Theoretical Study on the Capacity of Associative Memory with Multiple Reference Points

Enrique Mérida-Casermeiro[1], Domingo López-Rodríguez[1], Gloria Galán-Marín[2], and Juan M. Ortiz-de-Lazcano-Lobato[3]

[1] Department of Applied Mathematics, University of Málaga, Málaga, Spain
{merida,dlopez}@ctima.uma.es
[2] Department of Electronics and Electromechanical Engineering, University of Extremadura, Badajoz, Spain
gloriagm@unex.es
[3] Department of Computer Science and Artificial Intelligence, University of Málaga, Málaga, Spain
jmortiz@lcc.uma.es

Abstract. An extension to Hopfield's model of associative memory is studied in the present work. In particular, this paper is focused in giving solutions to the two main problems present in the model: the apparition of spurious patterns in the learning phase (implying the well-known and undesirable effect of storing the opposite pattern) and the problem of its reduced capacity (the probability of error in the retrieving phase increases as the number of stored patterns grows). In this work, a method to avoid spurious patterns is presented and studied, and an explanation to the previously mentioned effect is given. Another novel technique to increase the capacity of a network is proposed here, based on the idea of using several reference points when storing patterns. It is studied in depth, and an explicit formula for the capacity of the network is provided. This formula shows the linear dependence of the capacity of the new model on the number of reference points, implying the increase of the capacity in this model.

1 Introduction

Associative memory has received much attention for the last two decades. Though numerous models have been developed and investigated, the most influential is Hopfield's associative memory, based on his bipolar model (BH) [1]. This kind of memory arises as a result of his studies on collective computation in neural networks.

Hopfield's model consists in a fully-interconnected series of bi-valued neurons (outputs are either -1 or $+1$). Neural connection strength is expressed in terms of weight matrix $W = (w_{i,j})$, where $w_{i,j}$ represents the synaptic connection between neurons i and j. This matrix is determined in the learning phase by applying Hebb's postulate of learning [2], and no further synaptic modification is considered later.

Two main problems arise in this model: the apparition of spurious patterns and its low capacity.

Spurious patterns are stable states, that is, local minima of the corresponding energy function of the network, not associated to any stored (input) pattern. The simplest, but

J. Mira and J.R. Álvarez (Eds.): IWINAC 2007, Part I, LNCS 4527, pp. 292–302, 2007.

not the least important, case of apparition of spurious patterns is the fact of storing, given a pattern, its opposite, i.e. both X and $-X$ are stable states for the net, but only one of them has been introduced as an input pattern.

The problem of spurious patterns is very fundamental for cognitive modelers as well as practical users of neural networks.

Many solutions have been suggested in the literature. Some of them [3,4] are based on introducing asymmetry in synaptic connections.

However, it has been demonstrated that synaptic asymmetry does not provide by itself a satisfactory solution to the problem of spurious patterns, see [5,6]. Athithan [7] provided a solution based on neural self-interactions with a suitably chosen magnitude, if Hebb's learning rule is used, leading to the near (but not) total suppression of spurious patterns.

Crick [8] suggested the idea of unlearning the spurious patterns as a biologically plausible solution to suppress them. With a physiological explanation, they suggest that spurious patterns are unlearned randomly by human brain during sleep, by means of a process that is the reverse of Hebb's learning rule. This may result in the suppression of many spurious patterns with large basins of attraction. Experiments have shown that their idea leads to an enlargening of the basins for correct patterns along with the elimination of a significant fraction of spurious patterns [9]. However, a great number of spurious patterns with small basins of attraction do survive. Also, in the process of indiscriminate reverse learning, there is a finite probability of unlearning correct patterns, what makes this strategy unacceptable.

On the other hand, the capacity parameter α is usually defined as the quotient between the maximum number of patterns to load into the network, and the number of used neurons that achieve an acceptable error probability in the retrieving phase, usually $p_e = 0.01$ or $p_e = 0.05$. It was empirically shown that this constant is approximately $\alpha = 0.15$ for BH (very close to its actual value, $\alpha = 0.1847$, see [4]). The meaning of this capacity parameter is that, if the net is formed by N neurons, a maximum of $K \leq \alpha N$ patterns can be stored and retrieved with little error probability.

