112

Associative Memory and Its Statistical Neurodynamical Analysis

Shun-ichi Amari

University of Tokyo, Faculty of Engineering,
Bunkyo-ku, Tokyo 113, Japan

1. Introduction

A neural network is a complex system consisting of a large number of mutually connected elements having a simple input-output relation. Its behavior is highly non-linear, so that it is in general difficult to analyze its dynamical behavior of information processing. In order to elucidate a typical behavior, we study a network whose connection weights or synaptic efficacies of connections are randomly generated and then fixed. Given a probability law of connection weights, we have an ensemble of randomly generated networks. Statistical neurodynamics provides a theoretical method to search for macroscopic behaviors which are shared by all typical random networks in the ensemble, i.e. those networks generated by the same probability law.

Amari [1], [2] studied the macroscopic dynamical behaviors of randomly generated neural networks, and then proposed a mathematical method of statistical neurodynamics [3]. Amari et al. [4] proved some fundamental lemmas of statistical neurodynamics. The statistical neurodynamical method is shown to be useful for analyzing behaviors of associative memory models (Amari and Maginu [5]). The present paper proposes various versions of associative memory models, and gives their dynamical behaviors and capacities by the statistical neurodynamical method.

Correlation type associative memory models were proposed by Nakano [6], Kohonen [7], and Anderson [8], independently. A mathematical analysis of their dynamical behaviors was given by Amari [9] (see also Amari [10]). Since Hopfield [11] pointed out the spin-glass analogy, there have appeared a number of theoretical works (e.g., Amit et al. [12], Meir and Domany [13], McEliece et al. [14], Amari and Maginu [5]), which used the statistical neurodynamical method in some sense.

There are a number of versions of the associative memory model. They are, for example, a cross-correlation associative memory, an autocorrelation associative memory, a sequence recalling associative memory model [9], a bilateral associative memory (BAM) [15], etc. We analyze the dynamical behaviors and the memory capacities of these models by the statistical neurodynamical method. Their behaviors are different. For example, the memory capacity of an autocorrelation sequence generator is about twice that of an autocorrelation associative memory. We show where this difference comes from.

2. Various types of associative memory models

A neural element, which we treat here, has the following simple input-output relation : It receives n input signals $x_1, \cdots x_n$ and emits one output y, where

$$y = sgn\ (\sum w_i x_i - h)\ .$$

Here, w_i are called the connection weights or synaptic efficacies, sgn is the signature function taking a value $+1$ or -1 depending on the sign of its operand. Signals x_i and y also take on values $+1$ or -1. In the present paper, w_i 's are assumed to be randomly generated in some manner, and we put $h= 0$ for the sake of simplicity. We show various types of associative memory models in the following.

1) *Cross-correlation associative memory.*

Let us consider a network consisting of n neurons. It receives a vector input signal $\mathbf{x}=(x_1, x_2, \cdots, x_n)$ and emits a vector output signal $\mathbf{y}=(y_1, y_2, \cdots, y_n)$, where y_i is the output of the i-th neuron. Let w_{ij} be the synaptic efficacy of the j-th component of a signal $\mathbf{x}$ entering into the i-th neuron. See Fig. 1. Then, the input-output behavior of the network is written, in the component form, as

$$y_i=sgn\ (\Sigma\ w_{ij}\, x_j)\ . \tag{1}$$

We may write (1) as

$$\mathbf{y}=T\mathbf{x}=sgn\ (W\mathbf{x})\ , \tag{2}$$

where T is the non-linear operator determined by the synaptic efficacy matrix $W=(w_{ij})$.

When m pairs of vectors $(\mathbf{s}^1, \mathbf{q}^1),\cdots, (\mathbf{s}^m, \mathbf{q}^m)$ are given, a cross-associative memory is required to emit $\mathbf{q}^\mu$ when $\mathbf{s}^\mu$ is input ($\mu=1, 2, \cdots, m$). In other words, the input-output relation of the net is required to satisfy $T\mathbf{s}^\mu=\mathbf{q}^\mu$ for all $\mu=1,\cdots, m$. This implies that the $\mathbf{q}^\mu$ is recalled from $\mathbf{s}^\mu$. Moreover, it is expected to have a noise reduction property such that, when input $\mathbf{x}$ is a noisy version of $\mathbf{s}^\mu$, the output $T\mathbf{x}$ is equal to $\mathbf{q}^\mu$ or much closer to $\mathbf{q}^\mu$. We use the inner product

$$a = \frac{1}{n}\, \mathbf{s}^\mu \cdot \mathbf{x}$$

to evaluate the similarity of $\mathbf{x}$ to $\mathbf{s}^\mu$. In the correlation associative memory, synaptic efficacy matrix W is put equal to

$$w_{ij} = \frac{1}{n} \sum_{\mu=1}^{m} q_i^\mu s_j^\mu \qquad , \tag{3}$$

depending on the pairs (s^μ, q^μ) to be memorized. In the vector-matrix notion, this is written as

$$W=\frac{1}{n}\sum_{\mu=1}^{m} q^\mu (s^\mu)' \quad , \tag{4}$$

where q^μ and s^μ are assumed to be column vectors, and (s^μ)' denotes the transposition of s^μ.

