbol{J})=N} d\boldsymbol{J} \prod_{\mu=1}^{p} \theta(N^{-1/2}(\boldsymbol{\xi}^{(\boldsymbol{\mu})},\boldsymbol{J}) - k), \tag{3.4}$$

($\theta(x)$ is the Heaviside-function), $|S_N|$ is the Lebesgue measure of N-dimensional sphere of radius $N^{1/2}$. Then, the question of interest is the behaviour of $\frac{1}{N}\log\Theta_{N,p}(k)$ in the limit $N,p\to\infty$, $\frac{p}{N}\to\alpha$.

This problem has a very nice geometrical interpretation. For very large integer N consider the N-dimensional sphere S_N of radius $N^{1/2}$ centered in the origin and $p=\alpha N$ independent random half spaces Π_μ $(\mu=1,\dots p)$. Let $\Pi_\mu = \{\boldsymbol{J}\in\mathbf{R}^N : N^{-1/2}(\boldsymbol{\xi}^{(\boldsymbol{\mu})},\boldsymbol{J})\geq k\}$, where $\boldsymbol{\xi}^{(\boldsymbol{\mu})}$ are i.i.d. random vectors with i.i.d. Bernulli components $\xi_j^{(\mu)}$ and k is the distance from Π_μ to the origin. The problem is to find the maximum value of α such that the volume of the intersection of S_N with $\cap\Pi_\mu$ is not "too small" comparing with $|S_N|$, i.e. their ratio is of the order e^{-NC} with some bounded C. Let us remark here, that since $|S_N| \sim \pi^{1/2}\left(\frac{\pi e}{2}\right)^{N/2}$ as $N\to\infty$, it is natural to expect that the "normal behavior" of our ratio is just e^{-NC}, and so the words "too small" mean that the ratio tends to zero more fast than e^{-NC} with any positive C.

Gardner [G] had solved this problem by using replica trick, described in the previous section. As it was mentioned above this method is far from being rigorous from mathematical point of

view, but it plays very important role in the physical literature and gives results which usually are correct. Using this method Gardner has shown that for any $\alpha < \alpha_c(k)$, where

$$\alpha_c(k) \equiv (\frac{1}{\sqrt{2\pi}} \int_{-k}^{\infty} (u+k)^2 e^{-u^2/2} du)^{-1}, \tag{3.5}$$

we have so-called replica-symmetric solution of the problem. This means first of all that, if we define the Edward-Anderson order parameter as

$$q_N = N^{-1} \sum \langle J_i \rangle_\Theta^2, \tag{3.6}$$

with $\langle \ldots \rangle_\Theta$ being the uniform distribution on the intersection of S_N with $\cap \Pi_\mu$, then q_N possess the self-averaging property (2.8) and its limiting mean value can be found as a solution of the replica symmetric equation

$$q = \alpha(1-q) E \left\{ \left[H\left(\frac{u\sqrt{q}+k}{\sqrt{1-q}} \right) \right]^{-2} \right\} \tag{3.7}$$

where $\mathrm{H}(x) \equiv \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$ and u is a Gaussian normal random variable. Besides, there exists

$$\lim_{N,p\to\infty, p/N\to\alpha} N^{-1} E\{\log \Theta_{N,p}(k)\} = \mathcal{F}(\alpha, k) =$$
$$\equiv \min_{q:0\le q\le 1} \left[\alpha E \left\{ \log H\left(\frac{u\sqrt{q}+k}{\sqrt{1-q}} \right) \right\} + \frac{1}{2}\frac{q}{1-q} + \frac{1}{2}\log(1-q) \right]. \tag{3.8}$$

It is easy to check that equation (3.7) is just the minimum condition for the function in the right hand side of (3.8). For $\alpha \ge \alpha_c(k)$

$$\frac{1}{N} \log \Theta_{N,p}(k) \to -\infty, \quad as \quad N \to \infty.$$

It is interesting to observe, that $\alpha_c(0) = \frac{1}{2}$ (cf with the Hopfield model, where $\alpha_c = 0.138...$).