McEliece [10] showed that an upper bound for the asymptotic capacity of the network is $\frac{1}{2 \log N}$, if most of the input (prototype) patterns are to remain as fixed points. This capacity decreases to $\frac{1}{4 \log N}$ if every pattern must be a fixed point of the net.

By using Markov chains to study capacity and the recall error probability, Ho et al. [11] showed results very similar to those obtained by McEliece, since for them it is $\alpha = 0.12$ for small values of N, and the asymptotical capacity is given by $\frac{1}{4 \log N}$.

Kuh [12] manifested roughly similar estimations by making use of normal approximation theory and the theorems about exchangeables random variables.

In this work, a multivalued generalization of Hopfield's model (called MREM) is studied as an associative memory, and a technique to totally avoid the apparition of spurious patterns (in both models, BH and MREM) is explained in terms of the decrease of the energy function associated to patterns.

The main contribution of this paper consists in an extension of these models as associative memories to overcome the problem of the reduced capacity, by using a new technique which ensures the linear increase of the capacity.

2 The MREM Model

Let us consider a recurrent neural network formed by N neurons, where the state of each neuron $i \in \mathcal{I} = \{1, \ldots, N\}$ is defined by its output V_i taking values in any finite set $\mathcal{M} = \{m_1, m_2, \ldots, m_L\}$. This set does not need to be numerical.

The state of the network, at time t, is given by a N-dimensional vector, $\boldsymbol{V}(t) = (V_1(t), V_2(t), \ldots, V_N(t)) \in \mathcal{M}^N$. Associated to every state vector, an energy function, is defined:

$$E(\boldsymbol{V}) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} f(V_i, V_j) + \sum_{i=1}^{N} \theta_i(V_i) \tag{1}$$

where $w_{i,j}$ is the weight of the connection from the j-th neuron to the i-th neuron, $f : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ can be considered as a measure of similarity between the outputs of two neurons, usually verifying the conditions mentioned in [13]:

1. For all $x \in \mathcal{M}$, $f(x, x) = c \in \mathbb{R}$.
2. f is a symmetric function: for every $x, y \in \mathcal{M}$, $f(x, y) = f(y, x)$.
3. If $x \neq y$, then $f(x, y) \leq c$.

and $\theta_i : \mathcal{M} \to \mathbb{R}$ are the threshold functions. Since thresholds will not be used for content addressable memory, henceforth we will consider θ_i be the zero function for all $i = 1, \ldots, N$.

The introduction of this similarity function provides, to the network, of a wide range of possibilities to represent different problems [13,14]. So, it leads to a better and richer (giving more information) representation of problems than other multivalued models, as SOAR and MAREN [15,16], since in those models most of the information enclosed in the multivalued representation is lost by the use of the signum function that only produces values in $\{-1, 0, 1\}$.

The energy function characterizes the dynamics of the net, as happened in BH. In every instant, the net evolves to reach a state of lower energy than the current one.

In this work, we have considered discrete time and semi-parallel dynamics, where only one neuron is updated at time t. The next state of the net will be the one that achieves the greatest descent of the energy function by changing only one neuron output.

Let us consider a total order in $\mathcal{M}$. The potential increment when a-th neuron changes its output from V_a to $l \in \mathcal{M}$ at time t, is

$$U_{a,l}(t) = -\sum_{i=1}^{N} [w_{i,a} \cdot (f(V_i(t), l) - f(V_i(t), V_a(t)))] \tag{2}$$

(due to the similarity conditions imposed to f).

We use the following updating rule for the neuron outputs:

$$V_a(t+1) = \begin{cases} l, & \text{if } U_{a,l}(t) \geq U_{b,k}(t) \forall k \in \mathcal{M} \text{ and } \forall b \in \mathcal{I} \\ V_a(t), & \text{otherwise} \end{cases} \tag{3}$$

This means that each neuron computes in parallel the value of a L-dimensional vector of potentials, related to the energy decrement produced if the neuron state is changed.

The only neuron changing its current state is the one producing the maximum decrease of energy.

It has been proved that the MREM model with this dynamics always converges to a minimal state [13]. This result is particularly important when dealing with combinatorial optimization problems, where the application of MREM has been very fruitful [13,17].

