Neural Networks, Vol. 6, pp. 115–126, 1993
Printed in the USA.

0893-6080/93 $6.00 + .00

ORIGINAL CONTRIBUTION

Associative Memory With Nonmonotone Dynamics

MASAHIKO MORITA

University of Tsukuba

(*Received* 6 *January* 1992; *revised and accepted* 25 *June* 1992)

Abstract—*The dynamics of autocorrelation associative memory is examined, and a novel neural dynamics which greatly enhances the ability of associative neural networks is presented. This dynamics is such that the output of some particular neurons is reversed (for a discrete model) or the output function is not sigmoid but nonmonotonic (for an analog model). It is also shown by numerical experiments that most of the problems of the conventional model are overcome by the improved dynamics. These results are important not only for practical purposes but also for understanding dynamical properties of associative neural networks.*

Keywords—Dynamics of associative memory, Associative neural networks, Autocorrelation associative memory, Recalling process, Memory capacity, Spurious memory, Partial reverse method, Nonmonotone dynamics, Memory of correlated patterns.

1. INTRODUCTION

The autocorrelation type of associative memory (autocorrelation model) is one of the most fundamental models of neural networks. It was proposed about 20 years ago (Anderson, 1972; Kohonen, 1972; Nakano, 1972) and has been studied by many researchers since then (e.g., Amari, 1977; Amit, Gutfreund, & Sompolinsky, 1985; Gardner, 1986).

However, we do not yet understand its dynamics well because an associative neural network is generally a complex nonlinear system and is difficult to analyze mathematically. At the same time, the autocorrelation model presents major problems as an associative memory device.

1. Memory capacity is very low: The maximal number of patterns that can be exactly stored in a network with n neurons is only about $n/(2 \log n)$ (McEliece, Posner, Rodemich, & Venkatesh, 1987). Even if we permit a small amount of error, it is $0.15n$ at most.
2. Basins of attraction of stored patterns are small and so the recollection ability is not very great.
3. There are many spurious memories, namely equilibrium states different from stored patterns, whose number increases exponentially with n. Moreover, we cannot distinguish them from true memories if we only see the outputs of the neurons.
4. Stored patterns must be nearly orthogonal to each other. Otherwise, the performance of the model is greatly decreased (see Section 5).

Although many modified models have been proposed so far, not all of these problems have been settled. On the contrary, we may say that associative memory has hardly made progress in fundamental principles.

It was reported recently that the autocorrelation model shows strange dynamic behavior in recollection (Amari & Maginu, 1988), which seems to cause a decline in the capacity of the model. While examining this phenomenon, I noted that there is room for improvement in the conventional neural dynamics which had not changed essentially since McCulloch and Pitts (1943) proposed their neuron model.

This paper reports that by modifying the conventional recollection dynamics, we can improve the autocorrelation model greatly and solve the above problems. This research also aims to deepen our understanding of the dynamics of neural networks.

2. DYNAMICS OF ASSOCIATIVE MEMORY

2.1. Autocorrelation Associative Memory

First let us consider a discrete type of neural network consisting of n neurons with outputs ± 1. We represent the connection weight from the j-th neuron to the i-th by w_{ij} and the output of the i-th neuron at time t by $x_i(t)$. This network has 2^n possible states, and the cur-

The author thanks Professor K. Nakano, Professor S. Yoshizawa, and Professor S. Amari for helpful advice and discussions.

Requests for reprints should be sent to Masahiko Morita, Institute of Information Sciences and Electronics, University of Tsukuba, Tsukuba, Ibaraki 305, Japan.

rent state is described by an n-dimensional vector $X = (x_1, \ldots, x_n)^T$, where superscript T denotes the transpose. We will denote the network state at time t by X_t.

Let $S^1, S^2, \ldots, S^m$ be binary patterns (represented by column vectors) to be stored and s_i^μ ($= \pm 1$) be the i-th component of S^μ. We assume that n and m are sufficiently large and that S^μ is randomly chosen out of the 2^n possible patterns. Then these patterns are almost orthogonal to each other, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} s_i^\mu s_i^\nu \simeq 0 \ (\mu \neq \nu). \tag{1}$$

In an autocorrelation associative memory, the S^μ ($\mu = 1, \ldots, m$) are stored in the network using the weight matrix $W = [w_{ij}]$ given by

$$W = \frac{1}{n}\sum_{\mu=1}^{m} S^\mu S^{\mu T}. \tag{2}$$

We usually set $w_{ii} = 0$. Although there are various versions of the dynamics for recalling stored patterns, we deal here with the simplest one:

$$x_i(t+1) = \mathrm{sgn}\left(\sum_{j=1}^{n} w_{ij} x_j(t)\right), \tag{3}$$

where $\mathrm{sgn}(u) = 1$ when $u > 0$ and -1 when $u < 0$. That is, the network changes its state at discrete times $t = 0, 1, 2, \ldots$ according to

$$X_{t+1} = \mathrm{sgn}(WX_t), \tag{4}$$

where the function sgn operates componentwise on vectors.

During this process, the energy of the network

$$E = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} x_i x_j \tag{5}$$

decreases with time (Hopfield, 1982) so that the network state finally settles in an equilibrium corresponding to a local minimum of the energy.* Since E can be rewritten as

$$\begin{aligned} E &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{1}{n}\sum_{\mu=1}^{m} s_i^\mu s_j^\mu\right) x_i x_j \\ &= -\frac{1}{2n}\sum_{\mu=1}^{m}\left(\sum_{i=1}^{n} s_i^\mu x_i\right)^2, \end{aligned} \tag{6}$$

it is considered that E has a minimum at S^μ.

Consequently, when an initial state sufficiently close to a stored pattern is given, the network state X is expected to reach an equilibrium which is identical or very close to the stored pattern. Conversely, we can define the set of initial states from which X reaches an equilibrium $\bar{X}$; this set is called the basin of attraction of $\bar{X}$ and its size is an index of the recollection ability of the model.

* Strictly, this holds true only if the network acts asynchronously. There can be limit cycles of period 2 if the dynamics is synchronous. However, it is virtually irrelevant to the following discussions.

