Exact Solution of a Layered Neural Network Model

Ronny Meir and Eytan Domany
Department of Electronics, Weizmann Institute of Science, 76100 Rehovot, Israel
(Received 2 March 1987)

We study a deterministic parallel feed-forward neural network. Exact results for the response of the network are presented; given an initial state that has finite overlap with one stored random key pattern, we calculate the overlap on all subsequent layers (time steps). A region of good recall is separated by a first-order line from one of vanishing asymptotic overlap. Relaxation time to the limiting state is shown to diverge at the overloading transition.

PACS numbers: 87.30.Gy, 05.40.+j, 05.50.+q

When a spin-glass or neural network is set in an initial state, which then develops in time according to some dynamic rule, one of the most intriguing and important questions that can be asked concerns remanence.[1] For example, the overlap of the state at time t with the initial state converges for $t \to \infty$ to a constant, that measures the extent to which the system "remembers" its initial state. In a neural network (that models a memory) one typically starts with an initial state that has a finite overlap with a "key pattern," and studies the overlap of the final state with the same key pattern.[2,3] This overlap quantifies the extent to which the network is able to recall a learned key pattern. Analytic evaluation of remanence is usually a rather complicated problem that has not yet been accomplished. The exact solution of the equilibrium statistical mechanics of various spin-glass models[4] does not contain any dynamic information; neither do calculations that aim at obtaining the number of stable states of a system.[5]

In what follows we present analytic results for a quantity, m^*, analogous to remanence in a layered neural network. We find that as a parameter α is varied, m^* exhibits an "overloading" phase transition. For $\alpha > \alpha_c = 0.269$ we have $m^* = 0$, and $m^* > 0$ for $\alpha < \alpha_c$. The transition is first order. Furthermore, we calculated the manner in which the remanence relaxes with time to this asymptotic value. Relaxation is exponential, with a characteristic time scale τ that diverges at the transition, $\tau \sim (\alpha_c - \alpha)^{-1/2}$.

The model for which these results were obtained is a layered feed-forward network, that has been introduced recently.[6] Consider L layers; each contains N cells (spins), with a binary variable $S_i^{(l)} = \pm 1$ associated with cell i of layer l. Each cell is *connected to all* cells of the neighboring layers. The bonds are, however, *unidirectional*: The state of layer $l+1$ is determined by the state (at the previous time step) of layer l. Dynamics is fully *deterministic* (zero temperature) and *parallel*,

$$S_i^{(l+1)} = \operatorname{sgn}\left[\sum_{j=1}^{N} J_{ij}^{(l)} S_j^{(l)}\right]. \tag{1}$$

The couplings or bonds $J_{ij}^{(l)}$ are chosen by the popular prescription[2,3]

$$J_{ij}^{(l)} = \frac{1}{N}\sum_{\nu=1}^{\alpha N} \xi_{i,\nu}^{(l+1)} \xi_{j,\nu}^{(l)}, \tag{2}$$

where $\xi_{i,\nu}^{(l)}$ with $\nu = 1,2,\ldots,\alpha N$ are the stored key patterns.

Note that each key pattern carries a layer index. This is a central feature that characterizes the class of model neural networks studied in Ref. 6; it has conceptual as well as technical significance. Conceptually, it represents the fact that only the first layer of a network is in direct contact with the "external world," and hence only on the first (input) layer are the representations of the key patterns externally dictated. On all subsequent layers the system is free to choose an internal representation of any key pattern. Using an iterated learning procedure, we have shown[6] that the network is capable of perfect recall of key patterns. No such iterated learning is allowed in the network considered here (named "simple" in Ref. 6). We assume that the internal representations $\xi_{i,\nu}^{(l)}$ of the key patterns are randomly chosen; all $\xi_{i,\nu}^{(l)} = \pm 1$ with equal probability. It is precisely this fact, of the independent choice of representations on different layers, that technically allows analytic solution of our model. The solution yields information on the time development of the system: The first layer is set in an initial state, which determines the state of the next layer at the next (discrete) time step, and so on. Hence obviously our model can also be viewed as one with a single layer of cells, but time-dependent couplings. We emphasize that we do *not* address here the problem of the dynamics of the learning process; only the operational stage is studied.

