Parallel dynamics of the neural network with the pseudoinverse coupling matrix

1991 J. Phys. A: Math. Gen. 24 2201

(http://iopscience.iop.org/0305-4470/24/9/026)

J. Phys. A: Math. Gen. 24 (1991) 2201-2218. Printed in the UK

Parallel dynamics of the neural network with the pseudoinverse coupling matrix

Rolf D Henkel and Manfred Opper

Institut für Theoretische Physik III, Justus-Liebig-Universität Giessen, Heinrich-Buff-Ring 16, 6300 Giessen, Germany

Received 3 September 1990

Abstract. We investigate the parallel dynamics of the neural network with the pseudoinverse coupling matrix. Based on an exact dynamic theory we develop an approximate treatment for long timescales. The temporal development of the overlap, of correlation functions and of the remanent magnetization are investigated. We present results for deterministic and stochastic dynamics. Large-scale numerical simulations supplement our analytical findings.

1. Introduction

Neural networks are now regarded as a promising tool to extend the capabilities of present-day computers. After the revived interest into these large arrays of multiple-connected, simple processing units due to the work of Hopfield [1], a lot of different network structures have been put forward [2, 3]. One of the simplest types of neural networks consists of two-state neurons $S_i = \pm 1$, $(i = 1, \ldots, N)$, ('spins'), which change their internal states according to a prescribed update rule. The neurons are connected to each other via the synaptic coupling matrix J_{ij}, in which the knowledge of the neural network is coded.

One possible update rule—and the one we will be primarily concerned with in this paper—is the synchronous, parallel update of the neurons S_i at discrete time intervals [4], defined by

$$S_i(t+1) = \operatorname{sgn}(h_i(t)) \tag{1}$$

where

$$h_i(t) = \sum_j J_{ij} S_j(t)$$

are the internal fields at time t. Another type of dynamics is the asynchronous update, where the spins are updated according to (1), but in a fixed or random sequential order. Some results for this update will be given in this paper too.

If the synaptic couplings J_{ij} of the neural network are chosen properly, the network is able to learn a number of patterns $\xi_i^\nu = \pm 1$, $(i = 1, \ldots, N;\ \nu = 1, \ldots, p)$. Then, under the dynamics defined by (1), the network can restore noisy, corrupted input images of the learned pattern. It functions as an autoassociative memory.

We will consider in this paper the network with the pseudoinverse coupling matrix [5]. The pseudoinverse matrix is easy and fast to calculate, directly [6] or iteratively [7, 8]. It is a projection operator into the space of the learned pattern. The coupling matrix J_{ij}, with zeroed diagonal element, stabilizes the pattern up to a storage ratio of

0305-4470/91/092201+18$03.50

$\alpha = p/N = 1$. We have, for $N \to \infty$,

$$\sum_j J_{ij}\xi_j^\mu = (1-\alpha)\xi_i^\mu. \tag{2}$$

The dynamics of the network will be described in terms of the overlap between the state $S_i(t)$ of the neurons at time t and a pattern ξ_i^μ:

$$m(t) = \frac{1}{N}\sum_i \xi_i^\mu S_i(t).$$

Clearly, $m = 1$ is equivalent to $S_i = \xi_i^\mu$ for all i.

The actual performance of the network will depend on the value of the start overlap $m(0)$ at time $t = 0$ and on the storage ratio α of the network. Starting from an initially noisy state $m(0) < 1$, the network can, for $t \to \infty$:

(i) recognize the pattern, $m(t) \to 1$;

(ii) partly restore the pattern, $m(0) < m(t) < 1$; or

(iii) fail to recognize the pattern, $m(t) < m(0)$.

Usually the overlap assumes a continuum of values, depending on the initial conditions. This is caused by metastable states and 2-cycles, which trap the network dynamically before reaching the attractors.

Within equilibrium theories of statistical physics [9, 10], which describe only the thermodynamically stable attractors of the network, these details of the dynamics are not covered. Furthermore, it is hard to derive results for the transient behaviour of neural networks [11–13] from exact dynamic mean-field theories [14–16].

The neural network with pseudoinverse coupling matrix is of the 'fully connected' type, i.e. every neuron is connected to each other neuron. In this type of network long-time correlations build up and lead to a rapidly increasing number of order parameters in exact treatments. This renders exact approaches useless after a few time-steps [14, 15]. Often one considers a strongly diluted version of the neural network in question [2, 17, 18], i.e. a network where a given neuron is *not* connected to an extensive number of other neurons. This assumption simplifies matters drastically in terms of parameters to handle, but also the rich structure in the dynamic behaviour of the fully connected network is lost.

In this paper we will not resort to this 'diluted-network' approach. Instead, after careful investigation of the exact mean-field theory, we will develop a new approximate treatment [19] able to reproduce the rich dynamic behaviour of the fully connected neural network.

The paper is organized as follows. In section 2 we develop the exact dynamic mean field approach to the dynamics, which will serve as a starting point for our approximate theory presented in section 3. In section 4 we will discuss the dynamics of the neural network within the framework of our approximate treatment. In section 5, our dynamic approach will be extended to noisy dynamics. Supplementary numerical simulations will be given, including large-scale numerical simulation with noisy dynamics, along the discussion of the analytical results of sections 4 and 5.

2. The distribution of the internal fields—exact results

We base our analysis on the probability distribution of the internal field $h_i(t)$. Since we are interested in the generic dynamic behaviour of the network, we will average over the set of $p = \alpha N$ stored random patterns ξ_i^ν and over the initial conditions.

The internal fields become random variables with respect to the initial conditions setting $\xi_i^1 = 1$, we have

$$m(t+1) = \frac{1}{N}\sum_i \xi_i^1 S_i(t+1) = \int \mathrm{d}h\, P_t(h)\,\mathrm{sgn}(h) \tag{3}$$

which defines

$$P_t(h) = \frac{1}{N}\sum_i [\delta(h - h_i(t))]_{S_i(0)} \tag{4}$$

the averaged distribution function of the internal fields.

In order to calculate this probability distribution, we introduce the generating function†

$$[Z(\underline{l})]_{S_i(0)} = \left[\mathrm{Tr}_{S_i(t)} \int \prod_{it} \left(\mathrm{d}h_i(t)\Theta(S_i(t+1)h_i(t))\delta(h_i(t) - \sum_{j\neq i} J_{ij}S_i(t)) \right) \exp\left(\mathrm{i}\sum_{it} l_i(t)h_i(t) \right) \right]_{S_i(0)}.$$

Here $\underline{l}$ is an abbreviation for the vector $\underline{l} = (l_1, \ldots, l_N)$. The trace extends over all spins $S_i(t)$ for $t > 0$ and results in $Z(\underline{0})$ being normalized to unity.

