Physica D 68 (1993) 401–415
North-Holland

SDI: 0167-2789(93)E0192-E

PHYSICA D

Multifractality in forgetful memories

U. Behn[a], J.L. van Hemmen[b], R. Kühn[c], A. Lange[a] and V.A. Zagrebnov[d,1]

[a] *Institut für Theoretische Physik, Universität Leipzig, D-04109 Leipzig, Germany*
[b] *Physik-Department, TU München, D-85748 Garching, Germany*
[c] *Institut für Theoretische Physik, Universität Heidelberg, D-69120 Heidelberg, Germany*
[d] *Instituut voor Theoretische Fysica, Katholieke Universiteit Leuven, B-3001 Leuven, Belgium*

Received 31 December 1992
Accepted 2 April 1993
Communicated by F.H. Busse

Learning rules of a forgetful memory generate their synaptic efficacies through iterative procedures that operate on the input data, random patterns. We analyse invariant distributions of the synaptic couplings as they arise asymptotically and show that they exhibit fractal or multifractal properties. We also discuss their dependence upon the learning rule and the parameters specifying it, and indicate how the nature of the invariant distribution is related to the network performance.

1. Introduction

Learning in neural networks involves an adaptation of synaptic efficacies in response to external stimuli-patterns or associations to be established, e.g., as attractors of a recurrent network dynamics or as desired input/output relations in networks of feed-forward type.

Learning is usually called supervised, if adaptation aims at reducing observed differences between an actual and a desired network response. Typical examples are provided by various forms of perceptron learning, or learning by error back-propagation [1–5]. Alternatively, if the adaptation process is *not* mediated by an observed difference between actual and desired systems response, learning is called unsupervised; synapses are changed strictly in response to the newly presented stimulus. The various forms of Hebbian covariance learning rules [6–9] belong to this category.

In both cases, learning generates a dynamic flow in the space of couplings. For supervised learning, this flow should ideally come to a halt at what hopefully constitutes a good solution to the problem at hand. For unsupervised learning it does not as long as new stimuli are presented to the net.

In the present paper, we are connected with situations where the asymptotics of the dynamic flow generated by learning is amenable to a statistical analysis. Such is the case, if the flow is deterministic but evolves on a chaotic attractor, or if it is stochastically driven by an irreducible Markov process. To be specific, we shall here consider only the second possibility, typically associated with learning random patterns, and we analyse asymptotic invariant distributions of the couplings generated by such a learning process (provided they exist).

The type of learning rule we have in mind is of a general, iterative, local Hebbian form

$$J_{ij}^{\text{new}} = \Phi(\xi_i^{\text{new}}\xi_j^{\text{new}}; J_{ij}^{\text{old}}) , \tag{1}$$

[1] On leave of absence from Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, Dubna 141980, CIS-Russia.

where ξ_i^{new} and ξ_j^{new} are independent identically distributed random variables (i.i.d.r.v.) taking values in $\{1, -1\}$. If the mnemodynamic updating function Φ is such that it gives rise to synapses of bounded strength, the model so defined will typically exhibit the phenomenon of forgetfulness [8]. Learning rules proposed in connection with forgetful memories [8,10–14] will, indeed, be taken as paradigms of our considerations. Our main result is that invariant distributions of the couplings in these models quite generally exhibit fractal or multifractal properties. Similar phenomena were recently found to occur also in error back-propagation learning [15].

Formally, the learning rule (1) defines a discrete stochastic mapping on the real line. This kind of mapping has been studied previously in connection with the one-dimensional random field Ising model (RFIM) [16–21,24]. Asymptotically it generates an invariant distribution of effective fields. Changes in the singular nature of the invariant distribution have been suggested to be at the heart of drastic changes in the observable behaviour of the system [16,19,20]. Accordingly, our aim here is to characterise the singular nature of invariant distributions generated by (1), and to see whether it can be related to network performance.

The outline of the paper is as follows. In section 2, we introduce the discrete stochastic mappings associated with forgetful memories in more detail. In section 3, a description of their invariant distributions in terms of a symbolic dynamics [18,24] generated by iterative solutions of the Chapman–Kolmogorov equation is presented. Some qualitative aspects of the invariant measures, in particular, the nature of their support and their behaviour at end points of the support are analysed in section 4. An analysis of their singular properties in terms of an infinite set of generalized fractal (or Renyi) dimensions [22] is outlined in section 5. We present results for two typical cases, the linear palimpsestic scheme introduced by Nadal et al. [11] and by Mézard et al. [12], and the nonlinear encoding schemes studied by Hopfield [8], Parisi [10], Derrida and Nadal [13], and by van Hemmen et al. [14]. Finally, in section 6, the relation between the singular nature of an invariant J_{ij} distribution and the network performance is discussed.

2. Forgetful memories and discrete stochastic mappings

Learning rules proposed to describe forgetful memories [8,10–14] are quite generally of the form

$$J_{ij}^{(\mu)} = N^{-1}\phi(\epsilon_N \xi_i^\mu \xi_j^\mu + N J_{ij}^{(\mu-1)}) , \tag{2}$$

where $J_{ij}^{(\mu)}$ denotes the value of the coupling $j \to i$, after pattern μ, i.e., $\{\xi_i^\mu;\ i = 1, \ldots, N\}$, has been embedded in the net. This amounts to taking

$$\Phi(u, v) = N^{-1}\phi(\epsilon_N u + N v) \tag{3}$$

for some real function ϕ of a *single* variable in (1). Parisi's choice [10] was

$$\phi(x) = \operatorname{sgn}(x) \min(1, |x|) . \tag{4}$$

It constitutes a formal variant of a proposal originally due to Hopfield [8], namely, to let the generalized Hebb rule $\Delta J_{ij} \propto \xi_i^{\text{new}} \xi_j^{\text{new}}$ become operative *only* when the resulting J_{ij} does not exceed certain upper or lower bounds. We will refer to this case as the Hopfield–Parisi model. For a proper scaling of ϵ_N with system size N, viz., $\epsilon_N = \mathcal{O}(N^{-1/2})$, Parisi was able to demonstrate numerically that approximately

0.04 N patterns can constantly be held in store, a result which was subsequently confirmed analytically by van Hemmen et al. [4]. If new patterns are added, they are stored at the expense of old ones which are forgotten. That is, the system admits a stationary mode of operation, and does not run into the catastrophic forgetting of the original Hopfield model [8,9]. The Hopfield–Parisi model has also been studied in the highly dilute limit by Derrida and Nadal [13].

Different choices for ϕ in (3) give rise to different models, some of which have been studied in the past. For example,

$$\phi(x) = \lambda_N x \,, \tag{5}$$

with $\epsilon_N = 1$ and $\lambda_N = 1 - \mathcal{O}(N^{-1})$ corresponds to the so-called marginalist scheme studied by Nadal et al. [11] and by Mézard et al. [12]. Even though $\phi(x)$ does not saturate in this case, it does generate synaptic couplings which cannot exceed $N^{-1}\lambda_N/(1-\lambda_N)$ in absolute value, due to the iterative prescription (3). The storage capacity for this model can be as high as 0.05 N in the large-N limit [12].

