J Stat Phys (2010) 140: 784–796
DOI 10.1007/s10955-010-0020-y

The Replica Symmetric Approximation of the Analogical Neural Network

Adriano Barra · Giuseppe Genovese · Francesco Guerra

Received: 1 April 2010 / Accepted: 28 June 2010 / Published online: 13 July 2010

Abstract In this paper we continue our investigation of the analogical neural network, by introducing and studying its replica symmetric approximation in the absence of external fields. Bridging the neural network to a bipartite spin-glass, we introduce and apply a new interpolation scheme to its free energy, that naturally extends the interpolation via cavity fields or stochastic perturbations from the usual spin glass case to these models.

While our methods allow the formulation of a fully broken replica symmetry scheme, in this paper we limit ourselves to the replica symmetric case, in order to give the basic essence of our interpolation method. The order parameters in this case are given by the assumed averages of the overlaps for the original spin variables, and for the new Gaussian variables. As a result, we obtain the free energy of the system as a sum rule, which, at least at the replica symmetric level, can be solved exactly, through a self-consistent mini-max variational principle.

The so gained replica symmetric approximation turns out to be exactly correct in the ergodic region, where it coincides with the annealed expression for the free energy, and in the low density limit of stored patterns. Moreover, in the spin glass limit it gives the correct expression for the replica symmetric approximation in this case. We calculate also the entropy density in the low temperature region, where we find that it becomes negative, as expected for this kind of approximation. Interestingly, in contrast with the case where the stored patterns are digital, no phase transition is found in the low temperature limit, as a function of the density of stored patterns.

A. Barra (✉)
Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 2, 00185, Rome, Italy
e-mail: adriano.barra@roma1.infn.it

G. Genovese
Dipartimento di Matematica, Sapienza Università di Roma, Piazzale Aldo Moro 2, 00185, Rome, Italy
e-mail: giuseppe.genovese@mat.uniroma1.it

F. Guerra
Dipartimento di Fisica, Sapienza Università di Roma, and INFN, Sezione di Roma 1, Piazzale Aldo Moro 2, 00185, Rome, Italy
e-mail: francesco.guerra@roma1.infn.it

Keywords Spin-glasses · Neural networks · Replica symmetry

1 Introduction

The last years have witnessed a great progress in the study of disordered models, whose description is reached in the frame of statistical mechanics for complex system. As a consequence, the need of powerful tools for their analysis raises, which ultimately push further the global field of research suggesting new possible models where their applicability can be achieved.

Among these, interestingly, neural networks have never been analyzed from an interpolating, stochastic perturbation, perspective [13]. As a matter of fact, from the early work by Hopfield [17] and the, nowadays historical, theory of Amit, Gutfreund and Sompolinsky (AGS) [1–3], to the modern theory for learning [8], about the neural networks (thought of as spin glasses with a Hebb-like "synaptic matrix" [16]) several contributions appeared (e.g. [7, 18–21]), often following understanding of spin-glasses (e.g. [10–12, 22]), and the analysis at low level of stored memories has been achieved, in particular.

However in the high level of stored memories, fundamental enquiries are still in a quite initial stage. Furthermore, general problems, as the existence of a well defined thermodynamic limit, achieved for the spin glass case in [14, 15], are unsolved.

Previously, we began to study an "analogical version" of the standard Hopfield model, by taking the freedom of allowing the learned patterns to have continuous values, their probability distribution being a standard Gaussian $\mathcal{N}[0,1]$ [5].

Within this scenario, we proved the existence of an ergodic phase where the explicit expressions for all the thermodynamic quantities (free energy, entropy, internal energy) have been found to self-average around their annealed expression in the thermodynamic limit, in complete agreement with AGS theory. Moreover, the explicit expression of the rescaled fluctuations of these variables, and of the overlaps, have been given.

In this paper, again by using an analogy between neural networks and bipartite spin glasses, we move on by introducing a novel interpolating technique (essentially based on two different stochastic perturbations), which we use to give a complete description of the analogical Hopfield model phase diagram, in the replica symmetric approximation, and with any level of stored memories (i.e. patterns).

An important feature of our analysis is the absence of any spontaneous "magnetization", even in the broken ergodicity phase, or in the low temperature limit.

The paper is organized as follows: In Sect. 2 we introduce the analogical neural network with all its statistical mechanics package of definitions and properties. In Sect. 3 we define its replica symmetric approximation by means of our interpolating scheme, in the frame of a self-consistent mini-max principle for the free energy, ruled by numerical order parameters. In Sect. 4 we study the properties of the replica symmetric approximation, in reference to the ergodic phase, to the spin glass limit, to the low density limit for stored patterns, and to the low temperature behavior. Finally, Sect. 5 is dedicated to some conclusion and outlook for future developments, mainly related to the establishment of a fully broken replica symmetry regime.

2 Analogical Neural Network

We introduce a large network of N two-state neurons $\sigma_i = \pm 1$, $i \in (1, \dots, N)$, which are thought of as quiescent (sleeping), when their value is -1, or spiking (emitting a current signal to other neurons), when their value is $+1$. They interact throughout a symmetric synaptic matrix J_{ij} defined accordingly the Hebb rule for learning [16],

$$J_{ij} = \sum_{\mu=1}^{K} \xi_i^{\mu} \xi_j^{\mu}. \tag{1}$$

Each random variable, $\xi^{\mu} = \{\xi_1^{\mu}, \dots, \xi_N^{\mu}\}$, represents a learned pattern, and tries to bring the overall current in the network (or in some part) stable with respect to itself (when this happens, we say we have a retrieval state, see e.g. [1]). The analysis of the network assumes that the system has already stored K patterns (the learning procedure is not investigated here) and we are interested in the case in which this number increases proportionally (linearly) to the system size (high storage level), i.e. $N \to \infty$, $K \to \infty$ with $K/N \to \alpha$, where α is a free parameter of the theory, the density of stored patterns.