Related but different layered networks were widely studied: as models of associative memory,[7] as processing devices trained to recognize translationally invariant patterns[8] or to identify the parity of a sequence,[9] as well as models for the emergence of spatial-opponent and orientation selective cells.[10] The complexity of the algorithms obtained precludes analytical statements on these networks. Other related models include the Little model[2]; if in our model we force the representations of the key

patterns to be the same on all layers, and set $J_{ii}=0$, the Little model results, with our layer index corresponding to time in the latter. For the Little model only the overlaps following two time steps were analytically calculated.[11] Another related model[12] is a cellular atomaton whose dynamic rule is a random function of position.[13] Our model can be thought of as a cellular automaton in which the dynamic rule is a (random) function of time.

By exact solution of our model we mean that given an initial state that has overlap $m^{(1)}$ with a key pattern, we have a recursive formula that yields the overlap on any subsequent layer and/or time step, averaged over all key patterns ξ. We calculate this solution using the same methods as applied to the Little-model stable states,[14] and to dynamics of the Sherrington-Kirkpatrick and Little models.[11] Here only an outline of the method is given: Details will be presented elsewhere.[15]

Consider a random assignment of $\nu=1,2,\ldots,\alpha N$ key patterns $\xi_{i,\nu}^{(l)}$ on each of L layers of the network. Choose an initial state on the first layer, $S_i^{(1)}$, such that its overlap with one key pattern (say $\nu=1$) is $m^{(1)}=O(1)$ and with the other patterns $\nu\neq 1$, $m_\nu^{(1)}=O(1/\sqrt{N})$. The question we ask is what is the probability $P(m^{(L)}\mid m^{(1)})$ that the dynamic rules (1) and (2) produce on layer L a state $S^{(L)}$ that has overlap $m^{(L)}=O(1)$ with key pattern $\xi_{i,1}^{(L)}$ [and $O(1/\sqrt{N})$ with the others]? In order to "enforce" the dynamic rule, consider on each site of layers $l>1$ the quantity[14]

$$R_i^{(l)}=S_i^{(l)}\textstyle\sum_j J_{ij}^{(l-1)}S_j^{(l-1)}. \tag{3}$$

When $R_i^{(l)}>0$ spin i on layer l is aligned with the "field" produced by the spins of the previous layer, as required by (1). Hence when the quantity

$$Y[\{\xi\};S^{(1)},S^{(L)}]=\sum_{\{S^{(2)},\ldots,S^{(L-1)}\}}\prod_{l=1}^{L-1}\prod_{i=1}^{N}\theta(R_i^{(l+1)})=1, \tag{4}$$

the dynamic rule indeed takes us from the initial state $S^{(1)}$ to the final state $S^{(L)}$.

Summing Y over all N_ξ realizations of the key patterns $\{\xi\}$,[16] one obtains the fraction $f(S^{(1)},S^{(L)})$ of all $\{\xi\}$ configurations that take the system from $S^{(1)}$ to $S^{(L)}$. This fraction depends on the initial and final states only through their overlaps with the respective key patterns, $m^{(1)}$ and $m^{(L)}$. Thus the probability we wanted to calculate is given by

$$P(m^{(L)}\mid m^{(1)})=\exp[Ns(m^{(L)})](1/N_\xi)\textstyle\sum_\xi Y[\{\xi\},S^{(1)},S^{(L)}], \tag{5}$$

where $\exp[Ns(m^{(L)})]$ is the number of states $S^{(L)}$ with overlap $m^{(L)}$, and

$$s(m)=-\tfrac{1}{2}(1-m)\ln[\tfrac{1}{2}(1-m)]-\tfrac{1}{2}(1+m)\ln[\tfrac{1}{2}(1+m)].$$

The conditional probability (5) is evaluated with some fairly standard techniques.[14,15] First, each θ function is represented as a double integral,

$$\theta(R)=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\mu\int_0^{\infty}dx\exp[i\mu(R-x)],$$

the summation over the ξ and S variables is carried out, and $P(m^{(L)}\mid m^{(1)})$ is expressed in terms of a multiple integral,

