

Bump formations in attractor neural network and their application in image reconstruction

Kostadin Koroutchev^{*,†} and Elka Korutcheva^{**,‡}

^{*} *Escuela Politécnica Superior, Universidad Autónoma de Madrid, Cantoblanco, Madrid, 28049, Spain*

[†] *Also at: Institute for Communication and Computer Systems, Bulgarian Academy of Science, Sofia, Bulgaria.*

^{**} *Depto. de Física Fundamental, Universidad Nacional de Educación a Distancia, c/ Senda del Rey 9, 28080 Madrid, Spain*

[‡] *Also at: G.Nadjakov Inst.Solid State Physics, Bulgarian Academy of Science, Sofia, Bulgaria.*

Abstract. In this paper we analyze the bump formations in binary attractor neural networks with distance dependent connectivity. We show that by introducing a two stage learning procedure an increase of the critical storage capacity of the network is observed. The procedure has been tested on a network with $N = 64K$ neurons by using a selection of 3700 natural images. Our analysis shows that the bumps can be regarded as intrinsic characteristics of the image and the topology of the network and they can be used to improve the performance of the network by increasing its capacity.

Keywords: Attractor Neural Networks, Bump Formation

PACS: 84.35.+i, 64.60.Cn, 87.18.Sn, 07.05+.Mh

INTRODUCTION

Recently bump formations in recurrent neural networks have been analyzed in several investigations concerning linear-threshold units [1, 2], binary units [3, 4] and Small-World networks of Integrate and Fire neurons [5]. It has been shown that the localized retrieval is due to the short-range connectivity of the network, which could explain the behavior in structures of biological relevance [6].

The investigation of the spontaneous activity bumps in Small-World networks (SW) [7, 8] of Integrate-and Fire neurons [5] has recently shown that the network retrieves when its connectivity is close to the random and displays localized bumps of activity when the connectivity is close to the ordered. The two regimes are mutually exclusive in the range of the parameter governing the proportion of the long-range connections on the SW topology of the network.

The result related to the appearance of bump formations have been recently reported by us [3] in the case of binary model for associative network. We demonstrated that the spatially asymmetric retrieval states (SAS) can be observed when the network is constrained to have a different retrieval activity compared to that induced by the patterns.

The model, we studied in Ref.[3], has a symmetric and distance dependent connectivity for all neurons within an attractor neural network (NN) of Hebbian type formed by N binary neurons $\{S_i\}$, $S_i \in \{-1, 1\}$, $i = 1, \dots, N$, storing p binary patterns η_i^μ , $\mu \in \{1 \dots p\}$

CP887, *Cooperative Behavior in Neural Systems: Ninth Granada Lectures*

edited by J. Marro, P. L. Garrido, and J. J. Torres

© 2007 American Institute of Physics 978-0-7354-0390-1/07/\$23.00

with symmetric connectivity between the neurons $c_{ij} = c_{ji} \in \{0, 1\}$, $c_{ii} = 0$.

The model is defined by the following Hopfield's Hamiltonian [9]:

$$H = \frac{1}{N} \sum_{ij} J_{ij} S_i S_j, \quad (1)$$

where we have assumed a Hebbian learning rule:

$$J_{ij} = \frac{1}{c} \sum_{\mu=1}^p c_{ij} (\eta_i^\mu - a)(\eta_j^\mu - a). \quad (2)$$

The learned patterns are drawn from the following distribution:

$$P(\eta_i^\mu) = \frac{1+a}{2} \delta(\eta_i^\mu - 1) + \frac{1-a}{2} \delta(\eta_i^\mu + 1), \quad (3)$$

where the parameter a is the sparsity of the code [10]. The number of connections per neuron is assumed to be constant and equal to $C \equiv cN$.

In order to introduce an asymmetry between the retrieval and the learning states, a condition on the mean activity of the network has to be imposed:

$$H_a = NR \left(\sum_i S_i / N - a \right),$$

where the parameter R controls the affinity of the system toward the appearance of bump formations.

