Journal of Statistical Physics, Vol. 62, Nos. 1/2, 1991

Pair-Correlated Patterns in Hopfield Model of Neural Networks

Francisco A. Tamarit[1, 2] and Evaldo M. F. Curado[1]

Received June 7, 1990; final September 5, 1990

We study the retrieval properties of the Hopfield model of neural networks when the memorized patterns are statistically correlated in pairs. There is a finite correlation κ between the memories of each pair, but memories of different pairs are uncorrelated. The analysis is restricted to the case of an arbitrary but finite number of memories in the thermodynamic limit. We find that there are two retrieval regimes: for $0<T<(1-\kappa)$ the system recognizes the stored patterns and for $(1-\kappa)<T<(1+\kappa)$ the system is able to recognize pairs, but it is not able to distinguish between its two patterns.

KEY WORDS: Hopfield model; neural networks; correlated patterns.

In the recent years great effort has been devoted to the study of spin-glass models for associative memory and in particular to the Hopfield model.[1-3] As first proposed, this model is seriously limited by the fact that the memorized configurations have to be uncorrelated in order to provide associative memory, and a fair amount of analytical and numerical work has been done to overcome this restriction.[4-11] Most of this work focuses on the learning rule, attempting to find a suitable modification able to achieve the desired properties, or are concerned only with the storage capacity in the $T \to 0$ limit.

Although it is well known that a finite overlap among the stored configurations tends to destabilize them, little has been said about how the correlation modifies the equilibrium properties of the system before they are completely destabilized. In order to answer this question, we study in this paper the thermodynamics of the Hopfield model as originally

[1] Centro Brasileiro de Pesquisas Fisicas-CNPq, CEP 22290, Rio de Janeiro, Brazil.

[2] Facultad de Matemática, Astronomia y Fisica, Universidad Nacional de Córdoba, 5000 Córdoba, Argentina.

0022-4715/91/0100-0473$06.50/0 © 1991 Plenum Publishing Corporation

proposed, but with the stored patterns statistically correlated. The correlation is introduced in a simple manner, namely, the patterns are grouped in pairs, in such a way that the two patterns of each pair have a finite overlap, while two patterns of different pairs have zero overlap and are statistically uncorrelated. Besides the simplifications introduced in the calculations by considering this particular case instead of a more general one, only in this case does the system provide associative memory for any value of the correlation κ ($0 \leqslant \kappa \leqslant 1$). We restrict ourselves in this paper to the case of an arbitrary but finite number of stored configurations.

The Hopfield model for associative memory consists of a system of N neurons, each one modeled by an Ising spin variable which can assume the values $+1$ and -1, representing the active and passive states of the neuron, respectively. Each neuron is fully connected with the rest of the system through the symmetric synaptic matrix J_{ij} and the dynamics is a Monte Carlo process governed by the Hamiltonian

$$H = -\frac{1}{2} \sum_{i,j} J_{ij} S_i S_j \tag{1}$$

The synaptic matrix is constructed following Hebb's rule

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{r} \sum_{l=1}^{2} \xi_i^{\mu,l} \xi_j^{\mu,l}, \qquad i \neq j \tag{2}$$

where the ξ's are quenched, random variables which take the values ± 1 according to the following distribution:

$$P(\{\xi_i^{\mu,l}\}) = \prod_{i=1}^{N} \prod_{\mu=1}^{r} p(\xi_i^{\mu,1}, \xi_i^{\mu,2}) \tag{3}$$

with

$$\begin{aligned} p(\xi_i^{\mu,1}, \xi_i^{\mu,2}) = {} & \frac{1+\kappa}{4} \delta(\xi_i^{\mu,1}+1)\, \delta(\xi_i^{\mu,2}+1) \\ & + \frac{1-\kappa}{4} \delta(\xi_i^{\mu,1}+1)\, \delta(\xi_i^{\mu,2}-1) \\ & + \frac{1-\kappa}{4} \delta(\xi_i^{\mu,1}-1)\, \delta(\xi_i^{\mu,2}+1) \\ & + \frac{1+\kappa}{4} \delta(\xi_i^{\mu,1}-1)\, \delta(\xi_i^{\mu,2}-1) \end{aligned} \tag{4}$$