In the case where the network has stored $p = \alpha N$ random, uncorrelated patterns ξ_i^ν the distribution $P_t(h)$ becomes self-averaging and we can replace the site-average of (4) by the average over the couplings J_{ij}. Using similar methods as in [21], we show in appendix 1 that this average can be replaced in the limit $N \to \infty$ by an average over a Gaussian noise $\underline{w}$:

$$[Z(\underline{l})]_J \sim \left[\mathrm{Tr}_{S_i(t)} \int \prod_{it} (\mathrm{d}h_i(t))\delta\left(N^{-1}\sum_i h_i(t) - (1-\alpha)S_i(t) \right) \times \exp\left(\mathrm{i}\sum_{it} l_i(t)h_i(t) \right) \prod_{it} (\Theta(S_i(t+1)h_i(t))) \times \prod_{it} \left(\delta\left(h_i(t) + \alpha S_i(t) - w_i(t) - \sum_\tau K_{t\tau} f_i(\tau) - a_1(t) \right) \right) \right]_{\underline{w}}.$$

This generating function corresponds to a dynamic system for N uncoupled spins:

$$S_i(t+1) = \mathrm{sgn}(h_i(t)) \tag{5}$$

$$h_i(t) + \alpha S_i(t) = m(t) + \sum_\tau K_{t\tau} f_i(\tau) + w_i(t) \tag{6}$$

$$f_i(t) = h_i(t) - (1-\alpha)S_i(t). \tag{7}$$

The correlations of the Gaussians $w_i(t)$ are given by (overbars indicate averages)

$$\overline{w_i(t)w_i(\tau)} = \alpha\overline{a_\mu(t)a_\mu(\tau)} \tag{8}$$

where the auxiliary variables $a_\mu(t)$ are the random overlaps with the other patterns. They are obtained from

$$a_\mu(t) = -\alpha^{-1}\sum_\tau K_{t\tau} v_\mu(\tau). \tag{9}$$

† A similar function was used in [20] to study the parallel dynamics of diluted networks.

The $v_\mu(t)$ are additional Gaussian noise terms with zero mean and correlations fixed by the projections $f_i(t)$ in the space orthogonal to the patterns

$$\overline{v_\mu(t)v_\mu(\tau)} = \overline{f_i(t)f_i(\tau)}. \tag{10}$$

The kernel $K_{t\tau}$ is obtained from the equation

$$\sum_{t'} K_{tt'}\hat{K}_{t'\tau} = -\alpha\delta_{t\tau} \tag{11}$$

with the response function

$$\hat{K}_{t\tau} = \overline{\frac{\partial f_i(t)}{\partial w_i(\tau)}}. \tag{12}$$

Equations (5)–(12) together with (3) can be solved in principle to yield the exact field distribution at successive time-steps. Exact results for the first few time-steps will be given in appendix 2.

The J_{ij} average has made all spin sites equivalent, i.e. correlations between spins in the initial conditions are destroyed by the average over the pattern. Expanding (6), we find (dropping indices i and μ and indicating time-steps through subscripts) the following general structure for the internal field at time t:

$$h_t = \tilde{z}_t + \sum_{\tau<t} \tilde{K}_{t\tau}S_\tau. \tag{13}$$

Thus h_t is a sum of a Gaussian $\tilde{z}_t$ with a memory term $\Sigma_{\tau<t}\tilde{K}_{t\tau}S_\tau$. The $\tilde{K}_{t\tau}$ are sums of products of the original $K_{t\tau}$.

The two contributions of (13) have simple physical interpretations in the cavity approach of Mézard *et al* [22]. Adding a new spin together with its couplings to a system of N spins, there will be a part of the internal field coming from the magnetizations of the unperturbed system, the Gaussian field $\tilde{z}_t$, and a response term $\Sigma_{\tau<t}\tilde{K}_{t\tau}S_\tau$, caused by the polarization of the N spins due to the presence of the new spin at previous times τ. It is this structure of the internal field that will serve as the starting point of our approximate treatment presented in the following section. We expect that the structure of (13) holds for a broad class of networks [23, 24].

Clearly, each new time-step introduces within the exact theory a new Gaussian noise term w_t, including its correlations to the other noises, and new $\hat{K}_{t\tau}$ and $K_{t\tau}$. The number of equations to solve grows rapidly with time. Carrying out these exact calculations becomes unfeasible after a few time-steps. All interesting questions, however, such as areas of attraction, remanence effects, etc., require the consideration of longer timescales, typically of the order of 10–30 time-steps. In order to handle these longer timescales, one has to develop approximate treatments with drastically reduced numbers of dynamic parameters.

3. Approximating the internal fields—the double-peak dynamics

The internal field at an arbitary time is composed of a Gaussian noise term plus a memory term, depending on the values of the spin at previous times. It is this memory term, which is the source of all remanence effects encountered in the dynamics of neural networks. It manifests itself in a pronounced double-peaked structure of the probability distribution of the internal fields.

We introduce now the simplest non-Gaussian approximation—termed *double-peak dynamics* (DPD)—which is capable of approximating the exact probability distribution

of the internal fields by replacing the exact random variable h_t of (13) with

$$h_t = u_t + d_t S_{t-1}. \tag{14}$$

Here u_t is an effective Gaussian noise assumed to be uncorrelated with S_{t-1} and the total effect of the memory term $\Sigma_{\tau<t} \tilde{K}_{t\tau} S_\tau$ in the exact expression (13) is supposed to be summarized by the term $d_t S_{t-1}$. The Gaussian noise u_t is thus not identical to $\tilde{z}_t$ of (13) and d_t is a renormalized self-coupling, not equal to $\tilde{K}_{tt-1}$. The probability distribution $P_t(h)$ resulting from our ansatz (14) consists of two Gaussian peaks of width Δu_t^2, separated by a distance $2d_t$.

Using only the value of the spin one time-step before, S_{t-1}, in the memory term of our ansatz, is exact for times $t = 0$, 1. On the other hand, for large times t, strong correlations exist between S_{t-1} and S_{t-2}, S_{t-3}, etc. They are partially taken into account by the renormalization of d_t. As we will show in the following, the inclusion of a memory term in this simple fashion is sufficient to describe the dynamics of a neural network faithfully.

Our ansatz has only three parameters: $\overline{h_t}$, $\overline{\Delta h_t^2}$ and d_t. We will derive, in the following, recursion relations for these dynamic quantities. Surprisingly, for models with symmetric interactions, i.e. $J_{ij} = J_{ji}$, the renormalized self-coupling d_t can be determined self-consistently from our ansatz.