In Ref.[3] we have shown that this condition is necessary and sufficient condition for the observation of bump formations. Similar observations have been reported in the case of linear-threshold network [1], where in order to observe bump formations, one has to constrain the activity of the network. The same is true in the case of smoothly saturating binary network [2], when the highest activity level, that can achieved by the neurons, is above the maximum activity of the units in the stored pattern.

An interesting phenomenon, discussed in Ref.[3], is related to the dramatic drop of the capacity of the network in the regime of the appearance of the bumps, which is due to the reduction of the effective size of the networks until the bumps size.

On the other side, the spatially restricted activity means that the effective coding is restricted to very sparse coding. Usually, the capacity of the network, keeping patterns with sparsity a , is proportional to $1/a |\log a|$ and increases with the sparsity.

In a recent paper [4] we have used explicitly the sparseness of the code, imposed by the bump appearance, in order to increase the capacity of a two-dimensional network that stores natural images. By means of simulations, we have shown that the capacity of the network can be increased to the limits typical for sparsely connected network, e.g. to achieve capacities of order of C .

Based on the these results, we present here the theoretical bases of the analysis and show that the bump formations can be effectively used during the learning process in order to enhance the network performance.

To give an analytic expression of our results obtained by learning using bump formations [4], we investigated the case when the learned patterns are drawn from the following distribution, written in terms of the variables $\xi_i^\mu = \eta_i^\mu - a$.

$$P(\xi_i^\mu) = \frac{1+a}{2} \theta_i^\mu \delta[\xi_i^\mu - (2 - (1+a)\pi R_0^2)] + (1 - \frac{1+a}{2} \theta_i^\mu) \delta[\xi_i^\mu + (1+a)\pi R_0^2], \quad (4)$$

where $\theta_i^\mu \equiv \theta(R_0 - |r_i - r_\mu|)$. Here by r_μ we denote the center of the bump corresponding to the μ -th pattern, while r_i labels the coordinate of the i -th neuron. R_0 is the radius of the spot, which models the bump shape and we are supposing to train the network by the bumps, obtained after preprocessing the patterns (images).

THEORETICAL ANALYSIS

The theoretical analysis of the model, based on the pattern's distribution given by Eq. (3), has been explained in details in Ref.[3]. Here we briefly present the main points of this analysis and the most important results.

For the theoretical analysis of the spatially asymmetric states (SAS), we consider the decomposition of the connectivity matrix c_{ij} by its eigenvalues λ_k and eigenvectors $a_i^{(k)}$:

$$c_{ij} = \sum_k b_i^{(k)} b_j^{(k)}, \quad (5)$$

with

$$b_i^k \equiv a_i^{(k)} \sqrt{\lambda_k/c}. \quad (6)$$

Following the classical analysis of Amit et al. [11], we study the following binary Hopfield model [9]:

$$H = -\frac{1}{cN} \sum_{i,j,\mu} S_i \xi_i^\mu c_{ij} \xi_j^\mu S_j - \sum_{\nu=1}^s h^\nu \sum_i \xi_i^\nu S_i + NR \overline{S_i b_i^0}. \quad (7)$$

The second term in the Hamiltonian introduces a small external field h , which will tend later to zero. This term is necessary to take into account the finite numbers of overlaps s that condense macroscopically. The third term imposes an asymmetry in the neural network's states, which is controlled by the value of the parameter R . This term is responsible for the bump formations. In Eq. (7) the over line means a spatial averaging $\overline{(\cdot)} = \frac{1}{N} \sum_i (\cdot)$.