In standard literature, these patters are usually taken as i.i.d. random variables, taking values ± 1, each with equal probabilities $\frac{1}{2}, \frac{1}{2}$. Throughout this paper we will make a different choice, by considering pattern with a unit Gaussian distribution:

$$P(\xi_i^{\mu}) = \frac{1}{\sqrt{2\pi}} e^{-(\xi_i^{\mu})^2/2}. \tag{2}$$

The average over the quenched memories will be denoted by $\mathbb{E}$ and for a generic function of these memories $F(\xi)$ can be written as

$$\mathbb{E}[F(\xi)] = \int \prod_{\mu=1}^{K} \prod_{i=1}^{N} \frac{d\xi_i^{\mu} e^{-\frac{(\xi_i^{\mu})^2}{2}}}{\sqrt{2\pi}} F(\xi) = \int F(\xi) d\mu(\xi), \tag{3}$$

of course $\mathbb{E}[\xi_i^{\mu}] = 0$ and $\mathbb{E}[(\xi_i^{\mu})^2] = 1$.

The Hamiltonian of the model is defined as follows

$$H_N(\sigma, \xi) = -\frac{1}{N} \sum_{\mu=1}^{K} \sum_{i<j}^{N} \xi_i^{\mu} \xi_j^{\mu} \sigma_i \sigma_j. \tag{4}$$

For the infinite volume limit, we assume that K is an N dependent increasing sequence, with $\lim_{N\to\infty} K/N = \alpha \in \mathbb{R}^+$. In the following, we write often α in place of K/N, by neglecting terms irrelevant in the infinite volume limit. As it is usually done in statistical mechanics, we define the partition function, the "pressure" and the quenched free energy per site as

$$Z_{N,K}(\beta, \xi) = \sum_{\{\sigma\}} e^{-\beta H_N(\sigma,\xi)} = \sum_{\{\sigma\}} e^{\frac{\beta}{N} \sum_{\mu=1}^{K} \sum_{i<j}^{N} \xi_i^{\mu} \xi_j^{\mu} \sigma_i \sigma_j}, \tag{5}$$

$$A_{N,K}(\beta) = \frac{1}{N} \mathbb{E} \log Z_{N,K}(\beta, \xi), \tag{6}$$

$$f_{N,K}(\beta) = -\frac{1}{\beta} A_{N,K}(\beta), \tag{7}$$

where β is the inverse temperature. By splitting the summation $\sum_{i<j}^{N} = \frac{1}{2} \sum_{ij}^{N} - \frac{1}{2} \sum_{ij}^{N} \delta_{ij}$, we can write down for the partition function

$$\begin{aligned} Z_N(\beta; \xi) &= \sum_{\{\sigma\}} \exp\left(\frac{\beta}{2N} \sum_{\mu=1}^{K} \sum_{ij}^{N} \xi_i^{\mu} \xi_j^{\mu} \sigma_i \sigma_j - \frac{\beta}{2N} \sum_{\mu=1}^{K} \sum_{i}^{N} (\xi_i^{\mu})^2 \right) \\ &= \tilde{Z}(\beta; \xi) \times \left(e^{\frac{-\beta}{2N} \sum_{\mu=1}^{K} \sum_{i=1}^{N} (\xi_i^{\mu})^2} \right), \end{aligned} \tag{8}$$

where

$$\tilde{Z}(\beta;\xi)=\sum_{\{\sigma\}}\exp\left(\frac{\beta}{2N}\sum_{\mu=1}^{K}\sum_{ij}^{N}\xi_i^\mu\xi_j^\mu\sigma_i\sigma_j\right)=\sum_{\{\sigma\}}\exp\left(\frac{1}{2}\beta N\sum_{\mu=1}^{K}m_\mu^2(\sigma,\xi)\right), \tag{9}$$

and $m_\mu(\sigma,\xi)$ are the partial magnetizations, defined by

$$m_\mu(\sigma,\xi)=\frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu\sigma_i. \tag{10}$$

Notice that $\tilde{Z}$ is the only term of the partition function depending on the particular state of the network. The contribution of the other term is easily calculated in evaluating the pressure, in such a way that we can consider the problem with partition function $\tilde{Z}_{N,K}(\beta,\xi)$, and add to the corresponding pressure a term given by $-\alpha\beta/2$ [5], in the infinite volume limit. Let us now perform a Gaussian transform in order to linearize with respect to the magnetizations m_μ. The expression for the partition function (9) becomes

$$\tilde{Z}_N(\beta;\xi)=\sum_{\sigma}\int\prod_{\mu=1}^{K}d\mu(z_\mu)\exp\left(\sqrt{\frac{\beta}{N}}\sum_{\mu=1}^{K}\sum_{i=1}^{N}\xi_i^\mu\sigma_i z_\mu\right), \tag{11}$$

where $d\mu(z_\mu)$ is the standard unit Gaussian measure for all the z_μ.