2.2. Recalling Process

The actual recalling process, however, is not so simple that X consistently approaches the target pattern (say, S^1), as shown by Amari and Maginu (1988). Figure 1 shows how the overlap p between X and S^1 changes with time when initial states are given with various distances from S^1. Here, the overlap p is defined by

$$p = \frac{1}{n}\sum_{i=1}^{n} s_i^1 x_i, \tag{7}$$

and $p = 1$ implies $X = S^1$.

We can see that if the initial overlap p_0 is smaller than a critical value p_c, it initially increases but soon decreases (Figure 1a). In such a case, the network state converges to an equilibrium state which is very different from the stored patterns, i.e., a spurious memory is recalled.

The decrease in the overlap is always observed if the pattern-to-neuron ratio $r = m/n$ is greater than about 0.15, as shown in Figure 1b. No matter how large the initial overlap may be, X retreats from S^1. Even in this case, X once comes rather close to S^1, which suggests that the network has not lost the information on the stored patterns.

Why do these apparently strange phenomena occur? Amari and Maginu (1988) have analyzed them using a statistical neuro-dynamical method (see also Meir & Domany, 1987), but it is complicated and not easy to understand intuitively. Here we will examine the recalling process from a different angle.

First, we introduce a quantity

$$\sigma^2 = \sum_{\nu=2}^{m}\left(\frac{1}{n}\sum_{i=1}^{n} s_i^\nu x_i\right)^2, \tag{8}$$

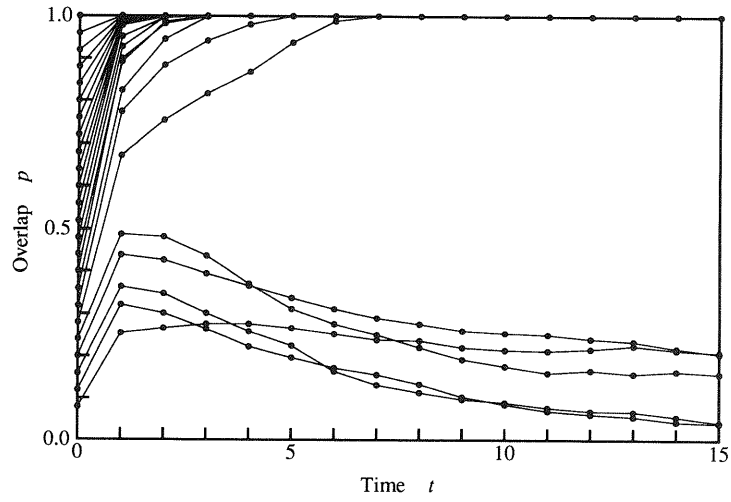
or the sum of squares of the overlaps with all the stored patterns except S^1. When X_0 is given at random, $\sigma \simeq r$ because $\sum_i s_i^\nu x_i$ ($\nu = 2, \ldots, m$) are independent subject to normal distribution with mean of 0 and variance of $1/n$. Similarly, $\sigma^2 \simeq r$ when $X = S^1$.

Figure 2 shows how this quantity σ^2 changes during the recalling process in some numerical experiments. The ordinate of the graph is p^2 and the abscissa is σ^2. Each segment of the lines represents one state transition of the network. It should be noted that the sum of p^2 and σ^2 increases with time since from eqn (6),

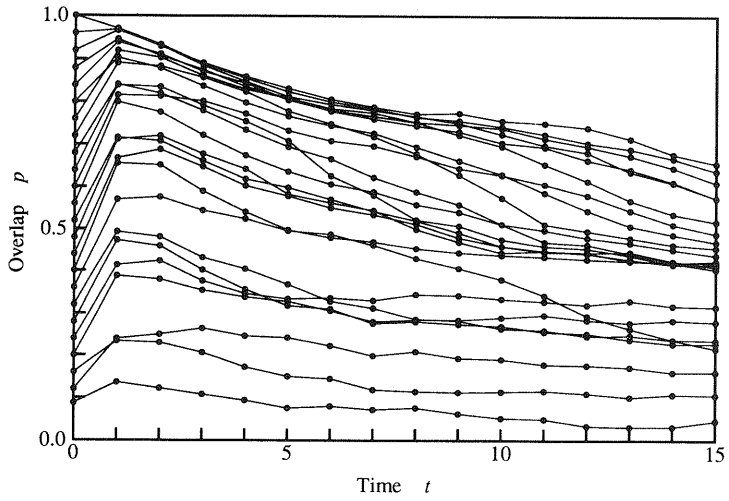
$$p^2 + \sigma^2 = -\frac{2}{n}E. \tag{9}$$

The results shown in Figure 2 indicate the following:

1. At first, both p^2 and σ^2 increase. If σ^2 does not grow too much, then it will become small again and S^1



(a)



(b)

FIGURE 1. Time courses of the overlap p between the network state X and the object pattern S^1 in the numerical experiment with $n = 1000$: (a) $r = 0.08$; (b) $r = 0.2$. Initial states are randomly chosen from among the states having a certain overlap with S^1.

is recalled; but if it once becomes too large, p^2 will decrease instead of σ^2 and recollection is unsuccessful (Figure 2a). Intuitively, this means that X does not go straight toward S^1 but goes "deviously" because it is attracted by other stored patterns, as shown schematically in Figure 3.

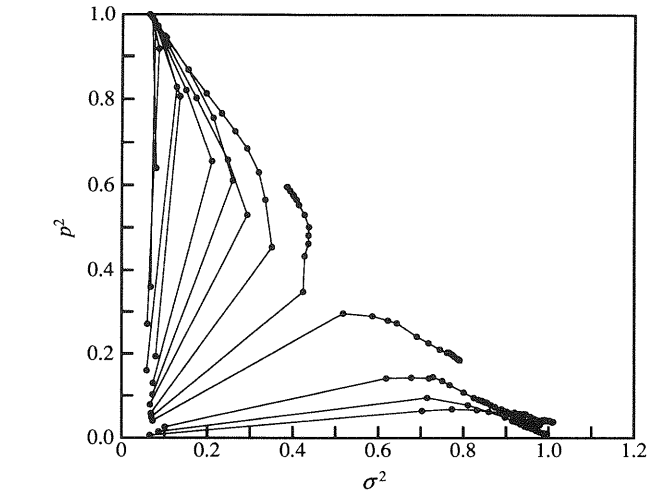
2. When $r = 0.2$ (>0.15), σ^2 always increases with time (Figure 2b). We may suppose that the stored pattern is a saddle point of the energy function (see Figure 4). The energy is lower at S^1 than at almost all the states around it, but there exist a very small number of states where the energy is still lower. Through these states, X retreats from S^1.

