ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 393-398

Storage Capacity and Dynamics of Nonmonotonic Networks

Bruno Crespi[a] and Ignazio Lazzizzera[b]
a. IRST, I-38050 Povo (Trento) Italy,
b. Univ. of Trento, Physics Dept., I-38050 Povo (Trento) Italy
INFN Gruppo collegato di Trento

Abstract. This work investigates the retrieval capacities of different types of nonmonotonic neurons. Storage capacity is maximized when the neuron response is a function with well defined geometrical characteristics. Numerical experiments demonstrate that storage capacity is directly related to the dynamical property of the iterative map that describes the network evolution. Maximum capacity is reached when the neuron dynamics are subdivided into two non-overlapping "erratic bands" around points $x_i = \pm 1$.

1. Introduction

We consider the storage capacity of fully connected Hopfield models [2, 1]. The discrete dynamics of the network elements, $X(t) = \{x_i(t),\ i = 1, .., N\}$, are assumed to be governed by the N-dimensional iterative map

$$x_i(t+1) = g\left(\sum_{j=1}^{N} T_{ij}\, x_j(t)\right) \qquad \text{with} \qquad T_{ij} = \frac{1}{N}\sum_{\mu=1}^{M} \xi_i^\mu\, \xi_j^\mu \tag{1}$$

where $\Xi^\mu = (\xi_1^\mu, \xi_2^\mu, ..., \xi_N^\mu)$, $\xi_i^\mu = \pm 1$, $\mu = 1, 2, ..., M$ are the configurations to be memorized.

The retrieval process is identified in the convergence of network configuration $X(t)$ towards one of the stable fixed points of iterative map (1). A successful recognition is achieved when the overlap between the final static configuration and one of the memorized configurations, e.g. $\mu = 1$, is close to unity, $m_1 \approx \frac{1}{N}\sum_i \xi_i^1\, x_i = 1$. In this work, functions $g(x)$ are normalized in such a way that $g(\pm 1) = \pm 1$. This normalization ensures that in the ideal case in which the stored vectors are orthogonal, $(\Xi^\mu, \Xi^\nu) = \delta_{\mu,\nu}$, configurations Ξ^μ are fixed points of iterative map (1).

In the traditional models, in which the neuron response is a monotone function, the number of configurations that can be stored and retrieved without errors – the *absolute capacity* – is $M_A \approx \frac{N}{2 \ln N}$. If a small fraction of errors is allowed, – the *relative capacity* – is $M_R \approx \alpha_c N$ where $\alpha_c \approx 0.14$ [3, 4]. In the $N \to \infty$ limit, critical values M_A and M_R separate two distinct phases of the system.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 393-398

It was shown that if the sigmoid is replaced with a nonmonotone function storage capacity is decisively improved [6, 7]. In [7], a theoretical study of two-stage dynamics with neuron response $f(x) = -a\,x + c\,sgn(x)$, with $a = (c+1)/2$ and $a > 0$, indicates an absolute capacity of about $N/\sqrt{2\,ln\,N}$ and computer simulations estimate a relative capacity of $0.3\,N$.

2. The model considered

As a first example of nonmonotone models, we consider function

$$g_{gd}(u) = u\,e^{-\frac{\beta}{2}(u^2-1)}. \tag{2}$$

that depends on a single parameter, β. Slope at $x = 1$ is negative for $\beta > 1$.

Figures 1 shows the retrieval performance for a $N = 100$ network with neuron response given by (2), as a function of parameter β for several values of loading parameter $\alpha = M/N$. Initial configurations have a $m_0 = 0.8$ overlap with the target. Each point, for $1 < \beta < 5$, is generated by averaging over 1000 samples. Performance is measured in terms of the overlap between the network output and the pattern to be retrieved. In the following simulations the system final output is binarized, i.e. final values x_i are set to ± 1 by means of $x_i \rightarrow \overline{x_i} = sgn(x_i)$. We indicate with $\overline{m}$ the overlap obtained using binarized variables $\overline{x_i}$.

The final configuration has values x_i that are contained in two bands around ± 1, the width of which depends on α and β. For low α the range of values of β that maximizes the overlap is wide, as α increases this range shrinks around value $\beta \approx 3$. Retrieval with few errors is obtained for $\alpha < 0.4$. Networks of larger size give the same results, enhancing the difference between different phases of the network behavior.