We further assume that the components of the vectors q^μ and s^μ are randomly and independently determined such that they are $+1$ and -1 with probability 0.5 each. Then W is a randomly generated matrix. We search for the macroscopic property which holds for almost all W determined in this way, as the number n of the neurons becomes infinitely large.

2) *Cascade associative memory.*

Let us consider a cascaded series of cross-correlation associative memory models, $N_1, N_2, \cdots N_l, \cdots$ (Fig. 2).

The cascaded associative memory was studied by Meir and Domany [13]. Let N_l be a network which receives input x^l and emits output x^{l+1},

$$x^{l+1}=T_l x^l , \tag{5}$$

where T_l is the non-linear transformation. It is written as

$$T_l x = sgn(W_l x), \tag{6}$$

where W_l is the synaptic efficacy matrix of N_l. A cascade associative system is a concatenation of networks N_1, N_2, $\cdots$, N_k, $\cdots$, such that an output x^l of the l-th network N_l becomes an input of the $(l+1)$st network N_{l+1}.

Let $S1=\{ s^1{}_1, \cdots, s^1{}_{k+1}, \cdots\}$, $S2=\{ s^2{}_1, \cdots, s^2{}_{k+1}, \cdots\}$, $Sm=\{ s^m{}_1, \cdots, s^m{}_{k+1}, \cdots\}$ be m sequences of vectors. When $s^\mu_{l+1} = T_l s^\mu_l$ holds for $\mu=1,\cdots, m$, the system recalls sequence S^μ by emitting $\{ s^\mu_{l+1}\}$ from N_l, when s^μ_1 is input to the first net N_1. When the input x_1 to N_1 is close to s^μ_1, it is expected that $x_{l+1}=T_l x_l$ becomes closer and closer to s^μ_{l+1}. This is the noise reduction property, which we study in terms of the similarity or direction cosine $a^\mu_l = (1 / n)\, s^\mu_l \cdot x_l$.

We assume that the matrix W_l is determined by

$$W_l = \frac{1}{n} \sum_{\mu=1}^{m} s^\mu_{l+1} (s^\mu_l)' , \tag{7}$$

where each component of s^μ_l is determined randomly and independently as before. Without loss of generality, we assume that x_1 is close to $s^1{}_1$, and search for the dynamical relation of $a_l = (1 / n)\, s^1{}_l \cdot x_l$, $l= 1, 2, \cdots$.

3) *Cyclic associative memory and BAM.*

A cyclic associative memory is obtained by adding a feedback connection from the output of a cascaded associative memory network to its input (Fig. 3). A cyclic associative memory is calld a k-AM, when it consists of k networks connected in the form of a ring. When a signal is transformed through component networks N_l sequentially one at a time, we have the following equation $\mathbf{x}^{(t)}{}_{l+1} = T_l \mathbf{x}^{(t)}{}_l$, where t is the number of times the circuit is circled and l is calculated with modulo k such that $\mathbf{x}^{(t)}{}_{k+1}$ is put equal to $\mathbf{x}^{(t+1)}{}_1$.

Let $S^\mu = \{\mathbf{s}^\mu{}_1, \cdots, \mathbf{s}^\mu{}_k\}$, $\mu = 1, \cdots, m$, be m sequences of vectors of length k. It is expected that $\mathbf{s}^\mu{}_{l+1} = T_l \mathbf{s}^\mu{}_l$ holds for all $\mu = 1, \cdots, m$, $l = 1, \cdots, k$ $(k + 1 = 1)$, where the connection matrix W_l of N_l is given by

$$W_l = \frac{1}{n} \sum_{\mu=1}^{m} \mathbf{s}^\mu_{l+1} (\mathbf{s}^\mu_l)' \ .$$

It is also expected that a k-AM has a good noise reduction property that, given an input $\mathbf{x}_1$, the signal $\mathbf{x}_l$ converges to $\mathbf{s}^1{}_l$ provided $\mathbf{x}_1$ is close to $\mathbf{s}^1{}_1$, by circulating the network.

When $k = 2$, we have a 2-AM which is called a BAM (bilateral associative memory) (Kosko [15] , Okajima et al. [16]). As we see in the following, the behavior of a BAM is slightly different from other k-AM ($k \geqq 3$).

4) *Autoassociative memory.*

Let us consider a network in which the output is fed back to its input (Fig. 4). This network can be regarded as a 1-AM. When its connections matrix is W, its dynamical behavior is given by

$$\mathbf{x}^{t+1} = T\mathbf{x}^t = sgn(W\mathbf{x}^t) \ ,$$

where $\mathbf{x}^t$ is the output (the state vector) of the network at time t.

