Basin of attraction in networks of multistate neurons

M. Bouten

Limburgs Universitair Centrum, B-3590 Diepenbeek, Belgium

A. Engel

Institut für Theoretische Physik, Georg-August-Universität, D-3400 Göttingen, Germany

(Received 19 June 1992)

The retrieval dynamics of an extremely diluted network of three-state neurons designed by the projection rule is investigated. Depending on the parameters of the model four different fixed points with different stability can exist. Only one of them corresponds to full retrieval. It is discussed how the parameters of the model must be chosen in order to get perfect retrieval and large basins of attraction. Finally it is shown that a network of three-state neurons storing binary patterns has better retrieval properties than a network of binary neurons storing the same patterns.

PACS number(s): 87.10.+e, 64.60.Cn, 75.10.Nr

Recently there has been much interest in the properties of neural-network models with grey-level neurons [1–4]. Unlike the bistable McCulloch-Pitts neuron which is either "on" or "off" these neurons allow for intermediate levels of activity. On the other hand their dynamics is still that of linear threshold elements. This keeps the line to neurophysiologically motivated models using graded neurons [5–9] which is lost if one considers more general multistate neuron models using Potts-type interactions [10–13].

Networks of grey-level neurons can function as associative memories storing grey-toned patterns much the same as networks of McCulloch-Pitts neurons storing binary patterns. The maximal storage capacity has been determined using the methods of Gardner [14] and a variant of the Adatron algorithm was put forward as a fast learning algorithm [4]. In the present paper we complete the picture by investigating the retrieval dynamics in detail.

To this end we consider an extremely diluted network [15] for which one can derive closed evolution equations for the macroscopic order parameters. For simplicity we consider three-state neurons with the possible activity levels -1, 0, and 1. Models of this type with Hebbian synaptic couplings have already been studied both analytically [1] and numerically [2]. However, as known from the case of binary neurons, the retrieval dynamics of networks designed by more sophisticated learning rules giving constant or optimal stabilities is qualitatively different from the Hebbian case. In particular an unstable fixed point of the order-parameter flow emerges characterizing the typical size of the basins of attraction.

The model consists of N three-state neurons $S_i = 0, \pm 1$. On average each neuron is connected with $C \ll \log N$ randomly chosen other neurons via synaptic couplings J_{ij}. The neurons are updated in parallel according to their local fields

$$h_i(t) = \frac{1}{\sqrt{C}} \sum_j J_{ij} S_j(t) \tag{1}$$

using the rule (see Fig.1)

$$S_i(t+1) = \begin{cases} 1 & \text{if} & \theta < h_i(t) \\ 0 & \text{if} & -\theta \le h_i(t) \le \theta \\ -1 & \text{if} & h_i(t) < -\theta. \end{cases} \tag{2}$$

For $\theta = 0$ we recover the usual dynamics of binary neurons. The synaptic couplings J_{ij} are determined such that $p = \alpha C$ patterns $\{\xi_i^\mu\}$ are fixed points of the dynamics (1),(2). The patterns are randomly generated from the distribution

$$P(\xi_i^\mu) = (1 - a)\, \delta(\xi_i^\mu) + \frac{a}{2}\, [\delta(\xi_i^\mu + 1) + \delta(\xi_i^\mu - 1)]. \tag{3}$$

In the following we will call a site i with $\xi_i^\mu = 0$ a passive site and one with $\xi_i^\mu = \pm 1$ an active site of pattern μ.

