Revista Mexicana de Física **36** *No. 3(1990) 415–423*

Spin glass model for a neural network: configuration space studies

C. Martínez, E. Cota and L. Viana
Laboratorio de Ensenada, Instituto de Física
Universidad Nacional Autónoma de México
Apartado postal 2681, 22800 Ensenada B.C., México.
(Recibido el 10 de octubre de 1989; aceptado el 17 de mayo de 1990)

Abstract. We consider a long range Ising Neural Network, where p unbiased patterns have been stored according to Hebb's rule and a a modified version designed to reflect training. We perform $T = 0$ MonteCarlo simulations in order to explore and compare the configuration space in both cases, as a function of the ratio of the number of stored patterns to the size of the system.

PACS: 87.30.G; 64.60.C: 75.10.H

1. Introduction

In the last few years, there has been a growing interest in studying Neural Networks (NN) due to their features as associative fault-tolerant memories. These studies have a double interest: on one side, to understand the mechanisms responsible for the storage and retrieval of information in biological systems, and on the other, the design of so called "Neural Computers" with learning and generalizing abilities. An exposition of the basic ideas from the biological and computational points of view has been given by Viana [1].

Physicists became interested in this field after the pioneering work of Hopfield [2], who made a mathematical analogy between an asembly of neurons and certain disordered magnetic materials called Spin Glasses (SG) [3]. This connection allowed the use of the methods developed in Statistical Mechanics to study some general properties of NN in the thermodynamic limit.

Spin Glass-like models for NN conceive neurons as two-state elements which interact with other neurons via their synapsis $\{J_{ij}\}$, and whose dynamics, in the absence of noise, is governed by an energy function given by

$$H = -\tfrac{1}{2} \sum J_{ij} S_i S_j, \tag{1}$$

where S_i denotes the state of the i-th neuron with $S_i = +1$ (-1) for a firing (quiescent) neuron; $J_{ij} = J_{ji}$ represents the "synaptic strength" and takes account of the interaction between neurons i and j, by taking a positive (negative) value for an excitatory (inhibitory) interaction. Due to the mixture of excitatory and inhibitory interactions, the global energy function possesses a high number of local minima.

The retrieval capabilities of these NN are a natural consequence of their dynamics, as they evolve spontaneously towards a minimum of the free energy which has a large 'overlap' [4] with the initial state. Therefore the final (or equilibrium) state of the network will depend on both the initial state and the specific set $\{J_{ij}\}$. In this way, learning is related to a suitable modification of the connection coefficients $\{J_{ij}\}$ which favours particular configurations by making them minima of the hamiltonian. In biological systems it has been found that learning indeed changes the values of the synaptic strengths [5].

One of the central problems in this field is to find a good prescription to determine the value for the set $\{J_{ij}\}$, such that the network presents the desired behaviour. It is not easy (if at all possible) to find an analytic "recipe" to construct an energy function whose only minima are those related to the states we intend to store ("pure memories"); for it has been found that in this process other minima appear, which have a negative influence on the capabilities of the system to retrieve stored information. These "spurious" minima are related to mixtures of "pure" memories and their number grows very rapidly with p, the number of stored patterns. However, more relevant than the number of spurious minima is the total percentage of configuration space occupied by their domains of attraction, as this parameter will affect the retrieval capabilities of the network.

Many different "learning rules" have been proposed, among them, the simplest and best studied is the Hebb rule. This learning prescription assumes that synaptic strengths change in response to experience in a way proportional to the correlation between the firing of the pre and post synaptic neurons. That is, J_{ij} is given by the sum of p random patterns $\{\xi^\mu\}$, according to

$$J_{ij} = \frac{1}{N} \sum_{\mu} \xi_i^\mu \xi_j^\mu \tag{2}$$

for $i \neq j$ and $J_{ii} = 0$; where ξ_i^μ with $\mu = 1, \ldots, p$ are "quenched" variables which correspond to p stored patterns and can take the values ± 1, with equal probability. This case has been studied analytically in the thermodynamic limit $N \to \infty$, for both p finite and infinite [6,7], where N is the number of elements in the system. This was done as a function of a load parameter defined as $\alpha = p/N$, by taking into account the cases $\alpha = 0$ and $\alpha \neq 0$, respectively. The results obtained are summarized below:

For $\alpha = 0$, it was found [6] that , for $0 < T < 1$, the energy ground states are all related to pure memories; where $T = 1$ is the critical level of noise below which the ordered states appear (T is equivalent to temperature in the case of SG). For $0.46 \lesssim T < 1$ these "pure" or "Mattis states" are the only minima; however, as the noise level decreases, spurious stable states show up. These spurious minima correspond to a mixture of $2s + 1$ pure memories [$s = 1, 2, \ldots, (p-1)/2$], and can be classified as symmetric and asymmetric states. The symmetric (asymmetric) spurious states are minima equidistant (non equidistant), according to the Hamming distance [8], to the pure patterns they are related with. As T is lowered, the number of such states increases until, for $T \to 0$, it grows exponentially with p, and $f_1 < f_3 < \ldots < f_{2s+1}$, where f_{2s+1} accounts for the free energy of any state composed by the symmetric mixture of $2s+1$ pure memories. We intuitively expect that states higher in free energy have smaller basins of attraction than those of lower energy. However, to the authors' knowledge, the percentage of the configuration space occupied by their basins of attraction has not been evaluated.

For a finite ratio α, the situation is as follows [7]: As α grows, the sum of the random overlaps among patterns becomes an important contribution to the energy function, until it eventually unstabilizes the stored patterns and there is a catastrophic and discontinuous breakdown for $\alpha \geq 0.138$. Beyond this point, retrieval of stored patterns is no longer possible.

A modified Hebb rule in which it is possible to reflect some degree of training has also been studied analytically [9]. According to these prescription the synaptic strengths are given by

$$J_{ij} = \frac{1}{N} \sum_{\mu} J_\mu \xi_i^\mu \xi_j^\mu \tag{3}$$

where strictly different weights J_μ were assigned to each of a finite number of stored patterns ($1 = J_1 > J_2 > \ldots > J_p > 0$). Due to the lack of symmetry between the stored patterns, this model presents the following properties: For $T \to 0$ the energy of the pure state related to J_μ is given by $f_{J\mu} = -.5J_\mu$. Therefore it is no longer true that all the minima related to "pure memories" are global minima of the energy. Moreover, some spurious states can have energies higher than those of pure states. Another consequence is that symmetric spurious states do not exist. Instead, all solutions involve an unequal overlap of an odd number of memories, whose relative value depends on the noise level T.

Although these analytical studies give us important information about the existence of spurious minima, they fail to provide any information about the relative size of the different basins of attraction; this is due to the nonergodicity of NN. Therefore, if we are interested in obtaining a more complete picture of the configuration space, it is necessary to use other complementary techniques such as computer simulations.

In this paper we carry out $T = 0$ Monte Carlo calculations for a long range Ising Neural Network, composed by N neuron-like elements, whose dynamics is exactly obeyed by Eq. (1). That is, we consider a network with no noise, and synaptic interactions given in turn by Hebb's rule [Eq. (2)] and the modified version [Eq. (3)]. We compare their configuration spaces in order to understand the role played by spurious memories in both cases.

2. Procedure

We analyze numerically the relative importance of pure and spurious memories for both learning prescriptions, through the relative size of their domains of attraction. This is given simply by the relative number of times the system retrieves such states starting from configurations generated at random. We perform this analysis as a function of $\alpha = p/N$. In the thermodynamic limit, the case for finite p would correspond to the limit $\alpha \to 0$.

We are interested in obtaining results valid for an ensemble of systems, each consisting of N neuron-like components, where a finite number p of random unbiased patterns have been learnt. Therefore, strictly speaking it would be necessary to take ensemble averages over such systems. However, since this would entail the use of huge amounts of computer time, we chose to perform the simulations in a "representative" set with p embedded random patterns with zero overlap, that is subjected to a global restriction given by:

$$q_{\mu\tau} = \frac{1}{N} \sum_i \xi_i^\mu \xi_i^\tau = \delta_{\mu\tau} \quad \text{for all } \mu\tau \tag{4}$$

This choice eliminates some of the effects produced by the finite size of the sample [10], *i.e.*, it approximates a feature present in an infinite NN with $\alpha = 0$, where the overlaps among the stored patterns are of order zero.

