Convergence Results in an Associative Memory Model

János Komlós Ramamohan Paturi
University of California, San Diego
La Jolla, CA 92093

In this paper, we consider Hopfield's [1] model of associative memory. This model consists of a system of fully interconnected neurons or threshold elements where each interconnection is symmetric and has certain weight w_{ij}. Each neuron can be in one of two states ± 1. Each neuron updates its state based on the weighted sum of the states of the other neurons. If this sum exceeds a certain threshold associated with the neuron, the state is set to $+1$. Otherwise, it is set to -1. This updating can take place in one of two modes: *synchronous*, or *asynchronous*. The state of the entire system is represented by an n-dimensional vector, where n is the number of neurons in the system. A state is called *stable* if no transition out of it is possible. We define the *energy* of the system in state x as $-1/2 \sum_{i,j} w_{ij} x_i x_j$. In this energy landscape, stable states correspond to local minima.

The items to be stored in the memory can be represented by certain states of the system. Given a set of states to be stored, one can, by an appropriate selection of the weights of the interconnections, create a *sphere of attraction* around each of the states such that any state in this sphere would eventually come closer to the center. It is this *error-correcting* behavior that gives the system associative memory character.

The model uses Hebb's rule to select the weights of the interconnections. According to this rule, we set the weights $w_{ij} = v_i v_j \quad i \neq j$, to remember a single vector $v = (v_1, v_2, \cdots, v_n)$. To store several vectors in the system, we simply add the corresponding weights together. We call each such stored vector a *fundamental* memory.

When we store a number of fundamental memories in the system, we expect each of them to be stable and to attract *all* the vectors within a ρn distance for some constant $\rho > 0$. Or more generally, we consider the system to be error-correcting, if every vector within a distance ρn from a fundamental memory eventually ends up within a distance of εn for some $\varepsilon < \rho$. We call this εn *residual error.* We are also concerned with the time it takes for the error-correction.

A reasonable minimal storage requirement is that we would like to store almost all sets of m vectors. Therefore, we will take a set of m *random vectors* as our set of fundamental memories, and expect the system to remember them with *probability near 1.*

When we have a number of fundamental memories, the retrieval of a memory will be disturbed by the *noise* created by the other fundamental memories. Yet, we hope that this noise is not overwhelming when the number of fundamental memories is not too large. Hence, the main question is to determine the amount of error-correction and the rate of convergence as a function of the number of fundamental memories, m.

Results The main mathematical difficulty is to show that the probability that noise is high is exponentially small. Exponentially small bounds on the probability are necessary to account for the exponential number of possibilities that arise when we want to attract all the vectors in a sphere of radius ρn. To this end, we prove the One-step Error-correction Lemma which gives a quantitative picture of the error-correcting behavior of the model (see [2]). The following is an informal version of the lemma.

Write $\alpha = m/n$. There are constants α_s and ρ_s such that for $\alpha \leq \alpha_s$ the following holds with probability near 1. If the system is started at any state at a distance ρn ($\rho \leq \rho_s$) from a fundamental memory (ρn 'errors'), then it will correct most of the ρn errors in one synchronous step, and be at a much smaller distance $\rho' n$ from the fundamental memory. This ρ' is about ρ^3 if $\rho > \alpha$, and about $\alpha\rho^2$ if $\rho < \alpha$.

All our results basically follow from this lemma. Here, we give a brief summary of our results.

α_s, α_a, ρ_s, ρ_a, ρ_b are absolute constants. All the statements hold with probability approaching 1 as $n \to \infty$.

- If $m \leq \alpha_s n$, and if the system is started within a distance of $\rho_s n$ from a fundamental memory, then, in about $\log(n/m)$ synchronous steps, it will end up within a distance $ne^{-n/4m}$ from the fundamental memory, that is, it will eventually get within a distance $ne^{-n/4m}$ of the fundamental memory and remains within that distance.

  When $m < n/(4 \log n)$, the system will *converge* to the fundamental memory in $O(\log \log n)$ synchronous steps.

- In the asynchronous case, if $m \leq \alpha_a n$, and if the system is started within a distance of ρ_a from a fundamental memory, then it will *converge* to a stable state within a distance of $ne^{-n/4m}$ from the fundamental memory.

  In particular, when $m < n/(4 \log n)$, the system will converge to the fundamental memory.

- For any fundamental memory v, the maximum energy of any state *within* a distance of $\rho_a n$ from v is less than the minimum energy of any state *at* a distance of $\rho_b n$ from v, and there are no stable states in the annuli defined by the radii $\rho_b n$ and $ne^{-n/4m}$ centered at the fundamental memories.

We also prove an exponential(in m) lower bound on the number of stable states in the system. Details can be found in [2].

For a general treatment of the model where the interconnections can be arbitrary, we refer the reader to our paper [3].

References

[1] Hopfield, J. J. (1982). Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proc. Natl. Acad. Sci. USA* **79** 2554–2558.

[2] Komlós, J. and Paturi, R. (1988). Convergence Results in an Associative Memory Model. *Neural Networks*, Vol. 1.

[3] Komlós, J. and Paturi, R. (1988). Effect of Connectivity in an Associative Memory Model. *IEEE Symposium on Foundations of Computer Science.*