The dynamics of a neuron configuration $S_i(t)$ correlated with the first pattern and uncorrelated with the other patterns is most clearly described by three macroscopic order parameters: the overlap at the active sites of pattern 1

$$m(t) = \frac{1}{aN} \sum_i \xi_i^1 S_i(t), \tag{4}$$

the activity at these sites

$$A_a(t) = \frac{1}{aN} \sum_i |\xi_i^1|\, |S_i(t)|, \tag{5}$$
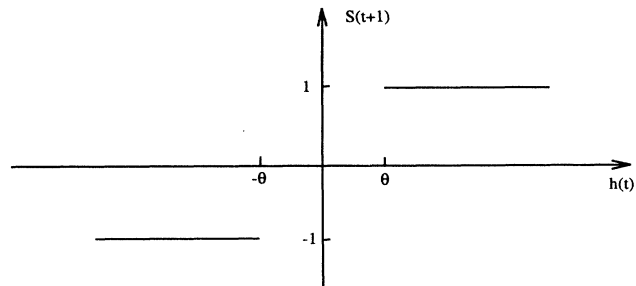

FIG. 1. The neuron activity function for symmetric three-state neurons.

and the activity at the passive sites of pattern 1

$$A_p(t) = \frac{1}{(1-a)N} \sum_i (1 - |\xi_i^1|) |S_i(t)|. \tag{6}$$

Specifying the learning rule $J_{ij} = J_{ij}(\xi_i^\mu)$ one can derive evolution equations for these order parameters in the usual way [15, 16]. Here we will consider the pseudoinverse rule

$$\frac{1}{\sqrt{C}} \sum_j J_{ij} \xi_j^\mu = \kappa \, \xi_i^\mu, \tag{7}$$

which makes the patterns $\{\xi_i^\mu\}$ fixed points of the dynamics (1), (2) if κ is taken larger than θ. Imposing the usual normalization condition $\sum_j J_{ij}^2 = C$, the stability parameter κ is related to the storage capacity α_c by [17]

$$\alpha_c = \frac{1}{1 + \kappa^2}. \tag{8}$$

From Eqs. (7), (1), and (2) we get the following dynamics of the order parameters (4)–(6):

$$m(t+1) = \frac{1}{2} \left[\text{erf}\left(\frac{\kappa m(t) + \theta}{\sigma(t)} \right) + \text{erf}\left(\frac{\kappa m(t) - \theta}{\sigma(t)} \right) \right], \tag{9}$$

$$A_a(t+1) = 1 - \frac{1}{2} \left[\text{erf}\left(\frac{\kappa m(t) + \theta}{\sigma(t)} \right) \right.$$
$$\left. - \text{erf}\left(\frac{\kappa m(t) - \theta}{\sigma(t)} \right) \right], \tag{10}$$

$$A_p(t+1) = 1 - \text{erf}\left(\frac{\theta}{\sigma(t)} \right), \tag{11}$$

where

$$\sigma^2(t) = 2[(1-a)A_p(t) + a(A_a(t) - m^2(t))]. \tag{12}$$

The three-dimensional flow (9)–(11) can be reduced to a two-dimensional one by introducing the overall activity

$$A(t) = (1-a)\,A_p(t) + a\,A_a(t). \tag{13}$$

We then get

$$m(t+1) = \frac{1}{2} \left[\text{erf}\left(\frac{\kappa m(t) + \theta}{\sigma(t)} \right) + \text{erf}\left(\frac{\kappa m(t) - \theta}{\sigma(t)} \right) \right], \tag{14}$$

$$A(t+1) = 1 - (1-a)\text{erf}\left(\frac{\theta}{\sigma(t)} \right)$$
$$- \frac{a}{2} \left[\text{erf}\left(\frac{\kappa m(t) + \theta}{\sigma(t)} \right) - \text{erf}\left(\frac{\kappa m(t) - \theta}{\sigma(t)} \right) \right], \tag{15}$$

where now $\sigma(t)$ is given by

$$\sigma^2(t) = 2\,[A(t) - am^2(t)]. \tag{16}$$

These equations are similar to the case of Hebb couplings

[1], the main difference being the appearance of $m(t)$ in $\sigma(t)$.