For fixed μ and l, $\{\xi_i^{\mu,1}\}$ with $i = 1,\ldots, N$ is a particular configuration of the system. They form a set of $2r$ configurations which have been learnt and according to Eq. (3) they are divided into r pairs denoted by the Greek superscript $\mu = 1, \ldots, r$ and each pattern of the pair is denoted by the Latin superscript $l = 1, 2$. The distribution (4) implies that

$$\langle\langle \xi_i^{\mu l} \xi_i^{\nu m} \rangle\rangle = \delta^{\mu\nu}(\delta^{lm}(1-\kappa)+\kappa) \tag{5}$$

$$\langle\langle \xi_i^{\nu l} \rangle\rangle = 0 \tag{6}$$

where $\langle\langle \cdots \rangle\rangle$ denotes the average with distribution (3). Equation (5) means that two patterns of different pairs are statistically independent, while the two patterns of a pair are statistically correlated. Note that unlike Amit *et al.* did in ref. 4, here the overlap is introduced through a statistically dependent distribution, without removing the condition (6). This model provides one of the simplest hierarchically correlated trees of patterns, consisting of r categories, each one with two memories.[5]

In order to see that in this case the effect of correlation between patterns does not destabilize them, we calculate the local field acting on the site i when the system is in the state $\{S_i\} = \{\xi_i^{\nu, 1}\}$, in the limit $N \to \infty$ and finite r,

$$h_i^{\nu 1} = \lim_{N \to \infty} \sum_j^N J_{ij} \xi_j^{\nu 1} = \xi_i^{\nu 1} + \kappa \xi_i^{\nu 2} \tag{7}$$

The first term in Eq. (7) represents a signal which tends to align S_i with the local field, while the second represents a noise. Nevertheless, the last one can never destabilize the signal, because its modulus is less than the modulus of the signal for any value of the correlation, that is, the $2r$ patterns will be stable for any value of κ $(0 \leqslant \kappa \leqslant 1)$. Although the system still provides associative memory, its performance (size of the basins of attraction, number and properties of the spurious states, storage capacity, etc.) will be modified when the correlation is introduced and we devote the rest of this work to analyzing how some of these properties change. Following the ideas of Amit *et al.*,[2] we introduce a stochastic noise represented by a finite temperature T that measures the level of synaptic noise and study the statistical mechanics of the Hamiltonian (1). We calculate the partition function

$$\mathbb{Z} = \mathrm{Tr}_s \exp(-\beta H) \tag{8}$$

with H given by Eq. (1), J_{ij} by Eq. (2), and the ξ's chosen according to the distributions (3) and (4). For a given realization of the ξ's, $\mathbb{Z}$ can be rewritten as

$$\mathbb{Z} = (N\beta)^r e^{-\beta r} \int \prod_\mu \prod_l \frac{dm^{\mu l}}{(2\pi)^{1/2}} \times \exp\left\{ \frac{N\beta \mathbf{m}^2}{2} + \sum_{i=1}^N \ln[2\cosh(\beta \mathbf{m} \cdot \boldsymbol{\xi}_i)] \right\} \tag{9}$$

where we have adopted the following notation:

$$\mathbf{m} = (m^{11}, m^{12}, m^{21},..., m^{r1}, m^{r2})$$
$$\boldsymbol{\xi}_i = (\xi_i^{11}, \xi_i^{12}, \xi_i^{21},..., \xi_i^{r1}, \xi_i^{r2})$$
$$\mathbf{m} \cdot \boldsymbol{\xi}_i = \sum_{\nu=1} \sum_{l=1} \xi_i^{\nu l} m^{\nu l}$$

