108.

Neural Networks, Vol. 1, pp. 63–73, 1988
Printed in the USA.

0893-6080/88 $3.00 + .00

ORIGINAL CONTRIBUTION

Statistical Neurodynamics of Associative Memory

SHUN-ICHI AMARI

University of Tokyo

KENJIRO MAGINU

Kyushu Institute of Technology

(*Received and accepted* 15 *October* 1987)

Abstract—*A new statistical neurodynamical method is proposed for analyzing the non-equilibrium dynamical behaviors of an autocorrelation associative memory model. The theory explains strange dynamical behaviors in recalling processes which are observed by computer simulations: Starting with an initial state close to a memorized pattern, the state monotonically approaches the memorized one. Starting with an initial state which is not so close to a memorized one, the state once approaches it but then goes away from it. The theory not only gives the relative and absolute capacity of the memory network without using the spin glass analogy, but it explains the non-equilibrium or transient dynamical behaviors of the recalling process by taking the long-term correlation effects into account. It thus explains the strange behaviors due to strange shapes of the basins of attractors.*

Keywords—Associative memory, Memory capacity, Recalling process, Basin of attractor, Statistical neurodynamics, Macroscopic state, Equilibrium.

INTRODUCTION

The autocorrelation type associative memory network, proposed more than fifteen years ago, has recently attracted much attention. In spite of the recent theoretical progress based on the statistical dynamical spin glass analogy, we have not yet succeeded in explaining its behaviors, in particular non-equilibrium dynamical properties in the recalling processes. Computer simulation shows interesting dynamical behaviors of recalling processes: When the network starts with an initial state which is within a critical distance from one of the memorized patterns, the state monotonically approaches the memorized one, and is trapped in a state very close to it. When the network starts with an initial state beyond a critical distance, the state once approaches the closest memorized one and their distance becomes shorter than the critical distance in many cases, but the state then goes back apart from it and is trapped in some spurious state. This is the normal behavior when the memorized pattern ratio is smaller than some critical value which is about 0.15. When the pattern ratio is larger than 0.15, the state may approach a memorized one, but always goes away from the memorized state and is trapped in a spurious one. This suggests that (a) the basin of the attractor of a memorized pattern has a strange shape, and (b) the distance is not a proper macroscopic state to explain the transient dynamical behaviors of the recalling process. We also need to derive the critical pattern ratio (the relative capacity) and the critical initial distance which guarantees the success of recalling. The present paper proposes a new statistical neurodynamical method without using the spin glass analogy. The method is applicable to a wide range of associative memory networks other than one treated in the present paper, in which we assume that memorized patterns are randomly generated and fixed, but the state transition is synchronous and deterministic.

The correlational associative memory has a long history going back to the early seventies. A number of important papers appeared in 1972, although these works are seldom referred to recently, but some results have been refound again. Nakano (1972), Anderson (1972), and Kohonen (1972) independently proposed the correlation type associative memory in this year. Amari (1972a) gave a mathematical analysis of the stability of memorized patterns. He studied not only the stability of equilibria of memorized patterns in symmetric connections, but also analyzed the dynamical behaviors of associative networks of asymmetric connections in which dynamical pattern sequences were memorized and recalled. Uesaka and Ozeki (1972) also gave the (relative) capacity of such a network. Since

Requests for reprints should be sent to Shun-ichi Amari, Faculty of Engineering, University of Tokyo, Bunkyo-ku Tokyo, 113 Japan.

then there have appeared a number of important contributions to the associative memory model. Little (1974) introduced stochastic behaviors and gave statistical dynamical treatment (see also Little & Shaw, 1978). Kohonen and Ruohonen (1973) proposed the generalized inverse type associative memory. Amari (1977a) proposed a natural implementation of the generalized inverse learning and analyzed its resolution and dynamical behavior. A similar method was used by Kinzel (1985) and Shinomoto (1987) (see also Kohonen, 1977 for various applications of the associative memory). Rolls (1987) proposed a model of the hypocampus having the associative memory mechanism.

Hopfield (1982, 1984) introduced asynchronous behaviors and gave the spin glass analogy. This opened a way to a new theoretical method, and interesting papers followed (e.g., Toulouse, Dehane, & Changeux, 1986, Peretto, 1984, Kinzel, 1985). Amit, Gutfreund, and Sompolinsky (1985a, 1985b) gave a statistical analysis of the associative memory model based on the equilibrium theory. However, this excellent theory is still unsatisfactory because it is applicable only to the network of symmetric connections with stochastic behaviors, and it does not elucidate the dynamic recalling processes but only gives the equilibrium distributions.

We develop in the present paper a new statistical neurodynamical method to elucidate the dynamical properties of the associative memory. The statistical neurodynamical method has been so far applied to the macroscopic dynamical analysis of randomly connected neuron networks. Randomly connected neuron networks were studied by many researchers long years ago, and their multistable characteristics were analyzed by Harth, Csermely, Beek, and Lindsay (1971) and by Amari (1971, 1972b), and their oscillatory behaviors were analyzed by Amari (1971, 1972b) (cf. Wilson & Cowan, 1972). The macroscopic state transition equations were derived in these papers from the microscopic neuronal state transition equations by using the central limit theorem and the law of large numbers. However, there is a mathematical difficulty in deriving these macroscopic state transition equations, because the long-term temporal correlations of random variables might not be neglected. This is the same difficulty we encounter when we derive the Boltzmann equation in statistical mechanics (see Kac, 1959), as was pointed out by Rozonoer (1969). Amari (1974) and Amari, Yoshida, and Kanatani (1977) studied this problem mathematically and solved it in some cases.

The macroscopic behavior of the associative network can also be analyzed by the statistical neurodynamical method. However, we need to evaluate the long-term temporal correlational effects, which can be neglected in the previous applications. Roughly speaking, the temporal correlations are explained as follows. Let $\mathbf{x}_{t+1} = T_W \mathbf{x}_t$ be the microscopic state transition equation, where T_W is the non-linear state transition operator of a net whose connection weights are given by the matrix $W = (w_{ij})$. When W is a randomly determined matrix, we can derive macroscopic equations concerning the state transition by considering the law of large numbers, provided $\mathbf{x}_t$ is independent of W. However, $\mathbf{x}_t = T_W \mathbf{x}_{t-1}$ also depends on the same random matrix W, so that we need to take the correlations of two (or more) T_w's in $\mathbf{x}_{t+1} = T_W(T_W \mathbf{x}_{t-1})$ into account, and so on. This is the problem we should solve for analyzing the behaviors of associative memory networks.

The present paper gives a mathematical method for studying associative memory networks, in which information is encoded in a distributed manner. The neural system also has another possibility of forming localized information representations in the form of a cortical map. There have been proposed systematic mathematical studies which elucidate the properties of self-organizing local representation of information. They consist of the dynamics of excitation patterns in neural fields (Amari, 1977b; Kishimoto & Amari, 1979), learning and self-organizing characteristics of neural systems (Amari & Takeuchi, 1978), and the geometrical and topological properties of formation of information maps in the brain (Amari, 1980, 1983; Takeuchi & Amari, 1979). The third interesting possibility is information representation by back error propagation learning rule (Rumelhart, Hinton, & Williams, 1986). It should be remarked that the generalized delta rule was proposed in the sixties (Amari, 1967), and was applied to learning of hidden units, although only a very small scale computer simulation was done.

After the final version of the present paper was completed, we found two interesting papers related to the subject, Meir and Domany (1987), and McEliece *et al.* (1987).