If function $f(x, y) = 2\delta_{x,y} - 1$, which equals 1 if and only if its two parameters coincide, and -1 in the rest of cases, is used and $\mathcal{M} = \{-1, 1\}$, MREM reduces to Hopfield's model. So, MREM is a powerful generalization of BH and other multivalued models, because it is capable of representing the information more accurately than those models.

2.1 MREM as Auto-associative Memory

Now, let $\{X^{(k)} : k = 1, \ldots, K\}$ be a set of patterns to be loaded into the neural network. Then, in order to store a pattern, $X = (X_i)_{i \in \mathcal{I}}$, components of the W matrix must be modified in order to make X the state of the network with minimal energy.

As pointed out in [13], since energy function is defined as in Eq. (1), we calculate $\frac{\partial E}{\partial w_{ij}} = -\frac{1}{2} f(V_i, V_j)$ and we modify the components of matrix W in order to reduce the energy of state $\boldsymbol{V} = X$ by the rule $\Delta w_{i,j} = -\alpha \frac{\partial E}{\partial w_{i,j}} = \frac{\alpha}{2} f(X_i, X_j)$ for some $\alpha > 0$. For simplicity, we can consider $\alpha = 2$, resulting:

$$\Delta w_{i,j} = f(X_i, X_j) \tag{4}$$

and considering that, at first, $W = 0$, that is, all the states of the network have the same energy and adding over all the patterns, the next expression is obtained:

$$w_{i,j} = \sum_{k=1}^{K} f(X_i^{(k)}, X_j^{(k)}) \tag{5}$$

Equation (5) is a generalization of Hebb's postulate of learning, because the weight w_{ij} between neurons is increased in correspondence with their similarity.

It must be pointed out that, when bipolar neurons and the product function $f(x, y) = xy$ are used, the well-known learning rule of patterns in the Hopfield's network is obtained. This is also achieved by the use of the function $f(x, y) = 2\delta_{x,y} - 1$. In the rest of this work we will consider the use of this function for our study.

Analogously to BH, the network is initialized with the known part of the pattern to be retrieved (called probe). The network dynamics will converge to a stable state minimizing the energy function, and it will be the answer of the network.

3 How to Avoid Spurious States

When a pattern X is loaded into the network, by modifying weight matrix W, not only the energy corresponding to state $\boldsymbol{V} = X$ is decreased. This fact can be explained in terms of the so-called associated vectors.

Definition 1. *Given a state V, its* associated matrix *is defined as $G_{\boldsymbol{V}} = (g_{i,j})$ such that $g_{i,j} = f(V_i, V_j)$.*

Its associated vector *is $A_{\boldsymbol{V}} = (a_k)$, with $a_{j+N(i-1)} = g_{i,j}$, that is, it is built by expanding the associated matrix as a vector of N^2 components.*

Lemma 1. *The increment of energy of a state V when pattern X is loaded into the network, by using Eq. (4), is given by:*

$$\Delta E(\boldsymbol{V}) = -\frac{1}{2} < A_X, A_{\boldsymbol{V}} >$$

where $< \cdot, \cdot >$ denotes the usual inner product.

Lemma 2. *Given a state vector $\boldsymbol{V}$, we have $A_{\boldsymbol{V}} = A_{-\boldsymbol{V}}$. So $E(\boldsymbol{V}) = E(-\boldsymbol{V})$.*

These two results explain why spurious patterns are loaded into the network.

It must be noted that, in MREM, the number of spurious patterns appearing after the load of a vector into the net is greater than the corresponding in BH.

An important remark has to be done at this point: With this notation, the expression of the energy function can be rewritten as:

$$E(\boldsymbol{V}) = -\frac{1}{2} \sum_{k=1}^{K} < A_{X^{(k)}}, A_{\boldsymbol{V}} >$$

Since all associated vectors are vectors of N^2 components taking value in $\{-1, 1\}$, their norms are equal, $||A_{\boldsymbol{V}}||_E = N$ for all $\boldsymbol{V}$. This result implies that what is actually stored in the network is the orientation of the vectors associated to loaded patterns.

From the above expression for the increment of energy, and using that components of associated vectors are either -1 or 1, the following expression for the decrease of energy when a pattern is loaded is obtained:

$$-\Delta E(\boldsymbol{V}) = \frac{1}{2}(N - 2d_H(\boldsymbol{V}, X))^2 \tag{6}$$

where $d_H(\boldsymbol{V}, X)$ is the Hamming distance between vectors $\boldsymbol{V}$ and X.