Multiplying (14) by S_{t-1} and averaging we find for d_t

$$\begin{aligned} d_t &= \overline{S_i(t-1)h_i(t)} - \overline{u_i(t)S_i(t-1)} \\ &= N^{-1} \sum_{i,j} \overline{S_i(t)J_{ij}S_j(t-1)} - \overline{u_i} m_{t-1}. \end{aligned} \tag{15}$$

Summing over j and using the symmetry of the matrix, the first term in (15) simplifies to

$$\begin{aligned} N^{-1} \sum_i \overline{S_i(t)h_i(t-1)} &= N^{-1} \sum_i \overline{\operatorname{sgn}(h_i(t-1))h_i(t-1)} \\ &= \int P_{t-1}(h)|h| \, \mathrm{d}h. \end{aligned}$$

This yields finally for d_t the equation

$$\begin{aligned} (1 - m_{t-1}^2) d_t &= \int P_{t-1}(h)|h| \, \mathrm{d}h - \overline{h_t} m_{t-1} \\ &= \overline{|h_{t-1}|} - \overline{h_t} m_{t-1}. \end{aligned} \tag{16}$$

The remaining task is to calculate the first two moments of the field distribution, $\overline{h_t}$ and $\overline{\Delta h_t^2}$. Only here the specific type of the neural network enters our analysis. For the neural network in question, with the pseudoinverse coupling matrix, $\overline{h_t}$ is given exactly by (cf (2))

$$\overline{h_t} = (1 - \alpha) m_t \tag{17}$$

and $\overline{\Delta h_t^2}$ can be approximated by (see appendix 3)

$$\begin{aligned} \overline{\Delta h_t^2} &= \alpha(1-\alpha)(1-m_t^2) \\ &\quad + \frac{d_t^2 (1 - m_{t-1}^2)^2}{\overline{\Delta h_{t-1}^2}} \left[(1-2\alpha)^2 + \alpha(1-\alpha)\left(1 - \frac{\alpha(1-\alpha)}{\overline{\Delta h_{t-1}^2}}(1 - m_{t-1}^2)\right)\right] \end{aligned} \tag{18}$$

The last equation is exact for $\alpha = \frac{1}{2}$ and exhibits symmetry around $\alpha = \frac{1}{2}$ for zero overlap.

4. Results

By iterating the DPD equations one directly obtains m_t and the probability distribution $P_t(h)$. From $P_t(h)$ various other quantities can be calculated, some of which will be presented in this section. We will discuss basins of attraction, temporal correlations and various remanence effects. Supplementary numerical data will serve as a validation of our approach.

A first, crucial check of the validity of our ansatz is the comparison of the DPD distribution with the actual field distribution at a late time. In figure 1 we display numerical data at time $t = 10$ and $\alpha = \frac{1}{2}$. The start overlap was $m_0 = \frac{1}{2}$. The pronounced double-peaked structure of the field distribution is clearly seen. The DPD prediction (smooth curve) and the numerical data agree quite well.

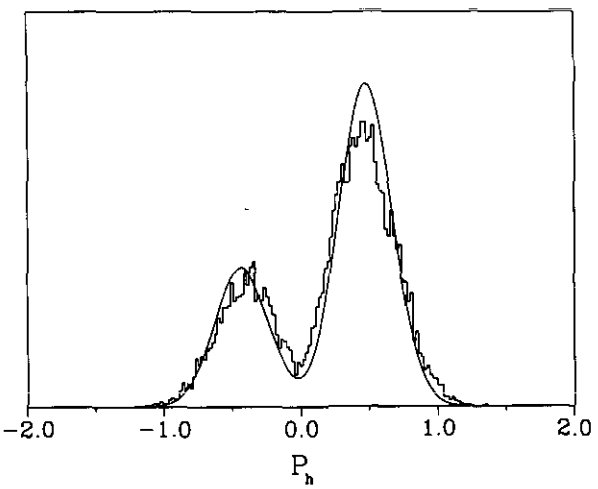

Figure 1. The distribution of the internal fields at $t = 10$. Shown are the prediction of the DPD and a distribution obtained from numerical simulations. The numerical data was obtained from 128-spin systems at $\alpha = \frac{1}{2}$, starting with an overlap of $m_0 = \frac{1}{2}$.

We turn now to the dynamics of the neural network. It depends strongly on the start overlap m_0. For an overview, we present in figure 2 the flow of the network dynamics at $\alpha = \frac{1}{2}$ and various start overlaps m_0. We observe fast relaxation, already at times $t \approx 10$–30 the dynamics have basically stopped. Note that the obtained final overlaps depend strongly on the start overlaps m_0; this is the first occurrence of a remanence effect within our DPD approach. The numerical data, which we have included in figure 2 for comparison with the DPD prediction, shows the same behaviour.

To summarize the behaviour of the network at different α-values, we display in figure 3 the overlap at $t = 30$ as a function of m_0. For all three m_t/m_0 curves there exists a critical overlap m_c (the 'edge of the cliff') above which the noisy input pattern gets completely restored: $m_t \to 1$. This defines the boundary of the area of attraction.

For values m_0 below m_c, the input pattern does not get restored, and the dynamic of the network depends strongly on α. At $\alpha = 0.5$, the network stays close to the start overlaps m_0 as already discussed. At lower α-values, the system flows always towards the pattern, but gets trapped before reaching $m = 1$ (cf the $\alpha = 0.3$ curve in figure 3). The dynamic traps are metastable states and 2-cycles [25].

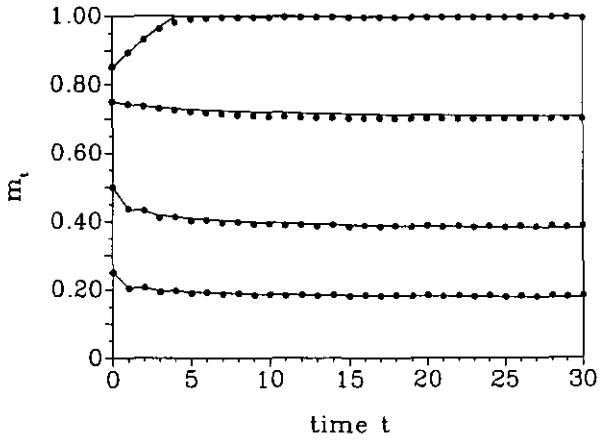

Figure 2. Trajectories m_t for several starting values m_0 at $\alpha = 0.5$. The lines are the prediction of the DPD, the dots numerical data from 256-spin systems, averaged over 1024 samples per trajectory.

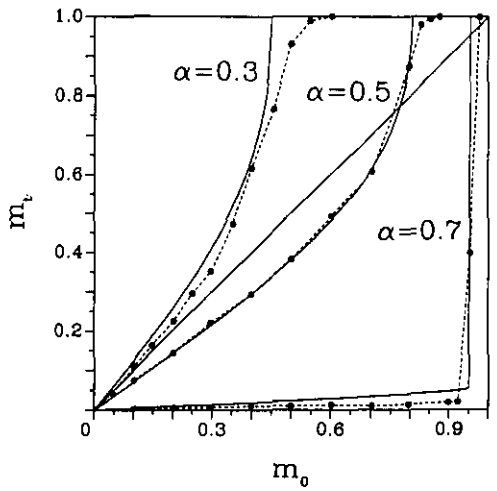

Figure 3. The overlap as a function of the initial overlap m_0. Displayed are the DPD predictions for $\alpha = 0.3$, 0.5 and 0.7 at time $t = 30$ (full lines). The supplementary numerical data (broken lines) were obtained from 256-spin systems. Each data point is an average over 100 samples.