Given m patterns $\mathbf{s}^1, \cdots \mathbf{s}^m$, we have an autoassociative memory, when we put

$$W = \frac{1}{n} \sum_{\mu=1}^{m} \mathbf{s}^\mu (\mathbf{s}^\mu)' \ .$$

It is expected that all $\mathbf{s}^\mu$ are equilibria of the dynamics, satisfying $\mathbf{s}^\mu = T\mathbf{s}^\mu$ for all μ. The dynamics of this model is interesting. Its equilibrium behavior was analyzed by Amit et al. [12], and its dynamical behavior was analyzed by Amari and Maginu [5].

5) *Associative sequence generator.*

Let us consider a sequence $S = \{\mathbf{s}^1, \mathbf{s}^2, \cdots, \mathbf{s}^m\}$ of patterns. Let us also consider a network with recurrent connections (1-AM) whose synaptic efficacy matrix is given by

$$W = \frac{1}{n} \sum_{l=1}^{m} \mathbf{s}^{l+1} (\mathbf{s}^l)' \quad ,$$

where $\mathbf{s}^{m+1} = \mathbf{s}^1$. It is expected that the network recalls the sequence S, $T\mathbf{s}^l = \mathbf{s}^{l+1}$. Moreover, given an initial state $\mathbf{x}^1$ which is within a neighborhood of some $\mathbf{s}^\mu$, the network recalls the sequence more and more precisely through dynamical state transitions. This is the noise reduction property. Such a model was proposed by Amari [9] and he gave a mathematical analysis of its stability, see also Cottrell [17].

Let S_μ be m sequences of length k_μ, $\mu = 1, 2, \cdots m$,

$$S_\mu = \{\mathbf{s}_\mu^1, \mathbf{s}_\mu^2, \cdots, \mathbf{s}_\mu^{k_\mu}\}.$$

When W is put equal to

$$W = \frac{1}{n} \sum_{\mu=1}^{m} \sum_{l=1}^{k_\mu} \mathbf{s}_\mu^{l+1} (\mathbf{s}_\mu^l)' \quad ,$$

given any $\mathbf{s}^l_\mu$ or $\mathbf{x}$ in its neighborhood as the initial state, the model is expected to recall the remaining sequence by the dynamics,

$$\mathbf{s}_\mu^{l+1} = T\, \mathbf{s}_\mu^l \quad .$$

We assume here that $k_\mu \geqq 3$.

3. Noise reduction property of cross associative memory

We study in the beginning the noise reduction property of the simplest cross-associative memory. Without loss of generality, we study the property of recalling $\mathbf{q}^1$ from a noisy version of $\mathbf{s}^1$. Given an input $\mathbf{x}$ whose similarity to $\mathbf{s}^1$ is measured by the direction cosine to $\mathbf{s}^1$, $a = a_1(\mathbf{x}) = (1/n)\, \mathbf{s}^1 \cdot \mathbf{x}$, we search for the direction cosine a' or the similarity of the output $\mathbf{y} = T\mathbf{x}$ to $\mathbf{q}^1$, $a' = (1/n)\, \mathbf{q}^1 \cdot \mathbf{y}$. To this end, we calculate the i-th component of the output $\mathbf{y}$ as follows.

$$y_i = sgn\left(\sum w_{ij} x_j\right) = q_i^1\, sgn\left(a + q_i^1 N_i\right) \tag{8}$$

where

$$N_i = \frac{1}{n} \sum_{\mu=2}^{m} \sum_{j=1}^{n} q_i^\mu s_j^\mu x_j \quad . \tag{9}$$

The term N_i represents the interference or crosstalk from the superposed pattern pairs in W other than ($\mathbf{s}^1, \mathbf{q}^1$). When $N_i = 0$ and $a > 0$, we have exact recalling $\mathbf{y} = \mathbf{q}^1$.

How large is the noise term N_i ? Since the vectors ($\mathbf{s}^\mu$, $\mathbf{q}^\mu$) are generated randomly, we can evaluate it by using probability theory. The term N_i is a sum of n(m - 1) random variables $q^\mu_i s^\mu_j x_j$ devided by n, so that the central limit theorem guarantees that N_i is normally distributed with mean 0 and variance $\sigma^2 = (m$

$-1)/n \doteq r$, where $r = m/n$ is the ratio of the number of the memorized pattern pairs to the number of neurons.

The probability that the output y_i is erroneous is given by

$$P = Prob\{a + q_i^1 N_i < 0\} = Prob\{N_i < -a\} = \Phi(\frac{a}{\sigma}) \quad , \tag{10}$$

where

$$\Phi(u) = \int_u^\infty \frac{1}{\sqrt{2\pi}} exp\{-\frac{t^2}{2}\} dt . \tag{11}$$

The direction cosine a' of the output is given by $a' = (1/n) \Sigma y_i q^1_i$. Since the probabolity of $y_i = q^1_i$ is $1 - \Phi$ and the probability of y_i not eaqual to q^1_i is Φ, by the law of large numbers, a' converge to $a' = 1 - 2\Phi(a/\sigma) = F(a/\sigma)$, where

$$F(u) = \int_{-u}^{u} \frac{1}{\sqrt{2\pi}} exp\{-\frac{t^2}{2}\} dt \quad . \tag{12}$$

This gives the answer to our simple problem.

