ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 27-32

Parallel dynamics of extremely diluted neural networks

D. Bollé , B. Vinck and V. A. Zagrebnov

Instituut voor Theoretische Fysica K.U. Leuven B-3001 Leuven, Belgium

Abstract. Using a probabilistic approach the parallel dynamics of neural networks with graded-response or analog neurons is studied at zero and at non-zero temperatures. Recursion relations are derived for the exact parallel dynamics of the extremely diluted asymmetric versions of these networks. An explicit analysis, including dynamical capacity-temperature diagrams and a study of retrieval performance in terms of the hamming distance, is carried out for the $Q = 3$-Ising model and for the piecewise linear model [1].

1. Motivation

In general, the aim of this paper is to study the ability of storage and retrieval of gray toned patterns in multi-state neural networks. More particularly we are interested in the dynamical performance of these networks measured in terms of the hamming distance.

2. The Model

Consider a network Λ of N neurons which can take values in the set $\mathcal{S} = \{-1 = s_1 < s_2 < ... < s_{Q-1} < s_Q = +1\}$. The p patterns to be stored in this network, $\{\xi_i^\mu \in \mathcal{S}\}, i \in \Lambda = \{1, 2, \ldots, N\}, \mu \in \mathcal{P} = \{1, 2, \ldots, p\}$ are supposed to be i.i.d.r.v. with zero mean, $E[\xi_i^\mu] = 0$, and variance $A = Var[\xi_i^\mu]$. The latter is a measure for the activity of the patterns.

Given a configuration $\boldsymbol{\sigma}_\Lambda = \{\sigma_j\}, j \in \Lambda$, the local field h_i in neuron $i \in \Lambda$ is

$$h_i(\boldsymbol{\sigma}_\Lambda) = \sum_{j \in \Lambda \setminus i} J_{ij} \sigma_j$$

where the synaptic couplings are given by

$$J_{ij} = \frac{1}{AN} \sum_{\mu \in \mathcal{P}} \xi_i^\mu \xi_j^\mu .$$

The zero temperature parallel dynamics is defined by a gain function g(.)

$$\sigma_i(t+1) = \mathrm{g}[h_i(\boldsymbol{\sigma}_{\Lambda \setminus i}(t))].$$

ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 27-32

At non zero temperature $\beta = T^{-1}$ the parallel dynamics of this network is defined by the transition probabilities

$$\Pr\{\sigma_i(t+1) = s_k \in \mathcal{S}|\sigma_{\Lambda\setminus i}(t)\} = \frac{\exp[-\beta\epsilon_i(s_k|\sigma_{\Lambda\setminus i}(t))]}{\sum_{s\in\mathcal{S}}\exp[-\beta\epsilon_i(s|\sigma_{\Lambda\setminus i}(t))]}$$

where the energy potential $\epsilon_i(s|\sigma)$ of neuron i is taken to be [2]

$$\epsilon_i(s|\sigma_{\Lambda\setminus i}) = -\frac{1}{2}(h_i(\sigma_{\Lambda\setminus i})s - bs^2) \ , b > 0.$$

3. Methods and Techniques

The retrieval quality of the network state at time t can be measured by the so called *hamming distances* which are nothing but the euclidean distances between the network state and the stored patterns :

$$d_H^\nu(\sigma(t)) = \frac{1}{N}\sum_{i\in\Lambda}(\sigma_i(t) - \xi_i^\nu)^2.$$

Observe that these hamming distances can be rewritten in terms of more familiar macroscopic quantities, viz. the overlaps and the network activity

$$m_\Lambda^\nu(t) = \frac{1}{AN}\sum_{i\in\Lambda}\xi_i^\nu\sigma_i(t)$$

$$a_\Lambda(t) = \frac{1}{N}\sum_{i\in\Lambda}(\sigma_i(t))^2.$$

We have

$$d_H^\nu(\sigma(t)) = A + a_\Lambda(t) - 2Am_\Lambda^\nu(t).$$

Assume now that initially the network configuration is correlated with only one stored pattern. In other words the set $\{\sigma_i(0)\}, i \in \Lambda$ is a collection of i.i.d.r.v. on $\mathcal{S}$ with mean $\mathrm{E}[\sigma_i(0)] = 0$, variance $a(0) \equiv \mathrm{Var}[\sigma_i(0)]$ and such that

$$\frac{1}{A}\ \mathrm{E}[\xi_i^\mu\sigma_i(0)] = \delta_{\mu\nu}m_0^\nu.$$

By the Law of Large Numbers and the Central Limit Theorem [3] the local field in i at time $t = 0$ can consequently be split up in a *signal* term and a *noise* term ([4], [5])

$$h_i(\sigma(0)) \equiv \lim_{N\to\infty} h_i(\sigma_\Lambda(0)) \stackrel{\mathcal{D}}{=} \xi_i^\nu m^\nu(0) + \sqrt{\alpha a(0)}\mathcal{N}(0,1).$$