By using the ‘‘replica formalism’’ [12], the replicated partition function is represented by decoupling the sites using the above expansion of the connectivity matrix c_{ij} , Eq. (5):

$$\begin{aligned} \langle\langle Z^n \rangle\rangle = e^{-\beta \alpha N n / 2c} & \left\langle \left\langle Tr_{S^p} \exp \left[\frac{\beta}{2N} \sum_{\mu \rho l} \sum_{ij} (\xi_i^\mu S_i^\rho b_l^i) (\xi_j^\mu S_j^\rho b_l^j) + \right. \right. \right. \\ & \left. \left. \left. \beta \sum_\nu h^\nu \sum_{ip} \xi_i^\nu S_i^\rho - \beta RN \sum_{ip} b_i^0 S_i^\rho / N \right] \right\rangle \right\rangle, \quad (8) \end{aligned}$$

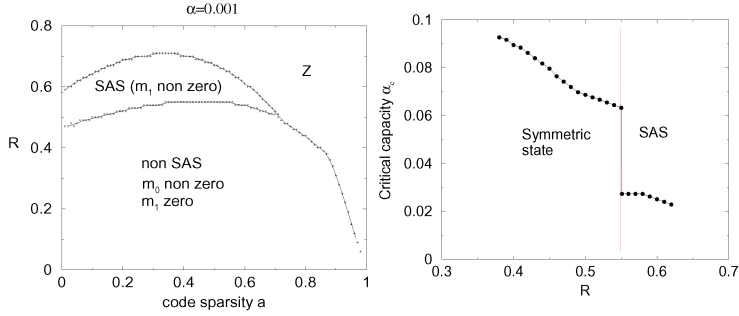


FIGURE 1. Left: Phase diagram for $\alpha = 0.001$ ($\alpha/c = 0.02$). The SAS region, where local bumps are observed, is relatively large. The Z region corresponds to trivial solutions for the overlap. Right: The critical storage capacity α_c as a function of R for $a = 0.4$, showing a drop on the transition to a bump state.

with $\alpha = p/N$ being the storage capacity and n -the number of replicas. Following the classical analysis [11], we use the variables $m_{\rho k}^{\mu} = \frac{1}{cN} \sum_i \xi_i^{\mu} S_i^{\rho} b_i^k$, which are the usual overlaps between neurons and patterns for each replica ρ and each eigenvalue k . Then we split the sums over the first s “condensed”, labeled by v , and the remaining (infinite) $p - s$ patterns, over which an average and a later expansion of the expressions in terms of their corresponding order parameters (OP) is done. In addition, we also use two other order parameters, $q_k^{\rho, \sigma} = \overline{(b_i^k)^2 S_i^{\rho} S_i^{\sigma}}$, which is related to the neuron activity, and the parameter $r_k^{\rho, \sigma}$, conjugate to $q_k^{\rho, \sigma}$, that measures the noise in the system. The indexes $\rho, \sigma = 1, \dots, n$ label the different replicas, while the role of the connectivity matrix is taken into account by the parameters b_i^k .

Finally, by using the replica symmetry ansatz [12] $m_{\rho, k}^v = m_k^v, q_k^{\rho, \sigma} = q_k$ for $\rho \neq \sigma$ and $r_k^{\rho, \sigma} = r_k$ for $\rho \neq \sigma$, followed by a suitable linearization of the quadratic in S -terms and the application of the saddle-point method [11], we obtain the following expression for the free energy per neuron:

$$\begin{aligned}
 f &= \frac{\alpha}{c} (1+a) \pi R_0^2 + \frac{1}{2} \sum_k (m_k)^2 - \alpha \beta (1+a) \pi R_0^2 \sum_k r_k (q_k - \mu_k) + \\
 &+ \frac{\alpha}{2\beta} \sum_k [\ln(1 - 2\beta(1+a)\pi R_0^2 (\mu_k - q_k)) - \\
 &- 2\beta(1+a)\pi R_0^2 q_k (1 - 2\beta(1+a)\pi R_0^2 (\mu_k - q_k))^{-1}] - \\
 &- \frac{1}{\beta} \int \frac{dz e^{-z^2/2}}{\sqrt{2\pi}} \ln 2 \cosh \beta \left(z \sqrt{2\alpha(1+a)\pi R_0^2 \sum_l r_l b_l^l b_l^l} + \sum_l m_l \xi_l b_l^l + R b_l^0 \right).
 \end{aligned}$$

Here $\mu_k = \lambda_k/cN$ and we have used the fact that the average over a finite number of patterns ξ^v can be self-averaged [3], [11].



