Chaotic behavior of a layered neural network

B. Derrida*

The Institute for Advanced Studies, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

R. Meir

Department of Electronics, Weizmann Institute of Science, Rehovot 76100, Israel

(Received 20 April 1988)

We consider the evolution of configurations in a layered feed-forward neural network. Exact expressions for the evolution of the distance between two configurations are obtained in the thermodynamic limit. Our results show that the distance between two arbitrarily close configurations always increases, implying chaotic behavior, even in the phase of good retrieval.

Several authors have recently predicted the existence of chaos in the dynamics of soft-spin models,[1] diluted asymmetric spin-glasses,[2] and some neural network models.[3] For diluted asymmetric neural models this chaotic behavior can be observed even in the good retrieval phase. It is therefore interesting to know whether such chaotic behavior is a general feature of such neural network models or is special to those which have been studied so far.

The number of neural network models with exactly soluble dynamics is, up to now, very limited. To our knowledge they can be classified into two classes: (a) diluted networks and (b) layered networks. For diluted networks it has been shown[3] that a chaotic behavior is always present. The purpose of this paper is to show that layered neural networks, studied by Meir and Domany[4] (MD), exhibit a very similar chaotic behavior.

The model we study is the following. Consider a system of L layers, each containing N spins S_i^l. Each spin is connected to *all* spins on its neighboring layers, while no connections exist within each layer. The bonds between layers are *unidirectional*, i.e., the network is feed-forward. We concentrate here on zero temperature, i.e., fully deterministic, and *parallel* dynamics. Thus the state of layer l is determined by the state of the previous layer according to the equation

$$S_i^l = \text{sgn}\left[\sum_{j=1}^{N} J_{ij}^{l-1} S_j^{l-1} \right] . \tag{1}$$

The couplings J_{ij}^l are chosen according to the prescription[5]

$$J_{ij}^l = \frac{1}{N} \sum_{v=1}^{p} \xi_{iv}^{l+1} \xi_{jv}^l , \tag{2}$$

where ξ_{iv}^l, $v=1,\ldots,p$ are the representations of the stored key patterns on each layer. In the model studied in this paper we take each ξ_{iv}^l to be ± 1 with equal probability. More details concerning this model can be found in Ref. 4. The main quantity of interest in neural networks is the overlap of the current configuration S_i^l with the patterns ξ_{iv}^l embedded in the network. Therefore it is

convenient to define a quantity m_μ^l which measures the "closeness" of the current state S_i^l on layer l to the pattern $\xi_{i\mu}^l$ on layer l:

$$m_\mu^l = \frac{1}{N} \sum_{i=1}^{N} S_i^l \xi_{i\mu}^l . \tag{3a}$$

Reference 4 gave an exact form for the layer-to-layer recursion relation of the quantity m_1^l, which measures the "time" evolution of the overlap between the state S_i^l and key pattern number 1 (the "condensed" pattern), assuming that the initial state had large overlap m_1^1 with pattern 1 on the first overlayer, and random overlap with all others. In turns out that the evolution of m_1^l is coupled to the variable q^l which measures the mean square of the overlap of S_i^l with the other patterns ($\mu > 1$)

$$q^l = \sum_{\mu > 1} (m_\mu^l)^2 . \tag{3b}$$

Denoting the state of the "new" layer with a prime, the recursion relations obtained in Ref. 4 can be written

$$m' = \text{erf}(m/\sqrt{2q}) ,$$
$$q' = \alpha + \frac{2}{\pi} e^{-m^2/q} . \tag{4}$$

Here $\alpha = p/N$ is the ratio between the number of patterns embedded in the network and the number of sites per layer. We have also suppressed the subscript 1 attached to m, so that an m with no subscript refers to pattern 1. Note that we have modified our definition of q, with respect to that appearing in MD, by multiplying the q of MD by α. Solving these equations two types of behavior were found. For $\alpha > \alpha_c \simeq 0.27$ *any* initial configuration, however close to the pattern ξ_1^1 on the first layer, always iterates to an asymptotic zero overlap. For $\alpha < \alpha_c$ two types of behaviors can be seen. If $m^1 < m_c^1(\alpha)$ then, as for $\alpha > \alpha_c$, any initial configuration ends up with $m^* = m^\infty = 0$. However, if $m^1 > m_c^1(\alpha)$ the asymptotic overlap m^* is very close to 1. In fact, for $\alpha \rightarrow 0$ one finds

$$m^* \approx 1 - \left[\frac{2\alpha}{\pi} \right]^{1/2} e^{-1/2\alpha} . \tag{5}$$