In order to test how good this approximation is, we stored information according to the Hebb rule [Eq. (2)], in 10 different samples (with no restrictions), for the values $p = 3$ and $N = 192$. We performed $T = 0$ Monte Carlo simulations and obtained an average of the relative size of their basins of attraction (Table 1). Subsequently, we made the same simulations on a sample constructed by following the same learning prescription, with the same size and load information, but with stored patterns which comply with the restriction given by Eq. (4) (Table 2). Comparison of results shown in Tables 1 and 2, indicate that the use of this "representative" set gives us results which approximate those obtained by taking averages over samples with random overlaps, thus confirming this approximation as a suitable one.

Once we accepted the use of this "representative" set, the procedure was as follows: We chose N, the total number of spins in the system and generated p configurations in such a way that there were no correlations among them. Next, we stored those patterns by fixing the set of values $\{J_{ij}\}$ according to either the original or modified versions of the Hebb prescription [Eqs. (2) and (3), respectively]. Afterwards, we proceeded to explore the configuration space: We generated a configuration at random updating the spins one at a time, by aligning them along their internal fields. We did so, in random order, until the network settled into a stable configuration for which $S_i(t+1) = S_i(t)$ for all i. We repeated this process 3000 times and calculated the percentage of times each of the stored patterns was retrieved. At the same time, the percentage of times a spurious memory was obtained was evaluated. The same procedure was followed for values of N ranging from 32 to 512 and p from 3 to 7.

	Basins of attraction size (% in configuration space)										
Pattern Number	1	2	3	4	5	6	7	8	9	10	$\langle\ \rangle$
1	26.50	25.90	25.36	26.76	26.46	27.36	25.60	27.03	25.26	25.30	26.15
2	26.93	27.33	26.13	26.66	25.73	26.23	26.86	25.60	26.23	24.83	26.25
3	26.23	25.16	26.30	27.36	28.36	25.83	27.30	24.23	26.86	27.73	26.54
$\langle\%\rangle$ pure $= 26.31\%$ $\sigma = .7565$											

TABLE I. Results obtained in 10 different samples with embedded patterns with random overlaps ($N = 192, p = 3$). For a description see text.

Pattern Number	Basins of attraction size (% in configuration space)
1	26.43
2	26.77
3	25.70
$\langle\%\rangle$ pure $= 26.30\%$ $\sigma = .445\%$	

TABLE II. Results obtained in one sample with three stored patterns without overlaps ($N = 192$). For a description see text.

3. Results

When interpreting our results it is important to keep in mind the following considerations: we studied the behaviour of this network as a function of α, and these results could, in principle, be extrapolated to the thermodynamic limit in both cases $\alpha = 0$ (p finite), and $\alpha \neq 0$ ($p \to \infty$). However, we expect our results to be better suited to the first case since we considered nonoverlapping patterns in our problem. For finite α, the sum of the random overlaps among patterns, which we are neglecting, becomes an important contribution to the energy function. This contribution is reflected in the discontinuous brakedown in the storage capabilities for $\alpha > 0.138$, as opposed to the gradual reduction on retrieval capabilities of a NN with orthogonal stored patterns, which is given through the continuous decrease in the size of the basins of attraction as α increases.

Figure 1 shows a typical run consisting of 3000 trials on a sample with $N = 320$ elements and $p = 5$ stored patterns, for Hebb (1.a), and modified Hebb (1.b) with $\langle J_\mu \rangle = 0.7$. The abscissas show, with open symbols, the relative number of times each "pure memory" was retrieved, and full circles correspond to the percentage of times any of the "spurious memories" was recalled. We can see from figure (1.b) that, as expected, the higher the value of J_μ the better the retrievability of the corresponding memory. It can also be appreciated that the relative importance of spurious memories decreases in case (b) with respect to case (a). In other words, if one introduces "training" into the model via the modified Hebb rule, one can expect overall improvement in the retrieval capabilities of the network, since the importance of some pure memories over others can be stressed and the percentage