After this explanation, we propose a solution for this problem:

Definition 2. *The* augmented pattern $\hat{X}$*, associated to $X \in \mathcal{M}^N$, is defined by appending to X the possible values of its components, that is, if $\mathcal{M} = \{m_1, \ldots, m_L\}$, then $\hat{X} = (X_1, \ldots, X_N, m_1, \ldots, m_L\}$. Particularly:*

- *In case of bipolar outputs, $\mathcal{M} = \{-1, 1\}$, and it is $\hat{X} = (X_1, \ldots, X_N, -1, 1)$.*
- *If $\mathcal{M} = \{1, \ldots, L\}$, then $\hat{X} = (X_1, \ldots, X_N, 1, 2, \ldots, L)$.*

By making use of augmented patterns, the problem of spurious patterns is solved, as stated in the next result, which is easy to prove:

Theorem 1. *The function Ψ that associates an augmented pattern to its corresponding associated vector is injective.*

Then, in order to store a pattern X, it will suffice to load its augmented version, which will be the unique state maximizing the decrease of energy.

It must be noted that it will only be necessary to consider N neurons, their weights, and the weights corresponding to the last L neurons, that remain fixed, and do not need to be implemented.

4 Associative Memory with Multiple Reference Points

In Hopfield's classical model, the unique reference point is the origin in $\mathbb{R}^N$, that is, patterns are not shifted or translated to achieve better results. As the network stores the orientations of the associated vectors, it could be useful to shift patterns by different amounts in order to be capable of distinguishing them more accurately.

In this work, let us consider $\mathcal{M} = \{1, \ldots, M\}$. So, to load the set $\{X^{(k)} : k = 1, \ldots, K\}$, we use as reference points $O^{(1)}, \ldots, O^{(Q)} \in \mathcal{M}^N$. This means that what the net is going to store is the set of augmented patterns related to $X^{(k)} - O^{(q)}$, for each k and q.

As $X_i^{(k)} - O_i^{(q)} \in \{1 - M, \ldots, -1, 0, 1, \ldots, M - 1\} = \mathcal{M}'$, the augmented pattern associated to $X^{(k)} - O^{(q)}$ will be (using the same notation for simplicity)

$$X^{(k)} - O^{(q)} = (X_1^{(k)} - O_1^{(q)}, \ldots, X_N^{(k)} - O_N^{(q)}, 1 - M, \ldots, M - 1\}$$

In addition, we will refer to the components of the above vector as

$$(X^{(k)} - O^{(q)})_i = \begin{cases} X_i^{(k)} - O_i^{(q)} & i \leq N \\ i - (N + M) & N + 1 \leq i \leq N + 2M - 1 \end{cases}$$

Let us also denote $L = 2M - 1$, the cardinal of the set $\mathcal{M}'$.

By extending what was exposed in Sec. 2, a new energy function is introduced:

$$E(\boldsymbol{V}) = -\frac{1}{2} \sum_{q=1}^{Q} \sum_{i=1}^{N+L} \sum_{j=1}^{N+L} w_{i,j}^{(q)} f((\boldsymbol{V} - O^{(q)})_i, (\boldsymbol{V} - O^{(q)})_j) \tag{7}$$

where

$$w_{i,j}^{(q)} = \sum_{q=1}^{Q} \sum_{k=1}^{K} f((\boldsymbol{V} - O^{(q)})_i, (\boldsymbol{V} - O^{(q)})_j)$$

The above expression can be rewritten in the following terms:

$$E(\boldsymbol{V}) = \sum_{q=1}^{Q} E_q(\boldsymbol{V})$$

The expression of this new energy function implies that each neuron will be able to perform a more complex process in each step, since it has to take into account Q reference points, available in its own local memory. Thus, we are increasing the complexity of the neuron model, that will lead to a higher performance of the net in terms of an increase of its capacity, as we will prove in the next Section.

We must observe that this new model, with $Q = 1$, reduces to MREM standard associative memory.

5 Capacity of the New Model

The capacity of the network is a measure of the amount of patterns that can be introduced into the network such that at the retrieving phase the probability of error does

not exceed a threshold, p_e. The aim of this section is to present a study of the network capacity (similar to [4]) that provides us with an exact or very approximate expression for the capacity parameter α.