In this paper we calculate the time evolution of the overlap between two configurations S_i^l and $\widetilde{S}_i^{\,l}$. Consider two initial states S_i^1 and S_i^1 on the first layer, and let them evolve according to the same rule given in (1) and (2). Define the overlap between the states on layer l as

$$Q^l = \frac{1}{N} \sum_{i=1}^{N} S_i^l \widetilde{S}_i^{\,l} \ . \tag{6}$$

Clearly, if the two configurations are identical, i.e., $Q^l = 1$, they remain identical at all times. In other words, $Q = 1$ is a fixed point. However, two thinks can happen: either $Q = 1$ is a stable fixed point of the dynamics, or it is unstable. If it is unstable, this means that two arbitrarily close initial configurations diverge away from each other.

In this paper, we obtain the analytic expression for the evolution of Q^1. We always find that $dQ'/dQ = \infty$ as $Q \rightarrow 1$, showing that the behavior is always chaotic.

As described in Eqs. (1) and (2) the time evolution of the two configurations is given by the following relations:

$$S_i' = \mathrm{sgn}(h_i) \ ,$$
$$\widetilde{S}_i' = \mathrm{sgn}(\widetilde{h}_i) \ , \tag{7}$$

where the fields h_i and $\widetilde{h}_i$ are given by

$$h_i = \sum_\mu \xi_{i\mu}' m_\mu \ ,$$
$$\widetilde{h}_i = \sum_\mu \xi_{i\mu}' \widetilde{m}_\mu \ . \tag{8}$$

Here m_μ is defined in Eq. (3a) and $\widetilde{m}_\mu$ is similarly defined, with $\widetilde{S}_i$ instead of S_i. The essential thing to note is that the quantities m_μ and $\widetilde{m}_\mu$ depend only on layers 1 to $(l-1)$ and *not* on the random variables $\xi_{i\mu}'$, of the new layer l, marked with a prime. Thus, it is easy to see that the random variables h_i and h_k for $i \neq k$ are *uncorrelated*, and similarly for $\widetilde{h}_i$ and $\widetilde{h}_k$. However, the fields h_i and $\widetilde{h}_i$ on the same site are correlated; treating these correlations is the basis of the calculation which follows. Let us note, that it is the independence of fields on different sites on the same layer, that makes this model exactly soluble. In the exactly soluble diluted networks,[3] it was also this independence which made the model soluble. However, the reason for the fields on different sites being independent in the diluted models is not the same: in Ref. 3 it was due to the fact that with probability approaching 1,

as the system size goes to infinity, the trees of ancestors of any two sites are disjoint.

To proceed further we must calculate the overlap between the two configurations. Due to the independence of the fields on different sites, we make the following replacement, valid in the limit $N \rightarrow \infty$:

$$\frac{1}{N} \sum_{i=1}^{N} S_i' \widetilde{S}_i' \rightarrow \langle S_i' \widetilde{S}_i' \rangle \ , \tag{9}$$

where the averages are with respect to the random variables $\{\xi_{i\nu}'\}$. Thus we have from Eqs. (7), (8), and (9)

$$Q' = \left\langle \mathrm{sgn}\left[\sum_\mu \xi_{i\mu}' m_\mu \right] \mathrm{sgn}\left[\sum_\mu \xi_{i\mu}' \widetilde{m}_\mu \right] \right\rangle \ . \tag{10}$$