AUTOCORRELATION ASSOCIATIVE MEMORY

Let us consider a formal neuron which receives n input signals $x_1, \ldots, x_n$ and emits one binary output signal z by calculating the sign of the weighted sum of inputs, as

$$z = \mathrm{sgn}(\sum_{i=1}^{n} w_i x_i). \tag{2.1}$$

Here, $\mathrm{sgn}(u) = 1$ when $u > 0$, and $\mathrm{sgn}(u) = -1$ when $u < 0$. The neuron is in the excited state when $z = 1$, and is otherwise in the quiescent state. Let us consider a network consisting of n mutually interconnected formal neurons, and let w_{ij} be the synaptic efficacy or the weight of connection from the jth neuron to the ith neuron. The present state of the network is represented by a column vector $\mathbf{x} = (x_1, \ldots, x_n)'$ whose ith component x_i is the state, that is, the output, of the ith neuron. We assume that every neuron changes its state at discrete times $t = 1, 2, \cdots$ synchronously, and let

$\mathbf{x}_t$ be the state of the net at time t. We then have the following state transition equation,

$$x_i^{t+1} = \mathrm{sgn}(\sum_j w_{ij} x_j^t),$$

where x_j^t is the jth component of the state vector $\mathbf{x}_t$ at time t. We may abbreviate this equation as

$$\mathbf{x}_{t+1} = T\mathbf{x}_t \tag{2.2}$$

Here T is the non-linear state transition operator defined by

$$T\mathbf{x} = \mathrm{sgn}(\mathbf{W}\mathbf{x}), \tag{2.3}$$

where $\mathbf{W} = (w_{ij})$ is the connection matrix and sgn is operated componentwise, that is, the ith component of $T\mathbf{x}$ is given by

$$(T\mathbf{x})_i = \mathrm{sgn}(\sum_j w_{ij} x_j).$$

A net has 2^n states. A state x is an equilibrium state, when it satisfies

$$\mathbf{x} = T\mathbf{x}.$$

The basin $B(\mathbf{x})$ of an equilibrium state $\mathbf{x}$ is the set of those states each of which falls in the state $\mathbf{x}$ after a finite number of state transitions, that is,

$$B(\mathbf{x}) = \{\mathbf{x}_0 \mid \text{there exists an } N \text{ such that } T^N\mathbf{x}_0 = \mathbf{x}\}.$$

Let $\mathbf{s}^1, \mathbf{s}^2, \ldots, \mathbf{s}^m$ be m state vectors. If we can choose the connections $\mathbf{W}$ such that all of them are equilibrium states,

$$T\mathbf{s}^\alpha = \mathbf{s}^\alpha, \quad \alpha = 1, 2, \ldots, m$$

we say that these vectors are "memorized" in the network in the form of its equilibria. A vector $\mathbf{s}$ is "recalled" from any vector $\mathbf{x}_0$ belonging to its basin $B(\mathbf{s}^\alpha)$, because the sequence of states $\mathbf{x}_t$ beginning with an initial state $\mathbf{x}_0 \in B(\mathbf{s}^\alpha)$,

$$\mathbf{x}_{t+1} = T\mathbf{x}_t, \quad t = 0, 1, 2, \cdots$$

converges to the memorized $\mathbf{s}^\alpha$ within a finite number of state transitions.

The autocorrelation associative memory uses the following connection matrix

$$\mathbf{W} = \frac{1}{n}\sum_{\alpha=1}^{m} \mathbf{s}^\alpha \mathbf{s}^{\alpha\prime},$$

where $\mathbf{x}'$ is the transposition of $\mathbf{x}$, in other words.

$$w_{ij} = \frac{1}{n}\sum_{\alpha=1}^{m} s_i^\alpha s_j^\alpha \tag{2.4}$$

in the component form, s_i^α being the ith component of vector $\mathbf{s}^\alpha$. We usually put $w_{ii} = 0$. When s^α are mutually orthogonal, this scheme works very well, because

$$W\mathbf{s}^\alpha = \mathbf{s}^\alpha$$

holds for any α, so that

$$T\mathbf{s}^\alpha = \mathbf{s}^\alpha.$$

However, when the memorized vectors are not orthogonal, $T\mathbf{s}^\alpha = \mathbf{s}^\alpha$ does not necessarily hold, because of mutual interference of the memorized patterns $\mathbf{s}^\alpha$. Moreover, the interference may produce many spurious memories $\mathbf{s}$ other than $\mathbf{s}^\alpha$ which also satisfy $T\mathbf{s} = \mathbf{s}$. There are some methods to overcome this difficulty. One is to encode memory items into randomly generated patterns $\mathbf{s}^\alpha$, so that the orthogonality holds approximately. Another way is to use the generalized inverse (see Amari, 1977a; Kohonen, 1977).

The present paper studies the dynamical characteristic of recalling processes when $\mathbf{s}^\alpha$ are randomly generated, or more precisely, s_i^α are independent random variables taking 1 and -1 with probability $\frac{1}{2}$ each. We search for asymptotic properties which hold when the number n of neurons and the number m of memory patterns are large. More precisely, we consider an ensemble $\mathcal{E}$ of networks whose connection matrices are determined from m randomly and independently generated patterns. Each network behaves in a deterministic manner. There are infinitely many such networks generated by the same probability law. We search for those properties which are valid for "almost all" networks in $\mathcal{E}$ as n and m tend to infinity. We put $r = m/n$, and call it the pattern ratio. This is because the behavior of the associative net might be determined solely depending on the ratio r, when n and m are large.

ABSOLUTE STABILITY

We first evaluate the probability $P_n(m)$ that any randomly generated pattern s^α is an equilibrium state of the net. When

$$\lim_{n\to\infty} P_n(m) = 1, \tag{3.1}$$

the net is capable of memorizing m patterns in the form of its equilibria. Therefore, the least upper bound C_a of $r = m/n$ for which Equation (3.1) holds, gives the absolute capacity of the associative memory net. This implies that at most $m = nC_a$ patterns can be memorized in a net. Unfortunately, it is known that $C_a = 0$ (see, e.g., Weisbuch, 1985). We prove it briefly, by evaluating $P_n(m)$ a little more precisely than before.

Given a memorized vector $\mathbf{s}^\alpha$, its next state $T\mathbf{s}^\alpha$ is written as

$$T\mathbf{s}^\alpha = \mathrm{sgn}\left(\frac{1}{n}\mathbf{s}^\alpha\mathbf{s}^\alpha \cdot \mathbf{s}^\alpha + \frac{1}{n}\sum\nolimits'_\beta \mathbf{s}^\beta\mathbf{s}^\beta \cdot \mathbf{s}^\alpha\right) = \mathrm{sgn}(\mathbf{s}^\alpha + \mathbf{N}),$$

where sgn is operated componentwise, $\cdot$ is the inner product, and $\sum'_\beta$ denotes the summation over β which is not equal to α, that is,

$$\sum\nolimits'_\beta = \sum_{\beta\neq\alpha}.$$

Here

$$N = \frac{1}{n} \sum'_\beta s^\beta s^\beta \cdot s^\alpha \quad (3.2)$$

represents the interference term originating from the superposition of many vectors s^β. The ith component N_i of N is written as

$$N_i = \frac{1}{n} \sum'_\beta \sum_j s_i^\beta s_j^\beta s_j^\alpha,$$

where s_i^β are randomly generated independent variables. Hence, by applying the central limit theorem, each component N_i of N is regarded as normally distributed with mean 0 and variance

$$n(m-1)/n^2 \doteqdot r.$$

The ith component of Ts^α is altered by the noise, when N_i is smaller than -1 (if s_i^α is equal to 1), or N_i is larger than 1 (if s_i^α is equal to -1). The probability that this error occurs is given by

$$p = \text{Prob}\{N_i < -1\} = \text{erf}(n/m),$$

where erf is the error integral

$$\text{erf}(u) = \int_{-\infty}^{-u} (2\pi)^{-1/2} \exp\left\{-\frac{t^2}{2}\right\} dt.$$

Since $Ts^\alpha = s^\alpha$ holds when none of the n components are altered by the noise, we have

$$P_n(m) = (1-p)^n = \{1 - \text{erf}(n/m)\}^n. \quad (3.3)$$

By evaluating this probability (Appendix A), we have

$$\lim_{n\to\infty} P_n(m) = 1,$$

when m/n is no larger than

$$r(n) = \frac{m}{n} = \frac{1}{2 \log n - \log\log n}. \quad (3.4)$$

This shows that the absolute memory capacity is 0,

$$C_a = 0.$$

Theorem 1. A pattern is exactly memorized as an equilibrium state of the net, when

$$m = \frac{n}{2 \log n - \log\log n} + \text{smaller order terms.}$$

It should be noted that the term loglog n should not be neglected even when $n = 10{,}000$. For example, when $n = 1{,}000$, $r(n) = 0.084$, while $1/(2 \log n) = 0.072$; when $n = 10{,}000$, $r(n) = 0.062$, while $1/(2 \log n) = 0.054$.

