International Journal of Neural Systems, Vol. 3 (Supp. 1992) 71–77
Proceedings of the Neural Networks: From Biology to High Energy Physics

MEMORY RETRIEVAL IN OPTIMAL SUBSPACES

G. Boffetta
Dip. di Fisica Generale e INFN, Universitá di Torino, Via P.Giuria 1
10125 Torino, Italy

R. Monasson
Laboratoire de Physique Théorique, E.N.S., 24 rue Lhomond
75231 Paris Cedex 05, France

and

R. Zecchina
Dip. di Fisica Teorica e INFN, Universitá di Torino, Via P.Giuria 1
10125 Torino, Italy
Dip. di Fisica, Politecnico di Torino, C.so Duca degli Abruzzi 24
10129 Torino, Italy

A simple dynamical scheme for Attractor Neural Networks with non–monotonic three state effective neurons is discussed. For the unsupervised Hebb learning rule, we give some basic numerical results which are interpreted in terms of a combinatorial task realized by the dynamical process (dynamical selection of optimal subspaces). An analytical estimate of optimal performance is given by resorting to two different simplified versions of the model. We show that replica symmetry breaking is required since the replica symmetric solutions are unstable.

1. Introduction

The definition of relaxation dynamics is the central problem for the study of the associative and computational capabilities in models of attractor neural networks. For spin models past effort was concentrated in studying a variety of synaptic structures or learning algorithms. More recently, biological models[1] try to give analytical control to neural networks which present realistic dynamical features. The model[2] described in this paper can be placed midway between spin models and biological models, and is far from being an attempt to propose a realistic picture of neural networks. More simply, we give a brief discussion concerning the possible computational and physical role played by local inhibitory interactions which lead to an effective non–monotonic transfer function for the neurons. In the last few years others models characterized by non-monotonic neurons have been proposed.[3–5]

For the Hebb learning rule we show, numerically, that the critical capacity increases with respect to the Hopfield case and that such result can be interpreted in terms of a twofold task realized by the dynamical process. By means of local inhibition, the system dynamically selects a subspace (or subnetwork) of minimal static noise with respect to the recalled pattern; at the same time, and in the selected subspace, the retrieval of the memorized pattern is performed. The dynamic behaviour of the network, for deterministic sequential updating, range from fixed points to chaotic evolution, with the storage ratio as control parameter, the transition appearing in correspondence to the collapse of the associative performance.

Resorting to two simplified versions of the model, we study their optimal performance by the replica method; in particular the role of non–monotonic functions and of subspaces dynamical selection are discussed.

Section **2** of the paper is devoted to illustrate the numerical study while Sec. **3** contains the analytical results on optimal storage capacity.

2. The Model and its Performance

The attractor neural network under study is composed of N three state $\pm 1, 0$ formal neurons. During the learning process only the ± 1 values are considered (the patterns are binary) and the 0-state should be considered as a dynamical *don't care* state, not belonging to the patterns code. The system is fully connected and its evolution is governed by sequential updating of the following equations

$$S_i = \begin{cases} \mathrm{sgn}(h_i) & if \quad |h_i| \leq \gamma \\ 0 & if \quad |h_i| > \gamma \end{cases} \tag{1}$$

$$h_i = \sum_{j=1}^{N} J_{ij} S_j \qquad i = 1, \dots, N \ , \tag{2}$$

where γ is the threshold of the local inhibitory feedback, and the $\{J_{ij}\}$ are the couplings.

Defining the **retrieval activity** as the fraction of neurons which are not not in the zero state

$$a = \frac{1}{N} \sum_i S_i^2, \tag{3}$$

the parameter that defines the retrieval quality is the **scaled overlap**

$$m^\mu = \frac{1}{Na} \sum_i \xi_i^\mu S_i \ , \tag{4}$$

where the $\{\xi_i^\mu = \pm 1,\ i = 1, N; \mu = 1, P\}$ are the memorized binary patterns. The scaled overlap can be thought simply as the overlap computed in the subspace M of the *active* neurons, $M \equiv \{i \ / \ S_i \neq 0,\ i = 1, N\}$.