Performing the averages over $\xi_{i\mu}'$, remembering that the m_μ's and $\widetilde{m}_\mu$'s are uncorrelated with them we obtain

$$Q' = \int \int dx\, dy \frac{1}{2\pi(q\widetilde{q} - s^2)^{1/2}}$$
$$\times e^{-(\widetilde{q}x^2 + qy^2 - 2sxy)/2(q\widetilde{q} - s^2)}$$
$$\times \mathrm{sgn}(m + x)\mathrm{sgn}(\widetilde{m} + y) \ , \tag{11}$$

where

$$q = \sum_{\mu > 1} m_\mu^2, \quad \widetilde{q} = \sum_{\mu (>1)} \widetilde{m}_\mu^2 \ ,$$
$$s = \sum_{\mu > 1} m_\mu \widetilde{m}_\mu \ , \tag{12}$$

and $m = m_1$, $\widetilde{m} = \widetilde{m}_1$. In deriving Eqs. (11) and (12) we have assumed that m_μ and $\widetilde{m}_\mu$ are $O(1/\sqrt{N})$ for $\mu > 1$. The recursion relations for m, $\widetilde{m}$, q, and $\widetilde{q}$ are given in Eq. (4).

To derive the recursion for s, we note that m_ν' and $\widetilde{m}_\nu'$ can be written as

$$m_\nu' = \frac{1}{N} \sum_i \mathrm{sgn}\left[\xi_{i\nu}' \xi_{i1}' m + m_\nu + \sum_{\substack{\mu \\ \mu \neq \nu, 1}} \xi_{i\nu}' \xi_{i\mu}' m_\mu \right] \ ,$$
$$\widetilde{m}_\nu' = \frac{1}{N} \sum_i \mathrm{sgn}\left[\xi_{i\nu}' \xi_{i1}' \widetilde{m} + \widetilde{m}_\nu + \sum_{\substack{\mu \\ \mu \neq \nu, 1}} \xi_{i\nu}' \xi_{i\mu}' \widetilde{m}_\mu \right] \ . \tag{13}$$

Forming the product $m_\nu' \widetilde{m}_\nu'$ and taking the sum over ν we get s' as defined in Eq. (12). It is convenient to separate the product $m_\nu' \widetilde{m}_\nu'$ into two contributions (the diagonal and the nondiagonal), as follows:

$$(m_\nu' \widetilde{m}_\nu')_\mathrm{d} = \frac{1}{N^2} \sum_i \mathrm{sgn}\left[\xi_{i\nu}' \xi_{i1}' m + m_\nu + \sum_{\substack{\mu,\nu \\ \mu \neq \nu, 1}} \xi_{i\nu}' \xi_{i\mu}' m_\mu \right] \mathrm{sgn}\left[\xi_{i\nu}' \xi_{i1}' \widetilde{m} + \widetilde{m}_\nu + \sum_{\substack{\mu,\nu \\ \mu! - \nu, 1}} \xi_{i\nu}' \xi_{i\mu}' \widetilde{m}_\mu \right] \ , \tag{14a}$$

$$(m_\nu' \widetilde{m}_\nu')_\mathrm{nd} = \frac{1}{N^2} \sum_{\substack{i,j \\ i \neq j}} \mathrm{sgn}\left[\xi_{i\nu}' \xi_{il}' m + m_\nu + \sum_{\substack{\mu \\ \mu \neq \nu, 1}} \xi_{i\nu}' \xi_{i\mu}' m_\mu \right] \mathrm{sgn}\left[\xi_{j\nu}' \xi_{j1}' \widetilde{m} + \widetilde{m}_\nu + \sum_{\substack{\mu \\ \mu \neq \nu, 1}} \xi_{j\nu}' \xi_{j\mu}' \widetilde{m}_\mu \right] \ . \tag{14b}$$

In order to calculate the averages of these quantities over $\xi_{i\nu}' = \pm 1$ we need the probability distributions of the variables appearing in these equations. Looking at the diagonal part (14a) first we define

$$x = \sum_{\substack{\mu \\ \mu \neq \nu, 1}} \xi_{i\nu}' \xi_{i\mu}' m_\mu, \quad y = \sum_{\substack{\mu \\ \mu \neq \nu, 1}} \xi_{i\nu}' \xi_{i\mu}' \widetilde{m}_\mu \ . \tag{15}$$

Calculating the probability distribution $P(x,y)$ of (x,y),

we obtain, to leading order in N,

$$P(x,y) = \frac{1}{2\pi(q\bar{q} - s^2)^{1/2}} \exp\left[-\frac{\bar{q}x^2 + qy^2 - 2sxy}{2(q\bar{q} - s^2)}\right],$$

$$(16)$$

where q, $\bar{q}$, and s are defined in Eq. (12).

