CORRELATED ATTRACTORS FROM UNCORRELATED STIMULI

L. F. Cugliandolo
Dipartimento di Fisica, Università di Roma, *"La Sapienza"*,
I-00185 Roma, Italia
and
INFN Sezione di Roma

July 1993

Abstract

Analytic results for a neural network model proposed to account for neurophysiological experiments in which attractors for uncorrelated patterns are found to be correlated are presented. Exact expressions for the associative attractors and their correlations are described.

1 Introduction

Griniasty *et al.* [1] have proposed a neural network model that describes phenomena observed in single unit recordings in the anterior ventral temporal cortex of performing monkeys [2]. The synaptic matrix considers the memorized patterns as a (cyclic) sequence and couples consecutive patterns with a strengh a in addition to the auto-associative term. The network has correlated associative attractors though the learnt patterns are uncorrelated.

The notation in this note follows ref.[1]. The number of stored patterns p is finite. In order to solve the model without numerical approximations, we have used the *Mathematica* [3] program. The relevant dynamics of the network when stimulated by a pure pattern is described by the iteration of the mean-field equations starting from a network state identical to the stimulating pattern. The overlap m_μ can be exactly computed in each time step and, interestingly enough, *exact* fixed points that correspond to the attractors can be found.

Numerical simulations and the storage capacity analysis of this type of networks are presented elsewhere [4].

2 The model with ± 1 neurons

The synaptic matrix is, as in ref [1],

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} [\xi_i^\mu \xi_j^\mu + a(\xi_i^{\mu+1} \xi_j^\mu + \xi_i^{\mu-1} \xi_j^\mu)] \; . \tag{2.1}$$

If $a < 1/2$ the network behaves as a Hopfield one (see ref. [1]).

If $a > 1/2$ and $p < 11$, two different behaviours appear depending on whether p is even or odd. If p is odd, after $\frac{p-1}{2}$ steps the network relaxes to the fixed points

$$\begin{array}{lllll} p & = & 3 & m_\mu & = & 1/2\,(1,1,1)\;, \\ p & = & 5 & m_\mu & = & 1/8\,(5,3,1,1,3)\;, \\ p & = & 7 & m_\mu & = & 1/32\,(19,13,3,1,13,13)\;, \\ p & = & 9 & m_\mu & = & 1/128\,(77,51,13,3,1,1,3,13,51)\;. \end{array}$$

They can be represented by

$$m_{\frac{p+1}{2}} = \frac{1}{2^{p-2}} \;, \tag{2.2}$$

$$m_\eta = 4m_{\eta+1} + \frac{(-1)^{\eta+(\frac{p+1}{2})}}{2^{p-2}} \;, \quad \eta = 2, \ldots, \frac{p+1}{2} - 1 \tag{2.3}$$

$$m_1 = m_2 + 2m_3 \;, \tag{2.4}$$

$$m_{\frac{p+1}{2}+\gamma} = m_{\frac{p+1}{2}-(\gamma-1)} \;, \quad \gamma = 1, \ldots, \frac{p+1}{2} \; . \tag{2.5}$$

If p is even, two-cycle solutions given by the formulæ

$$m_{\frac{p}{2}+1} = 0 \;,$$

$$
\begin{aligned}
m_\eta &= 4m_{\eta+1} - \frac{(-1)^{\eta+(\frac{p}{2}+1)}}{2^{p-3}}\ ,\ \ \eta = 2,\ldots,\frac{p}{2}\ ,\\
m_1 &= m_2 + 2m_3\ ,\\
m_{\frac{p}{2}+1+\gamma} &= m_{\frac{p}{2}+1-\gamma}\ ,\ \ \gamma = 1,\ldots,\frac{p}{2}\ ,
\end{aligned}
\tag{2.6}
$$

and

$$
\begin{aligned}
m_{\frac{p}{2}+1} &= 1\ ,\\
m_\eta &= 4m_{\eta+1} + 3\frac{(-1)^{\frac{p}{2}+1+\eta}}{2^{p-2}}\ ,\ \ \eta = 2,\ldots,\frac{p}{2}\ ,\\
m_1 &= m_2 + 2m_3\ ,\\
m_{\frac{p}{2}+1+\gamma} &= m_{\frac{p}{2}+1-\gamma}\ ,\ \ \gamma = 1,\ldots,\frac{p}{2}
\end{aligned}
\tag{2.7}
$$

are present (they appear because the dynamics is parallel).

When $p \geq 10$, the fixed points are reached after four iterations. The whole set of attractors is obtained by cyclic rotations of

$$
m^1_\mu = \frac{1}{2^7}\,(77, 51, 13, 3, 1, 0, \ldots, 0, 1, 3, 13, 51) \tag{2.8}
$$

and they are universal, *i.e.* the non-zero components do not depend neither on p nor on a. Surprisingly, the overlap vector has exactly 9 components different from zero symmetrically disposed around the component associated to the stimulating pattern.

These attractors are mutually correlated and the correlations only depend on the separation of the corresponding stimulating patterns in the memorized sequence, $d = \alpha - \beta$ ($\alpha > \beta$) and $C_d = C_{p-d}$. If $p \geq n + 2 + d$ (n is the number of non-zero components of m_μ, *i.e.* $n = 9$),

$$\begin{array}{lllllll} C_1 & = & C_{p-1} & = & 1360/2^{11} & \simeq & 0.66 \\ C_2 & = & C_{p-2} & = & 680/2^{11} & \simeq & 0.33 \\ C_3 & = & C_{p-3} & = & 252/2^{11} & \simeq & 0.12 \\ C_4 & = & C_{p-4} & = & 82/2^{11} & \simeq & 0.04 \\ C_5 & = & C_{p-5} & = & 23/2^{11} & \simeq & 0.01 \end{array} \tag{2.9}$$

while more distant attractors are not significantly correlated. They are independent of p and a, if $a \in (1/2, 1)$. Furthermore, if $p \geq 22$ and $10 < d < p-10$ attractors are not correlated at all: $C_d = 0$.