Taken F as a generic function of the neurons, we define the Boltzmann state $\omega_\beta(F)$, at a given level of noise or inverse temperature β, as

$$\omega_\beta(F)=\omega(F)=(Z_N(\beta;\xi))^{-1}\sum_{\{\sigma\}}F(\sigma)e^{-\beta H_N(\sigma,\xi)}, \tag{12}$$

and often we drop the subscript β for the sake of simplicity. The s-replicated Boltzmann measure is defined as $\Omega=\omega^1\times\omega^2\times\cdots\times\omega^s$ in which all the single Boltzmann states are independent and share identical values for the quenched noise ξ. For the sake of clearness, given a function F of the neurons of the s replicas, and using the symbol $a\in[1,\dots,s]$ to label replicas, such an average can be written as

$$\Omega(F(\sigma^1,\dots,\sigma^s))=\frac{1}{Z_N^s}\sum_{\{\sigma^1\}}\sum_{\{\sigma^2\}}\cdots\sum_{\{\sigma^s\}}F(\sigma^1,\dots,\sigma^s)\exp\left(-\beta\sum_{a=1}^{s}H_N(\sigma^a,\xi)\right). \tag{13}$$

We use the symbol $\langle.\rangle$ to mean $\langle.\rangle=\mathbb{E}\Omega(.)$. Notice that the Boltzmann states can be equivalently expressed in terms of $\tilde{Z}$ and the *Boltzmannfaktor* in (9).

In the thermodynamic limit, $N\to\infty$, $K\to\infty$ with $K/N\to\alpha$, it is assumed

$$\lim_{N\to\infty}A_{N,K}(\beta)=A(\alpha,\beta)=-\beta f(\alpha,\beta)$$

(therefore α is a second parameter of the theory in addition to β). We remind that $f(\alpha,\beta)=u(\alpha,\beta)-\beta^{-1}s(\alpha,\beta)$ is the free energy density, $u(\alpha,\beta)$ the internal energy density and $s(\alpha,\beta)$ the intensive entropy.

Reflecting the bipartite nature of the Hopfield model expressed by (11), we introduce two order parameters The first is the overlap between the replicated neurons (first party overlap), defined as

$$q_{ab} = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^a\sigma_i^b \in [-1,+1], \tag{14}$$

and the second the overlap between the replicated Gaussian variables z (second party overlap), defined as

$$p_{ab} = \frac{1}{K}\sum_{\mu=1}^{K} z_a^\mu z_b^\mu \in (-\infty,+\infty). \tag{15}$$

Both the two order parameters above play a considerable role in the theory as they can express thermodynamical quantities [5].

3 Replica Symmetric Free Energy

In this section we pay attention to the structure of the free energy: we want to obtain the latter via a sum rule in which we may isolate explicitly the order parameter fluctuations so to be able to neglect them achieving a replica-symmetric behavior.

Due to the equivalence among neural network and bipartite spin-glasses, we generalize the way cavity field and the stochastic stability techniques work on spin glasses to these structures by introducing a new interpolation scheme.

For the sake of clearness, in order to exploit the interpolation method adapted to the physics of the model, we introduce three free parameters in the interpolating structure (i.e. a, b, c) that we fix *a posteriori*, toward the establishment of the sum rule.

In a pure stochastic stability fashion [10], we need to introduce also two classes of i.i.d. $\mathcal{N}[0,1]$ variables, namely N variables η_i and K variables $\tilde{\eta}_\mu$, whose average is still encoded into the $\mathbb{E}$ operator and by which we define the following interpolating quenched pressure $\tilde{A}_{N,K}(\beta,t)$

$$\tilde{A}_{N,K}(\beta,t) = \frac{1}{N}\mathbb{E}\log\sum_\sigma\int\prod_\mu^K d\mu(z_\mu)\exp\left(\sqrt{t}\sqrt{\frac{\beta}{N}}\sum_{i,\mu}^{N,K}\xi_i^\mu\sigma_i z_\mu\right) \cdot\exp\left(a\sqrt{1-t}\sum_i^N\eta_i\sigma_i\right)\exp\left(b\sqrt{1-t}\sum_\mu^K\tilde{\eta}_\mu z_\mu\right)\exp\left(c\frac{(1-t)}{2}\sum_\mu^K z_\mu^2\right). \tag{16}$$

Notice that we have found convenient to introduce not only the cavity fields η_i and $\tilde{\eta}_\mu$ but also a term producing a continuous renormalization of the z variables. The existence of the z integrals requires $c < 1$, of course. Without loss of generality, as it will clear in the following, we may also assume $c \geq 0$. We remark that $t \in [0,1]$ interpolates between the value $t = 0$, where the interpolating quenched pressure becomes made of by non-interacting systems (a series of one-body problems) whose solution is straightforward, and the opposite limit, $t = 1$, that recovers the correct quenched free energy.
The plan is then to evaluate the t-streaming of such a quantity and then obtain the correct expression by using the fundamental theorem of calculus:

$$\frac{1}{N}\mathbb{E}\log\tilde{Z}_{N,K} = \tilde{A}_{N,K}(\beta,t=1) = \tilde{A}_{N,K}(\beta,t=0) + \int_0^1 dt'\left(\frac{d\tilde{A}_{N,K}(\beta,t)}{dt}\right)_{t=t'}. \tag{17}$$

When evaluating the streaming $d\tilde{A}_{N,K}(\beta,t)/dt$ we get the sum of four terms (A,B,C,D). Each comes as a consequence of the derivation of a corresponding exponential term appearing into the expression (16).