Theorem 1. The noise reduction property of a cross associative memory is given by

$$a' = F(\frac{a}{\sqrt{r}}) \quad . \tag{13}$$

This is the well-known result (see, e.g., Amari [10], Kinzel [18]). The problem is whether this result can be generalized to be applicable to the dynamics of autoassociative memory or many other versions of correlation type associative memory models.

4. Dynamics of cascade associative memory, k-AM, and sequence generator

We now search for the noise reduction property of a cascade associative memory. Since this is a concatenation of cross-associative memory models, one may think that the overall noise reduction property is given also by concatenation,

$$a_{l+1} = F(a_l/\sqrt{r}) \ , \ l = 1, 2, \cdots$$

where

$$a_l = (1/n) \mathbf{s}_l^1 \cdot \mathbf{x}^l$$

is the similarity of the output of the l-th layer to its expected output $\mathbf{s}^1{}_l$. However, this is not the case. As is seen from (7), the connection matrices W_{l-1} and W_l of two successive layers are not stochastically independent. This is because they depend partly on the same random vectors s^{μ}_{l}. Therefore, the probability distribution of the crosstalk term $N_{l,\,i}$ of the l-th layer is not so simple as before.

The i-th component of x_l is written as

$$x_{l,i} = sgn(\frac{1}{n} s^1_{l,i}\, \mathbf{s}^1_{l-1} \cdot \mathbf{x}_{l-1} + N_{l,i}) = s^1_{l,i}\, sgn(a_{l-1} + s^1_{l,i}\, N_{l,i}) \quad ,$$

where

$$N_{l,i} = \frac{1}{n} \sum_{\mu=2}^{m} s^{\mu}_{l,i}\, \mathbf{s}^{\mu}_{l-1} \cdot \mathbf{x}_{l-1} \tag{14}$$

$$= \frac{1}{n} \sum_{\mu} \sum_{j} Z^{\mu}_{lij} \quad ,$$

with

$$Z^{\mu}_{lij} = s^{\mu}_{l,i}\, s^{\mu}_{l-1,j}\, x^{l-1}_{j} \quad .$$

We assume that $N_{l,\,i}$ is normally distributed with mean μ_l and variances σ^2_l. Although $N_{l,\,i}$ is a sum of Z^{μ}_{lij}, its variance is not simply equal to r. This is because

$$x^{l-1}_{j} = s^1_{l-1,j}\, sgn(a_{l-2} + s^1_{l-1,j}\, N_{l-1,j})$$

depends on s^{μ}_{l-1}, so that Z^{μ}_{lij} are not independent.

Since x^{l-1} does not depend on s^{μ}_{l}, we have

$$E[\, Z^{\mu}_{lij}\,] = 0 \quad .$$

This shows that

$$\mu_l = 0 \quad ,$$

so that $N_{l,\,i}$ is distributed with mean 0.

In order to calculate the variance

$$\sigma^2_l = E[(\sum Z^{\mu}_{lij})^2] \quad ,$$

we note that

1) $$E[(Z^{\mu}_{lij})^2] = 1 \quad ,$$

2) $$E[Z^{\mu}_{lij}\; Z^{\mu'}_{lik}] = E[(s^{\mu}_{l+1,i}\, s^{\mu'}_{l+1,i})\, s^{\mu}_{l,j}\, s^{\mu'}_{l,k}\, x^l_j x^l_k] = 0 \quad ,$$

provided $\mu \neq \mu'$. The correlational terms are, therefore, given by

$$A = E[Z^{\mu}_{lij} Z^{\mu}_{lij'}\,],\;\; j \neq j' \quad .$$

This term can be rewritten as

$$A = E[s s' sgn(a + sM + N)\, sgn(a + s'M + N')]$$

where

$$s = s^{\mu}_{l,j},\;\; s' = s^{\mu}_{l,j'},\;\; a = a_{l-1} \quad ,$$

and

$$M = \frac{1}{n} \mathbf{s}^{\mu}_{l-1} \cdot \mathbf{x}^{l-1}$$

represents the common factor in $N_{l-1,j}$ and $N_{l-1,j'}$. This is a small random variable subject to $N(0, \sigma^2_{l-1} / m)$, which are independent of the random variables N, N' subject to $N(0, \sigma^2_{l-1})$, which are the remaining terms in $N_{l-1,j}$ and $N_{l-1,j'}$, respectively. Since M, N and N' do not depend on s and s', we have

$$A = \frac{1}{2} E[sgn(a+M+N)\, sgn(a+M+N')] - \frac{1}{2} E[sgn(a+M+N)\, sgn(a-M+N')]$$