$$P(m^{(L)}\mid m^{(1)})=C\int\prod_{l=1}^{L-1}dq^{(l)}d\hat{q}^{(l)}\int\prod_{l=2}^{L-1}dp^{(l)}d\hat{p}^{(l)}dm^{(l)}\exp\left\{i\alpha N\left[\sum_{l=1}^{L-1}q^{(l)}\hat{q}^{(l)}+\sum_{l=2}^{L-1}\hat{p}^{(l)}p^{(l)}\right]\right.$$
$$\left.+\alpha N\ln Z(\hat{q}^{(1)},\ldots,\hat{q}^{(L-1)},\hat{p}^{(2)},\ldots,\hat{p}^{(L-1)})+N\sum_{l=2}^{L}[s(m^{(l)})+f^{(l)}]\right\}, \tag{6}$$

where C is a constant and

$$f^{(l)}=\tfrac{1}{2}(1-m^{(l)})\ln I_+^{(l)}+\tfrac{1}{2}(1+m^{(l)})\ln I_-^{(l)} \tag{7}$$

with[17]

$$I_\pm^{(l)}=\frac{1}{(2\pi)^{1/2}}\int_{b_\pm^{(l)}}^{\infty}\exp(-y^2/2)dy;\quad b_\pm^{(l)}=\frac{-\alpha p^{(l)}\pm m^{(l-1)}}{(\alpha q^{(l-1)})^{1/2}}. \tag{8}$$

Note that the integration variables $m^{(l)}$ are the overlap on the intermediate layers. The function Z is given by

$$Z=\int\prod_{l=1}^{L-1}d\lambda^{(l)}d\hat{\lambda}^{(l)}\exp\left\{i\textstyle\sum\lambda^{(l)}\hat{\lambda}^{(l)}-i\sum\hat{q}^{(l)}(\lambda^{(l)})^2-\frac{1}{2}\sum(\hat{\lambda}^{(l)})^2-i\sum\hat{p}^{(l)}\hat{\lambda}^{(l)}\lambda^{(l-1)}\right\}. \tag{9}$$

In the limit $N\to\infty$ the integral is evaluated by the saddle-point method; the integrand has the form $\exp(NF)$; setting equal to zero the derivatives of F with respect to the integration variables yields saddle-point equations. We supplement these by $\partial F/\partial m^{(L)}=0$, corresponding to evaluation of the value of $m^{(L)}$ at which $P(m^{(L)}\mid m^{(1)})$ is maximal. The

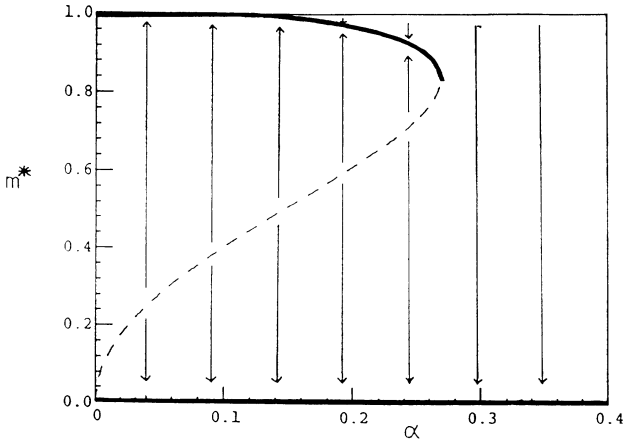

FIG. 1. Remanence, or asymptotic overlap m^* with a key pattern, vs α. The upper branch and the $m^*=0$ line (both solid lines) are stable fixed points of the dynamic recursions Eq. (10). The lower branch (dashed) is unstable. For $\alpha > \alpha_c = 0.269$ only the $m^*=0$ solution is accessible.

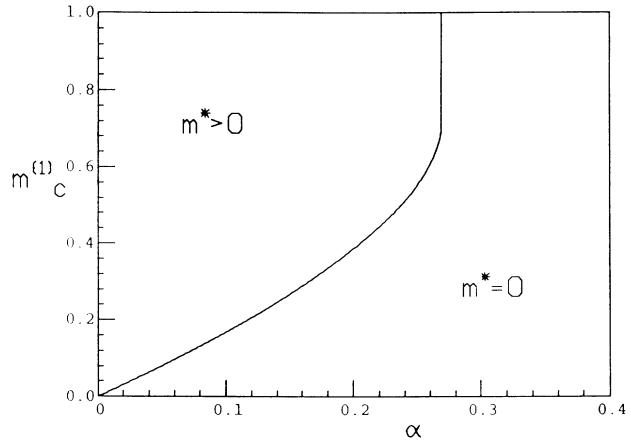

FIG. 2. The two phases, one with high remanence (memory) and one with $m^*=0$, are separated by a first-order line. For initial overlap $m^{(1)} < m_c^{(1)}$ vanishing, limiting m^* is obtained, even for $\alpha < \alpha_c$. In the phase with $m^* > 0$ the limiting overlap is given by the upper branch of Fig. 1.

distribution of $m^{(L)}$ goes in the limit $N \to \infty$ to a δ function around this value.