In order to characterize the order parameter flow (14),(15) it is useful to determine its fixed points. First we see that $m = 0$ is an invariant line corresponding to the dynamics of neuron configurations uncorrelated with the couplings. In this case we get from Eq. (15):

$$A(t+1) = 1 - \text{erf}\left(\frac{\theta}{\sqrt{2A(t)}} \right). \tag{17}$$

Therefore we always have a fixed point with $A = A_Z = 0$. It is an obvious consequence of the dynamics (2) and means that all neurons are simply turned off. For $\theta < \theta_c \cong 0.575$ there is, however, another attractor on the $m = 0$ line corresponding to a nonzero self-sustained activity A_S of the network, which is not correlated with the patterns. The value of A_S depends only on θ. It increases from $A_S \cong 0.25$ at $\theta \cong 0.575$ with decreasing θ and tends to one for $\theta \to 0$. For $\theta = 0$ it is of course nothing but the nonretrieval fixed point known from the dynamics of binary neurons. Since these two fixed points, which will be denoted by Z and S, respectively, concern configurations uncorrelated with the patterns, they are the same for all learning rules giving rise to synapses normalized to $\sum_j J_{ij}^2 = C$. In particular they are present also for Hebb couplings [1].

In addition to these nonretrieval fixed points there can exist two fixed points R and R^* with nonzero values of m. The first of them corresponds to full retrieval, i.e., $m = 1, A = a$. It exists for all $\theta > 0$ and for $\theta = 0$ if $a = 1$. The nature of the second fixed point can best be understood by considering the limit $\kappa \to \infty$ (or equivalently $\alpha \to 0$). Then also $m = 1$ is an invariant line. This means that at the active sites the neuron configuration is fully aligned with the pattern. On the other hand the neurons at the passive sites are decoupled from the rest and can either be off (this corresponds to the fixed point R) or can establish a self-sustained activity $A_p > 0$. Indeed for $\kappa \to \infty$ we find from (9) and (10) $m = A_a = 1$ and therefore (11) and (12) simplify to

$$A_p(t+1) = 1 - \text{erf}\left(\frac{\theta}{\sqrt{2(1-a)A_p(t)}} \right). \tag{18}$$

Comparing (18) with (17) we see that for $\theta < \theta_c\sqrt{1-a}$ it is possible to find a nonzero self-sustained activity at the passive sites and therefore a second attractor R^* on the $m = 1$ line. For finite κ the active and passive sites are not completely decoupled and a self-sustained activity at the passive sites will give rise to a slight reduction of m.

Let us now discuss the implications of the order parameter flow for the retrieval dynamics of the network. The model has three parameters: the pattern activity a, the pattern stability κ, and the width 2θ of the zero-activity interval of the neuron activity function. Depending on the value of these parameters quite different phase diagrams can be found. Rather than to list these possibilities we try to discuss those parameter regions which are interesting for associative recall of patterns, concentrat-
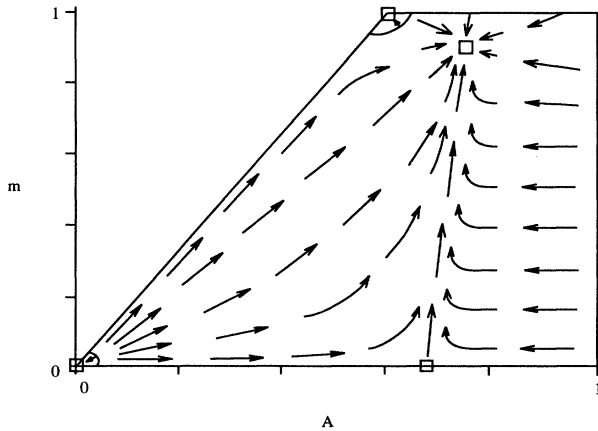
FIG. 2. Typical flow diagram for the overlap and the activity in a network of three-state neurons. The squares denote the fixed points discussed in the text. The parameters are $a = 0.6$, $\theta = 0.34$, $\kappa = 1.13$.

