Saddle point equations of a neural network with correlated attractors
M. Shcherbina, B. Tirozzi

To cite this version:

HAL Id: jpa-00246937
https://hal.archives-ouvertes.fr/jpa-00246937
Submitted on 1 Jan 1994

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Short Communication

Saddle point equations of a neural network with correlated attractors

M. Shcherbina (1) and B. Tirozzi (2)

(1) Institute for Low Temperature Physics and Engineering, 47 Lenin Avenue, Kharkov 310164, Ukraine
(2) Department of Mathematics, University of Rome "La Sapienza", P.le Aldo Moro, 5 — 00185, Roma, Italy

(Received 17 January 1994, received in final form 7 February 1994, accepted 25 February 1994)

Abstract. — We derive in a rigorous way, using a new technique, the saddle-point equations of a modified Hopfield model which stores sequences of patterns. We obtain the same equations as those of the replica-symmetry approach without using the $n \to 0$ limit but assuming selfaveraging of the Edwards-Anderson order parameter.

Introduction.

We consider a model proposed by Griniasty et al. [1–3] which generalizes the Hopfield model by adding to the Hebb interaction [6, 7] a term coupling patterns with different indexes $\mu$. Surprisingly, the behaviour of such a model is similar to the results of Miyashita’s experiment [4, 5] in which patterns presented in a certain order in time originate attractors with space correlations. In these models patterns having $\mu$ index different by one are coupled. We allow instead a coupling among all the pairs with different $\mu$ index through a generic Toeplitz matrix. The model is solved using an interpolating technique [8] which avoids the mathematically undefined $n \to 0$ limit, and the replica symmetry assumption is substituted with the self-averageness of the Edwards-Anderson parameter. We get the same equations as in [3] if we take our matrix to be equal to the one used there. In [3] an additional hypothesis was used that one can divide the patterns among condensing or not condensing. This hypothesis was verified by simulations in the case of their particular coupling. We do not need such an assumption that is not likely to be true in the general case in which all the patterns are coupled. In fact we obtain saddle-point equations which allow any decay law of $m^\mu$ with $\mu$. In this paper we derive the saddle point equations using two theorems. The proofs of these theorems will be published in a longer paper with more details.
1. The Hamiltonian and the basic assumptions.

We consider the following Hamiltonian:

\[ H = -\frac{1}{2N} \sum_{i,j=1}^{N} \sum_{\mu,\nu=1}^{P} a_{\mu-\nu} \xi_i^{\mu} \xi_j^{\nu} S_i S_j + \frac{\varepsilon}{N} \sum_{i=1}^{N} \sum_{\mu=1}^{P} c^{\mu} \xi_i^{\mu} S_i \]

\[ + \frac{\varepsilon_1}{N} \sum_{i=1}^{N} h_i S_i + \frac{\varepsilon_2}{N} \sum_{\mu=1}^{P} \gamma^{\mu} t^{\mu} \]  \( (1.1) \)

where \( \xi_i^{\mu} = \pm 1, \mu = 1 \ldots P, i = 1 \ldots N \) are the patterns stored into the system. These variables are taken as independent identically distributed random variables with zero mean. The index of the lattice site is denoted by \( i \) and \( \mu \) is the index of the pattern. The \( S_i \) are the neuronal activities with \( S_i = \pm 1 \). \( A = \{a_{\mu-\nu}\}_{\mu,\nu=1}^{P} \) is a Toeplitz matrix such that \( a_{\mu+p} = a_{\mu} \). It coincides with the matrix introduced in [1] in the case \( a_0 = 1, a_1 = a_{-1} = a, a_k = 0 \ (k \geq 2) \).

The term multiplying \( \varepsilon \) is introduced in order to have overlaps \( m^{\mu} \) different from zero. The phase transition is studied by sending \( \varepsilon \to 0 \). The coefficients \( c^{\mu} \) are taken in such a way that \( \sum_{i=1}^{N} (c^{\mu})^2 < +\infty \).

The \( \varepsilon_1 \) and \( \varepsilon_2 \) terms have been added in order to obtain some useful properties needed in the derivation of the theorems shown below. \( h_i, \gamma^{\mu} \) are independent identically distributed random with zero mean and variance 1.

The expectation with respect to the Gaussian variables \( h_i, \gamma^{\mu} \) and the patterns \( \xi_i^{\mu} \) will be denoted by \( E \).