3 Network of $0, 1$ neurons

The dynamics of the spike emission is

$$s_i(t + \delta t) = \Theta[\sum_{i \neq j=1}^{N} J_{ij} s_j(t) - \theta] \tag{3.1}$$

with J_{ij} as in ref.[1] and $P[\eta_i^\mu = 1] = f$.

The critical a, below which pure pattern states are stable, is

$$a_{cr} = \min(\frac{f + \theta}{2(1-f)}, \frac{1 - f - \theta}{2f}) \ . \tag{3.2}$$

The dependence of attractors on a and p, once the probability f and the threshold θ are chosen, becomes here more complicated.

If $a > a_{cr}$ and p is small enough, the overlap vectors corresponding to the new fixed points have all components different from zero. Increasing p (a fixed), a "critical" p_{cr} is reached after which it has a relatively small number of components different from zero concentrated around the starting pattern and the rest of them exactly zero. p_{cr} increases in steps with a and inside each sub-interval there is universality. The values of a dividing sub-intervals depend on f and θ.

Some examples are in order. The ± 1 neural activity model is recovered taking $f = 1/2$ and $\theta = 0$. Then, $a_{cr} = 1/2$ and the whole interval $(1/2, 1)$ presents universality. $p_{cr} = 10$, $n = 9$ and attractors and correlations are given by eqs.(2.8)-(2.9).

If, for instance, $f = 1/10$, $a = 1/10$ and $\theta = 0$: $a > a_{cr}$. The fixed points are

$$\begin{array}{lllll} p & = & 3 & m^1_\mu & = & (1-f)^2 \, (1,1,1) \; , \\ p & \geq & 4 & m^1_\mu & = & (1-f)^2 \, (1,1,0,\ldots,0,1) \; . \end{array}$$

Thus, $p_{cr} = 4$ and $n = 3$.

Increasing a so as to move to another sub-interval, the fixed points change.

Taking $a = 3/10$ they are

$$\begin{array}{rclcrcl} p & = & 3\,, & & m_\mu^1 & = & (1-f)^2\,(1,1,1)\,, \\ p & = & 4\,, & & m_\mu^1 & = & (1-f)^3\,(1,1,1,1)\,, \\ p & = & 5\,, & & m_\mu^1 & = & (1-f)^3\,(1+f,1+f,f,f,1+f)\,, \\ p & \geq & 6\,, & & m_\mu^1 & = & (1-f)^3\,(1+f,1+f,f,0,\ldots,0,f,1+f)\,, \end{array}$$

$p_{cr} = 6$ and $n = 5$.

In the first example, if $p \geq n+2+d = 5+d$ the correlations are

$$\begin{array}{rclclcl} C_1 & = & C_{p-1} & = & 171/271 & \simeq & 0.63\,, \\ C_2 & = & C_{p-2} & = & 81/271 & \simeq & 0.30\,, \\ C_3 = \ldots & = & C_{p-3} & = & 0\,, & & \end{array}$$

while in the second example if $p \geq n+2+d = 7+d$, they are

$$\begin{array}{rclclcl} C_1 & = & C_{p-1} & = & 206119/306119 & \simeq & 0.673\,, \\ C_2 & = & C_{p-2} & = & 106119/306119 & \simeq & 0.347\,, \\ C_3 & = & C_{p-3} & = & 16119/306119 & \simeq & 0.053\,, \\ C_4 & = & C_{p-4} & = & 279/306119 & \simeq & 0.002\,, \\ C_5 = \ldots & = & C_{p-5} & = & 0\,. & & \end{array}$$

In both examples the correlations are independent of p. Their values decay with increasing separation and they are exactly zero after a small distance. In general, if $p > 2n+3$ and $n+1 < d < p-(n+1)$ the correlations are independent of d and *exactly* zero.

The attractors and correlations are similar to those given by the examples in the whole range of values of parameters such that inequality (3.2) holds.

In conclusion, it has been shown that the model of ref. [1], both for ± 1 and $1,0$ neurons, has attractors with a finite similarity index (the components of the overlap) with a small number of nearest stimuli in the learning sequence (5 if $a > 1/2$ for the ± 1 and a function of a and θ, at fixed f, varying in steps for the $1, 0$ neurons, respectively).

ACKNOWLEDGEMENTS

We wish to thank D. J. Amit for stimulating and helpful discussions; J. Kurchan, E. Sivan, M. Tsodyks and M. A. Virasoro for useful discussions.

References

[1] M. Griniasty, M. V. Tsodyks, D. J. Amit; *"Convertion of temporal correlations between stimuli to spatial correlations between attractors"*, Neural Computation **5**, 1-17 (1993).

[2] Y. Miyashita; *"Neuronal correlate of visual associative long-term memory in the primate temporal cortex"*, Nature **335** 817 (1988).

[3] S. Wolfram; *Mathematica*, Computer Program, Wolfram Reasearch Inc., 1990

[4] L. F. Cugliandolo, M. V. Tsodyks; *"Capacity of networks with correlated attractors"*, submitted for publication on J. Phys **A**.