Once introduced the averages $\langle\cdot\rangle_t$ that naturally extend the Boltzmann measure encoded in the interpolating scheme (and reduce to the proper one whenever setting $t=1$), we can write them down as

$$A=\frac{1}{N}\sqrt{\frac{\beta}{N}}\frac{1}{2\sqrt{t}}\sum_{i,\mu}^{N,K}\mathbb{E}\xi_{i,\mu}\omega(\sigma_i z_\mu)=\frac{\beta}{2N}\mathbb{E}\sum_{\mu}^{K}\omega(z_\mu^2)-\frac{\alpha\beta}{2}\langle q_{12}p_{12}\rangle_t,$$

$$B=\frac{-a}{2N\sqrt{1-t}}\sum_{i}^{N}\mathbb{E}\eta_i\omega(\sigma_i)=-\frac{a^2}{2}\big(1-\langle q_{12}\rangle_t\big),$$

$$C=\frac{-b}{2N\sqrt{1-t}}\sum_{\mu}^{K}\mathbb{E}\tilde{\eta}_\mu\omega(z_\mu)=\frac{-b^2}{2N}\sum_{\mu}^{K}\mathbb{E}\omega(z_\mu^2)+\frac{\alpha b^2}{2}\langle p_{12}\rangle_t,$$

$$D=\frac{-c}{2N}\sum_{\mu}^{k}\omega(z_\mu^2),$$

where in the first three equations we used the well known property of the Gaussian expectation $\mathbb{E}(\xi F(\xi))=\mathbb{E}(\partial_\xi F(\xi))$ (integration by parts, or Wick theorem). Notice that we have written α, in place of K/N, by neglecting terms irrelevant in the infinite volume limit.

Summing up all contributions (A,B,C,D) we get

$$\frac{d\tilde{A}_{N,K}(\beta,t)}{dt}=(\beta-b^2-c)\frac{1}{2N}\mathbb{E}\sum_{\mu}^{K}\omega(z_\mu^2)-\frac{\alpha\beta}{2}\langle q_{12}p_{12}\rangle_t$$
$$-\frac{a^2}{2}(1-\langle q_{12}\rangle_t)+\frac{\alpha b^2}{2}\langle p_{12}\rangle_t. \tag{18}$$

So we see that if we introduce the new parameters $\bar{q}$ and $\bar{p}$, with $0\le\bar{q}\le 1, 0\le\bar{p}$, and choose

$$a=\sqrt{\alpha\beta\bar{p}},\qquad b=\sqrt{\beta\bar{q}},\qquad c=\beta(1-\bar{q}),$$

then we get

$$\frac{d\tilde{A}_{N,K}(\beta,t)}{dt}=-\frac{\alpha\beta}{2}\langle(q_{12}-\bar{q})(p_{12}-\bar{p})\rangle_t-\frac{\alpha\beta}{2}\bar{p}(1-\bar{q}), \tag{19}$$

where we have added and subtracted a term $\alpha\beta\bar{q}\bar{p}/2$, in order to center and complete the product of the two overlaps. Notice that there is an additional bound on $\bar{q}$, resulting from $c<1$, in the form $\bar{q}>1-\beta^{-1}$, which is effective only if $\beta>1$.

By inserting the expression (19) into (17) we obtain the wanted sum rule, provided the explicit expression of $\tilde{A}_{N,K}(\beta,t=0)$ is known. But this is easily obtained, because there is complete factorization with respect to the σ_i and the z_μ variables. Through a simple direct calculation we have

$$
\begin{aligned}
&\tilde{A}_{N,K}(\beta, t=0)\\
&= \frac{1}{N}\mathbb{E}\log\sum_{\sigma}\int\prod_{\mu}^{K} d\mu(z_\mu) e^{\sqrt{\alpha\beta\bar{p}}\sum_i^N \eta_i\sigma_i} e^{\sqrt{\beta\bar{q}}\sum_\mu^K \tilde{\eta}_\mu z_\mu} e^{\frac{\beta}{2}(1-\bar{q})\sum_\mu^{kN} z_\mu^2}\\
&= \frac{1}{N}\mathbb{E}\log\sum_{\sigma} e^{\sqrt{\alpha\beta\bar{p}}\sum_i^N \eta_i\sigma_i}\\
&\quad + \frac{1}{N}\mathbb{E}\log\int\left(\prod_{\mu}^{K}\frac{dz_\mu}{2\pi}\right) e^{-\frac{1}{2}\sum_\mu^K z_\mu^2(1-\beta(1-\bar{q}))} e^{\sqrt{\beta\bar{q}}\sum_\mu^K \tilde{\eta}_\mu z_\mu}\\
&= \log 2 + \int d\mu(\eta)\log\cosh\left(\eta\sqrt{\alpha\beta\bar{p}}\right)\\
&\quad + \frac{\alpha}{2}\log\left(\frac{1}{1-\beta(1-\bar{q})}\right) + \frac{\alpha\beta}{2}\frac{\bar{q}}{1-\beta(1-\bar{q})},
\end{aligned}
\tag{20}
$$

where η is $\mathcal{N}(0,1)$, and we have rescaled each z_μ variable according to the new Gaussian variance σ defined by

$$
\sigma^2 = (1-\beta(1-\bar{q}))^{-1}. \tag{21}
$$

As a consequence, by collecting all our results, and by taking into account the connection between Z and $\tilde{Z}$, we can write the final sum rule in the form

$$
\frac{1}{N}\mathbb{E}(\log Z_{N,K}(\beta,\xi)) + \frac{\alpha\beta}{2}\int_0^1 \langle (q_{12}-\bar{q})(p_{12}-\bar{p})\rangle_t dt = \bar{A}(\bar{p},\bar{q};\alpha,\beta), \tag{22}
$$

where the trial function $\bar{A}$ is defined through

$$
\begin{aligned}
\bar{A}(\bar{p},\bar{q};\alpha,\beta) &= \log 2 + \int d\mu(\eta)\log\cosh(\eta\sqrt{\alpha\beta\bar{p}})\\
&\quad + \frac{\alpha}{2}\log\left(\frac{1}{1-\beta(1-\bar{q})}\right) + \frac{\alpha\beta}{2}\frac{\bar{q}}{1-\beta(1-\bar{q})} - \frac{\alpha\beta}{2}\bar{p}(1-\bar{q}) - \frac{\alpha\beta}{2}.
\end{aligned}
\tag{23}
$$