Of course these illustrations are not exact because the state space is actually n-dimensional, but they are useful for understanding the dynamic behavior of the network.

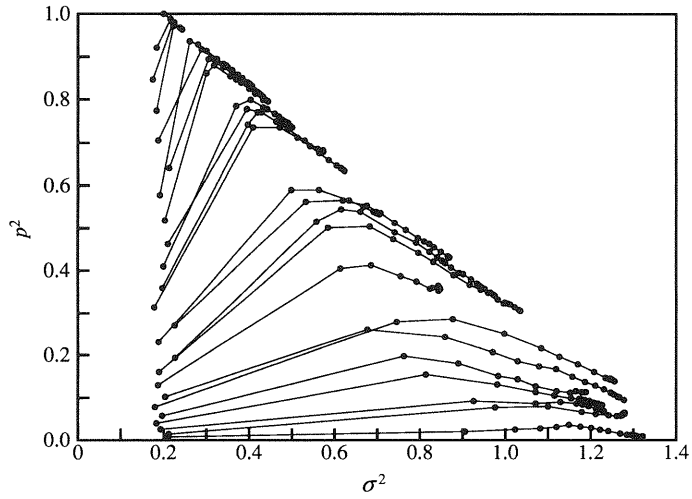
3. IMPROVING THE RECOLLECTION DYNAMICS

The preceding findings show that many of the problems of conventional associative memory are closely connected with the dynamical properties of the network. This implies that in order to solve the problems, we should improve the recollection dynamics rather than the manner of storing patterns, though the major effort has been directed to designing the weight matrix and the learning algorithm. That approach is promising since even the ordinary autocorrelation matrix given by eqn (2) contains more information than is actually used by conventional dynamics.

From this point of view, we will try to improve the autocorrelation model by modifying only the manner of recollection.



(a)



(b)

FIGURE 2. Processes of recollection in the autocorrelation model: (a) $r = 0.08$; (b) $r = 0.2$. The abscissa is the square sum of the overlaps of X with the stored patterns other than S^1. Each segment of lines represents one step of the state transition. Initial states are distributed along the line $\sigma^2 = r$, and S^1 is located near the point $(r, 1)$.

3.1. Partial Reverse Method

As a modification to neural dynamics, we often put "thermal noise" in neural networks since this makes it possible for the network state to avoid local minima of the energy. But this is not effective for associative memory models because such noise almost always decreases p and increases σ^2 when X is near S^1.

However, if we can move X not at random but in such a direction in Figure 2 that only σ^2 is reduced, we will be able to enhance the recollection ability of the network. Intuitively, this operation means making X distant from patterns S^ν ($\nu \neq 1$) while keeping the distance from S^1 constant (Figure 5). This is the basic idea of the "partial reverse method" (Morita, Yoshizawa, & Nakano, 1990a).

Since x_i takes 1 or -1, moving the network state X in a certain direction is "reversing," that is to say, changing the sign of, the outputs of certain neurons. The important point is how to determine such neurons. We must increase the energy of the network with the least possible decrease in p. I will show a concrete algorithm for realizing this, which may not be optimum, but is simple and effective.

3.2. Algorithm

One step of recollection is divided into two phases, I and II. Conventional dynamics is used in the first phase, and the outputs of some neurons are reversed in the second. These phases are iterated alternately.

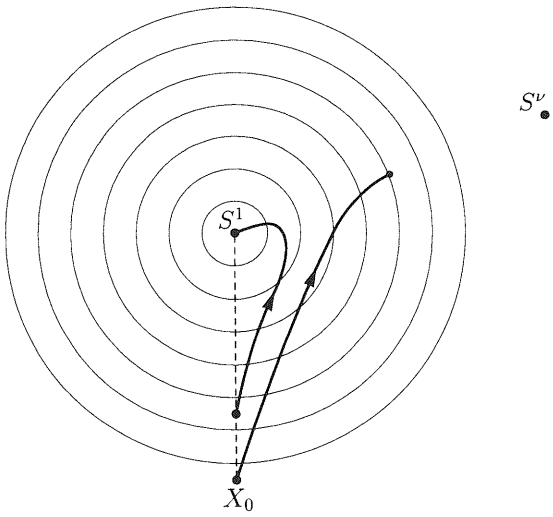


FIGURE 3. Schematic representation of the recalling process. The network state goes around to S^1 or reaches a spurious memory because it is attracted by other stored patterns S^ν ($\nu = 2, \ldots, m$).

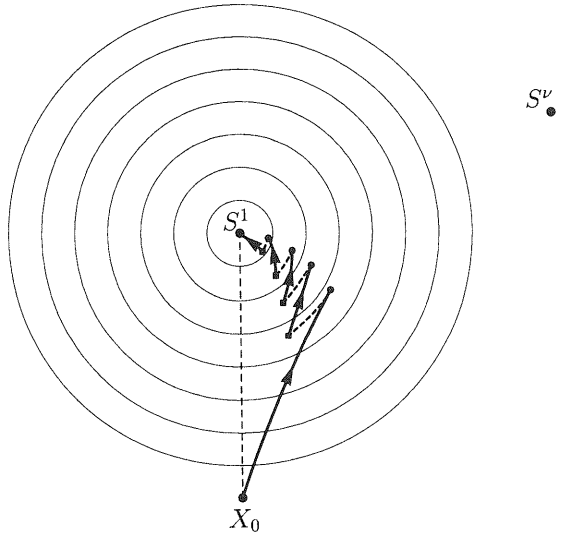


FIGURE 5. Basic idea of the partial reverse method. The network state is moved so that the distances from S^ν increase but the distance from S^1 does not.