FIGURE 2. The contours of Lena and the bump formation with sparsity $a = 0.005$ and $r = 15$. The bump represents a compact area.

The equations for the OP m_k , q_k and r_k are respectively:

$$m_k = \int \frac{dze^{-z^2/2}}{\sqrt{2\pi}} \xi_i b_i^k \tanh \beta \left(z \sqrt{2\alpha(1+a)\pi R_0^2 \sum_l r_l b_i^l b_i^l + \sum_l m_l \xi_i b_i^l + R b_i^0} \right),$$

and

$$q_k = \int \frac{dze^{-z^2/2}}{\sqrt{2\pi}} (b_i^k)^2 \tanh^2 \beta \left(z \sqrt{2\alpha(1+a)\pi R_0^2 \sum_l r_l b_i^l b_i^l + \sum_l m_l \xi_i b_i^l + R b_i^0} \right)$$

and

$$r_k = \frac{2q_k(1+a)\pi R_0^2}{(1 - 2\beta(1+a)\pi R_0^2(\mu_k - q_k))^2}. \quad (9)$$

Note that the the only difference in the above equations, compared to the corresponding equations obtained in Ref.[3], is the replacement of the expression $1 - a^2$ by the following one $2(1+a)\pi R_0^2$.

As we know from Ref.[3], in the case of distribution of the patterns given by Eq. (3), the numerical analysis of the corresponding OP's equations for $T = 0$, gives a stable region for the solutions corresponding to bump formations for different values of the load α , the sparsity a and the retrieval asymmetry parameter R , shown in Fig. 1(Left). The sparsity of the code a enhances the SAS effect, although it is also observed for $a = 0$.

The effect of the asymmetry parameter R is seen for intermediate values, where the bumpy region (SAS) is formed. For larger values of R and a , the only phase is the paramagnetic phase (Z).

The behavior of the critical storage capacity α_c shows for this case a drastic drop of α_c at the transition from homogeneous retrieval (symmetric) state to spatially localized bump state (Fig. 1, Right). In this case only the fraction of the network in the bump can be excited and the storage capacity drops proportionally to the size of the bump. This last result has been used by us in Ref.[4] in order to show that although the critical storage

capacity decreases by a half at the transition point, the learning using the corresponding bump formations can improve the storage capacity and the performance of the network.

In the following we will analyze this more interesting case, corresponding to the distribution, given by Eq. (4).

NUMERICAL ANALYSIS

In Ref.[4] we have performed computer experiments to show that the capacity of the network can be increased to the limits typical for sparsely connected network. For this aim we have used the natural image database of van Hateren [13] of 3700 images with decreased resolution restricted down to 256 by 256 points.

In that work we applied a two stage learning procedure. In the first stage we trained a Hebbian network with spatially dependent connectivity and probability of connections $p(r) \propto \text{const} + e^{-r^2/2\sigma^2}$ with r being the radius of the connectivity. By rising the value of the threshold, we observed a bump with determined sparsity, which is stable enough in order to reconstruct the complete image. An example of such a bump is given in Fig. 2. The location of the bump depends exclusively on the topology of the network and on the pattern. During the second stage, we used the bump in order to train a Hebbian network with exactly the same topology, that contains only bumps from the variety of the images. We refer this network as a Cumulative Recurrent Neural Network.

We have shown that the image can be restored effectively from its bump. This gives us an opportunity to save only the bump in a network and to associate another feed-forward network in order to restore the image in question. When the bumps are orthogonal enough, we have a good recovery of the initial patterns.