The choice

$$\phi(x) = \tanh(x) \,, \tag{6}$$

with $\epsilon_N = \mathcal{O}(N^{-1})$, provides an example of a general nonlinear iterative prescription which also gives rise to a working memory that allows for a stationary mode of operation. For this model, the number of patterns held in active memory is again approximately 0.04 N [14].

Finally, by taking

$$\phi(x) = x \tag{7}$$

and $\epsilon_N = 1$, one recovers the original Hopfield model [8]. Unlike the other encoding schemes, this one can give rise to synapses of unbounded strength. Furthermore, it does not break the permutation invariance encoded patterns, and it acquires no statistical stationarity as more and more patterns are embedded in the net. Consequently it leads to a catastrophic forgetting for $\alpha = p/N$ beyond $\alpha_c \simeq 0.14$ [8,9], where p is the number of stored patterns.

In order to study the statistical properties of the encoding processes introduced above, we concentrate on a single bond $j \to i$. Omitting bond indices, and defining $x_\mu = NJ_{ij}^{(\mu)}$, and i.i.d.r.v. $\sigma_\mu = \xi_j^\mu \xi_j^\mu = \pm 1$, $\mu = 1, 2, \ldots$, the encoding processes are seen to take the form

$$x_{\mu+1} = \phi(x_\mu + \epsilon\sigma_\mu) := \phi_{\sigma_\mu}(x_\mu) \,, \tag{8}$$

where, to simplify the notation, we have omitted the scaling of $\epsilon = \epsilon_N$ with system size N. Equation (8) describes a stationary Markov process, driven by the (random) patterns to be encoded in the couplings. The properties of the driven process $\{x_\mu\}$ – state space, invariant measure (if it exists), and so on – depend on the choice of the synaptic function ϕ in (8), with a wide range of possibilities, as shown in table 1. For instance, the number of states may be finite, as in the Hopfield–Parisi model, denumerable as in the Hopfield model, or uncountable as in the linear marginalist and the nonlinear encoding schemes, (5) and (6), respectively. In these two cases, the invariant measure of the encoding process turns out to exhibit fractal or multifractal properties. It is these two cases which we now consider in more detail.

Table 1
Number of states and characteristics of the invariant measure generated by learning rule (8).

Model	$\phi(x), \epsilon_N$	Number of states	Invariant measure
Hopfield [8]	$x, \epsilon_N = 1$	countable	random walk on $\mathbb{Z}$
Hopfield–Parisi [8,10,13,14]	$\phi(x) = \mathrm{sgn}(x)\min(1, \lvert x\rvert)$, $\epsilon_N = \mathcal{O}(N^{-1/2})$	finite	finite state Markov chain
Binary synapses	$\phi(x) = \mathrm{sgn}(x)\min(1, \lvert x\rvert)$, $\epsilon_N > 2$	2	concentrated on ± 1
Marginalist [11–13]	$\phi(x) = \lambda x$, $\epsilon_N = 1$ $\lambda = 1 - \mathcal{O}(N^{-1})$	uncountable	one-scale fractal which is thin for $\lambda < \frac{1}{2}$, fat for $\lambda \geq \frac{1}{2}$
Smooth [14]	$\phi(x) = \tanh(x)$, $\epsilon_N = \mathcal{O}(N^{-1})$	uncountable	multifractal which is thin for $\epsilon > 0.957$, fat for $\epsilon \leq 0.957$

3. Symbolic dynamics

In the case of the marginalist encoding scheme (5) and the smooth learning rule (6), the map (8) generates an uncountable number of states, and a corresponding hierarchy of bands and gaps. As for the RFIM [18,24], these can be encoded in a unique way by a symbolic dynamics. We denote the result of n iterations, starting from an initial value x by

$$y = \phi_{(\sigma)_n}(x) \equiv \phi_{\sigma_n}(\phi_{\sigma_{n-1}}(\ldots \phi_{\sigma_1}(x)\ldots)), \tag{9}$$

where $(\sigma)_n = (\sigma_n, \sigma_{n-1}, \ldots, \sigma_1)$ is the sequence of n signs characterizing the history of the stochastic learning process. Similarly, the n-iteration pre-image x of y, will be denoted by

$$x = \phi^{-1}_{(\sigma)_n}(y) \equiv \phi^{-1}_{\sigma_1}(\phi^{-1}_{\sigma_2}(\ldots \phi^{-1}_{\sigma_n}(y)\ldots)), \tag{10}$$

where as compared to (9) the σ_i appear in reverse order, so that $\phi^{-1}_{(\sigma)_n}(\phi_{(\sigma)_n}(x)) = x$.

Let (σ) denote an *infinite* string of signs. The attractor of the encoding process is given by the set $\{\phi_{(\sigma)}\}$, which for $\phi' < 1$ almost everywhere on $\mathbb{R}$ is uncountable and independent of the starting point x one might wish to choose,[#1] i.e., the dynamics acts *transitively* on $\{\phi_{(\sigma)}\}$. The attracting set $\{\phi_{(\sigma)}\}$ is contained in the interval $I = [\phi_{(-)}, \phi_{(+)}]$, where $(-)$ and $(+)$ denote infinite strings of minuses and pluses, respectively. The endpoints $\phi_{(+)}$ and $\phi_{(-)}$ of I are fixed points of ϕ_+ and ϕ_-, i.e.,

$$\phi_{(+)} = x^* = \phi_+(x^*) = \phi(x^* + \epsilon), \tag{11}$$

and similarly for $\phi_{(-)} = -\phi_{(+)}$. It is on the set $\{\phi_{(\sigma)}\} \subseteq I$ where the unique invariant probability measure μ of the stochastic encoding process lives.

One can construct μ from a sequence $\{\mu^{(n)}\}_{n\in\mathbb{N}}$ of iterates of the Chapman–Kolmogorov (or Frobenius–Perron) equation corresponding to the stochastic mapping (1). Since $\{\phi_{(\sigma)}\}$ attracts the whole real line, there is for each initial probability measure $\mu^{(0)}$ and for arbitrary $\epsilon > 0$ a compact $K_\epsilon \supset \{\phi_{(\sigma)}\}$ such that $\int_{K_\epsilon} \mathrm{d}\mu^{(n)}(x) \geq 1 - \epsilon$, *uniformly* for $n \geq N_\epsilon$. By Prohorov's theorem [23], the sequence $\{\mu^{(n)}\}_{n\in\mathbb{N}}$ is, therefore, *compact* with respect to weak convergence, i.e., there are subsequences $\{\mu^{(n_{k_\alpha})}\}_{k_\alpha}$ having weak limits μ_α. As limits of iterations, the $\{\mu_\alpha\}_\alpha$ form a collection of invariant measures sharing the *same* support, $\mathrm{supp}\,\mu_\alpha = \{\phi_{(\sigma)}\}$, on which the dynamics acts *transitively*. Hence, the μ_α are ergodic and, therefore, must all be the same. That is, $\{\mu_\alpha\} = \{\mu\}$, and the weak limit μ of $\{\mu^{(n)}\}_{n\in\mathcal{N}}$ is *unique*.

