External fields in attractor neural networks with different learning rules

1991 J. Phys. A: Math. Gen. 24 313

(http://iopscience.iop.org/0305-4470/24/1/037)

J. Phys. A: Math. Gen. 24 (1991) 313–326. Printed in the UK

External fields in attractor neural networks with different learning rules

A Rau, D Sherrington and K Y M Wong
Department of Theoretical Physics, University of Oxford, Oxford OX1 3NP, UK

Received 12 July 1990

Abstract. The effects of external fields on the retrieval properties of highly dilute attractor neural networks with general classes of learning rules are examined. It can be shown that external fields increase basins of attraction making even perfect retrieval possible for relatively high loading. The application of different classes of noise distributions on the stimulus field indicates that certain first-order transitions occuring are peculiarities of the type of noise. Optimally adapted networks in the presence of external fields extend the critical loading above which perfect retrieval is impossible. In the presence of external fields in the high-temperature regime, Hebb networks also retrieve better than rules with more optimal performances at low temperatures.

1. Introduction

Robustness against noise disruption is one of the major features of attractor neural networks. In the presence of a small amount of noise, either due to an external temperature or to the mutual disturbance of the patterns, the system is still able to drift towards a fixed point or limit cycle.

Originally only specific learning rules were studied to investigate the properties of these networks. The work of Gardner [1] made it possible to investigate optimal properties of networks without reference to a particular learning rule. She was able to show that in a highly dilute network the maximum storage capacity is 2, but that at far lower loading ($\alpha = 0.42$) the basins of attraction already start to shrink. This articulates itself in the appearance of unstable fixed points, which often hide the attractor with high overlap if we start with a low initial overlap. In this paper we will investigate a method of widening the basins of attraction through the application of external stimuli fields.

External stimuli have been studied previously by different groups [2–5], with the motivation either to imitate more realistic physiological situations, or to improve the ability of the network to retrieve. However, previous studies have considered either only particular learning rules or only specific noise distributions on the external stimulus field. Here we will consider different learning rules which represent limiting cases of certain universality classes [6] near saturation, i.e. when the volume of solutions in the space of interactions shrinks to zero. The noise distributions for the external stimulus field will also be taken from structurally different types of distributions. This analysis will enable us to gain insight into some general results accompanying the noise

0305-4470/91/010313+14$03.50

spectrum in external fields. We will show that certain first-order transitions occurring are peculiarities of the noise distributions.

We also analyse an optimally adapted neural network [7,8] in the presence of an external stimulus field. For this we consider a network which is optimally adapted in the presence of training noise and an external stimulus field. Thus the analysis of such a network gives the best possible attractor overlap, provided we have specified the noise distribution on the external stimulus field.

In addition we will investigate the influence of an external temperature on the retrieval process. Learning rules such as the Hebb rule yield higher storage capacity at high temperatures in the absence of external fields, when compared with learning rules more optimized at zero temperature, such as the one by Gardner. It turns out that similar phenomena exist in the presence of external fields; the Hebb rule retrieves better than the pseudo-inverse rule at high temperatures, although the converse is true at low temperatures.

2. The model

2.1. The general dynamical equations

Guided by our previous work [4] on external stimuli we consider a highly diluted network, similar in structure to that of Derrida *et al* [9], but with the synaptic weights prescribed by more general classes of learning rules. Dilution allows us to extend the first step dynamics by Kepler and Abbott [10] to any time step, since the high dilution limit enables us to neglect dynamical site correlations.

Using the approach of [10] we can determine the macroscopic overlap of the system after one time step provided we know the aligning field distribution in the system.

As in our previous work we consider the storage of a set of p binary patterns $\{\xi_i^\mu\}$ in a dilute neural network with parallel dynamics introduced through the updating rule

$$S_i(t+1) = \operatorname{sgn}\left(\frac{1}{\sqrt{C}}\sum_{j=i_1}^{i_C} J_{ij}S_j(t) + h_i^{\text{ext}}\right) \tag{1}$$

where h_i^{ext} is the external stimulus field, C the average number of connections between the N neurons and $S_i(t)$ the state of the binary McCulloch–Pitts neuron at site i and time t. The nodes feeding neuron i are given by $j = i_1, \ldots, i_C$. The interactions satisfy the spherical constraint $\sum_j J_{ij}^2 = C$. We will be interested in systems where $C \to \infty$ and

$$\lim_{N\to\infty} \frac{\ln N}{\ln C} = \infty \tag{2}$$

which is the criterion for high dilution [11].

The stimulus field $h_i^{\text{ext}} = h_i[\eta]$ is a functional of a certain noise distribution η. The sign of the field is assumed to have an average non-zero overlap m_0 with one of the patterns we want to retrieve, say pattern one. This overlap is of course equivalent to the overlap of the initial configuration $\{S_i(t=0)\}$ with $\{\xi_i^1\}$ and contains the information about how many of the bits in the field agree on average with the pattern.

Thus, during the dynamical evolution of the system the external field will persistently support the retrieval of one pattern.

