Storage Capacity of Kernel Associative Memories

B. Caputo and H. Niemann

Department of Computer Science, Chair for Pattern Recognition
University of Erlangen-Nuremberg,
Martenstrasse 3, 91058 Erlangen, Germany

Abstract. This contribution discusses the thermodynamic phases and storage capacity of an extension of the Hopfield-Little model of associative memory via kernel functions. The analysis is presented for the case of polynomial and Gaussian kernels in a replica symmetry ansatz. As a general result we found for both kernels that the storage capacity increases considerably compared to the Hopfield-Little model.

1 Introduction

Learning and recognition in the context of neural networks is an intensively studied field. A lot of work has been done on networks in which learning is Hebbian, and recognition is represented by attractor dynamics of the network. Particularly, the Hopfield-Little model (H-L, [6],[7]) is a network exhibiting associative memory based on the Hamiltonian

$$H = -\frac{1}{2N}\sum_{\mu=1}^{M}\sum_{i\neq j}^{N} s_i s_j \xi_i^{(\mu)}\xi_j^{(\mu)}, \tag{1}$$

where the s_i are N dynamical variables taking on the values ± 1 and the $\xi_i^{(\mu)}$ (with $\xi_i^{(\mu)} \pm 1$) are M fixed patterns which are the memories being stored. The storage capacity and thermodynamic properties of this model have been studied in detail within the context of spin glass theory [3]. Many authors studied generalizations of the H-L model which include interactions between p (> 2) Ising spins ([1],[5]), excluding all terms with at least two indices equal (symmetric terms). As a general result, higher order Hamiltonians present an increase in the storage capacity compared to the H-L model.

It was pointed out in [4] that if the symmetric term is included in the Hamiltonian (1), it can be written as a function of the scalar product between $\boldsymbol{\xi}^{(\mu)}$ and $\mathbf{s}$. The Euclidean scalar product can thus be substituted by a Mercer kernel [9], providing a new higher order generalization of the H-L energy. We call this new model Kernel Associative Memory (KAM). This new energy was used in [4] within a Markov Random Field framework for statistical modeling purposes. There are several reasons for considering this generalization. First, we will show

J.R. Dorronsoro (Ed.): ICANN 2002, LNCS 2415, pp. 51–56, 2002.

in this paper that this model presents a higher storage capacity with respect to the H-L model. Second, the higher order generalization of the H-L model via Mercer kernel gives the possibility to study a much richer class of models, due to the variety of possible Mercer kernels [9]. Here we study the storage capacity and thermodynamic properties of KAM for polynomial and Gaussian kernels, in a replica symmetry ansatz. To our knowledge, no previous works has considered the storage capacity of such a generalization of the H-L model.
The paper is organized as follows: Section 2 presents KAM; in Section 3 we compute the free energy and the order parameters within a replica symmetry ansatz, and in Section 4 we study in detail the zero temperature limit. The paper concludes with a summary discussion.

2 Kernel Associative Memories

The H-L energy (1) can be rewritten in the equivalent form

$$H = -\frac{1}{2}\sum_{\mu=1}^{M}\left[\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} s_i \xi_i^{(\mu)}\right]^2 - 1\right]. \tag{2}$$

This energy can be generalized to higher order correlations via Mercer kernels [9]:

$$H_{KAM} = -\frac{1}{2}\sum_{\mu=1}^{M} K\left(\frac{1}{\sqrt{N}}\mathbf{s}, \boldsymbol{\xi}^{(\mu)}\right). \tag{3}$$

The possibility to kernelize the H-L energy function was recognized first in [4]. We call this model Kernel Associative Memory (KAM). It is fully specified once the functional form of the kernel is given. In this paper we will consider *polynomial* and *Gaussian kernels*:

$$K_{poly}(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}\cdot\boldsymbol{y})^p, \qquad K_{Gauss}(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\rho||\boldsymbol{x}-\boldsymbol{y}||^2). \tag{4}$$

Our goal is to study the storage capacity of energy (3) for kernels (4), using tools of statistical mechanics of spin glasses. To this purpose, we note that the study of energy (3) can be done for both kernels (4) considering the general case

$$H = -\frac{1}{N^{1-p/2}}\sum_{\mu=1}^{M} F\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} s_i \xi_i^{(\mu)}\right), \qquad F(x) = A_p x^p. \tag{5}$$