This sum rule holds as it stands for all allowed values of the trial parameters $\bar{q}$ and $\bar{p}$. The sum rule gives a nice expression of the free energy in terms of the trial function $\bar{A}$, with a correction expressed in terms of the fluctuations of p_{12} and q_{12}, with respect to the trial values $\bar{p}$ and $\bar{q}$.

Our task now is to fix the trial parameters $\bar{p}$ and $\bar{q}$, so that the influence of the fluctuation term is kept to a minimum. Here the experience coming from the consideration of the elementary case of the bipartite mean field ferromagnet [9] is precious. In fact, we know that a minimax principle must be involved, as in any case of bipartite models.

To this purpose, we need to establish the following properties of the dependence of the trial function from the trial parameters.

First of all $\bar{A}(\bar{p},\bar{q})$ is concave in $\bar{p}$ for any allowed value of $\bar{q}$. The proof is simple. In fact, a direct standard calculation gives

$$
\partial_{\bar{p}}\bar{A}(\bar{p},\bar{q}) = \frac{\alpha\beta}{2}\left(\bar{q} - \int d\mu(\eta)\tanh^2(\eta\sqrt{\alpha\beta\bar{p}})\right). \tag{24}
$$

The right hand side is clearly decreasing in $\bar{p}$, and concavity is established. Moreover, the trial function, for a fixed $\bar{q}$, assumes its maximum value for a well defined value of $\bar{p}$, let us say $\bar{p}(\bar{q})$, where the derivative vanishes, and therefore

$$\bar{q} = \int d\mu(\eta) \tanh^2\Big(\eta\sqrt{\alpha\beta\,\bar{p}(\bar{q})}\,\Big). \tag{25}$$

It is very easy to understand the meaning of this expression. In fact, if we calculate the average $\langle q_{12}\rangle_{t=0}$ at $t = 0$, by taking into account that here the state is factorized, we find

$$\langle q_{12}\rangle_{t=0} = \int d\mu(\eta) \tanh^2\Big(\eta\sqrt{\alpha\beta\,\bar{p}}\,\Big). \tag{26}$$

From this expression, one can show easily that $\bar{p}(\bar{q})$ is increasing and convex in $\bar{q}$. Moreover, $\bar{p}(0) = 0$, and $\bar{p}(\bar{q})/\bar{q}$ is increasing in $\bar{q}$. We have also $\lim_{\bar{q}\to 0} \bar{p}(\bar{q})/\bar{q} = (\alpha\beta)^{-1}$.

Let us now consider the trial function, $\bar{A}(\bar{p}(\bar{q}), \bar{q})$, at the value where the maximum in $\bar{p}$ is reached. A simple calculation shows that

$$\frac{1}{2\bar{q}}\frac{d}{d\bar{q}}\bar{A}(\bar{p}(\bar{q}), \bar{q}) = \frac{\alpha\beta}{4}\left(\frac{\bar{p}(\bar{q})}{\bar{q}} - \frac{\beta}{(1-\beta(1-\bar{q}))^2}\right). \tag{27}$$

We can see that the right hand side is increasing in $\bar{q}$. Therefore, $\bar{A}(\bar{p}(\bar{q}), \bar{q})$, as a function of $\bar{q}^2$, is convex. The minimum is achieved, in general, at the value of $\bar{q}$ where the derivative vanishes. Therefore, we can uniquely define the replica symmetric approximation through the minimax principle

$$A_{RS}(\alpha, \beta) = \min_{\bar{q}} \max_{\bar{p}} \bar{A}(\bar{p}, \bar{q}; \alpha, \beta) = \bar{A}(\bar{p}(\alpha, \beta), \bar{q}(\alpha, \beta); \alpha, \beta). \tag{28}$$

The values $(\bar{p}(\alpha, \beta), \bar{q}(\alpha, \beta))$, where the minimax principle is realized, are uniquely defined as the (nontrivial) intersection of the two curves

$$\bar{q} = \int d\mu(\eta) \tanh^2\left(\eta\sqrt{\alpha\beta\,\bar{p}}\right), \quad \bar{p} = \frac{\beta\bar{q}}{(1-\beta(1-\bar{q}))^2}, \tag{29}$$

on the $(\bar{p}, \bar{q})$ plane.

Now, the meaning of the minimax principle is clear. The trial parameters are chosen in such a way to correctly reproduce the averages $\langle q_{12}\rangle_{t=0} = \bar{q}$, $\langle p_{12}\rangle_{t=0} = \bar{q}$, at the initial factorized state at $t = 0$. The max part provides for q_{12}, as mentioned before, while the min part provides for p_{12}, as a simple analogous calculation shows. As t increases from the value $t = 0$, the overlaps stay locked for a while, producing the vanishing of the correlation correction in the sum rule. Then, eventually the locking is lost, and a departure from the replica symmetric approximation is expected, as we will prove in the following at least for high values of β.