At α-values larger than 0.5, a quite different dynamic behaviour is found (figure 3 displays $\alpha = 0.7$). Here, m_t develops basically towards 0 or 1 only. This seems to indicate that in the high α-region, the traps encountered for $\alpha \leqslant 0.5$ do not exist or are dynamically unimportant.

This pronounced dependence of the network dynamics on the storage ratio α is unique to the synchronous dynamics. Under asynchronous dynamics—which we have investigated numerically—we find remanent overlap at all α-values, including values of $\alpha > 0.5$ [23].

Looking at the critical overlap m_c as a function of α, we obtain the area-of-attraction plot (figure 4). Starting at a given α with an overlap m_0 above the solid line in

figure 4, the input pattern gets restored. For start overlaps below m_c, the dynamics of the network depends on α as discussed. We find that the area of attraction shrinks quickly to a small patch around the patterns for values higher then $\alpha \approx 0.5$. Large areas of attraction are only found in the low α-region, and a full basin of attraction is obtained only in the $\alpha \to 0$ limit.

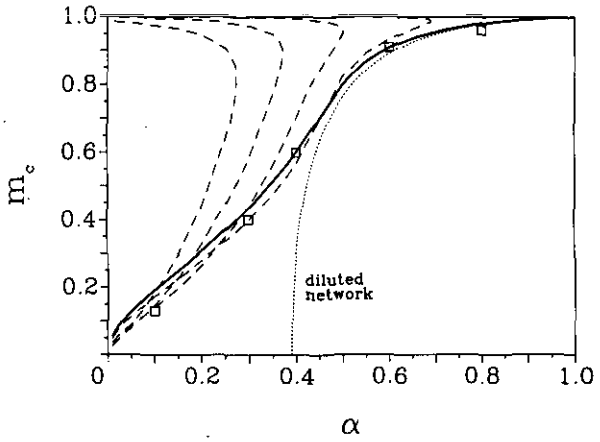

Figure 4. The area of attraction for a network with the pseudoinverse coupling matrix (full line). Included is a heuristic rule, discussed in the text, and numerical data for the area of attraction (□, from [29]). The broken lines show the area of attraction at temperatures 0.4, 0.3, 0.2 and 0.1 (from left to right). Note that no remarkable increase in the area of attraction is observed under noisy dynamics.

Included in figure 4 is the prediction for the area of attraction of the 'diluted-network' approach. This rule works well only in the vicinity of the patterns [18]. For low α-values, it predicts a full area of attraction, which is wrong. The diluted theory is a one-parameter theory and has only the trivial fixed points 0 and 1 in the case of the pseudoinverse coupling matrix. Thus, if the dynamics of the diluted network flows towards the pattern in the first time-step, it converges to $m = 1$. This is, however, always the case in the low α-region. The fact that the system may get trapped in metastable states or 2-cycles is missed within this approach.

Presenting the network with an input pattern with no overlap to one of the stored patterns, the overlap m_t is not a good dynamic parameter, since it stays zero all the time. We can gain nevertheless insight in the temporal development of the neural network by considering instead the correlation function $c_{t,t-2}$. It is given within our DPD approach as

$$c_{t,t-2} = \operatorname{erf}\left(\frac{d_{t-1}}{\sqrt{2\Delta u_{t-1}^2}}\right)$$

assuming $m_0 = 0$. We find (see figure 5) for the function $1 - c_{t,t-2}$ an approximate power-law decay similar to the numerical findings of Gardner *et al* [14] in the case of the SK model of spin glasses. From the DPD approach, the decay exponent for the pseudoinverse is given by ≈ -1.3. Numerically, we find a decay exponent of ≈ -1.4 as compared to $-\frac{3}{2}$ in the case of the SK model.

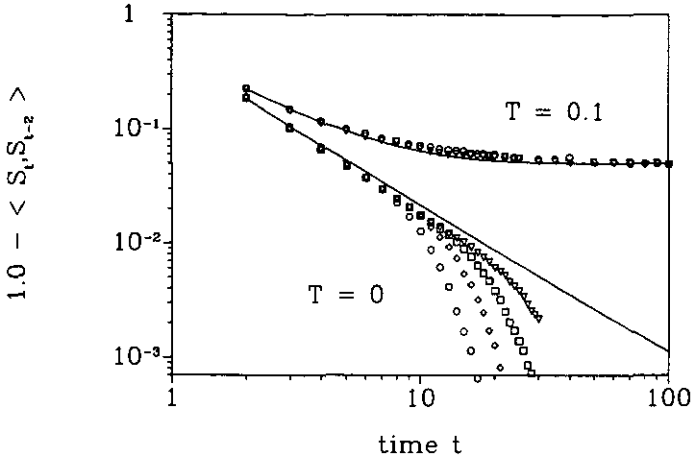

Figure 5. $1-\overline{S_i(t)S_i(t-2)}$ as a function of time. Lines are the prediction of the DPD. The numerical data for zero temperature was obtained with system sizes of 128 (○), 256 (◇), 512 (□) and 1024 (▽) spins. For temperature $T=0.1$, system sizes of 128 and 1024 spins were used. Sample sizes are 100 ($T=0.1$) and 1024 ($T=0.0$) samples per data point.

Our supplementary numerical data included in figure 5 departs from the power law decay of $1-c_{t,t-2}$ after a certain time. This time is strongly size dependent and increases with system size. We think this departure from the power law behaviour can be viewed as a good indicator for the onset of finite-size effects in time-dependent quantities of fully connected networks. Note that even for quite large systems ($N=1024$) the deviation from the power law occurs at relatively short times ($t\approx 20$).

Another quantity which can be investigated in the case of zero overlap of the patterns is the remanent magnetization $c_{t,0}$. It is the projection of the state vector $S_i(t)$ at time t onto the initial state $S_i(0)$. Using a gauge transformation of the dynamics, we can transform $S_i(0)$ onto the all-one pattern $\xi_i^1=1$ and identify $c_{t,0}$ with m_t in our usual dynamic approach. The transformation does not change the gauge-invariant quantities d_t and $\overline{h_t^2}$, i.e. d_t is still given by (16) and $\overline{h_t^2}$ by (18) with $m_t=0$. Since the all-one pattern ξ_i^1 is, however, no longer explicitly stored in the coupling matrix J_{ij}, the remaining dynamic parameter, namely $\overline{h_t}$, can not be calculated as in (17). We use instead the unbiased estimate $\overline{u_t}=0$, which is consistent with the first two time-steps.

Under this heuristic approximation, the remanent magnetization $c_{t,0}$ factorizes in time, e.g.

$$c_{2t,0}=\prod_{\tau=1}^{t} c_{2\tau,2\tau-2}.$$

We find that the remanent magnetization at even time-steps has a maximum at $\alpha=0.5$, and falls off towards zero at low and high α-values (figure 6). At odd time-steps, zero remanent magnetization for all α-values is predicted. Thus the remanent magnetization of the neural network with the pseudoinverse coupling matrix should show oscillatory behaviour like the SK model of spin glasses under parallel update [4].