Let us remark that, according to the results of Gardner, for this model there is no field of parameters here with so called broken replica symmetric solution, so there is a hope, that differently from the Hopfield model the Gardner model could be studied completely (i.e. in the whole field of parameters) by using the cavity method described in the previous section.

3.1 Rigorous results for the Gardner problem

The first rigorous result for the Gardner problem with Gaussian $\boldsymbol{\xi}^{(\mu)}$ was obtained by Talagrand [T2]. He proved large deviation type bounds for the fluctuations of $\log \Theta_{N,p}(k)$.

Complete rigorous solution for the Gardner problem was obtained in [ST2] (see also [ST3]), where the Gardner formulas (3.8) (3.7) for the free energy and the Edwards-Anderson order parameter were proved. To this end we use a version of the cavity method, but the problem is that we are not able to produce the equations for the order parameter in the case, when the "randomness" is not included in the Hamiltonian, but is contained in the form of the integration domain. That is why we used a rather common trick: substitute θ-functions by some smooth functions which depend on the small parameter ε and tend, as $\varepsilon \to 0$, to θ-function. We choose for these purposes $\mathrm{H}(x\varepsilon^{-1/2})$, where H is the erf-function, but the particular form of these smoothing functions is not very important for us. The most important fact is, that they are not zero in any point and so, taking their logarithms, we can treat them as a part of our Hamiltonian.

So we introduce the intermediate Hamiltonian

$$\mathcal{H}_{N,p}(\boldsymbol{J},k,h,z,\varepsilon) \equiv -\sum_{\mu=1}^{p} \log \mathrm{H}\left(\frac{k-(\boldsymbol{\xi}^{(\mu)},\boldsymbol{J})N^{-1/2}}{\sqrt{\varepsilon}}\right) + \frac{z}{2}(\boldsymbol{J},\boldsymbol{J}), \tag{3.9}$$

The partition function for this Hamiltonian is

$$\begin{aligned} Z_{N,p}(k,z,\varepsilon) &= |S_N|^{-1} \int d\boldsymbol{J} \exp\{-\mathcal{H}_\varepsilon(\boldsymbol{J},k,z,\varepsilon)\} \\ &= |S_N|^{-1} \int d\boldsymbol{J} \prod_{\mu=1}^{p} \mathrm{H}\left(\frac{k-(\boldsymbol{\xi}^{(\mu)},\boldsymbol{J})N^{-1/2}}{\sqrt{\varepsilon}}\right) \exp\{-z(\boldsymbol{J},\boldsymbol{J})/2\}. \end{aligned} \tag{3.10}$$

We denote also by $\langle \ldots \rangle$ the corresponding Gibbs averaging and

$$f_{N,p}(k,z,\varepsilon) \equiv \frac{1}{N} \log Z_{N,p}(k,z,\varepsilon).$$

One more difference of this model from the model (3.4) is that we introduce an additional parameter $z > 0$ to replace the integration over the sphere $(\boldsymbol{J},\boldsymbol{J}) = N$ in (3.4) by the integration in the whole space $\mathbf{R}^N$ in (3.10). It is proven in [ST2] that if we find the thermodynamic limit

$$\lim_{N,p\to\infty, p/N\to\alpha} E\{f_{N,p}(k,z,\varepsilon)\} = F(\alpha,k,z,\varepsilon)$$

and choose z^* from the condition

$$F(\alpha,k,z^*,\varepsilon) + \frac{z^*}{2} = \min_{z>0}\{F(\alpha,k,z,\varepsilon) + \frac{z}{2}\},$$

then

$$\lim_{N,p\to\infty, p/N\to\alpha} N^{-1} E\left\{ \log \sigma_N^{-1} \int_{(\boldsymbol{J},\boldsymbol{J})=N} d\boldsymbol{J} \exp\{-\mathcal{H}(\boldsymbol{J},k,0,\varepsilon)\right\} = F(\alpha,k,z^*,\varepsilon) + \frac{z^*}{2}.$$

We call the model (3.9)-(3.10) by the modified Gardner model. The free energy of this model can be found using the following theorem proven in [ST2]:

Theorem 4. *For $\alpha < 2$, ε small enough, and $z \le \varepsilon^{-1/3}$, there exists*

$$\begin{aligned} &\lim_{N,p\to\infty,\alpha_N\to\alpha} E\{f_{N,p}(k,z,\varepsilon)\} = F(\alpha,k,z,\varepsilon), \\ &F(\alpha,k,h,z,\varepsilon) \equiv \max_{R>0} \min_{0\le q\le R} \left[\alpha E\left\{\log \mathrm{H}\left(\frac{u\sqrt{q}+k}{\sqrt{\varepsilon+R-q}}\right)\right\}\right. \\ &\qquad \left. + \frac{1}{2}\frac{q}{R-q} + \frac{1}{2}\log(R-q) - \frac{z}{2}R\right], \end{aligned}$$

where u is a normal random variable.

As it was mentioned above, the proof of Theorem 4 is based on the the application of the cavity method to the Gardner problem. The key point of this application is the proof of the vanishing of the thermodynamic correlations between J_i and J_j in the limit $N \to \infty$ (cf (2.6)):

$$E\left\{\langle (J_i - \langle J_i\rangle)(J_j - \langle J_j\rangle)\rangle^2\right\} \to 0, \quad as \quad N\to\infty. \tag{3.11}$$

which follows from the Brascamp-Lieb [BL] inequalities, according to which for any integer n and any $\mathbf{x} \in \mathbf{R}^N$

$$\left\langle \left(\frac{(\dot{\boldsymbol{J}}, \mathbf{x})}{\sqrt{N}} \right)^{2n} \right\rangle \leq \frac{\Gamma(2n-1)}{z^n \Gamma(n-1)} \left(\frac{|\mathbf{x}|^2}{Nn} \right)^n . \tag{3.12}$$

It is interesting to remark that the Brascamp-Lieb inequalities follow from the classical geometrical theorem:

Theorem of Brunn-Minkowski

Let $M \subset \mathbb{R}^N$ be some convex set. Consider the family of hyper planes $\mathcal{L}(t) = \{x \in \mathbb{R}^N (x, e) = t\sqrt{N}\}$. Let $\mathcal{A}(t) = M \cap \mathcal{L}(t)$. Consider $R(t) \equiv [\mathrm{mes}\mathcal{A}(t)]^{1/N}$. Then

$$\frac{d^2 R(t)}{dt^2} \leq 0$$

and $\frac{d^2 R(t)}{dt^2} \equiv 0$ for $t \in [t_1', t_2']$ if and only if all the sets $\mathcal{A}(t)$ for $t \in [t_1', t_2']$ are homothetic to each other.

After the proof of Theorem 4 the next step is the limiting transition $\varepsilon \to 0$, i.e. the proof that the product of αN θ-functions in (3.4) can be replaced by the product of $\mathrm{H}(\frac{x}{\sqrt{\varepsilon}})$ with the small difference, when ε is small enough. Despite expectations, it is the most difficult step from the technical point of view. It is rather simple to prove, that the expression (3.8) is an upper bound or $\log \Theta_N, p(k)$. But the estimate from below is much more complicated. The problem is that to estimate the difference between the free energies corresponding to two Hamiltonians we, as a rule, need to have them defined in the common configuration space, or, at least, we need to know some a priori bounds for some Gibbs averages. In the case of the Gardner problem we do not posses this information. This leads to rather serious technical problems.

The final result has the form:

Theorem 5. *For any $\alpha < \alpha_c(k)$ there exists*

$$\lim_{N,p \to \infty, p/N \to \alpha} E\{N^{-1} \log \Theta_{N,p}(k)\} = \lim_{\varepsilon \to 0} \max_{z>0} F(\alpha, k, z, \varepsilon) = \mathcal{F}(\alpha, k),$$

where $\mathcal{F}(\alpha, k)$ is the Gardner expression.

For $\alpha > \alpha_c(k)$ $\quad E\{N^{-1} \log \Theta_{N,p}(k)\} \to -\infty$, as $N \to \infty$.