Following the work of [1, 10, 12] we calculate the averaged output overlap g_i at node i for an initial overlap m_0. It turns out that this quantity can be expressed in terms of the aligning field Λ_i^1 which is given by

$$\Lambda_i^1 = \frac{\xi_i^1}{\sqrt{C}} \sum_{j=1}^{C} J_{ij}\xi_j^1. \tag{3}$$

(Henceforth we drop the subscript i for convenience.) Since the local field for pattern one is Gaussian distributed with a mean of $(m_0\Lambda^1 + \xi^1 h[\eta])$ and width of $\sqrt{1-m_0^2}$, the average output overlap can be written as

$$\begin{aligned} g_{m_0}(\Lambda^1, h[\eta])) &= \int \mathrm{D}z \,\mathrm{sgn}\left((m_0\Lambda^1 + \xi^1 h[\eta]) + \sqrt{1-m_0^2}\, z \right) \\ &= \mathrm{erf}\left(\frac{m_0\Lambda^1 + \xi^1 h[\eta]}{\sqrt{2(1-m_0^2)}} \right) \end{aligned} \tag{4}$$

where $\mathrm{D}z = \exp(-z^2/2)\mathrm{d}z/\sqrt{2\pi}$.

The dynamics in highly connected networks is still a problem for more than the first few steps [13]. Thus, below we restrict ourselves to the consideration of a highly dilute network. The overlap of the system with pattern one at time $t+1$ can be given recursively in terms of the overlap at time t

$$m(t+1) = \left\langle \left\langle \mathrm{erf}\left(\frac{m(t)\Lambda + \xi^1 h[\eta]}{\sqrt{2(1-m^2(t))}} \right) \right\rangle_\eta \right\rangle_\Lambda . \tag{5}$$

Here $\langle \ldots \rangle_\eta$ denotes the average over the noisy external field and $\langle \ldots \rangle_\Lambda$ represents averaging over the aligning field distribution $\rho(\Lambda)$ which can be determined via (3) using the quenched distribution of pattern one.

As an aside we mention here that the deterministic dynamics of (1) can be replaced by finite-temperature dynamics using not a step but a sigmoid updating function. It can easily be shown [11] that the general dynamical rule $S(t+1) = \langle \mathrm{sgn}\,(h_{\mathrm{tot}}(t) + \tau) \rangle_\tau$ reduces to Glauber dynamics for $\rho(\tau) = \beta/2[\cosh(\beta\tau)]^{-2}$; $\rho(\tau)$ is the distribution against which we average. We shall instead use a Gaussian distribution [14] for $\rho(\tau)$ of width (temperature) T and centred around 0. The only change in (5) resulting from the introduction of T is that $[2(1-m^2(t))]^{1/2}$ must be replaced by $[2(1-m^2(t)+T^2)]^{1/2}$. Later on we will analyse the influence of $T > 0$ *in nuce*.

2.2. The different aligning field distributions

Kepler and Abbott [10] classified the aligning field distributions of different synaptic prescriptions into certain universality classes. We restrict our considerations here to the universal behaviour near saturation.

If the *a priori* distribution of aligning fields is bounded and differentiable in the region $[\kappa, +\infty]$ and vanishes outside this region then the aligning field distribution is identical to the one for the Gardner optimized learning rule, which is

$$\rho_{\mathrm{O}}(\Lambda) = \frac{\exp(-\Lambda^2/2)}{\sqrt{2\pi}}\Theta(\Lambda - \kappa) + \delta(\Lambda - \kappa)\int_{-\infty}^{\kappa} \mathrm{D}z \tag{6}$$

where κ is the stability determining α through

$$\alpha^{-1} = \int_{-\infty}^{\kappa} \mathrm{D}z\,(\kappa - z)^2. \tag{7}$$

Two other learning rules in Kepler and Abbott's classification scheme of universality classes are the Hebbian and pseudo-inverse learning rules. The pseudo-inverse matrix [15, 16] yields an aligning field distribution of

$$\rho_P(\Lambda) = \delta(\Lambda - \sqrt{1/\alpha - 1}) \tag{8}$$

whereas the Hebb learning rule leads to a Gaussian distribution [10] centred around $1/\sqrt{\alpha}$, i.e.

$$\rho_H(\Lambda) = \frac{1}{\sqrt{2\pi}} \exp[-1/2\,(\Lambda - 1/\sqrt{\alpha})^2]. \tag{9}$$

A further motivation for considering a Gaussian distribution is that, as Wong and Sherrington [8] have recently shown, the dilute Hopfield network with Hebbian synapses gives the maximum storage capacity $\alpha_c = 2/[\pi(1+T^2)]$ in the high-temperature regime ($T > 0.38$) provided no external field is present.

Due to the universal classes they represent, these three aligning field distributions are particularly suited to investigate general properties of external stimuli.

2.3. The noise in the stimulus

A perfect stimulus field would have the form $h\xi^1$ where h is an arbitrary field strength. Since, however, we want to consider only fields which have the same noisy overlap with one pattern as the initial configuration of the system, we have to introduce noise on the stimulus field. Here we can consider highly singular (δ-like) and continuous distributions.