because Prob$\{ s s' = 1 \}$ = Prob$\{ s s' = -1 \} = \frac{1}{2}$. By taking the expectation with respect to N and N', where M is fixed, i.e., by taking the conditional expection, we have

$$A = \frac{1}{2} E[F(\frac{a+M}{\sigma}) F(\frac{a+M}{\sigma}) - F(\frac{a+M}{\sigma}) F(\frac{a-M}{\sigma})],$$

where $\sigma = \sigma_{l-1}$. Since M is a random variable of order $1/\sqrt{m}$, we expand the above A in the Taylor series of M and take the expectation with repeat to M. We then have

$$A = \frac{4}{m} \{p(\frac{a}{\sigma})\}^2 ,$$

where

$$p(v) = \frac{1}{\sqrt{2\pi}} exp\{-\frac{v^2}{2}\} .$$

This proves that

$$\sigma_l^2 = r + 4\{p(\overline{a}_l)\}^2 ,$$

where

$$\overline{a}_l = \frac{a_l}{\sigma_l} .$$

The noise reduction property of a cascade network is then given by the following theorem, which was first obtained by Meir and Domany [13] by a different method.

Theorem 2. The direction cosine changes as

$$a_{l+1} = F(a_l / \sigma_l) , \tag{15}$$

$$\sigma^{l+1} = r + 2\{p(a_l/\sigma_l)\}^2 \tag{16}$$

$$p(u) = \frac{1}{\sqrt{2\pi}} exp\{-\frac{u^2}{2}\}$$

in the process of recalling S^{μ} in a cascade associative memory system.

The system of equations (15), (16) demonstrate interesting properties : There is a threshold r_c such that, when the pattern ratio $r = m / n$ is larger than r_c, a_l never converges to 1 as l becomes large, even if a_1 is very large or even equal to 1. In this case, any sequence S^μ cannot be recalled. On the other hand, when r is smaller than r_c, a_l converges to 1 as l becomes large, provided a_1 is larger than a threshold a_c. Hence, we may call r_c the capacity of the system. When $r < r_c$, there exists $a_c(r)$ such that, when $a_1 < a_c$, a_l converges to 1, but it otherwise does not converge to 1. Therefore, a_c denotes the size of the basin of attractor from which a good recall is possible. The two characteristic quantities r_c and $a_c(r)$ can be calculated from Theorem 2. See Meir and Domany [13].

When the recalling process fails to recall S^μ, it sometimes occurs that a_l once increases and becomes larger than the threshold $a_c(r)$. However, it decreases again. This interesting phemomenon is found also in the dynamics of auto-correlation associative memory (Amari and Maginu [5]). All of these characteristics are common to the auto-correlation associative memory analyzed by Amari and Maginu [5], although the dynamical equations are different.

The computer simulation of the noise reduction behavior of a cascaded network is shown in Fig. 5a where $r = 0.08$ and in Fig. 5b where $r = 0.2$. The theory is in good agreement.

In the case of a cyclic associative memory (k-AM), the noise reduction property would be a little different. In the case of a cascaded memory, in

$$\mathbf{x}^{l+1} = sgn(W_l \, sgn(W_{l-1}\mathbf{x}_{l-1})),$$

W_l is independent of $\mathbf{x}_{l-1}$. Therefore, the direct correlation between W_l and $\mathbf{x}_l$ emerges through that of W_l and W_{l-1}. This is not true in the case of a k-AM, because $\mathbf{x}_{l-1}$ has some correlations with W_l. This correlation, however, is very diffuse, so that it might be neglected as n tends to infinity. Therefore, we use the following assumption, which was also used in Amari and Maginu [5].

Assumption. The direct correlations of W_l and $\mathbf{x}_{l-1}$ can be neglected in the calculation of correlations of W_l and $\mathbf{x}_l$.

If this assumption is true, it is easy to show that the noise reduction property of a k-AM ($k \geqq 3$), or of an associative sequence generator is the same as a cascade associative memory. However, a 2-AM (BAM) and an autoassociative memory have different noise reduction properties.

We show a computer simulation of the behavior of a sequence generator (Fig. 6), of a 3-AM (Fig. 7), and of a BAM (Fig. 8). We can see that the simulation is also in good agreement with the theory even in the case of a BAM. However, the behavior of an autoassociative memory is different (Fig. 9).

It should be emphasized that the memory capacity of a sequence generator is 0.27, which is twice as that of an autoassociative memory, although both models use networks of the same recurrent architecture.

Conclusions

We have shown theoretical derivations of dynamical behaviors of recalling processes of various types of associative memory models, by using the statistical neurodynamical method.