The absolute stability is a strong criterion. Even if $C_a = 0$, it is not necessary to be so pessimistic. If we do not require absolutely precise recalling of s^α, but require only the convergence to one that is "sufficiently close" to the memorized s^α, we have another definition of the capacity (cf. Amit et al., 1985a, 1985b). This capacity is calculated through another more detailed statistical analysis of the dynamics of recalling processes. We show computer simulated experiments before we give the theoretical analysis.

DYNAMICAL BEHAVIORS OF ASSOCIATIVE NET—COMPUTER SIMULATION

Let us consider the dynamical process of recalling one of the memorized patterns, say s^1. Let $a(x)$ be the direction cosine between x and s^1 defined by

$$a(x) = \frac{1}{n} s^1 \cdot x = \frac{1}{n} \sum s_i^1 x_i.$$

It is connected with the normalized Hamming distance $d(x)$ between x and s^1 by

$$a(x) = 1 - 2d(x),$$

$$d(x) = \frac{1}{2n} \sum_i |x_i - s_i^1|.$$

Given an initial state x_0, we study how it approaches s^1 in the recalling process. That is, we study the dynamical change of

$$a_t = a(x_t), \quad d_t = d(x_t)$$

where

$$x_t = Tx_{t-1} = T^t x_0$$

is the state of the net at time t.

Figure 1 shows a typical result, where $n = 5{,}000$, $m = 400$, $r = 0.08$. The ordinate is time t, the abscissa is the direction cosine a_t, and the curves show the dynamical processes of recalling, starting at various initial a_0. Any curve is cut off at the time when the state x_t falls in an equilibrium state which is not necessarily equal to s^1. Figure 2 shows another typical result with $r = 0.2$.

We can observe the following facts from the simulation experiments.

1. The patterns s^α themselves are not equilibria, so that almost all processes fail to find the memorized one in the exact sense. This is in good agreement with the theoretical result $C_a = 0$, because, for $n = 5{,}000$, $m = 400$,

$$0.08 = \frac{m}{n} > \frac{1}{2 \log n - \log\log n} = 0.67$$

2. In spite that $n = 5{,}000$ is very large, the state converges very quickly to an equilibrium. This is the fact noticed and studied theoretically in statistical neurodynamics (Amari, 1974). No limit cycles are found in the experiments, in agreement with the asynchronous state transition case where the non-existence of limit cycles was proved (Hopfield, 1982).

3. In the case of $r = 0.08$ (Figure 1), we observe the following threshold phenomenon: There exists a threshold $h(r)$ such that any recalling process a_t con-

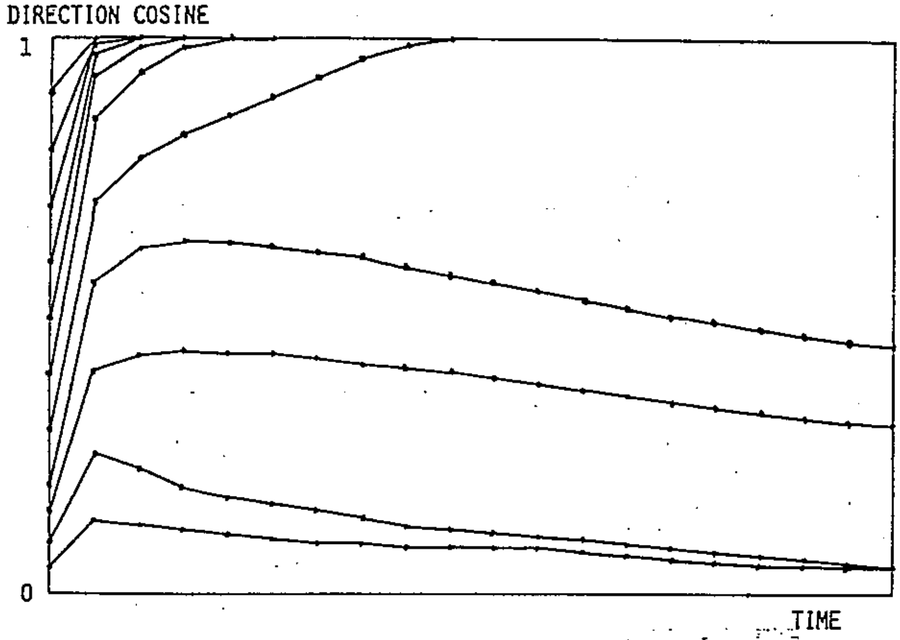

FIGURE 1. Dynamic behaviors of recalling processes; simulation with $n = 5000$, $m = 400$, $r = 0.08$.

verges to a value $\bar{a}(r)$ which is close to 1, when the initial direction cosine a_0 is greater than $h(r)$. When the initial a_0 is smaller than $h(r)$, the direction cosine goes up in the beginning but eventually decays and stops at various values, falling into equilibria. We may therefore say that the pattern s^1 is recalled correctly in the macroscopic or relative sense, if the initial direction cosine a_0 is larger than $h(r)$. The threshold $h(r)$ and the finial recalled $\bar{a}$ depend on r. This dynamical property holds for $r = 0 \sim 0.15$. However, the situation changes as r becomes larger than some critical value $\bar{r}$, which is about 0.15 in good agreement with Hopfield's observation (1982) or the theoretical result of Amit et al. (1985b).

4. Figure 2 shows a typical behavior in case with $r > \bar{r}$. The direction cosine never reaches a value close to 1. Even if it increases in the beginning of the recalling process, it soon decreases and stops at various values. Therefore, there is no thresholding effect which we see in the case of $r < \bar{r}$, and the recalling process fails to reach in a small neighborhood of s^1. One may say that the threshold value $h(r)$ becomes larger than 1 in this case. This is because too many patterns are memorized in the net and the net is overload.

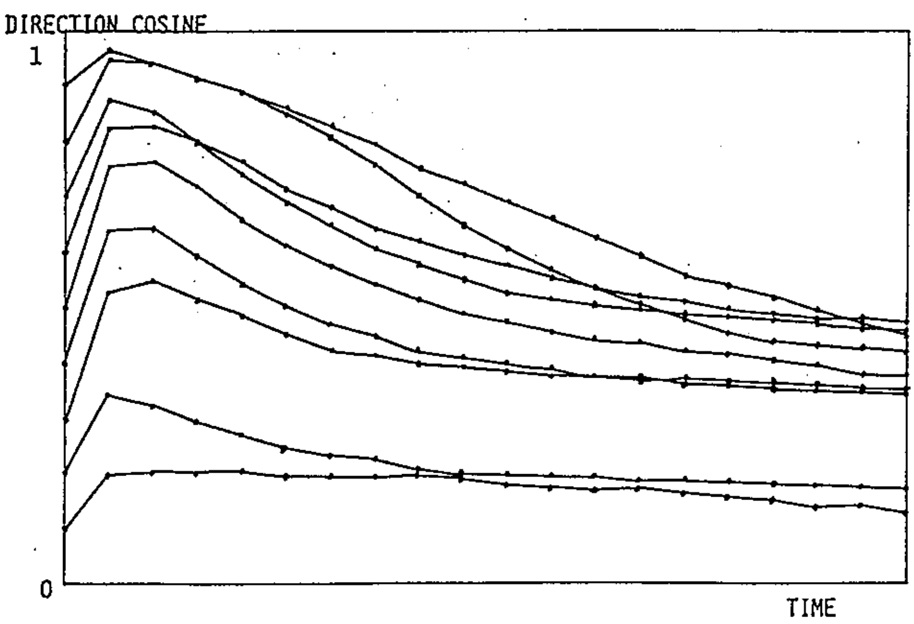

FIGURE 2. Dynamic behaviors of recalling processes; simulation with $n = 3000$, $m = 600$, $r = 0.2$.