For the singular distribution we consider a field which is $h\xi^1$ with probability a_1 and $\epsilon h\xi^1$ with probability $1 - a_1$, where ϵ is an arbitrary factor less than one. Relative to pattern one the external field distribution is

$$\rho_S(h^{\mathrm{ext}}) = a_1\,\delta(h^{\mathrm{ext}} - h\xi^1) + a_\epsilon\,\delta(h^{\mathrm{ext}} - \epsilon h\xi^1) \tag{10}$$

where a_1 and a_ϵ are chosen such that the average overlap of the stimulus field with the pattern is m_0, i.e. $a_1 = (m_0 - \epsilon)/(1 - \epsilon)$ and $a_\epsilon = 1 - a_1$. Equation (10) reduces to *discrete noise* for $\epsilon = -1$ and *hidden units* for $\epsilon = 0$ in [2]. We will consider primarily these special cases but also analyse how ϵ influences the retrieval in general.

The continuous noise distribution is taken to be $h(\xi^1 + \eta)$ where η is Gaussian distributed with zero mean and width δ. The distribution of the external field relative to pattern one is thus given by

$$\rho_G(h^{\mathrm{ext}}) = \frac{1}{\sqrt{2\pi}\delta} \exp\left(-\frac{(h^{\mathrm{ext}} - h\xi^1)^2}{2\delta^2}\right) \tag{11}$$

where the overlap m_0 of the sign of the field with pattern one is ensured via $\delta = 1/[\sqrt{2}\,\mathrm{erf}^{-1}(m_0)]$.

The factors h in (10) and (11) are arbitrary and thus not yet comparable. If however we want to compare discrete and Gaussian noise we have to ensure that (10) and (11) have the same width. This necessitates a rescaling of the h in (11) by a factor of $1/\sqrt{1+\delta^2}$ which has been done in all figures.

2.4. The final dynamical equations

The dynamics of the system is given after performing the various averages by

$$m(t+1) = f(m(t)). \tag{12}$$

In the case of the Hebbian learning rule with singular noise we can show that

$$f_{\mathrm{H,S}} = \sum_{q\in\{1,\epsilon\}} a_q \operatorname{erf}\left(\frac{1}{\sqrt{2\alpha}}\left(m(t) + qh\sqrt{\alpha}\right)\right). \tag{13}$$

For $\epsilon = -1$ this result reduces to the one given in [4], where we remark that the field strengths h here and in [4] vary by a factor of $\sqrt{\alpha}$ which is a result of the different normalizations of the interaction matrices. Since h is an arbitrary factor this has of course no consequences. For Gaussian noise we get

$$f_{\mathrm{H,G}} = \operatorname{erf}\left(\frac{m(t) + h\sqrt{\alpha}}{\sqrt{2\alpha(1+h^2\delta^2)}}\right). \tag{14}$$

The pseudo-inverse learning rule gives a dynamical function for singular noise of

$$f_{\mathrm{P,S}} = \sum_{q\in\{1,\epsilon\}} a_q \operatorname{erf}\left(\frac{m(t)\sqrt{1/\alpha - 1} + qh}{\sqrt{2(1-m^2(t))}}\right) \tag{15}$$

and for Gaussian noise yields

$$f_{\mathrm{P,G}} = \operatorname{erf}\left(\frac{m(t)\sqrt{1/\alpha - 1} + h}{\sqrt{2(1-m^2(t)+h^2\delta^2)}}\right). \tag{16}$$

For the Gardner learning rule we get

$$f_{\mathrm{O,S}} = \sum_{q\in\{1,\epsilon\}} a_q \left[\operatorname{erf}\left(\frac{m(t)\kappa + qh}{\sqrt{2(1-m^2(t))}}\right)\int_{-\infty}^{\kappa} \mathrm{D}y + \int_{\kappa}^{\infty} \mathrm{D}\Lambda \operatorname{erf}\left(\frac{m(t)\Lambda + qh}{\sqrt{2(1-m^2(t))}}\right)\right] \tag{17}$$

for singular noise and

$$f_{\mathrm{O,G}} = \operatorname{erf}\left(\frac{m(t)\kappa + h}{\sqrt{2(1-m^2(t)+h^2\delta^2)}}\right)\int_{-\infty}^{\kappa} \mathrm{D}y + \int_{\kappa}^{\infty} \mathrm{D}\Lambda \operatorname{erf}\left(\frac{m(t)\Lambda + h}{\sqrt{2(1-m^2(t)+h^2\delta^2)}}\right) \tag{18}$$

for Gaussian noise. Equation (7) again relates κ and α. The special case of equations (15) and (17) for $\epsilon = -1$ has independently been studied by Engel *et al* [5].

2.5. Optimally adapted networks

Following the work of Wong and Sherrington [7,8] we consider *training with noise* to enhance memory associativity for retrieval in the presence of a noisy stimulus field. We optimize the averaged output overlap g_i^μ at node i of the stored patterns $\{\xi_i^\mu\}$. In [8] the principle of adaptation was developed. The idea is that when a system optimizes its performance in a training environment, then its performance is optimized in the same retrieval environment.