References

1. Amari, S. [1971] : "Characteristics of randomly connected threshold element networks and network systems", Proc. IEEE, 59, 35-47.
2. Amari, S. [1972a] : "Characteristics of random nets of analog neuron-like elements", IEEE Trans., SMC-2, 643-657.
3. Amari, S. [1974] : "A method of statistical neurodynamics", Kybernetik, 14, 201-215.
4. Amari, S. Yoshida, K., and Kanatani, K. [1977] : "A mathematical foundation for statistical neurodynamics", SIAM J. App. Math., 33, 95-126.
5. Amari, S. and Maginu, K. [1988] : "Statistical neurodynamics of associative memory", Neural Networks, 1, 63-73
6. Nakano, K. [1972] : "Association--- a model of associative memory", IEEE Trans., SMC-2, 381-388
7. Kohonen, T. [1972] : "Correlation matrix memories", IEEE Trans., C-21, 353-359.
8. Anderson, J. A. [1972] : "A simple neural network generating interactive memory", Math. Biosciences, 14, 197-220.
9. Amari, S. [1972] : "Learning patterns and pattern sequences by self-organizing nets of threshold elements", IEEE Trans., C-21, 1197-1206.
10. Amari, S. [1977] : "Neural theory of association and concept-formation", Biol. Cybernetics, 26, 175-185.
11. Hopfield, J. J.[1982] : "Neural networks and physical systems with emergent collective computational abilities", P. Nat. Acad. Sci. U.S.A., vol. 79, 2445-2458.
12. Amit, D. J., Gutfreund, H. and Sompolinsky, H. [1986] : "Spin-glass models of neural networks", Phys. Rev., A2, 1007-1018.
13. Meir, R. and Domany, E. [1987] : "Exact solution of a layered neural network memory", Phy. Rev. Lett., 59, 359-362.
14. McEliece, R. J. et. al. [1987] : "The Capacity of the Hopfield associative memory", IEEE Trans., Inf. Theory, IT-33, 461-482.

15. Kosko, B. [1987] : "Adaptive bilateral associative memory", Appl. Oct., 26, 4947-4960.

16. Okajima, K., Tanaka, S. and Fujiwara, S. [1977] : "A heteroassociative memory network wirh feedback connection", Proc. 1st Int. Conf. Neural Networks, II-711-II-718.