Let us suppose that K patterns $X^{(1)}, \ldots, X^{(K)}$, have been loaded into the network, and that state vector $\boldsymbol{V}$ matches a stored pattern, $X^{(k_0)}$. Suppose that state $\boldsymbol{V}'$ coincides to $\boldsymbol{V}$ except in one component. Without loss of generality, this component can be assumed to be the first one, that is, $V_i = V_i'$ if $i > 1$ and $V_1 \neq V_1'$.

By denoting as $D = \Delta E = E(\boldsymbol{V}') - E(\boldsymbol{V})$ the energy increment between these two states $\boldsymbol{V}$ and $\boldsymbol{V}'$, the pattern $X^{(k_0)}$ is correctly retrieved when pattern $\boldsymbol{V}'$ is introduced into the net if $D > 0$, because this condition implies that $\boldsymbol{V}$ is a fixed point for the dynamics that are being used. So, in order to calculate the error probability in the retrieval phase, the probability $P(D < 0)$ must be computed.

But

$$D = \Delta E = \sum_{q=1}^{Q} E_q(\boldsymbol{V}') - \sum_{q=1}^{Q} E_q(\boldsymbol{V}) = \sum_{q=1}^{Q} (E_q(\boldsymbol{V}') - E_q(\boldsymbol{V})) = \sum_{q=1}^{Q} \Delta E_q \quad (8)$$

and thus we have to compute $D_q = \Delta E_q$.

To this end, we present some technical results which will guide us to the main result of this section, the capacity of the network with multiple reference points. Proofs for these results will be omitted due to the limitation in the length of this paper. The reader can refer to [18] for a detailed proof in the case of Hopfield's bipolar model with multiple reference points.

Lemma 3. *We have*

$$D_q = N + 3 - \sum_{i=2}^{N} \phi_i + \sum_{k \neq k_0} \sum_{i=2}^{N+L} \xi_i \quad (9)$$

where

1. *ϕ_i is a random variable (r. v.) with mean $E(\phi_i) = 1 - 4\frac{2M^2+1}{3M^3}$ and variance $V(\phi_i) = 8\frac{2M^2+1}{3M^3}\left(1 - 2\frac{2M^2+1}{3M^3}\right)$, for all $i \leq N$.*
2. *ξ_i is another r. v. with mean $E(\xi_i) = 0$ and variance $V(\xi_i) = 8\frac{2M^2+1}{3M^3}$, for every i.*

The exact formula for D is given in the following lemma, by making use of this last result applied to Eq. (8).

Lemma 4. *For $N \geq 30$ and $Q \geq 1$,*

$$D = Q(N+3) + \Omega$$

where Ω is a Gaussian r. v. with mean

$$\mu = Q(N-1)(4\frac{2M^2+1}{3M^3} - 1)$$

and variance given by

$$\sigma^2 = 8Q\frac{2M^2+1}{3M^3}\left((N-1)(1 - 2\frac{2M^2+1}{3M^3}) + (N+2M-2)(K-1)\right)$$

This result allows us to calculate

$$P(D<0)=P(Q(N+3)+\Omega<0)=P(\Omega<-Q(N+3))=$$

$$=P\left(\frac{\Omega-\mu}{\sigma}<\frac{-Q(N+3)-\mu}{\sigma}\right)$$

Since $Z=\frac{\Omega-\mu}{\sigma}$ is a Gaussian with mean 0, and variance 1, there exists one unique $z_\alpha\in\mathbb{R}$ such that $P(Z<z_\alpha)=p_e$. For example, for $p_e=0.05$, it is $z_\alpha=-1.645$, and for $p_e=0.01$, it is $z_\alpha=-2.326$. By using that $P(Z<z_\alpha)=p_e=P(D<0)$, we arrive at

$$\frac{-Q(N+3)-\mu}{\sigma}=z_\alpha \tag{10}$$

The next step is to use the proper definition of the parameter of capacity α. It is the quotient between the number of patterns and the number of neurons which achieve an error probability lower than p_e. So, $\alpha=\frac{K}{N}$, that is, $K=\alpha N$.