As we want to retrieve noisy patterns in the presence of an external stimulus field, we thus also train the system in the same surrounding. In (4) we derived the output overlap $g_{m_t}(\Lambda^\mu, h_t[\eta])$ for an initial (training) overlap m_t and an external (training) field $h_t[\eta]$ to be

$$g_{m_t}(\Lambda^\mu, h_t[\eta]) = \left\langle \operatorname{erf}\left(\frac{m_t\Lambda^\mu + \xi^1 h_t[\eta]}{\sqrt{2(1-m_t^2)}}\right)\right\rangle_\eta . \tag{19}$$

We start by maximizing the output overlap for the first time step, which is equivalent to optimizing the performance function $C = \sum_\mu g(\Lambda^\mu)$. Following the reasoning of [8] we can determine the average maximum performance per pattern $\int \mathrm{d}\Lambda\, \rho(\Lambda) g(\Lambda)$. The aligning field distribution is given by $\rho(\Lambda) = \int \mathrm{D}t\, \delta(\Lambda - \lambda(t))$, where $\lambda(t)$ is the inverse function of

$$t(\lambda) = \lambda - \gamma g'(\lambda). \tag{20}$$

The parameter γ is related to α by

$$\alpha^{-1} = \int \mathrm{D}t\, (\lambda(t) - t)^2 . \tag{21}$$

When $\lambda(t)$ is not single valued we take the λ which gives the largest value of $\left(g(\lambda) - (\lambda - t)^2/2\gamma\right)$, which is equivalent to a Maxwell construction.

As we have now specified the training procedure we define the retrieval. The retrieval output $f_{m_t}(m)$ is given by

$$f_{m_t}(m) = \int \mathrm{d}\Lambda\, \rho_{m_t}(\Lambda) g_m(\Lambda) \tag{22}$$

where m is the input overlap. Using the principle of adaptation which is formulated more precisely in [8] we optimize the retrieval performance of our system by choosing the initial overlap and the external field in the training and retrieving stages to be the same.

As we are interested here in dilute networks, where the complete dynamics is known, the fixed-point equation we finally want to solve self-consistently is given by

$$m^* = \int \mathrm{D}t \operatorname{erf}\left(\frac{m^* \lambda_{m^*}(t) + h}{\sqrt{2(1-(m^*)^2 + h^2\delta^2)}}\right). \tag{23}$$

Here Gaussian noise on the stimulus field was applied, since it shows less 'artificial' phase transitions than singular noise.

We note a strong similarity between (23) and the corresponding equation of [8] in the presence of Gaussian retrieval noise but no stimulus field; in the latter case $h\delta$ is replaced by the retrieval temperature T and there is no h in the numerator of the argument of the error function.

The stable fixed points of (23) give optimal retrievers, the unstable fixed points give the optimal basin boundaries of adaptation.

3. Results

3.1. The different fixed-point behaviour

After many iterations the system approaches a fixed point. We will only be interested in this long term dynamics. In the following we examine the behaviour of the different typical dynamical equations (13)–(18) in terms of their fixed points of the retrieval map (12). We will consider only loading levels where in the field-free system the basins of attraction are already smaller than one. This corresponds to an $\alpha > \alpha_G$ with $\alpha_G = 0.42$ for the Gardner optimized learning rule and an $\alpha > \alpha_P$ with $\alpha_P = 0.38$ for the pseudo-inverse learning rule. For the following figures we thus take $\alpha = 0.5$ and $m_0 = 0.4$. Considering the retrieval behaviour when we start from an initial overlap m_0, we will construct retrieval behaviour diagrams in the space of loading α and external field h. In general, we can consider the retrieval behaviour for an *arbitrary* initial overlap which may be different from m_0. However, our retrieval diagrams do not show the general basins of attraction for such general initial conditions. Here for each h there is only *one* output overlap, which is the stable fixed point m^* closest to m_0 which can be reached *without* crossing an unstable fixed point. Where necessary, the other stable and unstable fixed points are only shown to illustrate the structural differences between the systems.

In general, the optimal learning rules (6) and (8) make perfect retrieval possible at the cost of the radius of the basins of attraction. The Hebb rule however creates very wide basins of attraction at the expense of reduced retrieval quality. We will show in the next three subsections that in the presence of external fields these statements are still valid and, in addition, the basins for perfect retrieval are substantially enlarged, making retrieval starting from the initial overlap m_0 possible.

3.1.1. The pseudo-inverse rule. As a first example we examine the pseudo-inverse learning rule. We start with the case of singular noise (15), since it yields a very rich scenario. Figure 1 shows the fixed-point overlap $m^*(h)$ at intermediate loading for different values of $\epsilon \in \{-1, -0.5, 0\}$. In this and the following fixed-point diagrams, a multiple solution of m^* for a given h or α corresponds to an alternation of stable and unstable fixed points, starting downwards from the highest m^* as a stable fixed point.