[#1] This is why a starting value was left unspecified in our notation.

The support of the invariant measure may or may not have gaps in I, depending on whether the length $l_0 := \phi_+(-x^*) - \phi_-(x^*)$ of the interval

$$\Delta_0 = (\phi_-(x^*), \phi_+(-x^*)) \tag{12}$$

is positive or not; see fig. 1. For $l_0 > 0$, the interval Δ_0 is not visited by the encoding process, once it has been attracted by I, nor are all intervals generated from Δ_0 by successive iterations,

$$\Delta_{(\sigma)_n} = (\phi_{((\sigma)_n,-)}(x^*), \phi_{((\sigma)_n,+)}(-x^*)), \quad n \geq 1, \tag{13}$$

making for 2^n gaps at level n. Thus, for $l_0 > 0$ the support of the invariant measure is a fractal with Hausdorff dimension strictly less than one, which may be self-similar, as it is for $\phi(x) = \lambda x$, or not as in the case $\phi(x) = \tanh(x)$. For $l_0 \leq 0$ the support of the invariant measure is the full interval I. In this case, it is convenient to define, similarly to (12), the overlap $\Omega_0 = (\phi_+(-x^*), \phi_-(x^*))$, which generates a hierarchy of overlaps $\Omega_{(\sigma)_n} = (\phi_{((\sigma)_n,+)}(-x^*), \phi_{((\sigma)_n,-)}(x^*))$ in analogy to (13). This hierarchy of overlaps becomes of essential importance when computing the Renyi dimensions characterizing the invariant measure living on I (see below).

In order to obtain information about the nature of the invariant measure living on I, one iterates the Chapman–Kolmogorov equation [23] for the evolution of probabilities,

$$\begin{aligned} p^{(n+1)}(x) &= \int_{\mathbb{R}} \mathrm{d}y\, p^{(n)}(y) \cdot \tfrac{1}{2} \sum_{\sigma=\pm 1} \delta(x - \phi(y + \sigma\epsilon)) \\ &= \tfrac{1}{2} \sum_{\sigma=\pm 1} \frac{1}{\phi'(\phi^{-1}(x))}\, p^{(n)}(\phi_\sigma^{-1}(x)) = \tfrac{1}{2} \sum_{\sigma=\pm 1} \frac{1}{\phi_\sigma'(\phi_\sigma^{-1}(x))}\, p^{(n)}(\phi_\sigma^{-1}(x)), \end{aligned} \tag{14}$$

where we have exploited the invertibility of the ϕ's under consideration, and the circumstance that

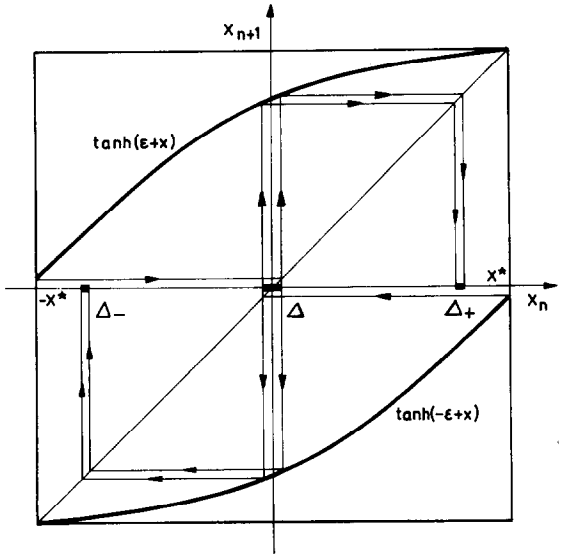

Fig. 1. Discrete stochastic mapping for the nonlinear encoding scheme (6) in the case of nonoverlapping bands. Only gaps of first and second generations are indicated.

$\phi'_\sigma(\phi_\sigma^{-1}(x))$ does not depend on σ (cf. eq. (8)) and is, in fact, equal to $\phi'(\phi^{-1}(x))$, in order to obtain the "symmetrized" hence easier – to-iterate final expression. Starting from an arbitrary initial probability density $p^{(0)}$ with support on some compact $I^{(0)}$, this iteration yields 2^n bands in the nth generation which can be encoded by a symbolic dynamics, too [18,24].

To wit, the first iterate of $p^{(0)}$ consists of two bands p_σ, $\sigma = \pm 1$, living on the images I_σ of the initial support $I^{(0)}$ under ϕ_σ, $\sigma = \pm 1$,

$$p_\sigma(x) = \begin{cases} \frac{1}{2} \dfrac{1}{\phi'_\sigma(\phi_\sigma^{-1}(x))} p^{(0)}(\phi_\sigma^{-1}(x)) & \text{for } x \in I_\sigma, \sigma = \pm 1 , \\ 0 & \text{otherwise} , \end{cases} \tag{15}$$

and

$$p^{(1)}(x) = \sum_{\sigma=\pm 1} p_\sigma(x) . \tag{16}$$

Similarly, each band $p_{(\sigma)_{n-1}}$ living on $I_{(\sigma)_{n-1}}$, the image of $I^{(0)}$ under $\phi_{(\sigma)_{n-1}}$, generates in the next generation the two bands

$$p_{(\sigma)_n}(x) = \frac{1}{2} \frac{1}{\phi'_{\sigma_n}(\phi_{\sigma_n}^{-1}(x))} p_{\sigma_{n-1}}(\phi_{\sigma_n}^{-1}(x)) , \quad \sigma_n = \pm 1 , \tag{17}$$

for $x \in I_{(\sigma)_n}$, and 0 otherwise. By iteration, we obtain [18,24]

$$p_{(\sigma)_n}(x) = p^{(0)}(\phi_{(\sigma)_n}^{-1}(x)) \prod_{\nu=1}^{n} [2\phi'_{\sigma_\nu}(\phi_{(\sigma_\nu,\sigma_{\nu+1},\ldots,\sigma_n)}^{-1}(x))]^{-1} , \quad x \in I_{(\sigma)_n} , \tag{18}$$

and

$$p^{(n)}(x) = \sum_{(\sigma)_n} p_{(\sigma)_n}(x) . \tag{19}$$

Equations (18) and (19) are an analog of a path integral representation, giving $p^{(n)}(x)$ as a sum of contributions from all 2^n encoding trajectories of length n that terminate in x.