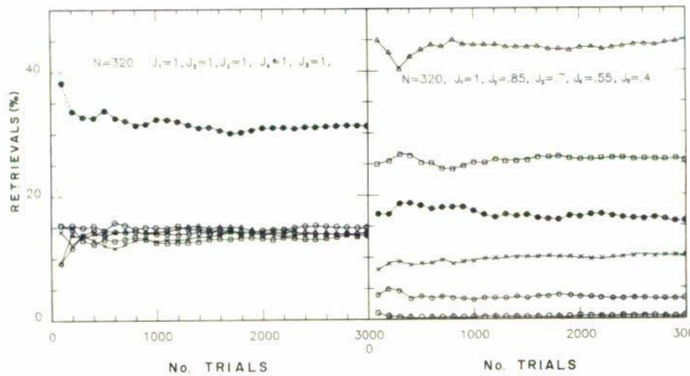

FIGURE 1. Typical run corresponding to 3000 trials (random initial configurations) with $N = 320$ and $p = 5$, showing the relative number of times each pure memory was retrieved (open symbols). Full circles correspond to spurious memories. (a) Equal weights ($J_\mu = 1$); (b) different weights J_μ, such that $\langle J_\mu \rangle = 0.7$. In both cases, the symbols used are Δ for J_1, $\square$ for J_2, $\times$ for J_3, an hexagon for J_4 and $\circ$ for J_5.

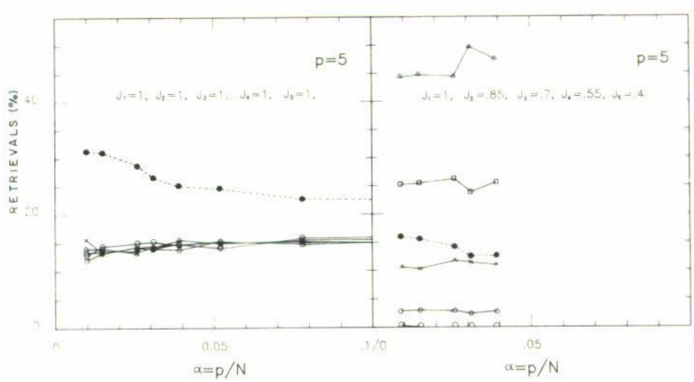

FIGURE 2. Percentage of times each memory was retrieved as a function of $\alpha = p/N$ for $p = 5$ and values of N ranging from 32 to 512. Full circles correspond to spurious states. (a) Equal weights ($J_\mu = 1$); (b) different weights J_μ, such that $\langle J_\mu \rangle = 0.7$

of configuration space occupied by spurious memories is significantly diminished. Similar graphs for other values of N and p can be found elsewhere [11].

Figure 2 is also related to the case $p = 5$ but involves values of N ranging from 32 to 512. This figure shows, as a function of $\alpha = p/N$, the percentage of times

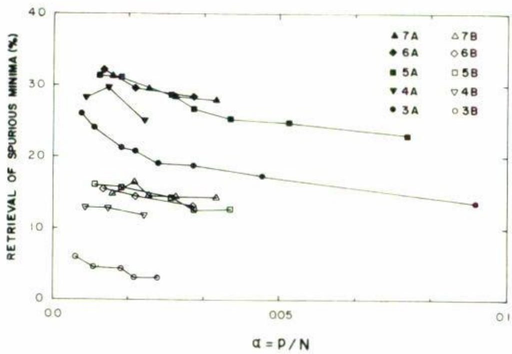

FIGURE 3. Percentage retrieval of spurious memories as a function of α for p from 3 to 7, for both equal and different weights. These two groups of cases are indicated as (pA) and (pB) respectively.

each memory was retrieved, after 3000 runs; full circles corresponding to spurious states. The cases with equal weights ($J_\mu = 1$), and different weights J_μ such that $\langle J_\mu \rangle = 0.7$, are shown in parts (a) and (b), respectively. Again we observe, for the range of values of α considered, that the percentage of configuration space occupied by spurious memories decreases in case (b) with respect to (a). In fact, we may extrapolate from the figure to the $\alpha = 0$ limit to obtain approximate values of 32% and 16% for cases (a) and (b), respectively, for this value of p.

This remark is more clearly seen in Fig. 3, where the percentage retrieval of spurious memories is shown as a function of α, for both equal and different weights. These two groups of cases are indicated as (pA) and (pB) respectively, for $p = 3$, 4, 5, 6 and 7. Here, it can be seen that the total "area" of the configuration space occupied by the domains of attraction of spurious memories, is considerably higher if all memories are stored with equal weight. Also, notice that as p increases, the percentage of configuration space occupied by spurious memories tends to a limit already for $p = 7$, for both Hebb and modified Hebb rules. By extrapolating to the $\alpha = 0$ limit, we obtain results similar to the values mentioned in the last paragraph.