In the limit $N \to \infty$ and finite r the order parameter $\mathbf{m}$ can be determined by the saddle point equation. Using the fact that both $\ln(\mathbb{Z})$ and $\mathbf{m}$ are self-averaging, we obtain

$$f(\beta) = \frac{1}{2}\mathbf{m}^2 - \frac{1}{\beta} \langle\langle \ln[2\cosh(\beta \mathbf{m} \cdot \boldsymbol{\xi})] \rangle\rangle \tag{10a}$$

$$\mathbf{m} = \langle\langle \boldsymbol{\xi} \tanh(\beta \mathbf{m} \cdot \boldsymbol{\xi}) \rangle\rangle \tag{10b}$$

In order to know whether a solution of Eq. (10) is stable or not, we also study the eigenvalues of the matrix A given by

$$A^{\mu l \nu m} = \frac{\partial^2 f(\beta)}{\partial m_{\mu l}\, \partial m_{\nu m}} = \delta^{\mu\nu}\delta^{lm} - \beta(\langle\langle \xi^{\nu l}\xi^{\mu m} \rangle\rangle - q^{\mu l \nu m}) \tag{11}$$

with

$$q^{\mu l \nu m} = \langle\langle \xi^{\mu l}\xi^{\nu m} \tanh^2(\beta \mathbf{m} \cdot \boldsymbol{\xi}) \rangle\rangle \tag{12}$$

Obviously, this model does not admit Mattis-like solutions, but new kinds of solutions appear which allow recognition. The system presents now three different regimes instead of two, as occurs in the $\kappa = 0$ case (Fig. 1):

(a) The *paramagnetic regime*: above $T_1 = (1 + \kappa)$ the only solution is the state $\mathbf{m} = 0$ with

$$f(\beta) = -T \ln 2$$

which we will call paramagnetic in analogy with the magnetic case. The matrix A has two eigenvalues

$$\lambda_1 = 1 - \beta(1 - \kappa)$$
$$\lambda_2 = 1 - \beta(1 + \kappa)$$

each one with degeneracy r. This solution is then stable for $T > T_1$, at which λ_2 becomes negative. In this regime the system is not able to recognize the stored patterns.

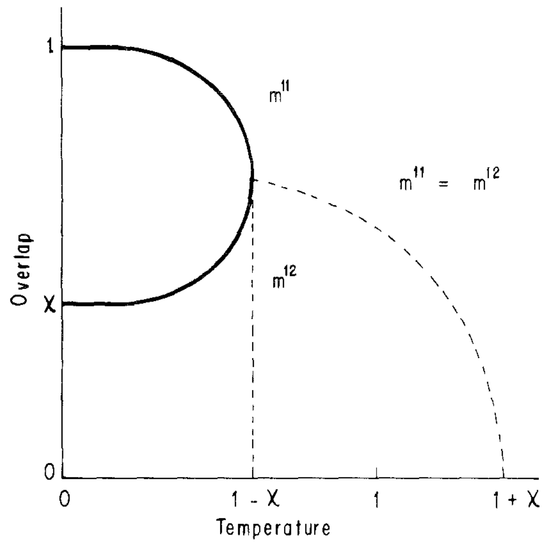

Fig. 1. The nonzero overlaps $m^{\nu 1}$ and $m^{\nu 2}$ vs. temperature for the retrieval solutions. The dashed lines correspond to the pair retrieval regime ($m^{\nu 1} = m^{\nu 2}$) and the continuous lines to the pattern retrieval regime ($m^{\nu 1} > m^{\nu 2}$).

(b) The *pair retrieval regime*: at T_1, $2r$ solutions emerge with the form

$$\mathbf{m} = (0, 0, ..., m^{\nu 1}, m^{\nu 2}, 0, ..., 0) \tag{13a}$$

$$m^{\nu 1} = m^{\nu 2} \tag{13b}$$

Averaging Eq. (10b), one obtains that

$$m = m^{\nu 1} = m^{\nu 2} = \frac{1+\kappa}{2} \tanh(\beta 2m) \tag{14}$$

The matrix A has, in this case four eigenvalues:

$$\lambda_1 = 1 - \beta(1+\kappa)[1 - \tanh^2(2\beta m)] = 1 - \left.\frac{dm}{dm'}\right|_{m=m^*} \tag{15a}$$