In order to deal with the nondiagonal term (14b) we must separate the random variables into two groups. In the first group M_1, we have the terms for which $\xi'_{i\nu}\xi'_{i\mu} = \xi'_{j\nu}\xi'_{j\mu}$, and in the second group M_2, terms for which $\xi'_{i\nu}\xi'_{i\mu} = -\xi'_{j\nu}\xi'_{j\mu}$. Since the patterns are random, the sizes of the two groups are equal. We now need the probability distribution of the four (correlated) random variables x_1, x_2, y_1, y_2 defined by

$$x_1 = \sum_{\mu \in M_1} \xi'_{i\nu}\xi'_{i\mu} m_\mu, \quad x_2 = \sum_{\mu \in M_2} \xi'_{i\nu}\xi'_{i\mu} m_\mu,$$

$$(17)$$

$$y_1 = \sum_{\mu \in M_1} \xi'_{i\nu}\xi'_{i\mu}\tilde{m}_\mu, \quad y_2 = \sum_{\mu \in M_2} \xi'_{i\nu}\xi'_{i\mu}\tilde{m}_\mu.$$

It is easy to see that the probability distribution for x_1, x_2, y_1, y_2 can be factorized:

$$P(x_1, x_2, y_1, y_2) = \rho(x_1, y_1)\rho(x_2, y_2),$$

$$(18)$$

with

$$\rho(x_i, y_i) = \frac{1}{\pi(q\bar{q} - s^2)^{1/2}} \exp\left[-\frac{\bar{q}x_i^2 + qy_i^2 - 2sx_iy_i}{q\bar{q} - s^2}\right].$$

$$(19)$$

Note that (19) differs from (16) by a factor of 2 in the exponential, due to the fact that each group M_i has $\alpha N/2$ terms.

At this stage we are ready to calculate the averages of Eqs. (14a) and (14b), which can be expressed in the following form:

$$\left\langle \sum_{\nu > 1} m'_\nu \tilde{m}'_\nu \right\rangle_d = \frac{1}{N} \sum_{\nu > 1} \frac{1}{2} \sum_{\epsilon = \pm 1} \langle \text{sgn}(m_\nu + \epsilon m + x)\text{sgn}(\tilde{m}_\nu + \epsilon\tilde{m} + y) \rangle,$$

$$(20a)$$

$$\left\langle \sum_{\nu > 1} m'_\nu \tilde{m}'_\nu \right\rangle_{nd} = \sum_{\nu > 1} \frac{1}{4} \sum_{\epsilon, \bar{\epsilon}} \langle \text{sgn}(m_\nu + \epsilon m + x_1 + x_2)\text{sgn}(\tilde{m}_\nu + \bar{\epsilon}\tilde{m} + y_1 - y_2) \rangle.$$

$$(20b)$$

We note that since m_ν and $\tilde{m}_\nu$ are expected to be $O(1/\sqrt{N})$, we may, to lowest order in N, neglect the terms containing m_ν and $\tilde{m}_\nu$ in the diagonal contribution to Eq. (20). However, this is *not* permissible in the nondiagonal part, since these terms contribute a term of $O(1)$ to s' (actually, the lowest-order term, obtained by setting $m_\nu, \tilde{m}_\nu = 0$, vanishes).

Bearing these observations in mind, we calculate the various averages to obtain the final expression for s':

$$s' = \alpha \int \int dx\, dy \frac{1}{2\pi(q\bar{q} - s^2)^{1/2}}$$

$$\times e^{-(\bar{q}x^2 + qy^2 - 2sxy)/2(q\bar{q} - s^2)}$$

$$\times \text{sgn}(m + x)\text{sgn}(\tilde{m} + y)$$

$$+ \frac{2s}{\pi(q\bar{q})^{1/2}} e^{-m^2/2q} e^{-\tilde{m}^2/2\bar{q}}.$$

$$(21)$$

This expression, together with (4), gives the evolution of s. Of course, from (11) one can get the evolution of the overlap Q. It takes a particularly simple form in the case where the two patterns have zero overlap with the condensed pattern. [From Eq. (4) we note that if the initial overlap of each pattern with the condensed pattern is 0, it will remain so at all times.] Setting $m, \tilde{m} = 0$ in Eq. (21) we obtain

$$s' = \alpha \left[1 - \frac{2}{\pi}\tan^{-1}(q\bar{q}/s^2 - 1)^{1/2}\right] + \frac{2s}{\pi(q\bar{q})^{1/2}}.$$

$$(22)$$

In Fig. 1 we present our exact results and compare them to computer simulations, which can be seen to agree very well with them.