The critical value $\bar{r}$ defines another capacity C_r of the associative memory in the relative or macroscopic sense,

$$C_r = \bar{r}.$$

If the memory pattern ratio is smaller than C_r, the recalling process is relatively successful if the initial direction cosine is larger than $h(r)$, in the sense that the final recollected pattern $\bar{\mathrm{s}}$ is very close to s^1. The recalled $\bar{\mathrm{s}}$ is sufficiently useful in the later information processing in the neural system. However, if r is larger than C_r, the recalling process fails even in the relative sense. This observation suggests that a bifurcation occurs at $\bar{r}$ (see Figure 3) concerning the macroscopic dynamical process to be studied theoretically.

The present paper proposes a method of statistical neurodynamics to explain the threshold phenomenon and to calculate $h(r)$ and C_r theoretically.

TRANSITION OF DIRECTION COSINE

We begin with a naive approach to the dynamics of direction cosine. Given an initial state x, the ith component of the next state $\mathrm{x}' = T\mathrm{x}$ is given by

$$x_i' = \mathrm{sgn}(u_i),$$

$$u_i = \frac{1}{n}\sum w_{ij}x_j = \frac{1}{n}\sum (s_i^\beta s_j^\beta x_j) = a(\mathrm{x}) + N_i. \quad (5.1)$$

The noise term

$$N_i = \frac{1}{n}\sum'_{\beta,j} s_i^\beta s_j^\beta x_j \quad (5.2)$$

is the sum of a large number of independent and identically distributed random variables $s_i^\beta s_j^\beta x_j$, so that it is asymptotically subject to the normal distribution $N(0, r)$ with mean 0 and variance r. Hence, we can put

$$x_i' = s_i^1 \mathrm{sgn}(a + \sqrt{r}\epsilon_i)$$

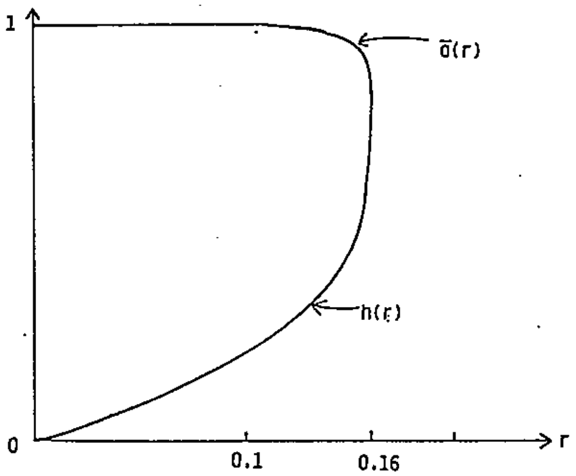

FIGURE 3. Equilibrium and threshold direction cosine versus pattern ratio r; theoretical curve.

where ϵ_i is subject to $N(0, 1)$. Although u_i are not mutually independent, the law of large numbers guarantees that the next direction cosine $a' = a(\mathrm{x}')$,

$$a' = \frac{1}{n}\sum s_i^1 x_i' = \frac{1}{n}\sum \mathrm{sgn}(a + \sqrt{r}\epsilon_i),$$

converges to the (ensemble) average of $\mathrm{sgn}(a + \sqrt{r}\epsilon_i)$, which is given by

$$a' = E[\mathrm{sgn}(a + \sqrt{r}\epsilon_i)] = 2\,\mathrm{Prob}\{a + \sqrt{r}\epsilon_i > 0\} - 1 = F\left(\frac{a}{\sqrt{r}}\right),$$

where E denotes the expection and

$$F(s) = \int_{-s}^{s} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u2}{2}\right\} du. \quad (5.3)$$

If we substitute x_t for x, then $\mathrm{x}' = \mathrm{x}_{t+1}$. This leads to the following dynamical equation

$$a_{t+1} = F(a_t/\sqrt{r}), \quad (5.4)$$

which was studied by Amari (1977a) and also by Kinzel (1985), and Shinomoto (1987). This equation is quite correct for $t = 0$. However, the simulated results show that this does not hold for $t = 1, 2, \ldots$. This is because x_t depends on s^β. This can easily be understood by writing the noise term as

$$N_i = (1/n)\sum'_\beta s_i^\beta\, \mathrm{s}^\beta \cdot \mathrm{x}_t,$$

where $\mathrm{x}_t = T\mathrm{x}_{t-1}$ is determined from x_{t-1} depending on s^β. This shows that N_i is not a sum of independent random variables but depends on s^β in a very complicated manner. Hence, when we evaluate the probability distribution of N_i, we need to take the correlations of s^β and x_t into account. This shows that the probability distribution might be different from $N(0, r)$, unless $t = 0$.

It has been proved in statistical neurodynamics of random nets with independently and identically distributed random variables w_{ij} that such long-term correlations can be neglected in a weak sense (Amari et al., 1977). However, this is not the case in the present associative memory, although the direction cosine $a(\mathrm{x})$ or the distance $d(\mathrm{x})$ satisfies the macroscopic state condition (Amari, 1974).

As we see in Figure 1, starting with $a_0 < h(r)$, $a_1 = a_0'$ may become larger than $h(r)$, but it eventually decreases. However, starting with an initial direction cosine equal to this a_0' which is larger than $h(r)$, the direction cosine converges nearly equal to 1 in almost all cases. This shows clearly that a_t is not a macroscopic state variable (cf. Amari, 1974; Amari et al., 1977), because a_{t+1} is not determined from a_t but determined depending on the past history. We explain this in Figure 4: Let S be the set of states whose direction cosine to s^1 is equal to a. Let $a' = F(a/\sqrt{r})$, $a'' = F(a'/\sqrt{r})$, and let S' and S'' be the sets of states whose direction cosines are a' and a'', respectively. Almost all states in S enter

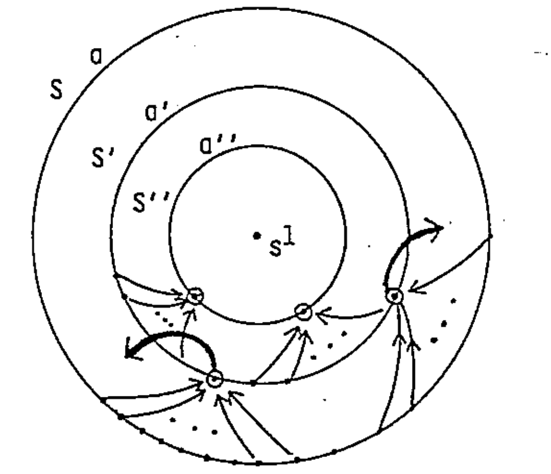

FIGURE 4. Transition of direction cosine.

into S', and almost all states in S' enter into S'' by the state transition. However, the image $TS = \{\mathbf{x}' | \mathbf{x}' = T\mathbf{x}, \mathbf{x} \in S\}$ is concentrated on a negligibly small special part of S', and those special states in S' never enter into S'' but behave quite differently from other members of S'. This is possible only when the number $|TS|$ of those states is negligibly small compared to the number $|S'|$. In other words, $\mathbf{x}_t = T^t\mathbf{x}_0$ belongs to the set of these special states. This consideration also shows that the attractor basin $B(\mathbf{s}^1)$ of $\mathbf{s}^1$ has a strange shape as shown in Figure 5.

STATISTICAL NEURODYNAMICS

Although the simple normal assumption (5.2) does not hold, it is still plausible to assume that N_i^t is normally distributed in the weak sense. The "weak sense" implies that it holds up to some time $t = T(n)$ when n is large (cf. Amari et al., 1977), $T(n)$ tending to infinity as $n \rightarrow \infty$. We assume that

$$N_i^t = \frac{1}{n} \sum_\beta' \sum_j s_i^\beta s_j^\beta x_j^t \qquad (6.1)$$

is normally distributed with mean 0 and variance σ_t^2. We show later that the mean b_t of N_i^t is not equal to 0. However, here we assume that $w_{ii} = 0$, $b_t = 0$ and obtain a simplified equation for the macroscopic dynamical behavior. Since b_t is small, this gives a good approximation.