The saddle-point equations can be solved[15]; the relevant part of the solution consists of a recursion relation that determines the overlap on layer l, $m^{(l)}$, and the variable $q^{(l)}$ in terms of $m^{(l-1)}$ and $q^{(l-1)}$. The recursions have the form

$$m^{(l+1)} = \mathrm{erf}[m^{(l)}/(2\alpha q^{(l)})^{1/2}] \tag{10}$$

and

$$q^{(l+1)} = 1 + (2/\alpha\pi)\exp[-(m^{(l)})^2/\alpha q^{(l)}].$$

The value of $m^{(1)}$ is set by the initial state, and $q^{(1)}=1$. We have previously[6] obtained (10) for $m^{(2)}$; however, its form is only an approximation for the solution on subsequent layers. The difference is due to the fact that for all i and $l > 1$, $S_i^{(l)}$ and $\xi_{i,\mu}^{(l)}$ are correlated. When these correlations are neglected and the recursion with $q=1$ is used as an approximation for the recursion for $l > 2$ as well,[6] a *continuous* transition results, at $\alpha_c = 2/\pi$. This approximation is modified in (10) by the fact that the "width" parameter $q^{(l)}$ also changes with l, as if the effective value of α got renormalized by the increase of layer index and/or time. The effects of this modification on the predictions derived from the exact solution (10) are rather dramatic. The long-time, large-L behavior of the overlap is determined by the fixed points of (10), $m^{(l)} = m^*$ (the recursion for q is parasitic, dragged by $m^{(l)}$). The solution of the fixed-point equation is plotted versus α in Fig. 1. The $m^*=0$ fixed point is always stable; for $\alpha < \alpha_c = 0.269$, however, two additional solutions exist. The branch with higher values of m^* is stable and the lower branch unstable. This fixed-point topology is similar to the critical manifold of the q-state Potts models,[18] with α corresponding to q.

For the relevant parameters of the problem, namely α and the initial overlap $m^{(1)}$, the dynamics governed by Eq. (10) gives rise to the phase diagram of Fig. 2. For $m^{(1)} > m_c^{(1)}$ the limiting overlap $m^* \neq 0$, and its value is given by the upper branch of Fig. 1. For $\alpha \ll 1$ this branch has the form $m^* \simeq 1 - (2\alpha/\pi)^{1/2}\exp(-1/2\alpha)$.

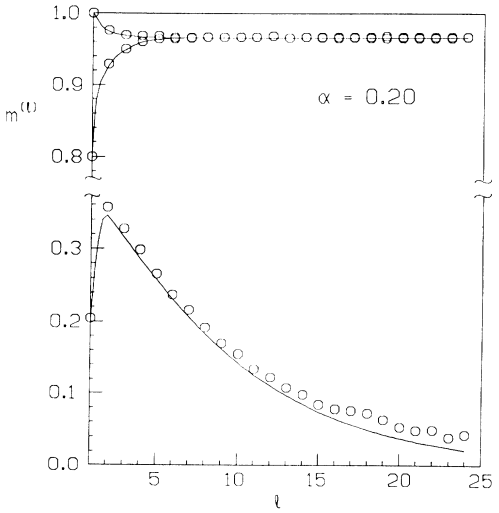

FIG. 3. Overlap $m^{(l)}$ as function of layer index (or time) l. The initial overlap for the two upper curves is above $m_c^{(1)}$ of Fig. 2; the circles represent simulations (with $N=200$) that agree perfectly with the analytic curves. The lower curve starts at an initial overlap of 0.2, below $m_c^{(1)}$. Deviations of simulations from the analytic curve are due to finite-size effects.