Alternatively, instead of considering the evolution of probability densities, one may investigate a Chapman–Kolmogorov equation for the evolution of probability distributions

$$F^{(n)}(x) = \int_{-\infty}^{x} \mathrm{d}\mu^{(n)}(y) ,$$

that is,

$$F^{(n+1)}(x) = \int_{\mathbb{R}} \mathrm{d}\mu^{(n)}(y) \tfrac{1}{2} \sum_{\sigma=\pm 1} \theta(x - \phi(y + \sigma\epsilon)) = \tfrac{1}{2} \sum_{\sigma=\pm 1} F^{(n)}(\phi_\sigma^{-1}(x)) . \tag{20}$$

Starting from an initial distribution $F^{(0)}(x) = \int_{-\infty}^{x} \mathrm{d}y\, p^{(0)}(y)$, one obtains a representation of $F^{(n)}$ as a sum of contributions coming from all 2^n n-step pre-images of x,

$$F^{(n)}(x) = \sum_{(\sigma)_n} F_{(\sigma)_n}(x) , \tag{21}$$

where

$$F_{(\sigma)_n}(x) = 2^{-n} F^{(0)}(\phi^{-1}_{(\sigma)_n}(x)) , \quad x \in I_{(\sigma)_n} , \tag{22}$$

in complete analogy to (18) and (19).

Since the encoding scheme is ergodic on $\{\phi_{(\sigma)}\}$, eqs. (19) and (21) have limits as $n \to \infty$. For the distribution it is a pointwise limit F (in the points of continuity of F), satisfying the invariance condition $F(x) = \frac{1}{2}\Sigma_\sigma F(\phi_\sigma^{-1}(x))$, while the corresponding limit of (19) exists only in the weak sense [23]: $\lim_{n\to\infty} \int_{\mathbb{R}} dx\, p^{(n)}(x)\, f(x) = \int_{\mathbb{R}} dx\, p(x)\, f(x)$ for $f \in C(\mathbb{R})$. The limiting distribution F defines the invariant measure μ through $d\mu(x) = F(x + dx) - F(x)$.

4. Qualitative analysis of the invariant measure

The most prominent feature of the invariant measure, i.e., its support $\{\phi_{(\sigma)}\} \subseteq I = [-x^*, x^*]$ has been considered in the previous section. As already pointed out, the size $l_0 := \phi_+(-x^*) - \phi_-(x^*)$ of the first gap decides whether the support is a fractal with Hausdorff dimension strictly smaller than one.

For the nonlinear encoding scheme (6) and (8), l_0 depends on ϵ, and turns negative when ϵ becomes smaller than the positive solution $\epsilon = \epsilon_c^{(1)} \simeq 0.957$ of $\epsilon = \tanh(2\epsilon)$. For the linear encoding scheme (5) and (8), the gaps in the support vanish when λ increases beyond $\lambda_c^{(1)} = \frac{1}{2}$.

The behaviour of the invariant measure at the fixed points $\pm x^*$ of $\phi_\pm$ is also amenable to analytic investigation. Let us begin by considering the behaviour of $p^{(n)}(x)$ given by (19) as $x \to x^*$. Without loss of generality, we generate the $p^{(n)}$ starting from a $p^{(0)}$ living on I. The first thing to note is that sufficiently close to x^*, the *only* contribution to $p^{(n)}(x)$ comes from the $(+)_n$-band

$$p_{(+)_n}(x) = p^{(0)}(\phi^{-1}_{(+)_n}(x)) \prod_{\nu=1}^{n} [2\phi'_+(\phi^{-1}_{(+)_\nu}(x))]^{-1} . \tag{23}$$

Since the pre-image of the fixed point x^* is x^* itself, we obtain

$$p_{(+)_n}(x^*) = p^{(0)}(x^*)[2\phi'_+(x^*)]^{-n} . \tag{24}$$

Thus, as $n \to \infty$,

$$p^{(n)}(x^*) \to \begin{cases} 0 , & \text{if } \phi'_+(x^*) > \frac{1}{2} , \\ \infty , & \text{if } \phi'_+(x^*) < \frac{1}{2} , \end{cases} \tag{25}$$

and the invariant measure will be zero or diverge at x^* for $\phi'_+(x^*) > \frac{1}{2}$ and $\phi'_+(x^*) < \frac{1}{2}$, respectively. By symmetry, the same occurs at $-x^*$.

We note in passing that, starting from an arbitrary initial distribution, one arrives after a sufficiently large number m of pre-iterations at a distribution which is concentrated on I, except for contributions exponentially small in m. Hence, the result (25) is true for arbitrary initial distributions.

For the nonlinear encoding scheme (6), we have $\phi'_+(x^*) = 1 - x^{*2}$, so that the "critical" $x_c^* = x^*(\epsilon_c)$ is given by $x_c^* = 1/\sqrt{2}$, corresponding to $\epsilon_c^{(2)} \simeq 0.174$. The linear palimpsestic scheme has $\phi'_+(x) \equiv \lambda$ so that we get a critical $\lambda_c^{(2)} = \frac{1}{2}$. For this scheme, $\lambda = \frac{1}{2}$ is also the value for which the gaps of the fractal support of the invariant measure close.

Next, we investigate the *scaling behaviour* of the invariant measure at the boundaries of the support, i.e., the behaviour of

$$P(l) = \lim_{n\to\infty} P^{(n)}(l) = \lim_{n\to\infty} \int_{x^*-l}^{x^*} \mathrm{d}x\, p^{(n)}(x)\,, \tag{26}$$

as $l \to 0$, which is expected to behave as $P(l) \sim l^{\alpha}$, with a scaling exponent α that is to be determined. To this end, we first consider $P^{(n)}(l)$ at fixed (large) n, in the small l limit. If l is chosen so small that $x^* - l$ is larger than the right boundary of the band encoded by $((+)_{n-1}, -)$, only the $(+)_n$ band will contribute to $P^{(n)}(l)$,

$$P^{(n)}(l) = \int_{x^*-l}^{x^*} \mathrm{d}x\, p_{(+)_n}(x)\,. \tag{27}$$

Using (18) for $(\sigma)_n = (+)_n$, we get

$$P^{(n)}(l) = \int_{x^*-1}^{x^*} \mathrm{d}x\, p^{(0)}(\phi^{-1}_{(+)_n}(x)) \prod_{\nu=1}^{n} [2\phi'_+(\phi^{-1}_{(+)_\nu}(x))]^{-1}\,. \tag{28}$$

Expanding the $\phi^{-1}_{(+)_\nu}(x)$ in (28) about x^*, we obtain

$$\phi^{-1}_{(+)_\nu}(x^* - \epsilon) = x^* - \frac{\epsilon}{\lambda^\nu} + \mathcal{O}\left(\left(\frac{\epsilon}{\lambda^\nu}\right)^2\right), \tag{29}$$

where $\lambda := \phi'_+(x^*) < 1$. Under the condition $\epsilon/\lambda^\nu \ll 1$, we can expand further

$$\phi'_+(\phi^{-1}_{(+)_\nu}(x^* - \epsilon)) = \lambda - \phi''_+(x^*)\frac{\epsilon}{\lambda^\nu} + \mathcal{O}\left(\left(\frac{\epsilon}{\lambda^\nu}\right)^2\right). \tag{30}$$