More definitely, in phase I, the weighted sum of the input is calculated:

$$u_i = \sum_{j=1}^{n} w_{ij} x_j(t). \tag{10}$$

The output in this phase is given by

$$x_i = \mathrm{sgn}(u_i). \tag{11}$$

In phase II, only the neurons which satisfy $|u_i| > h\,(>0)$ in eqn (10) emit their outputs, and the weighted sum v_i with respect to these outputs is calculated:

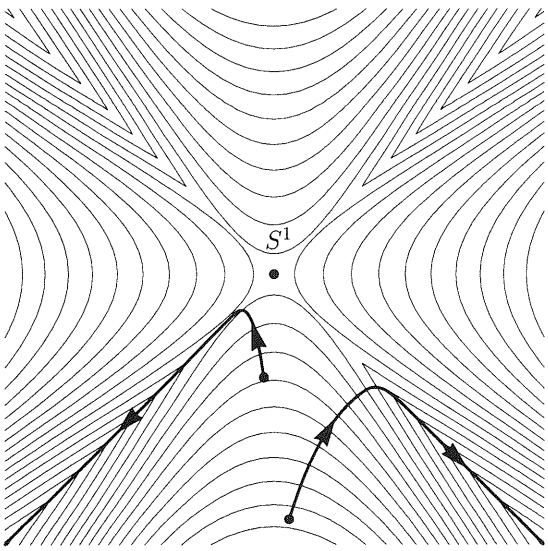


FIGURE 4. Schematic energy function around S^1 when $r > 0.15$. The gradient of E is negative toward S^1 almost everywhere except in the narrow ravines reaching S^1, through which X retreats from S^1.

$$v_i = \sum_{j=1}^{n} w_{ij} \phi(u_j), \tag{12}$$

where

$$\phi(u) = \begin{cases} -1 & (u < -h), \\ 0 & (-h \le u \le h), \\ 1 & (u > h). \end{cases} \tag{13}$$

The final output is determined by

$$x_i = \mathrm{sgn}(u_i - \lambda v_i), \tag{14}$$

λ being a positive constant. As a result, x_i obtained in phase I is reversed when

$$0 < \frac{u_i}{v_i} < \lambda \tag{15}$$

is satisfied.

More briefly, this dynamics can be expressed by

$$X_{t+1} = \mathrm{sgn}(W(X_t - \lambda \phi(W X_t))), \tag{16}$$

where the function ϕ operates componentwise on vectors. We should note that X does not always converge to an equilibrium.

Equation (16) implies that the influence of neurons with a large value of $|u_i|$ should be reduced. This is because such neurons are the cause of the increase in σ^2. Note that the energy can be rewritten as

$$E = -\frac{1}{2} \sum_{i=1}^{n} |u_i| \tag{17}$$

when the dynamics is conventional and the network is in an equilibrium state.

Here, the values of parameters h and λ must be determined carefully. Good performance is obtained in

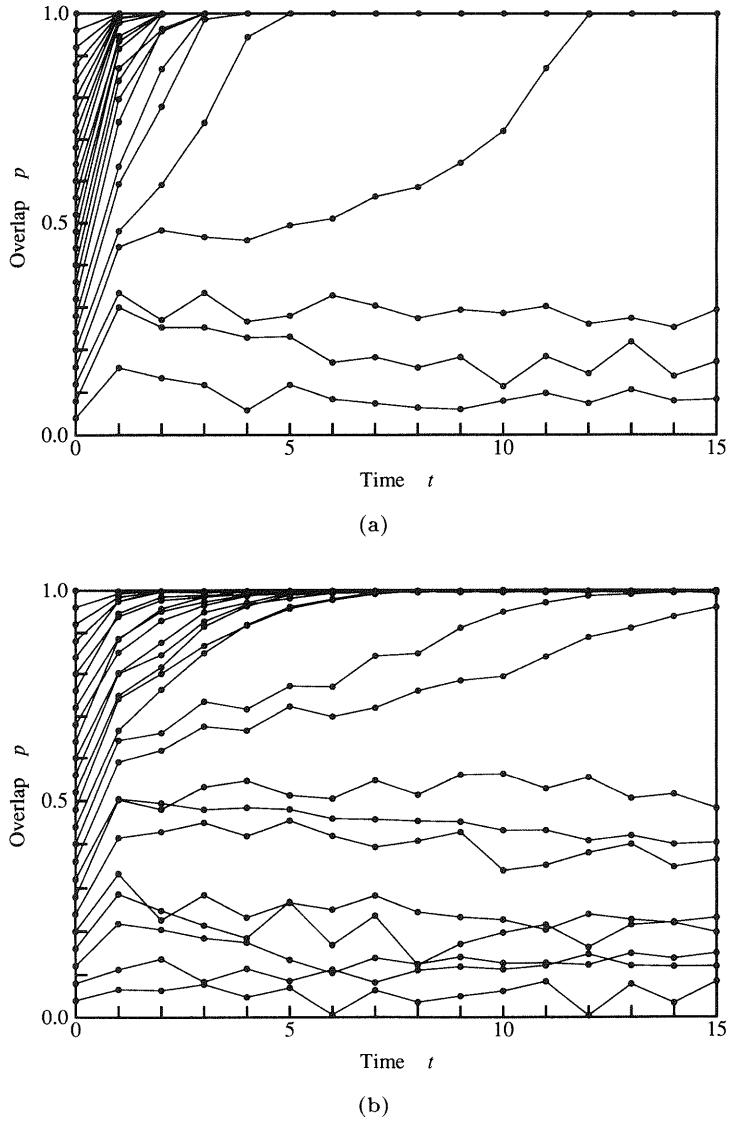


FIGURE 6. Time courses of the overlap p in the autocorrelation model with the partial reverse method: (a) $r = 0.08$; (b) $r = 0.2$. All the experimental conditions are the same as for Figure 1, except for the use of improved dynamics.

numerical experiments when $\lambda \simeq 2.7$ and h is between 1.8 and 2.0; if the pattern-to-neuron ratio r is known,

$$h = 1 + 2.0\sqrt{r} \tag{18}$$

is recommended. If values of the parameters are determined appropriately, the number of neurons which satisfy $|u_i| > h$ is small when X is close to S^μ. Reversing of the sign of x_i in phase II, therefore, seldom occurs when recollection is successful, provided r is not too large. Conversely, numerical experiments show that we may judge the recollection to be correct if no reversing occurs at an equilibrium state.