The case $\epsilon = -1$ corresponds to the one analysed in [5]. For $h < h_< \approx 0.25$ we observe that the strong attractor $m^* = 1$, which exists for $h < \sqrt{1/\alpha - 1} = 1$ ($\alpha = 0.5$), is hidden by an unstable fixed point. For $h < h_<$ an increase in h leads to a continuous increase of a weaker attractor m^*, to which the initial overlap m_0 converges. If h exceeds $h_<$ a first-order transition occurs when the weaker attractor m^* merges with the basin boundary overlap, thus opening up the $m^* = 1$ attractor. A further increase in the field beyond $h_> \approx 0.45$ leads to another first-order transition where now, as for small h, the perfect attractor is again hidden, leading to an m^* monotonically decreasing with h. For very large h the fixed-point overlap reaches m_0 (independently of α). The first barrier for low h can be interpreted as a blocking due to the narrowed basin of attraction in the field free case. For high h the blocking takes place due to the overemphasis of the erroneous external field.

The case $\epsilon = -0.5$ is similar to the one just described, only the band $[h_<, h_>]$ is wider, since the proportion of erroneous bits in the hint is smaller compared to $\epsilon = -1$.

The hidden units ($\epsilon = 0$) mark an interesting regime. Since the external field does not impose errors but only correct bits, we can retrieve perfectly for all $h > h_< \approx 0.25$,

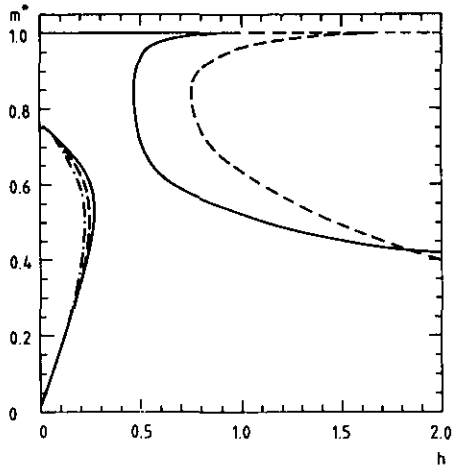

Figure 1. Fixed points of (12) for the pseudo-inverse rule with singular noise; $\alpha = 0.5$, $m_0 = 0.4$, $\epsilon = -1$ (—), $\epsilon = -0.5$ (- - -), $\epsilon = 0$ (— · —).

and $m^* = 1$ is a stable fixed point for all h. In a normal retrieval problem α and m_0 are usually unknown in advance. The absence of the upper band limit $h_>$ is therefore of interest for practical purposes, since it facilitates the choice of an appropriate h. The hidden unit stimulus is of great importance in pattern association problems where those bits which are known to be correct are supported whereas the other (hidden) ones receive no stimulus at all.

The change in the retrieval diagrams for increasing α can easily be predicted for the case of discrete noise. Increasing the load narrows the band of unblocked perfect retrieval. At a certain critical loading the band edges merge, leaving an attractor behaviour similar to the one for the Hebb rule [4].

Figure 2 shows m^* for hidden units and different levels of loading. For increasing α the $h_<$ beyond which perfect retrieval is unblocked increases. At a certain value α_b, which depends on m_0 ($\alpha_b \approx 0.655$ for $m_0 = 0.4$), there occurs a phase transition, which results in the appearance of an unstable fixed point hiding the strong attractor for all h. This scenario ($\alpha > \alpha_b$) is similar to the one observed by [2] for hidden units.

Up to now we have only analysed the influence of singular noise. For Gaussian noise, the breakdown of the unblocked perfect retrieval regime at high field h is not accompanied by the first-order transition at which an imperfect retrieval phase appears discontinuously. Instead there is a continuous weakening of the strong attractor (figure 3) which starts to become important at about the same h as the first-order transition in the case of discrete noise. Thus figure 3 illustrates, and we will support this by some other cases, the fact that some phase transitions occuring in systems with discrete noise cannot be observed in systems with Gaussian noise, and are due to the peculiarities of the discrete noise distribution.

3.1.2. The Hebb rule. The Hebb rule in the presence of discrete noise has been analysed in detail for a DGZ network in [4]. Typical attractor behaviours can be seen there for different loading levels and initial overlaps. The main aspect is that for intermediate h substantial improvement can be achieved. For intermediate α, m_0 (e.g. $\alpha = 0.5$, $m_0 = 0.4$) the influences of Gaussian and discrete noise are structurally similar. Contrary to the comment in [5] we point out that Hebb networks with discrete

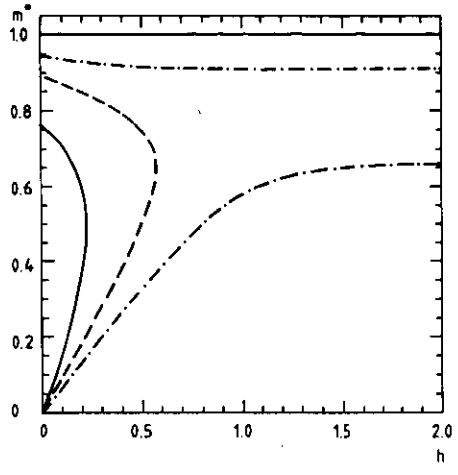

Figure 2. Fixed points of (12) for the pseudo-inverse rule with hidden units; $m_0 = 0.4$, $\alpha = 0.5$ (—), $\alpha = 0.6$ (– – –), $\alpha = 0.7$ (— · —). Note that $m^* = 1$ is a stable fixed point for all these cases.