In order to apply this to (28), we require

$$l/\lambda^n \leq \eta \ll 1\,. \tag{31}$$

To lowest nontrivial order in ϵ then, (28) reads

$$P^{(n)}(l) = \int_0^l \mathrm{d}\epsilon\, p^{(0)}(\phi^{-1}_{(+)_n}(x^* - \epsilon))(2\lambda)^{-n}\left(1 + \frac{\epsilon}{\lambda^n}\frac{\phi''_+(x^*)}{\lambda}\frac{1-\lambda^n}{1-\lambda}\right). \tag{32}$$

Finally, due to (29) and (31), we can exploit the mean value theorem to obtain

$$P^{(n)}(l) = p^{(0)}(\xi)(2\lambda)^{-n}\left(l + \frac{1}{2}\frac{l^2}{\lambda^n}\frac{\phi''_+(x^*)}{\lambda}\frac{1-\lambda^n}{1-\lambda}\right), \tag{33}$$

where $\xi = \phi^{-1}_{(+)_n}(x^* - \epsilon_0)$ for some $0 \leq \epsilon_0 \leq l$. To study the scaling behaviour of (26), we investigate (33) in the *scaling-limit*, $l \to 0$ as $n \to \infty$ according to

$$l = l(n) = \eta\lambda^n Q(n)\,, \tag{34}$$

where $0 < Q(n) \leq 1$ is some non-increasing algebraic function of n. For large n, we then get

$$P^{(n)}(l(n))/l(n)^\alpha \sim \frac{[\eta Q(n)]^{1-\alpha}}{(2\lambda^\alpha)^n} + \frac{1}{2}\frac{[\eta Q(n)]^{2-\alpha}}{(2\lambda^\alpha)^n}\frac{\phi''_+(x^*)}{\lambda}\frac{1-\lambda^n}{1-\lambda}\,. \tag{35}$$

As $n \to \infty$ this goes to zero or infinity, unless

$$\alpha = -\frac{\log 2}{\log \phi'_+(x^*)}, \tag{36}$$

no matter what Q. The scaling limit (35) achieves that (36) represents the *true* scaling exponent of the rightmost (leftmost) band.

We conjecture that the scaling behaviour of the full invariant measure at the boundaries of the support is dominated by the scaling behavior of the right- and leftmost bands. This conjecture is supported by numerical investigations (both numerical solutions of Chapman–Kolmogorov equations and box-counting simulations) of the coarse grained measure. At the boundaries of the support, one expects the coarse grained density $P(l)/l$ to scale according to $P(l)/l \sim l^{\alpha-1}$, and qualitative changes of such scaling behaviour at $\alpha = 1$ and $\alpha = 2$ are correctly predicted by (36). See figs. 2a–c. Namely, as $l \to 0$

$$P(l)/l \to \begin{cases} \infty, & \text{if } \alpha < 1, \text{ i.e., } \phi'_+(x^*) < \frac{1}{2}, \\ 0, & \text{if } \alpha > 1, \text{ i.e., } \phi'_+(x^*) > \frac{1}{2}. \end{cases} \tag{37}$$

Note that this behaviour is consistent with our previous analysis (25). Moreover, considering the "left derivative" of the coarse grained density at x^*, viz. $-\partial_l(P(l)/l)$, one observes

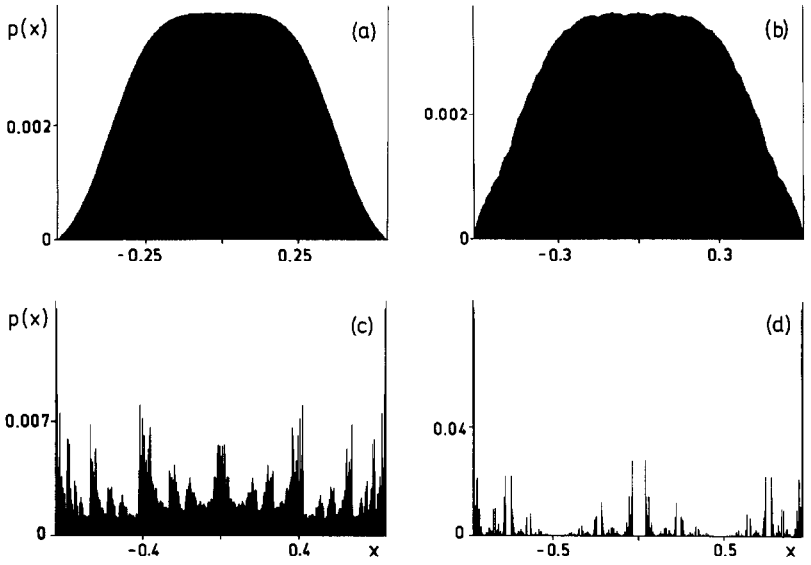

Fig. 2. Qualitatively different shapes of invariant measures. In (a)–(d) we have $\epsilon = 0.05$, 0.1, 0.4, and 1.0, respectively. The measure at the boundaries of the support is either zero, (a) and (b), or infinite, (c) and (d). The former case is further distinguished by the derivative of the outmost bands at the boundaries of the support, which either vanishes (a) or diverges (b). It is clearly visible that the multifractal measure in (a)–(c) covers the whole interval (fat fractal), whereas in (d) it is a thin fractal. The critical ϵ_c separating regions of qualitatively different behaviour are given in the main text. The histograms are calculated by a digital simulation of a trajectory of length 10^9. Note the different scales on both axes.

$$-\partial_l(P(l)/l) \sim -(\alpha-1)l^{\alpha-2} \to \begin{cases} 0\,, & \text{if } \alpha > 2, \text{ i.e., } \phi'_+(x^*) > 1/\sqrt{2}\,, \\ -\infty\,, & \text{if } 2 > \alpha > 1, \text{ i.e., } 1/\sqrt{2} > \phi'_+(x^*) > \frac{1}{2}\,, \\ \infty\,, & \text{if } 1 > \alpha, \text{ i.e., } \frac{1}{2} > \phi'_+(x^*)\,, \end{cases} \tag{38}$$

as $l \to 0$. By symmetry, analogous behaviour occurs at $-x^*$. The changes in the qualitative behaviour of the measure are clearly seen in fig. 2. If parameters are such that $P(l)/l \to 0$ as $l \to 0$, then for the linear palimpsestic scheme the change in derivatives occurs at $\lambda_c^{(3)} = 1/\sqrt{2}$. For the nonlinear encoding scheme, the corresponding critical values are $x_c^* = \sqrt{1-\sqrt{\frac{1}{2}}}$, or equivalently $\epsilon_c^{(3)} \simeq 0.064$.