For $A_p = 1$, p finite, equations (5) represent the KAM energy (3) for polynomial kernels (4). For $A_p = \rho^p/p!, p \to \infty$, equations (5) describe the behavior of the KAM energy (3) for Gaussian kernels (4). This can be shown as follows: note first that

$$\exp(-\rho||\boldsymbol{s}-\boldsymbol{\xi}^{(\mu)}||^2) = \exp(-\rho[\boldsymbol{s}\cdot\boldsymbol{s}+\boldsymbol{\xi}^{(\mu)}\cdot\boldsymbol{\xi}^{(\mu)} - 2\boldsymbol{s}\cdot\boldsymbol{\xi}^{(\mu)}]) = \exp(-2\rho[N-\boldsymbol{s}\cdot\boldsymbol{\xi}^{(\mu)}]).$$

The multiplicating factor can be inglobated in the ρ, and the constant factor can be neglected. The Gaussian kernel function thus becomes

$$\exp(-\rho||\boldsymbol{s}-\boldsymbol{\xi}^{(\mu)}||^2) \longrightarrow \exp(\rho[\boldsymbol{s}\cdot\boldsymbol{\xi}^{(\mu)}]) \simeq 1+\rho\boldsymbol{s}\boldsymbol{\xi}^{(\mu)}+\frac{\rho^2}{2!}(\boldsymbol{s}\boldsymbol{\xi}^{(\mu)})^2+\frac{\rho^3}{3!}(\boldsymbol{s}\boldsymbol{\xi}^{(\mu)})^3+\dots$$

A generalization of the H-L energy function in the form (5) was proposed first by Abbott in [1]. In that paper, storage capacity and thermodynamic properties were derived for a simplified version of (5), where all terms with at least two equal indices were excluded; a particular choice of A_p was done. Other authors considered this kind of simplifications for higher order extension of the H-L models (see for instance [5]). The analysis we present here include all the terms in the energy and is not limited to a particular choice of the coefficient A_p; thus is more general. To the best of our knowledge, this is the first analysis on the storage capacity and thermodynamic properties of a generalization of the H-L model via kernel functions.

3 Free Energy and Order Parameters

We study the overlap of a configuration s_i with one of the stored patterns, arbitrarily taken to be $\xi_i^{(1)}$,

$$m = \left\langle\left\langle \frac{1}{N}\sum_i^N \langle s_i\rangle \xi_i^{(1)} \right\rangle\right\rangle, \tag{6}$$

where the angle bracket $\langle\dots\rangle$ represents a thermodynamic average while the double brackets $\langle\langle\dots\rangle\rangle$ represents a quenched average over the stored patterns $\xi_i^{(\mu)}$. The quenched average over patterns is done using the replica methods [8]; in the mean-field approximation, the free energy depends on m and on the order parameters

$$q_{ab} = \left\langle\left\langle \frac{1}{N}\sum_i^N \langle s_i^a\rangle\langle s_i^b\rangle \right\rangle\right\rangle, r_{ab} = \frac{1}{MN}\sum_{\mu=2}^{M}\left\langle\left\langle \sum_i^N \langle s_i^a\rangle \xi_i^{(1)} \sum_j^N \langle s_j^b\rangle \xi_j^{(1)} \right\rangle\right\rangle, \tag{7}$$

with a, b replica indices. The calculation is analog to that done in [1],[2]. Considering a *replica symmetry ansatz* [8], $q_{ab} = q, r_{ab} = r$ and the free energy at temperature $T = 1/\beta$ is given by

$$f = (p-1)A_p m^p + \frac{\alpha\beta}{2}[r(1-q) - G(q)] - \frac{1}{\beta}\int Dz \ln[2\cosh\beta(\sqrt{\alpha r}z) + pA_p m^{p-1}], \tag{8}$$

with $M = \alpha N^{p-1}$ and $Dz = \frac{dz}{\sqrt{2\pi}}e^{-z^2/2}$. The function $G(q)$ is given by

$$G(q) = A_p^2 \sum_{c=1}^{p} \frac{1}{2}[1 + (-1)^{p+c}]\frac{1}{c!}\left[\frac{p!}{(p-c)!!}\right]^2 (1 - q^c) \tag{9}$$

The free energy leads to the following *order parameters*:

$$m = \int Dz \tanh \beta(\sqrt{\alpha r} z + pA_p m^{p-1}), r = -\frac{\partial G(q)}{\partial q}, q = \int Dz \tanh^2 \beta(\sqrt{\alpha r} z + pA_p m^{p-1}) \quad (10)$$

At all temperatures, these equations have a paramagnetic solution $r = q = m = 0$.