Our numerical simulations (see figure 6) confirm the DPD prediction for even time-steps, but zero remanent magnetization at odd time-steps is only found for $\alpha=\frac{1}{2}$. At low α-values, the predicted oscillation in the remanent magnetization is not found. For increasing α, however, the remanent magnetization shows the expected oscillatory

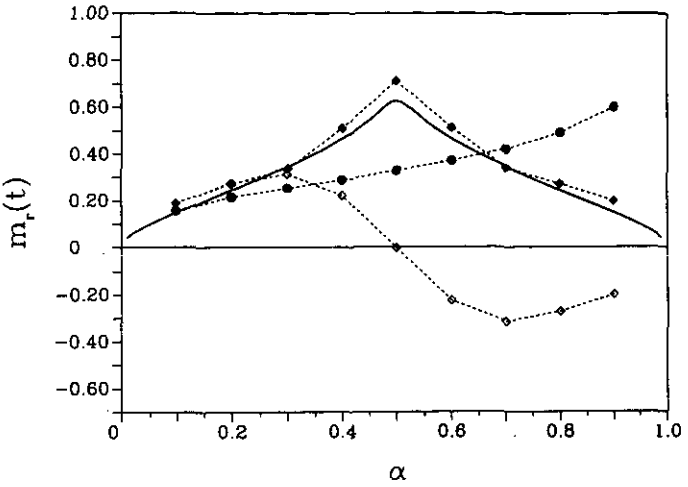

Figure 6. The remanent magnetization as a function of α. Full lines are the predication of the DPD. The numerical data for parallel update (◆, even time-steps, ◇ odd time-steps) shows symmetry/asymmetry around $\alpha = 0.5$. No symmetry is found in the numerical data for serial update (●). All numerical data points are averages over 1024 samples of 256-spin systems.

behaviour and, in the high α-region, even a negative remanent magnetization at odd time-steps is observed!

This oscillation of the remanent magnetization in the high α-region is a very interesting effect which could be used to design a fault-tolerant novelty detector. Normally, in order to decide if a noisy input pattern is one of the patterns stored in the neural network, one would have to check after the recall phase all $p = \alpha N$ overlaps of the patterns. Alternatively, it is sufficient just to monitor the remanent magnetization $c_{t,0}$ for a few time-steps during the restoration process. If the remanent magnetization shows an oscillatory behaviour, the input pattern is currently not part of the memory. If the remanent magnetization assumes a constant value, the input pattern is being successfully restored. We expect that this effect can be enforced by appropriate network design.

We have also included in figure 6 the remanent magnetization for the asynchronous update rule. It exhibits, as expected, a quite different behaviour than the one found for the synchronous update. No oscillation and no symmetry around $\alpha = \frac{1}{2}$ is observed. The remanent magnetization increases with α and suggests an abundance of dynamic traps in the high α-region. This view is supported by our observation of strong *remanence effects* in the overlap m_t for high αs under asynchronous dynamics.

5. Dynamics with external noise

It is easy to extend our DPD approach to noisy dynamics. The update is now defined via

$$S_i(t+1) = \operatorname{sgn}(h_i(t) + r_i(t))$$

where $r_i(t)$ are random variables considered to represent fast synaptic noise. Choosing

for $r_i(t)$,

$$r_i(t) = \frac{1}{2\beta} \ln\left(\frac{1-x_i(t)}{x_i(t)}\right)$$

with $x_i(t)$ being a uniformly distributed random variable $x_i(t) \in (0, 1)$ corresponds to the usual Monte Carlo dynamics for parallel update [26], which obey detailed balance. The noise parameter β can be regarded as an inverse temperature $T: \beta = 1/T$.

The field distribution $P_t(h)$ now also includes an average over the synaptic noises for times $\tau < t$. Since the additional noise at time t is uncorrelated to all other random processes, the changes in the DPD equation are minor. The overlap is now given by $m_t = \overline{\tanh(\beta h_{t-1})}$, and, in (18) and (16), one has to replace $\overline{|h_{t-1}|}$ by $\overline{h_{t-1} \tanh(\beta h_{t-1})}$. Clearly, for $\beta \to \infty$, the original DPD equations are recovered.

In figure 7 we display the behaviour of the network at $\alpha = 0.5$ and for temperatures T equal to 0.0, 0.1 and 0.2. With synaptic noise, the area of attraction becomes smaller for increasing temperature and the m_t versus m_0 plot develops towards a step function with increasing time. In fact, if the pattern is not restored, the overlap m_t always decays exponentially fast to zero. The time constant of the decay depends on α and T and can, however, be quite small for low temperatures. Nevertheless, the DPD indicates that the dynamic traps encountered in the noiseless dynamics (see section 4) play no role under noisy dynamics.

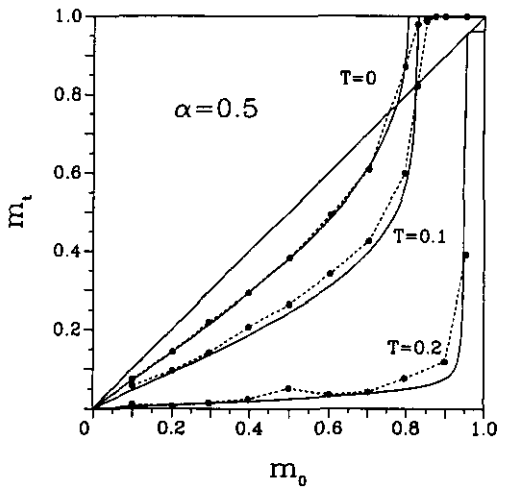

Figure 7. The overlap at time $t = 30$ for several temperatures T. Included are additional numerical data (broken lines), obtained from networks of 256 spins, averaged over 100 samples per data point.

With synaptic noise the retrieval states are no longer identical to the learned patterns. This is clearly seen in the $T = 0.2$ curve of figure 7, where the retrieval state has only an overlap of ≈ 0.96 with the pattern. This temperature, $T = 0.2$, is very close to critical temperature $T_c = 0.201$, where even the retrieval states are no longer dynamically stable.

This critical temperature T_c is a function of α as shown in figure 8. We find from our dynamic theory a strong first-order transition, as already observed within the static mean field approach of Kanter and Sompolinsky [10]. Our curve lies close to their mean field results. Since their statistical analysis is, however, based on a different

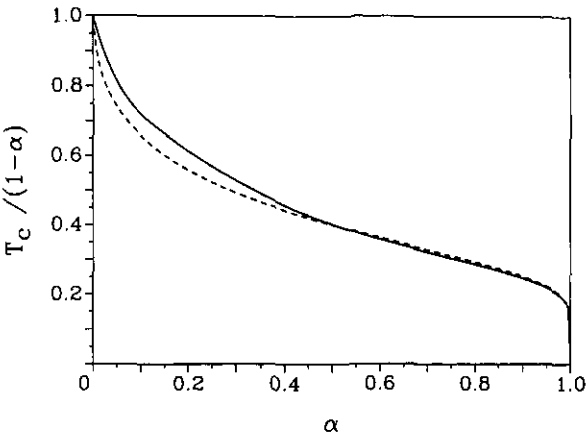

Figure 8. The critical temperature T_c, where the retrieval states become dynamically unstable, as a function of α (full line). For comparison, we have included the findings of Kanter and Sompolinsky (broken line, from [10]).