Thus, our scaling analysis has allowed us to examine the behaviour of the invariant measure at the boundaries of its support. On the other hand, it is known that the scaling behaviour of the invariant measure on the *whole* support can be characterised by an infinite sequence of generalized fractal dimensions. This will be the subject of the next section.

5. Generalized fractal dimensions

As just announced, further quantitative characterization of the invariant measures is possible by determining their generalized fractal (or Renyi) dimensions [22]

$$D_q = \frac{1}{q-1} \lim_{l \to 0} \frac{\ln \Sigma_{i=1}^{N(l)} P_i^q}{\ln l}\,. \tag{39}$$

In (39), the sum is over all $N(l)$ nonoverlapping cells on length l used to cover $\{\phi_{(\sigma)}\}$, and P_i denotes

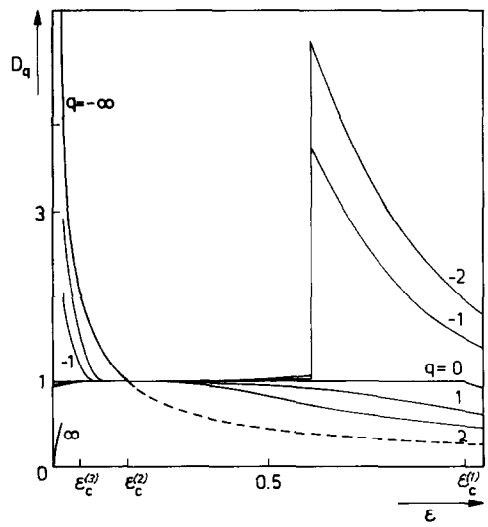

Fig. 3. Generalized fractal dimensions D_q of the invariant measure corresponding to the encoding scheme (6), for $q = 0, \pm 1, \pm 2$. The most remarkable feature is that the width of the multifractal spectrum becomes narrow near $\epsilon_c^{(2)}$ where the measure at the boundaries of the support jumps from zero to infinity. A higher resolution would reveal, that the drastic increase of D_{-1} and D_{-2} near $\epsilon = 0.6$ is continuous, and that the apparent cusps are smooth. The D_q are calculated in the thermodynamic formalism for a natural partition of the interval, comparing partition functions of generations $n = 13$ and $n = 14$. For $\epsilon < \epsilon_c^{(2)}$, $D_{-\infty}(\epsilon)$ is also displayed. For $\epsilon > \epsilon_c^{(2)}$, the dashed line shows the scaling exponent α of the right(left)most band, which approaches $D_\infty(\epsilon)$ as ϵ becomes large. For $\epsilon \to 0$, we have drawn $D_\infty(\epsilon) \to 0$ only schematically.

the total weight of the measure on cell i. The parameter q can take on any real value but is usually considered only for the integers $\mathbb{Z}$. For $q=0$, eq. (39) gives the so-called fractal or Hausdorff dimension of the support. By taking the limit $q \to 1$ in (39), one obtains the information dimension D_1, for $q=2$ the correlation dimension D_2, and so on. The D_q can be shown to be monotone non-increasing functions of q. The limiting values $D_{-\infty}$ and D_{∞} characterise contributions to the invariant measure originating from the most rare and the most dominant events, respectively; cf. eq. (39).

In cases, where the *cells at the boundary* of the support correspond to the most rare or most dominant events, our scaling analysis of the previous section would entail that $D_{-\infty}=\alpha$ or $D_{+\infty}=\alpha$, respectively, with α given by (36). Namely, if $\phi_+'(x^*)>\frac{1}{2}$ so that $\alpha>1$, then the cells at the boundary of the support correspond to the *most rare* events, hence $D_{-\infty}=\alpha$. On the other hand, as $\phi_+'(x^*)\to 0$, entailing that $\alpha \to 0$, then the cells at the boundary of the support represent the *most dominant* events, so that $D_{+\infty} \sim \alpha$.

For a one-scale fractal, with a measure which is homogeneous on a self-similar fractal support, all D_q are equal. Such is the case for the linear palimpsestic encoding scheme (5) when the gaps are open, $\lambda \le \frac{1}{2}$. Then

$$D_q = -\frac{\ln 2}{\ln \lambda}, \tag{40}$$

whatever q. Thus all D_q are equal to the Hausdorff dimension $D_0 = D_H$ of the support, which is a self-similar Cantor-set contained in I. If the support is not self-similar, or if bands are overlapping, this need no longer be the case.

To proceed with the computation of fractal dimensions in general, it is advantageous to drop the equipartition of the support in identical cells of size l as in (39) and use a *natural* partition $\{l_i\}$ generated by the mapping itself instead. In the thermodynamic formalism [22], one introduces a partition function

$$\Gamma(q,\tau,\{l_i\}) = \sum_i \frac{P_i^q}{l_i^\tau}, \tag{41}$$

and studies its scaling with increasing resolution, i.e., for $l_i \le l$, with $l \to 0$. With increasing resolution, $\Gamma(q,\tau,\{l_i\})$ will tend to zero or to infinity, *unless* $\tau=(q-1)D_q$. This observation is used to determine the D_q. It can be shown [25] that the generalized fractal dimensions so obtained agree with those determined from (39).

For parameters for which the support of the invariant measure is itself a fractal (with Hausdorff dimension strictly smaller than 1), the natural partition in the nth generation is given by the 2^n nonoverlapping images $I_{(\sigma)_n}$ of the initial interval I, each carrying the same total measure 2^{-n}. On the other hand, for parameters for which the $I_{(\sigma)_n}$ overlap, so that the support of the invariant measure is the whole interval I, this interval has to be repartitioned [24]. Here we use the 2^{n+1} endpoints of the $I_{(\sigma)_n}$, i.e., the images $\phi_{(\sigma)_n}(\pm x^*)$ of the end points $\pm x^*$ of I to construct $2^{n+1}-1$ nonoverlapping new intervals covering I. Each of these new intervals will generally carry weight coming from *several* bands $p_{(\sigma)_n}$. Explicit knowledge of the $p_{(\sigma)_n}(x)$ from (18) allows to take these contributions correctly into account. We determined the D_q by solving $\Gamma_n(\tau)-\Gamma_{n-1}(\tau)=0$ as an eigenvalue equation [26] for $\tau=(q-1)D_q$, where n denotes the generation used to determine Γ. In practice, $n=10,\ldots,15$ gives reasonably accurate results.

Figure 3 gives the multifractal spectrum D_q for the smooth nonlinear encoding scheme (6).