$$\lambda_2 = 1 - \beta(1-\kappa) \tag{15b}$$

$$\lambda_3 = 1 - \beta(1+\kappa)\left[1 - \frac{1+\kappa}{2}\tanh^2(2\beta m)\right] \tag{15c}$$

$$\lambda_4 = 1 - \beta(1-\kappa)\left[1 - \frac{1+\kappa}{2}\tanh^2(2\beta m)\right] \tag{15d}$$

where m^* is a stable solution of Eq. (14). As is known, for $T < T_1$, Eq. (14) has two nonzero stable solutions, and at these values the derivative is less than one, following that $\lambda_1 > 0$ for $T < T_1$. From Eq. (15b) it follows that $\lambda_2 > 0$ only for $T > T_2 = 1 - \kappa$. We calculated numerically λ_3 and λ_4 and verified that they are positive for $T_2 < T < T_1$, showing that these solutions are stable in this range of temperatures. In this regime the system is then able to recognize pairs, but it is not able to distinguish between its two patterns.

(c) The *pattern retrieval regime*: at T_2, where the pair retrieval solutions become unstable, a new kind of solution appears allowing pattern recognition and with the form

$$\mathbf{m} = (0, 0, \dots, m^{\nu 1}, m^{\nu 2}, 0, \dots, 0) \tag{16a}$$

$$m^{\nu 1} \neq m^{\nu 2} \tag{16b}$$

After averaging Eq. (10b), we obtain

$$m^{\nu 1} = \frac{1+\kappa}{2} \tanh[\beta(m^{\nu 1} + m^{\nu 2})] + \frac{1-\kappa}{2} \tanh[\beta(m^{\nu 1} - m^{\nu 2})] \tag{17a}$$

$$m^{\nu 2} = \frac{1+\kappa}{2} \tanh[\beta(m^{\nu 1} + m^{\nu 2})] - \frac{1-\kappa}{2} \tanh[\beta(m^{\nu 1} - m^{\nu 2})] \tag{17b}$$

Using the fact that $\tanh(x) \to \text{sign}(x)$ as $x \to \infty$, for $T \to 0$, Eqs. (17a), (17b), and (10a) tend to

$$m^{\nu 1} = 1 \tag{18a}$$

$$m^{\nu 2} = \kappa \tag{18b}$$

$$f(T=0) = -\tfrac{1}{2}(1 + \kappa^2) \tag{18c}$$

The matrix A now has four eigenvalues:

$$\lambda_1 = 1 - \beta(1+\kappa)\{1 - \tanh^2[\beta(m^{\nu 1} + m^{\nu 2})]\} \tag{19a}$$

$$\lambda_2 = 1 - \beta(1-\kappa)\{1 - \tanh^2[\beta(m^{\nu 1} + m^{\nu 2})]\} \tag{19b}$$

$$\lambda_3 = 1 - \beta(1+\kappa)(1 + q^{1111}) \tag{19c}$$

$$\lambda_4 = 1 - \beta(1-\kappa)(1 - q^{1111}) \tag{19d}$$

Let us define $u = (m^{\nu 1} + m^{\nu 2})$ and $v = (m^{\nu 1} - m^{\nu 2})$. Then

$$u = (1+\kappa)\tanh(\beta u) \tag{20a}$$

$$v = (1-\kappa)\tanh(\beta v) \tag{20b}$$

We can rewrite Eqs. (19a) and (19b) as

$$\lambda_1 = 1 - \left.\frac{du}{du'}\right|_{u=u^*} \tag{21a}$$

$$\lambda_2 = 1 - \left.\frac{dv}{dv'}\right|_{v=v^*} \tag{21b}$$

where u^* and v^* are solutions of Eqs. (20a) and (20b), respectively. Then we obtain $\lambda_1 > 0$ for $T < T_1$ and $\lambda_2 > 0$ for $T < T_2$. We analyzed λ_3 and λ_4 numerically and found that they are positive for $T < T_2$, concluding that these solutions are stable in the whole range between $T = 0$ and $T_2 = 1 - \kappa$.