The direction cosine a_{t+1} is given by

$$a_{t+1} = F(a_t/\sigma_t).$$

We calculate σ_{t+1} as a function of a_t and σ_t in Appendix B, by taking the correlations of w_{ij} and $\mathbf{x}_t$ into account,

$$\sigma_{t+1} = G(a_t, \sigma_t).$$

This shows that a vector (a_t, σ_t) forms a macroscopic state variable governing the recalling processes of the associative memory network.

Theorem 2. The simplified macroscopic state equation is given by

$$a_{t+1} = F(a_t/\sigma_t), \qquad (6.2)$$

$$\sigma_{t+1}^2 = r + 4[p(\bar{a}_t)]^2 + 4r\bar{a}_t p(\bar{a}_t) a_{t+1}, \qquad (6.3)$$

where

$$\bar{a}_t = a_t/\sigma_t, \qquad (6.4)$$

$$p(u) = (1/\sqrt{2\pi})\exp\{-u^2/2\}. \qquad (6.5)$$

As stated previously, N_i has a bias term depending on x_i^{t-1}, that is, N_i is subject to $N(s_i^1 x_i^{t-1} b_t, \sigma_t^2)$ conditioned on x_i^{t-1}. This gives a more precise macroscopic equation, which has three macroscopic variables (a_t, b_t, σ_t). Its derivation is given in Appendix C.

Theorem 3. The macroscopic state equation is given by

$$a_{t+1} = \langle F(\bar{a}_t) \rangle, \qquad (6.6)$$

$$b_{t+1} = (2r/\sigma_t)\langle p(\bar{a}_t) \rangle, \qquad (6.7)$$

$$\sigma_{t+1}^2 = r + 4\langle p(\bar{a}_t) \rangle^2 + 4r\bar{a}_t \langle p(\bar{a}_t) \rangle a_{t+1}, \qquad (6.8)$$

where the operator $\langle \ \rangle$ implies

$$\langle f(\bar{a}_t) \rangle = (1 + a_{t-1}) f[\bar{a}_t + (b_t/\sigma_t)]/2 + (1 - a_{t-1}) f[\bar{a}_t - (b_t/\sigma_t)]/2 \qquad (6.9)$$

BEHAVIORS OF MACROSCOPIC STATE EQUATIONS

The dynamical behaviors of the simplified macroscopic state equations (6.2), (6.3) are shown in Figure 6, where $r = 0.08$. The critical direction cosine is $h(r) = 0.16$ in this case. When the initial activity is above this value, the direction cosine a_t increases monotonically. When the initial activity is below $h(r)$, a_t still increases in the beginning but decays later. The behaviors are at least qualitatively in good agreement with the simulated results (Figure 1). One big difference is that the process stops within a finite number of time steps in simulation because the state is trapped in an equilibrium, while the theoretical curves run without time limit. This is because the number of the neurons is assumed to be infinitely large in theory. When n is finite, any state falls in an equilibrium within the

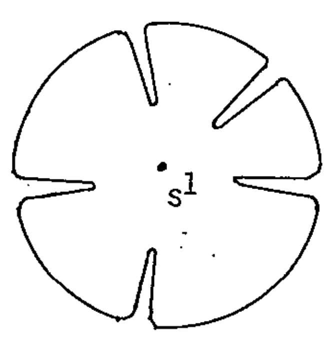

FIGURE 5. Basin of attractor.

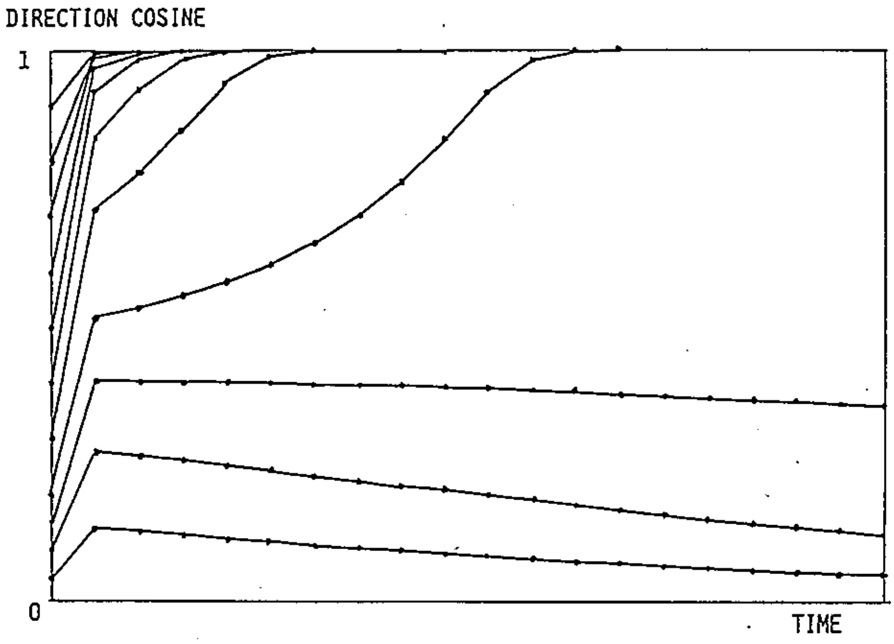

FIGURE 6. Dynamic behaviors of recalling processes: Theory with $r = 0.08$.

average transient time $T(n)$, which tends to infinity as $n \to \infty$. If we cut off the theoretical curves at a finite number of time steps, Figure 6 and Figure 1 seem to resemble each other.

Figure 7 shows the behaviors of the theoretical macroscopic equations with $r = 0.2$, which is larger than the relative capacity. The curves are similar to those in Figure 2, if we cut them off at a finite number of transient times, say $t = 10$.

Figure 3 shows the equilibrium states $\bar{a}(r)$ and the threshold $h(r)$ versus the pattern ratio r of the theoretical equations. There are two branches, the upper one of which is stable and is very close to 1. This shows that the recalled pattern is very close to the memorized one. The lower branch is unstable, showing the average diameter of the basin of attractor. It rapidly shrinks as r becomes large, and the two branches disappear at the bifurcation point, $r \doteqdot 0.16$. When r is larger than this value, there are no macroscopic equilibrium states except for $a = 0$. This point corresponds to the relative capacity of the associative memory.

The behaviors of the detailed equations (6.6)–(6.8) are also qualitatively similar, although they do not fit well quantitatively.

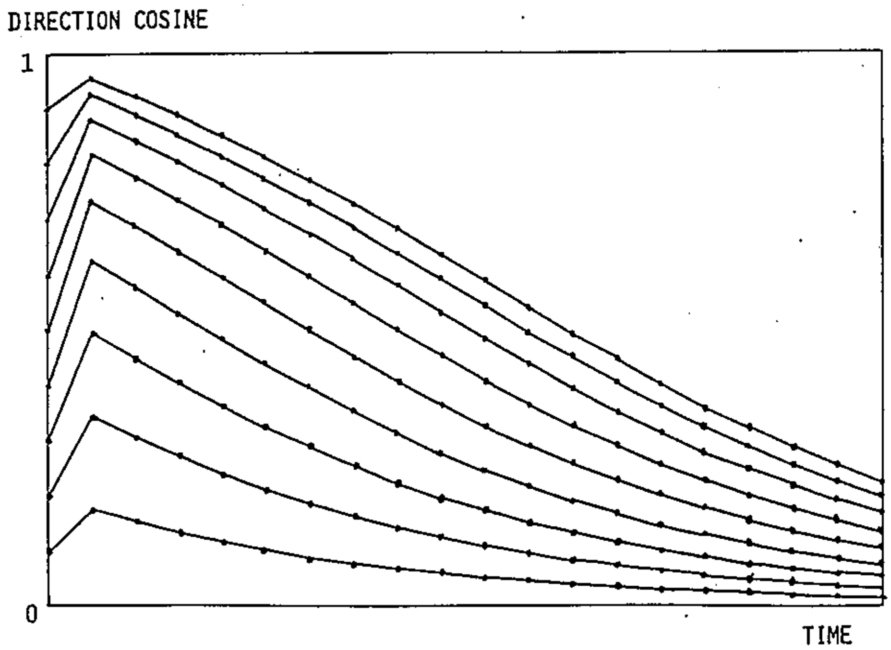

FIGURE 7. Dynamic behaviors of recalling processes: Theory with $r = 0.2$.