Qualitatively, the behaviour of the generalized dimensions can be understood as follows. The tails of the measure at the boundaries of the support determine $D_{-\infty}(\epsilon)$ for $\epsilon < \epsilon_c^{(2)}$ as discussed in section 5. In particular, $D_{-\infty}(\epsilon)$ diverges, as $\epsilon \to 0$. In the same limit, the mass of the distribution concentrates at the center of the support in a δ-function like fashion, so that $D_{\infty}(\epsilon)$ decreases to zero as $\epsilon \to 0$. In the opposite limit, $\epsilon \to \infty$, the invariant measure is a *thin* fractal, collapsing into two δ-functions at the endpoints of the support. This means that all fractal dimensions approach zero, implying that the width of the spectrum of fractal dimensions approaches zero, too.

Now we look at the fate of the multifractal as ϵ decreases from large to small values. For still very large ϵ, the fractal is thin and most bins are empty at any resolution. With decreasing ϵ, the fractal becomes more dense, and all fractal dimensions increase. If the gaps close, the fractal becomes fat, and $D_0 = 1$ for all $\epsilon < \epsilon_c^{(1)}$. However there are still very deep valleys and high peaks on all scales that cause D_q for $q > 0$ or $q < 0$ to be smaller or larger than D_0, respectively. Upon further decrease of ϵ the D_q for $q < 0$ decrease drastically and reach values close to 1, which announces that a dense nonfractal "background" on the whole support has emerged. Still, superimposed on this background, there are peaks on all scales, so that $D_q < 1$ for $q > 0$. At $\epsilon_c^{(2)}$, the behaviour of the measure at the boundaries of the support changes qualitatively, $p(\pm x^*)$ jumps from ∞ to zero as ϵ is decreased through this critical value, and $D_{-\infty} = 1$. For $\epsilon_c^{(3)} < \epsilon < \epsilon_c^{(2)}$, there are tails decreasing to zero with infinite slope as $x \to \pm x^*$. At $\epsilon_c^{(3)}$ there is yet another qualitative change, the slope jumps from ∞ to zero, i.e., $D_{-\infty} = 2$. The scaling behaviour of this slope determines $D_{-\infty}$ as discussed earlier.

6. Concluding remarks

In the previous sections, we have investigated invariant distributions of synaptic efficacies resulting from learning in forgetful memories. For two such learning rules, viz., (5) and (6), it has been demonstrated that invariant distributions generated by learning exhibit fractal of multifractal properties. As the parameters λ or ϵ characterizing the learning rules are varied, the invariant measures have been seen to undergo three "transitions", qualitative changes of their properties. One is associated with a closing of gaps in the support, the other two with changes of the measures at boundaries of their support.

Of course, it seems natural to ask whether properties of the invariant measures can be related to network performance, and, specifically, whether the transitions correspond to equally sharp transitions in network performance as measured, e.g., by the storage capacity. The answer to the first question, perhaps not too surprisingly, is "yes", whereas that to the more specific second question turns out to be "no". We now discuss why.

In a neural network storing a set of patterns $\{\xi_i^\rho\}$, as encoded in its synaptic matrix J_{ij}, the retrieval quality of any one of them depends on its *embedding strength* [27,28]

$$e_\nu = \frac{1}{N} \sum_{ij} J_{ij} \xi_i^\nu \xi_j^\nu \,. \tag{42}$$

In forgetful memories [8,10–14,28], the embedding strength of a given pattern depends on its *storage ancestry*. That is, if $J_{ij} = J_{ij}^{(\tau)}$ encodes patterns up to the one labeled τ, and if $\boldsymbol{\xi}^\nu$ was encoded n steps before $\boldsymbol{\xi}^\tau$ in the iterative scheme (2), then $e_\nu = e_{\tau - n} = e(n)$. The *decay* of the embedding strengths with storage ancestry n determines how fast a pattern is forgotten. The reason is that the retrieval quality degrades with decaying $e(n)$, and it will vanish once $e(n)$ falls below some level of noise present in the

system. This may be either thermal noise or static noise generated by interference between stored patterns due to nonzero mutual overlaps. The decay of $e(n)$ with n is always exponentially fast. The decay *rate* is determined by the average Lyapunov exponent γ of the stochastic mapping (8) associated with the learning rule (1),

$$e(n) \sim \exp\{\gamma n\} , \tag{43}$$

as $n \to \infty$. Now γ depends on both ϕ and the invariant measure μ created by the map with a given ϕ through

$$\gamma = \int_I \mathrm{d}\mu(x) \, \tfrac{1}{2} \sum_{\sigma = \pm 1} \ln \phi'_\sigma(x) = \int_I \mathrm{d}\mu(x) \ln \phi'_+(x) . \tag{44}$$

In this way, the invariant measure does have its effect on network performance in that the decay rate of embedding strengths and thereby the number of patterns that can be kept above noise levels (and thus retrievable) is determined by μ through (44). The average Lyapunov exponent does, in fact, *also* determine the decay rate of the static noise generated by each pattern, and a decrease of the decay rate of the embedding strengths implies a corresponding decrease of the decay rate of the noise. The optimal choice of these rates – i.e., that maximizing the storage capacity – through appropriate scaling of network parameters is, therefore, the outcome of a subtle compromise between two antagonistic optimization criteria, viz., letting the embedding strength of a pattern (but with it unavoidably also the "noise" it creates) decay as slowly as possible, versus letting the noise created by a pattern (but with it also the "signal") decay as quickly as possible [10–14,28].

For the linear encoding scheme (5), γ is easily evaluated. One has

$$\gamma = \int_I \mathrm{d}\mu(x) \ln \lambda_N = \ln \lambda_N . \tag{45}$$

Here we have used the fact that μ is normalized. Thus, for the decay rate to be small enough to guarantee extensive storage capacities, i.e., good retrieval for $n = \alpha N$ patterns with $\sigma > 0$ (here, α is *not* the scaling exponent introduced in sections 4 and 5), one must have $-\ln \lambda_N = \mathcal{O}(N^{-1})$ or, equivalently $\lambda_N = 1 - \mathcal{O}(N^{-1})$, which confirms the choice of Mézard et al. [12]. Note that with this choice the embedding strengths are slowly decaying functions on the α-scale. Note also that this scaling of λ_N with system size N entails that $\gamma \to 0$ as the thermodynamic limit is taken, implying that – in this limit – the memory is working just at the border of chaos.

For the nonlinear encoding scheme generated by the hyperbolic tangent in (6), a first rough estimate of γ, exploiting $\phi'_+(x^*) \le \phi'_+(x) \le 1$ for $x \in I = [-x^*, x^*]$ gives

$$\gamma = \int_I \mathrm{d}\mu(x) \ln(1 - \phi_+(x)^2) \le \ln(1 - x^{*2}) , \tag{46}$$

where x^* depends on ϵ through (11). For sufficiently small ϵ, x^* itself will be small so that (11) can be written $x^* \sim x^* + \epsilon - \frac{1}{3}(x^* + \epsilon)^3$. Hence $x^* = \mathcal{O}(\epsilon^{1/3})$, and thus $\gamma = \mathcal{O}(\epsilon^{2/3})$. This rough estimate, however, fails to take into account that, for small ϵ, the invariant measure is actually concentrated near the center of I, i.e., the origin, and therefore fails to produce the correct scaling of γ with ϵ. A sounder procedure is to exploit the smallness of x^*, hence x, to expand $\ln[1 - \phi_+(x)^2]$ in (46). This gives $\gamma \sim \langle x^2 \rangle = \int_I \mathrm{d}\mu(x) \, x^2$ as ϵ, and thereby x^*, go to zero. Detailed analysis [14] reveals $\langle x^2 \rangle \propto \epsilon$, as $\epsilon \to 0$.