Incidently, replacing the first X_t on the right side of eqn (16) with $\mathrm{sgn}(WX_t)$ and setting $\psi(u) = \mathrm{sgn}(u) - \lambda\phi(u)$, we obtain a simpler algorithm expressed by

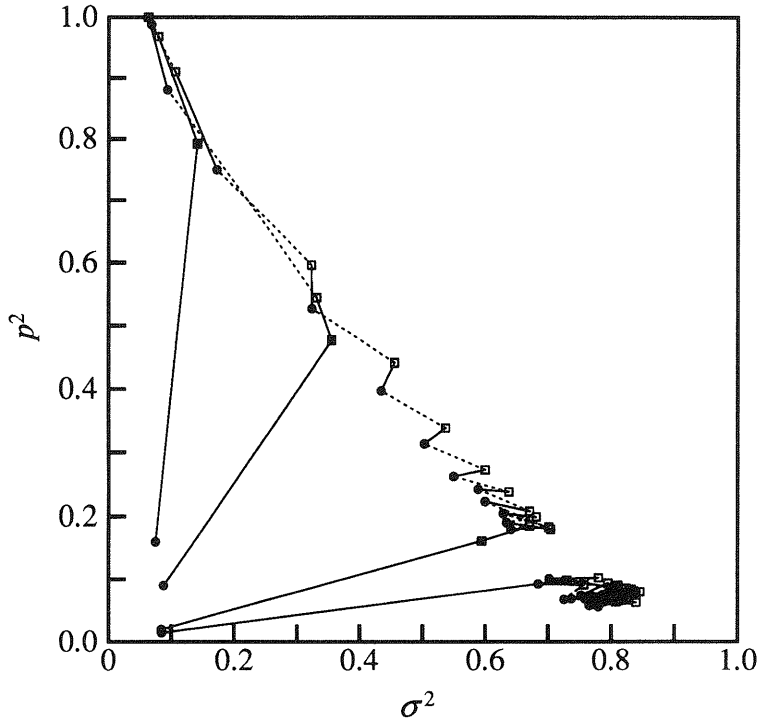
$$X_{t+1} = \mathrm{sgn}(W\psi(WX_t)), \tag{19}$$

which is an algorithm of applying two output functions $\mathrm{sgn}(u)$ and $\psi(u)$ alternately. However, this modified algorithm shows worse performance than the original, though it is better than conventional dynamics.

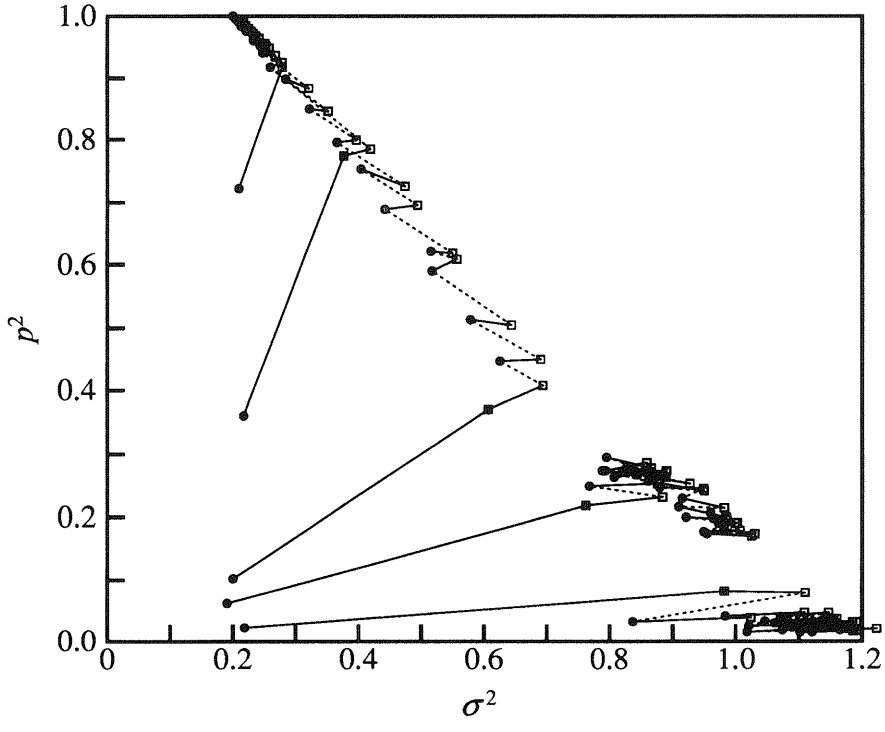
3.3. Numerical Experiments

Figure 6 shows the time courses of the overlap p when the partial reverse method is applied with parameters $\lambda = 2.7$; $h = 1.57$ for $r = 0.08$ and $h = 2.09$ for $r = 0.2$ according to eqn (18). From this figure, we can see the following:

1. The critical overlap p_c is smaller than for conventional dynamics. This means that the recollection ability of the network is raised.
2. The time (the number of steps) required for completing recollection is decreased.



(a)



(b)

FIGURE 7. Processes of recollection with the partial reverse method: (a) $r = 0.08$; (b) $r = 0.2$. The solid and broken lines represent the state transitions in the first and the second phases, respectively.

3. Correct recollection is possible even if $r > 0.15$, i.e., memory capacity increases. The upper limit of r is about 0.27, though this depends on the parameters.

Figure 7 shows the time courses of the pair (σ^2, p^2) as in Figure 2, where the solid and broken lines represent the state transitions in phases I and II, respectively. We see that X approaches S^1 in phase II rather than in phase I, contrary to the illustration in Figure 5.

4. NONMONOTONE DYNAMICS

Next, let us consider a network consisting of analog neurons with continuous-time dynamics:

$$\tau \frac{du_i}{dt} = -u_i + \sum_{j=1}^{n} w_{ij} y_j,$$
$$y_i = f(u_i). \tag{20}$$

Here, u_i denotes the instantaneous potential of the i-th neuron, y_i the output and τ a time constant; $f(u)$ is the output function, which is normally a monotonic sigmoid function (Figure 8a) given by

$$f(u) = \frac{1 - \exp[-cu]}{1 + \exp[-cu]}, \tag{21}$$

c being a positive constant. For convenience, we put $x_i = \mathrm{sgn}(u_i)$ and call $X = (x_1, \ldots, x_n)^T$ the network state as before, though the current state of the network is specified by vector $U = (u_1, \ldots, u_n)^T$ rather than X.