Given a set of P random independent binary patterns $\{\xi_i^\mu\}$, the Hebb-Hopfield learning rule corresponds to fix the synaptic matrix J_{ij} by the additive relation $J_{ij} = \frac{1}{N} \sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu$ (with $J_{ii} = 0$).

The effect of the dynamical process defined by (1) and (2) is the selection of subspaces M of active neurons in which the static noise is minimized (such subspaces will be hereafter referred to as *orthogonal* subspaces). Before entering in the description of the results, it is worthwhile to remember that, in Hopfield-like attractor neural networks, the mean of cross correlation fluctuations produce in the local fields of the neurons a static noise, referred to as cross-talk of the memories. Together with temporal correlations, the static noise is responsible of the phase transition of the neural networks from associative memory to spin-glass. More precisely, when the Hopfield model is in a fixed point ξ^σ which belongs to the set of memories, the local fields are given by $h_i \xi_i^\sigma = 1 + R_i^\sigma$ where $R_i^\sigma = \frac{1}{N} \sum_{\mu \neq \sigma} \sum_{j \neq i} \xi_i^\mu \xi_j^\mu \xi_i^\sigma \xi_j^\sigma$ is the static noise (Gaussian distribution with 0 mean and variance $\sqrt{\alpha}$). The well known critical value of $\alpha = P/N$ at which the phase transition takes place is 0.14, as computed analytically by the replica method[6] and confirmed numerically. If the patterns were exactly orthogonal the fluctuations of the noise term R_i^σ would be zero and the critical capacity one.

The preliminary performance study of the model under discussion have revealed several new basic features, in particular:

- the critical capacity, for the Hebb learning rule, results increased up to $\alpha_c \approx 0.28$;
- the mean cross correlation fluctuations computed in the selected subspaces is minimized by the dynamical process in the region $\alpha < \alpha_c$;
- in correspondence to the associative transition the system goes through a dynamic transition from fixed points to chaotic trajectories.

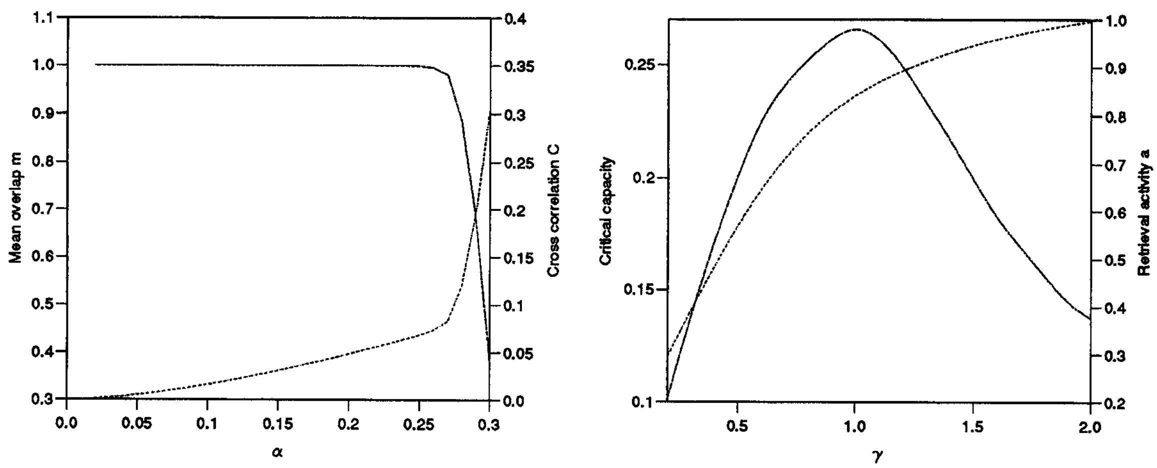


Fig.1-2. Mean overlap m (solid line) and cross correlation C (dashed line).
Critical capacity α_c (solid line) and retrieval activity a (dashed line) vs. threshold parameter γ.