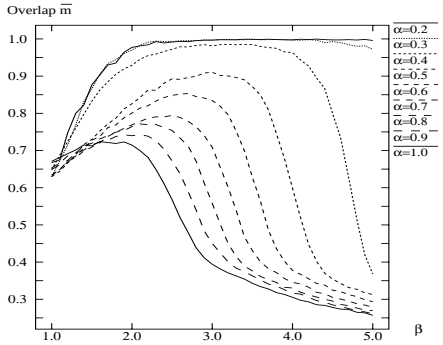

Figure 1: The retrieval performance for a $N = 100$ network.

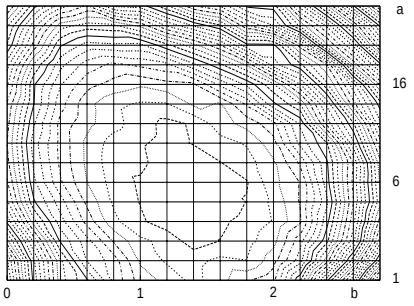

Figure 2: Contour plot of overlap $\overline{m}$ for the piecewise linear response for $\alpha = 0.4$.

Capacity enhancement with respect to monotone functions is a general behavior of nonmonotone response functions. Referring to [5, 6, 7], we have experimentally compared function (2) with two other functions: 1) the piecewise linear

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 393-398

function,

$$g_{pl}(x) = \begin{cases} a\,x & \text{for } |x| < -\frac{1+b}{a+b} \\ -b\,x \pm (1+b) & \text{for } \frac{1+b}{a+b} < \pm x < \frac{1+b}{b} \\ \mp b\,d \pm (1+b) & \text{for } \frac{1+b}{b} < d < \pm x \end{cases} \tag{3}$$

and 2) the Morita function,

$$g_m(x) = A\,\frac{1-e^{-c\,x}}{1+e^{-c\,x}} \cdot \frac{1+\kappa\,e^{c'\,(|x|-h)}}{1+\,e^{c'\,(|x|-h)}} \tag{4}$$

where A is a normalization factor such that $g_m(\pm 1) = \pm 1$ and $h = 1$ [1].

The shape of these functions depends on the parameter values. We performed a numerical study to individuate the parameter values that maximize retrieval capacity. Better results are achieved when positive (negative) values are mapped into positive (negative) values. For this reason we set $d = (1+b)/b$ in the piecewise linear function, and $\kappa = 0$ in the Morita function.

Once parameters d and k are set, the other parameters (a and b for the piecewise linear function and c and c' for the Morita function) are related with the positive and negative slopes of the functions. For these two parameters we searched the values that maximize the overlap between the system configuration and the pattern to be retrieved, averaged over a large number of samples (1000). For $\alpha = 0.4$, the optimal values are $a \approx 6$ and $b \approx 1.4$ for the piecewise linear function, see Fig. 2. The spacing between contour levels is 0.13, the highest contour level is 0.936 [2]. In the same way, for the Morita function we found $c \approx 6$, $c^{'} \approx 5$.

Note that even for $\alpha = 0.4$ there is a large plateau of parameter values that maximize storage capacity, i.e. the contour line that includes values within a small fraction of the maximum delimits a large area. For lower values of load parameter α this area enlarges while it shrinks for higher values. This behavior is common to all three functions. The optimal values slightly change with the loading parameter α. For example, for $\alpha = 0.46$ we have $a \approx 6$ and $b \approx 1.0$ for the piecewise linear response.

A direct comparison between the storage capacities of the three functions — the Gaussian derivative eq. (2), the Morita function, and the piecewise linear function — is shown in figures 3. For all the functions, parameter values that maximize storage capacity are used. The figure was obtained for a $N = 100$ network, starting from initial configurations that have a $m_0 = 0.8$ overlap with the target. Each point is generated by averaging over 1000 samples.

The simulation results indicate that retrieval performance depends on the geometric shape of the nonmonotone function. In fact, as shown in Fig. 4, the three functions have very similar shapes and performance when the optimal parameters are used. For function (2) we choose as optimal value the median of the intersections betweens the $m = 0.9$ line and the $\alpha = 0.4$ curve, i.e. $\beta \approx 3.2$.