We cannot immediately apply the partial reverse method to this model because the network acts continuously. However, since reducing the influence of neurons for which $|u_i|$ is very large was the essence of the previous algorithm, it is expected that a similar effect is achieved by using such a nonmonotone function as shown in Figure 8b instead of the conventional sigmoid function. We may assume that this nonmonotone output function operates to keep the variance of $|u_i|$, and thus σ^2, from growing too large.

The function in Figure 8b is given by

$$f(u) = \frac{1 - \exp[-cu]}{1 + \exp[-cu]} \cdot \frac{1 + \kappa \exp[c'(|u| - h)]}{1 + \exp[c'(|u| - h)]}, \tag{22}$$

where c' and h are positive constants and κ is a parameter which is usually negative (in the following experiments, $c = 50$, $c' = 15$, $h = 0.5$ and $\kappa = -1$). It should be noted that the form of the output function and values of the parameters are not very critical; the most essential factor is the nonmonotonic property of $f(u)$.

4.1. Numerical Experiments

Figure 9 shows the temporal changes in the overlap p between X and S^1 for conventional dynamics (a) and for nonmonotone dynamics (b), where the units of the abscissa are the time constant τ. The experiment was carried out with $n = 1000$ and $m = 200$ ($r = 0.2$).

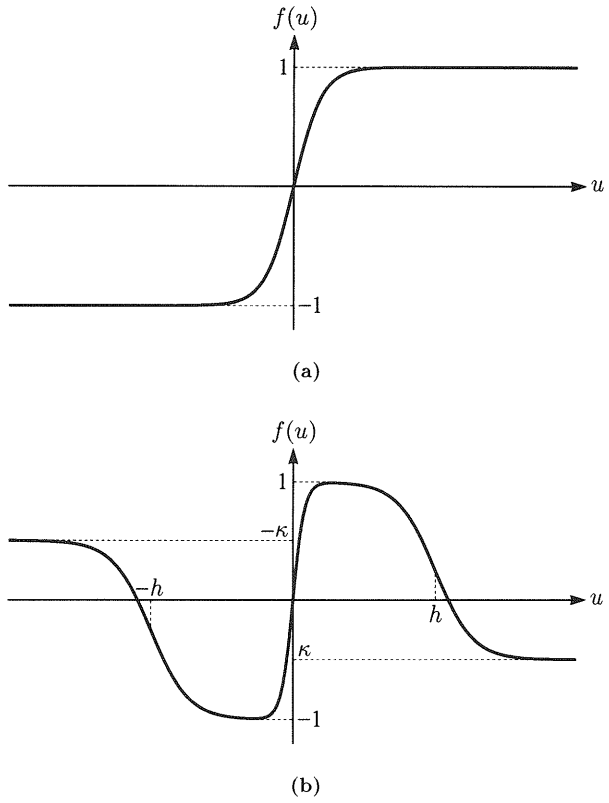


FIGURE 8. Output functions of the analog neuron model: (a) conventional sigmoid function; (b) nonmonotone function.

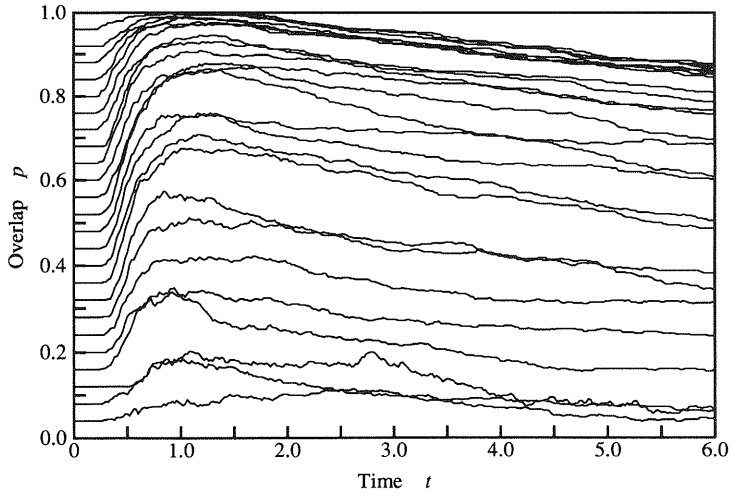
As we see from this figure, the nonmonotone model has marked merits:

1. The memory capacity is still larger than that for the partial reverse method; the upper limit of r is about 0.32[†] for the above parameter values.
2. Even if $p_0 < p_c$ and recollection is unsuccessful, spurious memories are not recalled; instead, the network state continues to wander—the variation in p appears chaotic—without reaching any equilibrium[‡]. This does not necessarily mean that there are no spurious memories, but implies that they are extremely rare.
3. The overlap p becomes exactly 1 when $p_0 > p_c$, i.e., the correct pattern is recalled without error. This indicates that the absolute memory capacity (the capacity where no recollection error is permitted) is proportional to n although it was on the order $n/\log n$ for the conventional autocorrelation model.

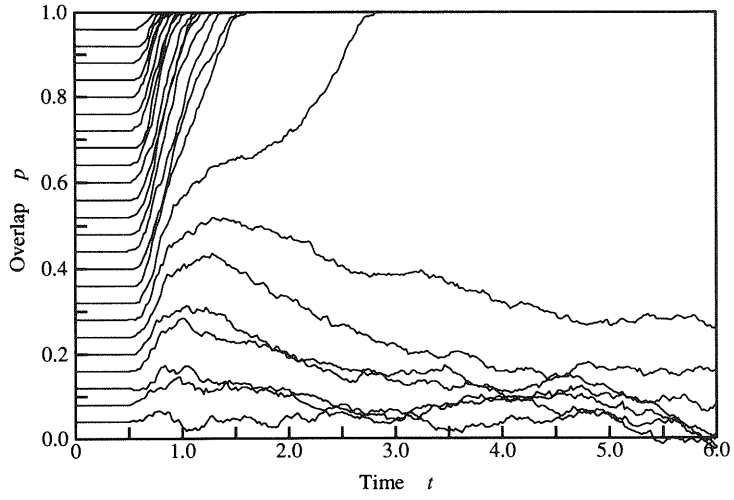
Furthermore, this model has another important property, which is dealt with in the next section.