The quantitative results concerning associative performance, are obtained by means of extended simulations. A typical simulation takes the memorized patterns as initial configurations and lets the system relax until it reaches a stationary point. The quantity describing the performance of the network as an associative memory is the mean scaled overlap m between the final stationary states and the memorized patterns, used as initial states. As the number of memorized configurations grows, one observes a threshold at $\alpha = \alpha_c \approx 0.28$ beyond which the stored states become unstable. The numerical results are reported in Fig.1 for a network of size $N = 1000$. We observe that since the recall of the patterns is performed with no errors (up to $\alpha \approx 0.25$), also the number of stored bits in the synaptic matrix results increased with respect to the Hopfield case (the fractional size of subspaces ($a \approx 0.8$) times the capacity is greater then 0.14).

As it is shown in Fig. 2 the size of subspaces and the network capacity are controlled by the threshold parameter γ. It is straightforward to notice that an optimal value $\gamma_{\text{opt}} \approx 1.0$ exists and that, for $\gamma \to \infty$, the Hopfield result ($\alpha_c = 0.14$) is recovered.

In order to show the static noise reduction, in Fig.1 is also reported the mean fluctuation of the cross correlations (or static noise) in the subspaces:

$$C = \frac{1}{P}\sum_{\sigma}\sum_{\mu\neq\sigma}\left(\frac{1}{aN}\sum_{i=1}^{N}\xi_i^\mu\xi_i^\sigma\epsilon_i^\sigma\right)^2 = \frac{1}{P}\sum_{\sigma}\sum_{\mu\neq\sigma}\left(\frac{1}{aN}\sum_{i\in M}\xi_i^\mu\xi_i^\sigma\right)^2 , \tag{5}$$

where $\epsilon_i^\sigma = 1$ if $i \in M$ in pattern σ and zero otherwise, as a function of α. It is clear from the figure that C does not follow a statistical law but undergoes a minimization that qualitatively explains the increase in the storage capacity. For $\alpha < \alpha_c$, once the system has relaxed in a stationary subspace, the model becomes equivalent (in the subspace) to a Hopfield network with a static noise term which is no longer random. The statistical mechanics of the combinatorial task of minimizing the *noise-energy* term (5) can be studied analytically by the replica method; the results are of general interest in that give an upper bound to the performance of networks endowed with Hebb-like synaptic matrices and with the possibility selecting optimal subnetworks for retrieval dynamics of the patterns.

As already stated, the behaviour of the neural network as a dynamical system is directly related to its performance as an associative memory. The system shows an abrupt transition in the dynamics, from fixed points to chaotic trajectories, in correspondence to the value of the storage ratio at which the memorized configurations become unstable. The only (external) control parameter of the model as a dynamical system is the storage ratio $\alpha = P/N$. Dynamic complex behaviour appears as a clear signal of saturation of the attractor neural network and does not depend on the symmetry of the couplings.

As a concluding remark of this section, we observe that the dynamic selection of subspaces seems to take advantage of finite size effects allowing the storage of correlated patterns also with the simple Hebb rule. Numerical tests on patterns generated according to the probability distribution $P(\xi) = \frac{(1-b)}{2}\delta(\xi - 1) + \frac{(1+b)}{2}\delta(\xi + 1)$, indicate $b = b_c \approx 0.1$ as a critical value for the bias: below such a value the critical capacity is retained constant (0.28) independently on the size N (tests have been done for $N = 200, 400, 800, 1000$). Analytical and numerical work is in progress on this point, devoted to clarify the performance with spatially correlated patterns[7].