Notice that only the trivial intersection $\bar{p} = 0, \bar{q} = 0$ is possible, in the region where $\beta(1+\sqrt{\alpha}) < 1$. In fact, for $\beta < 1$, we have at $\bar{q} = 0$

$$\frac{d}{d\bar{q}^2}\bar{A}(\bar{p}(\bar{q}), \bar{q})|_{\bar{q}=0} = \frac{1}{4}\left(1 - \frac{\alpha\beta^2}{(1-\beta)^2}\right). \tag{30}$$

Therefore, if $\alpha\beta^2 < (1-\beta)^2$, i.e. $\beta(1+\sqrt{\alpha}) < 1$, this derivative at the origin is positive, and the minimum is achieved at $\bar{q} = 0$, which implies also $\bar{p} = 0$. But this is the annealed region. Here, for $\bar{q}=0$, $\bar{p}=0$, our replica symmetric approximation, given by (23) and (28), coincides with the annealed expression for the free energy [5], which is considered to be correct in the infinite volume limit, in full agreement with the AGS results.

4 Properties of the Replica Symmetric Free Energy

We have already seen that our replica symmetric approximation gives the right answer in the annealed region.

It is also important to remark that a minimax principle for the free energy emerges also for the dichotomic model with an external field, in the region where the replica symmetric expression can be shown to hold rigorously in the infinite volume limit, as proved for example by Michel Talagrand in his book [22].

At first sight, it is surprising that the definition of the replica symmetric *Ansatz*, as given above for the analogical neural network, does not include a spontaneous magnetization term, forcing the spins to align along some given stored pattern, as it happens in the dichotomic case. As a matter of fact, we could include a term of this kind in the interpolation procedure, ruled by an additional trial parameter M. It is enough to single out one of the terms in (9), for example m_1^2, and let all other terms undergo the Gaussian transformation, as explained before. Then, we could insert a new term in the interpolating expression (16) of the form $tm_1^2+(1-t)Mm_1$, by following for this particular magnetization the standard procedure as in the Curie-Weiss model (see for example [4, 13]), which implies maximization with respect to M. We would end up with an apparently more general sum rule

$$\frac{1}{N}\mathbb{E}(\log Z_{N,K}(\beta,\xi))+\frac{\alpha\beta}{2}\int_0^1\langle(q_{12}-\bar{q})(p_{12}-\bar{p})\rangle_t dt$$
$$=\frac{\beta}{2}\int_0^1\langle(m_1-M)^2)\rangle_t dt+\bar{A}(\bar{p},\bar{q},M;\alpha,\beta), \tag{31}$$

where there is also the additional fluctuation of m_1 with respect to its trial value M. Now the trial function $\bar{A}$ is defined through

$$\bar{A}(\bar{p},\bar{q},M;\alpha,\beta)=\log 2+\int d\mu(\eta)\log\cosh(\eta\sqrt{\alpha\beta\bar{p}+\beta^2M^2})$$
$$+\frac{\alpha}{2}\log\left(\frac{1}{1-\beta(1-\bar{q})}\right)+\frac{\alpha\beta}{2}\frac{\bar{q}}{1-\beta(1-\bar{q})}-\frac{\alpha\beta}{2}\bar{p}(1-\bar{q})$$
$$-\frac{\alpha\beta}{2}-\frac{\beta}{2}M^2. \tag{32}$$

Let us notice that the additional order parameter M appears under the square root in the integral with respect to η, because in any case the independent Gaussian cavity fields η_i and the Gaussian fields ξ_i^1 in m_1 can be lumped together into new Gaussian cavity fields with the proper variance.

Along the minimax procedure, the treatment of the $\bar{p},\bar{q}$ variables is similar to the case without M. For the M derivative we have

$$\partial_{M^2}\bar{A}=\frac{\beta}{2}\big(\beta(1-\bar{q})-1\big)<0, \tag{33}$$

since in any case $\beta(1-\bar{q})<1$ must be true. Therefore, $\bar{A}$ is decreasing in M^2, and the optimal value in the max part of the variational principle is reached for $M=0$. Therefore, no M parameter is necessary. In a sense, it is the parameter $\bar{p}$ which replaces completely M.

In fact, as a test, we can consider the limit for $\alpha\to 0$ of the replica symmetric *Ansatz* (low charge limit). The interesting case is when $\beta>1$. Then $\bar{q}(\alpha,\beta)>1-\beta^{-1}$. The self-consistent equations (29) imply that $\bar{q}(\alpha,\beta)\to 1-\beta^{-1}$ in this limit. Moreover, $\bar{p}(\alpha,\beta)\to\infty$, in such a way that $\alpha\bar{p}(\alpha,\beta)\to\beta M^2(\beta)$, where $M(\beta)$ is the unique solution of the equation $\beta(1-\int d\mu(\eta)\tanh^2(\beta M(\beta)\eta))=1$. Moreover, we have

$$\lim_{\alpha\to 0}\bar{A}_{RS}(\alpha,\beta)=\log 2+\int d\mu(\eta)\log\cosh(\beta M(\beta)\eta)-\frac{\beta}{2}M^2(\beta). \tag{34}$$

Therefore, we see that in the low charge limit we recover exactly the expression for the magnetization order parameter, and for the free energy, of the neural net constructed with a finite fixed number K of stored patterns, in the infinite volume limit $N\to\infty$. We refer to [6], and references quoted there, for an extensive treatment of the low charge limit.