[1] The value of h was set to unity in order to lower the number of parameters while keeping a reasonable shape.

[2] The optimal values are defined as the approximate position of the maximum; they were found using a higher number of contour levels than that shown in the figure.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 393-398

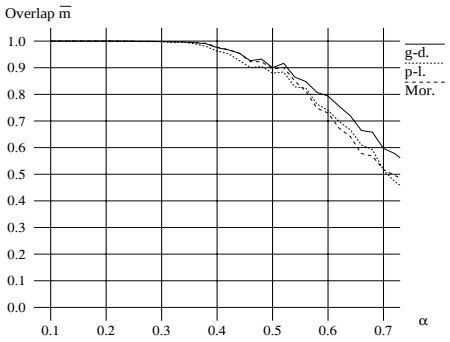

Figure 3: Performance for the three functions with optimal parameters.

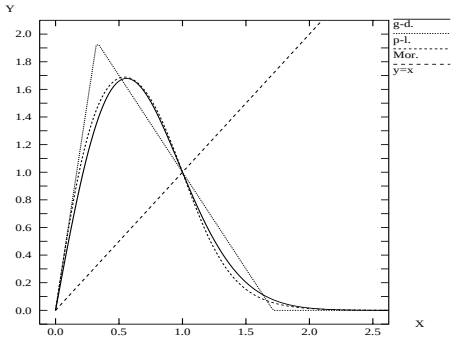

Figure 4: Shape of the three functions for "optimal" parameter values.

3. Relation with the dynamical properties of the response function

We consider function (2). Fig. 5 shows the m distribution as a function of parameter β for different values of the loading parameter $\alpha = M/N$. Note that in the computation of m, final values x_i are not binarized to ± 1 by means of $x \rightarrow sgn(x)$ as previously done. The distribution was obtained using 1000 samples for each value of β and α. The gray level of a pixel (β, m) is proportional to the number of samples that fall into $[\beta - 1/128, \beta + 1/128]$ and $[x - 1/128, x + 1/128]$. For

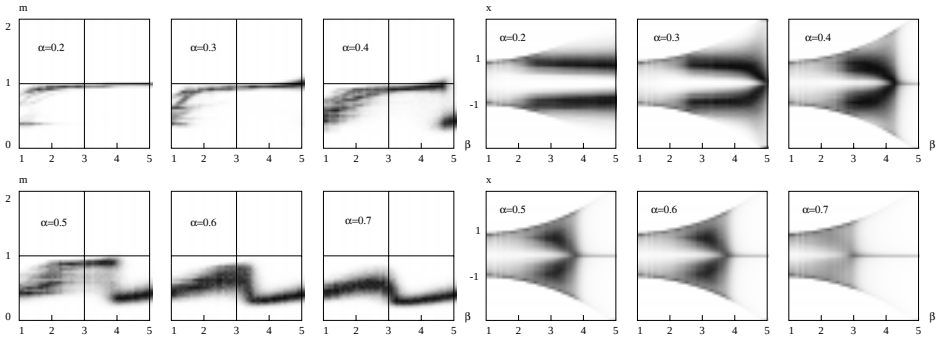

Figure 5: Distribution of overlap m in the range $1 < \beta < 5$ for six values of the load parameter α (initial overlap $m_0 = 0.8$).

Figure 6: Distribution of x in the range $1 < \beta < 5$ for six values of the load parameter α (initial overlap $m_0 = 0.8$).

fixed α, there is a sharp transition in performance when β becomes larger than a threshold value (for $\alpha = 0.2$ the transition point is at $\beta \approx 6$). The value of this threshold decreases as α increases.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 393-398

The low value of m for "high" β is caused by neuron values $x_i,\ i = 1, ..., N$, moving close to zero. This is evidenced by Fig. 6 that displays values x_i during the network evolution for different values of α as a function of β. Each iteration is generated from a random initial configuration that has a $m_0 = 0.8$ overlap with the target. The gray level of a pixel (β, x) is proportional to the number of times the iterated points $x_i,\ \ i = 1, ..., N$, fall into the ranges of values $[\beta - 1/128, \beta + 1/128]$ and $[x - 1/128, x + 1/128]$. In general, the network configuration has values x_i that are not strictly ± 1 but are contained within two bands around ± 1. The width of these bands depends on α and β.