By combining Eq. (10) and the above expression for K, we get the following result:

Theorem 2. *The capacity of the network as associative memory with multiple reference points is given by*

$$\alpha=\frac{1}{N}\left[1+\frac{1}{V}\left(\frac{T^2}{z_\alpha^2}-U\right)\right] \tag{11}$$

where

$$T=\sqrt{Q}\left((N+3)+(N-1)(4\frac{2M^2+1}{3M^3}-1)\right)$$

$$U=8(N-1)\frac{2M^2+1}{3M^3}(1-2\frac{2M^2+1}{3M^3})$$

and

$$V=8(N+2M-2)\frac{2M^2+1}{3M^3}$$

From this theorem, some important corollaries can be stated:

Corollary 1. *The capacity of the network is asymptotically increasing with the number of reference points.*

This result can be deduced from the fact that Eq. (11) can be rewritten in the following terms:

$$\alpha=R_0(M,N)+R_1(M,N)Q$$

with $R_1(M,N)>0$. So, given N and M, α is an increasing function for values of $Q\geq 1$. This implies that, by increasing the number of reference points, capacity greater than 1 may be achieved, as can be verified in Fig. 1. It must be remembered that the maximum capacity in BH was 1.

Corollary 2. *Given M and Q, there exists a positive constant $\alpha_{\min}$ such that $\alpha\geq\alpha_{\min}$ for all N.*

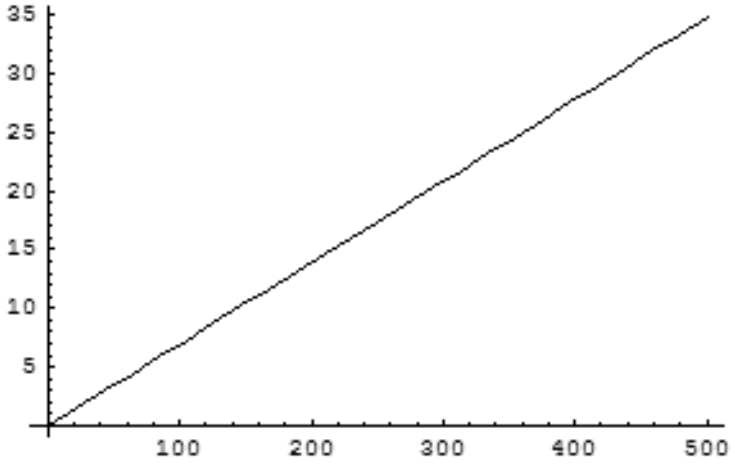

Fig. 1. Capacity of a network with $N = 40$ neurons and $M = 4$ states, as a function of Q, which varies from 1 to 500

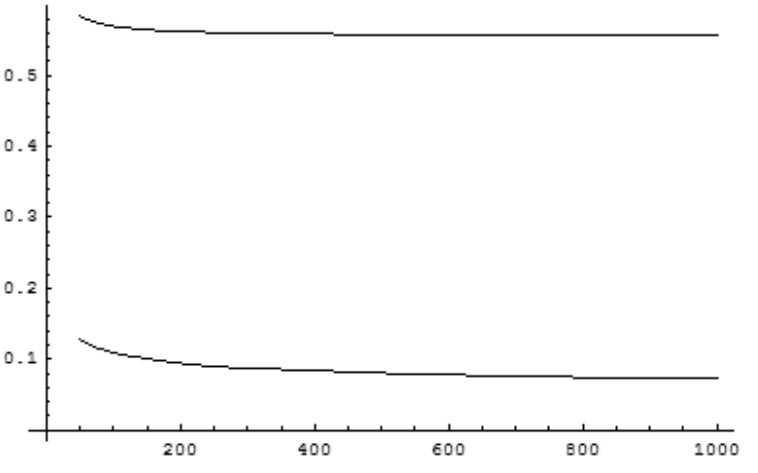

Fig. 2. Capacity of the bipolar network with $N \in \{50, \ldots, 1000\}$ neurons and $Q = 4$, compared with the bound given by McEliece (lower graph)

This corollary can be proved just by noting that the capacity is a decreasing function of N, the number of neurons, so we can fix M and Q and calculate the value $\alpha_{\min} = \lim_{N\to\infty} \alpha = \frac{2Q}{z_\alpha^2} \frac{2M^2+1}{3M^3}$. If we consider $M = 2$ and $Q = 2$ and $p_e = 0.01$, a value of $\alpha_{\min} = 0.2772$ is obtained, and for $p_e = 0.05$, a value of $\alpha_{\min} = 0.5543$. In Fig. 2, the value of α for a bipolar network is compared with the upper bound given by [10], and it is shown that $\alpha > \frac{1}{2 \log N}$, meaning a great improvement on the capacity of the

net, since this upper bound tends to 0 as N approaches ∞, and our technique ensures a minimum positive amount of capacity for the net.