Hamiltonian, valid for sequential update, the results are not directly comparable to our dynamic treatment. Nevertheless, close agreement is observed between the two theories. A similar agreement between equilibrium results for parallel and serial update was found in [27] for the case of the Hopfield coupling matrix.

The function $1-c_{t,t-2}$ can also be calculated for noisy dynamics within the DPD approach. We have included data for $\alpha=0.5$ and $T=0.1$ in figure 5. Note that $T=0$ is still lower than the critical temperature $T_c=0.201$ at this α-value. The supplementary numerical data given for system sizes of 128 and 1024 spins match the DPD prediction. Clearly, at least at the timescales displayed here, the prominent finite-size effects found under deterministic dynamics are not observed with noisy dynamics.

From our numerical simulations, we find for the remanent magnetization $c_{t,0}$ an exponential decay. This is in contrast to the SK model of spin glasses, where we observe a power law decay of the remanent magnetization under parallel noisy dynamics [28].

Finally, in figure 4, we also display the areas of attraction for a number of different temperature values (broken curves). For each temperature, at a given α-value, the lower branch of a curve gives the critical overlap m_c and the upper branch the corresponding overlap of the retrieval states. For all m_0-values between m_c and 1, the system flows finally towards the retrieval states. For all values below m_c, the overlap decays exponentially fast to zero.

As expected, the areas of attraction diminish with higher temperatures. This is directly caused by the destabilization of the patterns due to the synaptic noise. At intermediate temperatures, it could be possible that the corresponding destabilization of the metastable states would lead to an enlarged area of attraction. The increase in the area of attraction we observe at low α-values is, however, marginal.

6. Conclusion

We have thoroughly investigated the dynamics of the neural network with the pseudoinverse coupling matrix. Our approach was three-fold: an exact dynamic treatment was used to gain insight into the structure of the internal fields of the network. This provided

a base for the development of our DPD approach to network dynamics. The DPD approach was used to obtain the main results of this paper. In addition, extensive numerical simulations were performed to validate our approach.

The exact description of the dynamics of the neural network turned out to be impractical after a few time-steps, due to a rapidly increasing number of parameters. The same situation had been already encountered by Derrida and coworkers in their treatment of the SK and Hopfield models and seems to be characteristic for fully connected networks.

Nevertheless, the exact theory yielded the general structure of the internal field: a sum of a Gaussin noise term and a memory term, where the values of the neurons at previous times appear. In developing the DPD approach it was our goal to preserve this specific structure of the internal fields and simultaneously keep the theory as simple as possible. In our approach, only three dynamic parameters, namely $\overline{h_i}$, Δh_i^2 and the renormalized self-coupling d_t, have to be calculated at each time-step. We expect that the simplicity of our ansatz supports the application to other models as well, such as the Hopfield neural network or the SK model of spin glasses. An extension to other update rules, like soft spin dynamics, should be possible.

All three parameters of our theory are given by equations derived from first principles, the DPD approach has no adjustable parameters. Within our approach we were able to describe faithfully the temporal development of the neural network for the whole α- and temperature range. Obviously the reduction of the memory term to its simplest form in the DPD approach is sufficient to describe most of the dynamic effects in a neural network correctly.

Acknowledgments

The authors would like to thank W Kinzel and S Diederich for stimulating discussions. The numerical simulations were carried out on PCs, the CYBER 860 of the HRZ Giessen and the CRAY Y-MP of the HLRZ Jülich. This work was supported by grants of the Deutsche Forschungsgemeinschaft. It is part of the PhD Thesis of RDH.

Appendix 1

In this appendix we evaluate the generating function

$$[Z(\underline{l})]_J = \left[\mathrm{Tr}_{S_i(t)} \int \prod_{it} \left(\mathrm{d}h_i(t) \Theta(S_i(t+1)h_i(t)) \delta\left(h_i(t) - \sum_{j \neq i} J_{ij} S_i(t) \right) \right) \times \exp\left(\mathrm{i} \sum_{it} l_i(t) h_i(t) \right) \right]_J$$

in the limit $N \to \infty$.

To facilitate the averaging over J_{ij}, we rewrite $h_i(t)$ in terms of expansion coefficients $a_\mu(t)$:

$$h_i(t) + \alpha S_i(t) = a_1(t) + N^{-1/2} \sum_{\mu > 1} \xi_i^\mu a_\mu .$$

To fix the internal fields completely, we have to specify the components in the space

orthogonal to the patterns. This is done via the constraints

$$N^{-1/2}\sum_i \xi_i^\mu (h_i(t)-(1-\alpha)S_i(t))=0.$$

We arrive at

$$Z(\underline{l})=\mathrm{Tr}_{S_i(t)}\int \mathrm{D}h\,\mathrm{D}\hat{h}\,\mathrm{D}a\,\mathrm{D}\hat{a}\,\det(C_{\mu\nu})$$
$$\times\delta\Big(N^{-1}\sum_i (h_i(t)-(1-\alpha)S_i(t))\Big)\prod_{it}(\Theta(S_i(t+1)h_i(t)))$$
$$\times\exp\Big(\mathrm{i}\sum_{it} l_i(t)h_i(t)+L[h(t),\hat{h}(t),a(t),\hat{a}(t)]\Big) \tag{A1.1}$$

with

$$L[h(t),\hat{h}(t),a(t),\hat{a}(t)]$$
$$=\mathrm{i}\sum_{it}\hat{h}_i(t)\Big(h_i(t)+\alpha S_i(t)-a_1(t)-N^{-1/2}\sum_{\mu>1}a_\mu(t)\xi_i^\mu\Big)$$
$$+\mathrm{i}\sum_{\mu>1,\,t}\hat{a}_\mu(t)\Big(N^{-1/2}\sum_i\xi_i^\mu[h_i(t)+(1-\alpha)S_i(t)]\Big)$$

and

$$\int \mathrm{D}h\,\mathrm{D}\hat{h}\,\mathrm{D}a\,\mathrm{D}\hat{a}=\int\prod_{it}(\mathrm{d}h_i(t)\,\mathrm{d}\hat{h}_i(t))\prod_{\mu t}\mathrm{d}a_\mu(t)\prod_{\substack{\mu t\\ \mu>1}}\mathrm{d}\hat{a}_\mu(t).$$

The term $\det(C_{\mu\nu})$ is the functional Jacobian of our transformation, which ensures that $Z(\underline{0})=1$, still.