Thus, in order to guarantee extensive storage capacities for this model, the scaling $\epsilon = \epsilon_N = \mathcal{O}(N^{-1})$ has been adopted [14]. Thus, for this system too, $\gamma \to 0$ in the thermodynamic limit.

The transitions undergone by invariant measures occur for too large values of ϵ_N and $1 - \lambda_N$ as to be of any significance, however, for storage capacities in the (thermodynamic) large-system limit. The average Lyapunov exponents vary continuously across the transitions, so that the decay of embedding strengths is not significantly affected by the occurrence of these transitions.

To summarize, in the present paper, we have investigated invariant distributions of synaptic efficacies resulting from learning in forgetful memories. For two such learning rules, it was demonstrated that invariant distributions generated by learning in a stationary environment exhibit fractal or multifractal properties. Previous analyses have shown [10–14] that in order to ensure extensive storage capacities in the thermodynamic limit, the stochastic mappings associated with forgetful memories are of interest only in the smooth limits $\epsilon_N \to 0$ or $\lambda_N \to 1$, where (multi) fractality of the invariant distributions – while still present and pronounced – has in some sense a trivial origin [25]: virtually all mass of the invariant measures is concentrated near the center of the support, and there are exponential tails exhibiting essential singularities at the boundaries of the support. Nevertheless, for large but finite systems, there are always optimal nonzero values of the parameters ϵ_N and λ_N, respectively which ensure the largest capacity for the given rule and system size N, so non-trivial (multi) fractality of J_{ij}-distributions *is* a feature of memories which forget. It remains an open question, though, whether there is a *direct* connection between characteristics of the (multi) fractal properties of the measure, such as the information dimension or the width of the spectrum of fractal dimensions, and the performance of the memory. This question is under investigation.

Very recently, in a different context, the invariant measure for the membrane potential in a single-neuron model [29] has been numerically generated, and a similar behaviour as in fig. 2 can be observed.

Acknowledgements

This work has been carried out during the last four odd years during several stays of the authors at each others institutions. We would like to thank the University of Leipzig (J.L.v.H., R.K., V.A.Z.), the Joint Institute for Nuclear Research, Dubna (U.B., R.K., A.L.), and the Sonderforschungsbereich-123 at Heidelberg University (U.B., V.A.Z.) for their hospitality and support extended to us while this work was done.

References

[1] F. Rosenblatt, Principles of Neurodynamics (Spartan Books, New York, 1962).
[2] M.L. Minsky and S. Papert, Perceptrons, 3rd Ed. (MIT Press, Cambridge, MA, 1969; enlarged version 1988).
[3] W. Krauth and M. Mézard, J. Phys. A 20 (1987) L745.
[4] E. Gardner, J. Phys. A 21 (1988) 257.
[5] D.E. Rumelheart and J.L. McClelland, Parallel Distributed Processing, Vol. 1 (MIT Press, Cambridge, MA, 1986).
[6] D.O. Hebb, The Organization of Behavior (Wiley, New York, 1949).
[7] L.N. Cooper, in: Nobel Symposia, Vol. 24, eds. B. Lundqvist and S. Lundqvist (Academic Press, New York, 1973) pp. 252–264.
[8] J.J. Hopfield, Proc. Natl. Acad. Sci. USA 79 (1982) 2554.
[9] D.J. Amit, H. Gutfreund and H. Sompolinsky, Phys. Rev. A 35 (1987) 2293.
[10] G. Parisi, J. Phys. A 19 (1986) L617.

[11] J.P. Nadal, G. Toulouse, J.P. Changeux and S. Dehaene, Europhys. Lett. 1 (1986) 535.
[12] M. Mézard, J.P. Nadal and G. Toulouse, J. Phys. (Paris) 47 (1986) 1457.
[13] B. Derrida and J.P. Nadal, J. Stat. Phys. 49 (1987) 993.
[14] J.L. van Hemmen, G. Keller and R. Kühn, Europhys. Lett. 5 (1988) 663.
[15] G. Radons, H.G. Schuster and D. Werner, in: Parallel Processing in Neural Systems and Computers, R. Eckmiller, ed. (Elsevier, Amsterdam, 1990) p. 261; Phys. Lett. A 174 (1993) 293.
[16] G. Györgyi and P. Rujàn, J. Phys C 17 (1984) 4207.
[17] P. Szépfalusy and U. Behn, Z. Phys. B 65 (1987) 337.
[18] U. Behn and V.A. Zagrebnov, J. Stat. Phys. 47 (1987) 939; J. Phys. A 21 (1988) 2151; Phys. Rev. B 38 (1988) 7115.
[19] J. Bene and P. Szépfalusy, Phys. Rev. A 37 (1988) 1702;
J. Bene, Phys. Rev. A 39 (1988) 2090.
[20] T. Tanaka, H. Fujisaka and M. Inoue, Phys. Rev. A 39 (1989) 3170; Prog. Theor. Phys. 84 (1990) 584.
[21] U. Behn, V.B. Priezzhev and V.A. Zagrebnov, Physica A 167 (1990) 457.
[22] T.C. Halsey, M.H. Jensen, I. Procaccia and B.I. Shraiman, Phys. Rev. A 33 (1989) 1141.
[23] A.N. Shiryayev, Probability (Springer, Berlin, 1984).
[24] U. Behn and A. Lange, in: From Phase Transitions to Chaos, G. Györgyi, I. Kondor, L. Sasvári and T. Tél, eds. (World Scientific, Singapore, 1992) pp. 217–226.
[25] Hao Bai Lin, Elementary Symbolic Dynamics and Chaos in Dissipative Systems (World Scientific, Singapore, 1989).
[26] T. Tél, Phys. Rev. A 36 (1987) 1502; A 36 (1987) 2507; Z. Naturforsch. 43a (1988) 1154;
M.J. Feigenbaum, I. Procaccia and T. Tél, Phys. Rev. A 39 (1989) 5359.
[27] H. Sompolinsky, Phys. Rev. A 34 (1986) 2571.
[28] J.L. van Hemmen, D. Grensing, A. Huber and R. Kühn, J. Stat. Phys. 50 (1988) 231; 50 (1988) 259.
[29] P.C. Bressloff, Phys. Rev. A 45 (1992) 7549.