6 Conclusions and Future Work

In this paper, an extension to Hopfield's associative memory has been studied to overcome some of the most important problems or lacks it possesses: spurious patterns and low capacity.

A method to avoid the apparition of spurious patterns has been presented. This method also explains the well-known (and undesirable) phenomenon of storing the opposite of a pattern.

A new technique to increase the network capacity as a content-addressable memory has also been proposed, based on the use of multiple reference points, which contributes many new possibilities of study and research.

Our future work covers several aspects of these methods:

- Find the optimal configuration of $O^{(q)}$ for a given set of patterns (randomly distributed or with a specific distribution), that is, the distribution of $O^{(q)}$ which discriminates most the patterns and makes the net achieve the maximum possible capacity.
- Consider a mix of fixed and random reference points.

References

1. Hopfield, J.: Neural networks and physical systems with emergent collective computational abilities. Volume 79. (1982) 2254 – 2558
2. Hebb, D.O.: The Organization of Behavior. Wiley (1949)
3. Parisi, G.: Asymmetric neural networks and the process of learning. J. Phys. A: Math. and Gen. **19** (1986) L675 – L680
4. Hertz, J.A., Grinstein, G., Solla, S.A.: Heidelberg Colloquium on Glassy Dynamics. In:. Springer-Verlag (1987)
5. Treves, A., Amit, D.J.: Metastable states in asymmetrically diluted hopfield networks. J. Phys A: Math. and Gen. **21** (1988) 3155 – 3169
6. Singh, M.P., Chengxiang, Z., Dasgupta, C.: Analytic study of the effects of synaptic asymmetry. Phys. Rev. E **52** (1995) 5261 – 5272
7. Athithan, G., Dasgupta, C.: On the problem of spurious patterns in neural associative models. IEEE Transactions on Neural Networks **8** (1997) 1483 – 1491
8. Crick, F., Mitchinson, G.: The function of dream sleep. Nature **304** (1983) 111 – 114
9. Hemmen, J.L.v., Kuhn, R.: Collective Phenomena in Neural Networks. Springer-Verlag (1991)
10. McEliece, R.J., Posner, E.C., Rodemich, E.R., Venkatesh, S.S.: The capacity of the hopfield associative memory. IEEE Transactions on Information Theory **IT-33** (1987) 461 – 482
11. Ho, C.Y., Sasase, I., Mori, S.: On the capacity of the hopfield associative memory. In: Proceedings of IJCNN 1992. (1992) II196 – II201
12. Kuh, A., Dickinson, B.W.: Information capacity of associative memory. IEEE Transactions on Information Theory **IT-35** (1989) 59 – 68
13. Mérida-Casermeiro, E., Muñoz Pérez, J.: Mrem: An associative autonomous recurrent network. Journal of Intelligent and Fuzzy Systems **12** (2002) 163 – 173
14. Mérida-Casermeiro, E., Galán-Marín, G., Muñoz Pérez, J.: An efficient multivalued hopfield network for the travelling salesman problem. Neural Processing Letters **14** (2001) 203 – 216

15. Erdem, M.H., Ozturk, Y.: A new family of multivalued networks. Neural Networks **9** (1996) 979 – 989
16. Ozturk, Y., Abut, H.: System of associative relationships (soar). (1997)
17. Mérida-Casermeiro, E., Muñoz Pérez, J., Domínguez-Merino, E.: An n-parallel multivalued network: Applications to the travelling salesman problem. Computational Methods in Neural Modelling, Lecture Notes in Computer Science **2686** (2003) 406 – 413
18. López-Rodríguez, D., Mérida-Casermeiro, E., Ortiz-de Lazcano-Lobato, J.M.: Hopfield network as associative memory with multiple reference points. In Ardil, C., ed.: International Enformatika Conference. (2005) 62 – 67