In this formula $\alpha = p/N$ and $\mathcal{N}(0,1)$ denotes a Gaussian random variable with mean zero and unit variance. This implies that the network state at $t = 1$ can be expressed in terms of the macroscopic quantities at $t = 0$ as

$$\sigma_i(1) = \mathrm{g}[\xi_i^\nu m^\nu(0) + \sqrt{\alpha a(0)}\mathcal{N}(0,1)].$$

ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 27-32

Finally, we have to solve the feedback and correlation problem that show up in the fully connected network. In this setting it is not correct to assume that at $t = 1$ (or later) the set $\{\sigma_i(0)\}, i \in \Lambda$ is still a collection of i.i.d.r.v. on $\mathcal{S}$. We circumvent these difficulties by considering an *extremely diluted version* of the network (see [6], [7]). More specifically, we redefine the synaptic efficacies

$$\begin{aligned} J_{ij}(c) &\equiv \frac{c_{ij}}{c} N J_{ij} \quad i, j \neq i \in \Lambda, \\ \Pr(c_{ij}) &\equiv \tfrac{c}{N}\delta(c_{ij} - 1) + (1 - \tfrac{c}{N})\delta(c_{ij}), \text{ i.i.d.r.v..} \end{aligned}$$

Consequently, in the limit $N \to \infty$ almost all feedback loops in the graph $G_{\mathbb{N}}(c) = \{(i,j) : c_{ij} = 1, i, j \neq i \in \mathbb{N}\}$ are eliminated. Furthermore, with probability one any finite number of neurons have disjoint clusters of ancestors. The configuration $\{\sigma_i(1)\}, i \in \mathbb{N}$ is therefore again a collection of i.i.d.r.v..

4. Results

The signal-to-noise analysis allows us to express the main overlap and activity at $t = 1$ in terms of the main overlap and activity at $t = 0$. Because of the extreme dilution we can repeat the same formulae also at later times. It is therefore possible to derive expressions in the limit $t \to \infty$: the so called fixed-point equations. At $T = 0$ these read

$$\begin{aligned} m &= \frac{1}{A} \langle\langle \xi g(\xi m + \sqrt{\alpha a} Z) \rangle\rangle_{\xi, Z} \\ a &= \langle\langle g^2(\xi m + \sqrt{\alpha a} Z) \rangle\rangle_{\xi, Z}. \end{aligned}$$

At $T = \beta^{-1} \neq 0$ we find

$$\begin{aligned} m &= \left\langle\left\langle \frac{\sum_{s \in \mathcal{S}} \xi s \exp[\frac{-\beta}{2} s(\xi m + \sqrt{\alpha a} Z - bs)]}{\sum_{s \in \mathcal{S}} \exp[\frac{-\beta}{2} s(\xi m + \sqrt{\alpha a} Z - bs)]} \right\rangle\right\rangle_{\xi, Z} \\ a &= \left\langle\left\langle \frac{\sum_{s \in \mathcal{S}} s^2 \exp[\frac{-\beta}{2} s(\xi m + \sqrt{\alpha a} Z - bs)]}{\sum_{s \in \mathcal{S}} \exp[\frac{-\beta}{2} s(\xi m + \sqrt{\alpha a} Z - bs)]} \right\rangle\right\rangle_{\xi, Z}. \end{aligned}$$

In these expressions $\langle\langle .. \rangle\rangle_{\xi, Z}$ denotes an average over the pattern distribution and over the Gaussian random variable Z.

In Figs. 1 and 2 the performance of the network is analyzed using the following abbreviations in the accompanying tables :

- Description of solutions in (m, a)-plane :
  Z : zero solution $(m = 0, a = 0)$
  R : retrieval solution $(m \neq 0, a \neq 0)$
  S : sustained activity ("chaotic") solution $(m = 0, a \neq 0)$
  a : attracting point
  r : repelling point
  s : saddle-point

ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 27-32

- **Performance indicator :**
  D : line of optimal Hamming distance

	I	*II*	*III*	*IV*
R1	a	a	a	a
R2	s	s	s	-
S1	-	a	s	s
S2	-	r	r	s
Z	a	a	a	a

	I	*II*
R1	a	a
R2	s	-
S	a	s

	I
R	a
S	s

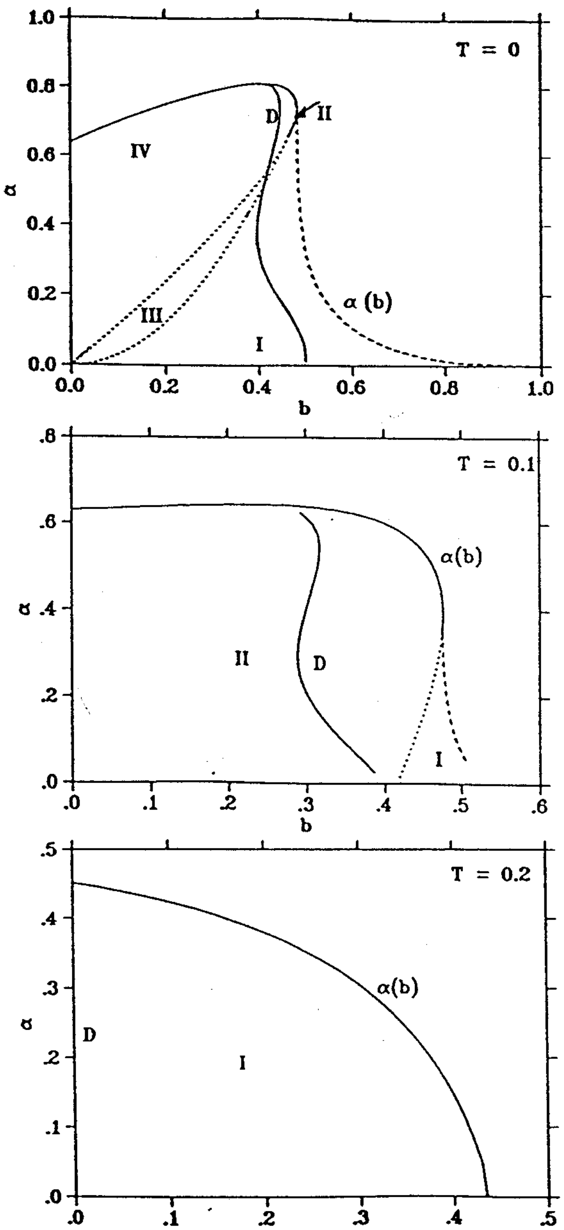

Fig. 1 : (α, b)-diagram for an extremely diluted network of 3 - Ising neurons storing uniformly distributed patterns, $g(x) = \mathrm{sgn}(x)\theta(|x| - b)$. Evolution of the retrieval characteristics with increasing T.

ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 27-32

	I	*II*
R	a	a
S	-	s
Z	s	r

	I
R	a
S	s

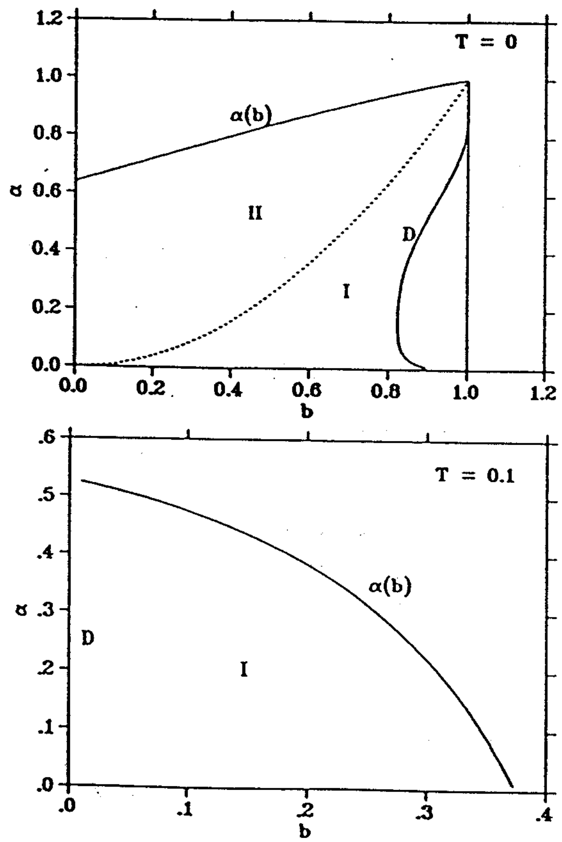

Fig. 2: (α, b)-diagram for an extremely diluted network of analog neurons storing uniformly distributed patterns, $g(x) = (|x/b+1| - |x/b-1|)/2$. Evolution of the retrieval characteristics with increasing T.

5. Conclusion

We have studied the retrieval quality for a class of extremely diluted multi-state networks in terms of the hamming distance, and optimized it as a function of the gain. At $T = 0$ we have observed fundamental differences in phase portrait between the graded response and the analog case. These may be related to differences in basin of attraction and/or retrieval time. At sufficiently high temperature there are no qualitative differences left between the graded response and the analog case.

This work has been supported in part by the Research Fund of the K.U. Leuven (grant OT/91/13).

ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 27-32

References

[1] D. Bollé , B. Vinck and V. A. Zagrebnov, *J. Stat. Phys.* 70 Vol. 5/6 (1993) to appear.

[2] H. Rieger, *J. Phys. A* 23:L1273 (1990).

[3] A. N. Shiryayev, *Probability* (Springer-Verlag, New York, 1984).

[4] A. E. Patrick and V. A. Zagrebnov, *J. Stat. Phys.* 63:59 (1991).

[5] V. A. Zagrebnov and A. S. Chvyrov, *Sov. Phys.-JETP* 68:153 (1989).

[6] B. Derrida, E. Gardner and A. Zippelius, *Europhys. Lett.* 4:167 (1987).

[7] R. Kree and A. Zippelius, in *Models of neural networks* E. Domany, J. L. van Hemmen and K. Schulten, eds. (Springer-Verlag, Berlin 1991):193.