The variables \( t^{\mu} \) are defined as in [8]:

\[ t^{\mu} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_i^{\mu} S_i, \quad t_1^{\mu} = \frac{1}{\sqrt{N}} \sum_{i=2}^{N} \xi_i^{\mu} S_i. \]  \( (1.2) \)

we define an interpolating Hamiltonian \( H(\tau) \).

Let \( \tau \) be a sequence of real parameters \( \{\tau_{-P}, \ldots, \tau_0, \tau_1, \ldots, \tau_P\} \). Then the Hamiltonian interpolating between the system with \( N - 1 \) neurons and \( N \) neurons is defined as:

\[ H(\tau) = -\frac{1}{2N} \sum_{i,j=2}^{N} \sum_{\mu,\nu} a_{\mu-\nu} \xi_i^{\mu} \xi_j^{\nu} S_i S_j + \frac{\varepsilon}{N} \sum_{i=1}^{N} \sum_{\mu=1}^{P} c^{\mu} \xi_i^{\mu} S_i \]

\[ + \frac{\varepsilon_1}{N} \sum_{i=2}^{N} h_i S_i + \frac{\varepsilon_2}{N} \sum_{\mu=1}^{P} \gamma^{\mu} t_1^{\mu} + \frac{1}{\sqrt{N}} \sum_{\mu,\nu} \varepsilon \tau_{\mu-\nu} a_{\mu-\nu} \xi_1^{\nu} t_1^{\mu}. \]  \( (1.3) \)

Let \( \tilde{h}_1 \) be defined as:

\[ \tilde{h}_1 = \varepsilon \sum_{\mu=1}^{P} c^{\mu} \xi_1^{\mu} + \varepsilon_1 h_1 + \frac{1}{\sqrt{N}} \sum_{\mu=1}^{P} \gamma^{\mu} \xi_1^{\mu} \]  \( (1.4) \)

then

\[ H(\tau, \eta) = H(\tau) + \eta \tilde{h}_1 \]

is the \( N \)-neuron Hamiltonian \( (1.1) \) when \( \tau = \{S_1, \ldots, S_1\} \) and \( \eta = S_1 \). The variables \( t_1^{\mu} \) introduced in (1.2) have been chosen in such a way to have this property. \( \langle \tau \rangle \) is the expectation with
respect to the Gibbs measure generated by $H(\tau)$. We also introduce the following quantities:

$$U_\mu = \frac{1}{N} \sum_{\nu=1}^{p} \langle t^{\nu+\mu} \rangle, \quad \dot{U}_\mu = \frac{1}{N} \sum_{\nu=1}^{p} \langle \dot{t}^{\nu+\mu} \rangle$$

$$\tau_\mu = \frac{1}{p} \sum_{\nu=1}^{p} (\langle t^{\nu+\mu} \rangle - E(t^{\nu+\mu}) E(t^\nu)), \quad \dot{t}^\mu = t^\mu - \langle t^\mu \rangle$$

where by definition, $t^{\mu+p} = t^\mu$, $\xi_{\mu+p}^{\nu} = \xi_{\mu}^{\nu}$.

We also define the following function, which is the main tool in our approach:

$$u(\tau) = \frac{1}{\beta} \ln \frac{\text{Tr} e^{-\beta H(\tau)}}{\text{Tr} e^{-\beta H(\emptyset)}}$$

where $\text{Tr} = \sum_{\{s_2, \ldots, s_N\}}$ represents the sum over the configurations of the system with $N-1$ lattice points.

The set of $\tau_\mu$ defined in (1.6) are used in order to get the analogues of the $r$ parameter appearing in [3, 8]:

$$r = \sum_{\mu, \nu} \tau_{\mu-\nu} a_\mu a_\nu$$

where

$$\tau_{\mu-\nu} = \begin{cases} \tau_{\mu-\nu} & \text{for } \mu \geq \nu + 1 \\ \tau_{\mu-\nu+p} & \text{for } \mu < \nu + 1 \end{cases} .$$

Let $f_N$ be the free-energy corresponding to the Hamiltonian (1.1), and let $\alpha = p/N$, $\alpha$ fixed, be the capacity of the system, and $q$ the Edwards–Anderson (EA) parameter.

Then we have the following results.