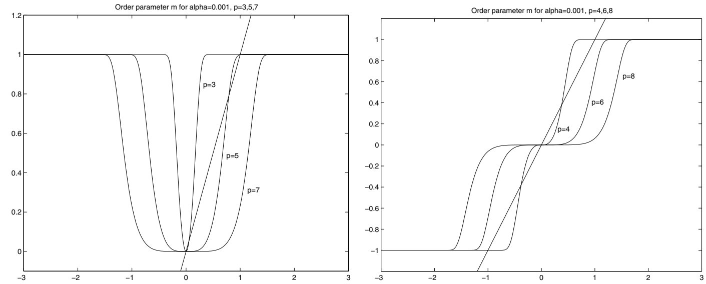

Fig. 1. Graphical representation of the solutions of the retrieval equation for $p = 3, 5, 7$ (left) and $p = 4, 6, 8$ (right), $\alpha = 0.001$.

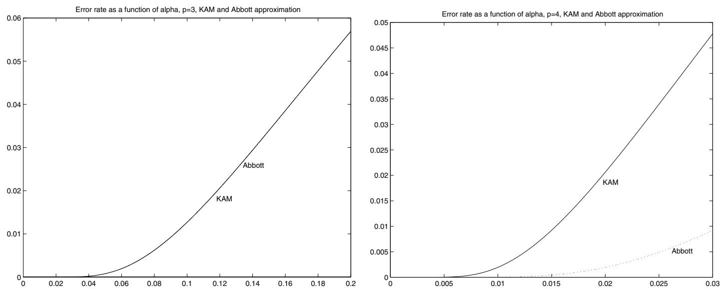

Fig. 2. The error rate $(1 - m)/2$ as a function of α, for KAM and Abbott model, for $p = 3$ (left) and $p = 5$ (right). For $p = 3$ the two curves are indistinguishable, for $p = 5$ the Abbott approximation gives the lower error rate.

4 The Zero–Temperature Limit: $\beta \to \infty$

In the zero-temperature limit $\beta \to \infty$, equations (10) simplifies to $q \to 1$,

$$m \to \mathrm{erf}\left(\frac{pA_p m^{p-1}}{\sqrt{\alpha r}}\right), \qquad r \to \frac{1}{2} A_p^2 \sum_{c=1}^{p} R_c(p) \quad (11)$$

with

$$R_c(p) = [1 + (-1)^{p+c}] \frac{c}{c!} \left[\frac{p!}{(p-c)!!} \right]^2 \tag{12}$$

Note that, in this approximation, m does not depend on A_p:

$$m = \text{erf}\left(\frac{pA_p m^{p-1}}{\sqrt{\alpha r}} \right) = \text{erf}\left(\frac{pm^{p-1}}{\sqrt{\alpha}} \frac{A_p}{A_p \sqrt{\sum_{c=1}^{p} R_c(p)/2}} \right). \tag{13}$$

Graphical solutions of equation (13) are shown in Figure 1, for several values of p (odd and even). From these graphics we can make the following considerations: first, higher order Hamiltonians (5) with p odd eliminate the symmetry between the stored memory and their complementary states, in which the state of each neuron is reversed (see Figure 1). Second, as p increases, the value of α for which there is more than one intersection ($m = 0$, the spin-glass solution), as to say the storage capacity, goes to zero (see Figure 1). Figure 2 shows the percentage of errors $(1 - m)/2$ made in recovering a particular memory configuration as a function of α. The percentage of errors for the KAM model, for $p = 3, 5$, is compared with the same quantity obtained by Abbott [1], considering a higher order Hamiltonian. In that case m is given by:

$$m^{Abbott} = \text{erf}\left[\left(\frac{p}{2\alpha p!} \right)^2 m^{p-1} \right] \tag{14}$$

in the zero temperature limit. In both cases (p=3,5) and for both models (KAM and Abbott), the overlap with the input pattern m remains quite close to 1 even for $\alpha \to \alpha_c$. This can be seen because the fraction of errors is always small. Thus the quality of recall is good. Nevertheless, the fraction of errors is smaller for the Abbott model as p increase.