It is interesting to mention one more problem which is very similar to the Gardner problem. It is so-called the Gardner-Derrida problem [DG] in which we seek the matrix $\{J_{ij}\}_{i,j=1}^N$, satisfying conditions (3.2) or (3.3) but assuming values $J_{ij} = \pm 1$. The geometrical interpretation here is that we are interested in the measure of the intersection of our random half spaces Π_μ with a discrete cube $\Sigma_N = [-1, 1]^N$. This problem was also solved by the replica trick (see [DG]) and similarly to the Gardner problem it was shown that the replica symmetric solution for this problem is true in the whole field of parameters (α, k). But till now the rigorous proof of these results with some version of the cavity method was found only for $\alpha << 1$ (see [T3], [T4]). This difference with a case of the Gardner problem is explained by the fact that in the former we can use the Brascamp-Lieb inequalities (3.12) to prove the vanishing of the thermodynamic correlations (3.11), while in the case of the Gardner-Derrida model these inequalities are not applicable.

3.2 CLT for the free energy and order parameters

An important ingredient of the analysis of the free energy of the model (3.9) in [ST2] was the proof of the fact that the variance of its order parameters (or the overlap parameters) disappears in the thermodynamic limit. In the paper [ST4] we study the behaviour of fluctuations of the overlap parameters, defined as

$$R_{l,m} = \frac{1}{N}(\boldsymbol{J}^{(l)}, \boldsymbol{J}^{(m)}), \quad (l,m = 1, \dots n), \tag{3.13}$$

where the upper indexes of the variables $\boldsymbol{J}$ mean that we consider n replicas of the Hamiltonian (3.9) with the same random parameters $\{\boldsymbol{\xi}^{(\mu)}\}_{\mu=1}^{p}$, but different $\boldsymbol{J}^{(1)}, \dots, \boldsymbol{J}^{(n)}$.

We introduce also the notations:

$$\begin{aligned} &\dot{q} = N^{1/2}(\langle R_{1,2}\rangle - q), \\ &T_{l,m} = \frac{1}{N^{1/2}}(\dot{\boldsymbol{J}}^{(l)}, \dot{\boldsymbol{J}}^{(m)}), \quad T_l = \frac{1}{N^{1/2}}(\dot{\boldsymbol{J}}^{(l)}, \langle \boldsymbol{J}\rangle). \end{aligned} \tag{3.14}$$

Here $\dot{\mathbf{J}} \equiv \mathbf{J} - \langle \mathbf{J}\rangle$ and $\langle \boldsymbol{J}\rangle = (\langle J_1\rangle, \dots \langle J_N\rangle) \in \mathbf{R}^N$, where $\langle \dots \rangle$ is the Gibbs averaging with respect to the Hamiltonian (3.9). (q, R) is the solution of the system of equations:

$$\begin{aligned} q &= (R-q)^2 \left[\frac{\alpha}{R-q+\varepsilon} E\left\{ \mathrm{A}^2\left(\frac{\sqrt{q}u + k}{\sqrt{R-q+\varepsilon}} \right) \right\} \right], \\ z &= \frac{\alpha}{(R-q+\varepsilon)^{3/2}} E\left\{ (\sqrt{q}u + k)\mathrm{A}\left(\frac{\sqrt{q}u + k}{\sqrt{R-q+\varepsilon}} \right) \right\} \\ &\quad - \frac{q}{(R-q)^2} + \frac{1}{R-q}, \end{aligned} \tag{3.15}$$

with

$$\mathrm{A}(x) = -\frac{1}{\sqrt{2\pi}} \frac{d}{dx} \log \mathrm{H}(x).$$

These equations are equivalent to $\frac{\partial \mathcal{F}}{\partial q} = 0$ $\frac{\partial \mathcal{F}}{\partial R} = 0$, for the function $\mathcal{F}(q, R; k, z, \varepsilon)$ which is defined by the expression in the r.h.s. of (1.8) before taking $\max_R \min_q$. It is proven in [ST2] that if $\alpha < 2$, $\varepsilon \le \varepsilon^*(\alpha, k)$ and $z \le \varepsilon^{1/3}$, the the system (3.15) has a unique solution.

To avoid additional technical difficulties in the proof of central limit theorems we assume that $\{\xi_i^{(\mu)}\}$ are independent normal random variables.