As a further test of consistency in the definition of the replica symmetric *Ansatz*, we can consider the spin glass limit of the neural net. Let us recall how it is defined. Start from the expression (5) of the partition function of the neural net. By keeping fixed the number of spins N, let us take $K\to\infty$, and also $\beta\to 0$, in such a way that $\beta\sqrt{K/N}\to\beta'$, where β' will act as inverse temperature of the limiting spin glass. By interpreting K/N as α, we could say in a sense that an $\alpha\to\infty$ limit is involved. Let us write the interaction exponent in (5) in the equivalent form

$$\beta\sqrt{\frac{K}{N}}\frac{1}{\sqrt{N}}\sum_{i<j}^{N}\left(\frac{1}{\sqrt{K}}\sum_{\mu=1}^{K}\xi_i^\mu\xi_j^\mu\right)\sigma_i\sigma_j. \tag{35}$$

In the limit, we have $\beta\sqrt{K/N}\to\beta'$, by definition. Moreover, as a consequence of the central limit theorem, for each couple (i,j), we have

$$\lim_{K\to\infty}\frac{1}{\sqrt{K}}\sum_{\mu=1}^{K}\xi_i^\mu\xi_j^\mu=J_{ij},$$

in distribution, where the J_{ij}'s are a family of independent unit Gaussian random variables. Therefore, the interaction exponent converges to $\beta'\sum_{i<j}^{N}J_{ij}\sigma_i\sigma_j/\sqrt{N}$, which is the Sherrington-Kirkpatrick mean field expression.

On the other hand, if we start from (28), and (29), and perform the limit $\alpha\to\infty$, $\beta\to 0$, $\beta\sqrt{\alpha}\to\beta'$, we find $\bar{q}(\alpha,\beta)\to\hat{q}(\beta')$, $\bar{p}(\alpha,\beta)\to 0$, $\bar{p}(\alpha,\beta)/\beta\to\hat{q}(\beta')$. Here $\hat{q}(\beta')$ is the nontrivial solution of $\hat{q}(\beta')=\int\tanh^2(\beta'\sqrt{\hat{q}(\beta')}\eta)d\mu(\eta)$, if $\beta'>1$, while $\hat{q}(\beta')=0$ if $\beta'\le 1$. As a consequence

$$\lim_{\alpha\to\infty}A_{RS}(\alpha,\beta)=\hat{A}(\beta')=\log 2+\int\log\cosh\left(\beta'\sqrt{\hat{q}(\beta')}\eta\right)d\mu(\eta)+\frac{\beta'^2}{4}(1-\hat{q}(\beta'))^2. \tag{36}$$

Therefore, we see that in the limit we recover the correct expression for the well known replica symmetric *Ansatz* for the spin glass, which is known to be rigorously an upper bound for the pressure [11].

Therefore, our replica symmetric *Ansatz* for the neural net is known to be the exact solution in the annealed region and in the low charge limit, while it turns out to be an upper bound for the pressure in the spin glass limit. An attractive conjecture is that it is always an upper bound in the whole parameter region. But in order to establish this rigorously it would be necessary to have a better control on the fluctuation correcting terms in the sum rule. Further research toward this objective is necessary.

We end this section, by reporting about the properties of the replica symmetric *Ansatz* in the low temperature limit, as $\beta \to \infty$. Our main results concern the ground state energy and its associated entropy, defined by

$$\hat{e}_{RS}(\alpha) = \lim_{\beta\to\infty} \partial_\beta \bar{A}_{RS}(\alpha, \beta) = \lim_{\beta\to\infty} \bar{A}_{RS}(\alpha, \beta)/\beta, \tag{37}$$

$$\hat{s}_{RS}(\alpha) = \lim_{\beta\to\infty} \big(\bar{A}_{RS}(\alpha, \beta) - \beta\partial_\beta \bar{A}_{RS}(\alpha, \beta)\big). \tag{38}$$

First of all, from the self-consistency equations (29), through a long and straightforward calculation, we can easily derive the following essential information on the low temperature limit for the order parameters, $\bar{q}(\alpha, \beta) \to 1$, $\beta(1 - \bar{q}(\alpha, \beta)) \to \sqrt{2/\alpha\pi}(1 + \sqrt{2/\alpha\pi})^{-1}$, $\bar{p}(\alpha, \beta) \to \infty$, $\bar{p}(\alpha, \beta)/\beta \to (1 + \sqrt{2/\alpha\pi})^2$. Then, by exploiting the definition (28), we derive the following expressions for the ground state energy and the entropy

$$\hat{e}_{RS}(\alpha) = \frac{1}{\pi} + \sqrt{\frac{2}{\pi}}\sqrt{\alpha}, \tag{39}$$

$$\hat{s}_{RS}(\alpha) = \frac{1}{2}\alpha\left(\log\left(1 + \sqrt{\frac{2}{\alpha\pi}}\right) - \sqrt{\frac{2}{\alpha\pi}}\right). \tag{40}$$

Notice that the entropy $\hat{s}_{RS}(\alpha)$ suffers the typical disease of a replica symmetric *Ansatz* of being negative. Therefore, the true solution of the model must involve replica symmetry breaking. However, in the low charge limit we have the correct behavior $\lim_{\alpha\to 0} \hat{s}_{RS}(\alpha) = 0$, while in the spin glass limit we have $\lim_{\alpha\to\infty} \hat{s}_{RS}(\alpha) = -1/(2\pi)$, which is the ground state entropy of the replica symmetric approximation for the spin glass. Analogously, for the ground state energy we have the correct low charge limit $\lim_{\alpha\to 0} \hat{e}_{RS}(\alpha) = 1/\pi$, and the correct spin glass replica symmetric limit $\lim_{\alpha\to\infty} \hat{e}_{RS}(\alpha)/\sqrt{\alpha} = \sqrt{2/\pi}$.