CONCLUDING REMARKS

We have proposed a new statistical neurodynamical method of studying the autocorrelation type associative network model. The theory can explain various interesting dynamical phenomena observed by computer simulation. The method can be applied to the analysis of a net of stochastic behavior, a net of non-symmetric connections, a net of bilateral connections, and so forth.

There remain a number of interesting theoretical problems to be studied further. They are, for example, the dependencies of the transient length or recalling time $T(n)$ on the number n of neurons, and the average number and the distribution of spurious equilibrium states. We need to develop the present method and combine it to the analysis given in Amari (1974).

REFERENCES

Amari, S. (1967). A theory of adaptive pattern classifiers. *Institute of Electrical and Electronics Engineers, Transactions Elec. Com.*, EC-16, 279–307.

Amari, S. (1971). Characteristics of randomly connected threshold element networks and network systems. *Proceedings of the Institute of Electrical and Electronics Engineers*, 59, 35–47.

Amari, S. (1972a). Learning patterns and pattern sequences by self-organizing nets of threshold elements. *Institute of Electrical and Electronics Engineers Transactions*, C-21, 1197–1206.

Amari, S. (1972b). Characteristics of random nets of analog neuron-like elements. *Institute of Electrical and Electronics Engineers Transactions Sys., Man and Cybernetics*, SMC-2, 643–657.

Amari, S. (1974). A method of statistical neurodynamics. *Kybernetik*, 14, 201–215.

Amari, S. (1977a). Neural theory of association and concept-formation. *Biological Cybernetics*, 26, 175–185.

Amari, S. (1977b). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27, 77–87.

Amari, S. (1980). Topological organization of nerve fields. *Bulletin of Mathematical Biology*, 42, 339–364.

Amari, S. (1983). Field theory of self-organizing neural nets. *Institute of Electrical and Electronics Engineers Transactions*, SMC-13, 741–748.

Amari, S., & Takeuchi, A. (1978). Mathematical theory on formation of category detecting nerve cells. *Biological Cybernetics*, 29, 127–136.

Amari, S., Yoshida, K., Kanatani, K. (1977). A mathematical foundation for statistical neurodynamics. *SIAM Journal of Applied Mathematics*, 33, 95–126.

Anderson, J. A. (1972). A simple neural network generating interactive memory. *Mathematical Biosciences*, 14, 197–220.

Amit, D. J., Gutfreund, H., & Sompolinsky, H. (1985a). Spin-glass models of neural networks. *Physical Reviews*, A2, 1007–1018.

Amit, D. J., Gutfreund, H., & Sompolinsky, H. (1985b). Storing infinite numbers of patterns in a spin-glass model of neural networks. *Physical Review Letters*, 55, 1530–1533.

Harth, E. M., Csermely, T. J., Beek, B., & Lindsay, R. D. (1970). Brain functions and neural dynamics. *Journal of Theoretical Biology*, 26, 93–120.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79, 2445–2558.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 81, 3088–3092.

Kac, M. (1959). *Probability and related topics in physical sciences*. New York: Wiley Interscience.

Kinzel, W. (1985). Learning and pattern recognition in spin glass models. *Z. Physik*, B60, 205–213.

Kishimoto, K., & Amari, S. (1979). Existence and stability of local excitations in homogeneous neural fields. *Journal of Mathematical Biology*, 7, 303–318.

Kohonen, T. (1972). Correlation matrix memories. *Institution of Electrical and Electronics Engineers Transactions*, C-21, 353–359.

Kohonen, T., & Ruohonen, M. (1973). Representation of associated data by matrix operators. *Institute of Electrical and Electronics Engineers Trans. Comput.*, 22, 701–702.

Kohonen, T. (1977). *Associative Memory*. New York: Springer.

Little, W. A. (1974). The existence of persistent states in the brain. *Math. Biosci.*, 19, 101–120.

Little, W. A., & Shaw, G. L. (1978). Analytic study of the memory storage capacity of a neural network. *Mathematical Biosciences*, 39, 281–290.

McEliece, R. J., Posner, E. C., Rodemich, E. R., & Venkatesh, S. S. (1987). The capacity of the Hopfield associative memory. *Institute of Electrical and Electronics Engineers Trans. Inf. Theory*, IT-33, 461–482.

Meir, R., & Domany, E. (1987). Exact solution of a layered neural network memory. *Physical Review Letters*, 59, 359–362.

Nakano, K. (1972). Association—a model of associative memory. *Institute of Electrical and Electronics Engineers Transactions*, SMC-2, 381–388.

Parisi, G. (1986). Asymmetric neural networks and the process of learning. *Journal of Physics*, A19, L675–L680.

Peretto, P. (1984). Collective properties of neural networks: A statistical physics approach. *Biological Cybernetics*, 50, 51–62.

Rolls, E. T. (1987). Information representation, processing, and storage in the brain: Analysis at the single neuron level. In J. P. Changeaux & M. Konishi (Eds.), *The neural and molecular basis of learning*, (pp. 503–540). New York: Wiley.

Rozonoer, L. I. (1969). Random logical nets: I. *Automatika Telemekhanika*, 5, 137–147.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing* (Vol. 1). Cambridge: MIT Press.

Shinomoto, S. (1987). A cognitive and associative memory. *Biological Cybernetics*, 57, 197–206.

Takeuchi, A., & Amari, S. (1979). Formation of topographic maps and columner microstructures. *Biological Cybernetics*, 35, 63–72.

Toulouse, G., Dehane, S., & Changeux, J.-P. (1986). Spin glass model of learning by selection. *Proceedings of the National Academy of Sciences of the United States of America*, 83, 1695–1698.

Uesaka, Y., & Ozeki, K. (1972). Some properties of associative type memories. *Journal of Institute of Electrical and Communication Engineers of Japan*, 55-D, 323–220.

Weisbuch, G. (1985). Scaling laws for the attractors of Hopfield networks. *Journal de Physique Letters*, 46, 623–630.

Willson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12, 1–24.

APPENDIX A: ABSOLUTE CAPACITY

From (3.3), we have

$$\log p_n(m) = n \log\left\{1 - \mathrm{erf}\left(\sqrt{\frac{n}{m}}\right)\right\}.$$

By putting

$$\frac{n}{m} = r(n),$$

we see that $\lim P_n(m) = 1$ requires $\lim r(n) = \infty$. When r is large, we have

$$\mathrm{erf}(\sqrt{r}) \doteqdot \frac{1}{\sqrt{r}} \mathrm{erf}\left\{-\frac{r}{2}\right\}.$$

This yields

$$\log P_n(m) \doteqdot -\frac{n}{\sqrt{r}} \exp\left\{-\frac{r}{2}\right\} = \exp\left\{-\frac{r}{2} + \log n - \frac{1}{2}\log r\right\}.$$

In order that $\lim P_n(m) = 1$ holds, $r(n)$ cannot be larger than

$$2 \log n = \log\log n + \text{smaller order terms.}$$

APPENDIX B: EVALUATION OF THE PROBABILITY DISTRIBUTION OF N_i

Without loss of generality, we assume that

$$\mathbf{s}^1 = (1, 1, \ldots, 1)'.$$

Then, the direction cosine $a(\mathbf{x})$ of $\mathbf{x}$ and $\mathbf{s}^1$ is given by $a(\mathbf{x}) = (1/n)\Sigma x_i$. We also assume that $w_{ii} = 0$ for all i. This makes the following calculations easier, although we can calculate σ_t^2 without this assumption in almost the same manner. The noise term is written as

$$N_i^t = \frac{1}{n} \sum' \sum s_i^\alpha s_j^\alpha x_j^t = \frac{1}{n} \sum_\alpha' \sum_j z_{ij\alpha}^t, \tag{A1}$$

where we put

$$z_{ij\alpha}^t = s_i^\alpha s_j^\alpha x_j^t, \quad \alpha \neq 1, \quad j \neq i. \tag{A2}$$