FIG. 4. Dependence of κ_B on θ. Above the line there are large basins of attraction of the patterns, below most initial configurations are attracted by the fixed point S (cf. Figs. 5 and 6).

ing in particular on the new features introduced by the three-state nature of the neurons. Let us for the time being fix $a = 0.6$ so that a pattern has approximately the same number of 0's as ± 1's.

Figure 2 shows a typical plot of the fixed points and separatrices of the order parameter flow. Most of the initial conditions are attracted by the fixed point R^* which gives only approximate retrieval. The basin of attraction of the full-retrieval fixed point R is on the other hand very small. It seems advantageous to increase the value of θ in order to suppress the fixed points S and R^*. Then it is very easy to decide from the output whether the network has identified the pattern or not, since in the latter case it simply turns off. Looking at the plot of the separatrices for such a situation, however, one realizes that the basin of attraction of the full-retrieval fixed point R would be small unless κ is rather large implying a low value of α (Fig. 3).

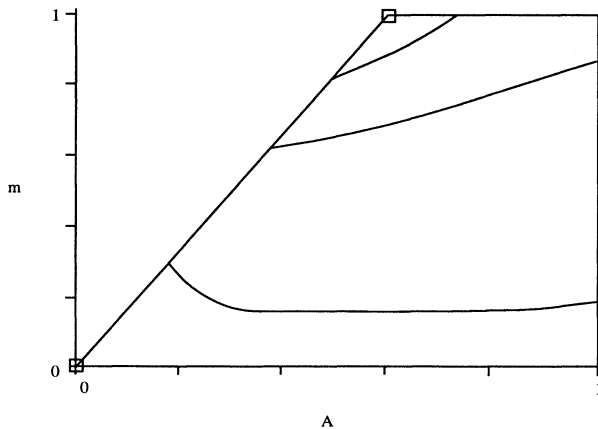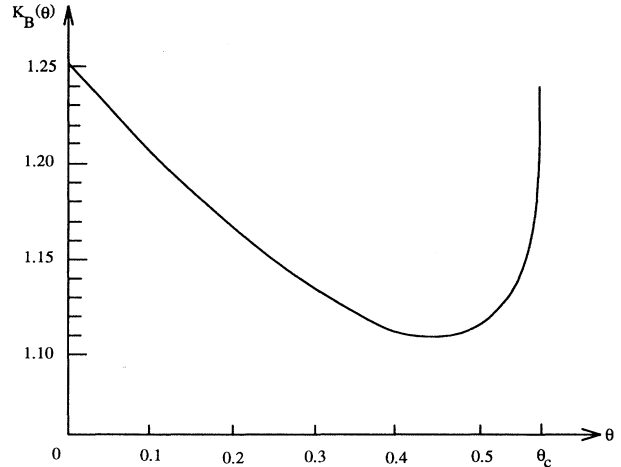Smaller values of θ can yield larger basins of attraction.

This is due to the fact that the fixed point S, though an attractor on the $m = 0$ line, becomes unstable with respect to positive values of m if κ is larger than a critical value $\kappa_B(\theta)$. This is similar to the instability of the non-retrieval fixed point $m = 0$ for binary neurons if $\kappa > 1.25$. From (14) and (16) it is easy to derive the expression for κ_B in terms of the activity $A_S(\theta)$:

$$\kappa_B(\theta) = \sqrt{\frac{\pi A_S}{2}} \exp\left\{\frac{\theta^2}{2A_S}\right\}. \tag{19}$$