17. Cottrell, M. [1988] : "Stability and attractivity in associative memory networks", Biol. Cybern, 58, 129-139.

18. Kinzel, W. [1985] : "Learning and pattern recognition in spin glass models", Z. Physik, B60, 205-213.

Figure captions

Fig. 1. Cross-associative memory

Fig. 2. Cascade associative memory

Fig. 3. Cyclic associative memory

Fig. 4. Autoassociative memory

Fig. 5. Dynamical behavior of cascade networks

Fig. 6. Dynamical behavior of sequence generators

Fig. 7. Dynamical behavior of 3-AM

Fig. 8. Dynamical behavior of BAM

Fig. 9. Dynamical behavior of autoassociative memory

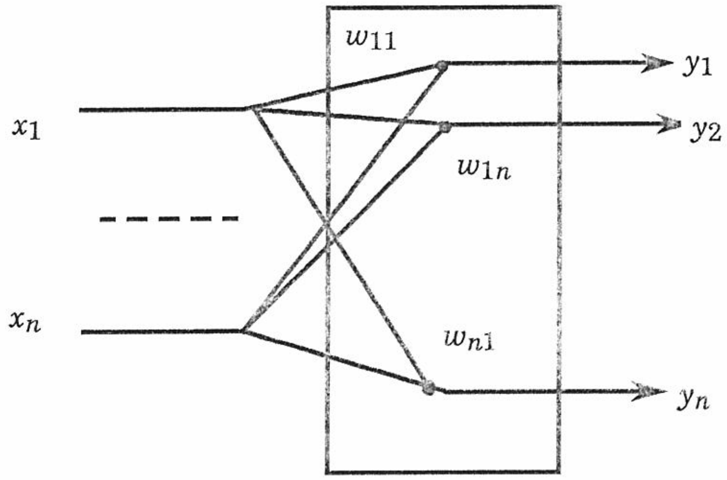

Fig. 1 Cross-associative memory

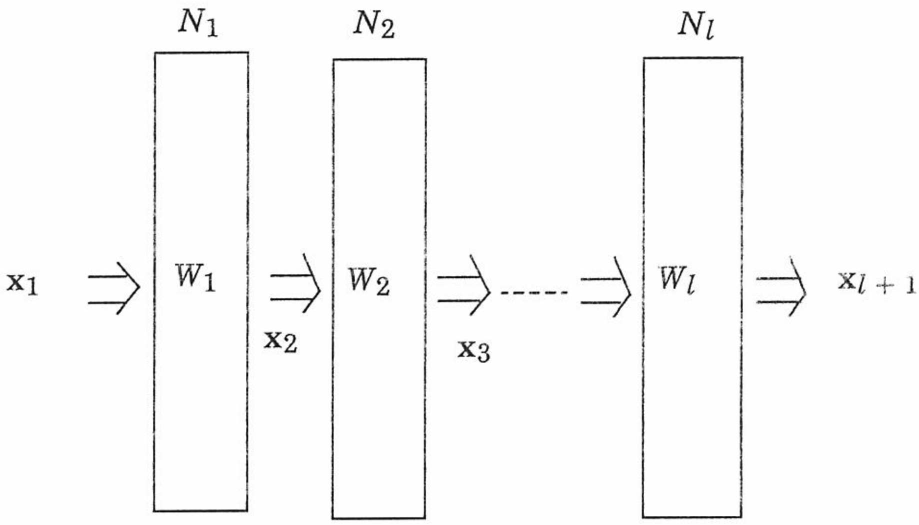

Fig. 2 Cascade associative memory

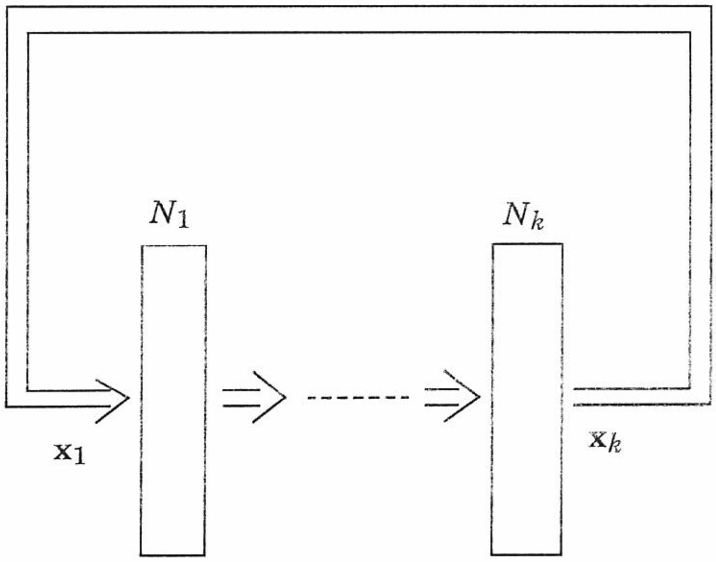

Fig. 3 Cyclic associative memory

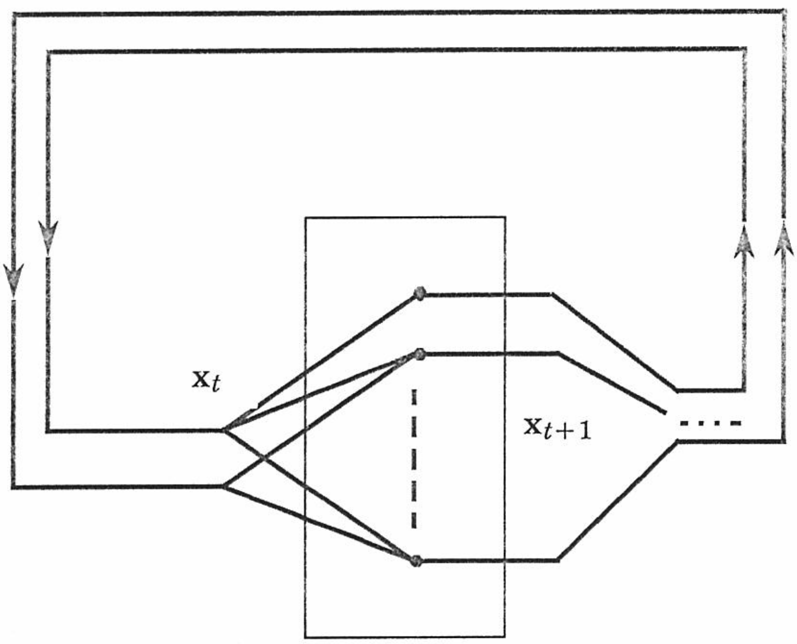

Fig. 4 Autoassociative memory

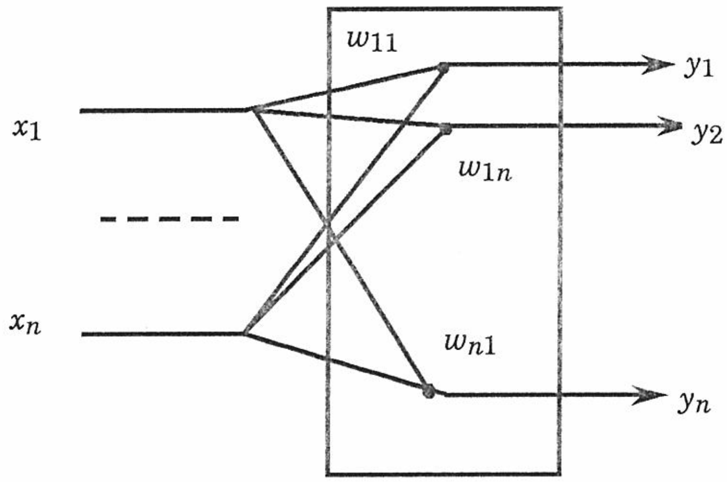

Fig. 1 Cross-associative memory

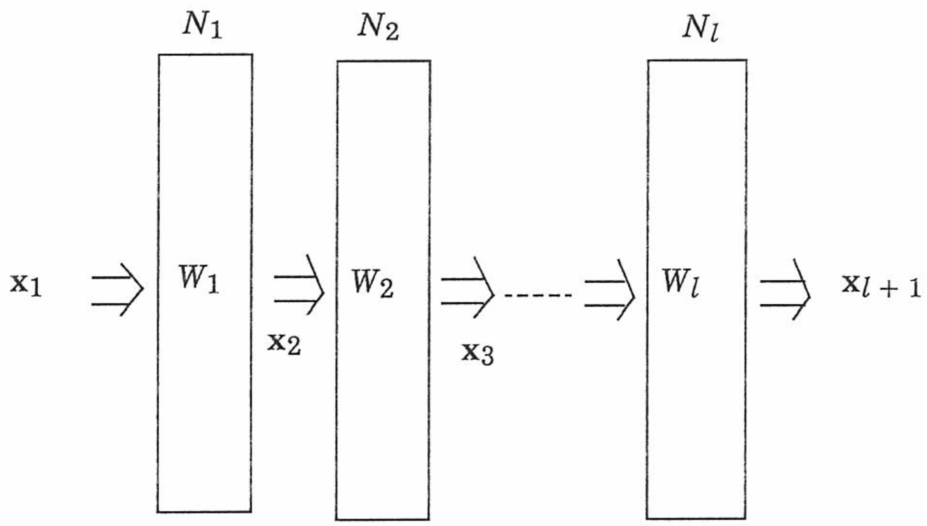

Fig. 2 Cascade associative memory

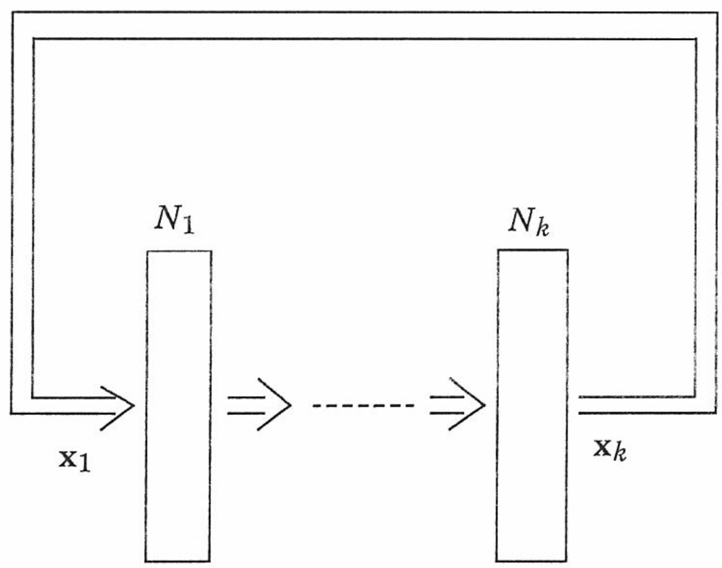

Fig. 3 Cyclic associative memory

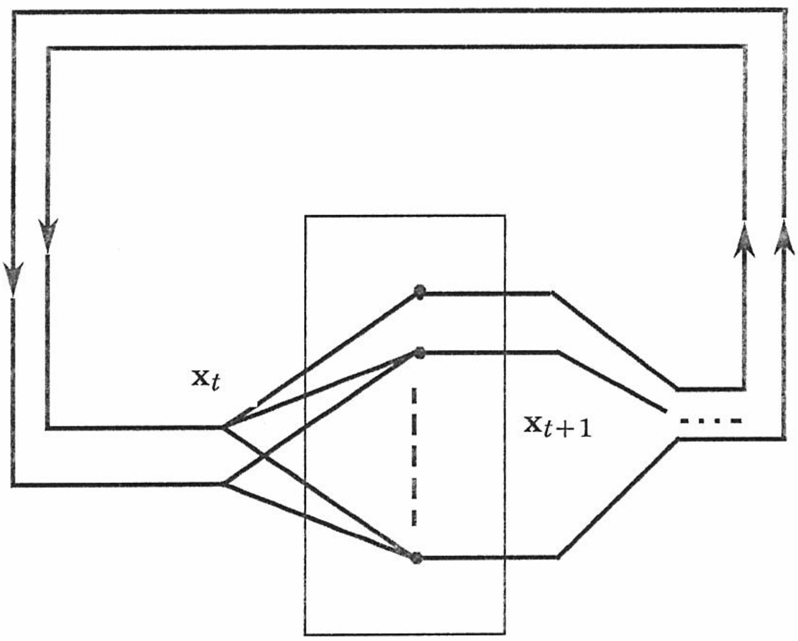

Fig. 4 Autoassociative memory