PHASE-SPACE GARDENING IN THE BINARY-COUPLINGS MEMORY NETWORK

Norman Hendrich
University of Hamburg · Department of Computer Science · D-2000 Hamburg 54
e-mail: hendrich@informatik.uni-hamburg.de

Abstract

The dynamics of the binary-couplings Hopfield/Gardner associative memory network is studied. A iterative learning rule is presented that allows to adjust the stabilities for each pattern at each neuron individually.

Simulations of the resulting network show that the basins of attraction of the patterns can be shaped as desired. The dependence of m_c on the stability κ is shown.

1. Introduction

Attractor neural networks of the Hopfield/Gardner type are interesting models of associative memory and they have been studied intensively during the past few years. The basic idea is to store $P = \alpha N$ binary patterns $\xi_i^\mu = \pm 1$, $\mu = 1, \ldots, P$, $i = 1, \ldots, N$ as the fixed points of the dynamics of a network with N neurons S_i. Usually a parallel dynamics is used,

$$S_i(t+1) = \operatorname{sgn}\Big(||J_{ij}||^{-1} \sum_j J_{ij} S_j(t) \Big),$$

and therefore the patterns are stored if $\xi_i^\mu = \operatorname{sgn}(||J_{ij}||^{-1} \sum_{j \neq i} J_{ij} \xi_j^\mu)$. To ensure finite basins of attraction a positive stability κ is introduced and the synaptic couplings J_{ij} are chosen to fulfil the stronger constraints

$$\kappa_{i\mu} = \xi_i^\mu \cdot h_{i\mu} = \xi_i^\mu \cdot \Big(||J_{ij}||^{-1} \sum_j J_{ij} \xi_j^\mu \Big) \geq \kappa > 0.$$

Many aspects of the networks are quite well understood by now. Most notably, the calculation of the optimal storage capacity (number P of patterns that can be stored) was possible for a variety of models [2], [3].

Recently several authors concentrated on the dynamics of the networks. The basic question concerns the size of the basins of attraction of the patterns stored. Due to the inherent complexity of an analytical treatment, most studies have relied on simulations. Only in the special case of strongly diluted systems, with less than $O(\ln N)$ couplings per neuron, an analytical solution has been obtained [4], and the basins of attraction were found to be optimal for storage ratios $\alpha \leq 0.41$ in the spherical model [4], [5]. A similar model was developed for fully connected models in [8].

2. The binary-couplings associative memory network

The storage capacity of the binary-couplings network ($J_{ij} = \pm 1$) was studied in [3], [11], [12] using the replica and zero-entropy methods and by numerical simulations and was found to be $\alpha_B \approx 0.83$.

From the viewpoint of information theory this result is very encouraging, as the relative information content of the network is finite — $0.83 \cdot N^2$ bits of information stored using a coupling matrix of N^2 bits. Note that the Hopfield model ($\alpha_H \approx 0.14$) and the spherical model ($\alpha_c = 2$) have vanishing

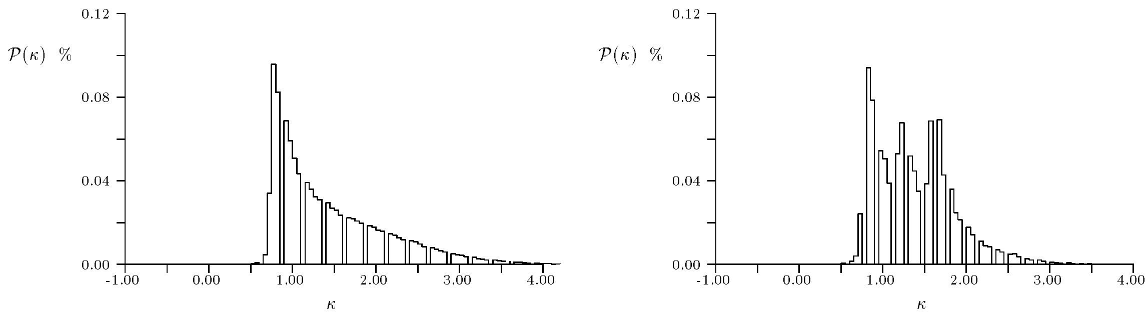

Figure 1: (a) Distribution of stabilities $\mathcal{P}(\kappa_{i\mu})$ at $\alpha = 0.20$ ($N = 1024$). (b) Distribution of stabilities at $\alpha = 0.15$, with desired stabilities $\kappa = 1.0, 1.5, 2.0$ ($N = 512$).

relative information content in the limit $N \to \infty$, because the synaptic values need to be encoded using $O(\ln N)$ bits.

However, while good learning algorithms are known for the spherical model, e. g. [10], the development of learning rules for the binary-couplings model proved to be very hard. The complete enumeration algorithm used to numerically determine the storage capacity α_B is obviously impractical even for very small networks.

Recently a new learning rule was introduced in [13] that starts with unconstrained couplings but restricts the values as the algorithm proceeds. This allows standard optimization methods to be used for the discrete optimization problem. Extrapolation shows that the algorithm succeeds to learn the given patterns for up to at least $\alpha \approx 0.5$.