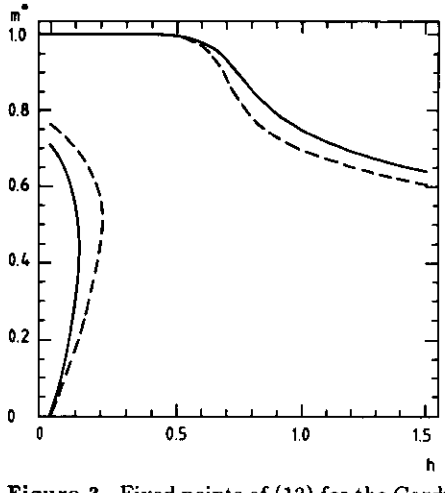

Figure 3. Fixed points of (12) for the Gardner (—) and pseudo-inverse rule (– – –) with Gaussian noise; $\alpha = 0.5$, $m_0 = 0.4$.

noise at very low loading also exhibit unstable fixed points, as shown in figure 4 (cf also [4], figure 1). If, however, we apply Gaussian noise in this case we find that the first-order transition is replaced by a continuously decreasing curve. As mentioned in subsection 3.1.1 the appearance of certain first-order transitions is a peculiarity of the noise distribution.

3.1.3. The Gardner rule. For both the case of discrete and Gaussian noise the Gardner rule gives results which are similar to the ones for the pseudo-inverse rule. Figure 3 compares the results for the two rules for Gaussian noise. For the Gardner rule and discrete noise the strong retrieval attractor exists for $\alpha < 2$, whereas for the pseudo-inverse rule however only for $\alpha < 1$. Nevertheless the two rules show almost the same

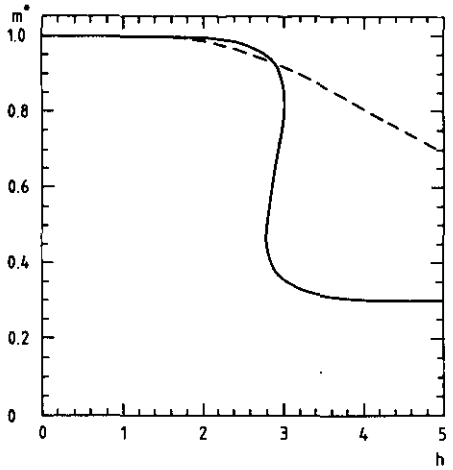

Figure 4. Fixed points of (12) for the Hebb rule with discrete (—) and Gaussian noise (– – –); $\alpha = 0.05$, $m_0 = 0.3$.

behaviour as the basins of attraction are extremely narrow for high loading.

3.2. Retrieval behaviour diagrams

In order to illustrate similarities and dissimilarities of the different learning rules, we compare here the retrieval behaviours of the systems in the α–h space. Since perfect ($m^* = 1$) and almost perfect retrieval lead essentially to the same performance, we consider a boundary factor m_b which, with an appropriate choice, gives a good qualitative guide to the retrieval properties of the system. We thereby distinguish three different regimes: strong (s) retrieval ($m^* > m_b$), weak (w) retrieval ($m_b > m^* > m_0$) and bad (b) retrieval ($m^* < m_0$), where m^* is the attractor overlap to which m_0 converges.

The following figures with $m_b = 0.9$ show these three different performance types in the α–h plane. For continuous transitions of the unblocked perfect retrieval phase (e.g. cases in figure 3), the strong and weak retrieval regimes are separated by the contour of constant attractor overlap m_b; for discontinuous transitions of the unblocked perfect retrieval phase and, in addition, for the bifurcation overlap at the transition below m_b (e.g. cases in figure 1), the strong and weak retrieval regimes are now separated by the generic first-order transition line. In this case, the separator of the two regimes is insensitive to the choice of m_b.

On the other hand, the separator of the weak and bad regimes is a contour of constant attractor overlap m_0: it is an arbitrary indication line to provide information on the retrieval performance, and is therefore different from a phase transition line.

Figure 5 compares the results for discrete and Gaussian noise for the pseudo-inverse rule. It can be observed that in the absence of h the basins start to shrink for $\alpha \approx 0.4$ ($m_0 = 0.4$). We also observe that for intermediate fields $h \approx 0.4$ the storage can be extended to a maximum $\alpha \approx 0.5$. The shape of the separating curve between region (s) and (w) implies that there is only a small band of h values which widen the basins sufficiently to achieve unblocked perfect retrieval for high loading. The results for Gaussian and discrete noises are similar, in that a change in m_0 results in a shift of the curves to the right for increasing m_0 and to the left for decreasing m_0. For the