Since $E[N_i^t] = 0$ is assumed, the variance of N_i^t is given by

$$\sigma_t^2 = E[(N_i^t)^2] = (1/n^2) \sum_{j,j',\alpha,\alpha'} E[z_{ij\alpha}^t z_{ij'\alpha'}^t].$$

This summation consists of the four types of terms classified by their indices:

1. terms with $j = j'$, $\alpha = \alpha'$: There are $n(m-1)$ such terms, and we put

$$v_1 = E[(z_{ij\alpha}^t)^2].$$

2. terms with $j = j'$, $\alpha \neq \alpha'$: There are $n(m-1)(m-2)$ such terms,

$$v_2 = E[z_{ij\alpha}^t z_{ij\alpha'}^t], \quad \alpha \neq \alpha'.$$

3. terms with $j \neq j'$, $\alpha = \alpha'$: there are $n(n-1)(m-2)$ such terms

$$v_3 = E[z_{ij\alpha}^t z_{ij'\alpha}^t], \quad j \neq j'.$$

4. terms with $j \neq j'$, $\alpha \neq \alpha'$: There are $n(n-1)(m-1)(m-2)$ such terms

$$v_4 = E[z_{ij\alpha}^t z_{ij'\alpha'}^t], \quad j \neq j', \quad \alpha \neq \alpha'.$$

When n is large, we have

$$\sigma_t^2 = rv_1 + r^2 n v_2 + rnv_3 + r^2 n^2 v_4, \tag{A3}$$

neglecting higher order terms, where we will show soon that $v_1 = 1$, $v_2 = 0$, v_3 is of order n^{-1} and v_4 is of order n^{-2}.

Since $z_{ij\alpha}^t = \pm 1$, we have

$$v_1 = 1.$$

Similarly, we have

$$v_2 = E[s_i^\alpha s_j^\alpha s_j^{\alpha'} (x_j^t)^2] = 0.$$

In order to calculate v_3, we single out the terms in x_j^t which include s_j^α and $s_{j'}^\alpha$, explicitly, as

$$x_j^t = \mathrm{sgn}(a_{t-1} + Q + s_j^\alpha R + (1/n) s_j^\alpha s_{j'}^\alpha x_{j'}^{t-1}),$$
$$x_{j'}^t = \mathrm{sgn}(a_{t-1} + Q' + s_{j'}^\alpha R + (1/n) s_{j'}^\alpha s_j^\alpha x_j^{t-1}),$$

where

$$Q = \frac{1}{n} \sum_\beta'' \sum_k' s_j^\beta s_k^\beta x_k^{t-1},$$

$$Q' = \frac{1}{n} \sum_\beta'' \sum_k' s_{j'}^\beta s_k^\beta x_k^{t-1},$$

$$R = \frac{1}{n} \sum_k' s_k^\alpha x_k^{t-1},$$

Σ_β'' denoting the summation over $\beta \neq 1, \alpha$, and Σ_k' denoting the summation over $k \neq i, j$. Then, we can assume that Q, Q', and R are mutually asymptotically independent normal random variables with mean 0. The variances of Q and Q' are σ_{t-1}^2, and the variance of R is σ_{t-1}^2/m. We calculate v_3 by taking the expectation of the conditional expectation

$$v_3 = E\, E[s_i^\alpha s_j^\alpha s_i^\alpha s_{j'}^\alpha x_j^t x_{j'}^t | s_j^\alpha, s_{j'}^\alpha],$$

where $E[\cdot | s_j^\alpha, s_{j'}^\alpha]$ is the conditional expectation conditioned on four possible values of $(s_j^\alpha, s_{j'}^\alpha)$. Here, we evaluate the direct correlations of s_j^α, $s_{j'}^\alpha$ and $\mathbf{x}^t$ in this manner, by neglecting the correlations with $\mathbf{x}^{t-1}$. However, these indirect correlations are not simply neglected, but are taken into account through the distribution of Q, Q' and R. In other words, the higher order correlations are also included through the renormalization of the distribution of N_i^t.

Let us denote the conditional expectation by

$$Y_{pq} = E[z_{ij\alpha}^t z_{ij'\alpha}^t | s_j^\alpha = p, s_{j'}^\alpha = q], \tag{A4}$$

where $p, q = \pm 1$. Then

$$v_3 = \sum_{p,q} \mathrm{Prob}\{s_j^\alpha = p, s_{j'}^\alpha = q\} Y_{pq}$$
$$= \frac{1}{4}(Y_{11} + Y_{-11} + Y_{1-1} + Y_{-1-1}),$$

because $\mathrm{Prob}\{s_j^\alpha = p, s_{j'}^\alpha = q\} = \frac{1}{4}$ for all four combinations of p and q. We fix x_j^{t-1} and $x_{j'}^{t-1}$ and calculate Y_{11} and Y_{1-1} conditioned on these variables. Then, we have

$$Y_{11} = E[x_j^t x_{j'}^t | s_j^\alpha = 1, s_{j'}^\alpha = 1]$$
$$= E[\mathrm{sgn}(a_{t-1} + Q + R + n^{-1} x_{j'}^{t-1})(a_{t-1} + Q' + R + n^{-1} x_j^{t-1})],$$
$$Y_{1-1} = E[\mathrm{sgn}(a_{t-1} + Q - R - n^{-1} x_{j'}^{t-1})(a_{t-1} + Q' - R - n^{-1} x_j^{t-1})].$$

In order to calculate $Y_{p,q}$, we express the normal random variables Q, Q', and R by using unit uncorrelated normal random variables u and v,

$$Q + R = \sigma_{t-1}(u + (2m)^{-1} v), \; Q' + R = \sigma_{t-1}(v + (2m)^{-1} u)$$

in the case of Y_{11}, and

$$Q + R = \sigma_{t-1}(u - (2m)^{-1} v), \; Q' - R = \sigma_{t-1}(v - (2m)^{-1} u)$$

in the case of Y_{1-1}. Then, we have

$$Y_{11} = E[\mathrm{sgn}(\bar{a} + u + \tilde{v} + \tilde{x}')(\bar{a} + v + \tilde{u} + \tilde{x})]$$

etc., where

$$\bar{a} = a_{t-1}/\sigma_{t-1}, \quad \tilde{x}' = n^{-1} x_{j'}^{t-1}/\sigma_{t-1}, \quad \tilde{x} = n^{-1} x_j^{t-1}/\sigma_{t-1},$$

$$\tilde{u} = (2m)^{-1} u, \quad \tilde{v} = (2m)^{-1} v.$$

Moreover, from

$$E[\mathrm{sgn}\, X] = 2\,\mathrm{Prob}\{X > 0\} - 1,$$

we have

$$Y_{11} = 2\,\mathrm{Prob}\{(\bar{a} + u + \tilde{v} + \tilde{x}')(\bar{a} + v + \tilde{u} + \tilde{x}) > 0\} - 1,$$

which can be calculated by integrating the two-dimensional normal density

$$p(u, v) = (2\pi)^{-1} \exp\left\{-\frac{1}{2}(u^2 + v^2)\right\}$$

over the region on which the term in the bracket is positive. Since $\tilde{u}$, $\tilde{v}$, $\tilde{x}$ and $\tilde{x}'$ are small order terms, the sum of Y_{11} and Y_{1-1}

$$Y_{11} + Y_{1-1} = 2\ \mathrm{Prob}\{(\bar{a} + u + \tilde{v} + \tilde{x}')(\bar{a} + v + \tilde{u} + \tilde{x}) > 0\}$$
$$- 2\ \mathrm{Prob}\{(\bar{a} + u + \tilde{v} - \tilde{x}')(a + v - \tilde{u} - \tilde{x}) > 0\}$$

can easily be calculated, because the first order terms of Y_{11} and Y_{1-1}, given by $\pm\mathrm{Prob}\{(\bar{a} + u)(\bar{a} + v) > 0\}$, are cancelled out. Therefore, we calculate the small order term, which is decomposed as

$$\begin{aligned} Y_{11} + Y_{1-1} = {} & 2\,[\mathrm{Prob}\{(\bar{a} + u + \tilde{v} + \tilde{x}')(\bar{a} + v) > 0\} \\ & - \mathrm{Prob}\{(\bar{a} + u - \tilde{v} - \tilde{x}')(\bar{a} + v) > 0\}] \\ & + 2\,[\mathrm{Prob}\{(\bar{a} + u)(\bar{a} + v + \tilde{u} + \tilde{x}) > 0\} \\ & - \mathrm{Prob}\{(\bar{a} + u)(\bar{a} + v - \tilde{u} - \tilde{x}) > 0\}]. \qquad (A5) \end{aligned}$$