A very simple but effective learning rule for the binary-couplings network was proposed by [9]. Their algorithm randomly selects a coupling at neuron S_i and adjusts the value of J_{ij} to minimize the cost-function $E(\kappa, J_i) = \sum_{\mu=1}^{P} (\kappa - \kappa_{i\mu})^2$. This is iterated until the energy E reaches a (local) minimum. Application of this algorithm results in a critical storage capacity of $\alpha_B \approx 0.4$.

In this paper a straightforward modification of the above rule was used that tries to minimize the cost-function

$$E(\kappa, J_i) = \sum_{\mu=1}^{P} \Theta(\kappa - \kappa_{i\mu}) \cdot (\kappa - \kappa_{i\mu}).$$

While this algorithm is less suitable for analytical studies, it tries to install minimum stabilities of at least κ (and it is the best candidate for hardware implementation of binary-couplings ANNs). The algorithm gives $\alpha_B \approx 0.4$.

An example of the distribution of stabilities $\mathcal{P}(\kappa_{i\mu})$ after learning is shown in figure 1a, where it is plotted for a network with $\alpha = 0.20$ and $\kappa = 0.75$ ($N = 1024$). The learning rule succeeds to install almost all stabilities $\kappa_{i\mu} > \kappa$ with an (exponential) tail of larger stabilities.

3. Shaping the basins of attraction

In the simulations presented here, the stabilities of all patterns were adjusted individually during learning. We use the symbol $\lambda_{i\mu}$ for the desired stability of pattern ξ_i^μ at neuron S_i and the cost function in the learning algorithm is further modified to $E(\{\lambda\}, J_i) = \sum_{\mu=1}^{P} \Theta(\lambda_{i\mu} - \kappa_{i\mu}) \cdot (\lambda_{i\mu} - \kappa_{i\mu})$.

In a first set of simulations the patterns were seperated into groups with identical stabilities. For example, figure 1b shows the distribution of stabilities after learning in a network with $\alpha = 0.15$, where λ_μ was set to 2.0, 1.5 and 1.0 for 20%, 20% and 60% of all patterns respectively ($N = 512$).

The simulations to determine the basins of attraction follow the basic steps presented in [1]. First $P = \alpha \cdot N$ uncorrelated random patterns ξ_i^μ are generated and the learning algorithm is applied until all patterns are stabilized.

Then random test patterns $\xi_i^{\mu,r}$ with initial overlap $m_0 = (1/N) \sum_i \xi_i^\mu \xi_i^{\mu,r}$ with pattern ξ_i^μ are generated and iterated under the dynamics until the network state stabilizes or reaches a cycle, when the final overlap m_f and other values are recorded. The average over many test patterns allows to

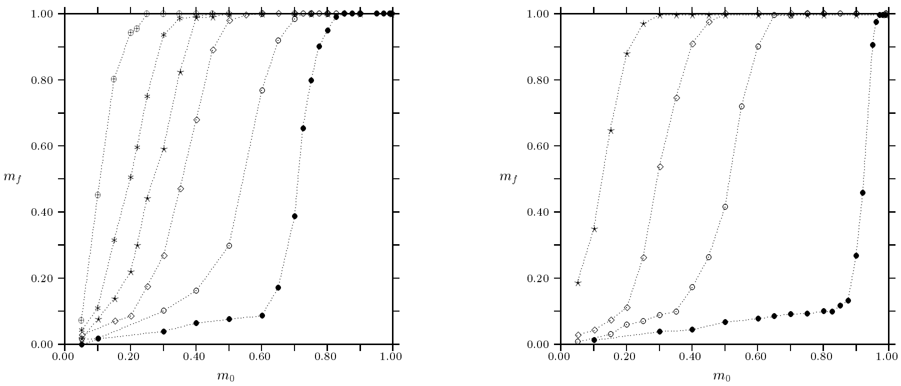

Figure 2: (a) Mean final overlap $m_f(m_0)$ at $\alpha = 0.15$. Stabilities are $\oplus$ 2.1, $*$ 1.8, $\star$ 1.6, $\diamond$ 1.4, $\circ$ 1.1 and $\bullet$ 0.9 (b) Mean final overlap $m_f(m_0)$ at $\alpha = 0.25$. Stabilities are $\star$ 2.3, $\diamond$ 1.8, $\circ$ 1.4 and $\bullet$ 0.7 ($N = 512$).

calculate the desired quantities, e. g. the values of $m_f(m_i)$, the number of network iterations to stabilize, etc.

Typical results for the recall properties of the networks are shown in figure 2. The left diagram shows the mean final overlap m_f as a function of initial overlap m_0 in a network with $\alpha = 0.15$ and desired stabilities $\lambda_\mu = 2.1$, 1.8, 1.6, 1.4, 1.1 and 0.9 (for 10%, 10%, 10%, 10%, 40% and 20% of patterns). The right diagram shows $mf(m_0)$ in a network with $\alpha = 0.25$ and desired stabilities $\lambda_\mu = 2.3$, 1.8, 1.4, 0.7 (for 10%, 10%, 10% and 70% of patterns).