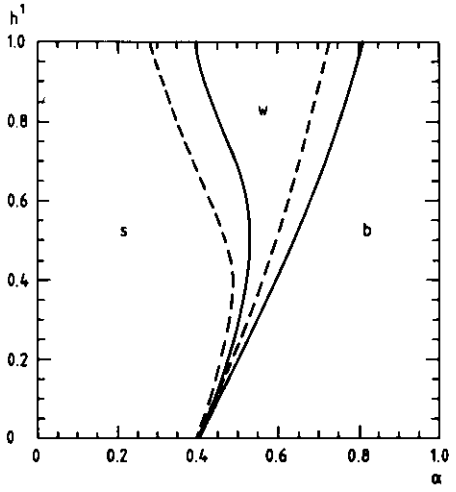

Figure 5. Retrieval phase diagram in the $\alpha - h$ space for the pseudo-inverse rule with Gaussian (—) and discrete noise (– – –) showing the different regimes (cf text); $m_0 = 0.4$.

case of discrete noise one can easily show that the region of strong retrieval is bounded from above by the curve $h_c(\alpha) = \sqrt{1/\alpha - 1}$ which determines the disappearance of the perfect attractor. It should be emphasized that the curve separating strong and weak retrieval is in the case of discrete noise a generic phase transition line, and is therefore independent of m_b provided that m_b is not too small. However, in the case of Gaussian noise it is merely a performance contour and is dependent on m_b.

Figure 6 shows the result for the Hebb rule with Gaussian noise. We observe that now due to the large basins of attraction the band $[h_<, h_>]$ is much wider. On the other hand the maximal loading level for strong retrieval is much smaller compared to, for example, the Gardner rule. To illustrate the dependence of the boundary value m_b we give the results for $m_b = 0.9$ and 0.95.

In figure 6 we also show the results for the pseudo-inverse rule with hidden units. Two remarks should be made. First, the upper limit of the band of h giving perfect retrieval disappears as implied by figure 2. As a result, the separator between the strong and weak regimes (again a generic phase line for sufficiently high m_b) becomes monotonic in α–h space. Furthermore, since the strong retrieval regime is completely blocked for all h when α exceeds $\alpha_b = 0.655$, this separator approaches the line $\alpha = 0.655$ asymptotically from below as $h \to \infty$. Secondly, the maximal α for which we can still retrieve perfectly is rather high (0.655, as compared to 0.5 in figure 5).

Summarizing these results, we conclude that external fields have a very similar influence on the network performance independently of the learning rule or the type of noise in the stimulus. However, many discontinuous transitions observed in the presence of discrete noise disappear when smoother, e.g. Gaussian, distributions are applied.

3.3. Results for optimally adapted retrievers

Before proceeding, a short comment on the numerical analysis of (23) should be made. Since (20), (21) and (23) all depend on m^* they need to be solved simultaneously, which is a very cumbersome procedure. In order to find $m_h^*(\alpha)$ we use the following

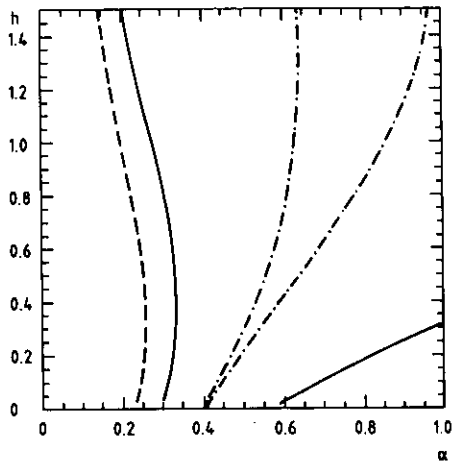

Figure 6. Retrieval phase diagram in the α–h space for the Hebb rule with Gaussian noise ($m_b = 0.9$ (—), 0.95 (- - -)) and for the pseudo-inverse rule with hidden units (— · —) showing the different regimes (cf text); $m_0 = 0.4$.

scheme. We fix h and m^* and solve (23) and (20) for γ. Once γ is found α can easily be determined via (21). In the next step m^* is fixed to another value and the procedure starts again. As an initial guess the zero temperature retriever curve given in [8] (figure 3 therein) is very helpful as it coincides with the present retriever curve for $h = 0$.

Figure 7 shows $m^*(\alpha)$ for different field strengths. Analogously to [8] we point out that solutions of (23) do not correspond to single retrieval mappings but should be seen as attractors of self-adaptation. We are primarily interested in how far one can increase the loading α_{opt}, if we require almost perfect retrieval with a radius of the basins of attraction which is almost one.

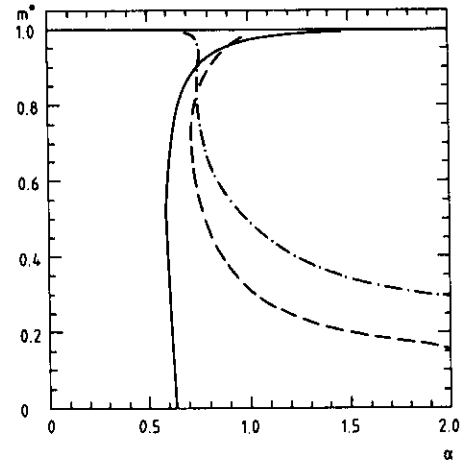

Figure 7. Attractors of adaptation for an optimized system ($m_0 = 0.4$) with Gaussian noise and different fields; $h = 0$ (—), $h = 0.2$ (- - -), $h = 0.4$ (— · —).