It is important to remark that the expression of the ground state energy and entropy is completely smooth in the whole parameter range $0 \le \alpha < \infty$. No phase transition associated with a kind of saturation in the retrieval procedure has been found.

This ends our discussion about the properties of the replica symmetric *Ansatz* in the analogical neural network.

5 Conclusion and Outlook for Future Development

We have considered a neural network model with quenched memories encoded by words with continuous Gaussian distributed values. By generalizing methods exploited in the spin glass case, we have introduced an interpolation procedure in order to reach the replica symmetric approximation for the model. This leads to an *Ansatz* for the free energy which involves two order parameters, related to the overlaps. The order parameters are uniquely defined by self-consistent equations, in terms of the two free parameters of the model, the temperature and the charge of the storage.

We have found that this replica symmetric approximation is indeed exact in the annealed region, and in the low charge limit, while it gives the known replica symmetric approximation in the spin glass limit. We have also studied the low temperature limit, by finding that the entropy is negative, as expected. In the low temperature region, no evidence of a phase transition, associated with the breaking of the retrieval power, has been found, in contrast with the well known results of the replica symmetric approximation in the dichotomic case.

Our method can be immediately expanded toward a full replica breaking scheme, by following the same procedure exploited in the spin glass case. At the replica symmetric level the two order parameters are related to the presumed values of the two overlaps, the first among the spin variables, the second among the auxiliary variables entering the transformation into a bipartite spin glass. When full replica symmetry is assured, then the order parameter will acquire a functional character, associated to the joint distribution of the overlaps. In a forthcoming paper, we will show how our scheme extends to the fully broken replica symmetry case, where now the order parameters include the fluctuations of the overlaps.

Acknowledgements Support from MiUR through the FIRB project RBFR08EKEV and INFN (Italian Institute for Nuclear Physics) is gratefully acknowledged.

A.B. work is supported by the SmartLife Project too (Ministry Decree 13/03/2007 n. 368) which is acknowledged together with GNFM for partially covering travel expenses.

References

1. Amit, D.J.: Modeling Brain Function: The World of Attractor Neural Network. Cambridge University Press, Cambridge (1992)
2. Amit, D.J., Gutfreund, H., Sompolinsky, H.: Spin glass model of neural networks. Phys. Rev. A **32**, 1007–1018 (1985)
3. Amit, D.J., Gutfreund, H., Sompolinsky, H.: Storing infinite numbers of patterns in a spin glass model of neural networks. Phys. Rev. Lett. **55**, 1530–1533 (1985)
4. Barra, A.: The mean field Ising model trough interpolating techniques. J. Stat. Phys. **132**, 12–32 (2008)
5. Barra, A., Guerra, F.: About the ergodic regime in the analogical Hopfield neural networks: moments of the partition function. J. Math. Phys. **50**, 125217 (2008)
6. Bovier, A., Picco, P.: Mathematical Aspects of Spin Glasses and Neural Networks. Birkhäuser, Basel (1998) and references therein
7. Bovier, A., van Enter, A.C.D., Niederhauser, B.: Stochastic symmetry-breaking in a Gaussian Hopfield-model. J. Stat. Phys. **95**, 181–213 (1999)
8. Coolen, A.C.C., Kuehn, R., Sollich, P.: Theory of Neural Information Processing Systems. Oxford University Press, Oxford (2005)
9. Genovese, G., Barra, A.: A certain class of Curie-Weiss models. arXiv:0906.4673 (2009)
10. Guerra, F.: About the overlap distribution in mean field spin glass models. Int. J. Mod. Phys. B **10**, 1675–1684 (1996)
11. Guerra, F.: Sum rules for the free energy in the mean field spin glass model. In: Mathematical Physics in Mathematics and Physics: Quantum and Operator Algebraic Aspects. Fields Institute Communications, vol. 30 Amer. Math. Soc., Providence (2001)
12. Guerra, F.: Broken replica symmetry bounds in the mean field spin glass model. Commun. Math. Phys. **233**, 1–12 (2003)
13. Guerra, F.: An introduction to mean field spin glass theory: methods and results. In: Bovier, A., et al. (eds.) Mathematical Statistical Physics, pp. 243–271. Elsevier, Oxford (2006)
14. Guerra, F., Toninelli, F.L.: The thermodynamic limit in mean field spin glass models. Commun. Math. Phys. **230**, 71–79 (2002)
15. Guerra, F., Toninelli, F.L.: The infinite volume limit in generalized mean field disordered models. Markov Process. Relat. Fields **9**, 195–207 (2003)
16. Hebb, D.O.: Organization of Behaviour. Wiley, New York (1949)
17. Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. **79**, 2554–2558 (1982)
18. Pastur, L., Scherbina, M., Tirozzi, B.: The replica symmetric solution of the Hopfield model without replica trick. J. Stat. Phys. **74**, 1161–1183 (1994)
19. Pastur, L., Scherbina, M., Tirozzi, B.: On the replica symmetric equations for the Hopfield model. J. Math. Phys. **40**, 3930–3947 (1999)

20. Talagrand, M.: Rigourous results for the Hopfield model with many patterns. Probab. Theory Relat. Fields **110**, 177–276 (1998)
21. Talagrand, M.: Exponential inequalities and convergence of moments in the replica-symmetric regime of the Hopfield model. Ann. Probab. **38**, 1393–1469 (2000)
22. Talagrand, M.: Spin Glasses: A Challenge for Mathematicians. Cavity and Mean Field Models. Springer, Berlin (2003)