For $\theta = 0$, one has $A_S = 1$ so that $\kappa_B = \sqrt{\pi/2} \cong 1.25$. For general values of θ, the function $\kappa_B(\theta)$ is shown in Fig. 4. The minimal value of κ_B is 1.111 and corresponds to $\theta \cong 0.45$. In Figs. 5 and 6 it is shown how the flow diagrams differ for values of κ slightly smaller and slightly larger than κ_B. For $\kappa < \kappa_B$, S is a stable fixed point and attracts almost all initial configurations (Fig. 5).
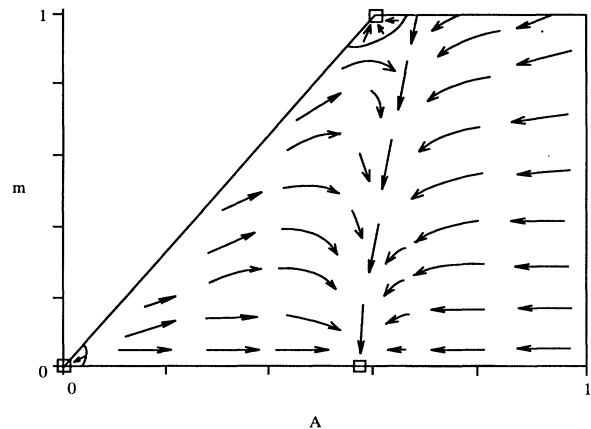

FIG. 3. Separatrices between the fixed points Z and R for $a = 0.6$, $\theta = 0.6$, and $\kappa = 1.1$, 1.2, 1.3 (from top to bottom).

FIG. 5. Flow diagram for $a = 0.6$, $\theta = 0.43$, $\kappa = 1.0 < \kappa_B(\theta)$. The basin of attraction of the retrieval fixed point R is small.

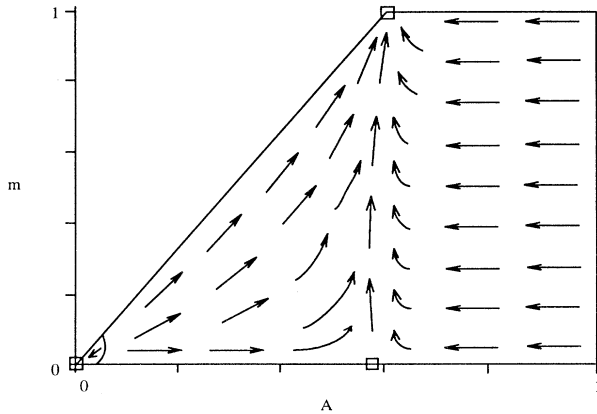FIG. 6. Flow diagram for $a = 0.6$, $\theta = 0.43$, $\kappa = 1.12 >$ $\kappa_B(\theta)$. The basin of attraction of R is large.

FIG. 7. Flow diagram for $a = 1$, $\theta = 0.43$, $\kappa = 1.12$. The cross on the $A = 1$ line gives the value of m at the unstable fixed point of a network of binary neurons.

For $\kappa > \kappa_B$, S becomes a saddle and now practically all initial conditions will eventually flow to the full-retrieval fixed point R (Fig. 6). Note that the value of κ_B for binary neurons is about 1.25 corresponding to $\alpha_B \cong 0.39$, whereas here an optimally chosen θ allows for $\kappa_B \cong 1.11$ corresponding to $\alpha_B \cong 0.45$.

In order to achieve large basins of attraction in a network of three-state neurons one can therefore use quite generally the following strategy. First θ should be smaller than $\theta_c \cong 0.575$ in order to establish the fixed point S. Otherwise the zero fixed point Z attracts too many initial configurations. Then θ and κ should be chosen such that $\kappa > \kappa_B(\theta)$ because only then is S a saddle and most of the initial conditions will flow to the full-retrieval fixed point R (Fig. 6). On the other hand θ must be large enough in order to suppress the fixed point R^* which, when it exists, has a large basin of attraction (Fig. 2). Hence depending on a a rather narrow interval $\theta_c\sqrt{1-a} < \theta < \theta_c$ for θ can be given, from which the value of θ with the smallest $\kappa_B(\theta)$ may be chosen.