Let A and A' be the regions given by

$$A = \{(u, v) | (\bar{a} + u + \tilde{v} + \tilde{x}')(a + v) > 0\},$$
$$A' = \{(u, v) | (\bar{a} + u - \tilde{v} - \tilde{x}')(a + v) > 0\}$$

in the (u, v)-plane, respectively. Then, the former half of equation (A5) is written as

$$\int_A p(u, v)dudv - \int_{A'} p(u, v)dudv = \int_{A-A'} p(u, v)dudv.$$

Since the signed incremental region $A = A - A'$ is bounded by the two lines in the (u, v)-plane

$$u = -\frac{1}{2m} v - \bar{a} - \tilde{x} = f(v),$$
$$u = \frac{1}{2m} v - \bar{a} + \tilde{x} = g(v),$$

and the sign of the region changes at $v = -\bar{a}$, the integral is written as

$$\int_{-\infty}^{-\bar{a}} dv p(v) \int_{g(v)}^{f(v)} p(u)du + \int_{-\bar{a}}^{\infty} dv p(v) \int_{f(v)}^{g(v)} p(u)du,$$

where $p(u)$ and $p(v)$ are the unit normal densities. Since

$$f(v) - g(v) = -\frac{1}{m} v - \frac{2}{n\sigma_{t-1}} x_{j'}$$

is a first order small term, the above integral is written as

$$\begin{aligned} p(\bar{a}) & \left[\int_{-\infty}^{-\bar{a}} (f(v) - g(v))p(v)dv + \int_{-\bar{a}}^{\infty} (g(v) - f(v))p(v)dv\right] \\ & = n^{-1}p(\bar{a})\left[\int_{-\infty}^{-\bar{a}} (-r^{-1}v - 2\sigma_{t-1}^{-1}x_{j'})p(v)dv \right. \\ & \qquad \left. + \int_{-\bar{a}}^{\infty} (r^{-1} + 2\sigma_{t-1}^{-1}x_{j'})p(v)dv\right] \\ & = n^{-1}p(\bar{a})\{2r^{-1}p(\bar{a}) + 2\sigma_{t-1}^{-1}x_{j'}F(\bar{a})\}. \end{aligned}$$

The latter half of the integral is obtained in a similar manner, giving

$$Y_{11} + Y_{1-1} = 4n^{-1}[2r^{-1}\{p(\bar{a})\}^2 + \sigma_{t-1}^{-1}(x_j + x_{j'})p(\bar{a})F(\bar{a})].$$

The term $Y_{-1-1} + Y_{-11}$ is the same as $Y_{11} + Y_{1-1}$. Among n x_j^{t-1}, nd_{t-1} are -1 and $n(1 - d_{t-1})$ are 1. Hence, averaging over x_j and $x_{j'}$, we have

$$nv_3 = 4r^{-1}\{p(\bar{a}_{t-1})\}^2 + 4\bar{a}_{t-1}p(\bar{a}_{t-1})F(\bar{a}_{t-1}),$$

where $\bar{a}_{t-1} = a_{t-1}/\sigma_{t-1}$.

The term v_4 can be calculated in a similar manner. By singling out the directly correlated terms in x_j^{t-1} and $x_{j'}^{t-1}$ as

$$z_{ij\alpha}^t = s_i^\alpha s_j^\alpha \mathrm{sgn}(a_{t-1} + P + n^{-1}s_i^\alpha s_j^\alpha x_i^{t-1}),$$
$$z_{ij'\alpha'}^t = s_i^\alpha s_j^\alpha \mathrm{sgn}(a_{t-1} + P' + n^{-1}s_i^{\alpha'} s_j^{\alpha'} x_i^{t-1}),$$

we see that P and P' are asymptotically independent subject to $N(0, \sigma_{t-1}^2)$. Then, the expectation of $z_{ij\alpha}^t z_{ij'\alpha'}^t$ is written as

$$\begin{aligned} E[z_{ij\alpha}^t z_{ij'\alpha'}^t] = \tfrac{1}{4}\{ & E[\mathrm{sgn}(\bar{a} + u + \tilde{x})(\bar{a} + v + \tilde{x})] \\ & + E[\mathrm{sgn}(a + u - \tilde{x})(\bar{a} + v - \tilde{x})] \\ & - E[\mathrm{sgn}(\bar{a} + u - \tilde{x})(\bar{a} + v + \tilde{x})] \\ & - E[\mathrm{sgn}(\bar{a} + u + \tilde{x})(\bar{a} + v - \tilde{x})]\} \end{aligned}$$

where

$$\tilde{x} = (n\sigma_{t-1})^{-1}x_i^{t-1}$$

is fixed, and u and v are unit independent normal random variables. This gives

$$nr^2v_4 = 4(\sigma_{t-1})^{-2}\{p(\bar{a}_{t-1})\}^2.$$

However, we can prove in Appendix C that $E[N_i^t] \neq 0$. Therefore, we need to subtract the term $\{E[N_i^t]\}^2$ from σ_t^2. The term $\{E[N_i^t]\}^2$ is equal to nr^2v_4. Hence, by summarizing these terms, we have

$$\sigma_{t+1}^2 = r + 4p(\bar{a}_t)^2 + 4r\bar{a}_t p(\bar{a}_t)F(\bar{a}_t).$$

APPENDIX C: MACROSCOPIC EQUATION INCLUDING BIAS TERM

The noise term N_i^t was assumed to be subject to $N(0, \sigma_t^2)$ in the previous simplified equation. However, a careful study shows the expectation of N_i^t is not 0. It depends on x_i^{t-1} (or more precisely $s_i^1 x_i^{t-1}$), so that we assume that N_i^t is subject to $N(x_i^{t-1}b_t, \sigma_t^2)$ conditionally on x_i^{t-1}. We then derive the dynamical equations to determine (a_t, b_t, σ_t) which are the macroscopic state variables.

Let

$$\bar{a}_{t+} = (a_t + b_t)/\sigma_t, \quad \bar{a}_{t-} = (a_t - b_t)/\sigma_t.$$

For those neurons for which $x_i^{t-1} = +1$, its activity a (i.e., the direction cosine to s^1) at time $t + 1$ is given by $F(\bar{a}_{t+})$, while the activity is given by $F(\bar{a}_{t-})$ at time $t + 1$ for those for which $x_i^{t-1} = -1$. Therefore, the total activity a_{t+1} is given by

$$a_{t+1} = \langle F(\bar{a}_t)\rangle,$$

where $\langle\ \rangle$ implies the following averaging operation,

$$\langle f(\bar{a}_t)\rangle = [(1 + a_{t-1})f(\bar{a}_{t+}) + (1 - a_{t-1})f(\bar{a}_{t-})]/2.$$

In order to calculate b_t, we decompose $z_{ij\alpha}^t$ as

$$z_{ij\alpha}^t = s_i^\alpha s_j^\alpha \mathrm{sgn}(a_{t-1} + Q_j^{t-1} + (1/n)s_i^\alpha s_j^\alpha x_i^{t-1})$$

by singling out the term including s_i^α, s_j^α from N_i^{t-1}. By taking the conditional expectation of $z_{ij\alpha}^t$ conditioned on x_i^{t-1}, s_i^α, s_j^α, we have the following relation

$$b_{t+1} = (2r/\sigma_t)\langle p(\bar{a}_t)\rangle.$$

Calculations of σ_t^2,

$$\sigma_t^2 = E[(N_i^t)^2] - E[N_i^t]^2$$

proceed almost similarly as in Appendix B, giving

$$\sigma_{t+1}^2 = r + 4\langle p(\bar{a}_t)\rangle^2 + 4(ra_t/\sigma_t)\langle p(\bar{a}_t)\rangle a_{t+1}.$$