**Theorem 1.**

a) $f_N$ is self-averaging (SA)

b) $m^\mu$ is SA

**Theorem 2.** If $\sup |\tau_\mu| \leq 1$ and of in some range of the parameters $\beta, \varepsilon, \varepsilon_1, \varepsilon_2, \alpha (\varepsilon, \varepsilon_1, \varepsilon_2 \in (0, \delta))$ $q$ is SA then

$$u(\tau) = \sum_{\mu} a_\mu \tau_\mu \sum_{\nu=1}^{p} \frac{\xi_1^{\nu}}{\sqrt{N}} \langle t_{1+\nu}^{\mu} \rangle_0$$

$$+ \sum_{\mu, \nu} \frac{\beta}{2} a_\mu a_\nu \tau_{\mu-\nu} \dot{U}_{\mu-\nu} + R_N ,$$

where $r = \sum_{\mu, \nu} \tau_{\mu-\nu} a_\mu a_\nu$ are SA and $E\{R_N^2\} \to 0$ for $N \to \infty$

$\langle \cdot \rangle_0$ is the expectation $\langle \cdot \rangle_\emptyset$ with $\tau = 0$. These two theorems will be shown in a more complete paper.
2. Derivation of the saddle–point equations

Here we will show how it is possible to get the saddle point equations using Theorems 1 and 2. Let us define the quantity

$$ V_\mu = \sum_{\nu=1}^{p} \left( \frac{\xi_\nu}{\sqrt{N}} \langle t_1^{\mu+\nu} \rangle_0 - \frac{\xi_\nu}{\sqrt{N}} E\langle t_1^{\mu+\nu} \rangle_0 \right) \quad (2.1) $$

Then the variables $V_\mu$ are Gaussian with zero mean and

$$ E_\xi \{ V_\mu V_\nu \} = \alpha E_\xi \{ r_{\mu-\nu} \} \quad (2.2) $$

Here the symbol $E_\xi$ means expectation with respect to the random variable $\xi_1$. Using the same argument as in [8] we get in the $N \to \infty, \varepsilon_1, \varepsilon_2 \to 0$ limit

$$ \langle S_1 \rangle = \tgh \beta ((A_\mu, \xi_1) + \varepsilon \sum_{\mu=1}^{p} c^\mu \xi_1^\mu + \sqrt{\alpha r} v) $$

where $(A_\mu, \xi_1) = \sum_{\mu, \nu} a_{\mu-\nu} a_{\mu} \xi_1^\nu$ and $V = \frac{1}{\sqrt{\alpha r}} \sum_{\mu} a_{\mu} V_\mu$ is a Gaussian random variable with zero mean and variance 1. Then, in the limit $N \to \infty, \varepsilon_1, \varepsilon_2 \to 0$, we have the first saddle point equations:

$$ q = E \int \frac{d\nu}{\sqrt{2\pi}} \tgh^2 \beta ((A_\mu, \xi_1) + \varepsilon \sum_{\mu=1}^{p} c^\mu \xi_1^\mu + \sqrt{\alpha r} v). \quad (2.3) $$

$$ m^\mu = E \left\{ \xi_1^\mu \int \frac{d\nu}{\sqrt{2\pi}} \tgh \beta ((A_\mu, \xi_1) + \varepsilon \sum_{\mu=1}^{p} c^\mu \xi_1^\mu + \sqrt{\alpha r} v) \right\}. \quad (2.4) $$

In the formulae (2.3) and (2.4) the symbol $\xi_1$ represents the vector $\xi_1^1, \ldots, \xi_1^p$. The derivation of the equation for $r$ is more lengthy and we give only some hint. We get the following equation:

$$ E \{ r_\mu \} = \{ \alpha q \delta_{\mu-0} + \lambda (AE \{ r \})_\mu + \beta q (AEU)_\mu \} \quad (2.5) $$

Then we can compute

$$ r = E \{ r \} = \sum_{\mu, \nu} E \{ r_{\mu-\nu} \} a_{\mu} a_{\nu} = q \sum_{\mu, \nu} (1 - \lambda A)^{-1}_{\mu, \nu} a_{\mu} a_{\nu} $$

$$ + \beta q (1 - q) \sum_{\mu, \nu} ((1 - \lambda A)^{-2} A)_{\mu-\nu} a_{\mu} a_{\nu} $$

$$ = \frac{1}{p} q \frac{A^2}{(1 - \lambda A)^2} \int_{-\infty}^{+\infty} \frac{c^2(x)}{(1 - \beta (1 - q) c(x))^2} dx \quad (2.6) $$

where $c(x)$ is the function

$$ c(x) = a_0 + 2 \sum_{i=1}^{p} a_i \cos lx \quad (2.7) $$

Thus we get the third saddle–point equation:

$$ r = \frac{q}{2\pi} \int_{-\pi}^{+\pi} \frac{c^2(x)}{(1 - \beta (1 - q) c(x))^2} dx. \quad (2.8) $$
References