The main result of the paper [ST2] is

Theorem 6. *Consider any $\alpha < 2$, $k > 0$, $\varepsilon \le \varepsilon^*(\alpha, k)$ and $z \le \varepsilon^{-1/3}$. Then for any integer n the families of random variables $\{\sqrt{N}(R_{l,m} - E\langle R_{l,m}\rangle)\}_{l<m\le n}$, converges in distribution, as $N, p \to \infty, p/N \to \alpha$, to the Gaussian family of random variables $\{v_{l,m}\}_{l<m\le n}$, with the covariance matrix:*

$$\begin{aligned} E\{v_{l,m}v_{l,m}\} &= A^*, \\ E\{v_{l,m}v_{l,m'}\} &= B^* \ (m \ne m'), \\ E\{v_{l,m}v_{l',m'}\} &= C^* \ (m, m', l, l' \ are \ different). \end{aligned} \tag{3.16}$$

In particular,

$$\begin{aligned} \lim_{N,p\to\infty, p/N\to\alpha} E\left\{\langle T_{1,2}^{2n}\rangle\right\} &= \frac{\Gamma(2n-1)}{\Gamma(n-1)} A_*^n \\ \lim_{N,p\to\infty, p/N\to\alpha} E\left\{\langle T_1^{2n}\rangle\right\} &= \frac{\Gamma(2n-1)}{\Gamma(n-1)} B_*^n \\ \lim_{N,p\to\infty, p/N\to\alpha} E\{\dot{q}^{2n}\} &= \frac{\Gamma(2n-1)}{\Gamma(n-1)} C_*^n, \end{aligned} \tag{3.17}$$

where the constants A^*, B^*, C^*, A_*, B_*, C_* depend on $\alpha, k, z, \varepsilon$ and all odd moments for these random variables tend to zero.

Remark 2. *In fact it follows from our proof that* $\{T_{l,m}\}_{l<m\leq n}$ *and* $\{T_l\}_{l\leq n}$ *in some sense do not depend on the random variables* $\{\xi_i^{(\mu)}\}$*, i.e. if we consider* P*- some product of* $\{T_{l,m}\}_{l<m\leq n}$ *and* $\{T_l\}_{l\leq n}$*, then*

$$\lim_{N,p\to\infty,p/N\to\alpha} E\{(\langle P\rangle - E\langle P\rangle)^2\} = 0. \tag{3.18}$$

Similar result for the free energy (3.9) of the modified Gardner model was obtained in [**?**].

Theorem 7. *Consider the modified Gardner model with i.i.d. normal variables* $\{\xi_i^{(\mu)}\}_{i=1,\dots,N,\mu=1,\dots,p}$*. Then for any* $\alpha < 2$*,* $k > 0$*,* $\varepsilon \leq \varepsilon^*(\alpha, k)$ *and* $z \leq \varepsilon^{-1/3}$ *the random variable*

$$v_{N,p} = N^{1/2}(f_{N,p} - E\{f_{N,p}\})$$

converges in distribution, as $N, p \to \infty, p/N \to \alpha$*, to a Gaussian random variable with zero mean and the variance*

$$V^2 = \alpha E\left\{\left(\log \mathrm{H}\left(\frac{u\sqrt{q}+k}{\sqrt{\varepsilon+R-q}}\right)\right)^2\right\} - \alpha E^2\left\{\log \mathrm{H}\left(\frac{u\sqrt{q}+k}{\sqrt{\varepsilon+R-q}}\right)\right\}$$

Similar results for the fluctuations of the overlap parameters and of the free energy were obtained in [T4] for the Gardner-Derrida model for small α and for the Sherrington-Kirkpatrick model for the high temperature. We would like to mention also the work [GuT], where the fluctuations of the overlap parameters for the Sherrington-Kirkpatrick model in the high temperature region were studied by the method of characteristic functions.