The comparison between figures 5 and 6 indicates that, for a fixed memory load, the system achieves maximum storage capacity when the dynamics form around points ± 1 two well-separated bands in which iterated points move erratically, i.e. evenly covering the available space. On the contrary when the iterated points accumulate around the fixed points, storage capacity is lowered.

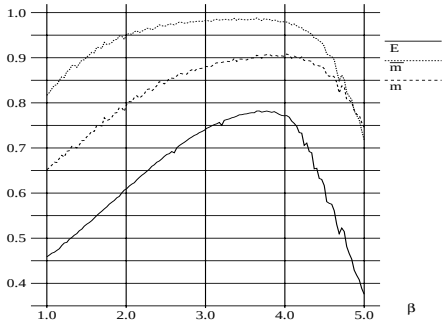

Figure 7: Comparison between retrieval capability, in terms of final overlaps m and $\overline{m}$, and the measure of the irregularity of the iteration process (as defined by eq. 10) for $\alpha = 0.4$. The initial overlap is $m_0 = 0.8$.

This irregular behavior is quantitatively characterized in Fig. 7, where overlap m and overlap $\overline{m}$ are plotted against a measure of the spread of the iterated points, as a function of β (similar results were obtained for other values of α). To define an appropriate measure, interval $[-3, 3]$ is discretized into $W = 10^6$ segments (of length $\epsilon = 10^{-6}$). The measure is defined as the ratio between the total number of values produced, NVP, and the number of segments visited during the iteration process,

$$E = \frac{\sum_{i=0}^{W} \delta_i}{NVP} \quad \in \ [0,\ 1] \tag{5}$$

where $\delta_i = 1$ if during iteration one or more values fall inside interval $[-3 + i\,\epsilon,\ -3 + (i+1)\,\epsilon]$, otherwise $\delta_i = 0$. The total number of values, NVP, is given by $N \sum_s^S I_s$, where $N = 100$ is the number of neurons, $S = 1000$ is the number of test samples, and I_s is the number of iterations corresponding to sample s.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 393-398

Measure E tends toward unity if the iterated points spread uniformly in interval $[-3, 3]$. On the contrary it decreases toward zero if points accumulate around particular values. The result is stable under variations of the number of intervals W (for large W, of order NVP). Similar results are obtained for other values of load parameter α.

4. Conclusions

The capacity of associative memories with different types of non-monotonic neurons were numerically investigated to determine the "optimal" shape of the neuron response function. It was shown that starting from different analytical expressions, such as a piecewise linear function, the Morita function, and a Gaussian derivative function, capacity is maximized when the functions approximate a well defined shape. The capacity enhancement is connected with the dynamical properties of the iterative map that describes the discrete network evolution. While in monotone models neurons are constrained to quasi-binary values close to ± 1, in the nonmonotone models they can assume values in a wider range. We have shown that this property is directly related to the capacity increase. In fact, the capability of retrieving patterns is maximized when the dynamics of the individual neuron are confined in two wide but well-disconnected bands around ± 1. The optimal shape of the neuron function corresponds to the case in which the two bands are uniformly covered by the iterated points.

References

[1] D.J. Amit. *Modelling Brain Functions.* Cambridge University Press, New York, 1989.

[2] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities". *Proc. Natl. Acad. Sci. USA*, Vol. 79, pp. 2554-2558, April 1982.

[3] I. Kanter and H. Sompolinsky, "Associative recall of memory without errors". *Phys. Rev. A*, Vol. 35, No. 1: 380–392, 1987.

[4] L. F. Abbott and Y. Arian. "Storage capacity of generalized networks". *Phys. Rev. A*, Vol. 36, No. 10: 5091–5094, 1987.

[5] S. Yoshizawa, M. Morita, and S. Amari "Capacity of associative memory using a nonmonotonic neuron model". *Neural networks*, Vol. 6, pp. 167-176 (1993).

[6] M. Morita. "Associative memory with Nonmonotonic Dynamics". *Neural networks*, Vol. 6, pp. 115-126 (1993).

[7] H. Yanai, and S. Amari "Auto-associative memory with two-stage dynamics of nonmonotonic neurons". *IEEE Transactions on Neural networks*, Vol. 7, No. 4, pp. 803-815 (1996).