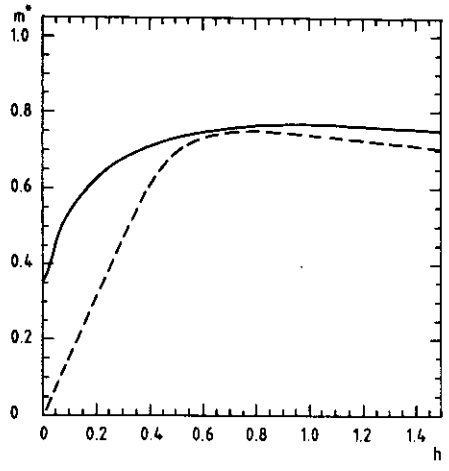

Figure 8. Asymptotic overlap for Hebb rule (—) and pseudo inverse (– – –) rule at temperature $T = 0.6$; $\alpha = 0.45$, $m_0 = 0.5$.

The initial overlap for the example we consider is again taken to be $m_0 = 0.4$. We observe that for a small field of e.g. $h = 0.2$ the critical loading can be extended to $\alpha_{\text{opt}} \approx 0.7$ and for $h = 0.4$ to $\alpha_{\text{opt}} \approx 0.75$. However we point out that, in the same way as mentioned before, high stimulus fields and Gaussian noise tend to lower the perfect ($m^* = 1$) attractor of self-adaptation continuously with h. The corresponding results for maximally stable networks (MSN, or Gardner rule) yields critical loadings where the basins start to shrink at $\alpha_{\text{MSN}}(h = 0.2) = 0.57$ and $\alpha_{\text{MSN}}(h = 0.4) = 0.63$, which are smaller than the ones for the optimally adapted network.

In conclusion, from our analysis we can say that the principle of adaptation turns out to have general utility. Indeed, this principle guarantees that the best performance is obtained among the set of all dilute networks [8].

3.4. Finite-temperature behaviour

Finite-temperature dynamics without external fields in dilute neural networks has been considered in great detail by [17]. Recent work [8] on optimized neural networks showed that in the high-temperature regime Hebbian synapses give the maximum storage capacity. These results were obtained in the absence of an external field. Our investigations show that this behaviour persists for external fields. Figure 8 shows two typical retrieval curves. It can be seen that the one for the Hebb rule exceeds the one for the pseudo-inverse rule. The noise-optimal behaviour of the Hebb rule makes the retrieval with external field more stable against disruptions through noise.

4. Conclusions

We have studied the retrieval properties of a highly dilute network trained with very general classes of learning rules. We imposed a noisy external stimulus, where also the noise distributions were taken from structurally different classes. The radius of the basins of attraction is in general widened for intermediate field strengths. For optimal learning rules, retrieval is possible even beyond their usual field-free storage limits.

Through the application of different types of external noise we could show that certain first-order phase transitions occur only for special noise distributions and should thus be taken cautiously for general considerations.

We extended the investigation of [8] to external fields, thus showing that the principle of adaptation can be applied successfully to optimize the performance of a network.

The introduction of temperature yielded that the Hebb rule retrieves in the high-temperature regime also in the presence of an external field better than other learning rules which are more suitable for zero-temperature retrieval, such as the pseudo-inverse rule. This is another illustration of the principle of specialization formulated in [8].

Acknowledgments

We thank the SERC and the Studienstiftung des deutschen Volkes for financial support. We also acknowledge several fruitful discussions with T Watkin.

References

[1] Gardner E 1989 *J. Phys. A: Math. Gen.* **22** 1969
[2] Amit D J, Parisi G and Nicolis S 1990 *Network* **1** 75
[3] Engel A, Englisch H and Schütte A 1989 *Europhys. Lett.* **8** 393
[4] Rau A and Sherrington D 1990 *Europhys. Lett.* **11** 499
[5] Engel A, Bouten M, Komoda A and Serneels R 1990 *Phys. Rev.* A **42** 4998
[6] Abbott L T and Kepler T B 1988 *J. Phys. A: Math. Gen.* **22** 2031
[7] Wong K Y M and Sherrington D 1990 *J. Phys. A: Math. Gen.* **23** L175
[8] Wong K Y M and Sherrington D *J. Phys. A: Math. Gen.* **23** 4659
[9] Derrida B, Gardner E and Zippelius A 1987 *Europhys. Lett.* **4** 167
[10] Kepler T B and Abbott L F 1988 *J. Physique* **49** 1657
[11] Kree R and Zippelius A 1990 *Preprint*
[12] Krauth W, Nadal J-P and Mézard M 1988 *J. Phys. A: Math. Gen.* **21** 2995
[13] Gardner E, Derrida B and Mottishaw P 1987 *J. Physique* **48** 741
[14] Peretto P 1984 *Biol. Cyb.* **50** 51
[15] Personnaz L, Guyon I and Dreyfus G 1985 *J. Physique* **46** L359
[16] Kanter I and Sompolinsky H 1987 *Phys. Rev.* A **35** 380
[17] Amit D J, Evans M, Horner H and Wong K Y M 1990 *J. Phys. A: Math. Gen.* **23** 3361