Averaging, we obtain, after some algebra,

$$[Z(\underline{l})]_J=\int\prod_{t\tau}\Big(\frac{N\,\mathrm{d}U_{t\tau}\,\mathrm{d}\hat{U}_{t\tau}}{2\pi}\,\frac{N\,\mathrm{d}C_{t\tau}\,\mathrm{d}\hat{C}_{t\tau}}{2\pi}\,\frac{N\,\mathrm{d}K_{t\tau}\,\mathrm{d}\hat{K}_{t\tau}}{2\pi}\Big)\exp(F(U,\hat{U},C,\hat{C},K,\hat{K}))$$

with

$$F=iN\sum_{t\tau}(\hat{U}_{t\tau}U_{t\tau}+\hat{C}_{t\tau}C_{t\tau}+i\hat{K}_{\tau t}K_{t\tau})+\ln(\tilde{Z}(U,\hat{U},C,\hat{C},K,\hat{K}))$$

$$\tilde{Z}=[\det(C_{\mu\nu})]_J\,\mathrm{Tr}_{S_i(t)}\int \mathrm{D}h\,\mathrm{D}\hat{h}\,\mathrm{D}a\,\mathrm{D}\hat{a}\,\delta\Big(N^{-1}\sum_i h_i(t)-(1-\alpha)S_i(t)\Big)$$
$$\times\prod_{it}(\Theta(S_i(t+1)h_i(t)))\exp(\tilde{L}[h,\hat{h},a,\hat{a},m,S,U,\hat{U},C,\hat{C},K,\hat{K}])$$

and

$$\tilde{L}=\mathrm{i}\sum_{it}\hat{h}_i(t)(h_i(t)+\alpha S_i(t)-a_1(t))+l_i(t)h_i(t)$$
$$+\tfrac{1}{2}\sum_{t\tau}\Big(U_{t\tau}\sum_i\hat{h}_i(t)\hat{h}_i(\tau)+C_{t\tau}\sum_{\mu>1}\hat{a}_\mu(t)\hat{a}_\mu(\tau)+2K_{t\tau}\sum_i i\hat{h}_i(t)f_i(\tau)\Big)$$
$$-\mathrm{i}\sum_{t\tau}\Big(\hat{U}_{t\tau}\sum_{\mu>1}a_\mu(t)a_\mu(\tau)+\hat{C}_{t\tau}\sum_i f_i(t)f_i(\tau)+\hat{K}_{\tau t}\sum_{\mu>1}a_\mu(t)i\hat{a}_\mu(\tau)\Big).$$

In the limit $N \to \infty$ the integrals can be evaluated via steepest descent. The integration variables take on the stationary point values

$$\hat{U}_{t\tau} = -\frac{\mathrm{i}}{2N} \sum_i \langle \hat{h}_i(t) \hat{h}_i(\tau) \rangle_{\tilde{Z}} \tag{A1.2}$$

$$U_{t\tau} = N^{-1} \sum_{\mu>1} \langle a_\mu(t) a_\mu(\tau) \rangle_{\tilde{Z}} \tag{A1.3}$$

$$\hat{C}_{t\tau} = -\frac{\mathrm{i}}{2N} \sum_\mu \langle \hat{a}_\mu(t) \hat{a}_\mu(\tau) \rangle_{\tilde{Z}} \tag{A1.4}$$

$$C_{t\tau} = N^{-1} \sum_i \langle f_i(t) f_i(\tau) \rangle_{\tilde{Z}} \tag{A1.5}$$

$$\hat{K}_{t\tau} = -N^{-1} \sum_i \langle \mathrm{i}\hat{h}_i(\tau) f_i(t) \rangle_{\tilde{Z}} \tag{A1.6}$$

$$K_{t\tau} = N^{-1} \sum_\mu \langle a_\mu(t) \mathrm{i}\hat{a}_\mu(\tau) \rangle_{\tilde{Z}}. \tag{A1.7}$$

Similar to the treatment of the SK model in [21], we set $\hat{U}_{t\tau} = \hat{C}_{t\tau} = 0$. This is a self-consistent solution of equations (A1.2)-(A1.7). Furthermore, non-zero $\hat{U}_{t\tau}$ or $\hat{C}_{t\tau}$ would violate the normalization of $[Z(\underline{0})]_J$.

In order to decouple the remaining part of $\tilde{L}$ we introduce Gaussian random fields $w_i(t)$ and $v_\mu(t)$ with zero mean and variances given by

$$[w_i(t) w_i(\tau)] = U_{t\tau}$$

and

$$[v_\mu(t) v_\mu(\tau)] = C_{t\tau}.$$

and obtain finally the averaged functional in the form

$$[Z(\underline{l})]_J = \left[\mathrm{Tr}_{S_i(t)} \int \mathrm{D}h \, \mathrm{D}a \prod_{it} (\Theta(S_i(t+1) h_i(t))) \right.$$
$$\times \delta\left(N^{-1} \sum_i h_i(t) - (1-\alpha) S_i(t) \right) \exp\left(\mathrm{i} \sum_{it} l_i(t) h_i(t) \right)$$
$$\times \prod_{it} \left(\delta(h_i(t) + \alpha S_i(t) - w_i(t) - \sum_\tau K_{t\tau} f_i(\tau) - a_1(t)) \right)$$
$$\left. \times \prod_{\mu>1,t} \left(\delta\left(\sum_\tau \hat{K}_{t\tau} a_\mu(\tau) - v_\mu(t) \right) \right) \right]_{w,v}.$$

Appendix 2

In this appendix we give—for reference—results derived from our exact dynamic theory for the first few time-steps.

The internal field at time time $t = 0$ has the structure

$$h_0 = (1-\alpha)(m_0 + w_0)$$

thus $P_0(h)$ is a Gaussian distribution with mean

$$\overline{h_0} = (1-\alpha) m_0$$

and variance

$$\overline{\Delta h_0^2} = \alpha(1-\alpha)(1-m_0^2).$$

For time $t=1$, the internal field is given by

$$h_1 = (1-\alpha)(m_1 + w_1 + (1-\alpha)K_{10}((m_0 + w_0 - S_0))) \tag{A2.1}$$

with

$$K_{10} = -2\frac{\alpha}{1-\alpha}P_0(0)$$

and

$$\overline{\Delta h_1^2} = \alpha(1-\alpha)(1-m_1^2) - (1-2\alpha)\left(2(1-\alpha)K_{10}m_0m_1 + (1-2\alpha)\frac{(1-\alpha)^3}{\alpha}K_{10}^2(1-m_0^2)\right).$$

Equation (26) has the structure

$$h_1 = z + d_1 S_0 \tag{A2.2}$$

where z is a Gaussian, uncorrelated to S_0. This structure is identical to our DPD ansatz and the corresponding probability distribution $P_1(h)$ consists accordingly of two Gaussian peaks of width Δz^2, separated by a distance $2d_1$.

Note that already at the second time-step a memory term appears in the structure of the internal fields. This is a quite general result, valid for other networks as well [23]. Therefore, approximations to the dynamics neglecting memory terms can in general not be expected to yield correct results, even for the second time-step.