Returning to the general recursion relations expressed in Eqs. (4), (11), and (21), one can study the stability of the fixed point $Q = 1$, corresponding to the two patterns becoming identical. Consider two configurations with the same projection $(m = \tilde{m})$ on pattern 1, and with the same q $(q = \bar{q})$. One can calculate the stability of the fixed point $Q = 1$ (which corresponds to $s = q$). Using (4) and (21), one can show that for $(q - s)$ small

$$q' - s' = \frac{4\alpha}{\pi} \left[\frac{q - s}{q + s}\right]^{1/2} e^{-m^2/(q+s)} + O(q - s).$$
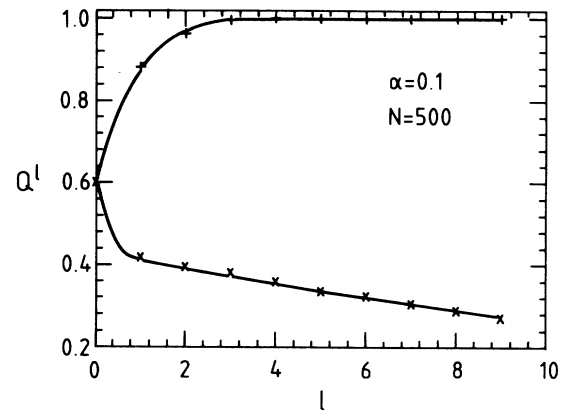
$$(23)$$

FIG. 1. Comparison between the analytic results, shown by continuous lines, and numerical simulations, for $\alpha = 0.1$. The upper curve corresponds to a situation where the initial overlaps m and $\tilde{m}$ were above $m_c^1(\alpha)$, while the lower curve corresponds to initial overlaps $m, \tilde{m} = 0$, i.e., below $m_c^1(\alpha)$. The initial values are $Q = 0.6$, $m = 0.8$, and $\tilde{m} = 0.64$, for the upper curve, and $Q = 0.6$, $m = \tilde{m} = 0$ for the lower curve. The simulation results were averaged over 2500 patterns, with $N = 500$ spins per layer.

We see that the fixed point $q = s$ is unstable. This implies that the fixed point $Q = 1$ is unstable too, because from (11) one has for $(q' - s')$ small

$$1 - Q' \approx \frac{1}{\alpha}(q' - s') . \tag{24}$$

This shows that the distance between two very close configurations always increases.

For $\alpha \to 0$ it is possible to give the analytic form for the fixed point value of the overlap Q. Expanding Eq. (11) at the fixed point, we find

$$Q^* \simeq 1 - \frac{8}{\pi^2} e^{-1/\alpha} . \tag{25}$$

We note that this is exactly the value obtained by Derrida et al.[3] for the diluted network. This equality holds also for the asymptotic value of m^*, as given in Eq. (5). As

found in the case of the diluted networks[3]

$$1 - m^* \gg 1 - Q^* ,$$

where m^*, the asymptotic value of m for $\alpha \to 0$ is given in Eq. (5). This implies that the two patterns approach one another more closely than either one approaches the stored pattern. Of course, as α increases, one expects m^* and Q^* to differ from those found in Ref. 3 since the equations giving the evolution of m and Q are different.

---

*Permanent address: Service de Physique Théorique, Centre d'Etudes Nucléaires de Saclay, F 91191 Gif-sur-Yvette, France.

[1]H. Sompolinsky, A. Crisanti, and H. J. Sommers (unpublished).

[2]B. Derrida, J. Phys. A **20**, L721 (1987).

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

[4]R. Meir and E. Domany, Phys. Rev. Lett. **59**, 359 (1987); Europhys. Lett. **4**, 645 (1987); Phys. Rev. A **37**, 608 (1988); R. Meir, J. Phys. (Paris) **49**, 202 (1988).

[5]J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).