3. Optimal Storage Capacity

It is of basic interest to understand whether a dynamical scheme which allows for dynamic selection of subnetworks provides a neural network model with enhanced optimal capacity with respect to the classical spin models. Assuming that nothing is known about the couplings, one can consider the J_{ij} as dynamical variables and study the fractional volume in the space of interactions that makes the patterns fixed points of the dynamics. Following Gardner and Derrida[8], we describe the problem in terms of a cost-energy function and study its statistical mechanics: for a generic choice of the $\{J_{ij}\}$, the cost function E_i is defined to be the number of patterns such that a given site i is wrong (with respect to (1))

$$E_i(\{J_{ij}\},\{\epsilon_i^\mu\}) = \sum_{\mu=1}^{P}\left[\epsilon_i^\mu(\Theta(h_i^\mu\xi_i^\mu + \gamma) - \Theta(h_i^\mu\xi_i^\mu)) + (1-\epsilon_i^\mu)\Theta(\gamma^2 - h_i^{\mu 2})\right] , \tag{6}$$

where Θ is the step function, the $h_i^\mu = \frac{1}{\sqrt{N}} \sum_j J_{ij}\xi_j^\mu \epsilon_j^\mu$ are the local fields, γ is the threshold of the inhibitory feedback and with $\epsilon_i^\mu = \{0,1\}$ being the variables that identify the subspace M ($\epsilon_i^\mu = 1$ if $i \in M$ and zero otherwise).

In order to estimate the optimal capacity, one should perform the replica theory on the following partition function

$$Z = \mathrm{Tr}_{\{\epsilon_i^\mu / \sum_i \epsilon_i^\mu = aN\}} \int \prod_{i \neq j} dJ_{ij} \delta\Big(\sum_{j(\neq i)} J_{ij}^2 - N \Big) \mathrm{e}^{-\beta \sum_i E_i} \; . \tag{7}$$

Since the latter task seems unmanageable, as a first step we resort to two simplified version of the model which, separately, retain its main characteristics (subspaces and non–monotonicity); in particular:

(i) we assume that the $\{\epsilon_i^\mu\}$ are quenched random variables, distributed according to $P(\epsilon_i^\mu) = (1-A)\delta(\epsilon_i^\mu) + A\delta(\epsilon_i^\mu - 1)$, $A \in [0,1]$;

(ii) we consider the case of a two-state (± 1) non–monotonic transfer function.

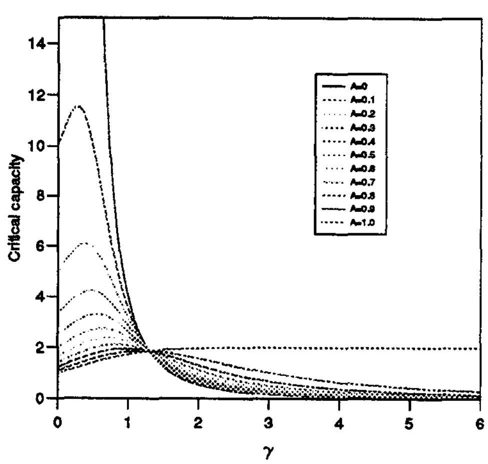
For lack of space, here we list only the final results for the critical storage capacity and for the stability analysis in the replica symmetric (R.S) assumption, and give the critical capacity value, for the second case, computed with one step in replica symmetry breaking.

The expressions of the R.S. critical capacity for the models are, respectively:

$$\alpha_c^{R.S.}(\gamma; A) = \left\{ 2(1-A)\int_0^\gamma D\zeta(\gamma-\zeta)^2 + \frac{A}{2} + A\int_\gamma^\infty D\zeta(\gamma-\zeta)^2 \right\}^{-1} \tag{8}$$

$$\alpha_c^{R.S.}(\gamma) = \left\{ \int_0^{\frac{\gamma}{2}} D\zeta\zeta^2 + \int_{\frac{\gamma}{2}}^\infty D\zeta(\gamma-\zeta)^2 \right\}^{-1} , \tag{9}$$

where $D\zeta = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\frac{-\zeta^2}{2}} d\zeta$ (for (9) see also Ref. 4). As Figs. 3 and 4 show, the values of critical capacity one finds are much higher than the monotonic perceptron capacity ($\alpha_c = 2$).