References

[AGS] Amit D., Gutfreund H. and Sompolinsky H. Statistical Mechanics of Neural Networks. Annals of Physics **173**, 30-47 (1987)

[BGP] A. Bovier, V. Gayrard and P. Picco: Large deviation principles for the Hopfield and the Kac-Hopfield model, Probab. Theory Rel. Fields **101**, 511-546 (1995)

[BG] A. Bovier, V. Gayrard. Hopfield models as generalized random mean field models. In "Mathematical aspects of spin glasses and neural networks", A. Bovier and P. Picco (eds.), Progress in Probability 41, 1-89 (Birkhäuser, Boston 1998)

[BL] Brascamp H.J., Lieb E.H. On the Extension of the Brunn-Minkowsky and Pekoda-Leindler Theorems, Includings Inequalities for Log Concave functions, and with an Application to the Diffusion Equation. J.Func.Anall. **22**,366-389 (1976)

[DG] B.Derrida, E.Gardner Optimal Stage Properties of Neural Network Models. J.Phys.A: Math.Gen. **21**, 271-284 (1988)

[FST] J.Feng, M.Shcherbina, B.Tirozzi . On the critical capacity of the Hopfield model. Commun.Math.Phys. V.216, p.139-177, (2001).

[G] E.Gardner: The Space of Interactions in Neural Network Models. J.Phys.A: Math.Gen. **21**, 257-270 (1988)

[GuT] F.Guerra and F.L.Toninelli. Central Limit Theorem for Fluctuations in the High Temperature Region of the Sherrington- Kirkpatrick Spin Glass Model. J.Math.Phys. **43**, 6224-6237 (2002)

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

[KP] Koch H., Piasko J., Some Rigorous Results on the Hopfield Neural Network Model. J. of Stat. Phys., **55**, 5/6, 903-928, (1993)

[L] D.Loukianova. Lower bounds on the restitution error of the Hopfield model. Prob. Theor. Relat. Fields, **107** 161-176 (1997)

[McEPRV] R.J. MacEliece, E.C.Posner, E.R.Rodemich, S.S. Venkatesir. The capacity of the Hopfield associative memory. IEEE Trans.Inform. Theory **33**, 461-468 (1987)

[MPV] Mezard M., Parisi G., Virasoro M.A. Spin Glass Theory and Beyond. Singapur: World Scientific, 1987.

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

[PF] L.Pastur, A.Figotin. Exactly Soluble Model of the Spin Glass. Soviet.Phys.J.E.T.P., **25**, 348-353, (1977)

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

[PST1] 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)

[PST2] L.Pastur, M.Shcherbina, B.Tirozzi. On the replica symmetric equations for the Hopfield model. J.Math.Phys. V.40 (1999)

[S] M.Shcherbina. More about the absence of selfaverageness of order parameter in SK-model. Preprint Rome University-I, 1991.

[ST1] M. Shcherbina, B. Tirozzi. The Free Energy of a Class of Hopfield Models. J. of Stat. Phys., **72** 1/2, 113-125, (1993)

[ST2] M. Shcherbina, B. Tirozzi. Rigorous Solution of the Gardner Problem. Commun.Math.Phys., (2003)

[ST3] M.Shcherbina, B.Tirozzi. On the Volume of the Intersection of a Sphere with Random Half Spaces. CRAS Ser.I **334** p.803-806, (2002)

[ST4] M. Shcherbina, B. Tirozzi. Central Limit Theorems for Order Parameters of the Gardner Problem. Markov processes and related fields, N4, p.583-602 (2003)

[ST5] M. Shcherbina, B. Tirozzi. Central limit theorems for the free energy of the modified Gardner problem. To appear in Markov processes and related fields (2004)

[T1] M. Talagrand. Rigorous Results for the Hopfield Models with Many Patterns. *Prob. Theor. Rel. Fields*, **110**, 109-176, (1998)

[T2] Talagrand M.: Self Averaging and the Space of Interactions in Neural Networks. Random Structures and Algorithms **14**, 199-213 (1998)

[T3] Talagrand M. Intersecting Random Half-Spaces: Toward the Gardner-Derrida Problem. Ann.Probab.,**28**, 725-758 (2000)

[T4] Talagrand M. Spin glasses: a challenge for mathematicians. Mean field models and cavity method. Springer-Verlag, 2002.