For time $t=2$ we have carried out the calculations only for $\alpha = \frac{1}{2}$ and $m_0 = 0$. The general structure of the field is

$$h_2 = (1-\alpha)(m_2 + w_2) + (1-\alpha)^2 K_{21}(m_1 + w_1 - S_1) + (1-\alpha)^2((1-\alpha)K_{21}K_{10} + K_{20})(m_0 + w_0 - S_0). \tag{A2.3}$$

Since [23]

$$(1-\alpha)K_{21}K_{10} + K_{20} = 2K_{10}P_1(0)(1-2\alpha) = 0$$

for $\alpha = \frac{1}{2}$, (A2.3) simplifies to

$$h_2 = \tilde{w} + \tilde{d}S_1. \tag{A2.4}$$

where $\tilde{w}$ is the Gaussian composed of w_1 and w_0.

The structure of (A2.4) looks identical to our ansatz (14), but the correlation between $S_1 = \mathrm{sgn}((1-\alpha)w_0)$ and w_1 (note that w_0 and w_1 are correlated) leads to a distribution for $P_2(h)$ which is no longer double Gaussian. We find

$$P_2(h) = \frac{1}{2}\frac{1}{\sqrt{2\pi\overline{\Delta\tilde{w}^2}}}\left[(1+\mathrm{erf}(\Delta_+(h)))\exp\left(-\frac{1}{2}\frac{(h-\tilde{d})^2}{\overline{\Delta\tilde{w}^2}}\right) + (1+\mathrm{erf}(\Delta_-(h)))\exp\left(-\frac{1}{2}\frac{(h+\tilde{d})^2}{\overline{\Delta\tilde{w}^2}}\right)\right] \tag{A2.5}$$

with $\Delta_\pm(h)$ given by

$$\Delta_\pm(h) = \frac{1}{\sqrt{2(1-\rho^2)}}\frac{\rho(h\mp\tilde{d})}{\sqrt{\overline{\Delta\tilde{w}^2}}}.$$

The parameters of this distribution are

$$\tilde{d} = 0.275\,76$$

$$\rho = 0.618\,54$$

$$\overline{\Delta \tilde{w}^2} = 0.091\,67.$$

Appendix 3

Using the projector property of the coupling matrix one obtains

$$\sum_{ijk} J_{ij}J_{ik}S_j(t)S_k(\tau) = \alpha(1-\alpha)\sum_k S_k(t)S_k(\tau) + (1-2\alpha)\sum_k h_k(t)S_k(\tau).$$

This gives

$$\begin{aligned}\overline{h_i^2(t+1)} &= \alpha(1-\alpha) + (1-2\alpha)\overline{h_i(t+1)\,\mathrm{sgn}(h_i(t))} \\ &\approx \alpha(1-\alpha) + (1-2\alpha)\overline{h_i(t+1)}\;\overline{\mathrm{sgn}(h_i(t))} \\ &\quad + (1-2\alpha)\underbrace{\frac{\overline{\Delta h_i(t+1)\Delta h_i(t)}}{\overline{\Delta h_i^2(t)}}\left(\overline{|h_i(t)|} - \overline{\mathrm{sgn}(h_i(t))\;h_i(t)}\right)}_{= a_i(t+1)}\end{aligned}$$

where we have used the approximation

$$\overline{\Delta a(x)\Delta b(y)} \approx \frac{\overline{\Delta x \Delta y}}{\overline{\Delta x^2}\;\overline{\Delta y^2}}\;\overline{a(x)\Delta x}\;\overline{b(y)\Delta y}.$$

This is an expansion in terms of small correlations $\overline{\Delta x \Delta y}$ for correlated random variables x and y.

With

$$\begin{aligned}&\overline{\Delta h_i(t+1)\Delta h_i(t)} \\ &\quad = \overline{h_i(t+1)h_i(t)} - \overline{h_i(t+1)}\;\overline{h_i(t)} \\ &\quad = (1-2\alpha)\overline{|h_i(t)|} - (1-\alpha)^2 m(t+1)m(t) \\ &\qquad + \alpha(1-\alpha)\,\overline{\mathrm{sgn}(h_i(t))\,\mathrm{sgn}(h_i(t-1))}\end{aligned}$$

and

$$\overline{\mathrm{sgn}(h_i(t))\,\mathrm{sgn}(h_i(t-1))} \approx m(t+1)m(t) + \frac{a_i(t)}{\overline{\Delta h_i^2(t)}}\left(\overline{|h_i(t)|} - (1-\alpha)m(t+1)m(t)\right)$$

we finally obtain, after some algebra, (18). Note that this equation is exact for $\alpha = 0.5$, regardless of the above used approximations.

References

[1] Hopfield J J 1982 *Proc. Natl Acad. Sci., USA* **79** 2554
[2] Derrida B, Gardner E and Zippelius A 1987 *Europhys. Lett.* **4** 167
[3] Domany E, Kinzel W and Meir R 1989 *J. Phys. A: Math. Gen.* **22** 2081
[4] Little W A 1974 *Math. Biosci.* **19** 101
Little W A and Shaw G L 1978 *Math. Biosci.* **39** 281
[5] Kohonen T 1988 *Self-Organization and Associative Memory* (Berlin: Springer)
[6] Personnaz L, Guyon I and Dreyfus G 1986 *Phys. Rev.* A **34** 4217
[7] Diederich S and Opper M 1987 *Phys. Rev. Lett.* **58** 949
[8] Opper M 1989 *Europhys. Lett.* **8** 389
[9] Amit D J, Gutfreund H and Sompolinsky H 1985 *Phys. Rev. Lett.* **55** 428
[10] Kanter I and Sompolinsky H 1987 *Phys. Rev.* A **35** 380
[11] Amari S and Maginu K 1988 *Neural Networks* **1** 63
[12] Kohring G A 1989 *Europhys. Lett.* **8** 697
[13] Patrick A E and Zagrebnov A *Preprint*
[14] Gardner E, Derrida B and Mottishaw P 1987 *J. Phys. (Paris)* **48** 741
[15] Krauth W, Nadal J P and Mézard M 1988 *J. Phys. A: Math. Gen.* **21** 2995
[16] Horner H, Bohrmann D, Frick M, Kinzenbach H and Schmidt A 1989 *Z. Phys.* B **76** 381
[17] Amit D J, Evans M R, Horner H and Wong K Y M *Preprint*
[18] Opper M, Kleinz J, Kõhler H and Kinzel W 1989 *J. Phys. A: Math. Gen.* **22** L407
[19] Henkel R D and Opper M 1990 *Europhys. Lett.* **11** 403
[20] Kree R and Zippelius A *Preprint*
[21] Sompolinsky H and Zippelius A 1982 *Phys. Rev.* B **25** 6860
[22] Mézard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)
[23] Henkel R D 1990 *PhD Thesis*, JLU Giessen
[24] Geszti T and Pázmándi F 1989 *Phys. Scri.* **T25** 152
[25] Frumkin A and Moses E 1986 *Phys. Rev.* A **34** 714
[26] Amit D J 1989 *Modelling Brain Function* (Cambridge: Cambridge University Press)
[27] Fontanari J F and Koeberle R 1987 *Phys. Rev.* A **36** 2475
[28] Henkel R D and Opper M In preparation
[29] Krätzschmar J and Kohring G A 1990 *J. Phys. (Paris)* **51** 223