[†] It can be as high as about 0.4 if the output function is suitably chosen (Yoshizawa, Morita, & Amari, in press). Also, the capacity increases when we set w_{ii} to some positive value.

[‡] If n is not large enough, one of the stored patterns or its reverse pattern can be recalled but the probability of this decreases exponentially as n increases.



(a)



(b)

FIGURE 9. Time courses of the overlap p for (a) conventional and (b) nonmonotone dynamics ($r = 0.2$). Time (the abscissa) is scaled by the time constant τ.

5. MEMORY FOR CORRELATED PATTERNS

In the preceding discussions, we assumed that all the stored patterns are selected independently at random so that they are not correlated with each other. This is because as the correlation between the patterns become large, the probability that

$$\mathrm{sgn}(WS^1) = S^1 \tag{23}$$

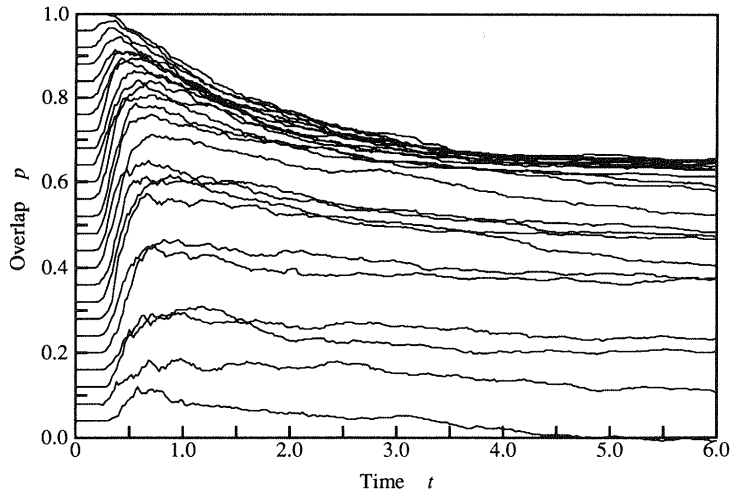
decreases greatly, and thus the network does not have an equilibrium in the neighborhood of S^1 as long as conventional dynamics is used. But this assumption is very unsatisfactory since it strongly restricts the representation of memory: No matter how closely related two things are, they must be represented by uncorrelated patterns.

In the preceding experiment, however, S^1 was an equilibrium state although eqn (23) did not hold exactly. This suggests that the above condition is not necessarily essential to the network with nonmonotone dynamics.

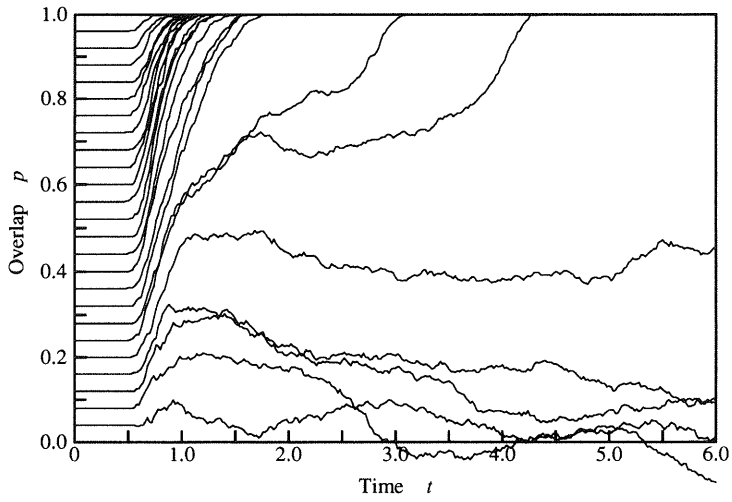
To examine this, we will deal with a special case where stored patterns are distributed in clusters and have a high correlation (Morita, Yoshizawa, & Nakano, 1990b). First we choose k patterns $C^1, C^2, \ldots, C^k$ at random, and generate $k \times l$ patterns $S^{11}, S^{12}, \ldots, S^{kl}$ independently so that

$$\frac{1}{n}\sum_{i=1}^{n} s_i^{\mu\nu} c_i^{\mu} = a \tag{24}$$

is satisfied for every $\mu = 1, \ldots, k$ and $\nu = 1, \ldots, l$, where a is a non-negative parameter representing the correlation between C^{μ} and $S^{\mu\nu}$. We can obtain $S^{\mu\nu}$ by selecting $n(1-a)/2$ components of C^{μ} randomly and reversing their sign.



(a)



(b)

FIGURE 10. Time courses of the overlap p when clustered patterns ($a = 0.6$) are stored: (a) for conventional dynamics; (b) for nonomonotone dynamics.

Then we store these patterns $S^{\mu\nu}$ using the autocorrelation matrix given by eqn (2), where we put $S^1 = S^{11}$, $S^2 = S^{12}$, ..., $S^m = S^{kl}$. If a is small, $S^{\mu\nu}$ are uncorrelated and distributed uniformly in state space; otherwise, they are distributed in k clusters with C^μ at the center of each cluster, and any pair of patterns in the cluster has an overlap of about a^2.

5.1. Numerical Experiments

Numerical experiments were carried out using $n = 1000$, $k = 50$ and $l = 4$ ($r = 0.2$).

Figure 10 shows the temporal changes in the overlap p when $a = 0.6$. This experiment was done in the same manner as for uncorrelated patterns except that the initial state X_0 necessarily had overlaps of about $a^2 p_0$ with $S^{1\nu}$ ($\nu \neq 1$).

As expected, the network with conventional dynamics cannot recall the correct pattern at all (Figure 10a); additionally, p decreases more rapidly than in Figure 9a when p_0 is large.

On the other hand, the model with nonmonotone dynamics works well (Figure 10b). Moreover, strangely enough, the critical overlap is smaller than that for uncorrelated patterns. To be more precise, the critical overlap p_c of this model changes with the correlation a as shown in Figure 11 (solid line): p_c decreases with an increase in a, provided that a is less than about 0.65; if a increases further, then p_c grows very sharply and finally exceeds 1, or the state S^1 becomes unstable.