From figures 1b and 2b, it can be observed that, for the modified Hebb's rule, memories with larger weights have a higher percentage of retrieval. This means that by varying the specific values of J_μ it is possible to modulate the percentage of times each memory is retrieved, *i.e.*, it is possible to simulate training. This is shown explicitly in figure 4, for the case $p = 7$, $N = 192$ and $\langle J_\mu \rangle = 0.7$.

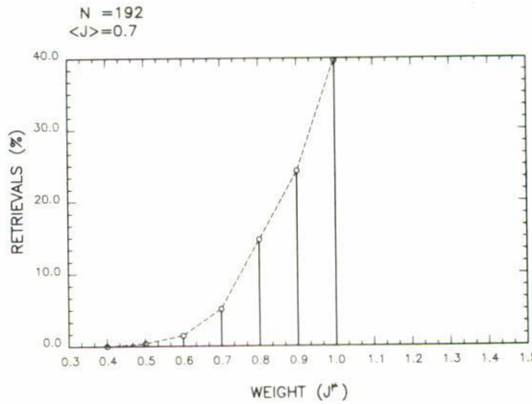

FIGURE 4. Typical graph showing the percentage retrieval of each memory as a function of its weight J_μ for the case: $p = 7$, $N = 192$ and $\langle J_\mu \rangle = 0.7$

4. Conclusions

We have carried out a comparative numerical study of the retrieval capabilities of a NN using Hebb's learning rule and a modified Hebb's rule where different weights are assigned to the stored patterns. Using a "representative" set with non-overlapping stored patterns to reduce size effects, we performed $T = 0$ Monte Carlo simulations to evaluate the 'area' of configuration space occupied by the basins of attraction of spurious memories, as a function of the loading parameter α. We find from our results that by varying the specific values of the weights J_μ in the modified Hebb's rule it is indeed possible to simulate training. In addition, we consistently find that using the modified Hebb's prescription reduces significantly the percentage of configuration space occupied by spurious memories, which translates into an improvement in the retrieval capabilities of the network. Further studies, concerned mainly with the evaluation of more realistic learning rules (from the biological point of view) including the consideration of asymmetry in the synaptic strenghts and dilution are currently underway.

References

1. L Viana, *Memoria Natural y Artificial*, "La Ciencia desde México", Fondo de Cultura Económica No. 88 (1990).
2. J.J. Hopfield *Proc. Natl. Acad. Sci. USA* **79** (1982) 2554.
3. For recent reviews on Spin Glasses see K. Binder, A.P. Young, *Rev. Mod. Phys.* **58** (1987) 801, and M. Mezard, G. Parisi and M.A. Virasoro *Spin Glass Theory and Beyond*, Lecture Notes in Physics No. 9, World Scientific (1987).
4. W. Kinzel, *Z. Phys.* **B60** (1985) 205.

5. E.R. Kandel, 'Small systems of neurons', in *The Brain* (Scientific American Book), W.H. Freeman and Co., San Francisco 1979.
6. D.J. Amit, H. Gutfreund and H. Sompolinsky, *Phys. Rev.* **A32** (1985) 1007.
7. D.J. Amit, H. Gutfreund and H. Sompolinsky, *Phys. Rev. Lett.* **55** (1985) 1530.
8. J.A. Hertz and R.G. Palmer; Duke University lecture notes to be published (1989), Secc. 2.2.1
9. L. Viana, *J. Phys. (France)* **49** (1988) 167.
10. However some other size effects are not removed, since our sample still has $\sum_{i,\mu}\xi_i^\mu \neq 0$.
11. C. Martínez, B.Sc. Thesis, Faculty of Science UNAM 1989.

Resumen. Consideramos una red neuronal de Ising de largo alcance en la cual han sido almacenados p patrones de acuerdo con la regla de aprendizaje de Hebb y con una modificación de ésta diseñada para reflejar entrenamiento. Llevamos a cabo simulaciones de Monte Carlo a $T = 0$, con objeto de explorar y comparar el espacio de configuraciones de ambos modelos, como función de la relación entre el número de patrones almacenados y el tamaño del sistema.