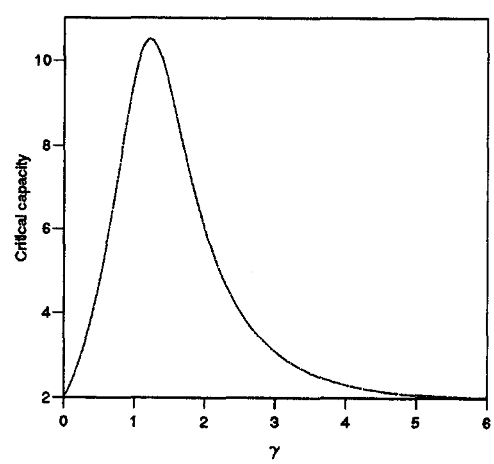


Fig.3. Critical capacity $\alpha_c^{R.S.}$ (8) vs. γ, for several A. Fig.4. Critical capacity $\alpha_c^{R.S.}$ (9) vs. γ.

To check the reliability of such results, we have analyzed the transverse stability of the symmetric saddle-point, finding the following stability conditions:

$$\alpha_c^{R.S.}\left[A(\frac{1}{2}+H(\gamma))+(1-A)(1-2H(\gamma))\right]<1 \tag{10}$$

$$\alpha_c^{R.S.}/2<1\,. \tag{11}$$

The first condition is satisfied only in a region of the (A,γ) plane where $\alpha_c < 2$, while the second is **never** satisfied since (9) gives $\alpha_c > 2$ for any γ. The R.S. solutions are thus unstable and replica symmetry breaking is required.

The necessity of breaking the replica symmetry is not surprising since for non-monotonic transfer functions the space of couplings is not connected, in analogy with multi-layered networks. Such analogy is even stronger if one realizes that two-state non-monotonic model is equivalent to a parity machine with three hidden units (with thresholds $-\gamma, 0, \gamma$) which have identical input synaptic vectors.

All the details concerning the computation with one step in replica symmetry breaking of the critical capacity and stabilities distribution can be found in Ref. 9. Here we just quote the final quantitative result concerning optimal capacity for the non-monotonic two-state model: numerical evaluation of the saddle-point equations (for unbiased patterns) gives $\alpha_c(\gamma^{\rm opt}) \approx 4.6$ with $\gamma^{\rm opt} \approx 1$, the corresponding R.S. value from (9) being $\alpha_c^{R.S.} \approx 10.5$. Since the number of allowed internal representation of the parity machine on which the non-monotonic model can be mapped is 4, the value of critical capacity should be compared with that of a two hidden units parity machine ($\alpha_c = 4$).

4. Conclusion

We have given a short discussion of some basic results concerning a dynamical scheme for neural networks which performs a combinatorial task together with associative recall. A more complete analysis of the computational capabilities and physical properties will be given in Refs. 7 and 9.

Acknowledgments

We would like to thank S. Solla for discussions.

References

1. D.J. Amit and M.V. Tsodyks, "Quantitative study of attractor neural network retrieving at low spikes rates: I and II," *Network* **2**, (1991).
2. R. Zecchina, "Computational and physical role of local inhibition in attractor neural networks: a Simple Model," in *Parallel Architectures and Neural Networks*, E. R. Caianiello (ed.) (World Scientific, 1992).
3. M. Morita, S. Yoshizawa and H. Nakano, "Analysis and improvement of the dynamics of autocorrelated associative memory," *IEICE Trans.* **J73-D-II**, 232 (1990).
4. K. Kobayashi, "On the capacity of a neuron with a non-monotone output function," *Network* **2**, 237 (1991).
5. S. Yoshizawa, M. Morita and S. Amari, "Capacity of associative memory using a non-monotonic neural model," (preprint 1992).

6. D.J. Amit, M. Gutfreund and H. Sompolinsky, "Storing infinite numbers of patterns in a spin-glass model of neural networks," *Phys. Rev. Lett.* **55**, 1530 (1985).
7. G. Boffetta, N. Brunel, R. Monasson and R. Zecchina (in preparation, 1992).
8. E. Gardner and B. Deridda, "Optimal storage properties of neural network models," *J. Phys.* **A21**, 271 (1988).
9. G. Boffetta, R. Monasson and R. Zecchina, "Symmetry breaking in non-monotonic neural networks," (in preparation, 1992).