In order to keep the network in the regime $C \propto N$, the connectivity of the system must be large enough. In these simulations we used $C = 300$, that fits well into the computer memory and for which value, the correlation radius r can vary from 10 to the size of the image.

For large values of r , no bump formations has been observed. On the other side, when r is very small, the network effectively converts into locally connected network. In this case the capacity is small, but not too small because the patterns are spatially correlated and the bump is well formed. Finally, for intermediate levels of the correlation radius r , one can achieve capacity close to the theoretical capacity for sparse random network.

The critical capacity versus the load is shown in Fig. 3. One can see that it has a maximum near $r = 20$ and in general is very high. It is of the same order as the critical capacity of a very sparse network, up to the correlation radius where the bumps disappear and the capacity drops sharply to zero. This behavior of the critical storage capacity is very promising in order to improve the performance of the network by using the bumpy character of the network activity.

CONCLUSION

In this paper we have shown that the storage capacity of a binary attractor neural network can be increased by using a learning procedure based on the bump formations obtained

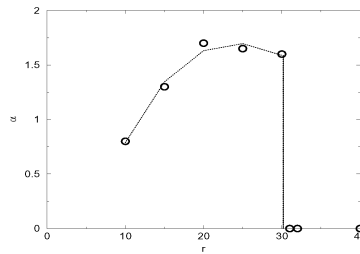


FIGURE 3. Non-trivial behavior of the critical storage capacity during the bump formation phase for $a = 0.001$. When $r = 31.3$, α_c drops to zero.

after preprocessing the original images. Theoretically this is described by the same form of the equations, governing the behavior of the system when using the distribution of the original patterns, with the only difference that now a suitable renormalization by factors depending on the characteristics of the spot (bump) is naturally imposed. The results are the manifestation of the idea that the bumps can be regarded as intrinsic characteristics of the image and the topology of the network.

ACKNOWLEDGMENTS

The authors acknowledge the financial support from the Spanish Grants DGI.M. CyT. FIS 2005-1729 and TIN 2004-07676-G01-01.

REFERENCES

1. Y. Roudi and A. Treves, *J. Stat. Mech.* **1**, P07010 (2004).
2. Y. Roudi and A. Treves, *Phys. Rev. E* **73**, 061904 (2006).
3. K. Koroutchev and E. Korutcheva, Preprint ICTP, Trieste, Italy, IC/2004/91, 1 (2004); K. Koroutchev and E. Korutcheva, *Central Eur. J. Phys.*, **3**, 409–419 (2005); K. Koroutchev and E. Korutcheva, *Phys. Rev. E* **73**, 026107 (2006).
4. K. Koroutchev and E. Korutcheva, accepted for publication in the Proceedings of ICANN06, LNCS 2006.
5. A. Anishchenko, E. Bienenstock and A. Treves, q-bio.NC/0502003.
6. V. Breitenberg and A. Schulz, *Anatomy of the Cortex*, Springer, Berlin, 1991.
7. D. J. Watts and S.H. Strogatz, *Nature*, **393**, 440–442 (1998).
8. D. J. Watts, *Small Worlds: The Dynamics of Networks Between Order and Randomness*, in Princeton Review in Complexity, Princeton University Press, Princeton, New Jersey, 1999.
9. J. Hopfield, *Proc. Natl. Acad. Sci. USA*, **79**, 2554–2558 (1982).
10. M. Tsodyks and M. Feigel'man, *Europhys. Lett.*, **6**, 101–105 (1988).
11. D. Amit, H. Gutfreund and H. Sompolinsky, *Ann. Phys.*, **173**, 30–67 (1987).
12. M. Mézard, G. Parisi and M. A. Virasoro, *Spin-glass theory and beyond*, World Scientific, Singapore, 1987.
13. J. H. van Hateren and A. van der Schaaf, in: *Proc. R. Soc. Lond.* **B265**, 359–366 (1998).

Copyright of AIP Conference Proceedings is the property of American Institute of Physics and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.