Even if $\alpha_c \to 0$ for large p, the total number of memory states allowed is given by $M = \alpha N^{p-1}$, thus it is expected to be large. In the limit $p \to \infty$, α_c and M can be calculated explicitly [2]: for large values of the argument of the erf function, it holds

$$m \approx 1 - \frac{\sqrt{\alpha/2 \sum_{c=1}^{p} R_c(p)}}{\sqrt{\pi} p m^{p-1}} \exp\left[-\frac{p^2 m^{p-1}}{\alpha/2 \sum_{c=1}^{p} R_c(p)} \right]. \tag{15}$$

For stability considerations, $m \approx 1$, thus the second term in equation (15) must go to zero. This leads to the critical value for α

$$\alpha_c = \frac{2p^2}{[\sum_{c=1}^{p} R_c(p)] \ln p} \approx \frac{2p^2}{p! 2^p \ln p \,[\sum_{c=1}^{p} c(1 + (-1)^{c+p})]}. \tag{16}$$

The total number of memory states that is possible to store is (using Stirling's approximation):

$$M = \alpha N^{p-1} \to \left(\frac{eN}{p} \right)^{p-1} \sqrt{\frac{2p}{\pi}} \frac{e}{2^p \ln p \sum_{c=1}^{p} (1 + (-1)^{c+p})}. \tag{17}$$

This result must be compared with the one obtained by Abbott [1]: $\alpha_c^{Abbott} = \frac{p}{2p!lnp}$; it is easy to see that $\alpha_c = \alpha_c^{Abbott} \cdot \frac{p}{2^p \sum_{c=1}^{p} c[1+(-1)^{c+p}]}$; thus, the introduction of terms in which more than one indices is equal in the generalized Hamiltonian leads to a decrease in the storage capacity; the decrease will be higher the higher is the number of these terms introduced in the energy formulations, according with the retrieval behavior (see Figure 2).

5 Summary

In this paper we presented kernel associative memories as a higher order generalization of the H-L model. The storage capacity of the new model is studied in a replica symmetry ansatz. The main result is that the storage capacity is higher than the H-L's one, but lower than the storage capacity obtained by other authors for different higher order generalizations. This work can be developed in many ways: first of all, the mean field calculation presented here is only sensitive to states that are stable in the thermodynamic limit. We expect to find in simulations correlations between spurious states and the memory patterns even for $\alpha > \alpha_c$. Simulations should also provide informations about the size of the basin of attraction. Second, replica symmetry breaking effects should be taken in account. Third, it should be explored how kernel properties can be used in order to reduce the memory required for storing the interaction matrix [5]. Finally, this study should be extended to other classes of kernels and to networks with continuous neurons. Future work will be concentrated in these directions.

References

1. L. F. Abbot, Y. Arian, "Storage capacity of generalized networks", *Physical Review A*, 36 (10), pp 5091-5094, November 1987.
2. D. J. Amit, *"Modeling Brain Function"*, Cambridge University Press, Cambridge, USA, 1989.
3. D. J. Amit, H. Gutfreund, H. Sompolinsky, "Spin glass models of neural networks", *Physical Review A*, 32, 1007, 1985.
4. B. Caputo, H. Niemann, "From Markov Random Fields to Associative Memories and Back: Spin Glass Markov Random Fields", *Proc. of IEEE workshop on Statistical and Computational Theories of Vision*, Vancouver, CA, July 2001.
5. E. Gardner, "Multiconnected neural network models", *Journal of Physics A*, 20, pp 3453-3464, 1987.
6. 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.
7. W. A. Little, G. L. Shaw, "Analytical study of the memory storage capacity of a neural network", *Math. Biosci.*, 39, 281, 1981.
8. M. Mezard, G. Parisi, M. Virasoro, *Spin Glass Theory and Beyond*, World Scientific, Singapore, 1987.
9. V. Vapnik, *Statistical learning theory*, J. Wiley, New York, 1998.