As we decrease $m^{(1)}$ (for fixed $\alpha < \alpha_c$), and cross $m_c^{(1)}(\alpha)$, the boundary of this phase, m^* jumps discontinuously to zero; the transition is first order. Such behavior was seen in another model recently.[19]

It is of interest to note that even though the transition is first order, the model exhibits "critical slowing down." Relaxation to the limiting value of m^* is exponential; $m^{(l)} - m^* \sim \exp(-l/\tau)$. The relaxation rate τ is determined by the recursion relations (10), linearized near m^*. Since as $\alpha \to \alpha_c$ two branches merge, one stable and one unstable, the fixed point at α_c must be marginally stable, and hence τ must diverge. Indeed, we find[15] that $\tau \sim (\alpha_c - \alpha)^{-1/2}$.

We have checked the exact solution (10) against numerical simulations. The main purpose of this is to find out the importance of finite-size effects. As evident from Fig. 3, for α not too close to the transition region excellent agreement with the exact solution (valid for $N \to \infty$) is obtained, even for N as low as 200. Finite-size effects become important as the phase boundary is approached. This can be seen to some extent in Fig. 3; while the upper two curves (corresponding to initial overlaps $m^{(1)}$ well within the $m^* \neq 0$ phase) exhibit perfect agreement with simulations, development from $m^{(1)} = 0.2$ does deviate slightly from the exact solution. This is due to the fact that near $m_c^{(1)}$ some members of the simulated ensemble flow to the "wrong" phase. However, as N increases, the relative weight of these "errors" decreases. The lower curve of Fig. 3 shows another interesting effect; even though the final overlap is 0, initially the overlap increases. Similar increase was found for the first time steps of the Little model.[11]

We have extended the calculations outlined above to initial states that have finite overlap with more than one key pattern, and to finite "temperature" dynamics.

We thank H. Orland, S. Levit, and B. Derrida for most helpful discussions, and B. Derrida for communicating his results prior to publication. This research was supported by the U.S.-Israel Binational Science Foundation, the Israel Academy of Sciences, and the Minerva Foundation.

Note added.—After completion of this work we received a preprint by Derrida, Gardner, and Zippelius, who solved a highly diluted asymmetric neural network, and found a continuous transition.

[1]W. Kinzel, Phys. Rev. B **33**, 5086 (1986).

[2]W. A. Little, Math. Biosci. **19**, 101 (1975).

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

[4]D. J. Amit, H. Gutfreund, and H. Sompolinsky, Ann. Phys. **173**, 30 (1987).

[5]C. De Dominicis, M. Gabay, T. Garel, and H. Orland, J. Phys. (Paris) **41**, 923 (1980).

[6]E. Domany, R. Meir, and W. Kinzel, Europhys. Lett. **2**, 175 (1986).

[7]T. Hogg and B. Huberman, J. Statist. Phys. **41**, 115 (1985), and Phys. Rev. Lett. **52**, 1024 (1984).

[8]M. Kerszberg and A. Bergman, to be published.

[9]D. E. Rummelhart, G. E. Hinton, and R. J. Williams, *Parallel Distributed Process* (MIT Press, Cambridge, MA, 1986), Vol. 1, p. 318.

[10]R. Linsker, Proc. Nat. Acad. Sci. U.S.A. **83**, 7508, 8390 (1986).

[11]E. Gardner, B. Derrida, and P. Mottishaw, J. Phys. (Paris) (to be published).

[12]S. A. Kauffman, Physica (Amsterdam) **100**, 145 (1984).

[13]B. Derrida and Y. Pomeau, Europhys. Lett. **1**, 45 (1986); B. Derrida and G. Weisbuch, J. Phys. (Paris) **47**, 1297 (1986).

[14]E. Gardner, J. Phys. A **19**, L1047 (1986).

[15]R. Meir and E. Domany, unpublished.

[16]Note that $S^{(1)}$ and $S^{(L)}$ are held fixed; $\xi_{l,\nu}^{(l)}$ are summed over, except for $\nu = 1$, $l = 1$, and $l = L$. The summation over these patterns is done so that the initial and final overlaps maintain their preassigned values.

[17]For $l = L$ Eq. (8) is to be used with $p^{(L)} = 0$.

[18]B. Nienhuis, A. N. Berker, E. K. Riedel, and M. Schick, Phys. Rev. Lett. **43**, 737 (1979).

[19]I. Kanter and H. Sompolinsky, Phys. Rev. A **35**, 380 (1987).