There is another set of solutions with the general form

$$\mathbf{m} = m_{n,k}(\overbrace{1, 1,..., 1}^{k \text{ pairs}}, \overbrace{-1,...}^{n-k \text{ pairs}}, \overbrace{0,..., 0}^{r-n \text{ pairs}})$$

There are $2^n\binom{r}{k}\binom{r-k}{n-k}$ equivalent solutions for fixed k, and they appear at

$$T_{nk} = \left[1 + \left(\frac{2k}{n} - 1\right)\kappa\right]$$

with

$$f_{nk}(\beta) \simeq -\frac{3n^2 T_{nk}}{\Delta}(T_{nk} - T)^2$$

where

$$\Delta = \langle\langle (Z_k + \bar{Z}_{n-k})^4 \rangle\rangle$$

$$Z_k = \sum_{\nu=1}^{k} (\xi^{\nu 1} + \xi^{\nu 2})$$

$$\bar{Z}_{n-k} = \sum_{\nu=k+1}^{n} (\xi^{\nu 1} - \xi^{\nu 2})$$

Note that $T_2 \leqslant T_{nk} \leqslant T_1$. All the solutions with $n = k$ appear at $T_1 = 1 + \kappa$, and in particular, the pair retrieval solutions of the previous subsection correspond to the case $n = k = 1$, which also have the lowest energy among all of them. That means the system undergoes a second-order phase transition from the paramagnetic regime to the pair retrieval regime at T_1. These solutions correspond, in the limit $\kappa \to 0$, to the symmetric solutions with an even number of nonzero components in ref. 3. It can be shown that these solutions are unstable near and at $T = 0$, and that those solutions with $n = k$ are unstable in the whole interval.

We found other solutions, but their energies at $T = 0$ are always higher then the energy of the retrieval solutions. Among these, the solutions with the general form

$$\mathbf{m} = (m_1, m_1, m_1, m_2, 0,..., 0)$$

have the lowest energy at $T = 0$, where

$$m_1 = \frac{1}{2}\left(1 + \frac{\kappa}{3}\right), \qquad m_2 = \frac{\kappa}{2}(1 - \kappa)$$

These solutions correspond, in the limit $\kappa \to 0$, to the symmetric solutions with three nonzero components given in ref. 2. All this means that, although we could not prove that the ground state is associated with the solutions (17), we did not find any solution with free energy less than (18c). Nevertheless, we know that they are the ground states at least near $\kappa = 0$ and $\kappa = 1$, the two limits in which we recover the statements of ref. 2.

Although we have focused our analysis on a simple case, the results show how the performance of the model is modified when correlation between patterns is introduced. The major feature of the model is the existence of three different regimes: a low-temperature retrieval one in which the system recognizes the stored patterns, an intermediate-temperature pair retrieval regime (that disappears when the correlation $\kappa \to 0$) in which the system recognizes the pairs but is not able to distinguish between the patterns, and the paramagnetic regime at high temperature, in which the system does not provide associative memory. Another interesting feature of the model is the splitting of the temperatures at which the symmetric solutions appear.

ACKNOWLEDGMENTS

We are deeply indebted to C. Tsallis and S. Coutinho for useful discussions. One of us (F. A. T.) thanks CNPq (Brazil) for financial support.

REFERENCES

1. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**:2554 (1982).
2. D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Phys. Rev. A* **35**:2293 (1987).
3. D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Phys. Rev. A* **32**:1007 (1985).
4. D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Ann. Phys.* **173**:30 (1987).
5. N. Parga and J. Virasoro, *J. Phys.* **47**:1857 (1986).
6. H. Gutfreund, *Phys. Rev. A* **37**:570 (1988).
7. L. Personnaz, I. Guyon, and G. Dreyfus, *J. Phys. Lett.* **46**:L359 (1985).
8. M. V. Feigelman and L. B. Ioffe, *Int. J. Mod. Phys. B* **1**:51 (1987).
9. S. Bacci, J. Alfaro, C. Wiecko, and N. Parga, preprint.
10. B. Derrida, E. Gardner, and A. Zippelius, *Europhys. Lett.* **4**:167 (1987).
11. J. F. Fontanari and R. Köberle, *J. Phys. A* **21**:2477 (1988).