For comparative purposes, an experiment was carried out on a pseudoinverse type of model (Kohonen, 1988), whose weight matrix is given by

$$W = \Sigma(\Sigma^T\Sigma)^{-1}\Sigma^T, \tag{25}$$

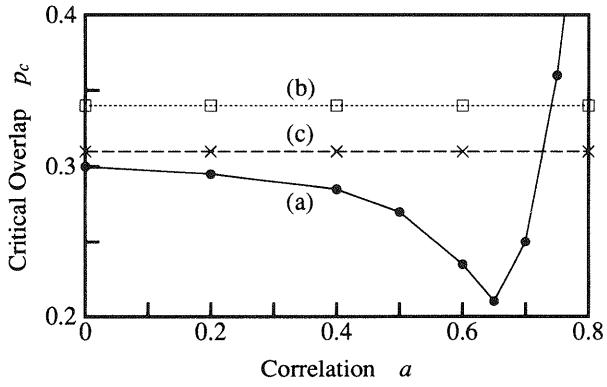


FIGURE 11. Changes in the critical overlaps of (a) the autocorrelation model with nonmonotone dynamics and of the pseudoinverse models with (b) conventional and (c) nonmonotone dynamics, as a function of the correlation a.

where $\Sigma = (S^1, \ldots, S^m)$. The critical overlaps for conventional (dotted line) and nonmonotone (broken line) dynamics are plotted in Figure 11.

We see that p_c does not depend on a in either case. Furthermore, it is larger than that above when $a < 0.7$. This means that although the pseudoinverse model can memorize strongly correlated patterns, it cannot make good use of the correlation, or, in a sense, the structure of memory (note that the total amount of information on the stored patterns decreases as a increases).

6. DISCUSSION

Then why does the autocorrelation model with nonmonotone dynamics show good performance for correlated patterns? A precise answer to this question cannot be given because mathematical analysis is quite difficult; I will give a rather intuitive explanation instead.

Figure 12 shows a process of recollection when $p_0 \simeq p_c$ and thus the network takes a long time to finish recalling successfully. In this figure, p' (dotted line) is the overlap between the network state X and the center C^1 of the cluster:

$$p' = \frac{1}{n} \sum_{i=1}^{n} x_i c_i^1 . \tag{26}$$

We see that p' increases more rapidly than p at first (note that $p'_0 \simeq ap_0$), and then decreases to 0.6 ($=a$). Intuitively, this indicates the following. When X is far enough from all of the stored patterns, each cluster behaves like an attractor of a dynamical system; when X approaches a cluster, it is attracted to just one pattern in the cluster (Figure 13).

The point is how a single pattern is separated from the others in the same cluster. To examine this, let us introduce a quantity $d_{\mu\nu}$ defined by

$$d_{\mu\nu} = \frac{\sum_i s_i^{\mu\nu} y_i}{\sum_i x_i y_i} , \tag{27}$$

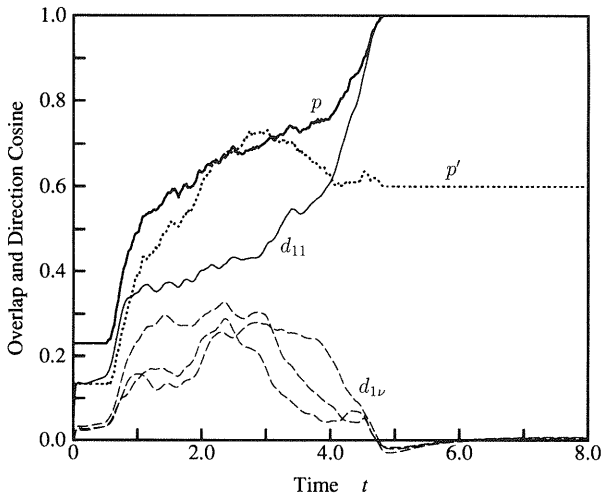


FIGURE 12. A process of recollection of the nonmonotone dynamics model ($a = 0.6$). Time courses in p (thick line), p' (dotted line), d_{11} (solid line), d_{12}, d_{13}, and d_{14} (broken lines) are plotted. The initial overlap $p_0 = 0.23$.

which is similar to the direction cosine between $S^{\mu\nu}$ and the output vector $Y = (y_1, \ldots, y_n)^T$. Although d_{11} is similar to p ($d_{11} = 1$ when $p = 1$), they are not equivalent because $y_i = f(u_i)$ can be very different from $x_i = \mathrm{sgn}(u_i)$ if $f(u)$ is nonmonotonic.

Using these quantities, we can write the input vector $V = WY$ in the form

$$V = \frac{\alpha}{n} (d_{11} S^{11} + d_{12} S^{12} + \ldots + d_{kl} S^{kl}) - rY, \tag{28}$$

where $\alpha = \sum_i x_i y_i$. The terms $d_{12}S^{12}, \ldots, d_{kl}S^{kl}$ on the right side of this equation correspond to noise. If $f(u)$ is the conventional sigmoid function, the noise terms $d_{1\nu}S^{1\nu}$ ($\nu = 2, \ldots, l$) necessarily grow when X approaches S^{11}, since $d_{1\nu}$ is nearly equal to the overlap between X and $S^{1\nu}$.

However, when the nonmonotone output function is used, they decrease to almost zero, as shown in Figure

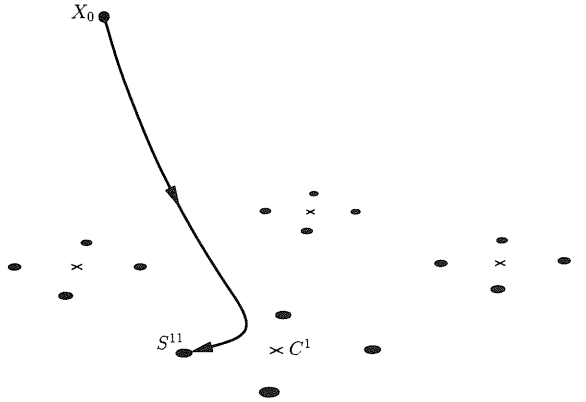


FIGURE 13. Schematic recalling process for the clustered patterns. At first, X approaches the center of the cluster, and then goes to the individual pattern.