Finally it is interesting to discuss the case $a = 1$, i.e., storage of binary patterns in a network of ternary neurons. Figure 7 shows a flow diagram for $\theta = 0.43$ and $\kappa = 1.12$. Practically all initial configurations with $m(t = 0) > 0$ are attracted by the retrieval fixed point R. This is also the case for initial states with activity $A = 1$, even though the value of κ is smaller than the value $\kappa_B = 1.25$ for binary neurons. In a network of binary neurons storing the same patterns the initial overlap must exceed a critical value m_c in order to en-
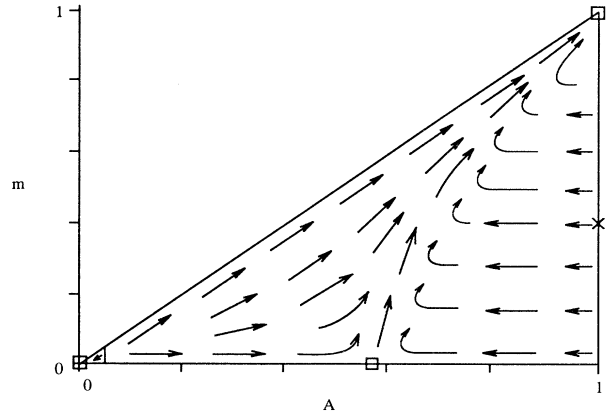
sure that the initial configuration will converge to the retrieval fixed point. The value of m_c for the parameters given is indicated in Fig. 7 by the cross on the $A = 1$ line. As far as the basin of attraction is concerned it is hence advantageous to store binary patterns in a network of multilevel neurons. The reason for this is easy to understand qualitatively. In the case of binary neurons, the neurons in the initial configuration that are antiparallel to the pattern contribute significantly to the local fields and can block the way to the retrieval attractor. In the case of ternary neurons these neurons are very likely to be turned off in the first time steps, i.e., there is a quick process of reducing the activity from its starting value 1 (cf. Fig. 7). Then the misaligned neurons do not contribute to the local fields and the correlations with the pattern can slowly increase. Only at the end, when m tends to 1, will the activity grow again and finally reach its initial value. The zero-activity interval $-\theta < h(t) < \theta$ in the neuron-activity function reduces the degree of frustration and may allow for retrieval also in situations where for $\theta = 0$ no retrieval is possible.

It is straightforward to extend the above analysis to networks of three-state neurons designed by the optimal Gardner prescription [14]. Since for binary neurons the basins of attraction show a very similar dependence on α for both learning rules we expect that also in the case of ternary neurons the results for the Gardner rule will be qualitatively the same as those for the projection rule obtained in the present paper.

[1] J. S. Yedidia, J. Phys. A **22**, 2265 (1989).
[2] C. Meunier, D. Hansel, and A. Varga, J. Stat. Phys. **55**, 859 (1989).
[3] H. Rieger, J. Phys. A **23**, L1273 (1990).
[4] S. Mertens, H. M. Köhler, and S. Bös, J. Phys. A **24**, 4941 (1991).
[5] J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A **81**, 3088 (1984).
[6] A. Treves, Phys. Rev. A **42**, 2418 (1990).
[7] M. Shiino and T. Fukai, J. Phys. A **23**, L1009 (1990).
[8] C. M. Marcus, F. R. Waugh, and R. M. Westervelt, Phys. Rev. A **41**, 3355 (1990).
[9] S. Mertens, J. Phys. A **24**, 337 (1991).

[10] I. Kanter, Phys. Rev. A **37**, 2739 (1988).
[11] J. Cook, J. Phys. A **22**, 2057 (1989).
[12] J. P. Nadal and A. Rau, J. Phys. I (France) **1**, 1109 (1991).
[13] D. Bolle, P. Dupont, and J. van Mourik, J. Phys. A **24**, 1065 (1991).
[14] E. Gardner, J. Phys. A **21**, 257 (1988).
[15] B. Derrida, E. Gardner, and A. Zippelius, Europhys. Lett. **4**, 167 (1987)
[16] E. Gardner, J. Phys. A **22**, 1969 (1989).
[17] W. Krauth, M. Mezard, and J. P. Nadal, Complex Systems **2**, 387 (1988).