It is obviously possible to adjust the size of the basins of attraction by this method. In the same network patterns with large stabilities $\lambda_\mu > 1.0$ are recalled within very large basins of attraction, while patterns with stabilities smaller than $\lambda_\mu \approx 0.8$ have tiny basins of attraction. A detailed study of this phase-space gardening for both the spherical and the binary-couplings model will be presented elsewhere [6].

4. The network with linear distribution of stabilities

To further illustrate the dependence of the recall of a pattern on its stability, another simple choice for the distribution of stabilities, the linear function

$$\lambda_{i\mu} = \mu \cdot (\kappa_{\max}/N)$$

was used. That is, a pattern ξ_i^μ is stabilized proportional to its index μ and independent of the neuron i. After learning the functions $m_f(m_0)$ and $f_p(m_0)$ were determined for all patterns ξ_i^μ.

To present the results in a simple form, the size of the basin of attraction of a pattern is defined as the value m_c of initial overlap which guarantees a final overlap of larger than $m_f > 0.97$. Alternative definitions of m_c are possible, e. g. in [1] the critical initial overlap is defined by $f_p(m_c) = 0.5$.

Figure 3 shows the result of plotting the size of the basins of attraction of a pattern, m_c, versus its stability κ_μ for storage ratios $\alpha = 0.10$, 0.15, 0.20, and 0.25 ($N = 512$).

Note that no patterns are recalled (with $m_f > 0.97$) if their stability is less than 0.80 and that this value seems to independent of α. The dependence of the size of the basins of attraction m_c on λ_μ at fixed α appears to be linear up to quite large basins of attraction (i.e. small m_c).

5. Summary

A simple iterative learning rule for the binary-couplings attractor neural network was presented that allows to adjust the desired stabilities $\lambda_{i\mu}$ for each pattern and neuron independently up to a storage ratio $\alpha \approx 0.4$. Two different choices of the distribution of desired stabilities were studied in this paper.

It was shown that the learning algorithm allows to decide the size of the basins of attraction of the individual patterns. While only uncorrelated patterns were presented here, the generalization to

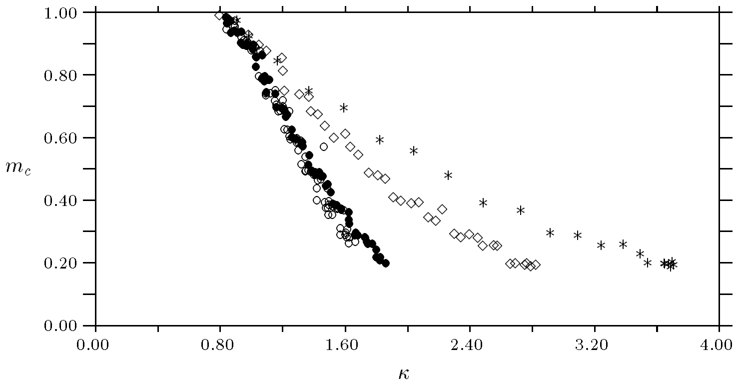

Figure 3: Critical initial overlap m_c as a function of λ_μ in networks with linear distribution of λ_μ. Storage ratios $\alpha = *\ 0.10$, $\diamond\ 0.15$, $\bullet\ 0.20$, and $\circ\ 0.25$.

correlated patterns and distributions where a pattern is stored more strongly on some neurons than others (leading to anisotropic basins of attraction) is straightforward.

The possibility to shape the basins of attraction of the stored information patterns makes the networks more plausible as models of human associative memory. It might even lead to new applications, as conventional RAM-based associative memories lack a storage-efficient means to adjust the recall of the stored patterns [7].

References

[1] B. M. Forrest, *Journal of Physics* A 21, 245-255 (1988).

[2] E. Gardner, *Europhysics Letters* 4, 481–485 (1987).

[3] E. Gardner, B. Derrida, *Journal of Physics* A 21, 271–284 (1988).

[4] E. Gardner, *Journal of Physics* A 22, 1969–1974 (1989).

[5] N. Hendrich, *Journal of Physics* A 24, 2877–2887 (1991).

[6] N. Hendrich, submitted to *Network* (1993).

[7] N. Hendrich, Information theoretical aspects of attractor networks, in preparation (1993).

[8] T. B. Kepler and L. F. Abbott, *Journal de Physique (France)* 49, 1657–1662 (1988).

[9] H. Koehler, S. Diederich, W. Kinzel and M. Opper, *Z. Physik B* 78, 333.

[10] W. Krauth and M. Mézard, *Journal of Physics* A 20, L745–L752 (1987).

[11] W. Krauth and M. Mézard, *Journal de Physique (France)* 50, 3057–3066 (1989).

[12] W. Krauth and M. Opper, *Journal of Physics* A 22, L519–L523 (1989).

[13] C. J. Pérez, J. Carrabina and E. Valderrama, *Network* 3, 165–176 (1992).